



# CLASSIFICAÇÃO DE PARECERES DE CONTAS EM FUNÇÃO DO ÍNDICE DE EFETIVIDADE DA GESTÃO MUNICIPAL

*Classification of provisional opinions through the Municipal Management Effectiveness Index*

## Otoniel Arruda Costa

Bacharel em Ciências Navais/Administração pela Escola Naval, especialista em Normas Internacionais de Contabilidade pela FEARP-USP (MBA) e especialista em Data Science e Analytics (MBA). Especialista em Sistemas de Intendência pelo Centro de Instrução Almirante Wandenkolk (Marinha do Brasil) e servidor do Tribunal de Contas do Estado de São Paulo.

**Orcid:** <https://orcid.org/0000-0001-5346-7374>

**Lattes:** <https://lattes.cnpq.br/2101791643993928>

## Henrique Raymundo Gióia

Mestre e doutorando em Economia Aplicada pela ESALQ-USP.

**Orcid:** <https://orcid.org/0000-0002-8639-1481>

**Lattes:** <http://lattes.cnpq.br/5957983550791571>

**E-mail:** [henriquerg@usp.br](mailto:henriquerg@usp.br)

## 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 efetuassem 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.

## 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.

## 1. INTRODUÇÃO

O Tribunal de Contas da União (TCU), os tribunais de contas estaduais (TCE), do Distrito Federal (TCDF) e dos Municípios (TCM) possuem, de acordo com a Constituição Federal do Brasil, a missão de auxiliar o Poder Legislativo na fiscalização contábil, financeira, orçamentária, operacional e patrimonial da União, Estados, Distrito Federal e Municípios. Para o atendimento dessa obrigação, os tribunais de contas (TCs) são responsáveis por emitir pareceres (chamados de Parecer Prévio ou Parecer de Contas) sobre as prestações de contas da gestão dos chefes do Poder Executivo para posterior apreciação pelas respectivas Casas legislativas. Para emitir esses pareceres, são desenvolvidas atividades de análise de tais prestações de contas pelos corpos técnicos dessas cortes de contas. Para tanto, alguns dos principais instrumentos legais que os norteiam são a Lei Federal nº 4.320/64 e a Lei Federal nº 101/2000 (Lei de Responsabilidade Fiscal). Ocorre que parte desses trabalhos técnicos são baseados predominantemente em análises de atos e fatos administrativos e contábeis já consumados pelos órgãos públicos (controle *a posteriori*). Nesse contexto, Silva Junior (2008) e Willeman



(2016) destacam a baixa efetividade que o controle posterior possui na preservação do patrimônio público. Willeman (2016) acresce que há necessidade de revisão do modelo de controle externo brasileiro, de modo a conferir maior efetividade aos processos de prestação de contas e aos de auditorias operacionais. Nesse contexto de melhorias, encontra-se o Índice de Efetividade da Gestão Municipal (IEG-M). Criado em 2014 pelo Tribunal de Contas do Estado de São Paulo (TCE-SP, 2020), o IEG-M é uma técnica de fiscalização utilizada pelos TCs que visa avaliar a gestão dos órgãos em suas esferas de atuação sob a perspectiva da qualidade do gasto governamental para os cidadãos. Araújo, Bezerra Filho e Motoki (2019) assinalam que o IEG-M possui adesão dos 27 TCs dos Estados e do Distrito Federal. Isto posto, o IEG-M é representado por uma série de cálculos matemáticos em sete dimensões temáticas: educação (i-Educ), saúde (i-Saúde), gestão fiscal (i-Fiscal), planejamento (i-Plan), meio ambiente (i-Amb), defesa civil (i-Cidade) e governança em tecnologia da informação (i-Gov-TI). Dessa forma, de acordo com o manual do IEG-M do exercício de 2020 do TC objeto de estudo, em cada dimensão temática os municípios podem ser avaliados em cinco faixas de resultados, a saber: A (altamente efetivo); B+ (muito efetivo); B (efetivo); C+ (em fase de adequação), e C (baixo nível de adequação). Além dessas dimensões, o somatório de todos esses indicadores resulta em um índice final, chamado IEG-M (que, neste estudo, será chamado de IEG-M geral ou final para desambiguação). Para o cálculo das notas de cada dimensão, existem duas fontes de dados da gestão. A primeira fonte são respostas a questionários sobre políticas públicas que os municípios respondem em plataformas online gestadas pelos TCs. Após esse envio de dados, servidores desses órgãos fiscalizadores fazem a aferição da fidedignidade e veracidade das respostas apresentadas. A outra fonte se refere aos dados de gestão das prefeituras processados diariamente pelos TC. Em etapa posterior, com base nessas informações, são realizados os cálculos dos IEG-M dos municípios. Para cada dimensão, os pontos dos diversos quesitos (cada um possui uma pontuação diferente) são somados e divididos por mil. Já a nota final (IEG-M) é definida pelas notas das dimensões ponderadas por pesos. Uma vez obtidos os valores dos IEG-M dos municípios, as cortes de contas analisam as prestações de contas, emitem os relatórios de autoria e, por fim, emitem os pareceres prévios acerca das contas de cada um desses órgãos. De acordo com o TC examinado nesta pesquisa, todos os municípios do Estado são auditados anualmente, incluindo-se a aferição anual dos IEG-M de todos eles.

