CLASSIFICATION OF PROVISIONAL OPINIONS THROUGH THE MUNICIPAL MANAGEMENT EFFECTIVENESS INDEX

Classificação de pareceres de contas em função do Índice de Efetividade da Gestão Municipal

Otoniel Arruda Costa
Bachelor of Naval Sciences/Administration from the Naval School, specialist in International Accounting Standards from FEARP-USP (MBA), and specialist in Data Science and Analytics (MBA). Specialist in Intendancy Systems from the Admiral Wandenkolk Instruction Center (Brazilian Navy) and civil servant at the State Court of Audit of São Paulo.
Orcid: https://orcid.org/0000-0001-5346-7374
Lattes: https://lattes.cnpq.br/2101791643993928

Henrique Raymundo Gióia
Master's degree and doctoral student in Applied Economics from ESALQ-USP.
Orcid: https://orcid.org/0000-0002-8639-1481
Lattes: http://lattes.cnpq.br/5957983550791571
E-mail: henriquerg@usp.br

ABSTRACT

One of the main constitutional attributions of the courts of accounts (CA) is to evaluate the management of public administration bodies. Thus, the Municipal Management Effectiveness Index (IEGM in Portuguese) was created in 2014 with the purpose of measuring the quality of the expenses made by these same agencies. This new audit format carried out by these courts aims to provide a greater return to society than that obtained with the *posteriori* inspections. The objective of the research was to create a model through which the seven dimensions of the IEGM, in addition to the final score of this index, could predict whether the municipality would receive a provisional favorable or unfavorable opinion from Sao Paulo State Court of Accounts (TCE-SP). With this, it aims to identify municipalities with greater management difficulties and, in this way, allow more intense preventive inspection actions on them. With the use of machine learning techniques, it was possible to create a classification tree that made this prediction. Among the results of the research is the achievement of an accuracy between 63 and 71%, of a sensitivity higher than 76%, which allows to detect potentially failed municipalities, and of a specificity above 75%. Another advantage of the model was the identification of the variables that had greater information gain or lower entropy, especially the IFISCAL. Among the limitations is the fact that the main package to identify the optimal hyperparameters, the *mlr*, is in the deactivation phase to make way for a new package.

Keywords: Accuracy; ROC curve; R language; Sensitivity; Unfavorable.

RESUMO

Uma das principais atribuições constitucionais dos tribunais de contas é a de avaliar a gestão dos órgãos da administração pública. Assim, o Índice de Efetividade da Gestão Municipal (IEG-M) foi criado em 2014 com a finalidade de mensurar a qualidade dos gastos realizados por esses mesmos órgãos. Esse novo formato de auditoria procedido por essas cortes visa proporcionar um maior retorno para a sociedade do que aquele obtido com as fiscalizações a posteriori. O objetivo desta pesquisa foi criar um modelo por meio do qual as sete dimensões do IEG-M, além da nota final desse índice, pudessem prever se o município receberia um parecer prévio favorável ou desfavorável da Corte de Contas de São Paulo (TCE-SP). Com isso, visou-se identificar municípios com maiores dificuldades de gestão e, dessa forma, permitir a promoção de ações fiscalizatórias preventivas mais intensas sobre eles. Com a utilização das técnicas de aprendizado de máquina, foi possível criar uma árvore de classificação que efetuasse essa previsão. Entre os resultados da pesquisa está a obtenção de uma acurácia entre 63 a 71%, de uma sensibilidade superior a 76%, o que permite detectar municípios potencialmente reprovados, e de uma especificidade acima de 75%. Outra vantagem do modelo foi a identificação das variáveis que possuíam maior ganho de informação ou menor entropia, com destaque para o iFISCAL. Entre as limitações à pesquisa nesse tema inclui-se o fato de que o principal pacote para identificar os hiperparâmetros ótimos, o mlr, está em fase de desativação para dar lugar a um novo pacote.

Palavras-chave: *acurácia; curva ROC; desfavorável; linguagem R; sensibilidade.*

1. INTRODUCTION

The Federal Court of Accounts (TCU), the State Courts of Accounts (CA), the Federal District Court of Accounts (TCDF) and the Municipal Court of Accounts (TCM) have, in accordance with the Federal Constitution of Brazil, the mission of assisting the Legislative Power in the accounting, financial, budgetary, operational and asset supervision of the Union, States, Federal District, and Municipalities. To meet this obligation, CAs are responsible for issuing opinions (called Provisional Opinions or Auditors' Opinions) on the accountability of the management of the Heads of the Executive Branch for subsequent consideration by the respective Legislative Houses. To issue these opinions, the technical bodies of these CAs develop actions to analyze these accounts. To this end, some of the main legal instruments that guide them are Federal Law No. 4,320/64 and Federal Law No. 101/2000 (Fiscal Responsibility Law). It happens that part of these technical works are based predominantly on analysis of acts and administrative and accounting facts already consummated by public agencies (*posteriori* control). In this context, Silva Junior (2008) and Willeman (2016) highlight the low effectiveness that the subsequent control has in the preservation of public assets. Willeman (2016) adds that there is a need to review the Brazilian external control model, in order to confer greater effectiveness in the processes of accountability and operational audits. The IEGM is an example of these improvements. Created in 2014 by the Sao Paulo State Court of Accounts (TCE-SP, 2020), the IEGM is an monitoring technique used by Brazilian CAs to evaluate the management of the agencies in their spheres of action in matters of quality of government spending for the benefit of citizens. Araújo, Bezerra Filho and Motoki (2019) point out that the IEGM has the adhesion of the 27 CAs of the States and Federal District. That said, the IEGM is represented by a series of mathematical calculations in seven thematic dimensions: education (i-Educ), health (i-Health), fiscal management (i-Fiscal), planning (i-Plan), environment (i-Amb), civil defense (i-City) and governance in information technology (i-Gov-TI). Thus, according to the IEGM manual of the 2020 exercise of the CA object of this study, in each thematic dimension the municipalities can be evaluated in five ranges of results, namely: A (highly effective); B+ (very effective); B (effective); C+ (in the adaptation phase), and C (low level of adequacy). In addition to these dimensions, the sum of all these indicators results in a final index, called IEGM (referred to in this study as the general or final IEGM so as to avoid confusion). For calculating the grades of each dimension, there are two sources of management data. The first source comprises responses to questionnaires about public policies answered by the municipalities in online platforms provided by the CA. After the data is sent, officials of these audit bodies assess the reliability and veracity of the answers presented. The other source refers to the management data that city administrations send to CA databases. In a later stage, based on this information, the municipalities' IEGM is calculated. For each dimension, the points of the various items (each has a different score) are added and divided by a thousand (/1000). The final result or index (IEGM) is defined by the sum of the grades of the weighted dimensions. Once the IEGM values are obtained, the CA analyzes the rendering of accounts, issues the reports of authorship and, lastly, issues the previous opinions about the accounts of each of these bodies. As for the CA examined in this research, it audits all the State's municipalities each year, which includes measuring each of the municipalities' IEGM.

As for the preponderant factors for the assessment of the accounts of public agencies, Rodrigues (2022) analyzed the relationship between the IEGM and previous opinions in the municipalities of the state of Rio Grande do Norte. In that study it was not possible to identify a positive correlation between the results of the IEGM and the opinion for the approval of the accounts of the municipalities of that State, since a limiting aspect of the research was the high rate of disapproval of accounts. In addition, the study concluded that the financial surplus, public revenues, and population size were not determinant for the outcome of the final IEGM. On the other hand, the research of Maeda and Varela (2017) showed that there is a link between the IEGM (final grade) and the accounts opinions in the state of São Paulo in the 2014 fiscal year. In another study, Macieira (2016) examined the determinant variables for the issuance of provisional opinions by the TCE-SP between 2008 and 2013. For the author, the quantitative variable of the population has a positive correlation in relation to the previous unfavorable opinion, while the revenues have a negative correlation. In view of these findings, the importance of the IEGM in the oversight performance of the CA is attested. However, the current literature is scarce regarding predictive analyses of external control. In other perspective, with the adoption of mathematical and statistical models it is possible to direct efforts to improve this scenario. One aspect of these innovations is the possibility of applying machine learning techniques to evaluate the results of city administrations. For Géron (2021), machine learning is the programming science in which computers can learn from data.

Thus, the objective of this research is to develop a machine learning model that makes it possible to predict whether municipalities will have their accounts approved or rejected due to the grades of the seven dimensions of the IEGM and the final IEGM, in order to enable the Court of Accounts in question to identify the dimensions of the IEGM that have greater predictive power over their previous opinions. Therefore, it aims to enhance the planning and execution process of audit work for these CA.

2. MATERIAL AND METHODS

In the development of the study, the structure proposed by Volpato (2015, p. 11) was adapted. Thus, the first section describes the object of the research. In the following section, the methodology is presented according to its classifications and, subsequently, the target population. Section 2.4 shows how data was collected. Section 2.5 is dedicated to data analysis. And in the last section, limitations of the study are expressed.

2.1 Characterizing the object of the research

For the accomplishment of this study, the state of São Paulo and all its municipalities were selected (except for the state capital, which has its own CA). These federative entities – as they are called – have, according to Paulo and Alexandrino (2013), important characteristics granted by the Constitution: capacity for self-organization and self-legislation (state constitutions and organic laws), self-government (elections for governor, mayors, and members of the Legislative Power) and self-administration, which means the administrative, legislative, and tax capacity independently of other entities (Paulo; Alexandrino, 2012). On the other hand, Brazilian states and municipalities are subject to accounting, financial, budgetary, operational and asset supervision, as provided for in articles 70 to 75 of the Federal Constitution. In the case of the federative entity analyzed in this study, this oversight power is posited in articles 32 to 36 of the São Paulo State Constitution. The norm highlights the provision granting this CA the authority to issue provisional opinions on the rendering of accounts of the municipalities under its jurisdiction. In the State in question, there are 645 (six hundred and forty-five) municipalities and their respective administrations. All these bodies – except for the state capital – are audited annually by the CA. Among the techniques employed in this audit work is the IEGM. This consists of a methodology for calculating management indicators in which, according to Rodrigues (2022), TCE-SP (2020) and Amorim (2017), the quality of public spending is assessed. To this end, public policies are evaluated in the following dimensions, according to the IEGM Manual of the 2017 fiscal year of the TCE-SP:

- I-Educ: This dimension measures the provision of services at various educational levels, especially in early childhood and elementary education, evaluates the promotion of pedagogical programs, such as those related to reading, writing and evaluation tests of students, teachers and other education professionals. Logistical aspects, such as food, uniforms and school transportation, are also evaluated. The existence of diagnostic works on the school infrastructure is verified; for example, whether there are surveys about the demand for vacancies in teaching units, whether there are schools with interrupted operation and whether units need repairs. Anti-bullying policies, performance of the councils, and accessibility are also evaluated, among other matters.

