ARTIFICIAL INTELLIGENCE ON LEGAL LANGUAGE PROCESSING: USING DEEP LEARNING TO FIND THE REGULATORY LAW FRAMEWORK FOR THE THIRD SECTOR

Mauricio Barros de Jesus[1]
André da Silva Goes[1]
Leonardo de Guimarães Santiago[1]
Marcelo Augusto Pedreira Xavier[1]
Sólon Bevilacqua[2]

[1]Tribunal de Contas do Estado de Goiás (TCE/GO), Goiânia – Goiás, Brazil
[2]Universidade Federal de Goiás (UFG), Goiânia – Goiás, Brazil

This paper deals with the application of artificial intelligence algorithms in processing legal language to identify a complete set of rules applicable to a given legal theme. In this study, we sought to delimit the regulatory framework that involves the Third Sector, based on the data set on the Brazilian regulation flow (RegBR). From the bibliographic research, machine learning techniques were applied to automate the classification of each sentence within the analyzed normative acts, allowing us to identify to what extent a norm applies to the selected topic. The BERT model with fine-tuning by a Brazilian legal dataset was highly effective, reaching 94% of precision (F1-Score and AUC). The results include a total found of 2,359 rules spread in 611 normative acts on the 1,330,190 sentences distributed in 51 thousand regulations contained in the dataset, demonstrating how the applied techniques can contribute to the improvement of the themes involved.

Keywords: third sector; regulation; deep learning; natural language processing; law.

DOI: https://doi.org/10.21874/rsp.v74i2.8091 ISSN: 0034-9240 | e-ISSN: 2357-8017
Submetido: 29 de setembro de 2022. Aceito: 13 de outubro de 2022.

INTELIGÊNCIA ARTIFICIAL NO PROCESSAMENTO DE LINGUAGEM JURÍDICA: APLICAÇÃO DE *DEEP LEARNING* PARA DEFINIÇÃO DO MARCO REGULATÓRIO DO TERCEIRO SETOR

O presente artigo trata da aplicação de algoritmos de inteligência artificial no processamento de linguagem jurídica, a fim de possibilitar a identificação de um conjunto completo de normas aplicável a uma determinada temática legal. Neste estudo, buscou-se delimitar o marco regulatório que envolve o Terceiro Setor, a partir do conjunto de dados sobre o fluxo regulatório brasileiro (RegBR). A partir de pesquisa bibliográfica, foram aplicadas técnicas de aprendizagem de máquina para automatizar a classificação de cada sentença contida nos atos normativos analisados, permitindo identificar em que medida uma norma se aplica ao tema selecionado. O modelo BERT com ajuste fino com trechos de leis brasileiras foi altamente eficaz, atingindo 94% de precisão (F1-Score e AUC). Como resultados, foram identificadas 2.359 regras espalhadas em 611 normas, extraídas entre 1.330.190 dispositivos legais distribuídos em 51 mil regulações, demonstrando que as técnicas aplicadas podem contribuir para o aperfeiçoamento das temáticas envolvidas.

Palavras-chave: terceiro setor; regulação; aprendizagem profunda; processamento de linguagem natural; direito.

INTELIGENCIA ARTIFICIAL EN PROCESAMIENTO DEL LENGUAJE EN EL DERECHO: APLICACIÓN DEL *DEEP LEARNING* PARA DEFINIR EL MARCO REGULATORIO DEL TERCER SECTOR

Este artículo trata sobre la aplicación de algoritmos de inteligencia artificial en el procesamiento del lenguaje jurídico para permitir la identificación de un conjunto completo de normas aplicables a un determinado tema. En este estudio, buscamos delimitar el marco regulatorio que involucra al Tercer Sector, a partir del conjunto de datos sobre el flujo regulatorio brasileño (RegBR). A partir de la investigación bibliográfica se aplicaron técnicas de aprendizaje automático para automatizar la clasificación de cada oración dentro de los actos normativos analizados, permitiendo identificar en qué medida aplica una norma al tema seleccionado. El modelo BERT perfeccionado con extractos de las leyes brasileñas fue muy eficaz y logró un 94 % de precison (F1-Score y AUC). Fue posible encontrar un total de 2.359 reglas esparcidas en 611 actos normativos, retiradas entre 1.330.190 sentencias distribuidas en 51 mil regulaciones, demostrando así cómo las técnicas aplicadas pueden contribuir a la mejora de los temas estudiados.

Palabras clave: tercer sector; regulación; aprendizaje profundo; procesamiento natural del lenguaje; derecho.

Acknowledgments

The authors would like to thank the Accounts Court of Goiás (TCE/GO), which under a partnership with the Federal University of Goiás (UFG) for Professional Master's in Public Administration (PROFIAP), made this research possible and kindly provided your datacenter with the GPU's, without which the results would take much longer.

1. INTRODUCTION

The modern state has social welfare as one of its elementary characteristics. Consequently, the Government's activity must ensure the provision of essential services of public interest, such as health and education. However, most liberal conceptions open a gap for individuals to act alongside the State in social utility functions, thus founding a new frontier for public policy. As this study intends to demonstrate, the so-called Third Sector (TS) brings together institutions that are not part of the State or the market. Because they do not belong to the public or private sectors, these non-profit civil society organizations are often subsidized by the government and become their partners in the execution of public policies.

Therefore, through normative acts, the government establishes a legal framework to regulate and enable these activities. However, in a country with a *Civil Law*[1] tradition, such as Brazil, the regulations are extensive and difficult to understand even for those who dedicate their studies to law. Consequently, artificial intelligence techniques, especially those focused on natural language processing, emerge to be tools to help human beings understand, select, and manage large volumes of textual information, such as extensive rules and normative acts.

It is in this context that the research problem of this work is included. The RegBR (https://infogov.enap.gov.br/regbr) is the first national tool for regulatory flow analysis, developed by the National School of Public Administration [Enap] (2022). This tool centralizes, classifies, and analyzes federal government norms, allowing us to understand the history of the creation of normative acts in Brazil in the last 60 years. Until March 2022, the database had 50,999 regulations related to important social aspects such as the COVID-19 pandemic; Labor Laws; the creation of micro and small enterprises, and the Brazilian Disarmament Statute.

The main objective of this paper is to answer the following question: How to categorize the records of the Brazilian federal government acts list contained in the RegBR dataset, the rules about the Third Sector, and identify those sentences that are pertinent to this theme?

When answering this question, this work aims to obtain the legal framework for the TS's regulation in Brazil. For that, it used machine learning algorithms in which the computer learns to classify and qualitatively analyze the text content based on data observations, using the neural network developed by Google for the academic community under the name of Bidirectional Encoder Representations from Transformers - BERT. In order for this, we studied regulatory theories in the literature about the TS to create artificial intelligence capable of helping to delimit the legal regulatory framework applicable to TS's organizations within the RegBR database.

In the first section, we approach the main theories of regulation and present the key aspects of the law that are essential to constructing norms related to the TS and the concepts

[1] Different of Common Law, it is a legal system of Roman-Germanic origin, where written law is observed as the main source of justice.

applicable to machine learning. Then, we detail the methodology adopted in this paper and discuss the results achieved, concluding with suggestions for future research.

2. BACKGROUND AND LITERATURE REVIEW

The growing participation of non-governmental organizations (NGOs) that make up the TS as public service providers and stakeholders in the legislative process has been considered the most critical change in regulation in the last three decades (Talesh, 2022). During this period, several government initiatives were launched to regulate the activities of NGOs and establish mechanisms to defend the probity and integrity of TS institutions (Crack, 2018).