Quanto aos fatores preponderantes para o julgamento das contas dos órgãos públicos, Rodrigues (2022) analisou a relação entre o IEG-M e os pareceres prévios nos municípios do estado do Rio Grande do Norte. Em seu estudo, não foi possível identificar correlação positiva entre os resultados do IEG-M e o parecer pela aprovação das contas dos municípios daquele Estado, visto que um aspecto limitador da pesquisa foi o elevado índice de desaprovação de contas. Além disso, o estudo concluiu que o superávit financeiro, as receitas públicas e o tamanho da população não foram determinantes para o resultado do IEG-M final. Por outro lado, a pesquisa de Maeda e Varela (2017) mostrou que existe relação entre o IEG-M (nota final) e os pareceres de contas no estado de São Paulo no exercício de 2014. Em outro estudo, Macieira (2016) examinou as variáveis determinantes para a emissão dos pareceres prévios pelo TCE-SP entre 2008 e 2013. Para o autor, a variável quantitativa da população possui correlação positiva em relação ao parecer prévio desfavorável, ao passo que as receitas possuem correlação negativa. Diante dessas constatações, verifica-se a importância do IEG-M na atuação fiscalizatória dos TCs. Contudo, a literatura atualmente existente é escassa no tocante a análises preditivas do controle externo. Por outro lado, com a adoção de modelos matemáticos



e estatísticos é possível o direcionamento de esforços para aprimorar essa atuação. Uma vertente dessas inovações é a possibilidade de aplicação de técnicas de aprendizado de máquina (do inglês, *machine learning*) para que sejam avaliados os resultados da gestão. Para Géron (2021), o aprendizado de máquina é a ciência de programação em que computadores podem aprender com dados.

Assim sendo, o objetivo dessa pesquisa é o desenvolvimento de um modelo de aprendizado de máquina em que se permita prever se os municípios terão suas contas aprovadas ou rejeitadas em função das notas das sete dimensões do IEG-M e do IEG-M final, de modo a possibilitar à Corte de Contas em questão identificar as dimensões do IEG-M que possuem maior poder preditivo sobre seus pareceres prévios. Visa-se, assim, contribuir com o processo de planejamento e de execução dos trabalhos de auditoria desses tribunais de contas.

## 2. MATERIAL E MÉTODOS

No desenvolvimento do estudo, adaptou-se a estrutura proposta por Volpato (2015, p. 11). Desse modo, a primeira subseção descreve o sujeito objeto da pesquisa. Na subseção seguinte, apresenta-se a metodologia, de acordo com suas classificações e, na sequência, a subseção relativa à população-alvo. Na subseção 2.4, faz-se a exposição da forma de coleta dos dados. A subseção 2.5, por sua vez, dedica-se à análise de dados e, na última subseção, são expressas as limitações do estudo.

### 2.1 Caracterização do sujeito objeto da pesquisa

Para a realização do presente estudo, selecionou-se o Estado de São Paulo e todos os municípios a ele pertencentes. Esses entes federativos – como são denominados – possuem, segundo Paulo e Alexandrino (2013), importantes características outorgadas pela Constituição: capacidade de auto-organização e autolegislação (constituições estaduais e leis orgânicas), autogoverno (eleições para governador, prefeitos e para os membros do Poder Legislativo) e autoadministração, que significa a capacidade administrativa, legislativa e tributária de forma independente dos demais entes (Paulo; Alexandrino, 2012). Por outro lado, os estados e municípios brasileiros estão sujeitos à fiscalização contábil, financeira, orçamentária, operacional e patrimonial, conforme previsto nos artigos 70 a 75 da Constituição Federal. No caso do ente federativo deste estudo, esse poder fiscalizatório está contido nos artigos 32 a 36 da Constituição Estadual de São Paulo. Destaca-se na norma a regra que outorga a esse TC a competência para emitir o parecer prévio da prestação de contas dos municípios sob sua jurisdição. No Estado em referência, encontram-se 644 (seiscentos e quarenta e quatro) municípios e respectivas prefeituras. Todos esses órgãos são fiscalizados anualmente pelo TC, exceto a capital do Estado, que possui um TC próprio. E entre as técnicas empregadas neste trabalho de auditoria está o IEG-M, que consiste em uma metodologia de apuração dos indicadores da gestão em que se verifica, de acordo com o Rodrigues (2022), TCE-SP (2020) e Amorim (2017), a qualidade do gasto público. Para tanto, são avaliadas as políticas públicas nas seguintes dimensões, de acordo com Manual do IEG-M do exercício de 2017 do TCE-SP:



