Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of São Paulo

André Ippolito and Augusto Cezar Garcia Lozano
Tax Intelligence Office, Under-secretariat of Municipal Revenue, Secretariat of Finance, São Paulo City Hall, 190 Libero Badaró Street, São Paulo, Brazil

Keywords: Data, Government, Decision-making, Machine Learning, Fiscal, Audit, Tax, Crime, Random Forests, Ensemble, Compliance, Revenue.

Abstract: With the advent of Big Data, several industries utilize data for analytical and competitive purposes. The government sector is following this trend, aiming to accelerate the decision-making process and improve the efficiency of operations. The predictive capabilities of Machine Learning strengthen the decision-making process. The main motivation of this work is to use Machine Learning to aid decision-making in fiscal audit plans related to service taxes of the municipality of São Paulo. In this work, we applied Machine Learning to predict crimes against the service tax system of São Paulo. In our methods, we structured a process comprised of the following steps: feature selection; data extraction from our databases; data partitioning; model training and testing; model evaluation; model validation. Our results demonstrated that Random Forests prevailed over other learning algorithms in terms of tax crime prediction performance. Our results also showed Random Forests' capability to generalize to new data. We believe that the supremacy of Random Forests is due to the synergy of its ensemble of trees, which contributed to improve tax crime prediction performance. With better predictions, our audit plans became more assertive. Consequently, this rises taxpayers' compliance with tax laws and increases tax revenue.

1 INTRODUCTION

With the large volume of data currently available, companies of various types of industries are utilizing data for competitive reasons (Provost and Fawcett, 2013). Computers now have more processing power and current algorithms can perform deeper analysis than before. This scenario has enabled the automation of data analysis, which in turn improves decision-making. Decision-making based on data, or data-driven decision-making, is the process of making decisions based on data analysis rather than intuition. In the corporate world, the practice of decision-making based on data has strong correlation with productivity growth, financial return and rise of market value.

The government area is progressively exploring data to benefit from data analysis and data-driven decision-making (Matheus et al., 2018). Governments collect data from a myriad of areas, such as traffic, energy and social security. Some of the goals related to the analysis of this data are faster and more precise decision-making, resulting in increased efficiency and effectiveness of operations.

One kind of data analysis that empowers decision-making is predictive analytics. Historical data stored in corporate databases allow risk prediction and trade opportunities discovery by means of predicitive analytics. The results of these analysis guide decisions. Machine Learning operationalizes the core techniques and algorithms of predictive analytics (Mitchell, 1997).

The main motivation of this case study is the use of Machine Learning to help decision-making in government taxes audit plans. Most of these plans aim to increase taxpayers' compliance and, therefore, can leverage government taxes revenue. Compliance means conforming to a rule, such as a policy or a law (Lin, 2016). Compliance has applicability to several areas. Some examples are compliance in healthcare, sales or taxes. Research on tax compliance leads to the conclusion that an individual pays taxes due to fear of the economic consequences of detection and punishment (Alm, 2019). Machine Learning can predict taxpayers' actions that do not comply with tax laws, such as crimes against the tax system.

We ground the work of this case study on data

Ippolito, A. and Lozano, A.
Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of São Paulo.
DOI: 10.5220/0009564704520459
In *Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020) - Volume 1*, pages 452-459
ISBN: 978-989-758-423-7

from fiscal audits of the municipality of São Paulo, related to service taxes. The city of São Paulo has an outstanding role in the Brazilian economic scenario. With regard to the Brazilian Gross Domestic Product (GDP), in the year of 2016, São Paulo had a contribution of 33.71% (São Paulo State Government, 2019). In 2019, São Paulo's revenue from municipal taxes represented 20% of the revenue collected from all the Brazilian municipalities (São Paulo Commercial Association, 2019). The predominant participation of São Paulo's tax revenue in the total income of the municipality corroborates its importance (São Paulo City Hall, 2019). In 2018, it amounted to 56.72% of the total revenue, turning out to be the main resource that comprised the total income. Among all municipal taxes, service taxes were the most relevant, corresponding to 49.97% of the revenue, followed by property taxes, which contributed with 33.45%.

