This paper is a post-print paper accepted in The First IEEE International Conference on Artificial Intelligence Testing. The final version is available through IEEE Xplore: https://ieeexplore.ieee.org/document/8718230
Citation information: C. Augusto, J. Morán, C. De La Riva and J. Tuya, "Test-Driven Anonymization for Artificial Intelligence," 2019 IEEE International Conference On Artificial Intelligence Testing (AITest), Newark, CA, USA, 2019, pp. 103-110. doi: 10.1109/AITest.2019.00011

Test-driven Anonymization for Artificial Intelligence

Cristian Augusto
Department of Computing
University of Oviedo
Gijón, Spain
augustocristian@uniovi.es

Jesús Morán
Department of Computing
University of Oviedo
Gijón, Spain
moranjesus@uniovi.es

Claudio de la Riva
Department of Computing
University of Oviedo
Gijón, Spain
claudio@uniovi.es

Javier Tuya
Department of Computing
University of Oviedo
Gijón, Spain
tuya@uniovi.es

***Abstract*—In recent years, data published and shared with third parties to develop artificial intelligence (AI) tools and services has significantly increased. When there are regulatory or internal requirements regarding privacy of data, anonymization techniques are used to maintain privacy by transforming the data. The side-effect is that the anonymization may lead to useless data to train and test the AI because it is highly dependent on the quality of the data. To overcome this problem, we propose a test-driven anonymization approach for artificial intelligence tools. The approach tests different anonymization efforts to achieve a trade-off in terms of privacy (non-functional quality) and functional suitability of the artificial intelligence technique (functional quality). The approach has been validated by means of two real-life datasets in the domains of healthcare and health insurance. Each of these datasets is anonymized with several privacy protections and then used to train classification AIs. The results show how we can anonymize the data to achieve an adequate functional suitability in the AI context while maintaining the privacy of the anonymized data as high as possible.**

Keywords—Anonymization, Software Testing, Artificial Intelligence, k-Anonymity.

I. INTRODUCTION

Artificial Intelligence (AI) is a broad multidisciplinary field aimed to reproduce human reasoning. AI has several techniques to provide intelligence based on different approaches such as extracting patterns from data or mimicking the biological processes. These techniques are used to develop a tool that usually learns from previous data to make predictions or forecasts. For example, an AI tool can learn from historical data to predict if a bank should grant a credit or not based on both the age and other information of the client's account. Fig. 1 represents the development and usage of the AI tools. First, the developer implements the AI tool (AI_D) that learns the patterns from the historical data (dataset *D*). In the case of the bank credits, the AI tool can learn that young clients tend not to pay the credits more than older clients. Then, the bank grants the credits more easily to older clients than the younger ones. Age is not the only factor to be considered to grant a credit or not, and the AI tool may learn more sophisticated patterns based on the client earnings. Then a properly developed AI tool is able to grant/reject a credit like an experimented banker because it learns from the historical data the characteristics of the successful/failed credits.

During development of an AI tool, the dataset (*D*) is usually divided into two parts as in Fig. 2: one part called "training data" is used to learn (*D* Train data), and the other part called "testing data" is used to test the functional suitability (*D* Test data). Firstly the developer selects the AI technique and its parameters. The supervised AI techniques learn only from training data. This process commonly called "training" obtains an AI tool/model (AI_D) that may learn the patterns/formulas of the data to make correct predictions. In order to check if the developer has selected the right AI technique with good parameters, the AI model is usually tested against known data (*D* Test data). This test dataset contains historical data with the expected outputs, for example some historical bank information about successful/failed credits. Then the tester hides the expected outputs of testing data (*D* Test data) and checks if the AI model is able to predict them properly [1]. As shown in Fig. 2, the AI model could not pass the tests, in which case the developer should make changes such as tuning the AI technique and parameters.

Once the AI model passes the tests, it can be used in production to take advantage of the experience learned from the historical data. Then the AI model (AI_D) is able to predict outcomes for a specific domain, not only for the historical data (*D*), but also for the new production data. For example, the AI model of the bank is able to grant/reject the credits both for historical clients, and also for new clients. The ideal AI model of the bank domain grants credits only for those new clients that will pay the credits and rejects the credits for those that will not. An AI model (AI_D) that passes the tests can fail dramatically in production due to a poor-quality dataset used to train and test. This situation can happen when the data are anonymized for privacy reasons. If the historical

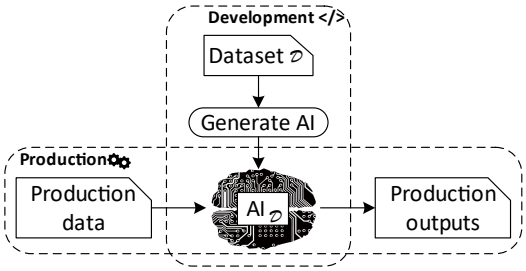

Fig. 1. Development and execution of AI models.

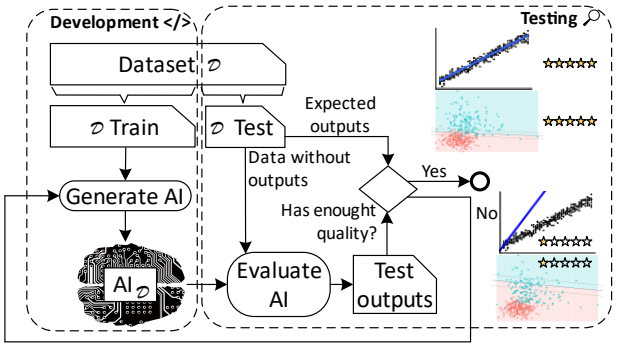

Fig. 2. Training and testing AI models.

data (D) is biased, then the AI model learns incorrect patterns and will fail in production because it does not recognise the unbiased/real patterns. These faults are difficult to detect because the AI model is correct according to the historical data used in training-testing but it is also useless because both model and historical data are biased/incorrect. Then if the AI model does not learn from the genuine data (i.e. alteration of the data), the model could be incorrect and useless for production.

