



UNIVERSIDAD DE OVIEDO

Centro Internacional de Postgrado

**PROGRAMA DE DOCTORADO: INGENIERÍA DE PRODUCCIÓN,
MINERO – AMBIENTAL Y DE PROYECTOS**

TESIS DOCTORAL

**INFLUENCIA Y CARACTERIZACIÓN
DE LOS MÉTODOS DE
CROWDFUNDING EN LA
REALIZACIÓN DE PROYECTOS**

Aladino Fernández Blanco

Abril, 2021



AUTORIZACIÓN PARA LA PRESENTACIÓN DE TESIS DOCTORAL

Año Académico:2020/2021

1.- Datos personales del autor de la Tesis	
Apellidos: FERNÁNDEZ BLANCO	Nombre: ALADINO
DNI/Pasaporte/NIE:	Teléfono:

2.- Datos académicos	
Programa de Doctorado cursado: Ingeniería de Producción, Minero-Ambiental y de Proyectos	
Órgano responsable: UNIVERSIDAD DE OVIEDO - CENTRO INTERNACIONAL DE POSTGRADO	
Departamento/Instituto en el que presenta la Tesis Doctoral: CENTRO INTERNACIONAL DE POSTGRADO	
Título definitivo de la Tesis	
Español/Otro Idioma: Influencia y caracterización de los métodos de Crowdfunding en la realización de proyectos.	Inglés: Influence and characterization of Crowdfunding methods in the performing projects
Rama de conocimiento: INGENIERÍA Y ARQUITECTURA	

3.- Autorización del Director/es y Tutor de la tesis	
D/D ^a : JOAQUÍN VILLANUEVA BALSERA	DNI/Pasaporte/NIE:
Departamento/Instituto: EXPLOTACION Y PROSPECCION DE MINAS	
D/D ^a :	DNI/Pasaporte/NIE:
Departamento/Instituto/Institución:	
Autorización del Tutor de la tesis	
D/D ^a : FRANCISCO ORTEGA FERNÁNDEZ	DNI/Pasaporte/NIE:
Departamento/Instituto/Institución: EXPLOTACION Y PROSPECCION DE MINAS	

Autoriza la presentación de la tesis doctoral en cumplimiento de lo establecido en el Art. 32 del Reglamento de los Estudios de Doctorado, aprobado por el Consejo de Gobierno, en su sesión del día 20 de julio de 2018 (BOPA del 9 de agosto de 2018)

Oviedo, 15 de abril de 2021

Director/es de la Tesis

Tutor de la Tesis

Fdo.: Joaquín Villanueva Balsera

Fdo.: Francisco Ortega Fernández

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN INGENIERÍA DE PRODUCCIÓN, MINERO-AMBIENTAL Y DE PROYECTOS.



RESOLUCIÓN DE PRESENTACIÓN DE TESIS DOCTORAL

Año Académico:2020 / _2021

1.- Datos personales del autor de la Tesis	
Apellidos: Fernández Blanco	Nombre: Aladino
	Teléfono:

2.- Datos académicos	
Programa de Doctorado cursado: Programa Oficial de Doctorado en Ingeniería de Producción, Minero-Ambiental y de Proyectos	
Órgano responsable: Universidad de Oviedo - Centro Internacional de Postgrado	
Departamento/Instituto en el que presenta la Tesis Doctoral: Explotación y Prospección de Minas	
Título definitivo de la Tesis	
Español/Otro Idioma: Influencia y caracterización de los métodos de Crowdfunding en la realización de proyectos.	Inglés: Influence and characterization of Crowdfunding methods in the performing projects
Rama de conocimiento: Ingeniería y Arquitectura	
Señale si procede: <input type="checkbox"/> Mención Internacional <input type="checkbox"/> Idioma de presentación de la Tesis distinto al español <input type="checkbox"/> Presentación como compendio de publicaciones	

3.- Autorización del Presidente de la Comisión Académica	
D/D ^a : Jose Carlos Rico Fernández	DNI/Pasaporte/NIE:
Departamento/Instituto: Construcción e Ingeniería de Fabricación	

Resolución: La Comisión Académica del Programa de Doctorado en **Ingeniería de Producción, Minero-Ambiental y de Proyectos** en su reunión de fecha **16-06-2021**, acordó la presentación de la tesis doctoral a la Comisión de Doctorado, previa comprobación de que la tesis presentada y la documentación que la acompaña cumplen con la normativa vigente, según lo establecido en el Art.32.8 del Reglamento de los Estudios de Doctorado, aprobado por el Consejo de Gobierno, en su sesión del día 20 de julio de 2018 (BOPA del 9 de agosto de 2018)

Además, informa:

	Favorable	Desfavorable
• Mención Internacional	<input type="checkbox"/>	<input type="checkbox"/>
• Idioma	<input type="checkbox"/>	<input type="checkbox"/>
• Presentación como compendio de publicaciones	<input type="checkbox"/>	<input type="checkbox"/>



Justificación

La tesis plantea un conjunto de modelos que permiten caracterizar los factores de éxito o fracaso de los proyectos financiados mediante crowdfunding, así como clasificar y predecir el éxito. Para tal objetivo se parte de un conjunto de datos procedente de proyectos reales y contrastado usado como referencia también en otras publicaciones científicas. La aplicación de estos modelos proporciona un paradigma que informará de los atributos sobre los que se puede actuar de modo que se pueda aumentar las posibilidades de éxito del proyecto financiado con crowdfunding. Se ha aplicado el método científico y una metodología contrastada para modelar un conjunto de datos mediante técnicas de *machine learning*. Uno de los aspectos más destacables es la posibilidad de que un creador de un proyecto sea capaz de influir sobre el entorno del proyecto para conseguir el éxito de financiación. El trabajo añade valor a la investigación en Dirección de Proyectos con el análisis realizado en este tipo de financiación. El trabajo realizado, así como los resultados cumplen con los criterios científicos de calidad, avalado por la publicación en revista indexada.

Oviedo, 16 de junio de 2021

Presidente de la Comisión Académica del Programa de Doctorado

Fdo.: Jose Carlos Rico Fernández

Contra la presente Resolución, podrá interponer recurso de alzada ante el Rectorado, en el plazo de un mes, a partir del día siguiente al de la presente notificación, de conformidad con el art. 122 de la Ley 39/2015, de 1 de octubre, de Procedimiento Administrativo Común de las Administraciones Públicas

SR. DIRECTOR DEL CENTRO INTERNACIONAL DE POSTGRADO



RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Influencia y caracterización de los métodos de Crowdfunding en la realización de proyectos	Inglés: Influence and characterization of Crowdfunding methods in the performing projects
2.- Autor	
Nombre: ALADINO FERNANDEZ BLANCO	DNI/Pasaporte/NIE:
Programa de Doctorado: Ingeniería de Producción, Minero-Ambiental y de Proyectos	
Órgano responsable: Universidad de Oviedo - Centro Internacional de Postgrado	

RESUMEN (en español)

El crowdfunding es un medio de financiación de proyectos altamente innovadores que nació para dar respuesta a la dificultad de financiar proyectos durante la última crisis económica mundial. El secreto de su éxito radica en dividir el riesgo en partes más pequeñas, de tal manera que una multitud potencialmente inversora pueda colaborar en el proyecto con una pequeña cuantía económica a la espera de una recompensa, producto del proyecto. El número de proyectos que consiguen financiación usando este medio ha ido aumentando hasta llegar a un punto donde se produce un retroceso. La motivación de este estudio radica en proporcionar una herramienta capaz de potenciar los proyectos hacia el éxito o al menos alejarlos del fracaso, para incrementar la tasa de éxito y aportar un alto grado de sostenibilidad a este tipo de financiación.

Se estudiarán los factores principales que intervienen en el crowdfunding, con el fin de satisfacer por un lado al creador, al ser capaz de llevar su proyecto al éxito y por otro lado complacer a los colaboradores del proyecto al materializarse su recompensa.

Para llevarlo a cabo se ha adquirido una base de datos de la plataforma "Kickstarter", la más conocida del sector. Tras su procesado, resulta un conjunto de datos muy representativo del periodo de estudio sobre el que se aplicarán técnicas de minería de datos.

El estudio se realiza en dos pasos. En el primero se caracteriza el éxito de los proyectos a través de la identificación de los factores principales que intervienen en el éxito, en segundo lugar, se realiza una clasificación y predicción de éxito en los proyectos, con el fin de identificar cuales son las características que tienen mayor capacidad para inducir al éxito un proyecto y así asistir al creador a dirigir sus esfuerzos sobre las más influyentes.

A través de técnicas de clusterización y asociación se identifican seis clústeres de proyectos, cada uno con características diferenciadoras del resto y que suponen una herramienta de gran ayuda para el creador por su capacidad para potenciar proyectos hacia el éxito. Estos seis clústeres caracterizan estados finales de proyectos, resultando tres estadios diferentes: dos de éxito, uno de advertencia y tres de fracaso. Así, un creador puede ubicar su proyecto en el sistema asociando las características de su proyecto al clúster que mejor se ajuste y posteriormente puede decidir si es necesario modificar las características de su proyecto en base a la adaptación hacia un clúster con mayor potencial de éxito.

Finalmente se han aplicado técnicas de clasificación y predicción del éxito de los proyectos en base a sus características más relevantes, demostrando tanto su compatibilidad con las técnicas anteriores como su carácter de respuesta ágil cuando se aplican de manera unilateral.



RESUMEN (en Inglés)

Crowdfunding is a method of financing highly innovative projects that was born in response to the difficulty of financing projects during the last global economic crisis. Away from banking institutions, the secret of its success consists in dividing the risk into smaller parts, so that a crowd of potential investors can collaborate in the project with a small amount of money while waiting for a reward, the product of the project.

The number of projects that obtain funding using this method has been increasing since its inception but has reached a point where there has been a decline and the ratio of successful to unsuccessful projects is not improving. The motivation of this study consists in providing a tool capable of promoting projects towards success or at least keeping them away from failure, in order to increase the success rate and provide a high degree of sustainability to this type of financing.

To achieve this, it will study the main factors involved in crowdfunding in order to satisfy the creator, by being able to succeed with his project, and on the other hand, to please the project's collaborators when their reward materialises.

To carry it out, it was acquired a database from the "Kickstarter" platform, the best known in the sector, providing a highly representative set of data for the study period. Its unstructured nature has led to the application of different techniques to clean and give structure to the data, finally obtaining a balanced working dataset that is sufficiently representative for the subsequent application of data mining techniques.

The study is carried out in two steps, the first one characterises the success of the projects through the identification of the main factors involved in success. The second, it performs a classification and prediction of success in the projects, in order to identify which are the characteristics that have greater capacity to induce the success of a project and thus assist the creator to direct his efforts on the most influential ones.

Through clustering and association techniques we identify six clusters of projects, each of which has characteristics that distinguish them perfectly and they are a great tool in order to help the creator due to their capacity to promote projects towards success. These six clusters describe the end states of projects, resulting in three different stages: two of success, one of warning and three of failure. So, a creator can place a project in the system by associating the characteristics of his project to the cluster that best fits and can subsequently decide whether it is necessary to modify the characteristics of his project based on the adaptation towards a cluster with greater potential for success.

Finally, it has been applied techniques for classifying and predicting the success of projects based on their most relevant characteristics, demonstrating both their compatibility with the previous techniques and their agile response character when applied unilaterally.

**SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO
EN INGENIERÍA DE PRODUCCIÓN, MINERO-AMBIENTAL Y DE PROYECTOS**

AGRADECIMIENTOS

Llegado el momento final, no me gustaría presentar este proyecto de tesis doctoral sin dejar de agradecer la ayuda y comprensión de todos los que han estado conmigo durante el transcurso de esta investigación. Gran parte del interés e ilusión que he puesto en el desarrollo de este trabajo les pertenece.

En primer lugar, me gustaría agradecer la ayuda a todos los miembros del Área de Proyectos de Ingeniería, allí he visto el buen hacer de un equipo, he encontrado motivación y apoyo constante, siempre abiertos y dispuestos a ayudarme.

Especialmente quiero agradecer a mi director Joaquín Villanueva por estar a mi lado desde el comienzo sin importar el día ni la hora y por su paciencia en la resolución de problemas que han surgido, también a Vicente Rodríguez por ayudarme en los momentos más duros, a Francisco Ortega por hacer que me apasione la investigación y por la mentoría necesaria que me ha permitido combinar estudios y trabajo.

A mi familia, por todo el apoyo y amor que he recibido, por entender los periodos de tiempo que no he podido estar con ellos, especialmente con mis padres y mi hermana, también agradecer a mi abuelo por empeñarse en sufragar con ilusión parte de los gastos de este trabajo y a mi suegro por apoyarme en todo momento.

Finalmente, a Henar, gracias por animarme, entenderme y estar a mi lado en todo momento y a Simón por comprender que a veces no he podido jugar con él. A vosotros os pertenece una parte muy grande de este trabajo.

Muchas gracias a todos.

RESUMEN

El crowdfunding es un medio de financiación de proyectos altamente innovadores que nació para dar respuesta a la dificultad de financiar proyectos durante la última crisis económica mundial. El secreto de su éxito radica en dividir el riesgo en partes más pequeñas, de tal manera que una multitud potencialmente inversora pueda colaborar en el proyecto con una pequeña cuantía económica a la espera de una recompensa, producto del proyecto.

El número de proyectos que consiguen financiación usando este medio ha ido aumentando hasta llegar a un punto donde se produce un retroceso. La motivación de este estudio radica en proporcionar una herramienta capaz de potenciar los proyectos hacia el éxito o al menos alejarlos del fracaso, para incrementar la tasa de éxito y aportar un alto grado de sostenibilidad a este tipo de financiación

Se estudiarán los factores principales que intervienen en el crowdfunding, con el fin de satisfacer por un lado al creador, al ser capaz de llevar su proyecto al éxito y por otro lado complacer a los colaboradores del proyecto al materializarse su recompensa.

Para llevarlo a cabo se ha adquirido una base de datos de la plataforma “Kickstarter”, la más conocida del sector. Tras su procesado, resulta un conjunto de datos muy representativo del periodo de estudio sobre el que se aplicarán técnicas de minería de datos.

El estudio se realiza en dos pasos. En el primero se caracteriza el éxito de los proyectos a través de la identificación de los factores principales que intervienen en el éxito, en segundo lugar, se realiza una clasificación y predicción de éxito en los proyectos, con el fin de identificar cuales son las características que tienen mayor capacidad para inducir al éxito un proyecto y así asistir al creador a dirigir sus esfuerzos sobre las más influyentes.

A través de técnicas de clusterización y asociación se identifican seis clústeres de proyectos, cada uno con características diferenciadoras del resto y que suponen una herramienta de gran ayuda para el creador por su capacidad para potenciar proyectos hacia el éxito. Estos seis clústeres caracterizan estados finales de proyectos, resultando tres estadios diferentes: dos de éxito, uno de advertencia y tres de fracaso. Así, un creador puede ubicar su proyecto en el sistema asociando las características de su proyecto al clúster que mejor se ajuste y posteriormente puede decidir si es necesario modificar las características de su proyecto en base a la adaptación hacia un clúster con mayor potencial de éxito.

Finalmente se han aplicado técnicas de clasificación y predicción del éxito de los proyectos en base a sus características más relevantes, demostrando tanto su compatibilidad con las técnicas anteriores como su carácter de respuesta ágil cuando se aplican de manera unilateral.

ABSTRACT

Crowdfunding is a method of financing highly innovative projects that was born in response to the difficulty of financing projects during the last global economic crisis. Away from banking institutions, the secret of its success consists in dividing the risk into smaller parts, so that a crowd of potential investors can collaborate in the project with a small amount of money while waiting for a reward, the product of the project.

The number of projects that obtain funding using this method has been increasing since its inception but has reached a point where there has been a decline and the ratio of successful to unsuccessful projects is not improving. The motivation of this study consists in providing a tool capable of promoting projects towards success or at least keeping them away from failure, in order to increase the success rate and provide a high degree of sustainability to this type of financing.

To achieve this, it will study the main factors involved in crowdfunding in order to satisfy the creator, by being able to succeed with his project, and on the other hand, to please the project's collaborators when their reward materialises.

To carry it out, it was acquired a database from the "Kickstarter" platform, the best known in the sector, providing highly representative set of data for the study period. Its unstructured nature has led to the application of different techniques to clean and give structure to the data, finally obtaining a balanced working dataset that is sufficiently representative for the subsequent application of data mining techniques.

The study is carried out in two steps, the first one characterises the success of the projects through the identification of the main factors involved in success. The second, it performs a classification and prediction of success in the projects, in order to identify which are the characteristics that have greater capacity to induce the success of a project and thus assist the creator to direct his efforts on the most influential ones.

Through clustering and association techniques we identify six clusters of projects, each of which has characteristics that distinguish them perfectly and they are a great tool in order to help the creator due to their capacity to promote projects towards success. These six clusters describe the end states of projects, resulting in three different stages: two of success, one of warning and three of failure. So, a creator can place a project in the system by associating the characteristics of his project to the cluster that best fits and can subsequently decide whether it is necessary to modify the characteristics of his project based on the adaptation towards a cluster with greater potential for success.

Finally, it has been applied techniques for classifying and predicting the success of projects based on their most relevant characteristics, demonstrating both their compatibility with the previous techniques and their agile response character when applied unilaterally.

ÍNDICE

1	INTRODUCCIÓN	5
2	OBJETIVO	7
3	ENTORNO DEL CROWDFUNDING	9
3.1	CROWDFUNDING	9
3.1.1	<i>Origen</i>	9
3.1.2	<i>Que es el crowdfunding</i>	10
3.1.3	<i>Características principales</i>	11
3.1.4	<i>Recorrido histórico del Crowdfunding.</i>	13
3.1.5	<i>Que tipo de proyectos es posible financiar</i>	15
3.1.6	<i>Entorno de funcionamiento del crowdfunding</i>	16
3.1.7	<i>Tipos de crowdfunding</i>	20
3.1.8	<i>Alternativas al crowdfunding</i>	22
3.1.9	<i>Otras formas de financiación</i>	32
3.1.10	<i>Características básicas de los sistemas</i>	32
4	PERSPECTIVAS DEL PROBLEMA	35
4.1	CARACTERIZACIÓN DEL ÉXITO.....	35
4.2	PREDICCIÓN DEL ÉXITO	36
4.3	FRACASO.....	37
4.3.1	<i>Para el creador</i>	37
4.3.2	<i>Para el colaborador</i>	38
4.4	CONTROL DEL AVANCE DE LA CAMPAÑA	39
4.4.1	<i>Patrones</i>	39
5	METODOLOGÍA	41
5.1	CRISP-DM	41
5.1.1	<i>Estructura</i>	42
5.1.2	<i>Fases</i>	42
6	TÉCNICAS EMPLEADAS	46
6.1	PCA (PRINCIPAL COMPONENTS ANALYSIS)	47
6.2	SELF-ORGANIZING MAPS.....	49
6.3	CLUSTERIZACIÓN Y EL MÉTODO K-MEANS.....	54
6.4	MULTIVARIATE ADAPTATIVE REGRESSION SPLINE	56
6.5	DECISION TREE	57
6.6	NUBE DE ETIQUETAS.....	60
7	CARACTERIZACIÓN DEL ÉXITO EN LOS PROYECTOS	63
7.1	PROCESADO DEL CONJUNTO DE DATOS	63
7.1.1	<i>Recolección de datos</i>	64
7.1.2	<i>Descripción de los datos</i>	67
7.1.3	<i>Adaptación de los datos</i>	73
7.1.4	<i>Exploración de los datos</i>	78
7.2	IDENTIFICACIÓN DE ZONAS DE ÉXITO MEDIANTE CLUSTERIZACIÓN	89
7.2.1	<i>Representación mediante Self-Organizing Maps</i>	89
7.2.2	<i>Clustering</i>	91
7.2.3	<i>Éxito en los clústeres identificados</i>	97

7.2.4	Indicadores de estado.....	98
7.3	EVALUACIÓN DE RESULTADOS	104
8	CLASIFICACIÓN Y PREDICCIÓN DEL ÉXITO DE LOS PROYECTOS.....	108
8.1	CLASIFICACIÓN	108
8.2	PREDICCIÓN	110
8.3	INTERACCIÓN ENTRE MODELOS	115
9	CONCLUSIONES.....	118
10	LÍNEAS DE FUTURO	120
11	REFERENCIAS	121
ANEXOS	134

ÍNDICE DE TABLAS

TABLA 1: PRINCIPALES DIFERENCIAS ENTRE LA FINANCIACIÓN TRADICIONAL Y EL CROWDFUNDING.....	11
TABLA 2: CONFIGURACIÓN DE VISUALIZACIÓN PROGRAMA VOSVIEWER.....	18
TABLA 3: DATA SETS COMPARADOS POR SUS VARIABLES MÁS SIGNIFICATIVAS	66
TABLA 4: CARACTERIZACIÓN DE VARIABLES.....	68
TABLA 5: DISTRIBUCIÓN DE SUBCATEGORÍAS POR CATEGORÍA.....	69
TABLA 6: NÚMERO DE DATOS DE CADA VARIABLE EN EL DATA SET	72
TABLA 7: CARACTERIZACIÓN DE LA VARIABLE STATUS	73
TABLA 8: CONTENIDO DE PROYECTOS EN CADA SUBCATEGORÍA.....	74
TABLA 9: PRINCIPALES CARACTERÍSTICAS DE LAS CATEGORÍAS	79
TABLA 10: VARIABLES DEL DATA SET USADAS PARA EL ESTUDIO.....	80
TABLA 11: CORRELACIÓN DE VARIABLES.....	81
TABLA 12: PRUEBA KOLMOGOROV-SMIRNOV CON CORRECCIÓN DE SIGNIFICACIÓN DE LILLIEFORS.....	84
TABLA 13: PRUEBA DE MÚLTIPLES FACTORES	85
TABLA 14: PRINCIPALES CARACTERÍSTICAS DE LOS CLÚSTERES IDENTIFICADOS	93
TABLA 15: IDENTIFICACIÓN DE VALORES ALTOS Y BAJOS DE VARIABLE POR CLÚSTER.....	96
TABLA 16: PORCENTAJE DE ÉXITO POR CATEGORÍA Y CLÚSTER.....	97
TABLA 17: NÚMERO TOTAL DE PROYECTOS POR CATEGORÍA Y CLÚSTER	98
TABLA 18: DISTRIBUCIÓN DE VALORES DE SRI POR CATEGORÍA Y CLÚSTER	100
TABLA 19: DISTRIBUCIÓN DE VALORES DE SOA POR CATEGORÍA Y CLÚSTER.....	101
TABLA 20: DISTRIBUCIÓN DE VALORES DE FOA POR CATEGORÍA Y CLÚSTER.....	101
TABLA 21: IMPORTANCIA DE LAS VARIABLES: MODELO PREDICCIÓN PLEDGED	111
TABLA 22: ERRORES EN LA PREDICCIÓN	112
TABLA 23: MATRIZ DE CONFUSIÓN CON CLÚSTERES	117
TABLA 24: CARACTERÍSTICAS COMPARADAS DE LOS SISTEMAS DE FINANCIACIÓN DE PROYECTOS.....	135
TABLA 25: LOADINGS.....	137
TABLA 26: VARIANZA EXPLICADA.....	138

ÍNDICE DE FIGURAS

FIGURA 1: INFRAESTRUCTURA DEL CROWDFUNDING (THE WORLD BANK, 2013)	13
FIGURA 2: LÍNEA TEMPORAL DE EVOLUCIÓN DEL CROWDFUNDING.....	15
FIGURA 3: NÚMERO DE ARTÍCULOS PUBLICADOS POR AÑO.....	17
FIGURA 4: MAPEO DE TÉRMINOS DURANTE EL PERIODO 2012-2017	19
FIGURA 5: TIPOS DE CROWDFUNDING.....	20
FIGURA 6: RELACIÓN TIEMPO-BENEFICIO EN LOS MECANISMOS DE FINANCIACIÓN DE PROYECTOS (MUÑOZ, 2011)	33
FIGURA 7: INVERSIÓN INDIVIDUAL EN CADA FASE DE LA EMPRESA (X.NET AND CAPARRÓS, 2014)	34
FIGURA 8: ESQUEMA DE LOS CUATRO NIVELES DE ABSTRACCIÓN DE LA METODOLOGÍA CRISP-DM	42
FIGURA 9: FASES DEL MODELO DE REFERENCIA CRISP-DM	43
FIGURA 10: FASES DE ANÁLISIS DEL PROBLEMA DEL MODELO DE REFERENCIA CRISP-DM.....	43
FIGURA 10: FASES DE ANÁLISIS DE DATOS DEL MODELO DE REFERENCIA CRISP-DM	44
FIGURA 10: FASES DE PREPARACIÓN DE LOS DATOS DEL MODELO DE REFERENCIA CRISP-DM	44
FIGURA 10: FASES DE MODELADO DEL MODELO DE REFERENCIA CRISP-DM.....	44
FIGURA 10: FASES DE EVALUACIÓN DEL MODELO DE REFERENCIA CRISP-DM	45
FIGURA 10: FASES DE IMPLANTACIÓN DEL MODELO DE REFERENCIA CRISP-DM	45
FIGURA 10: PERCEPTRÓN COMO UNIDAD BÁSICA.....	50
FIGURA 11: ORGANIZACIÓN DE LAS CAPAS EN UNA RED NEURONAL ARTIFICIAL	52
FIGURA 12: MODELO KOHONEN DE CONEXIONES SINÁPTICAS (HAYKIN, 2009)	53
FIGURA 13: MATRIZ DE DISTANCIAS UNIFICADAS (U-MATRIX)	54
FIGURA 14: ELEMENTOS ESTRUCTURALES DEL ALGORITMO CART.....	59
FIGURA 15: NUBE DE ETIQUETAS DE LA PÁGINA WEB DE ARCELORMITTAL ESPAÑA.....	62
FIGURA 16: NÚMERO DE PROYECTOS EXITOSOS Y FRACASADOS DESDE EL INICIO DE KICKSTARTER (BIDAUX, 2019).....	64
FIGURA 17: CONTEO DE PALABRAS MÁS FRECUENTES	75
FIGURA 18: TÉRMINOS MÁS USADOS EN EL TÍTULO DEL PROYECTO	76
FIGURA 19: GRÁFICA GOAL VS PLEDGED.....	78
FIGURA 20: ENFRENTAMIENTO DE VARIABLES CON COMPORTAMIENTO PARTICULAR	82
FIGURA 21: EVOLUCIÓN DEL ÉXITO Y FRACASO	83
FIGURA 22: PROPORCIÓN DE LA VARIANZA ACUMULADA PARA CADA UNA DE LAS COMPONENTES.....	87
FIGURA 23: MAPA AUTOORGANIZADO DE CADA VARIABLE A ESTUDIO, TAMBIÉN INCLUYE U-MATRIX	90
FIGURA 24: POSICIÓN E IDENTIFICACIÓN DE CLÚSTERES.....	92
FIGURA 25: PROYECCIÓN DE ÉXITO Y FRACASO SOBRE EL MAPA DE CLÚSTERES DE LA FIGURA 24	94
FIGURA 26: IDENTIFICACIÓN DE LA ZONA DE ÉXITO.....	95
FIGURA 27: ZONAS DE ÉXITO Y FRACASO SUPERPUESTAS A LA FIGURA 24	95
FIGURA 28: DISTRIBUCIÓN DE LOS VALORES DE SOA DE CADA CATEGORÍA POR CADA CLÚSTER	102
FIGURA 29: DISTRIBUCIÓN DE LOS VALORES DE FOA DE CADA CATEGORÍA POR CADA CLÚSTER	103
FIGURA 30: ETIQUETADO DE CLÚSTERES	105
FIGURA 31: CART DECISION TREE.....	109
FIGURA 32: PORCENTAJE DE ACIERTOS CON UN MARGEN DE ERROR RELATIVO EN FASE DE PRUEBA	112
FIGURA 33: ANÁLISIS DE SENSIBILIDAD POR PAREJAS DE VARIABLES PREDICTORAS.....	114
FIGURA 34: ANÁLISIS DE SENSIBILIDAD CON UNA VARIABLE DE ENTRADA	115
FIGURA 35: DECISION TREE COMPUESTO	116
FIGURA 36: RESIDUOS DE PLEDGED ESTIMADOS EN CADA CLÚSTER POR EL MODELO	117
FIGURA 37: SISTEMA DE DECISIÓN PARA LAS ALTERNATIVAS AL CROWDFUNDING DE RECOMPENSA	136

1 INTRODUCCIÓN

Hoy en día disponer de financiación para una idea o proyecto nuevo no es fácil. Los proyectos o ideas de carácter innovador son muy difíciles de financiar, ya que su baja credibilidad deja a la vista la probabilidad de no obtener retorno de la inversión (Bustos Contell, 2013). Ante un entorno de crisis económica, estos proyectos producen una aversión para un inversor no especializado, por el riesgo que conlleva lo desconocido y la incertidumbre de la propia novedad (Lee et al., 2015). Las entidades bancarias lejos de facilitar la obtención de crédito, usan sus herramientas administrativas para asegurarse que detrás de ese proyecto, haya un riesgo mínimo y así no agravar el problema de falta de liquidez, más aún, cuando las condiciones del entorno no son las más apropiadas para generar confianza en un inversor, ya que factores externos pueden contribuir negativamente e incluso llevar al fracaso a un proyecto técnicamente exitoso (Bakker, 2013). Por ello, estos proyectos resultan ser los más frágiles y castigados.

Liberar del estancamiento a proyectos con factor de innovación importante no sólo beneficia al creador y al patrocinador, sino que también contribuye al crecimiento económico de un país, al influir en diversos factores, principalmente relacionados con la oferta, la demanda y la eficiencia (Boldeanu and Constantinescu, 2015). No se entiende un crecimiento sostenido si no está apoyado en el desarrollo tecnológico y la innovación, que mejore la capacidad de adaptación a base de incorporar nuevas técnicas e incrementar la eficiencia de los procesos.

Todos los métodos de financiación conocidos hasta ahora carecen de una conexión directa entre proyecto e inversor, por lo que el financiador no tiene la necesidad de conocer las características del proyecto. Como primera alternativa a la financiación bancaria directa aparecen figuras financieras más especializadas como el *Capital Riesgo* o los *Business Angels*, destinados a invertir o financiar en empresas en fases iniciales, donde la credibilidad y los flujos de caja no son tan altos como para asegurar que serán un valor seguro (Soto-Moya, 2016), en muchos casos este aumento de recursos lleva consigo la pretensión por parte del prestamista de tener cierta presencia en el consejo de administración, para que, sin ánimo de intervenir en las decisiones cotidianas de la empresa, si puedan participar activamente en la toma de decisiones estratégicas (Dibrova, 2015).

El interés por cubrir a un mayor número de proyectos dirige a las diferentes entidades a identificar la potencialidad de ciertos proyectos mediante la contratación de agentes especializados. Bien sea, mediante un asesor externo o recibiendo el apoyo de inversores con experiencia previa en ese tipo de proyectos (Hall and Lerner, 2009) y serán los que estudien la viabilidad de cada caso, para verificar el posible riesgo de la inversión y poder impulsar los proyectos con mayor probabilidad de éxito. En este punto se puede afirmar ya, que no existe una relación directa entre credibilidad y probabilidad de éxito.

Para rebajar la sensación de riesgo es frecuente que el poseedor de la idea haga uso de todos los recursos propios posibles, con el fin de que las necesidades a cubrir por una entidad externa sean las mínimas, técnicas conocidas como *Bootstrapping* (Waleczek et al., 2018), también ayuda buscar parte de la financiación inicial entre amigos y familiares, gracias a que estos pueden conocer de primera mano el proyecto y quién lo llevará a cabo, lo que se conoce como FFF (Family, Friends and Fools) (Kotha and George, 2012a).

Llegado el punto en el que el proyecto se encuentra estancado por la ausencia de recursos que permitan continuar con el desarrollo del ciclo de vida del proyecto (PMBOK, 2013), se hace latente la necesidad de multiplicar las oportunidades de financiación dando a conocer el proyecto a un número grande de inversores y no sólo limitarlo a aquellos propuestos por las diferentes entidades. Una vez concedido apoyo financiero y mirando mas allá de la obtención de los recursos económicos, el proyecto también se beneficiará de una mejor imagen y mayor confianza de cara a terceras partes (Gerber and Hui, 2013), ya que un proyecto financiado actúa como un aval o seguro en el que confiaron otros inversores.

El procedimiento más eficiente para dar a conocer las necesidades del proyecto es hacer uso de internet. Gracias a la evolución de las tecnologías de la información y la comunicación es posible acceder y compartir toda clase información de manera inmediata. Con el surgimiento de internet se amplía enormemente la capacidad de multiplicar el impacto de cualquier actividad, siempre que sea difundida en el espacio adecuado, bien a través de foros, redes sociales, blogs...etc. (Behringer and Sassenberg, 2015). Por otro lado, el uso de internet como medio implica aceptar el hecho de navegar en un entorno muy activo y susceptible al cambio, estimulado por factores como: tendencias colectivas, cambios en plataformas Web, actualización de normativas, etc. Siendo capaz de amenazar seriamente proyectos con factor de innovación importante en sus productos, como los dispositivos electrónicos (Wilhelm et al., 2015).

2 OBJETIVO

El crowdfunding permite financiar un proyecto apoyándose en una multitud de colaboradores interesados en una recompensa que surge como producto del proyecto. Esta desvinculación de entidades capaces de dotar de financiación a los proyectos hacia una multitud potencialmente inversora va más allá del entorno de crisis económica que lo vio nacer, para establecerse y continuar libremente como una forma de adquisición de recursos económicos independiente y alternativa a las actuales.

La escasez de herramientas que vinculen un determinado proyecto hacia un método de financiación u otro son escasas y poco fiables, debido a la diferente naturaleza de los proyectos y al riesgo percibido que en muchas ocasiones dista mucho del riesgo real.

A pesar de la lenta recuperación económica después de la crisis, los proyectos de crowdfunding siguen siendo una forma potencial de desarrollar proyectos altamente innovadores al margen del contexto social o económico, sin embargo, no todos los creadores que han elegido este método de financiación llevan sus proyectos al éxito, una gran parte y cada vez más, presentan sus proyectos a las plataformas en aras de ser financiados y no son capaces de alcanzar la financiación.

Este estudio pretende proporcionar recursos para garantizar la sostenibilidad del crowdfunding como medio de financiación de proyectos, en base a facilitar a los creadores una herramienta basada en indicadores del proyecto y de la campaña, para que puedan conocer la trayectoria del proyecto en base a unos estados finales, de características conocidas y etiquetados como de éxito o fracaso, de esta manera, un creador puede conocer el estado al que se dirige su proyecto, rectificar la trayectoria si es necesario y reforzar o asegurar la financiación.

Este trabajo plantea dar solución al problema basándose en datos históricos de proyectos, para lo cual se adquiere una base de datos muy representativa de proyectos de crowdfunding sobre la que se aplican técnicas de minería de datos, posteriormente se extraerán y clasificarán los factores más importantes que intervienen en el éxito o fracaso de este tipo de proyectos y mediante diferentes técnicas de clasificación y agrupamiento se estudiará cómo los proyectos se pueden asociar a grupos con características identificativas y diferenciables, de tal manera que permita de antemano conocer el punto final de su trayectoria.

El conjunto de datos se compone de un número elevado de casos reales con un número limitado de variables implicadas, lo que permite realizar estudios con alta fiabilidad en sus resultados. Con el fin de extraer la máxima información se opta por una estrategia de dos pasos:

- Realizar una agrupación de proyectos de éxito en base a su naturaleza y sus factores más influyentes, con el fin de identificar elementos comunes que aporten capacidad para definir la estrategia adecuada dependiendo de la

naturaleza del proyecto. De esta manera se aporta una sólida base para que el creador pueda definir una táctica que dirija a un proyecto hacia el éxito o al menos lo aleje del fracaso.

- A través del conjunto de datos adquirido y mediante el uso de un algoritmo basado en árboles de decisión se podrá valorar la efectividad del proyecto y a partir de ellos realizar estudios what-if que permitan su mejora. De esta manera se alcanzará una predicción de éxito o fracaso de los proyectos, en función de la selección de las variables de entrada más relevantes y categorizará el estado final de un proyecto basándose en la capacidad para inducir al éxito de cada categoría.

Ambas estrategias de estudio están encaminadas a reforzar la sostenibilidad del crowdfunding como medio de financiación de proyectos innovadores, más allá de las condiciones particulares de cada proyecto y del entorno de crisis que derivó en su origen. Por ello, este estudio proporciona información para mejorar la viabilidad del crowdfunding usando como punto de vista la mejora del ratio de éxito de los proyectos, de esta manera se induce a creador un potencial a realizar los cambios necesarios en su proyecto en aras de alcanzar el éxito.

3 ENTORNO DEL CROWDFUNDING

3.1 CROWDFUNDING

3.1.1 ORIGEN

La búsqueda de recursos haciendo uso de internet se beneficia de la interconexión de las partes implicadas y sus relaciones multidireccionales, por lo que resulta ser un ecosistema ideal para el surgimiento de nuevos modelos de desarrollo, capaces de aprovechar el potencial de una multitud para alcanzar un objetivo concreto, cambiando de esta manera la perspectiva habitual del mercado, no siendo el inversor quien porte el problema para buscar solución.

Este nuevo escenario receptivo a compartir para evolucionar va a dar lugar a la aparición del crowdsourcing (Schwienbacher and Larralde, 2010), anglicismo que proviene de la unión de las palabras, “crowd” (multitud) y “outsourcing” (externalización), y que plantea una exteriorización de los problemas o retos, exponiéndolos a una gran multitud a través de internet y con la finalidad de búsqueda de soluciones o ideas en vías de llevarse a cabo (Assis Neto and Santos, 2018), fue citado por primera vez por Jeff Howe en *Wired Magazine*, una revista americana dedicada a la alta tecnología (Howe, 2006).

Mediante el *crowdsourcing*, una empresa puede ofertar un problema o proyecto para cubrir una necesidad propia y específica. Esta oferta aparecerá publicada (generalmente en internet) para que una cantidad de personas, de número indeterminado y generalmente muy grande, realice sus propuestas a la solución.

Este modelo de ayuda bidireccional propone un intercambio de un proyecto por una cuantía económica (generalmente inferior a la del mercado), un puesto de trabajo o una recompensa de cualquier otro tipo (Assis Neto and Santos, 2018). Una vez distribuido el problema ante una multitud, el crowdsourcing facilita enormemente la obtención de feedback que aporta gran cantidad de información para resolución de problemas e incluso puede llegar a generar conocimiento nuevo, dando lugar a la *inteligencia colectiva* (Mačiulienė and Skaržauskienė, 2015).

La idea principal del crowdsourcing consiste en involucrar al consumidor, en base a su interés por el producto, para que aporte una posible solución a un problema, rompiendo el esquema dual tradicionalmente separado de productor y consumidor pasivo. Pasando ahora a ser un *trabajador consumidor*, dispuesto a llevar parte de la carga de trabajo que anteriormente pertenecía al productor (Kleemann and Voß, 2008).

Tras la idea de crowdsourcing fundamentada básicamente en el interés de una multitud por una recompensa (Brabham, 2008), emerge la idea de ofrecer ayuda económica en lugar de know-how como nuevo medio de financiación, con la principal ventaja sobre los medios de financiación tradicionales que no es un número limitado de personas las

que deciden si el proyecto puede ser viable o no, sino que en su lugar lo hace una multitud potencialmente inversora.

3.1.2 QUE ES EL CROWDFUNDING

Partiendo de la idea de Jeff Howe, en agosto de 2006, Michael Sullivan acuñó en su blog *fundavlog* el término de *crowdfunding* (Gobble, Mary Anne M., 2012), otro anglicismo de origen similar al *crowdsourcing*, cuya etimología proviene de las palabras: “crowd-” (multitud) y “-funding”, (financiación), traducido al español se entiende por: micromecenazgo, financiación colectiva o financiación en masa.

La idea de búsqueda de recursos económicos de manera colectiva para desarrollar un proyecto no es nueva, así que tampoco se puede asociar directamente su existencia a internet. La primera vez que se tiene constancia de esta adquisición de recursos con los elementos principales del crowdfunding, fue cuando Francia regaló a Estados Unidos la famosa Estatua de la Libertad, donde Joseph Pulitzer, editor y propietario del conocido periódico estadounidense “New York World”, comenzó a publicar en su periódico artículos sobre la necesidad de hacer un pedestal para la Estatua de la Libertad. Mediante una buena campaña de marketing a través de su prensa, convenció a una multitud de clase media y alta para recaudar la cantidad necesaria de dinero y construir el pedestal con un valor total de \$100.000, a cambio Pulitzer se comprometió a publicar la lista de todas aquellas personas que lo habían hecho posible (U.S. National Park Service, 2016), fuese cual fuese su aportación económica.

Se puede definir *crowdfunding* como una forma alternativa de financiación de ideas o proyectos, nacida para acoger aquellos con difícil financiación por su carácter claramente novedoso o innovador, en el que cualquier persona puede hacer una aportación económica para obtener una recompensa o simplemente y de manera altruista por la satisfacción de haber colaborado en un proyecto. Así se consigue un doble objetivo: conseguir la financiación para su desarrollo y de forma simultánea conseguir el cliente, puesto que cliente y financiador son parte indisoluble del método en este tipo de proyectos.

Mediante crowdfunding los fondos necesarios son adelantados, frente a la tradicional venta de producto, lo que disminuye el riesgo y posibilita el desarrollo del proyecto (Negruşa et al., 2014). Esta relación se realiza de forma preferente a través de plataformas específicas, que ponen en contacto a personas interesadas en este tipo de productos, con emprendedores en condiciones de ofrecerlos, de otra forma no llegarían a estar en contacto, provocando necesidades no cubiertas y expectativas no realizadas

Armin Schwenbacher y Benjamin Larralde proponen una definición formal del crowdfunding que es aceptada hoy en día por la mayoría de autores: “Una convocatoria abierta, fundamentalmente a través de Internet, para la provisión de recursos financieros, ya sea en forma de donación o a cambio de algún tipo de recompensa y/o

los derechos de voto con el fin de apoyar las iniciativas para propósitos específicos” (Schwienbacher and Larralde, 2010).

Mediante el crowdfunding un creador o un empresario puede plantearse la financiación externa de su proyecto a través de una convocatoria abierta en internet, donde cada individuo dona una cantidad pequeña de dinero, en lugar de solicitar una cantidad grande a un grupo más o menos pequeño de inversores sofisticados (Belleflamme et al., 2014).

Las aportaciones realizadas por financiadores o mecenas deben llegar a una cantidad mínima de dinero en un tiempo determinado, llegado a este punto, el portador de la idea o proyecto recibirá la financiación recaudada y el proyecto puede llevarse a cabo.

Cabe destacar que este tipo de financiación se enmarca en un contexto de proyectos diferente de los habituales, es por ello, que un gran número de proyectos con una fuerte base científica se dirijan a este tipo de modelo durante el periodo de investigación, tanto por su capacidad para autofinanciarse como por divulgación que permite el presentarlo a una multitud (Wheat et al., 2013).

3.1.3 CARACTERÍSTICAS PRINCIPALES

El crowdfunding se basa en la aportación de pequeñas cantidades de dinero por parte de una multitud, sin más intermediarios que una plataforma virtual y sus gestores, por lo que hay dos rasgos identificativos que hacen que esta vía de financiación de proyectos se diferencie mucho del resto de métodos, por un lado, no tiene intermediarios y por otro, la financiación de un proyecto se puede obtener de modo altruista, es decir sin estudio de viabilidad previa (Schwienbacher and Larralde, 2010).

En la Tabla 1 se encuentran enfrentadas las características más remarcables de la financiación tradicional y las del crowdfunding.

	Financiación tradicional	Crowdfunding
Tipo de proyecto	Necesidad de selección del modelo más apropiado al proyecto	Válido para cualquier proyecto innovador
Estudio de viabilidad	Necesario análisis de riesgos y estudio de viabilidad financiera	No necesario
Financiador	Entidad independiente	Cliente
Coste de la financiación	Porcentaje fijado de antemano	Proporcional al éxito de la financiación
Riesgo	Concentrado en la entidad financiera	Distribuido entre financiadores
Protección del proyecto	Alta confidencialidad	Alto riesgo de copia por exposición pública
Posibilidad de ampliación	Si, con requisitos de solvencia	No

Tabla 1: Principales diferencias entre la financiación tradicional y el crowdfunding

El éxito de este fenómeno actual se debe a la evolución de internet y a la creación de las primeras redes sociales a mediados de la década de los 90, estas han ido mejorando en capacidad y en velocidad para unir, primero a pequeños grupos de personas y después gran una multitud dispuesta a hacer los primeros intercambios usando internet como medio (Schwienbacher and Larralde, 2010).

Un factor clave para que este tipo de financiación tenga éxito es el hecho de cuidar los vínculos con la multitud desde el inicio del proyecto, manteniéndola informada a través de la plataforma de crowdfunding o haciendo uso de las redes sociales (Zheng et al., 2014), por ello, es de gran importancia generar un clima de confianza hacia los posibles interesados en el proyecto, cuidando el medio de presentación del proyecto.

Hay que tener en cuenta que, de manera opuesta a la financiación tradicional, en lugar de presentar los resultados al público tras terminar el proyecto, el crowdfunding necesita el apoyo y conocimiento público antes de empezar el mismo. Un factor clave para que este tipo de financiación tenga éxito, es el hecho de cuidar desde el inicio del proyecto, los vínculos con la multitud (Mollick, 2014), tanto a través de la plataforma de crowdfunding como a través de las redes sociales.

Desde su comienzo, la financiación tipo crowdfunding no sigue un estándar definido, sino que es un proceso flexible y con innumerables posibilidades de llevarlo a cabo. A pesar de que es posible el desarrollo de una campaña de crowdfunding sin hacer uso de internet como medio de publicidad, hoy en día no tiene sentido prescindir de las capacidades de distribución de la Web, sin embargo se ha demostrado que la combinación de una campaña de crowdfunding, con comunicación boca a boca a través de redes sociales como Facebook o Twitter, contribuye a aumentar la probabilidad de recibir fondos en periodos cortos de tiempo (Moqri and Bandyopadhyay, 2017). Incluso es posible llevar a cabo una campaña de financiación de este tipo sin una plataforma ad hoc como en el caso del *Do it Yourself crowdfunding*, donde a través de una página web personal se puede publicitar la campaña de recaudación (Quilageo, 2015), con el principal inconveniente de que no proporciona la suficiente confianza al carecer de cierto control y regulación, es por ello que en este tipo de crowdfunding sea más común cuando el impulsor alcanza un alto grado de popularidad previo (Johnson et al., 2017).

Aunque es posible el desarrollo de una campaña de crowdfunding mediante una plataforma privada (Olson, 2015), las plataformas públicas evitan entre otros, la falsedad en los proyectos o el fraude en los pagos, para lo que se apoyan en diferentes formas de mostrar tanto la seguridad como la transparencia, mediante informes anuales, cifras de éxito, cantidad de empleos creados, etc. (The World Bank, 2013)

En la Figura 1 se muestran los lazos bidireccionales entre comunidad, empresa y capital atravesando siempre el marco de las plataformas de crowdfunding.

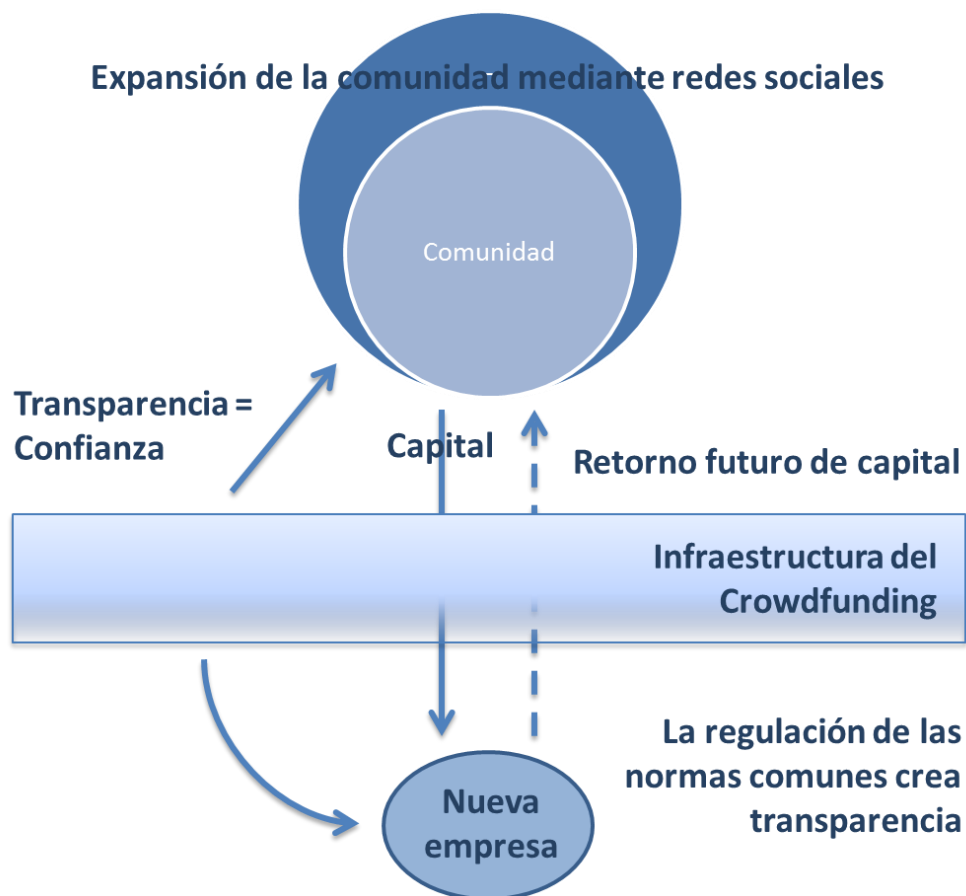


Figura 1: Infraestructura del Crowdfunding (The World Bank, 2013)

3.1.4 RECORRIDO HISTÓRICO DEL CROWDFUNDING.

Antes de llegar al modelo de financiación del *crowdfunding* a través de una plataforma Web, los intercambios de información han evolucionado desde las primeras redes sociales basadas en la comunicación *emisor-receptor*, propios de la Web 1.0, hasta los intercambios de información colectivos, simultáneos y multidireccionales de hoy en día pertenecientes a la conocida Web 2.0 (Newman et al., 2016).

En 1994 se forma GeoCities, una plataforma de alojamiento Web gratuita, considerada como la primera plataforma social, donde el usuario podía compartir sus experiencias asociándolas a un grupo atendiendo a la naturaleza de su contenido. En 1997 Adam Hinkley, un joven programador australiano de 17 años, desarrolla una aplicación llamada *Hotline Connections* para su sistema operativo MAC OS, en el que todavía sin servidores, los usuarios podían compartir contenidos de ordenador a ordenador, a modo de red, con el problema adicional que un ordenador apagado no podía compartir información (DeMartino, 2015). Comenzaba lo que hoy se conoce como aplicaciones P2P (peer-to-peer).

Pero no fue hasta 1997 cuando se forman SixDegrees.com y AOL Instant Messenger, las cuales, partiendo de un perfil personal y una lista de amigos permitía a los usuarios contar sus experiencias de manera similar a como se conoce hoy en día (Castrataro, 2011).

Paralelamente, también en 1997 el grupo británico *Marillion* fue ejemplo mundial de financiación al recaudar a través de internet la cifra de \$60,000 para hacer una gira por Estados Unidos (Gamble et al., 2017).

En 1999 la plataforma Napster se convierte en un referente para compartir archivos. Esta vez a través de un servidor o FTP, una multitud de usuarios podían subir sus archivos a una plataforma Web (generalmente contenido multimedia), para su uso o descarga compartida. Debido al enorme éxito de la plataforma (aprox. 60 millones de personas) y por los consecuentes problemas de derechos de autor, dos años más tarde la plataforma Napster deja de funcionar (Madden, 2009).

En 2003 artistShare comienza a funcionar como plataforma para la recaudación de fondos de proyectos de ámbito musical a través de internet (artistShare, 2018), Sin embargo no fue hasta 2006, cuando aparece *Sellaband* simultáneamente a los conceptos de *crowdfunding* y *crowdsourcing*. Fue la primera página de crowdfunding con una interfaz de usuario y unas características que se mantienen hoy en día (Agrawal et al., 2015).

En 2008 se crea en San Francisco *Indiegogo* la primera Web con la que se puede financiar gran tipología de proyectos mediante *crowdfunding*, aunque no fue hasta abril de 2009 cuando nace la plataforma *Kickstarter* que da popularidad a este fenómeno gracias a sus proyectos más exitosos (Mollick, 2014).

El desencadenante del estado actual del crowdfunding sigue las tres ideas iniciales propuestas en el proyecto de Pulitzer; *Fraccionar el coste de un proyecto en pequeñas partes accesibles a futuros promotores, difundirlo para darlo a conocer al mayor número de personas posibles y finalmente ofrecer una recompensa suficientemente atractiva a cambio* (Leimeister, 2012).

No es hasta el año 2012 cuando aparece la primera regulación del crowdfunding en Estados Unidos, lugar que lo vio nacer y en el que se desarrolla con más actividad desde su inicio. Este tipo de control de la financiación llegará a España tres años más tarde.

La facilidad de acceso a este modelo ha crecido paralelamente gracias al desarrollo de internet, así en 2018 fue posible recaudar 4 mil millones de dólares mediante crowdfunding para el proyecto EOS, una criptomoneda basada en blockchain sin más recompensa para el usuario que su capital convertido a la criptomoneda Ethereum y que ha conseguido llegar a 1,28 millones de usuarios (Daily Hodl Staff, 2020). En la Figura 2 se observa la evolución temporal del crowdfunding atendiendo a sus principales hitos.

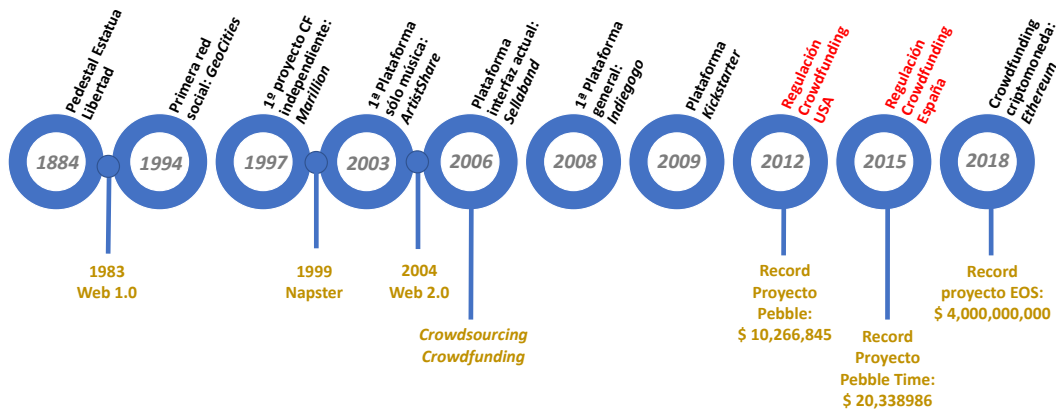


Figura 2: Línea temporal de evolución del crowdfunding

3.1.5 QUE TIPO DE PROYECTOS ES POSIBLE FINANCIAR

La diversidad de los proyectos financiables es tan amplia, como la imaginación de un impulsor para crearlos. Para dotar a este abanico de proyectos de cierta seguridad, las plataformas de crowdfunding estudian la viabilidad de cada proyecto antes de ser publicado. Esto ha propiciado no solamente el aumento de plataformas genéricas, sino que también aparezcan plataformas especializadas en proyectos de una determinada naturaleza, beneficiándose ambas partes del mayor conocimiento de la temática, dando impulso a proyectos que en una plataforma genérica no tendrían cabida.