- I-Saúde (Health): It assess the public health infrastructure in the municipalities by verifying the number of basic units, equipment and professionals. In addition, it measures whether there is promotion and implementation of public health programs through consultations and surgeries. It evaluates whether information systems are used as a means of interaction between the municipal and the state and federal levels of the Unified Health System (SUS), among others.

- I-Plan: This index aims to verify whether the Pluriannual Plan (PPA), the Budget Guidelines Act (LDO) and the Annual Budget Act (LOA) – which are laws provided for in the Federal Constitution as instruments of mandatory service by the Public Power with regard to the promotion of public policies to citizens – have been elaborated and followed and whether the proposed results were achieved. In addition, questions about the formal structure of the planning sector regarding the existence and effective use of information systems are foreseen in this dimension, also considering the degree of training of servers in this sector. Checks are adopted regarding the holding of public hearings for the survey of problems of the municipalities by the prefectures and also regarding the monitoring of the execution of the budget by those interested in public management (mayors, internal controls, citizens, and others).

- I-Fiscal: It evaluates the level of collection of taxes and other types of revenues, the realization of expenses and indebtedness, assets and liabilities, personnel expenses, among others.

- I-Amb (Environment): This aspect aims to measure the solid waste management policies of the municipalities, the management of environmental preservation, if present, within the audited regions, the management of water and sewage, the promotion of educational initiatives in the environmental area, as well as the adherence to those already existing in other spheres of government, the evaluation of environmental licensing policies, among others.

- I-Cidade (City): It measures the structure of civil defense in the municipality, the management of urban mobility and public roads, the level of security in schools and health posts, among others.

- I-Gov-TI (Information Technology): It seeks to assess the structuring of information technology in municipal administration, the level of transparency of government data, including personnel expenses, revenues and expenses, assets and liabilities, as well as public tenders, among others.

According to the TCE-SP, the IEGM can be summarized as showed in Figure 1.

Figure 1 – IEGM Dimensions

Source: TCE-SP (2022).

Once the grades in each dimension are calculated, the IEGM score of the municipality is obtained by means of the following mathematical formula, the result of which is called IEGM.

IEGM = [(i-Plan + i-Fiscal + i-Educ + i-Health) x 20 + i-Amb x 10 + i-City x 5 + i-Gov-TI x 5] / 100. The processing and dissemination of these results are defined independently within the scope of each CA in Brazil. In the case under study, these results are kept by the CA and disclosed only at the time of the rendering of accounts of the municipalities. Both the grades of the dimensions and the final grade of the IEGM are taken into account in the assessments.

As mentioned in this topic, Brazilian municipalities have the power of self-administration. However, they are subject to scrutiny by the CA, which, among their various constitutional powers, is to issue opinions on the accounts of mayors to the respective Legislative Houses, which will then make a final decision on the management of these political agents. In this context, the IEGM was established as a technique that allows the evaluation of several dimensions of municipal administration, in order to add elements of measurement of the quality of public expenditures and not only evaluations of administrative facts already consummated, which prove to be ineffective for the process of accountability to society. Lastly, the results of these IEGM represent one of the components taken into account in the opinions issued on the management of the municipalities.

2.2 Research methodology

This study is framed within the concept of Jurimetry, defined by Garcia (2021) as the application of statistical sciences within the legal domain. Consequently, it adopts a quantitative research approach, as emphasized by the quantifiable nature of the predictive model construction (Fonseca, 2002; Silveira; Córdova, 2009). In terms of its nature, the research is classified as applied (Silveira; Córdova, 2009) due to its potential practical utility. Regarding objectives, it aligns with an explanatory framework (Gil, 2007; Silveira; Córdova, 2009), aiming not only to identify values conducive to formulating an optimal classification model for account opinions based on the IEGM but also to elucidate, in the context of existing literature, the dimensions of the IEGM that most significantly influence the predictive process of municipal account judgments. In terms of methodology, the study adopts an experimental design, involving the formulation of problem statements and hypotheses that define the variables affecting the phenomenon under investigation (Triviños, 1987; Silveira; Córdova, 2009), which are then subjected to various statistical tests.

2.3 Definition of the target population

Regading the target population, all 644 municipalities within the State of São Paulo were selected, all falling under the jurisdiction of the same CA. Since data on the IEGM for this federative entity were readily available, it was possible to examine the behavior of the entire population. All IEGM data collected refers to the 2017 fiscal year, and all municipalities within this entity have been assessed since the implementation of this management evaluation methodology. However, it is important to note a caveat. Due to federative autonomy, some Brazilian states adopted this indicator for different fiscal years, which may affect the replication of this study with other CAs. Additionally, certain states did not mandate all municipalities to participate, which can further limit the study's replicability. For example, in the State of Amazonas, the obligation to complete IEGM questionnaires became mandatory only from 2019, as stated in item I, article 3 of Resolution No. 03/2019. Consequently, due to these constraints related to certain CAs and inspection exercises, some replications of this experiment may rely solely on sample analysis, which may represent a limitation of the predictive power of the models and, by extension, their conclusions.

2.4 Data collection

The data collection instrument used was documentary research conducted on the World Wide Web. All necessary data for the research were exclusively sourced from the official website of the CA, including both individualized IEGM data and individualized account opinion data for the 2017 fiscal year. Therefore, no internal data from the studied organization was collected. One limitation to note is that the list of previous opinions does not include information on whether the trial results have already considered any appeals filed by the municipalities or the final judgments of these appeals. To develop the predictive model of account opinions, only two files needed to be collected: one with IEGM data for all municipalities and another with account opinion data, both relating to the 2017 fiscal year. These files were obtained in PDF format and converted into spreadsheet format using an online document format converter. Subsequently, the spreadsheet with IEGM results for the corresponding municipalities was filled with the account opinion data through the "Opinion" column. Table 1 provides an illustration of this merging of spreadsheets. It is important to note that the "Position" and "Municipality" columns serve purely as counting and organizational tools and do not imply any form of ranking among the municipalities analyzed.

Table 1 – Results of the IEGM and the Preliminary Opinion of the CA by municipality

Pos.	Municipality	iEduc	iSaude	iPlan	iFiscal	iAmb	iCidade	iGovTI	IEGM	Provisional Opinion
1	City 1	C+	B	B	B	B	C	C+	B	Favourable
(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)
N	City N	A	B+	B	C+	C	B	B+	B	Unfavorable

Source: the authors.

After this phase of data collection, they were evaluated.

2.5 Data analysis

Once the goal of the study was defined, the subject and methodology were known, the target population was determined and data were collected, it was initially necessary to provide context for the entire process of examining the problem and, within it, develop the solution. To achieve this, the following steps were followed, adopting and adapting the flows proposed by Géron (2021):

1. The problem was studied:

   1.1 The independent and dependent variables of the object of the study were established.

2. The methodology for addressing the problem was outlined.
3. The *software* used for developing the research's solution was selected.
4. Data was treated and the script was developed and trained.
5. Inspection was carried out to identify potential data updates and debug errors, making the necessary corrections.
6. Algorithm and results were evaluated.
7. The Solution (classification model) was made available.

As already discussed in the preceding topic, the subject of the problem was defined as the municipalities, with the aim of estimating or predicting in advance the outcomes of the account rendering judgments based on the evaluations of their respective IEGM. The established objective is to enhance audit and inspection processes and, in this way, contribute to the municipal administration by indicating possible points of improvement in local management.

With that in mind, the dimensions of the IEGM and their respective grades were established as independent variables, namely: i-Educ, i-Health, i-Plan, i-Fiscal, i-Amb, i-City, i-Govti, and IEGM (the latter in relation to the final score of the weighted indicators). In each of these dimensions, municipalities were assigned scores indicating the quality of management or public spending, ranging from “A” to “C” grades: “A”, “B+”, “B”, “C+”, or “C”. The dependent variable, on the other hand, is the outcome of the judgment of the municipalities’ accounts, which can be categorized as “favorable” (also known as ‘accepted’ or ‘approved’ provisional opinion) or “unfavorable” (also referred to as ‘rejected’ provisional opinion).

The methodology adopted was the “classification or decision tree”. For Géron (2021), classification or decision trees (known within the concept of *Classification and Regression Tree* [CART]) are algorithms that allow to classify data and events (from a statistical point of view) through machine learning (also called training). In other words, from historical data, a system can make generalizations through established algorithms. According to the author, the goal is to achieve accurate predictions on new data based on the training process. In this context, the experiment consisted of creating a classification tree that, based on the results of the IEGM in each dimension of each of the municipalities, would be able to predict whether their accounts would be assessed in a favorable or unfavorable manner by the CA.

In the decision tree, there is a process of branching the data that contributes most to the predictive power of this model. These instances that aid in prediction are referred to as nodes, branches, and leaves (Lauretto, 2010). Each node is constructed from attributes or values that, recursively, contribute to the predictive power of the information under analysis (in this case the result of the judgment of the accounts in favorable or unfavorable). The first node is called the “root node” and the last of a branch is called the “terminal node” or “leaf.” According to Stankevix (2019), the terminal node is where the criteria on which the model response is obtained are found. For the same author, the branches constitute the connection between two nodes.

In this study, each of these nodes indicated which combinations of IEGM dimensions (independent variables) produced a “favorable” or “unfavorable” (dependent variable) result. The purpose of the model was to predict with the utmost accuracy whether a municipality would have its annual administration rejected or approved in the Court’s provisional opinion.

It should be noted, as Cristiano (2017) asserts, that accuracy is not a widely used measure, as it varies with prevalence (a metric that analyzes changes in study variables over time). However, the author suggests that accuracy may be relevant when evaluating the rate of false negatives in diagnostic tests. For instance, if a patient receives a diagnosis of being healthy but is actually a carrier of the disease, this poses risks to their health.