Although regulation necessarily remains a public function, the role of TS entities in this process is more evident. It has become recurrent in academic research due to the economic relevance acquired and the importance given to the responsibility conferred by the provision of social services traditionally carried out by public institutions (Phillips & Hebb, 2010).

2.1 Regulation Theories

Regulation is a theme present in people's daily life. Its concept is related to a set of governmental laws, originating from the State or public organizations. This includes the application of rules that directly or indirectly influence the daily lives of citizens, companies, and ST organizations, and their monitoring of social, business, and political actors (Levi-Faur, 2011).

In Brazil, this theme took up many public debates. The 1995's Brazilian State management reform, idealized by Bresser-Pereira, highlighted the distinction between the exclusive activities of the State and social activities (Brasil, 1995). From then on, increasingly the Government has become a financier to the non-State organizations working in carrying out social policies.

Thus, regulation deals with the intervention in the behavior or activities of individual or corporate actors. Is it over there consisting of the license to drive, the definition of public policies, the permission for the construction or expansion of residential or commercial property, and the establishment of restrictions or barriers for certain economic sectors (Hahn & Tetlock, 2008; Koop & Lodge, 2017).

The expectation of the regulatory process is to solve real and relevant problems, such as the development of public policies, protection against financial fraud, prevention of work accidents, environmental issues, and public safety. On the other hand, it is certain that regulation demands counterparts and can generate unwanted effects. These consequences include increases in taxes and fees, increases in the value of goods and services, reduction of workers' salaries, limitation of social security benefits, and especially less privacy or mitigation. of personal freedom (Ellig, 2018; Saab & Silva, 2021).

As said, the command and control or top-down approach are characterized by a system of rules arising from the government's actions. This is, more precisely, from public institutions with the adoption of traditional instruments of law-making by state bodies, such as legislatures, courts, and administrative agencies. The regulation law included formal rules and stipulations, contradictory methods, and applicable means of dispute resolution and was the most widely studied theory in the last century unfolding in several theoretical lines (Aranha, 2014; Advocacia-Geral da União [AGU], 2018).

One of these currents discusses how interest groups participate directly in government processes. Constant influence groups are formed by businessmen, public servants, and politicians, and set the course for the enactment of laws and regulatory rules that meet their corporate interests. In this context, they seek to understand the role of business power as interest groups in the formation of coalitions to lobby in the negotiation of favorable rules, the construction or establishment of an agenda in its strategic favor, or through direct influence over government decision-makers (Talesh, 2022).

In turn, public choice theory brings together a more transactional view of the legislative process. It means to recognize that the regulatory decision-making process is a necessary product of the exchange of political advantages between elected representatives, interest groups, and regulatory agencies, through the application of economic models in decision-making in the legislative and administrative scope. From this perspective, politicians and voters are considered rational utility maximizers operating in a competitive electoral and legislative market, issuing rules that expand the benefits to legislators who, in turn, make them available as a governmental benefit to the group of voters (Aranha, 2014).

Also, in this area, regulatory capture studies the mechanisms and approaches that companies tend to have to influence legislation and regulatory policy, giving rise to the capture of specialized agencies and implying a conflict between private and public interests. It is considered that in this circumstance, regulation is captured by the industry to act for its benefit, leaving the regulators, as an institutional survival strategy, to meet private demands by instituting policies that are favorable to them (Aranha, 2019; Talesh, 2022).

The theories outlined above that study interest groups, public choice, or the capture of agencies try to explain the regulatory process in a vertical context where legislators and regulators try to coerce or, in some cases, encourage organizations to comply with edited norms, while organizations engage in rational and strategic choices about whether or not to comply with legislation and how to influence legal mandates. In this view, the law is exogenous (produced by the government) yet open to organizational influence (Levi-Faur, 2011; Talesh, 2022).

However, nowadays, regulation with a bias in command and control in which there is the possibility of capture no longer seems legitimate, forcing a turn to regulation focused on governance, with more emphasis on social and environmental policies through participatory management. stakeholders, the State, and other regulatory bodies (Dias, Farias, Coura, Demo, & Athayde, 2022).

This regulatory model, classified by Talesh (2022) as risk management or self-regulation, transfers the risk and responsibility to the actor responsible for the actions. It encourages private institutions or the TS to develop models of self-organization with a view to satisfactorily achieving better results translated into public goods or services and ensuring more effective compliance with applicable social, environmental, labor, and consumer standards.

The increase in involvement, delegation, and deference to non-state actors, therefore, was noticeable from the consolidation of regulation with a focus on governance. This is identified as the most important change in the regulation process in the last three decades, which permeates the creation of rules exclusively by governmental entities to gradually count on the involvement of non-governmental organizations, whether in the construction of the regulatory process or the execution of public policies (Phillips & Smith, 2014; AGU, 2018).

In this light, private actors and civil society are seen as resources and instruments for the formulation of co-produced public policies. They are instead of reducing them to passivity and merely subjects of public regulation, in a perspective of joint efforts to build education of a collective will be based on a diversity of interests, the definition of a system of rules that shape and regulate the actions of social and political actors; or the political direction of social and economic relations based on soft and cooperative policy instruments, such as persuasion, voluntary coordination, and public performance benchmarking procedures (Ansell & Torfing, 2016).

2.2 Third Sector Regulation

Regarding the regulation of the TS, we observe a concern with those who develop social activities in the education, health, and social assistance sectors. The first focus of the regulators is related to the social control and transparency of transactions and administrative acts by civil society organizations, being identified as one of the main requirements that must permeate the legislation affecting the sector (Abrucio, 2006; Montano & Pires, 2019; Phillips & Smith, 2014; Sano & Abrucio, 2008).

A second trend is that, in response to various scandals, regulators have become more focused on compliance and on strengthening institutional control mechanisms, largely due to serious situations identified in countries such as Canada in which philanthropic institutions were detected using false receipts to receive tax benefits. The situation also occurred in England, where The Charity Commission has been criticized for failing to take steps to prevent abuses of this nature or for failing to introduce rules to ensure that non-profit organizations do not have individuals with a history of fraud serving on their boards or senior management (Phillips & Smith, 2014).

In Brazil, the literature mentions several circumstances related to acts of corruption involving social organizations. This can be observed in the case of the municipality of Rio de

Janeiro/RJ, where deviations of more than R$ 700 million were found in the municipal health area and in Natal/RN, where that culminated in the conviction and imprisonment of directors of the third sector institution involved in other fraud (Vieira, 2016).

In this step, the accountability of TS entities has been defined in various ways by academics, NGOs, and standards-setting bodies over the years. The debates revolve around the definition of how the accountability process will take place. In general terms, the regulation on accountability should establish transparency requirements about the organization's structure, what it is committed to doing and the progress achieved; to open and promote communication between the interested parties, with the aim of continuous improvement (Crack, 2018).

Regulators also turned their view to the formal aspect, filling gaps and duplications in regulatory regimes. This happened with the review and expansion of criteria for the qualification of institutions and conferring criteria for the assessment of economic and financial sustainability about the ability to obtain alternative sources of funding, and establishment of the entity's existence and minimum experience to formalize partnerships with the government (Montano & Pires, 2019; Phillips & Smith, 2014).

Another way to increase legal compliance and ensure better self-regulation is through good governance. Regulators are, therefore, articulating more directly about the performance of non-state actors that constitutes good governance practice, demanding more information about these practices. It is common for the command to provide detailed information on several relevant operational issues, including executive compensation, board governance, fundraising costs, the establishment of a recognition policy, awarding seals or rankings of excellence, and rewarding good practices observed in TS institutions (Phillips & Smith, 2014).

