Simultaneous Classification and Regression for Zakat Under-Reporting Detection

Mohamed Maher Ben Ismail * and Nasser AlSadhan

Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia; nsadhan@ksu.edu.sa
* Correspondence: mbenismail@ksu.edu.sa

Abstract: Tax revenue represents an essential budget source for most countries around the world. Accordingly, the modernization of relevant technological infrastructure has become a key factor of tax administration strategy for improving tax collection efficiency. In particular, the fiscal consolidation of the Kingdom of Saudi Arabia has been supported by considerable development in tax policy and administration, aimed at raising more taxes from non-oil activities. In fact, non-Saudi investors are liable for income tax in Saudi Arabia. On the other hand, Saudi citizen investors (and citizens of the GCC countries) are liable for Zakat, an Islamic assessment. Typically, taxpayers are in charge of preparing and accurately reporting their Zakat declaration. This allows tax authorities to overview and audit their business activities. However, despite administration efforts to increase taxpayer compliance, considerable revenue remains at under-reporting risk. In this paper, we introduce a novel intelligent approach to support tax authority efforts in detecting under-reporting among Zakat payer declarations. In particular, the proposed solution aims at improving detection accuracy and determining the fraud cases that correspond to a higher revenue at risk. Specifically, we formulate Zakat under-reporting detection as a supervised machine learning task through the design of a deep neural network that performs simultaneous classification and regression tasks. In particular, the proposed network contains an input layer, five hidden layers, and two output layers for classification and regression. Zakat declarations are mapped into the predefined "under-reporting" or "actual declaration" classes. Moreover, the revenue at risk caused by the predicted fraud cases is learned by the designed model. This allows the proposed approach to prioritize the auditing of specific Zakat payers based on the corresponding predicted revenue at risk. A real dataset including 51,919 Zakat declarations was used to validate and assess the designed model. Further, the Synthetic Minority Oversampling Technique (SMOTE) boosted the proposed model performance in terms of classification and prioritization.

Keywords: Zakat; fraud detection; machine learning; neural networks

1. Introduction

Each year, the Saudi Zakat Tax and Customs Authority (ZATCA) [1] receives millions of tax declarations from individuals, government agencies, financial institutions, companies, and various other entities. Those electronic forms include financial data relevant to taxpayer actvities. Additionnally, ZATCA collects third-party data from several government agencies to consolidate taxpayer profiles and cross-checks their financial information. Typically, ZATCA relies on big data for its tax enforcement activities. In particular, Zakat under-reporting represents a major risk faced by the Saudi tax authority. In fact, the Zakat base represents the net worth of the entity as calculated for Zakat purposes [2]. Then, Zakat is charged on the company's Zakat base at 2.5%. To bridge the widening Zakat gap, ZATCA governments enforce various legal penalties and regulations [2]. In fact, Zakat non-compliance is perceived as a delinquent act. Thus, taxpayers are required to prove their compliance in order to avoid legal consequences.

The earliest solutions considered to improve tax compliance relied exclusively on auditor efforts. However, this strategy is costly and constrained by the large number of taxpayers in addition to the reduced audit capacity of the tax administration. Moreover, big data being collected by tax authorities and stored in their databases are not efficiently exploited to enhance the detection rate of tax under-reporting. In other words, most of the case selection strategies rely on the intuition, domain knowledge, and experience of auditors with no intelligent mining of the existing data [3]. Further, taxpayers have been continuously developing new techniques of tax evasion that are relatively difficult to detect, requiring the deployment of advanced robust fraud detection methods [4,5].

The recent advances in Artificial Intelligence and its application to data science promoted the interest of tax administrations around the world to design intelligent solutions to support their conventional approaches to determine fraudulent behavior and optimize the management of the avilable auditing resources and data collection capabilities. In particular, the rich tax data collected by tax adiminstrations triggered the development of advanced analytics models intended to investigate and mine tax fraud patterns [6]. Namely, machine learning (ML) techniques have been adapted and associated with a large amount of tax datasets to improve risk description and detection performance [7]. Particularly, Value Added Tax (VAT) under-reporting detection has been formulated as a supervised learning task. In other words, historical VAT data have been used to train classification models able to map unseen VAT declarations into one of the predefined classes: under-reporting or actual declaration [8].

Despite the promising solutions introduced to address VAT under-reporting risk [9], to the best of our knowledge, no intelligent approaches have been proposed to alleviate Zakat fraud concerns. Moreover, although existing supervised machine learning techniques [10] solve the categorization problem, the challenge for tax administration remains the prioritization of cases to be audited. For instance, of 100,000 declarations positively categorized using a classification model as under-reporting risks, and given the audit capacity of a tax administration, only a subset of the flagged samples will be selected and sent to the operations department for audit. Naturally, the classification confidence or probability is used to determine the most risky instances. However, this priority measure does not take into consideration the revenue at risk, which is among the main key performance indicators for tax administrations. In other words, a highly accurate supervised learning model may not yield more tax income for the government.

This research aims at addressing this challenge as well as enhancing the overall under-reporting detection performance, through the design and development of a supervised machine learning model that predicts Zakat under-reporting and the revenue at risk. Specifically, a deep neural network is designed to classify Zakat declarations into "under-reporting" or "actual declaration" classes, and predict the expected tax gap. Moreover, the proposed model would support administration efforts to pinpoint the declarartions and/or taxpayers that were assigned to the class "under-reporting" and that correspond to a higher revenue at risk. Specifically, the proposed model generates a confidence value that encloses the likelihood of belonging to the class "under-reporting" as well as the cooresponding revenue at risk. Thus, a high confidence value is associated with both under-reporting risk and high revenue at risk. Accordingly, the main contributions of this research can be summarized as: (i) designing a deep neural network performing simultaneous classification and regression, (ii) enabling Zakat declarations to categorize taxpayers into compliant or non-compliant, and (iii) determining the shortlist of Zakat payers that should be audited first in order to maximize the expected Zakat income. Further, the model hyper-parameters are investigated to determine the optimal settings. Namely, activation, the batch size, epochs, and the number of layers are investigated during the fine tuning phase.

The rest of this manuscript is organized as follows: Section 2 surveys the related works relevant to tax fraud prediction using machine learning techiques. In Section 3, the proposed solution is depicted, while the experiments settings, findings and discussion are

outlined in Section 4. Finally, the research conclusions and future work are presented in Section 5.

