

Inferring about fraudulent collusion risk on Brazilian public works contracts in official texts using a Bi-LSTM approach

Marcos Cavalcanti Lima **Leonardo R. de Carvalho** **Felipe L. de Souza Mendes**
Dept. of Computer Science Dept. of Computer Science Dept. of Computer Science
University of Brasília (UnB) University of Brasília (UnB) University of Brasília (UnB)
Brasília - DF - Brazil Brasília - DF - Brazil Brasília - DF - Brazil
marcosc.mcl@pf.gov.br leouesb@gmail.com felipelopes@gmail.com

Roberta da Costa Silva **Aleteia Araujo** **Flávio de Barros Vidal**
Dept. of Computer Science Dept. of Computer Science Dept. of Computer Science
University of Brasília (UnB) University of Brasília (UnB) University of Brasília (UnB)
Brasília - DF - Brazil Brasília - DF - Brazil Brasília - DF - Brazil
robertacs42@gmail.com aleteia@unb.br fbvidal@unb.br

Abstract

Public works procurements move US\$ 10 billion yearly in Brazil and are a preferred field for collusion and fraud. Federal Police and audit agencies investigate collusion (bid-rigging), over-pricing, and delivery fraud in this field and efforts have been employed to early detect fraud and collusion on public works procurements. The current automatic methods of fraud detection use structured data to classification and usually do not involve annotated data. The use of NLP for this kind of application is rare. Our work introduces a new dataset formed by public procurement calls available on Brazilian official journal (*Diário Oficial da União*), using by 15,132,968 textual entries of which 1,907 are annotated risky entries. Both bottleneck deep neural network and Bi-LSTM shown competitive compared with classical classifiers and achieved better precision (93.0% and 92.4%, respectively), which signs improvements in a criminal fraud investigation.

1 Introduction

In the last five years, Brazil's federal government invested (Brazil, 2020b) in 23,352 public works contracts adding up to R\$ 283.8 billion (approx. US\$ 49.3 billion in May 2020). Those works consist of all sorts of projects from oil refineries to ports, from soccer stadiums to power plants, from tunnels to dams, and are developed on a continental-sized territory, generating an endless and growing quantity of information regarding those projects. Thereupon, public works procurements are a preferred field for collusion and fraud (OCDE, 2007).

Brazilian Federal Police have been working on fraud investigations on public works for the last four decades and develop its investigation based on a highly skilled group of experts formed by civil, electrical, mechanical, computer engineers, and accountants (APCF, 2020). The types of fraud investigated are mainly collusion (bid-rigging), over-pricing, and delivery fraud (quality and quantity of services and materials). We will bring the knowledge accrued during those decades to enhance our data understanding (Lopes, 2011).

As described in Foundation (2020), public works contracts are made via procurement, that is the process of public administration uses to make all its contracts. Every procurement step is usually publicized by a call for application, and any interested people (or enterprises) around the world can obtain data from all available government journals of Brazil (the prominent public information journal in Brazil is the *Diário Oficial da União - DOU*). Despite being easy to access, tables, texts and documents do not bring any other annotated data for classification, even less for fraud detection. Those types of datasets have been studied, for example, by named-entity recognition in (Alvarez-Rodríguez et al., 2015) and linking open data, as in (Alvarez et al., 2011).

On the other hand, detecting and proving fraud on construction procurements is a laborious task, consuming around one month of forensic expert work per procurement/contract. Furthermore, it is essential to detect and combat fraud since a procurement first step because, as observed (Signor et al., 2019; Lopes, 2015), over-pricing is hardly obtained without collusion as most prices are set

during procurement. So, it has been the object of many studies (Kawai and Nakabayashi, 2014; Signor et al., 2019; Anysz et al., 2019; Sun and Sales, 2018; Vallim, 2020), but none of them used unstructured data to produce its goals.

Based on these presented statements, this work is focused on present a new dataset with textual information extracted from Brazilian Public government journals with an annotated *ground-truth* by forensics experts. It is, also, included in this work an initial classification methodology using a Bi-LSTM model and all early results are compared with main *state of the art* techniques.

The manuscript is organized as follows: In Section 2 are presented the related works about fraudulent collusion on public works contracts in official texts. Section 3 explains our proposed methodology. Section 4 contains information about the results and discussion. Section 5 provides conclusion points and introduce further works.

2 Related Works

It will be presented two parts: the first one, about fraud detection efforts, and the second one, about NLP classification.