An emerging trend in the regulatory aspect of the TS is the demand for information on the effectiveness of the activities carried out. An example of this is the rules that address this issue in the United Kingdom require proof of the impact of activities, through an annual report, in order to obtain a public benefit, such as funding or tax incentives.

Although such reports are still incipient in content, primarily due to the technical difficulty identified in small institutions, they are beginning to gain acceptance as a means of promoting accountability. With them emerges the possibility of defining a set of compromises between the parts, including the objectives and targets to be achieved over some time and the indicators to be used to assess the achievement of targets (Barbosa ***et al.***, 2015; Phillips & Smith, 2014).

Greater attention to results also encourages the government and the third sector to strengthen self-regulation. With that, TS can establish standards of ethical behavior and a code of conduct. An example is a Canadian regulation that instituted a complete certification process involving peer self-assessment and reviews with seventy-three standards to be observed in non-state entities in partnership with the government (Phillips & Smith, 2014).

Regarding financing, the literature points to expanding new forms of private financing, with the denomination of social investment instead of philanthropy, in a substitutive or complementary character to public financing. In terms of indirect financing through tax incentives, the practice is becoming more contested, largely due to aspects of fiscal restraint supported by spending cuts and restructuring of public services.

On the other hand, funding cuts are accompanied by government support and advice in developing alternative funding models. This can be observed in the creation of a sustainability fund that helps medium-sized institutions to become less dependent on government subsidies and open up new sources of funding (Mcmullin, 2021; Phillips & Smith, 2014).

As soon, the context studied and presented so far herein converged and supported the machine process learning, shown in detail in the next section, in an attempt to identify and classify among the 51,000 standards cataloged in RegBR those that, with a planned level of assertiveness, are relevant to the TS.

2.3 Machine learning

Machine learning is a subarea of artificial intelligence. It has evolved from studying mathematical pattern recognition in computer engineering science. Likewise, deep learning is a branch of human-based machine learning that attempts to model high-level data abstractions supported by computer artificial neural networks.

The main objective of the data mining process is to find patterns to predict behavior from the analysis of historical data. For this, it is possible to train a machine learning model in three ways: supervised, unsupervised, and reinforcement learning (Luger, 2013).

When we use supervised learning, there are pre-labeled classes in the training dataset. The label is used to train the algorithm and verify the classification accuracy obtained from the machine learning model by applying evaluation measures. On the other hand, in unsupervised learning, the algorithm itself must explain the results. In reinforcement learning, the model learns by exploring the possibilities of the environment in which it is inserted, seeking, by trial and error, which options have the highest reward or lowest penalty (Luger, 2013).

In addition, there are four groups of machine learning tasks: classification, regression, clustering, and association rule learning. Machine learning models learn to classify text based on observations of pre-labeled examples as training data, learning associations between texts and their labels (Minaee ***et al.***, 2020).

Classification tasks belong to the supervised learning group in which, from a set of previously labeled data, a mathematical model is trained to classify new information not labeled. Regression tasks do the same as classification tasks, but the result is a numeric value. Clustering tasks create groups of instances with common characteristics. In turn, association tasks identify rules that associate item sets (Amaral, 2016).

Classical natural language processing (NLP) tasks algorithms include Naïve Bayes, support vector machines (SVM), gradient boosting, and random forests. However, Minaee *et al.* (2020) describe several limitations of these models, such as the dependence of the pre-labeled dataset, which sometimes is small; more complex analysis to obtain good performance; the strong dependence on domain knowledge; and the limitation of generalizing to new tasks. Therefore, it is possible to conclude that years of deep neural approaches have been explored to address these limitations.

In this way, it proposed a recipe to decide which neural network model fits for specific NLP tasks. It depends on a lot of aspects, such as the nature of the target task and domain, the availability of in-domain labels, and the latency and capacity constraints of applications (Minaee *et al.*, 2020).

For this work, as will be discussed in the next section, we built a relatively small pre-labeled dataset, in a specific legal domain, with different document sizes and context between words is important. Thus, a pre-trained BERT model was chosen as a classifier.

Devlin, Chang, Lee, & Toutanova (2019) introduced the model Bi-directional Transformer Encoder Representations (BERT) for short, as pre-training language representation that achieves state-of-the-art results in a wide variety of NLP tasks.

They built the BERT´s architecture based on Vaswani *et al.* (2017) implementation of a multi-layer Bidirectional Transformer Encoder. The final model was run on NLP tasks and performed better than other state-of-the-art models, such as ELMO and OpenAI GPT (Devlin *et al.*, 2019).

The bases of the BERT model are pre-training and fine-tuning tasks. For pre-training, the model is trained with a massive unlabeled dataset from scratch. Since BERT uses transfer learning, it is possible to use a pre-trained model as the starting point. After that, the pre-trained model is loaded for fine-tuning, and all parameters are tuned using a specific dataset from the specific downstream tasks (Devlin *et al.*, 2019, Zheng *et al.*, 2021; Chalkidis *et al.*, 2020).

One of the advantages of BERT is the bidirectional transformer model that uses a deeper network architecture and is pre-trained with a massive text corpus to learn lexical text representations by predicting words conditioned on their context (Minaee *et al.*, 2020).

There have been numerous works on improving original BERT, such as RoBERTa, ALBERT, and DistillBERT (Minaee *et al.*, 2020). This work used the BERTIMBAL[2] model, a pre-trained BERT model for Brazilian Portuguese that achieves state-of-the-art performances on three downstream NLP tasks, such as Named Entity Recognition, Sentence Textual Similarity, and Recognizing Textual Entailment.

[2] Version BERT-large-portuguese-cased model from Hugginface repository. Retrieved from https://huggingface.co/neuralmind/bert-base-portuguese-cased

2.4 Related Works

Chalkidis *et al.* (2020) introduce LEGAL-BERT, a family of BERT models to legal NLP research, computational law, and legal technology application. The paper compares the performance of BERT with generic corpora fine-tuned with a specific domain and the BERT pre-trained from scratch with specialized legal domain corpora. They used 12 GB of English legal text from several fields, such as legislation, court cases, and contracts. The authors conclude that this BERT skilled version performs best results in various end-tasks, including multi-label classification.

Soh, Lim, & Chai (2019) compare traditional topic models and more recent, sophisticated, and computationally intensive techniques that use deep learning on the legal area classification task. They used 6,227 Singapore Supreme Court Judgments documents as a legal dataset, written in English. In conclusion, the traditional models LSA perform better than neural network models, including BERT and GloVe CNN.

Probably the lowest performance of the BERT model can be attributed to pre-processing tasks because the authors utilized only the first 512 tokens of each document as input to the BERT classifier since the model of the self-attention architecture only accepts that same number of tokens as input.

On the other hand, Shaffer & Mayhew (2019) split the documents into paragraphs and used all tokens as input, which avoids loss of information. They designed a method to link the US Supreme Court with the US Constitution. They compared the results of a rule-based system, a linear model, and a neural architecture for matching pairs of paragraphs. The deep learning approach performs better than other models. For that, they applied an embedded BERT model and a bidirectional LSTM.

Silva, Neves, Jesus, & Modanez (2022) used NLP techniques to outline the jurisprudence of the Federal Accounts Court (TCU) related to fraud in public purchases. A dataset of 300 labeled judgments was used, which contained 11 different types of corrupting pathologies. It classified 309,000 TCU decisions using Logistic Regression, Support Vector Machine, Random Forest, LSTM, bi-LSTM, and Convolutional Neural Network (CNN) models.

Although the test stage achieved an accuracy of 82.69%, in the evaluation stage the results were inconclusive. It suggests that one of the possible causes of the low quality of the final result is that the same TCU decision may include more than one class and they proposed that a new text segmentation approach could perform better (Silva *et al.*, 2022).