2. Related Works

To the best of our knowledge, to date, no research has dealt with Zakat fraud detection using machine learning techniques. Accordingly, this section covers relevant tax fraud detection techniques introduced by researchers and/or adopted by fiscal administrations around the globe.

Typically, rule-based systems have been designed to address challenges related to various tax fraud detection. However, the resulting solutions proved to be limited by the millions of taxpayers (individual and business) to be investigated in addition to the subjective intuition and knowledge of auditors when selecting suspicious cases [11]. This alternative exhibits two main drawbacks: (i) expensive maintenance and update costs of knowledge-based approaches, and (ii) dependence on previous experience, which affects its ability to recognize recent fraudulent behavior. On the other hand, fraudsters keep developing tactics to evade paying taxes. This makes auditor intuition and experience insufficient to track them.

Recently, data science and Artificial Intelligence emerged as the most promising alternatives for addressing complex analytics challenges. Specifically, they have been used to leverage the machine's ability to learn from available data with no explicit programming. The works in [11,12] introduced fraud-focused advanced data analytics and machine learning. Although they were not focused on tax fraud detection, multiple studies [13,14] adapted supervised and unsupervised machine learning techniques. One should note that more contributions relied on unsupervised machine learning due to the scarcity of labeled data. Particularly, the authors in [13] relied on unsupervised learning to group similarly valued tax declarations into homogeneous clusters. Then, they adjusted the resulting probability distribution of each obtained cluster. Finally, the detection of suspicious patterns was achieved based on the quantiles of cluster-adjusted distribution. Despite the reported promising results, the main limitations of the work were data scarcity and the reduced number of features used to represent it. Moreover, the absence of labeled gold data affected the trustworthiness of the performance achieved by the model.

The pioneering deployment of supervised machine learning techniques to address fiscal fraud detection was achieved in [12]. Specifically, the authors investigated the C5.0 decision tree algorithm to build a predictive model for tax evasion detection in Italy. In addition, random forests [15], rule-based classification [16] and Bayesian networks were also considered to resolve fraudulent tax behavior detection. The researchers in [17] depicted a Value Added Tax (VAT) screening framework to determine non-compliant VAT declarations. In particular, the Apriori algorithm was employed to mine association rules from historical data of business entities with a confirmed fraudulent behavior. The resulting model was assessed using non-compliant VAT declarations collected in Taiwan from 2003 to 2004. Similarly, the Apriori algorithm was adapted in [18,19] to mine hidden patterns underlying fraudulent tax behaviors. Specifically, it was associated with Principal Component Analysis (PCA) [20] and Singular Value Decomposition (SVD) [21], as dimensionality reduction techniques, in order to determine the relevant fraud indicators and learn a fraud scale to rank Brazilian taxpayers based on the risk they represent for the tax administration. In [22], unsupervised and supervised machine learning techniques were coupled to detect taxpayers' suspicious behavior. In particular, two unsupervised learning techniques were adopted to discover clusters of business entities that exhibit similar tax-related behavior. Namely, Neural Gas [23] and Self-Organizing Maps [24] were investigated and associated with two datasets including tax declarations of micro to small and medium to large companies collected in Chile from 2005 to 2007. Additionally, three supervised learning techniques, namely, neural network, decision tree and Bayesian network were used to build classification models able to detect fraudulent taxpayer behavior. Similarly, the research in [21] included a comparison between Artificial Neural Networks (ANN) [25], Support

Vector Machines (SVM) [26], and K-Nearest Neighbors in the context of credit card fraud detection. The reported results showed that ANN outperforms the other models. In [27], the authors introduced a fraud detection approach for Mellon Bank. ANN proved to be more accurate and to enhance the timeliness and overall detection performance. ANN overtook decision trees and Naive Bayes classifiers when associated with financial data to detect fraud. In summary, the multilayer perceptron (MLP) model has been recommended for fraud detection tasks [28].

In [29], the researchers introduced a multilayer perceptron neural network model to detect fraud in personal income tax forms. The reported findings show that the multilayer perceptron can be considered as an efficient classifier to predict fraudulent taxpayers, and estimate the taxpayer's likelihood of cheating tax. One should note that the latter approach can be generalized to recognize fraud patterns for other types of taxes. A Hybrid Unsupervised Outlier Detection (HUNOD) model was presented in [22] to mine risky tax behaviors. In particular, user knowledge was fed into a combination of representational learning and clustering to detect outliers in personal income tax data. The authors claim that the interpretability of the detected outliers is performed through the training of explainable-by-design surrogate models over outliers validated internally. Recently, in [30], the authors coupled Artificial Neural Networks with a real dataset to detect factors related to income tax fraud. Their approach was designed to reduce time, effort, and cost taken by auditors in the manual identification of cases to be audited. This was the first study to adopt Artificial Neural Networks for income tax fraud detection in Rwanda. Similarly, financial prediction was tackled in [31] using deep convolutional neural networks (DCNN) and multilayer perceptron (MLP). In particular, the authors, used an 8-layer MLP and a 13-layer DCNN for their credit scoring model. The models were assessed using Australian and German credit scoring data. The reported experiments proved that DCNN achieved a considerably higher performance compared to MLP. The researchers in [32] outlined a transfer learning approach to build a tax evasion detection model. Specifically, exploited conditional adversarial networks to encode a collection of labelled tax evasion records by extracting the relevant features. The transfer learning approach is then conducted by fine-tuning the trained model using five tax datasets collected in five Chinese regions. In [33], a large-scale dataset of electronic records of taxable transactions collected in Mexico was analyzed. The authors concluded that the interaction patterns of evaders differ from those corresponding to typical taxpayer behavior. Based on this finding, they built deep neural network and random forest [15] models to classify unseen records as suspicious or evasion-free cases.