One of the main actions that potentially help to increase São Paulo's contribution to tax revenue is the implementation of tax audit plans that are oriented to tax compliance. We have practical results in the Brazilian municipality of São Paulo, which demonstrate that audits originated by tax compliance actions, aiming to orientate taxpayers on how to comply with tax laws, incremented our service taxes revenue in 15%. The main cause of this increment is the rise in risk perception, since taxpayers realize they are under surveillance. Sometimes only sending messages to taxpayers telling them that they will be object of scrutiny can increase compliance (Alm, 2019).

We believe that Machine Learning can help our compliance-oriented audit plans to be more assertive, due to the predictive power of its techniques and algorithms. Thus, Machine Learning application in audit plans can reverberate, leading to higher amounts of tax revenue. Predictions of tax crimes can permit our local government to plan fiscal audits precisely, before crimes are committed, forcing taxpayers to comply with tax laws and regulations.

Some governments apply Machine Learning in crime prediction. Police in Venice (Bernasconi, 2018) and in Chicago (Fingas, 2017) utilize Machine Learning to predict crimes like robberies, shootings and murders. The Internal Revenue Service (IRS) of the United States of America (Olavsrud, 2019) applies Machine Learning to detect identity theft and pre-refund fraud in the tax system. In comparison, our work aims to predict different types of crimes that are specific to the service tax system of the municipality of São Paulo, such as denial to provide documents to fiscal authorities. Other governments use Machine Learning to tax fraud prediction. The Government of Chile (González and Velásquez, 2012) and of Spain (López et al., 2019) have case studies based on Neural Networks. In our work, we apply and compare more Machine Learning algorithms, like Random Forests, Logistic Regression and Ensemble Learning.

In this work, we apply Machine Learning techniques and algorithms with the goal of predicting service tax crimes against the tax system of the municipality of São Paulo. As input, we use data from our fiscal audits. In general terms, our methods encompass the following steps: feature selection; data extraction from our fiscal audits database; data partitioning; model training and testing; model evaluation; model validation. The results of our case study highlight Random Forests' tax crime prediction performance and also its capability of adapting to new data. We are not aware of any work with the goal to predict crimes against São Paulo's service tax system, based on Random Forests.

This paper is organized as follows: Section 2 reviews the related works, Section 3 provides theoretical background on fiscal authorities, tax audits, crimes against the tax system and Machine Learning, Section 4 explains our methods, Section 5 presents and discusses the results of our case study, Section 6 concludes the paper and suggests future work.

2 RELATED WORKS

Some related works about the use of Machine Learning in crime prediction and tax fraud detection deserve highlight. One example is the use of Machine Learning by the Italian Police with the goal of predicting crimes (Bernasconi, 2018). In this case, Machine Learning extracts patterns from data about time and localization of previous crimes. It triggers alerts, outputting where and when a crime has high probability to occur. This conduced to more precise prediction of crimes, redounding in the arrest of a man at a hotel bar in Venice just before he was about to commit a robbery.

Other case study that applied Machine Learning to predict crimes comes from the Chicago Police (Fingas, 2017). The solution analyses crime statistics, social and economic data, climate and localization registries and data from shot sensors. Whenever Machine Learning predicts a crime with high probability, the solution sends an alert to the police officers' smartphones. Chicago Police reports reduction in the number of shootings and murders after the use of Machine Learning.

Some governments have applied Machine Learning for tax fraud prediction. One example is the IRS of the United States of America, which implemented the Return Review Program (RRP) system. The main objective of the RRP is to detect fraud, identifying fraudulent returns at a lower false detection rate (McKenney, 2017). RRP aims to detect identity theft and pre-refund fraud in the tax system and it applies predictive techniques and models (Olavsrud, 2019). Reports from the IRS state that an RRP pilot of 2014 was able to improve fraud detection by 59.4% (McKenney, 2017).

There are other case studies in the government area related to tax fraud prediction. One example is originated from the Tax Administration of Chile (González and Velásquez, 2012), which applied Decision Trees, Neural Networks and Bayesian Networks to detect taxpayers who use false invoices. In their case study, Neural Networks prevailed over the other algorithms, correctly detecting 92% of the fraud cases.

