2016 15th IEEE International Conference on Machine Learning and Applications

Deep Learning Anomaly Detection as Support Fraud Investigation in Brazilian Exports and Anti-Money Laundering

Ebberth L Paula*, Marcelo Ladeira†, Rommel N. Carvalho†‡ and Thiago Marzagão‡
*Coordination of Research and Investigation (COPEI)
Secretariat of Federal Revenue of Brazil (RFB), Brasilia, DF, Brazil
Email: ebberth.paula@receita.fazenda.gov.br
†Department of Computer Science (CIC)
University of Brasilia (UnB), Brasilia, DF, Brazil
Email: {mladeira,rommelnc}@unb.br
‡Department of Research and Strategic Information (DIE)
Ministry of Transparency, Monitoring and Control (MTFC), Brasilia, DF, Brazil
Email: {rommel.carvalho,thiago.marzagao}@cgu.gov.br
§Tyrell Inc., 123 Replicant Street, Los Angeles, California 90210–4321

Abstract

Normally exports of goods and products are transactions encouraged by the governments of countries. Typically these incentives are promoted by tax exemptions or lower tax collections. However, exports fraud may occur with objectives not related to tax evasion, for example money laundering. This article presents the results obtained in implementing the unsupervised Deep Learning model to classify Brazilian exporters regarding the possibility of committing fraud in exports. Assuming that the vast majority of exporters have explanatory features of their export volume which interrelate in a standard way, we used the AutoEncoder to detect anomalous situations with regards to the data pattern. The databases used in this work come from exports of goods and products that occurred in Brazil in 2014, provided by the Secretariat of Federal Revenue of Brazil. From attributes that characterize export companies, the model was able to detect anomalies in at least twenty exporters.

1. Introduction

Several authors ([1], [2], [3], and [4]) indicate money laundering cases with the use of foreign trade, thus taking advantage of the difficulties of the countries to exchange information massively, to operate the 'clean' money. The US Immigration and Customs Enforcement [5] define Trade-Based Money Laundering as "an alternative remittance system that allows illegal organizations the opportunity to earn, move and store proceeds disguised as legitimate trade. Value can be moved through this process by false-invoicing, over-invoicing and under-invoicing commodities that are imported or exported around the world". In Brazil, the law is explicit as to the application of money laundering to those who import or export goods that do not correspond to their true value [6].

According to the Egmont Group[1], "Money laundering is the process by which the criminal transforms resources from illegal activities in assets with an apparently legal source. This practice generally involves multiple transactions, to hide the source of financial assets and allow them to be used without compromising the criminals. The concealment is thus the basis for all washing operations involving money from a criminal history".

Brazilian exports are directed annually to nearly 200 countries. Thousands of invoices with tax suspension on goods destined for export are issued daily. About 50,000 legal entities directly or indirectly operated in shipping goods and merchandise abroad annually. The *Mercosur Common Nomenclature* (NCM)[2], used for the tax classification of goods, distinguishes between 9,600 types of goods and merchandise, each subject to specific legislation. Most of the variables are nonlinearly correlated and temporally dependent. It is difficult for humans to distinguish the normal state from the abnormal state only by looking at the raw data. For this reason, training a machine to learn the normal state and displaying the reconstruction error as the anomaly score is valuable.

This paper presents results of applying unsupervised deep learning AutoEncoder in databases of foreign trade of the Secretariat of Federal Revenue of Brazil with the objective of identifying exporting corporations whose explanatory variables of their export operations in 2014 show signs of divergence (anomalies) compared to regular patterns found.

This article is structured as follows: Section 2 presents the work related to money laundering, fraud and error detection on imports. It also presents the state-of-the-art data mining techniques for high-cardinality attributes with non-

1. International group created to promote worldwide the treatment of suspected communications related to money laundering. http://www.egmontgroup.org/

2. The Mercosur Common Nomenclature (MCN) was adopted by the countries that integrate the Argentina, Brazil and Uruguay Block to foster international trade growth, make the creation and comparison of statistics easier, in addition to elaborating freight tariffs and providing other relevant information to international trade. http://bit.ly/29wHa1T

978-1-5090-6167-9/16 $31.00 © 2016 IEEE
DOI 10.1109/ICMLA.2016.73

linear relationships. Section 3 presents the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology used in this study. Section 4 addresses the current scenario in Brazil for fraud detection in exports and combating money laundering. Section 5 presents the understanding of the data and their preparation for modeling. Section 6 addresses the modeling process and evaluation of the model. Finally, Section 7 presents the conclusion and future work.