Brazilian Federal Police has aspiring to improve its fraud detection mechanisms. Therefore, as presented in Vallim (2020) made a CBR model of paving services in the Paraná State approaching paving works contracts, which are one of the most budget consuming services in a state or city level and focus of criminal activities. This model used procurement type, enterprises, contract, and georeferenced data, and aimed to classify collusion cases, all of them based in a manual approach.

Another way to prove and identify public procurement collusion is by the use of statistics and probability. Those methods were explored on several Federal Police's studies and were based on joint behavior analysis of competitors who act to achieve bid-rigging. It was successfully applied to oil-related contracts using Operation Car Wash¹ information (Signor et al., 2019, 2020a) and for infrastructure projects (Signor et al., 2020b) with capped first-price auctions.

Brazilian Comptroller General of the Union (CGU), a national auditing public agency, also has several initiatives to reach a reliable classifier

¹Operation Car Wash is an ongoing nationwide corruption investigation led by Brazilian Federal Police, and it is focused on Petrobrás procurements. It is called "the biggest corruption scandal in history" (Watts, 2017).

for public procurement fraud. In Ralha and Silva (2012) was elaborated a unsupervised evaluator that, using *a priori* rules, evaluate the possibility of a certain group winning a given tender. They used structured data to bring suspect groups to be evaluated by experts. The work developed in Balaniuk et al. (2012) focused on the evaluation of fraud risk in government agencies using a Naïve Bayes Classifier for audit planning by the use of structured data and patterns of fraudulent activity. Sun and Sales (2018) used traditional neural networks and deep neural networks (DNN) to elaborate an early alarm system. The CGU studies usually have as features and fraud indicators: the number of bids, estimated cost and price relations, relations between public and private parts, political links of political parties, etc. Carvalho and Carvalho (2016) achieved good results using Bayesian Models with structured data from penalties database. They used data enrolled from the federal civil servants, servants' roles and income, number of accounts judged irregular and number of regularity certificates on an agency unit, and affiliated servants of each management unit.

Anysz et al. (2019) uses ANN and structured data on Poland's highways public procurements. They used the number of enterprises, price differences, contract order in the same place, and set of propositions to assess its fraud risk.

Works using TF-IDF in procurement documentation, as presented by Rabuzin and Modrušan (2019), tested Logistic Regression, SVM and Naïve Bayes on potential corruption. Their model had no annotated data, so it was focused in finding one bid tenders which "could be potentially suspicious."

Natural Language Processing is not often used to classify public procurement documents for risk or fraud. The technology is used for assessing fraud risk in health care claims (Popowich, 2005; Van Arkel et al., 2013) and financial reports (Seemakurthi et al., 2015; Goel and Uzuner, 2016). Public works publications data are not uniform enough to be structured, and, even if it is possible, it would be extremely laborious and it might be done at the cost of losing some unknown or undetected features.

All these studies suggest that fraud or risk public procurement classification has been developed based mostly on structured data and the use of NLP for this specific kind of classification is rare or nonexistent.

Regarding NLP classification methods, Braz

et al. (2018) proposed a Bi-LSTM based model to classify Brazilian supreme court documents as part of VICTOR project. de Araujo et al. (2020) made a confrontation of SVM and UMLFiT with bag-of-words features to classify the source of all calls on Brazilian Distrito Federal’s official journal and conclude that SVM was still competitive with more modern methods.

Kowsari et al. (2019) identified a wide use of TF-IDF as feature for text classification and, as architecture, the use of deep, convolutional and deep belief neural networks, and Bi-LSTM. As good examples, Chen et al. (2018) used a DNN model with a 2D TF-IDF feature to classify Twitter comments concerning cyberbullying and hate speech. Finally, Chen et al. (2017) classified costumers reviews using a Bi-LSTM network followed by a 1D CNN with word embedding features.

3 Proposed Methodology

The proposed methodology is presented in Figure 1 and detailed on along with the next subsections. The workflow presented in Figure 1 is formed by two actors: Dataset and Classification, respectively.

