



# Transfer Learning Applied to Government Auditing: A Focused Approach on Financial Statements in Maranhão, Brazil

HELOISA GUIMARÃES COELHO

novembro de 2024

# **Transfer Learning Applied to Government Auditing**

## **A Focused Approach on Financial Statements in Maranhão, Brazil**

**Heloisa Guimarães Coelho**  
**Student No.: 1222616**

**A dissertation submitted in partial fulfillment of the  
requirements for the degree of Master of Artificial Intelligence  
Engineering**

**Supervisor: Dr. Maria Goreti Carvalho Marreiros**  
**Principal Coordinator Professor, Institute of Engineering, Polytechnic of Porto**

**Co-Supervisor: Dr. Luis Fernando Maia Santos Silva**  
**Full Professor, Federal Institute of Maranhão, Caxias Campus, Brazil**

**Evaluation Committee:**

President:

Dr. Carlos Fernando da Silva Ramos

Principal Coordinator Professor, Institute of Engineering, Polytechnic of Porto

Members:

Dr. Luiz Felipe Rocha de Faria

Coordinator Professor, Institute of Engineering, Polytechnic of Porto

Dr. António Constantino Lopes Martins

Adjunct Professor, Institute of Engineering, Polytechnic of Porto



# Dedictory

To my good Lord, God, from whom I always receive a special blessing and without whom this dissertation would not have been completed. To my Holy Mother the Virgin Mary who is always praying for me from heaven and Saint Joseph and my Guardian Angel who are always protecting me.

To my family, in particular my parents and brother, who are my best friends in this world, for all the love and support they are always giving me.



# Abstract

Since Brazil's return to democracy, dozens of laws, decrees and normative instructions have been drafted with the purpose of regulating and improving the mechanisms for controlling and monitoring municipal public resources. These regulations are specifically aimed at the process of accountability by elected officials, who currently rely on the help of accountants responsible for preparing and submitting financial statements to the Courts of Auditors.

However, according to data from the TCU (Federal Court of Accounts), in 2023, Maranhão was the Brazilian State with the highest number of rejected accounts. There are several reasons that can lead to these processes being challenged, including incorrect application of resources, flaws in documentation, human errors, among others. In practice, the routine of accountants includes repetitive and mechanical activities that requires considerable time to prepare and review documents, hence often leading to errors in classification and issuing of documentation.

In this context, this dissertation investigates the use of Transfer Learning (TL) to improve automation and accuracy in the classification of financial commitment notes, an initial document in the public expenditure cycle, with a specific focus on the context of the state of Maranhão. To this end, BERTimbau, a pre-trained language model for Brazilian Portuguese, was fine-tuned to assist government accountants in reducing classification errors and ensuring compliance with local and national financial regulations.

The CRISP-DM methodology, widely used in data science, was adopted to structure the development of the project. The dataset used, consisting of several classifications of commitment notes for the year 2023, was thoroughly analyzed and pre-processed. For the fine-tuning process of the model, two samples with a similar number of data were selected, varying only the number of possible classifications, due to the high degree of imbalance between the classes.

Even in a multiclass context with datasets with a reduced number of classes, the results obtained indicate that the BERTimbau model presents strong performance in the classification task, achieving 98% accuracy with an error rate of 0.10 in the test set, highlighting the effectiveness of BERTimbau in public financial auditing applications. These results highlight the effectiveness of BERTimbau for public financial auditing applications. It is therefore concluded that TL models have great potential to optimize and improve financial auditing processes, with positive implications for wider adoption in Brazil.

**Keywords:** Government Auditing, Financial Auditing, Financial Statements, Transfer Learning, Fine-Tuning, BERTimbau



# Resumo

Desde a redemocratização brasileira, dezenas de leis, decretos e instruções normativas foram elaboradas com o propósito de regulamentar e aprimorar os mecanismos de controle e fiscalização dos recursos públicos municipais. Essas regulamentações visam especialmente o processo de prestação de contas por parte dos gestores eleitos, que contam atualmente com a ajuda de contabilistas responsáveis pela elaboração e envio das demonstrações financeiras aos Tribunais de Contas.

Contudo, segundo dados do TCU (Tribunal de Contas da União), em 2023 o Maranhão foi o estado brasileiro com o maior número de contas reprovadas. São diversos os motivos que podem causar a impugnação destes processos, desde a incorreta aplicação dos recursos, vícios na documentação, erros humanos, entre outros. Na prática, a rotina dos contabilistas inclui atividades repetitivas e mecânicas, o que torna comuns erros de classificação e emissão de documentos, devido ao tempo considerável necessário para elaboração e revisão desses registros.

Perante esse contexto, esta dissertação investiga o uso de Transfer Learning (TL) para melhorar a automação e a precisão na classificação das notas de empenho financeiro, documento inicial no ciclo da despesa pública, com foco específico no contexto do estado do Maranhão. Para tal, foi realizado o fine-tuning do BERTimbau, um modelo de linguagem pré-treinado para o português do Brasil, com o intuito de auxiliar contabilistas governamentais na redução de erros de classificação e garantir a conformidade com as regulamentações financeiras locais e nacionais.

A metodologia CRISP-DM, amplamente utilizada em ciência de dados, foi adotada para estruturar o desenvolvimento do projeto. O conjunto de dados utilizado, composto por várias classificações de notas de empenho referentes ao ano de 2023, foi minuciosamente analisado e pré-processado. Para o processo de fine-tuning do modelo selecionou-se duas amostras com um número semelhante de instâncias, variando apenas a quantidade de classificações possíveis, devido ao alto grau de desbalanciamento entre as classes.

Mesmo em um contexto multiclasse com conjuntos de dados com quantidade de classes reduzida, os resultados obtidos indicam que o modelo BERTimbau apresenta excelente desempenho na tarefa de classificação, alcançando 98% de acurácia com taxa de erro de 0.10 no conjunto de teste, o que destaca a eficácia do BERTimbau em aplicações de auditoria financeira pública. Esses resultados sublinham a eficácia do BERTimbau para aplicações de auditoria financeira pública. Conclui-se, assim, que modelos de TL possuem grande potencial para otimizar e aprimorar processos de auditoria financeira, com implicações positivas para uma adoção mais ampla no Brasil.

**Palavras-chave:** Government Auditing, Financial Auditing, Financial Statements, Transfer Learning, Fine-Tuning, BERTimbau



# Acknowledgement

Firstly, I would express my deep gratitude to God and all my friends in Heaven, who always support me and encourage me with their prayings to keep going. To my family, especially my parents, brother and dear aunts, for all the love and support which they are always giving me. It was through their struggles and efforts that I had the opportunity to get here.

To ISEP for welcoming me so kindly, to all my professors for all the knowledge they provided me, especially Professor Dr. Carlos Ramos for all the support and my dear Supervisor Professor Dr. Goreti Marreiros, to whom I am especially grateful for the advice and trust.

I would like to Thank my Co-Supervisor, Professor Dr. Luis Fernando Maia for all the guidance, support and encouragement. His advice and trust were essential throughout my academic journey. To my beloved institution IFMA and its dedicated team, both at the Rectorate in São Luís and at Caxias Campus, for their efforts to make possible for me to fulfill my dream of studying abroad and having extraordinary experiences.

To the Maranhão Foundation for the Protection of Research and Scientific and Technological Development (FAPEMA) and the Innovation Laboratory of IFMA Caxias Campus for the opportunity to develop this project and all the financial support.

To the two accountants who promptly contributed to this project with clarifications on public accounting in the Maranhão context, as well as the necessary authorizations for the development of the AI model.

To all my friends, from whom I always find encouragement, prayers and good laughs. You are a gift from God in my life. Thank you very much for everything.

Finally, my sincere gratitude to everyone involved in the success of the work presented in this dissertation.



# Contents

<b>List of Acronyms</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Context and Motivation . . . . .	1
1.2 Problem Statement . . . . .	3
1.3 Objectives and Research Questions . . . . .	4
1.4 Scientific Contributions . . . . .	4
1.5 Document Structure . . . . .	5
<b>2 State-of-the-art</b>	<b>7</b>
2.1 Research Methodology . . . . .	7
2.2 Research Findings . . . . .	8
2.3 Discussion . . . . .	10
2.4 Chapter Remarks . . . . .	12
<b>3 Methodology</b>	<b>15</b>
3.1 Project Development . . . . .	15
3.2 Technologies . . . . .	17
3.3 Evaluation . . . . .	19
3.4 Chapter Remarks . . . . .	20
<b>4 Data Overview</b>	<b>21</b>
4.1 Business Understanding . . . . .	21
4.2 Data Understanding . . . . .	23
4.2.1 Data Adopted . . . . .	23
4.2.2 Data Extraction . . . . .	23
4.2.3 Data Exploration . . . . .	24
4.2.4 Data Protection and Ethical Aspects . . . . .	30
4.3 Chapter Remarks . . . . .	30
<b>5 Proposed Solution and Evaluation</b>	<b>33</b>
5.1 Data Preparation . . . . .	33
5.1.1 Data Preprocessing . . . . .	33
5.1.2 Data Balancing . . . . .	34
5.2 Model Specifications . . . . .	35
5.3 Fine-Tuning Process and Evaluation . . . . .	37
5.4 Chapter Remarks . . . . .	42
<b>6 Conclusion</b>	<b>43</b>
6.1 Accomplished Objectives . . . . .	43
6.2 Limitations and Future Work . . . . .	44

6.3 Chapter Remarks . . . . .	44
<b>Bibliography</b>	<b>45</b>
<b>A Government Functional Codes</b>	<b>51</b>

# List of Figures

1.1	Percentage of disapproved accounts by Brazilian macro-region in 2023 . . .	2
1.2	Number of accounts judged irregular by UF in 2023. . . . .	2
2.1	Adapted PRISMA diagram . . . . .	9
3.1	Life cycle of CRISP-DM . . . . .	16
3.2	DSPA survey results . . . . .	17
3.3	Overview of the proposed model and its phases . . . . .	18
4.1	Workflow in Maranhão's Government Auditing by TCE-MA . . . . .	21
4.2	Diagram of AI integrated in the ConnTec system . . . . .	22
4.3	Adapted example of a Commitment Note from Maranhão . . . . .	24
4.4	Percentage of Commitment Notes by Government Function Code . . . . .	25
4.5	Percentage of Commitment Notes by Government Function Code . . . . .	26
4.6	Commitment Notes categorized by Type . . . . .	27
4.7	Distribution of Commitment Notes by Class . . . . .	28
5.1	Final dataset with attributes ordered . . . . .	34
5.2	Token Count of the column description . . . . .	36
5.3	Training and validation accuracy results for DS10Codes p/ epoch . . . . .	38
5.4	Training and validation accuracy results for DS15Codes p/ epoch . . . . .	38
5.5	Training and validation loss results for DS10Codes p/ epoch . . . . .	39
5.6	Training and validation loss results for DS15Codes p/ epoch . . . . .	39
5.7	Confusion Matrix for for DS10Codes . . . . .	41
5.8	Confusion Matrix for for DS15Codes . . . . .	42



# List of Tables

2.1	Final Research Queries . . . . .	7
2.2	Inclusion Criteria . . . . .	8
2.3	Exclusion Criteria . . . . .	8
2.4	Selected studies for the review . . . . .	10
2.5	Overview of ML Approaches in Government Financial Audits . . . . .	13
3.1	Confusion Matrix Structure . . . . .	19
4.1	Dataset Structure and Attributes . . . . .	28
5.1	Datasets Detailed p/ Function Code . . . . .	35
5.2	Function Codes and Descriptions . . . . .	35
5.3	Summary of the hyperparameters used with the model . . . . .	37
5.4	No. examples p/ dataset . . . . .	37
5.5	Number of Samples and Training time p/ dataset . . . . .	37
5.6	Dataset Accuracy Metrics . . . . .	38
5.7	Datasets Loss Metrics . . . . .	39
5.8	Precision, Recall, and F1 Score Metrics . . . . .	40
5.9	Precision, Recall, F1-Score for each Code in the DS10Codes and DS15Codes datasets. . . . .	40
5.10	Balanced Metrics in the DS10Codes and DS15Codes datasets. . . . .	41
A.1	Government Function and Sub-Function Codes for Commitment Notes . . . . .	51



# List of Acronyms

AI	Artificial Intelligence.
ANN	Artificial Neural Network.
BERT	Bidirectional Encoder Representations from Transformers.
CPF	Social Security Number.
CRISP-DM	Cross-Industry Standard Process for Data Mining.
DCGAS	Dutch Central Government Audit Service.
DSPA	Data Science Process Alliance.
FN	False Negative.
FP	False Positive.
GDPR	General Data Protection Regulation.
IBGE	Brazilian Institute of Geography and Statistics.
IFMA	Federal Institute of Maranhão.
KNN	K-Nearest Neighbors.
Labcontas	Control Information Laboratory.
LGPD	General Data Protection Law.
LRF	Fiscal Responsibility Law.
ML	Machine Learning.
NLP	Natural Language Processing.
OCR	Optical Character Recognition.
OCSVM	One Class Support Vector Machine.
PGE/RJ	State Prosecutor's Office of Rio de Janeiro.
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
RF	Random Forest.
SGD	Stochastic Gradient Descent.

SICONV	System of Management of Agreements and Transfer Contracts.
SINC-Fiscal	System for Control of Fiscal Information.
SIOFINET	Sistema de Programação e Execução Orçamentária e Financeira.
SVM	Support Vector Machine.
TCE-SP	Court of Accounts of the State of São Paulo.
TCE-GO	Court of Accounts of the State of Goiás.
TCE-MA	Court of Accounts of the State of Maranhão.
TCU	Federal Court of Accounts.
TL	Transfer Learning.
TN	True Negative.
TP	True Positive.
UF	Federative Unit.
VRAM	Video Random-Access Memory.
XGBoost	Extreme Gradient Boosting.
XML	eXtensible Markup Language.

# Chapter 1

## Introduction

This chapter is dedicated to present the context and motivation of this dissertation proposal. It encompasses the problem addressed, followed by the proposed approach, objectives, and research questions that have been defined, along with scientific contributions. Additionally, it outlines the document's structure.

### 1.1 Context and Motivation

Government audit refers to the systematic examination and evaluation of public sector organizations or entities, serving a critical function in maintaining the integrity, transparency, and accountability of public finances [1]. Multiple types of audit are conducted within this scope to evaluate diverse aspects of financial management and regulatory compliance [2].

One essential type is the financial audit, where financial statements play a central role in government audit processes, offering key insights into the financial health and performance of public entities [3]. By analyzing these statements, auditors aim at identifying irregularities, assess financial risks, and verify compliance with accounting standards and regulatory obligations [4].

In Brazil, according to data from the Brazilian Institute of Geography and Statistics, which is abbreviated IBGE, there are 5.570 municipalities, of which 217 are located in the state of Maranhão [5]. All municipalities in the national territory are managed by elected public officials through popular elections, and these officials act as heads of the municipal executive power, primarily responsible for managing the city's interests [6].

Among the responsibilities of public officials is the obligation to report expenditure of taxpayers' money to oversight bodies throughout the year, a duty established in the Federal Constitution (arts. 70 and 71). This involves providing information about their administration, with the expectation of understanding the results achieved in the allocation of financial resources in relation to the goals set by public authorities [7]. This process aims at ensuring not only transparency in the mandate but also self-evaluation [8].

In this context, the main guideline is the Fiscal Responsibility Law - LRF (Complementary Law No. 101/2000 [9]), which establishes norms and regulations for the public financial accountability process. It states that any situation leading to non-compliance with the rules can be classified as a crime, resulting in fiscal and criminal penalties as stipulated in the Law of Administrative Improbability (Law No. 8,429/1992) [10] and Decree-Law No. 201/1967 [11].

Unfortunately, a large number of public officials are facing legal proceedings and convictions. According to the administrative control body of the federal government, the Federal Court of Accounts (TCU) [12], in 2023 there were more than 6,800 condemned public officials in Brazil, ineligible for re-election, distributed among its five macro-regions<sup>1</sup>. As illustrated in the Figure 1.1, the Brazil's Northeast is the region with the highest number of public officials with disapproved accounts (39.5%), followed by the Southeast region with a total of (23.7%) [12].

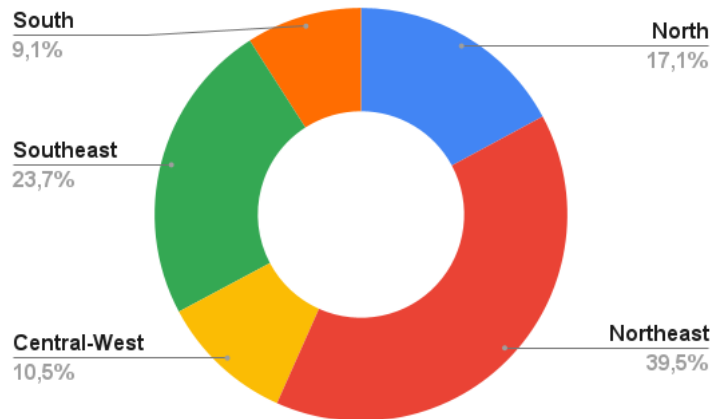


Figure 1.1: Percentage of disapproved accounts by Brazilian macro-region in 2023.

The number of disapproved accounts has been increasing each year [14], and looking at the national ranking of public officials convicted by Brazilian Federative Unit (UF)<sup>2</sup> in 2023, Maranhão (MA) leads with a total of 1,290 disapproved financial statements, followed by the states of São Paulo (SP) and Rio de Janeiro (RJ) with 966 and 719 disapprovals respectively. Thus, MA has 33% more disapproved statements when compared to the second place [12], as shown in the Figure 1.2.