3. METHOD

After defining the research problem, main and specific objectives were established. A literature review was accomplished through searches in Google Scholar and Scopus databases, looking for articles published from 2018 forward related to machine learning and NLP applied to legal problems.

To select the first dataset to train the model, we used brazilian portuguese keywords such as: "Terceiro Setor"; "Organizações Sociais"; "Organizações da Sociedade Civil"; "OSCIP"; "controle social"; "prestação de serviços públicos"; "colaboração"; "assimetria de informações"; "terceirização"; "parceria"; "co-participação"; "co-produção"; "co-gestão"; etc.

As a result, articles were selected based on reading the abstract and the conclusion, seeking the most significant adherence to this research's objectives, and observing the number of citations. In addition, business specialists collected a set of 17 legal rules applicable to the TS.

The data mining tasks were developed considering the CRISP-DM (Cross Industry Standard Process for Data Mining) reference model: Business Understanding; Data Understanding; Data Preparation; Modeling; Evaluation; and Deployment (Shearer, 2000). All its stages are well organized, structured, and defined, helping to conduct and evaluate research in this area.

In the Business Understanding stage, the relevant issues were delved into to identify the universe of rules applicable to the TS, as well as possible databases. This step was discussed in previous sections, and the other stages will be described in the following sections.

Afterward, with the entire RegBR dataset classified, we calculated the percentage of sentences applicable to the theme of interest in each document. This result can be represented by the expression $X = TSts\ /\ TS$ where *TSts* is the *third sector's sentences*, and *TS* is the *Total Sentences*. This metric was used to define the document as Focus on Third-Sector; Addresses Third-Sector; Mentions to Third-Sector; and Non-Third-Sector, in an adaptation of the Likert scale that in its original form, that consists of five points, however, over time, the researchers changed the number of points used in your works (Bermudes; Santana; Braga & Souza, 2016).

Given the impossibility of a law containing one hundred percent of sentences related to the TS , we reduced the number of points on the scale to four levels, leaving the no sentence for rules with mention of TS arriving at a more consistent approach or legislation focused non-governmental organizations, as described in Table 01.

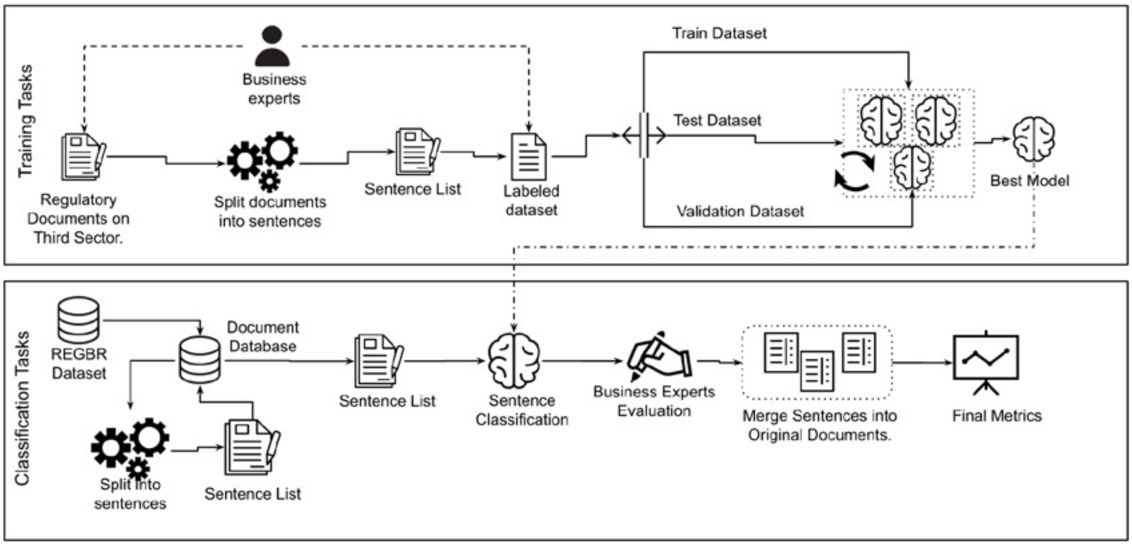
Table 01 - Adherence of the law to the Third Sector according to the sentences

TS Percent	Classification of the Law
0%	Non-Third Sector
0% - 30%	Mentions to Third Sector
30% - 60%	Addresses Third Sector
60% - 100%	Focus on the Third Sector

Note: *This* metric represents % sentences about the Third Sector on Total sentences of the law.
Source: elaborated by the authors.

Next, other steps of the methodological process used in this research are detailed. Figure 01 describes the overview of machine learning tasks that will be discussed in the following sections.

Figure 01 - Overview of machine learning tasks



Source: elaborated by the authors.

3.1 Data Understanding and Data Preparation

In the Data Understanding stage, the existing databases on the topic were evaluated. As mentioned above, the main challenge of classification tasks using supervised techniques is acquiring a valid and labeled dataset to train machine learning models (Ko & Seo, 2009).

In fact, there is not available and labeled dataset on the regulation of the TS in Brazilian regulatory law. Nevertheless, in Brazil, some laws are specific about this sector, and others mention it partially. Consequently, a minimal previous dataset using the legal documents identified by human specialists was built in the Business Understanding step.

In the Data Preparation stage, the RegBR dataset was loaded into a MongoDB NoSQL database and the first statistics were collected. Each document on RegBR was split into sentences using NLTK Sentence Tokenizer and stored in the database. We optimized the sentence generation algorithm for Brazilian law vocabulary. In addition, we took care that all sentences had a maximum of 512 tokens without stopwords to optimize the training tasks and avoid loss by BERT padding/truncate process, similar to Shaffer & Mayhew (2019).

No preprocessing or additional treatment was carried out on the database provided by RegBR. The data just was loaded as published. Therefore, eventually, any inconsistency in the data may be associated with the source of information obtained. Table 02 resumes the entire dataset:

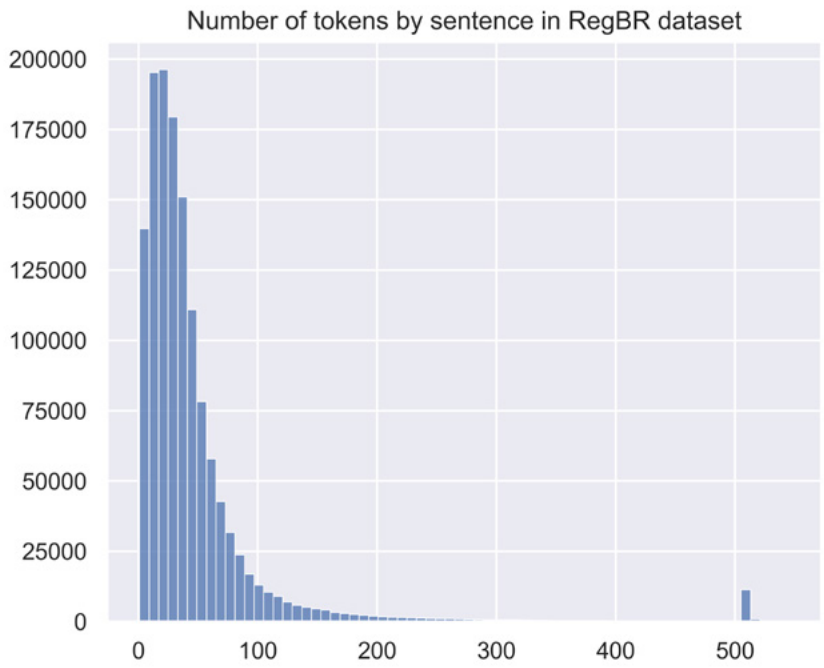
Table 02 - Summary of RegBR Dataset

Documents	50,999
Sentences	1,330,190
Tokens mean by Sentence	46.18
Tokens STD by Sentence	64.6
Tokens Min	1
Tokens Max	512

Source: elaborated by the authors.