2. Related Works

In this section we present some of the most relevant works related to the application of data mining techniques in the field of combating money laundering and fraud.

Applications developed for the financial system represent most of the articles that use data mining techniques for money laundering detection, even when searching for papers from more than ten years ago. To the best of our knowledge, there are no applications involving trade-based money laundering detection. Nevertheless, there are works that use artificial intelligence for this purpose via *Financial Crimes Enforcement Network* in 1995 [7] and 1998 [8]. Unfortunatelly, these articles do not specify the databases used.

Larik and Haider [9] approach the problem of dirty money entering the financial system with a hybrid approach for detecting anomalies in financial transactions. This approach employs unsupervised clusters to meet normal standards of behavior for clients in conjunction with the use of statistical techniques to identify the diversion of a particular transaction of the corresponding expected behavior in their group. A variant of the Euclidean Adaptive Resonance Theory (EART) is suggested to group clients into different clusters. The perspective of the authors, unlike what is discussed in this paper, is a financial institution with a focus on transactions.

Khan et al. [10] present a Bayesian network approach (BN) to analyze transactions of customers of a financial institution in order to detect suspicious patterns. Based on transaction history, the proposed approach assigns a baseline from which the transaction becomes suspect. The problem with this approach when transposed to this work domain is the absence of a relevant historical period.

Raza and Haider [11] join the two approaches mentioned above to create what they called *Suspicious Activity Reporting using Dynamic Bayesian Network* (SARDBN), a combination of clustering with *dynamic Bayesian network* (DBN) to identify anomalies in sequences of transactions. The authors created an index called *Anomaly Index Rank and using Entropy* (AIRE), which measures the degree of abnormality in an operation and compares it with a predefined threshold value to mark the transaction as normal or suspicious. This index is similar to the baseline proposed by Khan et al. [10]. However, this division into two phases appear to suffer less of the problems outlined in the previous section, because the clustering first evaluates all the customers and the AIRE evaluates transactions of a given client individually.

Rajput et al. [12] address the problem by proposing ontologies and rules written in *Semantic Web Rule Language* (SWRL). Such an approach, according to the authors, require less computation and allows the reuse of the knowledge base in similar areas.

In the money laundering domain, Sharma and Panigrahi [13] show that the technical data mining and logistic models, neural networks, Bayesian networks, and decision trees have been extensively applied to provide solutions to the problems of fraud detection and classification. From the study of forty-five articles on fraud in the financial system, the authors present four groups of approaches in mining commonly used data. Table 1 presents a summary of the survey.

TABLE 1. Approaches to fraud detection in the finance domain

Method	% of papers
Regression models	40%
Neural Network	31%
Fuzzy Logic	16%
Genetic algorithms and specialist systems	13%

Finally, Jambeiro and Wainer [14], [15], when examining the use of Bayesian methods in a practical interest of pattern classification problem for the Secretariat of Federal Revenue of Brazil from a similar basis (databases of imports-trade and NCM) to the one proposed in this work (databases of exports-trade and NCM), showed empirically that more advanced Bayesian strategies for the treatment of high cardinality of attributes (pre-processing for cardinality reduction and substitution of conditional probability tables, Bayesian networks, default tables, decision trees and decision graphs) although they bring specific benefits, do not result in overall performance gain in our target domain. Their work then turned to propose a new Bayesian classification method, named Hierarchical Pattern Bayes (HPB). “The HPB runtime is exponential in the number of attributes, but is independent of its cardinality. Thus, in areas where the attributes are few, but have high cardinality, it is much faster” than traditional algorithms.

2.1. State-of-the-art

In this work we chose to use Deep Neural networks AutoEncoders. This tool, in addition to dealing with the problems faced by Jambeiro [14], [15], allows unsupervised (AutoEncoder) and semi-supervised detection of anomalies. When compared to most related works of the financial system, it has the advantage of performing nonlinear generalizations.

Deep Learning has emerged as one capable of reaching the state-of-the-art algorithm for various domains: Szegedy et al. [16] propose a deep convolutional neural network architecture that achieves the new state-of-the-art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014. Jaiswal et al. [17] achieve state-of-the-art performance on the FERA-2015 Challenge dataset recognizing

spontaneous facial expressions. Liang et al. [18] achieve state-of-the-art of Atari Games using shallow reinforcement learning with a recently introduced Deep Q-Networks (DQN) algorithm - a combination of deep neural networks and reinforcement learning.