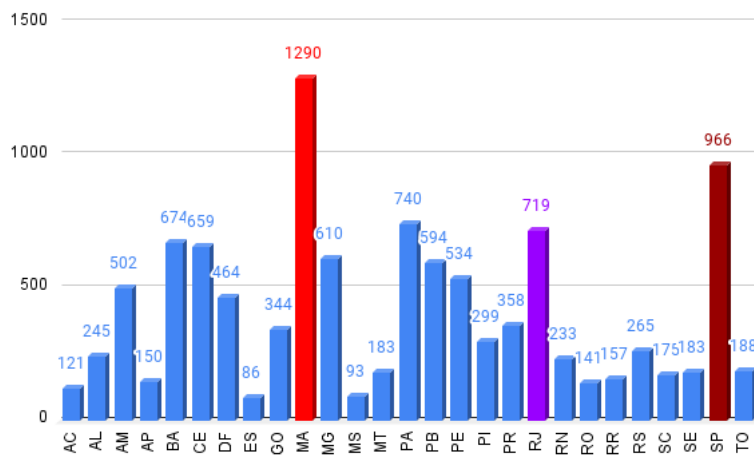


Figure 1.2: Number of accounts judged irregular by UF in 2023

<sup>1</sup>Brazil has five macro-regions: North, Northeast, Southeast, Central West and South [13]

<sup>2</sup>Brazil has 26 states and 1 federal district [15]

Given this scenario, it is crucial for both civil society and public officials that the management of financial resources and the corresponding preparation of their financial statements are carried out correctly. Therefore, these officials often hire accountants to ensure the proper management of resources and their presentation during the accountability process to oversight bodies, such as the Court of Accounts of the State of Maranhão (TCE-MA), responsible for evaluating the application of public resources in the state [16].

Furthermore, in the context of the state of Maranhão, until the completion of this project, there are currently no regional tools using Artificial Intelligence (AI) techniques to assist these accountants in the analysis and verification of the data that make up financial statements reports.

Considering that the construction of such reports requires a great deal of attention, as they involve on average 20 different types of documents [17], and any oversight can directly interfere with the audit of these accounts, the application of Machine Learning (ML) algorithms, specifically Natural Language Processing (NLP) techniques, in the state of Maranhão has been studied. Carmo and colleagues [18] present three spaces in the TCE-MA where the presence of solutions with ML models would be beneficial, among them the audit of public accounts, with the objective of identifying, mainly, irregularities in financial statements, a field with its first results applying Transfer Learning (TL) as solution, as shown in the study by Coelho and colleagues [19].

This dissertation aims to apply ML, specifically TL, to a sample of the documents created by accountants and sent to TCE-MA, with the goal of assisting in the evaluation of the data and contributing to the reduction of errors in the public financial statements of the state of Maranhão. The application of ML in this context has the potential to improve the process by providing greater agility and helping to minimize errors, as algorithms can be trained to analyze and verify data, alerting about possible violations to ensure compliance with relevant norms, all with increased productivity, precision, and reduced costs [20], particularly the TL models, which are pre-trained on vast datasets, capture deep contextual and semantic relationships in language, allowing them to effectively adapt to classification tasks [21]. This capability gives these models the ability to classify financial statements, identify irregularities, and understand complex regulatory language with high accuracy.

## 1.2 Problem Statement

In Brazil, several documents make up the financial statement, among them those that refer to the execution of public expenditure. This involves the use of public funds collected through taxes or other sources to subsidize actions, services, and public investments [22].

According to Law No. 4,320/1964 [23], the execution of public expenditure must be carried out in three phases: commitment, the stage where money is reserved for the payment of goods or services; liquidation, the stage responsible for verifying if the terms of contract that were agreed upon are being complied; payment, as the name suggests, is the transfer of funds to the seller or service provider contracted, given that the previous stages were correctly fulfilled [22, 23].

In Maranhão, the submission of these data related to the execution of public expenditure is subject of evaluation every two months and generally done through the System for Control of Fiscal Information (SINC-Fiscal)<sup>3</sup>, developed by the Information Technology Department

---

<sup>3</sup>Available in: <https://apps.tce.ma.gov.br/sinc-site/fiscal>

of TCE-MA. The system has undergone constant updates and currently receives such data in XML (eXtensible Markup Language) file format in order to facilitate the generation of reports [17].

However, even with the creation of systems like SINC-Fiscal, which incorporate tools to simplify the submission of information to TCE-MA, the preparation of this financial statement presents challenges of complexity and bureaucracy. This is due to the need to classify numerous fields of documents according to the standards established by Brazilian regulations, as it involves a text classification problem with multiple categories.

Therefore, this dissertation proposes to investigate the impact of applying ML algorithms on those text classification, since this is a Multiclass text classification problem.

### 1.3 Objectives and Research Questions

The main goal of this dissertation is the development of a solution to help the accountants on the preparation of government financial statements in the state of Maranhão, Brazil, to help in reducing the number of disapproved accounts. To achieve this purpose, five more Objectives (OB) were defined:

- **OB1:** Investigate the current state of the art in ML applied on Government audit with focus on financial statements;
- **OB2:** Develop an ML model able to make multiclass text classification applied in the context of the government financial statement in the state of Maranhão;
- **OB3:** Test the model developed using real data to adjust the hyperparameters;
- **OB4:** Validate the model with accountants experienced with government financial statements in the state of Maranhão;
- **OB5:** Make available a ML solution which helps to reduce the number of disapproved accounts in Maranhão.

To achieve the established objectives and guide the research which will fundament this dissertation, two Research Questions(RQ) were carefully defined:

- **RQ1:** How has ML been involved in the Government audit with focus on financial statements?
- **RQ2:** What are the ML methods currently employed in the Government audit with focus on financial statements?

### 1.4 Scientific Contributions

This study is part of the research initiatives conducted at the Innovation Laboratory of the Federal Institute of Maranhão (IFMA). The project has already produced initial results, showcasing the potential of Transfer Learning (TL) models, specifically BERT Multilingual, for enhancing the classification accuracy and automation of financial commitment notes in government audit processes.

- **COELHO, H. G.**, OLIVEIRA, L. V., SILVA, L. F. M. S., NEPOMUCENO, R. S. (2022). SOFTWARE PARA CONTABILIDADE DE ÓRGÃOS PÚBLICOS AUXILIADO POR APRENDIZADO DE MÁQUINA. In: Anais do Universo IFMA 2022. A, 2022. Available in: universolFMA2022/anais/590033[19].

The latest results were also presented in a Conference and will be indexed.

- **COELHO, H. G.**, MARREIROS, G., SILVA, L. F. M. S., (2024). Machine Learning applied in Government Audit with focus on financial statement: A Systematic Literature Review. Accepted to Practical Applications of Agents, Multi-Agent Systems (PAAMS) Conference, presented and to be indexed.

## 1.5 Document Structure

The present document is organized into multiple chapters, each covering key aspects of the research in order to enhance its understanding.

The current chapter, Chapter 1 - Introduction, provides an overview of the context and motivation behind this dissertation, outlining the problem, objectives, defined research questions, and scientific contributions.

The Chapter 2 is dedicated to the State-of-the-art, which is a detailed exploration of the integration of ML within Government Auditing focused on financial statements, in order to answer the proposed research questions of this study.

Chapter 3 - Methodology describes the CRISP-DM approach, the methodology adopted, along with the technologies and the evaluations techniques selected for this study.

Chapter 4 - Data Overview offers a detailed description of the study's context and its specificity within the state of Maranhão, including data characteristics and relevance.

The Chapter 5 - Proposed Solution and Evaluation is divided in three sections, covering the data preparation process, the model fine-tuning, and the evaluation of results.

Chapter 6 - Conclusions, summarizes the key findings and the limitations of the study and the objectives achieved. Suggestions for future research and potential advancements in applying ML models for public sector accountability are also presented.

Finally, the Appendix: Supporting materials, including supplementary tables and information are provided to facilitate understanding.



## Chapter 2

# State-of-the-art

The present chapter is dedicated to examining the current state of art in the integration of ML within Government Auditing, with a specific focus on financial statements. Subsequently, the aim is to furnish the necessary information to address the formulated research questions. In the subsequent sections, the formulated methodology is presented, followed by a delve into the findings, and engaging in a comprehensive discussion.

### 2.1 Research Methodology

To answer the research questions which was defined on the chapter 1.3 and help to better understand this field of study, a literature review was conducted and managed inspired by the latest version of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology, 2020 [24]. Created to improve the quality of medical literature, PRISMA has become a widely employed methodology in systematic reviews across various scientific fields [24]. Its popularity is attributed to its ability to produce transparent reviews with a well-defined methodology, making replication straightforward [25].

After an initial and careful exploration of the field, some keywords were chosen. Considering this dissertation proposed a solution to Brazil, a country which has Portuguese as its native language, the search terms were used both in English and Portuguese. These keywords were combined with boolean operators (i.e. AND, OR, NOT used in search queries to connect or modify the relationship between terms. AND narrows the search, OR broadens it, and NOT excludes specific terms [26]) to formulate a customized Research Query and get more precise results.

In both situations, conducting the search in English or Portuguese aimed to utilize the most accurate expressions to represent the focus of the search, namely government audits with a focus on financial auditing. This explains why the Portuguese Research Query contains more expressions than the English Research Query, as presented in Table 2.1.

Table 2.1: Final Research Queries

Language	Research Query
English	((("government audit" AND "financial audit") OR "government accounts audit") AND ("natural language processing" OR "machine learning"))
Portuguese	((("prestação de contas públicas" OR "fiscalização orçamentária" OR "fiscalização de gastos públicos" OR "auditoria de contas públicas" OR "auditoria financeira") AND ("processamento de linguagem natural" OR "machine learning" OR "aprendizado de máquina"))

In terms of databases, several renowned search sources, including IEEE Xplore, ACM Digital Library, ArXiv, and ScienceDirect, were selected for use in this dissertation. However, given the dissertation's specific focus on financial statements from government audits, initial searches using these databases did not return valuable findings, then Google Scholar was selected as the single search base as it indexes other sources, resulting in significant findings. Google Scholar emerged as a key tool for accessing a multitude of databases relevant to the dissertation's targeted subject.

Moreover, some inclusion and exclusion criteria were defined to conduct the review. The inclusion criteria (IC), outlined in Table 2.2, aimed to identify studies with characteristics that directly address the research questions. Key characteristics included the language of the paper (IC1), the date of publication (IC2), and whether the article presents the implementation or application of a ML approach in financial statements from government auditing (IC3).

Table 2.2: Inclusion Criteria

<b>Inclusion Criteria</b>
<b>IC1:</b> Available in English or Portuguese language
<b>IC2:</b> Published from 2018 onwards
<b>IC3:</b> Introduced or applied a ML approach for government audit with focus in financial statement

Conversely, the exclusion criteria (EC), detailed in Table 2.3, were designed to eliminate studies not aligned with the primary focus of this project. Studies were excluded if they were duplicates (EC1), had restricted access due to paywalls (EC2), were surveys or literature reviews (EC3), focused on company auditing rather than government auditing (EC4), involved government auditing but did not focus on financial statements (EC5), or did not employ any ML techniques (EC6).

Table 2.3: Exclusion Criteria

<b>Exclusion Criteria</b>
<b>EC1:</b> Duplicated publication
<b>EC2:</b> Access restricted due to paywall
<b>EC3:</b> Survey or review
<b>EC4:</b> Focus on company auditing
<b>EC5:</b> Does not focus on financial statement
<b>EC6:</b> Does not use ML as a research technique

Considering the scope of the review, both sets of criteria were tailored to maximize the quality and relevance of the results, facilitating the subsequent discussions.

## 2.2 Research Findings

Following the outlined methodology, a total of 98 studies were initially obtained. The English research query yielded 31 results, while the Portuguese query produced 67 results. Subsequent to the application of exclusion criteria, it was aimed at eliminating duplicate publications and identifying studies irrelevant to the dissertation theme; 42 studies advanced to the second screening phase, that involved a detailed examination of the main aspects of

## 2.2. Research Findings

each study to determine its potential to address the research questions. Figure 2.1 presents the adapted PRISMA diagram with the detailed process, culminating in the selection of 9 studies for inclusion in the following discussion.

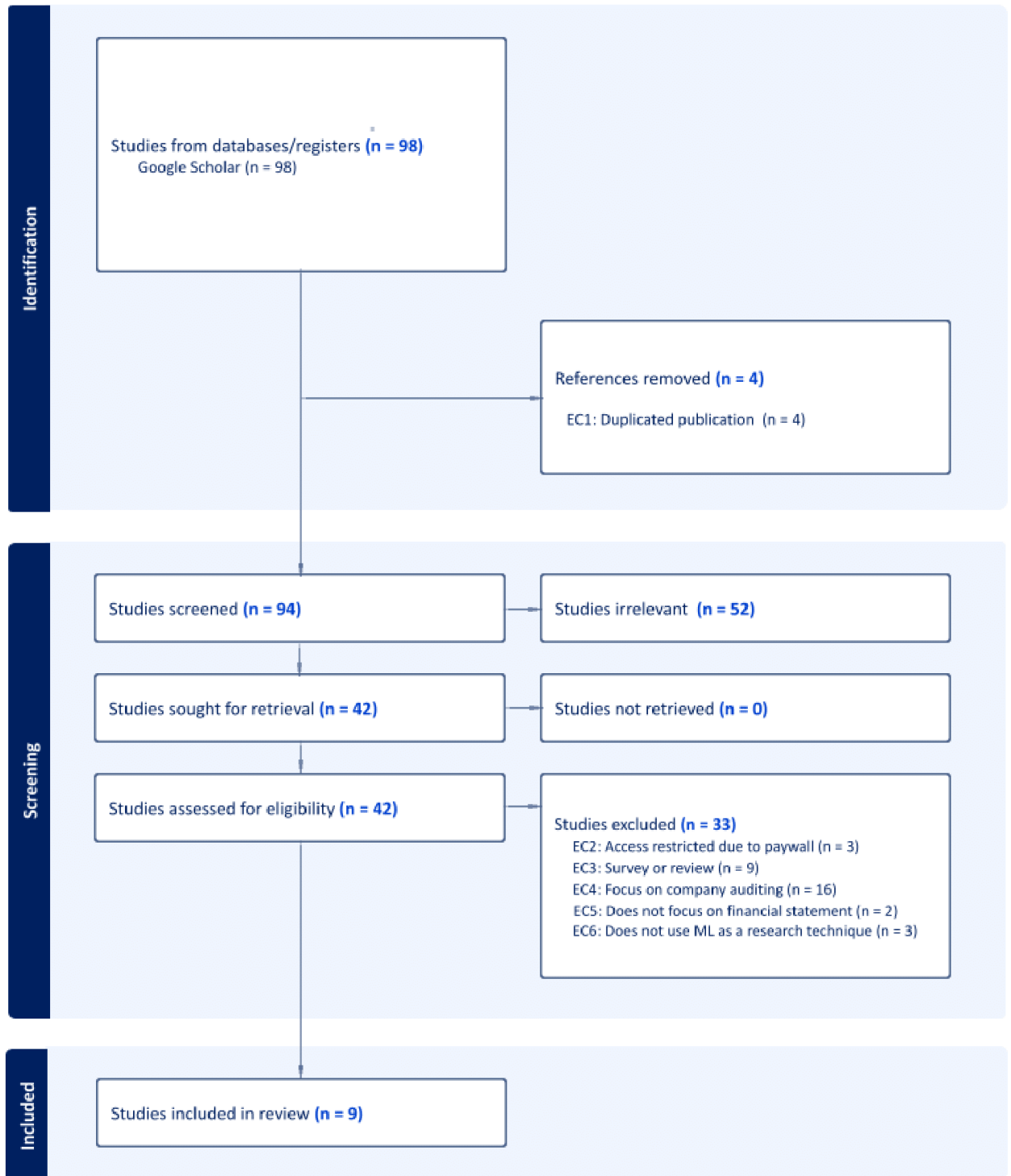


Figure 2.1: Adapted PRISMA diagram

Table 2.4 presents a summary of the selected articles with important information as authors, title and year of publication.

Table 2.4: Selected studies for the review

Author	Title	Year
K. Mabelane et al. [27]	An Analysis of Local Government Financial Statement Audit Outcomes in a Developing Economy Using Machine Learning	2022
L. H. Hamelers [28]	Detecting and explaining potential financial fraud cases in invoice data with Machine Learning	2021
A. O. Andrade et al. [29]	Prediction and Visualisation of SICONV Project Profiles Using Machine Learning	2022
J. J. Souza [30]	A necessidade de regulação do uso da inteligência artificial nas ações de controle do tribunal de contas da união	2021
R. O. C. Hildebrand [31]	A experiência do Tribunal de Contas da União com inteligência artificial	2021
T. R. Guimarães et al. [32]	Análise dos fatores que contribuem para a recuperação dos créditos de ICMS inscritos em dívida ativa no estado do Rio de Janeiro	2023
T. S. Cruz et al. [33]	Automatizando a fiscalização de gastos públicos por meio da classificação automática de empenhos utilizando aprendizado de máquina	2021
G. D. Lins et al. [34]	Parametrização de despesas municipais e detecção de anomalias	2021
P. N. S. Carmo et al. [18]	Aprendizado de máquina em ações de controle no Tribunal de Contas do estado do Maranhão	2018

## 2.3 Discussion

In one of his renowned works on ML, Mitchell [35] defines the primary goal of this branch of Artificial Intelligence (AI) as the development of computer programs that autonomously enhance their performance by successfully addressing problems derived from past experiences. In this process many things can contribute to having the best results, some of them are the quantity and quality of the data.

Examining the selected articles, a common theme emerges as ML is employed in diverse government contexts, unified by a substantial reliance on data. Regardless of the country, the prevailing aim among authors is to utilize ML tools to assist government auditors in identifying financial statement fraud.

For example, Mabelane and colleagues [27] conducted a study using ML to analyze and predict audit outcomes for South African municipalities. After an extensive literature review, the authors selected three prominent algorithms—Decision Trees, Artificial Neural Network (ANN), and Logistic Regression. The ANN demonstrated superior results, achieving a Precision-Recall curve of 0.7074, signifying its suitability for predicting audit opinions in South African municipalities.

### 2.3. Discussion

---

Hamelers [28], in collaboration with the Dutch Central Government Audit Service (DCGAS), formulated a study to employ Isolation Forest, Local Outlier Factor, and One Class Support Vector Machine (OCSVM) algorithms to detect fraudulent invoices. All algorithms, categorized as unsupervised ML, were tested, and the Isolation Forest algorithm successfully identified 72% of the invoices, aligning with the project's primary objective.