After that, the frequency of the number of tokens for each sentence was plotted in a histogram. This result was demonstrated in Figure 02. Most of the 1,330,190 sentences contain between 20 and 50 tokens, but there is a considerable number of documents that contain a lot more tokens. This distribution demonstrates that the sentences have varying lengths, ranging from 1 token to 512 tokens, with a standard deviation of approximately 64.6.

Figure 02 - Histogram of frequency of the token number by sentence



Source: elaborated by the authors.

Hence, the first challenge was to build a dataset about TS's regulation. Business experts collected 17 regulatory documents and an equal number of other laws on various topics, such as health care, communication, and public security. The legal acts were split into sentences using NLTK Sentence Tokenizer with the same algorithm used to split RegBR dataset documents into sentences.

The 34 documents were divided into 6,040 sentences manually annotated by business specialists, with 519 orations labeled as TS and 5,521 as non-TS class. The tokens "sociedade civil", "organização da sociedade", "organização social", "parceria", and "administração pública" are the most relevant terms on corpora.

All these words accurately translate terms that are known to be related to the TS, at least in the vocabulary of the analyzed law dataset. These words are equivalent to nouns and expressions that in English mean "civil society", "organization of society", "social organization", or "partnership", that is, basically as entities that participate and form this segment of the social economy. We removed stopwords and punctuations to analyze a corpus of 519 TS's sentences and plotted a word cloud as shown in Figure 03.

Figure 03 - Word cloud of TS's sentences dataset

Source: elaborated by the authors.

The terms evidenced above are closely related to the keywords used for the purpose of recognizing the laws of TS. For example, the words "sociedade civil", "organização da sociedade", "organização social", or "parceria" were all indicated by the analyst experts as terms that relate to the regulation of this sector. Although the combined term "administração pública", which means *public administration*, has not been expressly included by the experts. In our analysis we understand the model's prediction correctly, considering that much of the legislation addresses partnerships between NOGs and the public sector.

After the business expert validation, the annotated dataset was split into the Training/Testing Dataset (TDS) and Validation Dataset (VDS). The information is unbalanced, with 8.5% positive samples. We randomly built the TDS with 80% of the minority class and 98% of the majority class. The VDS was built with 20% of the minority class and 2% of the majority class to obtain a balanced dataset. It is essential to avoid problems with evaluation metrics. Table 03 describes the dataset used to train, test, and validate the deep learning NLP model.

Table 03 - Business expert annotated datasets overview

Dataset	Size	Third-sector	Non-third-sector
Entire Train/Test/ Validation	6,040	519	5,521
Training/Test (TDS)	5,963	442	5,521
Validation (VDS)	262	113	149

Source: elaborated by the authors.

3.2 Modeling and Evaluation

The pre-trained BERTIMBAL version BERT-large-portuguese-cased model was fine-tuned with training with the TDS dataset. It used the hyper-parameter suggested by Devlin *et al.* (2019) for fine-tuning downstream tasks:

- batch size = 16;
- learning rate = 2e-5;
- epochs= 4;
- dropout rate of 0.1.

In our research, we used the PyTorch library. McCormick & Ryan (2019) inspired the source code, and the dataset and the developed model are online on GitHub (https://github.com/mbjesus/regbr/). It applied the Adam Optimization Algorithm from PyTorch worked in a GPU Nvidia.

For validation tasks, it used the metric F1-Score and Area Under the ROC Curve (AUC). F1-Score is useful to evaluate binary classification tasks, which label is 'positive' or 'negative'. It combines the harmonic mean between precision and recall. Precision is the number of true positives (TP) divided by the number of true positives plus false positives (FP). The Recall is the true positives divided by true positives plus false negatives (FN) (Ferri, Hernández-Orallo, & Modroiu, 2009; Liu, Zhou, Wen, & Tang, 2014). These metrics can be represented by the expressions below:

Figure 04 - Metrics for validation tasks

$$F1 = \frac{2 \times Recall \times Precision}{Recall \times Precision}$$

$$Precision = \frac{TP}{TP + FP}$$

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

Source: elaborated by the authors.

The Receiver Operating Characteristic (ROC) is a curve on a two-dimensional depiction between the True Positive rate on the Y axis ($t = TP / P$) and the False Positive rate on the X axis ($f = FP / P$) as proposed by Liu *et al.* (2014). The Area Under the ROC Curve (AUC) is a measure between 0 and 1: the closer to 1, the better the classifier, as it has a higher true positive rate than a false positive rate.

After each epoch, the fine-tuned model was evaluated with the VDS dataset. The training loss rate decreased with each epoch, demonstrating that it could learn from the training data. However, the Evaluation loss rate remained practically constant.

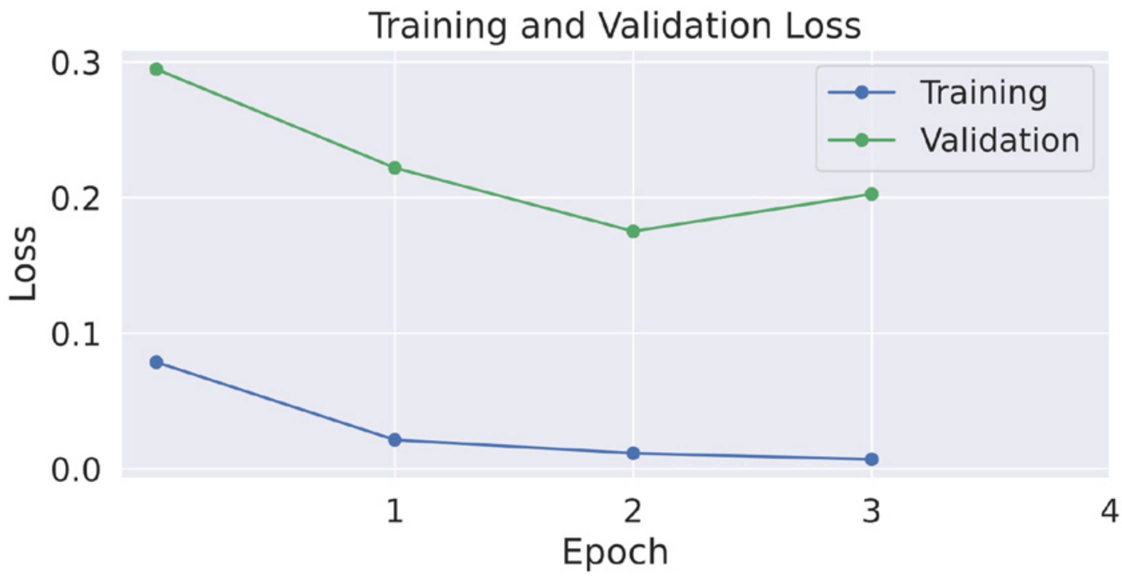
F1-Score and AUC metrics increased slightly. These numbers reinforce that the four training epochs proposed by Devlin *et al.* (2019) for fine-tuning are sufficient. Table 04 details the metrics of the training and validation stages, and Figure 05 presents the comparison of training and validation loss.

Table 04 - Fine-tuned model evaluates results with the VDS dataset

Epoch	Training Loss	Validation Loss	Validation F1-measure	Validation AUC
1	0.078	0.294	0.911	0.925
2	0.021	0.221	0.949	0.952
3	0.011	0.175	0.954	0.957
4	0.007	0.202	0.949	0.948

Source: elaborated by the authors.