This problem affects AI models developed from anonymized data. Recently some AI models have been generated from third party data, such as those datasets published under open data/government, released by the organizations or sold by data-centric companies, among others [2]. The private information of these datasets must not be released to enforce laws or internal privacy requirements, and the datasets must be anonymized as in Fig. 3. The anonymization can be done with different techniques such as the generalization or suppression of the sensitive data to avoid the malicious re-identification of the individuals. The degree of anonymization can be measured with the k-anonymity, among others. K-anonymity guarantees that if we know the data that identify an individual, there are at least k-individuals with the same identifiers. Bigger k values make the leak of sensible data more difficult because the identifiers known by the attacker are related with the sensitive data of k-individuals. During the anonymization, the private data (D) are transformed removing information or generalizing the data (D_{anom}). Then the AI model (AI_{anom}) that learns from anonymized data is not learning directly from original data, but rather from transformed data. During this transformation, the data loses information that could mislead the learning process of the AI models. Then the AI model could be correct according to the anonymized data, but useless because during the anonymization the data lost relevant information to learn.

The organizations that release a dataset to be used in AI must not only anonymize the private data, but also guarantee that these anonymized data are still useful for the AI models. For example, Netflix released a useful dataset for a $1M competition aimed to create an AI tool [3]. This dataset was not anonymized enough because it was possible to leak sensitive information such as the apparent political preferences of the users only with a small amount of information and the IMDB database [4]. There are other similar cases, such as the dataset about New York taxis. In this dataset, the taxis could be identified sometimes through the medallion and other information [5]. Another privacy breach of this dataset is that sometimes it is possible to obtain the trips and fares of the celebrities given the metadata of the photos taken near the taxi (place, time and others) [6]. Then personal information such as the address of the celebrities or other sensitive data derived from the trips could be leaked.

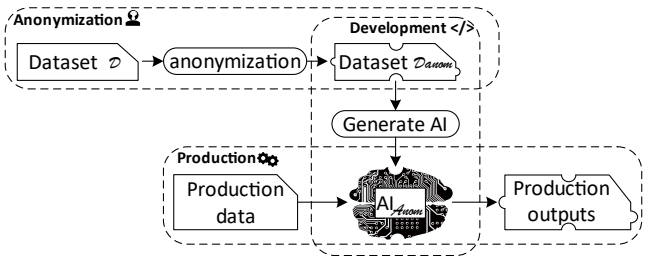

Fig. 3. Development and execution of AI models from anonynized data.

On the other hand, the anonymized dataset (D_{anom}) could become useless for the AI models in comparison with the non-anonymized dataset (D). For example, the anonymized dataset of rare diseases is not universally useful because it is only sometimes appropriate for some AI models, not very good for other AI models, and useless for other AI models due to the anonymization [7]. After the anonymization, the dataset lost information that can make that AI model learn the wrong patterns/formulas. Then the anonymization must preserve the privacy of the information while keeping the anonymized dataset useful for the AI models (functional suitable).

In this paper, we propose a test-driven anonymization approach for AI models aimed to trade-off the privacy (non-functional quality) and the functional suitability (functional quality). This approach anonymizes the data several times to obtain, during the testing, a good trade-off between non-functional and functional quality of the AI model. The contributions of this paper are:

1) A test-driven anonymization technique to trade-off the privacy of the dataset and the functional suitability of the AI models that learns from the anonymized dataset.
2) A quality measure for anonymization composed by the functional and non-functional quality prioritized by the user requirements in AI.
3) Evaluation through 2 real-life datasets of healthcare and health insurance using AI classification.

The remainder of this paper is organized as follows. The test-driven anonymization approach proposed is defined in Section II and evaluated in Section III. The related work is discussed in Section IV. Finally, the conclusions and future work are described in Section V.

II. Test Driven Anonymization

The anonymized data must not only guarantee privacy (non-functional quality), but also functional suitability for the AI models (functional quality). Test-Driven Anonymization (TDA) is an approach targeted towards the development of an AI model to anonymize the data until the tests achieve a trade-off between both quality characteristics. In a first stage, the dataset is anonymized several times and used to train an AI tool. Then the AI tools are tested with the original data until obtaining a trade-off against both AI functional suitability (Q_F) and the privacy (Q_{NF}) of the dataset.

Fig. 4 depicts an example of four non-linear regression models to predict the value of function Y = f(X). The models are developed using the original dataset (1) and anonymized datasets obtained through three anonymization efforts (2-4). As the degree of anonymization of the dataset increases after each effort, the functional suitability of the AI models decreases because these models learn the wrong patterns due to lost information. The first model (1) follows a linear-logistic-sigmoid function that achieves good functional suitability (R^2 = 0.979) and low privacy (1-anonymity) because it fits the original dataset. In the second AI model (2) the dataset is anonymized to 5-anonymity, and the AI model is still useful because it also fits the real data in linear-