These examples illustrate the versatility of ML, showcasing its effectiveness in achieving established goals securely and with favorable outcomes. ML, as a key method for imparting knowledge to computers [36], has found application in various governmental sectors globally, including the Brazilian government.

Andrade and colleagues [29] exemplify this in their study where the K-Nearest Neighbors (KNN) algorithm achieved the highest accuracy ( $0.991 \pm 0.002$ ) among 25 models tested. The goal of this study was applying unsupervised ML to make predictions and facilitate the data visualization from a system called SICONV (System of Management of Agreements and Transfer Contracts), a Brazilian Government initiative that records the entire life cycle of agreements.

Furthermore, another highlighted fact in the Brazilian Government is the use of robots by TCU. Sofia, Mônica, Adele, Alice, Ágata, Carina and Zello (chatbot) are acronyms for robots created or personalized by the TCU to help auditors in their analysis. All of these robots are part of Labcontas (Control Information Laboratory) which currently combines 80 databases. Every robot was thought to be used in a specific part of the Government auditing process and to deal with sensitivity information, that's because the organization does not publish many details about the development process. However, observing the services provided by those robots, it is safe to say they can recognize concepts of NLP (Natural Language Processing), text mining, text classification and ML in their types of learning such as active, unsupervised and supervised [30, 31, 37–40].

The technological advancement extends beyond the federal government, with some Brazilian states incorporating ML in their auditing activities. One example found was the study conducted by Guimarães [32] which had as main objective to contribute to improving the management of the State of Rio de Janeiro's active debt. To achieve its goal, ML was applied to produce the classification of the probability of recovering the registered credits, allowing the State Prosecutor's Office of Rio de Janeiro (PGE/RJ) to allocate recovery efforts more efficiently, while recognizing these credits more objectively in accounting terms. It used a ML model known as XGBoost (Extreme Gradient Boosting) which is a model derived from the decision tree model. According to the author, the algorithm uses tree classification to achieve better prediction results with greater operational efficiency, a great approach for the context of the study achieving a 98% success rate.

Another government body that has adopted ML models in its activities is the Court of Accounts. Cruz [33] conducted a study using data from the Court of Accounts of the State of Goiás (TCE-GO). The data, generated through the SIOFINET (Portuguese: *Sistema de Programação e Execução Orçamentária e Financeira*) system, contained all commitments made between January 2015 and April 2020. That data was selected with the idea to develop a ML model that automatically identifies public commitments that have been classified under an incorrect expenditure category. Support Vector Machine (SVM), Random Forest (RF), KNN and Stochastic Gradient Descent (SGD) were evaluated, and SVM produced the best results achieving a hit rate of 90% according to the Micro F1 evaluation metric.

Similarly, in the state of São Paulo, Lins [34] used data from TCE-SP (Court of Accounts of the State of São Paulo) to develop a study that had as main purpose applying ML methods for parameterizing municipal expenses and detecting anomalies, aiming at increasing the efficiency of the audit processes of municipal expenses. The authors tried to build the predictive model using data of 644 municipalities in the state of São Paulo over 10 years (2009 - 2019) and testing many algorithms, such as Gradient Booster, Regression Model, Random Forest Regressor, ANN and mathematical approaches to evaluate the success rate. After many tests and adjustments, the Random Forest Regressor model achieved a success rate greater than 90% according to the cross validation evaluation metric.

Although there is no official information regarding the adoption of AI in the state of Maranhão's auditing process, Carmo and colleagues [18] published a study that outlined potential scenarios where ML could be applied to aid auditors. This includes recognizing texts in unstructured documents, identifying suspect information in commitment documents, predicting companies prone to infractions, and identifying suspect companies involved in government projects. The highlighted areas encompass technologies like Optical Character Recognition (OCR), Text Mining, and others, demonstrating the potential for constructing supervised and unsupervised ML models.

## 2.4 Chapter Remarks

This chapter was idealized to investigate the current scenario of ML applied in Government Auditing, specifically with a focus on financial statements. It serves as the foundation for the development of this dissertation, employing a systematic review based on the PRISMA methodology. The design of the systematic review incorporates carefully chosen inclusion and exclusion criteria, along with research queries tailored for the present scenario. The selection of the PRISMA methodology aimed to provide comprehensive information that were essential for addressing the research questions formulated in Chapter 1.3 and directly answered below.

### **RQ1: How has ML been involved in Government audit with a focus on financial statements?**

ML has significantly shaped the landscape of Government audit, particularly in the examination of financial statements. Notably, ML has been extensively employed to enhance the effectiveness of audit processes across various contexts, both within and outside Brazil. Examples include studies in South African municipalities, the Dutch Central Government Audit Service (DCGAS), and the Brazilian government. In South Africa, ML was utilized to analyze and predict audit outcomes, with the ANN demonstrating notable success [27]. The DCGAS employed ML algorithms like Isolation Forest to detect fraud in invoices, showcasing the adaptability of ML in diverse government audit scenarios [28]. In Brazil, federal institutions such as the Court of Accounts and state-level entities in Goiás, São Paulo, and Rio de Janeiro have embraced ML for tasks ranging from predicting audit outcomes to efficiently managing active debt [29–34, 37]. Other states, such as Maranhão, already identify some spaces and activities in which ML can perfectly be adapted and used [18]. Those examples just show that the adoption of ML in government audit reflects a transformative trend, offering improved accuracy, efficiency, and decision-making capabilities.

### **RQ2: What are the ML methods currently employed in Government audits with a focus on financial statements?**

## 2.4. Chapter Remarks

Various ML methods have been employed in Government audits, demonstrating the versatility and applicability of these techniques. These approaches have been tested and applied in different audit contexts, showcasing their effectiveness in identifying fraud, predicting audit outcomes, and improving the efficiency of audit processes, as presented in Table 2.5.

Table 2.5: Overview of ML Approaches in Government Financial Audits

<b>Paper Reference</b>	<b>Context in Financial Audit Field</b>	<b>ML Approach</b>
[27]	ML was used to analyze and predict audit outcomes for South African municipalities	Decision Trees, ANN, and Logistic Regression, with the ANN presented the best results
[28]	Unsupervised ML algorithms were tested to find one capable of detecting fraudulent invoices from the Dutch Government.	Isolation Forest, Local Outlier Factor, and OCSVM, with Isolation Forest was able to identify fraudulent invoices according to the auditors
[29]	The project had as goal to apply unsupervised ML to make predictions and facilitate the data visualization from a system called SICONV from Brazilian Government	Among 25 techniques tested, the KNN algorithm was chosen with the best results
[30] and [31]	The articles presented various situations where the TCU-Brazil applied ML algorithms	Models from NLP, text mining, text classification, and various types of ML such as active, unsupervised, and supervised.
[32]	An ML model was implemented to contribute to the improvement of the management of the State of Rio de Janeiro's active debt in Brazil	An ML model called XGBoost
[33]	An ML model that automatically identifies public commitments misclassified under incorrect expenditure categories in TCE-Goiás, Brazil.	The algorithms SVM, RF, KNN, and SGD were evaluated, with SVM achieving the best results in this context
[34]	The study proposes using ML methods for parameterizing municipal expenses and detecting anomalies to increase the efficiency of the audit processes of municipal expenses in TCE - São Paulo, Brazil	Gradient Booster, Regression Model, Random Forest Regressor, ANN, and mathematical approaches to evaluate the success rate, where the Random Forest Regressor model achieved the highest success rate.
[18]	The article highlights several areas where ML could be applied within TCE - Maranhão	ML models used for OCR, text mining, text classification, and other ML approaches

The choice of algorithms often depends on the specific audit objectives and the nature of financial statements being analyzed. The success rates achieved, as evidenced in studies across multiple regions, underscore the potential of ML methods to contribute significantly to the advancement of government audit practices with a focus on financial statements.

In summary, this chapter not only explores the current state of ML in Government Auditing but also provides important insights to guide the subsequent steps of this dissertation. The

systematic review, based in one adaptation of PRISMA methodology, ensures a methodical and exhaustive exploration of the applications and methods of ML currently present in Government audit, with a special focus on financial statements.

# Chapter 3

## Methodology

This chapter aims to present a comprehensive overview of the methodology employed in this study. Starting with an outline of the selected approach, followed by the technologies and evaluation techniques selected, it details each phase of the process, explaining the rationale behind all choices.

### 3.1 Project Development

The methodology chosen to guide the development of this project was the CRISP-DM (Cross-Industry Standard Process for Data Mining). The approach was developed in the late 1990s by a group of experts in Data Mining Market [41] to provide a method for carrying out large data mining projects designed to be more cost-effective, reliable, repeatable, manageable, and efficient [42].

The CRISP-DM deals with the entire process of an AI model development, seeking to understand from the business perspective to the deployment of the solution [41, 43]. The methodology is organized in six phases well defined and interlinked:

1. **Business Understanding:** this initial step seeks to understand the business perspective, its objectives and requirements [41]. This knowledge should be converted into a data mining problem definition and goals to achieve in the project [44];
2. **Data Understanding:** this phase begins with initial data collection and involves activities to take a closer look at the data, identifying quality issues, uncovering initial insights, and detecting interesting subsets for forming hypotheses about hidden information [41, 45].
3. **Data Preparation:** also called “data munging” and usually is the longest part of the project. This stage is dedicated to prepare the final dataset, to be inputted in the modeling tools, from the initial raw data [41]. Depending the context of the project, data preparation generally includes: Merging datasets and/or records; Selecting a sample of the data; Aggregating records; Creating new attributes; Sorting data for modeling; Handling blank or missing values; Splitting into training and test datasets, and others [46].
4. **Modeling:** this step focus to apply the distinct chosen modeling techniques, by adjusting their parameters for optimal performance[47]. Since multiple techniques may address the same data mining problem and some require specific data formats, it is often necessary to revisit the data preparation phase [41].

5. **Evaluation:** this step is centered to determine which model most effectively addresses the business objectives, outlines subsequent actions and reviews the steps taken to create it [48]. A primary goal is to identify any significant business issues that may not have been adequately addressed, and by the end of this phase, a decision should be made regarding the utilization of the data mining results [41].
6. **Deployment:** the creation of the model typically does not conclude the project. The knowledge gained must be organized and presented in a way that the customer can effectively use it [42]. Depending on the requirements, the deployment phase may range from generating a straightforward report to implementing a complex, repeatable data mining process [41].

The implementation of those six phases are not rigid [41]. The Figure 3.1 provides a visual representation of the life cycle of a data mining project developed in this context. It is important to highlight that this reference model represents the cyclical nature of data mining, where the process continues even after a solution is deployed, each insight can inform and improve future processes [41, 49].

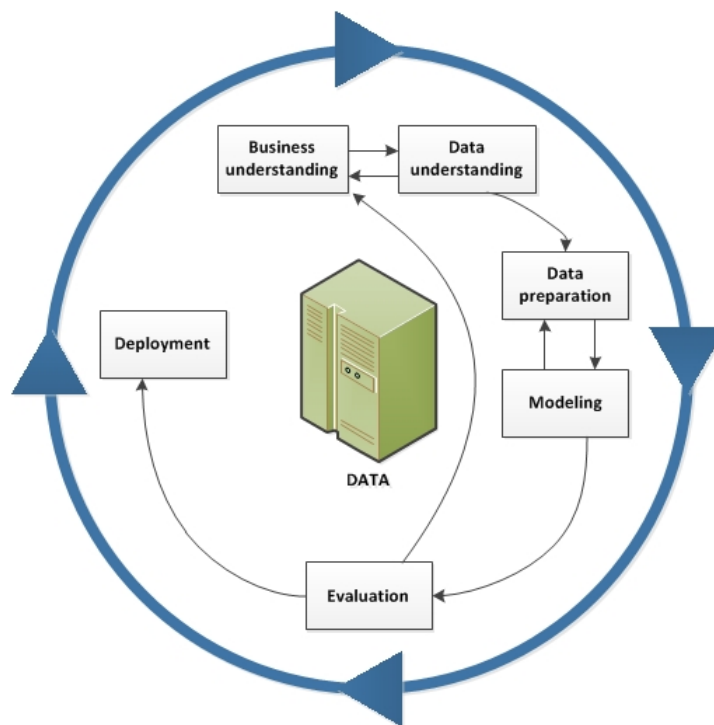


Figure 3.1: Life cycle of CRISP-DM; Image from [49]

Furthermore, the CRISP-DM is a very popular methodology adopted in the Data Science area [43]. According to the study conducted by DSPA (Data Science Process Alliance), a community dedicated to supporting the development of AI projects [50], nearly fifty percent of the interviewees answered that they use the CRISP-DM as a methodology in their projects, shown in Figure 3.2. The study also showed other approaches currently adopted like Scrum (a very popular framework for software development), Kanban (a visual workflow management technique focused on continuous delivery), “My Own” (referencing that the user prefers a custom methodology) and others.

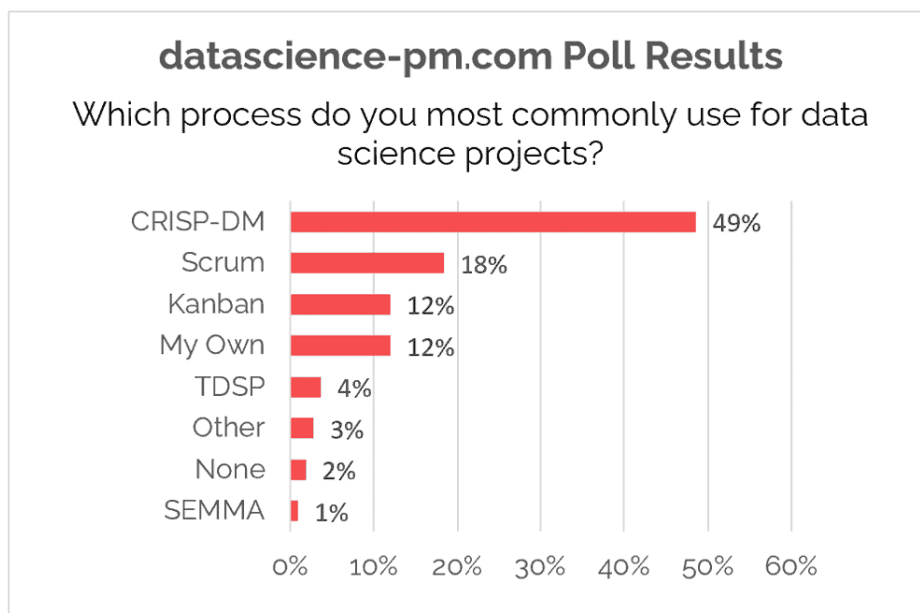


Figure 3.2: DSPA survey results; Image from [51]

Given this scenario, the CRISP-DM methodology is directly aligned with the proposed objectives of this study, since its phases range from the ideation of the model to its availability to the end user, in our case the accountants of the state of Maranhão in Brazil.

## 3.2 Technologies

Chapter 2, State-of-the-art, presented several ML approaches that are being applied in Government Auditing, aligned with the context, objectives and type of data adopted. Most of these studies used traditional ML methods, obtaining good results, but requiring a large dataset and robust hardware for model development. For this study, some of these ML methods were also considered, however, given the context and specific challenges of this study, since the data has a language originating from Brazilian public accounting, with many specific terms and expressions, and also seeking to achieve the best model performance with the data obtained, testing a new approach in the area, TL was chosen instead of traditional approaches.

Currently, TL is one of the most used methods in NLP, because it allows leveraging pre-trained models, which can provide robust starting points, especially when working with complex data structures or limited labeled data [21, 52]. Those features meet the need of this study, offering efficiency and potentially greater accuracy compared to developing a model from scratch, as well as eliminating the necessity of building a specific language model just for this task [52].

Considering that the present dissertation has been designed for a Brazilian context, focused on the State of Maranhão, BERTimbau [53], a Brazilian Portuguese-language adaptation of the BERT (Bidirectional Encoder Representations from Transformers) [54] in its “base” version with 12 Layers and 110M Params, was selected for this study due to its effectiveness in handling nuanced linguistic features specific to Brazilian Portuguese [53, 55]. BERT and their numerous variants are based on Transformer Encoder architecture, which confers

the ability to recognize and preserve the sentence context, as well as capture long-range dependencies within text through self-attention mechanisms [53, 56], an important feature for this study, given the specific context, language and data limitations, such as the number of examples per code, per year (the dataset having examples only from the year 2023) and others. This framework enables models like BERTimbau to learn contextual relationships in language more effectively than traditional methods [55].

Additionally, it was decided to also use the BERTimbau Tokenizer, which breaks the sentences into subword units, allowing it to handle a rich variety of words in the language and represent even rare or complex words accurately, preserving linguistic nuances crucial for fine-grained language tasks [53]. An overview of the model and their phases can be viewed in Figure 3.3.

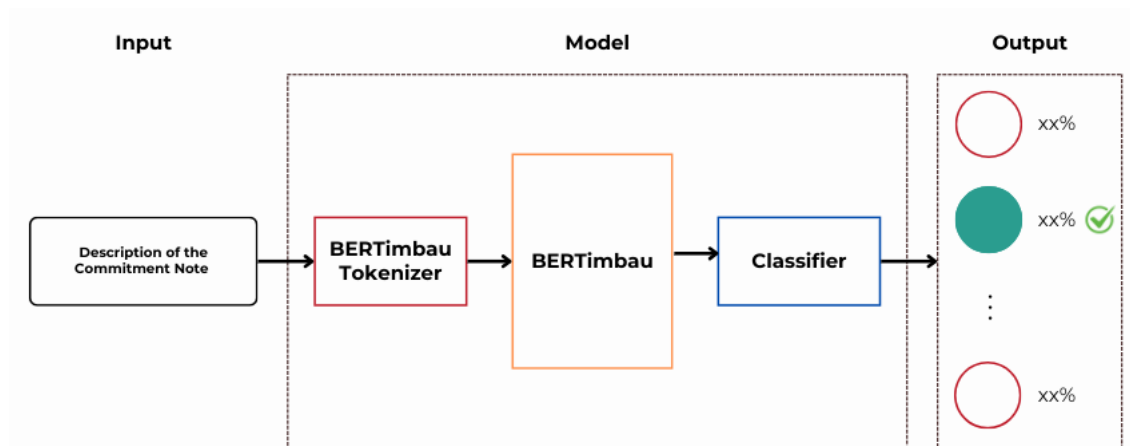


Figure 3.3: Overview of the proposed model and its phases

To support the development process, aiming to facilitate this phase, Python<sup>1</sup> was chosen as the programming language, once it is one of the most used programming languages in ML projects, due to its widespread by the community and its extensive support [57]. Key Python libraries were employed, including NumPy<sup>2</sup> and Pandas<sup>3</sup> for data management, Matplotlib<sup>4</sup> for data visualization, and Scikit-learn[58] and PyTorch<sup>5</sup> for model development. Additionally, the Hugging Face Transformers framework was integrated to provide seamless access to the BERTimbau model<sup>6</sup>.