Semi-supervised learning was also used to tackle tax evasion detection. In [34], a semi-supervised approach was introduced for VAT audit case selection. Precisely, a gated mixture variational autoencoder network [35] was adapted to extract relevant features and map them into some predefined classes. Another solution based on positive and unlabeled (PU) learning techniques was depicted in [36]. One should note that PU techniques are suitable for data collections including a small subset of data that are positively annotated while the remaining records are not labeled. The method uses: (i) one-class probabilistic classification to generate pseudo-labels and assign them to unlabeled data, (ii) random forest [15] to determine relevant features and (iii) LightGBM [37] as a predictive model to classify unseen records. Additionally, the authors in [38] investigated PU learning tax evasion detection. Their method integrated features obtained by embedding a transaction graph into an Euclidean space. The work was then extended further by the researchers in [39] who introduced a graph-embedding algorithm for transaction graphs that, prior to generating pseudo-labels for unlabeled records, extracts network-based features. Finally, a multilayer perceptron (MLP) neural network is built using pseudo-annotated data to detect unseen tax evasion instances.

As outlined above, neither rule-based solutions [10] nor unsupervised-learning-based approaches [13,14] yielded satisfactory achievements when used to address tax fraud detection. The reported results were typically constrained by the expensive maintenance and update cost of knowledge-based rules as well as the np-hardness of the clustering problem. Alternatively, the supervised-learning-based solutions proved to be promising despite the scarcity of labeled data [12]. In particular, machine learning techniques such as decision tree, random forest, Bayesian networks, Support Vector Machines (SVM) and K-Nearest Neighbors were investigated to mine fraudulent tax behavior [15,26,35]. Moreover, dimensionality reduction techniques such as Principal Component Analysis (PCA) [20] and Singular Value Decomposition (SVD) [40] were deployed as dimensionality reduction techniques in order to identify relevant fraud indicators/attributes. More recently, ANN-based approaches proved to be more effective than shallow-model-based solutions in detecting tax fraud cases [21,27,31]. One should note that semi-supervised learning along with outlier detection techniques were also investigated to enhance VAT audit case selection [34–39]. However, the obtained results do not show drastic improvements in terms of detection performance.

3. Proposed Approach

Given the criticality of the detection task in the context of tax under-reporting risk. This research aims to introduce an under-reporting detection approach based on deep learning techniques. Specifically, it formulates the prediction of Zakat under-reporting cases as a classification task. Moreover, it estimates the revenue at risk as the outcome through a regression task, which yields an objective prioritization of the auditing operations to be conducted by tax administrations. In other words, the proposed model classifies Zakat declarations into the "under-reporting" and "actual declaration" categories, which represent the positive class and the negative class, respectively. Furthermore, it determines which cases among those assigned to the positive class should be given higher priority for auditing by the tax administration. In fact, the priority sorting is conducted based on the confidence degree generated by the proposed model. Specifically, the proposed model assigns a high confidence value to under-reporting which corresponds to an expected high revenue at risk.

In the following, we depict the design details of the proposed system. Figure 1 overviews the proposed network designed to extract relevant low-level features and learn their mapping into the predefined classes. As such, a collection of Zakat declarations including applicable attributes is fed into the system for the training phase. Note that these training instances are labelled, which makes them suitable for the training of the proposed deep neural networks. Note that the supervision information (labels) represents previous auditing results. Precisely, each training instance is associated with a ground truth class label as well as a Zakat revenue calculated as the difference between the pre-audit and post-audit Zakat amounts. The considered deep neural network consists of an input layer, a set of hidden layers, and two output layers. The latter are designed to perform simultaneously: (i) a regression task to predict the expected revenue at risk, and (ii) a classification task to assign each declaration to "under-reporting" or "actual declaration" category using a sigmoid layer fed with the same input. The proposed dual classification and regression task yields an objective prioritization of the auditing effort based on the confidence value generated by the designed model.
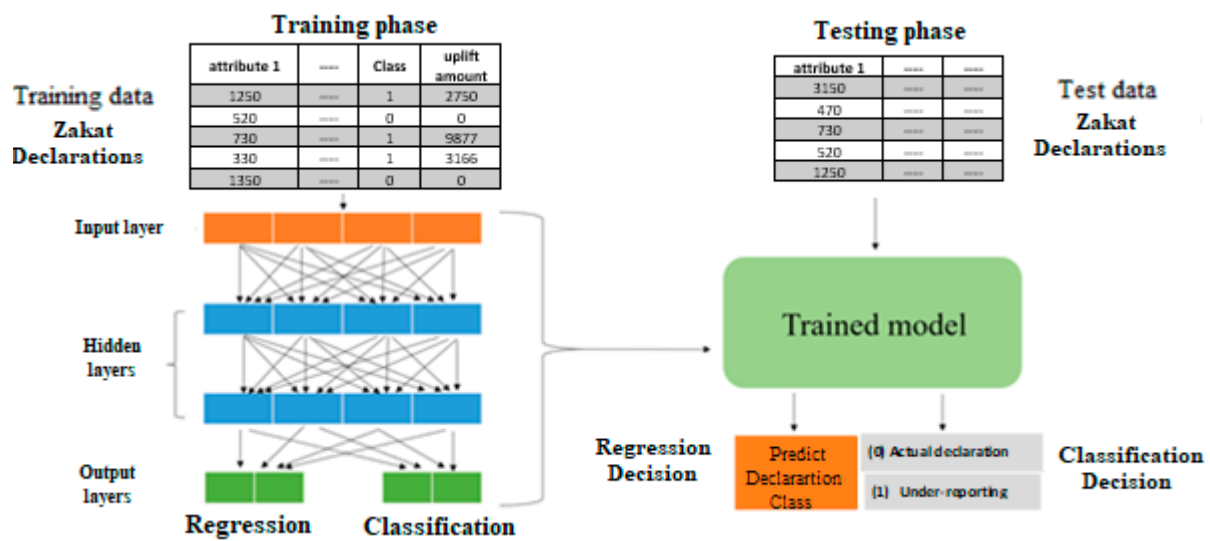
Figure 1. The proposed simultaneous classification and regression based approach.

The training of the proposed model relies on the optimization of the following loss function:

$$L = \beta(MSE) + (1 - \beta)(Binary\ cross\ entropy) \tag{1}$$

where β represents the coefficient of the loss function that controls the tradeoff between linear regression layer and the sigmoid layer loss functions. Particularly, the loss function corresponding to the regression layer is formulated as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(h(x_i) - y_i)^2 \tag{2}$$

where N is the number of instances, $h(x_i)$ expresses the predicted value for input x_i, and y_i represents the actual value corresponding to the input x_i.