- i-Educ: mede a prestação de serviços em diversos níveis educacionais, sobretudo no ensino infantil e fundamental, avalia a promoção de programas pedagógicos, tais como aqueles relacionados com a leitura, com a escrita e com as provas de avaliação de alunos, de docentes e de outros profissionais da educação. Aspectos logísticos, tais como alimentação, uniformes e transporte escolar, também são avaliados. Verifica-se a existência de trabalhos diagnósticos sobre a infraestrutura escolar; por exemplo, se ocorrem levantamentos acerca da demanda por vagas em unidades de ensino, se existem escolas com funcionamento interrompido e se unidades necessitam de reparos. Também são avaliados: enfrentamento ao bullying, atuação dos conselhos, acessibilidade, entre outros quesitos.
- i-Saúde: afere a infraestrutura de saúde pública nos municípios por meio de verificação da quantidade de unidades básicas, de equipamentos e de profissionais. Além disso, mede se há promoção e realização de programas de saúde pública por meio de consultas e cirurgias. Avalia se são utilizados sistemas de informação como meio de interação do nível municipal com os níveis estadual e federal do Sistema Único de Saúde (SUS), entre outros aspectos.
- i-Plan: destina-se a verificar se o Plano Plurianual (PPA), a Lei de Diretrizes Orçamentárias (LDO) e a Lei Orçamentária Anual (LOA) – que são leis previstas na Constituição Federal como instrumentos de atendimento obrigatório pelo Poder Público no que se refere à promoção de políticas públicas aos cidadãos – foram elaborados, cumpridos e os resultados propostos alcançados. Além disso, estão previstos nesta dimensão questionamentos sobre a estrutura formal do setor de planejamento no tocante à existência e ao uso efetivo de sistemas de informação, também considerando-se o grau de treinamento de servidores desse setor. São adotadas verificações quanto à realização de audiências públicas para o levantamento de problemas dos municípios pelas prefeituras e quanto ao acompanhamento da execução do orçamento pelos interessados na gestão pública (prefeitos, Controles Internos, cidadãos e outros).
- i-Fiscal: avalia o nível de arrecadação de tributos e de outros tipos de receitas, de realização de despesas e de endividamento, dos precatórios e da dívida ativa, das despesas com pessoal e com encargos sociais, entre outros quesitos.
- i-Amb: visa mensurar as políticas de gestão dos resíduos sólidos dos municípios; a gestão das áreas de preservação ambiental quando eventualmente existentes nas regiões sob auditoria; a gestão de águas e esgoto; a promoção de iniciativas educacionais na área ambiental, bem como a adesão às já existentes em outras esferas de governo; a avaliação das políticas de licenciamento ambiental, entre outros aspectos.
- i-Cidade: mede a estrutura da defesa civil no município, a gestão da mobilidade urbana e das vias públicas, o nível da segurança em escolas e em postos de saúde, entre outros quesitos.
- i-GovTI: busca aferir a estruturação de tecnologia da informação na administração municipal; o nível de transparência dos dados governamentais, entre eles as despesas de pessoal, receitas e despesas, ativos e passivos, bem como as licitações públicas, entre outros aspectos.



De acordo com o TCE-SP, o IEG-M pode ser resumido na forma da Figura 1.

**Figura 1 – Dimensões do IEG-M**



Fonte: TCE-SP (2022).

Uma vez calculadas as notas em cada dimensão, a nota do IEG-M do município é obtida por meio da seguinte fórmula matemática, cujo resultado é denominado IEG-M.

$$\text{IEG-M} = [ (i\text{-Plan} + i\text{-Fiscal} + i\text{-Educ} + i\text{-Saúde}) \times 20 + i\text{-Amb} \times 10 + i\text{-Cidade} \times 5 + i\text{-GovTI} \times 5 ] / 100.$$

O tratamento e divulgação desses resultados são definidos de forma independente no âmbito de cada TC do Brasil. No caso em estudo, esses resultados são guardados pelos TCs e divulgados tão somente no momento do julgamento das contas dos municípios. Tanto as notas das dimensões quanto a nota final do IEG-M são levadas em consideração no julgamento de contas.

Como se verifica, os municípios brasileiros possuem o poder de autoadministração. Contudo, estão sujeitos ao escrutínio dos tribunais de contas, que têm, entre suas diversas competências constitucionais, a de emitir os pareceres de contas dos prefeitos às respectivas Casas legislativas para decidirem definitivamente sobre a gestão desses agentes políticos. Nesse contexto, o IEG-M foi estabelecido como uma técnica que permite avaliar diversas dimensões dessa gestão, de modo a agregar elementos de aferição da qualidade dos gastos públicos e não apenas avaliações de fatos administrativos já consumados, os quais que se mostram pouco efetivos para o processo de prestação de contas à sociedade. Por fim, os resultados desses IEG-M representam um dos componentes levados em consideração nos pareceres de contas sobre a gestão dos municípios.

## 2.2 Metodologia de pesquisa

Este trabalho está circunscrito pelo conceito de Jurimetria, o qual, segundo Garcia (2021), é a aplicação das ciências estatísticas no direito. Desse modo, trata-se de uma pesquisa com abordagem quantitativa, visto que os resultados da construção do modelo preditivo podem ser quantificados (Silveira; Córdova, 2009; Fonseca, 2002). Quanto à natureza, a pesquisa se classifica em aplicada (Silveira; Córdova, 2009), pois seus resultados podem ser utilizados na prática. Com relação aos objetivos, classifica-se em explicativa (Silveira; Córdova, 2009; Gil, 2007), pois, além de investigar os valores que permitem a formulação do modelo de

classificação mais adequado de parecer de contas em função do IEG-M, visa compreender, à luz da literatura existente, quais dimensões do IEG-M mais contribuem para o processo de previsão do resultado do julgamento das contas municipais. Quanto aos procedimentos, o estudo se caracteriza por estudo experimental, pois envolve a formulação do problema e das hipóteses que delimitam as variáveis que atuam no fenômeno estudado (Silveira; Córdova, 2009; Triviños, 1987), aplicando-se a elas diferentes testes estatísticos.