Another tool adopted was the Google Colaboratory platform, also know as Colab, in order to improve development and testing phases, as it provides access to GPUs and TPUs, which are particularly beneficial for developing ML models, especially through TL. For this study Colab's default GPU - a NVIDIA Tesla K80 with 15GB of RAM and 12GB of VRAM (Video Random-Access Memory) - was utilized.

<sup>1</sup>More info in: <https://www.python.org/>

<sup>2</sup>More info in: <https://numpy.org/>

<sup>3</sup>More info in: <https://pandas.pydata.org/>

<sup>4</sup>More info in: <https://matplotlib.org/>

<sup>5</sup>More info in: <https://pytorch.org/>

<sup>6</sup>More info in: <https://github.com/neuralmind-ai/portuguese-bert>

### 3.3 Evaluation

In every teaching procedure, it is necessary to observe the process to identify whether the methods used are generating the expected results. In the case of ML, this is an essential task. Therefore, to measure the level of success of the proposed model, six techniques were chosen, which will be defined below according to studies [59] and [60].

It is important to emphasize that, in computational terms, model predictions can be categorized into four types: True Positive (TP), False Negative (FN), True Negative (TN), and False Positive (FP). These classifications form the basis of the confusion matrix, which compares the actual class with the predicted outcome. Table 3.1 exemplifies a sample confusion matrix.

Table 3.1: Confusion Matrix Structure

Confusion Matrix		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

It is important to highlight that the confusion matrix is not limited to situations in which the classification has only two possible classes, and can also be applied to multiclass problems. Additionally, the information contained in the confusion matrix provides support for the calculation of other metrics, such as accuracy, recall, precision and F1-Score. These metrics, like the confusion matrix, are not restricted by the number of classes and can be effectively computed for multiclass classification tasks. Therefore, all of these metrics were also adopted in this study and are described below.

- **Accuracy**

Accuracy is responsible for computing the precision of predictions, that is, the number of correct predictions made by the model, regardless of the class. It is calculated by dividing the total number of correct predictions by the total number of items, as shown in Equation 3.1.

$$Accuracy = \frac{(TP + TN)}{(TN + FP + FN + TP)} \quad (3.1)$$

- **Recall**

Recall or True Positive Rate, is the method that represents the proportion of positive cases that were correctly classified, as shown in Equation 3.2.

$$Recall = \frac{TP}{(TP + FN)} \quad (3.2)$$

- **Precision**

Precision, also known as positive predictive value, seeks to evaluate the number of correct classifications (TP) over all values that should be true (TP + FP), as can be seen in Equation 3.3.

$$Precision = \frac{TP}{(TP + FP)} \quad (3.3)$$

- **F1-Score**

F1 score or F-Measure, is defined by the harmonic mean of precision and sensitivity, and obtained based on Equation 3.4.

$$F1 = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (3.4)$$

In summary, to properly evaluate the progress of software that uses ML, it is important to observe from several angles to ensure that the results obtained are consistent with the stipulated objectives. The metrics presented here can be used together to observe the progress of building an ML model, an idea that will be applied in the present research.

### 3.4 Chapter Remarks

In this chapter the methodology adopted in this research was defined with a comprehensive explanation of each step involved. The study adopts the CRISP-DM as methodology, once it is a flexible and widely accepted framework in industry, making it well-suited for guiding the development of this dissertation. The knowledge of each phase of CRISP-DM, from data understanding to deployment, will support the process of development of this dissertation.

In addition, the chosen model, BERTimbau, and its associated tokenizer were introduced as the primary tools for processing and analyzing Brazilian Portuguese text. These technologies in combination with the programming language Python and its libraries, including NumPy, Pandas, and PyTorch, were responsible to support the development of the solution proposed in this dissertation.

Additionally, the evaluation metrics were also discussed, as they are crucial for assessing the effectiveness of the model and determining whether the study is achieving the proposed objectives. The metrics provide a foundation for measuring model performance and identifying areas for improvement. All of the information presented above serve as a guiding framework for the following chapters, supporting a rigorous and well-documented research process.

# Chapter 4

## Data Overview

This chapter presents the development of the first two steps, named Business and Data Understanding, of the CRISP-DM methodology adopted in this research. Both phases of the methodology were adapted to provide robust knowledge about the entire context of the dissertation and its particularities.

### 4.1 Business Understanding

The Brazilian Government Auditing, especially in the state of Maranhão, involves many scenarios, starting from the organization of public financial data by mayors and accountants to its analysis by TCE-MA auditors. Focusing on the elaboration of the financial statements, the professionals responsible to conduct this task have to deal with many phases as shown in the diagram presented in Figure 4.1, which includes the use of systems to generate reports without aid from intelligent tools.

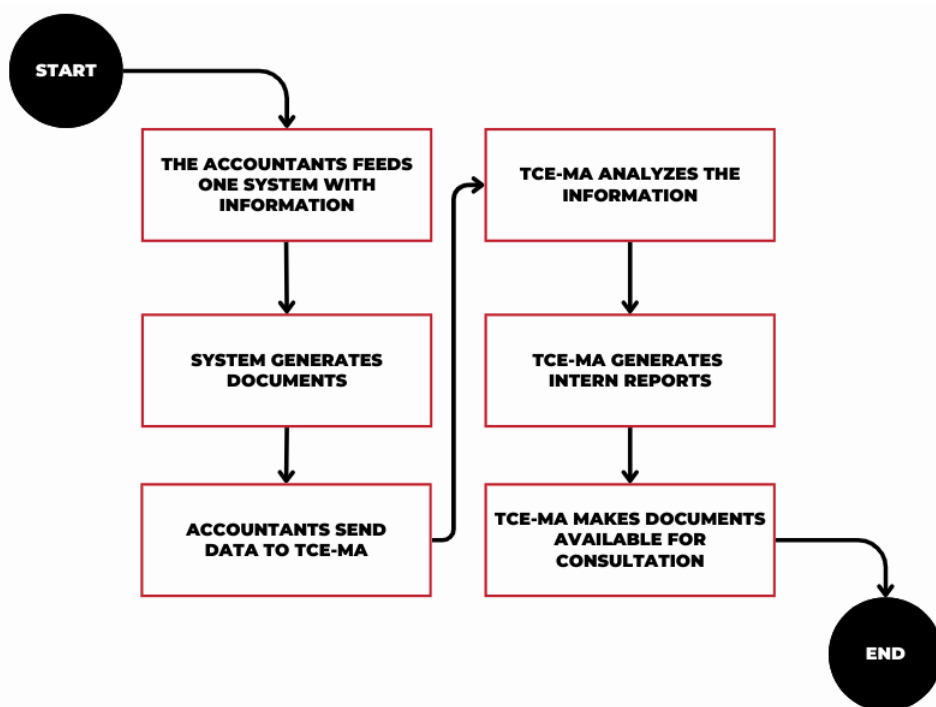


Figure 4.1: Workflow in Maranhão's Government Auditing by TCE-MA

The diagram above was constructed based on the reports of two experiment accountants in Government Auditing of Maranhão. In an informal conversation to familiarize with the theme of this dissertation, they mentioned that to elaborate these documents to TCE-MA sometimes it is necessary to memorize many rules, which may require years of experience and a constant risk of making mistakes. In their conception, a beginner accountant allegedly would not produce reliable information for the TCE-MA.

Furthermore, the submission to the competent bodies may happen yearly, every two or three months, or depending on the type of government accountability. As previously discussed in the sections 1.1 and 1.2, the SINC-Fiscal is one of the systems of TCE-MA responsible for receiving part of the documents that composes the Government Auditing in the state of Maranhão. As the amount of documents received per year in this system is high, after many updates and improvements, this tool receives the data in XML format using a specific layout which provides support for the auditing and subsequent production of reports [61].

All of these actions are results from the Strategic Planning designed by the TCE-MA for an 8-year period (2019 to 2027) which established many steps to refine the agency's performance in the state [62]. A key goal of the Information Technology Department is to improve its management, thus TCE-MA plans to insert in the future many types of technology in its operations including AI, Big Data and others, where their degree of success can be measured by its effectiveness [62].

On observing TCE-MA and the accountants involved with Government Auditing, this dissertation identified the opportunity to explore the application of ML in the area, but with focus on helping accountants produce the financial statements for the TCE-MA. Knowing that there are no current systems in the area using AI for this purpose, the Innovation Laboratory of IFMA Caxias Campus is conducting a project to produce an updated system to this area called ConnTec<sup>1</sup>. The main intention of this project is to integrate the AI model in this system to analyze the information before submission to TCE-MA, as illustrated in Figure 4.2.

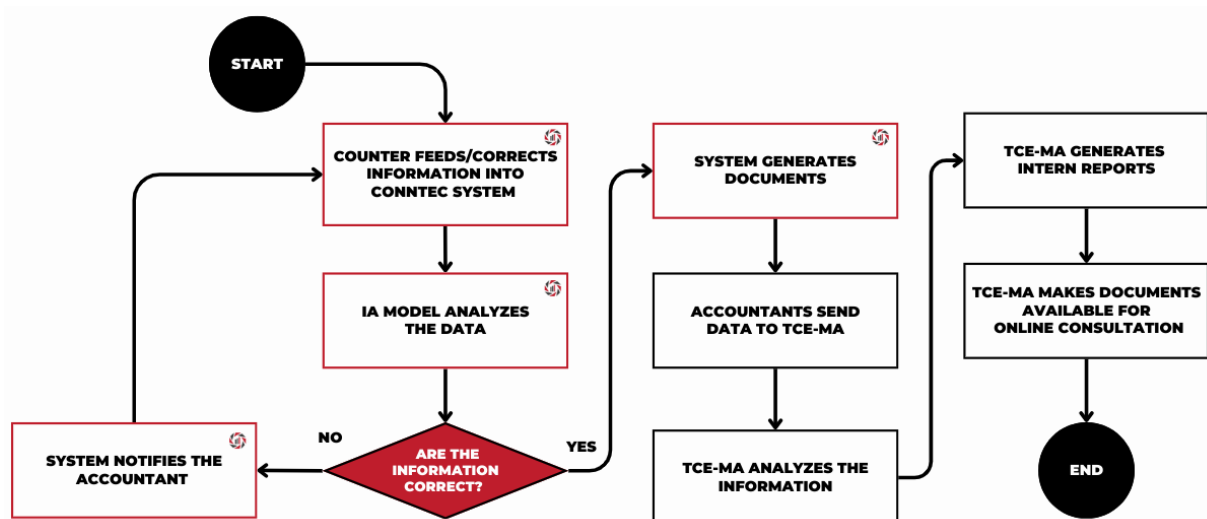


Figure 4.2: Diagram of AI integrated in the ConnTec system

<sup>1</sup>More info in <https://conntec.net/>

Once the ML models can help identify errors in the data based on a history, the combination of the model produced in this study with the ConnTec system will help accountants from Maranhão to produce their documents more accurately.

## 4.2 Data Understanding

The following subsections details the steps taken to achieve a deep understanding of the data adopted in this study. Each phase discusses a different aspect of the data, from its ideation and extraction to exploration.

### 4.2.1 Data Adopted

The SINC-Fiscal currently comprises an external system of TCE-MA specialized in receiving information about budgetary, financial and asset management of public bodies and entities subject to inspection by TCE-MA [17], among them the city halls of Maranhão. This system is accessed by mayors or accountants responsible for the submission of reports.

Many documents compose this part of the Government Auditing, such as commitments notes, settlements, expenses, payments, and others to support the integration of these documents. All of those documents must follow a predetermined structure designed by TCE-MA and after the auditing is available to be accessed<sup>2</sup>.

Given this context and after evaluating the documents along with two experienced accountants, the commitment notes were selected to be the focus of this work. They are part of the public expenditure process and are the most suitable document for some validation by AI since their correct answer depends on many possibilities predetermined by the laws of the Brazilian Government.

The decision to work with this type of document was made based on the public expenditure execution structure, as discussed in chapter 1.2. It is composed of three stages: commitment, liquidation and payment [23]. The commitment is the first stage that starts the process and consequently is the first to be evaluated. If there is any pending issue this may result in restrictions after a certain period, which may compromise essential services that would no longer be offered to the population [63].

### 4.2.2 Data Extraction

Currently, SINC-Fiscal has an API which allows the execution of many actions of this type of Government Auditing, among them access to the data submitted in the system in the year of 2023. This restriction is due to the fact that in previous years the submission was done differently so that there was no integration between the databases.

Due to this situation, the extraction of the data from the TCE-MA database was conducted with the help of the accountants who are contributing to this dissertation, and provided the TCE-MA authorization to access the data. Then, the acquired dataset was stored in Google Drive to facilitate the analysis which were made using the Google Collab Platform.

---

<sup>2</sup>more information in [17]

### 4.2.3 Data Exploration

As result of the Data Extraction phase, 717.360 examples of Commitment Notes, from all 217 municipalities from the State of Maranhão in 2023, were composing the initial dataset for this project. In total, 27 attributes constituted the dataset received with information involving the city hall, laws and the reason for the expense, as shown in Figure 4.3.

```
<?xml version="1.0" encoding="UTF-8"?>
<commitments>
  <commitment>
    <class>1</class>
    <cnpjManagingUnit>40453315000109</cnpjManagingUnit>
    <actionCode>1234</actionCode>
    <trackingCode>1234</trackingCode>
    <functionCode>10</functionCode>
    <expenditureNatureCode>123456</expenditureNatureCode>
    <expenditureElementCode>48</expenditureElementCode>
    <programCode>1234</programCode>
    <subfunctionCode>301</subfunctionCode>
    <budgetaryUnitCode>123456</budgetaryUnitCode>
    <cpfAuthorizer>55720151052</cpfAuthorizer>
    <creditor>40453315000109</creditor>
    <commitmentDate>2022-01-10</commitmentDate>
    <sourceOfFunds>1234</sourceOfFunds>
    <history>textwithsize500</history>
    <contractId>textwithsize20</contractId>
    <commitmentNumber>12345678901</commitmentNumber>
    <commitmentType>0</commitmentType>
    <amount>20.2</amount>
    <functionCodeDescription>Health</functionCodeDescription>
    <subfunctionCodeDescription>Basic Care</subfunctionCodeDescription>
    <managingUnit>Municipal Government</managingUnit>
    <budgetaryUnitDescription>MUNICIPAL HEALTH FUND</budgetaryUnitDescription>
    <file>None</file>
    <entityName>Municipal Government - City Name</entityName>
    <period>4th Bimester</period>
    <fiscalYear>2023</fiscalYear>
  </commitment>
</commitments>
```

Figure 4.3: Adapted example of a Commitment Note from Maranhão

According to the accountants, some columns are foreign keys of other documents received by TCE-MA, such as the Managing Unit CNPJ, Budgetary Unit Code, Action Code, Program Code and CPF Authorizer. All of those information are inserted together, because they provide details of the context, but for internal organization their data are submitted in different tables having only the id code in the Commitment Note.

It is important to note that among the 27 attributes some of them are designed for internal control of TCE-MA like the Contract ID, Commitment Date, Number, File Name, submission period and year of government exercise. The fields Class and Commitment Type receive abbreviations of possible answers. Furthermore, the Creditor and CPF Authorizer fields have as answer the Social Security Number (Social Security Number - *Cadastro de Pessoa Fisica*

## 4.2. Data Understanding

in portuguese), of the person responsible for the document, those information are personal and sensitive in Brazil, which will receive an especial attention in this study.

Moreover, the other columns are key information to identify and understand the Commitment Notes context, such as the Government Function and Sub-Function Codes, Expenditure Nature Code, Expenditure Element Code, Source of Funds and Tracking Code. In addition, those fields have a few hundred possibilities of answer and the right combination between them, since they are information regulated by Brazilian laws.

However, with only the History and the Amount fields there is no preliminary response, since one is dedicated to provide a description of the document and the other the total amount to be spent respectively, which has to receive more attention, since they can reveal important information about the purpose of the expenditure for the municipality.

To better understand the structure and characteristics of the dataset, a comprehensive exploration of the main attributes indicated by the accounts was carried out, providing important insights into their distribution, balance across classes, and others. Analyzing the Government Function Code field (functionCode), the adopted dataset presents 28 codes out of the 29 available (see Appendix A for more information), of which approximately 85% are concentrated in only 4 codes, as shown in Figure 4.4. These codes refer to actions related to Health, Education, Administration and Social Assistance, some of the activities that have the greatest demand from the population in a government. On the other hand, areas such as Agrarian Organization, Foreign Relations, National Defense and Industry together do not exceed 0.5% of the total present in the dataset.