Por esta razón, muchos proyectos son denegados por las plataformas incluso antes de ser analizados, al no cumplir normas propias o al ir contra su código de conducta. Así, la financiación de una película de cine erótico, no sería posible llevarla a cabo en la sección “Film & Video” de la plataforma genérica más conocida mundialmente www.KickStarter.com, pero sin embargo, sí lo es en una plataforma especializada en este tipo de cine como www.gogofantasy.com

Hay algunos ejemplos de éxito, de naturaleza muy diversa que han usado este tipo de financiación y que contribuyeron a resaltar y dar fama mundial al crowdfunding. Proyectos que a priori denotan poca credibilidad pero que han sido de gran impacto social y han contribuido a dar fe y confianza en este tipo tan particular de financiación. Algunos ejemplos son:

Campaña electoral del candidato a presidente Barack Obama: En 2008 finaliza la campaña política de Barack Obama para la presidencia de EEUU. A tal efecto, se crea una campaña de crowdfunding denominada “Yes, We Can”, publicitada a través de las redes sociales que logró recaudar un total de 631 millones de dólares, donde el 50% de las aportaciones era de menos de 200\$ (Hemer, 2011). Esta campaña nace con la intención de captar donantes altruistas y también sorteando premios como: una cena con George Clooney, cena con Obama y Bill Clinton, conocer a Michael Jordan, etc. dependiendo de la cantidad aportada (Dan Marom, 2012). En España, el partido político

“Podemos” fundado en 2014, alcanzó un 57,2% de financiación mediante crowdfunding privado para iniciar su andadura política.

Proyectos de entretenimiento: A través de la plataforma afincada en Brooklyn Kickstarter, el 19 de mayo de 2012 se finalizaba la campaña para financiar un smartwatch llamado *Pebble*, que con una meta de 100.000\$, logra recaudar 10.266.845\$ (Kickstarter, 2012). En España a finales de 2013 a través de la plataforma *Lánzanos*, una empresa de juegos de mesa presenta *Heroquest 25 aniversario*, un proyecto para recordar un antiguo juego de tablero que logra recaudar 679.927€ batiendo el record de recaudación por una plataforma en España (Lánzanos, 2018).

Proyectos de ciencia: El 10 de julio de 2013 un par de ingenieros de la Universidad de Michigan, finalizan con éxito la campaña de “*Roboroach*” recaudando 12.339\$, no sin diversas críticas por parte de diversas asociaciones, se trata de un circuito colocado sobre una cucaracha que mediante estímulos en sus neuronas permite dirigirla a través de un Smartphone (Kickstarter, 2013).

Existen plataformas especializadas y centralizadas en un tipo muy focalizado de proyectos como *Ecocrowdfunding.com* dedicada a proyectos que contribuyen a compensar la Huella de Carbono y mejorar la sostenibilidad de los medios. Incluso en Ucrania existe una plataforma que permite financiar tanto proyectos civiles como militares (People’sProject, 2018)

Cualquier tipo de proyecto es posible que encuentre su financiación mediante este medio siempre y cuando esté enmarcado dentro de la legislación vigente del país y las normas o código de conducta de cada plataforma.

3.1.6 ENTORNO DE FUNCIONAMIENTO DEL CROWDFUNDING

Hasta ahora, en el entorno de proyectos conocido, la motivación que lleva al desarrollo de un proyecto, más allá del beneficio económico que proporciona, toma como punto de partida una necesidad específica o un problema a solucionar, operando generalmente en un entorno “Pull” que atiende a cubrir dichas necesidades. La propia naturaleza de los proyectos de crowdfunding no busca tanto resolver un problema sino crear una necesidad a través de un proyecto diferente y llamativo, como cita la propia Kickstarter: “Kickstarter está diseñado específicamente para proyectos creativos”.

Bajo la perspectiva de un creador de proyectos de crowdfunding, la identificación de una necesidad a cubrir o un problema a resolver reside en su capacidad para identificarlo y para desarrollar el producto posteriormente, técnica asociada a entornos de mercado “Push”. Una vez identificados problema y solución, el proyecto debe darse a conocer a través de la plataforma para crear esa necesidad en posibles colaboradores (Gerber and Hui, 2013).

Hasta ahora el director del proyecto era el responsable de mantener el control del proyecto durante su ciclo de vida, en base a mantener bajo control los cinco grupos de

procesos más importantes: Inicio, Planificación, Ejecución, Monitoreo y Control y Cierre (PMBOK, 2013), por ello es necesario conocer si la concepción habitual de proyectos se mantiene para el crowdfunding.

Con el fin de analizar la actividad científica sobre el crowdfunding y mejorar la comprensión del marco en el que se encuadra, se realiza un análisis bibliométrico realizado haciendo uso de la plataforma WOS (Web Of Science) y del software de mapeo VOSviewer (versión 1.6.8). El análisis bibliométrico permite identificar y agrupar las palabras más relacionadas de tal manera que los términos son mutuamente dependientes cuando su utilización es muy común, así se puede observar gráficamente los conceptos o temas que los diferentes autores asocian con el crowdfunding.

Para llevar a cabo este tipo de estudio es necesario recopilar datos de artículos publicados en un periodo de tiempo representativo para el crowdfunding, por ello se adquiere información de artículos publicados en todas las bases de datos de WOS durante el periodo 2012–2017. Se elige 2012 como año de inicio por ser el año en el que se publica el primer artículo sobre el crowdfunding (*A Legal Study on the Crowdfunding as Means of obtaining capital*) y 2017 por ser representativo de la actualidad. La producción científica sobre el crowdfunding ha ido aumentando año a año durante el periodo de recolección de artículos, como muestra la Figura 3.

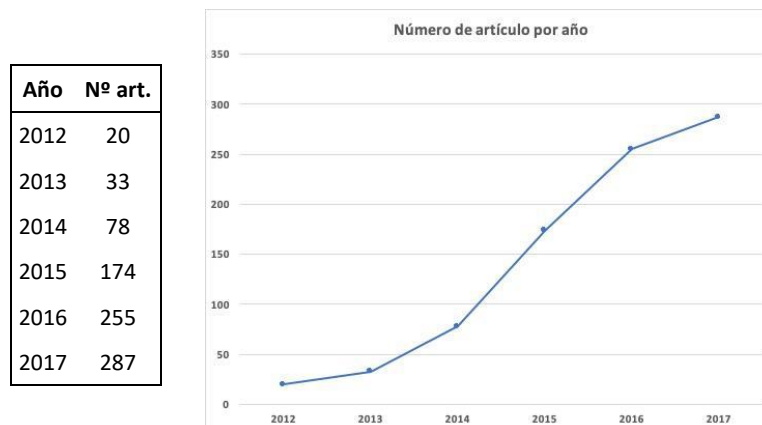


Figura 3: Número de artículos publicados por año

Con el fin de realizar una visualización de los términos más estudiados relacionados con el crowdfunding, se toman los 50 artículos más citados que contienen la palabra “crowdfunding” en el Título y/o Abstract, resultando así un total de 300 artículos científicos.

La visualización de palabras se realiza haciendo uso del programa de mapeo VOSviewer (versión 1.6.8), con el que es posible mapear la coocurrencia de palabras más significativas que estén contenidas dentro del Título y Abstract. También se ha considerado el contado total de palabras (Full counting) en lugar del contado binario, ya que, con este último se reduce de gran manera el número de palabras para reducir coocurrencias.

Se realiza una selección del umbral de un mínimo de 7 ocurrencias de un término dando lugar a un total de 1.274 términos, de los cuales 27 de ellos superan el umbral establecido.

Una vez recopilados los datos necesarios, se introducen los datos en el software VOSviewer y se eliminan las palabras menos relevantes. El propio programa selecciona por defecto y como óptimos diversos patrones de configuración, en función al número de datos, para su correcta visualización en pantalla, de esta manera el tamaño máximo de los términos que aparecen siempre se encuentra escalado respecto del mínimo, como se observa en la Tabla 2.

• Método de normalización: Asociación fuerte	• Método de normalización: LinLog (atracción lineal, repulsión logarítmica) / Modularidad
• Escala de visualización: 1.5	• Atracción: 5
• Peso: Ocurrencias	• Repulsión: 1
• Variación de tamaño: 0.5	• Visualización superpuesta: Pesos: Ocurrencia / Puntuación: Promedio de citas
• Ancho del núcleo: 1.5	
• Tamaño de etiquetas: 1.5	

Tabla 2: Configuración de visualización programa VOSviewer

En función de los datos restantes, VOSviewer crea un mapa de términos, ubicados de tal forma que la distancia entre dos pares proporciona una indicación del número de coocurrencias de los términos. En general, cuanto menor es la distancia entre dos términos, mayor es el número de coocurrencias entre ellos y por tanto más relacionados están, ya que aparecen con más frecuencia en el mismo texto.

Este mapeo de términos pretende ayudar a entender los esfuerzos científicos por ubicar este tipo de financiación de proyectos respecto a métodos de financiación tradicionales, desde la aparición de los primeros estudios de la materia hasta cinco años más tarde. Las palabras que presentan más coocurrencia con “crowdfunding” se conforman entorno a tres clústeres definidos claramente en la Figura 4: “entrepreneur”, “project” y “platform”. *Los colores Verde, amarillo y rojo indican si el score es: bajo, medio o alto respectivamente.

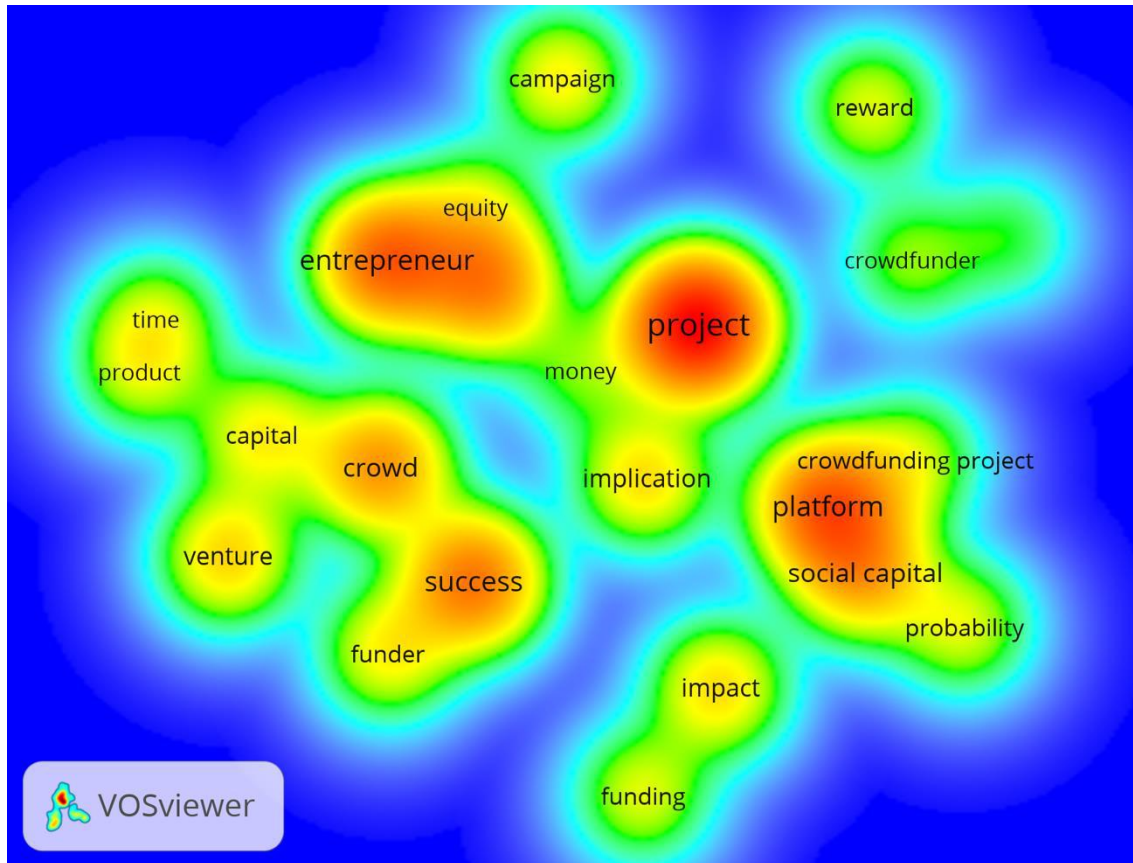


Figura 4: Mapeo de términos durante el periodo 2012-2017

El término “project” que protagoniza el clúster principal, tiene un núcleo claramente diferenciado ya que aparece fuertemente asociada al crowdfunding, lo que induce a concebir el crowdfunding hacia la perspectiva tradicional de un proyecto. El segundo núcleo por importancia está representado por “platform”, haciendo referencia a las plataformas de crowdfunding y el tercer grupo lo preside la palabra “entrepreneur”, asociado a personas que descubren una oportunidad para desarrollar un proyecto nuevo y obtener ganancias.

Resulta destacable el término “success” que representa el cuarto núcleo más importante, está más asociado a la palabra “crowd” que a la palabra “project”, debido a que es más común estudiar el éxito desde el punto de vista de la multitud que desde la plataforma o las características del proyecto.

A pesar de la importancia de la calidad de un proyecto para alcanzar el éxito, existe un hueco importante en el estudio de los lazos que unen proyecto con el éxito, a pesar de la importancia de la multitud en el éxito resulta de gran interés el estudio de los factores de un proyecto que determinan o dirigen un proyecto al éxito.

El alcance de este análisis bibliométrico no radica en abrir un nuevo campo de estudio, sino que está enclavado en comprender a través de la literatura científica, como se relacionan los proyectos de crowdfunding con el creador, con los colaboradores y con su entorno, así como determinar cualitativamente en que medida se asemejan a los

proyectos tradicionales. Estas técnicas proporcionan un excelente marco para comprender como encaja el crowdfunding en el ecosistema de financiación existente.

3.1.7 TIPOS DE CROWDFUNDING

Actualmente el crowdfunding se encuentra estructurado atendiendo a criterios basados en la futura recompensa que puede obtener el financiador si la recaudación es exitosa, unos modelos son más populares que otros, debido en gran parte a la popularidad de los proyectos más exitosos y otros menos conocidos debido a la naturaleza que hay detrás del inversor.

Desde el punto de vista del creador, el beneficio económico no siempre es la principal motivación, también se reconocen aspectos como emprendedor social, productor independiente o soñador atrevido (Ryu and Kim, 2018).

Independientemente del tipo de crowdfunding elegido, el posible beneficio que puede obtener un inversor/financiador ha de ser lo más atractivo posible, pero cuidando no pasar por alto el fin principal, que no es otro que obtener la financiación necesaria para desarrollar su actividad. El impulsor del proyecto deberá elegir un ratio riesgo/beneficio lo suficientemente llamativo de cara a la multitud para que una persona se vea interesada y pueda plantearse la inversión o financiación.

Aunque muchos autores difieren en los tipos de crowdfunding existentes, la mayoría defienden cuatro tipos, cada uno de ellos basados en la retribución a obtener por el financiador: crowdfunding de donación, de préstamo, de recompensa y de inversión (Alonso et al., 2013), distribuidos como se observa en la Figura 5.

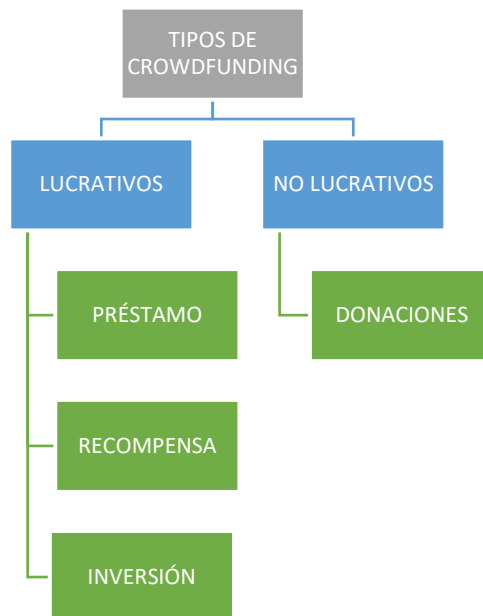


Figura 5: Tipos de crowdfunding

De los cuatro tipos, el de donación es el único no lucrativo, donde los mecenas aportan una cierta cantidad económica solamente por la satisfacción personal de participar o

ayudar al desarrollo de un proyecto y sin ambición de obtener una retribución económica o un producto consecuencia del proyecto. Por ello, resulta un medio muy conocido y eficaz para adquisición de recursos con fines sociales o benéficos, incluso algunas plataformas permiten hospedar proyectos que carecen de fotografías o vídeos (Xu, 2018). Por otro lado, este tipo de proyectos está influenciado por el periodo de campaña (Solomon et al., 2015) y algunos creadores ofrecen un regalo simbólico por haber participado y así aumentar la potencialidad de los proyectos (Lange et al., 2017), estos regalos normalmente corresponden a una visita *in situ* al lugar de realización del proyecto, un certificado de haber colaborado o simplemente están compuestos de un objeto, en cualquier caso, es imperativo que sean de poco valor para evitar pérdida de recursos y evitar pago de tasas administrativas (Metrejean and McKay, 2018).

El crowdfunding de préstamo, también conocido como *Crowdlending* o *Debt-based crowdfunding*, al igual que el resto de los tipos de crowdfunding está basado en recaudar fondos de una multitud de personas, pero en este caso el único fin es obtener una recompensa económica para la realización del proyecto (Paschen, 2017). En crowdlending el promotor del proyecto generalmente es una empresa aunque puede ser una persona o un grupo de personas las que pueden ofertar una necesidad de capital para llevar a cabo un proyecto, así los impulsores del proyecto podrán aportar una cierta cuantía económica que posterior a la realización del proyecto les será devuelta con incentivos. Se ha demostrado sus capacidades, sobre todo en la PYME donde posee claras ventajas respecto a otros modelos de financiación (Roig Hernando, 2016). En crowdlending se conoce como préstamos P2B (Peer to Business) los realizados de persona a empresa mientras que los llamados P2P (Peer to Peer) son los realizados de persona a persona (Leimeister, 2012).

Se ha de tener en cuenta que el crowdlending es una actividad regulada por *Comisión Nacional del Mercado de Valores* en España (CNMV, 2020) y por la JOBS Act en Estados Unidos (U.S. Securities And Exchange Commission, 2016).

El crowdfunding de inversión, denominado comúnmente Equity Crowdfunding, responde a un promotor o empresario que quiere lanzar un producto, servicio o idea nueva en la que además está dispuesto a aumentar parte del capital mediante pequeñas aportaciones de la multitud, cediendo a cambio parte de los derechos sobre la empresa, aunque también es posible que el inversor busque algún privilegio sobre el producto fabricado (Vulkan et al., 2016).

El Equity Crowdfunding es un modelo muy cercano al Capital Riesgo (CR), aunque mucho menos restrictivo ya que el inversor no tiene porque ser un experto en la materia, por eso puede ser la opción ideal para aquellas empresas, que aún no han dado suficientes beneficios en fase inicial y necesitan hacer una inversión como el caso de muchas Start-ups, también para aquellos que no han podido acceder al CR porque la empresa está en fase demasiado temprana o llevan a cabo proyectos con alto nivel de riesgo.

Para minimizar el riesgo en la medida de lo posible, las plataformas de Equity Crowdfunding cuentan con un sistema de ponderación de empresas que facilita a las personas interesadas tanto su selección como la cantidad a invertir (Hornuf and Schwienbacher, 2018). Al igual que el crowdfunding de préstamo, también es una actividad regulada por la *Comisión Nacional del Mercado de Valores en España* (CNMV, 2020).

El crowdfunding de recompensa o *Reward-based Crowdfunding* se caracteriza por el premio o recompensa (no financiera) que reciben los usuarios al contribuir en la financiación de un proyecto específico. Estas aportaciones son generalmente de valor variable y reducido, en muchos casos es considerado como una preventa (Crosetto and Regner, 2014) a un precio más atractivo que el posterior en el mercado (Mollick, 2014). De los cuatro tipos de crowdfunding existentes, el de *recompensa* es el más conocido debido a la repercusión social que ha adquirido a través de sus proyectos más ambiciosos, también es el más interesante por su gran capacidad para dar salida a proyectos con un factor de riesgo muy alto o muy baja credibilidad. por ello este trabajo se centra en este tipo de financiación.

Atendiendo a la naturaleza del proyecto y a diferencia de los tipos anteriores, las recompensas ofertadas en este tipo de crowdfunding pueden ser de dos tipos, ninguna de ellas del tipo financiero. Por un lado las recompensas del tipo tangible, donde la aportación se corresponde con el valor de un producto físico, tales como los dispositivos electrónicos y por otro lado las recompensas intangibles donde el producto físico adquirido no se corresponde con el valor de su fabricación como puede ser una entrada para un espectáculo (Messeni Petruzzelli et al., 2018).

Para atraer rápidamente a futuros inversores es frecuente el uso de los llamados “early birds”, recompensas limitadas por una cantidad claramente inferior al resto, que ayudan a proporcionar sensación de prisa al comienzo de la campaña de recaudación (Wessel et al., 2019).

3.1.8 ALTERNATIVAS AL CROWDFUNDING

A pesar de que muchos proyectos novedosos nacen directamente para ser financiados mediante crowdfunding, antes de optar por este tipo de financiación es necesario estudiar que alternativa es la más adecuada para cada fase del proyecto, de no ser así existe una elevada probabilidad de quiebra por la mala gestión de recursos financieros (Bustos Contell, 2013) tanto por exceso como por defecto.

En etapas iniciales del proyecto cuando la incertidumbre es muy alta, es lógico pensar que las inyecciones de capital a fondo perdido proporcionan una gran ayuda sin generación de deuda, lo que dota al proyecto de una cierta escalabilidad financiera permitiendo un mejor control de los recursos. En la Tabla 24 adjunta en el Anexo se observan comparadas, las principales características de los métodos más comunes de financiación de proyectos. Se incluyen desde los que permiten adquirir capital con una

deuda mínima, hasta los que necesitan que el proyecto o empresa genere un gran volumen de capital para que se concedan. También permite observar la posición que ocupan los diferentes métodos de crowdfunding respecto de la financiación tradicional.

Bootstrapping y Family, Friends & Fools (FFF)

Bootstrapping y Family, Friends and Fools son dos modelos de obtención de recursos (económicos y/o materiales) propio de etapas pre-proyecto o proyecto sin desarrollar, donde la idea aún está en fase de desarrollo, por lo que también se suele denominar *capital de arranque* o de *puesta en marcha*.

Los fondos más cercanos e inmediatos de los que se puede disponer son los propios, así como lugar y útiles de trabajo, por eso la técnica del *Bootstrapping* consiste en comenzar el proyecto usando los propios para capitalizar nuestro negocio. De esta manera se puede iniciar un negocio sin necesidad de acudir a una entidad para la concesión de un crédito o similar, por otro lado y de manera didáctica también aportará experiencia sobre la gestión e importancia de la asignación de recursos, ayudando a superar problemas de valoración de los mismos (Kotha and George, 2012b).

Dentro de los recursos materiales, los de utilización más común son: herramientas, vehículos o maquinaria y local o espacio de trabajo, dentro de los recursos humanos entra la mano de obra, tanto de ayuda como de formación, hasta adquirir el know-how suficiente que permita el desarrollo controlado de la actividad, en muchos casos a través de habilidades empresariales adquiridas en la universidad o bien con las habilidades personales (Gieure et al., 2020). La formación del promotor del proyecto se debe de considerar como un recurso, ya que cobra gran importancia a la hora de adquirir otros nuevos a través de fondos públicos o bien a través del establecimiento de alianzas con institutos de ciencia y tecnología (de Faria et al., 2019).

Para comenzar un proyecto con estas técnicas de autofinanciación, se necesita cambiar el punto de vista y reorientar el enfoque del proyecto. En lugar de preparar un proyecto que cubra los aspectos fundamentales en aras de que una entidad compruebe su viabilidad y lo financie, el promotor de la idea debe de orientarlo de cara al mercado, de manera que su producto resulte atractivo para una multitud. Realizar un producto o servicio de gusto por el cliente puede revertir en una retribución económica que permita la perfecta actividad empresarial (Kok and Biemans, 2009). Pasado este punto, el proyecto se puede autofinanciar reiteradamente, aunque eso si, con un crecimiento generalmente muy lento.

Mediante el uso de FFF, la mayoría de los colaboradores esperan retribución, económica o de otra naturaleza, en caso de ser bienes materiales no tienen porque ser de la misma cuantía que los bienes prestados, pero generalmente en estos casos no importa el plazo de devolución de esos intereses generados (muypymes.com, 2015).

La mayoría de los especialistas aconsejan que se reinviertan los beneficios constantemente en las primeras etapas, que se hagan las suficientes campañas de

marketing mediante redes sociales. Cuando los participantes de este modelo de financiación aportan sus propios recursos al proyecto, en muchas ocasiones tienen posibilidad de convertirse en socios de la nueva empresa, este modo de financiación está motivado por la relación de confianza que se establece con los emprendedores. En la mayoría de los casos tienen por objeto no perder la inversión realizada y no la de obtener grandes plusvalías.

Ayudas y subvenciones (AS)

Son aportaciones a fondo perdido de las administraciones que están sujetas a la disponibilidad presupuestaria y están dirigidas a emprendedores y empresas para fomentar el cumplimiento de algún objetivo determinado. Tienen como objetivo el atraer nuevas ayudas económicas una vez difundido en los medios apropiados (Bianchi et al., 2019). Por ello, en muchos casos se utiliza la prestación por desempleo para este fin, ya que es posible recibir el abono por desempleo en un único pago (SEPE, 2020).

Son, sin ninguna duda la vía de financiación más atractiva para los emprendedores, ya que por su naturaleza, no llevan asociado ningún coste financiero y no tienen que ser devueltos ni a corto ni a largo plazo (Caloffi et al., 2018). Por otro lado, en ocasiones su cobro se produce con un retraso considerable, lo que puede conllevar graves problemas si hay una previsión disponer de ese dinero en una determinada fecha.

Lo más común es acudir a entidades destinadas a fomentar la implantación y desarrollo de nuevos proyectos empresariales generalmente de ámbito local que a parte de conceder la ayuda económica en muchos casos también ofrecen mentoría (CDTI, 2020) (CEEI, 2020)

Fondos de Capital semilla (CS)

Capital Semilla (CS) responde a una cantidad económica generalmente de carácter muy bajo, en las etapas tempranas (*early stage*) de una idea o empresa. La inmensa mayoría que se beneficia de este tipo de financiación son proyectos de carácter innovador generados por Start-ups, Spin-offs o particulares, en cualquier caso empresas de mucho potencial que no pueden generar flujo de caja inicial.

La decisión de invertir no se basa tanto en unas altas probabilidades de generar flujos de caja si no en la confianza en la idea, en el proyecto y en el equipo que lo llevará a cabo. Los proyectos altamente innovadores se beneficiarán de la flexibilidad que aporta en etapas iniciales (Martín-Barrera et al., 2017), por ello es frecuente que las entidades que los conceden no obtengan beneficios si el proyecto o empresa tampoco lo hace (Xstartups, 2020), en previsión de mejora.

Por lo general, este capital inicial se utiliza para desarrollar una idea de negocio hasta el punto de poder presentarla eficazmente a las empresas de capital de riesgo que tienen grandes cantidades de dinero para invertir.

Concurso de emprendimiento (CE)

Están promovidos normalmente por asociaciones de empresarios, bancos, escuelas de negocios o cualquier entidad destinada a promover la generación de nuevas ideas para premiarlas con algún tipo de recompensa, económica, mediática o de cualquier otra naturaleza.

Plataformas como www.impulsame.es incentivan a los ganadores de los concursos con premios como: inversión en capital, asesoramiento y mentorización, alojamiento empresarial o un servicio gratuito de gestiones administrativas, fiscales y contables (Impúlsame, 2020).

En este caso particular de financiación las ideas o proyectos se exponen como *Pitching*, consiste en una breve presentación de un proyecto de un determinado tipo, frente a unos inversores o patrocinadores que seleccionan aquellos proyectos más atractivos o interesantes para financiarlos, patrocinarlos u ofrecer algún tipo de beneficio.

Créditos Rápidos Online (CRO)

Promovido por entidades financieras y muy popular hace una década, tanto por la facilidad de obtención como por la rapidez, hoy en día y debido al periodo de crisis económica que se vive cuenta con menos usuarios, debido a las comisiones e intereses que generan, en muchos casos superan el 25% TAE.

Puede ser una alternativa viable para hacer frente a imprevistos del proyecto o empresa siempre que se cumplan unos requisitos mínimos, a pesar de su gran coste, una de sus principales ventajas radica en gran velocidad de concesión, solamente la tarjeta de crédito es más rápida. Otras de sus ventajas radica en la forma de solicitud, donde a parte de contar con un simulador de crédito directo, que permite al usuario elegir tiempo y cantidad (COFIDIS, 2020), no es necesario acercarse a una entidad bancaria, ni hay que dar explicaciones del uso al cual irá destinado, incluso muchas plataformas permiten la concesión de este tipo de créditos en días festivos y periodo vacacional.

Tarjeta de crédito (TC)

Las Tarjetas de Crédito son herramientas financieras muy útiles, con las que se puede obtener una cierta cantidad económica con la que financiar proyectos personales y empresariales de manera rápida, de hecho, es el mecanismo de obtención de crédito más rápido que existe. Esta inmediatez tiene una parte negativa que hace que una entidad corra un mayor riesgo y por lo tanto se resume en un mayor coste financiero o tasa de interés para el que la utiliza, por lo que se deben asignar correctamente los límites de la tarjeta (Sohn et al., 2014).

A parte de la facilidad para cubrir imprevistos, también puede utilizarse de manera muy eficiente en casos de oportunidades de negocio inmediatas, por ejemplo, para conseguir financiación rápida para participar en una exposición o feria, para reinvertir en un proyecto o negocio de nueva creación. En la actualidad es muy usada en proyectos o

negocios de bajo capital, sustituyendo en muchos casos a los préstamos debido a las bonificaciones de uso que posee (BBVA, 2020).

Préstamos P2P (P2P)

Los préstamos P2P (*peer-to-peer*) o persona a persona, tienen lugar a través de una plataforma Web, de modo que un inversor sin más intermediarios puede contribuir a financiar una compra de cualquier naturaleza o una inversión a un particular, a cambio de un interés en su devolución. La plataforma en muchos casos y dependiendo de la cantidad económica solicitada, se encarga de hacer un estudio a través de una serie de agencias para evaluar entre otras cosas, la morosidad o la demora en pagos anteriores, para establecer una mínima garantía de devolución (Suryono et al., 2019).

Las Webs destinadas a este modelo de financiación no trabajan con plazos e intereses fijos para cualquier inversión, sino que estos suelen variar función del riesgo del solicitante y la cuantía económica solicitada.

A pesar de que este medio de financiación va dirigido a particulares en aras de desarrollar un proyecto personal y no a empresas, como es el caso del crowdlending, ambos están sujetos a la Ley 5/2015, de 27 de abril, de Fomento de la Financiación Empresarial (Jefatura del Estado, 2015),

Microcréditos (MC)

Son instrumentos financieros similares a los préstamos, con la particularidad de que están especialmente diseñados para satisfacer las necesidades de personas o colectivos sin recursos con el objetivo de que puedan emprender actividades por cuenta propia que generen ingresos. Surgieron dentro de las organizaciones no gubernamentales para el desarrollo de poblaciones y colectivos desfavorecidos, de ahí que su popularidad varíe en función de la geografía (Islam and O’Gorman, 2019).

El uso de microcréditos en pro de la empresa es más reciente, pero ha crecido de forma notable en el último decenio, ya que generalmente los intereses son más bajos que el resto de los créditos ofertados por la entidad. Permiten impulsar el desarrollo local a través del fomento de la actividad económica susceptible de comercialización (Valentin Mballa, 2017).

Por norma general, aunque no requieren garantías propiamente dichas, la concesión de microcrédito está supeditada a que una fundación o institución nos avale o garantice la viabilidad de nuestra empresa, por lo que es necesario presentar un plan de empresa con las características del proyecto y su viabilidad.

Business Angels (BA)

Se denominan también *inversores informales*, son inversores individuales que se encargan de financiar empresas en fase inicial, su principal trabajo suele ser el de asesoramiento, ya que aportan una valiosa experiencia personal para los emprendedores, con el fin de incrementar las posibilidades de éxito y obtener una

remuneración a medio plazo. Es común que estos renuncien a la recuperación periódica de capital a cambio de poder obtener una importante plusvalía con la venta de sus participaciones, de ahí el interés de estos inversores por alcanzar proyectos que estén muy apoyados por el creador (Mittiness et al., 2012).

Actualmente suele clasificarse en cuatro grupos diferentes: Empresario, Financiero, Trabajador o Asesor o consultor, dependiendo del rol que adopte el Business Angel, aunque también puede ocurrir el caso de que se mezclen características de uno u otro grupo. Esta asociación de trabajador-empresa conlleva que el inversor tenga su residencia cerca de la zona de actuación, con el fin de establecer un control directo con la empresa, generalmente no suelen estar a más de 100 kilómetros.

Innovación abierta (IA)

Está enmarcada dentro de los proyectos de I+D y proviene directamente del *crowdsourcing*, esta moderna herramienta de financiación de proyectos permite exteriorizar la investigación y el desarrollo de una empresa a través de una convocatoria donde se buscan las mejores ideas. Su desarrollo es posible gracias a unas plataformas Web en las que empresas, con un espacio limitado, exponen los problemas a resolver. Las personas que acceden a la plataforma pueden hacerlo creando un perfil en modo "Solver" (solucionador) y elegir si lo desean el proyecto más cercano a sus expectativas (Innocentive, 2020).

Es una herramienta habitual y muy útil en proyectos relacionados con el sector TIC debido a los sucesivos cambios que se producen en este tipo de proyectos, también es usada en las empresas para reducir las inversiones directas en I+D, aunque muchas otras la rechazan por ser una pérdida de control sobre esta parte del conocimiento de la empresa. A pesar de que este proceso basa su éxito en la transferencia de conocimientos y relaciones (Bacon et al., 2019). Por ese motivo es muy importante la protección de su propiedad intelectual mediante algún tipo de licencia.

Patrocinador (P)

Esta forma de financiación también conocida como *Financiación con Proveedores* no es válida para cualquier empresa o proyecto, por su particularidad de estar orientada a la comercialización de productos, ya que el proyecto o empresa debe de estar preparado para dar fruto a corto plazo para financiarse con los productos de los proveedores.

Las ventajas para el proveedor provienen de satisfacer mayor demanda de su producto y entrar en nuevos mercados a través de las últimas ideas, por el contrario hay un cierto riesgo debido a que las ideas demasiado novedosas tienen menos probabilidad de ser aplicadas (Homfeldt et al., 2019).

Es muy importante establecer un buen modelo de negocio para lograr trabajar con el capital de los proveedores, así cómo mantener las relaciones, ya que, en muchos casos resulta difícil la interacción entre las dos partes implicadas, proveedor y creador (Kim

and Zhu, 2018). Una vez generada la cohesión y durante el ciclo de vida de este modelo de financiación, ambas se nutrirán de ideas de la otra aportando nuevas posibilidades de negocio.

Intercambio de servicios (IS)

Comúnmente conocido como *trueque* o *Intercambio B2B (Business-to-Business)*, fue muy popular hace algunas décadas sobre todo entre personas con pocos recursos económicos, aunque actualmente pasa a ser un modelo de obtención de recursos para emprender un proyecto o una actividad empresarial.

El intercambio de servicios crea un vínculo entre el minorista y la empresa donante que lleva a la más pequeña a pedir mayor cantidad de producto. Por otro lado, son más ventajosos los modelos de intercambio en los que no existe comisión o esta es muy baja (Hua et al., 2020).

Una de las características más remarcadas es que siempre se lleva a cabo sin recurrir a dinero en metálico, solamente se realiza intercambio de bienes y servicios.

Aunque no es necesario realizarlo mediante una plataforma Web, simplemente a través de la comunicación verbal es posible, hay plataformas especializadas en este tipo de transacción conocidas como plataformas de intercambio *B2B* que facilitan mucho la tarea de intercambio por hacer de nexo entre los usuarios, hay varias plataformas destinadas a este fin, aunque muchas con metodologías y funcionamiento totalmente diferentes se basan en el mismo intercambio de bienes y servicios. Dos ejemplos representativos son: <http://www.trocobuy.com> (basada en disposición y devolución) y otra como <http://www.bancodeltiempo.sanjavier.es> (basada en intercambios de tiempo).

El caso particular de la Web Trocobuy.com presenta un modelo basado en la obtención de un crédito comercial, de cantidades que oscilan entre 3.000€ y 300.000€ de dinero no efectivo, para comprar y contratar productos y/o servicios solamente a través de la plataforma, con el único coste de un interés (3,75%), esta vez con dinero “real” por cada transacción que realicemos. Por lo que, con los recursos obtenidos ya se puede empezar la actividad empresarial.

Préstamo participativo (PP)

Aplicado solamente a PYMES de cualquier sector excepto el inmobiliario y financiero, que estén en una situación saneada y equilibrada, constituyen una modalidad de préstamo solamente reconocido como actividad financiera en España y Francia, que por sus características y su regulación legal se convierte en una vía de financiación híbrida entre recursos propios y los recursos ajenos. Dicho de otra manera, se consideran como patrimonio neto a la hora de determinar la solvencia de la empresa, además de ser deducible, por lo que puede retrasar el reconocimiento de quiebra por endeudamiento, si este fuese el caso.

Se caracteriza porque la entidad prestamista puede beneficiarse de la empresa financiada, además de cobrar una cantidad a modo de interés, también un vencimiento a largo plazo y el interés que percibe el prestamista va en función de la actividad de la empresa, el pago de los intereses está subordinado a otro tipo de créditos, es decir, la prioridad se establece en hacer el pago a otro tipo de créditos sino es posible hacer este y se considera patrimonio neto de la empresa. Por ello de cara al prestamista es entendido como un recurso propio de la empresa, hecho por el cual se suelen considerar como socios del proyecto y de cara al empresario como una deuda (Bustos Contell, 2013).

A pesar de que muchos incluyen hasta la posibilidad de carencia en el pago, por estar considerado como *deuda subordinada* se exige que las amortizaciones de deuda por parte de la empresa a la entidad que emite el préstamo participativo, se compensen en igual medida que el incremento de los fondos propios (Bravo Nufrio, 2017).

Patrocinadores (P)

Son varios los motivos que pueden llevar a una empresa a prestar ayuda a otra a cambio de algún tipo de beneficio. Entendido por muchas empresas como un instrumento más, perteneciente al campo de las acciones para promover la Responsabilidad Social Empresarial (RSE). Se establece una relación entre dos empresas o empresa y organización, para dar ayuda o llevar a cabo un proyecto, sin que ambas actividades estén ligadas.

Se entiende por financiación con un patrocinador, a la prestación económica o material que se realiza a una empresa para el desarrollo de un proyecto, este no tiene por que coincidir de manera estricta con su actividad principal, a cambio de obtener algún tipo de rendimiento (Ministerio de Educación, 2020).

Los tipos de proyectos que abarca este tipo de financiación están enmarcados en los ámbitos de las actividades que movilizan al público y generan audiencia en los medios de información. Por ello es muy común que estos proyectos suelen estar enmarcados en el campo de la imagen y la comunicación. Aunque hoy en día no es raro encontrarlos en el ámbito de aplicación social y cultural abarca deporte, patrimonio arquitectónico, enseñanza, salud, medioambiente, actuaciones solidarias, medios de comunicación y grandes eventos como juegos olímpicos, congresos, etc.

Son muy comunes las inclusiones de patrocinio en empresas en grandes eventos deportivos que mueven a mucha masa de gente, al igual que eventos relacionados con la cultura. Por eso marcas muy conocidas como la de bebida RedBull pueden patrocinar empresas por una cantidad muy elevada de bienes económicos y materiales. El 14 de octubre de 2012, el paracaidista y saltador BASE Felix Baumgartner realiza un salto al vacío desde la estratosfera en un globo diseñado a tal efecto gracias a RedBull gracias a el patrocinio de este proyecto con un coste de 30 millones de dólares (“DreaMatch Solutions,” 2013).

Las Sociedades de Garantía Recíproca (SGRs)

Cada vez menos populares debido a la consecuente dependencia de las entidades bancarias, Estas SGR sociedades, no se conciben como un medio de financiación en si, sino funcionando a modo de aval o garantía como un medio o soporte para facilitar el acceso o mejorar las condiciones de un crédito de una pyme en una entidad financiera. De capital variable y sin ánimo de lucro están sujetas a la supervisión e inspección del Banco de España.

Estas sociedades (actualmente 18 en España) funcionan a modo de aval o garantía y beneficia tanto a PYMES por facilitar la concesión de crédito como a entidades financieras que poseen la garantía necesaria para dar crédito, respondiendo las SGR por la empresa si fuese necesario. A su vez las PYMES también se ven beneficiadas con el valor añadido de ser estudiadas y seguidas para ser poseedoras de este tipo de apoyo, estableciéndose un compromiso de adquisición de incluso una o más cuotas sociales de la SGR, ayudando a esta e incrementando sus recursos para poder seguir realizando su actividad (CESGAR, 2015)

Family offices (FO)

Las primeras señales de *Family Office*, se dieron en Suiza y en EEUU, cuando el crecimiento industrial derivó en un aumento del poder adquisitivo de algunos sectores hasta tal punto que fue necesario contratar a personas para gestionar ese patrimonio.

Hoy en día más especializado el sector, existen gestores o empresas dedicadas al tratamiento del patrimonio de familias no “tan adineradas”, los conocidos como *Multi Family Office* (MFO) que, con un patrimonio, no tan elevado como los anteriores, buscan conservar e incrementan este patrimonio a base de invertir en: activos financieros, inmobiliarios y operaciones de actividad económica. La mejor forma para conseguir financiación de un FO es a través de una Web destinada a este servicio que funciona a modo de clúster (Cluster Family Office, 2016).

Su acceso no es muy difícil si se tiene un buen proyecto, porque estos gestores de patrimonio están familiarizados con inversiones en start-ups y en empresas en diferentes fases de desarrollo y se mueven en canales de inversión, como las citadas redes de *Business Angels* de escuelas de negocio y de instituciones públicas privadas, o los fondos de inversión de *Capital Riesgo* entre otros.

Capital riesgo (CR)

El capital riesgo (CR) es un instrumento financiero, sujeto a la supervisión de la CNMV considerados como *capital neto*, a diferencia del préstamo participativo que es considerado como un *pasivo* y por tanto no es deducible.

Están diseñados para estructurar la toma de participación por parte de administraciones, es decir, los socios realizan una inyección de capital en la empresa con la única finalidad de obtener una plusvalía vendiendo sus acciones cuando realice su

desinversión. Es por ello que el porcentaje de participación de una empresa con *capital riesgo* no supera el 50%, normalmente oscila entre el 30 y el 45% aunque este porcentaje es suficiente para que pueda participar activamente en la toma de decisiones estratégicas de la empresa desde el consejo de administración.

Promovido por entidades financieras o corporaciones industriales en nuevos proyectos empresariales o en empresas que deben afrontar un proceso de crecimiento o reestructuración, están destinados a empresas o proyectos que puedan obtener beneficios relevantes de medio a largo plazo, ya que estos socios renuncian a la recuperación periódica de su capital y al cobro de un interés a cambio de poder obtener una importante plusvalía con la venta de sus participaciones, con un horizonte temporal, generalmente predeterminado entre cuatro y seis años.

Normalmente, aunque la normativa no lo contemple, los profesionales del CR separan en dos etapas la financiación, denominando *venture capital (VC)*, cuando la inversión va dirigida a empresas inmaduras o en etapas tempranas y *private equity* cuando se realiza en empresas consolidadas (Bustos Contell, 2013).

Comúnmente el *capital riesgo* genera unos gastos de estudio, denominados *due diligence*, consistentes en una revisión exhaustiva de la empresa por parte de un agente y una posterior estructuración si fuese necesario por parte del inversor.

Más allá de la financiación y de cara al prestatario, el capital riesgo también es considerado como un facilitador del para dar garantía o aval a otros medios de financiación.

Mercado Alternativo Bursátil (MAB)

Este conocido modelo de financiación es aplicable solamente cuando una empresa cuenta con un determinado capital mínimo (1.200.000 €) y es necesario que ningún accionista posea más del 25% del capital social, se corresponde con la última fase de crecimiento fuerte y sostenido de una empresa.

Una de las principales ventajas de entrar en el *Mercado Alternativo Bursátil (MAB)* es la credibilidad de un proyecto ante futuros inversores, con lo que existe la posibilidad de ampliar capital con desembolso, posibilidad de desinvertir posteriormente si se opta por una oferta pública de venta.

La salida a el mercado de valores a pesar de estar bajo la supervisión de la Comisión Nacional del Mercado de Valores (CNMV), también porta una serie de inconvenientes como la normativa regulatoria, la transparencia de datos y las prácticas de buen gobierno corporativo que hay que cumplir (CNMV, 2020).

La cuantía económica alcanzable da una idea del tamaño de la financiación, ya que son del orden de 2 a 25 millones de euros, siendo lo común moverse en financiaciones de 12 a 15 millones de euros. Para ello, las empresas deben estar en estadios de generar al

menos de 2 a 3 millones de euros de EBITDA y ventas del orden de 8 millones de euros (Bolsa de Madrid, 2020).

3.1.9 OTRAS FORMAS DE FINANCIACIÓN

Existen algunas herramientas que, sin llegar a ser un método de financiación en sí, facilitan el acceso sobre todo a proyectos o empresas de riesgo bajo o con una cierta viabilidad. Pueden llegar a resultar muy útiles ya que ayudan a funcionar en conjunto con la financiación principal.

Para fomentar una marca comercial o un producto, es muy común intentar acceder a un emplazamiento publicitario tipo *Branded Content* o *Product Placement*, técnicas que llevan usándose de forma muy efectiva desde los años 80 y que consisten en incluir de forma implícita o explícita aquello que queremos promocionar a través de un medio audiovisual de cierta relevancia o bien con la ayuda de un personaje público que tiene mayor capacidad para atraer atención de posibles inversores (Pino Romero and Castelló Martínez, 2015).

Mediante el *Leasing* un emprendedor puede adquirir un recurso material en base a un contrato de arrendamiento a largo plazo con opción a compra asumiendo los gastos derivados de mantenimiento o uso.

Por otro lado, las técnicas de *Factoring* permiten adelantar el cobro de una factura a una empresa, en base a ceder las facturas generadas por sus ventas a una compañía externa, generalmente un banco, que gestionará su cobro aplicando una comisión. En caso opuesto se encuentran las técnicas de *Confirming* consistentes en un servicio de pago a proveedores (GEDESCO, 2020).

3.1.10 CARACTERÍSTICAS BÁSICAS DE LOS SISTEMAS

En la Tabla 24 se muestran distintos tipos de financiación de proyectos, unos más conocidos y tradicionales como el *capital riesgo* o el *mercado alternativo bursátil* y otros más modernos como el *crowdfunding*.

Una situación crítica y relativamente común en las empresas en fase de crecimiento, se produce cuando, incluso siendo una empresa rentable, esta decide dar un paso demasiado largo de cara a su crecimiento, valga como ejemplo el paso de pequeña a mediana empresa, donde se ha de realizar una gran inversión que requiera gran cantidad de fondos (Furr, 2011). A pesar de que las empresas más comedidas optan por una financiación de pasos más cortos, con el fin de evitar la tentación de tener un exceso de fondos con la consiguiente dificultad de racionamiento, en muchos casos no es así, conduciendo esta gestión inadecuada de los paquetes financieros a una alta probabilidad de quiebra, ya que, con esta dificultad de gestión añadida, es muy probable que no permita obtener unos plazos razonables de amortización económica y financiera. Por ello previniendo esta situación conocida como *muerte de éxito* es muy coherente

que mucho opten por autofinanciar la deuda, usar el capital riesgo y el préstamo participativo durante todo el proyecto de empresa (Bustos Contell, 2013).

Existen múltiples mecanismos que se utilizan para financiar proyectos. Muchos de ellos son a través de dinero público como las entidades bancarias y otros mediante asociaciones privadas. A lo largo de este apartado se presentan los tipos de financiación que actualmente se utilizan para financiar proyectos que comparten una naturaleza similar por ser de carácter novedoso o innovador y en los que la financiación debido a su posible falta de credibilidad es difícil por los medios tradicionales conocidos.

En la Figura 6 se observa la trazabilidad ideal de crecimiento de una empresa que comienza con el modelo de una Start-up. Las necesidades más comunes de financiación para llevar a cabo un proyecto siempre van en aumento, a medida que la empresa crece, van variando los vehículos de financiación más adecuados a su velocidad y riesgo empresarial (Muñoz, 2011).

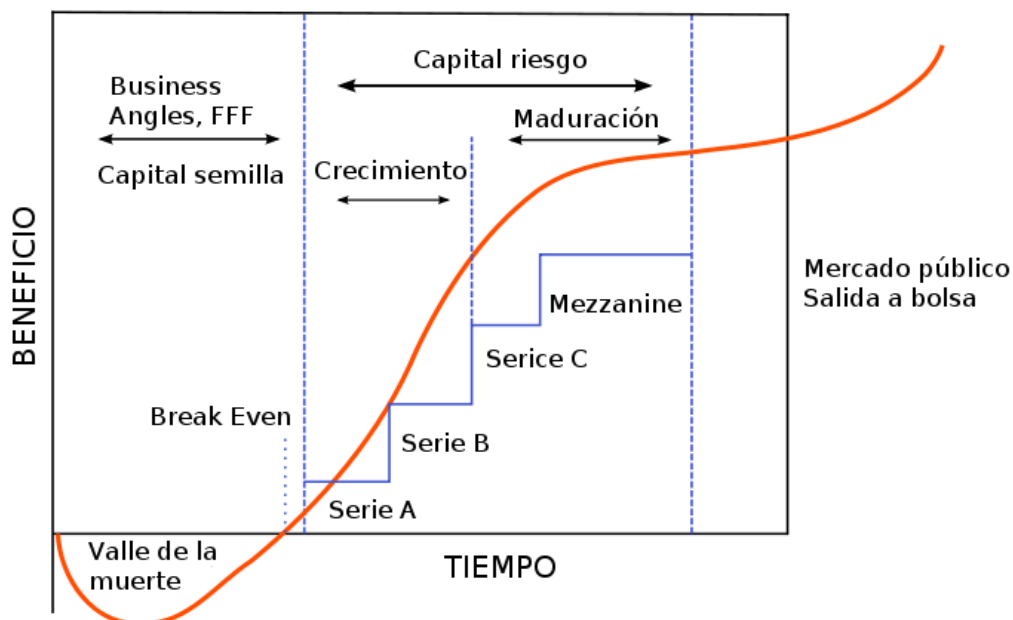


Figura 6: Relación tiempo-beneficio en los mecanismos de financiación de proyectos (Muñoz, 2011)

Las empresas emergentes y altamente innovadoras suelen tener una trayectoria muy marcada, como en el caso de las Start-ups, donde los periodos de financiación pasan por unos puntos similares en casi todos los casos y las necesidades de financiación en cada fase de la empresa son similares, teniendo en cuenta el crecimiento de los beneficios y las necesidades de inversión en función del tiempo.

Para dar comienzo a un proyecto o empresa, lo más común es iniciar la actividad con fondos propios (FFF), como capital de inicio, también llamado: *Capital de Arranque* o *Capital Semilla* (Andrieu and Peter Groh, 2020). Durante el primer año de vida de la empresa, por norma general no se suele generar ningún beneficio, ya que se están invirtiendo todos los recursos en el desarrollo del proyecto, este intervalo temporal se

conoce como *Valle de la muerte*. Una vez se atraviesa esta zona de negativa, si la idea es buena y el proyecto tiene perspectiva de salir con beneficio, resultará más fácil conseguir pasar a la siguiente fase para conseguir financiación apoyándose en una financiación del tipo Business Angel si se dan las condiciones necesarias (Cipollone and Giordani, 2019), la experiencia e involucración que aportan estos últimos pueden resultar de gran ayuda para encaminar la empresa hacia el comienzo de los beneficios.

Pasado el punto en el que la empresa sale de la zona de depresión, denominado como “Break Even”, la empresa comienza a generar beneficios y crece en necesidades y recursos. En esta fase es donde los Business Angels empiezan a recuperar su inversión y sería necesaria una nueva inversión denominada normalmente Venture Capital o Capital Riesgo (Andrieu and Peter Groh, 2020). En Estados Unidos es común categorizar la evolución de una empresa para definir diferentes series de financiación, así, dentro de este periodo se puede encontrar las series, A, B, C, ...etc.

Después de la inyección financiera del Capital Riesgo, si la empresa sigue con buena tasa de crecimiento estaría preparada para dar el salto al último escalón de crecimiento que suele ser la salida a bolsa (Muñoz, 2011).

En la Figura 7 se observa el lugar que ocupa el crowdfunding, teniendo en cuenta la cantidad invertida por cada contribuyente según el tipo de financiación, así los inversores en este tipo de proyectos o productos lo hacen por término medio en la franja entre los 20 y los 100€ (X.net and Caparrós, 2014).

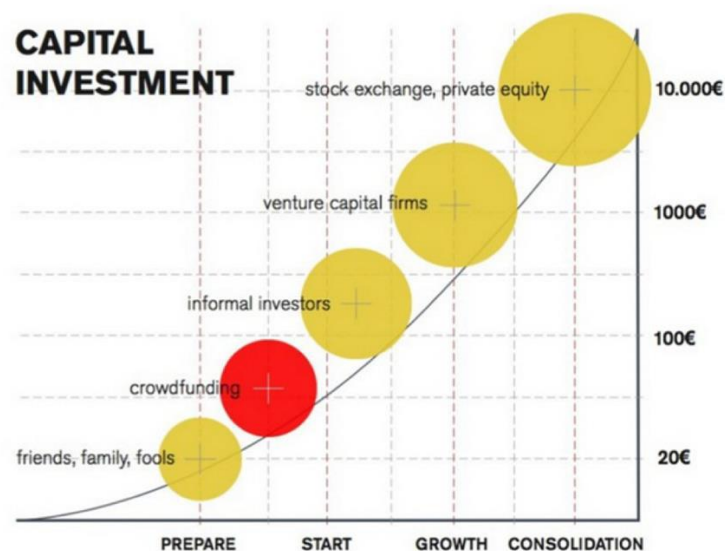


Figura 7: Inversión individual en cada fase de la empresa (X.net and Caparrós, 2014)

4 PERSPECTIVAS DEL PROBLEMA

La dificultad para desarrollar proyectos de alto contenido en innovación se hace más evidente a medida que se endurecen las condiciones del entorno. Sin embargo, aumenta la necesidad de búsqueda de soluciones que permitan adaptarse a las nuevas condiciones.

El crowdfunding de recompensa permite desarrollar proyectos altamente innovadores en entornos no propicios para otros tipos de financiación. Esta capacidad motiva que este estudio se concentre en el crowdfunding de recompensa.