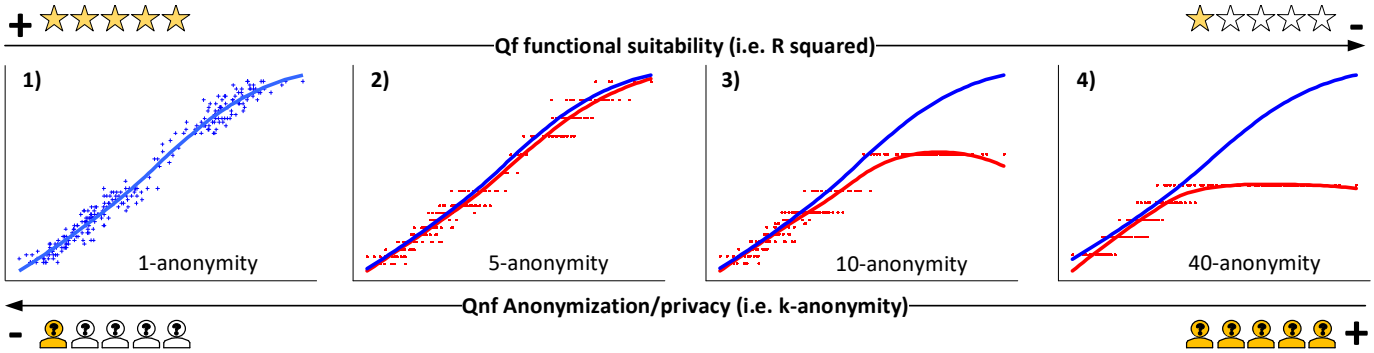

Fig. 4 Anonymization used in non-lineal regression model.

logistic-sigmoid function. In the third AI model (3) the anonymization is increased again to 10-anonymity. This model fits the anonymized data but does not fit the original data, especially the right part of the X values. It follows a logarithmic function instead of the linear-logistic-sigmoid which leads to a useless model. Finally, in the last AI model (4) the anonymization is increased to 40-anonymity, and the difference with the target model is higher.

Although this fourth AI model (4) fits the anonymized data, it is useless because it does not fit the original data well, especially in the middle and right part of the X values. Then this model can fail with production data (similar to the original data), but the fault of this model can also be masked during testing against the anonymized data. Fig. 5 provides more insight into the above by measuring the R^2 values: the AI model trained with the original data (1) has good functional suitability ($R^2 = 0.979$) and then can be useful also with production data. The middle and right parts of Fig. 5 contain the fourth AI model (4b y 4c) trained with the anonymized data (40-anonymity) and evaluated against anonymized and real data, respectively. Although this model fits correctly with the anonymized data (4a) ($R^2 = 0.92$), the model is faulty because it does not fit the original data correctly (4b) ($R^2 = 0.11$). This AI model will fail when it is used with production data because it has learned a biased pattern due to the information lost during the anonymization.

These faults caused by over-anonymization are difficult to find, especially when the developers only have access to the anonymized data because they also evaluate the model with anonymized data as in the middle of Fig. 5 (4b). Then these faults can be masked during testing because both the AI model and the test data (D_{anom} Test with inputs and expected outputs) are wrong due to the information lost during anonymization (incorrect test oracle).

To overcome this problem, the test-driven anonymization (TDA) approach presented in this paper evaluates the functional suitability of the AI models with the original data (D Test) instead of the anonymized data as depicted in Fig. 6. The original dataset (D) is anonymized with different degrees of anonymization and then these anonymized data are used to train an AI model (AI_{anom}) with AI methods such as regression and classification. Then the quality of each anonymization is evaluated through a composition of both the privacy and the functional suitability of the AI models. To measure the functionality suitability (Q_F) the AI model is not tested against the anonymized data, but against the original data using any standard measurement such as R^2, accuracy or precision, among others. On the other hand, the anonymization of the dataset (Q_{NF}) may also be obtained with any standard measurement such as k-anonymity or l-diversity, among others. Then a quality metric (Q) is obtained by the weighted sum of both quality metrics: $Q = Q_F + \alpha \cdot Q_{NF}$. The α value is a parameter that represents the user requirement to prioritize the anonymization (non-functional quality) over the functional suitability (functional quality), or the opposite. For example, in some anonymization, more functionality may be preferable over the anonymization (low α), or the opposite, more anonymization over functionality (high α). As the original dataset (D) is anonymized several times (with different degrees of anonymization), the best anonymization (D_{anom}) is that which maximizes the quality (Q), achieving a trade-off

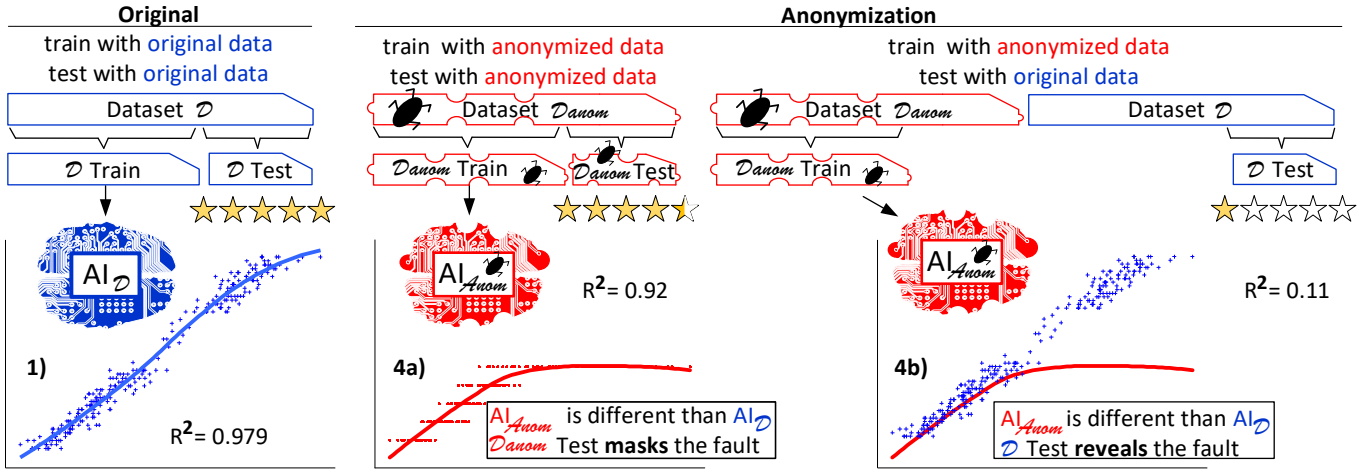

Fig. 5 Testing with anonymized and original data.