The future of deep learning is unsupervised learning, but it has been overshadowed by the successes of purely supervised learning [19]. Semi-supervised learning follows the same path. Problems with or without a small subset of the observations having a corresponding class label are of "immense practical interest in a wide range of applications where unlabeled data is abundant, but obtaining class labels is expensive or impossible to obtain for the entire data set" [20].

3. Methodology

This study used as reference model the Cross Industry Standard Process for Data Mining (CRISP-DM) [21], since it is a well-known data mining reference model. The CRISP-DM methodology is flexible and allows the creation of a model that fits the specific needs of projects. It is observed that the execution sequence of the phases is not rigid and depends on the results achieved in each phase (see Figure 1).

Fig. 1. Phases of the CRISP-DM Process Model

The life cycle of the mining project on this methodology consists of six phases:

1) *Business understanding* This initial phase focuses on understanding the goals and project requirements from a business perspective, then converting that knowledge into a definition of the data mining problem and a preliminary plan designed to achieve the objectives.
2) *Data understanding* The data understanding phase starts with the initial data collection and continues with activities that allow the familiarization with the data, the identification of data quality problems, the discovery of the first insights into the data and/or detection of interesting subsets to form hypotheses about the unknown information. Sections 1 and 4 of this paper summarize the results of this step of the methodology.
3) *Data preparation* The data preparation phase concentrates all activities necessary to the construction of the final data set to be used in the modeling phase. Data preparation tasks are typically performed several times and not in any prescribed order. The tasks include selecting, cleaning, constructing, integrating and formatting data for modeling purposes.
4) *Modeling* At this stage, several modelling techniques are chosen and applied and its parameters are adjusted to the optimum values. Usually, there are many different techniques for the same data mining problem. Some techniques have specific requirements regarding the form of the data. Thus, it is often necessary to go back to previous phases to perform adjustments.
5) *Evaluation* In this phase, it is important to evaluate and review the steps performed to create the final model (or models), before final deployment, to make sure it achieves the business objectives. It is important to try to determine if there are any important business issues not yet considered. At the end of this phase, it is important to decide if the results are satisfactory and whether the final model should be used or not.
6) *Deployment* The knowledge obtained with the models generated must be applied in the Organization and this knowledge must be disseminated and presented to users in a way that they can use it.

4. Scenario

In Brazil, in 2012, the Law 9613/98 (amended by Law 12,683/12) [6], brought important advances in preventing and combating money laundering with the extinction of the exhaustive list of predicate criminal offenses. Now any criminal offense is considered a precedent to money laundering. This law establishes a framework to combat money laundering and related crimes in which the Secretariat of Federal Revenue of Brazil plays an important role in fiscal intelligence.

In this context, the Secretariat of Federal Revenue of Brazil is responsible, among other related duties, to "plan, coordinate and implement the tax intelligence activities in the fight against laundering and concealment of assets, rights and values" [22]. The cases that may relate to money laundering crimes are selected for investigation from various mechanisms such as complaints, audits, lawsuits, cross-checking, among others.

It is intended that the presented data mining techniques will join the currently existing mechanisms for selection of suspected exports frauds. Besides detecting anomalies related to fraud and money laudering, the analysis of complaints against companies can also benefit from models generated by these techniques. The predictive variables for the company in question may be submitted to the model for evaluation of their suspicion.

5. Data Understanding and Data Preparation

It was identified eighty attributes that proved sufficient to characterize fraudulent exports based on the experience of the author and empirical studies conducted. These attributes are distributed in ten different dimensions:

1) *Registration* Registration data that allow unequivocal identification of the exporter and its license to operate in foreign trade.
2) *Foreign Trade* Export volumes and values, commercial classification of goods and products, origin and destination of goods and products.
3) *Tax Collection* Amounts charged and paid in fees and taxes in the years in which the exporting company conducted export activities.
4) *Financial Transactions* Transacted values in Brazilian financial institutions, consolidated per year, bank accounts (debit and credit), credit cards, and foreign exchange transactions (purchase, sale, and transfer).
5) *Tax Withheld at Source* Amounts related to taxes that companies are required to hold upon payment for services.
6) *Employees* Amounts collected by the exporting companies in the form of social security of its employees.
7) *Electronic Invoices* Information about electronic invoices emitted when the company purchases goods and products for commercialization and industrialization and about electronic invoices emitted when the company sells goods and manufactured products for export.
8) *Supplementary Obligations* Information regarding compliance with the obligation to deliver different types of declarations to the tax authorities.
9) *Inspection Operations* Information regarding tax and customs inspections already conducted in exporting companies.
10) *Others* Information concerning surveillance operations already carried out in the exporting companies.