### 2.3 Definição da população-alvo

Quanto à população-alvo, selecionaram-se os 644 municípios do Estado de São Paulo. Tendo em vista que a totalidade dos dados do IEG-M desse ente federativo estava disponível, foi possível estudar o comportamento de toda a população. Os IEG-M coletados se referem ao exercício de 2017 e todos os municípios desse ente foram avaliados desde a implantação desta metodologia de avaliação de gestão. Nesse aspecto, cabe uma ressalva. Considerando-se o poder de autonomia dos entes federativos do Brasil, alguns estados brasileiros adotaram esse indicador em diferentes datas, o que pode impactar a reprodução do resultado deste estudo com o de outros TCs. Além disso, alguns estados não exigiam que todos os municípios o fizessem de forma obrigatória, o que constitui outro fator limitante na replicação do estudo. É o caso do Estado do Amazonas, que, por meio da redação do inciso I do art. 3º da Resolução nº 03/2019, deixou expresso que a obrigatoriedade de preenchimento dos questionários do IEG-M tornar-se-ia obrigatória a partir de 2019. Portanto, diante dessa restrição em relação a alguns TCs e em relação a determinados exercícios de fiscalização, algumas replicações deste experimento estão sujeitas a análise unicamente amostral, o que pode implicar limitação no poder preditivo dos modelos obtidos e, conseqüentemente, em suas conclusões.

### 2.4 Coleta dos dados

O instrumento da coleta de dados foi a pesquisa documental na rede mundial de computadores. Os dados necessários à realização da pesquisa foram obtidos exclusivamente da página oficial do TC na internet, tanto no caso dos dados individualizados do IEG-M, quanto no caso dos dados individualizados dos pareceres de contas relativamente ao exercício fiscal de 2017. Por isso, não existem dados internos à organização estudada. Como limitação, não consta na listagem de pareceres prévios informação sobre se os resultados dos julgamentos já consideraram os eventuais recursos interpostos pelos municípios, nem sobre o trânsito em julgado das respectivas sentenças. Isto posto, para a elaboração do modelo preditivo de parecer de contas foi necessária a coleta de apenas dois arquivos: um com os dados dos IEG-M de todos os municípios e outro com os dados dos pareceres de contas destes, ambos relativos ao exercício de 2017. Os arquivos foram obtidos em formato PDF e transformados em formato de planilha eletrônica por meio de uso de sítio conversor de formatos de documentos da rede mundial de computadores. Posteriormente, à planilha que continha os resultados do IEG-M dos correspondentes municípios agregou-se aquela que possuía os pareceres prévios por meio da coluna “Parecer”. A Tabela 1 exemplifica o resultado dessa junção de planilhas. As colunas “Posição” e “Município” possuem caráter meramente de contagem e de ordem e não se referem a qualquer tipo de ranqueamento entre os municípios analisados.

**Tabela 1** – Resultados do IEG-M e o Parecer Prévio do TC por município – exercício de 2017

Pos.	Município	iEduc	iSaude	iPlan	iFiscal	iAmb	iCidade	iGovti	IEG-M	Parecer
1	Cidade 1	C+	B	B	B	B	C	C+	B	Favorável
(..)	(..)	(..)	(..)	(..)	(..)	(..)	(..)	(..)	(..)	(..)
N	Cidade N	A	B+	B	C+	C	B	B+	B	Não favorável

Fonte: os autores.

Terminada essa fase de obtenção de dados, passou-se à análise deles.

## 2.5 Análise de dados

Uma vez definido o objetivo do estudo, conhecidos o sujeito e a metodologia, definida a população-alvo e coletados os dados, foi preciso inicialmente contextualizar todo o processo de exame do problema, e dentro dele desenvolver a solução. Para tanto, obedeceram-se às seguintes etapas, adotando-se e adaptando-se os fluxos propostos por Géron (2021):

1. Estudou-se o problema:
  - 1.1 Estabeleceram-se as variáveis independentes e dependente do objeto do estudo;
2. Definiu-se a metodologia de solução do problema;
3. Houve a escolha do aplicativo para desenvolver sua solução;
4. Trataram-se os dados, desenvolveu-se e treinou-se o algoritmo;
5. Realizou-se inspeção para a verificação da eventual necessidade de atualização dos dados e para a depuração de erros, realizando-se as devidas correções;
6. Avaliou-se a solução; e
7. Disponibilizou-se a solução (modelo de classificação).

Com já abordado no tópico anterior, definiu-se o sujeito do problema, que são os municípios e sobre os quais se deseja estimar ou prever com antecedência quais seriam os resultados dos julgamentos das prestações de contas a partir dos resultados da avaliação dos respectivos IEG-M. O objetivo estabelecido foi o de incrementar os processos de auditorias e fiscalizações e, dessa forma, contribuir com a gestão municipal indicando pontos de melhoria à gestão das prefeituras municipais.