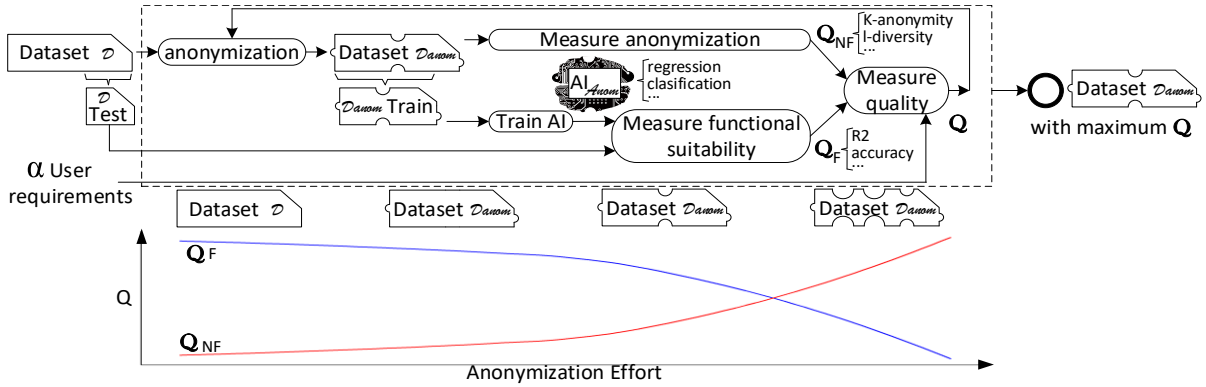

Fig. 6 Test-Driven Anonimyzation for AI models.

between functional suitability (Q_F) and privacy/anonymization (Q_{NF}) according to the user requirements (α).

For example, Fig. 7 indicates the maximum values of the anonymization quality (Q) for the anonymization efforts represented in Fig.4, using three different values of the user requirements α: more prioritization of functional than non-functional quality (left), balanced prioritization (middle) and more prioritization of non-functional than functional quality (right). In the first anonymization of the dataset (anonymization effort 1 of Fig. 7), the non-functional quality (Q_{NF}) increases from 1-anonymity to 5-anonymity while the functional quality (Q_F) is slightly decreased from R^2 = 0.979 to R^2 = 0.971. The anonymized dataset of the first effort is useful for machine learning and also protects the sensible data with anonymization. After the second anonymization effort, the dataset is less useful for machine learning because R^2 decreases to 0.68 and 0.1 in the second and third anonymization efforts, respectively.

On the left of Fig. 7, we prioritize the functional suitability of the AI model. Then the quality values are similar to the functional quality values with low influence of the non-functional quality values. For this user requirement the best anonymization is obtained in the first effort represented in Fig. 4 (2) with 5-anonymity and R^2 = 0.971. In the opposite way, on the right of Fig. 7 we prioritize the privacy/anonymization of the dataset. Then the quality values are the non-functional quality values, slightly influenced by the functional quality values. For this user requirement the best anonymization is obtained in the third effort represented in Fig. 4 (4) with 40-anonymity and R^2 = 0.11. This anonymized dataset is useless to be used in AI but has a high privacy/anonymization. The middle of Fig. 7 represents the user requirements having a balance of functional and non-functional, i.e. it does not prioritize one quality characteristic over the other characteristic. Then both characteristics have similar influence on the quality, obtaining a maximum quality in the first effort represented in Fig. 4 (2) with 5-anonymity an R^2 = 0.971. This anonymized dataset has sufficient privacy/anonymization and is also useful for AI as it fits the model, despite the model being trained using anonymized data.

The TDA approach is summarized in the following pseudocode:

```
INPUT:
  D    data to be anonymized
OUPUT:
  Danom) anonymized data that trade-off quality
Function:
  LOOP: #anonymization efforts
  |  Danom <-anonymize(D)
  |
  |  #non-functional quality i.e. k-anonymity
  |  QNF <- getAnonymizationDegree(Danom)
  |
  |  #functional suitability in AI i.e. accuracy
  |  AIanom  <- trainAI(Danom Train data)
  |  #testing against non-anonymized data
  |  QF <- testAI(AIanom,D Test data)
  |
  |  #Trade-off of both qualities
  |_  QEFFORT <- QF + α · QNF
  RETURN Danom with maximum QEFFORT
```

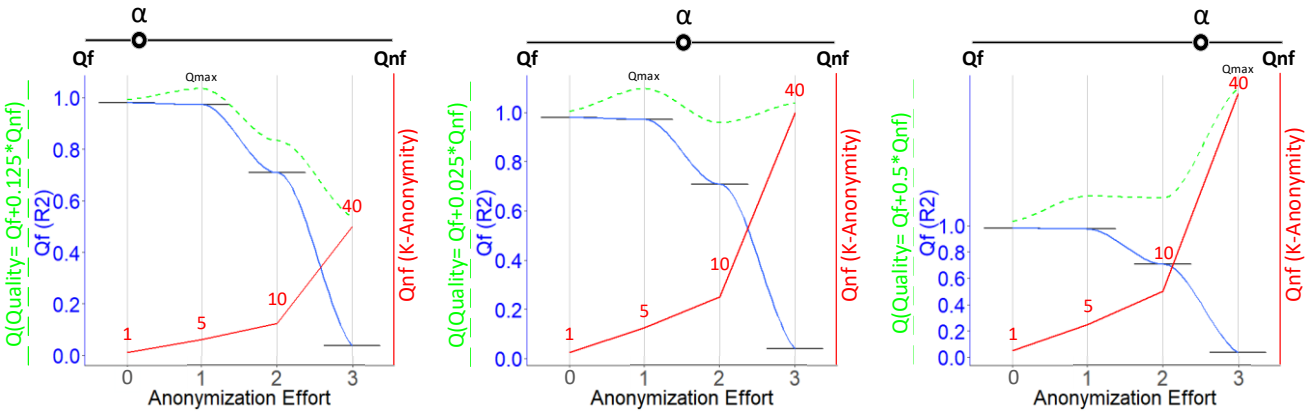

Fig. 7 Test driven anonymization for a non-linear regression model.