Further, the binary cross entropy is exploited for the optimization of the classification model:

$$Binary\ cross\ entropy = \frac{1}{N}\sum_{i=1}^{N} -(y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)) \tag{3}$$

where p_i represents the probability of the first category, and $(1 - p_i)$ represents the probability of the second category.

The backpropagation algorithm considered for optimizing the loss function while training the proposed model and updating the network weights is detailed Algorithm 1.

Algorithm 1: Backpropagation

---

for *i* in reversed network:
    Layer ←network$_i$
    if *i* is last layer then:
        for *j* in Layer:
            Error$_C$ ← ***Binary cross entropy***
            Error$_R$ ← ***MSE***
            Loss ← ***β(Error$_C$)*** + ***(1 − β)(Error$_R$)***
    update the network weights

For the testing phase, an unseen a Zakat declarations dataset is conveyed to the trained model to classify cases into "under-valuation" or "actual declaration", and predict each declaration revenue at risk using the regression layer.

Further, one can note that the designed architecture was fine-tuned empirically. In other words, the final network architecture as well as the hyper-parameter settings were

optimized through comprehensive experiments. In particular, the appropriate number of layers along with the number of neurons per layer were determined empirically. Additionally, the network hyper-parameters, such as the batch size, optimizer, the number of epochs and the learning rate, were investigated during the training phase. Accordingly, the architecture proposed in Figure 1 encloses an input layer with five hidden layers. The first four hidden layers are, respectively, followed by dropout layers. In fact, the latter layers represent a regularization technique that prevents the proposed neural networks from overfitting. Specifically, these dropout layers are intended to modify the network itself to avoid overfitting. On the other hand, the last hidden layer is followed by two output layers. The first one is meant to perform the regression task while the second layer is dedicated for the classification purpose. Figure 2 illustrates the structure of the proposed architecture.

input	input:	[(None, 94)]
InputLayer	output:	[(None, 94)]

hidden_1	input:	(None, 94)
Dense	output:	(None, 94)

dropout_1	input:	(None, 94)
Dropout	output:	(None, 94)

hidden_2	input:	(None, 94)
Dense	output:	(None, 1000)

dropout_2	input:	(None, 1000)
Dropout	output:	(None, 1000)

hidden_3	input:	(None, 1000)
Dense	output:	(None, 500)

dropout_3	input:	(None, 500)
Dropout	output:	(None, 500)

hidden_4	input:	(None, 500)
Dense	output:	(None, 250)

dropout_4	input:	(None, 250)
Dropout	output:	(None, 250)

hidden_5	input:	(None, 250)
Dense	output:	(None, 50)

reg_output	input:	(None, 50)
Dense	output:	(None, 1)

class_output	input:	(None, 50)
Dense	output:	(None, 1)

Figure 2. The proposed Model Architecture.

Accordingly, Table 1 details each layer of the proposed architecture. As it can be seen, the input layer is fed with 94-dimentional data and conveys them to the next hidden layer

which is also composed of 94 neurons. Then, a dropout layer is placed to randomly omit 10% of the neurons in order to increase the resulting model generalization capability and avoid overfitting. On the other hand, the next hidden layer encloses 1000 neurons, followed by a 10% dropout layer, while the third hidden layer encloses 1000-dimensional features and yields 500-dimensional features through 500 neurons prior to a dropout layer. Similarly, the 250-neuron fourth hidden layer is coupled with the last 10% dropout layer, and yields 250-dimensional features to be processed by 50 neurons. Note that a l2 regularizer was associated with a ReLU activation function for all hidden layers. Further, the prediction is synchronously performed through the considered output layers. In particular, as illustrated in Figure 2, a "reg_output" activation function is dedicated for the revenue at risk prediction, while a "class_output" layer that consists of a sigmoid function is intended to classify Zakat declarations into the "under-reporting" and "actual declaration" categories.

Table 1. Details of the proposed network layers.

Layer	Layer	Neurons	Activation Function	Regularizer	Dropout Ratio	Input Size	Output Size
Input	Input	-	-	-	-	94	94
hidden_1	Dense	94	Relu	L2	-	94	94
dropout_1	Dropout	-	-	-	10%	94	94
hidden_2	Dense	1000	Relu	L2	-	94	1000
dropout_2	Dropout	-	-	-	10%	1000	1000
hidden_3	Dense	500	Relu	L2	-	1000	500
dropout_3	Dropout	-	-	-	10%	500	500
hidden_4	Dense	250	Relu	L2	-	500	250
dropout_4	Dropout	-	-	-	10%	250	250
hidden_5	Dense	50	Relu	L2	-	250	50
Reg_output	Output	1	Linear	-	-	50	1
Class_output	Output	1	sigmoid	-	-	50	1

4. Experiments

This section outlines the experiment settings and the dataset used to develop the proposed approach. Moreover, the data preparation and pre-processing techniques considered for this research, in addition, to the training strategy adopted to build the model are revealed. Furthermore, the standard performance measures used to assess the regression and classification results achieved by the model are defined.

In this research, Keras library [41] was coupled with the TensorFlow platform [42] and a Spyder [43] open-source environment to implement the proposed work. On the other hand, the hardware specifications include 16 GB dual-channel RAM, an Intel i7 9700 CPU, and Intel UHD GPU 630. Zakat declarations used in this research were provided by the Saudi Zakat, Tax, and Customs Authority (ZATCA) subject to releasing only information related to data and research findings that are considered as non-sensitive by ZATCA. The rationale that supports this decision is to avoid publishing information, relevant to ZATCA strategies, that can be exploited by some taxpayers to adapt their fraudulent behavior. Specifically, the data used in this research experiment consist of Zakat filings randomly selected from the Zakat declarartions collected by ZATCA between January 2018 and April 2022. Table 2 presents a high-level description of this data collection. The attributes enclosed in the ZATCA [1] dataset consist of relevant fields extracted from Zakat forms. Moreover, derived features were designed to enrich the considered dataset. However, due to confidentiality, the authors of this research can share a limited amount of details about the variables that were engineered based Zakat declaration forms. One should note that

this does not affect the effectiveness of the proposed approach that can be exploited by other researchers and fed with different set of variables.

Table 2. Data summary.