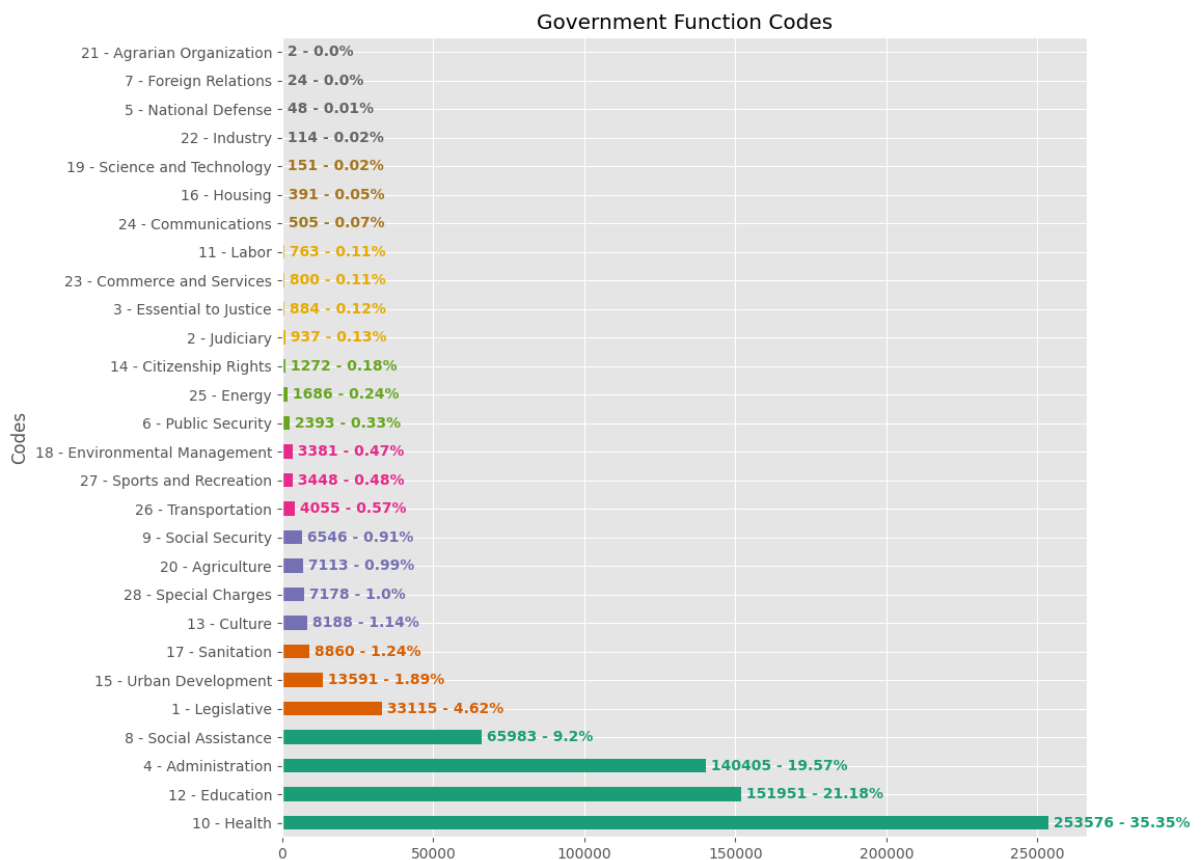


Figure 4.4: Percentage of Commitment Notes by Government Function Code

According to the Brazilian regulations for the year 2023 followed by the TCE-MA [64], the Sub-Function Code (subfunctionCode) has 115 possible codes (see Appendix A for more information). According to Figure 4.5, the analyzed dataset presents 91 of them, with the first four (122 - General Administration, 361 - Elementary Education, 301 - Primary Care and 302 - Hospital and Outpatient Care) being directly related to the four most used function codes. This is because the Government Function Code reflects the institutional mission of the agency and basically corresponds to ministries such as Culture, Education, Health, while the Sub-Function Code is the partition of the Function Code and, in general, combines with the typical code of the area or sector, although it is not restricted to it, it aims at detail the purpose of the expense [65].

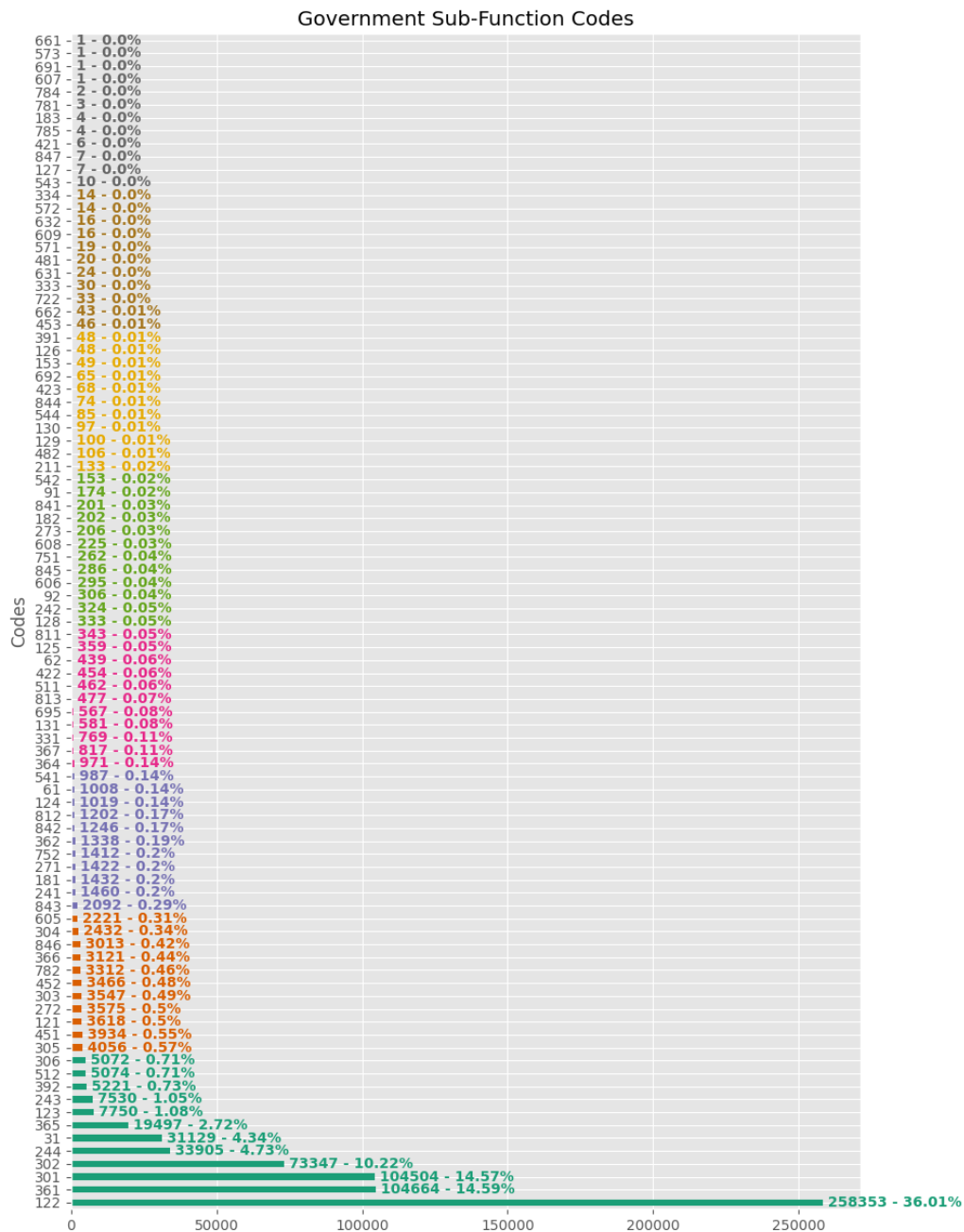


Figure 4.5: Percentage of Commitment Notes by Government Function Code

## 4.2. Data Understanding

Another attribute analyzed was the Commitment Type (`commitmentType`), which is responsible for describing the type of document. This attribute can take the following values: Ordinary, intended to cover expenses whose exact value is known; Estimated, intended to cover expenses whose value cannot yet be determined; Global, for contractual and other expenses that can be paid in installments; Reinforcement, when the amount committed is insufficient to cover the expense to be incurred; Cancellation, if the value of the commitment exceeds the amount of the expense incurred; or Reversal, when there is a need to cancel the commitment note completely due to some failure [66, 67]. In the analyzed dataset, more than 90% of the commitment notes were classified as Ordinary, while there were no examples of the Reinforcement type, as presented in Figure 4.6.

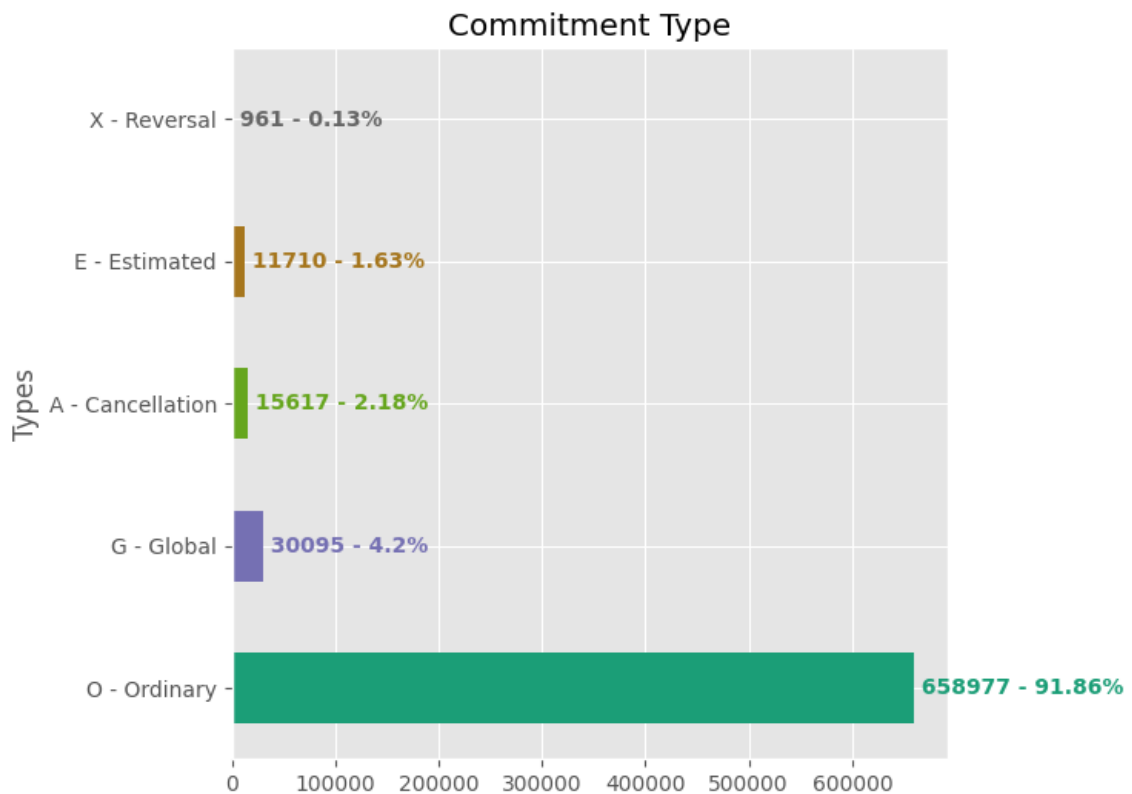


Figure 4.6: Commitment Notes categorized by Type

The TCE also has a field called `class`, which is responsible for discriminating what the commitment note refers to: payment, public contracting or others, for more specific government needs. In the studied data set, approximately 64% of the commitment documents were classified as others, while 20% refer to payment and approximately 15% to public contracts, as illustrated in Figure 4.7. According to the accountants, this field is important because it discriminates against the purpose of the commitment note, thus facilitating its analysis by the TCE.

Table 4.1 provides a detailed description of the Commitment Note attributes, based on TCE-MA regulations [68], highlighting in bold the most relevant attributes identified after the analyses which will probably compose the final training dataset due to their relevance in the context studied.

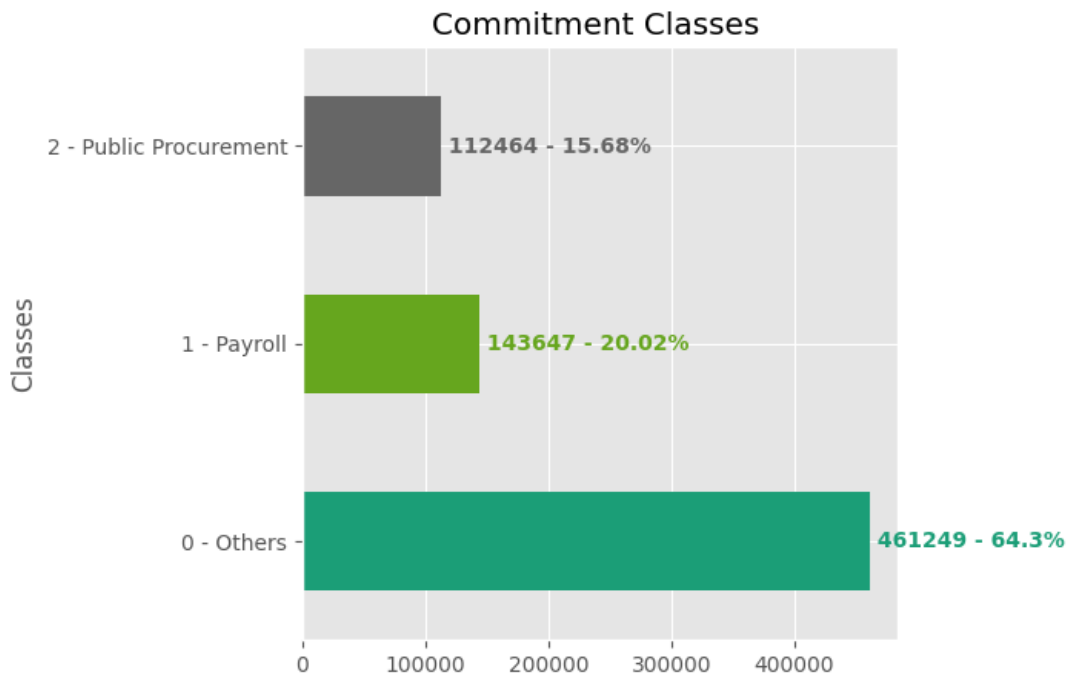


Figure 4.7: Distribution of Commitment Notes by Class

Table 4.1: Dataset Structure and Attributes

Name	Description	Observation/Source
cnpjManagingUnit	Managing unit CNPJ	managing units table
managingUnit	Managing unit name	managing units table
budgetaryUnitCode	Budgetary unit code	budgetary units table
<b>budgetaryUnitDescription</b>	Budgetary unit name	budgetary units table
<b>class</b>	Commitment classification	0-Others 1-Payroll 2-Public Procurement
contractId	Unique code for administrative contract identification within the Managing Unit	Sinc-Contracts System
creditor	Creditor CPF/CNPJ	...NN
<b>functionCode</b>	Function code, according to expenditure functional classification	Appendix II (Layout of MSC) Ordinance STN No. 642, Sept. 20, 2019 [64]
<b>functionCodeDescription</b>	Description of the Function Code	Appendix II (Layout of MSC) Ordinance STN No. 642, Sept. 20, 2019 [64]
<b>subfunctionCode</b>	Sub-function code, according to expenditure functional classification	Appendix II (Layout of MSC) Ordinance STN No. 642, Sept. 20, 2019 [64]

#### 4.2. Data Understanding

<i>(Continued from previous page)</i>		
<b>Name</b>	<b>Description</b>	<b>Observation/Source</b>
<b>subfunctionCodeDescription</b>	Description of the Sub-function code	Appendix II (Layout of MSC Ordinance STN No. 642, Sept. 20, 2019 [64])
programCode	Annual budget law program code	programs table
actionCode	Annual budget law action code	actions table
expenditureNatureCode	Expenditure nature code (ND): Economic Category, Nature Group, and Application Mode.	Appendix II (Layout of MSC Ordinance STN No. 642, Sept. 20, 2019 [64])
expenditureElementCode	Expenditure element code, according to budget expenditure classification by nature	Appendix II (Layout of MSC Ordinance STN No. 642, Sept. 20, 2019 [64])
commitmentNumber	Commitment number	...NN
<b>commitmentType</b>	Commitment type	O-Ordinary E-Estimated G-Global R-Reinforcement A-Annulment X-Reversal
commitmentDate	Commitment issuance date	YYYY-MM-DD
<b>history</b>	Commitment history	...CC
cpfAuthorizer	Expense authorizer CPF responsible for commitment issuance	expense authorizers
sourceOfFunds	Source of Funds code (FR)	Appendix II (Layout of MSC Ordinance STN No. 642, Sept. 20, 2019 [64])
trackingCode	Budget execution tracking code (CO)	Appendix II (Layout of MSC Ordinance STN No. 642, Sept. 20, 2019 [64])
<b>amount</b>	Amount committed per resource source/additional resource information	...NNNN.NN
file	File name of the document	..CC
entityName	City name of the document	..CC
period	Bimonthly of the document sent	..CC
fiscalYear	Year of the document	YYYY

The data format adhere to the following structure:

- ...**NN** = Numeric values: individual digits from 0 to 9.
- ...**CC** = Characters: any graphical symbol typically found on a computer keyboard.

- ... **DD** = Day of the month, ranging from 01 to 28, 29, 30, or 31.
- ... **MM** = Month of the year, ranging from 01 to 12.
- ... **YYYY** = Fiscal year, in four digits, without thousands separators. Example: 2021.
- **DECIMALS(16.2)** = Values without special characters, without thousands separators, and with a period (".") as a decimal separator. This format includes up to fourteen integers and two decimal places. For instance, one thousand two hundred thirty-four reais and fifty-six cents should be written as: 1234.56

In conclusion, even with some observations, the acquired dataset has quality enough to this study, since its data is complete and provide valuable information for training a ML model able to help the accountants to construct their Commitment notes for Government Auditing in Maranhão.

#### 4.2.4 Data Protection and Ethical Aspects

ML is one of the most relevant and emerging areas of AI and was born from the need to automatically process the large volumes of data produced in different ways. In other words, to implement algorithms capable of detecting patterns and, from them, making predictions, optimizations or suggestions [35, 69]. Currently there are many types of ML models, but all their success depends on the dataset. The data is one of the most important parts in the process of building a purpose-built ML model.