III. Evaluation

The Test-driven anonymization approach (TDA) is evaluated through the following three research questions:

RQ1) Based on the user requirements, is the anonymization approach able to prioritize the functional or non-functional quality?

RQ2) Can the anonymization approach achieve a trade-off between the functional and non-functional quality?

RQ3) Do the anonymized data preserve privacy while still being useful for AI models?

These research questions are answered through two real-life datasets in different domains:

1) Breast-Cancer (Wisconsin) [8]: public data about digitalized tumor images of surgical procedures. This dataset has several attributes such as the tumor radius, texture and symmetry. The AI model aims to predict the diagnosis of the tumor (benign or malign) given its radius and symmetry.

2) Medical Cost Personal Dataset (Medical Cost) [9]: public data about the bills of an insurance carrier. This dataset has several attributes such as age, sex, and body mass index (BMI). The AI model aims to classify a person into smoker or non-smoker given the BMI and the insurance charges.

Each of these datasets are anonymized using the generalization technique 6 times in a row (anonymization efforts). Each AI model is trained with Random Forest and tested 100 times per each anonymization effort, leading to a total of 1,200 AI models. For each model, the quality of both the AI model and the anonymization are evaluated as shown in the previous section but using accuracy and prediction instead of R^2.

Research question **RQ1** will be answered after analysing the distribution of the anonymizations that maximizes the quality for different user requirements. Research question **RQ2** will be answered taking into account the trends of functional and non-functional quality over the different anonymization efforts, and the maximum value of their composition $Q = Q_F + \alpha \cdot Q_{NF.}$ Finally, research question **RQ3** will be answered analysing the functional and non-functional measurements in the anonymization that maximizes the quality.

These research questions are answered in Subsection III.A and the discussion is in Subsection III. B.

A. Results

1) Research Question 1

The test-driven anonymization approach (TDA) obtains the anonymized data that maximizes the quality in accordance with the user requirements (α) according to: $Q = Q_F + \alpha \cdot Q_{NF.}$ (see Section II). The parameter α is intended to determine the priority of the functional quality over the non-functional quality. Low values give more priority to the functional quality (i.e. more privacy) and higher values give more priority to the non-functional quality (i.e. stronger anonymization).

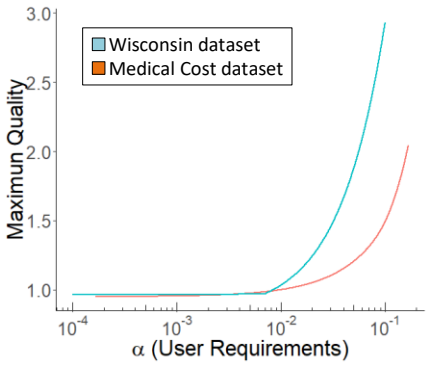

Fig. 8 Maximun Quality for the different user requiements (α)

We observed in Fig. 7 that the best anonymization that we can achieve depends on the user requirements α, and the different scales of each of the quality measures. The functional quality is measured by the accuracy/precision with values between zero and one. On the other hand, the non-functional quality is measured by k-anonymity with values higher than one.

Fig. 8 depicts the values of the quality of the best anonymization achieved according to different values of α for each of the datasets. We observe two different trends: (1) a rather stable value of quality until a given value of α, and (2) a sudden exponential increase of quality. At first glance, we could draw the (incorrect) conclusion that the highest values of α are the best choice. But this is due to the exponential increasing of k-anonymity measure (see again Fig. 7), which means that we are giving a too low priority to functional quality. Unless we are not interested in keeping privacy (measured as the non-functional quality), the range of values leading to trend (2) is not usable to achieve a trade-off between functional and non-functional quality.

Taking into account the above considerations, the answer to the **RQ1** is that the anonymization approach is able to prioritize either functional or non-functional quality based on the user requirements α. But, in order to achieve a trade-off we should keep the values of α in the usable range of values that gives the stable trend (1), as described in the previous paragraph.

2) Research Questions 2 and 3

Figures 9 and 10 depict the results of the anonymization using TDA on each of the two datasets using a value of α taking into account the aforementioned criteria. Each figure has the same structure: the top (A) indicates the distribution of the data used by the AI model in the dataset without anonymization and with two different anonymization degrees. The middle (B) depicts the AI model generated (always tested with the original data as shown in Section II). The bottom of the figures details the quality for different degrees of anonymization and its components (functional and non-functional quality) after 6 anonymization efforts. The functional quality measurements may vary slightly because a model trained with the same anonymized dataset can fit these data in different ways due to non-determinism.

During the anonymization of the Wisconsin dataset (Fig. 9), the predictions of the benign tumors slightly fit more closely to lower symmetric and radius of the tumor. Then as the functional quality of the AI model decreases, anonymization/privacy (non-functional) increases. In this case, the trade-off (maximum value of quality) is achieved

by an anonymization effort of 3, which leads to a value 5-anonymity while still having good accuracy. Although the point with no anonymization (1-anonymity) has a slightly better accuracy, the trade-off achieves a similar value, but with 5-anonymity.

During the anonymization of the Medical Cost dataset (Fig. 10), the predictions of non-smokers fit more slightly to lower charges at insurance carrier. Then the functional quality of the AI model again decreases while the privacy (non-functional quality) increases. In this case, the trade-off (maximum value of quality) is achieved by the first anonymization (anonymization effort 1), which leads to a value 2-anonymity while still achieving good precision.