Size	51,919
Audited returns	100%
Positive cases	32.84%
Negative cases	67.16%
Business categories of taxpayers	26
Business size of taxpayers	5
Number of features	44

In order to handle the outliers that may affect the model performance, the Winsorizing [44] function was deployed on the training data. Additionally, the StandardScaler function was employed to normalize the data attributes, which exhibit highly variant scales. Moreover, the dataset exhibits some data distribution imbalance. Furthermore, an oversampling technique, namely, the Synthetic Minority Oversampling Technique (SMOTE) [45] was employed to address the considerable data imbalance and generates instances based on their closest neighbors from the minority class. Similarly, an under-sampling of the majority class was randomly performed. This yielded equally balanced class distributions.

To build the proposed model, we split the ZATCA dataset into a 60% training set and a 20% validation set to adjust the hyper-parameters, while the remaining 20% set was dedicated for testing the model's performance. Further, for the conducted experiments, several measures were used to evaluate the classification performance. Namely, the accuracy, recall, precision, and the F1-measure were calculated based on the confusion matrix shown in Table 3, where the row corresponds to actual values, and the column reports the predicted values.

Table 3. Confusion matrix.

		Predicted Class	
		Under-Report	Actual
Actual Value	Under-Report	TN	FP
	Actual	FN	TP

In particular, TP (True Positive) represents the number of VAT under-reporting cases which the model classified correctly. On the other hand, TN (True Negative) reports the number of correctly predicted actual reporting cases, while FP (False Positive) and FN (False Negative) refer to the number of misclassified cases for both classes. Accordingly, the accuracy is defined as the ratio of correctly categorized instances. It is obtained using:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

In addition, the recall represents the ratio of correctly classified under-reporting declarations over all records from this class. It is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

Similarly, the precision reports the ratio of correctly categorized declarations over all instances assigned to the under-reporting class. It is obtained as follows:

$$Precision = \frac{TP}{TP + FP} \qquad (6)$$

Finally, the F1-measure is obtained as a combination for precision and recall using:

$$F1 - measure = \frac{2 * Recall * Precision}{Recall + Precision} \qquad (7)$$

To assess the regression performance, appropriate metrics such as the Mean Square Error (MSE), the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) were used in this research experiments. Specifically, the *MSE* is defined as an absolute measure of the model's fit goodness. It is calculated using:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(h(x_i) - y_i)^2 \qquad (8)$$

where N represents the number of instances in the dataset, while y_i and $h(x_i)$ and correspond to the predicted value for the input x and the actual value, respectively. Additionally, the RMSE measures the performance of the proposed model as a square root of the *MSE* value:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(h(x_i) - y_i)^2} \qquad (9)$$

Finally, the *MAE* is calculated as the sum of the error absolute values:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|h(x_i) - y_i| \qquad (10)$$

Since the proposed model contains two output layers, we have set weights for the loss of each output layer. In other words, the alpha value (β) is set as the weight of the regression layer loss and $(1 - \beta)$ as the weight of the classification layer loss. Therefore, different values were assigned to β in order to adjust the model and get the best performance for the proposed model. Table 4 proves that the proposed model achieved the best performance using the dataset without resampling when β was set to 0.4. On the other hand, when associating the considered dataset after applying SMOTE and hybrid SMOTE with random under-sampling (RU), the proposed model achieved the best performance for β equal to 0.3.

Table 4. Results obtained using the proposed model using different values of β.

Data	β	Validation						
		Classification				Regression		
		Accuracy	Precision	Recall	F1-Measure	MSE	RMSE	MAE
No resampling	β = 0.3	0.96	0.72	0.77	0.74	0.390	0.625	0.276
	β = 0.4	0.94	0.64	0.75	0.67	0.416	0.645	0.304
SMOTE	β = 0.4	0.98	1.00	0.96	0.98	0.547	0.739	0.335
	β = 0.5	0.96	0.72	0.76	0.74	0.457	0.689	0.289
SMOTE + RU	β = 0.3	0.99	0.99	0.99	0.99	0.528	1.727	0.308
	β = 0.4	0.97	0.99	0.95	0.97	0.547	0.739	0.335

Table 5 shows that the proposed model using the dataset without resampling achieved the lowest performance using a batch size of 512 and a learning rate of 1×10^{-4}, while the model, using the dataset after applying SMOTE, achieved better performance with respect to Precision, Recall, and F1-score. Moreover, it yielded a lower regression error

when using a batch size of 256 and a learning rate of 1×10^{-3}. Additionally, as it can be seen, the proposed model yielded better performance for a batch size of 256 when applying hybrid SMOTE along with RU. One should note that while validating the model, the ReduceLROnPlateau method was used to reduce the learning rate when the loss function stops improving. The parameter 'Patience' was set to 15. This means that the learning rate is reduced if no improvement is recorded for 15 epochs and min_lr to 1×10^{-7}. Moreover, to prevent overfitting, 'EarlyStopping' was used and 'patience' was set to 100. In other words, the training stops when the validation loss does not improve for 100 epochs. One should note that, for N_t training instances and a batch size of b_s, we need N_t/b_s iterations to complete one epoch.

Table 5. Results obtained using the proposed model for different batch sizes and learning rates.

| Data	Learning Rate	Batch Size	Validation						
			Classification				Regression		
			Accuracy	Precision	Recall	F1-Measure	MSE	RMSE	MAE
No resampling	1×10^{-3}	256	0.96	0.72	0.77	0.74	0.390	0.625	0.276
	1×10^{-4}	512	0.94	0.64	0.75	0.67	0.390	0.624	0.299
SMOTE	1×10^{-3}	256	0.98	1.00	0.96	0.98	0.502	0.708	0.294
	1×10^{-4}	512	0.97	0.80	0.74	0.77	0.475	0.689	0.289
SMOTE + RU	1×10^{-3}	256	0.99	0.99	0.99	0.99	0.478	0.691	0.326
	1×10^{-4}	512	0.99	0.99	0.99	0.99	0.484	0.696	0.290

To further tune the performance of the proposed model after applying SMOTE and hybrid SMOTE with RU to the entire dataset, we investigated the model with different values of β and recorded the attained performance as depicted in Table 6. As such, the proposed model achieved better regression performance when resampling the dataset. In particular, associating 15% resampling with SMOTE and RU yielded the best classification and regression performance. The classification performance, which was initially good, did not improve further. This proves that the additional data improved the generalization of the regression model without affecting the classification capability of the model. The experiment results reported above show that training the proposed model on a larger and more balanced dataset improves the regression's performance. In other words, data resampling improved model generalization.