Para dirigir al éxito a un proyecto de crowdfunding de este tipo, es necesario realizar observaciones desde los diferentes caminos posibles, ya que, a medida que se desarrolla la campaña de financiación puede resultar tan interesante la búsqueda del éxito, el alejarse del fracaso o simplemente identificar si la financiación va por buen camino.

Tanto para el creador del proyecto como para el colaborador es importante conocer y entender el significado del éxito y fracaso en una campaña de este tipo. Por ello, en este apartado se caracteriza el significado del éxito y del fracaso desde el punto de vista del creador y colaborador, también los mecanismos que permiten cierto grado de predicción del éxito y seguimiento de la campaña de financiación.

4.1 CARACTERIZACIÓN DEL ÉXITO

A pesar de que para un proyectista, el éxito de un proyecto se mide con el control del alcance, tiempo, costo, calidad, riesgo y recursos (PMBOK, 2013), de cara al impulsor de un proyecto, la percepción del éxito puede no tener dependencia de estos factores, ya que la motivación de adquirir un producto novedoso compensa en muchos casos el hecho de no cumplir los plazos de entrega, relativamente frecuentes en este tipo de financiación (Tran et al., 2017).

Cuando un creador pretende buscar financiación mediante crowdfunding de recompensa, tiene libertad de elegir el valor de algunas variables, tales como el mes de publicación, valor de cada recompensa o número de ellas, sin embargo, otras de vital importancia como la cantidad mínima para comenzar el proyecto (Goal), no es posible modificación una vez iniciada la financiación (Kickstarter, 2019). La determinación del presupuesto del proyecto o Goal, ha de realizarse en la *etapa de planificación*, dentro de la *gestión de costos del proyecto* (PMBOK, 2013) y a pesar de contener los costes totales del proyecto, su valor debe ser lo más ajustado posible, ya que condiciona de manera significativa la cercanía del éxito (Li et al., 2018) (Cordova et al., 2015), para reducir este valor, es común aprovechar todos los recursos propios disponibles usando técnicas de *Bootstrapping* (Lahm, 2005).

Para un creador, la importancia de la variable Goal radica en la facilidad de alcance de la cifra propuesta. Si se alcanza, la financiación se considera exitosa y el beneficio económico alcanzado será transmitido al creador del proyecto en aras de llevarse a

cabo, en caso contrario, la financiación se considera fracasada y no se transmitirá ninguna cantidad, este tipo de funcionamiento corresponde a plataformas de crowdfunding de recompensa basadas en el modelo “All-Or-Nothing” (AON). Este método de pago al creador, es comúnmente utilizado en este tipo de proyectos novedosos, ya que permite aumentar la confianza del impulsor del proyecto al saber que si no se alcanza suficiente capital su dinero no será retirado (Cumming et al., 2015).

Es posible que se supere el Goal antes de que finalice el tiempo de recaudación, en este caso la campaña de recaudación del proyecto puede alcanzar valores mayores de los esperados, ejemplo de ello es el proyecto *Pebble* (Kickstarter, 2012) en el que la cantidad máxima alcanzada (Pledged) superó en 102 veces al Goal. El número de veces que se supera la variable Goal determina el grado de éxito de la recaudación y comúnmente se habla de proyectos de alto éxito.

4.2 PREDICCIÓN DEL ÉXITO

El uso de internet como medio de difusión del proyecto implica aceptar el hecho de navegar en un entorno muy activo y con cambios constantes, estimulados tanto por el por las tendencias colectivas, cambios en plataformas Web, actualización de normativas, etc. Hsieh et al. ya estudiaron como los movimientos sociales influyen en el éxito de los proyectos de crowdfunding ayudando a entender como el contexto social afecta al éxito de los proyectos de este tipo (Hsieh et al., 2019). Estas condiciones de entorno forman un contexto clave para generar confianza en el inversor ya que pueden influir negativamente e incluso poner en serio riesgo un proyecto bien ejecutado (Bakker, 2013) (Wilhelm et al., 2015).

El poderoso efecto de internet y en concreto las redes sociales han demostrado ser una herramienta muy útil en campañas de marketing de productos, de cara a estimular a una multitud la adquisición de un producto nuevo (Yerasani et al., 2019). Por otra parte, el estudio de los factores influyentes en el éxito se ve obstaculizado por la dificultad de control de gran número de variables interrelacionadas.

Hui et al. recopilaron datos a través de entrevistas a 47 creadores de proyectos y encontraron importantes posibilidades de mejora, sobre todo identificaron grandes necesidades en áreas de mentoría y feedback ya que exponen cómo factor muy importante la comunicación con la comunidad (Hui et al., 2014).

También se ha estudiado la relación entre el éxito y los logotipos, fotos e imágenes presentes en la campaña de financiación. Haciendo uso de encuestas, una base de datos y un experimento, se han revelado importantes capacidades para influir en los impulsores, resultando ser los de mayor complejidad visual los que pueden dar una señal de innovación y ayudar a impulsar aún más la campaña (Mahmood et al., 2019).

Desde el punto de vista de la descripción del proyecto Mitra and Gilbert consideran para su estudio la variable de respuesta binaria: proyecto exitoso o proyecto fracasado y

aplicando técnicas de text mining analizan como ciertas frases y textos que acompañan a la descripción del proyecto son capaces de inducir al éxito o al fracaso (Mittra and Gilbert, 2014).

En el artículo *“Crowdfunding and Online Social networks”* se expone cómo el número de clicks en redes sociales está fuertemente relacionado con los patrocinadores, usando datos de clicks en redes sociales, combinándolos con una base de datos de proyectos de la plataforma descubrieron que el número de patrocinadores aumenta proporcionalmente a la raíz cuadrada del número de clicks (Hekman and Brussee, 2013).

4.3 FRACASO

En crowdfunding se entiende por fracaso al proyecto que cumpliendo con las normas de la plataforma no alcanza la cuantía mínima para su desarrollo. Este estado resulta negativo para ambas partes, ya que implica tanto al creador (beneficiario económico) como a los colaboradores (beneficiarios del producto). Sin embargo, un proyecto puede resultar exitoso para el creador porque alcanzó el capital necesario, pero puede que el producto carezca de las especificaciones técnicas comprometidas al principio del proyecto, convirtiéndose entonces en una campaña exitosa para el creador, pero fraudulenta para el consumidor. Por ello es necesario disociar las consecuencias para el creador y para los colaboradores.

4.3.1 PARA EL CREADOR

A priori, la mayor preocupación para el creador reside en no alcanzar la financiación necesaria. El tipo de financiación AON característica de este modelo, no permite conseguir ninguna cuantía económica por parte de la plataforma si no se alcanza la cantidad mínima, al contrario de otros modelos como el *“Keep-It-All”* (KIA) que admiten conseguir la cuantía económica alcanzada hasta la fecha fin del proyecto, si esta al menos supera un umbral fijado por la plataforma, de esta manera en muchos casos se establecerían varios grados de fracaso dependiendo de la cantidad alcanzada, incluso en algunos casos permite desarrollar el proyecto con normalidad (Cumming et al., 2015).

En el caso AON, el fracaso tiene las mismas consecuencias económicas tanto para proyectos que han estado a punto de alcanzar su capital mínimo como los que han recibido pocos o ningún apoyo.

A priori no hay ningún indicio de que las causas que definen el éxito sean capaces de definir el fracaso de manera simétrica, sin embargo los esfuerzos del creador están muy condicionados al éxito y de manera parcial al fracaso, ya que la experiencia de un creador en este tipo de proyectos ayuda notablemente a corregir o alejar del éxito los proyectos (Tran et al., 2016).

Por ello, una de las razones más comunes capaces de llevar al fracaso a este tipo de proyectos reside en la incapacidad del creador para hacer llegar su proyecto hacia

futuros inversores (Mollick, 2014), razón por la cual se recomienda enérgicamente el uso de redes sociales como Twitter (An et al., 2014).

Más allá de la no obtención de financiación y a pesar de que el mismo proyecto puede ser posicionado en otros medios similares al mismo tiempo, cuando un proyecto no consigue la financiación necesaria sigue expuesto en la plataforma, mostrando la etiqueta de “fracasado”, acompañando a esta etiqueta también se ofrece información de todas las características del proyecto y del proceso de financiación hasta su fin. Estos atributos expuestos a una multitud pueden condicionar una idea novedosa, si no se encuentra convenientemente protegida (Bacon et al., 2019), también puede desfavorecer el perfil del creador, en detrimento de proyectos posteriores que pueda realizar.

4.3.2 PARA EL COLABORADOR

Fraude

El distanciamiento físico entre colaborador y creador imposibilita cualquier tipo de auditoría in situ, por lo que es responsabilidad del interesado la consulta de dudas a través de las herramientas que ofrece la plataforma, tales como actualizaciones del proyecto, preguntas, etc. También es de obligado cumplimiento la grabación de un vídeo donde se muestre al menos un prototipo funcional del producto.

A pesar de que los colaboradores presentan cierta tolerancia al fraude (Ellman and Hurkens, 2019), el fraude no es muy común en este tipo de proyectos, ya que existen unas normas de obligado cumplimiento que minimizan el riesgo de fraude y aportan confianza al futuro colaborador (Kickstarter, 2019). Ningún proyecto presentado puede pasar a la fase de financiación si al menos no ha conseguido al menos un número determinado de likes en Facebook o Twitter. Revisar el perfil del creador en redes sociales puede aportar confianza ya que permite al futuro patrocinador comprobar la red social del entorno del creador, a pesar de que es posible el establecimiento de falsos likes (Wessel et al., 2016).

Recompensas de baja calidad

Uno de los proyectos de mayor éxito de crowdfunding fue el smartwatch Pebble (Kickstarter, 2012), que a pesar de superar ampliamente el objetivo y considerarse como un caso ejemplar de éxito en crowdfunding, la enorme demanda de este smartwatch ha provocado que se superase la capacidad productiva prevista por el creador, teniendo que desviar la producción a China para satisfacer a casi 69.000 interesados.

A pesar de que las plataformas advierten de los riesgos inherentes a los productos, muchos de los patrocinadores del proyecto Pebble se quejaron de la baja calidad de las últimas unidades en ajuste y en acabados (Montes de Oca, 2019). Es responsabilidad del creador alcanzar la capacidad productiva necesaria para su proyecto y así evitar este tipo de problemas (Koch, 2016).

Tiempo de entrega

El retraso en la entrega es muy común en este tipo de proyectos y no parece ser un gran problema para los patrocinadores ya que a medida que se desarrolla el proyecto los interesados son más conscientes de las fases del proyecto. Las plataformas evitan la responsabilidad contractual directa por la entrega de recompensas, por lo que una vez que el proyecto ha finalizado la fase de financiación la plataforma exime la responsabilidad de la entrega, quedando así dependiendo solamente del creador (Mollick, 2014).

4.4 CONTROL DEL AVANCE DE LA CAMPAÑA

La naturaleza y las características de este tipo de proyectos es tan amplia que dificulta el uso de herramientas concretas para cada caso y que controlen el desarrollo de la campaña de financiación, también resulta de especial dificultad conocer en cada momento si el proyecto está en vías de ser financiado, va encaminado al fracaso o admite alguna modificación para reorientarlo hacia el éxito. Por ello resulta necesario conocer que pasos deben tomarse en cada momento y posteriormente comprobar si la respuesta acerca más al éxito o lo aleja.

Basándose en la importancia que tiene el buen comienzo de una campaña de crowdfunding, Etter et al. proponen un mecanismo orientado a predecir el destino de la campaña de financiación observando periodos temporales pertenecientes al inicio de la financiación (Etter et al., 2013) que sin duda aportan una gran ayuda en un estadio inicial.

Las plataformas tienen herramientas a disposición de los creadores que permiten observar los apoyos que recibe la campaña a tiempo real, aunque la interpretación de los datos tiene que ser llevada a cabo por el creador, salvo que se contrate un servicio de asesoría externo (Boostyourcampaign, 2020), muchos de estos servicios están basados en satisfacer el marketing y publicidad de la campaña en redes sociales, también existen otras herramientas online externas que se encargan de tomar datos de diarios de la campaña y muestran los datos de manera diferente o bien crean indicadores para facilitar la labor al creador (Clark, 2020). A través de estas herramientas, el creador puede comparar su campaña con otras exitosas de naturaleza similar, para superponer las trayectorias de recaudación y comprobar en cada punto hacia donde se dirige su tendencia.

4.4.1 PATRONES

Existen algunos indicadores del avance de la campaña ampliamente usados por su sencillez y rapidez de uso, a pesar de que son de carácter generalista y puede haber elevadas diferencias con la campaña activa. Su control se basa en comparar la campaña en estado de financiación con patrones observados en campañas de crowdfunding que han resultado exitosas.

Una regla comúnmente usada por los creadores en periodos iniciales de recaudación es conocida como la regla del “30-90-100” que recuerda de modo nemotécnico que una campaña que alcance el 30% del objetivo la primera semana de campaña, el proyecto cuenta con un 90% de probabilidad de alcanzar el 100% de la cantidad necesaria (Acconcia, 2014).

Otro patrón descriptivo del avance de la campaña y ampliamente utilizado es la curva en forma de “U”, también conocida como “Bathtub” o curva de la bañera. Este patrón de comportamiento está basado en que la probabilidad de éxito de un proyecto de crowdfunding mejora considerablemente si los aportes iniciales son fuertes, ya que, en estados iniciales es cuando se manifiesta de manera más importante el *efecto espectador* (Chan et al., 2020). Este efecto justifica que tras las primeras aportaciones realizadas a la campaña y motivados por la novedad, le siguen otros apoyos basados en la muestra de confianza que ofrecen los primeros. Este comportamiento también está caracterizado por una zona intermedia donde las contribuciones a la campaña se reducen ostensiblemente hasta una zona final donde vuelven a repuntar los apoyos, debido a la sensación de miedo que proporciona la fecha límite y las posibilidad de quedarse sin el producto (Kuppuswamy and Bayus, 2013).

5 METODOLOGÍA

En trabajos donde es necesario el estudio relacional de un gran número de datos es muy común el uso de técnicas de data science, sobre todo en sectores donde las interacciones son muy complejas como el caso de las redes sociales (Sohrabi and Akbari, 2016). Su potencialidad radica en la capacidad de extraer información más allá de la mera relación univariante lineal. La gran ventaja de usar técnicas de data mining para este trabajo radica en que aportan en gran medida información que explica la relación de cada factor influyente en crowdfunding sobre el éxito final.

Se han valorado distintas estrategias metodológicas orientadas a la realización de proyectos de minería de datos, valorando experiencias exitosas previas, estabilidad en el tiempo y difusión, razón por la que se ha optado por organizar el estudio con la metodología CRISP-DM, a continuación, se describe sus características, estructura y fases.

5.1 CRISP-DM

Para estructurar este proceso se llevará a cabo el procedimiento CRISP-DM (*Cross Industry Standard Process for Data Mining*) seleccionado cómo la metodología más apropiada por su flexibilidad y porque está suficientemente probada para convertir datos en conocimiento de una manera organizada (Chapman et al., 2000). Consta de seis fases claramente definidas que por su carácter iterativo aseguran que cada proceso de análisis se beneficie de la experiencia previa.

La información se ha convertido en uno de los recursos más valiosos de las organizaciones. Con el paso del tiempo el volumen de datos almacenados ha ido aumentando en relación con el progreso de las TIC en todos los sectores. Esta historia en forma de datos ha derivado en la creación de diversas técnicas que faciliten su entendimiento. Este es el caso de la minería de datos (*Data Mining*), concebida para extraer conocimiento útil y comprensible, previamente desconocido en grandes volúmenes de datos. La extracción de esta información útil es un proceso complejo, que requiere la aplicación de una metodología que permita una utilización ordenada y eficiente de las técnicas y herramientas disponibles, ya que por lo general los datos no se encuentran refinados y para poder analizarlos con fiabilidad es necesario que exista una cierta estructuración y coherencia entre los mismos. Por ello, una de las metodologías más avanzadas del proceso de minería de datos es CRISP-DM, respaldado y desarrollado a finales de 1996 por un consorcio de algunas de las mayores compañías en la industria de la minería de datos: NCR, SPSS, DaimlerChrysler y OHRA (Chapman et al., 2000).

Una de las grandes ventajas de CRISP-DM es la gran capacidad de adaptación a proyectos de datos ya que es capaz de relacionar todas las fases del proceso permitiendo el cruzado de datos en todo momento y con la flexibilidad para elegir

libremente las herramientas a usar, esta metodología permite estructurar un proyecto de minería de datos en diversas fases interrelacionadas entre si, resultando un proyecto de *Data Mining* iterativo e interactivo.

5.1.1 ESTRUCTURA

La metodología CRISP-DM proporciona dos documentos distintos como herramienta de ayuda en el desarrollo del proyecto de Data Mining: el modelo de referencia y la guía del usuario.

El documento del modelo de referencia describe de forma general las fases, tareas generales y salidas de un proyecto de *Data Mining*. La guía del usuario proporciona información más detallada sobre la aplicación práctica del modelo de referencia a proyectos de *Data Mining* específicos, proporcionando consejos y listas de comprobación sobre las tareas correspondientes a cada fase.

La metodología CRISP-DM consta de cuatro niveles de abstracción organizados de forma jerárquica en tareas que van desde un nivel general a casos más específicos como se observa en la Figura 8.

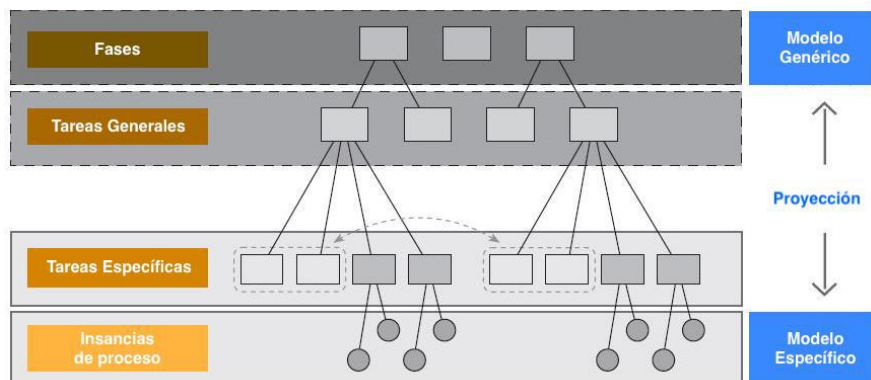


Figura 8: Esquema de los cuatro niveles de abstracción de la metodología CRISP-DM

5.1.2 FASES

La metodología CRISP-DM estructura el ciclo de vida de un proyecto de *Data Mining* en seis fases, que interactúan entre ellas de forma iterativa durante el desarrollo del proyecto. Cada fase, a su vez se encuentra estructurada en varias tareas generales de segundo nivel, que posteriormente se proyectan a tareas específicas donde se describen las acciones para desarrollarlo. De esta manera, si el segundo nivel contiene la tarea general de “generación de nuevas variables”, el tercer nivel deberá contener los trabajos necesarios para trasladarlo a un caso específico como puede ser por ejemplo “variables discretas” y/o “variables continuas”. El cuarto nivel, recoge el conjunto de acciones, decisiones y resultados sobre el proyecto de Data Mining específico como se observa en la Figura 9.

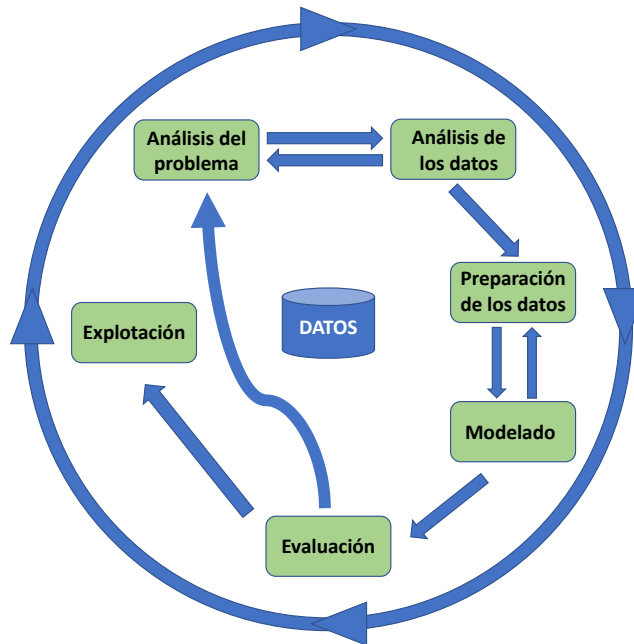


Figura 9: Fases del modelo de referencia CRISP-DM

Las flechas indican relaciones más habituales entre las fases, aunque se pueden establecer relaciones entre cualquier fase. El círculo exterior simboliza la naturaleza cíclica del proceso de modelado.

En la fase de *Análisis del Problema* se deben entender los objetivos del proyecto y los requerimientos desde una perspectiva del negocio, ya es imposible tomar decisiones sin tener un conocimiento adecuado de lo resulta interesante, sorprendente o relevante con respecto al problema a resolver, posteriormente este conocimiento se convierte en la definición de un problema de explotación de información, para diseñar un plan preliminar y lograr dichos objetivos.

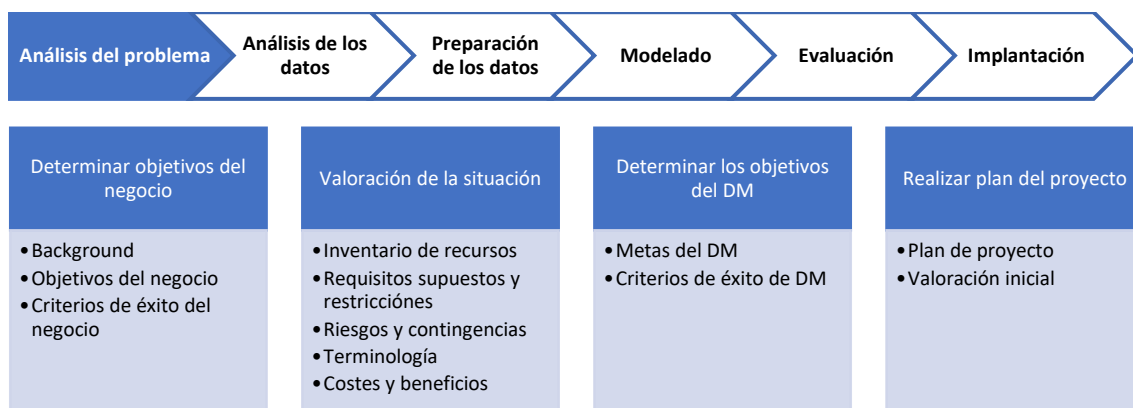


Figura 10: Fases de análisis del problema del modelo de referencia CRISP-DM

En el *Análisis de los datos* se incluye la recolección inicial de datos y un análisis exploratorio para familiarizarse con ellos, para entender su origen, naturaleza y fiabilidad. Es importante la adecuada selección de herramientas de procesado que permitan identificar las primeras pautas o detectar subconjuntos interesantes, sólo es posible si se tiene una buena idea global de los datos.

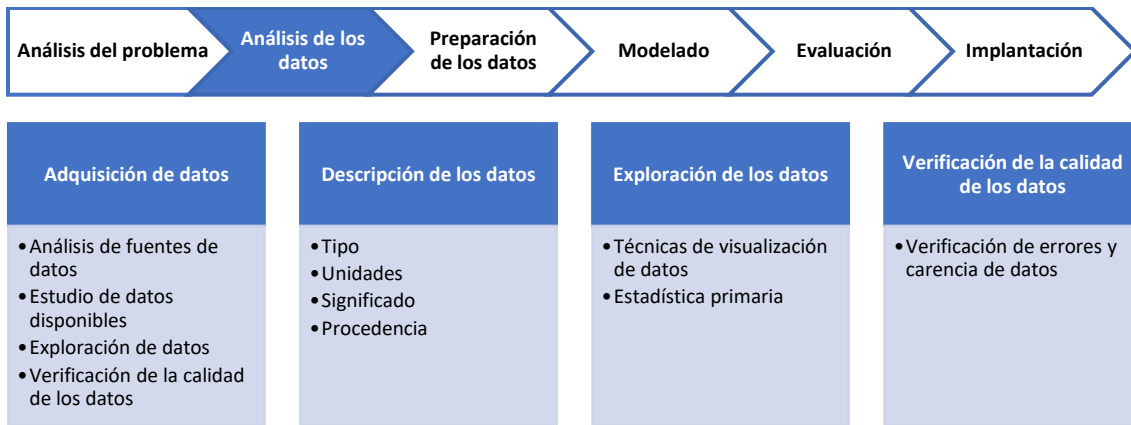


Figura 11: Fases de análisis de datos del modelo de referencia CRISP-DM

En la Fase de *Preparación de los Datos* el objetivo fundamental es facilitar la labor de construir modelos fiables y precisos a partir de los datos iniciales, extrayendo nuevas características, corrigiendo errores en los datos y realizando adaptaciones necesarias para la aplicación de técnicas de modelado de datos de la siguiente fase.

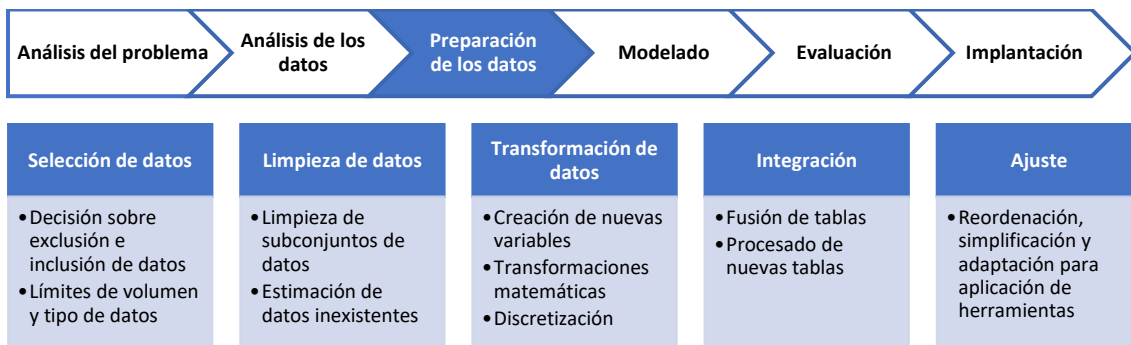


Figura 12: Fases de preparación de los datos del modelo de referencia CRISP-DM

En la fase de *Modelado*, los datos ya se encuentran adaptados a las herramientas necesarias para la creación de un modelo donde buscar soluciones al problema y es en esta fase donde se especifica. Puede ser mediante un modelo predictivo, descriptivo o una clasificación de datos en conjunto de grupos, así como el análisis de ciertas propiedades, dependencias o estimadores de la función de densidad de probabilidad de los datos. Desde esta fase muchas veces es necesario volver a la fase de *Preparación de los Datos*.

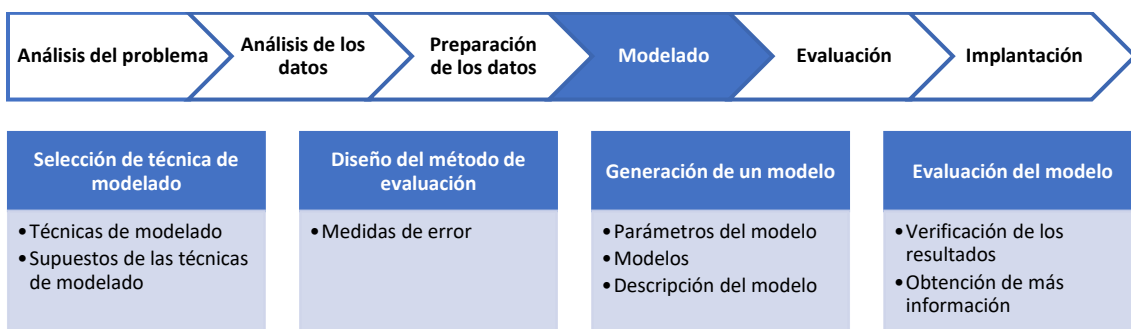


Figura 13: Fases de modelado del modelo de referencia CRISP-DM

La fase de *Evaluación* es la que precede a la solución a desarrollar, es necesario que el punto de vista del problema sea el original propuesto para determinar si la solución encontrada es lo suficientemente buena como para llevarla a cabo. Aquí se evalúan y revisan anteriores modelos, así como nuevas ideas o modelos secundarios que pueden complementar el problema a resolver.

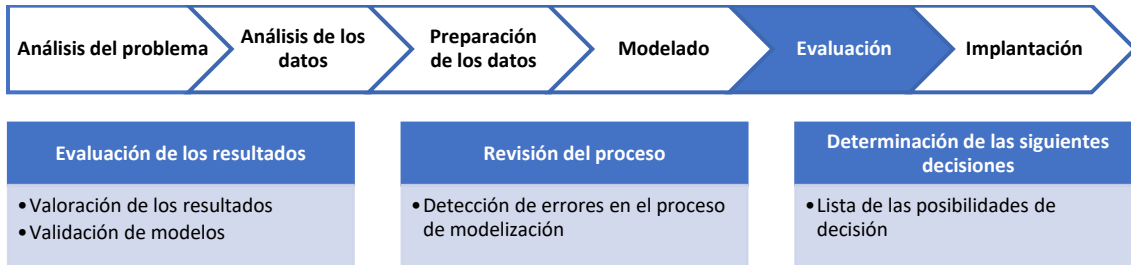


Figura 14: Fases de evaluación del modelo de referencia CRISP-DM

En la fase de *Implantación* se aplica la solución para resolver el problema original, siguiendo una lista de pasos creados que serán revisados a medida que se vayan aplicando, finalmente se evalúan las conclusiones basándose en los informes generados durante el desarrollo del proyecto de *Data Mining*.

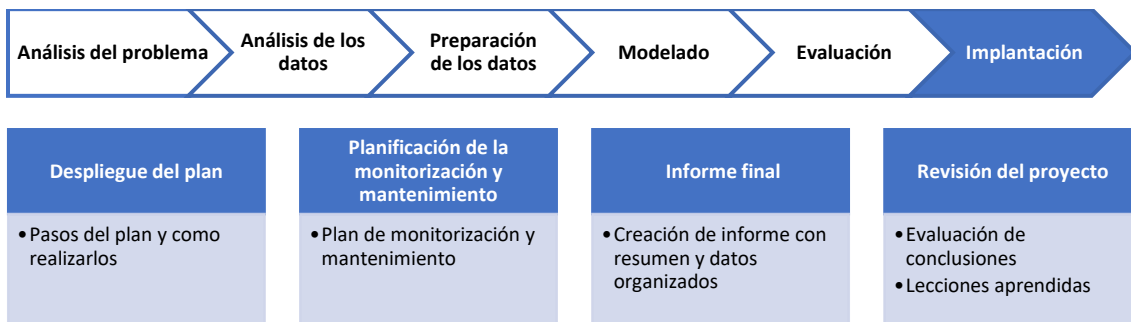


Figura 15: Fases de implantación del modelo de referencia CRISP-DM

6 TÉCNICAS EMPLEADAS

Para alcanzar el objetivo de este trabajo es necesario aplicar una serie de técnicas encaminadas a caracterizar el conjunto de datos, basadas en identificar las variables más representativas y diferenciadoras que permitan mejorar la comprensión y estructura de los datos. Las técnicas que se van a aplicar están contenidas dentro de la etapa de análisis de KDD (Knowledge Discovery in Databases) (Han and Kamber, 2011) y facilitan el reordenamiento y agrupación de los datos, que puede derivar en una posterior identificación de patrones, grupos con características comunes o simplemente identificar las variables más influyentes del sistema.

Un rasgo común que se ha tenido en cuenta cuando se han seleccionado las técnicas usadas en este trabajo es que su uso esté ampliamente extendido y contrastado por la comunidad científica internacional, también deben de poseer la flexibilidad suficiente para caracterizar y procesar datos cada una desde un enfoque diferente. Algunas lo hacen teniendo en cuenta el punto de vista del conjunto de datos y otras valoran el entorno de las variables. Facilitando en todo caso la aplicación de criterios de asociación grupales que permiten estudiar las variables más representativas de manera aislada y en su conjunto.

Para estadios iniciales de este trabajo se usan técnicas con capacidad para valorar el entorno de los datos desde el punto de vista de la investigación científica, esta técnica será de gran ayuda para comprender que palabras o términos resultan más influyentes para la comunidad científica. A pesar de que el procesado de variables no numéricas no se encuentra dentro del alcance de este trabajo, se extraerán las palabras más frecuentes dentro de los títulos de proyectos para contemplar la posibilidad de aportar más información a otras variables.

Una vez dentro de las variables numéricas, se realizará una identificación de las variables más influyentes en cada categoría y así valorar si el éxito de cada categoría está relacionado con ciertas variables asociadas, posteriormente se procederá a una reducción del número de dimensiones del sistema, entre otras cosas, para comprobar si existe posibilidad de prescindir de alguna variable redundante que limite el número de combinaciones entre variables.

También se agruparán los datos en clústeres atendiendo la categoría a la que pertenecen, así se podrá evaluar de manera gráfica tanto zonas asociadas a éxito dentro de una categoría como también buscar si existen zonas de éxito común para todas las categorías.

6.1 PCA (PRINCIPAL COMPONENTS ANALYSIS)

Contexto de uso

Para mejorar el conocimiento de un conjunto de datos es necesario hacer un estudio exploratorio usando técnicas propias de estadística descriptiva, el análisis de las componentes principales permite reducir las dimensiones del sistema.

Gracias a las aportaciones del científico británico Karl Pearson, considerado como uno de los fundadores de la estadística matemática en el siglo XIX, logró desarrollar el ajuste por mínimos cuadrados ortogonales (Pearson, 1901), técnicas que en los años 30 del siglo XX colaboraron a desarrollar un modelo que permite identificar las componentes principales en un sistema de este tipo (Hotelling, 1933).

En el momento de recoger información de una muestra de datos, lo más habitual es tomar el mayor número de variables posible, sin embargo, si se toma un número muy elevado de variables aumentan en gran medida los coeficientes de correlación con lo que es muy difícil visualizar las diferentes relaciones entre variables.

En estadios iniciales cuando aún no existe conocimiento suficiente de los datos, existe el problema añadido de tomar un elevado número de variables, pudiendo darse el caso de que una o varias midan lo mismo desde diferentes puntos de vista. Llegados a este punto, es necesario reducir el número de variables, teniendo en cuenta que cuanto mayor sea la variabilidad de los datos (varianza) mayor es la cantidad de información aportada.

El método de Análisis de Componentes Principales tiene como objetivo reducir la información que aporta un conjunto de variables, denominadas *originales*, para dar lugar a otras denominadas *componentes principales*, creando de esta manera un sistema en el que estas componentes están interrelacionadas entre sí y pueden ser ordenadas en función de la información que llevan. Generalmente este método se aplica sobre variables tipificadas para evitar los problemas asociados a las diferentes escalas.

En la vida real, se supone con mucha frecuencia que una variable en estudio tiene una distribución normal, ya que, muchas características se miden a través de la asociación de diversas causas que actúan simultáneamente sobre el suceso. Por ejemplo: se considera con distribución normal la altura de las personas, debido a que su valor se debe a diferentes causas, ya sea alimentarias, genéticas o ambientales.

El análisis de componentes principales no requiere suponer la normalidad de los datos, aunque si los datos cumplen los criterios de normalidad las componentes se pueden interpretar de manera más profunda. Esto matemáticamente se justifica a través del Teorema Central del Límite que demuestra que la suma de variables independientes de varianza no nula se aproxima a una curva normal. La situación ideal en este análisis sería que un número bajo de componentes principales explicaran mucha varianza.

Cálculo de las componentes principales

Las nuevas variables serán combinación lineal de las anteriores y se irán añadiendo en orden según importancia, referida esta a la variabilidad total de la muestra.

Así, se buscan $m < p$ variables que deben de ser combinación lineal de las originales (p), y deben de ser linealmente independientes entre ellas y que recojan la mayor variabilidad de datos posible.

De esta manera se toman p variables, (x_1, x_2, \dots, x_p) , para tratar de encontrar a partir de ellas las variables (y_1, y_2, \dots, y_p) independientes entre si y que además sus varianzas vayan aumentando progresivamente. De esta manera cada y_j donde $j = 1, \dots, p$ es combinación lineal de las variables originales (x_1, x_2, \dots, x_p) .

$$y_j = a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jp}x_p = \mathbf{a}'_j \mathbf{x} \quad (1)$$

$$\mathbf{a}'_j = (a_{1j}, a_{2j}, \dots, a_{pj}) \text{ y } \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \quad (2)$$

Como se pretende maximizar la varianza se aumentan los coeficientes a_{ij} con el fin de mantener la ortogonalidad de la transformación. Estos valores de a se denominan *loadings*, son los que definen la componente y pueden interpretarse como la importancia que tiene cada variable en cada componente, por ello se impone $\mathbf{a}'_j = (a_{1j}, a_{2j}, \dots, a_{pj}) = 1$, o lo que es lo mismo, hacer que el módulo sea 1, de tal manera que:

$$\mathbf{a}'_j \mathbf{a}_j = \sum_{k=1}^p a_{kj}^2 = 1 \quad (3)$$

Se extraen los factores:

$$Var(y_1) = Var(\mathbf{a}'_1 \mathbf{x}) = \mathbf{a}'_1 \Sigma \mathbf{a}_1 \quad (4)$$

Esta función se maximiza utilizando el método de los *multiplicadores de Lagrange*, teniendo en cuenta la restricción inicial: $\mathbf{a}'_1 \mathbf{a}_1 = 1$.

λ es un autovalor de Σ . La matriz de covarianzas es de orden p , si también toma valores positivos, tendrá p autovalores distintos: $\lambda_1, \lambda_2 \dots \lambda_p$

Dando lugar a:

$$Var(y_1) = Var(\mathbf{a}'_j \mathbf{x}) = \mathbf{a}'_j \sum \mathbf{a}_1 = \mathbf{a}'_j \lambda \mathbf{a}_1 = \lambda \mathbf{a}'_1 \mathbf{a}_1 = \lambda * 1 = \lambda \quad (5)$$

De esta manera la primera componente principal queda definida según:

$$y_1 = \mathbf{a}'_1 \mathbf{x} = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \quad (6)$$

A partir de la primera componente se calcula la segunda, teniendo en cuenta que estas deben ser linealmente independientes entre ellas, es decir: $Cov(y_2, y_1) = 0$.

Todas las componentes de y (p componentes), se pueden expresar en forma matricial, donde el producto de la matriz formada por los autovectores multiplicada por el vector x que contiene las variables originales $x_1, x_2 \dots x_p$.

$$y = Ax \quad (7)$$

Donde:

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}, A = \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \dots & a_{pp} \end{pmatrix}, x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \quad (8)$$

Y se obtiene que:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_p \end{pmatrix} \quad (9)$$

$$\Lambda = Var(Y) = A'Var(X)A = A' \sum A \quad (10)$$

6.2 SELF-ORGANIZING MAPS

Contexto de uso

La posible relación o interconexión entre varios conjuntos de datos deja de ser evidente cuando el tamaño de estos sobrepasa cierto límite. Intentando entender el funcionamiento computacional de conexión entre las neuronas cerebrales, surge el primer modelo neuronal moderno. De manera matemática McCulloch and Pitts aportan una primera explicación al funcionamiento de interconexión neuronal, ayudándose por la lógica binaria crean un modelo que más tarde favorecería el desarrollo de lo que hoy en día se conoce como Redes Neuronales Artificiales (McCulloch and Pitts, 1943).

Casi quince años después inspirado en la teoría de la plasticidad sináptica (adaptación de las neuronas durante el proceso de aprendizaje) Frank Rosenblatt realizó una mejora sobre el modelo proporcionado por McCulloch and Pitts desarrollando el *perceptrón*, una neurona artificial considerada como una unidad básica, que a través de

su algoritmo es capaz de generar un criterio para seleccionar un subgrupo de un grupo más grande como se observa en la Figura 16 del perceptrón (Rosenblatt, 1958).

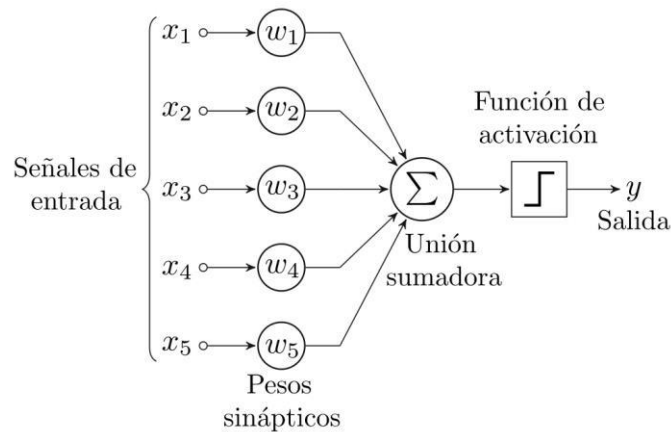


Figura 16: Perceptrón como unidad básica

Fueron McCulloch and Pitts quienes aportaron la primera explicación de la eficiencia en la transmisión de señales al cuerpo de la neurona, para ello se asocia un factor multiplicativo a cada línea de entrada a la neurona, de esta manera, si una neurona tiene n entradas que proceden de otras neuronas el modelo viene representado por la ecuación:

$$Input\ total = peso\ linea\ 1 \times input\ 1 + \dots + peso\ linea\ n \times input\ n \quad (11)$$

Es decir:

$$Input\ total = \sum_{i=1}^n w_i x_y \quad (12)$$

En minería de datos es común el reconocimiento de patrones, ya que, permite simplificar el conocimiento en base a comparar medidas con un modelo para conocer su estado actual.

Existen multitud de algoritmos destinados a este fin, pero gracias a su carácter gráfico y no jerárquico, una de las herramientas más adecuadas para llevar a cabo el reconocimiento de patrones en minería de datos es mediante el estudio de redes neuronales artificiales (ANN) (Coates and Ng, 2012), una de las ANN más destacadas es la SOM (Self-Organizing Map) o mapas autoorganizados que permiten mediante una representación gráfica observar la simplificación de un conjunto de datos en base a utilizar una función de vecindad para preservar las propiedades topológicas del espacio de entrada. Esta técnica desarrollada en 1982 por Teuvo Kohonen (Kohonen, 1982) es hoy en día una de las principales técnicas de análisis de datos usada en literatura científica.

Esta red neuronal sigue el principio de “maximizar la similaridad intraclase y minimizar la similaridad interclase” (Han and Kamber, 2011) en base a un proceso de entrenamiento que toma como entrada cada uno de los proyectos y mediante un proceso competitivo basado en distancias euclídeas, se genera por cada neurona un nuevo vector representante de los proyectos llamado centroide. Esta operativa forma parte de un proceso de aprendizaje no supervisado, porque sólo se conocen los datos de entrada y no existen datos de salida que correspondan a un determinado input, es decir, la red no recibe ninguna información de cual es la salida, de hecho, se dice que esta red es capaz de auto organizarse.

A través de esta descripción, las redes neuronales artificiales tratan de simular el proceso biológico tal como lo harían sus homólogas, que se acaba de describir, en un equipo computacional basándose en el siguiente modelo:

- Se procesa la información en elementos simples llamados neuronas o nodos.
- Los inputs o señales, x_i , $i = 1, 2, \dots, n$; son transmitidos entre las neuronas a través de conexiones.
- Cada conexión lleva asociado un peso w_{ij} que representa la intensidad de la señal recibida correspondiente al input i en la neurona j . La mayor o menor intensidad de cada input representa la influencia en el resultado final.
- Cada neurona j suma todas las señales recibidas, $\sum w_{ij} x_i$ que proceden de otras neuronas y se transforman mediante una función de activación f , generalmente no lineal, para determinar la señal de salida output que transmite a otras neuronas, donde el proceso se repite. Tras la interacción de las neuronas que componen la red, la señal o respuesta de la red se transmite al exterior.
- El aprendizaje se realiza mediante el ajuste de pesos que ponderan las conexiones entre las neuronas de la red. El algoritmo de aprendizaje es el método por el que se obtienen los ajustes óptimos evitando tener que supervisar todo el proceso. Este método consta de una serie de pasos sencillos e iterativos.
- Para procesar la información las neuronas se organizan en capas: la capa de entrada transmite las variables input utilizadas, la capa de salida presenta el output final y entre medias pueden existir una o más capas ocultas que procesan la información.

La arquitectura de esta red neuronal artificial está determinada por tres capas virtuales que definen su estructura funcional. La capa más externa se denomina *capa de entrada* ya que recibe los inputs del exterior y generalmente no procesa las señales de entrada, sino que éstas son enviadas a la siguiente capa, la *capa oculta*. Las neuronas pertenecientes a esta última transforman la señal recibida por la capa de entrada hacia la siguiente capa, pero durante este proceso el aprendizaje de la red transmitirá información a las neuronas de la capa o capas ocultas para que modifiquen sus pesos

mediante el proceso denominado *entrenamiento*. Por último, la *capa de salida* que posee las neuronas que envían la respuesta al exterior tras realizar una nueva transformación de la señal recibida por las neuronas de la capa oculta. En la Figura 17 se muestra la arquitectura en tres capas, característica de este sistema.

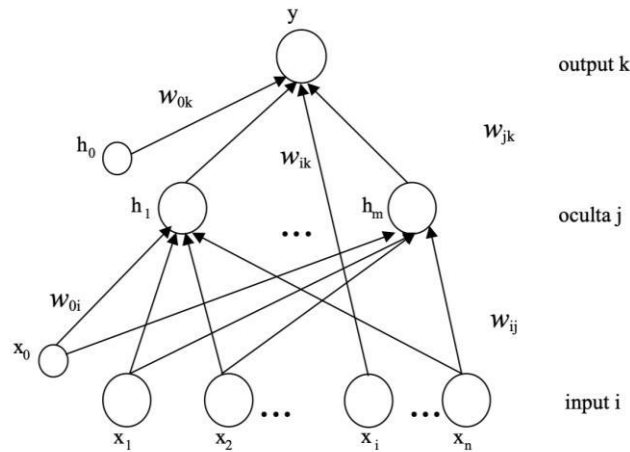


Figura 17: Organización de las capas en una red neuronal artificial

El objetivo de esta red es aprender a asociar un vector input, X , a un vector output, Y , a partiendo de la interacción entre neuronas, W . Esto es, dada la función de activación f , la red es entrenada para buscar una matriz de pesos W tal que:

$$f \quad (13)$$

$$\mathbb{R}^n \rightarrow \mathbb{R}^m \quad (14)$$

$$f(W, X) = Y \quad (15)$$

$$y_k = f_0 \left[\sum_{i \rightarrow k} w_{ik} x_i + \sum_{j \rightarrow k} w_{jk} f_h \left(\sum_{i \rightarrow j} w_{ij} x_i \right) \right] \quad (16)$$

A partir de una muestra $\{X(p), Y(p)\}, p = 1, 2, \dots, H$ donde $X(p) \in \mathbb{R}^n, Y(p) \in \mathbb{R}^m$. Donde i denota la información en capa de entrada, j la capa oculta, k la capa de salida, h la función de activación en las neuronas de la capa oculta y 0 en las neuronas de la capa de salida (Haykin, 2009).

Visualización

Estas redes creadas basan su funcionamiento en un aprendizaje en el que las neuronas de salida de la red compiten entre si para que se activen o se retiren, de tal manera que sólo quede una neurona de salida o una neurona por cada grupo. La neurona de salida de este proceso competitivo se denomina neurona ganadora.

A pesar de que, en este tipo de mapas autoorganizados, las neuronas se colocan en los nodos de una red, esta puede ser de varias dimensiones, aunque lo más común es que sea unidimensional o bidimensional.

Las neuronas se van sintonizando selectivamente dependiendo los estímulos de la entrada o clases de patrones de entrada durante el proceso de aprendizaje competitivo. Las ubicaciones de las neuronas sintonizadas, es decir, las neuronas ganadoras, se ordenan entre sí de tal manera que se crea un sistema de coordenadas significativo para las diferentes características de entrada sobre la red (Haykin, 2009). De esta manera se va formado un mapa topográfico caracterizado por los patrones de entrada de las neuronas en la red, quedando ordenado según las características estadísticas intrínsecas contenidas en los patrones de entrada como se observa en la Figura 18.

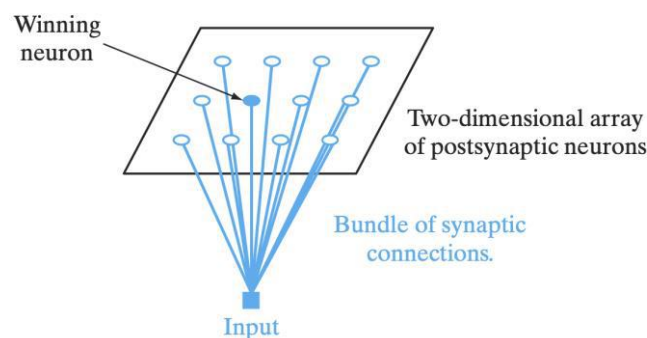


Figura 18: Modelo Kohonen de conexiones sinápticas (Haykin, 2009)

Cada neurona de la red está completamente conectada con todos los nodos fuente ubicados en la capa de entrada, una vez que la red se ha iniciado correctamente el mapa se crea basándose en tres procesos: competitividad, cooperación y adaptación sináptica.

La neurona ganadora y su radio de influencia disminuye monótonamente a medida que avanza el entrenamiento (Vesanto, 2002), así el mapa SOM creado logra que las neuronas próximas sintonicen con patrones similares, quedando reflejado sobre el mapa una imagen con cierto orden topológico presente en el espacio de entrada.

El número de paneles en la capa de salida se esquematiza en un diagrama denominado Matriz de Distancias Unificadas (U-matrix) donde la representación de similaridad se muestra en color o en escala de grises como se observa en la Figura 19.

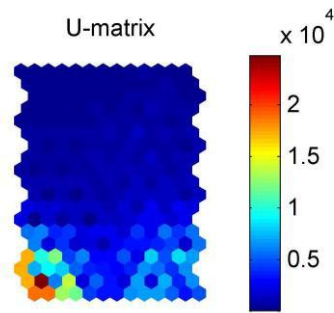


Figura 19: Matriz de distancias unificadas (U-matrix)

En este caso, cada neurona ganadora viene representada por un hexágono, por lo que nos da la posibilidad de establecer seis vecindades. El color de cada hexágono representa (según la escala) la distancia promedio entre el vector de referencia (ubicado en la posición del mapa) y las neuronas ganadoras vecinas. En la Figura 19 se observa que las zonas en color azul representan distancias cortas (mayor similitud) y en rojo distancias largas (menor similitud).

6.3 CLUSTERIZACIÓN Y EL MÉTODO K-MEANS

La complejidad de análisis de una gran cantidad de datos se reduce si estos se dividen en grupos (clústeres) que mantengan la información que proporcionan, así como las relaciones establecidas entre ellos. Esto permite una abstracción de los datos individuales hacia clústeres donde residen los datos. Se facilita de esta manera el entendimiento de las observaciones al caracterizar grupos frente a objetos individuales.

K-means es un algoritmo de clasificación (no supervisada) que agrupa objetos en un número K de grupos fundamentados por sus características. El principio de funcionamiento de este algoritmo es minimizar la suma de las distancias entre cada objeto y el centroide de su grupo o clúster. Por ello K-means resuelve el problema optimizar las distancias de cada objeto al centroide de su clúster

La finalidad es establecer grupos significativos que mantengan la estructura natural de los datos, aunque también es posible aplicar este algoritmo previo a otra fase, como comprender clases o grupos de datos asociados según características comunes.

Esta agrupación de datos en clústeres se puede considerar en cierta manera como una clasificación (Žalik, 2008), ya que se crea un etiquetado de objetos en base a su clase, esto deriva en que en ocasiones se defina el análisis de clústeres como método no supervisado, ya que el proceso ignora la variable respuesta que indica a que grupo pertenece cada observación.

Funcionamiento

Para desarrollar esta técnica es necesario en primer lugar seleccionar un número determinado de grupos (K) con los que se va a trabajar (Macqueen, 1967). La selección del número de ellos depende fundamentalmente de los datos, para optimizar el número

de clústeres se ha de tener en cuenta que cada clúster debe contener observaciones con la menor varianza interna posible. Posteriormente a cada objeto de los datos se asocia su centroide más cercano y finalmente se reorganiza la posición del centroide de cada grupo para tomar como nuevo centroide la posición promedio de los objetos que pertenecen a ese grupo, con el fin de encontrar los mejores clústeres considerados por tener una suma de varianzas internas menor de todo el conjunto de clústeres. Es sistema se reorganiza el sistema iterando varias veces hasta que los centroides se establecen en una posición.

Los datos se introducen en el sistema en forma de vectores reales de d dimensiones (x_1, x_2, \dots, x_n) y el algoritmo construye K grupos con el fin de minimizar las distancias entre los objetos dentro de cada grupo a su centroide, siendo S el conjunto de datos, $S = \{s_1, s_2, \dots, s_K\}$. Obteniendo de esta manera K clústeres con su centroide μ_i .

$$\min_s E(\mu_i) = \min_s \sum_{i=1}^K \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (17)$$

Para el posterior proceso de iteración se impone la condición necesaria de extremo a esta función, de tal manera que:

$$\frac{\partial E}{\partial \mu_i} = 0 \Rightarrow \mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad (18)$$

Distancias

En todos los métodos de clusterización la distancia entre puntos juega un papel importante para cuantificar la diferencia o similitud entre datos, considerando siempre que cuanto más similares sean dos observaciones, más próximas estarán entre si.

De esta manera, cada clúster debe de contener objetos muy similares entre si y a su vez lo más diferentes posible del resto de clústeres, dando lugar a *la máxima homogeneidad intraclúster y la máxima heterogeneidad interclúster*.

El algoritmo K-means comienza con la búsqueda del centroide (centro geométrico del clúster), a través de una selección aleatoria de K observaciones sobre el conjunto de datos. Posteriormente, a cada una de las observaciones restantes se le asigna su centroide más cercano, haciendo uso en este caso de la distancia Euclídea entre el objeto y la media del grupo.

La elección del número de observaciones depende fundamentalmente de los datos; generalmente, valores grandes asignados a K reducen el efecto de ruido en la clasificación, pero por otro lado crean límites entre clases parecidas. Por ello un buen valor de K puede ser seleccionado mediante una optimización de uso.

La distancia Euclídea entre dos observaciones p y q se define cómo la longitud del segmento rectilíneo que los une, en coordenadas cartesianas la distancia Euclídea se realiza mediante la aplicación del teorema de Pitágoras. De esta manera la distancia Euclídea entre dos puntos p y q , definidos por sus coordenadas cartesianas (x,y) , viene dada por la ecuación:

$$d_{euc}(p, q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2} \quad (19)$$

Después de la asignación, el algoritmo calcula el nuevo valor medio para volver a recalcular los centroides y comprobar que cada observación puede estar más cerca de un grupo diferente. Estos pasos de actualización de centroide y asignación de clústeres se repiten de manera iterativa hasta que los clústeres asignados dejen de cambiar, en este momento queda establecido un mapa de clústeres.