Figure 05 - Comparison of training and validation loss

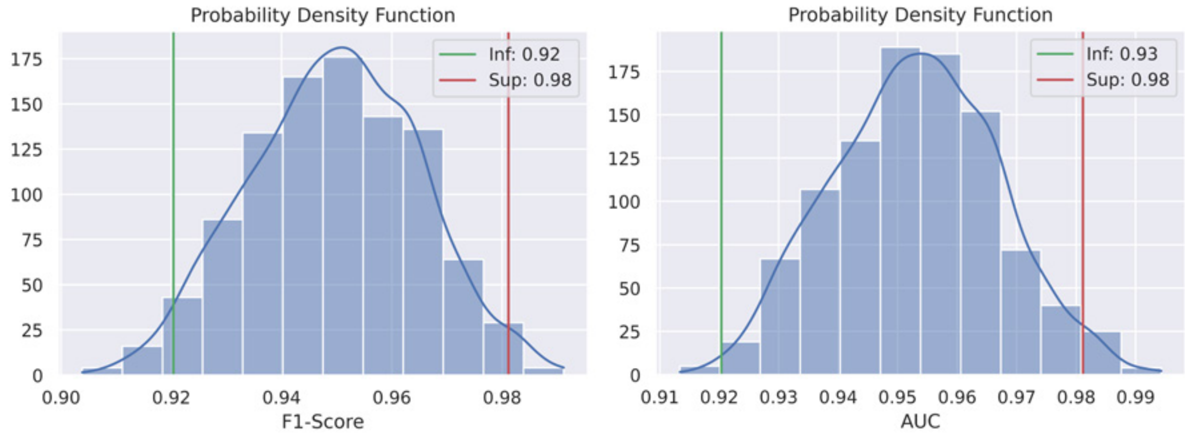


Source: elaborated by the authors.

Bootstrap is a computationally intensive resampling technique that allows obtaining parameters of a population through successive samplings in an initial sample. In this technique, an initial sample is taken from the population. From this selection, several others are produced, and an empirical distribution of the target variable is generated without necessarily knowing any previous distribution of this variable in the population. We assume that the empirical distribution is close to the population distribution; therefore, it is possible to estimate complex population parameters from it (Hastie, Tibshirani, Friedman, & Franklin, 2005).

We build a Bootstrap over 1000 samples without repetition from the VDS dataset and calculate the F1-Score and AUC density distribution. As demonstrated in Figure 06, the distribution is close to the Normal Distribution. Consequently, we can estimate the confidence interval with α = *5%* for these metrics based on the Limit Central Theorem. The confidence interval for F1-Score was 92% and 98% and for AUC was 93% and 98%.

Figure 06 - Bootstrap resampling results



Note: 95.0 confidence interval 92% and 98%
Source: elaborated by the authors.

4. DEPLOYMENT AND RESULTS

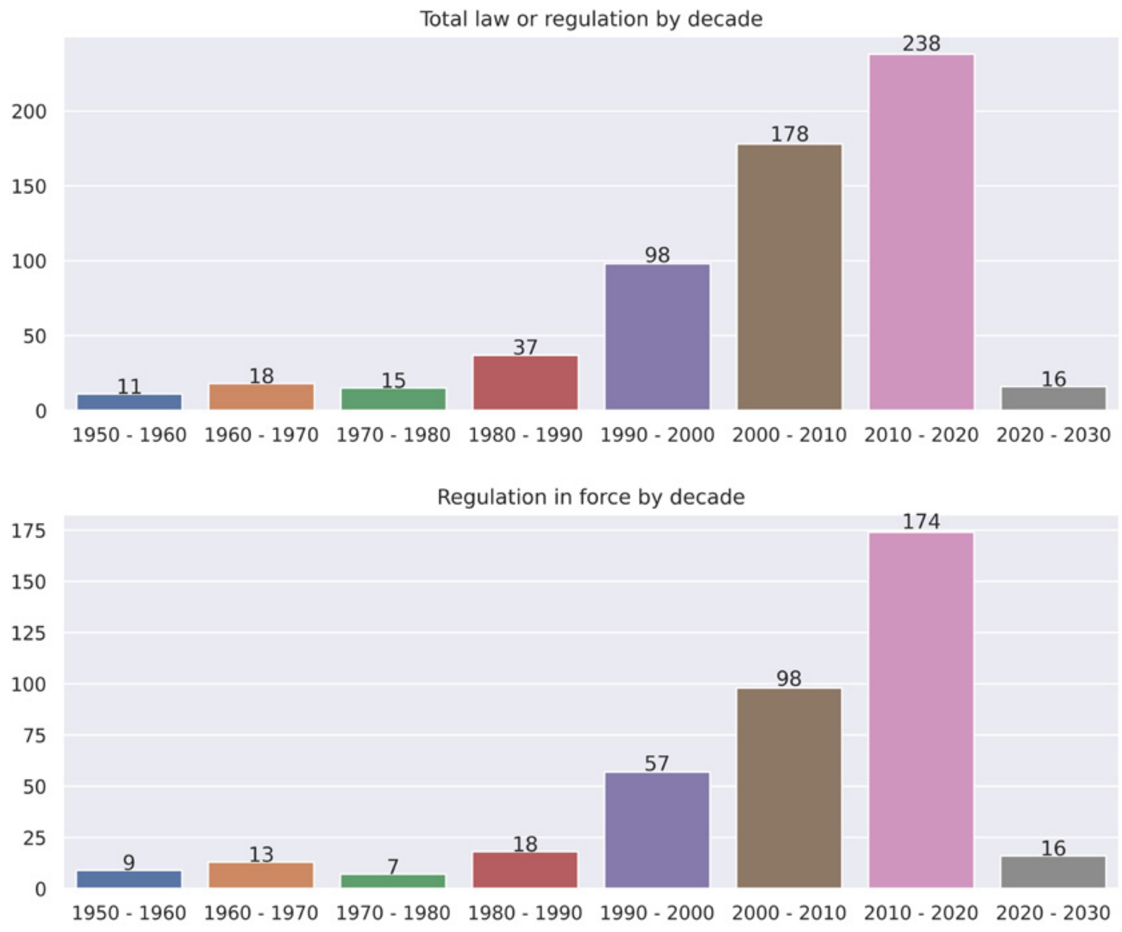
Finally, in the Deployment stage, 1,330,190 sentences were classified, and the BERT model labeled 2,538 sentences as TS. Business experts analyzed the results and found 27 incorrect classifications, and we removed these sentences from the TS dataset. After that, we merge the sentences to rebuild the original document to compute the metrics. The result was used to build the legal framework for the third sector.

These complete results of automated classification were revised by humans again. The conclusion is that the computational model obtained an assertiveness of 93.94%. In the end, 710 regulations were found with the prediction of the algorithm, of which 611 were correctly predicted. The total of rule sentences predicted was 2,511, where 2,359 was about TS. For comparison, the first search using keywords returns just 17 regulations against more than 610 with machine learning.

This slight imprecision is explained by the fact that many expressions detected by the A.I. are related to other forms of a covenant with the government. An example of this is the mention of "contratos de gestão" (management agreement) that can refer to social organizations but also refer to executive agencies (a type of contract of goals between public departments). Another situation is public-private partnerships aimed at executing infrastructure projects, more similar to contracting and outsourcing processes, not being characterized as a third sector.

The first law identified by the machine learning algorithm was published in 1952, and many other regulations to the Third Sector have been published, of which 392 remain in force, according to the RegBR database. Notably, 47 laws were published between the 1950s and 1900s and are still in force. Figure 07 shows the total number of acts by decade and the current laws that remain.