As the first contribution of this research consists of classifying Zakat declarations into the "under-reporting" and "actual declaration" classes, recall is more important than precision for the classification task. On the other hand, for the prioritization of auditing achieved through the regression task, the revenue at risk represents the main performance metric for this supervised learning task. Table 6 reports the results obtained using the best proposed model. Namely, the ROC curve in Figure 3a and the expected revenue at risk vs. the auditing rate in Figure 3b were achieved using the proposed model, SMOTE + RU and a sampling rate of 15%. Note that the learning rate, the batch size and β were set to 1×10^{-3}, 256 and 0.3, respectively. As it can be seen, auditing 40% of positive cases yields 79% of the expected revenue. In other words, the proposed system requires a 40% auditing coverage to collect 79% of Zakat revenue. This proves that a tradeoff between the performance of the classifier and the prioritization of auditing tasks has been successfully established. In fact, this meets the objective of the proposed model which does not only classify Zakat declarations but also prioritizes the under-reporting risk based on the expected revenue at risk for a better governance of the auditing resources. Further investigation of the results reported in Figure 3b showed that the leveled segment of the graph is caused by considerable subset of test instances that were correctly assigned to the "under-reporting"

class with high probability confidences; however, the associated expected revenue at risk was relatively small.

Table 6. Results achieved by the proposed model using different α values after data resampling by 10% and 15% using SMOTE and RU.

Data	β	Validation						
		Classification				Regression		
		Accuracy	Precision	Recall	F1-Measure	MSE	RMSE	MAE
Resample by 10% SMOTE	β = 0.4	0.98	1.00	0.96	0.98	0.547	0.739	0.335
	β = 0.5	0.96	0.72	0.76	0.74	0.457	0.689	0.289
Resample by 15% SMOTE	β = 0.4	0.98	1.00	0.96	0.98	0.479	0.722	0.319
	β = 0.5	0.96	0.72	0.76	0.74	0.417	0.663	0.274
Resample by 10% SMOTE + RU	β = 0.3	0.99	0.99	0.99	0.99	0.502	0.707	0.296
	β = 0.4	0.97	0.99	0.95	0.97	0.497	0.689	0.311
Resample by 15% SMOTE + RU	β = 0.3	0.99	0.99	0.99	0.99	0.402	0.583	0.264
	β = 0.4	0.99	0.99	0.99	0.99	0.427	0.587	0.263
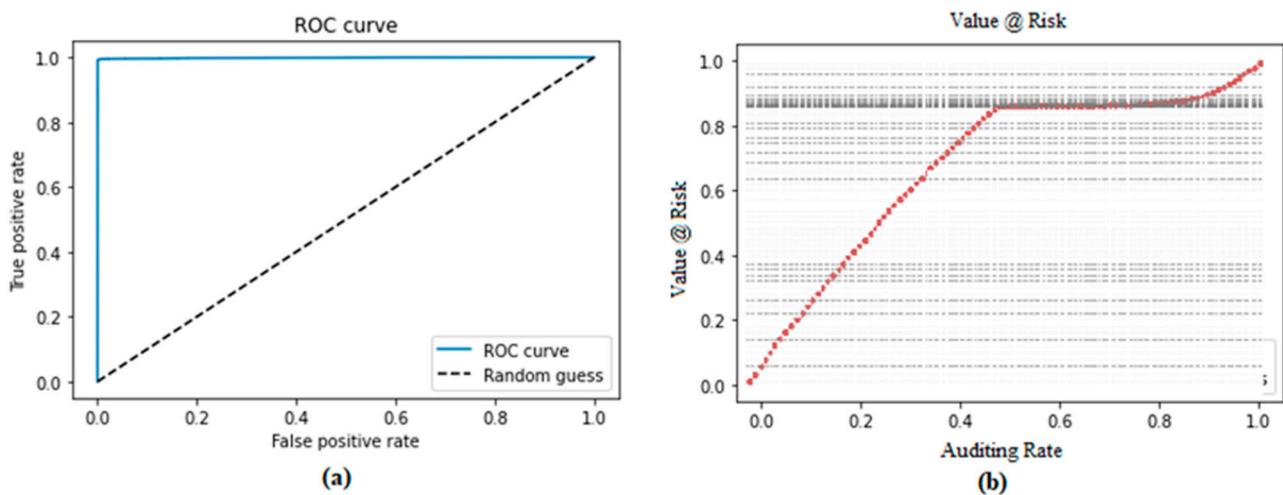
Figure 3. Results obtained using the proposed model: (**a**) ROC curve and (**b**) revenue at risk vs. auditing rate.

Finally, in Figure 4, we compare the performance of the best-proposed model with five typical machine learning models, namely k-Nearest Neighbors (KNN), the Naïve Bayes (NB) classifier, Logistic Regression (LR) [46], CART decision tree and Linear Discriminant Analysis (LDA). Particularly, the proposed model outperforms CART with an increase of 6% and 5% in terms of accuracy and F1-score, respectively. Moreover, it overtakes the LDA model with an improvement of 17% and 11% in terms of accuracy and F1-score, respectively. One should mention that in addition to this improvement recorded at the classification level, the proposed model allows effective prioritization of the detected under-valuation cases as illustrated in Figure 3.
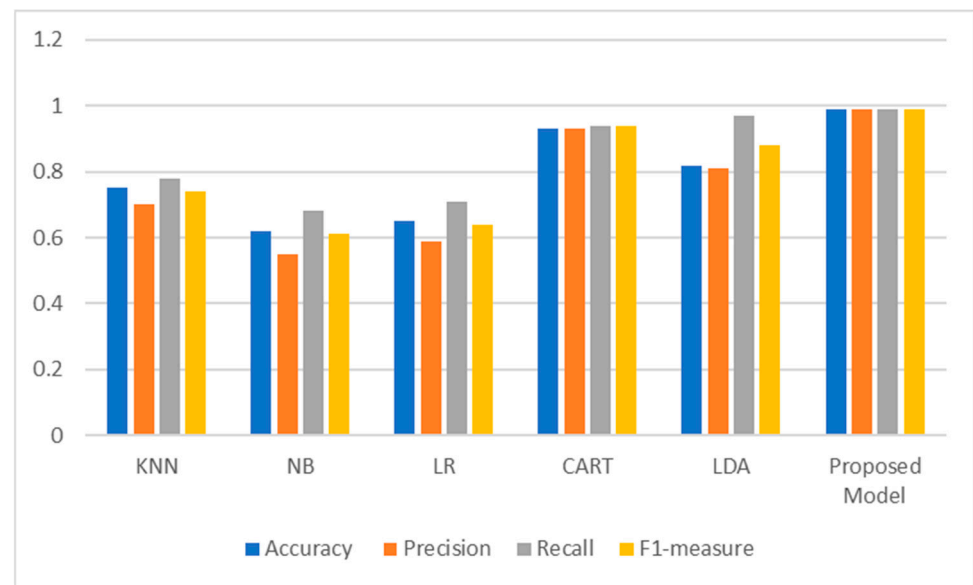
Figure 4. Classification performance obtained different using different models.