In order to regulate the use of these data in the digital environment, the General Data Protection Regulation (GDPR) [70] in the European Union and the LGPD (LGPD)[71] in Brazil were created. They are comprehensive legal rules that regulate the collection, storage, and processing of personal data. Both laws aim to protect individuals' privacy and data rights, imposing strict requirements on how organizations handle sensitive information. The GDPR applies across the European Union context, while the LGPD governs data processing in Brazil. These regulations mandate that personal data must be processed lawfully, transparently, and for specific purposes, requiring researchers and organizations to take significant precautions when handling any personal information and, at any time, be prepared to stop its use, since the data owner can make such a request.

Given this context and the scope of the dissertation, certain decisions regarding the data were made. Since the study is conducted in both the European Union and Brazil, it was necessary to ensure compliance with the GDPR and the LGPD. As part of this and to adhere to these regulations and safeguard data privacy, all personal information received in the dataset was excluded, named Creditor and CPF Authorizer columns. The exclusion of those columns do not affect the efficiency of the dataset adopted, once they make reference to other tables which will not be used in this dissertation. Furthermore the other columns of the dataset do not have any reference to personal data, as discussed in the data exploration, subsection 4.2.3.

### 4.3 Chapter Remarks

This chapter was dedicated to present an in-depth analysis of the dataset utilized in this study. Firstly, a complete overview of the scenario in Maranhão Government was presented, based on the reports from two experienced accountants from Maranhão and aligned with the regulations set by TCE-MA. This contextual background not only sheds light on the

### 4.3. Chapter Remarks

---

specifics of public sector accounting practices in the state of Maranhão but also explores the future application of the model developed in this dissertation and the practical setting in which it is expected to operate.

Then, the complete process employed to address the entire understanding of the data was presented. Beginning with an explanation of the dataset's relevance within TCE-MA, followed by an explanation about the procedures used for data extraction and concludes with an overview of each field, its context and relevance in a Commitment Note. Another important topic addressed was the ethical aspects, highlighting the study's adherence to both Brazilian and European Union data protection laws, given the sensitive nature of the dataset and the study's broader implications.



## Chapter 5

# Proposed Solution and Evaluation

This chapter is dedicated to presenting all the phases of the development of the proposed model, grouped by the preparation of two different datasets and the final model configurations, followed by the training process and its results with a detailed evaluation. Each step was carefully designed to align with the Data Preparation, Modeling, and Evaluation phases of the CRISP-DM methodology.

### 5.1 Data Preparation

In the next subsections are presented the steps taken during the preparation of two different datasets to train the BERTimbau model and evaluate its results, addressing the needs of the data received.

#### 5.1.1 Data Preprocessing

The initial step in data preparation involved analyzing which information would be included in the final dataset. After a careful investigation as shown in section 4.2, certain columns were removed based on specific criteria. This decision considered factors such as the dataset's large volume and the hardware limitations (detailed in section 3.2). The excluded columns included: attributes that serve as foreign keys to other tables received by TCE-MA (Managing Unit CNPJ, Budgetary Unit Code, Action Code and Program Code); columns addressing personal information (CPF Authorizer and Creditor) in order to respect the GDPR and LGPD orientations; and fields used for internal control within TCE-MA systems (Managing Unit Name, Contract ID, Commitment Date, Number, File Name, submission period and year of government exercise), as these provide only general information of the note and lack specific relevance to this study.

Additionally, after conducting these analysis and following the recommendations of the accountants it was decided to develop the ML model using the attributes History, Class, Commitment Type, and Budgetary Unit Name to classify the appropriate Government Function Code of the Commitment Note. This objective was set based on the accountant's experience which highlighted challenges in selecting the correct code when creating the Commitment Note. Other attributes regulated by Brazilian Government laws, such as Amount, Government Sub-Function, Expenditure Nature Code, Expenditure Element Code, Source of Funds and Tracking Code also require special attention when creating the Commitment Note. However, these attributes presented a significant data imbalance, which could be explored in future research.

Another preprocessing step taken was the replacement of the abbreviations in the attributes Class and Commitment Type, which made it possible to have more significant information about the Commitment Note context, once these fields provide brief, but important description of the document. After this, the columns of the dataset were ordered based on the importance of the attribute according to the exploration made in the dataset, as shown in Figure 5.1.

```
<?xml version="1.0" encoding="UTF-8"?>
<commitments>
  <commitment>
    <history>textwithsize500</history>
    <budgetaryUnitDescription>MUNICIPAL HEALTH FUND</budgetaryUnitDescription>
    <commitmentType>0</commitmentType>
    <class>1</class>
    <functionCode>10</functionCode>
    <functionCodeDescription>Health</functionCodeDescription>
  </commitment>
</commitments>
```

Figure 5.1: Final dataset with attributes ordered

Finally, it is important to note that the initial exploration of the dataset revealed that the History field usually contains key words which could help in the classification process, then this field is the first one to avoid losing much information if the sentence is bigger than the max length defined to the model.

### 5.1.2 Data Balancing

An essential consideration in preparing the dataset was addressing data imbalance, which can significantly impact the model's performance and accuracy. In this dissertation the dataset received was unbalanced in almost every category, as discussed in subsection 4.2.3. Given this situation, after deciding which columns to work and the field to classify (Government Function Code), it was applied the technique of undersampling configured to randomly select the examples of codes with a larger number of rows. The application of the undersampling techniques have the potential to improve the performance of the model with unbalanced datasets, reducing the training cost [72].

With 85.3% of the dataset concentrated in only 4 of the 29 Function Codes, two datasets with different sample sizes were prepared for model training: one with 10 codes (DS10Codes) and the other with 15 codes (DS15Codes). This selection was based on currency counts, where the smallest count defined the sample size for the others.

However, even with the adoption of the Under Sampling technique, the Dataset A (DS10Codes) was reduced by 50% to attend to the limitations of the hardware used in this project and facilitate the analysis compared with dataset B (DS15Codes). As a result of those choices, the number of Commitment Notes in the dataset changed from 717.360 to 35.560 in dataset A and 35.895 in dataset B, as detailed in Table 5.1.

Table 5.1: Datasets Detailed p/ Function Code

ID	No. Function Codes	No. examples p/ Code	Percentage p/ Code	Total no. examples
DS10Codes	10	3.556	10.0%	35.560
DS15Codes	15	2.393	6.67%	35.895

Furthermore, Table 5.2 presents the chosen codes and their descriptions, ordered by their frequency in the dataset, of which the last five codes were present only in the Dataset DS15Codes.

Table 5.2: Function Codes and Descriptions

#	Code	Description
1	10	Health
2	12	Education
3	4	Administration
4	8	Social Assistance
5	1	Legislative
6	15	Urban Development
7	17	Sanitation
8	13	Culture
9	28	Special Charges
10	20	Agriculture
11	9	Social Security
12	26	Transportation
13	27	Sports and Recreation
14	18	Environmental Management
15	6	Public Security

## 5.2 Model Specifications

The first step taken in the process of fine-tuning the BERTimbau was to merge the fields History, Class, Commitment Type, Managing Unit Name and Budgetary Unit Name to only one sentence separated by spaces, composing a column called **description**. Another change made was in the attribute of the Functional Government Codes, which had the codes replaced by an integer from 0 to 9 for DS10Codes and 0 to 14 for DS15Codes, in order to address the necessity of the model. Those two columns composed the final dataset to building the model.

Another important decision taken is related to the selection and configuration of hyperparameters, since they are crucial aspects of the model development process, significantly impacting its performance. In this dissertation, the hyperparameters for the BERTimbau-based model were defined in alignment with the recommendations of its authors [53] along with initial exploratory tests. These tests were conducted to evaluate the model behavior under various configurations and ensure that the selected values supported efficient training and optimal results.

The first step was dedicated to define the **maximum sequence length** accepted by the model. For this purpose, it analyzed the sequence length of the description field after tokenization with the BERTimbal tokenizer, which revealed that almost all of the final sentence was around 300 tokens, as is shown in Figure 5.2. Based on this information the maximum sequence length of training was set to 300 to avoid losing information.

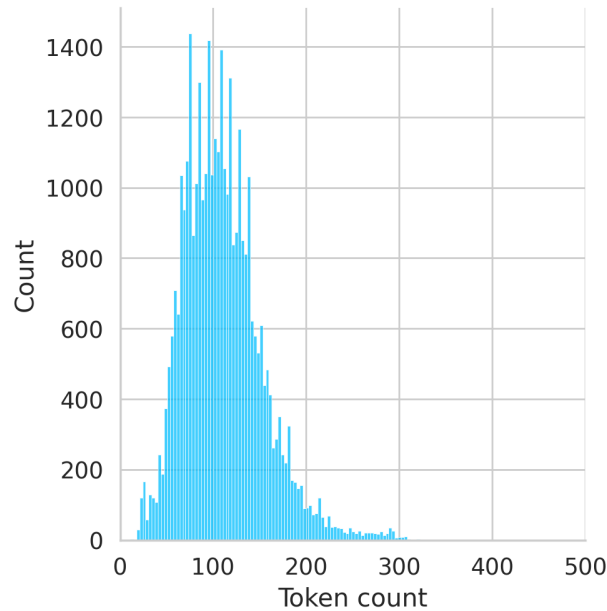


Figure 5.2: Token Count of the column description

Next, others hyperparameters were set, starting with the number of **epochs** set to 10, which balances sufficient training iterations without overfitting, with a **batch size** of 32, ensuring an efficient trade-off between computational resource usage and training stability. The **optimizer** chosen was AdamW, due to its widespread use in the area and the good results presented in the initial tests.

The **learning rate** was set to  $2 \times 10^{-5}$ , a value commonly used for fine-tuning transformer models, coupled with a **linear learning rate scheduler** with warmup to gradually increase the learning rate at the beginning of training. This approach avoids large initial updates that can destabilize the model, especially given the complexity of BERT-based architectures. However, the **warmup rate** was set to zero, indicating the choice of not taking additional warmup steps beyond those implied by the linear schedule, in order to observe the model with a more direct contact with the data. As for the **objective loss** employed, the technique employed was the Cross-Entropy Loss, which is widely used in classification tasks to measure the difference between predicted probabilities and true class labels.

In summary, all defined hyperparameters were aligned with the multiclass classification goal of this study, producing valuable results. Table 5.3 provides an overview of these hyperparameters used during the fine-tuning process.

Finally, the training set is defined for both datasets with 80% of the data and the remaining 20% distributed between 10% for validation and 10% for testing. Table 5.4 presents the number of examples per dataset in each category.

Table 5.3: Summary of the hyperparameters used with the model

Hyperparameters	Value
Epochs	10
Batch size	32
Optimizer	AdamW Algorithm
Learning Rate	2e-5
Learning rate scheduler	Linear with warmup
Warmup proportion	0
Training steps	Total steps
Objective Loss	Cross Entropy Loss
Maximum sequence length	300

Table 5.4: No. examples p/ dataset

Category	No. examples in DS10Codes	No. examples in DS15Codes
Train - 80%	28.448	28.716
Validation - 10%	3.556	3.589
Test - 10%	3.556	3.589

### 5.3 Fine-Tuning Process and Evaluation

With the configurations detailed above, the BERTimbau model was fine-tuned with both previously defined datasets. Observing the total training time of the model on these different situations it was noted that the dataset with 10 codes had a slight increase in time in comparison to the other dataset with 15 codes. Many reasons could be responsible for this phenomenon, one of them is the number of samples per class during the training process. That is because the DS10Codes had 3.556 examples per Function Code and DS15Codes had 2.393 examples, which means that the first dataset had about one thousand more examples if compared with the second one, approximately 32.7% of difference between them. In order to facilitate the comprehension, Table 5.5 presents a summary of the situation with the total training time and the quantity of samples per dataset.

Table 5.5: Number of Samples and Training time p/ dataset

Metric	DS10Codes	DS15Codes
Number of samples per Dataset	3.556	2.393
Wall time	3h 8min 52s	2h 56min 59s

Another important metric addressed was accuracy to measure classification performance. As detailed in Table 5.6, the BERTimbau model presented a consistent and robust performance on both datasets under evaluation during all phases, training, validation, and testing. These results demonstrate that the constructed model was able to develop its ability to learn patterns from the training phase and apply this knowledge in the validation and testing phases. Furthermore, the model successfully avoided cases of overfitting and underfitting, a potentially critical achievement in fine-tuning models such as BERTimbau.

A comparison between the training and validation accuracy results across epochs for both datasets used is presented in Figures 5.3 and 5.4. Analyzing the results on both datasets, already in the initial training and validation phases, the model achieved an accuracy above

Table 5.6: Dataset Accuracy Metrics

Metric	DS10Codes	DS15Codes
Training Accuracy	0.997	0.995
Validation Accuracy	0.976	0.973
Testing Accuracy	0.982	0.977

90%, with stability around epoch 5. This consistent behavior demonstrates the effectiveness of the chosen hyperparameters, as well as the quality of the data preprocessing techniques applied.

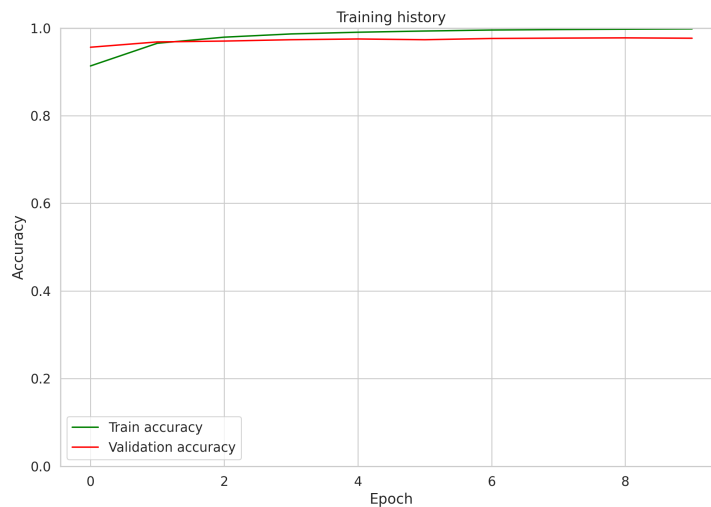


Figure 5.3: Training and validation accuracy results for DS10Codes p/ epoch

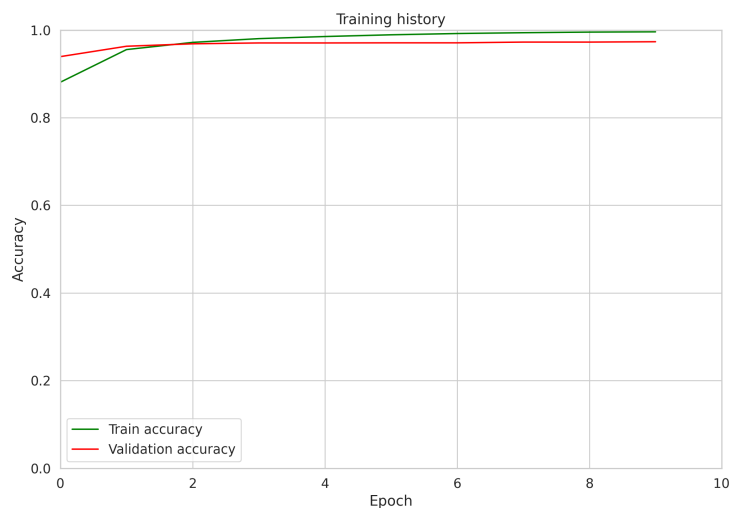


Figure 5.4: Training and validation accuracy results for DS15Codes p/ epoch

Similarly, the minimization of errors, called Loss, produced very promising results on both datasets. Table 5.7 presents a summary of the Loss obtained in the training, validation and testing phases, which is observed, in both datasets, the metric remarkably close to zero in the training phase, indicating that the BERTimbau model was able to learn the patterns within the training data with high confidence. However, a slight increase in the Loss, values around

### 5.3. Fine-Tuning Process and Evaluation

0.1, was observed in the validation and testing phases. These differences suggest that the model, even with high accuracy, was more adapted to the training data than to unseen data. This behavior is not uncommon, especially in large-scale models like BERT<sub>im</sub> applied to multiclass classification tasks.

Table 5.7: Datasets Loss Metrics

Metric	DS10Codes	DS15Codes
Training Loss	0.007	0.015
Validation Loss	0.126	0.159
Testing Loss	0.103	0.144

Figures 5.5 and 5.6 also provide a comparison of the training and validation loss over the epochs per adopted dataset. In both situations, a high decrease in the loss is observed in the training phase with stability only in the final epochs, while in the validation phase the decrease is smoother, stabilizing in the initial epochs.

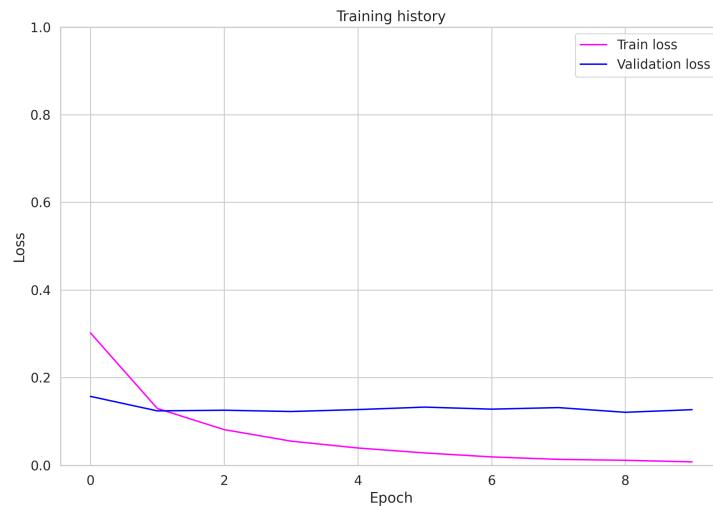


Figure 5.5: Training and validation loss results for DS10Codes p/ epoch

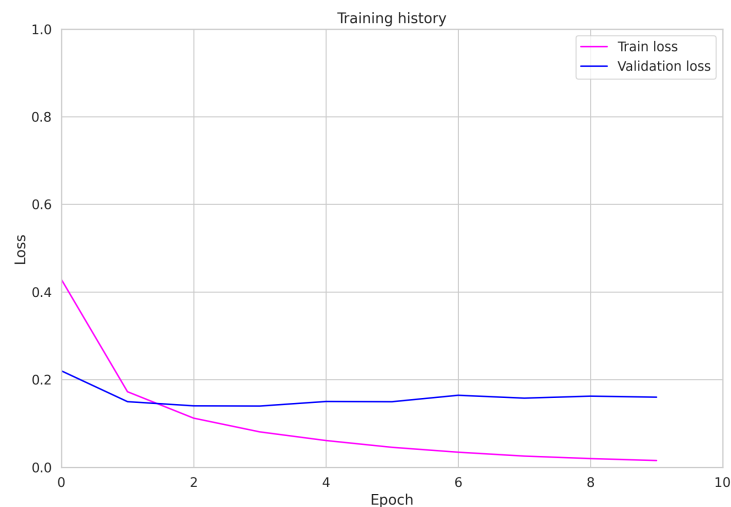


Figure 5.6: Training and validation loss results for DS15Codes p/ epoch

Furthermore, the Precision, Recall, and F1-Score metrics were also carefully analyzed to provide a detailed assessment of the model's performance in correctly identifying instances in different classes. As shown in Table 5.8, the model obtained consistent Precision, Recall, and F1-Score numbers with both datasets used during the training and validation phases. Observing these results, it can be noted that the values were similar in both contexts, highlighting the robustness of the BERTimbau model in dealing with several possible classification scenarios, ensuring not only the correctness of the predictions, but also minimizing errors across multiple classes.