Figure 07 - Total number of acts found by algorithm per decade x into force acts

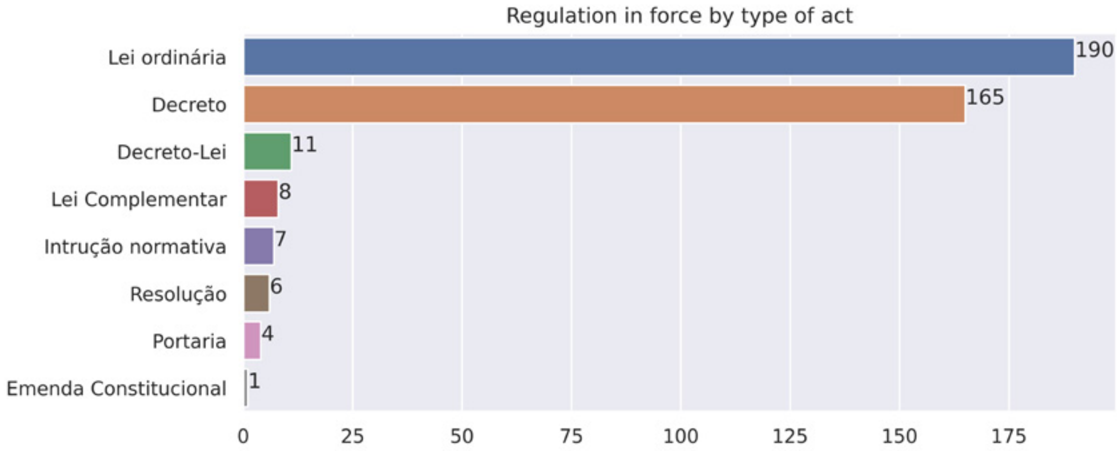


Source: elaborated by the authors.

These results are coherent with the Brazilian historical context. In the 1990s, after the promulgation of the Constitution on October 5, 1988, there was a considerable increase in the regulation of the TS. In this decade, 98 rules were published, among which 57 acts remain in force. The trend of increasing TS regulation was more evident in the following years. More laws were published in the 2000s, and in the 2010s, the number of regulations peaked.

Another metric extracted from the items selected using artificial intelligence is the type of normative act. In this regard, the result shows that the vast majority of regulations have the legal form (“Lei Ordinária”), with 190 regulatory actions in this classification. This number is closely followed by regulatory decrees (“Decretos”), a lower normative type whose edition is attributed to the head of the Brazilian Executive, totaling 165 legal documents in this format. Figure 08 below shows the complete result for this classification:

Figure 08 - Regulation acts about Third-Sector identified by normative type



Source: elaborated by the authors.

Finally, from the proposed classifications, it was possible to extract the metric referring to mentions of the TS. Table 05 summarizes the results.

Table 05 - Overview of RegBR dataset categorization

Focus on Third Sector	3
Addresses Third Sector	29
Mentions to Third Sector	579
Non-Third-Sector	50,338

Note: The complete result and the computational model are online and open-source, published at GitHub (https://github.com/mbjesus/regbr/)
Source: elaborated by the authors.

The research has demonstrated how artificial intelligence can help legal professionals to select and manage large volumes of information. In this way, it contributes to deep learning research and demonstrates practical application to the improvement of the analyzed legislation. The proposed method and techniques can be replicated on legal databases worldwide and any other topic of interest.

5. DISCUSSION AND CONCLUSIONS

The present work used artificial intelligence to perform the automated processing of legal language and analyze the Brazilian regulatory flow by identifying and classifying the sentences and norms related to the Third Sector. The partnership between the government and individuals in the execution of public policies is, therefore, the central theme of the object of study.

To regulate these civil society institutions, rules are necessary to coordinate these social activities. The precise definition of a regulatory framework bringing together sparse normative acts such as Brazil can use technology to improve the contribution of law to social reality.

The bibliographic research revisited the regulatory theories and the literature on the analyzed theme. We delve deeper into the regulation theory to identify essential aspects for constructing norms related to the research topic to extract applicable concepts in machine learning. We also revisit the state-of-the-art in natural language processing to build a background on artificial intelligence.

The results achieved the proposed objectives, demonstrating that the analysis and classification promoted with the BERT model with fine-tuning with a Brazilian legal dataset were highly effective, reaching 94% F1-Score and 94% de AUC.

The model processed 1,330,190 sentences and correctly classified 2,359 as TS. This result allowed us to find other 611 laws applicable to the TS, from a set of 50,999 rules in the RegBR dataset. From this information, it was possible to evaluate metrics illustrating how this regulation was developed over time.

Thereby, the work contributes to the improvement of research in machine learning and the law for TS. For future work, the aim is to extend the classification to jurisprudential decisions of public accounts, to understand how the public-private collaboration has been inspected by the Brazilian Courts. Furthermore, semantic analysis of the norms can be applied to verify if it limits or grants a right to the TS.

REFERENCES

Abrucio, F. L. (2006). Contratualização e organizações sociais: reflexões teóricas e lições da experiência internacional. *Debates GV Saúde, 1*(10), 24–27.

Advocacia-Geral da União (2018). *Manual de Boas Práticas Regulatórias*. Brasília: AGU. Retrieved from https://drive.google.com/file/d/1lKjBXT7WQpXWAdM_3xZClsD-BnqvU8-X/view

Amaral, F. (2016). *Aprenda Mineração de Dados: teoria e prática*. Alta Books: Rio de Janeiro.

Ansell, C., & Torfing, J. (2016). Introduction: theories of governance. In: Ansell C., Torfing, J. (editors). *Handbook on Theories of Governance*. Elgar Online. https://doi.org/10.4337/9781782548508.00008

Aranha, M. I. (2014). *Manual de Direito regulatório: fundamentos de Direito regulatório (*2nd ed., Coleford, Gloucestershire, United Kingdom: Laccademia Publishing Limited.

Aranha, M. I. (2019). Teoria jurídica da regulação: entre escolha pública e captura. *RDP - Revista Direito Público - Edição Especial: 15 anos. 16* (2019), 11–37.

Barbosa, N. B., Lima, H. de, Trindade, A. M. V. da, Dias, N. V., Castro, M. V. de, & Santos, C. K. dos. (2015). As Organizações Sociais de Saúde como forma de gestão público-privada em Goiás – o caso HUANA. *Revista Do Serviço Público, 66* (1), 121 - 144. https://doi.org/10.21874/rsp.v66i1.686

Bermudes, W. L., Santana, B. T., Braga, J. H. O., & Souza, P. H. (2016). Tipos de escalas utilizadas em pesquisas e suas aplicações. *Revista Vértices*, *18* (2), 7-20.