6.4 MULTIVARIATE ADAPTATIVE REGRESSION SPLINE

Dentro de los análisis de regresión utilizados en estadística y más allá de los modelos típicamente lineales, se encuentra esta técnica no paramétrica comúnmente denominada como *MARS* por su acrónimo en inglés (Multivariate Adaptative Regression Spline). A pesar de que el término *MARS* es una marca registrada con licencia de Salford Systems (Salford Systems, 2020), el algoritmo de da lugar a su uso puede desarrollarse a través de diversos programas, por ejemplo: en “*Python*” se reconoce bajo el nombre de “*Py-earth*” (“*Python*,” 2020) y en “*R*” se reconoce bajo el nombre de “*Earth*” (Milborrow, 2019).

Este algoritmo modela automáticamente las no linealidades y las interacciones entre variables, aportando una clara ventaja que radica en la capacidad intrínseca para reducir las dimensiones de un sistema hacia modelos más simples y fáciles de interpretar (Friedman, 1991), así como aportar muy buenas capacidades para seleccionar las variables más relevantes del sistema.

Estructura

Este algoritmo funciona bajo la premisa “divide y vencerás”, así los datos de entrenamiento se dividen en porciones lineales (splines) de diferente gradiente (pendiente). No hace suposiciones sobre las relaciones funcionales subyacentes entre las variables dependientes e independientes, sino que en general, los splines se conectan suavemente entre sí formando unas curvas fragmentadas (polinomios), también conocidas como funciones base (FB), para dar como resultado un modelo flexible que puede manejar tanto el comportamiento lineal como el no lineal (Minitab, 2018).

Los puntos de conexión/interfaz entre las piezas se llaman nudos. Marcando el final de una región de datos y el comienzo de otra, los nudos candidatos se colocan en posiciones aleatorias dentro del rango de cada variable de entrada.

Este modelo puede describirse matemáticamente como:

$$y_t = f(x_t) = \beta_0 + \sum_{i=1}^k \beta_i B(x_{it}) \quad (20)$$

Donde y_t es la variable respuesta en el instante t y β_i son los parámetros del modelo para las respectivas variables x_{it} , que van de $i = 1, \dots, k$. El valor β_0 representa al intercepto y las funciones bases $B(x_{it})$ son funciones que dependen de las respectivas variables (x_{it}), en donde cada $B(x_{it})$ puede escribirse como $B(x_{it}) = \max(0, x_{it} - c)$ o bien $B(x_{it}) = \max(0, c - x_{it})$, c es un valor umbral y k representa el número de explicativas, que incluye interacciones de las variables predictoras. Los puntos de partición del espacio y los parámetros de los modelos se obtienen a partir de los datos analizados (Lewis and Stevens, 1991). El número de funciones base resultantes indica la complejidad del modelo.

Funcionamiento

El funcionamiento del algoritmo de *MARS* está basado en dos pasos. Primero construye un gran número de funciones base, es decir, cada función en el espacio puede representarse como una combinación lineal de las funciones de base. Estas funciones se seleccionan para sobreajustar los datos inicialmente, momento en el que se permite que las variables entren como continuas, categóricas u ordinales, siendo el mecanismo por el cual se definen los intervalos de las variables, y pueden interactuar entre sí o restringirse para entrar sólo como componentes aditivos.

En segundo lugar, las funciones de base se suprimen en el orden de menor contribución utilizando el criterio de validación cruzada generalizada (VCG), de tal manera que la medida de importancia de una variable puede evaluarse observando la disminución del VCG calculado cuando se elimina una variable del modelo.

De esta manera, es capaz de trabajar con cualquier forma arbitraria de la función y las interacciones, utilizando para ello el procedimiento de construcción de modelos en dos etapas mencionado. *MARS* es capaz de rastrear de manera confiable las estructuras de datos muy complejas que a menudo se ocultan en los datos de alta (Friedman, 1991)

6.5 DECISION TREE

Los Decisión trees (árboles de decisión) son una herramienta destinada a crear un modelo capaz de predecir el valor de una variable destino, teniendo en cuenta el valor de las variables de entrada. Se utilizan desde hace siglos siendo especialmente apropiados para la expresión de modelos estratégicos, matemáticos, lógicos y otros.

Este modelo de clasificación, conocido como ID3 (*Iterative Dichotomiser 3*) o “*inducción mediante árboles de decisión*” fue desarrollado por John Ross Quinlan y permite desarrollar un modelo en forma de árbol invertido de proporciones mínimas en el que la información expuesta se mantenga organizada y entendible por cualquier persona, donde siguiendo una secuencia de preguntas es capaz de evaluar la mejor respuesta posible (Quinlan, 1986). Haciendo uso de los árboles de decisión se pueden distinguir objetos de diferentes clases e identificar las características que los describen.

A partir del ID3, Quinlan desarrolló un nuevo algoritmo que permite obtener los resultados de forma más compacta y que denominó C4.5 y al igual que el ID3 forma parte de la familia de los TDIDT (Top Down Induction Trees). En ambos casos el principio de funcionamiento es similar, salvo que en el ID3 la variable de salida ha que ser numérica, y en el C4.5 requiere que sea categórica (Salzberg, 1994).

También a partir del ID3 se desarrolla la estrategia CART que cuenta con gran capacidad para detectar interdependencias entre parámetros y que a su vez puede usarlos más de una vez en diversas partes del modelo formando una estructura flexible y con gran capacidad.

Al ser un algoritmo no paramétrico, tiene la ventaja de no depender de ningún tipo de distribución estadística y con poca influencia de outliers, a pesar de que necesita más datos de entrenamiento respecto de sus homólogos (Nisbet et al., 2009).

Estructura

El nombre árbol de decisión proviene de la forma que adquiere en su representación, ya que es similar a la forma de un árbol en posición invertida. La construcción del árbol se basa en una división binaria recursiva que a medida que se va extendiendo en su desarrollo va perdiendo generalidad y ganando particularidad.

El clasificador utilizado CART es un árbol de decisión que se construye desde la raíz hasta las hojas respetando el principio de la Navaja de Ockham, es decir, dadas dos soluciones correctas para un determinado problema, debemos elegir la solución más simple (Dougherty, 2012).

Su estructura básica de un árbol CART está formada los elementos que se observan en la Figura 20, donde se observa que comenzando desde la Raíz o Nodo Inicial se establece la primera condición ($A > n_A$), donde A muestra el primer factor condicionante del conjunto de datos y n_A el valor numérico que puede tomar ese factor, de esta manera, una vez tomada la decisión se continúa hacia otros nodos a través de las ramas y así sucesivamente hasta que el un set de posibles respuestas desemboque en un nodo terminal. Por tanto, los elementos diferenciadores de esta técnica son:

Raíz o Nodo Inicial: Identifica la variable a clasificar y establece la condición de nodo de más alto nivel.

Ramas: Elementos de decisión que asocian nodos y hojas, dependientes del atributo condicionante.

Nodo: Identifica la variable a clasificar en base a unas decisiones tomadas en un nodo anterior y siempre le precede otro nodo.

Hoja o Nodo Terminal: Variable clasificada en base a una serie de atributos establecidos en nodos anteriores, carece de consecuentes.

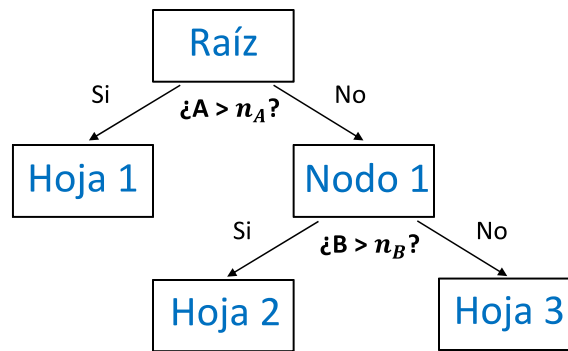


Figura 20: Elementos estructurales del algoritmo CART

Metodología de trabajo

El procedimiento de trabajo se inspira en la técnica de partición recursiva por la que se rige el CART (*Classification and Regression Tree*) y en la modelación aditiva generalizada (Hastie and Tibshirani, 1986) y que facilita en gran medida la búsqueda de transformaciones e interacciones de variables para interpretar pautas y relaciones que pueden quedar ocultos ante una compleja estructura de datos.

A través del conjunto de datos inicial, el algoritmo trabaja de manera inductiva para generar una estructura de reglas a partir de subconjuntos de casos extraídos del conjunto de datos de entrenamiento. A medida que va generando esta estructura el propio algoritmo comprueba la bondad de los datos mediante criterios capaces de proporcionar la precisión de la clasificación. De esta manera se realizan todas las pruebas posibles teniendo en cuenta las posibles divisiones del conjunto de datos y de todas ellas, el algoritmo seleccionará las pruebas que mayor ganancia de información proporcionen. También se analiza el conjunto de datos de entrenamiento para construir una herramienta que sea capaz de clasificar correctamente los ejemplos de entrenamiento y los de prueba.

Cada caso contiene una serie de valores que miden ciertas propiedades del sistema denominados atributos. Para los atributos discretos, el algoritmo considera una prueba con n resultados (siendo n el número de valores posibles), mientras que para los atributos continuos realiza una prueba binaria sobre cada valor que toma el atributo. Posteriormente es en cada nodo donde el sistema debe decidir que prueba realizar para continuar.

Este algoritmo también posee unos casos base, es decir, que no utiliza la recursión para producir una respuesta y son:

- Todas las muestras en la lista pertenecen a la misma clase, en este caso, el algoritmo crea un nodo de hoja para el árbol de decisión que indica optar por esta clase.
- Ninguno de los atributos proporciona ninguna ganancia de información. Llegado a este caso, el algoritmo crea un nodo de decisión más alto en el árbol, aplicando el valor esperado de la clase.
- Si se encuentra una clase no vista antes se crea un nodo hacia arriba del árbol usando el valor esperado.

Si no se cumple ninguno de estos, para cada atributo se calcula la ganancia de información, posteriormente el conjunto de datos se divide en subconjuntos nuevamente tomando el atributo que proporciones la menor ganancia de información, por último, se crea un nodo que contiene el atributo, de esta manera el proceso se repite con el resto de los atributos hasta que se llegue a uno de los casos base.

Una de las ventajas que proporciona este algoritmo es la posibilidad de “poda” una vez realizado, es decir, se puede disminuir en gran medida la cantidad de información se aporta capacidad al usuario de eliminar el set de respuestas de un nodo en base a convertir estos a hoja. Esta funcionalidad simplifica el árbol, teniendo en cuenta que en numerosas ocasiones usando árboles de gran tamaño no todas las ramas proporcionan información útil para el usuario y de esta manera es posible decidir por no ramificar más a partir de un punto para ganar simplicidad y teniendo en cuenta que se pierde un pequeño porcentaje de acierto (Maimon and Rokach, 2010).

6.6 NUBE DE ETIQUETAS

La extracción de información a través de los datos es una forma prometedora de crear valor y aumentar los recursos de investigación por ello el Big Data o análisis de ingentes cantidades de datos se ha posicionado como una de las técnicas más reconocidas para la extracción de conocimiento a través de los datos, aunque por el contra las técnicas aplicadas para ello son complejas y no siempre se genera un valor efectivo para los adoptadores, la nube de etiquetas ha surgido como herramienta para cubrir esa brecha por su simple uso y elevado potencial, ya que permite reducir un texto estructurado o desestructurado a las palabras más importantes de una manera gráfica (Elia et al., 2020).

La nube de etiquetas o nube de palabras es una técnica de visualización de datos, perteneciente a las técnicas de text mining no lingüísticas donde se representan visualmente los datos de un texto. A pesar de que suele utilizarse para representar metadatos de palabras clave de sitios web, también puede ser usada para visualizar texto en forma libre. Para ello se utilizan palabras sueltas y a partir de ellas se genera

una representación gráfica de una palabra sobre la base de las frecuencias de uso de la palabra en ese dominio. De tal manera que se puede interpretar como la prominencia de la palabra dentro de un dominio específico. Las nubes de palabras son una forma rápida de resumir grandes cantidades de datos (Kaptein, 2012), también tiene capacidad de representar la ideología respaldada por un determinado conjunto de datos.

Este característico tipo de visualización de datos puede ayudar a los interesados a realizar un análisis exploratorio de texto identificando los campos semánticos a través de las palabras que aparecen con más frecuencia para conocer el interés de un determinado grupo de personas.

Antes de proceder a la visualización de datos es necesario realizar un filtrado de palabras que no aporten información al conjunto de datos, de esta manera: artículos, pronombres, preposiciones, conectores..., y demás elementos destinados a dar coherencia a un texto serán eliminados, simplificando el conjunto de palabras a sólo aquellas que tengan capacidad de aportar información o puedan ser relacionadas mediante campos semánticos como: nombres, verbos, adjetivos, etc.

Para identificar las palabras más repetidas en un texto se visualizan por tamaños, así las palabras más frecuentes serán de mayor tamaño, a pesar de que es posible elegir el ratio de tamaños, generalmente el tamaño es proporcional al número de veces que aparece ese término, de esta manera siempre van a coexistir términos de gran tamaño y de pequeño tamaño con cierta gradación. Gradaciones pequeñas van a dar lugar a palabras muy representativas en la nube por su tamaño respecto del resto y gradaciones bajas van a dificultar más la extracción de términos más comunes.

Existe un número variado de formas de representar la nube de palabras sin que esto afecte al resultado final, eligiéndose por norma general la que resulte más compacta o atractiva visualmente. Un número excesivo de palabras generalmente provoca confusión, ya que las palabras de menor tamaño tienden a ocupar los huecos de las de mayor tamaño. Un número muy bajo de términos puede no ofrecer una visión objetiva de las palabras menos importantes, generalmente el número idóneo se sitúa entre 50 y 150 términos dependiendo del escalado de las palabras.

La Figura 21 muestra la nube de etiquetas extraída de la página web de ArcelorMittal (<https://spain.arcelormittal.com>) usando el programa online Nube de palabras (www.nubedepalabras.es), en el que se muestra las 123 palabras más representativas contenidas en la página web de ArcelorMittal España.

7 CARACTERIZACIÓN DEL ÉXITO EN LOS PROYECTOS

Con el objetivo de identificar el éxito en este tipo de proyectos se recurre a la adquisición de datos históricos de proyectos, con el precepto de que los datos estén perfectamente contrastados por otros autores y supongan un periodo representativo de este tipo de financiación.

La adquisición de datos históricos se centra en la fase de cierre de un proyecto, donde se debe de realizar un análisis retrospectivo para seleccionar que decisiones han sido cruciales, que acciones se han tomado para resolver el problema o cuales llevan el proyecto al éxito. Para ello se realiza un análisis a posteriori o AAR (After Action Review) de los resultados del proyecto, tanto si resultan exitosos como fracasados, siguiendo las directrices del Asociación de Gestión de Proyectos (PMI - Project Management Institute), para determinar las causas (PMBOK, 2013). Resulta crucial conocer las decisiones tomadas de cara a estandarizar las acciones que aseguren el éxito de los proyectos o al menos los alejen del fracaso.

Es posible extraer datos históricos directamente desde la plataforma hospedadora de proyectos mediante un programa de indexación de contenido Web, pero la configuración del propio programa en cada equipo y las modificaciones realizadas por la propia plataforma en ese transcurso temporal, abren la posibilidad de que los datos extraídos sean diferentes que los usados por otros autores, por ello el proceso de búsqueda del conjunto de datos ideal debe de ser riguroso para no perder fidelidad en la información y debe de estar apoyado en la capacidad para extraer conocimiento nuevo a la par que complementar el ya existente sobre el campo.

7.1 PROCESADO DEL CONJUNTO DE DATOS

Para la realización de este estudio se considera de especial interés dirigir la extracción de datos hacia las plataformas de crowdfunding, ya que son las primeras interesadas en el éxito de los proyectos, por ser sus ganancias proporcionales a la recaudación de los proyectos exitosos. Con la finalidad de aumentar el ratio de éxito, intentan evitar proyectos con ciertos riesgos asociados, realizando frecuentes cambios y actualizaciones de sus normas (Belleflamme et al., 2015).

Gran parte de los autores que han dirigido sus estudios hacia las plataformas de crowdfunding, lo han hecho para combinarlos con datos de redes sociales. En el artículo: *“Recommending Investors for Crowdfunding Projects”* se recogen datos durante tres meses de la plataforma Kickstarter para combinarlos con datos de Twitter y alcanzar un 84% de precisión para predecir potenciales inversores de Twitter (An et al., 2014). Hopp et al. combinan datos extraídos de la plataforma Kickstarter y datos de Facebook asociados a estos proyectos para descubrir como los eventos caracterizados como positivos, publicados en esta red social, influyen directamente sobre potenciales impulsores y sobre el éxito de los proyectos (Hopp et al., 2019).

También se han recogido datos a mano de una conocida plataforma de crowdfunding francesa *KissKissBankBank* para lograr evaluar cómo los factores de éxito varían a medida que avanza la campaña, también han encontrado paralelismo con las plataformas de crowdfunding americanas, las cuales parecen estar impulsadas por factores de éxito similares (Petitjean, 2018).

7.1.1 RECOLECCIÓN DE DATOS

La selección de la plataforma Kickstarter como objeto de estudio, se debe a su gran distribución y sus capacidades, su situación como dominante y que el resto de plataformas dirigen sus estrategias de benchmarking hacia esta, incluso Estados Unidos creó un proyecto de ley para regular el crowdfunding basándose en la plataforma Kickstarter (Franzen, 2012). Diez años después de su nacimiento sigue siendo uno de los portales de crowdfunding más importantes, ya que, está entre las 600 páginas Web más visitadas del mundo (Alexa, 2020).

El periodo de estudio se ha seleccionado basándose principalmente en dos factores: que esté bien contrastado por otros autores y además que sea muy representativo para el crowdfunding. Este periodo, comprende desde el inicio de financiación en la plataforma en el año 2009, hasta el año 2012 donde se produce una clara detención del número de proyectos que resultan exitosos. Muestra de ello es la Figura 22 donde se observa el aumento inicial del número de proyectos exitosos y fracasados desde el inicio, con un crecimiento sostenido que aumenta con el número de proyectos. En el año 2013 se detiene, y lejos de mantener la proporción éxito / fracaso y comienza a aumentar el número de proyectos fracasados sin que lo haga el de los exitosos, hasta el año 2015 donde se ubica el punto con menor ratio de éxito de toda la serie de datos recogidos (21,48% de proyectos exitosos). En 2018 aún se mantienen los valores de fracaso superiores a los del año 2013 como se observa en la Figura 22 (BIDAUX, 2019).

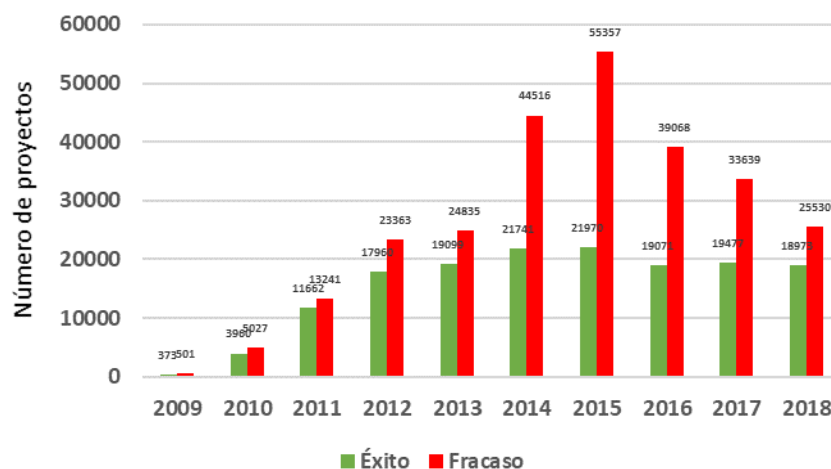


Figura 22: Número de proyectos exitosos y fracasados desde el inicio de Kickstarter (BIDAUX, 2019)

Un detalle a tener en cuenta es que el conjunto de proyectos de éxito y fracaso de cada año, no suman el total de proyectos presentados a la plataforma, sino que existe un

número relativamente elevado de proyectos que tras haber sido presentados a Kickstarter han sido rechazados por diversos motivos; Bien porque no han alcanzado las exigencias de la plataforma, han abandonado, han sido cancelados por el creador o han controlado mal las restricciones del proyecto: calidad, tiempo y costo (PMBOK, 2013).

Estos proyectos rechazados por la plataforma se pueden considerar como fracasados ya que no han conseguido financiación dentro del ciclo de vida de este proceso. Por tanto, teniendo en cuenta todos los proyectos presentados, el ratio de éxito para cada periodo temporal sería aún más bajo. A pesar de que actualmente Kickstarter guarda con gran recelo estos datos, declaró que para los años 2010 y 2011 se han presentado 11.130 y 27.086 proyectos respectivamente, en lugar de los 8.987 y 24.903 que han acabado la financiación con éxito o fracaso, esto supone un 16,15% y un 7,46% más sobre la suma de exitosos y fracasados (Kickstarter, PBC, 2012).

Para la adquisición de datos sobre los que trabajar, también se ha tenido en cuenta el rango del periodo de estudio para no perder fidelidad en los resultados. Los datos obtenidos durante periodos más largos de tiempo conllevan el problema de la evolución de las plataformas, tanto para adaptarse al marco legal, como para realizar numerosos cambios que permitan aumentar el número de proyectos de éxito y la sostenibilidad del modelo. De esta manera, Kickstarter modifica las categorías y subcategorías que pueden acoger un proyecto en el año 2014, añadiendo 2 nuevas categorías y 94 nuevas subcategorías (Nitsuh Abebe, 2014).

Para la realización de este trabajo se ha estudiado la posibilidad seleccionar un data set que sea los más representativo posible del periodo de estudio, para ello, además de contener las fechas clave que comprende la fase de crecimiento y ser contrastada por otros autores, ha de contener un elevado número de proyectos y debe de proporcionar variables numéricas que aporten suficiente información.

Uno de los data sets evaluados se ha obtenido a través de la Web de Webrobots. Este sitio web está dedicado al rastreo y control de “arañas Web” y proporciona un data set de Kickstarter en formato JSON. Esta extracción de datos se realiza con la aplicación de técnicas de *Web scraping*, consistentes en la extracción automatizada y metódica de información de sitios Web que simulan la navegación de una persona (Matsudaira, 2012), el programa que lo realiza, denominado *araña Web* extrae información de unos campos definidos y con una frecuencia planificada de antemano para crear indexarlos en un fichero. A pesar de la capacidad de extracción de datos que proporciona el algoritmo de la araña Web de Webrobots (28 variables por proyecto), se ha desechado la posibilidad de uso debido a que el periodo de recolección de datos comprende desde el año 2014 hasta el año 2019. Por otro lado, se ha visto que en este periodo de tiempo se produce un gran incremento de proyectos fracasados que podrían perturbar el resultado. La metodología de indexación no es la más apropiada para este tipo de estudio, ya que actualiza el data set cada vez que incorpora un elemento nuevo, modificando los datos ya indexados.

También se ha estudiado el uso de un data set extraído por Mickaël Mouillé para la web Kaggle.com, extraída usando *Python* como software. Después de su análisis se ha decidido que, a pesar de contener 13 variables, solamente 7 de ellas son atributos modelables, este número se ha considerado insuficiente para ampliar las capacidades de este estudio donde las técnicas a emplear tienen mucha capacidad para extraer información numérica a través de la interrelación de variables. No obstante, puede ser considerada de importancia para futuros usos.

El data set seleccionado finalmente ha sido adquirido usando una araña Web en este caso programada en lenguaje de PHP, valores separados por comas (.csv). Los datos se extrajeron de la web Kickstarter en un periodo temporal comprendido entre el 5 de marzo de 2009 al 8 de enero del 2012.

En la Tabla 3 se observan las principales variables analizadas de los 3 data sets con más potencial para este estudio.

Data set	Número de proyectos	Total de variables	Variables numéricas	Periodo
Mickaël Mouillé	378.660	15	9	2010 - 2017
Webrobots.com	145.627	28	13	2014 - 2019
Seleccionado	45.815	17	11	2009 - 2012

Tabla 3: Data sets comparados por sus variables más significativas

Este conjunto de datos seleccionado ya ha sido utilizado de manera parcial por diversos autores, y con diferentes enfoques. El profesor Mollick usó este periodo temporal de la misma plataforma para observar, entre otros, que la mayoría de los proyectos de éxito se entregan más tarde de lo previsto y que la geografía está vinculada con la naturaleza y la tasa de éxito de los proyectos (Mollick, 2014). Por otro lado Roma et al. también usaron el mismo periodo temporal para estudiar en este caso algunos factores influyen directamente sobre el éxito de los proyectos tecnológicos, demostrando que diversos factores como: la cantidad a conseguir, la existencia de patentes o una nutrida red de vínculos en redes sociales influyen positivamente en el éxito (Roma et al., 2017).

También se usó el data set de manera parcial para realizar un estudio desde las perspectiva de las redes sociales, demostrando que los creadores de proyectos de crowdfunding tienen mayores tasas de éxito si anteriormente han apoyado alguna campaña de financiación de este tipo (Zvilichovsky et al., 2013)

En el artículo “Identifying On-time Reward Delivery Projects with Estimating Delivery Duration on Kickstarter” los autores usan esta base de datos ampliada para la elaboración de un modelo predictivo para ayudar a conocer el estado de entrega de las recompensas (Tran et al., 2017).

La flexibilidad de uso de este conjunto de datos y la complementariedad con las soluciones obtenidas por otros autores, hacen del conjunto de datos seleccionado como la opción más apropiada para este tipo de estudio. La gran cantidad de variables numéricas y la calidad de los datos resulta ser una ventaja competitiva sobre los otros conjuntos de datos, los cuales necesitan más estructuración en su contenido tanto por ser una base de datos “viva” en el caso del archivo JSON, donde un proyecto puede cambiar de estado durante la recolección de datos, como con la dificultad de extraer información asociada al elevado número de variables no numéricas, en el caso del conjunto de datos de Mickaël Mouillé.

Por lo cual el conjunto de datos seleccionado se muestra como la mejor alternativa de cara a la naturaleza de este estudio.

7.1.2 DESCRIPCIÓN DE LOS DATOS

Los datos extraídos se encuentran encabezados por un número identificador único para cada proyecto y a su vez, cada uno está caracterizado en base a una de las 12 categorías posibles: Art, Comics, Dance, Design, Fashion, Food, Games, Music, Photography, Publishing, Technology and Theater (Pi, 2012).

7.1.2.1 Descripción de las variables

El data set extraído contiene 17 variables, 9 de ellas son cualitativas y 8 son cuantitativas como se observa en la Tabla 4 de caracterización de variables.

Variable	Tipo	Rango	Unidad
Project id	Cualitativa Nominal	39409 – 2147460119	adimensional
Name	Cualitativa Nominal	-	-
URL	Cualitativa Nominal	-	-
Category	Cualitativa Nominal	-	-
Subcategory	Cualitativa Nominal	-	-
Location	Cualitativa Nominal	-	-
Status	Cualitativa Politémica	-	-
Goal	Cuantitativa discreta	0,01 – 5250000	\$
Pledged	Cuantitativa discreta	0 – 10266845	\$
Funded percentage	Cuantitativa continua	0 – 10000	%
Backers	Cuantitativa discreta	0 – 87142	adimensional
Funded date	Cualitativa Nominal	5/3/09 – 8/1/12	d/m/a

Variable	Tipo	Rango	Unidad
Levels	Cuantitativa discreta	0 - 80	adimensional
Reward levels	Cualitativa Politémica	0 - 999	\$
Updates	Cuantitativa discreta	0 - 142	adimensional
Comments	Cuantitativa discreta	0 – 19311	adimensional
Duration	Cuantitativa continua	0 – 91,96	Días

Tabla 4: Caracterización de variables

Project ID: (Número de identificación único de cada proyecto), No tiene transcendencia desde el punto de vista de este estudio por ser un atributo equivalente al nombre o a la descripción del proyecto. Este número identificativo no sigue ninguna correlación con la fecha del proyecto, aunque resulta de gran comodidad para buscar un proyecto por su número ya que solamente contiene caracteres numéricos.

Name: (Nombre del proyecto), es el que da título a la campaña de financiación, puede ser un nombre propio, siglas o acrónimos e incluso pueden incluir una breve descripción del proyecto generalmente en una o dos líneas.

URL: (Dirección de hospedaje en la Web). Todos los proyectos están albergados en la plataforma por lo que es posible acceder directamente a un proyecto determinado a través de este parámetro, es un dato único de cada proyecto al igual que Project ID y Name. En este caso siempre comienza por: “<http://www.kickstarter.com/projects/...>”.

Category: (Categoría). Dentro de la plataforma los proyectos se encuentran agrupados por categorías donde comparten la misma naturaleza del proyecto, el fin principal de la categoría es permitir a un usuario de la plataforma o un patrocinador potencial dirigir su interés hacia un determinado tipo de proyectos, de la manera más rápida e intuitiva posible, existen 12 categorías: Art, Comics, Dance, Design, Fashion, Food, Games, Music, Photography, Publishing, Technology and Theater.

Subcategory: (Subcategoría). Es un subgrupo de Category formado por 30 elementos que permite un mayor conocimiento del tipo del proyecto a través de una subclasificación según la tipología del proyecto. De las Categorías mostradas, solamente 6 de ellas tienen subcategorías: Art, Design, Games, Music, Publishing y Technology. Resulta ser de gran ayuda para consultar proyectos de una naturaleza muy concreta. Cabe destacar que estas categorías también contienen una subcategoría del mismo nombre que la categoría y se encuentran distribuidas como se muestra a continuación en la Tabla 5.

Art	Design	Games	Music	Publishing	Technology
Art	Design	Games	Music	Publishing	Technology

Art	Design	Games	Music	Publishing	Technology
Conceptual Art	Crafts	Board	Classical Music	Art Book	Open Hardware
Digital Art	Graphic Design	Video Games	Country	Children's Book	Open Software
Illustration	Product Design		Electronic Music	Fiction	
Mixed Media			Hip-Hop	Journalism	
Painting			Indie Rock	Nonfiction	
Performance Art			Jazz	Periodical	
Public Art			Pop	Poetry	
Sculpture			Rock		
			World Music		

Tabla 5: Distribución de Subcategorías por Categoría

Location: (Localización). País o estado donde se encuentra el lugar de realización o ubicación del proyecto. Es un parámetro importante ya que el país donde se realiza el proyecto puede condicionar su éxito dentro o fuera de el mismo. Los impuestos y tasas de envío son capaces de afectar al éxito del proyecto, ya que cualquier país del mundo puede desarrollar proyectos en este portal.

Status: (Estado). Es uno de los factores más importantes a considerar en el estudio de un data set de este tipo de proyectos, ya que, con él se identifica el estado de la campaña de la financiación en el momento de la recogida de datos. Existen cinco posibilidades de esta variable:

- **Canceled:** Referido a proyectos cancelados por el creador. El indebido control de las limitaciones del proyecto puede llevar a su cancelación por parte del creador si este cree que no va a conseguir desarrollar el proyecto manteniendo la calidad, el coste y durante el tiempo necesario.
- **Failed:** Son proyectos que han completado el periodo de financiación en la plataforma y no han llegado a recaudar la cantidad mínima para su desarrollo (*Goal*). Estos proyectos al no ser desarrollados no obtienen ninguna cuantía económica y por tanto a los patrocinadores se le será devuelta la cantidad aportada.
- **Live:** Proyectos que no han acabado el periodo de financiación durante la fase de recogida de datos en la plataforma. No es un factor determinante para caracterizar ninguna una campaña de financiación, ya que el estado "Live" lo determina el algoritmo de la araña Web usada.

- **Successful:** Referido a proyectos que han acabado su financiación de manera exitosa, es decir, han alcanzado su meta (*Goal*) o la han superado. Estos proyectos permanecen aún presentados en la plataforma y quedan pendientes de envío.
- **Suspended:** Proyectos suspendidos por la plataforma, bien por no seguir sus normas de conducta o por considerarse fraudulentos. Apoyos o comentarios recibidos a una campaña por parte del creador o bien la presentación de un producto ajeno como propio, son causas comunes de suspensión de un proyecto.

Goal: (Meta) Cantidad mínima de recaudación para comenzar el proyecto expresada en dólares. Esta cantidad económica debe de ser puesta de antemano por el creador y es determinante para que el proyecto se lleve a cabo. La importancia en su elección es máxima ya que por otro lado no es limitante, es decir, una vez alcanzada si aún existe tiempo de financiación, esta puede verse superada de manera indeterminada dependiendo de la demanda del producto, por ello es crucial que el creador tenga en cuenta que a pesar de que puede ser muy beneficioso la superación del *Goal*, este tiene que ser capaz de desarrollar el proyecto para un número mayor de aportadores.

Pledged: (Prometido) Cantidad máxima alcanzada expresada en dólares en el momento de la recogida de datos. Cabe destacar que, *Pledged* se refiere a dinero virtual recaudado y retenido por la plataforma hasta que acabe la campaña de financiación, momento en el que si ha alcanzado o superado el *Goal* será entregado al creador para el desarrollo de su proyecto. Al contrario de lo que pueda parecer, una cuantía de *Pledged* elevada no aporta ninguna información esencial ya que depende directamente de la meta de financiación.

Funded Percentage: (Porcentaje de financiación) Porcentaje de superación del *Goal*, resultado de dividir *Pledged* entre *Goal* y expresado en tanto por uno, es por tanto una variable dependiente, la única del data set. Aporta información acerca de las veces que se superó el *Goal* e indica la relación entre ambas variables, aunque no haya terminado el periodo de financiación, de esta manera, un proyecto fracasado tendrá un *Funded Percentage* inferior a cero. Este parámetro además de aportar información de éxito o fracaso para cada proyecto también caracteriza los proyectos (de manera cualitativa) de alto éxito, dependiendo de las veces que este parámetro supere a la variable *Goal*.

Backers: (Aportadores) Numero de personas que contribuyen a la realización del proyecto. Cada aportador es único y tiene un perfil creado en la plataforma donde hará una selección entre los productos (recompensas) ofrecidos por el creador, una vez elegidos (tipo y cantidad) y seleccionados se le retendrá la cantidad de dinero correspondiente a la recompensa y a las comisiones de método de pago, con un compromiso de pago cuando acabe el periodo de financiación si esta resulta exitosa. Existe la posibilidad de que un *Backer* retire su contribución si lo desea.

Funded Date: (Fecha fin de la Financiación) Fecha de finalización de la campaña de financiación. Una vez concluida no es posible realizar más aportaciones a la campaña y el dinero comprometido será entregado al creador si al menos ha alcanzado el *Goal*. Esta fecha es seleccionada por el creador al inicio del proyecto por lo que debe de ser cautelosamente elegida ya que condiciona que una campaña pueda seguir recaudando dinero. Desde la fecha de inicio de campaña hasta esta no pueden pasar más de 60 días.

Levels: (Número de niveles) Número de niveles recompensas ofertadas por el creador. A pesar de que es un parámetro que se fija de antemano, puede ser modificado durante la campaña siempre y cuando sea beneficioso para el *Backer*, es decir, la recompensas ofrezcan más cantidad o calidad por la misma cuantía económica o bien exista la posibilidad de acceder a nuevas recompensas. El número de *Levels* necesario es indeterminado, aunque debe de tenerse en cuenta que un correcto escalado facilita la selección de la recompensa adecuada a cada aportador, desde las más bajas a las más altas.

Reward Levels: (Niveles de recompensa) Indica la cuantía económica de cada nivel de recompensa ofrecido por el creador, está expresado en dólares y sigue un orden creciente. Por lo general cuanto mayor sea el *Goal* mayor número de niveles de recompensa nos encontramos. El nivel de recompensa más frecuente es un donativo al proyecto, generalmente de baja cuantía.

Updates: (Actualizaciones) Comunicación de los avances del proyecto. Cualquier cambio en el proyecto puede ser comunicado mediante una actualización de la campaña de crowdfunding. Indica en que parte del ciclo de vida del proyecto se encuentra y complementa la información de la descripción inicial del proyecto. También llaman la atención sobre otros proyectos ya que indica que el proyecto se encuentra vivo.

Comments: (Comentarios) Número de comentarios realizados para cada proyecto. Dentro de la plataforma, la sección *Comments* deja espacio para la comunicación bidireccional creador-colaborador en la que se establece un canal para resolver dudas sobre el desarrollo del proyecto, las técnicas empleadas o cualquier otro dato relacionado con el proyecto a la vez que aporta feedback al creador.

Duration: (Duración) Duración máxima de la campaña de recaudación, a partir de esta fecha no es posible realizar más contribuciones al proyecto y si no se ha alcanzado el *Goal* el proyecto se queda sin financiación. A pesar de que sería posible establecer tiempos superiores Kickstarter pone un máximo de tiempo de 60 días y añade que las campañas con mayor probabilidad de éxito son de 30 días.

7.1.2.2 Dificultades detectadas en el data set

El conjunto de datos necesita ser procesado para mejorar la calidad de los datos desde la perspectiva del modelado, ya que muchas celdas se encuentran con el contenido desajustado y no todas las celdas tienen contenido (*missing data*). Del total de celdas

(778.855), solamente contienen información 514.833, esto es un total del 66,10% de celdas con información. Desde este punto de partida no se realiza ningún tipo de operación ya que el data set no está compensado, esta distribución provoca desequilibrios al comparar variables con diferente número de datos, lo que implica que los algoritmos aplicados al data set tienden a favorecer la clase con la mayor proporción de observaciones (conocida como clase mayoritaria), lo cual puede derivar en métricas de exactitud sesgadas. El uso de este tipo de datos no equilibrados es muy común en otras áreas de estudio como en medicina, en muchos casos se parte de una base de datos de este tipo para aplicar un algoritmo determinado y llegar a unos datos suficientemente representativos y equilibrados (Belarouci and Chikh, 2017), posteriormente su estudio no supone ningún problema de cara a perder fidelidad en los resultados. En la Tabla 6 se observan más detalladamente el número de valores vacías de cada variable, así como el porcentaje de contenido que la representa.

Variable	Missing data	Contenido	Variable	Missing data	Contenido
Project ID	0	100,00%	Funded percentage	19.359	57,75%
Name	0	100,00%	Backers	19.359	57,75%
URL	6.465	85,89%	Funded date	19.359	57,75%
Category	6.465	85,89%	Levels	19.359	57,75%
Subcategory	18.028	60,65%	Reward levels	19.398	57,66%
Location	20.066	56,20%	Updates	19.359	57,75%
Status	19.359	57,75%	Comments	19.359	57,75%
Goal	19.359	57,75%	Duration	19.359	57,75%
Pledged	19.369	57,72%			

Tabla 6: Número de datos de cada variable en el data set

Otro de los inconvenientes asociados a estos datos son los valores atípicos (*outliers*) tanto por ser demasiados bajos como demasiado elevados. Valores que no son representativos de la población y que deben de ser diferenciados convenientemente ya que pueden distorsionar seriamente el comportamiento de los contrastes estadísticos. Un ejemplo de ello es la financiación de Pebble, un proyecto consistente en un smartwatch con una meta de financiación (Goal) de 100.000\$ y que finalmente logró alcanzar una cantidad final (Pledged) de 102.668.45\$ (Kickstarter, 2012).

A pesar de que la variable Status se considera como la variable de salida más importante, la mayoría de los autores toman de ella solamente los proyectos exitosos (Schwienbacher and Larralde, 2010) y (Mollick, 2014), los cinco estados posibles que puede adoptar y las diferencias entre ellos supone la realización de una recategorización

de la variable Status. En la Tabla 7 se observa las diferencias citadas entre los diferentes modos de Status.

	Successful	Failed	Cancelled	Suspended	Live	Vacías
Número de celdas	13.233	10.707	29	3	2.481	19.362
Porcentaje	28,88%	23,37%	0,063%	0,0065%	5,41%	42.26%

Tabla 7: Caracterización de la variable Status

7.1.3 ADAPTACIÓN DE LOS DATOS

El conjunto de datos obtenido en crudo se encuentra en formato desestructurado, así que se realiza la conversión de los datos a columnas para dar lugar a un nuevo formato más comprensible y compatible con el procesamiento de dichos datos. Finalmente resulta un espacio de trabajo distribuido en una hoja de cálculo de 45.815 filas que representan a cada uno de los proyectos y 17 columnas que caracterizan los atributos asociados a cada proyecto, proporcionando tanto información descriptiva como del estado de la financiación.

7.1.3.1 Estructura de los datos

Para tratar los datos de las celdas vacías se ha llevado a cabo un criterio de categorización asociado a: celdas vacías predecibles, no predecibles, prescindibles y no prescindibles. En los casos donde la variable ausente es dependiente, se ha completado el contenido de la celda asociándolo a la variable independiente con el algoritmo necesario. Así, por ejemplo, el contenido de una celda de la variable Goal, se puede predecir a través de la variable Pledged y de la variable Funded Percentage, ya que se conoce que Funded Percentage es el número de veces que Pledged supera a la variable Goal para un proyecto determinado.

En los casos en los que la variable no es predecible, por ser una variable independiente, se ha recurrido a comprobar si es prescindible o imprescindible.

Se consideran variables prescindibles a: Project Id, Name, Url, Location y Subcategory ya que no representan ningún problema de cara a la clasificación o identificación de un proyecto, por ello no supone ningún problema asociar la casilla vacía a cualquier tipo de referencia identificativa. Sin embargo, para el caso de los datos ausentes que suponen un dato imprescindible, se ha decidido la eliminación del proyecto completo para no perder fidelidad y calidad en los datos.

Se ha priorizado el uso de la variable *Category* sobre *Subcategory* ya que la variable *Category* define de manera más genérica la naturaleza de este tipo de proyectos y permite identificar directamente a que sector va asociado.

En la Tabla 8 se muestra el contenido de proyectos de cada subcategoría, así como la categoría a la que pertenece, cabe destacar el valor más elevado en la subcategoría

Music (categoría Music) con 2.649 proyectos y el más bajo con 132 proyectos para Conceptual Art (Art).

Category	Subcategory	Proyectos	Category	Subcategory	Proyectos
Art	Art	959	Music	Classical Music	416
	Conceptual Art	160		Country	811
	Digital Art	132		Electronic Music	240
	Illustration	176		Hip-Hop	346
	Mixed Media	401		Indie Rock	1508
	Painting	465		Jazz	321
	Performance Art	436		Music	2649
	Public Art	529		Pop	571
	Sculpture	321		Rock	1473
					World Music
Design	Crafts	230	Publishing	Art Book	294
	Design	257		Children's Book	573
	Graphic Design	156		Fiction	961
	Product Design	966		Journalism	397
Games	Board & Games	520		Nonfiction	903
	Games	254		Periodical	271
	Video Games	857		Poetry	187
Technology	Open Hardware	175		Publishing	585
	Open Software	244			
	Technology	348			

Tabla 8: Contenido de proyectos en cada subcategoría

Observando los datos de la variable *Location* en el total de proyectos del data set (45.815), un 52,42% del total de proyectos, es decir, 23.560 proyectos tienen como país de origen Estados Unidos, no es de extrañar ya que el origen de la propia plataforma Kickstarter también lo es. Se han realizado estudios previos sobre la influencia del lugar de realización de los proyectos donde se ha demostrado que algunos lugares geográficos permanecen asociados a una categoría predominante (Mollick, 2014).

La variable *Name* (nombre del proyecto) no se considera imprescindible ya que no aporta información acerca de la calidad o los factores más importante de los proyectos, a pesar de que también ha sido objeto de estudio para determinar su influencia sobre el éxito de los proyectos. Wang et al. realizaron un análisis de sentimiento y comprobaron que la orientación en la descripción del proyecto hacia aspectos positivos o está fuertemente asociado con el éxito (Wang et al., 2017).

El nombre del proyecto proporciona cierta información a un potencial colaborador, por ello, este campo también ha sido objeto de estudio utilizando técnicas de text mining de carácter exploratorio. En este caso se realiza un estudio exploratorio sobre la frecuencia de repetición de términos que aparecen en el nombre de los proyectos, se han seleccionado los diez términos más repetidos como se observa en la Figura 23.

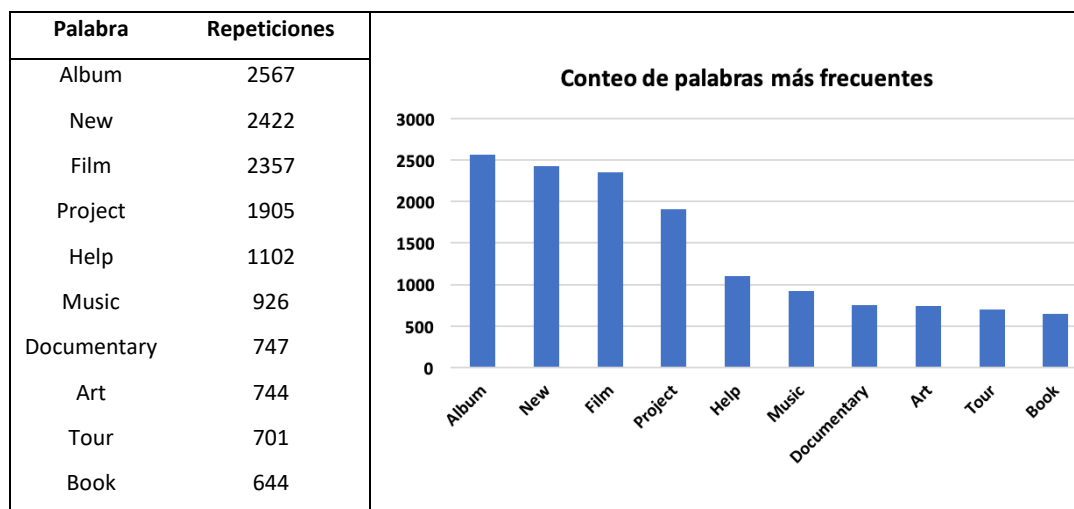


Figura 23: Conteo de palabras más frecuentes

Posteriormente se representan mediante la nube de palabras donde el tamaño se asocia al número de veces que se repite, se han filtrado las palabras que no aportan información, como conectores, artículos, etc. Para hacer más sencilla la visualización se ha decidido reducir la nube de palabras a los 100 términos más repetidos como se observa en la Figura 24. Para realizar este análisis gráfico de los resultados se utilizó la aplicación en línea Wordle (Feinberg, 2014), con la que también se ha decidido dotar a las palabras de diferentes tonalidades azules aleatoriamente para facilitar la visualización.

cancelados por la plataforma debido a una irregularidad detectada posterior a la recogida de datos, por tanto, los proyectos vivos se descartan porque son susceptibles de alterar la fidelidad de los resultados finales.

El lugar donde se inicia el proyecto no se considera relevante para este estudio ya que está enfocado al proyecto en sí y sus características, asumiendo que las características culturales y socioeconómicas de cada país se diluyen con la globalización de la red. Por tanto, los datos faltantes en las celdas de la variable *Location* se consideran prescindibles. En el artículo “The Economic Geography of Civic Crowdfunding” se evalúa empíricamente cómo en una plataforma de crowdfunding de financiación pública se ve afectada en la recaudación de capital dependiendo de la distancia al proyecto (Brent and Lorah, 2019), descartando la demografía de un vecindario como razón para alterar la capacidad de recaudar fondos.

7.1.3.2 Modificaciones en las Variables

De las 17 variables de las que se componía el data set nativo se han seleccionado para continuar el estudio: *Goal*, *Pledged*, *Funded percentage*, *Backers*, *Status*, *Category*, *Levels*, *Updates*, *Comments* y *Duration*. La variable *Funded percentage* se ha renombrado como *Rate_Pledged_Goal*, se ha adaptado el contenido a 2 cifras decimales ya que se considera que es el más adecuado para las escalas numéricas contenidas en el conjunto de datos.

Por otro lado, para aumentar la capacidad multivariante de este trabajo, se han modificados algunos atributos y se han creado otros derivados de los anteriores, con el objetivo de aportar una mejor comprensión del modelo. Así se ha modificado el contenido de *Reward Levels* y se ha creado nuevas variables: *Pledged_Backer*, *Min_RL*, *Max_RL*, *Range_RL*. La terminación RL hace referencia a *Reward Levels*.

La variable *Reward Levels* posee en cada caso el contenido de las distintas aportaciones posibles a cada proyecto y ahora la nueva variable *Levels* solamente cuantifica en cada caso el valor de los niveles, así, por ejemplo, un proyecto con valores de *Reward Levels* da de: “\$1, \$25, \$75, \$100”, pasará ahora a tener un contenido del número aportaciones posibles, es decir, 4, con el objetivo de realizar una evaluación de la influencia del número de niveles de recompensa.

A partir de la variable *Levels* también se han obtenido otras tres nuevas variables. *Min_RL* la cual hace referencia al mínimo valor de contribución posible para cada proyecto, *Max_RL* hace se refiere al máximo valor de contribución posible para cada proyecto y *Range_RL* es la diferencia entre *Max_RL* y *Min_RL*.

La variable *Pledged_Backer* se encuentra referida a la cantidad media aportada por cada contribuidor para llegar a Pledged final del proyecto. Es decir, esta variable resulta de dividir la variable *Pledged* entre la variable *Backer* para cada proyecto.

7.1.4 EXPLORACIÓN DE LOS DATOS

El data set finalmente obtenido después del proceso de limpieza, balanceado de datos y ajustes pertenecientes a la fase de análisis de datos de CRISP-DM, se reduce a 23.941 proyectos, número suficientemente representativo para aplicar técnicas estadísticas multivariantes. El ratio de éxito obtenido es de 55,28%, es decir, 13.233 proyectos resultaron exitosos.

En la Figura 25 se observa la distribución de los proyectos según su acumulación de capital (Pledged) y su Goal (ambos expresados en miles de \$ y escala logarítmica), de ahí que la recta que separa éxito de fracaso en lugar de ser horizontal aparezca con una inclinación de 45°. Para mejorar aún más la observación de los datos se ha usado una escala de reducción a través de la opacidad de los puntos con un factor de 1/100. En el gráfico se diferencian los proyectos exitosos en color cian y los fracasados en color rojo. La línea imaginaria que separa ambos colores pertenece a los puntos donde el Pledged y el Goal toman el mismo valor, es decir, donde los proyectos resultan exitosos con el mínimo margen, de ahí que sólo la ocupen proyectos de éxito.

La dispersión de éxito y fracaso de los proyectos apoya la explicación del profesor Mollick, dado que los proyectos que fracasan lo hacen con un margen muy amplio y los que resultan exitosos lo hacen con un margen más reducido, en línea con las observaciones del profesor Ethan Mollick, el centroide que representa el conjunto de los proyectos fracasados se encuentra más alejado de la línea de éxito-fracaso, sin embargo en el caso de los exitosos, los proyectos se encuentran distribuidos de manera lineal y paralelamente a la línea de éxito, de tal manera que su centroide se encuentra cercano a la línea éxito-fracaso (Mollick, 2014).

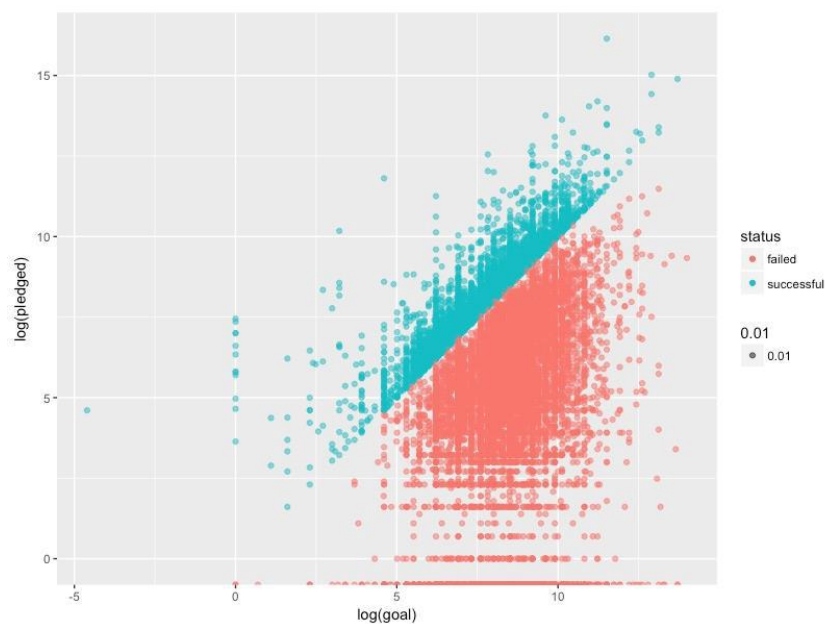


Figura 25: Gráfica Goal vs Pledged

Todos los proyectos del conjunto de datos se encuentran definidos por categorías y caracterizados principalmente según se observa en la Tabla 9. Cabe resaltar la categoría *Music* por su elevado número de proyectos (7238) y el Ratio de Éxito de la categoría *Dance* (72,71%).

Category	Número de Proyectos	Ratio de Éxito	Número de Aportadores
Art	3.304	57,35%	132.920
Comics	868	53,69%	77.752
Dance	590	72,71%	21.922
Design	1.416	47,81%	317.998
Fashion	927	33,55%	29.437
Food	1.167	51,67%	69.866
Games	929	36,49%	379.453
Music	7.238	66,37%	387.692
Photography	1.266	45,97%	39.479
Publishing	3.631	41,23%	145.083
Technology	700	39,71%	97.696
Theater	1.905	71,02%	78.703

Tabla 9: Principales características de las categorías

Cada una de las Categorías define el proyecto por sus principales características y por los valores de sus variables más representativas. A pesar del amplio rango de valores de algunas variables se ha decidido mantener los outliers tanto altos como bajos para contemplar la posibilidad de que puedan formar un grupo a parte, suficientemente representativo.

A pesar de que ciertas variables son definidas por el creador antes de presentar el proyecto a la plataforma, otras pueden ser modificadas a medida que se desarrolla la financiación del proyecto, incluso en algunos casos puede ser modificadas a criterio del creador siempre y cuando supongan una mejora para los contribuidores y sea admitida por la propia plataforma. Finalmente, el impacto de las variables restantes, las de modificación más indirecta por parte del creador, son consecuencia de las anteriores y ayudan a describir cuantitativamente el éxito o fracaso de un proyecto. En la Tabla 10 se observan todas las variables sometidas a estudio, tanto las nativas del data set como las de nueva creación.

Variable	Rango	Media	Desviación
Goal	0,01-1.200.000	7.746,54	24.917,23
Pledged	0-10.266.845	5.326,50	76.966,38
Rate_Pledge_Goal*	0-10.000	1,80	68,46
Bakers	0-87.142	79,90	931,21
Levels	0-80	7,82	4,24
Updates	0-142	4,20	6,37
Comments	0-19.311	10,46	223,98
Month	1-12	6,01	3,32
Duration	1-92	39,73	17,56
Min_RL*	0-5.000	8,32	43,78
Max_RL*	0-10.000	1.587,40	2.409,29
Range_RL*	0-9.999	1.579,08	2.409,26
Pledged_Backer*	0-3.757	60,73	75,18

**Variables de nueva creación*

Tabla 10: Variables del data set usadas para el estudio.

Atendiendo al valor que toman las variables durante el periodo de financiación se distinguen tres tipos:

- **Variables fijas:** El creador del proyecto debe asignar su valor antes del periodo de financiación: Goal, Month, Levels, Duration, Min_RL, Max_RL, Range_RL.
- **Variables de desarrollo:** Sólo se pueden modificar con los esfuerzos del creador durante el desarrollo de la campaña: Comments, Updates y Backers.
- **Variables indicador:** Son dependientes de las dos anteriores y dan una idea del impacto final de la financiación: Pledged, Rate_Pledge_Goal, Pledged_Backer.