In this same sense, corroborating the research, this study was based on two other important concepts in the prediction process: sensitivity and specificity. Sensitivity refers to the correctness of an event about which one wishes to predict the occurrence when the event comes to occur. Specificity refers to the accuracy of predicting a non-event, which is subsequently confirmed in reality. Accuracy, in turn, is obtained by the sum of these two quantities. Within the same model different combinations are possible of sensitivities and specificities or of true and false events and non-events. The Manual of the caret package shows how the various combinations of these quantities are generated and calculated, through what is called the Confusion Matrix (Figure 2).

Figure 2 – Confusion matrix

	Reference	
Predicted	Event	No Event
Event	A	B
No Event	C	D

The formulas used here are:

$$Sensitivity = A/(A + C)$$
$$Specificity = D/(B + D)$$

Source: Caret Package Manual (2022).

As it turns out, value ‘A’ represents the prediction of an event that has been confirmed in the real world, while ‘B’ denotes the prediction of an event that, on the contrary, did not occur. Result ‘C’ indicates the prediction that the event would not have occurred, but in reality, it did. Lastly, ‘D’ signifies an actual non-event that was accurately predicted as such. Sensitivity is given by A/(A+C) and specificity is D/(B+D).

The comparative analysis of sensitivity and specificity also involved the calculation of the false positive rate, that is, situations that are characterized by the occurrence of a non-event when the occurrence of an event was predicted (letter B of Figure 2). This rate is known as “1-specificity” because it is obtained precisely by the difference 1-D/(B+D). In this research, the event was defined, in compliance with the manual of the rpart package and according to the calculations made – independently and autonomously from the researcher’s control – that the event was set as being unfavorable opinion (“*parecer desfavorável*” in Portuguese) and the non-event was defined as favorable opinion (“*parecer favorável*” in Portuguese).

As previously clarified, the objective of constructing the model was to maximize accurate predictions of which municipal accounts would be rejected, while minimizing incorrect predictions that accounts would be approved when, in fact, they would have been rejected. This is crucial because such errors could result in the exclusion of certain municipalities from inspection procedures during audit planning, leading to societal costs in terms of efficiency, effectiveness, and economy of public policies. The goal was to develop a model with the highest possible accuracy rates, sensitivity, and specificity.

In addition to this analysis, the plotting of the various combinations of sensitivities and “1-specificities” in the Cartesian plane allowed the generation of a type of graph called “ROC curve”, whose area is called AUC (Area Under Curve). According to Fávero and Belfiore (2021), the larger the AUC area of the ROC curve, the higher the quality of the model. Cristiano (2017) adds that when the values of sensitivity and “1-specificity” approach equality, the lower the discriminating power of the diagnostic test. Thus, the algorithm development and consequently, the model under study, assessed these combinations of accuracy, sensitivity, and specificity.

After these considerations, data analysis was performed in the RStudio software (version 2022.12.0 Build 353), which, according to Alcoforado (2021), is a statistical and graphic computer system developed with the purpose of allowing the user to build predictive models through machine learning, among other functionalities, which are made possible through the R language. This system has some basic functions and others that must be installed on the computer on which the data analysis models are developed. The author points out that they are provided in "packages," which can be purchased from an RStudio support network called *The Comprehensive R Archive Network* (CRAN). After these considerations, the data were imported – from the spreadsheet format (in the case of XLS) to R – and the new table obtained was saved in a file of type *. RDA. As a first action of data treatment, it was defined that each dimension of the IEGM would be a column (independent variables). The opinion of accounts (dependent variable) was fixed in the rightmost column of this same table. The municipalities, in turn, were characterized as observations in the table. From this point on, we moved on to the construction of the model, as already defined, through the method of machine learning by classification tree.

As an initial task (step 1), several code generating packages (also called algorithms or scripts) were researched on the World Wide Web for model construction, such as *stackoverflow* and *towardscience*. Among the packages researched – whose functionalities can be found in each of the respective manuals available in CRAN – the following were adopted:

1. *Dplyr*: It was used in the process of organizing the database, known as *datawrangling*. One of its functions, *mutate*, was utilized to reclassify IEGM indicators – converting them from character to factor variables.

2. *Rpart*: Package responsible for having generated the model (or tree). It is also equipped with the command that facilitated the creation of graphs based on the proposed model. Furthermore, it enabled the adjustment of hyperparameters, such as the *complexity parameter* (CP).

3. *Rpart.plot*: Package that provides extended functionality similar to the rpart package, including functions like *rpart.predict*, having been used in conjunction with *rpart*.

4. *Metrics*: Allowed to obtain the area under the ROC curve through the *auc* command.

5. *Mlr*: It is the package that enabled the optimized combination of hyperparameters (process called *tuning*). This resulted in maximizing the accuracy of the model in the classification technique.

6. *Ggplot2*: This package was used for graphics generation.

7. *Plotly*: Its main feature is creating graphics as ggplot2 does.

8. *Tidyverse*: This is a compilation of several packages that share a common design.

9. *Scales*: This package was used for the configuration of graphs, such as axes and legends, and was used together with ggploty2.

10. *Caret*: It was used for the training and testing functions of the model in this classification tree.

11. *PROC*: This is another package that allowed the calculation of the area under the ROC curve.

12. *Rmisc*: Allows the calculation *summarySE*, that is, calculates mean, standard deviation, among others.

The initial package required for the creation of the classification tree adopted in this work was rpart. This package has one of the most essential functions in the construction of this type of model. To obtain this tree, criteria were adopted that could maximize the ability to generalize or predict (hyperparameters). Figure 3 shows hyperparameters from the *getParametSet* ("*classif.rpart*") command. Note that the "Type" column indicates whether the variable is an integer or discrete, while the "Def" column displays the default values adopted by R when not defined by the user. The "Constr" column shows the range of possible numbers for each hyperparameter. Lastly, the "Tunable" column indicates whether the element is tunable or not.

Figure 3 – Hyperparameters of a classification tree

```
> getParamSet("classif.rpart")
                   Type len  Def   Constr Req Tunable Trafo
minsplit        integer   -   20 1 to Inf   -    TRUE     -
minbucket       integer   -    - 1 to Inf   -    TRUE     -
cp              numeric   - 0.01   0 to 1   -    TRUE     -
maxcompete      integer   -    4 0 to Inf   -    TRUE     -
maxsurrogate    integer   -    5 0 to Inf   -    TRUE     -
usesurrogate   discrete   -    2    0,1,2   -    TRUE     -
surrogatestyle discrete   -    0      0,1   -    TRUE     -
maxdepth        integer   -   30  1 to 30   -    TRUE     -
xval            integer   -   10 0 to Inf   -   FALSE     -
parms           untyped   -    -        -   -    TRUE     -
> |
```

Source: RStudio.

In the analysis of the functions and package of tree development, Stankevix (2019), Bernardo (2022), elucidated the meaning of the main hyperparameters based on the *rpart* Manual. The minsplit refers to the smallest number of samples required on the root node. By default, it is set to 20. The *minbucket* is the fewest observations on the terminal node. The *maxdepth* is the size of the tree's growth, which can reach 30. The *complexity parameter* is the work measure necessary for the execution of an algorithm (Barbosa, Toscani, and Ribeiro, 2002). Lauretto (2010) explains that when excessive growth occurs, the tree may include attributes that contribute little to predictive power, or may even be irrelevant for predicting results for new data, a phenomenon known as *overfitting*. To mitigate this phenomenon, the algorithm can be attributed, according to the author, with a *complexity parameter* to find the most suitable tree for the problem at hand. The researcher emphasizes that the ideal tree is one that does not grow excessively to include nodes that contribute minimally from an informational standpoint, nor is it cut to the extent that nodes with predictive power are discarded. Lastly, *xval*, which is the number of cross-validations, according to the *rpart* Manual. Gerón (2021) defines *cross-validation* as the division of the training sample into a smaller one, which serves as validation, in a process called *k-fold* method.

After these considerations, the model proposed by Bernardo (2022) was used for the elaboration of step 1. In this package, the author used the *mlr package*, which automates the calculations of the hyperparameters and shows the results regarding each variation in them. It was through this algorithm that the script was developed for the predictive model of opinion on accounts. That said, the procedure unfolded as follows:

1. In step 1, the packages established by Bernardo (2022) were loaded, namely: *dplyr, rpart, rpart.plot, metrics, mlr, ggploty2, plotly*.

2. Subsequently, the spreadsheet was loaded with the data of the IEGM notes and the respective accounts opinions. In this step, the characters were converted to factors.

3. The size of the samples was defined for training (75%) and testing (25%).

4. The *rpart* command was run to generate the tree.

5. The first hyperparameter for testing was established, which was the *maxdepth* (1:30) – *makeParamSet* command.

6. The cross-validation parameter was defined in 3 iterations (*makeResampleDesc*).

7. Through the *tuneParams* and *generateHyperParsEffectData* commands, the graph of the accuracy variation as a function of *maxdepth* was generated, as shown in Figures 4 and 5.

8. Lastly, with the *setHyperPars* command, it was possible to store the best tuning results, which could be displayed from the creation of the objects *best_parameters*, *best_model*, and *d.tree.mlr.test* and, later, by executing the *predict* and *accuracy* commands, to obtain the predictions and the corresponding accuracy of the model.

9. The tuning was repeated, this time incorporating four hyperparameters (*Maxdepth*, *Minbucket*, *Minsplit and CP*), which are represented by Table 2 and Figures 6 and 7. For the purposes of analysis, the intervals were arbitrarily selected. As for *maxdepth*, the interval aimed to mitigate the overgrowth of the tree, generating *overfitting* and, at the same time, avoid the disposal of variables (IEGM dimensions) with predictive power, as Lauretto (2010) ensures.

10. In step 2, the model – the classification tree itself – was generated based on the hyperparameters obtained in step 1. For its execution, the following packages were loaded: *tidyverse*, *gtools*, *rmisc*, *scales*, *pROC*. The separate execution of these packages was intended to prevent overlapping commands that are common in two or more packages. In addition, the elaboration method adopted for the development of the tree was class-based, as detailed in the manual of the *rpart* package. Other possible classes would be “anova” or “Poisson”, but applicable to other machine *learning* analysis.

11. After the package loading was completed, and with reference to the hyperparameters calculated in the previous step, the second phase of the script was executed, which involved the development of the tree (Figure 8) based on the tuned hyperparameters.