One of the proposed models to be evaluated in the next phase of the Crisp-DM was Deep Learning AutoEncoder (see Section 6). For detection of anomalies in this model it is necessary that the "predictive attributes" reflect the phenomenon on which anomalies are sought. Thus, these eighty attributes went through two changes: 1) using Gradient Boosted Machines (GBM), we identified eighteen attributes able to explain 80% of the variability of the volumes exported by the companies; 2) for the unsupervised learning model to effectively detect anomalies related to exports, these eighteen attributes were then relativized from the formula shown below in eighteen indices, which were then used to learn the unsupervised model.

The relativization of predictive attributes is responsible for creating indexes that effectively reflect the participation of the attributes in the phenomenon in which anomalies are sought: the amount exported. For example, given the exploratory attribute *financial transactions*, the relativization transforms this attribute in *amount exported by financial transactions unit*.

Thus, the formula below indicates that given i explanatory attributes x, $Index_{x_i}$ indicates the Amount of exports for the record of a company for each unit of $ExplanatoryAttribute_{x_i}$.

$$Index_{x_i} = \frac{ExportAmount_{registry}}{ExplanatoryAttribute_{x_i}}$$

6. Modeling and Evaluation

For Data modeling we used *0xdatas H_2O software*[3] connected to *R* by *H_2O R package* [23].

H_2O is a Java Virtual Machine that is optimized for doing "in memory" processing of distributed, parallel machine learning algorithms on clusters. In this research we used just one node with 3 CPUs and 6 GB of memory allocated to H_2O.

6.1. Comparing Models

H_2O offers an array of machine learning algorithms. Deep Learning AutoEncoder [24] (encoding stage in Figure 2) and linear principal component analysis (PCA) [25] are the options available for reducing dimensionality.

To detect anomalies, Deep Learning AutoEncoder can handle this task through its decode stage (see details in Section 6.2) and, likewise, the results of dimensionality reduction obtained using the PCA method can be decoded in a deep network using only the decode stage. Thus, differences will be observed only in the coding phase. We investigated the performance differences for dimensionality reduction between the two models proposed.

In both models the same dimensionality reductions were applied. 819,990 records were processed corresponding to companies operating directly or indirectly in exports in the Brazilian market in 2014. The processing time using AutoEncoder was substantially lower, about 20 times faster.

These results are supported by Sakurada and Yairi [26]: PCA is computationally more expensive than AutoEncoder because it "basically requires to hold all the training samples". These authors demonstrates that AutoEncoders detect subtle anomalies which PCA fails to and they "can detect anomalies even with relatively high latent dimensions while linear PCA can not".

Another point in favor of AutoEncoders is its non-linear generalizability due to the presence of non-linear functions in both the encoder and in the decoder [24].

6.2. AutoEncoder

AutoEncoder proved to be the most appropriate method for anomaly detection task. It was much faster: PCA requires more computation power than AutoEncoder. PCA basically requires to hold all the training samples, which is also computationally

3. A Open Source Software for data analysis, Apache 2.0 licensed, available in http://www.h2o.ai/

expensive. AutoEncoders can detect subtle anomalies which linear PCA fails to detect and can avoid complex computation that PCA requires without degrading the quality of detecting performance.

According to Goodfellow et al. [24] AutoEncoders are neural networks that are trained to make copies of their entries in their outputs. Internally, they have a hidden layer h which is a *code* used to describe the input. These networks can be seen as consisting of two parts: An encoding function $h = f(x)$ and a decoding function $r = g(h)$ that produces the reconstruction. This architecture is presented in Figure 2. However AutoEncoders should not learn to copy perfectly, otherwise they would be useless. Restrictions in the inner layers (hidden layers) network allow such copying is only an approximation. This ultimately forces the AutoEncoder network to prioritize the most important aspects to make the copy. Thus, most often it learns the most useful properties of the data.