Isto posto, definiram-se como variáveis independentes as dimensões do IEG-M e respectivas notas, a saber: i-Educ, i-Saúde, i-Plan, i-Fiscal, i-Amb, i-Cidade, i-Govti, IEG-M (este último relativamente à nota final dos indicadores ponderados). Assim, em cada uma dessas dimensões, os municípios tiveram as notas, em ordem decrescente na qualidade da gestão ou do gasto público: “A”, “B+”, “B”, “C+” ou “C”. Já a variável dependente são os resultados dos julgamentos das contas dos municípios, que poderiam ser “favoráveis” ou “desfavoráveis”.

A metodologia adotada foi a “árvore de classificação ou de decisão”. Para Géron (2021), árvores de classificação ou de decisão (conhecidas dentro do conceito de *Classification and Regression Tree – CART*) são algoritmos que permitem classificar dados e eventos (do ponto de vista estatístico) por meio de um processo de aprendizado de máquina (também chamado de treinamento). Em outras palavras, a partir de dados históricos, um sistema é capaz de fazer generalizações por meio de algoritmos estabelecidos. Segundo o autor, o objetivo é acertar as previsões em dados novos a partir do processo de treinamento. Nesse contexto, o experimento consistiu em criar uma árvore de classificação que, a partir dos resultados do IEG-M em cada dimensão de cada um dos municípios, fosse capaz de prever se suas contas seriam ou não rejeitadas pelo TC.

Na árvore, ocorre um processo de se ramificar os dados que mais contribuem para o poder preditivo desse tipo de modelo. Essas instâncias que contribuem para a previsão são chamadas de nós, ramos e folhas (Lauretto, 2010). Cada nó é construído a partir de atributos ou valores que, recursivamente, contribuem para o poder preditivo da informação objeto de análise (neste caso, o resultado do julgamento das contas como favorável ou desfavorável). O primeiro nó é denominado “nó-raiz” e o último de um ramo é chamado de “nó terminal” ou “folha”. Segundo Stankevix (2019), é no nó terminal que se encontram os critérios sobre os quais se obtém a resposta do modelo. Para o mesmo autor, os ramos se constituem na ligação entre dois nós.

No presente estudo, cada um desses nós indicou quais combinações de dimensões do IEG-M (variáveis independentes) produziram um resultado “favorável” ou “desfavorável” (variável dependente). O objetivo do modelo foi prever com a máxima acurácia, ou seja, com acerto, se um município teria suas contas rejeitadas quando elas realmente fossem rejeitadas e prever a aprovação quando realmente fossem aprovadas pela Corte.

Ressalva-se, tal como assevera Cristiano (2017), que a acurácia não é uma medida muito utilizada, pois ela varia com a prevalência (métrica que analisa eventuais alterações das variáveis do estudo ao longo tempo). Por outro lado, para autora, a acurácia pode ser relevante quando se deseja verificar a taxa de falsos negativos em testes diagnósticos. Isso porque, por exemplo, se um paciente recebe um diagnóstico de que está saudável, mas é de fato portador da doença, isso implica riscos à sua saúde.

Nesse mesmo sentido, corroborando a citada pesquisa, este estudo baseou-se em dois outros conceitos importantes no processo de previsão, que são a sensibilidade e a especificidade. Por sensibilidade entende-se o acerto de um evento sobre o qual se deseja prever a ocorrência, quando o evento efetivamente vier a ocorrer. E por especificidade se entende quando é previsto um não evento e essa situação de previsão se confirma no mundo real. A acurácia, por sua vez, é obtida pelo somatório dessas duas grandezas. Dentro de um mesmo modelo são possíveis



diferentes combinações de sensibilidades e especificidades ou de verdadeiros e falsos eventos e não eventos. O Manual do pacote caret mostra como são geradas e calculadas as diversas combinações dessas grandezas, por meio do que se denomina Matriz de Confusão (Figura 2).

**Figura 2 – Matriz de confusão**

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

The formulas used here are:

$$\text{Sensitivity} = A / (A + C)$$

$$\text{Specificity} = D / (B + D)$$

Fonte: Manual do pacote caret (2022).

Como se verifica, o valor “A” é a previsão do evento que se confirmou no mundo real; “B” é a previsão de evento que, por outro lado, não ocorreu no plano fático. O resultado “C” é a previsão de que não teria ocorrido o evento, mas que, na realidade, ocorreu. E “D” é um não evento real que foi previsto como não evento. A sensibilidade é dada por  $A/(A+C)$ , e a especificidade por  $D/(B+D)$ .

A análise comparativa de sensibilidade e especificidade envolveu também o cálculo da taxa de falsos positivos, ou seja, situações que se caracterizam pela ocorrência de não evento quando foi prevista a ocorrência de um evento (letra B da Figura 2). Essa taxa é conhecida como “1-especificidade”, pois é obtido justamente pela diferença  $1-D/(B+D)$ . Na presente pesquisa, foi definido, em observância ao manual do pacote rpart e de acordo com os cálculos efetuados – de forma independente e autônoma do controle do pesquisador –, que o evento é o parecer desfavorável, e que o não evento é o parecer favorável.