12. The elaborated tree presented *cutoff* points (Figure 9), that is, percentages on which the R used as a parameter to calculate the probability that an event would be categorized as "unfavorable" (event) and "favorable" (non-event). From these cutoffs, the different values of accuracy, sensitivity, and specificity for the training and test samples were calculated. See Figure 10.

13. In stage 3, the observations (IEGM of the municipalities) were classified into "favorable" and "unfavorable" from the predict command. In Figures 11 and 12 it is possible to visualize two of the different possible confusion matrices in this model.

14. After this step, a table was generated with the predicted and observed values and the respective percentages of probability of "favorable" or "unfavorable" result in the judgment of accounts, according to the developed model. Figure 13 shows the result of this procedure.

15. Lastly, still in stage 3, the ROC curve was generated, both in the training and test sets. See figures 14 and 15.

16. From the developed model, the AUC areas of the ROC curve were calculated.

As can be seen, all the steps proposed by Géron (2021) were followed. The research problem, as well as the independent and dependent variables, were defined. The methodology was defined, and the software for solution development was selected. With the choice of RStudio, the packages used were established, and the scripts were developed based on the manuals of the packages and references available on the internet. Given that no errors were detected in the development of the model, the next chapter will focus on evaluating the preliminary results. However, before proceeding, the limitations of the research will be described.

2.6 Limitations of the study

One of the limitations of the study was the number of fiscal years with available account opinion data. For the CA under study, only data from 2017 were located on the World Wide Web, which prevented the analysis of other fiscal years for the same subject. This difficulty can be observed in relation to other CTs in Brazil, given the diversity of forms of work in each CA of the Federation. Lino and Aquino (2018) analyzed the differences between CAs. According to the authors, part of these bodies does not inspect their jurisdictions *in loco* annually, with rotation between the municipalities of each state annually. Araújo, Bezerra Filho and Motoki (2019) – in the analysis of the IEGM of the period between 2015 and 2017 – observed that some states responded to the IEGM questionnaire with data that were partly different from the reality found locally. Therefore, it is necessary to adopt a rigorous data validation process. It is worth noting that some states initially adopted the IEGM without requiring municipalities under their jurisdiction to adhere to it. This is the case of the state of Amazonas, which began to require all of them to participate only in 2019. Another limitation refers to the *mlr* package. During the execution of the model, the message was issued that the package was being deactivated. Future developments will only be possible with the *mlr3* package, which may entail risks of bugs between versions, as described in an alert message displayed by the application.

3. RESULTS AND DISCUSSION

First, in step 01, the algorithm proposed by Bernardo (2022) used the *mlr* package to obtain hyperparameters for generating the optimal tree. As shown in Figure 4, by varying a single hyperparameter (*maxdepth*), the highest mean accuracy in the execution of the *cross-validation* procedure was 74.33%. The *maxdepth* value that made it possible to obtain this optimal model was determined to be two.

Figure 4 – Tuning Classification Tree Hyperparameters - *maxdepth* x *acc.test.mean*

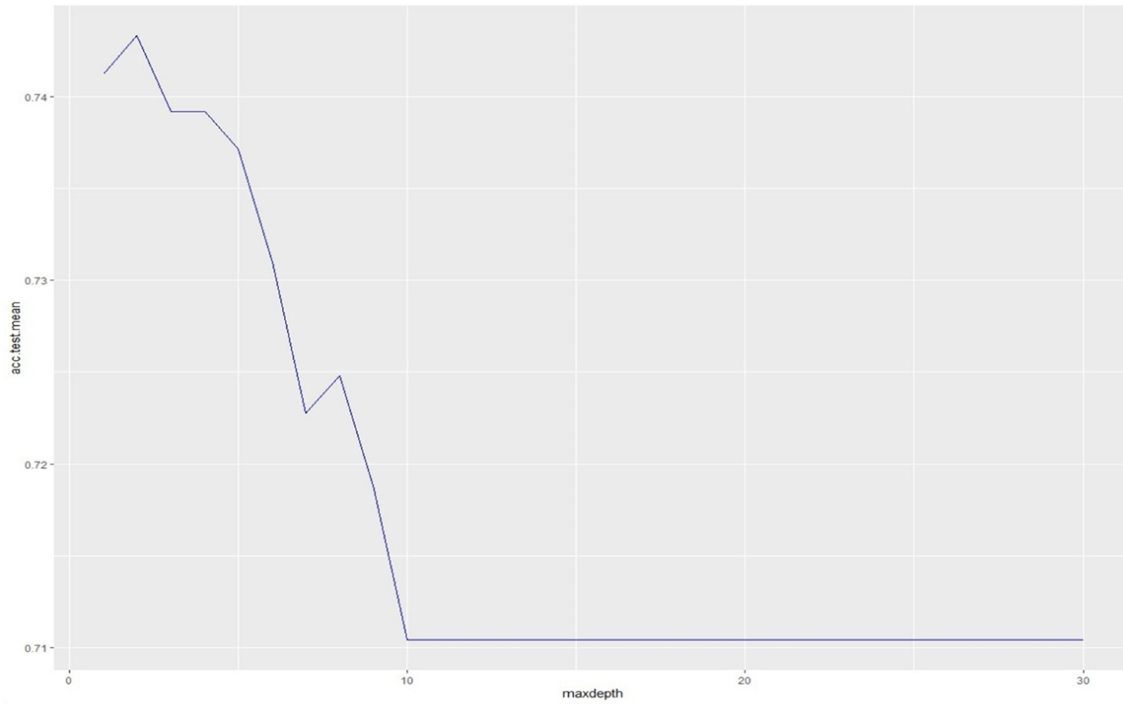
```
Console   Terminal ×   Background Jobs ×
R 4.2.2 · ~/04-TCC/02-Algoritmos-TCC/
[Tune-y] 24: acc.test.mean=0.7104067; time: 0.0 min
[Tune-x] 25: maxdepth=25
[Tune-y] 25: acc.test.mean=0.7104067; time: 0.0 min
[Tune-x] 26: maxdepth=26
[Tune-y] 26: acc.test.mean=0.7104067; time: 0.0 min
[Tune-x] 27: maxdepth=27
[Tune-y] 27: acc.test.mean=0.7104067; time: 0.0 min
[Tune-x] 28: maxdepth=28
[Tune-y] 28: acc.test.mean=0.7104067; time: 0.0 min
[Tune-x] 29: maxdepth=29
[Tune-y] 29: acc.test.mean=0.7104067; time: 0.0 min
[Tune-x] 30: maxdepth=30
[Tune-y] 30: acc.test.mean=0.7104067; time: 0.0 min
[Tune] Result: maxdepth=2 : acc.test.mean=0.7433033
```

Source: the authors.

Note: The expression "[Tune-X] [number]" indicates to the user which and how many hyperparameter combinations have been tested so far. The expression "[Tune-Y] [number]" indicates the result of the iteration of the momentum. "[Tune-X] Result" shows the optimal parameters and the average accuracy

Figure 5 shows the variation in mean accuracy of the model validation set as a function of *maxdepth*. It is possible to notice that the average accuracy curve declined as *maxdepth* increased, eventually stabilizing with a *maxdepth* of ten. This phenomenon occurs because as the tree expands, its generalization power diminishes, resulting in lower predictive accuracy for data beyond those in the training sample.

Figure 5 – *Graph Maxdepth* x *acc.test.mean* (validation set in *cross-validation* procedure)



Source: the authors.

After these procedures, the maximum mean accuracy obtained with this model, now in the test sample, was calculated at 75.15%. Then, by increasing the number of hyperparameters to four, new optimal parameters were calculated, and a new accuracy result was obtained. According to Table 2, *Maxdepth* equals 5, *Minbucket*, 10; *Minsplit*, 5; and *CP* was 0.01 representing the optimal hyperparameters to be used in the model.

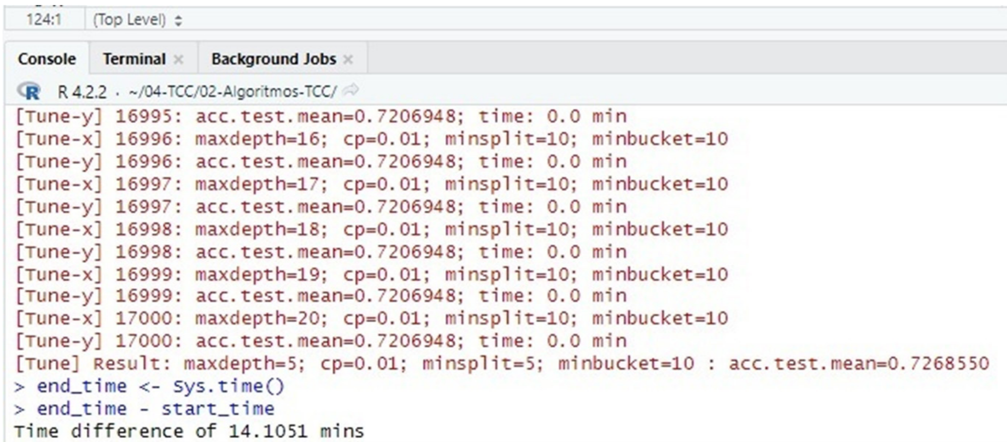
Table 2 – Tuned values range and the optimal value obtained by the algorithm

Hyperparameter	Tuning values range	Optimal value obtained
Maxdepth	04 a 20	5
Minbucket	01 a 10	10
Minsplit	01 a 10	5
CP	0.001 a 0.01	0.01

Source: the authors.

In performing this procedure (Figure 6), as already shown, R processed 17,000 combinations and reached the highest average accuracy of 72.68% in the different models resulting from the combinations (*cross-validation*) of hyperparameters. The processing time of this command in R was approximately fourteen minutes.