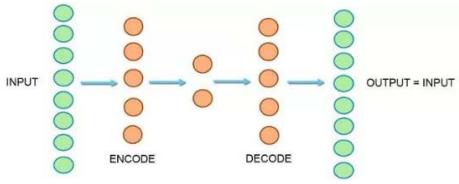

Fig. 2. Layers in a AutoEncoder network

Anomaly detection using dimensionality reduction is based on the assumption that the data has variables correlated with each other and that can be embedded into a lower dimensional subspace in which normal samples and anomalous samples appear significantly different [25].

In this work, we use 18 neurons (predictive attributes) as input layer and the same 18 neurons (predictive attributes) as the output layer. The goal here is that the network learn to copy input data to the output.

As hidden layers we used one with 6 neurons, one with 3 neurons and one with 6 neurons. So, the middle layer is a 3-dimensional representation of an 18-dimensional input. The objective here was to force the network to gradually reduce the dimensionality of the input data into a format in 3 dimensions. This prevents the learning to perfectly copy the entry, as the network will have to deal with a learning process in a few dimensions. The choice of the hidden layer size with 6-3-6 was made after various tests and graphical analysis of the middle layer. It is possible (and probable) that other combinations of hidden layers would reach similar results.

Figure 3 shows a graphical representation of the middle layer. We separate by color the twenty most anomalous records, i.e. twenty records in which the network had more difficulty to create a copy. These records will be those that we consider more likely to be suspected of fraud. This graphical view also allows us to realize that the middle layer was able to create a linear separation of records, focusing on the right part of the graphics the vast majority of records (corresponding to the records where there is a pattern of behavior) and in the left, more dispersed, anomalous records considered suspects.

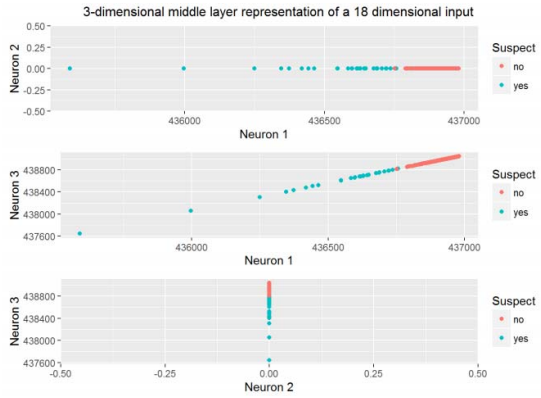

Fig. 3. 18 dimensional input in 3-dimensional representation of middle layer

The adjustment of the amount of epochs[4] was done by trial and error. A very small number could greatly decrease the network sensitivity. A large number tends to overfitting. The epochs were adjusted to 50 and the activation function used was ReLU ("Rectifier" in H_2O).

All other parameters were left at default values (per-weight adaptive learning rate, no L1/L2 regularization, no Dropout). Attempted settings of these parameters, despite having effects on the ability to learn to copy the data and thus influence the value of errors when comparing the input and output of the network, did not change the order of found anomalous records. Thus, we opted for the simplest model, namely the maintenance of defaults parameters.

6.2.1. Performance Analysis. We proceeded tests to verify the performance gains using different amounts of processors. These tests are intended to serve as reference of computational power needed for future works which involve the same database, but with greater granularity.

We conducted performance tests (on one cluster) varying the amount of processors to anomaly detection task with AutoEncoder. Four tests were conducted in a *Linux Ubuntu 16.04 LTS*: 1, 2, 3 and 4 allocated processors and 12GB of ram memory. Was used a *Intel Core i5-3317U CPU @ 1.70GHz* $\times 4$. Table 2 shows the results obtained.

TABLE 2. Performance Tests - varying the amount of processors to anomaly detection task with AutoEncoder

Number of allocated processors	Performance in milliseconds
1	54785
2	50512
3	49211
4	49001

4. Number of epochs represents "how many times the dataset should be iterated (streamed)" [23]

6.3. Evaluation

Once the model was trained we used the mean squared error (MSE) as a measure of how distant our predictions were from the real data. MSE measures the average of the squares of the errors, that is, the difference between the estimator and what is estimated. In this case, consider x_i the value of n neurons in the input layer and $\widehat{x_i}$ the value of n neurons in the output layer. The MSE value for each record containing n attributes of an exporting company is given by the formula below:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x}_i)^2$$

The higher the MSE value, the more anomalous, in relation to the pattern found in the data, a particular record is.