Another example of tax fraud prediction based on Machine Learning comes from the Spanish Institute of Fiscal Studies (López et al., 2019). Their study applied Neural Networks to data from the Spanish Revenue Office, with the goal of identifying taxpayers who evade tax. Their model yielded 84% of correct predictions.

Our work differentiates from these case studies. Firstly, comparing to the police cases and IRS, the type of crimes we aim to predict are peculiar and more specific to the service taxes scenario, also embracing more crimes, like denial to provide invoices to fiscal authorities. Secondly, in comparison to the Tax Administration of Chile and the Spanish Institute of Fiscal Studies, we applied and evaluated more Machine Learning algorithms, such as Random Forests, Logistic Regression and Ensemble Learning. Besides, these governments' tax systems are distinct from São Paulo's tax system. This implies that the laws taxpayers have to comply with, when executing services in the city of São Paulo, are also different and, consequently, the types of taxpayers' behaviors and crimes are distinct too. Our results showed the supremacy of Random Forests, with regard to tax crime prediction performance. To our knowledge, there are not works that aimed to predict crimes against São Paulo's service tax system, using Random Forests.

3 THEORETICAL BACKGROUND

In the following subsections, we define fiscal authorities, tax audits and crimes against the tax system, according to Brazilian's laws and regulations, also exemplifying some of these crimes. In the sequence, we explain the main concepts and foundations of Machine Leaning, describing the main features of the algorithms that we applied in this case study.

3.1 Fiscal Authorities

Fiscal authorities play an essential role in engendering tax revenue. According to Brazilian tax laws, fiscal authorities are the individuals authorized to collect taxes for the municipality, having the exclusive competency to constitute the tax credit. In order to constitute this credit, fiscal authorities have to operationalize tax audits.

3.2 Tax Audits

A tax audit is an inspection process that mainly comprises verifying the inception of the tax obligation, calculating the amount of tax that is due, identifying the taxpayer and proposing, if applicable, the tax penalty. Fiscal authorities have the functional responsibility for charging taxes and fines. Their activities, under their jurisdiction, have precedence over other administrative sectors.

3.3 Crimes against the Tax System

According to Brazilian tax laws, crimes against the tax system are crimes that aim to suppress or reduce taxes. These crimes are committed when taxpayers omit information, make false declarations, defraud fiscal documents, create false or inexact documents, and deny providing documents or invoices to fiscal authorities.

3.4 Machine Learning

In Machine Learning, algorithms learn from experience with respect to a task and performance measure, if its performance measure at the task improves with experience (Mitchell, 1997). Machine Learning algorithms learn from data, acquiring experience by adjusting its parameters based on data features (a.k.a. attributes), characteristics and patterns, in order to obtain its best performance measure.

One of the usual tasks in Machine Learning is classification. Algorithms specialized in this task have to specify which category some input data belong to (Goodfellow et al., 2019). Classification can be supervised or unsupervised. In the former, data has labels previously categorizing elements into the classes we are seeking. In the latter, there are no labels for categories and it is necessary to apply specific algorithms that learn the categories from data, forming groups based on similarities of data elements (a.k.a. instances).

Perfomance measures for classification are based on the following values, which we describe using the context of our case study:

- True Positives (TP): number of criminals correctly classified;
- False Positives (FP): number of taxpayers that are not criminals but are incorrectly classified as such;
- False Negatives (FN): number of criminals not classified as such;
- True Negatives (TN): number of not criminals correctly classified.

Some usual performance measures are accuracy (ACC), recall (R), precision (P), F-measure (F) and specificity (S) (Hossin and Sulaiman, 2015). Given a dataset of N taxpayers, these measures are calculated according to the following formulas:

$$ACC = \frac{TP + TN}{N} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$F = \frac{2\,x\,R\,x\,P}{R + P} \tag{4}$$

$$S = \frac{TN}{TN + FP} \tag{5}$$

In our study, we applied supervised classification, using the following algorithms: Neural Networks, Naive Bayes, Decision Trees, Ensemble Learning, Random Forests and Logistic Regression.