Dessa forma, como já esclarecido, o objetivo da construção do modelo foi prever com o máximo de acertos quais contas dos municípios seriam rejeitadas e minimizar previsões incorretas de que as contas seriam aprovadas quando, na verdade, teriam sido rejeitadas. Isso porque esse tipo de erro induziria, no momento do planejamento das auditorias, à exclusão de determinados municípios em procedimentos fiscalizatórios, o que levaria a um custo para a sociedade do ponto de vista da eficiência, eficácia, economicidade e efetividade das políticas públicas. Com efeito, mirou-se na obtenção de um modelo cujas taxas de acurácia, sensibilidade e especificidades são as maiores possíveis.

Além dessa análise, a plotagem das diversas combinações de sensibilidades e “1-especificidades” no plano cartesiano possibilitou a geração de um tipo de gráfico denominado “curva ROC”, cuja área é denominada de AUC (area under curve). Segundo Fávero e Belfiore (2021), quanto maior a área AUC da curva ROC, maior a qualidade do modelo. Cristiano (2017) acresce que, quando os valores de sensibilidade e “1-especificidade” se aproximam da igualdade, menor o poder discriminante do teste diagnóstico. Dessa forma, o desenvolvimento do algoritmo e, conseqüentemente, do modelo em estudo, avaliou essas combinações de acurácia, sensibilidade e especificidade.



Feitas essas considerações, a análise dos dados foi realizada no aplicativo RStudio (versão 2022.12.0 Build 353), o qual, segundo Alcoforado (2021), é um sistema computacional estatístico e gráfico desenvolvido com a finalidade de permitir ao usuário a construção de modelos preditivos por meio de aprendizado de máquina, entre outras funcionalidades, as quais se viabilizam por meio da linguagem R. Esse sistema possui algumas funções básicas e outras que devem ser instaladas no computador em que forem desenvolvidos os modelos de análise de dados. A autora assinala que elas são fornecidas em “pacotes”, os quais podem ser adquiridos em uma rede de suporte do RStudio chamada de The Comprehensive R Archive Network (CRAN). Feitas essas considerações, foi realizada a importação dos dados – do formato de planilha eletrônica (no caso XLS) para o R – e a nova tabela obtida foi salva em arquivo do tipo \*.RDA. Como primeira medida de tratamento de dados, definiu-se que cada dimensão do IEG-M seria uma coluna (variáveis independentes). Já o parecer de contas (variável dependente) foi fixado na coluna mais à direita dessa mesma tabela. Os municípios, por sua vez, foram caracterizados como observações da tabela. A partir desse ponto, passou-se à construção do modelo, como já definido, por meio do método de aprendizado de máquina por árvore de classificação.

Como tarefa inicial (etapa 1), foram pesquisados na rede mundial de computadores diversos pacotes geradores de códigos (também denominados algoritmos ou scripts) para construção de modelos, tais como *stackoverflow* e *towardscience*. Dentre os pacotes pesquisados – cujas funcionalidades podem ser encontradas em cada um dos respectivos manuais disponíveis no CRAN –, foram adotados os seguintes:

1. *Dplyr*: foi utilizado no processo de organização da base dados, conhecido como *datawrangling*. Uma das funções utilizadas foi o *mutate*, para alterar a classificação dos indicadores do IEG-M – de caractere (*character*) para fator (*factor*).
2. *Rpart*: pacote responsável por ter gerado o modelo (ou árvore). Possui também o comando que permitiu gerar gráficos a partir do modelo proposto. Também possibilitou o ajuste dos hiperparâmetros, como o custo de complexidade (*cp*).
3. *Rpart.plot*: pacote que possui funções que lhe conferem a característica de função estendida do pacote *Rpart*, como o *rpart.predict*, tendo sido utilizado em conjunto com o *rpart*.
4. *Metrics*: permitiu obter a área sob a curva ROC, com o comando *auc*.
5. *MLr*: é o pacote que possibilitou a combinação otimizada de hiperparâmetros (processo denominado de *tuning*), resultando em maximização da acurácia do modelo na técnica de classificação.
6. *Ggplot2*: este pacote foi utilizado para a geração de gráficos.
7. *Plotly*: idem ao pacote *ggplot2*.
8. *Tidyverse*: trata-se de uma compilação de diversos pacotes que compartilham um design comum.



9. Scales: este pacote foi utilizado para a configuração de gráficos; por exemplo, eixos e legendas, tendo sido utilizado junto com o ggploty2.
10. Caret: foi utilizado para as funções de treinamento e teste do modelo na presente árvore de classificação.
11. pROC: este é outro pacote que permitiu o cálculo da área sob a curva ROC.
12. Rmisc: permite o cálculo *summarySE*, ou seja, calcula média, desvio-padrão, entre outros valores.

O pacote inicial necessário para a criação da árvore de classificação adotado nesse trabalho foi o rpart. Esse pacote possui uma das funções mais essenciais na construção desse tipo de modelo. E para a obtenção dessa árvore, foram adotados critérios que pudessem maximizar a capacidade de generalização ou predição (hiperparâmetros). Na Figura 3, é possível a visualização desses hiperparâmetros, a partir do comando `getParamSet("classif.rpart")`. Note-se que a coluna “*type*” informa se a variável é inteira ou discreta, a coluna “*Def*” (do inglês default ou padrão) mostra os valores-padrões adotados pelo R, quando não definidos pelo usuário. A coluna “**Constr**” mostra o intervalo de números possíveis em cada hiperparâmetro e, por fim, se o elemento é sincronizável ou não.