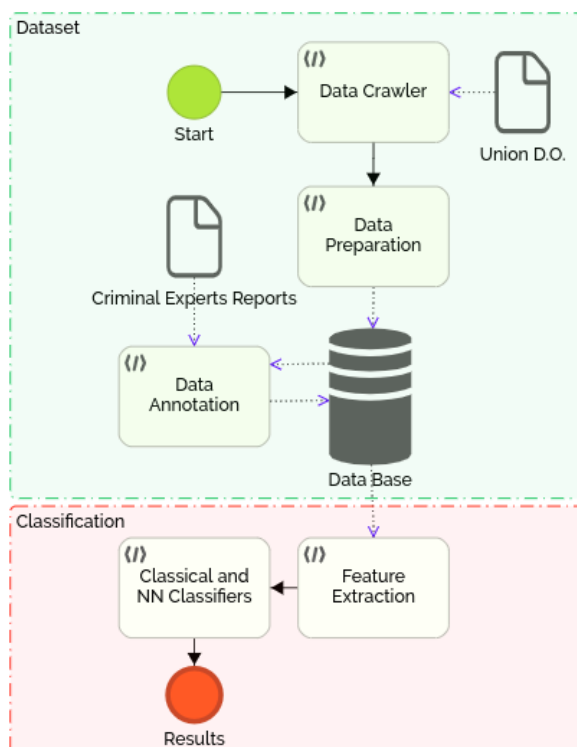


Figure 1: Workflow of the proposed methodology.

3.1 The Dataset Building Stage

The proposed dataset is a big set of text fragments extracted from DOU. All public procurement processed by the public administration uses this journal to make all its contracts. The Brazilian’s laws no. 8666/1993 and no. 10520/2002 (Brasil, 1993, 2002) oblige all agencies to follow a strict set of rules for any agreement, and it is even more detailed for construction projects. By force of those Brazilian’s laws, and by the constitutional publicity principle in this country, the main steps of the procurement process must be published as calls of application on the official media. Consequently, this data is accessible and reliable to serve as the main source of our proposed database.

Another indispensable characteristic of data must be its completeness (Weidema et al., 1996) and this is achieved knowing that, on Brazil, DOU Brazil (2020a) is a journal where all federal acts are publicized, it is divided into three sections and we can find on the third section: procurement calls, public tenders, contracts and it’s addenda, public agreements, etc. There are several ways to obtain data of public procurements, as open tender systems, transparency portals. Still, in the field of Brazilian public works, they are spread by many sites and agencies and available in different formats, tables, documents, and detail levels amid the three levels of Brazilian State: federal, state, and cities.

Diário Oficial’s publications, despite its relatively low level of detail, are very consistent and bring all vital information about the procurements as value, type, location, parts, object, etc.. They list without exception all public works of federal administration or with its direct financial support. This information raises the data reliability (Weidema et al., 1996) to be used by the criminal investigation and academic research.

3.1.1 Data Statistics

The database was obtained by a crawler algorithm developed by Ferreira (2018). It was applied to public data accessible at the site of Brazilian government enterprise for official publication as defined as *Imprensa Nacional* (Brazil, 2020a). A register sample of the dataset available for public download is shown in Figure 2.

Thus to form the database, the third section was downloaded from January 1998 to January 2018 into a PDF format and converted to a text format.



Figure 2: Sample register of the dataset.

Since February 2018, its publications are available on XML format and organized with a field for a public agency, type of document, etc. but without a specific field for procurement data, e.g. type of tender, deadlines, values, and scope. The dataset received information up to the last workday of February 2020 in a total of 15,132,968 entries. The Table 1 shows the dataset’s primary data statistics.

Table 1: Dataset Statistics.

Data	
Entries	15,132,968
Max. length	1,040,513 characters or 156,703 words
Mean length	761.0 char. or 110.2 words

Publications were organized after the crawling stage on JSON files that include sequential identification, date, and raw text. After February 2018, tables are shown in text with HTML formatting. The length of the text field varies from a dozen to thousands of characters with titles of subsections and signatures names being the shortest and public tenders with names lists the longest. Although it is not a structured text, it maintains individual traces

of order as it follows a formal and direct way of communication.

3.1.2 Data Annotation

The annotation of the public data was made with the use of a knowledge network of Brazilian forensic experts and do not represent an official assessment upon any person or public and private entity. The procurements, contracts, work, and/or agreements were annotated as having a fraud risk based on expert analysis. In their analysis its considered multiple indicators about date, place, type, parts (agency, enterprises), value, prices, execution, relation with other publications, and any other information linked to the process. So, it can not be concluded by a publication presence in this database about its legal or criminal status. Thereupon, we do not assess them as suspicion of fraud or not, but as a risk of fraud, and so publications were marked as having $risk = 1$.

All other publications included in the proposed dataset were marked, at first, as having $risk = 0$. Despite that, a procurement process is never said to be risk-free or fraud-free due to the nature of the criminal investigation (or an audit process). A total of 1,907 publications were marked as a risk of fraud, representing 0.012% of the dataset. The annotated data covers publications related to construction projects varying thousands to hundreds of million dollars in all Brazilian States.