3.4.1 Neural Networks

Neural Networks (Hardesty, 2017) are models represented by interconnected nodes that form a net. These networks usually solve classification problems. The neural network receives data values and features. It calculates weights for these features, with the objective of minimizing the error between the predicted classification and the actual classification. These weights are initially set to random values. In the sequence, an iterative process begins to adjust these weights, in order to minimize the error between the predicted classification and its true labels.

3.4.2 Naive Bayes

Naive Bayes (Zhang, 2019) is a probabilistic algorithm used for classification. It assumes independence among data features. This means that the presence of a particular feature is not probabilistically related to other features used in a model. It is mainly based on conditional probabilities and its application to classification enables to calculate the probability of classification in one of the classes of the data, given the value of features.

3.4.3 Decision Trees

Decision Trees (Poole and Mackworth, 2017) are structure-based models for classification. They are represented as hierarchies that form trees, in which nodes represent data features. Arcs coming from a node represent possible values of a feature. Moving down to the lowest level of the tree's hierarchy, leaves are reached, representing possible classifications of data elements. The starting node of a decision tree corresponds to the data feature that partitions data elements into the most homogeneous groups as possible. This homogeneity is measured by means of entropy (Mitchell, 1997) such that the less entropy a group has the more homogeneous are its data elements. The following node of the tree is the remaining data feature that best partitions the data in homogeneous groups. The process of selecting the features that represent the tree's nodes continues in this manner until all the features are represented in the tree.

3.4.4 Ensemble Learning

Ensemble Learning (Rokach, 2010) combines the results of multiple learning algorithms, aiming to yield better performance than can result from the isolated application of any of its constituent algorithms. For combination of multiple classifiers, the most commonly applied methods are simple fusion methods (Kuncheva, 2002). These methods combine the output of a committee of classifiers. If this output is given by probabilities, the final classification is measured calculating the maximum, average or median probability of all classifiers, for example. If the classifiers' output is discrete, e.g. binary classification, the committee classification results from a majority vote, in which the most frequent class label among the classifiers is the final classification. Usually, for better results, it is beneficial that the committee is composed of heterogeneous classifiers.

3.4.5 Random Forests

Random Forests (Breiman, 2001) are based on the construction of various decision trees. These trees are combined, such that a random forest is an ensemble of decision trees. In practice, a training set of data elements is drawn randomly for each tree. In addition, a subset of the input data features is sampled at random. Each tree is grown with these sets without pruning. To minimize the error in classification, trees of the ensemble must be the least similar as possible, to augment the synergy among the trees. Each tree predictor of a forest outputs a class and majority vote determines the Random Forests' prediction. Its basic premise is that an ensemble of decision trees will outperform any of the individual trees solely considered.

3.4.6 Logistic Regression

Logistic Regression (Russell and Norvig, 2010) is a classification algorithm based on Statistics and in its basic form classifies data elements in two classes (binary classification). It uses a logistic function, which is based on the natural logarithm. The logistic function outputs numbers between 0 and 1 that are interpreted as the probability of a data element belonging to a class. Logistic Regression fits weights to data features, minimizing the error between the predicted class and the true class.

4 MATERIAL AND METHODS

This case study aimed at applying Machine Learning to fiscal data, with the goal to predict crimes against the service tax system. We used historic data from our fiscal audits plan of action and from our face-to-face fiscal audits. We applied some of the main Machine Learning models and algorithms, which were compared with respect to performance measures.

Our objective was to obtain a predictive model that is able to assertively foresee which business taxpayers will commit crimes against the tax system. For this purpose, our basis to calibrate our model was the historic fiscal data and the Machine Learning algorithms that we applied to the data.

To apply our proposed methods, we used the open source tool KNIME (Berthold et al., 2009). This tool enabled us to explore data, apply Machine Learning algorithms, compare their performance measures and select the best-performing one. We operationalized these tasks with the use of workflows that KNIME's functionalities permit to build.