The anonymization of the two datasets can obtain a trade-off between the functional and non-functional quality. The answer to **RQ2** is that TDA can achieve this trade-off point that maximizes the quality of the anonymization for AI models. The Wisconsin dataset is anonymized preserving the privacy (5-anonymity) while keeping the data useful for AI achieving high accuracy. The Medical Cost dataset does not achieve high privacy (only 2-anonymity), but the data are still useful for AI because they achieve high precision. The answer to **RQ3** is that the TDA can preserve the privacy of the data (non-functional quality) while keeping these anonymized data useful for the AI models (functional quality)

B. Discussion

The evaluation of the two case studies indicates that TDA is able to anonymize the data based on user requirements to trade-off the privacy (non-functional quality) and the functional suitability of the AI models (functional quality). These anonymized datasets are useful for AI models, achieving high accuracy/precision (functional quality). The Wisconsin dataset achieves a sufficient anonymization degree (5-anonymity), but the Medical Cost dataset achieves slightly lower (2-anonymity). There are a number of limitations and threats to validity of these results that are discussed below.

The conclusion threats are those issues that could affect the concluding of the evaluation. The AI models used in the case studies are generated in a non-deterministic way and this randomness could bias the conclusions. To mitigate the problem, 1,200 AI models are generated during the evaluation. This evaluation is also reproducible because the non-determinism is handled through seeds.

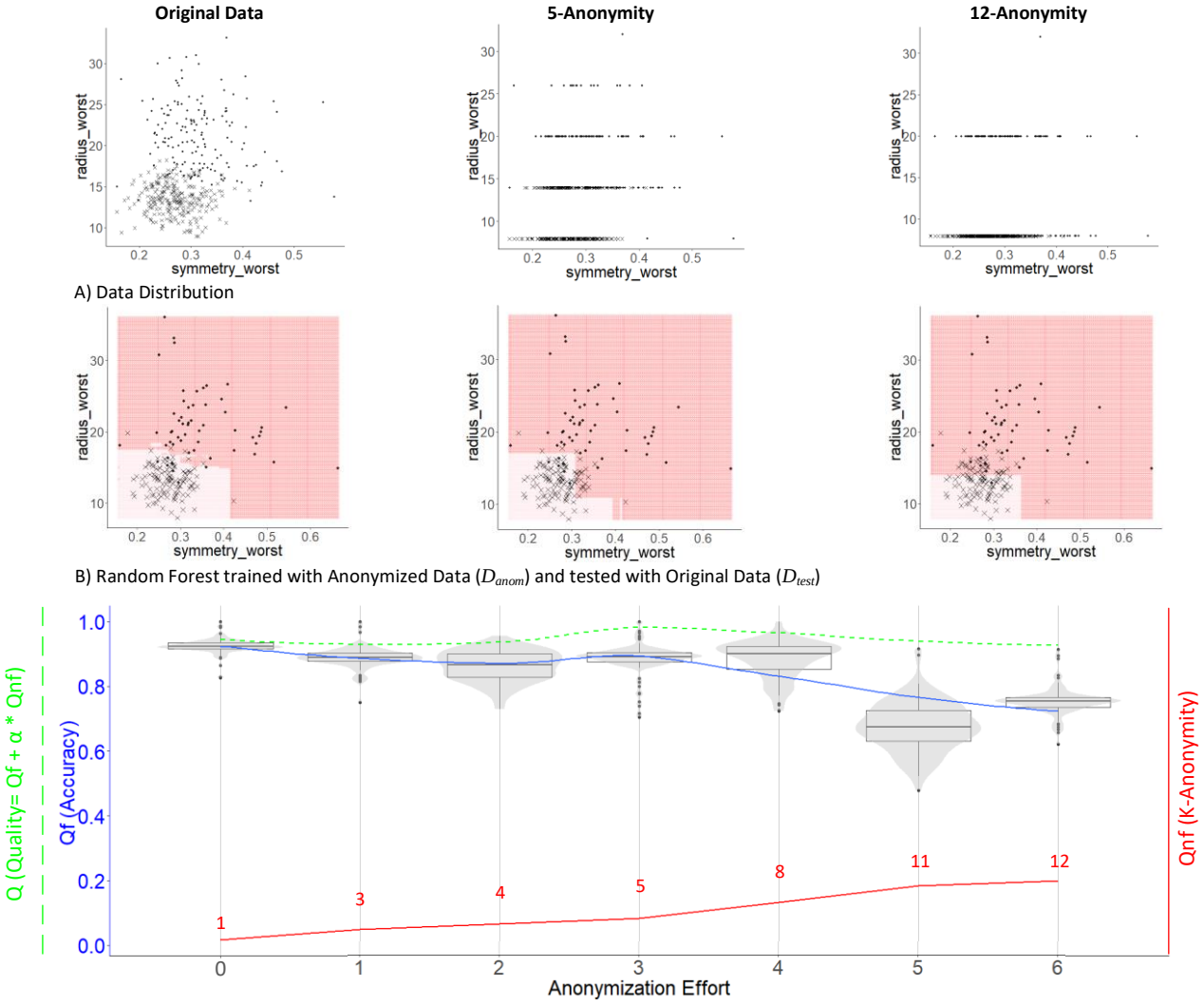

Fig. 9 Wisconsin Dataset Q_F (Model Accuracy) vs Q_{NF} (K-Anonymity of the dataset).