As expected, the proposed dataset is very unbalanced due to the time demanded of an expert to fully analyze a public work procurement process. This results in a rate of 7,935 not annotated to 1 labeled as risk 1.

Instructions for downloading Deep-Vacuity dataset, go to URL: <http://www.cic.unb.br/~fbvidal/deepvacuity/dataset/index.html>.

3.2 The Classification Stage

The classification stage of this dataset tries to emulate a criminal expert assessment about the possibility of fraud in a given procurement. Experts, interviewed by the authors, said that value, agency, enterprises, location, date, type of construction and the correlation between that information usually lead to a good guess about the procurement risk of fraud. These gathered variables are the structured data models described in Section 2.

3.2.1 Training and Testing Subsets

To deal with the imbalanced dataset and successfully generalize the models, we created ten subsets with randomly chosen 1,907 entries of the not annotated publications to balance with annotated data. One way to classify the data is to consider all randomly chosen publications as having *risk* = 0. Although there is an error in that assumption, it can be assumed as low as the rarity of *risk* 1 class. On top of that, the ten created subsets were divided into a training subset (with validation) and a test subset in ten-fold cross-validation archiving 100 training sets. Figure 3 illustrates how it was fulfilled.

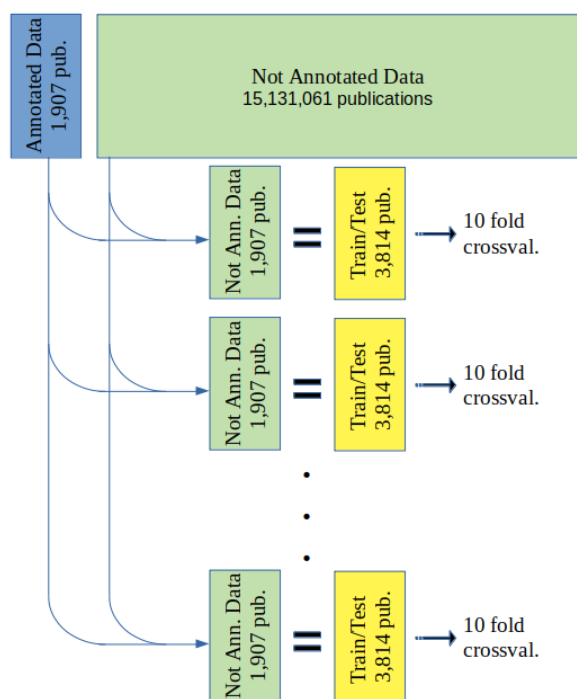


Figure 3: Training and Testing Subsets.

3.2.2 Comparison with Sparse Linear Classifiers

To create a baseline for comparisons, a wide range of classical linear supervised classifiers using sparse features was performed on the dataset, modeled using a Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction, that including the methods as follows:

- Stochastic Gradient Descent with Elastic Net with L1 Penalty (Zadrozny and Elkan, 2002; Zhang et al., 2002)
- Stochastic Gradient Descent with Elastic Net with L2 Penalty (Zadrozny and Elkan, 2002; Zhang et al., 2002)

- Linear SVC with L1-based feature selection (Fan et al., 2008)
- Naïve: Bernoulli, Complement and Multinomial (Manning et al., 2008; Rennie et al., 2003)
- Nearest Centroid (Tibshirani et al., 2002)
- Passive-Aggressive (Crammer et al., 2006)
- Perceptron (Freund and Schapire, 1999)
- Random Forest (Breiman, 2001)
- Ridge Classifier (Rifkin and Lippert, 2007)
- kNN with 10 neighbors (Altman, 1992)

The implementation of the Passive-Aggressive classifier method (Crammer et al., 2006) describes it as an online algorithm signifying that, for each prediction outcome for an instance of a sequential observation, the prediction mechanism is adjusted based on its correctness. A parameter of regularization controls this adjustment for the Passive-Aggressive method. Similarly, the Elastic Net penalty is one of the regularization parameters for the SGD method (Bottou, 2010). Figure 4 presents different F1 scores for the SDG using the parameters L1 and L2, being the Elastic Net, a compound of them, as defined by Zou and Hastie (2005). Those were the classifier methods with a higher F1 score (see Section 4). Benczúr et al. (2018) recommend an online learning strategy for scenarios with large data streams, because this sort of learning is based on each new event and its patterns.

3.2.3 Deep Neural Network Models

A current obstacle in science is to achieve good results using complex deep neural networks with few annotated data, but it is already a major advance in procurement fraud investigation, once deep neural networks use a supervised learning method.