Figure 6 – Tuning of classification tree hyperparameters - *maxdepth* / cp / *minsplit* / *minbucket* x *acc.test.mean* (validation set)

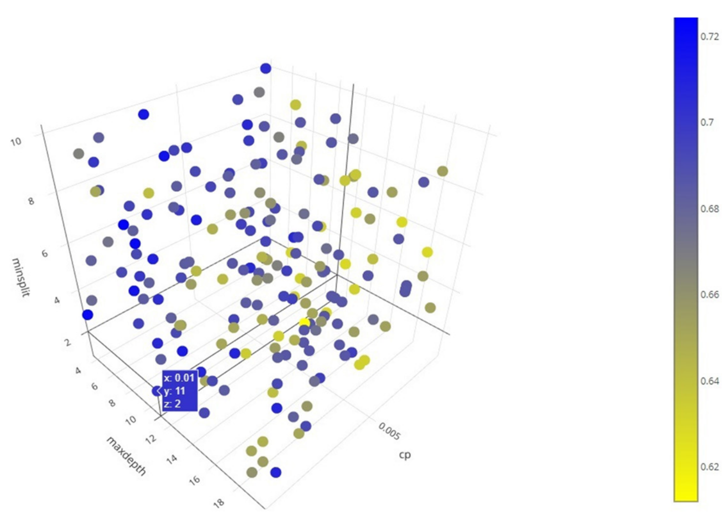


```
124:1   (Top Level) ÷
Console   Terminal ×   Background Jobs ×
R 4.2.2 · ~/04-TCC/02-Algoritmos-TCC/
[Tune-y] 16995: acc.test.mean=0.7206948; time: 0.0 min
[Tune-x] 16996: maxdepth=16; cp=0.01; minsplit=10; minbucket=10
[Tune-y] 16996: acc.test.mean=0.7206948; time: 0.0 min
[Tune-x] 16997: maxdepth=17; cp=0.01; minsplit=10; minbucket=10
[Tune-y] 16997: acc.test.mean=0.7206948; time: 0.0 min
[Tune-x] 16998: maxdepth=18; cp=0.01; minsplit=10; minbucket=10
[Tune-y] 16998: acc.test.mean=0.7206948; time: 0.0 min
[Tune-x] 16999: maxdepth=19; cp=0.01; minsplit=10; minbucket=10
[Tune-y] 16999: acc.test.mean=0.7206948; time: 0.0 min
[Tune-x] 17000: maxdepth=20; cp=0.01; minsplit=10; minbucket=10
[Tune-y] 17000: acc.test.mean=0.7206948; time: 0.0 min
[Tune] Result: maxdepth=5; cp=0.01; minsplit=5; minbucket=10 : acc.test.mean=0.7268550
> end_time <- Sys.time()
> end_time - start_time
Time difference of 14.1051 mins
```

Source: the authors.

The highest mean accuracy of the validation set of the *cross-validation* procedure, as shown in Figure 6, was 72.68%, and the one that returned the best result of the model, according to the algorithm, was 73.24%, in the test sample. Graphically, the tuning of the four hyperparameters resulted in the outcome shown in Figure 7. Of the 17,000 combinations obtained, only 200 were used to represent the three-dimensional graph. R allows the coordinates (*CP*, *maxdepth*, *minsplit*) to be displayed when positioning the cursor at each point. The color of each point represents the accuracy to be obtained in the model, according to the vertical bar on the right side. In Figure 7, the highlighted point (0.01, 11, 2) is merely illustrative. In the same figure, as noted, it is possible to plot up to three tuned hyperparameters on the Cartesian axes, although the *makeParamSet* function allows tuning a larger number of hyperparameters. In this case, the blue dots represent the combinations that generate the best accuracy, according to the caption to the right of the graph. Yellow dots, on the other hand, represent lower accuracy.

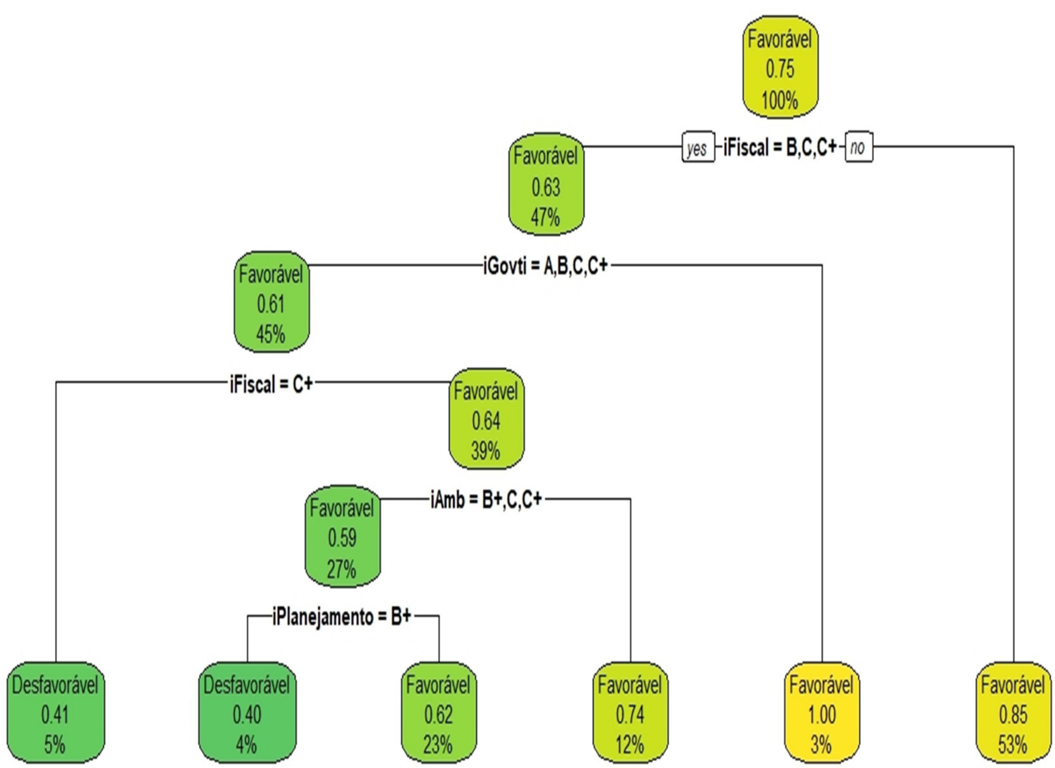
Figure 7 – *Maxdepth/minsplit/cp* x *acc.test.mean* combinations (validation set)



Source: the authors.

The advantage of the script executed in step 1 was to establish and tune the hyperparameters (Figure 3) provisional to the development of the model in the *rpart* package. However, a warning should be made that the processing time in the *tuneparams* command was longer as more hyperparameters were established in the *makeParamSet function*. With only one hyperparameter, there were only 30 iterations, a procedure that was completed in a few seconds. On the other hand, with four hyperparameters (*maxdepth*, *minbucket*, *minsplit* and *cp*), the processing time of the seventeen thousand iterations reached the time of 14 minutes until its effective completion. Another disadvantage of this stage is that the accuracy was calculated by R without varying the *cutoff* percentages. Checking the table of accuracy and respective *cutoff* points (Figure 10), it was found that R adopted 41% as *cutoff* for the prediction of results of favorable or unfavorable opinions. Thus, executing the commands in step 2 enabled the variation of accuracy values to be obtained as a function of different cutoff points. This analysis was supplemented with sensitivity and specificity values to determine the most appropriate *cutoff* for the model. Then, from these data obtained in step 1, the second step of the algorithm was performed, which was generating the classification tree, as presented in Figure 8.

Figure 8 – Classification Tree



Source: the authors.

At the top of the tree is the root node. Below it, there are two other nodes. One of them is a terminal node and the other one splits into two other internal nodes. Within each node the predominant result variable is identified (favorable or unfavorable). The first percentage number that appears on each node, just below the node's predominant result variable, indicates the total percentage of the sample that has the predominant result variable of the model (favorable). The number below indicates the percentage of the condition that is expressed in the horizontal line of the corresponding branch. In the case of this research, for example, it was observed in the root node that, out of the entire sample, 75% of the cities received a favorable opinion. Below, 47% of the municipalities received the grade "B", "C", or "C +" (see box with the text "yes") in IFISCAL. 63% of the 47% who owned these three IFISCAL notes received a favorable opinion. As for the municipalities that did not receive the grades "B", "C", or "C +", that is, received grades "A" or "B +" in IFISCAL, they represented 53% of the sample, and 85% had their accounts approved. Turning to the left side of the first branch, the municipalities that have grades "B", "C", "C+", but do not have IGOVTI with grades "A", "B", "C" and "C+" received the grade "B+" (to the exclusion of the other notes). In this situation this branch makes the connection with the terminal node "Favorable, 1.00, 3%". Thus, this node (IFISCAL "B", "C", "C+" and "IGOVTI "B+") represents 3% of the total population of 644 municipalities. However, all the municipalities that make up this sample (terminal node) had the accounts approved (1.00 = 100%).

It is worth noting that this configuration of the tree must be understood under the concept of entropy and information gain. Géron (2021) and Lauretto (2010) assert that entropy is the degree of uncertainty that exists in each node. Thus, if in a given node there are samples of municipalities represented by the two distinct decisions – favorable and unfavorable – then there is a certain degree of entropy in that node. Conversely, nodes that have samples represented only by municipalities with exclusively favorable or unfavorable opinions do not have entropy. As can be seen in Figure 8, in the first node on the left, there were 63% of favorable decisions, indicating a degree of entropy. On the other hand, in the case of the terminal node that comes just below, with 100% of the "favorable" decisions, it is said that entropy is equal to zero, null, or non-existent. This concept is important because it shows which variables contribute most to the model's predictive accuracy. Thus, variables that generate lower entropy have what is called "information gain", that is, they have greater predictive power. Therefore, variables that generate greater information gains are closer to the root node. In the case of this study, the IFISCAL is the variable that allows the greatest gain of information in the model. Soon after, the IGOV is the variable that generates the most information gain, and so on.

From these premises, the obtained tree can be described in full in the following form. It should be noted that the expressions "municipalities", "agencies", "administrations", "cities" and "samples", when referring to the data analyzed, are considered synonymous. That said, the following results were verified:

a. 75% of the municipalities had their accounts approved.

b. 47% of the administrations that had IFISCAL with notes "B", "C" and "C +", and of these, 63% had their accounts associated with a "favorable" opinion.

c. 53% of the agencies had IFISCAL with notes "A" or "B +", so that 85% had their accounts judged "favorable".

d. 45% of the sample has IGOVTI with notes “A”, “B”, “C”, “C+” and IFISCAL with notes “B”, “C” and “C+”.

e. 5% of the cities had a grade of “C” in IFISCAL and 41% of them had the accounts considered “favorable”. It is recalled that these municipalities are contained in the IGOVTI notes mentioned in subparagraph (d).

f. 39% of agencies received a grade of “B” and “C” in IFISCAL, and 64% of them had their accounts approved (“favorable”) by the CA surveyed.

g. The administrations mentioned in paragraph f, 27% had IAMB with grades “B+”, “C”, “C+”, and 59% of these had the accounts considered “favorable” by that Court of Accounts.

h. In relation to paragraph g, 4% achieved the grade “B +” in IPLAN, and 40% had the accounts considered “favorable”.

i. Those cities that had other scores in the IPLAN, which represented 23% of the sample, 62% had the accounts endorsed by the CA.

j. 12% of the agencies had “A” or “B” grades in the IAMB, and 74% had the accounts considered “favorable”.

k. Lastly, as previously stated, 3% of the municipalities had IFISCAL with notes “B”, “C” and “C +” and, at the same time, IGOVTI with note “B+”. This was the only case where the entropy was zero.