Figura 3 – Hiperparâmetros de uma árvore de classificação

```
> getParamSet("classif.rpart")
      Type len  Def  Constr Req Tunable Trafo
minsplitt integer - 20 1 to Inf - TRUE -
minbucket integer - - 1 to Inf - TRUE -
cp numeric - 0.01 0 to 1 - TRUE -
maxcompetet 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 -
> |
```

Fonte: RStudio.

Na análise das funções e pacote do desenvolvimento de árvores, Stankevix (2019) e Bernardo (2022) elucidaram o significado dos principais hiperparâmetros com base no Manual do rpart. O minsplitt é o menor número de amostras no nó-pai. Por padrão, adota-se como 20. Já o minbucket é o menor número de observações no nó terminal. O maxdepth é o tamanho do crescimento da árvore, o qual pode chegar a 30. Já o **complexity parameter** (em português, parâmetro ou custo de complexidade – cp) é a medida de trabalho necessário para a execução de um algoritmo (Barbosa, Toscani e Ribeiro, 2002). Lauretto (2010) explica que quando ocorre um excessivo crescimento, a árvore pode vir a apresentar atributos que agregam pouco poder de previsão, ou até mesmo não servir para prever resultados para dados novos (fenômeno



chamado de *overfitting*). Para reduzir esse fenômeno, pode-se atribuir ao algoritmo, segundo o autor, esse custo de complexidade para encontrar a árvore mais adequada ao problema a ser resolvido. O pesquisador ressalta que a árvore a ser desenvolvida é aquela que não cresce tanto a ponto de incorporar nós que pouco agregam do ponto de vista informacional, nem que seja cortada a ponto de se desperdiçar nós com poder preditivo. Por último, o *xval*, que é o número de validações cruzadas, de acordo com o Manual do *rpart*. Gerón (2021) define a validação cruzada como a divisão da amostra de treino em uma menor, que serve de validação, podendo-se fazer uso, para tanto, do método *k-fold*.

Feitas essas considerações, adotou-se o modelo proposto por Bernardo (2022) para a elaboração da etapa 1. Nesse pacote, o autor utilizou o pacote *mlr*, que automatiza os cálculos dos hiperparâmetros e mostra nestes os resultados a respeito de cada variação. E foi a partir desse algoritmo que se elaborou o algoritmo referente ao modelo preditivo de parecer de contas. Isto posto, o procedimento foi o seguinte:

1. Na etapa 1, foram carregados os pacotes estabelecidos por Bernardo (2022), a saber: *dplyr*, *rpart*, *rpart.plot*, *metrics*, *mlr*, *ggploty2*, *plotly*.
2. Posteriormente, carregou-se (*load*) a planilha com os dados das notas do IEG-M e os respectivos pareceres de contas. Nessa etapa, os caracteres foram convertidos em fatores.
3. Definiu-se o tamanho das amostras em treino (75%) e teste (25%).
4. Foi executado o comando *rpart* para gerar a árvore.
5. Estabeleceu-se o primeiro hiperparâmetro para teste, que foi o *maxdepth* (1:30) – comando *makeParamSet*.
6. Definiu-se o parâmetro de *cross-validation* em 3 iterações (*makeResampleDesc*).
7. Por meio dos comandos *tuneParams* e *generateHyperParsEffectData*, foi gerado o gráfico da variação da acurácia em função do *maxdepth*, conforme Figuras 4 e 5.
8. Por fim, com o comando *setHyperPars*, foi possível armazenar os melhores resultados da sincronização, os quais puderam ser exibidos a partir da criação dos objetos *best\_parameters*, *best\_model* e *d.tree.mlr.test* e, posteriormente, executando-se os comandos *predict* e *accuracy*, para obter as previsões e a respectiva acurácia do modelo.
9. A sincronização foi repetida, mas aumentando-se para quatro hiperparâmetros (*maxdepth*, *minbucket*, *minsplit* e *CP*), os quais estão representados pela Tabela 2 e pela Figuras 6 e 7. Para efeitos de análise, os intervalos foram arbitrariamente escolhidos. Quanto ao *maxdepth*, o intervalo visou mitigar o demasiado crescimento da árvore, gerando *overfitting* e, ao mesmo tempo, evitar o descarte de variáveis (dimensões do IEG-M) com poder preditivo, conforme preceitua Lauretto (2010).