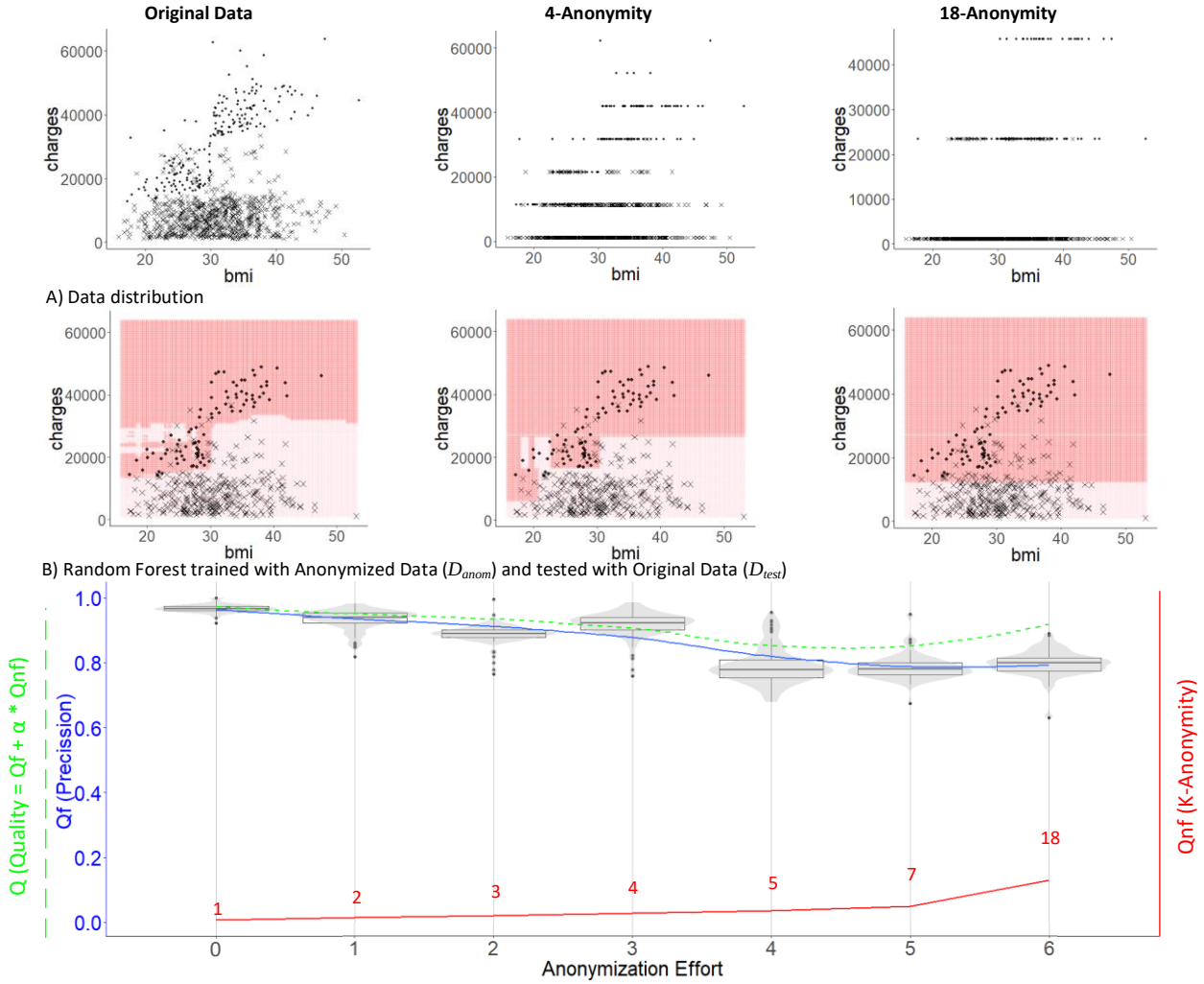

Fig. 10 Medical Cost Dataset Q_F (Model Precision) vs Q_{NF} (K-Anonymity of the dataset).

The internal threats are the issues related to the relationships between the information evaluated. During the evaluation, two public datasets are anonymized, but these datasets could also have been anonymized before being released. However, it is unlikely that these datasets were anonymized previously because apparently the data cannot identify the individuals.

The external threats are those issues related to the generalization of the results. The evaluation is performed through two case studies selected by consecutive sampling. Ideally, the case studies should be selected with random sampling, but sometimes this is not possible in software engineering, which is our case. To mitigate this problem, the two case studies were selected from real-life datasets instead of using synthetic data.

The datasets used during the evaluation are anonymized with generalization over one variable and their privacy is measured through k-anonymity. The functional quality is measured through accuracy/precision in a supervised AI model that classifies a binary variable from another two variables. TDA can be used with other anonymization techniques (i.e. suppression), anonymization measurements (i.e. l-diversity), AI models (i.e. linear regression), AI evaluation measurements (i.e. R^2), and another AI parametrization (i.e. forecasting one variable from another three variables). The technique is also described in Section II using a regression AI model, but in future work we plan to thoroughly experiment with other AI models beyond supervised machine learning.

The construct threats are those issues related with the theoretical concepts of the evaluation. TDA anonymizes the data to maximize a quality metric composed by the privacy/anonymization and the functional suitability. Other quality measurements are not considered such as the cost and the time employed in anonymization.

IV. Related Work

AI Testing is a field that has advanced in the last few years. Several of these advances employ the AI techniques to prioritize the test cases [10] and automatize the test oracle [11], among others. Our approach, TDA, is focused on the anonimization of the data, keeping them useful for the AI.

Anonymization aims to preserve the privacy of the dataset through different techniques such as k-Anonymity and generalization [12], [13]. We provide an overview on the related work on the fields of privacy preserving and anonymization that need to be supported with techniques that ensure the individual privacy [14]. There are some methods that modify the original data to protect each individual's identity [15]–[19]. These modifications can be done through

supressing or generalizating the sensible attributes. Those techniques are related to the field of Privacy-Preserving Data Publishing (PPDP). During the anonymization, the data preserves the privacy but also lose information.