In addition to the results in each of the nodes and branches – considering that the terminal node presented six sheets – we obtained the values of the criteria necessary for the formulation of the model responses, that is, the possible *cutoffs*. These terminal sheets or nodes were also presented in full in R, as transcribed in Figure 9. It is possible to observe that R itself highlights the leaves or terminal nodes, which are indicated by asterisks (*).

Figure 9 – Representation of the tree in full in R.

```
1) root 644 163 Favoravel (0.2531056 0.7468944)
  2) iFiscal=B,C,C+ 305 112 Favoravel (0.3672131 0.6327869)
    4) iGovti=A,B,C,C+ 288 112 Favoravel (0.3888889 0.6111111)
      8) iFiscal=C+ 34  14 Desfavoravel (0.5882353 0.4117647) *
      9) iFiscal=B,C 254  92 Favoravel (0.3622047 0.6377953)
       18) iAmb=B+,C,C+ 174  71 Favoravel (0.4080460 0.5919540)
         36) iPlanejamento=B+ 25  10 Desfavoravel (0.6000000 0.4000000) *
         37) iPlanejamento=A,B,C,C+ 149  56 Favoravel (0.3758389 0.6241611) *
       19) iAmb=A,B 80  21 Favoravel (0.2625000 0.7375000) *
    5) iGovti=B+ 17   0 Favoravel (0.0000000 1.0000000) *
  3) iFiscal=A,B+ 339  51 Favoravel (0.1504425 0.8495575) *
```

Source: the authors.

Indeed, each of the intervals was utilized in the model's prediction by comparing the actual IEGM data with those obtained in the generated tree. This process helped define the probabilities of the opinion being "favorable" or "unfavorable." Based on these intervals, the 6 *cutoff* points were defined (column "Established *cutoff* values" – Table 3).

Table 3 – Model cutoff points

Leaf	Intervals of probability	Established *cutoff* values
1	0.0000 – 0.4000	0.3900
2	0.4000 – 0.4117	0.4100
3	0.4117 – 0.6241	0.6200
4	0.6241 – 0.7375	0.7300
5	0.7375 – 0.8495	0.8400
6	0.8495 – 1.0000	0.9900

Source: the authors.

Regarding the model's performance, the observation of these *cutoff* points (Figure 10) revealed two distinct situations that warrant investigation by decision-makers. The first situation pertains to the *cutoff* point of 73%, which exhibited an accuracy of 71.97%, sensitivity (correct prediction of "unfavorable" judgments) of 61.53%, and specificity (correct prediction of "favorable" judgments) of 75.42%. On the other hand, for the *cutoff* of 84%, the accuracy was lower (63.69%), but the sensitivity achieved was higher (76.92%), while the specificity was 59.32%. Therefore, it's evident that although the *cutoff* point of "73%" has a higher accuracy, its sensitivity indicates a lower predictive power of unfavorable provisional opinion.

Figure 10 – Accuracy, Sensitivity and Specificity of the training and test samples

	Percentuais_de_Corte	acctr	sensitr	espectr	acctes	sensites	espectes
1	39%	0.7453799	0.00000000	0.00000000	0.7515924	0.00000000	0.00000000
2	41%	0.7618070	0.09677419	0.98898072	0.7324841	0.07692308	0.94915254
3	62%	0.7720739	0.23387097	0.95592287	0.7388535	0.15384615	0.93220339
4	73%	0.7022587	0.54032258	0.75757576	0.7197452	0.61538462	0.75423729
5	84%	0.6509240	0.66129032	0.64738292	0.6369427	0.76923077	0.59322034
6	99%	0.2854209	1.00000000	0.04132231	0.2611465	1.00000000	0.01694915

Source: Source: the authors.

Note: The column names with the final "tr" refer to the training sample and the final column names "tes" refer to the test sample.

Figure 11 shows the confusion matrix for the 73% *cutoff* and Figure 12 shows the confusion matrix for an 84% *cutoff.*

Figure 11 – Confusion matrix for training and test data with a *cutoff* point of 73%

```
c_treino        Desfavoravel Favoravel
  Desfavoravel            67        88
  Favoravel               57       275
> tabtes <- table(c_teste, teste$Parecer)
> tabtes

c_teste         Desfavoravel Favoravel
  Desfavoravel            24        29
  Favoravel               15        89
```

Source: the authors.

Figure 12 – Confusion matrix for training and test data with a *cutoff* point of 84%

```
c_treino        Desfavoravel Favoravel
  Desfavoravel            82       128
  Favoravel               42       235
> tabtes

c_teste         Desfavoravel Favoravel
  Desfavoravel            30        48
  Favoravel                9        70
```

Source: the authors.

Then, a table was generated to compare the predicted values, observed values, and their respective probabilities (Figure 13) for each municipality analyzed.

Figure 13 – Comparative table displaying results of "observed" versus "predicted" values along with probabilities of each occurrence according to the model

2023-JAN-07-Script-20.R* | aval_treino | aval_teste

Filter

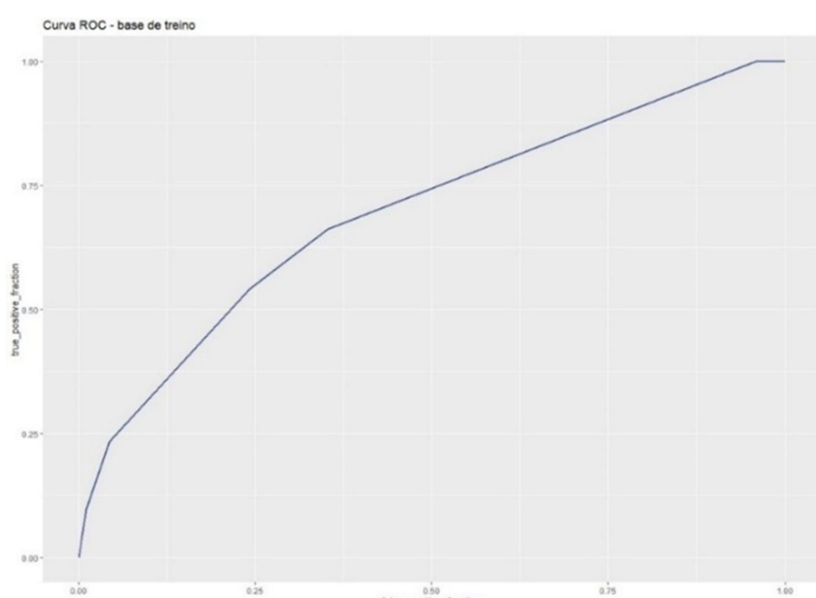
	obs	pred	Favoravel	Desfavoravel
1	Favoravel	Desfavoravel	0.6241611	0.3758389
2	Desfavoravel	Desfavoravel	0.4000000	0.6000000
3	Desfavoravel	Desfavoravel	0.6241611	0.3758389
4	Favoravel	Favoravel	0.8495575	0.1504425
5	Desfavoravel	Desfavoravel	0.6241611	0.3758389

Source: the authors.

Therefore, an adequate analysis would be to generate the confusion matrix of the two *cutoff* points (73% and 84%) – as already performed in Figures 11 and 12. For the *cutoff* point of 84%, in the training sample, of the 124 municipalities with unfavorable opinion, 42 were mistakenly classified as favorable opinion (33.87% of error). From the test sample, there were 9 accounts judged in the favorable class out of a total of 39 unfavorable accounts (23.08%). With the reduction of the *cutoff* point in the forecast from 84% to 73%, of the 124 that were trained in the model, 57 were erroneously classified as "favorable", that is, 15 units more than the test sample at the point of "84%". In the test sample, the error increased from 9 to 15 prefectures. In fact, that represents 21 more forecasting errors. Of the 644 municipalities, this represents only 3.73% of the study population. But if only the bodies that received an unfavorable opinion by the Court are taken into account – out of a total of 157 – this rate rises to 13.38%. Therefore, the choice of the *cutoff* point is a decision that must be made by those responsible for planning the audits. The use of other sources of information can contribute to the increase of the predictive power of the model or to the mitigation of errors and, thus, enable better decision-making from the institutional point of view.