For the feature extraction of the text, the TF-IDF approach was already used with linear classifiers and showed promising initial results. It was tested and implemented with deep neural networks, and initial tests showed better results. The vocabulary size was around 32 thousand words. The deep neural network models were build using the Tensorflow (Abadi et al., 2015). To evaluate the proposed dataset was chosen the Deep Neural Networks (DNN) and Bidirectional Neural Networks

models, all of them presented in Section 2 previously. Two DNN models were developed, the first model is defined as: $D(512, 0.4, \text{relu}) \rightarrow D(2048, 0.4, \text{relu}) \rightarrow D(4096, 0.4, \text{relu}) \rightarrow D(8192, 0.4, \text{relu}) \rightarrow D(4096, 0.4, \text{relu}) \rightarrow D(2048, 0.4, \text{relu}) \rightarrow D(512, 0.4, \text{relu}) \rightarrow D(256, 0.4, \text{relu}) \rightarrow D(1, 0, \text{sigmoid})$, where $D(u, dr, a)$ indicates a dense layer with u nodes on output space with dr dropout rate and a as described as the activation function. The second is a bottleneck model defined as $D(8192, 0.4, \text{relu}) \rightarrow D(4096, 0.4, \text{relu}) \rightarrow D(2048, 0.4, \text{relu}) \rightarrow D(1024, 0.4, \text{relu}) \rightarrow D(256, 0.4, \text{relu}) \rightarrow D(1024, 0.4, \text{relu}) \rightarrow D(2048, 0.4, \text{relu}) \rightarrow D(4096, 0.4, \text{relu}) \rightarrow D(8192, 0.4, \text{relu}) \rightarrow D(256, 0.4, \text{relu}) \rightarrow D(1, 0, \text{sigmoid})$.

Many shapes and parameters of the **Bi-LSTM** model (Schuster and Paliwal, 1997) network were tested, but the one with better results was defined as: $BL(192, 0.25) \rightarrow BL(128, 0.25) \rightarrow BL(96, 0.25) \rightarrow TD(128, 0, \text{relu}) \rightarrow FL \rightarrow D(1024, 0, \text{relu}) \rightarrow D(2, 0, \text{relu})$, where $BL(u, dr)$ is a Bi-LSTM layer with u and dr as defined above, TD is a time distributed dense layer similar to D , and FL is a layer that flattens the input. Outputs of the forward and backward RNNs were combined by concatenation in a length of 256 words.

4 Results

From every hundred tests ran for each model described above, it was computed precision, recall, and F1 Score with respective standard deviation. Those results, shown in Table 2, were plotted in Figure 4. The zoomed rectangle indicates most of the linear classifiers in a relative small F1 Score range from 91.4% to 93.4%. It can be observed that neural networks trended to produce a more significant standard deviation.

The best precision was achieved by the bottleneck network with 92.8%, the best recall was achieved by Naïve Complement with 97.2%, but the best value for F1 Score was by Passive-Aggressive with 93.4%.

Although the smaller F1-Score, both neural networks classifiers had higher precision, in other words, they classify less false positives entries. And knowing that the final model can raise a flag for further investigations and, on a finite resources police, that 1% difference equaling to an average monthly 569 entries, it is preferable to increase precision

Table 2: Comparative Result of all trained models.

Method	Mean		
	Precision	Recall	F1 Score
D2V-SVM	90.1%	87.4%	88.7%
D2V-DNN	88.8%	85.1%	86.6%
GL.Bi-LSTM	91.0%	90.9%	90.9%
Bi-LSTM	91.4%	93.4%	92.4%
Bottleneck	92.8%	93.2%	93.0%
DNN	91.7%	93.2%	92.4%
Elastic-Net	90.2%	96.3%	93.2%
L1-penalty	89.6%	95.2%	92.3%
L2-penalty	90.1%	96.4%	93.1%
LinearSVC	88.6%	94.6%	91.4%
NaïveBern.	89.4%	96.6%	92.9%
NaïveCompl	86.2%	97.2%	91.3%
NaïveMulti	88.6%	96.3%	92.3%
NearestCen.	85.4%	90.4%	87.8%
Pass.-Agg.	90.7%	96.3%	93.4%
Perceptron	89.2%	95.4%	92.2%
Random-for.	88.6%	95.3%	91.8%
Ridge-Class.	89.7%	96.3%	92.9%
kNN	80.9%	95.5%	87.6%
DNNClass.	88.1%	90.6%	89.2%

and lower recall than the opposite.