The MSE values are placed in ascending order and the distribution of the 170 highest values shown in Figure 4 indicate a clear change in behavior around the 20 last records.

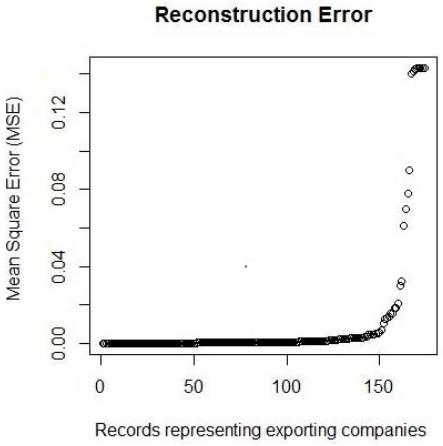

Fig. 4. The one hundred and seventy largest MSE values.

In order to carry out the evaluation of the records relating to major anomalies found, the attributes of the fifty companies that presented the highest MSE were presented to third party experts in exports fraud. Preliminarily, they considered the system as efficient, since it identified some fraud cases already known by the experts. The remaining cases will be evaluated for a conclusive opinion on the effectiveness of the model.

7. Conclusion and Future Works

This paper presented an unsupervised model for detecting fraud suspects in exports. Using the *0xdatas H_2O software* connected to *R* by *H_2O R package*, the performance of two-dimensionality reduction models were evaluated under the same conditions. The tests showed a performance to reduce dimensionalities about 20 times faster using Deep Learning AutoEncoder compared with PCA. The choice of AutoEncoder algorithm is supported by previous studies that indicate the detection of anomalies is more accurate and have a better power nonlinear generalization. *0xdatas H_2O software* provides other methods of analysis unsupervised but with linear linear approach. These methods can be tested in the future for comparison with this work.

The greatest difficulty in the use of unsupervised techniques is the evaluation of the results against the business objectives to be achieved. The evaluation of third party experts is subjective and therefore can be devoid of factors perceived by the data mining algorithm. In this work, the selection of suspected cases of fraudulent exports through unsupervised Deep Learning proved to be preliminarily promising, but a more thorough assessment should be made by experts. The in-depth investigation of cases identified is not trivial and takes time. Their conclusions will be disclosed in due course. Depending on the results, adjustments in the number of hidden layers and the number of neurons may prove necessary and lead to better results. Similarly, the decrease in the number of epochs may reduce a possible overfitting that has allowed even records with indications of fraud to have a low value of MSE.

Acknowledgments

The authors would like to thank the tax auditors Leon Solon da Silva, Marcelo Renato Lingerfelt and Nildomar Jose Medeiros for their help and support in making this work possible.

References

[1] Grupo de Egmont, *100 Casos de Lavagem de Dinheiro*. COAF, 2001. [Online]. Available: http://www.coaf.fazenda.gov.br/menu/pld-ft/publicacoes/100_Casos.pdf

[2] Conselho de Controle de Atividades Financeiras, *Casos e Casos - I Coletnea de Casos Brasileiros de Lavagem de Dinheiro*. COAF, 2011. [Online]. Available: www.coaf.fazenda.gov.br

[3] P. He, "A typological study on money laundering," *Journal of Money Laundering Control*, vol. 13, no. 1, pp. 15–32, Jan. 2010. [Online]. Available: http://www-emeraldinsight-com.ez54.periodicos.capes.gov.br/doi/full/10.1108/13685201011010182

[4] J. Madinger, *Money Laundering: A Guide for Criminal Investigators, Third Edition*. CRC Press, Dec. 2011.

[5] O. Greene, "Trade-Based Money Laundering," Jul. 2015, acesso em: 01/05/2016. [Online]. Available: https://www.dhgllp.com/Portals/4/ResourceMedia/publications/Risk-Advisory-Trade-Based-Money-Laundering.pdf

[6] Brasil, "Lei 9613, de 03 de maro de 1998." [Online]. Available: http://www.planalto.gov.br/ccivil_03/LEIS/L9613.htm

[7] T. E. Senator, H. G. Goldberg, J. Wooton, M. A. Cottini, A. F. Umar Khan, C. D. Klinger, W. M. Llamas, M. P. Marrone, and R. W. H. Wong, "The financial crimes enforcement network AI system (FAIS) : identifying potential money laundering from reports of large cash transactions," *The AI magazine*, vol. 16, no. 4, pp. 21–39, 1995. [Online]. Available: http://cat.inist.fr/?aModele=afficheN&cpsidt=2985240

[8] H. G. Goldberg and T. E. Senator, "Restructuring Databases for Knowledge Discovery by Consolidation and Link Formation," in *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 1995, pp. 136–141.