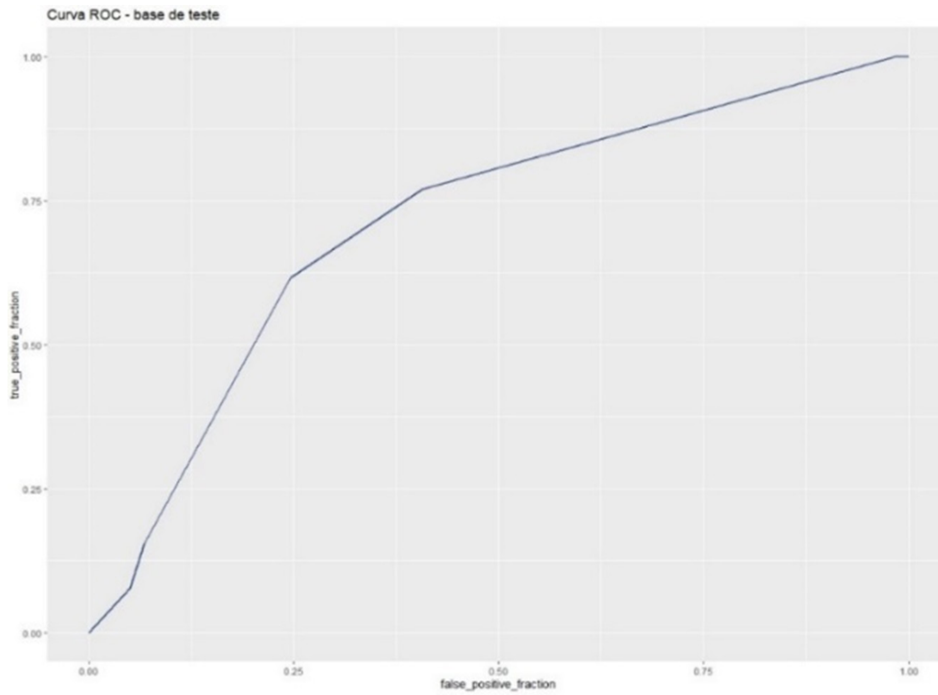
Regarding the ROC curve, two results were obtained. In the training set, the area obtained was 0.6937, and in the test set, it was 0.7106. This value does not depend on the *cutoff* point of the model already established, as seen in the calculations. In this research, different ROC curve formulations were not examined. An important aspect of the ROC curve, as shown in Figures 11, 12, 14 and 15, is that the *cutoff* points "73%" and "84%" represent the inflection points of the curves plotted in the Cartesian plane. The ordered pairs (1-specificity, sensitivity) of the training and test curves reflect what is contained in Figure 10 and in the confusion matrices of Figures 11 and 12. The pairs (040, 0.76), of the *cutoff* point of "84%", show a higher point on the axis of the ordinates than the point (0.24, 0.61). In this way, it is possible to graphically visualize the greater sensitivity or predictive power of the event, that is, greater performance in the detection of rejected accounts. According to the studies of Cristiano (2017), when the diagnostic test uses a point of the Cartesian plane in the middle of the space of the ROC curve, the true positives are slightly higher than the false positives, resulting in what is called the "strict" criterion of choice of the diagnostic test. In addition, it was observed in this research that the ROC curve in the test set (0.7106) has a slightly larger area than the area of the training set (0.6937).

Figure 14 – ROC curve – Training set



Source: the authors.

Figure 15 – ROC curve – Test set



Source: the authors.

In the development of research on sensitivity and specificity in the ROC curve, Cristiano (2017) focused on two aspects of evaluating the proposed model. One of them was the search for the best ROC curve and the other represented the search for the best *cutoff* point. The latter aspect was also adopted in this study. To this end, the constitutional function of a CA should be remembered, which is to assess the management of executive branch heads. Thus, a mistaken assessment that the municipality would receive a favorable opinion could induce the Court to stop intensifying enforcement efforts on a given municipality under mismanagement. Thus, although there are other techniques for planning the monitoring of the agencies, the premise of the model should be the identification of the largest number of cities with problems in the execution of public policies. Therefore, the approach of determining the best *cutoff* point was the solution adopted.

As for previous researches, the model developed in this research corroborates the analysis of Macieira (2016), since the revenues have a negative correlation with the unfavorable opinion. This means that the higher the revenue, the less likely it is that the municipal finances will be rejected. Thus, as in the IEGM the revenues are captured by IFISCAL – a variable that showed the highest information gain, being positioned just after the root node – it can be observed that there is alignment with the conclusions of that author. In the research comparing the IEGM with opinions obtained by municipalities in Rio Grande do Norte, while no correlation was found between the IEGM and unfavorable opinions, and a positive correlation with favorable account opinions couldn't be concluded, Rodrigues (2022) demonstrated that cities with a fiscal surplus during the evaluated period exhibited better results in both the IEDUC and overall IEGM, compared to municipalities with deficits. Another important conclusion of that study is that the volume of revenues was not relevant to the result of the final IEGM. The author highlighted that the research's limiting factor was the fact that 156 out of the 158 municipalities had their accounts rejected in the examined fiscal year. In this context, the prominence of IFISCAL as the variable with the highest predictive power in the model aligns with the findings of Macieira (2016) and Rodrigues (2022) regarding the role of surplus in IEGM results. While IEDUC and ISAUDE dimensions, related to priority public policies and specifically regulated by the Federal Constitution, were initially not included in the optimal classification tree obtained in this study, when the *maxdepth* was alternated to 6 and 7 respectively, these dimensions began to feature in the new trees obtained. It is worth noting that the aforementioned studies did not indicate a correlation between each of the seven dimensions of the IEGM (in isolation) and account opinions, which explains why these dimensions were not positioned closer to the root node due to their lower information gains. Therefore, increasing *maxdepth*, although not serving predictive purposes, can aid in understanding the behavior of potential predictor variables and contribute to improving the dimensions of the IEGM.

4. FINAL CONSIDERATIONS

The results of the study showed that it was possible to develop a model capable of predicting the result of the judgment of the prefectures' finances based on the IEGM scores of the respective municipalities. This was accomplished through the development of a classification tree methodology. Among the dimensions of the IEGM, the one that presented the greatest information gain or predictive power was that of IFISCAL. Another aspect found in this study is that achieving high accuracy with the model alone may not align with the objectives of the organization. In this case, one of the main missions of the CA in question is to find municipalities with management issues. Therefore, the sensitivity variable, when identifying municipalities with a likelihood of receiving unfavorable judgments accurately, becomes a crucial consideration in determining the *cutoff* point. In addition, the choice of hyperparameters, notably the *maxdepth* (depth of the tree) is relevant, in order to enable the model's generalization capacity. The tuning procedure is one of the appropriate means to carry out this task. However, it should be noted that, although the IEGM represents an important component of the current format of judgment of accounts by Brazilian courts of accounts, it is not the only aspect taken into consideration for this task. There are other parameters – objective and subjective – to evaluate the management of municipalities and their mayors' work in general. Therefore, it is proposed that new studies could improve the subject by developing additional analyses or models. The use of the *cross-validation* method with pruning, through the *printcp* command of the *rpart* package, or the use of regression models are examples of these other aspects. It is also recommended to replicate this research in other states and fiscal years to evaluate whether the optimal trees obtained differ from the current one and to identify the reasons for these potential variances. Given that the IEGM is assessed months before the municipalities' accounts are judged and sometimes extends beyond one year, considering appeal periods until the final judgment, adopting this predictive model offers the advantage of aiding in the planning of subsequent years' inspection work.

REFERENCES

ALCOFORADO, Luciane Ferreira. **Utilizando a linguagem R**: conceitos, manipulação, visualização, modelagem e elaboração de relatórios. Rio de Janeiro: Atlas Books, 2021.

AMORIM, André de Carvalho. A utilização do Índice de Efetividade da Gestão Municipal (IEG-M) como ferramenta de avaliação das políticas públicas. **Conteúdo Jurídico**, jul. 2017. Artigos. Available at: https://conteudojuridico.com.br/consulta/Artigos/50556/a-utilizacao-do-indice-de-efetividade-da-gestao-municipal-IEG-M-como-ferramenta-de-avaliacao-de-politicas-publicas. Consulted on: 23 jan. 2023.

AMAZONAS. Tribunal de Contas do Estado do Amazonas. **Resolução nº 03, de 14 de maio de 2019**. Inclui os lançamentos para o Índice de Eficiência da Gestão Municipal na formulação das prestações de contas anuais municipais; regula o exercício do contraditório por meio eletrônico; amplia a coleta de dados e documentos de defesa por meio digital; altera as resoluções nº 33/2012, 13/2015, 27/2013, 10/2016; autoriza modificações em manuais de controle externo e dá outras providencias. Manaus: TCE/AM, 2019. Available at: https://www2.tce.am.gov.br/?page_id=20964#68-130-2019. Consulted on: 23 jan. 2023.

ARAÚJO, Leonardo Coêlho de; BEZERRA FILHO, João Eudes; MOTOKI, Fabio Yoshio Suguri. CPT407 - Gestões Municipais no Brasil. Um estudo a partir do Índice da Efetividade da Gestão Municipal (IEG-M) publicado pelo Instituto Rui Barbosa. *In*: **Congresso ANPCont**, 13., São Paulo, 2019. Available at: https://anpcont.org.br/pdf/2019_CPT407.pdf. Consulted on: 6 nov. 2022.

BALLINGS, Michel; POEL, Dirk Van den. Comprehensive R Archive Network. **Package 'auc'**. Documentation. Reference Manual. Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/AUC/index.html. Consulted on: 23 jan. 2023.

BERNARDO, Ivo. Decision tree hyperparameter tuning in R using mlr: learn how to use mlr to perform hyperparameter gri search in R. **Towards Datascience**, 8 jun. 2022. Available at: https://towardsdatascience.com/decision-tree-hyperparameter-tuning-in-r-using-mlr-3248bfd2d88c. Consulted on: 23 jan. 2023.

BISCHL Bernd; LANG, Michel; KOTTHOFF, Lars; SCHRATZ, Patrick; SCHIFFNER, Julia; RICHTER, Jakob; JONES, Zachary; CASALICCHIO, Giuseppe; GALLO, Mason; BOSSEK, Jakob; STUDERUS, Erich; JUDT, Leonard; KUEHN, Tobias; KERSCHKE, Pascal; FENDT, Florian; PROBST, Philipp; SUN, Xudong; THOMAS, Janek; VIEIRA, Bruno; BEGGEL, Laura; AU, Quay; BINDER, Martin; PFISTERER, Florian; COORS, Stefan; BRONDER, Steve; ENGELHARDT, Alexander; MOLNAR, Christoph. **Package 'mlr**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/mlr/index.html. Consulted on: 23 jan. 2023.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília: Presidência da República, [2024]. Available at: https://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Consulted on: 6 nov. 2022.

BRASIL. **Lei Complementar nº 101, de 04 de maio de 2000**. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. Brasília: Presidência da República, [2024]. Available at: https://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm. Consulted on: 6 nov. 2022.

BRASIL. **Lei nº 4.320, de 17 de março de 1964**. Estatui Normas Gerais de Direito Financeiro para os balanços da União, Estados, Municípios e Distrito Federal. Brasília: Presidência da República, [1982]. Available at: http://www.planalto.gov.br/ccivil_03/leis/L4320compilado.htm. Consulted on: 6 nov. 2022

CRISTIANO, Mariana Vitória de Menezes Bordalo. **Sensibilidade e especificidade na curva ROC**: um caso de estudo. 2017. Dissertação (Mestrado em Gestão de Sistemas de Informação Médica) – Faculdade de Medicina da Universidade do Porto, Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Leiria, Leiria, Portugal, 2017. Available at: https://iconline.ipleiria.pt/bitstream/10400.8/2927/1/Disserta%C3%A7%C3%A3o_MarianaCristiano.pdf. Consulted on: 23 jan. 2023.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de Análise de Dados**: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. 1. ed. Rio de Janeiro: GEN. LTC, 2021.