Several works [7], [12], [20] are focused on the information lost during the anonymization and its effect on the AI and data mining. Unlike them, we propose a Test-Driven approach to anonymize the data aimed to achieve a trade-off between the privacy and the functional suitability of th AI tools that used the anonymized dataset. The closest researches to our work are [21], that maximizes the utility of the data for given privacy restraints, and [22] that makes an evaluation trying to maximize privacy and usefulness of the data. However, their scope is different, and it may be used with AI techniques and Privacy-Preserving methods. Additionally, our approach is intended to reach a trade-off point to maximize both goals, rather than maximize only one of them (the utility for a given k-anonymity). Another difference with [21] is that we use the same datasets for training and testing: our approach measures the functional quality of the AI models using the original dataset for testing.

V. Conclusions and Future Work

We have introduced a Test-Driven Anonymization approach (TDA) for artificial intelligence tools. During the testing, the anonymization of the data achieves a trade-off between the privacy (non-functional quality) and the functional suitability of the artificial intelligence tools (functional quality). TDA is intended to be used by the data provider to release useful datasets for developing AI tools, also maintaining the privacy. This approach is evaluated by means of two real-life datasets. Once the dataset is anonymized, it can be published or shared in line with the data protection laws while keeping them useful for artificial intelligence tools.

There are a number of open questions that we can summarize in three main lines for future work. The first is to extend the evaluation to more datasets and evaluate for each one to what extent the functional quality that we can obtain meets the minimum privacy level that the dataset complies with. The second is concerned with the techniques used to develop the model embedded in the AI tool and the measures used to evaluate the functional and non-functional quality. This would require comprehensive evaluations to check to what extent the TDA approach is dependent on them. The third line of work would focus on the automation of the approach to select the optimal α parameter as well as the trade-off, perhaps using search-based algorithms.

Acknowledgment

This work was supported in part by the Spanish Ministry of Economy and Competitiveness under TestEAMoS (TIN2016-76956-C3-1-R) project and ERDF funds.

References

[1] A. Ng, *Machine Learning Yearning*. 2017.

[2] J. Roski, G. W. Bo-Linn, and T. A. Andrews, "Creating value in health care through big data: opportunities and policy implications," *Health Aff. (Millwood)*, vol. 33, no. 7, pp. 1115–1122, 2014.

[3] Y. Koren, "The BellKor solution to the Netflix Grand Prize," 2009.

[4] A. Narayanan and V. Shmatikov, "Robust De-anonymization of Large Sparse Datasets," in *2008 IEEE Symposium on Security and Privacy (sp 2008)*, 2008, pp. 111–125.

[5] M. Douriez, H. Doraiswamy, J. Freire, and C. T. Silva, "Anonymizing NYC Taxi Data: Does It Matter?," in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2016, pp. 140–148.

[6] A. Tockar, "Riding with the stars: Passenger privacy in the nyc taxicab dataset," *Neustar Res. Sept.*, vol. 15, 2014.

[7] H. MacLeod, J. Abbott, and S. Patil, "Small Data Privacy Protection: An Exploration of the Utility of Anonymized Data of People with Rare Diseases," in *roceedings of the 2017 Workshop on Interactive Systems in Healthcare (WISH'17)*, 2017.

[8] "Wisconsin Dataset," *ICS UCI Archive*. [Online]. Available: http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.data. [Accessed: 20-Dec-2018].

[9] Brett Lantz, "Medical Cost Personal Datasets." [Online]. Available: https://www.kaggle.com/mirichoi0218/insurance. [Accessed: 21-Dec-2018].

[10] R. Lachmann, S. Schulze, M. Nieke, C. Seidl, and I. Schaefer, "System-Level Test Case Prioritization Using Machine Learning," 2016, pp. 361–368.

[11] S. R. Shahamiri, W. M. N. Wan Kadir, S. Ibrahim, and S. Mohd Hashim, "Artificial Neural Networks as multi-networks automated test oracle," *Autom. Softw. Eng.*, vol. 19, pp. 303–334, 2012.

[12] L. Sweeney, "Achieving K-anonymity Privacy Protection Using Generalization and Suppression," *Int J Uncertain Fuzziness Knowl-Based Syst*, vol. 10, no. 5, pp. 571–588, Oct. 2002.

[13] P. Samarati and L. Sweeney, "Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression," p. 19.

[14] C. C. Aggarwal and S. Y. Philip, "A general survey of privacy-preserving data mining models and algorithms," in *Privacy-preserving data mining*, Springer, 2008, pp. 11–52.

[15] P. Samarati and L. Sweeney, "Generalizing Data to Provide Anonymity when Disclosing Information (Abstract)," in *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, New York, NY, USA, 1998, pp. 188–.

[16] B. C. M. Fung, K. Wang, and P. S. Yu, "Top-down specialization for information and privacy preservation," in *21st International Conference on Data Engineering (ICDE'05)*, 2005, pp. 205–216.

[17] N. Mohammed, B. C. M. Fung, P. C. K. Hung, and C. Lee, "Anonymizing healthcare data: a case study on the blood transfusion service," in *KDD*, 2009.

[18] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," in *2007 IEEE 23rd International Conference on Data Engineering*, 2007, pp. 106–115.

[19] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "L-diversity: privacy beyond k-anonymity," in *22nd International Conference on Data Engineering (ICDE'06)*, 2006, pp. 24–24.

[20] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving Data Publishing: A Survey of Recent Developments," *ACM Comput Surv*, vol. 42, no. 4, pp. 14:1–14:53, Jun. 2010.

[21] V. S. Iyengar, "Transforming Data to Satisfy Privacy Constraints," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2002, pp. 279–288.

[22] J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A. W.-C. Fu, "Utility-based anonymization using local recoding," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 785–790.