7.1.4.1 Correlación de variables

En la Tabla 11 se muestra la correlación de variables pertenecientes a este estudio, donde se observa la elevada correlación lineal entre las variables: Backers, Pledged y Comments, esta relación ya ha sido observada anteriormente aunque de manera parcial (Lu et al., 2014) y (Wang et al., 2017). La aparente correlación total entre las variables Range_RL y Max_RL y la inexistente con Rate_Pledge_Goal y el resto de las variables, denota la necesidad de uso de técnicas más avanzadas que aporten más información, por lo que serán objeto de estudio posterior con el uso de técnicas multivariantes.

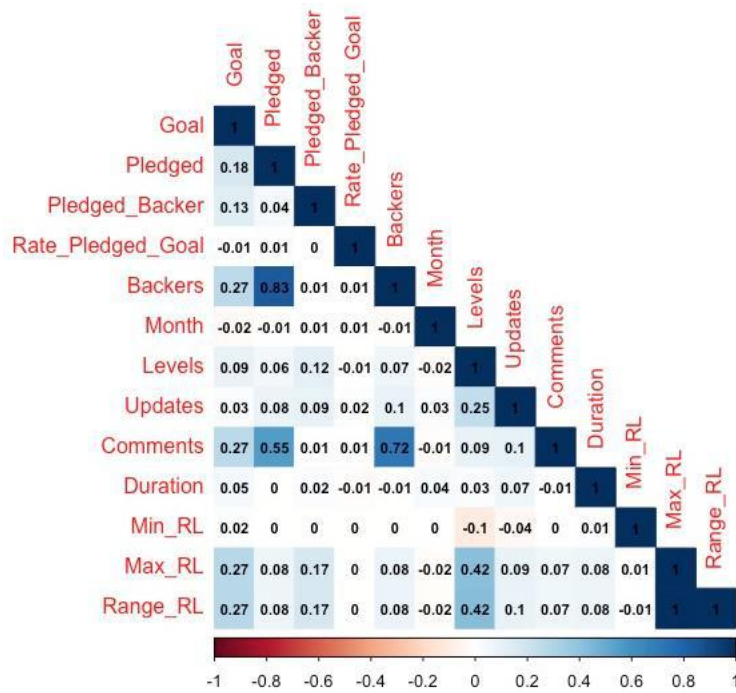
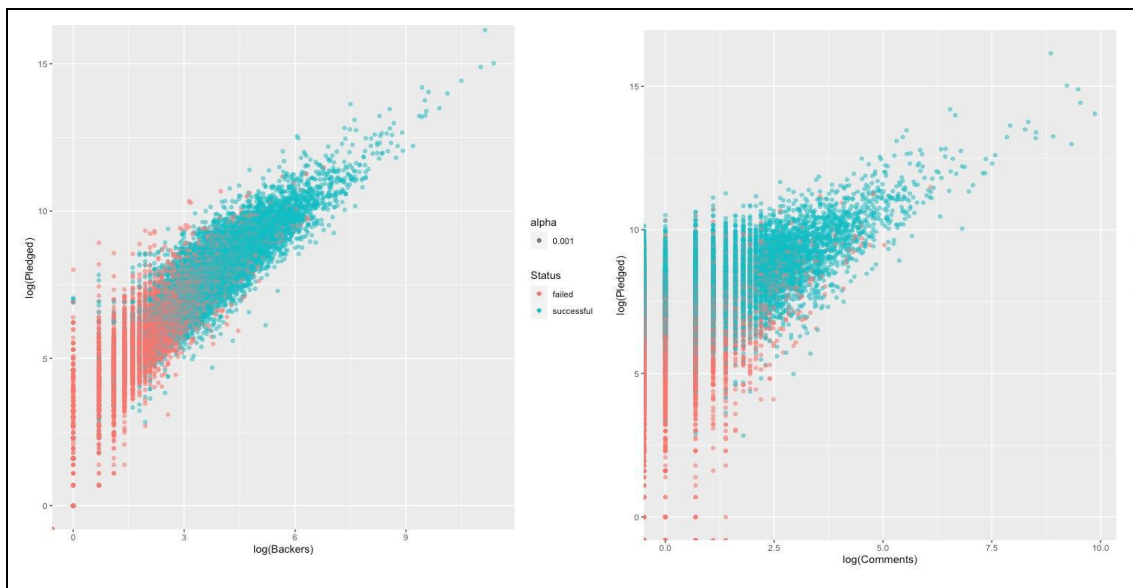


Tabla 11: Correlación de variables

A partir de la tabla de correlación de variables del apartado anterior, se han elegido las cuatro correlaciones más importantes para graficarlas y facilitar visualmente su comprensión, para ello, se han enfrentado: Backers vs Pledged (correlación 0,83), Backers vs Comments (0,72), Levels vs Max_RL (0,42) y Pledged vs Comments (0,55). Con el fin de evitar la saturación de datos en la visualización se ha seleccionado de color cian los proyectos exitosos y de rojo los que han resultado fallidos, también se selecciona un factor de reducción de 1/1000 y reescalado los ejes a escala logarítmica. El resultado de este enfrentamiento por parejas de variables se observa en la Figura 26.



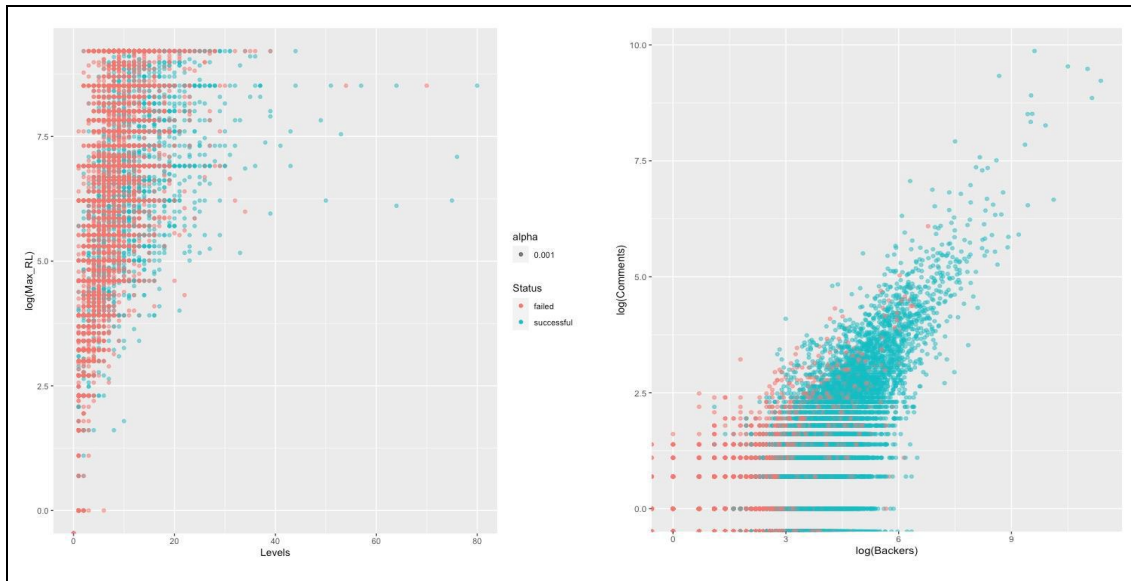


Figura 26: Enfrentamiento de variables con comportamiento particular

Las cuatro parejas de variables muestran una clara correlación positiva, donde a medida que aumentan los valores de ambos ejes, también aumenta el desorden en la linealidad de los datos. Al enfrentar las variables Backers vs Pledged se observa una linealidad muy marcada y con una tendencia muy acusada a medida que las variables van aumentando su valor. En principio, resulta lógico pensar que están asociados altos valores de recaudación en el proyecto (Pledged) con una cantidad elevada de colaboradores (Backers). En el segundo caso, Comments vs Pledged, se parte de una mayor varianza de valores en los estadios más bajos de Pledged antes de alcanzar cierta tendencia, a priori resulta difícil relacionar un número bajo de comentarios con la cantidad máxima alcanzada. Ya que se observa que los proyectos con pocos comentarios también tienen capacidad de alcanzar altos niveles de Pledged.

Por otro lado, al enfrentar Levels vs Max_RL se observa que la variable Max_RL presenta una elevada sensibilidad respecto de Levels, es decir, incrementos muy pequeños en la variable Levels provoca incrementos elevados en la variable Max_RL, sin poder precisar niveles de ambas óptimos para el éxito o fracaso de los proyectos. Aunque, de un primer vistazo parece lógico pensar que para que un proyecto pueda ofrecer recompensas de elevado valor, debe de hacerse ofreciendo un escalado desde la mínima hasta la máxima recompensa.

Finalmente, al contrastar Backers vs Comments, se observa que a pesar de su tendencia positiva en ambas variables tiene, comportamiento muy similar a la gráfica Comments vs Pledged, diferenciándose de esta última por la variabilidad y dispersión final de los datos, que provoca la difícil interpretación de la tendencia para los valores medios y altos de ambas variables.

A pesar de que hay variables con un índice de correlación muy bajo cabe mencionar que al enfrentar algunas de ellas tienen un comportamiento característico, que puede resultar de gran ayuda para comprender mejor la correspondencia entre unas y otras.

Hay dos casos observados en los que no existe mucha relación entre cada par de variables, pero si tienen un comportamiento representativo si se separa éxito de fracaso para cada una de ellas.

Por otra parte, en la Figura 27 se observa que las parejas de variables: Updates vs Rate_Pledged_Goal y Goal vs Pledged (ya observada en la Figura 25), proporcionan información sobre la evolución del éxito de manera diferente. En el primer caso, a pesar de que la variable Updates se mueve en valores con un dominio menor que Rate_Pledged_Goal, los proyectos fallidos que se encuentran muy cerca de ser financiados tienen más Updates que los que han quedado lejos de la financiación. Por otro lado, en los proyectos de éxito no parece que afecte en gran medida el número de Updates, existe a priori cierta correspondencia entre bajos valores de Rate_Pledged_Goal y una elevada amplitud de Updates.

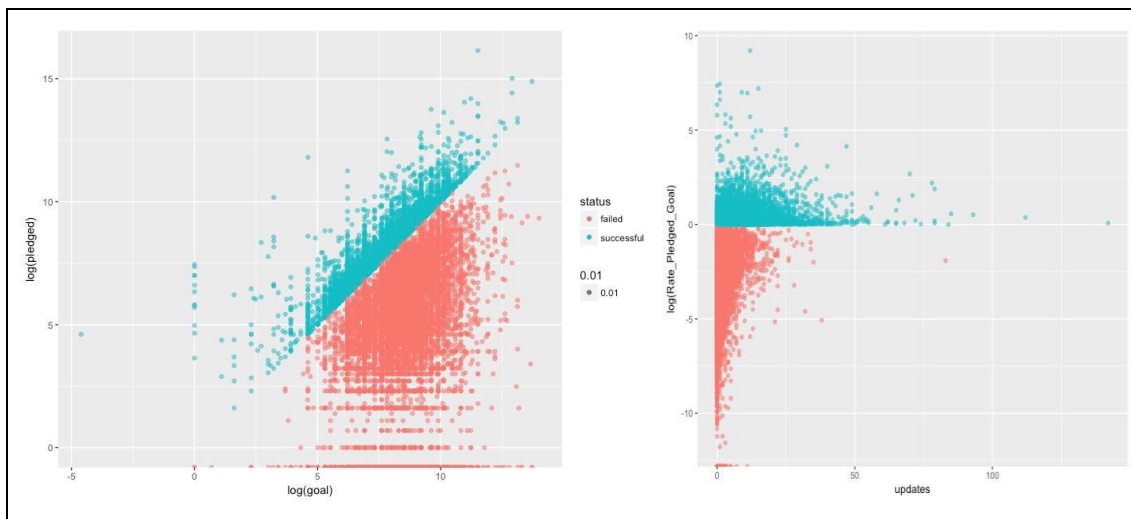


Figura 27: Evolución del éxito y fracaso

7.1.4.2 Importancia de las variables

Con el fin de analizar si las variables provienen de la misma distribución y así identificar si existen diferencias significativas entre los clústeres, se realiza una prueba previa para comprobar si cumple los criterios necesarios para realizar un análisis de la varianza (ANOVA) y así poder diferenciar si la variabilidad de los datos es imputable a las condiciones de los grupos o es debida al azar.

El primero de los criterios para cumplir es que las poblaciones sigan una distribución normal, para ello se realiza la prueba de Kolmogorov-Smirnov para determinar la bondad de ajuste y también se aplica la prueba de Lilliefors en la corrección de significación y verificar si los datos siguen una distribución normal. Con esta prueba se contrasta la hipótesis nula de que los datos provienen de una población con distribución normal, en la Tabla 12 donde se muestran el estadístico, los grados de libertad y el nivel de significación para cada una de las variables.

	Estadístico	Grados de libertad	Significación
Goal	0,378	23941	0,000
Pledged	0,472	23941	0,000
Pledged_Backer	0,210	23941	0,000
Rate_Pledged_Goal	0,490	23941	0,000
Backers	0,466	23941	0,000
Levels	0,132	23941	0,000
Updates	0,255	23941	0,000
Comments	0,481	23941	0,000
Duration	0,182	23941	0,000
Min_RL	0,432	23941	0,000
Max_RL	0,320	23941	0,000
Range_RL	0,320	23941	0,000

Tabla 12: Prueba Kolmogorov-Smirnov con corrección de significación de Lilliefors

Como se observa en la Tabla 12 el nivel de significación es $< 0,05$, por tanto, se rechaza la hipótesis nula (H_0), en este punto se puede determinar que las variables no siguen una distribución normal.

Dado que no se asume la normalidad de la muestra se opta por utilizar pruebas no paramétricas de varios factores y así facilitar la obtención de diversos contrastes de hipótesis, todas las variables son de entrada excepto Rate_Pledged_Goal que se ha tomado como variable dependiente de salida. En la Tabla 13 se muestran los datos obtenidos tras un estudio de varios factores, donde el software ha realizado las pruebas convenientes en cada caso para comprobar la viabilidad de ciertas hipótesis.

Prueba de hipótesis	Hipótesis nula	Prueba	Significación	Decisión
1	Las categorías de Month se producen con probabilidades iguales	Prueba de Chi-cuadrado para una muestra	0,000 ¹	Rechazar la hipótesis nula
2	Las categorías de Category se producen con probabilidades iguales	Prueba de Chi-cuadrado para una muestra	0,000 ¹	Rechazar la hipótesis nula
3	Las categorías de Subcategory se producen con probabilidades iguales	Prueba de Chi-cuadrado para una muestra	0,000 ¹	Rechazar la hipótesis nula

Prueba de hipótesis	Hipótesis nula	Prueba	Significación	Decisión
4	Las categorías definidas por Status = Successful y Failed se producen con las probabilidades de 0,5 y 0,5	Prueba binomial para una muestra	0,000 ¹	Rechazar la hipótesis nula
5	Las categorías de Clúster se producen con probabilidades iguales	Prueba de Chi-cuadrado para una muestra	0,000 ¹	Rechazar la hipótesis nula
6	La distribución de Goal es normal y con media 7.746,540 y una desviación estándar de 24.917,226	Prueba de Kolmogorov-Smirnov para una muestra	0,000 ¹	Rechazar la hipótesis nula
7	La distribución de Pledged es normal y con media 5.327 y una desviación estándar de 76.966,382	Prueba de Kolmogorov-Smirnov para una muestra	0,000 ¹	Rechazar la hipótesis nula
8	La distribución de Pledged_Backer es normal y con media 60,732 y una desviación estándar de 75,184	Prueba de Kolmogorov-Smirnov para una muestra	0,000 ¹	Rechazar la hipótesis nula
9	La distribución de Rate_Pledged_Goal es normal y con media 1,798 y una desviación estándar de 68,462	Prueba de Kolmogorov-Smirnov para una muestra	0,000 ¹	Rechazar la hipótesis nula
10	La distribución de Backers es normal y con media 80 y una desviación estándar de 931,213	Prueba de Kolmogorov-Smirnov para una muestra	0,000 ¹	Rechazar la hipótesis nula
11	La distribución de Levels es normal y con media 8 y una desviación estándar de 4,236	Prueba de Kolmogorov-Smirnov para una muestra	0,000 ¹	Rechazar la hipótesis nula

Se muestran significaciones asintóticas. El nivel de significación es 0,05. ¹Lilliefors corregida

Tabla 13: Prueba de múltiples factores

Los resultados conseguidos con todos los atributos determinan que hay diferencias estadísticamente significativas entre la distribución de cada variable y los clústeres. Por lo tanto, se puede analizar el comportamiento de los atributos que definen cada clúster.

7.1.4.3 Análisis de componentes principales (PCA)

Una de las ventajas de aplicar esta técnica en este trabajo radica en su capacidad para reducir el sistema de variables en base a reconocer aquellas que midan lo mismo desde diferente punto de vista para reducir la cantidad de información redundante.

Al enfrentar las trece variables dos a dos se obtienen 78 casos diferentes, resultado de aplicar la combinación $C(13,2) = \binom{13}{2} = 78$. Por tanto, a pesar de no ser un número excesivamente elevado de variables, sus combinaciones para correlacionarlas si lo son, de ahí su dificultad para compararlas, por ello, para reducir la dimensionalidad del conjunto de datos y mejorar el entendimiento de la dispersión se realiza un Análisis de

las Componentes Principales para identificar las direcciones en las que la varianza es mayor y diferenciar los vectores o direcciones más representativas dentro del sistema.

Estas técnicas ya han sido aplicadas anteriormente por otros autores a proyectos de la plataforma Kickstarter, concretamente para proyectos de la categoría Games también empleada en este estudio, se ha demostrado entre otras cosas la profunda relación existente entre éxito y colaboradores del (Song et al., 2019), también para el reconocimiento de patrones de cara a aumentar el conocimiento sobre el crowdfunding (Fan-Osuala et al., 2018).

Para realizar este proceso se centran las variables y se vuelven a escalar, de tal manera que tengan una media igual a 0 y una desviación estándar igual a 1, si se obviase este proceso, las variables Goal y Pledged dominarían la mayoría de las componentes principales, ya que ambas tienen una media muy elevada respecto del resto de variables, de igual manera la varianza toma valores muy dispares entre ellas.

Para determinar el número de componentes principales se fija la atención en la matriz de datos generada de dimensiones $n \times p$, donde n es el número de observaciones y p el número de variables. El menor de los dos valores determinados por $(n - 1, p)$ determinará el número de componentes principales resultando en este caso (23.940, 13), es decir, 13 es el número máximo de componentes principales que podemos tomar, aunque como se verá más adelante mediante observación de su comportamiento, solamente será necesario poner atención en las más representativas.

Se importa el data set al programa R (R Core Team, 2016) para realizar el proceso de centrado y reescalado de variables así como para obtener la tabla de *loadings*, estos valores ayudan a comprender el peso de cada variable en cada componente de tal manera que se pueden crear las componentes principales en función de ellos, así en la Tabla 25 se muestran vectores de *loadings* de cada componente principal etiquetadas como $PC1, PC2, \dots, PC13$, de esta manera, la primera componente principal quedaría definida como:

$$\begin{aligned} PC1 &= -0,30077 \textit{ Goal} - 0,3877 \textit{ Pledged} \\ &- 0,13562 \textit{ Pledged_Backer} \\ &- 0,00330 \textit{ Rate_Pledged_Goal} \\ &- 0,42364 \textit{ Backers} + 0,01721 \textit{ Month} \quad (21) \\ &- 0,27553 \textit{ Levels} - 0,15259 \textit{ Updates} \\ &- 0,37951 \textit{ Comments} \\ &- 0,04805 \textit{ Duration} + 0,01561 \textit{ Min_RL} \\ &- 0,39686 \textit{ Max_RL} - 0,39715 \textit{ Range_RL} \end{aligned}$$

Los valores asignados en la PC1 para las variables Goal, Pledged, Backers, Levels, Comments, Max_RL y Range_RL son de peso similar y de valor elevado con respecto a otros valores asignados como a Pledged_Backer (0,13562), esto significa que la primera componente recoge mayoritariamente la información correspondiente a Goal, Pledged,

Backers, Levels, Comments, Max_RL y Range_RL y muy poca información del resto de variables, para las siguientes componentes principales el análisis se realizará de manera similar. Para complementar esta información también se añade la Tabla 26, para proceder al análisis de datos consistente en evaluar la proporción de varianza explicada acumulada y así seleccionar el número de componentes principales mínimo, a partir del cual el incremento deja de ser sustancial. Para ello se representan los valores de la proporción de la varianza acumulada por cada componente principal como se muestra en la Figura 28.

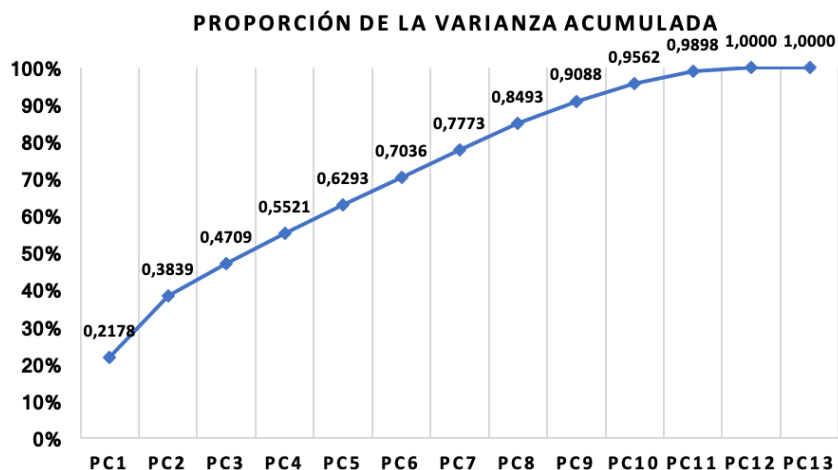


Figura 28: Proporción de la varianza acumulada para cada una de las componentes

La curva mostrada en la Figura 28 tiene un comportamiento bastante plano, donde la varianza acumulada de la primera componente (PC1) recoge el 21,78% de la información, la segunda componente aumenta hasta explicar el 38,39%, la diferencia en las sucesivas componentes es muy poco notable, hasta la décima componente donde pasado ese punto carece de interés de estudio, ya que, disminuye aún más la diferencia de varianzas con cada componente. Cabe resaltar que hasta la séptima componente (PC7) no se supera el 75% de la varianza explicada (77,73%).

Ante esta situación no resulta clara la elección del número de componentes principales del sistema, si bien la primera y segunda componente son las que capturan más porcentaje de varianza, ambas explican la varianza del 38,39% de los datos, la diferencia tan poco significativa en base a añadir el resto de las componentes principales con el fin de entender la varianza del resto de datos, dificulta en gran medida su estudio dado que la situación ideal en este análisis sería que un número bajo de componentes principales explicaran mucha varianza.

En este apartado se ha seleccionado el conjunto de datos de trabajo más apropiado para este estudio basándose principalmente en el elevado número de variables numéricas que posee, posteriormente se han descrito, agrupado y caracterizado. Paralelamente, también han sido sometidos a una limpieza y preparado para que se adapten

perfectamente en las siguientes fases de estudio sin que por ello se pierda calidad ni precisión en los datos.

Después de todo este proceso, el conjunto de datos original se reduce en un 47,77% dando lugar a un nuevo paquete de datos que consta de 23.941 proyectos y que forman un conjunto suficientemente representativo en cantidad y en calidad para el periodo de estudio de interés del crowdfunding.

De las 17 variables que contenía en conjunto de datos original se han reducido a 8, prescindiendo de variables que no aportaban información a un estudio de este tipo como se ha visto en el apartado 7.1.3.1. También se han añadido 5 variables de nueva creación que aumentan el potencial de este estudio aportando nueva información y complementando el valor de otras variables, se obtienen finalmente 13 variables que permiten caracterizar la totalidad del conjunto de datos.

A pesar de que la tipología de los proyectos se encuentra estructurada en categorías y subcategorías, se ha prescindido el uso de subcategorías porque no se consideran de relevancia para este tipo de estudio como se comenta en el apartado 7.1.3.1.

Llegados a este punto, todos los datos obtenidos y todas las variables de estudio se encuentran preparados para que puedan ser objeto de trabajo en diversas áreas de estudio. En este caso se va a someter al nuevo conjunto de datos a diversos algoritmos adaptados a clasificar o agrupar proyectos, de tal manera que bajo unas condiciones establecidas se puedan obtener conclusiones relevantes asociadas a un grupo con características comunes. Para llevarlo a cabo se va a usar una identificación de clústeres o grupos, que posteriormente puedan ser caracterizados en base a aplicar diversas técnicas sobre los resultados obtenidos.

7.2 IDENTIFICACIÓN DE ZONAS DE ÉXITO MEDIANTE CLUSTERIZACIÓN

La aplicación de técnicas de reconocimiento de clústeres, conocidas como clustering, consiguen en el conjunto final de datos de trabajo, mostrar una descripción visual de los proyectos en forma de grupos, apoyándose en el uso de las 13 variables consideradas para este estudio.

A través de los resultados obtenidos con estas técnicas, se pueden realizar observaciones directamente sobre el conjunto de datos, ya que, cada proyecto estará identificado según el grupo al que pertenece, también se tomarán decisiones sobre la representación gráfica obtenida, ya que facilita en gran medida la comparación entre las diferentes variables para el conjunto de datos.

Más allá de la utilidad directa que proporcionan los resultados obtenidos en este apartado, también resulta de gran interés porque que la estructura del conjunto de datos se adapta a mejor posteriores procesos de clasificación y predicción del éxito en los proyectos de crowdfunding.

7.2.1 REPRESENTACIÓN MEDIANTE SELF-ORGANIZING MAPS

Para profundizar más en la asociación de proyectos a clústeres y ampliar las capacidades de este estudio, se realiza un mapa autoorganizado también conocido como SOM (Self-Organizing Maps) en el que se representan las variables en función de una serie de neuronas de forma hexagonal al igual que en el K-means. La capacidad de abstracción de los datos de esta técnica y su carácter no jerárquico justifica su uso para discernir con más claridad el comportamiento de las variables clave (Han and Kamber, 2011).

Este tipo de mapas autoorganizados ya se ha usado con anterioridad, obteniendo buenos resultados para identificar patrones de comportamiento que influyen en el éxito o fracaso de los proyectos (Rodríguez Montequín et al., 2018).

Mediante el uso de esta técnica de representación se aumenta la capacidad de disminuir las dimensiones de un conjunto de datos y también permite la identificación visual de zonas comunes para las variables dadas, y así comparar uno o varios atributos a través del grid para la misma neurona, resultando más sencillo realizar agrupaciones de variables de una gran cantidad de datos para una posterior caracterización.

7.2.1.1 Representación gráfica

Cada una de las 13 variables está representada por un conjunto de 70 neuronas donde cada una representa un conjunto de proyectos como se explica en el apartado 7.2.2, por lo que cada variable se puede caracterizar gráficamente por diferentes zonas, atendiendo a los valores altos o bajos que tome la variable en esa zona.

Al conjunto de las 13 variables, se añade un gráfico más con la U-matrix, realizado a partir de los resultados del SOM, en él se representa la distancia entre el representante de la neurona y sus vecinos más cercanos, de esta manera se entiende que valores bajos

de distancia, constituyen un alto grado de similitud entre neuronas de esa región como se puede observar en el mapa SOM de la Figura 29.

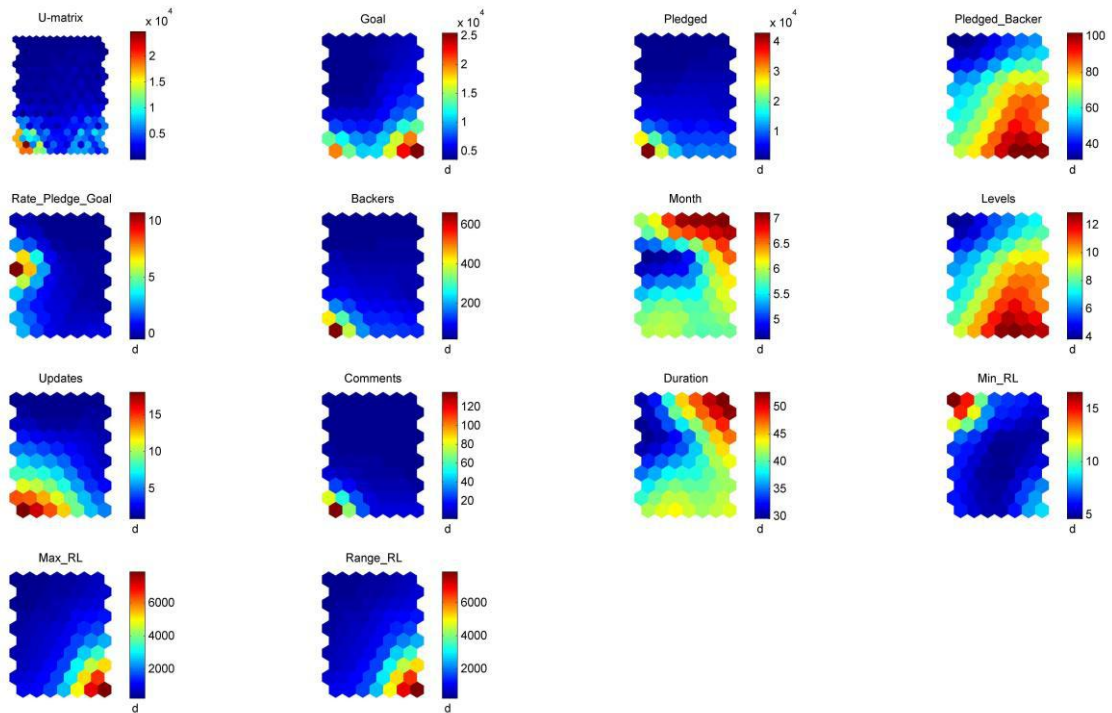


Figura 29: Mapa autoorganizado de cada variable a estudio, también incluye U-matrix

7.2.1.2 Interpretación de los resultados

En la Figura 29 muestran los mapas de las variables acompañados de una escala de color degradado desde azul a rojo, que permite identificar en cada mapa la similitud entre neuronas. Así, neuronas en color azul representan los valores más bajos de distancia o mucha similitud y en color rojo los más altos o poca similitud, independientemente de la escala de valores que tome

Para comprender mejor las distribuciones de los diferentes mapas se proporciona la U-matrix en el que se representa el valor de cada neurona, en particular es la distancia promedio entre la neurona y sus vecinos más cercanos, por tanto, en el gráfico de U-matrix de la Figura 29 se observa que la zona con la mayor distancia promedio se ubica claramente en la esquina inferior izquierda, es decir, de manera general donde los proyectos presentan menores similitudes.

Se observa que los mapas de las variables: Backers, Pledged, Comments y Updates, toman los valores más altos en la esquina inferior izquierda donde existe menos similitud entre neuronas. Entre tres de ellos Backers, Pledged y Comments el aspecto entre los mapas es similar, es probable que tengan un comportamiento similar en determinados casos.

Resulta destacable la descentralización de la variable Goal, al tomar valores altos en dos núcleos diferentes. La zona con valores de Goal más altos (esquina inferior derecha) está

relacionada con zonas de valor alto en Max_RL y Range_RL, en menor medida con Pledged_Backer y Levels. Por otro lado, la zona con valores de Goal no tan elevados (esquina inferior izquierda) corresponden a valores elevados de Backers, Updates, Pledged y Comments, esta conjunción de variables explica en gran medida la importancia de las actualizaciones de proyecto durante la campaña (Xu et al., 2014).

El atributo Rate_Pledge_Goal o cantidad de éxito, a pesar de ser calculado partiendo de la información de Pledged y de Goal, no existe relación directa en los mapas. Sin embargo, los valores más altos de Rate_Pledge_Goal están asociados a los más bajos de Duration y Month, estas últimas tienen un patrón de distancias muy similar que puede ser precedente de un comportamiento similar.

7.2.2 CLUSTERING

En esta fase de agrupado, se someterá al conjunto de datos que incluye las 13 variables, al control de un algoritmo que posicionará cada proyecto en un clúster determinado entre los previamente identificados por el mismo, así, la metodología de trabajo que usa establecerá el número de clústeres y los proyectos contenidos en cada uno.

Para realizarlo, el programa establecerá un dominio de trabajo compuesto por una serie de subgrupos, denominados neuronas, donde cada neurona representa a un grupo de proyectos con características similares estimadas por el algoritmo, el conjunto de estas neuronas representadas de forma gráfica permitirá observar los diferentes clústeres identificados, en base a una numeración y a un código de color.

En este caso, el algoritmo selecciona como ideal una matriz de 7x10 neuronas, con lo que cada variable queda definida por un grid de 70 neuronas. Posteriormente y para agrupar en K conjuntos se ha elegido el algoritmo "K-means" como modelo de agrupación más apropiado por su carácter no jerárquico además de carácter sencillo y rápido. Para identificar el valor óptimo de K se utiliza el índice de Davies-Bouldin que se puede interpretar como la distancia de cada individuo respecto al nuevo clúster identificado, de tal manera que, cuanto más bajo es el índice más homogéneo son los elementos de cada clúster (Davis and Bouldin, 1979), siguiendo este criterio, para los datos introducidos al sistema, el algoritmo distingue seis clústeres como valor óptimo de K.

El algoritmo perfila las neuronas de forma hexagonal, con la ventaja sobre otras formas de que cada centroide es capaz de compartir vecindad con otros seis, además también facilita la visualización al no dejar espacios interneurona excepto en el caso de los bordes (Spanakis and Weiss, 2016), como se puede observar más adelante en la Figura 30 se muestran los seis clústeres obtenidos, resultado de aplicar K-means sobre el grid de representantes identificados con SOM, diferenciados por colores, y con un número para una posterior referencia.

Los colores no tienen trascendencia en los resultados obtenidos, simplemente identifican con más facilidad el tipo de clúster y los diferencian del resto. Sin embargo, la numeración ha tenido trascendencia en el cálculo de clústeres, dado que, durante el proceso iterativo de asociación de proyectos a clústeres, el algoritmo selecciona como número 1 al primer clúster encontrado, número 2 al segundo clúster encontrado y así sucesivamente. De tal manera que finalmente se obtiene un grid de 70 neuronas donde el color y el número de cada neurona representan al clúster al que pertenece.

Como ejemplo, el clúster número 5 está situado en la parte superior izquierda de la Figura 30, está etiquetado de color naranja y ha sido el quinto en ser descubierto por el algoritmo, consta de 10 neuronas y cada una representa el centroide de proyectos contenidos en ella, es decir, dentro de cada neurona es donde los proyectos presentan mayor similitud respecto de las neuronas contiguas.

A través de estas técnicas de clustering se consigue identificar 6 clústeres de proyectos con características claramente diferenciadoras, los proyectos pertenecientes a un clúster presentan una similitud mucho más elevada que los proyectos de dos clústeres diferentes, por tanto, a partir de este punto es posible estudiar como influyen las variables en proyectos de un mismo clúster y no sobre todo el conjunto de datos.

Esta agrupación de proyectos va a permitir dirigir el estudio hacia las variables más influyentes de los proyectos representados uno de los clústeres de la Figura 30, de tal manera, que finalmente sea posible caracterizar ese clúster con una serie de factores identificativos y únicos de los proyectos que contiene.



Figura 30: Posición e identificación de clústeres

7.2.2.1 Taxonomía de clústeres

En este apartado se entrará a analizar el contenido de cada clúster representado en la Figura 30, que dará como resultado un mejor conocimiento de los proyectos que contiene y de las características que caracterizan al clúster.

Una vez definidos los proyectos que forman parte de cada clúster, se pueden extraer para su análisis por separado. Una primera caracterización de cada clúster revelará el ratio de éxito, número de neuronas y de proyectos que contiene, como se observa en la Tabla 14, en ella se muestra como se distribuyen las neuronas en cada clúster y se

observa que el número de neuronas es independiente del número de proyectos que contiene, es decir, no hay ninguna relación entre ambos.

Clúster	Neuronas	Proyectos	Ratio de éxito
1	14	5367	40,13%
2	14	5114	68,71%
3	11	2474	91,43%
4	14	2740	65,04%
5	10	5137	39,77%
6	7	3109	47,57%

Tabla 14: Principales características de los clústeres identificados

El número de proyectos que contiene cada clúster varía desde el *clúster 1*, que con 5367 proyectos es el que mayor cantidad acoge y el clúster 3 por ser el de menor número de proyectos con 2474. En cuanto a la proporción de proyectos exitosos respecto de los fracasados (ratio de éxito), el *clúster 3* domina claramente contando con un 91,43% de los proyectos exitosos. Sin embargo, a pesar de que el *clúster 1* posee la mayor cantidad de proyectos, el ratio de éxito es el menor de toda la serie solamente con un 40,13% de éxito, mostrando que en un principio no existe relación entre el número de proyectos que contiene cada clúster y el ratio de éxito.

Resultan destacables los *clústeres 2, 3 y 4* por ser los que superan la media de éxito de todo el conjunto de datos (55,28% de éxito) como se observó en el apartado 7.1.4 y entre los tres clústeres acogen al 43,13% del total de proyectos.

Estos resultados procedentes de una primera observación a la Tabla 14 en principio no determinan una relación afín entre el número de proyectos de cada clúster y el éxito de estos, sin embargo, los resultados conseguidos con todos los atributos determinan que hay diferencias estadísticamente significativas entre la distribución de cada variable y los clústeres.

El paso previo para estudiar las particularidades que define a cada clúster, es examinar su comportamiento grupal a través del análisis de su conducta ante el éxito o fracaso de los proyectos, posteriormente se pasará a analizar los atributos que los definen y se distinguirán las variables que los caracterizan.

7.2.2.2 Análisis de Éxito y Fracaso

Sobre la clusterización K-means de la Figura 30 se realiza una proyección del éxito y del fracaso, para observar si existen zonas representativas del éxito o del fracaso que cubran a uno o varios clústeres. Para ello en la Figura 31 mostrada a continuación, se muestran las dos proyecciones: de éxito representado en color verde y de fracaso en color rojo,

además se toma como elemento representativo el tamaño de cada neurona, siendo su tamaño proporcional al número de proyectos que contiene.

Para entender el significado descriptivo de la Figura 31 basta con superponer virtualmente ambas proyecciones y verificar las zonas que tienen elevado número de proyectos exitosos y bajo número de proyectos fracasados o viceversa.

Si se toma como ejemplo la neurona superior derecha en la proyección de éxito de la Figura 31, se observa que acoge muchos proyectos de éxito (gran tamaño), a priori puede considerarse una zona con elevado éxito, sin embargo, si se toma la neurona homóloga en la proyección de fracaso, se observa que tiene un tamaño similar a la anterior, por tanto, en ese punto, el número de proyectos de éxito y fracaso podría considerarse como similar.

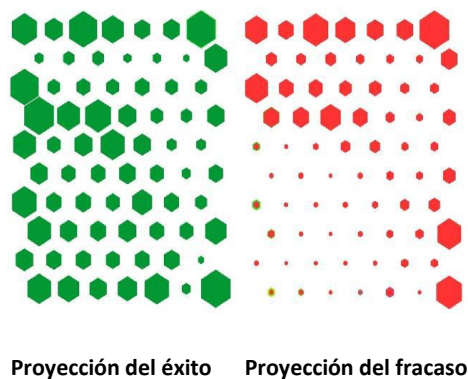


Figura 31: Proyección de éxito y fracaso sobre el mapa de clústeres de la Figura 30

La justificación de uso de ambas proyecciones de la Figura 31 y no solamente una (éxito o fracaso) reside en que a pesar de que es posible una descripción unilateral de la distribución de proyectos exitosos o fracasados, carece de sentido describir una sin tener en cuenta la otra.

A través de una observación hacia ambas proyecciones de la Figura 31, se observa claramente como la zona inferior izquierda posee un elevado número de proyectos de éxito y pocos fracasados, el tamaño de las neuronas de la proyección de éxito en esta zona es muy superior a la de fracaso, llegados a este punto se puede afirmar que existe una zona en la que el ratio de éxito de los proyectos es superior al resto. De la misma forma, también se distingue una zona superior en ambas proyecciones en la que, a pesar de que el número de proyectos de éxito parece elevado debido al tamaño de las neuronas, el gran tamaño de las neuronas de la proyección de fracaso hace reducir en exceso el ratio de éxito de los proyectos.

En la Figura 32 se distingue la zona con alto número de proyectos exitosos y bajo número de proyectos fracasados. Para dirigirse a esta zona se ha determinado la adopción de un nombre que la represente para referencias futuras, así pasa a denominarse *Zona de éxito*.

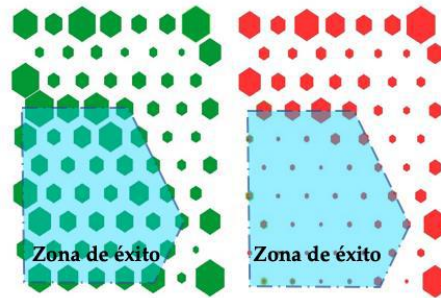


Figura 32: Identificación de la zona de éxito

Al igual que ocurre con la *Zona de éxito*, para distinguir la zona de elevada tasa de fracaso identificada anteriormente, se decide usar la denominación de *Zona de fracaso*, ambas zonas (éxito y fracaso) se observan superpuestas sobre el K-means en la Figura 30, lo que hace más representativo el mapa de clústeres.

A pesar de que estas dos zonas identificadas cubren la mayoría del K-means, para el resto áreas sin identificar no se realiza ninguna determinación a través de este método debido a que no se encuentra dentro del alcance de este trabajo.



Figura 33: Zonas de éxito y fracaso superpuestas a la Figura 30

La *Zona de fracaso* identificada en la parte superior del K-means, comprende el clúster 1 y el clúster 5, en ella se encuentra el 58,90% proyectos fracasados respecto del total, por ello, no es de extrañar que esta zona esté asociada a los dos clústeres con mayor número de proyectos y menor tasa de fracaso. Por otro lado, la *Zona de éxito* comprende los clústeres 2, 3 y una parte del clúster 4, en esta zona cabe destacar que solamente los proyectos que ocupan el clúster 2 y 3 suman el 43,64% proyectos exitosos respecto del total, por lo que el clúster 2 y el clúster 3 forman por si solos una parte muy representativa para este estudio.

El etiquetado de zonas es una herramienta muy útil de cara a una rápida identificación del éxito en los clústeres, pero resulta conveniente hacer un estudio más detallado con el fin de profundizar aún más en esas zonas, debido a esto se decide realizar una primera caracterización de clústeres en base a las variables del sistema.

A pesar de que cualquier variable del conjunto de datos se puede encontrar en cada proyecto y por lo tanto en cada clúster, los valores que toman estas son diferentes dependiendo de su posición en el sistema. Para complementar el significado de cada variable en cada clúster, se ha decidido realizar la Tabla 15 que permite mostrar como se clasifican las variables en cuanto a su valor (más alto o bajo) dentro de cada clúster.

Clúster	Valores	Variables estudiadas
1	Más bajo	<i>Rate_Pledged_Goal</i>
	Más alto	<i>Duration</i>
2	Más bajo	
	Más alto	<i>Rate_Pledged_Goal</i>
3	Más bajo	
	Más alto	<i>Pledged, Comments, Updates, Backers</i>
4	Más bajo	<i>Min_RL</i>
	Más alto	<i>Levels, Pledged_Backer</i>
5	Más bajo	<i>Pledged, Levels, Duration, Goal, Comments, Updates, Max_RL, Backers, Pledged_Backer, Range_RL</i>
	Más alto	<i>Min_RL</i>
6	Más bajo	
	Más alto	<i>Goal, Max_RL, Range_RL</i>

Tabla 15: Identificación de valores altos y bajos de variable por clúster

Esta distribución de variables por clúster se realiza en función del valor promedio de cada variable en el clúster, así, la variable *Duration* toma su valor promedio más bajo en el clúster 1 y la variable *Goal* toma su valor promedio más bajo en el clúster 6 junto a *Max_RL*. Todas las variables elegidas para caracterizar el conjunto de datos se ven representadas en la Tabla 15 y como cada variable solamente puede aparecer dos veces (valor promedio más alto y bajo) hay casos en los que quedan espacios sin cubrir, tales como el valor más bajo de los clústeres 2, 3 y 6, estos espacios pertenecen a situaciones en las que ninguna variable toma su valor promedio más bajo o alto en ese clúster.

De la Tabla 15 destaca el clúster 5 por la gran cantidad de variables que toman sus valores promedio más bajos, cabe recordar que este clúster pertenece a la *Zona de fracaso* al igual que el clúster 1. También el clúster 2 porque solamente representa el valor promedio más alto de la variable *Rate_Pledged_Goal*.

Aún es necesario profundizar más en el conocimiento de cada clúster antes de determinar si los valores promedio más altos o bajos ubicados en la *Zona de fracaso* determinan el descenso del ratio de éxito en los proyectos. Por ello, el paso siguiente es estudiar el comportamiento de cada categoría sobre el total de clústeres observados en la Figura 30 para determinar la relación existente entre las posiciones de las variables y el éxito o fracaso de los proyectos.

7.2.3 ÉXITO EN LOS CLÚSTERES IDENTIFICADOS

Una vez observada la relación entre clúster y atributos del proyecto, es necesario ampliar el conocimiento entre las diferentes categorías proyecto y cada clúster, para ello, se realiza una tabla comparativa donde se muestra el porcentaje de proyectos exitosos contenidos en cada clúster en base a su categoría. En la Tabla 16 se muestran el porcentaje de éxito de cada categoría en cada clúster, presentando una gradación de colores que abarca desde el color verde para valores cercanos al 100% de éxito, el blanco para valores cercanos al 50% y rojo para los valores cercanos a 0.

	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6
Art	43,65%	72,95%	91,53%	64,41%	46,76%	44,64%
Comics	28,44%	66,47%	92,20%	42,67%	34,07%	39,34%
Dance	63,95%	88,10%	90,91%	86,05%	64,74%	44,19%
Design	22,22%	52,22%	91,54%	48,12%	29,29%	42,17%
Fashion	19,69%	41,74%	82,76%	53,76%	16,23%	33,60%
Food	33,33%	67,42%	91,23%	58,38%	30,05%	42,40%
Games	10,71%	42,62%	83,96%	30,14%	16,03%	24,07%
Music	52,29%	79,97%	96,79%	77,00%	46,26%	64,74%
Photography	32,14%	59,20%	88,11%	51,56%	30,24%	38,98%
Publishing	26,08%	57,47%	87,84%	48,42%	27,28%	29,95%
Technology	20,00%	50,00%	93,46%	38,00%	16,36%	30,22%
Theater	61,98%	83,68%	92,73%	76,19%	67,40%	52,88%

Tabla 16: Porcentaje de éxito por categoría y clúster

Como visión general en la Tabla 16 se distingue el clúster 3 como el que mayor porcentaje de éxito tiene en todas las categorías, con un porcentaje de éxito que varía entre el 82,76% de la categoría Fashion hasta el 96,79% de la categoría Music. También se puede observar que los clústeres 1, 5 y 6 presentan baja tasa de éxito, más acusada en los clústeres 1 y 5. Los clústeres 2 y 4 son los que poseen el mayor número categorías con valores de éxito cercanos al 50%, lo que conlleva mayor dificultad para clasificarlos, ya que en algunas categorías el éxito alcanza valores superiores al 75%.

Pese a que el mayor porcentaje de éxito se concentra en el clúster 3, este tiene la particularidad de el éxito para algunas categorías se encuentra distribuido en los clústeres adyacentes y no solamente en el clúster 3, como es el caso de los proyectos asociados a las artes escénicas (Dance, Music y Theater) donde el éxito de estas categorías se dispersa hasta los clústeres 3 y 4 y en el caso particular de la categoría Dance incluso hacia los clústeres 1 y 5 manteniendo unos niveles de éxito por encima

del 63%, este rasgo distintivo encontrado dificulta la asociación categoría-clúster y será necesario realizar un análisis posterior.

La distribución del éxito de la Tabla 16 aporta valiosa información de cara a la caracterización de clústeres, pero no tiene en cuenta el número de proyectos en cada caso, de esta manera, una categoría puede tener un elevado porcentaje de éxito en un clúster y sin embargo acoger muy pocos proyectos en ese caso. Para complementar esta información se realiza la Tabla 17 que proporciona la cantidad de proyectos por categoría y clúster, combinadas ambas mediante un conector, permiten cuantificar la distribución de éxito por categoría y clúster teniendo en cuenta el éxito total de cada categoría.

	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6
Art	669	706	295	354	879	401
Comics	225	167	205	75	135	61
Dance	147	168	33	43	156	43
Design	261	316	260	133	280	166
Fashion	193	230	58	93	228	125
Food	213	267	114	173	183	217
Games	168	183	187	73	156	162
Music	1.658	1.473	592	1.165	1.258	1.092
Photography	336	250	143	128	291	118
Publishing	882	736	370	285	964	394
Technology	160	134	107	50	110	139
Theater	455	484	110	168	497	191

Tabla 17: Número total de proyectos por categoría y clúster

Como ya se había comentado en la Tabla 9, la categoría Music es la que más proyectos acoge de todo el conjunto de datos, por tanto, la distribución de éxito comentada en la Tabla 16 afecta a un número demasiado elevado de proyectos como para hacer una valoración a priori que resulte representativa. Por otro lado, sería erróneo pensar que un elevado porcentaje de éxito en clúster resulta concluyente ya que no tiene en cuenta el número de proyectos, como ejemplo, la categoría Dance de la Tabla 16 tiene un éxito de 90,91% en el clúster 3 y solamente ampara a 33 proyectos.

7.2.4 INDICADORES DE ESTADO

Llegados a este punto es necesario el desarrollo de unos indicadores que a modo de herramienta complementen la información anterior, haciendo de nexo entre la Tabla 15 y Tabla 16 que proporcione información complementaria teniendo en cuenta el número de proyectos de éxito en cada categoría y clúster, también debe de tener capacidad para valorar el éxito de una categoría en función del éxito total de la misma.

7.2.4.1 Ratios de éxito y fracaso

En primer lugar, se han definido dos indicadores que proporcionan el ratio de éxito y de fracaso uno respecto del otro. El SRI (Success Rate Index) valora el número de proyectos exitosos respecto de los fracasados y el FRI (Fail Rate Index) valora los proyectos fracasados respecto de los exitosos, en ambos casos para cada categoría y clúster como se observa en la Ecuación 1.

$$SRI_{i,j} = \frac{S_{i,j}}{F_{i,j}}$$

$$FRI_{i,j} = \frac{1}{SRI_{i,j}}$$

$S = \text{Successful projects}$
 $F = \text{Failed projects}$
 $i = \text{Categoría} \quad i \in [1,12]$
 $j = \text{Clúster} \quad j \in [1,6]$

Ecuación 1: Valoración de los ratios de éxito y fracaso

Esta relación entre proyectos exitosos (S) y fracasados (F) depende de la categoría y el clúster al que pertenece cada proyecto, de esta manera se toman datos para cada una de las 12 categorías y en cada uno de los 6 clústeres en ambos casos: SRI y FRI. El valor de FRI se obtiene calculando la inversa del SRI por lo que no es necesario volver a definir su fórmula.

El SRI queda siempre definido entre valores mayores que 0 como se observa en la Ecuación 2 dependiendo si la mayoría de los proyectos en el clúster son exitosos, fracasados o existe el mismo número de exitosos y fracasados.

$\frac{S}{F} > 1 \quad SRI \in (1, \infty)$

$\frac{S}{F} < 1 \quad SRI \in [0,1)$

$S = F \quad SRI = 1$

Ecuación 2: Dominio del SRI

De esta manera los valores de SRI quedan distribuidos por clúster como se observa en la Tabla 18.

SRI	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6
Art	0,77	2,70	10,80	1,81	0,88	0,81
Comics	0,40	1,98	11,81	0,74	0,52	0,65
Dance	1,77	7,40	10,00	6,17	1,84	0,79
Design	0,29	1,09	10,82	0,93	0,41	0,73
Fashion	0,25	0,72	4,80	1,16	0,19	0,51
Food	0,50	2,07	10,40	1,40	0,43	0,74

SRI	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6
Games	0,12	0,74	5,23	0,43	0,19	0,32
Music	1,10	3,99	30,16	3,35	0,86	1,84
Photography	0,47	1,45	7,41	1,06	0,43	0,64
Publishing	0,35	1,35	7,22	0,94	0,38	0,43
Technology	0,25	1,00	14,29	0,61	0,20	0,43
Theater	1,63	5,13	12,75	3,20	2,07	1,12

Tabla 18: Distribución de valores de SRI por categoría y clúster

7.2.4.2 Relatividad del éxito y fracaso

Hasta ahora el éxito de una categoría en un clúster dependía de la proporción de proyectos exitosos y fracasados, el método es muy rápido y práctico, pero no tiene en cuenta el entorno para su cálculo y se considera una evaluación relativa muy útil para proyectos pertenecientes a un grupo cerrado e independiente como puede ser: clúster, categoría, data set..., pero a medida que se generaliza para comparar con otros clústeres pierde eficacia.

Este concepto cambia la perspectiva de trabajo, ya que, sin perder el punto vista del clúster al que pertenece, ahora pasa a hacerlo respecto de la categoría. Se amplía así el concepto de éxito, ofreciendo una información cuantitativa que permite individualizar el éxito o fracaso teniendo en cuenta el entorno.

Tomando como punto de partida el SRI, se generan dos indicadores denominados SOA (Success ratio Over the Average) y FOA (Fail ratio Over the Average). Ambos permiten conocer si una categoría tiene elevado éxito (SOA) o elevado fracaso (FOA) respecto del éxito o fracaso del total de la categoría. También muestran una visión más amplia de la relatividad del éxito y conceden el beneficio de evaluar un clúster respecto de una categoría no solamente dependiendo del ratio de éxito que tenga (SRI), sino que también del número de proyectos que contiene y cómo beneficia ese clúster a la categoría. La forma de obtener ambos indicadores se define como se observa en la Ecuación 3.

$$SOA_{i,j} = \frac{SRI_{i,j} - \overline{SRI}_i}{\overline{SRI}_i}$$

$$FOA_{i,j} = \frac{FRI_{i,j} - \overline{FRI}_i}{\overline{FRI}_i}$$

$\overline{SRI} = \text{media de SRI}$

$\overline{FRI} = \text{media de FRI}$

Ecuación 3: Éxito y fracaso sobre la media

Para obtener estos indicadores es necesario calcular la media de SRI o FRI en cada caso y así se obtienen dos tablas de SOA y FOA donde se clasifican las diferentes categorías por clúster como se observa en la Tabla 19 y Tabla 20 de valores de SOA y FOA respectivamente.