The use of classical deep artificial neural networks classifiers proved that it is possible to use it on natural language processing applications to classify the procurement publications dataset and reach a reliable model to sort out risky procurements. Among the shelf feature extraction models, TF-IDF showed to be a better abstraction for the dataset and the better classical and neural networks classifiers obtained F1-Score results over 93%. Both bottleneck deep neural network and Bi-LSTM proved to be competitive with traditional classifiers and achieved better precision, which is more desirable (over recall) in a criminal fraud investigation.

Another contribution is towards in to use a basic set of LSTM models on a dataset without temporal correlation, as traditionally used by many other approaches. In the dataset used in our experiments, there is not any temporal correlation (or other features as dependency discourse parsing with word base, for example) among those collect data. Despite this feature, the used LSTM models achieved high performance, when compared with all baseline techniques. This issue opens new opportunities to possible developments using these models for solve situations when an application for text analy-

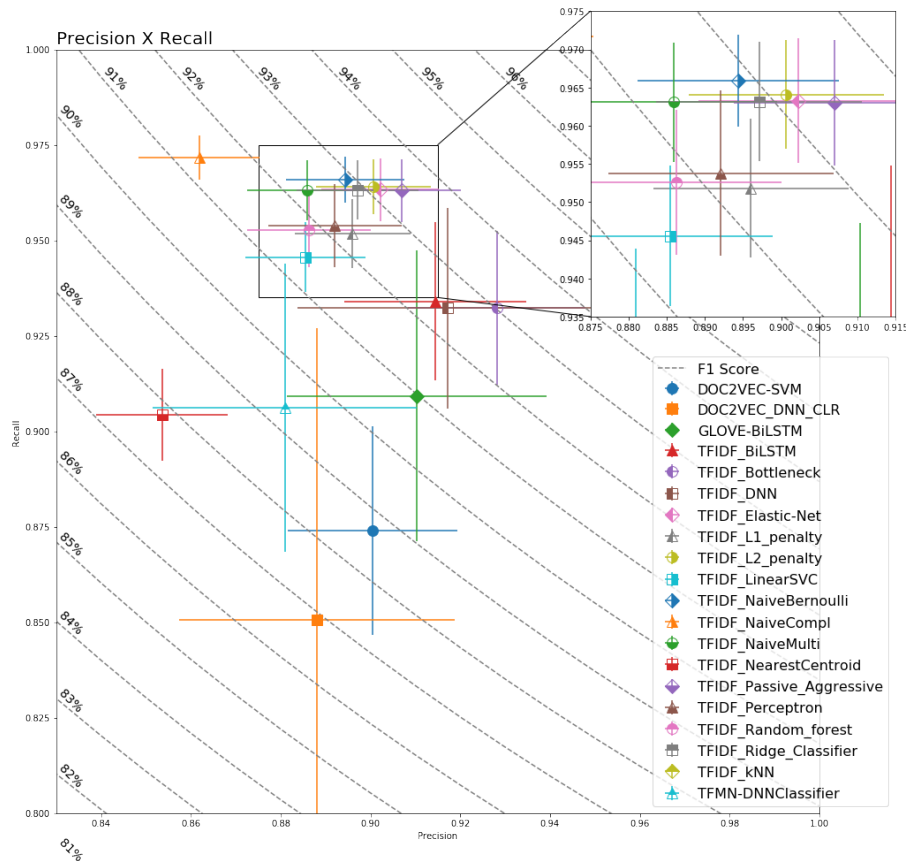


Figure 4: Precision, Recall and F1 Score of Models.

sis is achieved their limits (or bounded by boundary conditions), as a temporal correlation in the text fragment to be classified, for example.

5 Conclusions and Future Works

This work presented a preliminary evaluation of the DNN based on the LSTM model applied to fraudulent collusion detection from public works contracts in official texts. This proposed approach was compared with several state-of-the-art text classifiers (baseline classifiers), and the proposed method achieved competitive results for precision, recall, and F1 Score.

The proposed annotated dataset of the Brazilian public procurement calls allows researchers to explore a new form of fraud risk classification based on natural language processing and expert knowledge integration on its labeled data. The dataset covers already more than 22 years of publications, including more than 15 million entries, and it will be available for academic researches.

On the other hand, despite the full range of customization of neural networks, it is possible to achieve even better results. It will be studied in

the next works, new techniques to improve and customize feature extraction for the specific dataset and all LSTM models. For example, it is expected that Data Augmentation techniques should improve outcomes because of the small amount of annotated data available in the proposed dataset. Still, it is one of the many ways to achieve performance improvements in the classification process.