As illustrated in Figure 1, the process of our methods comprises the following steps:

1) Feature selection
2) Data extraction from our fiscal audits database
3) Data partitioning
4) Model training and testing
5) Model evaluation
6) Model validation

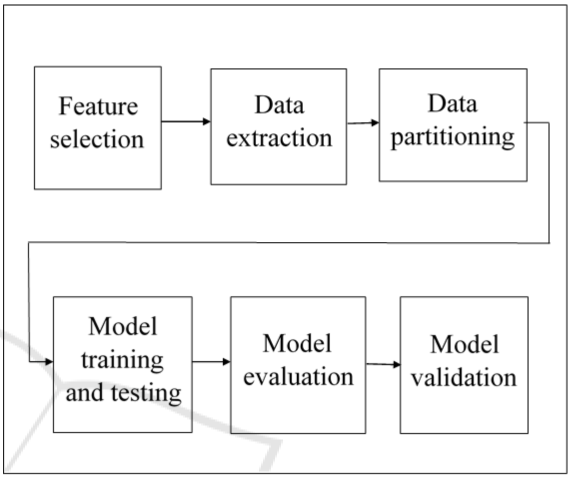

Figure 1: The process of our methods is composed of six steps: we select features; extract data from our database based on these features; partition data into train and test subsets; train and test the model applying different algorithms; evaluate and validate the model.

4.1 Feature Selection

In step 1, based on our data exploration findings and expertise, we selected the attributes that composed our model. We selected three features:

- anual tax value declared in invoices in which the taxpayer declared himself as a firm of professionals specialized in a unique service, such as a company of lawyers or a company of architects;
- anual tax value declared in invoices in which the taxpayer declared himself as a small business that is permitted to use simplified procedures to comply with tax service obligations;
- a binary field containing the information whether or not the taxpayer committed a crime against the tax system in previous tax audits (class label).

In Table 1, we give an example of an instance of our dataset with such attributes.

Table 1: Example of an instance of our dataset.

tax payer ID	anual tax value as a specialized company	anual tax value as a small business	crime
999999999	150,000	90,000	1

4.2 Data Extraction

In step 2, we selected data based on our feature selection. We collected data from face-to-face fiscal audits and plans of action of 2016, 2017 and 2018, which are stored in our relational database.

Originally, a fiscal audit is modelled in our database as an entity, with fields containing data related to the identification of the taxpayer, the date of the beginning and end of the fiscal audit, the audit goals and whether or not the taxpayer comitted a crime. In our database, a plan of action is also an entity, with fields that identify the taxpayer and the estimated tax fine he would be submitted to, based on the service income he declares in his invoices and also the declaration of his specific economic conditions (e.g. simple business), which implies in different tax obligations to comply with.

We extracted data applying Structured Query Language (SQL) scripts. In the sequence, we converted the resulting dataset to a spreadsheet format. There were no missing values and we did not need to transform data types, since the algorithms applied in the case study are all adequate to binary and numerical data.

4.3 Data Partitioning

In step 3, we separated our dataset into two subsets: one subset for calibration of our model (training and test data), comprised of data from 2016 and 2017; other subset to validate our model, comprised of data from 2018. The first subset has 151 cases (instances), comprised of 91 crime cases (60%) and 60 cases that are not crimes (40%). Tax values for this subset ranged from 6,030.27 reals (Brazilian currency) to 1,718,637.72 reals, having an average value of 200,589.25 reals. The validation subset has 66 cases, 36 of them are crimes (55%), while 30 are not crimes (45%). In this subset, tax values range from 50,799.00 reals to 5,936,188.34 reals, having an average value of 250,035.40 reals.

4.4 Model Training and Testing

In step 4, we trained and tested our model, using a 10-fold cross validation method, applying Machine Learning algorithms to our training and test subset of data. We applied six algorithms: Neural Networks, Naive Bayes, Decision Trees, Logistic Regression, Random Forests and Ensemble Learning. The last one is an ensemble of the other five algorithms, such that the resulting prediction corresponds to the prediction with highest probability among the classifiers of the ensemble.

4.5 Model Evaluation

In step 5, we evaluated the results obtained by each algorithm applied to the model, comparing them in terms of their performance measures: recall, precision, F-measure, accuracy and specificity.

4.6 Model Validation

In step 6, in order to validate our model, we applied the best-evaluated algorithm to our validation subset of data. In this step, we also calculated the performance measures of the algorithm.