FONSECA, João José Saraiva da. **Metodologia da pesquisa científica**. Fortaleza: Universidade Estadual do Ceará, 2002. Apostila.

GARCIA, Gilson Piqueras. **Tribunais de Contas e jurimetria**: contas rejeitadas e seus possíveis reflexos eleitorais. Revista do TCU, Brasília, n. 147, v. 1, p. 147-156 – 77, jan./jun. 2021. Available at: https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/1698. Consulted on: 7 nov. 2022.

GERÓN, Aurélien. **Mãos à obra**: aprendizado máquina com Scikit-Learn, Keras & TensorFlow. 2. ed. Rio de Janeiro: Atlas Books, 2021.

HAMNER, Ben; FRASCO, Michael; LEDELL, Erin. **Package ‘metrics’**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/Metrics/index.html. Consulted on: 23 jan. 2023.

HOPE, Ryan M. **Package ‘rmisc’**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/Rmisc/index.html. Consulted on: 23 jan. 2023.

KUHN, Max; WING, Jed; WESTON, Steve; WILLIAMS, Andre; KEEFER, Chris; ENGELHARDT, Allan; COOPER, Tony; MAYER, Zacary; KENKEL, Brenton; R Core Team; BENESTY, Michael; LESCARBEAU, Reynald; ZIEM, Andrew; SCRUCCA, Luca; TANG, Yuan; CANDAN, Can; HUNT, Tyler. Package ‘caret’. Documentation. **Reference Manual**. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/caret/index.html. Consulted on: 23 jan. 2023.

LAURETTO, Marcelo de Souza. **Árvores de Decisão**. Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, nov. 2010. Available at: https://edisciplinas.usp.br/pluginfile.php/4469825/mod_resource/content/1/ArvoresDecisao_normalsize.pdf. Consulted on: 23 jan. 2023.

LINO, André Feliciano; AQUINO, André Carlos Busanelli de. A diversidade dos Tribunais de Contas regionais na auditoria de governos. **Revista de Contabilidade e Finanças da USP**, São Paulo, v. 29, n. 76, p. 26-40, jan./abr. 2018.

MACIEIRA, Fabrício Carvalho. Determinantes do parecer prévio do Tribunal de Contas do Estado de São Paulo em relação às contas das prefeituras municipais. *In*: Congresso USP de Iniciação Científica em Contabilidade, 13., **Anais** [...]. São Paulo, jul. 2016. p. 1-17. Available at: https://congressousp.fipecafi.org/anais/16UspInternational/54.pdf. Consulted on: 06 nov 2022.

MAEDA, Leandro Jun; VARELA, Patrícia Siqueira. Auditoria Operacional e de Gestão: Associação entre o IEG-M/TCESP e os Pareceres de Prestação de Contas de Municípios Paulistas. *In*: Congresso USP Iniciação Científica em Contabilidade, 13., **Anais** [...]. São Paulo, jul. 2017. p. 1-17. Available at: https://congressousp.fipecafi.org/anais/17UspInternational/ArtigosDownload/296.pdf. Consulted on: 26 feb. 2023.

MILBORROW, Stephen. **Package ‘rpart.plot’**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/rpart.plot/index.html. Consulted on: 23 jan. 2023.

PAULO, Vicente; ALEXANDRINO, Marcelo. **Direito Constitucional Descomplicado**. 9. ed. Rio de Janeiro: Forense; São Paulo: Ed. Método, 2012.

ROBIN, Xavier; TURCK, Natacha; HAINARD, Alexandre; TIBERTI, Natalia; LISACEK, Frédérique; SANCHEZ, Jean-Charles; MÜLLER, Markus; SIEGERT, Stefan; DOERING, Matthias; BILLINGS, Zane. **Package ‘pROC’**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/pROC/index.html. Consulted on: 23 jan. 2023.

RODRIGUES, Maria Clara Ismael. Relação entre o IEG-M e o parecer prévio das prestações de contas dos municípios pelo TCE/RN. **Semana Acadêmica Revista Científica**, Fortaleza, ed. 226, v. 10, p. 1-28, 2022. DOI: http://dx.doi.org/10.35265/2236-6717-226-12237. Available at: https://semanaacademica.org.br/system/files/artigos/23_maria_clara_relacao_entre_o_iegm_e_o_parecer_previo_maria_clara_0_2.pdf. Consulted on: 2 nov. 2022.

SACHS, Michael C.; CORTY, Robert W.; CROUCH, Luis. **Package 'plotROC'**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/plotROC/index.html. Consulted on: 23 jan. 2023.

SÃO PAULO. **Constituição Estadual, de 05 de outubro de 1989**. São Paulo: Assembleia Legislativa do Estado de São Paulo, 1989. Available at: https://www.al.sp.gov.br/repositorio/legislacao/constituicao/1989/compilacao-constituicao-0-05.10.1989.html. Consulted on: 6 nov. 2022.

SÃO PAULO. Tribunal de Contas do Estado de São Paulo. Um em cada 4 prefeitos teve contas desaprovadas pelo TCESP. **TCESP**, 13 fev. 2020. Notícias. Available at: https://www.tce.sp.gov.br/6524-cada-4-prefeitos-teve-contas-desaprovadas-pelo-tcesp. Consulted on: 2 nov. 2022.

SÃO PAULO. Tribunal de Contas do Estado de São Paulo. **Índice de Efetividade da Gestão Municipal**. 2017. Available at: https://www.tce.sp.gov.br/sites/default/files/noticias/anuario_-_iegm-2017_1.pdf. Consulted on: 2 nov. 2022.

SÃO PAULO. TCE-SP. **Índice de Efetividade da Gestão Municipal**. **Manual 2021. Dados do exercício 2020**. Available at: https://www4.tce.sp.gov.br/transparencia/resultado_iegm. Consulted on: 6 nov. 2022.

SÃO PAULO. TCE-SP. **Índice de Efetividade da Gestão Municipal**. 2022. Available at: https://painel.tce.sp.gov.br/pentaho/api/repos/%3Apublic%3Aieg_m%3AIEG-M.wcdf/generatedContent?userid=anony&password=zero. Consulted on: 23 jan. 2023.

SIEVERT, Carson; PARMER, Chris; HOCKING, Toby; CHAMBERLAIN, Scott; RAM, Karthik; CORVELLEC, Marianne; DESPOUY, Pedro; BRÜGGEMANN, Salim; PLOTLY TECHNOLOGIES INC. **Package 'plotly'**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/plotly/index.html. Consulted on: 23 jan. 2023.

SILVA JÚNIOR, Bernardo Alves da. **O exercício do poder cautelar pelos tribunais de contas**. Revista do TCU, Brasília, n. 113, v. 1, p. 33-40, set./dez. 2008. Available at: https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/363. Consulted on: 6 nov. 2022.

SILVEIRA, Denise Tolfo; CÓRDOVA, Fernanda Peixoto. Unidade 2 – A Pesquisa Científica. *In*: GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (org.). **Métodos de Pesquisa**. Porto Alegre: Editora da UFRGS, 2009. Available at: http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf. Consulted on: 6 nov. 2022.

STANKEVIX, Gabriel. Árvore de Decisão em R: classificação e aplicação na Base de Baseball (MLB). **Medium**, out. 2019. Available at: https://medium.com/@gabriel.stankevix/arvore-de-decis%C3%A3o-em-r-85a449b296b2. Consulted on: 23 jan. 2023.

THERNEAU, Terry; ATKINSON, Beth; RIPLEY, Brian. **Package 'rpart'**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/rpart/index.html. Consulted on: 23 jan. 2023.

TRIVIÑOS, Augusto Nibaldo Silva. **Introdução à pesquisa em ciências sociais**: a pesquisa qualitativa em educação. São Paulo: Atlas, 1987.

VOLPATO, Gilson Luiz. O método lógico para redação científica. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, Rio de Janeiro, v. 9, n. 1, p. 1-14, jan./mar. 2015. DOI: https://doi.org/10.29397/reciis.v9i1.932. Available at: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/932. Consulted on: 08 may 2024.

WARNES, Gregory R.; BOLKER, Bem; LUMLEY, Thomas; MAGNUSSON, Arni; VENABLES, Bill; RYODAN, Genei; MOELLER, Steffen; WILSON, Ian; DAVIS, Mark; Jain, Nitin; CHAMBERLAIN, Scott. **Package 'gtools'**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/gtools/index.html. Consulted on: 23 jan. 2023.

WICKHAM, Hadley; FRANÇOIS, Romain; HENRY, Lionel; MÜLLER, Kirill; VAUGHAN, Davis; POSIT SOFTWARE, PBC. **Package 'dplyr'**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/dplyr/index.html. Consulted on: 23 jan. 2023.

WICKHAM, Hadley; CHANG, Wiston; HENRY, Lionel; PEDERSEN, Thomas Lin; TAKAHASHI, Kohske; WILKE, Claus; WOO, Kara; YUTANI, Hiroaki; DUNNINGTON, Dewey; BRAND, Teun van den; POSIT, PBC. **Package 'ggplot2'.** The Comprehensive R Archive Network. 2022. Documentation. Reference Manual. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/ggplot2/index.html. Consulted on: 23 jan. 2023.

WICKHAM, Hadley; PEDERSEN, Thomas Lin. **Package 'scales'**. Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/scales/index.html. Consulted on: 23 jan. 2023.

WICKHAM, Hadley; RStudio. **Package 'tidyverse'.** Documentation. Reference Manual. The Comprehensive R Archive Network. 2022. Available at: https://The Comprehensive R Archive Network.r-project.org/web/packages/tidyverse/index.html. Consulted on: 23 jan. 2023.

WILLEMAN, Marianna Montebello. O **desenho institucional dos Tribunais de Contas e sua vocação para a tutela da *accountabilty* democrática**: perspectivas em prol do direito e à boa administração pública no Brasil. 2016. Tese (Doutorado em Direito) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2016. Available at: https://www2.dbd.puc-rio.br/pergamum/tesesabertas/1221597_2016_completo.pdf. Consulted on: 6 nov. 2022.