[9] A. S. Larik and S. Haider, "Clustering based Anomalous Transaction Reporting," *Procedia Computer Science*, vol. 3, pp. 606–610, 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S187705091000476X

[10] N. S. Khan, A. S. Larik, Q. Rajput, and S. Haider, "A Bayesian Approach for Suspicious Financial Activity Reporting," *International Journal of Computers and Applications*, vol. 35, no. 4, pp. 181–187, Jan. 2013. [Online]. Available: http://www.tandfonline.com/doi/abs/10.2316/Journal.202.2013.4.202-3864

[11] S. Raza and S. Haider, "Suspicious activity reporting using dynamic bayesian networks," *Procedia Computer Science*, vol. 3, pp. 987–991, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050910005375

[12] Q. Rajput, N. S. Khan, A. Larik, and S. Haider, "Ontology Based Expert-System for Suspicious Transactions Detection," *Computer and Information Science*, vol. 7, no. 1, Jan. 2014. [Online]. Available: http://www.ccsenet.org/journal/index.php/cis/article/view/30883

[13] A. Sharma and P. K. Panigrahi, "A Review of Financial Accounting Fraud Detection based on Data Mining Techniques," *International Journal of Computer Applications*, vol. 39, no. 1, pp. 37–47, Feb. 2012, arXiv: 1309.3944. [Online]. Available: http://arxiv.org/abs/1309.3944

[14] J. J. Filho, "Tratamento Bayesiano de Interaes entre atributos de Alta Cardinalidade," Ph.D. dissertation, Unicamp, Sep. 2007. [Online]. Available: http://www.bibliotecadigital.unicamp.br/document/?code=vtls000426153&print=y

[15] J. J. Filho and J. Wainer, "Using a Hierarchical Bayesian Model to Handle High Cardinality Attributes with Relevant Interactions in a Classification Problem," in *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, ser. IJCAI'07. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007, pp. 2504–2509. [Online]. Available: http://dl.acm.org/citation.cfm?id=1625275.1625679

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper With Convolutions," 2015, pp. 1–9. [Online]. Available: http://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html

[17] S. Jaiswal and M. F. Valstar, "Deep learning the dynamic appearance and shape of facial action units," Lake Placid, USA, 2016. [Online]. Available: http://eprints.nottingham.ac.uk/31301/

[18] Y. Liang, M. C. Machado, E. Talvitie, and M. Bowling, "State of the Art Control of Atari Games Using Shallow Reinforcement Learning," *arXiv:1512.01563 [cs]*, Dec. 2015, arXiv: 1512.01563. [Online]. Available: http://arxiv.org/abs/1512.01563

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature14539

[20] D. P. Kingma, S. Mohamed, D. Jimenez Rezende, and M. Welling, "Semi-supervised Learning with Deep Generative Models," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3581–3589. [Online]. Available: http://papers.nips.cc/paper/5352-semi-supervised-learning-with-deep-generative-models.pdf

[21] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, *CRISP-DM 1.0 Step-by-step data mining guide*. IBM, Aug. 2000. [Online]. Available: ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf

[22] Receita Federal do Brasil, "Portaria RFB n 671, de 07 de fevereiro de 2014."

[23] S. Aiello, T. K. a. P. Maj, and w. c. f. t. H. a. team, "h2o: R Interface for H2o," Jun. 2016. [Online]. Available: https://cran.r-project.org/web/packages/h2o/index.html

[24] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," 2016, book in preparation for MIT Press. [Online]. Available: http://www.deeplearningbook.org

[25] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, Jul. 2009. [Online]. Available: http://doi.acm.org/10.1145/1541880.1541882

[26] M. Sakurada and T. Yairi, "Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction," in *Proceedings of the MLSDA 2014 2Nd Workshop on Machine Learning for Sensory Data Analysis*, ser. MLSDA'14. New York, NY, USA: ACM, 2014, pp. 4:4–4:11. [Online]. Available: http://doi.acm.org/10.1145/2689746.2689747