5 RESULTS AND DISCUSSION

After applying our methodology to the dataset comprising the fiscal data of 2016 and 2017, the algorithms achieved the performance measures listed in Table 2. We verified that Random Forests yielded the highest scores in the majority of the performance metrics utilized (precision, accuracy and specificity), considering the algorithms used in the case study. The resulting model corresponds to an ensemble of 100 decision trees.

We validated the model adjusted by Random Forests against fiscal data of 2018. The resulting performance metrics are listed in Table 3.

We believe that Random Forests results are due to its power to ensemble multiple decision trees, which enables the algorithm to adjust a model that best captures the synergy of a multitude of trees. Besides, when the adjusted model was applied to unseen fiscal data of 2018 (validation step), the results indicated the model's capability to generalize to new data.

Table 2: Performance measures of the algorithms in the evaluation step.

Machine Learning Algorithm	Recall	Precision	F-measure	Accuracy	Specificity
Random Forests	0.516	**0.870**	0.648	**0.662**	**0.883**
Naive Bayes	**0.890**	0.609	**0.723**	0.589	0.133
Decision Trees	0.681	0.674	0.678	0.609	0.500
Logistic Regression	0.571	0.650	0.608	0.556	0.533
Ensemble Learning	0.670	0.642	0.656	0.576	0.433
Neural Networks	0.538	0.790	0.641	0.636	0.783

Table 3: Performance measures of Random Forests in the validation step.

Recall	Precision	F-measure	Accuracy	Specificity
0.889	0.640	0.744	0.667	0.400

6 CONCLUSIONS AND FUTURE WORK

Contemporarily, as technology evolved, computers can provide more processing power and algorithms are capable of performing deeper analysis than before. This context enables data analysis automation, which facilitates decision-making. Governments are progressively benefiting from data analysis, seeking to improve efficiency and effectiveness of its operations.

Predictive analytics is one branch of data analysis that empowers decision-making. The core of techniques and algorithms that permits predicitive analytics emanates from Machine Learning. This core can support the decision-making process in government taxes audit plans. Tax crimes prediction with Machine Learning allows local governments to precisely plan fiscal audits before crimes against the tax system happen, making taxpayers comply with tax laws and regulations. Consequently, this naturally leverages tax revenue of local governments.

This case study aimed to apply Machine Learning to predict service tax crimes against the tax system of the municipality of São Paulo. We extracted data from our fiscal audits of 2016, 2017 and 2018. In our methodology, we implemented the following steps: feature selection; data extraction from our fiscal audits database; data partitioning; model training and testing; model evaluation; model validation. For training, testing and evaluating our model, we used fiscal data of 2016 and 2017 and utilized the following algorithms: Neural Networks, Naive Bayes, Decision Trees, Logistic Regression, Random Forests and Ensemble Learning.

Our results demonstrated that Random Forests excelled the other algorithms with regard to tax crime prediction performance metrics. Besides, we also validated the model adjusted by Random Forests, applying it to previously unseen data (fiscal data of 2018). We believe that the results achieved are justified by the fact that Random Forests are an ensemble of decision trees. These combination of multiple trees helps to strengthen the performance results, such that the synergy of the various decision trees contributes to make the predictions more precise. We are not aware of any work with the objective of predicting crimes against São Paulo's service tax system, based on Random Forests.

We also conclude that the use of Machine Learning contributes to the success of our fiscal audit plans. The main reason for this contribution is the fact that Machine Learning enables us to predict crimes against the law system, adjusting models that can more precisely make these predictions. These correct predictions guide the decision of planning fiscal audits more assertively.

In addition, in our work, a software tool that is embedded in a computer implements the process of crime prediction. Thus, we are able to enrich crime prediction with fastness and automation. This enrichment accelerates the decision process of our audit plans, which in turn causes the rise of fiscal presence and feeling of surveillance by taxpayers. Therefore, this results in more compliance to tax laws, leading to an increase in tax revenue.

Besides, we structured the process of our solution in a workflow, which contributes with ease of maintenance and evolution, regardless of the complexities of code implementation.