SOA	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6
Art	-0,74	-0,09	2,65	-0,39	-0,70	-0,73
Comics	-0,85	-0,26	3,40	-0,72	-0,81	-0,76
Dance	-0,62	0,59	1,15	0,32	-0,61	-0,83
Design	-0,88	-0,54	3,55	-0,61	-0,83	-0,69
Fashion	-0,81	-0,44	2,78	-0,08	-0,85	-0,60
Food	-0,81	-0,20	3,02	-0,46	-0,83	-0,72
Games	-0,90	-0,37	3,46	-0,63	-0,84	-0,73
Music	-0,84	-0,42	3,38	-0,51	-0,87	-0,73
Photography	-0,75	-0,24	2,88	-0,44	-0,77	-0,67
Publishing	-0,80	-0,24	3,06	-0,47	-0,79	-0,76
Technology	-0,91	-0,64	4,11	-0,78	-0,93	-0,85
Theater	-0,62	0,19	1,95	-0,26	-0,52	-0,74

Tabla 19: Distribución de valores de SOA por categoría y clúster

FOA	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6
Art	0,17	-0,14	-0,23	-0,08	0,12	0,16
Comics	0,45	-0,30	-0,46	0,01	0,23	0,08
Dance	0,02	-0,07	-0,08	-0,06	0,02	0,17
Design	0,82	-0,27	-0,62	-0,20	0,36	-0,08
Fashion	1,42	-0,70	-1,63	-1,12	2,27	-0,24
Food	0,32	-0,26	-0,41	-0,17	0,45	0,08
Games	4,18	-1,78	-2,76	-0,95	1,54	-0,24
Music	0,06	-0,04	-0,07	-0,03	0,09	0,00
Photography	0,43	-0,31	-0,60	-0,18	0,53	0,14
Publishing	0,68	-0,50	-0,84	-0,32	0,58	0,40
Technology	0,59	-0,48	-0,82	-0,26	0,99	-0,02
Theater	0,04	-0,05	-0,08	-0,03	0,01	0,11

Tabla 20: Distribución de valores de FOA por categoría y clúster

Ambas tablas muestran valores superiores e inferiores a cero, dependiendo de la incidencia de la categoría en cada clúster y ganando más interés cuando estos valores están más alejados del cero, porque serán más representativos.

7.2.4.3 Interpretación de los resultados

Para observar los valores de ambas tablas y con el fin de establecer un criterio, el punto de interés reside en los valores superiores a cero, de esta manera, una categoría que posea su valor más alto de SOA en un determinado clúster, tendrá un éxito superior a la media total de la categoría en ese clúster. Paralelamente, para observar los valores de FOA es conveniente poner atención a los valores mayores que cero sobre todo a los más elevados, ya que, mostrarán que una categoría tiene una tasa de fracaso en el clúster

superior a la media de la categoría, por ejemplo, observando la Tabla 19 la categoría Design solamente tiene valores positivos en el clúster 3 (3,55) lo que implica claramente que el éxito de esa categoría es superior a la media en clúster 3. En el caso de los valores de FOA de la Tabla 20 la categoría Games toma el valor más alto en el clúster 1 (4,18) siendo el único valor positivo, lo que determina que la categoría Games en el *clúster 1* tiene una tasa de fracaso muy elevada respecto de la media de la categoría, concretamente tiene una tasa de fracaso 4,18 veces superior al resto de la categoría.

En otras ocasiones no resulta tan determinante el contraste entre clústeres, debido a que las diferencias no son tan significativas como para considerarlas concluyentes, en cualquier caso, los valores de SOA y FOA proporcionan una valiosa información más allá de aumentar el conocimiento sobre las categorías, ya que, contribuyen a la caracterización y posterior etiquetado de clústeres de una forma más precisa.

Para otorgar a las tablas facilidad de comprensión, se muestran los datos de SOA y FOA expresados en tanto por ciento como se observa en la Figura 34 y la Figura 35.

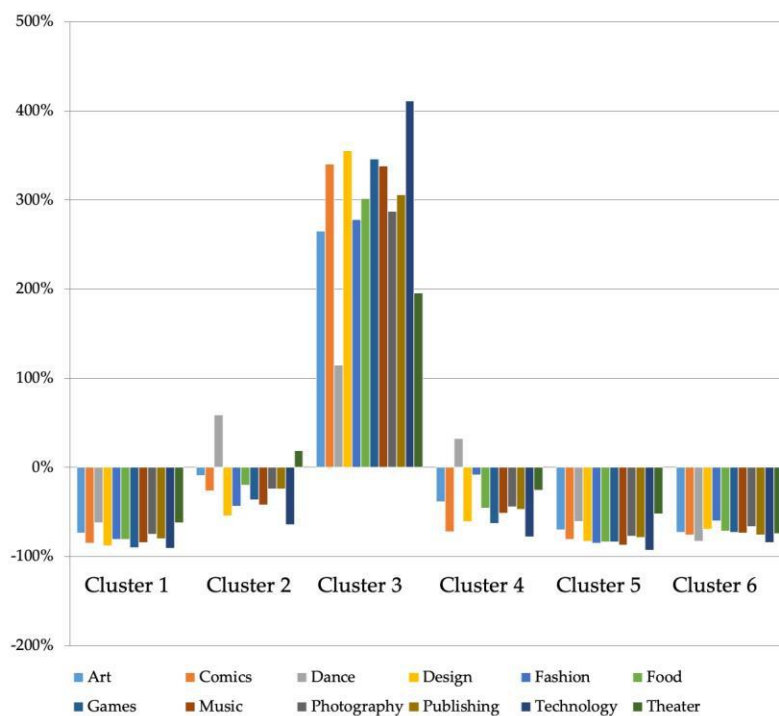


Figura 34: Distribución de los valores de SOA de cada categoría por cada clúster

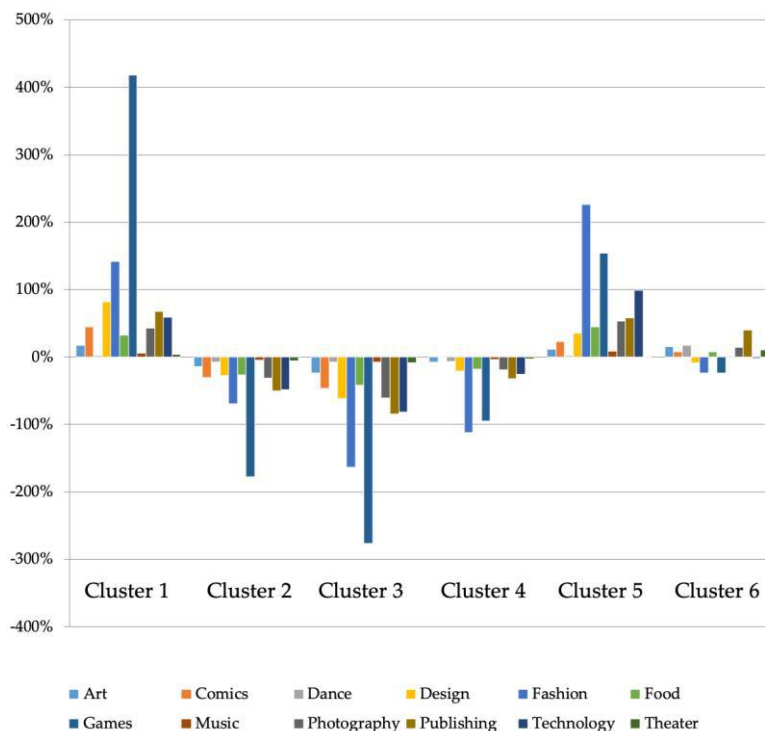


Figura 35: Distribución de los valores de FOA de cada categoría por cada clúster

En la Figura 34 se observa como vuelve a resaltar el clúster 3 como el de mayor éxito, todas las categorías tienen su mayor SOA en el clúster 3, a pesar de ello, algunas categorías tienen valores positivos de SOA en otros clústeres, como el caso de Dance en los clústeres 2 y 4 o Theater en el clúster 2.

La categoría ganadora del clúster 3 es claramente Technology, siendo el éxito en ese clúster más de cuatro veces superior sobre la categoría. La categoría con menos éxito en este clúster es Dance ya que en el clúster 3 sólo supera en 1,14 veces el éxito en ese clúster.

Por otro lado, observando la Figura 35 los clústeres 1 y 5 son los únicos que toman todos los valores de categoría positivos, siendo estos valores los menos recomendados de cara al éxito de los proyectos. Cabe recordar que como criterio de observación de este análisis se realiza en base a los valores positivos tanto de SOA como de FOA.

En el clúster 1 sobresalen con valores positivos suficientemente representativos la categoría Games con una tasa de fracaso 4,18 y Fashion con tasa de fracaso de 1,42 veces superior al resto de la categoría. En los clústeres 2, 3 y 4 todos los valores son menores o iguales a 0 por lo que implica que a priori, no comprometen en gran medida el éxito de los proyectos, en ellos se observa un comportamiento similar donde las categorías Games y Fashion toman los valores más bajos significativamente respecto del resto de categorías.

En el clúster 5 destacan tres categorías como las más perjudicadas por sus valores: Fashion (2,27), Games (1,54) y Technology (0,99), para el resto de las categorías de este clúster los valores se mantienen cercanos a cero sin grandes diferencias entre ellos.

Finalmente, el clúster 6 es el que más ambigüedad presenta ya que todas las categorías se mueven en valores de FOA entre 0,40 y -0,24. Este equilibrio de valores dificulta la conveniencia o no de un proyecto en ese clúster por lo que se debe realizar un análisis más profundo para comprobar si son valores representativos pese a su particularidad.

7.3 EVALUACIÓN DE RESULTADOS

Las técnicas de modelado aplicadas hasta ahora derivan en la búsqueda de un prototipo que permita caracterizar la trayectoria de un proyecto en base a unas zonas de destino con características conocidas. Para el desarrollo de este método de trabajo se ha tenido en cuenta la naturaleza fluida y cambiante de los proyectos de crowdfunding, evitando en todo caso el establecimiento de un sistema rígido donde el orden de los pasos y el peso de cada uno de ellos pueda comprometer la gran flexibilidad de la que se beneficia el crowdfunding. Las diferentes variables asociadas a este modelo de financiación condicionan esta flexibilidad.

En el apartado 7.1 durante el procesado del conjunto de datos se obtienen los primeros hallazgos interesantes sobre la interacción y asociación entre variables, ya que, el establecimiento de las variables fijas en el periodo precampaña va a condicionar las futuras relaciones con el resto de las variables, haciendo que el vínculo entre ellas sea más rígido o flexible, como el caso de la variable fija "Goal" donde cantidades objetivo demasiado elevadas limitarán en exceso la colaboración en el proyecto. El conjunto de variables denominadas "de desarrollo" han mostrado ser las más capaces sobre el resto y sobre las que el creador debe de tener mejor control ya que condicionarán posteriormente los buenos resultados de las variables indicador. Por ello uno de los preceptos fijados para la aplicación de las técnicas ha sido que el creador pueda modificar incluso las variables fijas antes de la campaña de financiación.

A medida que las variables de desarrollo y variables indicador van cambiando conforme transcurre la campaña de financiación, el sistema va ganando rigidez y perdiendo capacidad de adaptación, por lo que, desde el punto de vista de un creador, a medida que se desarrolla el proyecto también se va limitando la capacidad para que el creador modifique su estrategia, sobre todo en cuanto a antelación y proporción necesarias. Por tanto, el modelo final resulta ser adaptativo e iterativo, independientemente de la fase en la que se encuentre el proyecto y siempre teniendo en cuenta que un cambio en una variable puede condicionar al resto de variables.

Las técnicas de clusterización K-means han permitido asociar el conjunto de datos a diversos grupos de proyectos atendiendo a características intrínsecas de ellos y caracterizándolos de cara a su futuro éxito, de esta manera se revelan dos zonas que agrupan varios clústeres, cada una de las cuales comprende éxito y fracaso.

Como el número de proyectos de éxito es independiente del clúster en el que se encuentra, ha sido necesario definir nuevos indicadores denominados SOA y FOA que valoran respectivamente el éxito o el fracaso en cada clúster, independientemente del número de proyectos que contenga. El resultado de estos nuevos indicadores combinado con la caracterización de éxito en los clústeres del apartado 7.2.3 ha permitido desarrollar un mapa de clústeres que facilita la asociación de un proyecto a un clúster, en base las características de este.

Sobre el mapa de clústeres obtenido anteriormente en la Figura 30, se asocia un nombre a cada clúster para una rápida identificación del tipo de proyectos que contiene. Como muestra el mapa de la Figura 36 se ha renombrado de clústeres pasando de un número a una etiqueta, para ello se utilizan tres colores: *verde* para indicar que es un clúster de éxito, *naranja* para indicar cierta precaución y *rojo* para señalar directamente los clústeres a evitar, cada uno de un color y un nombre asociado. En algunos casos las etiquetas van acompañadas de una cruz roja (clúster más desfavorable) o con un V verde (clústeres más favorables). Resultando finalmente que los clústeres 2 y 3 aparecen con la etiqueta identificativa en color verde indicando que son clústeres con alto éxito, los clústeres 1, 5 y 6 poseen la etiqueta de color rojo e indican alto grado de fracaso, finalmente el clúster 4 de color naranja indicando precaución por encontrarse entre ambos estados, rojo y verde.

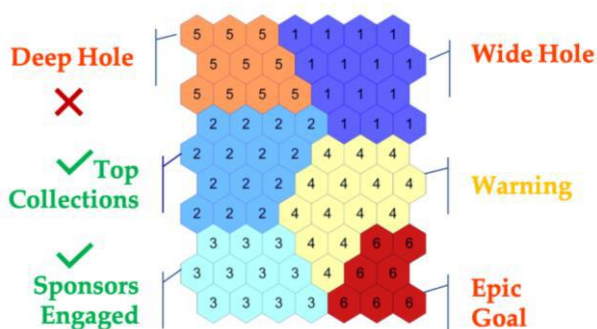


Figura 36: Etiquetado de clústeres

Dentro de la zona de fracaso, los clústeres 1 y 5 contienen ciertas similitudes, ambos poseen tasas de éxito muy bajas, próximas al 40%, por lo que son los clústeres más importantes para evitar. Ambos cuentan con unos valores de FOA muy altos, es decir, la media de éxito de la categoría se sitúa muy lejos de ellos. Las características intrínsecas de estos clústeres los hacen totalmente inapropiados para las categorías Games y Fashion, ambas categorías poseen en estos clústeres los valores más altos de FOA.

Como se observa en la anterior figura, el clúster 1 pasa a llamarse “Wide Hole” indicando que es un clúster con una tasa de éxito muy baja y con cabida para muchos proyectos. Estos proyectos se caracterizan por tener la duración media más alta y como resultado los valores más bajos de superación de objetivo. Paralelamente clúster 5, pasa a

denominarse “Deep Hole”, informando que los proyectos asociados a las características de este clúster poseen una serie de atributos perjudiciales y no es posible modificarlos convenientemente para dirigirlo hacia un clúster de éxito, ya que requiere cambios demasiado importantes en las variables fijas que deberían haberse hecho antes de entrar a la fase de financiación en la plataforma. Entre estos rasgos negativos se incluyen una cantidad mínima de aporte demasiado elevada y no proporcionada con el objetivo del proyecto, así como los promedios más bajos de Backers, Comments, Updates y Duración. “Deep Hole” presenta una clara desventaja para la categoría Technology presentando los valores más altos de FOA.

Uno de los dos clústeres ideales para albergar un proyecto es el clúster 2, debido a que tiene el segundo valor promedio más alto de SRI (2,47) y sus valores de FOA nunca muestran valores positivos, las categorías Dance y Theater son las únicas que presentan valores de SOA positivos, aunque ligeramente, lo que indica que tienen en este clúster un éxito significativo con respecto al promedio de la categoría. Uno de los rasgos más característicos de este clúster es que contiene proyectos con valores medios de Rate_Pledged_Goal más elevados. Debido a que sus proyectos superan ampliamente la cantidad mínima para iniciar la financiación, se consideran proyectos de alto éxito y el clúster 2 pasa a denominarse ahora “Top Collections”.

El segundo de los clústeres apropiados para albergar proyectos con alto ratio de éxito es el clúster 3, es el clúster con más éxito de los identificados, el 91,43% de los proyectos que contiene resultan exitosos. Uno rasgos identificadores del éxito en este clúster es la importancia de mantener altos valores que caracterizan las variables de desarrollo (Backers, Updates y Comments). El vínculo establecido entre el creador y estas tres variables establece la nueva denominación del clúster 3 pasando ahora a denominarse como "Sponsors Engaged". Las características de este clúster son indiscutibles cuando se trata de priorizar el éxito de los proyectos, pero no así para maximizar los beneficios. A pesar de que todas las categorías tienen sus mejores cifras de SOA en este clúster, la categoría más beneficiada es Technology donde la complejidad de los proyectos se beneficia de la buena comunicación entre creador y colaboradores.

A pesar de que el clúster 4 tiene un ratio de éxito del 65,04%, puede considerarse como riesgo aceptable teniendo en cuenta el carácter innovador de este tipo de proyectos. Tiene las cifras promedio más altas de aportación por colaborador, es decir, los colaboradores del proyecto están dispuestos a realizar aportaciones buscando recompensas de alto precio, por otro lado, cuenta con el mayor número de Levels y el menor Min_RL ofrecido, por lo que las recompensas no están correctamente establecidas, incluso es frecuente que algunas no sean seleccionadas por ningún colaborador. Solamente la categoría Dance tiene valores positivos de SOA debido a la dispersión del éxito de las artes escénicas comentado en la Tabla 16. Las características de este clúster hacen que no se pueda afianzar el éxito sin los esfuerzos del creador, a pesar de ello cuenta con una tasa de éxito suficiente para que el creador pueda

modificar con seguridad las características del proyecto y mejorar su potencial, por todo ello este clúster cambia su denominación a “Warning”.

El clúster 6 cuenta con una tasa de éxito del 47,57%, lo que resulta insuficiente de cara a afianzar el éxito, los valores de FOA son comparables a los obtenidos en “Wide Hole” y “Deep Hole”. Estos proyectos poseen un valor demasiado alto para una variable fija (Goal), la cantidad objetivo es demasiado ambiciosa por lo que no es posible la rectificación una vez iniciada la campaña de financiación, debido a las características proyectos que contiene, este clúster se denomina ahora “Epic Goal”.

Esta identificación y etiquetado de clústeres a través de sus atributos permite a un creador o director de un proyecto utilizar un instrumento que facilita la estimación de la viabilidad económica y financiera de un proyecto de crowdfunding. De esta manera, se pueden dedicar recursos y esfuerzos a mejorar la calidad y los índices de beneficio del proyecto.

La flexibilidad obtenida a través de este sistema admite no sólo a proyectos en fases previas a la financiación, sino que también a proyectos que ya hayan iniciado la campaña de financiación, de esta manera cualquier proyecto puede entrar en el sistema para ser ubicado en una de las seis zonas reconocidas y a su vez, si este tiene suficiente capacidad de cambio, determinada entre otras por las variables fijas y de desarrollo, tiene capacidad de mejora en base a la zona más factible y con mejores condiciones entre las seis reconocidas.

8 CLASIFICACIÓN Y PREDICCIÓN DEL ÉXITO DE LOS PROYECTOS

Uno de los objetivos de este trabajo consiste en descubrir cuales son los principales factores que inducen al éxito de los proyectos, como ya se ha visto en apartados anteriores, el conjunto de datos posee un número de variables numéricas suficientemente representativo como para hacer un estudio cuantitativo de las variables implicadas, pero no se ha realizado un estudio de la capacidad de inducción al éxito o al fracaso de cada variable, independientemente del valor que tome.

Este cambio de perspectiva se plantea debido al elevado número de factores y relaciones por los que se rigen este tipo de proyectos, aportando principalmente a los creadores otro punto de vista de cara a orientar los recursos o esfuerzos hacia las variables que tienen más capacidad de inducción al éxito del proyecto. Por ello, las ventajas de uso de estos mecanismos de predicción y clasificación radican en la agilidad con la que un creador puede dirigir su trabajo hacia los hitos más importantes del proyecto, así como decidir una estrategia basada en unos escenarios patrón.

Para llevarlo a cabo este proceso, se va a hacer uso del mismo conjunto de datos adquirido y procesado, como se explica en los anteriores apartados: 7.1.1, 7.1.2 y 7.1.3, con ánimo de adaptar estas nuevas observaciones al estudio cualitativo también aportado, se van a usar las mismas variables pero esta vez van a ser sometidas a unos algoritmos que las estructuran y ordenan, para que posteriormente se puedan clasificar y caracterizar por orden de importancia.

De cara a tomar decisiones ágiles es necesario vincular las acciones a las variables más influyentes sobre los resultados, prescindiendo de realizar esfuerzos innecesarios sobre aquellas que no aporten valor en el resultado final, por ello se ha decidido someter los datos a un mecanismo de clasificación que aclara y reduce la toma de decisiones. También se va a someter a los datos a un proceso de predicción para profundizar aún más en el conocimiento de los factores que acercan al éxito teniendo en cuenta las interacciones entre variables predictoras.

8.1 CLASIFICACIÓN

El uso de árboles de clasificación en proyectos de crowdfunding permite evaluar de forma ágil y rápida los diferentes caminos posibles o los valores tomados a través de los atributos del conjunto de datos que llevan al éxito o al fracaso, facilitando el planteamiento de diferentes escenarios en base a la posición de un proyecto dentro del sistema.

El creador del proyecto de crowdfunding puede calcular la probabilidad de éxito, a partir de la decisión tomada en cada momento y también desde el punto de vista probabilístico puede elegir el camino apropiado para establecer prioridades futuras. Se ha de tener presente, que a medida que se descienden niveles en el sistema, las variables van a tener menor capacidad de influencia sobre el resto, por lo que el orden

de discriminación será mucho menor y se reducirá la capacidad de realizar cambios en el proyecto.

En este tipo de clasificadores no usan como objetivo una variable numérica para establecer la clasificación, sino que lo hacen con una etiqueta seleccionada para ingresar al proceso de clasificación, en este caso se selecciona la variable “Status”, tomada como dicotómica por sus dos salidas posibles (Éxito / Fracaso).

El algoritmo seleccionado para realizar esta clasificación es CART (Classification And Regression Tree), uno de los principales criterios para su selección radica en la robustez ante outliers y su sencilla interpretabilidad (Breiman et al., 2017). El propio algoritmo selecciona aquellas variables y rangos que considera más significativos para la clasificación.

El árbol obtenido tras someter a los datos al algoritmo CART alcanza una precisión de resultados del 91,24% y da como resultado una estructura jerárquica de cuatro niveles, donde cada nodo aporta información referente a sus niveles inferiores y la decisión de llegar a un nodo u otro está basada en la selección de la rama apropiada. El algoritmo del CART selecciona para las ramas del sistema las variables Backers y Goal por ser las más representativas en cuanto a capacidad de discriminación. A través de la matriz de confusión se obtienen valores de acierto (valor predictivo positivo) del 89,34% en los proyectos de éxito y un 93,96% (valor predictivo negativo) en los proyectos fracasados.

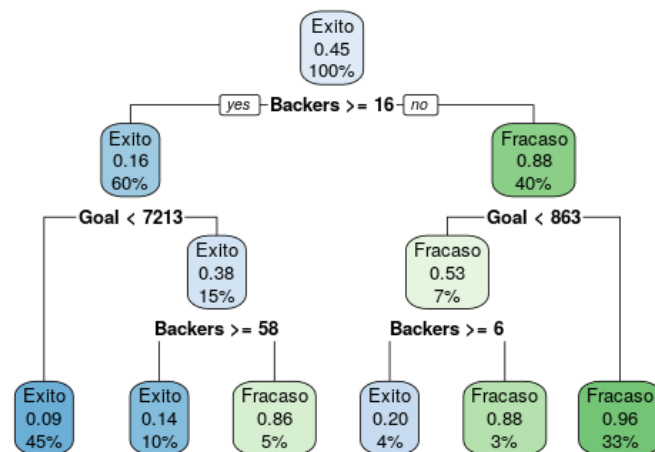


Figura 37: CART Decision tree

El árbol de decisión obtenido en la Figura 37 etiqueta cada nodo como de éxito o fracaso dependiendo del porcentaje de proyectos fracasados (en tanto por uno), también asocia un color en función de la probabilidad de fracaso (azul para el éxito y verde para el fracaso), y finalmente muestra el porcentaje de datos que comprende ese nodo.

Analizando los nodos terminales o nodos hoja obtenidos, resulta de gran interés el primer nodo terminal etiquetado con éxito, ya que alcanza casi la mitad de los datos (45%), y también tiene una probabilidad de fracaso muy baja (9%). Para llegar a este

nodo no es necesario que el proyecto alcance demasiados colaboradores, pero debe de contar con una cantidad objetivo suficientemente baja, en este caso ha de contar al menos con 16 Backers y el Goal no ha de ser muy elevado (7.213\$ como máximo), cabe señalar que el Goal promedio para los proyectos exitosos de todo el conjunto de datos se establece en 4.941\$.

Por el contrario, sino se alcanza el número suficiente de Backers, se pasa a un escenario negativo donde la probabilidad de fracaso resulta más elevada y solamente se puede evitar si los colaboradores del proyecto no descienden de 6 personas y el Goal, consiguientemente ha de ser suficientemente bajo, como valor máximo 863\$, a pesar de ello, este estadio sólo lo alcanzan el 4% de los proyectos.

Esta clasificación determina un modelo en el que la inducción al éxito de los proyectos con ciertas garantías ha de tener un número mínimo de colaboradores y una cantidad objetivo suficientemente baja. De no ser posible ajustar el Goal hasta esta cuantía, será necesario hacer llegar el proyecto a mayor número de Backers, al menos 58 colaboradores, con esta medida se alcanza una probabilidad de éxito del 86%, que a pesar de que es suficientemente alta este punto sólo llegan el 10% de los proyectos.

8.2 PREDICCIÓN

Para comprender que variables son las más influyentes en cada categoría, se realiza un estudio complementario utilizando un algoritmo basado en la realización de Multivariate Adaptive Regression Splines. Esta herramienta flexible, muy usada para resolver problemas no lineales de regresión, automatiza el proceso de búsqueda de modelos predictivos a través de la selección de las variables más relevantes y previniendo los sobreajustes mediante un autotest.

Al igual que en el apartado anterior, para someter el conjunto de datos al algoritmo de predicción se parte de los datos adquiridos, procesados y equilibrados, como se explica en los anteriores apartados: 7.1.1, 7.1.2 y 7.1.3.

Una de las bases de funcionamiento de este tipo de financiación es hacer llegar el proyecto al mayor número de colaboradores posibles (Leimeister, 2012), de esta manera se puede alcanzar una cuantía económica suficientemente elevada y a ser posible superior al Goal, por ello una de las variables más interesantes para someter a predicción es Pledged y es la variable seleccionada para someterla al proceso de predictivo.

Se ha decidido establecer un límite de Pledged antes de someter los datos al algoritmo, así se evitarán proyectos que superen el millón de dólares. A pesar de que solamente 7 proyectos superan esta cantidad, más allá de los problemas que pueden ocasionar los outliers, la razón para usar este prefiltro reside en que la gestión de cuantías económicas de esa magnitud no es fácil y requiere un número elevado de recursos, por lo que se

aleja de la idiosincrasia del crowdfunding y se adapta mejor a otros modelos de financiación de proyectos.

En el proceso de entrenamiento con el algoritmo de predicción de Pledged usando las Multivariate Adaptive Regression Splines, se han realizados distintas prueba con conjuntos de atributos significativos. Dado que los resultados conseguidos no tienen una mejora significativa se decide prescindir de la variable Category para evitar que los resultados en el modelo particularicen en exceso y se otorgue más importancia a una categoría que a otra, con el fin de conseguir un modelo generalista en vez de particularizado a cada categoría.

El hecho de poner un límite máximo a la variable de salida Pledged en cada uno de los proyectos no supone ningún detrimento en la calidad del modelo, tampoco lo hace el hecho de excluir las categorías que los clasifican, sino que permite que los predictores no encasillen el modelo hacia condiciones particulares.

Una vez que el algoritmo ha recorrido el conjunto de datos, se obtienen valores del ajuste del modelo con validación cruzada de R^2 (CVRsq) de la predicción, así como sus desviaciones (sd). Como se observa en la Tabla 21 el algoritmo despliega información de los seis atributos más importantes para el proceso de predicción.

	CVRsq	sd		
	0,583	0,254		
	nsubsets	gcv	rss	
Backers	25	100	100	
Goal	23	48,8	48,8	
Min_RL	21	37,1	37,2	
Duration	19	35,6	35,7	
Comments	18	31,8	31,9	
Range_RL	7	8,1	8,3	

Tabla 21: Importancia de las variables: modelo predicción Pledged

El código formulado por Multivariate Adaptive Regression Splines proporciona información de las variables más influyentes en el proceso de predicción y establece un orden jerárquico basado en el número de subconjuntos que dependen de la variable (nsubsets), considerándose la variable con más importancia cuantos más subconjuntos dependan de ella. También ofrece información normalizada de la aplicación de la validación cruzada generalizada (gcv) que puede considerarse como una estimación en el error de predicción y finalmente la suma de residuos al cuadrado (rss) para medir la discrepancia entre los datos y el modelo.

La variable Backers se posiciona como la más influyente para predecir la cantidad máxima alcanzada en el proyecto, seguida por Goal y Min_RL, Duration y Comments, a partir de este punto decae significativamente la importancia de la variable Range_RL de la que solamente dependen siete subconjuntos.

Uno de los parámetros más importantes para evaluar los resultados de esta técnica son el análisis de los errores en la predicción, ya que indican la capacidad de ajuste del modelo. Para ello en la Tabla 22 se muestran los errores: relativo, absoluto, valor de control, entrenamiento y prueba.

relativo	absoluto	control	entrena	prueba
0,1%	702,43	7,22	58,04	58,88
0,2%	1404,85	17,04	76,84	78,13
0,3%	2107,28	27,60	84,38	85,00
0,4%	2809,71	37,80	88,63	89,24
0,6%	4214,56	89,05	93,27	93,58
1,0%	7024,27	92,86	96,70	96,82
3,0%	21072,81	97,82	99,28	99,14
5,0%	35121,35	98,93	99,68	99,62

Tabla 22: Errores en la predicción

En el primer caso, con un error relativo inferior al 0,1% corresponde a un error absoluto relativamente bajo (702,43\$) sobre todo teniendo en cuenta el amplio margen de Pledged en los proyectos del conjunto de datos. El modelo es capaz de identificar correctamente cerca de un 58,88% de los casos, frente a un 7% en el casos del modelo de control. Por otro lado, los datos de entrenamiento y prueba obtienen resultados similares, en el caso del error relativo inferior al 0,4% se obtiene una diferencia entre ellos de 0,61%, para el resto de los casos es inferior al 0,5% y disminuyendo a medida que se amplía el margen de error relativo, lo que indica que el modelo se comporta de manera estable, generaliza bien y aporta fiabilidad al modelo como se observa en la Figura 38, que se muestra la gráfica de la evolución del modelo para el conjunto de datos de prueba.

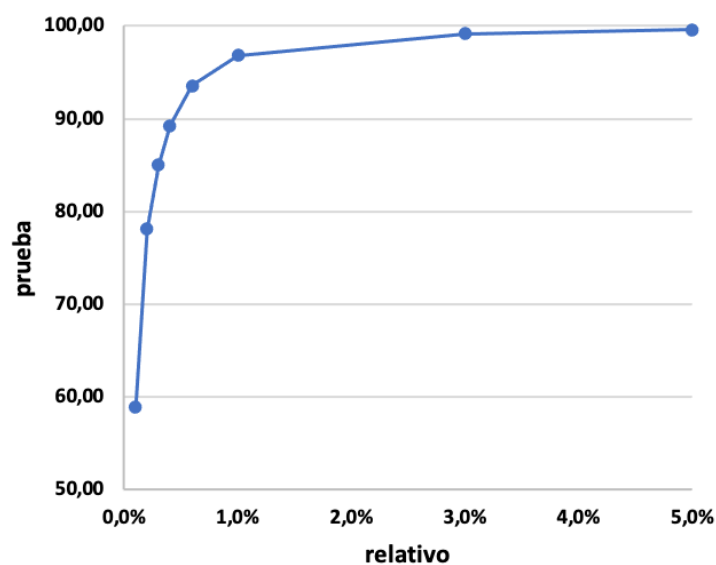


Figura 38: Porcentaje de aciertos con un margen de error relativo en fase de prueba

Como resultado se obtiene la Ecuación 4 con la que es posible predecir el valor de Pledged de un proyecto haciendo uso de cada una de las variables de entrada. La Ecuación 4 consta de 26 términos que contienen los 6 predictores presentados anteriormente en la Tabla 21. De estos 26 sumandos que componen la ecuación, 25 están formados por factores compuestos por la variable de entrada y por un valor comparativo del que se tomará el máximo (siempre ≥ 0) y que compondrán el resultado final de cada sumando.

$$\begin{aligned}
 \text{Pledged} = & \\
 & 4.2e+04 \\
 & - 138 * \max(0, 201 - \text{Backers}) \\
 & + 67 * \max(0, \text{Backers} - 201) \\
 & - 0.83 * \max(0, 27000 - \text{Goal}) \\
 & - 0.092 * \max(0, \text{Goal} - 27000) \\
 & - 95 * \max(0, 147 - \text{Comments}) \\
 & - 40 * \max(0, \text{Comments} - 147) \\
 & + 0.00049 * \max(0, 201 - \text{Backers}) * \max(0, \text{Goal} - 28000) \\
 & + 0.0041 * \max(0, 201 - \text{Backers}) * \max(0, 28000 - \text{Goal}) \\
 & + 0.00069 * \max(0, \text{Backers} - 201) * \max(0, \text{Goal} - 35000) \\
 & - 0.003 * \max(0, \text{Backers} - 201) * \max(0, 35000 - \text{Goal}) \\
 & - 126 * \max(0, \text{Backers} - 201) * \max(0, \text{Duration} - 53) \\
 & - 1.5 * \max(0, \text{Backers} - 201) * \max(0, 53 - \text{Duration}) \\
 & + 44 * \max(0, \text{Backers} - 201) * \max(0, \text{Duration} - 49) \\
 & + 81 * \max(0, \text{Backers} - 201) * \max(0, \text{Duration} - 55) \\
 & + 3.2 * \max(0, \text{Backers} - 201) * \max(0, \text{Min_RL} - 6) \\
 & + 0.018 * \max(0, 20000 - \text{Goal}) * \max(0, \text{Comments} - 147) \\
 & + 0.009 * \max(0, \text{Goal} - 20000) * \max(0, \text{Comments} - 147) \\
 & - 2.2e-05 * \max(0, \text{Backers} - 201) * \max(0, \text{Goal} - 35000) * \max(0, \text{Duration} - 31) \\
 & + 0.00055 * \max(0, \text{Backers} - 201) * \max(0, \text{Goal} - 35000) * \max(0, 31 - \text{Duration}) \\
 & + 0.00011 * \max(0, \text{Backers} - 201) * \max(0, 35000 - \text{Goal}) * \max(0, \text{Duration} - 59) \\
 & + 0.00012 * \max(0, \text{Backers} - 201) * \max(0, 35000 - \text{Goal}) * \max(0, 59 - \text{Duration}) \\
 & - 7.5e-05 * \max(0, \text{Backers} - 201) * \max(0, 35000 - \text{Goal}) * \max(0, \text{Min_RL} - 10) \\
 & + 8.9e-05 * \max(0, \text{Backers} - 201) * \max(0, 35000 - \text{Goal}) * \max(0, 10,00 - \text{Min_RL}) \\
 & - 2.9e-08 * \max(0, \text{Backers} - 201) * \max(0, 35000 - \text{Goal}) * \max(0, 59 - \text{Duration}) * \max(0, \text{Range_RL} - 7495) \\
 & - 9,00E-09 * \max(0, \text{Backers} - 201) * \max(0, 35000 - \text{Goal}) * \max(0, 59 - \text{Duration}) * \max(0, 7495 - \text{Range_RL})
 \end{aligned}$$

Ecuación 4: Multivariate Adaptive Regression Splines para Pledged

La Tabla 21 aporta información acerca de la importancia de las variables predictoras atendiendo a la capacidad de alteración del resultado final, sin embargo, no aportan información para comprender como afectan sus interacciones de cara a la variación de la variable Pledged. Por otro lado, el algoritmo de Multivariate Adaptive Regression Splines a través de la Ecuación 4 permite conocer cómo influyen cada uno de los atributos en la predicción de la variable objetivo Pledged.

Para aumentar el conocimiento sobre las interacciones más importantes entre variables se realiza un análisis de sensibilidad donde se enfrentan por pares las variables según su

mayor capacidad de modificar de Pledged, el resultado de este análisis de sensibilidad se presenta en la Figura 39.

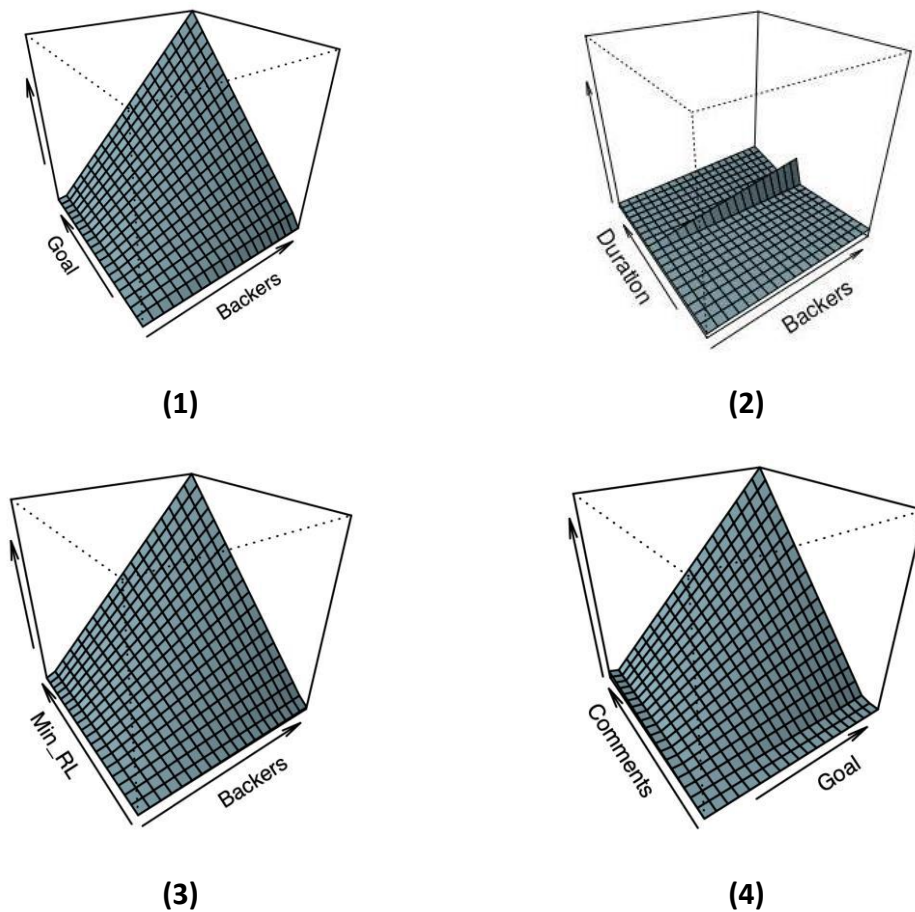


Figura 39: Análisis de sensibilidad por parejas de variables predictoras

En cada uno de los cuatro gráficos mostrados en la Figura 39 se presenta la interacción entre pares de variables predictoras y la variable de salida, las dos variables enfrentadas se muestran con su nombre y con unas flechas que indican en qué sentido crecen los valores, también en el tercer eje (vertical) se muestra el comportamiento de Pledged.

En los casos 1 y 3 muestran un comportamiento similar de la variable de salida, por lo que el estudio comparativo se realiza de manera similar. Sirva como ejemplo el caso número 1, donde un aumento de Backers y Goal proporciona un aumento de Pledged, acentuado cuanto mayor sea el valor de una variable predictor, ya que, el aumento de la otra multiplicará la variable de salida.

Un caso similar a los anteriores salvo por un ligera variación ocurre en el gráfico 4, donde Goal y Comments proporcionan una buena respuesta hacia Pledged a medida que aumentan sus valores, excepto en estadios iniciales de ambas donde el crecimiento de Pledged es prácticamente inexistente.

El enfrentamiento entre las variables Backers y Duration del gráfico 2 muestran el comportamiento más particular de los presentados. El aumento de ambas variables

proporciona un aumento muy leve de la variable Pledged, sin embargo, presenta una falta de estabilidad para una duración determinada ya que el aumento de la variable Backers proporciona un crecimiento de Pledged más significativo que el resto de los casos.

Para complementar los datos de la Figura 39 y obtener una mayor claridad de resultados, se realiza la Figura 40 donde se observa mediante diagramas bidimensionales cuales son las respuestas más destacadas o singulares ante el Pledged.

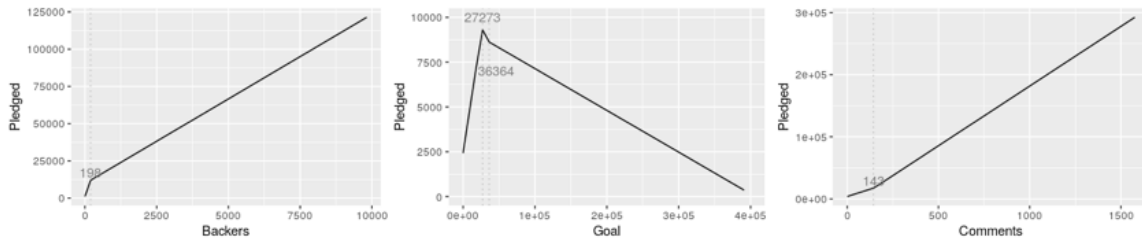


Figura 40: Análisis de sensibilidad con una variable de entrada

Como se puede observar, el comportamiento de Backers en la Figura 40 proporciona un gran incremento en el Pledged en los primeros estadios, cuando se alcanzan los 198 Backers la curva pierde pendiente y continúa con un comportamiento creciente y estable.

La respuesta de Pledged ante la variable Goal es mucho más determinante para los valores más bajos de Goal que para los altos, a partir de un valor determinado de Goal (27273 \$) los valores de Pledged comienzan a disminuir progresivamente.

La curva de Comments tiene rasgos similares con la curva de Backers, proporciona un aumento de Pledged continuo hasta su valor máximo, de una pendiente más moderada en sus estadios iniciales.

8.3 INTERACCIÓN ENTRE MODELOS

Uno de los rasgos comunes identificados al aplicar estas técnicas de clasificación y predicción reside en el solapamiento de variables obtenido. De esta manera, las variables más importantes identificadas con el algoritmo de Multivariate Adaptive Regression Splines también son las que conllevan la determinación del camino a tomar en el *Decision tree*.

El árbol clasificador del éxito mostrado en la Figura 37 resulta una herramienta esencial para un creador por su simplicidad y rapidez, aumentando sus capacidades de cara a una rápida toma de decisiones. Los dos factores condicionantes implicados (Goal y Backers) proveen a esta herramienta de un amplio espectro de uso, dado que, (atendiendo a la clasificación de variables del apartado 7.1.4) Goal es una variable fijada en la fase de prefinanciación y por tanto gran capacidad de inducción al éxito en fases iniciales del ciclo de vida del proyecto, por otro lado, la variable Backers es

potencialmente modificable con los esfuerzos del creador a medida que se desarrolla el proyecto. El vínculo entre ambas variables permite alcanzar el mejor rango de Backers atendiendo al Goal del proyecto dada la elevada correlación existente entre los altos valores de recaudación y el número de colaboradores visto en la Tabla 11.

Con objeto de clasificar el éxito en base a los resultados identificados en otras fases de la investigación, como el caso del SOM y las técnicas de clustering, se ha decidido realizar otro árbol CART denominado "Decision tree compuesto" en el que se han integrado además de las variables anteriores, la pertenencia de los proyectos a los seis clústeres clasificados en la Tabla 14, esto permite obtener una clasificación del éxito desde otro punto de vista aportando también fiabilidad para ambos modelos.

El nuevo árbol creado añade los clústeres: 2, 3 y 4 por su elevada importancia de cara a la clasificación del éxito como se observa en la Figura 41.

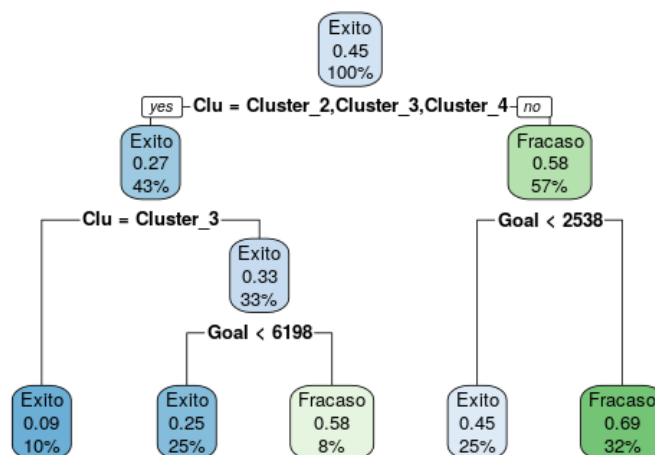


Figura 41: Decision tree compuesto

En el caso del modelo predictivo, la selección de variables también aporta información importante acerca de la capacidad de cada variable para llevar un proyecto al éxito, las tres primeras variables seleccionadas por el algoritmo CART, pertenecen también al conjunto de variables más importantes en anteriores partes de este estudio, ya que, valores de Goal elevados exigen mayor número de Backers y un correcto escalado de recompensas para que el creador pueda ofrecer unas recompensas más competitivas como se ha visto en el apartado 7.1.4.1 de este estudio.

Solamente añadiendo los seis clústeres y la variable Goal, el algoritmo es capaz de alcanzar una precisión de clasificación cercana al 70% como se observa en la matriz de confusión de la Tabla 23, el árbol resultante mostrado en la Figura 41 confirma la separación de los clústeres que agrupan el éxito y el fracaso.

Prediction	Reference	
	Éxito	Fracaso

Éxito	10058	4412
Fracaso	3140	6282

Tabla 23: Matriz de confusión con clústeres

El comportamiento del algoritmo predictivo para la variable Pledged se observa en la Figura 42, donde se muestran (por colores de clúster) las diferencias entre los valores de la variable Pledged y los valores estimados por el modelo. A partir de estos datos se construye la recta de tendencia también del mismo color, estos valores también están acompañados de una línea discontinua azul que señala el resultado ideal de predicción.

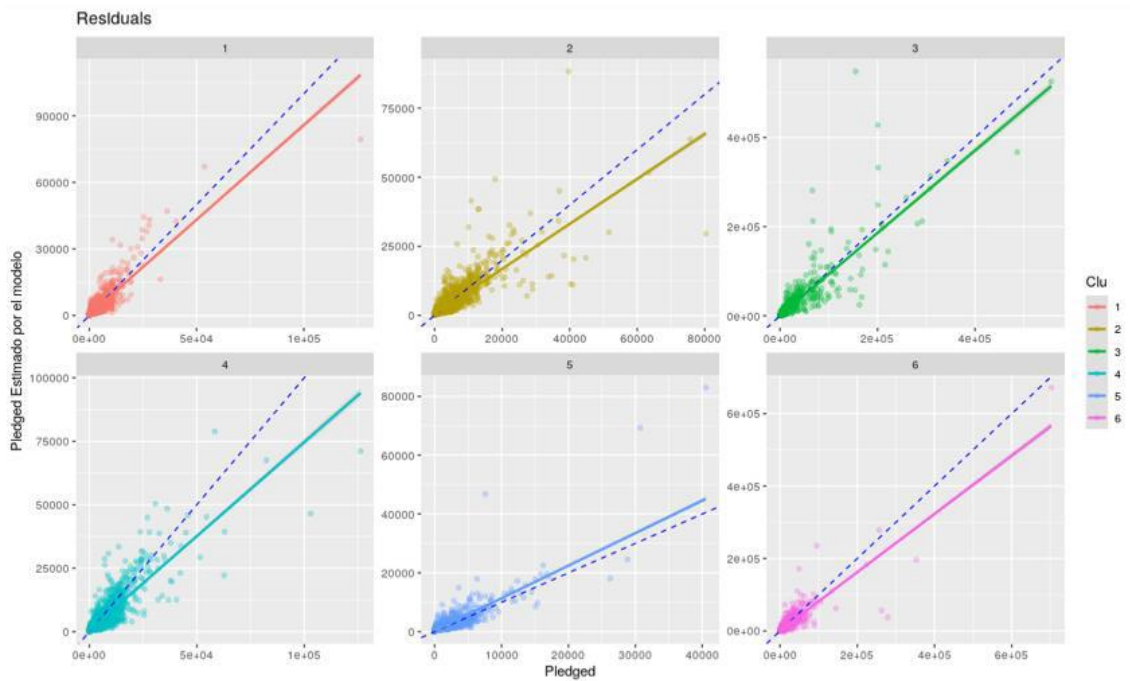


Figura 42: Residuos de Pledged estimados en cada clúster por el modelo

El comportamiento del modelo para los clústeres 2 y 3, etiquetados como de alto éxito en el apartado 7.3, genera buenos resultados con un comportamiento predictivo muy cercano al ideal. En el caso opuesto de los clústeres con alta tasa de fracaso, como son el 1 y el 5, el comportamiento es similar al caso anterior, ya que mantiene elevados valores de predicción. Esta valoración también es extensible al resto de clústeres con unos valores estimados por el modelo cercanos al ideal, mostrando en todos los casos buena calidad de predicción tanto para proyectos de éxito como para fracasados.

9 CONCLUSIONES

Para comprender los factores que llevan al éxito a los proyectos de crowdfunding se posiciona como un método muy fiable trabajar con un conjunto de datos contrastados de la plataforma hospedadora, por ello se ha dedicado especial esfuerzo en la selección de la plataforma de crowdfunding.

Mantener el punto de vista del creador para conocer su capacidad de influencia sobre el proyecto ha sido uno de los aspectos que han resultado más útiles y recursivos a lo largo de este estudio, ya que es el creador quien puede modificar ciertas variables del proyecto para orientarlo hacia un estadio de éxito o fracaso. En base a esto, se realiza una primera clasificación de variables atendiendo a la capacidad de inducción sobre el éxito de los proyectos, entre ellas se distinguen en tres grupos claramente diferenciados. La ventaja de esta categorización de variables reside en que permite al creador dirigir la atención hacia unas u otras dependiendo del estado de la financiación, evitando así interferencias entre variables y esfuerzos improductivos.

Las técnicas de representación gráfica mediante mapas autoorganizados y clusterización han demostrado adaptarse muy bien a este tipo de estudio, por aportar gran capacidad para descubrir relaciones entre variables de manera no jerárquica. Su aplicación ha permitido encontrar relaciones muy significativas entre variables con comportamiento similar. Se ha revelado la gran importancia de la variable Goal al presentarse en dos zonas diferentes con valores altos, siendo la zona con los valores menos elevados la que corresponde a valores elevados de: Backers, Pledged, Comments y Updates.

Mediante la técnica de K-means se han identificado seis clústeres de proyectos, cada uno de ellos con características diferenciadoras del resto, esto ha permitido posicionarlos en un mapa de clústeres y posteriormente solapar zonas donde se encuentran los proyectos exitosos y fracasados, resultando finalmente dos zonas que comprenden varios clústeres y que se han denominado Zona de éxito y Zona de fracaso.

Uno de los problemas encontrados de cara a valorar el éxito y fracaso en los clústeres, radica en que no existe ninguna relación entre el número de proyectos de éxito que contiene el clúster y la categoría a la que pertenecen, para ello se han creado dos indicadores que cuantifican el éxito y el fracaso de una categoría en un clúster respecto del éxito total de la categoría. Estos indicadores han resultado clave para la realización de este trabajo, ya que proporcionan una valiosa información para caracterizar la conveniencia de proyectos de una categoría a un clúster.

De los seis clústeres identificados, dos de ellos poseen características positivas para albergar elevado número de proyectos de éxito y son el destino al que cualquier creador debería dirigir su proyecto para afianzar el éxito. Tres clústeres poseen características perjudiciales para los proyectos lo que denota su bajo ratio de éxito, el clúster restante tiene un ratio de éxito insuficiente para afianzar el éxito, pero tiene características que

posibilitan que el proyecto realice los cambios necesarios para su reubicación en el clúster más conveniente.

La ventaja para un creador del uso de este sistema radica en la capacidad de poder integrar las características de un proyecto para posteriormente adaptarlo hacia el clúster más conveniente. Para ello, cada uno de los clústeres se ha renombrado para que un creador pueda asociar de manera ágil el nombre con las características principales del mismo.

Las técnicas de clasificación y predicción complementan perfectamente a las de clusterización, además tienen capacidad para aplicarse de manera unilateral con buenos resultados.

El árbol de clasificación ha demostrado ser una herramienta poderosa a la par de sencilla para ayudar al creador a dirigir su proyecto hacia el éxito de manera ágil, ya que proporciona diferentes alternativas probabilísticas donde la toma de decisiones se realiza en base a un estadio futuro del proyecto usando solamente dos variables. Por otro lado, el estudio del éxito de los proyectos mediante el algoritmo de predicción Multivariate Adaptive Regression Splines permite tener un control más exhaustivo de la trayectoria del proyecto al hacer uso de las variables más representativas, también ha demostrado gran fiabilidad para predecir el fracaso.

10 LÍNEAS DE FUTURO

Una vez descritas las conclusiones y los resultados a los que se ha llegado con este trabajo se procede a enumerar distintos puntos y líneas de futuro que han sido identificados a lo largo del estudio:

- El uso de un conjunto de datos históricos ha dado buenos resultados para conocer la trayectoria de los proyectos, por ello se abre la oportunidad de adaptar datos de proyectos actuales para someterlos a las técnicas aplicadas.
- Las redes sociales han demostrado poseer un capacidad única para extender conocimiento e ideas, resultaría de gran interés entender que tipo de interacción tienen con las tecnologías aplicadas en este estudio.
- La meta de los proyectos de crowdfunding es alcanzar el éxito en la financiación, pero se desconoce que ocurre con los que no alcanzan el éxito ¿se van a otra plataforma o medio de financiación? o por el contrario se cancelan.
- El valor y el número de recompensas ofrecidas está establecido por el creador, atendiendo a que deben ser lo más accesibles posible, pero no existe ninguna herramienta que permita obtener la cantidad de recompensas, así como su valor ideal.
- Desde los inicios del crowdfunding las subcategorías han ido variando hasta establecerse. Sería de gran interés conocer la influencia del éxito en función de la subcategoría en la que se posicione el proyecto.
- Desarrollar un sistema basado en agentes inteligentes, incorporando el conocimiento adquirido en este trabajo en relación con los atributos que definen los proyectos de crowdfunding. Este sistema sería capaz de actuar por si solo para llevar a cabo la publicación y actualización de los parámetros de un proyecto, emulando las decisiones de los promotores o los inversores, en un entorno de participación o competición entre proyectos.

11 REFERENCIAS

Acconcia, V., 2014. La regla 30-90-100 del crowdfunding [WWW Document]. Vanacco Crowdfunding Consult. URL <https://vanacco.com/la-regla-30-90-100-del-crowdfunding/> (accessed 1.31.19).

Agrawal, A., Catalini, C., Goldfarb, A., 2015. Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions: Crowdfunding. *J. Econ. Manag. Strategy* 24, 253–274. <https://doi.org/10.1111/jems.12093>

Alexa, 2020. Find, Reach, and Convert Your Audience [WWW Document]. Alexa Find Reach Convert Your Audience. URL https://www.alexa.com/siteinfo/kickstarter.com#section_traffic (accessed 2.9.20).

Alonso, J.P., Michudis, G.A., Barreiros, L.E., Borinsky, M.H., Buis, E.J., Herrera, M., Scotti, L.B., Weinschelbaum, B., Albanese, S., Cayuso, S., Etcheverry, R., Pullés, F.G., Gherzi, C., Goldin, A., Krause, M., Lipszyc, D., Niño, L., Baldi, R.R., 2013. PENSAR EN DERECHO NRO. 3 AÑO 2 362.