Table 5.8: Precision, Recall, and F1 Score Metrics

	DS10Codes			DS15Codes		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
<b>Validation</b>	0.976	0.976	0.976	0.973	0.973	0.972
<b>Testing</b>	0.982	0.982	0.981	0.977	0.977	0.977

Table 5.9 presents a comprehensive overview of the model performance for each class within the test datasets, DS10Codes and DS15Codes, detailing the Precision, Recall and F1-Score in both studied scenarios. Analyzing the results, one can observe a consistently high performance in most classes, with notable F1-Scores close to or at 1.0 for several codes, such as Legislative, Social Assistance and Health, in both adopted datasets.

Table 5.9: Precision, Recall, F1-Score for each Code in the DS10Codes and DS15Codes datasets.

Code	Code Description	DS10Codes			DS15Codes		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
01	<b>Legislative</b>	1.00	1.00	1.00	1.00	1.00	1.00
04	<b>Administration</b>	0.97	0.89	0.93	0.96	0.84	0.90
08	<b>Social Assistance</b>	0.99	0.99	0.99	0.99	0.99	0.99
10	<b>Health</b>	1.00	0.99	1.00	1.00	0.99	1.00
12	<b>Education</b>	0.98	0.99	0.99	0.99	0.97	0.98
13	<b>Culture</b>	0.99	0.98	0.98	0.96	0.99	0.97
15	<b>Urban Development</b>	0.95	0.99	0.97	0.93	0.97	0.95
17	<b>Sanitation</b>	0.99	0.99	0.99	0.98	0.98	0.98
20	<b>Agriculture</b>	0.98	0.99	0.99	0.98	1.00	0.99
28	<b>Special Charges</b>	0.97	0.99	0.98	0.95	1.00	0.97
06	<b>Public Security</b>	-	-	-	0.98	1.00	0.99
09	<b>Social Security</b>	-	-	-	0.99	0.99	0.99
26	<b>Transportation</b>	-	-	-	0.96	0.98	0.97
27	<b>Sports and Recreation</b>	-	-	-	1.00	0.98	0.99
18	<b>Environmental Management</b>	-	-	-	0.99	0.99	0.99

### 5.3. Fine-Tuning Process and Evaluation

Table 5.10 also provides the overall precision, macro average and weighted average of the results for both studied cases, which remained consistently high at 0.98, indicating a robust overall performance of the built model.

Table 5.10: Balanced Metrics in the DS10Codes and DS15Codes datasets.

Metric	DS10Codes	DS15Codes
Accuracy	0.98	0.98
Macro average	0.98	0.98
Weighted average	0.98	0.98

All of those results align with current authors in the field of BERT models, specifically BERTimbau's performance with Brazilian Portuguese language. Even in a setting with limited examples per code, the model demonstrates a good performance, especially with the Dataset with more codes (DS15Codes). This behavior is largely due to BERTimbau's prior training, which provides foundational knowledge valuable in scenarios where data scarcity would typically limit the effectiveness of traditional models [73, 74], given the amount of examples per code and the results obtained.

In the case of this study, despite significant challenges such as data imbalance and hardware limitations restricting the number of samples in the datasets for fine-tuning, the BERTimbau model demonstrated remarkable adaptability and efficiency even in the early training phases, indicating that the context was conducive to effective learning. The Confusion Matrices shown in Figures 5.7 and 5.8 reinforce these findings. These visible representations prove that the model, for both datasets, achieved a high level of accuracy across almost all classes, with minimal misclassifications. For both datasets, the analyses of the results showed some cases in which the model classified the descriptions of code 4, called Administration, in other options, such as 15: Urban Development and 28: Special Charges. There are some possibilities that can trigger this phenomenon, it is believed that the most appropriate for both situations is that the descriptions used in the model refinement set have similar words, as they are, in general, from very similar contexts within the Brazilian Government. However, even with this situation, the BERTimbau model presented good performance with both datasets.

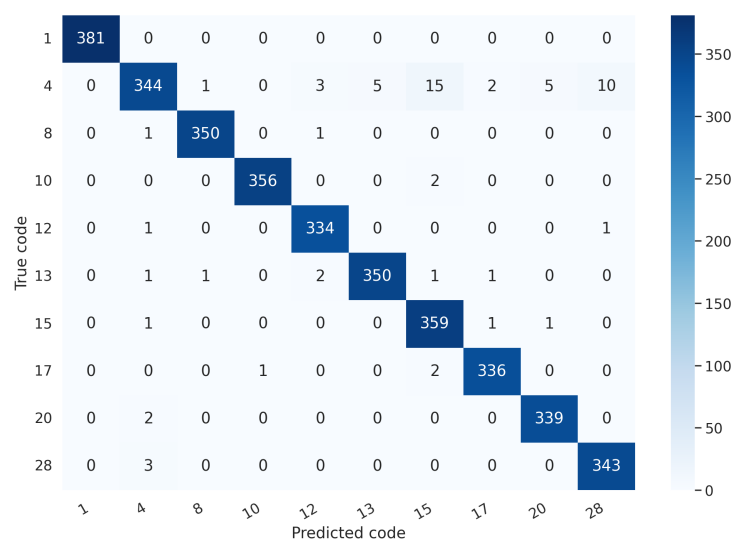


Figure 5.7: Confusion Matrix for for DS10Codes

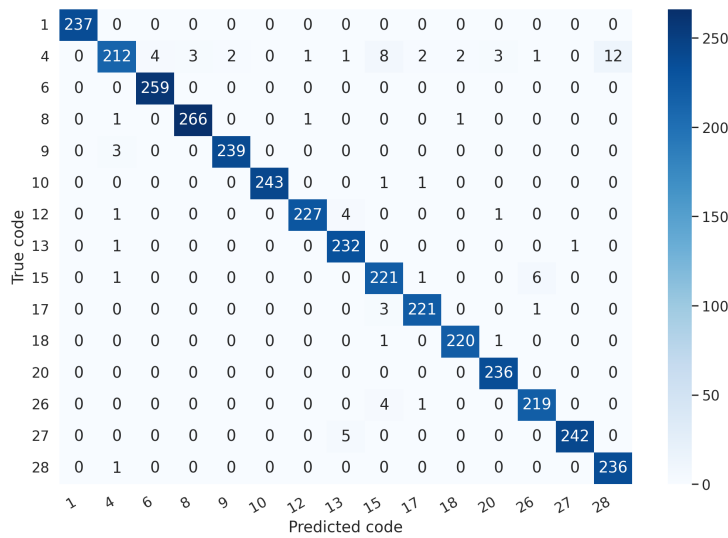


Figure 5.8: Confusion Matrix for for DS15Codes

In conclusion, all the results presented and discussed in this chapter highlight the adaptability and effectiveness of the proposed approach using the BERTimbau model for fine-tuning with specific datasets. Despite the challenges for both studied situations, the model showed strong performance across multiple metrics, achieving consistent results. All these findings emphasize the potential of leveraging TL techniques to address complex tasks and provide valuable insights for feature and application improvements.

## 5.4 Chapter Remarks

In this chapter the proposed model was created and evaluated with the specifications presented in the previous chapters. The initial phase was dedicated to preparing the dataset to train the model and, given the data specifications and the context of the dissertation, two different datasets were selected for fine-tuning the model. Respecting the limitations of the adopted hardware and defining the Government Function Code as the field to classify the Commitment Notes by the model, the most important fields were selected and the datasets were prepared differentiating them by the number of codes to classify a Note.

In addition, the final hyperparameters defined to train the BERTimbau were presented. Considering the length of the sequence of sentences in the dataset, the size of the samples to train, test and validate the model and the suggestions provided by the authors of BERTimbau, the final configurations were selected and presented.

All previous steps contributed to the fine-tuning of BERTimbau, which presented good results with the two selected datasets. Metrics such as Accuracy, Precision, Recall and F1-Score achieved high success rates and the Loss produced low numbers, indicating that the model was well adapted to the context, especially one with more Function Codes to classify and less data per code to fine-tuning.

# Chapter 6

## Conclusion

This chapter summarizes the conclusions obtained in this dissertation, presenting discussions about the objectives achieved, contributions, limitations found and suggestions for future work.

### 6.1 Accomplished Objectives

As discussed in the introduction, chapter 1, the main goal of this dissertation was the development of a solution to help the accountants to prepare government financial statements in the state of Maranhão, Brazil, thus aiding to reduce the number of disapproved accounts. To address this goal some specific objectives were defined:

- **OB1:** Investigate the current state of the art in ML applied on Government audit with focus on financial statements;
- **OB2:** Develop an ML model able to make multiclass text classification applied in the context of the government financial statement in the state of Maranhão;
- **OB3:** Test the model developed using real data to adjust the hyperparameters;
- **OB4:** Validate the model with accountants experienced with government financial statements in the state of Maranhão;
- **OB5:** Make available a ML solution which helps to reduce the number of disapproved accounts in Maranhão.

Analyzing the Objectives proposed, it is possible to affirm that the **OB1** was fulfilled, providing valuable insights to this dissertation and the scientific community, once its results were validated and presented in PAAMS (Practical Applications of Agents, Multi-Agent Systems, and Digital Twins) Conference 2024 occurred in Salamanca, Spain, and will be published on the proceedings of the conference 2024, supported by Springer Nature.

Similarly, the **OB2** and **OB3** were successfully accomplished, because they contributed directly to the main goal in a practical way, once the fine-tuning of both models were totally conducted with data from cities of the State of Maranhão, hence making possible to refine the model to the studied context.

However, it was not possible to implement **OB4**, as mentioned in section 4.1 the model built in this dissertation integrates a project developed by the Innovation Laboratory of IFMA Campus Caxias in Maranhão, Brazil, which has as main goal to produce a software called ConnTec, that was not completed yet to this date. Consequently the **OB5** was also

impaired, since the model could not be used outside of the ConnTec system, because it belongs to a very specific context of Maranhão's Government.

Finally, observing the entire context, even with those impairments the main objective of this dissertation was successfully achieved, once the fine-tuned BERTimbau model presented valuable results for both contexts, especially with the dataset containing more options and less examples per class to classify a Commitment Note (DS15Codes).

## 6.2 Limitations and Future Work

Despite the success of the BERTimbau model developed, the study had some limitations during the process. The first one was the very unbalanced data received, which had classes with only two samples to train and about 85% of the data concentrated on only four classes. Another limitation found was the hardware, even with a GPU disponible for tests its memory RAM was not large, which limited the number of samples on DS10Codes.

In addition, as discussed above, the main limitation found was the impossibility to validate the model with the experienced accountants to analyze the model behavior in real situations and complete the final step of the CRISP-DM methodology, named deployment.

As a result of this context, in future work it is recommended to include the remaining Government Functional Codes on the dataset used for fine-tuning the model to have a model able to classify the Commitment Notes among the 29 existing possibilities (see appendix A). Knowing that some Functions are more used than others it is suggested to combine the fine-tuning process with some method to alleviate the imbalance on the data, such as Class Balanced Loss used by Hammes and Freitas on their study [75], for example.

Other pertinent activity to be included on the project is the inclusion of the columns, named Government Sub-Function Code, Expenditure Nature Code, Expenditure Element Code, Source of Funds, Tracking Code and Amount (to investigate if the value is consonant with the Brazilian laws) on the analyzes of the model. Since the fields are considered key information on the Commitment Note creation with many possibilities supported by the Brazilian laws.

Lastly, a suggestion for future work is to experiment with an alternative version of BERTimbau, known as DistilBERTimbau. Recent studies, such as those by Barbon and Akabane [76], indicate that DistilBERTimbau achieves high performance while reducing computational costs, once it is a distilled version of the model with optimized layers.

## 6.3 Chapter Remarks

This dissertation presented a comprehensive analysis and development process for a model designed to support Maranhão state accountants in creating commitment notes in compliance with Brazilian law. This study is part of a broader initiative supported by the IFMA Caxias Innovation Laboratory, which seeks to assist the Government of Maranhão in reducing the number of financial statements disapproved due to errors during their preparation.

The project encountered numerous challenges, particularly with the unbalanced data. Nevertheless, it provided valuable insights for the field through the adopted methodology and the resulting model. This work is envisioned as a foundational step in integrating AI into Maranhão's government processes, especially in the area of financial auditing.

# Bibliography

- [1] DataSnipper. *Types of Governmental Audits: A Comprehensive Overview*. url: <https://www.datasnipper.com/resources/types-governmental-audits#:~:text=Governmental%20audits%20are%20systematic%20examinations,and%20effectiveness%20of%20internal%20controls..>
- [2] AICPA. *What is a Governmental Audit?* url: <https://us.aicpa.org/interestareas/governmentalauditquality/information-on-governmental-audits>.
- [3] Europe Union. *Supreme Audit Institutions in the EU and its member states – An overview*. doi:10.2865/686084.
- [4] THEIIA. *The Role of Auditing in Public Sector Governance*. [https://www.ca-ilg.org/sites/main/files/file-attachments/auditing\\_in\\_public\\_sector.pdf?1441835528](https://www.ca-ilg.org/sites/main/files/file-attachments/auditing_in_public_sector.pdf?1441835528). 2006.
- [5] Instituto Brasileiro de Geografia e Estatística - IBGE. *Cidades e Estados - Maranhão*. IBGE, 2023. url: <https://www.ibge.gov.br/cidades-e-estados/ma.html>.
- [6] Rodrigo Moreira da Silva. "Regime político brasileiro". In: *Estudos eleitorais* v. 8.n. 1 (2013), pp. 55–80. url: <http://bibliotecadigital.tse.jus.br/xmlui/handle/bdtse/1307>.
- [7] Senado Federal do Brasil. *Constituição da República Federativa do Brasil*. Brasília, DF: Senado Federal, 1988.
- [8] Tribunal de Contas da União - TCU. *Contexto geral das prestações de contas*. Tribunal de Contas da União, 2022. url: <https://portal.tcu.gov.br/contas/contas-e-relatorios-de-gestao/contexto-geral-das-prestacoes-de-contas.htm>.
- [9] Brasil. "Lei Complementar nº 101, de 4 de Maio de 2000". In: *Diário Oficial [da] República Federativa do Brasil* (2000).
- [10] Brasil. "Lei nº 8.429, de 2 de Junho de 1992". In: *Diário Oficial [da] República Federativa do Brasil* (1992).
- [11] Brasil. "Decreto-Lei nº 201, de 27 de Fevereiro de 1967". In: *Diário Oficial [da] República Federativa do Brasil* (1967).
- [12] Tribunal de Contas da União - TCU. *Lista de contas julgadas irregulares*. Tribunal de Contas da União, 2022. url: <https://contasirregulares.tcu.gov.br>.
- [13] Instituto Brasileiro de Geografia e Estatística - IBGE. *The Geography of Brazil*. IBGE, 1970. url: <https://www.ibge.gov.br/en/geosciences/territorial-organization/territorial-analyses/18381-geography-of-brazil.html?edicao=26683>.
- [14] Confederação Nacional dos Municípios - CNM. *Pesquisa da CNM mostra que 383 prefeitos já perderam o mandato*. Confederação Nacional dos Municípios, 2020. url: <https://www.cnm.org.br/cms/>.
- [15] SELA. *Brazil*. url: <https://www.sela.org/en/member-states/brazil/>.
- [16] Tribunal de Contas do Estado do Maranhão. *Carta de serviços ao usuário do TCE-MA*. São Luís, MA, 2022. url: <https://www.tcema.tc.br/>.
- [17] Tribunal de Contas do Estado do Maranhão. *SINC - Módulo Fiscal*. São Luís, MA, 2023. url: <https://app.tcema.tc.br/hotsites/fiscal>.