Brasil (1995). *Plano Diretor da Reforma do Aparelho do Estado Cadernos Mare*. Brasília: Presidência da República 1995. Retrieved from http://www.biblioteca.presidencia.gov.br/publicacoes-oficiais/catalogo/

fhc/plano-diretor-da-reforma-do-aparelho-do-estado-1995.pdf

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The Muppets straight out of Law School. *Findings of the Association for Computational Linguistics: EMNLP*. https://doi.org/10.48550/arXiv.2010.02559

Crack, A. M. (2018). The Regulation of International NGOS: Assessing the Effectiveness of the INGO Accountability Charter. *Voluntas 29*, 419–429. https://doi.org/10.1007/s11266-017-9866-9

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: *Pre-training of deep bidirectional transformers for language understanding*. https://doi.org/10.48550/arXiv.1810.04805

Dias, G., Farias, J., Coura, K., Demo, G., & Athayde, A. (2022). Reflexões sobre a falta de água em algumas regiões brasileiras e a coprodução na gestão pública:: rumo a uma gestão hídrica mais efetiva. *Revista do Serviço Público, 73* (1), 32 - 54. Recuperado de https://revista.enap.gov.br/index.php/RSP/article/view/4769

Ellig, J. (2018). Why and How Independent Agencies Should Conduct Regulatory Impact Analysis. *Cornell Journal of Law and Public Policy, 28* (1), 1–34. http://doi.org/10.2139/ssrn.3169413

Escola Nacional de Administração Pública (2022). *RegBR. Dados Atualizados em Maio/2022 com foco em normativos regulatórios*. Brasília: Enap. Retrieved from https://infogov.enap.gov.br/regbr

Ferri, C., Hernández-Orallo, J., & Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, *30* (1), 27-38. https://doi.org/10.1016/j.patrec.2008.08.010

Hahn, R. W. & Tetlock, P. C. (2008). Has Economic Analysis Improved Regulatory Decisions? *Journal of Economic Perspectives, 22* (1), 67–84. https://doi.org/10.1257/jep.22.1.67

Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer, 27* (2), 83–85. https://doi.org/10.1007/BF02985802

Ko, Y., & Seo, J. (2009). Text classification from unlabeled documents with bootstrapping and feature projection techniques. *Information Processing* & *Management, 45* (1), 70-83. https://doi.org/10.1016/j.ipm.2008.07.004

Koop, C., & Lodge, M. (2017). What is regulation? An interdisciplinary concept analysis. *Regulation & Governance, 11* (1), 95–108. https://doi.org/10.1111/rego.12094

Levi-Faur, D. (2011). Regulation and Regulatory Governance. In *Handbook on the Politics of Regulation*. [s.l.] Edward Elgar Publishing, *2011*, 3–21. https://doi.org/10.4337/9780857936110.00010

Liu, Y., Zhou, Y., Wen, S., & Tang, C. (2014). A Strategy on Selecting Performance Metrics for Classifier Evaluation. *International Journal of Mobile Computing and Multimedia Communications (IJMCMC), 6*(4), 20-35. http://doi.org/10.4018/IJMCMC.2014100102

Luger, G. F (2013). *Inteligência Artificial* (6th ed, Daniel Vieira, Trans.) São Paulo: Pearson Education do Brasil.

McCormick, C., & Ryan, N. (2019). *BERT Fine-Tuning Tutorial with PyTorch*. Retrieved from https://mccormickml.com/2019/07/22/BERT-fine-tuning/

McMullin, C. (2021). Challenging the necessity of New Public Governance: Co-production by third sector organizations under different models of public management. *Public Administration, 99* (1), 5–22. https://doi.org/10.1111/padm.12672

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2020). *Deep Learning Based Text Classification: A Comprehensive Review*. https://doi.org/10.48550/arXiv.2004.03705

Montano, M. R., & Pires, D. D. O. O (2019). Marco Regulatório Do Terceiro Setor: A Oferta Da Educação Infantil No Município De Porto Alegre. *Práxis Educacional, 15* (31), 16-17. https://doi.org/10.22481/praxis.v15i31.4656

Phillips, S., & Smith, S. (2014) A Dawn of Convergence?: Third sector policy regimes in the "Anglo-Saxon" cluster. *Public Management Review, 16* (8), 1141–1163. https://doi.org/10.1080/14719037.2014.965272

Phillips, S., & Hebb, T. (2010). Financing the third sector: Introduction. *Policy and Society, 29* (3), 181–187. https://doi.org/10.1016/j.polsoc.2010.07.001

Saab, F., & Silva, S. A. M. (2021). Evaluating the quality of Regulatory Impact Analysis: a literature review. *Revista do Serviço Público*, 72(b), 34 - 57. Retrieved from https://revista.enap.gov.br/index.php/RSP/article/view/5875

Sano, H., & Abrucio, F. L. (2008). Promessas e resultados da Nova Gestão Pública no Brasil: o caso das organizações sociais de saúde em São Paulo. *Revista de Administração de Empresas, 48* (3), 64–80. https://doi.org/10.1590/S0034-75902008000300007

Shaffer, R., & Mayhew, S. (2019). Legal Linking: Citation Resolution and Suggestion in Constitutional Law. In *Proceedings of the Natural Legal Language Processing Workshop 2019*, 39–44. Minneapolis, Minnesota: Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-2205

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing, 5* (4), 13-22.

Silva, T. T. S. E., Neves, C. B. D., Jesus, M. B. D., & Modanez, W. (2022). *Inteligência artificial e Jurisprudência: delimitação jurisprudencial nas decisões do TCU do conceito aberto de cláusula restritiva ao caráter*. Editora Dialética.

Soh, J., Lim, H., & Chai, I. (2019). Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments. In *Proceedings of the Natural Legal Language Processing Workshop 2019*, 67–77. Minneapolis, Minnesota. Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-2208

Talesh, S. (2022). Public law and regulatory theory. In *Handbook on Theories of Governance*. [s.l.] Edward Elgar Publishing. 104–117. https://doi.org/10.4337/9781800371972.00016

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L. Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, 6000–6010. https://doi.org/10.48550/arXiv.1706.03762

Vieira, J. P. T. O. (2016). *Fragilidades Dos Mecanismos De Controle No Município Do Rio De Janeiro.* Rio de Janeiro. FGV EBAPE - Dissertações, Mestrado Profissional em Administração Pública [366]. Retrieved from http://hdl.handle.net/10438/16060

Zheng, L., Guha, N., Anderson, B., Henderson, P., & Ho, D. (2021). When does pretraining help? assessing self-supervised learning for law and the CaseHOLD dataset of 53,000+ legal holdings. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law (ICAIL '21)*. Association for Computing Machinery, New York, NY, USA, 159–168. https://doi.org/10.1145/3462757.3466088

Mauricio Barros de Jesus
https://orcid.org/0000-0002-6999-3560
Master in Applied Computing with an emphasis on Data Science, Universidade de Brasília (UnB). Computer Engineer, Universidade Federal de Goiás (UFG).
mbjesus@tce.go.gov.br

André da Silva Goes
https://orcid.org/0000-0003-2718-3380
Student of the Professional Master's Program in Public Administration, Universidade Federal de Goiás (PROFIAP/UFG). Specialist in Auditing and Government Management and Bachelor's degree in Computer Science, Pontifícia Universidade Católica de Goiás (PUC-GO).
asgoes@tce.go.gov.br

Leonardo de Guimarães Santiago
https://orcid.org/0000-0002-0656-8110
Student of the Professional Master's program in Public Administration, Universidade Federal de Goiás (PROFIAP/UFG). Specialist in Public Sector Economics and Public Sector Auditing (WPós), Pontifícia Universidade Católica de Goiás (PUC-GO).
leonardogsantiago@gmail.com

Marcelo Augusto Pedreira Xavier
https://orcid.org/0000-0001-5974-9364
Specialist in External Control and Public Governance, Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP). Student of the Master's Program in Human Rights, Universidade Federal de Goiás (PPGIDH/UFG).
marceloaugustoweb1@gmail.com

Sólon Bevilacqua
https://orcid.org/0000-0002-0050-3527
Professor and Coordinator of the Professional Master's Degree in Public Administration, Universidade Federal de Goiás (PROFIAP/UFG). Master in Administration, Universidade Federal de Uberlândia (UFU). Doctor in Psychology, Pontifícia Universidade Católica de Goiás (PUC-GO). Specialist in Production Engineering and Graduated in Administration, Universidade Federal do Rio Grande do Sul (UFRGS).
solon@ufg.br