5. Conclusions

Recently, the Kingdom of Saudi Arabia (KSA) has started exploiting tax revenue to increase government investments in ambitious new initiatives or prevent drastic budget cuts. In particular, KSA has raised more taxes from non-oil activities to support fiscal consolidation. Moreover, it has witnessed modernization of the technological infrastructure of its tax administration in order to improve tax collection efficiency. In KSA, non-Saudi investors are liable for income tax. On the other hand, Saudi citizen investors (and citizens of the GCC countries) are liable for Zakat, an Islamic assessment. Typically, taxpayers are in charge of preparing and accurately reporting their Zakat declaration, which allows tax authorities to overview and audit their business activities.

Despite government efforts to increase taxpayer compliance, considerable revenue remains at under-reporting risk. Therefore, in this research, we outlined an intelligent approach to support tax authority efforts in detecting under-reporting among Zakat payer declarations. In particular, the proposed solution aims at improving detection accuracy and determining fraud cases that correspond to a higher revenue at risk. Specifically, we formulate Zakat under-reporting detection as a supervised machine learning task. Consequently, we designed a deep neural network that performs simultaneous classification and regression of Zakat declarations into the under-reporting or actual declaration classes and predicts the revenue at risk caused by this fraud, if any. In particular, the proposed network contains an input layer, five hidden layers, and two output layers for classification and regression tasks. This enables the proposed model to prioritize the auditing of specific taxpayers based on the predicted revenue at risk. The proposed model was validated and assessed using a real dataset including 51,919 Zakat declarations and standard performance metrics. Further, SMOTE improved the proposed model performance, and yielded a classification accuracy of 99% and an MAE of 26% for the regression task. Moreover, SMOTE enabled the proposed model to outperform relevant state-of-the-art supervised machine learning models.

As future work, we plan to expand the collection of Zakat declarations by considering relevant third-party data. Moreover, a dynamic approach to determine the best network architecture can be integrated in the solution. This would make the proposed work relevant to other fraud detection datasets and applications. Furthermore, a sequence of label classification techniques as well as regression frameworks will be considered for thorough empirical comparison of the proposed approach with relevant state-of-the-art solutions.

References

1. Zakat Tax and Customs Authority. Available online: https://zatca.gov.sa/ (accessed on 4 February 2023).
2. ZATCA. Rules for Calculating Zakat on a Deemed Basis. Available online: https://zatca.gov.sa/en/RulesRegulations/Taxes/Pages/CalculateZakat2.aspx (accessed on 4 February 2023).
3. Uyar, A.; Nimer, K.; Kuzey, C.; Shahbaz, M.; Schneider, F. Can e-government initiatives alleviate tax evasion? The moderation effect of ICT. *Technol. Forecast. Soc. Chang.* **2021**, *166*, 120597. [CrossRef]
4. Dias, A.; Pinto, C.; Batista, J.; Neves, E. Signaling tax evasion, financial ratios and cluster analysis. *BIS Q. Rev.* **2016**, *51*, 1–34.
5. Wu, R.S.; Ou, C.S.; Lin, H.; Chang, S.I.; Yen, D.C. Using data mining technique to enhance tax evasion detection performance. *Expert Syst. Appl.* **2012**, *10*, 8769–8777. [CrossRef]
6. Chica, M.; Hernandez, J.M.; Manrique-de Lara-Penate, C.; Chiong, R. An evolutionary game model for understanding fraud in consumption taxes [research frontier]. *IEEE Comput. Intell. Mag.* **2021**, *16*, 62–76. [CrossRef]
7. González, P.C.; Velásquez, J.D. Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Syst. Appl.* **2023**, *40*, 1427–1436. [CrossRef]
8. Chan, T.; Tan, C.-E.; Tagkopoulos, I. Audit lead selection and yield prediction from historical tax data using artificial neural networks. *PLoS ONE* **2022**, *17*, e0278121. [CrossRef]
9. González-Martel, C.; Hernández, J.M.; Manrique-de-Lara-Peñate, C. Identifying business misreporting in VAT using network analysis. *Decis. Support Syst.* **2021**, *141*, 13464. [CrossRef]
10. Vanhoeyveld, J.; Martens, D.; Peeters, B. Value-added tax fraud detection with scalable anomaly detection techniques. *Appl. Soft Comput.* **2020**, *86*, 105895. [CrossRef]
11. Fawcett, T.; Provost, F. Adaptive fraud detection. *Data Min. Knowl. Discov.* **1997**, *1*, 291–316. [CrossRef]
12. Bonchi, F.; Giannotti, F.; Mainetto, G.; Pedreschi, D. Using data mining techniques in fiscal fraud detection. In *Proceedings of the International Conference on DataWarehousing and Knowledge Discovery*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 369–376.
13. de Roux, D.; Perez, B.; Moreno, A.; Villamil, M.D.P.; Figueroa, C. Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 215–222.
14. Baghdasaryan, V.; Davtyan, H.; Sarikyan, A.; Navasardyan, Z. Improving Tax Audit Efficiency Using Machine Learning: The Role of Taxpayer's Network Data in Fraud Detection. *Appl. Artif. Intell.* **2022**, *36*. [CrossRef]
15. Tin Kam, H. Random Decision Forests (PDF). In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995.
16. Tung, A.K.H. Rule-based Classification. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer: Boston, MA, USA, 2009. [CrossRef]
17. Basta, S.; Fassetti, F.; Guarascio, M.; Manco, G.; Giannotti, F.; Pedreschi, D.; Spinsanti, L.; Papi, G.; Pisani, S. High quality true-positive prediction for fiscal fraud detection. In Proceedings of the 2009 IEEE International Conference on Data Mining Workshops, Miami, FL, USA, 6 December 2009; ICDMW'09; IEEE Computer Society: Washington, DC, USA, 2009; pp. 7–12. [CrossRef]
18. da Silva, L.S.; Rigitano, H.; Carvalho, R.N.; Souza, J.C.F. Bayesian networks on income tax audit selection—A case study of Brazilian tax administration. In Proceedings of the 13th UAI Bayesian Modeling Applications Workshop (BMAW 2016) Co-Located with the 32nd Conference on Uncertainty in Artificial Intelligence (UAI 2016), New York, NY, USA, 25 June 2016; Carvalho, R.N., Laskey, K.B., Eds.; CEUR-WS.org, CEUR Workshop Proceedings; Volume 1663, pp. 14–20.
19. Matos, T.; de Macedo, J.A.F.; Monteiro, J.M. An empirical method for discovering tax fraudsters: A real case study of Brazilian fiscal evasion. In Proceedings of the 19th International Database Engineering and Applications Symposium, Association for Computing Machinery, New York, NY, USA, 13–15 July 2015; IDEAS'15; pp. 41–48. [CrossRef]

20. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef] [PubMed]

21. Asha, R.B.; Suresh Kumar, K.R. Credit card fraud detection using Artificial Neural Networks. *Glob. Transit. Proc.* **2021**, *2*, 35–41.

22. Savić, M.; Atanasijević, J.; Jakovetić, D.; Krejić, N. Tax evasion risk management using a Hybrid Unsupervised Outlier Detection method. *Expert Syst. Appl. Int. J.* **2022**, *193*, 116409. [CrossRef]

23. Fritzke, B. A Growing Neural Gas Network Learns Topologies. In *Advances in Neural Information Processing Systems 7*; MIT Press: Cambridge, MA, USA, 1995.

24. Kohonen, T. The self-organizing map. *Proc. IEEE* **1990**, *78*, 1464–1480. [CrossRef]

25. Hardesty, L. *Explained: Neural Networks*; MIT News Office: Cambridge, MA, USA, 2017.

26. Cortes, C.; Vapnik, V. Support-vector networks (PDF). *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

27. Ghosh, S.; Douglas, L.R. Credit card fraud detection with a neural-network. In Proceedings of the Twenty-Seventh Hawaii International Conference, Wailea, HI, USA, 4–7 January 1994.

28. Mubarek, A.M.; Eşref, A.C. Multilayer perceptron neural network technique for fraud detection. In Proceedings of the S2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 5–8 October 2017.

29. Pérez López, C.; Delgado Rodríguez, M.; de Lucas Santos, S. Tax fraud detection through neural networks: An application using a sample of personal income taxpayers. *Future Internet* **2019**, *11*, 86. [CrossRef]

30. Murorunkwere, B.F.; Tuyishimire, O.; Haughton, D.; Nzabanita, J. Fraud Detection Using Neural Networks: A Case Study of Income Tax. *Future Internet* **2022**, *14*, 168. [CrossRef]

31. Neagoe, V.-E.; Ciotec, A.-D.; Cucu, G.-S. Deep convolutional neural networks versus multilayer perceptron for financial prediction. In Proceedings of the 2018 International Conference on Communications (COMM), Bucharest, Romania, 14–16 June 2018; IEEE: Piscataway, NJ, USA, 2018.

32. Wei, R.; Dong, B.; Zheng, Q.; Zhu, X.; Ruan, J.; He, H. Unsupervised conditional adversarial networks for tax evasion detection. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 1675–1680. [CrossRef]

33. Zumaya, M.; Guerrero, R.; Islas, E.; Pineda, O.K.; Gershenson, C.; Iñiguez, G.; Pineda, C. Identifying tax evasion in Mexico with tools from network science and machine learning. *Corrupt. Netw. Concepts Appl.* **2021**, 89–113. [CrossRef]

34. Kleanthous, C.; Chatzis, S. Gated mixture variational autoencoders for value added tax audit case selection. *Knowl.-Based Syst.* **2020**, *188*, 105048. [CrossRef]

35. Jinwon, A.; Sungzoon, C. Variational autoencoder based anomaly detection using reconstruction probability. *Spec. Lect. IE* **2015**, *2*, 1–18.

36. Wu, Y.; Zheng, Q.; Gao, Y.; Dong, B.; Wei, R.; Zhang, F.; He, H. TEDM-PU: A tax evasion detection method based on positive and unlabeled learning. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 1681–1686. [CrossRef]

37. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.

38. Mi, L.; Dong, B.; Shi, B.; Zheng, Q. A tax evasion detection method based on positive and unlabeled learning with network embedding features. In *Neural Information Processing*; Yang, H., Pasupa, K., Leung, A.C.S., Kwok, J.T., Chan, J.H., King, I., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 140–151.

39. Gao, Y.; Shi, B.; Dong, B.; Wang, Y.; Mi, L.; Zheng, Q. Tax Evasion Detection With FBNE-PU Algorithm Based on PnCGCN and PU Learning. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 931–944. [CrossRef]

40. Shen, J. On the singular values of Gaussian random matrices. *Linear Alg. Appl.* **2001**, *326*, 1–14. [CrossRef]

41. Available online: https://keras.io/ (accessed on 31 October 2022).

42. Available online: https://www.tensorflow.org/ (accessed on 4 February 2023).

43. Available online: https://www.spyder-ide.org/ (accessed on 4 February 2023).

44. Lee, B.K.; Lessler, J.; Stuart, E.A. Weight trimming and propensity score weighting. *PLoS ONE* **2011**, *6*, e18174. [CrossRef] [PubMed]

45. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]

46. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2000; ISBN 978-0-471-35632-5.