Acknowledgments

This work was supported by Brazilian Federal Police and University of Brasilia under the research project "Pesquisa Aplicada de Inovações Tecnológicas no domínio da Perícia Criminal Federal".

References

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon

- Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. [TensorFlow: Large-scale machine learning on heterogeneous systems](#). Software available from tensorflow.org.
- Naomi S Altman. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185.
- Jose María Alvarez, José Emilio Labra, Ángel Marín, and José Luis Marín. 2011. Semantic methods for reusing linking open data of the european public procurement notices. In *ESWC PhD Symposium*.
- Jose María Alvarez-Rodríguez, Michalis Vafopoulos, and Juan Llorens. 2015. [Enabling policy making processes by unifying and reconciling corporate names in public procurement data. the corfu technique](#). *Computer Standards & Interfaces*, 41:28 – 38.
- Hubert Anysz, Andrzej Foremny, and Janusz Kulejewski. 2019. Comparison of ann classifier to the neuro-fuzzy system for collusion detection in the tender procedures of road construction sector. In *IOP Conference Series: Materials Science and Engineering*, volume 471, page 112064. IOP Publishing.
- APCF. 2020. Perícia criminal. <http://apcf.org.br/pericia-criminal/pericia-criminal>. Visited: 2020-05-25.
- Pedro Henrique Luz de Araujo, Teófilo Emidio de Campos, and Marcelo Magalhães Silva de Sousa. 2020. Inferring the source of official texts: can svm beat ulmfit? In *International Conference on Computational Processing of the Portuguese Language*, pages 76–86. Springer.
- Remis Balaniuk, Pierre Bessiere, Emmanuel Mazer, and Paulo Cobbe. 2012. [Risk based Government Audit Planning using Naïve Bayes Classifiers](#). In *Advances in Knowledge-Based and Intelligent Information and Engineering Systems*, Spain.
- András A. Benczúr, Levente Kocsis, and Róbert Pálovics. 2018. [Online machine learning in big data streams](#).
- Leon Bottou. 2010. Stochastic gradient descent. <https://leon.bottou.org/projects/sgd>. Visited: 2020-05-24.
- Brasil. 1993. Lei n. 8.666, de 21 de junho de 1993. http://www.planalto.gov.br/ccivil_03/leis/18666cons.htm. Accessed: 2019-11-06.
- Brasil. 2002. Lei n. 10.520, de 17 de julho de 2002. http://www.planalto.gov.br/ccivil_03/LEIS/2002/L10520.htm. Accessed: 2019-11-06.
- Fabricio Ataiades Braz, Nilton Correia da Silva, Teófilo Emidio de Campos, Felipe Borges S Chaves, Marcelo HS Ferreira, Pedro Henrique Inazawa, Victor HD Coelho, Bernardo Pablo Sukiennik, Ana Paula Goncalves Soares de Almeida, Flavio Barros Vidal, et al. 2018. Document classification using a bi-lstm to unclog brazil’s supreme court. *arXiv preprint arXiv:1811.11569*.
- Brazil. 2020a. Imprensa nacional. <http://www.in.gov.br>. Visited: 2020-04-23.
- Brazil. 2020b. Painel de obras. <http://transferenciasabertas.planejamento.gov.br/>. Visited: 2020-05-19.
- Leo Breiman. 2001. Random forests. *Machine learning*, 45(1):5–32.
- Ricardo Silva Carvalho and Rommel Novaes Carvalho. 2016. Bayesian models to assess risk of corruption of federal management units. In *BMA@ UAI*, pages 28–35.
- Junyi Chen, Shankai Yan, and Ka-Chun Wong. 2018. Verbal aggression detection on twitter comments: Convolutional neural network for short-text sentiment analysis. *Neural Computing and Applications*, pages 1–10.
- Tao Chen, Ruifeng Xu, Yulan He, and Xuan Wang. 2017. Improving sentiment analysis via sentence type classification using bilstm-crf and cnn. *Expert Systems with Applications*, 72:221–230.
- Koby Crammer, Ofer Dekel, Joseph Keshet, Shai Shalev-Shwartz, and Yoram Singer. 2006. Online passive-aggressive algorithms. *Journal of Machine Learning Research*.
- Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. Liblinear: A library for large linear classification. *Journal of machine learning research*, 9(Aug):1871–1874.
- Hugo Honda Ferreira. 2018. *Processamento de Linguagem Natural e Classificação de Textos em Sistemas Modulares*. Monografia (Bacharel em Informática), Universidade de Brasília, Brasília, Brazil.
- Open Knowledge Foundation. 2020. Procurement. <http://index.okfn.org/dataset/procurement/>. Visited: 2020-05-19.
- Yoav Freund and Robert E Schapire. 1999. Large margin classification using the perceptron algorithm. *Machine learning*, 37(3):277–296.
- Sunita Goel and Ozlem Uzuner. 2016. Do sentiments matter in fraud detection? estimating semantic orientation of annual reports. *Intelligent Systems in Accounting, Finance and Management*, 23(3):215–239.