- [18] Paulo Nayron Soares do Carmo et al. "Aprendizado de Máquina em Ações de Controle no Tribunal de Contas do Estado do Maranhão". In: *VII Jornada de Informática do Maranhão* (2018).
- [19] Heloisa Guimaraes Coelho et al. "Software para Contabilidade de Órgãos Públicos auxiliado por Aprendizado de Máquina". In: *Anais do Universo IFMA* (2022). url: <https://www.even3.com.br/anais/universo2022/590033-software-para-contabilidade-de-orgaos-publicos-auxiliado-por-aprendizado-de-maquina/>.
- [20] Julie Watson. *How Artificial Intelligence Will Impact the Accounting Industry?* Ace Cloud Hosting, 2023. url: <https://www.acecloudhosting.com/blog/artificial-intelligence-impact-accounting/>.
- [21] Zehui Zhao et al. "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations". In: *Expert Systems with Applications* v. 242 (May 2024), p. 122807. issn: 0957-4174. doi: 10.1016/j.eswa.2023.122807. url: <http://dx.doi.org/10.1016/j.eswa.2023.122807>.
- [22] Portal da Transparência do Governo Federal - Brasil. *Execução da despesa pública*. Controladoria-Geral da União - CGU, 2022. url: <https://www.portaltransparencia.gov.br/url/e73c9f34>.
- [23] Brasil. "Lei nº 4.320, de 17 de Março de 1964". In: *Diário Oficial [da] República Federativa do Brasil* (1964).
- [24] Matthew J Page et al. "PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews". In: *bmj* v. 372 (2021).
- [25] Catrin Sohrabi et al. *PRISMA 2020 statement: What's new and the importance of reporting guidelines*. 2021.
- [26] MIT. *Libguides: Database Search Tips: Boolean operators*. url: <https://libguides.mit.edu/c.php?g=175963&p=1158594>.
- [27] Keletso Mabelane et al. "An Analysis of Local Government Financial Statement Audit Outcomes in a Developing Economy Using Machine Learning". In: *Sustainability* v. 15.n. 1 (2023). issn: 2071-1050. doi: 10.3390/su15010012. url: <https://www.mdpi.com/2071-1050/15/1/12>.
- [28] L.H. Hamelers. *Detecting and explaining potential financial fraud cases in invoice data with Machine Learning*. 2021. url: <http://essay.utwente.nl/85533/>.
- [29] Adriano de Oliveira Andrade et al. "Prediction and Visualisation of SICONV Project Profiles Using Machine Learning". In: *Systems* v. 10.n. 6 (2022). issn: 2079-8954. doi: 10.3390/systems10060252. url: <https://www.mdpi.com/2079-8954/10/6/252>.
- [30] Jéssica Jane de Souza. "A necessidade de regulação do uso da inteligência artificial nas ações de controle do tribunal de contas da união". MA thesis. Uninter, 2021.
- [31] Rodrigo Otávio Coelho Hildebrand. "A experiência do Tribunal de Contas da União com inteligência artificial". In: *Trabalho de Conclusão de Curso (Bacharelado em Direito)* (2021). url: <https://bdm.unb.br/handle/10483/29757>.
- [32] Tiago Rodrigues Guimarães et al. "Análise dos fatores que contribuem para a recuperação dos créditos de ICMS inscritos em dívida ativa no estado do Rio de Janeiro". MA thesis. Universidade do Estado do Rio de Janeiro, UERJ, 2023. url: <http://www.bdt.d.uerj.br/handle/1/20528>.
- [33] Thauan da Silva CRUZ et al. "Automatizando a fiscalização de gastos públicos por meio da classificação automática de empenhos utilizando aprendizado de máquina". In: *Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação)* (2021). url: <http://repositorio.ifg.edu.br:8080/handle/prefix/1094>.
- [34] Genaro Dueire Lins, Diego Asensi Tavares Marcelo, et al. "Parametrização de despesas municipais e detecção de anomalias". MA thesis. Fundação Getúlio Vargas, FGV, Rio

- de Janeiro, 2021. url: <http://bibliotecadigital.economia.gov.br/handle/123456789/526430>.
- [35] Tom M. Mitchell. *Machine Learning*. 1st ed. McGraw-Hill Education, 1997. isbn: 978-0070428072.
- [36] Teresa Bernarda Ludermir. "Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências". In: *Estudos Avançados* v. 35.n. 101 (Apr. 2021), pp. 85–94. issn: 0103-4014. doi: 10.1590/s0103-4014.2021.35101.007. url: <http://dx.doi.org/10.1590/s0103-4014.2021.35101.007>.
- [37] Marcos Bemquerer Costa and Patrícia Reis Leitão Bastos. "Alice, Monica, Adele, Sofia, Carina e Ágata: o uso da inteligência artificial pelo Tribunal de Contas da União". In: *Controle Externo: Revista do Tribunal de Contas do Estado de Goiás, Belo Horizonte*, ano v. 2 (2020), pp. 11–34.
- [38] Tribunal de Contas da União. *Zello | Portal TCU — portal.tcu.gov.br*. <https://portal.tcu.gov.br/data/pages/8A81881F77D527280177D58177E75C93.htm>. [Accessed 18-01-2024]. 2020.
- [39] Tribunal de Contas da União. *Artificial Intelligence boosts productivity in the Federal Court of Accounts | Portal TCU — portal.tcu.gov.br*. [https://portal.tcu.gov.br/en\\_us/imprensa/news/artificial-intelligence-boosts-productivity-in-the-federal-court-of-accounts.htm](https://portal.tcu.gov.br/en_us/imprensa/news/artificial-intelligence-boosts-productivity-in-the-federal-court-of-accounts.htm). [Accessed 19-01-2024]. 2018.
- [40] Tribunal de Contas da União. *Em seminário internacional, presidente do TCU defende o uso de ciência e tecnologia no combate à corrupção | Portal TCU — portal.tcu.gov.br*. <https://portal.tcu.gov.br/imprensa/noticias/em-seminario-internacional-presidente-do-tcu-defende-o-uso-de-ciencia-e-tecnologia-no-combate-a-corrupcao.htm>. [Accessed 19-01-2024]. 2023.
- [41] Pete Chapman et al. "CRISP-DM 1.0: Step-by-step data mining guide". In: *SPSS* (1999).
- [42] Rüdiger Wirth and Jochen Hipp. "CRISP-DM: Towards a standard process model for Data Mining". In: *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*. Vol. v. 1. Manchester. 2000, pp. 29–39.
- [43] Fernando Martínez-Plumed et al. "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories". In: *IEEE Transactions on Knowledge and Data Engineering* v. 33.n. 8 (2021), pp. 3048–3061. doi: 10.1109/TKDE.2019.2962680.
- [44] Alex Souza. *Metodologia Crisp-DM: Uma Abordagem Abrangente Para projetos de Dados*. July 2023. url: <https://medium.com/blog-do-zouza/metodologia-crisp-dm-uma-abordagem-abrangente-para-projetos-de-dados-d7e7135b907e>.
- [45] IBM. *CRISP-DM: Data Understanding Overview*. Feb. 2023. url: <https://www.ibm.com/docs/en/spss-modeler/18.4.0?topic=understanding-data-overview>.
- [46] IBM. *CRISP-DM: Data Preparation Overview*. Feb. 2023. url: <https://www.ibm.com/docs/en/spss-modeler/18.4.0?topic=preparation-data-overview>.
- [47] IBM. *CRISP-DM: Modeling Overview*. Feb. 2023. url: <https://www.ibm.com/docs/en/spss-modeler/18.4.0?topic=modeling-overview>.
- [48] Nick Hotz. *What is CRISP DM?* Apr. 2024. url: <https://www.datascience-pm.com/crisp-dm-2/>.
- [49] IBM. *CRISP-DM Help Overview*. Feb. 2023. url: <https://www.ibm.com/docs/en/spss-modeler/18.4.0?topic=dm-crisp-help-overview>.
- [50] Nick Hotz and Jeff Saltz. *DSPA: About us*. url: <https://www.datascience-pm.com/about/>.

- [51] Jeff Saltz. *CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects*. Apr. 2024. url: <https://www.datascience-pm.com/crisp-dm-still-most-popular/>.
- [52] Khushboo Taneja and Jyoti Vashishtha. "Comparison of Transfer Learning and Traditional Machine Learning Approach for Text Classification". In: *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, Mar. 2022, pp. 195–200. doi: 10.23919/indiacom54597.2022.9763279. url: <http://dx.doi.org/10.23919/INDIACom54597.2022.9763279>.
- [53] Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. "BERTimbau: pretrained BERT models for Brazilian Portuguese". In: *Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, October 20–23, 2020, Proceedings, Part I 9*. Springer. 2020, pp. 403–417.
- [54] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. arXiv: 1810.04805 [cs.CL]. url: <https://arxiv.org/abs/1810.04805>.
- [55] Isabel Nadine de Santana, Raphael Souza de Oliveira, and Erick Giovani Sperandio Nascimento. "Text Classification of News Using Deep Learning and Natural Language Processing Models Based on Transformers for Brazilian Portuguese". In: *Proceedings of the 26th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2022*. Vol. v. 3. WMSCI 2022. International Institute of Informatics and Cybernetics, July 2022, pp. 134–139. doi: 10.54808/wmsci2022.03.134. url: <http://dx.doi.org/10.54808/WMSCI2022.03.134>.
- [56] Tong Xiao and Jingbo Zhu. *Introduction to Transformers: an NLP Perspective*. 2023. arXiv: 2311.17633 [cs.CL]. url: <https://arxiv.org/abs/2311.17633>.
- [57] University San Diego. *7 reasons to learn Python for Data Science*. Oct. 2024. url: <https://onlinedegrees.sandiego.edu/python-for-data-science/>.
- [58] Lars Buitinck et al. "API design for machine learning software: experiences from the scikit-learn project". In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*. 2013, pp. 108–122.
- [59] Rishabh Choudhary and Hemant Kumar Gianey. "Comprehensive Review On Supervised Machine Learning Algorithms". In: *2017 International Conference on Machine Learning and Data Science (MLDS)*. 2017, pp. 37–43. doi: 10.1109/MLDS.2017.11.
- [60] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* v. 12 (2011), pp. 2825–2830.
- [61] ESCEX - Escola Superior de Controle Externo TCE-MA. *Apresentação do Sistema de Informações para Controle (SINC)*. São Luís, MA, 2021. url: <https://www.youtube.com/live/u0Yp9pHpwgk?si=sQbtNvOyT-HNVIHF>.
- [62] TCE-MA. *Planejamento estratégico*. 2022. url: [https://www.tcema.tc.br/transparencia\\_novo/index.php/institucional/planejamento-estrategico](https://www.tcema.tc.br/transparencia_novo/index.php/institucional/planejamento-estrategico).
- [63] Alisson Kemis Araújo et al. "A IMPORTÂNCIA ESTRATÉGICA DO EMPENHO, LIQUIDAÇÃO E PAGAMENTO PARA A ORGANIZAÇÃO DO SETOR PÚBLICO". In: *EmpíricaBR-Revista Brasileira de Gestão, Negócio e Tecnologia da Informação* v. 1.n. 1 (2018), pp. 148–166.
- [64] Brazilian Government. *Secretaria do Tesouro Nacional (STN) - versão: 2.0 Documentação*. url: <https://siconfi.tesouro.gov.br/siconfi/pages/public/conteudo/conteudo.jsf?id=12503>.
- [65] BRAZIL. *Função e subfunção*. url: <https://portaldatransparencia.gov.br/entenda-a-gestao-publica/funcao-e-subfuncao>.

- [66] Brasil. “Decreto nº 64.752, de 27 de Junho de 1969”. In: *Diário Oficial [da] República Federativa do Brasil* (1969).
- [67] UFSB. 2021. url: [https://ufsb.edu.br/propa/images/CCF\\_DIRPLAN/alterar\\_empenho\\_comprasnet.pdf](https://ufsb.edu.br/propa/images/CCF_DIRPLAN/alterar_empenho_comprasnet.pdf).
- [68] TCE Maranhão. *ANEXO I SISTEMA DE INFORMAÇÕES PARA CONTROLE MÓDULO FISCAL 2023*. url: <https://app.tcema.tc.br/publicacao/#/documentohtml/14991?compilado=true>.
- [69] Maria Cristina Cordeiro Sousa. “Uma análise do algoritmo K-means como introdução ao aprendizado de máquinas”. In: *Monografia Graduação - (Curso de Matemática)* (2019). url: <http://hdl.handle.net/11612/4511>.
- [70] European Union. *GDPR: General Data Protection Regulation*. May 2018. url: <https://gdpr-info.eu/>.
- [71] Brasil. “Lei Geral de Proteção de Dados Pessoais (LGPD)”. In: *Diário Oficial [da] República Federativa do Brasil* (2018).
- [72] Guilherme Fonseca, Washington Cunha, and Leonardo Rocha. “Análise Comparativa de Métodos de Undersampling em Classificação Automática de Texto Baseada em Transformers”. In: *Revista Eletrônica de Iniciação Científica em Computação* v. 22.n. 1 (June 2024), pp. 1–10. doi: 10.5753/reic.2024.4643. url: <https://journals-sol.sbc.org.br/index.php/reic/article/view/4643>.
- [73] Roberta Rodrigues de Lima et al. “An Empirical Comparison of Portuguese and Multilingual BERT Models for Auto-Classification of NCM Codes in International Trade”. In: *Big Data and Cognitive Computing* v. 6.n. 1 (2022). issn: 2504-2289. url: <https://www.mdpi.com/2504-2289/6/1/8>.
- [74] F.C. Souza, R.F. Nogueira, and R.A. Lotufo. “BERT models for Brazilian Portuguese: Pretraining, evaluation and tokenization analysis”. In: *Applied Soft Computing* v. 149 (Dec. 2023), p. 110901. issn: 1568-4946. doi: 10.1016/j.asoc.2023.110901. url: <http://dx.doi.org/10.1016/j.asoc.2023.110901>.
- [75] Luiz Hammes and Larissa Freitas. “Utilizando BERTimbau para a Classificação de Emoções em Português”. In: *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*. Evento Online: SBC, 2021, pp. 56–63. doi: 10.5753/stil.2021.17784. url: <https://sol.sbc.org.br/index.php/stil/article/view/17784>.
- [76] Rafael Silva Barbon and Ademar Takeo Akabane. “Towards Transfer Learning Techniques—BERT, DistilBERT, BERTimbau, and DistilBERTimbau for Automatic Text Classification from Different Languages: A Case Study”. In: *Sensors* v. 22.n. 21 (2022). issn: 1424-8220. doi: 10.3390/s22218184. url: <https://www.mdpi.com/1424-8220/22/21/8184>.



## Appendix A

# Government Functional Codes

In this appendix, Table A.1 presents a comprehensive list, presented of all applicable Government Function and Sub-Function Codes classifications for a Commitment Note, based on the guidelines and regulations of the Brazilian Government for the year of 2023 [64].

Table A.1: Government Function and Sub-Function Codes for Commitment Notes

<b>Government Function Code</b>	<b>Government Sub-Function Code</b>
01 - Legislative	031 - Legislative Action 032 - External Control
02 - Judiciary	061 - Judicial Action 062 - Defense of Public Interest in Judicial Process
03 - Essential to Justice	091 - Defense of Legal Order 092 - Judicial and Extrajudicial Representation
04 - Administration	121 - Planning and Budgeting 122 - General Administration 123 - Financial Administration 124 - Internal Control 125 - Regulation and Supervision 126 - Information Technology 127 - Territorial Planning 128 - Human Resource Development 129 - Revenue Administration 130 - Concession Administration 131 - Social Communication
05 - National Defense	151 - Air Defense 152 - Naval Defense 153 - Land Defense
06 - Public Security	181 - Policing 182 - Civil Defense 183 - Information and Intelligence
07 - Foreign Relations	211 - Diplomatic Relations 212 - International Cooperation
08 - Social Assistance	241 - Assistance to the Elderly 242 - Assistance to Disabled Persons 243 - Assistance to Children and Adolescents 244 - Community Assistance
09 - Social Security	271 - Basic Social Security

<i>(Continued from previous page)</i>	
<b>Function Code</b>	<b>Sub-Function Code</b>
	272 - Statutory Regime Social Security 273 - Complementary Social Security 274 - Special Social Security
10 - Health	301 - Primary Care 302 - Hospital and Ambulatory Assistance 303 - Prophylactic and Therapeutic Support 304 - Sanitary Surveillance 305 - Epidemiological Surveillance 306 - Food and Nutrition
11 - Labor	331 - Worker Protection and Benefits 332 - Labor Relations 333 - Employability 334 - Job Promotion
12 - Education	361 - Elementary Education 362 - Secondary Education 363 - Professional Education 364 - Higher Education 365 - Early Childhood Education 366 - Youth and Adult Education 367 - Special Education 368 - Basic Education
13 - Culture	391 - Historical, Artistic, and Archaeological Heritage 392 - Cultural Diffusion
14 - Citizenship Rights	421 - Custody and Social Reintegration 422 - Individual, Collective, and Diffuse Rights 423 - Assistance to Indigenous Peoples
15 - Urban Development	451 - Urban Infrastructure 452 - Urban Services 453 - Urban Collective Transport
16 - Housing	481 - Rural Housing 482 - Urban Housing
17 - Sanitation	511 - Rural Sanitation 512 - Urban Sanitation
18 - Environmental Management	541 - Environmental Preservation and Conservation 542 - Environmental Control 543 - Recovery of Degraded Areas 544 - Water Resources 545 - Meteorology
19 - Science and Technology	571 - Scientific Development 572 - Technological Development and Engineering 573 - Dissemination of Scientific and Technological Knowledge
20 - Agriculture	601 - Promotion of Crop Production 602 - Promotion of Animal Production 603 - Plant Health Protection 604 - Animal Health Protection 605 - Supply

Appendix A. Government Functional Codes

<i>(Continued from previous page)</i>	
<b>Function Code</b>	<b>Sub-Function Code</b>
	606 - Rural Extension 607 - Irrigation 608 - Promotion of Agricultural Production 609 - Agricultural Defense
21 - Agrarian Organization	631 - Agrarian Reform 632 - Colonization
22 - Industry	661 - Industrial Promotion 662 - Industrial Production 663 - Mining 664 - Industrial Property 665 - Standardization and Quality
23 - Commerce and Services	691 - Trade Promotion 692 - Commercialization 693 - Foreign Trade 694 - Financial Services 695 - Tourism
24 - Communications	721 - Postal Communications 722 - Telecommunications
25 - Energy	751 - Energy Conservation 752 - Electric Energy 753 - Fossil Fuels 754 - Biofuels
26 - Transport	781 - Air Transport 782 - Road Transport 783 - Rail Transport 784 - Water Transport 785 - Special Transport
27 - Sports and Recreation	811 - High-Performance Sports 812 - Community Sports 813 - Recreation
28 - Special Charges	841 - Refinancing of Domestic Debt 842 - Refinancing of External Debt 843 - Domestic Debt Service 844 - External Debt Service 845 - Other Transfers 846 - Other Special Charges 847 - Transfers for Basic Education
99 - Reserves	997 - RPPS Reserve 999 - Contingency Reserve