An, J., Quercia, D., Crowcroft, J., 2014. Recommending Investors for Crowdfunding Projects. *ArXiv14097489 Phys. Stat.*

Andrieu, G., Peter Groh, A., 2020. Strategic exits in secondary venture capital markets. *J. Bus. Ventur.* 105999. <https://doi.org/10.1016/j.jbusvent.2019.105999>

Assis Neto, F.R., Santos, C.A.S., 2018. Understanding crowdsourcing projects: A systematic review of tendencies, workflow, and quality management. *Inf. Process. Manag.* 54, 490–506. <https://doi.org/10.1016/j.ipm.2018.03.006>

artistShare, 2018. artistShare [WWW Document]. artistShare. URL <http://www.artistshare.com> (accessed 11.29.18).

Bacon, E., Williams, M.D., Davies, G.H., 2019. Recipes for success: Conditions for knowledge transfer across open innovation ecosystems. *Int. J. Inf. Manag.* 49, 377–387. <https://doi.org/10.1016/j.ijinfomgt.2019.07.012>

Bakker, G., 2013. Money for nothing: How firms have financed R&D-projects since the Industrial Revolution. *Res. Policy* 42, 1793–1814. <https://doi.org/10.1016/j.respol.2013.07.017>

BBVA, 2020. Préstamos o tarjetas de crédito: qué elegir a la hora de financiar [WWW Document]. BBVA Préstamos O Tarjetas Crédito Qué Elegir Hora Financ. URL <https://www.bbva.com/es/prestamos-tarjetas-credito-elegir-hora-financiar/> (accessed 1.27.20).

Behringer, N., Sassenberg, K., 2015. Introducing social media for knowledge management: Determinants of employees' intentions to adopt new tools. *Comput. Hum. Behav.* 48, 290–296. <https://doi.org/10.1016/j.chb.2015.01.069>

Belarouci, S., Chikh, M.A., 2017. Medical imbalanced data classification. *Adv. Sci.*

- Technol. Eng. Syst. J. 2, 116–124. <https://doi.org/10.25046/aj020316>
- Belleflamme, P., Lambert, T., Schwienbacher, A., 2014. Crowdfunding: Tapping the right crowd. *J. Bus. Ventur.* 29, 585–609. <https://doi.org/10.1016/j.jbusvent.2013.07.003>
- Belleflamme, P., Omrani, N., Peitz, M., 2015. The economics of crowdfunding platforms. *Inf. Econ. Policy* 33, 11–28. <https://doi.org/10.1016/j.infoecopol.2015.08.003>
- Bianchi, M., Murtinu, S., Scalera, V.G., 2019. R&D Subsidies as Dual Signals in Technological Collaborations. *Res. Policy* 48, 103821. <https://doi.org/10.1016/j.respol.2019.103821>
- BIDAUX, T., 2019. Games and Crowdfunding in 2018 [WWW Document]. URL <http://icopartners.com/2019/01/games-and-crowdfunding-in-2018/> (accessed 2.5.20).
- Boldeanu, F.T., Constantinescu, L., 2015. The main determinants affecting economic growth. *Bull. Transilv. Univ. Braşov, Economic Sciences* 8, 10.
- Bolsa de Madrid, 2020. Bolsa de Madrid [WWW Document]. Bolsa Madr. URL <http://www.bolsamadrid.es/esp/asp/Portada/Portada.aspx> (accessed 2.24.20).
- Boostyourcampaign, 2020. Kickstarter & Indiegogo Marketing Services [WWW Document]. Boost Your Campaign. URL <https://www.boostyourcampaign.com> (accessed 4.18.20).
- Brabham, D.C., 2008. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Converg. Int. J. Res. New Media Technol.* 14, 75–90. <https://doi.org/10.1177/1354856507084420>
- Bravo Nufrio, R., 2017. ¿Qué son los préstamos participativos? ¿Cuáles son sus ventajas? [WWW Document]. LegalToday. URL <http://www.legaltoday.com/practica-juridica/fiscal/fiscal/que-son-los-prestamos-participativos-cuales-son-sus-ventajas> (accessed 1.27.17).
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and Regression Trees, Classification and Regression Trees. Routledge, New York. <https://doi.org/10.1201/9781315139470>
- Brent, D.A., Lorah, K., 2019. The economic geography of civic crowdfunding. *Cities* 90, 122–130. <https://doi.org/10.1016/j.cities.2019.01.036>
- Bustos Contell, E., 2013. La financiación de la pyme: préstamo participativo vs. capital riesgo. Aeca, Madrid.
- Caloffi, A., Mariani, M., Rossi, F., Russo, M., 2018. A comparative evaluation of regional subsidies for collaborative and individual R&D in small and medium-sized enterprises. *Res. Policy* 47, 1437–1447. <https://doi.org/10.1016/j.respol.2018.04.022>
- Castrataro, D., 2011. A social history of crowdfunding [WWW Document]. Soc. Media Week SMW. URL <https://socialmediaweek.org/blog/2011/12/a-social-history-of-crowdfunding/> (accessed 11.28.18).
- CDTI, 2020. Ayudas a la I+D+I [WWW Document]. URL

https://www.cdti.es/index.asp?MP=100&MS=898&MN=1&r=1280*800 (accessed 1.25.20).

CEEI, 2020. CEEI Financiación [WWW Document]. URL <http://www.ceei.es/index.asp> (accessed 1.25.20).

CESGAR, 2015. SGR CESGAR. Sociedades de Garantía Recíproca [WWW Document]. URL <http://www.cesgar.es> (accessed 2.24.20).

Chan, C.S.R., Parhankangas, A., Sahaym, A., Oo, P., 2020. Bellwether and the herd? Unpacking the u-shaped relationship between prior funding and subsequent contributions in reward-based crowdfunding. *J. Bus. Ventur.* 35, 105934. <https://doi.org/10.1016/j.jbusvent.2019.04.002>

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., 2000. Step-by-step data mining guide 76.

Cipollone, A., Giordani, P.E., 2019. Entrepreneurs meet financiers: Evidence from the business angel market. *J. Bus. Ventur.* 34, 105913. <https://doi.org/10.1016/j.jbusvent.2018.11.003>

Clark, A., 2020. Kicktraq [WWW Document]. Kicktraq. URL <https://www.kicktraq.com> (accessed 4.18.20).

Cluster Family Office, 2016. Cluster Family Office [WWW Document]. URL <https://clusterfamilyoffice.com/clientes-family-office/> (accessed 2.24.20).

CNMV, 2020. Comisión Nacional del Mercado de Valores [WWW Document]. CNMV. URL <https://www.cnmv.es/portal/Menu/Sobre-CNMV.aspx> (accessed 2.24.20).

Coates, A., Ng, A.Y., 2012. Learning Feature Representations with K-Means, in: Montavon, G., Orr, G.B., Müller, K.-R. (Eds.), *Neural Networks: Tricks of the Trade*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 561–580. https://doi.org/10.1007/978-3-642-35289-8_30

COFIDIS, 2020. Crédito Directo Cofidis [WWW Document]. Cofidis Cuenta Con Nosotros. URL https://www.cofidis.es/es/landing/landing-rv-sem.html?codepromo=SEMGENCRAP&gclid=EAIaIQobChMIxZqF9ezl5wIVmNxRCh08zw9oEAAYASAAEgKL4fD_BwE (accessed 2.22.20).

Cordova, A., Dolci, J., Gianfrate, G., 2015. The Determinants of Crowdfunding Success: Evidence from Technology Projects. *Procedia - Soc. Behav. Sci.* 181, 115–124. <https://doi.org/10.1016/j.sbspro.2015.04.872>

Crosetto, P., Regner, T., 2014. Crowdfunding: Determinants of success and funding dynamics. *Jena Economic Research Papers*.

Cumming, D.J., Leboeuf, G., Schwienbacher, A., 2015. Crowdfunding Models: Keep-it-All vs. All-or-Nothing. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.2447567>

Daily Hodl Staff, 2020. Ethereum App Adoption Rising, With 1.28 Million Active Users in 2019: Report [WWW Document]. Dly. Hodl News Insight Digit. Econ. URL

<https://dailyhodl.com/2020/01/22/ethereum-app-adoption-rising-with-1-28-million-active-users-in-2019-report/> (accessed 3.27.20).

Dan Marom, 2012. A Framework for Political Crowdfunding: Lessons From President Obama [WWW Document]. Crowdsourcing Week. URL <https://crowdsourcingweek.com/blog/a-framework-for-political-crowdfunding-lessons-from-president-obama/> (accessed 12.1.18).

Davis, D.L., Bouldin, D.W., 1979. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* 1, 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>

de Faria, A.M., Oliveira Junior, M. de M., Borini, F.M., 2019. Public funding for innovation: The importance of individual resources of the entrepreneur and the relational resources of the firm. *Technol. Soc.* 59, 101159. <https://doi.org/10.1016/j.techsoc.2019.101159>

DeMartino, I., 2015. How an Australian Teen Nearly Created a Decentralized Reddit in 1997 (Op-Ed) [WWW Document]. COINTELEGRAPH Future Money. URL <https://cointelegraph.com/news/how-an-australian-teen-nearly-created-a-decentralized-reddit-in-1997>

Dibrova, A., 2015. Business Angel Investments: Risks and Opportunities. *Procedia - Soc. Behav. Sci.* 207, 280–289. <https://doi.org/10.1016/j.sbspro.2015.10.097>

Dougherty, G., 2012. Pattern recognition and classification. An introduction, *Pattern Recognition and Classification: An Introduction*. <https://doi.org/10.1007/978-1-4614-5323-9>

DreamMatch Solutions [WWW Document], 2013. . RedBull El Patrocinio Deport. URL <https://www.dreamatch.mx/2013/08/06/red-bull-patrocinio-deportivo/> (accessed 2.22.20).

Elia, G., Polimeno, G., Solazzo, G., Passiante, G., 2020. A multi-dimension framework for value creation through Big Data. *Ind. Mark. Manag.* S0019850120302212. <https://doi.org/10.1016/j.indmarman.2020.03.015>

Ellman, M., Hurkens, S., 2019. Fraud tolerance in optimal crowdfunding. *Econ. Lett.* 181, 11–16. <https://doi.org/10.1016/j.econlet.2019.04.015>

Etter, V., Grossglauser, M., Thiran, P., 2013. Launch hard or go home!: predicting the success of kickstarter campaigns. *ACM Press*, pp. 177–182. <https://doi.org/10.1145/2512938.2512957>

Fan-Osuala, O., Zantedeschi, D., Jank, W., 2018. Using past contribution patterns to forecast fundraising outcomes in crowdfunding. *Int. J. Forecast.* 34, 30–44. <https://doi.org/10.1016/j.ijforecast.2017.07.003>

Feinberg, J., 2014. Wordle [WWW Document]. URL <http://www.wordle.net> (accessed 4.15.20).

Franzen, C., 2012. Senate Passes Amended Crowdfunding Bill, Emulating Kickstarter Model. Talk. Points Memo.

- Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. *Ann Stat.* 19, 1–67. <https://doi.org/10.1214/aos/1176347963>
- Furr, N., 2011. #1 Cause of Startup Death? Premature Scaling [WWW Document]. *Forbes.com*. URL <https://www.forbes.com/sites/nathanfurr/2011/09/02/1-cause-of-startup-death-premature-scaling/> (accessed 1.23.20).
- Gamble, J.R., Brennan, M., McAdam, R., 2017. A rewarding experience? Exploring how crowdfunding is affecting music industry business models. *J. Bus. Res.* 70, 25–36. <https://doi.org/10.1016/j.jbusres.2016.07.009>
- GEDESCO, 2020. Financiación para empresas a corto, medio y largo plazo [WWW Document]. GEDESCO. URL <https://www.gedescos.es> (accessed 2.24.20).
- Gerber, E.M., Hui, J., 2013. Crowdfunding: Motivations and deterrents for participation. *ACM Trans. Comput.-Hum. Interact.* TOCHI 20, 34.
- Gieure, C., Benavides-Espinosa, M. del M., Roig-Dobón, S., 2020. The entrepreneurial process: The link between intentions and behavior. *J. Bus. Res.* S0148296319307635. <https://doi.org/10.1016/j.jbusres.2019.11.088>
- Gobble, Mary Anne M., 2012. Everyone is a venture capitalist: the new age of crowdfunding. - Free Online Library. *Res.-Technol. Manag.* <https://doi.org/10.5437/08956308X5504001>
- Hall, B.H., Lerner, J., 2009. *The Financing Of R&D And Innovation*.
- Han, J., Kamber, M., 2011. *Data mining: concepts and techniques*, 3rd ed. ed. Elsevier, Burlington, MA.
- Hastie, T., Tibshirani, R., 1986. Generalized Additive Models. *Stat. Sci* 1, 297–310. <https://doi.org/10.1214/ss/1177013604>
- Haykin, S.S., 2009. *Neural networks and learning machines*, 3rd ed. ed. Prentice Hall, New York.
- Hekman, E., Brussee, R., 2013. Crowdfunding and Online Social networks. Presented at the 2nd Consortium on Applied Research and Professional Education, At Manchester, p. 22.
- Hemer, J., 2011. A Snapshot on Crowdfunding 43.
- Homfeldt, F., Rese, A., Simon, F., 2019. Suppliers versus start-ups: Where do better innovation ideas come from? *Res. Policy* 48, 1738–1757. <https://doi.org/10.1016/j.respol.2019.04.002>
- Hopp, C., Kaminski, J., Piller, F., 2019. Accentuating lead user entrepreneur characteristics in crowdfunding campaigns – The role of personal affection and the capitalization of positive events. *J. Bus. Ventur. Insights* 11, e00106. <https://doi.org/10.1016/j.jbvi.2018.e00106>
- Hornuf, L., Schwienbacher, A., 2018. Market mechanisms and funding dynamics in equity crowdfunding. *J. Corp. Finance* 50, 556–574.

<https://doi.org/10.1016/j.jcorpfin.2017.08.009>

Hotelling, H., 1933. Analysis of Complex Statistical Variables into Principal Components. *J. Educ. Psychol.* 24, 498–520. <https://doi.org/10.1037/h0071325>

Howe, J., 2006. THE RISE OF CROWDSOURCING [WWW Document]. WIRED. URL <https://www.wired.com/2006/06/crowds/> (accessed 9.11.18).

Hsieh, H.-C., Hsieh, Y.-C., Vu, T.H.C., 2019. How social movements influence crowdfunding success. *Pac.-Basin Finance J.* 53, 308–320. <https://doi.org/10.1016/j.pacfin.2018.11.008>

Hua, G., Zhang, Y., Cheng, T.C.E., Wang, S., Zhang, J., 2020. The newsvendor problem with barter exchange. *Omega* 92, 102149. <https://doi.org/10.1016/j.omega.2019.102149>

Hui, J.S., Greenberg, M.D., Gerber, E.M., 2014. Understanding the role of community in crowdfunding work, in: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*. Presented at the the 17th ACM conference, ACM Press, Baltimore, Maryland, USA, pp. 62–74. <https://doi.org/10.1145/2531602.2531715>

Impúlsame, 2020. Impúlsame [WWW Document]. URL <http://impulsame.es> (accessed 1.27.20).

Innocentive, 2020. The Open Innovation Marketplace [WWW Document]. Innocentive Open Innov. Marketpl. URL <https://www.innocentive.com> (accessed 1.29.20).

Islam, K., O’Gorman, M., 2019. Microcredit contract design: A macroeconomic evaluation. *World Dev.* 124, 104634. <https://doi.org/10.1016/j.worlddev.2019.104634>

Jefatura del Estado, 2015. Ley 5/2015, de 27 de abril, de fomento de la financiación empresarial.

Johnson, K., Wesker, M., Shanahan, E., Ellenoff, D.S., 2017. Review of Crowdfunding Regulation 2017. Interpretations of existing regulation concerning crowdfunding in Europe, North America and Israel.pdf.

Kaptein, R., 2012. Learning to Analyze Relevancy and Polarity of Tweets 10.

Kickstarter, 2019. Kickstarter Centro de Ayuda para Creadores [WWW Document]. Kickstarter - Cent. Ayuda. URL <https://help.kickstarter.com/hc/es/categories/115000492154-Para-creadores> (accessed 1.27.19).

Kickstarter, 2013. The RoboRoach: Control a living insect from your smartphone! [WWW Document]. Kickstarter. URL <https://www.kickstarter.com/projects/backyardbrains/the-roboroach-control-a-living-insect-from-your-sm/updates> (accessed 12.1.18).

Kickstarter, 2012. Pebble: E-Paper Watch for iPhone and Android [WWW Document]. Kickstarter. URL <https://www.kickstarter.com/projects/getpebble/pebble-e-paper-watch-for-iphone-and-android/description> (accessed 12.1.18).

- Kickstarter, PBC, 2012. 2011: The Stats - Kickstarter. Kickstarter Blog.
- Kim, D.-Y., Zhu, P., 2018. Supplier dependence and R&D intensity: The moderating role of network centrality and interconnectedness. *J. Oper. Manag.* 64, 7–18. <https://doi.org/10.1016/j.jom.2018.11.002>
- Kleemann, F., Voß, G.G., 2008. Un(der)paid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing 4, 22.
- Koch, J.-A., 2016. The Phenomenon of Project Overfunding on Online Crowdfunding Platforms – Analyzing the Drivers of Overfunding. Presented at the Proceedings of the 24th European Conference on Information Systems (ECIS 2016), Turkey, p. 16.
- Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43, 59–69. <https://doi.org/10.1007/BF00337288>
- Kok, R.A.W., Biemans, W.G., 2009. Creating a market-oriented product innovation process: A contingency approach. *Technovation* 29, 517–526. <https://doi.org/10.1016/j.technovation.2008.12.004>
- Kotha, R., George, G., 2012a. Friends, family, or fools: Entrepreneur experience and its implications for equity distribution and resource mobilization. *J. Bus. Ventur.* 27, 525–543. <https://doi.org/10.1016/j.jbusvent.2012.02.001>
- Kotha, R., George, G., 2012b. Friends, family, or fools: Entrepreneur experience and its implications for equity distribution and resource mobilization. *J. Bus. Ventur.* 27, 525–543. <https://doi.org/10.1016/j.jbusvent.2012.02.001>
- Kuppuswamy, V., Bayus, B.L., 2013. CROWDFUNDING CREATIVE IDEAS.
- Lahm, R.J., 2005. BOOTSTRAPPING BUSINESS START-UPS: A REVIEW OF CURRENT BUSINESS PRACTICES. *Finance and Accounting Special Topics (Entrepreneurship)* 9.
- Lange, A., Price, M.K., Santore, R., 2017. Signaling quality through gifts: Implications for the charitable sector. *Eur. Econ. Rev.* 96, 48–61. <https://doi.org/10.1016/j.euroecorev.2017.04.007>
- Lánzanos, 2018. HeroQuest 25 Aniversario [WWW Document]. Lánzanos. URL (accessed 12.1.18).
- Leimeister, J.M., 2012. Crowdsourcing: Crowdfunding, Crowdvoting, Crowdcreation. *Control. Manag.* 56, 388–392. <https://doi.org/10.1365/s12176-012-0662-5>
- Lewis, P.A.W., Stevens, J.G., 1991. Nonlinear Modeling of Time Series Using Multivariate Adaptive Regression Splines (MARS). *J. Am. Stat. Assoc.* 86, 864–877. <https://doi.org/10.1080/01621459.1991.10475126>
- Li, H., Chen, X., Zhang, Y., Hai, M., 2018. Prediction of Financing Goal of Crowdfunding Projects. *Procedia Comput. Sci.* 139, 108–113. <https://doi.org/10.1016/j.procs.2018.10.225>
- Lu, C.-T., Xie, S., Kong, X., Yu, P.S., 2014. Inferring the impacts of social media on

- crowdfunding. ACM Press, pp. 573–582. <https://doi.org/10.1145/2556195.2556251>
- Mačiulienė, M., Skaržauskienė, A., 2015. Emergence of collective intelligence in online communities. *J. Bus. Res.* <https://doi.org/10.1016/j.jbusres.2015.10.044>
- Macqueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Multivar. Obs., Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 17* (281-297).
- Madden, M., 2009. In the decade since Napster's launch, selling recorded music has become as much of an art as making the music itself. 18.
- Mahmood, A., Luffarelli, J., Mukesh, M., 2019. What's in a logo? The impact of complex visual cues in equity crowdfunding. *J. Bus. Ventur.* 34, 41–62. <https://doi.org/10.1016/j.jbusvent.2018.09.006>
- Maimon, O., Rokach, L., 2010. *Data Mining and Knowledge Discovery Handbook*, 2nd ed.
- Martín-Barrera, G., Zamora-Ramírez, C., González-González, J.M., 2017. Impact of flexibility in public R&D funding: How real options could avoid the crowding-out effect. *Renew. Sustain. Energy Rev.* 76, 813–823. <https://doi.org/10.1016/j.rser.2017.03.086>
- Matsudaira, K., 2012. *Data Mining the Web Via Crawling*. Commun. ACM BLOG@CACM.
- McCulloch, W., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity 5, 19.
- Messeni Petruzzelli, A., Natalicchio, A., Panniello, U., Roma, P., 2018. Understanding the crowdfunding phenomenon and its implications for sustainability. *Technol. Forecast. Soc. Change.* <https://doi.org/10.1016/j.techfore.2018.10.002>
- Metrejean, C.T., McKay, B.A., 2018. Donation-based crowdfunding and nontaxable gifts.
- Milborrow, S., 2019. *Earth: Multivariate Adaptive Regression Splines* [WWW Document]. *Earth Multivar. Adapt. Regres. Splines*. URL <https://cran.r-project.org/web/packages/earth/index.html> (accessed 6.19.20).
- Ministerio de Educación, 2020. *Media Publicidad* [WWW Document]. *Inst. Nac. Tecnol. Educ. Form. Profr.* URL <http://recursos.cnice.mec.es/media/publicidad/bloque10/pag7.html> (accessed 2.22.20).
- Minitab, 2018. *Introducing MARS*.
- Mitra, T., Gilbert, E., 2014. The language that gets people to give: phrases that predict success on kickstarter. ACM Press, pp. 49–61. <https://doi.org/10.1145/2531602.2531656>
- Mittens, C., Sudek, R., Cardon, M.S., 2012. Angel investor characteristics that determine whether perceived passion leads to higher evaluations of funding potential. *J. Bus. Ventur.* 27, 592–606. <https://doi.org/10.1016/j.jbusvent.2011.11.003>
- Mollick, E., 2014. The dynamics of crowdfunding: An exploratory study. *J. Bus. Ventur.*

29, 1–16. <https://doi.org/10.1016/j.jbusvent.2013.06.005>

Montes de Oca, B., 2019. What happened to Pebble? | Startup Forensics [WWW Document]. URL <https://slidebean.com/blog/pebble-s-demise-startup-forensics> (accessed 4.17.20).

Moqri, M., Bandyopadhyay, S., 2017. Please Share! Online Word of Mouth and Charitable Crowdfunding, in: Fan, M., Heikkilä, J., Li, H., Shaw, M.J., Zhang, H. (Eds.), *Internetnetworked World*. Springer International Publishing, Cham, pp. 162–169. https://doi.org/10.1007/978-3-319-69644-7_16

Muñoz, J., 2011. Tipos de financiación en una startup [WWW Document]. URL <https://www.eoi.es/blogs/mbaptmad/2011/12/13/tipos-de-financiacion-en-una-startup/> (accessed 1.23.20).

muypymes.com, 2015. «Friends, family and fools»: el primer recurso del emprendedor que busca financiación [WWW Document]. MuyPymes. URL <https://www.muypymes.com/2015/08/31/friends-family-and-fools-el-primer-recurso-del-emprendedor-que-busca-financiacion> (accessed 1.25.20).

Negrușă, A.L., Rus, R.V., Sofică, A., 2014. Innovative Tools Used by Business Networks and Clusters in Communication. *Procedia - Soc. Behav. Sci.* 148, 588–595. <https://doi.org/10.1016/j.sbspro.2014.07.084>

Newman, R., Chang, V., Walters, R.J., Wills, G.B., 2016. Web 2.0—The past and the future. *Int. J. Inf. Manag.* 36, 591–598. <https://doi.org/10.1016/j.ijinfomgt.2016.03.010>

Nisbet, R., Elder, J., Miner, G., 2009. *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press, Inc., USA.

Nitsuh Abebe, 2014. Introducing Two New Categories: Journalism and Crafts. *Kickstarter Blog*. URL <https://www.kickstarter.com/blog/introducing-two-new-categories-journalism-and-crafts?lang=es> (accessed 2.8.20).

Olson, J., 2015. How we raised \$40k for Hours with a completely private crowdfunding campaign [WWW Document]. *Medium*. URL <https://medium.com/@jerols/6-things-we-learned-from-running-a-private-crowdfunding-campaign-33ac835de4dd> (accessed 3.27.20).

Paschen, J., 2017. Choose wisely: Crowdfunding through the stages of the startup life cycle. *Bus. Horiz.* 60, 179–188. <https://doi.org/10.1016/j.bushor.2016.11.003>

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*.pdf. *Philos. Mag.* 2, 559–572.

Pengnate, S. (Fone), Riggins, F.J., 2020. The role of emotion in P2P microfinance funding: A sentiment analysis approach. *Int. J. Inf. Manag.* 54, 102138. <https://doi.org/10.1016/j.ijinfomgt.2020.102138>

People'sProject, 2018. Peoples's Project, Ukraine's military and civil crowdfunding [WWW Document]. Peoples Proj. URL https://www.peoplesproject.com/en/?ab_category=social-en

Petitjean, M., 2018. What explains the success of reward-based crowdfunding campaigns as they unfold? Evidence from the French crowdfunding platform KissKissBankBank. *Finance Res. Lett.* 26, 9–14. <https://doi.org/10.1016/j.frl.2017.11.005>

Pi, J., 2012. *kickstarterscrape06022012.csv* [WWW Document]. Kickstarter Scrape. URL https://fusiontables.google.com/DataSource?docid=1tO8U__v4_oATOq6uo9dl-rPyg-dga3rnkZSbw9M#rows:id=1 (accessed 5.12.15).

Pino Romero, C. del, Castelló Martínez, A., 2015. La comunicación publicitaria se pone de moda: branded content y fashion films. *Rev. Mediterránea Comun.* 6, 105–128. <https://doi.org/10.14198/MEDCOM2015.6.1.07>

PMBOK, 2013. *Guía de los fundamentos para la dirección de proyectos (Guía del PMBOK)*.

Python [WWW Document], 2020. . Python. URL <https://www.python.org> (accessed 6.19.20).

Quilageo, 2015. *DIY Crowdfunding. How it can solve the 6 big crowdfunding problems*.

Quinlan, J.R., 1986. Induction of decision trees. *Mach. Learn.* 1, 81–106. <https://doi.org/10.1007/BF00116251>

R Core Team, 2016. *R: A language and environment for statistical computing* [WWW Document]. URL <http://www.R-project.org/>

Rodríguez Montequín, V., Villanueva Balsera, J., Cousillas Fernández, S.M., Ortega Fernández, F., 2018. Exploring Project Complexity through Project Failure Factors: Analysis of Cluster Patterns Using Self-Organizing Maps. *Complexity* 2018, 1–17. <https://doi.org/10.1155/2018/9496731>

Roig Hernando, J., 2016. Crowdfunding: The collaborative economy for channelling institutional and household savings. *Res. Int. Bus. Finance* 38, 326–337. <https://doi.org/10.1016/j.ribaf.2016.03.004>

Roma, P., Messeni Petruzzelli, A., Perrone, G., 2017. From the crowd to the market: The role of reward-based crowdfunding performance in attracting professional investors. *Res. Policy* 46, 1606–1628. <https://doi.org/10.1016/j.respol.2017.07.012>

Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65, 386–408.

Ryu, S., Kim, Y.-G., 2018. Money is not everything: A typology of crowdfunding project creators. *J. Strateg. Inf. Syst.* 27, 350–368. <https://doi.org/10.1016/j.jsis.2018.10.004>

Salford Systems, 2020. *MARS® - Multivariate Adaptive Regression Splines* [WWW Document]. URL <https://www.salford-systems.com/products/mars> (accessed 6.19.20).

Salzberg, S.L., 1994. C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Mach. Learn.* 16, 235–240. <https://doi.org/10.1007/BF00993309>

Schwiebacher, A., Larralde, B., 2010. *Crowdfunding of Small Entrepreneurial Ventures*.

SSRN Electron. J. <https://doi.org/10.2139/ssrn.1699183>

SEPE, 2020. Capitaliza tu prestación [WWW Document]. Capital. Tu Prestac. URL <https://www.sepe.es/HomeSepe/autonomos/prestaciones-para-emprendedores-y-autonomos/capitaliza-tu-prestacion.html> (accessed 1.25.20).

Sohn, S.Y., Lim, K.T., Ju, Y., 2014. Optimization strategy of credit line management for credit card business. *Comput. Oper. Res.* 48, 81–88. <https://doi.org/10.1016/j.cor.2014.03.013>

Sohrabi, M.K., Akbari, S., 2016. A comprehensive study on the effects of using data mining techniques to predict tie strength. *Comput. Hum. Behav.* 60, 534–541. <https://doi.org/10.1016/j.chb.2016.02.092>

Solomon, J., Ma, W., Wash, R., 2015. Don't wait!: How timing affects coordination of crowdfunding donations, in: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, pp. 547–556.

Song, Y., Berger, R., Yosipof, A., Barnes, B.R., 2019. Mining and investigating the factors influencing crowdfunding success. *Technol. Forecast. Soc. Change* 148, 119723. <https://doi.org/10.1016/j.techfore.2019.119723>

Soto-Moya, M. del M., 2016. Incentivos a la financiación empresarial: el caso particular de los business angels en España. *Coop. Desarro.* 24.

Spanakis, G., Weiss, G., 2016. AMSOM: Adaptive Moving Self-organizing Map for Clustering and Visualization:, in: *Proceedings of the 8th International Conference on Agents and Artificial Intelligence*. Presented at the 8th International Conference on Agents and Artificial Intelligence, SCITEPRESS - Science and Technology Publications, Rome, Italy, pp. 129–140. <https://doi.org/10.5220/0005704801290140>

Suryono, R.R., Purwandari, B., Budi, I., 2019. Peer to Peer (P2P) Lending Problems and Potential Solutions: A Systematic Literature Review. *Procedia Comput. Sci.* 161, 204–214. <https://doi.org/10.1016/j.procs.2019.11.116>

The World Bank, 2013. *Crowdfunding's Potential for the Developing World*. The World Bank.

Tran, T., Dontham, M.R., Chung, J., Lee, K., 2016. How to Succeed in Crowdfunding: a Long-Term Study in Kickstarter. *ArXiv160706839 Cs*.

Tran, T., Lee, K., Vo, N., Choi, H., 2017. Identifying On-time Reward Delivery Projects with Estimating Delivery Duration on Kickstarter. *ArXiv171004743 Cs*.

U.S. National Park Service, 2016. Pulitzer- In Depth - Statue Of Liberty National Monument (U.S. National Park Service) [WWW Document]. URL <http://www.nps.gov/stli/learn/historyculture/pulitzer-in-depth.htm> (accessed 2.3.16).

U.S. Securities And Exchange Commission, 2016. Jumpstart Our Business Startups (JOBS) Act [WWW Document]. US Secur. Exch. Comm. - Jump. Our Bus. Startups JOBS Act. URL <https://www.sec.gov/spotlight/jobs-act.shtml> (accessed 1.24.19).

Valentin Mballa, L., 2017. Desarrollo local y microfinanzas como estrategias de atención a las necesidades sociales: un acercamiento teórico conceptual. *Rev. Mex. Cienc. Políticas Soc.* 62, 101–127. [https://doi.org/10.1016/S0185-1918\(17\)30005-3](https://doi.org/10.1016/S0185-1918(17)30005-3)

Vesanto, J., 2002. Data Exploration Process Based on the Self-Organizing Map. 951-666-596-9.

Vulkan, N., Åstebro, T., Sierra, M.F., 2016. Equity crowdfunding: A new phenomena. *J. Bus. Ventur. Insights* 5, 37–49. <https://doi.org/10.1016/j.jbvi.2016.02.001>

Waleczek, P., Zehren, T., Flatten, T.C., 2018. Start-up financing: How founders finance their ventures' early stage. *Manag. Decis. Econ.* 39, 535–549. <https://doi.org/10.1002/mde.2925>

Wang, W., Zhu, K., Wang, H., Wu, Y.-C.J., 2017. The Impact of Sentiment Orientations on Successful Crowdfunding Campaigns through Text Analytics.pdf. *IET Softw.* 11. <https://doi.org/10.1049/iet-sen.2016.0295>

Wessel, M., Adam, M., Benlian, A., 2019. The impact of sold-out early birds on option selection in reward-based crowdfunding. *Decis. Support Syst.* 117, 48–61. <https://doi.org/10.1016/j.dss.2018.12.002>

Wessel, M., Thies, F., Benlian, A., 2016. The emergence and effects of fake social information: Evidence from crowdfunding. *Decis. Support Syst.* 90, 75–85. <https://doi.org/10.1016/j.dss.2016.06.021>

Wheat, R.E., Wang, Y., Byrnes, J.E., Ranganathan, J., 2013. Raising money for scientific research through crowdfunding. *Trends Ecol. Evol.* 28, 71–72. <https://doi.org/10.1016/j.tree.2012.11.001>

Wilhelm, M., Hutchins, M., Mars, C., Benoit-Norris, C., 2015. An overview of social impacts and their corresponding improvement implications: a mobile phone case study. *J. Clean. Prod.* 102, 302–315. <https://doi.org/10.1016/j.jclepro.2015.04.025>

X.net, Caparrós, S., 2014. Experiencias de CROWDFUNDING en el Estado español y Cataluña: PRINCIPALES CARACTERÍSTICAS, RETOS Y OBSTÁCULOS Inspiración y recomendaciones para un instrumento más sólido de financiación transversal colectiva, pública y privada de la cultura.

Xstartups, 2020. Financia tu Startup con un Enisa [WWW Document]. *Financ. Para Startups*. URL https://xstartups.es/?gclid=EA1aIQobChMI45T0sf2h5wIVC1PTCh06lgxvEAAYASAAEgljwvD_BwE (accessed 1.26.20).

Xu, A., Yang, X., Rao, H., Fu, W.-T., Huang, S.-W., Bailey, B.P., 2014. Show me the money!: an analysis of project updates during crowdfunding campaigns, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 591–600.

Xu, L.Z., 2018. Will a digital camera cure your sick puppy? Modality and category effects in donation-based crowdfunding. *Telemat. Inform.* 35, 1914–1924.

<https://doi.org/10.1016/j.tele.2018.06.004>

Yerasani, S., Appam, D., Sarma, M., Tiwari, M.K., 2019. Estimation and maximization of user influence in social networks. *Int. J. Inf. Manag.* 47, 44–51. <https://doi.org/10.1016/j.ijinfomgt.2018.12.016>

Žalik, K.R., 2008. An efficient k'-means clustering algorithm. *Pattern Recognit. Lett.* 29, 1385–1391. <https://doi.org/10.1016/j.patrec.2008.02.014>

Zheng, H., Li, D., Wu, J., Xu, Y., 2014. The role of multidimensional social capital in crowdfunding: A comparative study in China and US. *Inf. Manage.* 51, 488–496. <https://doi.org/10.1016/j.im.2014.03.003>

Zvilichovsky, D., Inbar, Y., Barzilay, O., 2013. Playing Both Sides of the Market: Success and Reciprocity on Crowdfunding Platforms. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.2304101>

ANEXOS

Ecuación 5: Multivariate Adaptive Regression Spline para Status

$$\begin{aligned}
 \text{Fracaso} = & \\
 & -0.4 \\
 & - 6.2e-06 * \max(0, 6180 - \text{Goal}) \\
 & - 3.6e-07 * \max(0, \text{Goal} - 6180) \\
 & + 0.00084 * \max(0, 30 - \text{Pledged_Backer}) \\
 & + 3.4e-05 * \max(0, \text{Pledged_Backer} - 30) \\
 & + 1.5 * \max(0, \text{Rate_Pledged_Goal} - 1.1) \\
 & + 1.3 * \max(0, 1.4 - \text{Rate_Pledged_Goal}) \\
 & - 1.5 * \max(0, \text{Rate_Pledged_Goal} - 1.4) \\
 & - 0.00021 * \max(0, 47 - \text{Backers}) \\
 & - 1.8e-08 * \max(0, 6180 - \text{Goal}) * \max(0, \text{Pledged_Backer} - 138) \\
 & + 7.7e-08 * \max(0, 6180 - \text{Goal}) * \max(0, 138 - \text{Pledged_Backer}) \\
 & - 5.7e-09 * \max(0, 6180 - \text{Goal}) * \max(0, \text{Backers} - 30) \\
 & - 2.8e-07 * \max(0, 6180 - \text{Goal}) * \max(0, 30 - \text{Backers}) \\
 & - 0.047 * \max(0, 0.39 - \text{Rate_Pledged_Goal}) * \max(0, 47 - \text{Backers}) \\
 & + 0.012 * \max(0, 1.4 - \text{Rate_Pledged_Goal}) * \max(0, 28 - \text{Backers})
 \end{aligned}$$

Tabla 24: Características comparadas de los sistemas de financiación de proyectos

TIPO DE FINANCIACIÓN	TIPO DE PROYECTO	SECTOR	PROMOTOR	TAMAÑO	PLAZO DEVOLUCIÓN	CARACTERÍSTICAS	PARTICIPANTES	PLATAFORMA	¿COMBINABLE?	¿SÓLO APORTA €?	POPULARIDAD
Family, Friends & Fools Bootstrapping	Cualquier	Cualquier	Persona física	Variable muy bajo	No hay	Fondo perdido	Uno o varios	Red de conocidos redes sociales	Con cualquier tipo	€, ciertos recursos y asesoría	Alta
Ayudas y Subvenciones	Cualquier	Cualquier excepto inmobiliario y financiero	Administraciones públicas	Variable medio	No hay	Fondo perdido, más probable a mas desfavorecidos	Emprendedores y PYME	emprendedores.es CDTI	No	Si	Alta
Capital semilla	Start-up sin constituir	Cualquier	Cualquiera principalmente BA	Misma cantidad que los fondos propios 25 m - 300 m€	4 -6 años	Permiten carencias de 1 a 2 años	PYMES	ENISA, CDTI...	Depende del equilibrio de la estructura financiera	Si	Media
Concurso	Innovación	Cualquier	Cualquier entidad PYME	Cualquiera normal < 6 m €	No hay	Gran esfuerzo para su concesión	PYMES Persona física o grupo	Difusión publicitaria salones de emprendimiento elevator pitch...	Si, una vez concedido	€, lecciones aprendidas y facilita el acceso a crédito	Baja
Créditos rápidos on-line	Cualquier	Cualquier	Entidad financiera	500 - 6 m €	Desde 7 días a 6 meses	Dinero en horas Interés anual >20%	Cualquiera	COFIDIS, MEDIATIS...	Si	Si	Alta
Tarjeta de crédito	Cualquier	Cualquier	Entidad Bancaria	0 - 10 m €	< 2 meses	Financiación instantánea	Cualquiera	Entidad Bancaria	Si	Si	Alta
Préstamos P2P	Compra, reforma, proyecto...	Cualquiera	Cualquier persona a través de la plataforma WEB	600 €- 8 m €	6, 12, 24, 36 o 48 meses	Permiten retraso en el pago	Sólo persona física	www.comunitae.com	No. Se estudia la morosidad	Si	Baja
Microcréditos	En inicio, consolidación o ampliación	Cualquier	Entidad financiera ICO	10-25 m €	Variable generalmente 3-5 años	No exige avales ni garantías	Persona física o grupo	Entidad bancaria o fundación www.ico.es www.unsolmon.org	No	Si	Alta
Business Angel	Innovación Start-ups	Cualquier	Red de BA	25-250 m€ cada BA	5-7 años	Zona cercana al BA Conoce el sector	PYME Persona física o grupo	Asociación: INNOVAN, ASBAN...	Sólo con otro B.A.	€ y También mentoría asesoría y contactos	Media
Innovación abierta	Solamente I+D	Innovación	Empresas, fundaciones entidades bancarias	Variable y amplio	Único	Protección de propiedad intelectual	Cualquiera	Plataforma WEB www.innocentive.com innovacion.universia.net	Si, una vez concedido	€, bienes materiales o espacio	Baja
Intercambio de servicios	Proyecto de empresa	Cualquier	Empresarios a través de página Web	3 m - 300 m€	Ilimitado	Ningún gasto solo 3,75% por operación, permite refinanciar	PYMES	www.trocobuy.com	si	No aporta € sólo productos y servicios	Baja
Préstamo Participativo	Innovación Start-ups	Cualquiera excepto inmobiliario y financiero	ENISA	300 m - 1,5 M€ nunca más que el valor de la empresa	hasta 9 años	Interés variable, deducible deuda subordinada, carencia hasta 7 años	PYME con un proyecto	Asociación: ENISA, LUZARO...	Con cualquier tipo de financiación	Si	Media
Patrocinador	Cualquier	Cualquier	Persona física	Variable	No hay	Depende de la publicidad	Cualquier empresa	Visita + presentación	Si	€, productos, servicios recursos	Media
Sociedad de Garantía recíproca	Consolidado, no de nueva creación	Cualquier	Entidades financieras	0	No hay	Sin ánimo de lucro	PYMES	Entidades Financieras	Sólo se puede usar combinado con otro método	No aporta € sólo concede aval	Media
Family Offices	Cualquier	Financiero, inmobiliario y otros	Empresa de asesoramiento financiero del patrimonio familiar	Similar al Capital Riesgo	Los mismos que Capital Riesgo	Pueden trabajar como B. Angels o como C. Riesgo	PYMES	Portales Web y clústeres eticapatrimonios.com wefamilyoffices.com	Si	€ y También mentoría asesoría y contactos	Baja
Capital Riesgo	Proyecto consolidado y en fase de crecimiento inicial	Cualquier excepto inmobiliario y MAB	Sociedades o Fondos de CR, Business Angels	10-30 M€	No hay	Se suele adquirir menos del 50% de la empresa	PYMES	Portales Web: https://www.ascrri.org	Si	También aporta confianza a futuros inversores	Alta
Mercado Alternativo Bursátil	Consolidado y de alto retorno	Cualquier	ENISA o Empresa con capital >1,2 M €	Max de 1,5 M€	2 años	interés y comisión 0% Notoriedad Liquidez	Sociedades Anónimas	ENISA	No	Si	Alta
Emplazamiento Publicitario	Cualquier	Marketing Audiovisual	Marca comercial	0	No hay	Necesita integración natural en un medio audiovisual	Empresa consolidada	Cualquier medio audiovisual www.youtube.com www.vimeo.com....	Si	También reconocimiento	Media
Reward Crowdfunding	Cualquier	Cualquiera	Plataforma WEB	Depende de la plataforma WEB	Depende de la plataforma WEB	Sin intermediarios Sin estudio de viabilidad	Personas físicas, grupos o cualquier tipo de empresa * <3 m € proyecto y <6 m € plataforma	Plataforma web www.lanzanos.com www.kickstarter.com www.indiegogo.com	Sólo si ya ha sido concedida	€ y Publicidad tambien en muchos casos Crowdsourcing	Media
Equity Crowdfunding	Cualquier	Cualquier	Plataforma WEB	Lo máximo permitido por ley	No hay	Sujeto a APL Inversión mínima 500 €	Start-up PYME	Plataformas WEB sociosinversores.com	Si	Si	Baja
Donation Crowdfunding	Cualquier	Social / cultural	Plataforma WEB	Variable bajo	No hay	Fondo perdido	Persona física, ONG's o cualquier entidad	Entidades bancarias y plataformas Web migranodearena.org	Si	Si	Alta
Crowdlending	Cualquier	Cualquier	Plataforma WEB	Variable bajo	De meses a 3-4 años	Sujeto al APL	Empresas y PYMES	Plataformas Web arboribus.com	Si	€ y Transparencia en operaciones rentabilidad atractiva	Baja

Figura 43: Sistema de decisión para las alternativas al crowdfunding de recompensa

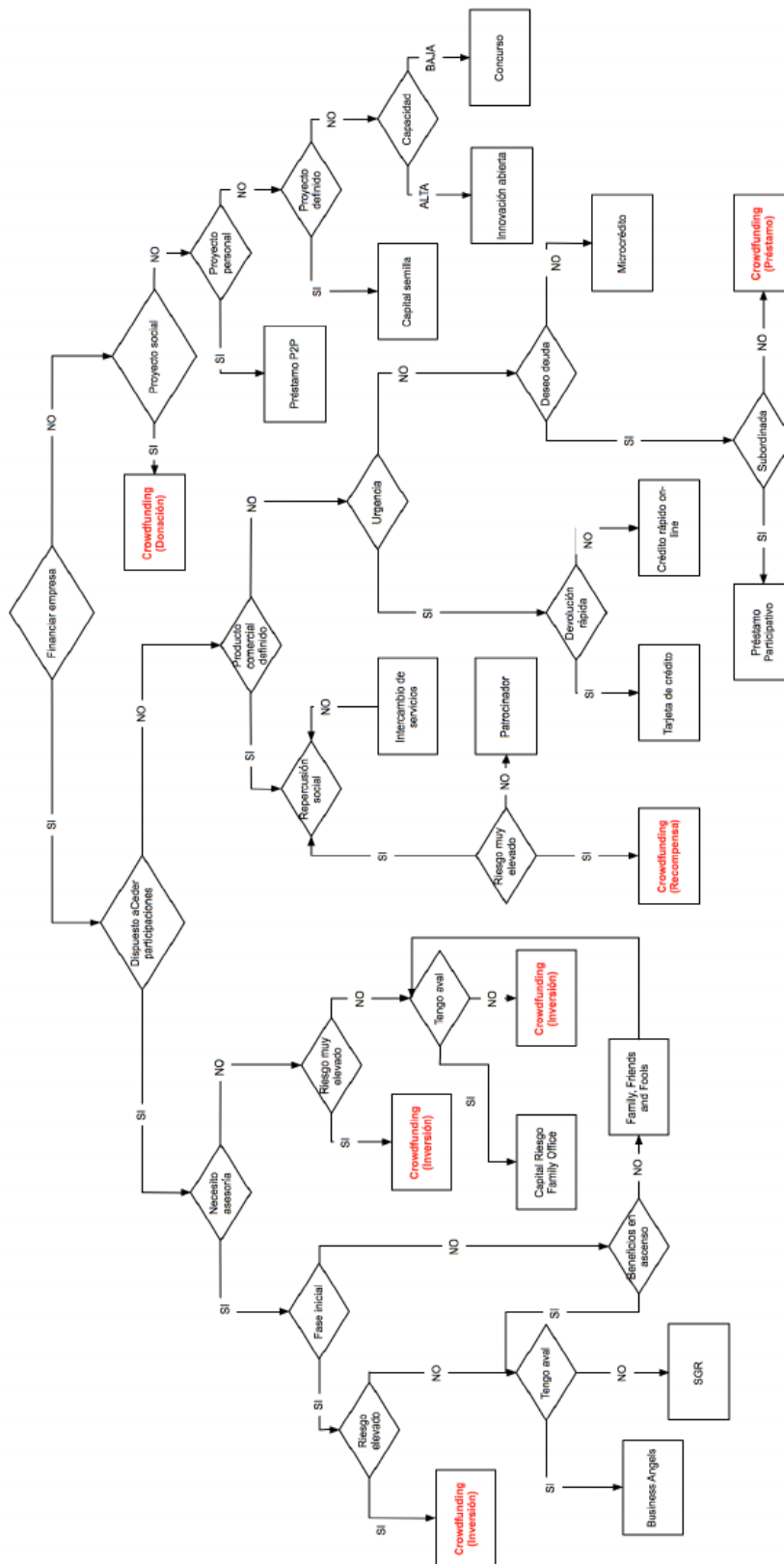


Tabla 25: Loadings



	Range_	Max_R	Min_RL	Duratio	Comm	Update	Levels	Month	Backer	Rate_PI	Pledge	Pledge	Goal
PC1	-0,3972	-0,3969	0,0156	-0,0480	-0,3795	-0,1526	-0,2755	0,0172	-0,4236	-0,0033	-0,1356	-0,3877	-0,3008
PC2	0,4518	0,4513	-0,0256	0,0772	-0,3698	0,0559	0,2874	-0,0018	-0,4253	-0,0153	0,1540	-0,3976	0,0117
PC3	-0,1715	-0,1800	-0,4698	0,1409	0,0107	0,6601	0,3476	0,2156	-0,0033	0,0835	0,0768	0,0132	-0,2853
PC4	0,0385	0,0301	-0,4602	-0,5976	0,0115	-0,1182	0,1810	-0,5774	0,0134	-0,0081	-0,1837	0,0089	-0,1198
PC5	-0,0216	-0,0238	-0,1255	0,2408	0,0142	-0,0062	0,0327	-0,0962	0,0108	-0,9514	-0,0966	0,0008	0,0023
PC6	-0,1689	-0,1642	0,2618	-0,1468	-0,0253	0,2789	-0,0380	-0,4172	-0,0477	-0,1015	0,7500	-0,0365	0,1553
PC7	-0,0095	-0,0063	0,1768	0,5884	0,0046	0,1654	0,0389	-0,6553	0,0026	0,2256	-0,3273	-0,0055	-0,0784
PC8	0,0451	0,0569	0,6503	-0,3962	0,0171	0,3545	0,2027	0,0638	0,0205	-0,1587	-0,3150	0,0292	-0,3437
PC9	0,1018	0,1022	0,0198	0,1593	-0,0413	-0,3565	0,0375	-0,0332	0,1146	0,0017	0,3654	0,2731	-0,7752
PC10	0,2492	0,2462	-0,1649	-0,0804	-0,1252	0,4072	-0,7899	-0,0447	0,0378	-0,0338	-0,0353	0,1032	-0,1472
PC11	-0,0615	-0,0611	0,0201	-0,0134	-0,7933	0,0112	0,1315	0,0044	0,1300	-0,0018	-0,0663	0,5325	0,2004
PC12	-0,0013	-0,0013	-0,0022	-0,0057	0,2665	0,0096	-0,0081	-0,0018	-0,7779	-0,0025	-0,0296	0,5659	0,0499
PC13	-0,7070	0,7071	-0,0128	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Tabla 26: Varianza explicada

	Proporción de varianza acumulada	Proporción de varianza	Desviación típica
PC1	0,2178	0,2178	1,6828
PC2	0,3839	0,1661	1,4692
PC3	0,4709	0,0871	1,0638
PC4	0,5521	0,0812	1,0271
PC5	0,6293	0,0772	1,0018
PC6	0,7036	0,0743	0,9828
PC7	0,7773	0,0737	0,9790
PC8	0,8493	0,0719	0,9671
PC9	0,9088	0,0596	0,8800
PC10	0,9562	0,0474	0,7849
PC11	0,9898	0,0336	0,6612
PC12	1,0000	0,0102	0,3634
PC13	1,0000	0,0000	0,0000

Article

Key Factors for Project Crowdfunding Success: An Empirical Study

Aladino Fernandez-Blanco *, Joaquin Villanueva-Balsera *, Vicente Rodriguez-Montequin 
and Henar Moran-Palacios

Project Engineering Department, University of Oviedo, 33004 Oviedo, Spain; montequi@api.uniovi.es (V.R.-M.); moranhenar@uniovi.es (H.M.-P.)

* Correspondence: aladinofb@gmail.com (A.F.-B.); jmvillanueva@uniovi.es (J.V.-B.);
Tel.: +34-985-104-272 (A.F.-B. & J.V.-B.)

Received: 26 November 2019; Accepted: 10 January 2020; Published: 14 January 2020



Abstract: Crowdfunding is a response to the financing problem of innovative projects in an environment of severe economic crisis. Its competitive advantage lies in its independence from banking institutions and the distribution of risk among a certain number of funders. Since its inception, the number of successfully completed projects has grown to a point where it has started to suffer a downturn that puts its sustainability at risk. This study concerns this particular period of downturn, in order to identify attributes that characterize it, and to define behavioral stereotypes that may be associated with new projects. On a wide data set from sufficiently contrasted projects, and through the use data mining techniques, we extracted the most influential factors in determining the success or failure of the projects, that will subsequently be grouped together using clustering techniques. Six groups of projects have been identified, each with their own characteristics that define them, two of them clearly guide the projects to success and another one allows the modification its characteristics to move away from failure. This achieved strategy allows us to estimate which potential group would be the result of a new project.

Keywords: crowdfunding; innovative projects; risk funding; project management

1. Introduction

Nowadays, it is not easy to finance a novel idea or project. The different entities capable of giving financial support to new projects—far from facilitating credit access—use their administrative tools to make sure that behind each project there is a minimum risk, as low credibility shows the probability of no return on investment, thus avoiding the problem of lack of liquidity during the project's life cycle [1].

The advance of information and communication technologies allows access to unlimited and immediate sources of information, multiplying the impact of any activity simply by disseminating it across the appropriate forums, social networks or specific platforms [2]. This new environment is particularly suitable for giving access to resource acquisition and improving the market for projects that would otherwise remain stagnant or backward [3].

“Crowdfunding” has emerged as a way of financing ideas and turning them into projects, especially those difficult to finance because of their innovative character. Anyone can contribute economically, either to obtain a reward, or altruistically—for the satisfaction of having collaborated on a project. Thus, a double objective is achieved: obtaining the finance for its development and simultaneously gaining clients, since client and promoter are an indissoluble part of the method in this type of project.

By means of crowdfunding, unlike the traditional sale of products, the necessary funds are advanced, which reduces risk and makes the project development possible. The contact is carried out through specific platforms that connect people interested in a particular type of product with entrepreneurs offering them. Otherwise, they would not come into contact, causing unmet needs and unrealized expectations [4]. So, a crowdfunding campaign is considered “successful” when the project reaches the minimum necessary funds (budgeted in advance) and is usually considered “very successful” if the funds reached exceed the budgeted amount.

The rise of these platforms has been unquestionable since 2006, when the Sellaband platform was created, aimed at raising funds for musical projects. It was the first online platform with an interface just as it is known nowadays. Since then, its number has not stopped growing; 2011 ended with 348 platforms [5] and 2017 closes with more than 2000 active platforms [6] and a large amount of money raised; Kickstarter, the industry’s most popular platform, has accumulated a total revenue of just over \$4 Billion [7].

Despite this positive data, crowdfunding has not reached maturity, since the failure rate is far from improving over the time. According to a study by Professor Mollick, based on an extensive database of Kickstarter projects between 2009 and 2012, the average success rate during that period was 48.1% [8]. The data that the platform shows in its statistics are even worse, starting with success rates in 2010 and 2011 of 43% and 46% respectively, and finishing with the last couple of years 2015 and 2016 with 36% and 32% success. 2016 stands out as the first year in which there is a recess in the number of projects proposed for funding, 14.5% less than the previous year [9]. The success rate results of Indiegogo, the second most popular crowdfunding platform, are even lower, as it closed 2015 with a success rate of less than 10% [10].

The sudden popularity of crowdfunding attributed to the impact of the most successful projects, has led to generalize this model to any new project with the risk of maladjustment. In spite of the improvements introduced in the platforms, the apparent facility to obtain financing leads creators to pursue this model, even when there are no real conditions to be successful. To increase success, platforms try to avoid projects with certain associated risks by frequently updating their standards, as they are the most interested part in managing successful projects that make a profit [11].