- Kei Kawai and Jun Nakabayashi. 2014. Detecting large-scale collusion in procurement auctions. *Available at SSRN 2467175*.
- Kamran Kowsari, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. 2019. Text classification algorithms: A survey. *Information*, 10(4):150.
- Alan de Oliveira Lopes. 2011. Superfaturamento de obras públicas. *Livro Pronto, São Paulo, Brazil*.
- Alan de Oliveira Lopes. 2015. O efeito pedagógico de operações da polícia federal: Um estudo de caso da operação "caixa de pandora". *Revista Brasileira de Ciências Policiais*, 6(1):67–85.
- Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to information retrieval*. Cambridge university press.
- OCDE. 2007. Bribery in public procurement methods, actors and counter-measures.
- Fred Popowich. 2005. Using text mining and natural language processing for health care claims processing. *ACM SIGKDD Explorations Newsletter*, 7(1):59–66.
- Kornelije Rabuzin and Nikola Modrušan. 2019. Prediction of public procurement corruption indices using machine learning methods.
- Célia Ghedini Ralha and Carlos Vinícius Sarmiento Silva. 2012. [A multi-agent data mining system for cartel detection in brazilian government procurement](#). *Expert Systems with Applications*, 39(14):11642 – 11656.
- Jason D Rennie, Lawrence Shih, Jaime Teevan, and David R Karger. 2003. Tackling the poor assumptions of naive bayes text classifiers. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 616–623.
- Ryan M Rifkin and Ross A Lippert. 2007. Notes on regularized least squares.
- Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681.
- Prasad Seemakurthi, Shuhao Zhang, and Yibing Qi. 2015. Detection of fraudulent financial reports with machine learning techniques. In *2015 Systems and Information Engineering Design Symposium*, pages 358–361. IEEE.
- Regis Signor, Peter ED Love, Alexanders TN Belarmino, and Oluwole Alfred Olatunji. 2020a. Detection of collusive tenders in infrastructure projects: Learning from operation car wash. *Journal of Construction Engineering and Management*, 146(1):05019015.
- Regis Signor, Peter ED Love, Acir Oliveira Jr, Alan O Lopes, and Pedro S Oliveira Jr. 2020b. Public infrastructure procurement: Detecting collusion in capped first-priced auctions. *Journal of Infrastructure Systems*, 26(2):05020002.
- Regis Signor, Peter ED Love, João José CB Vallim, Alexandre B Raupp, and Oluwole Olatunji. 2019. It is not collusion unless you get caught: the case of operation car wash and unearthing of a cartel. *Journal of Antitrust Enforcement*.
- Ting Sun and Leonardo J. Sales. 2018. [Predicting Public Procurement Irregularity: An Application of Neural Networks](#). *Journal of Emerging Technologies in Accounting*, 15(1):141–154.
- Robert Tibshirani, Trevor Hastie, Balasubramanian Narasimhan, and Gilbert Chu. 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences*, 99(10):6567–6572.
- João José de Castro Baptista Vallim. 2020. [Uso do Modelo de Raciocínio Baseado em Casos Para Monitoramento de Conluio em Licitações de Obras de Pavimentação Urbana](#). Master's thesis, Universidade Federal do Paraná, Curitiba.
- John H Van Arkel, James J Wagner, Corrine L Schweyen, Saralyn M Mahone, Terrill J Curtis, Scott HAGINS, et al. 2013. Predictive modeling processes for healthcare fraud detection. US Patent App. 13/536,414.
- Jonathan Watts. 2017. Operation car wash: Is this the biggest corruption scandal in history. *The Guardian*, 1(06):2017.
- Weidema, Bo Pedersen, Wesnaes, and Marianne Suhr. 1996. Data quality management for life cycle inventories—an example of using data quality indicators. *Journal of cleaner production*, 4(3-4):167–174.
- Bianca Zadrozny and Charles Elkan. 2002. Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 694–699.
- Tong Zhang, Fred Damerau, and David Johnson. 2002. Text chunking based on a generalization of winnow. *Journal of Machine Learning Research*, 2(Mar):615–637.
- Hui Zou and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320.