As future work, we consider the application of Machine Learning to predict tax fines values due to tax laws violations. This can be achieved by adjusting, testing, evaluating and validating regression models, for example, based on historical data of tax fines values. Aside from this, other techniques and algorithms can be added to the workflow of our solution, such as Principal Component Analysis (Johnson and Wichern, 2008) and Deep Learning (Goodfellow et al., 2019). A shortcoming of our work is the fact that we did not consider the influence of time in tax crime prediction. We can incorporate the effects of time adding a variable that represents the date of the tax law violation, to analyze if it improves the prediction capability of the model.

REFERENCES

Alm, J. (2019). What motivates tax compliance. In *Tulane Economics Working Paper Series*.

Bernasconi, F. (2018). Venezia, ecco il software che predice tutti i reati. In *Il Giornale*. Available at: http://www.ilgiornale.it/news/cronache/venezia-ecco-software-che-predice-tutti-i-reati-1603366.html. Accessed in January, 2020.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., and Wiswedel, B. (2009). KNIME - the Konstanz Information Miner: Version 2.0 and Beyond. *SIGKDD Explor. Newsl.*, 11(1).

Breiman, L. (2001). Random forests. In *Machine Learning, Volume 45, Issue 1*. Springer.

Fingas, J. (2017). Chicago police see less violent crime after using predictive code. In *Engadget*. Available at: https://www.engadget.com/2017/08/06/chicago-police-see-less-crime-after-predictive-code/. Accessed in January, 2020.

González, P. C. and Velásquez, J. D. (2012). Characterization and detection of taxpayers with false invoices using data mining techniques. In *Expert Systems with Applications*. Elsevier.

Goodfellow, I., Bengio, Y., and Courville, A. (2019). *Deep Learning*. MIT Press.

Hardesty, L. (2017). Explained: Neural networks. In *MIT News*. Available at: http://news.mit.edu/2017/explained-neural-networks-deep-learning-0414. Accessed in January, 2020.

Hossin, M. and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. In *International Journal of Data Mining and Knowledge Management Process*.

Johnson, R. A. and Wichern, D. W. (2008). *Applied Multivariate Statistical Analysis*. Pearson, 6^{th} edition.

Kuncheva, L. I. (2002). A theoretical study on six classifier fusion strategies. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Lin, T. C. W. (2016). Compliance, technology, and modern finance. In *Brooklyn Journal of Corporate Financial and Commercial Law, Vol. 11*. Temple University Legal Studies Research Paper.

López, C. P., Rodríguez, M. J. D., and Santos, S. L. (2019). Tax fraud detection through neural networks: An application using a sample of personal income taxpayers. In *Future Internet*. MDPI.

Matheus, R., Janssen, M., and Maheshwari, D. (2018). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. In *Government Information Quarterly*. Elsevier.

McKenney, M. E. (2017). The return review program increases fraud detection; however, full retirement of the electronic fraud detection system will be delayed. In *Treasury Inspector General for Tax Administration*. Available at: https://www.treasury.gov/tigta/auditreports/2017reports/201720080fr.pdf. Accessed in January, 2020.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, 1^{st} edition.

Olavsrud, T. (2019). IRS combats fraud with advanced data analytics. In *CIO*. Available at: https://www.cio.com/article/3435364/irs-combats-fraud-with-advanced-data-analytics.html. Accessed in January, 2020.

Poole, D. and Mackworth, A. (2017). *Artificial Intelligence: Foundations of Computational Agents*. Cambridge University Press, 2^{nd} edition.

Provost, F. and Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. In *Big Data*. MDPI.

Rokach, L. (2010). Ensemble-based classifiers. In *Artificial Intelligence Review*.

Russell, S. J. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Pearson, 3^{rd} edition.

São Paulo City Hall (2019). Relatório técnico do balanço geral de 2018. Accounting Department.

São Paulo Commercial Association (2019). Impostômetro. Available at: https://impostometro.com.br. Accessed in January, 2020.

São Paulo State Government (2019). Conheça São Paulo. Sistema Estadual de Análise de Dados (SEADE). Available at: https://www.seade.gov.br/wp-content/uploads/2019/01/Conheca_SP_2019_jan29.pdf. Accessed in January, 2020.

Zhang, Z. (2019). Naive bayes explained. In *Towards Data Science*. Available at: https://towardsdatascience.com/naive-bayes-explained-9d2b96f4a9c0. Accessed in January, 2020.