Leaving aside considerations of taste, aesthetics or fashion, it is essential to know, both for platforms and for entrepreneurs, which criteria represent a greater guarantee of success, in order to avoid unnecessary work, resource consumption and bad image.

The greatest uncertainty for a creator of these projects is whether the required financing will be guaranteed or not, due the lack of reliable methods or tools to direct or reorient a project towards success, as is discussed in Section 2.2. The main motivation for this study is the search for a tool based on the key factors of crowdfunding, these will allow a creator to adapt or modify the trajectory of a project based on a series of “pattern” scenarios with known characteristics. This interest is also stimulated because there is no tool capable of identifying which factors should be modified to guide the success or failure of a particular type of project. This contribution not only benefits the sponsors of the project by ensuring a product, but will also benefit the creator and the platform that hosts the project, contributing to the project funding sustainability. One of the foundations that allows the sustainable growth of this financing model is the constant search for satisfaction in both directions, through the reciprocity provided by this environment [12].

To solve this problem, we performed a retrospective analysis or AAR (After Action Review) of the project results, whether successful or failed, following the guidelines of the Project Management Institute (PMI), to determine the causes. Knowing how, why or what decisions have been taken to achieve the success of the project is crucial if we are to standardize the actions that ensure success or keep the project from failure [13].

We will use a data mining tools on a data set described in Section 4.3, to induce new knowledge from a historical compilation of other projects [14]. For this purpose, it is necessary to perform some

preliminary steps of exploration and cleaning in order to limit the problem to our needs. Later, there will be an AAR analysis of the data resulting from the closing of projects in crowdfunding environments.

This process will allow for the grouping of common factors in successful and failed projects, as shown in Sections 5.1 and 5.2, giving rise to six different types of project stereotypes, characterized by their most influential factors. The advantage for the creator of a project is clear, as comparative models are provided on which to associate any crowdfunding project, discussed in Section 6, allowing us to set a strategy and modify the project variables based on the most appropriate model, even after the funding period has started.

2. Theoretical Framework

2.1. What Is Crowdfunding?

Taking as a starting point the previous idea of “crowdsourcing”—with which it is possible to outsource a work to a group of people through an open call [15]—Michael Sullivan cited the term “crowdfunding” for the first time in his blog Fundavlog, based on the principle of collective collaboration [16]. Later and more accurately, it was defined as: “An open call, primarily through the Internet, for the provision of financial resources, either in the form of a donation or in exchange for some kind of reward and/or voting rights in order to support initiatives for specific purposes” [17].

The advantages of using a crowdfunding platform are several; from speeding up and making bureaucratic procedures cheaper [18], to carrying out matchmaker functions and serving as an advertising window [19]. They are also in charge of supplying the precious resource to the promoter of the idea, only if a certain economic amount is reached [17].

In order to achieve this type of financing, it is necessary to offer a new product or one with a high added value because it is innovative, or other factors differentiate it from the rest, or it is offered at a price clearly below the market price [8]. For this reason, the use of techniques to maximize resources as much as possible—like Bootstrapping techniques—is inexcusable [20].

It is possible to develop crowdfunding techniques through different means, although nowadays it could not be imagined without the use of the internet, both for its ease of access and distribution capacity [8]. The project stakeholders can benefit from the internet while using new communication channels and multiplying the potential of groups and collaborations, simplifying the development of many tasks [21].

2.2. Literature Review

The problem of predicting success or failure not only involves the creator, but also the platforms; for this reason, it is increasingly common to direct the study towards the behavior of the platforms, as their benefits are proportional to the success of the projects [11].

One of the first important contributions using data sets from Kickstarter and supported by the theoretical framework known to date, is a common behavior observed that in most successful projects, it has been observed that the projects that exceed the minimum amount of money for development, do so with a very small profit margin. On the other hand, projects that fail are usually far from reaching this minimum amount, as evidenced by [8] and [22].

It has also been found that a high money collection at the start of the funding campaign is linked to the success of the projects, which underlines the importance of promoting the project from its inception [23]. This behavior is explained by the “U” shaped pattern, observed in a high number of successful projects [24]. As well as the large initial accumulation, there is usually a strong increase in contributions happening at the end of the campaign in these cases, attributed to the fear of the product running out toward the end of the collection. To explain this, Kuppuswamy and Bayus highlight the strong importance of backers for the success of projects adding an explanation linking this fact to the spectator effect [25], which claims that a potential promoter does not contribute to a project of its interest, under the assumption that others will provide the necessary funds [26].

There are also certain features not directly associated with the project, such as the personal attributes of the creator, which are able to condition the decision of the crowd and increase the success rate [27]. Using a crowdfunding data set to connect with external data from social networks such as Facebook or Twitter, Mitra and Gilbert focused their research on the language used in the projects, showing how using persuasion and certain phrases and words reduced the failure rate significantly [28].

The influence of time periods on crowdfunding campaigns has also been studied in depth, Etter et al. extracted characteristics from a data set of 16,000 projects to create success predictors and observe the variation in success rate based on time series, finally reaching 76% accuracy, using a predictor only valid for the first four hours of the campaign [23]. Other authors focus the interest of their studies on shorter periods with the purpose of a taxonomic study [29] or the creation of tools that could lead to the project's success [30].

The specific use of project updates during publication as well as promotion activities in social networks have shown a high correlation with success [29]—especially in the initial stages, where social impact is most important [22]. It is not surprising that one of the most common reasons that lead to the failure of projects is the inability of creators to connect with investors [8], so there are proposals that recommend investors use Twitter, since the importance of this social network as a powerful tool for improving the success of projects has been observed, as evidenced by [31] and [32].

3. Crowdfunding as a Project-Financing Method

3.1. Fundamentals of Crowdfunding

The economic growth of a country is not understood unless it is supported by technological development and innovation, in order to gain adaptability by incorporating new techniques and increasing the efficiency of processes [33]. Even in periods of financial restriction, investment in innovation can be a stimulus for companies aimed at reducing production [34].

Environmental conditions are key to generating confidence in an investor, as external factors can contribute negatively and even lead to the failure of a technically successful project [35]. In an environment of economic crisis, projects with high levels of innovation cause an aversion for a non-specialized investor, due to the risk of the unknown and the uncertainty of the novelty itself [3]. Thus, these projects are often the most fragile and punished.

These economic constraints have led to the birth of the crowdfunding functional ecosystem [36], thanks to a considerable reduction in the sense of risk by dividing the economic target into many smaller parts, as many as backers. In contrast to traditional financing based on project portfolio, in a new or an emerging company have not an organisational structure to manage the financing [37].

The figure of the single investor characteristic of traditional of project finance models becomes a multitude of potential investors, who can collaborate in the development of a project with a small economic contribution, with the advantage that there is no intermediary, only a virtual platform and its managers. This small contribution becomes actual only if a minimum amount is reached, which is fixed beforehand. Otherwise, the amount contributed is returned. This system, known as “All-Or-Nothing”, contrasts with other known models and is used on most platforms of this type because it requires more involvement on the part of the entrepreneur and reduces the risk for the project collaborator [38]. On the other hand, it also provides a sign of confidence to potential contributors to the project since this system reveals the funds retained from the rest of contributors [39].

Table 1 compares the main attributes that differentiate traditional financing methods and crowdfunding.

Table 1. Main differences between traditional funding and crowdfunding.

	Traditional Funding	Crowdfunding
Type of project	Need to select the most appropriate funding model to project	Valid for any innovative project
Viability study	Necessary risk analysis and financial feasibility study	Not necessary
Funder	Independent entity	Customer
Cost of fund	Percentage fixed in advance	Proportional to the funding success
Project protection	High confidentiality	High risk of rights violations due to public exposure
Expandable	Yes, with solvency requirements	No

3.2. Challenges of Crowdfunded Projects

To carry out ambitious projects or innovation through mass funding may be a not so obvious need for a creator, due to the skepticism caused by the direct loss of control, with the inclusion of new elements such as the uncertainty of non-compliance [40]. In addition to the intrinsic danger of using the internet as a medium, which involves accepting the challenge of navigating a very active environment susceptible to change, constantly stimulated by collective trends, updates, etc., this fact is also able to threaten projects with important innovation factor in their products, such as highly technological ones [41].

Part of the solution to these problems is mitigated by making use of the high capacity for sharing on blogs [42]. Blogs open the door to a multitude of users, from enthusiasts to specialists, who can provide solutions to various problems [43], and even generate valuable feedback during the development of the project, thus making it possible to improve the adaptation of the product to the environment and the user, complementing behavioral, technical and contextual skills [44].

On the other hand, sharing ideas or knowledge without any type of protection, intellectual or any other, increases the risk of plagiarism [19], which is further aggravated if the project does not reach the minimum necessary amount, since it cannot be funded despite having been publicized both on the platform itself and on social networks [45], once the funding is cancelled, it is not possible to present the project on the same platform, because one of the admission criteria is the originality of the projects.

4. Research Methodology

4.1. Research Settings

The use of data science techniques allows us to relate each factor with the final success, which is particularly useful in sectors where it is necessary to study a huge amount of data [46]. Its potential lies in the capability to extract information beyond a simple relationship between variables, using statistical analysis software among others [47]. To perform this process, we will use the CRISP-DM methodology (Cross Industry Standard Process for Data Mining), because it is flexible and sufficiently tested to allow for the conversion of data into knowledge in an organized way [48]. It consists of six iterative phases ensuring that each analysis process benefits from previous experience.

Once the data set has been adapted to the needs of this study, the projects will be grouped based on their similarity. This new structure allows a new characterization while expanding information to create self-organized maps, which by their visual and abstract nature help to identify areas of projects that share properties and then analyze them individually. Subsequently, it is necessary to use indicators based on success and failure that will contribute greatly to the characterization of each of the stereotypes.

The application of the methodology used is focused on providing fidelity to the techniques used, and avoiding the use of variables that add uncertainty to the results. In this way, the value of the existing information is increased by adding new links between the factors that condition success, greatly facilitating the characterization of a project according to its properties and most influential factors.

4.2. Data Collection

The selection of Kickstarter as a data source is attributed to its leadership and wide popularity, as it appears in the top five hundred most searched websites in the world [49], with a contribution to the success of more than 167,000 projects since its creation [50].

We use a public raw data set, which has been used partially by several authors and includes a significant time window for crowdfunding, in which there is a sustained growth in the number of projects funded [9]. This data set is characterized by gathering information without a clearly defined objective and includes 45,815 projects developed between March 2009 and January 2012, arranged into 12 categories: Art, Comics, Dance, Design, Fashion, Food, Games, Music, Photography, Publishing, Technology and Theater [51].

For this study, only the projects that have been performed are selected, those that did not reach the end are not part of the study and therefore those that are alive or have been cancelled by the platform or the creator are left out and without opportunity as explained at the end of Section 3.2. As a result, a data set with 23,941 projects highly representative for this study and in which 55.28% of the projects are successful is obtained. Table 2 shows the number of projects by category, success rate and number of backers.

Table 2. Main features of the categories.

Category	Number of Projects	Success Rate	Number of Backers
Art	3304	57.35%	132,920
Comics	868	53.69%	77,752
Dance	590	72.71%	21,922
Design	1416	47.81%	317,998
Fashion	927	33.55%	29,437
Food	1167	51.67%	69,866
Games	929	36.49%	379,453
Music	7238	66.37%	387,692
Photography	1266	45.97%	39,479
Publishing	3631	41.23%	145,083
Technology	700	39.71%	97,696
Theater	1905	71.02%	78,703

However, some of the projects that collect over the minimum necessary do it a very small margin [8], so high collection values have not been deleted in order to analyze whether this subset can form a group with common characteristics, unlike other authors who do not consider them relevant for their study [52].

4.3. Variables Review

Eight numerical attributes are selected from the original data set, specifically those which can provide more information without limiting the model development. In order to enhance the magnitude of the study, new attributes capable of providing new information have also been created. In summary, the data set consists of 13 useful attributes, shown in Table 3 with their description.

According to its value during the financing period, three types of attributes are identified:

- Fixed variables. The project creator must assign their values before the funding campaign: Goal, Month, Levels, Duration, Min_RL, Max_RL, Range_RL.
- Development variables. They are only modified with the efforts of the creator during the development of the campaign: Comments, Updates and Backers.
- Indicator variables. They are dependent on the previous ones and give an idea of the final impact of the financing: Pledged, Rate_Pledge_Goal, Pledged_Backer.

Table 3. Data dictionary of the data set.

Variable	Range	Description
Goal	0.01–1,200,000	Minimum amount required in dollars
Pledged	0–10,266,845	Maximum amount reached in dollars
Rate_Pledge_Goal ¹	0–10,000	Achieved Goal Percentage
Backers	0–87,142	No. of contributors to each project
Levels	0–80	No. of contribution levels
Updates	0–142	No. of project updates
Comments	0–19,311	No. of feedbacks from each contributor
Month	1–12	Campaign end month
Duration	1–92	Total duration of funding in days
Min_RL ¹	0–5000	Minimum possible contribution in dollars
Max_RL ¹	0–10,000	Maximum possible contribution in dollars
Range_RL ¹	0–9999	Difference between Max_RL and Min_RL
Pledged_Backer ¹	0–3757	Average amount provided by Backer

Notes: ¹ New attributes result of the transformation of the existing features.

Mitra and Gilbert assessed a binary variable for their study, which identified which projects were successful or failed [28]. In this case, success is defined as the projects with a pledged value equal to or greater than the goal, and failure as the opposite case.

As a first contact with the data, the linear correlation of the attributes of the data set is studied. Figure 1 shows the correlation of variables selected for this study.

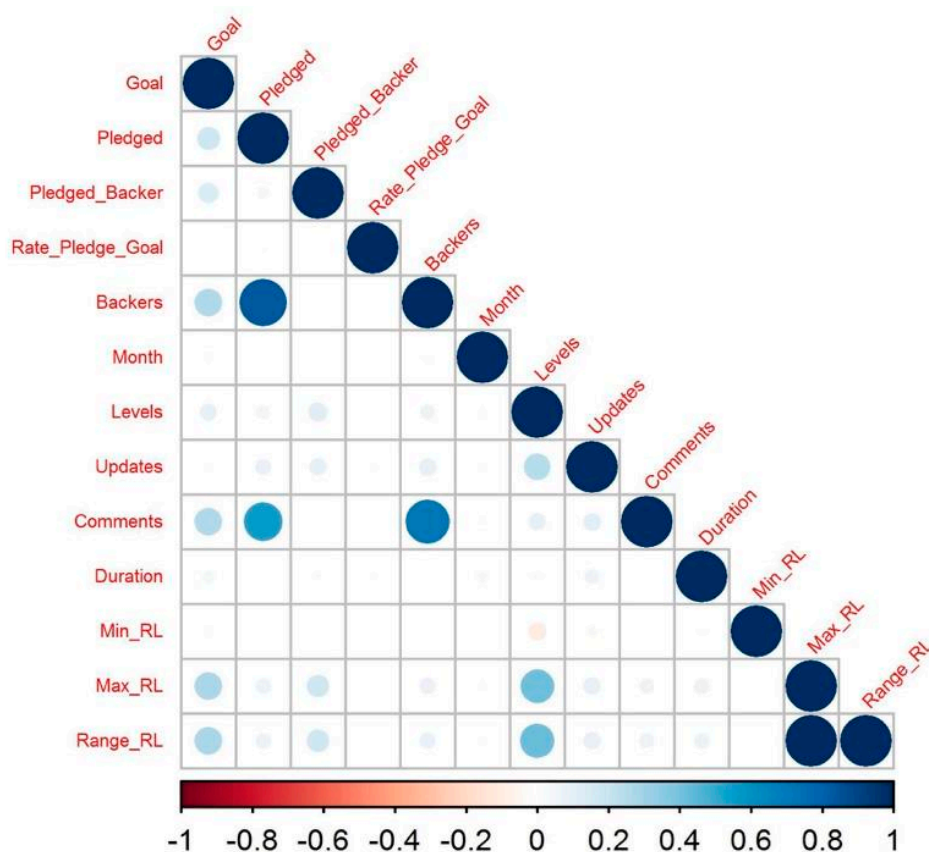


Figure 1. Correlation for key variables.

The high correlation between Backers, Pledged and Comments has already been observed previously, albeit partially [22,53].

The apparent correlation between Range_RL and Max_RL, as well as the absence of it between Rate_Pledge_Goal and the rest of the variables, denotes the need to use more advanced methods that provide more information than linear ones, such as the use of multivariate techniques.

5. Clustering Methodology

5.1. Justification of Use

The generalization in data collection allows the development of sophisticated techniques for extracting knowledge from information and data. These techniques are embedded within the KDD (Knowledge Discovery in Databases) analysis stage and are known as data mining techniques. The use of these techniques enables subsequent clustering of projects for labeling, following the principle of “maximizing the intraclass similarity and minimizing the interclass similarity” [14].

This approach is based on the behavior of artificial neural networks, where the process starts with a project as an input and, through a competitive process associated with each neuron, generates a new vector called centroid, which is representative of all the projects related by this process.

Subsequently, in order to group the input in k sets, the “K-means” algorithm is chosen as the most appropriate clustering model due to its non-hierarchical nature [54]. The Davies–Bouldin index is used to identify the optimum value of k . This index can be interpreted as the distance of each case to the newly identified cluster. The index decreases when the items in each cluster are more homogeneous [55]. Following this criterion, the algorithm distinguishes seventy neurons distributed in six clusters as the optimal value of k .

In Figure 2, each of the six clusters obtained is shown by colors. This is the result of applying the K-means technique on the grid of representative centroids identified using SOM.



Figure 2. K-means cluster map.

5.2. Analysis of Success and Failure

Starting from the neuron grid generated for the k-means cluster map, the number of projects labeled as success or failure is displayed. To allow the understanding of the chart, the size of each cell has been drawn proportionally to the number of projects it contains. In this way, Figure 3 shows the two projections made on Figure 2, one of success in green and failure in red. By superimposing these on the k-means cluster map it is possible to identify which zones or which clusters have more success or failure.

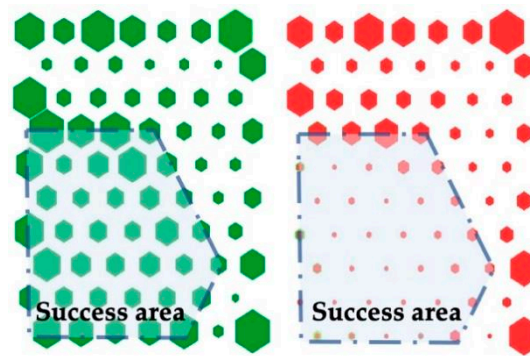


Figure 3. Success (green) and Failure (red) projection.

When comparing the two charts, there is an area of high interest, since it contains a higher number of successful projects and a reduced number of failed projects—the area in the lower left of the chart. This section is defined as a “Success Area”, and can be associated with clusters 3 and 2 in Figure 2. Both clusters make up 43.64% of the total of successful projects, with a relation ratio of 76% of success and only 24% of failure, which is clearly differentiated in the success–failure ratio in the initial data set.

Likewise, another section of the chart can be defined as a “Failure Area” (top section), that includes clusters 1 and 5 where 58.90% of all failed projects are located, with a 40% success rate and 60% failure rate, representing the inverse behavior to the initial set.

5.3. Cluster Taxonomy

Once the neurons that belong to each cluster have been identified, a first characterization of clusters is carried out based on the number of projects that continue and the proportion of successful and failed ones—the authors commonly call it success rate.

Table 4 shows the projects classified by cluster and success rate. Clusters 2, 3 and 4 are highlighted for surpassing the success rate, while the rest of the clusters are below.

Table 4. Distribution of the success rate for each cluster.

Cluster	Number of Projects	Success Rate
1	5367	40.13%
2	5114	68.71%
3	2474	91.43%
4	2740	65.04%
5	5137	39.77%
6	3109	47.57%

Since the sample does not show a normal distribution, it was decided that we should use the non-parametric Kruskal–Wallis test, in order to test whether the samples presented the same distribution, and identify if there were significant differences between clusters.

The results obtained with all the attributes determined that there were statistically significant differences between the distribution of each variable among the clusters. Therefore, the behavior of the attributes that define each cluster can be analyzed.

Table 5 shows the location in each cluster of the maximum and minimum average values for each variable—the empty cells indicate that there is no variable taking the highest or lowest values in that cluster.

Table 5. Distribution of lowest and highest average variable values. (-, empty cells).

Cluster	Values	Studied Variables
1	Lowest Highest	Rate_Pledged_Goal Duration
2	Lowest Highest	- Rate_Pledged_Goal
3	Lowest Highest	- Pledged, Comments, Updates, Backers
4	Lowest Highest	Min_RL Levels, Pledged_Backer
5	Lowest Highest	Pledged, Levels, Duration, Goal, Comments, Updates, Max_RL, Backers, Pledged_Backer Min_RL
6	Lowest Highest	- Goal, Max_RL

Using the duration attribute as a guideline, the projects with the highest average duration are classified in cluster 1, whereas those with the lowest duration are classified in cluster 5. The same reasoning can be applied to the rest of the attributes.

5.4. Self-Organizing Map Analysis

The presence of similar behavior between projects is analyzed by the generation of clusters using Self-Organizing Maps (SOM), a technique introduced by Teuvo Kohonen [56], which has been previously used to determine the success or failure of projects, based on determining the project characteristics of groups [57].

The SOM model allows us to cluster a new project in a grid area and associate it with the trend of success or failure. However, it is necessary to characterize each cluster in order to know the suitability of a project according to the cluster to which it belongs.

In order to extend the capabilities of the study, starting from the results of the SOM, a graph is created using the U-matrix, in which the distance between the centroid of each neuron and its closest neighbors is represented. It can be read that the low values represent a high degree of similarity between neurons in that region.

Figure 4 consists of the SOM model U-matrix followed by thirteen charts, one for each variable. In the chart of the variables, each cell color represents the value taken by that variable in the centroid of the neuron. This rendering allows the comparison of one or several attributes through the grid.

The maps for Backers, Pledged, Comments and Updates take the highest values in the lower left corner, where there is also less similarity or greater distance between the centroids, as indicated by the U-matrix.

The decentralization of the Goal variable should be highlighted—taking high values in two different groups. The zone with higher Goal values is related to zones of high value in Max_RL and Range_RL and, although to a lesser extent, to zones of high value in Pledged_Backer and Levels. On the other hand, the area with not so high Goal values corresponds to high values of Backers, Updates, Pledged and Comments, which complement the importance of the Updates during the collection campaign, a trend observed by other authors [29].

The attribute Rate_Pledge_Goal—or success quantity—although it is calculated from the information of Pledged, shows no relationship in the maps between them. This means that this new variable provides additional information to that initially provided by Pledged.

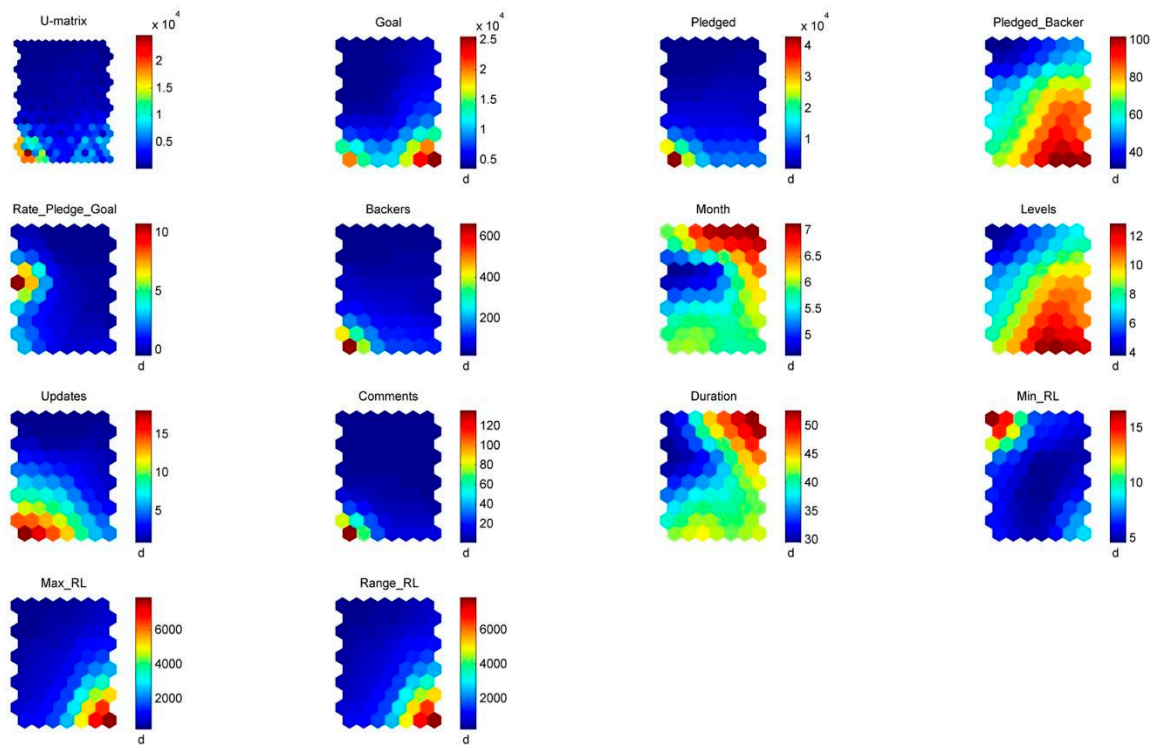


Figure 4. Map of Variable distribution.

5.5. Cluster Success Characterization

After analyzing the cluster-attribute performance, the relationship between the different project categories and each cluster is considered. For this purpose, Table 6 shows the percentage of successful projects in each cluster by category, with green indicating values over 50% and red indicating values below.

Table 6. Distribution of success by category and cluster.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Art	43.65%	72.95%	91.53%	64.41%	46.76%	44.64%
Comics	28.44%	66.47%	92.20%	42.67%	34.07%	39.34%
Dance	63.95%	88.10%	90.91%	86.05%	64.74%	44.19%
Design	22.22%	52.22%	91.54%	48.12%	29.29%	42.17%
Fashion	19.69%	41.74%	82.76%	53.76%	16.23%	33.60%
Food	33.33%	67.42%	91.23%	58.38%	30.05%	42.40%
Games	10.71%	42.62%	83.96%	30.14%	16.03%	24.07%
Music	52.29%	79.97%	96.79%	77.00%	46.26%	64.74%
Photography	32.14%	59.20%	88.11%	51.56%	30.24%	38.98%
Publishing	26.08%	57.47%	87.84%	48.42%	27.28%	29.95%
Technology	20.00%	50.00%	93.46%	38.00%	16.36%	30.22%
Theater	61.98%	83.68%	92.73%	76.19%	67.40%	52.88%

Notes: Highlights that the success of projects associated with the performing arts (Dance, Music and Theater) is spread among neighboring clusters, hindering category–cluster association.

This success distribution provides valuable information for subsequent cluster characterization but does not consider the number of projects in each cell. In order to do this, indicators are defined to quantify the success distribution by category and cluster considering the overall category success.

Table 7 provides the total number of projects by category and cluster which, together with Table 6, allows for a better understanding of the distribution of success.

Table 7. Distribution of projects by category and cluster.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Art	669	706	295	354	879	401
Comics	225	167	205	75	135	61
Dance	147	168	33	43	156	43
Design	261	316	260	133	280	166
Fashion	193	230	58	93	228	125
Food	213	267	114	173	183	217
Games	168	183	187	73	156	162
Music	1658	1473	592	1165	1258	1092
Photography	336	250	143	128	291	118
Publishing	882	736	370	285	964	394
Technology	160	134	107	50	110	139
Theater	455	484	110	168	497	191

At first, two indexes are defined, which allow for the monitoring of both successful and failed projects. SRI (Success Rate Index) assesses the number of successful projects in relation to those that have failed, whereas FRI (Fail Rate Index) assesses the number of failed projects with respect to those that have been successful. In both cases, they are calculated for each category and cluster.

$$SRI_{i,j} = \frac{S_{i,j}}{F_{i,j}} \quad (1)$$

$$FRI_{i,j} = \frac{1}{SRI_{i,j}} \quad (2)$$

Let i denote the list of twelve project categories, and let j denote the six clusters. S and F correspond to the number of successful projects and the number of failed projects, respectively.

In order to increase the significance of SRI and FRI, two other indexes are defined: SOA (Success ratio Over the Average) and FOA (Fail ratio Over the Average). High SOA values in a cluster indicate that a category has a success rate above the overall average; similarly, high FOA values indicate that a category has a failure rate above the overall average. They help us to understand whether a category is highly successful or not, depending on the total success or failure by category.

$$SOA_{i,j} = \frac{SRI_{i,j} - \overline{SRI}_i}{\overline{SRI}_i} \quad (3)$$

$$FOA_{i,j} = \frac{FRI_{i,j} - \overline{FRI}_i}{\overline{FRI}_i} \quad (4)$$

Let \overline{SRI}_i denote average of success rate index, and \overline{FRI}_i denote average of fail rate index in the i category.

Table A1 in Appendix A shows the distribution of the mean values obtained for SOA and FOA by category and cluster, and Table A2 in Appendix A shows the mean, median and standard deviation taken by the SRI, FRI, SOA and FOA indicators in each cluster. It should be noted that SOA and FOA are shown as percentages because they are relative values calculated according to SRI and FRI, respectively.

Figure 5 represents SOA indexes, where cluster 3 stands out for having a success rate above average in all project categories. The technology category stands out for having the highest values of SOA, surpassing the success rate of the cluster by more than four times in this particular category. With regard to the rest of the clusters, attention should be drawn to Dance in clusters 2 and 4, and Theater in cluster 2.

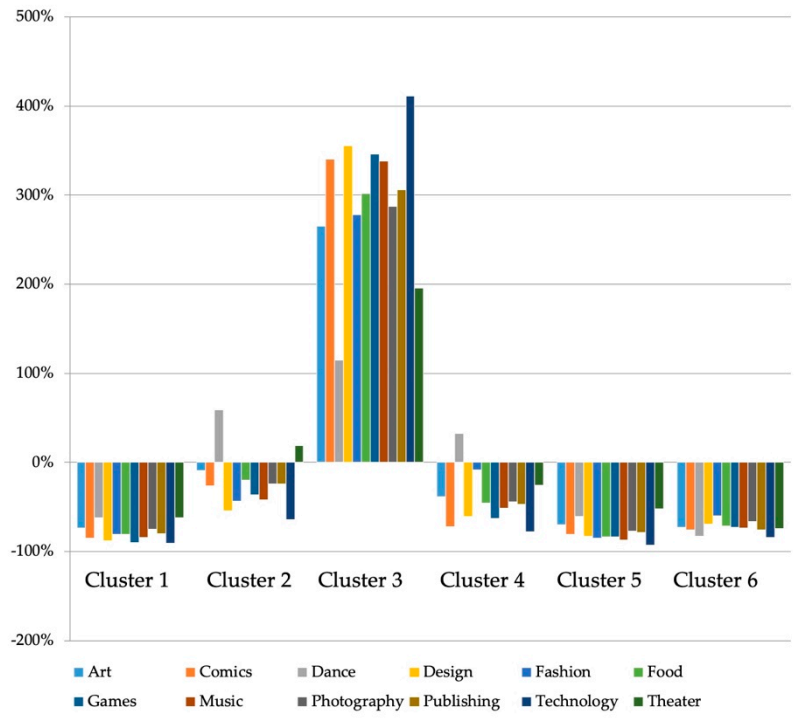


Figure 5. Success ratio Over the Average by category.

In contrast with the previous indicator, in the case of FOA in Figure 6, clusters 1 and 5 stand out as the most unfavorable, especially for the Games category—which has the highest failure rate—as well as Fashion category in cluster 5.

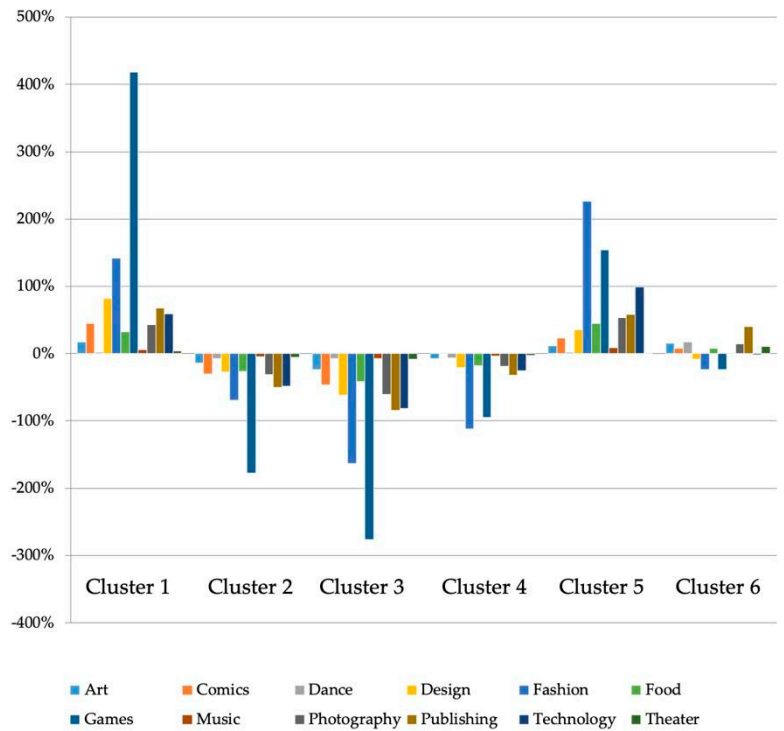


Figure 6. Fail ratio Over the Average by category.

6. Discussion

Six clusters have been identified by similarities in the characteristics of the attributes. They can be summarized in the two major groups identified in Figure 3 as a success area. The greatest number of projects that achieve the success and the least number of failures are in that area.

Subsequently, relative indices have been calculated which make it possible to identify the extent of success and failure with respect to that achieved by the category average.

The strategy used to carry out the discussion of the paper consists of analyzing the attributes that define each one of the clusters, taking information from Table 5, as well as the behavior of the SOA and FOA indicators in each one of the clusters, information that is taken from Figures 5 and 6.

Merging these two information sources allows us to know the characteristics of each group, and which success or failure index would be expected by each category inside them.

It starts by defining the behavior of each cluster by the representative attributes in each one.

Observing the Pledged, Comments, Updates and Backers projections in the maps of Figure 4, the similarities in the distribution of these variables are appreciated, since the highest values are grouped in the lower left corner of each map, which also corresponds to cluster 3 in Figure 7. Cluster 3 is associated with a likely success of the project, since the SOA index is very positive for all categories, as well as no FOA index being positive.

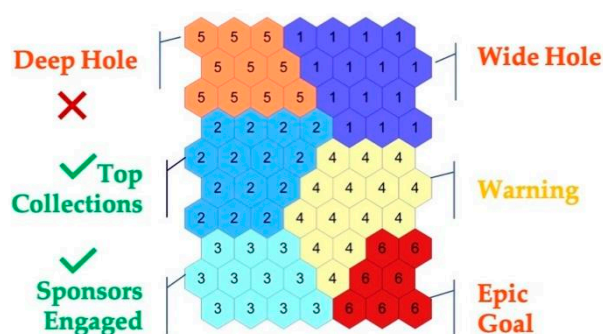


Figure 7. Cluster characterization.

This relationship underscores the importance of maintaining high values in these attributes in order to set the goal of success. Three of these attributes are classified as “Development Variables”, which mean that they can be modified during the development of the funding campaign to strengthen the project and induce success. For all these reasons, this is known as the “Sponsors Engaged”.

The characteristics of cluster 3 are indisputable when aiming to prioritize the success of projects, but not to maximize profits. Even for categories associated with the performing arts, where success is scattered, it is in cluster 3 where the highest SOA values are found.

As shown in Table 5, cluster 2 contains the highest average collection values for Rate_Pledged_Goal, as well as the second highest average SRI value (2.47). Therefore, cluster 2 is renamed “Top Collections”. This is reflected in the fact that the FOA index never shows positive values, but only Dance and Theater categories have a slight activation of the SOA index, having a significant success with respect to the average of the category.

Cluster 4 has a success rate of 65.04%, which can be considered as an acceptable risk, keeping in mind the innovative nature of this type of projects. It has the third greatest average SOA value and the highest values of Pledged_Backer as shown in Table 5. This is then referred to as “Warning”.

Most projects contained in clusters 1 and 5 fail, with success rates of around 40%. They also have the highest average FOA values, so it has been decided to highlight them as groups to be avoided. These two clusters have certain similarities, since they contain a similar number of projects and success rates of 40.13% and 39.77% respectively. Both clusters are not particularly suitable for hosting projects in the Games and Fashion categories, as both have the highest FOA values.

The projects located in cluster 1 have the highest average duration and the lowest Rate_Pledged_Goal values. Cluster 5 is characterized as having the lowest averages of Pledged, Comments, Updates, Backers and Duration. This situation is of a disadvantage to the Technology category, presenting the highest values of FOA. In order to differentiate the failure of these clusters, the duration attributes will be used, naming them “Wide Hole” and “Deep Hole” respectively.

Cluster 6 has a success rate lower than 48% as can be seen in Table 4, as well as SOA values comparable to those obtained in Cluster 1 and 5. Although it contains projects with a high risk, its most representative feature is its high Goal, as shown in Table 5. Thus, Cluster 6 has been renamed as “Epic Goal”.

After this discussion, we have found the attributes that define the success of each cluster. We suggest that these attributes are a consequence of the behavior of the cluster. For example, belonging to cluster 3 indicates that a project will have many followers, but there is no guarantee that the consequent will be able to deduce the antecedent—that is, the presence or absence of that attribute does not determine cluster membership.

7. Conclusions

Crowdfunding arises as a response to the problem of financing innovative projects in an environment of strong economic crisis, the paralysis of the growth of this model motivates the interest of this study towards a help tool that makes this model more sustainable. For do this, it was selected a sufficiently representative data set of projects known from other authors, and modern data mining techniques were applied. As a result, our conclusions may add knowledge to this means of acquiring resources in this very representative period of time.

The association of projects in clusters has been decisive in understanding how they work, by grouping and characterizing them by their particularities. An examination of the data set has revealed that it can be distributed over six different clusters. This assignment of projects to clusters allows any project to be subsequently incorporated into its corresponding cluster, making it easier for a creator to define a strategy or reorient a project in order to drive it to success, based on its position in the system.

Clusters known as “Top Collections” and “Sponsors Engaged” are the most suitable to host a project with the maximum potential for success, being characterized by a collection much higher than expected and a strong commitment between the Backers and project through good two-way communication. The name “Top collections” makes it easier to identify the cluster with the highest amount collected above the goal, and “Sponsors Engaged” identifies a cluster with projects more closely linked to communication with backers.

The cluster called “Warning” has a sufficient enough success rate to allow the creator to safely modify the characteristics of his project and improve its potential for success.

The “Deep Hole” and “Wide Hole” clusters are characterized by a very high failure rate. The first of them contains projects with the lowest values in certain basic attributes for the good development of crowdfunding—Comments and Updates, among others. The second cluster groups projects with the longest average duration. Both are considered as clusters to be avoided by any creator, which require the maximum effort to relocate them to the success zone. To facilitate the identification of the most inappropriate clusters to a creator, the designation “Hole” has been chosen. The “Deep hole” cluster contains projects with characteristics that are very difficult to modify to redirect the project to success, and the “Wide hole” cluster is understood as a cluster with a greater number of failed projects, though its characteristics allow a greater margin of modification than “Deep Hole”.

The relevance of having identified these six clusters and their attributes allows project managers to use a tool that facilitates the estimation of the economic and financial viability of the crowdfunding project being undertaken. In this way, resources and efforts can be dedicated to improving the quality and benefit ratios of the project.

Author Contributions: Conceptualization, A.F.-B. and J.V.-B.; Formal analysis, H.M.-P.; Investigation, A.F.-B.; Methodology, V.R.-M.; Supervision, V.R.-M.; Writing—original draft, A.F.-B. and J.V.-B.; Writing—review and editing, A.F.-B. and H.M.-P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Science, Technology and Innovation Plan of the Principality of Asturias (Spain) Ref: FC-GRUPIN-IDI/2018/000225, which is part-funded by the European Regional Development Fund (ERDF).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Distribution of of SOA and FOA mean values by category and cluster.

	1		2		3		4		5		6	
	SOA	FOA	SOA	FOA	SOA	FOA	SOA	FOA	SOA	FOA	SOA	FOA
Art	−0.74	0.17	−0.09	−0.14	2.65	−0.23	−0.39	−0.08	−0.70	0.12	−0.73	0.16
Comics	−0.85	0.45	−0.26	−0.30	3.40	−0.46	−0.72	0.01	−0.81	0.23	−0.76	0.08
Dance	−0.62	0.02	0.59	−0.07	1.15	−0.08	0.32	−0.06	−0.61	0.02	−0.83	0.17
Design	−0.88	0.82	−0.54	−0.27	3.55	−0.62	−0.61	−0.20	−0.83	0.36	−0.69	−0.08
Fashion	−0.81	1.42	−0.44	−0.70	2.78	−1.63	−0.08	−1.12	−0.85	2.27	−0.60	−0.24
Food	−0.81	0.32	−0.20	−0.26	3.02	−0.41	−0.46	−0.17	−0.83	0.45	−0.72	0.08
Games	−0.90	4.18	−0.37	−1.78	3.46	−2.76	−0.63	−0.95	−0.84	1.54	−0.73	−0.24
Music	−0.84	0.06	−0.42	−0.04	3.38	−0.07	−0.51	−0.03	−0.87	0.09	−0.73	0.00
Photography	−0.75	0.43	−0.24	−0.31	2.88	−0.60	−0.44	−0.18	−0.77	0.53	−0.67	0.14
Publishing	−0.80	0.68	−0.24	−0.50	3.06	−0.84	−0.47	−0.32	−0.79	0.58	−0.76	0.40
Technology	−0.91	0.59	−0.64	−0.48	4.11	−0.82	−0.78	−0.26	−0.93	0.99	−0.85	−0.02
Theater	−0.62	0.04	0.19	−0.05	1.95	−0.08	−0.26	−0.03	−0.52	0.01	−0.74	0.11

Table A2. Statistic values of SRI, FRI, FOA and SOA by cluster.

	SRI			FRI			FOA			SOA		
	Mean	Median	Sd	Mean	Median	Sd	Mean	Median	Sd	Mean	Median	Sd
Cluster 1	0.66	0.44	0.55	2.73	2.31	2.15	76.39%	43.70%	114.71%	−79.41%	43.70%	114.71%
Cluster 2	2.47	1.72	2.06	0.67	0.60	0.43	−40.97%	−28.81%	47.47%	−22.19%	−28.81%	47.47%
Cluster 3	11.31	10.60	6.61	0.11	0.09	0.05	−71.74%	−53.26%	77.97%	294.87%	−53.26%	77.97%
Cluster 4	1.82	1.11	1.66	0.94	0.90	0.62	−28.31%	−17.89%	36.58%	−42.02%	−17.89%	36.58%
Cluster 5	0.70	0.43	0.63	2.54	2.32	1.74	59.92%	40.41%	68.87%	−77.91%	40.41%	68.87%
Cluster 6	0.75	0.69	0.40	1.63	1.46	0.71	4.70%	7.89%	17.84%	−73.34%	7.89%	17.84%

References

- Bustos Contell, E. *La Financiación de la Pyme: Préstamo Participativo vs. Capital Riesgo*; Aeca: Madrid, Spain, 2013; ISBN 978-84-15467-68-7.
- Behringer, N.; Sassenberg, K. Introducing social media for knowledge management: Determinants of employees' intentions to adopt new tools. *Comput. Hum. Behav.* **2015**, *48*, 290–296. [[CrossRef](#)]
- Lee, N.; Sameen, H.; Cowling, M. Access to finance for innovative SMEs since the financial crisis. *Res. Policy* **2015**, *44*, 370–380. [[CrossRef](#)]
- Negruşa, A.L.; Rus, R.V.; Sofică, A. Innovative Tools Used by Business Networks and Clusters in Communication. *Procedia-Soc. Behav. Sci.* **2014**, *148*, 588–595. [[CrossRef](#)]
- Belleflamme, P.; Lambert, T.; Schwienbacher, A. Crowdfunding: Tapping the right crowd. *J. Bus. Ventur.* **2014**, *29*, 585–609. [[CrossRef](#)]
- Galkiewicz, D.P.; Galkiewicz, M. *Crowdfunding Monitor 2018: An Overview of European Projects Financed on Startnexus and Kickstarter Platforms between 2010 and mid-2017* 2018; Bermag: Szczecin, Poland, 2018.
- Pardes, A. *Wired Magazine*; Conde Nast: New York, NY, USA, 2019.
- Mollick, E. The dynamics of crowdfunding: An exploratory study. *J. Bus. Ventur.* **2014**, *29*, 1–16. [[CrossRef](#)]

9. Bidaux, T. Kickstarter in 2017-Year in Review. Available online: <http://icopartners.com/2018/01/kickstarter-2017-year-review/> (accessed on 1 October 2018).
10. Zhao, Q.; Chen, C.-D.; Wang, J.-L.; Chen, P.-C. Determinants of backers' funding intention in crowdfunding: Social exchange theory and regulatory focus. *Telemat. Inform.* **2017**, *34*, 370–384. [[CrossRef](#)]
11. Belleflamme, P.; Omrani, N.; Peitz, M. The economics of crowdfunding platforms. *Inf. Econ. Policy* **2015**, *33*, 11–28. [[CrossRef](#)]
12. Zvilichovsky, D.; Inbar, Y.; Barzilay, O. *Playing Both Sides of the Market: Success and Reciprocity on Crowdfunding Platforms*; Elsevier: Amsterdam, The Netherlands, 2015.
13. PMBOK. *Guía de los Fundamentos Para la Dirección de Proyectos (Guía del PMBOK)*; Project Management Institute: Newtown Square, PA, USA, 2013; ISBN 978-1-62825-009-1.
14. Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*, 3rd ed.; Elsevier: Burlington, MA, USA, 2011; ISBN 978-0-12-381479-1.
15. Howe, J. The rise of crowdsourcing. *Wired Mag.* **2006**, *14*, 1–4.
16. Mary Anne, M.G. Everyone is a venture capitalist: The new age of crowdfunding. *Res.-Technol. Manag.* **2012**, *55*, 4–7.
17. Schwienbacher, A.; Larralde, B. *Crowdfunding of Small Entrepreneurial Ventures. Handbook of Entrepreneurial Finance*; Oxford University Press: Oxford, UK, 2010.
18. Tran, T.; Dontham, M.R.; Chung, J.; Lee, K. How to Succeed in Crowdfunding: A Long-Term Study in Kickstarter. Available online: https://www.researchgate.net/publication/305637974_How_to_Succeed_in_Crowdfunding_a_Long-Term_Study_in_Kickstarter. (accessed on 15 February 2017).
19. Gerber, E.M.; Hui, J. Crowdfunding: Motivations and deterrents for participation. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **2013**, *20*, 34. [[CrossRef](#)]
20. Lahm, R.J., Jr.; Little, H.T., Jr.; Hall, G. Bootstrapping business start-ups: A review of Current Business Practices. In Proceedings of the Conference on Emerging Issues in Business and Technology: Finance and Accounting: Special Topics (Entrepreneurship) Paper presentation, Citeseer, The Pennsylvania State University, Philadelphia, PA, USA, 2005.
21. Mishra, A.; Chandrasekaran, A.; MacCormack, A. Collaboration in Multi-Partner R&D Projects: The Impact of Partnering Scale and Scope. *J. Oper. Manag.* **2015**, *33–34*, 1–14.
22. Lu, C.-T.; Xie, S.; Kong, X.; Yu, P.S. *Inferring the Impacts of Social Media on Crowdfunding*; ACM Press: New York, NY, USA, 2014; pp. 573–582.
23. Etter, V.; Grossglauser, M.; Thiran, P. *Launch Hard or Go Home! Predicting the Success of Kickstarter Campaigns*; ACM Press: New York, NY, USA, 2013; pp. 177–182.
24. Strickler, Y. Shortening the Maximum Project Length. The Kickstarter Blog. Available online: <https://www.kickstarter.com/blog/shortening-the-maximum-project-length> (accessed on 12 June 2016).
25. Kuppuswamy, V.; Bayus, B.L. Crowdfunding Creative Ideas: The Dynamics of Project Backers in Kickstarter. *SSRN Electron. J.* **2013**, *5*, 1–37. [[CrossRef](#)]
26. Fischer, P.; Krueger, J.I.; Greitemeyer, T.; Vogrincic, C.; Kastenmüller, A.; Frey, D.; Heene, M.; Wicher, M.; Kainbacher, M. The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychol. Bull.* **2011**, *137*, 517–537. [[CrossRef](#)] [[PubMed](#)]
27. Hekman, E.; Brussee, R. Crowdfunding and online social networks. *Retrieved May* **2013**, *15*, 2014.
28. Mitra, T.; Gilbert, E. *The Language that Gets People to Give: Phrases that Predict Success on Kickstarter*; ACM Press: New York, NY, USA, 2014; pp. 49–61.
29. Xu, A.; Yang, X.; Rao, H.; Fu, W.-T.; Huang, S.-W.; Bailey, B.P. Show me the money! An analysis of project updates during crowdfunding campaigns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2014; pp. 591–600.
30. Greenberg, M.D.; Pardo, B.; Hariharan, K.; Gerber, E. Crowdfunding support tools: Predicting success & failure. In *Proceedings of the CHI'13 Extended Abstracts on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2013; pp. 1815–1820.
31. An, J.; Quercia, D.; Crowcroft, J. Recommending investors for crowdfunding projects. In *Proceedings of the 23rd International Conference on World Wide Web*; ACM: New York, NY, USA, 2014; pp. 261–270.

32. Chung, J.; Lee, K. *A Long-Term Study of a Crowdfunding Platform: Predicting Project Success and Fundraising Amount*; ACM Press: New York, NY, USA, 2015; pp. 211–220.
33. Boldeanu, F.T.; Constantinescu, L. The main determinants affecting economic growth. *Bull. Transilv. Univ. Brasov. Econ. Sci. Ser. V* **2015**, *8*, 329.
34. Malamud, S.; Zucchi, F. Liquidity, innovation, and endogenous growth. *Eur. Cent. Bank* **2016**, *132*, 519–541.
35. Bakker, G. Money for nothing: How firms have financed R&D-projects since the Industrial Revolution. *Res. Policy* **2013**, *42*, 1793–1814.
36. InfoDev. Crowdfunding’s Potential for the Developing World. Available online: <http://www.infodev.org/crowdfunding> (accessed on 6 June 2016).
37. Beraldi, P.; Violi, A.; Ferrara, M.; Ciancio, C.; Pansera, B.A. Dealing with complex transaction costs in portfolio management. *Ann. Oper. Res.* **2019**, 1–16. [[CrossRef](#)]
38. Cumming, D.J.; Leboeuf, G.; Schwienbacher, A. Crowdfunding models: Keep-it-all vs. all-or-nothing. In Proceedings of the Paris December 2014 Finance Meeting EUROFIDAI-AFFI Paper. 2015. Available online: <http://dx.doi.org/10.2139/ssrn.2447567> (accessed on 5 July 2016).
39. Solomon, J.; Ma, W.; Wash, R. Don’t wait! How timing affects coordination of crowdfunding donations. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*; ACM: New York, NY, USA, 2015; pp. 547–556.
40. Aubert, B.A.; Kishore, R.; Iriyama, A. Exploring and managing the “innovation through outsourcing” paradox. *J. Strateg. Inf. Syst.* **2015**, *24*, 255–269. [[CrossRef](#)]
41. Wilhelm, M.; Hutchins, M.; Mars, C.; Benoit-Norris, C. An overview of social impacts and their corresponding improvement implications: A mobile phone case study. *J. Clean. Prod.* **2015**, *102*, 302–315. [[CrossRef](#)]
42. Noel, L. Using Blogs to Create a Constructivist Learning Environment. *Procedia-Soc. Behav. Sci.* **2015**, *174*, 617–621. [[CrossRef](#)]
43. Digout, J.; Azouri, M.; Decaudin, J.-M.; Rochard, S. Crowdsourcing, Outsourcing to Obtain a Creativity Group. *Arab Econ. Bus. J.* **2013**, *8*, 6–15. [[CrossRef](#)]
44. IPMA. Individual Competence Baseline for Project, Programme & Portfolio Management. Available online: <https://www.pma.at/files/downloads/440/ipmaicb4.pdf> (accessed on 27 September 2018).
45. Valančienė, L.; Jęglevičiūtė, S. Valuation of Crowdfunding: Benefits and Drawbacks. *Econ. Manag.* **2013**, *18*. [[CrossRef](#)]
46. Sohrabi, M.K.; Akbari, S. A comprehensive study on the effects of using data mining techniques to predict tie strength. *Comput. Hum. Behav.* **2016**, *60*, 534–541. [[CrossRef](#)]
47. R Core Team R: A Language and Environment for Statistical Computing. Available online: <http://www.R-project.org/> (accessed on 27 September 2018).
48. Peter, C.; Janet, C.; Randy, K.; Tom, K.; Thomas, R.; Shearer, C.R.H.; Rober, W. CRISP-DM 1.0 Step-by-Step Data Mining Guide. Available online: <https://www.bibsonomy.org/bibtex/d20b2a4a6204b04306c3fa3fcfa8ddbfb> (accessed on 5 October 2018).
49. Gedda, D.; Nilsson, B.; Săthén, Z.; Søilen, K.S. Crowdfunding: Finding the Optimal Platform for Funders and Entrepreneurs. *Technol. Innov. Manag. Rev.* **2016**, *6*, 10. [[CrossRef](#)]
50. Kickstarter Estadísticas de Kickstarter. Available online: <https://www.kickstarter.com/help/stats> (accessed on 9 August 2019).
51. Public Data Set. Available online: https://fusiontables.google.com/DataSource?docid=1tO8U_v4_oATQ6uo9dl-rPyg-dga3rnkZSbw9M#rows:id=1 (accessed on 9 August 2019).
52. Zheng, H.; Li, D.; Wu, J.; Xu, Y. The role of multidimensional social capital in crowdfunding: A comparative study in China and US. *Inf. Manag.* **2014**, *51*, 488–496. [[CrossRef](#)]
53. Wang, N.; Li, Q.; Liang, H.; Ye, T.; Ge, S. Understanding the importance of interaction between creators and backers in crowdfunding success. *Electron. Commer. Res. Appl.* **2018**, *27*, 106–117. [[CrossRef](#)]
54. Coates, A.; Ng, A.Y. Learning Feature Representations with K-Means. In *Neural Networks: Tricks of the Trade*; Montavon, G., Orr, G.B., Müller, K.-R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7700, pp. 561–580. ISBN 978-3-642-35288-1.
55. Davis, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *1*, 224–227. [[CrossRef](#)]

56. Ong, J.; Sibte, S. Data Mining Using Self-Organizing Kohonen maps: A Technique for Effective Data Clustering & Visualisation. In Proceedings of the International Conference on Artificial Intelligence, IC-AI '99, Las Vegas, NV, USA, 28 June–1 July 1999.
57. Rodríguez Montequín, V.; Villanueva Balsera, J.; Cousillas Fernández, S.M.; Ortega Fernández, F. Exploring Project Complexity through Project Failure Factors: Analysis of Cluster Patterns Using Self-Organizing Maps. *Complexity* **2018**, *2018*, 1–17. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).