10. Na etapa 2, foi gerado o modelo – a árvore de classificação propriamente dita – com base nos hiperparâmetros obtidos na etapa 1. Para sua execução, foram carregados os seguintes pacotes: tidyverse, gtools, rmisc, scales, pROC. A execução desses pacotes em separado visou evitar a sobreposição de comandos que são comuns em dois ou mais pacotes. Ademais, o método de elaboração adotado para o desenvolvimento da árvore foi o class, conforme detalhamento do manual do pacote rpart. Outras classes possíveis seriam “anova” ou “Poisson”, mas aplicáveis a outras análises de aprendizado de máquina.
11. Concluído o carregamento de pacotes e tomando-se como referência os hiperparâmetros calculados na etapa anterior, executou-se a segunda etapa do algoritmo, que foi o desenvolvimento da árvore (Figura 8) com base nos hiperparâmetros sincronizados.
12. A árvore elaborada apresentou pontos de corte (cutoff, em inglês), Figura 9, ou seja, percentuais sobre os quais o R utilizou como parâmetro para calcular a probabilidade de que um evento fosse categorizado em “desfavorável” (evento) e “favorável” (não evento). E, a partir destes pontos de corte, calcularam-se os diferentes valores de acurácia, de sensibilidade e de especificidade das amostras de treino e teste. *Vide* Figura 10.
13. Já na etapa 3, classificaram-se as observações (IEG-M dos municípios) em “favorável” e “desfavorável” a partir do comando predict. Nas Figuras 11 e 12, é possível visualizar duas entre as diferentes matrizes de confusão possíveis neste modelo.
14. Depois desse passo, gerou-se uma tabela com os valores previstos e observados e as respectivas porcentagens de probabilidade de resultado “favorável” ou “desfavorável” no julgamento de contas, de acordo com o modelo desenvolvido. A Figura 13 permite visualizar o resultado desse procedimento.
15. Por fim, ainda na etapa 3, gerou-se a curva ROC, tanto na base de treino quanto na de teste (Figuras 14 e 15).
16. A partir do modelo desenvolvido, calcularam-se as áreas AUC da curva ROC.

Como é possível observar, foram seguidas todas as etapas propostas por Géron (2021). Identificaram-se o problema e as variáveis independentes e dependente; pesquisou-se a metodologia e definiu-se o aplicativo para o desenvolvimento da solução. A partir da escolha do RStudio, foram estabelecidos os pacotes utilizados e desenvolvidos os algoritmos com base nos manuais dos pacotes e referências disponíveis na internet. Considerando-se que não se detectaram erros no desenvolvimento do modelo, efetuou-se a avaliação dos resultados preliminares, apresentada na seção 3. Antes, contudo, foram descritas as limitações da pesquisa.

## 2.6 Limitações do estudo

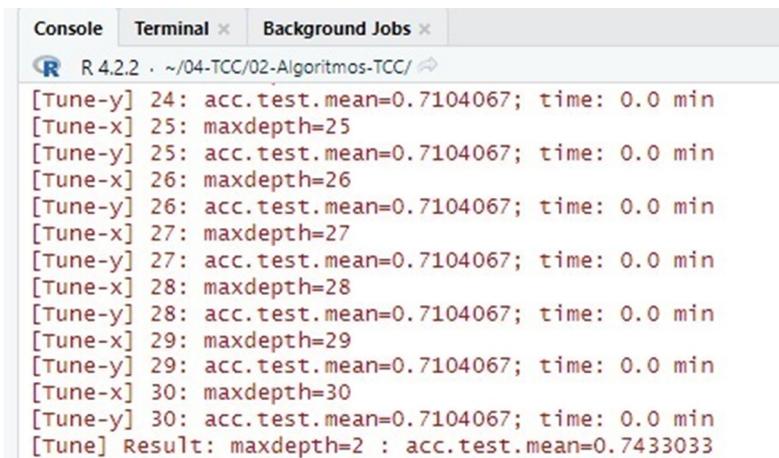
Uma das limitações do estudo foi a quantidade de exercícios fiscais com dados de parecer de contas disponível. Para o TC em estudo, foram localizados na rede mundial de computadores apenas os dados do exercício de 2017, o que impediu a análise de outros exercícios para

o mesmo sujeito. Essa dificuldade pode ser observada em relação a outros TCs do Brasil, vista a diversidade de formas de trabalho em cada TC da Federação. Lino e Aquino (2018) analisaram as diferenças existentes entre os TCs. Segundo os autores, parte desses órgãos não fiscaliza seus jurisdicionados *in loco* anualmente, havendo rodízio entre os municípios do ente. Araújo, Bezerra Filho e Motoki (2019) – na análise dos IEG-M do período entre 2015 e 2017 – observaram que alguns estados responderam ao questionário do IEG-M com dados que destoavam, em parte, da realidade encontrada. Por isso, se faz necessária a adoção de um rigoroso processo de validação dos dados. Lembra-se que alguns estados adotaram o IEG-M inicialmente sem a obrigatoriedade de adesão pelos municípios por eles jurisdicionados. É o caso do estado do Amazonas, que passou a exigir que todos eles participassem somente em 2019. Outra limitação se refere ao pacote *mlr*. Durante a execução do modelo, foi emitida a mensagem de que o pacote se encontrava em desativação. Futuros desenvolvimentos serão possíveis apenas com o pacote chamado *mlr3*, com riscos de bugs entre as versões, conforme descrito na mensagem de alerta.

### 3. RESULTADOS E DISCUSSÃO

Primeiramente, na etapa 1, o algoritmo proposto por Bernardo (2022), por meio do pacote *mlr*, permitiu a obtenção de hiperparâmetros para a geração da árvore ótima. Conforme indicado na Figura 4, variando-se um único hiperparâmetro (*maxdepth*), a maior acurácia média na execução do procedimento de *cross-validation* foi de 74,33%. Já o *maxdepth* que possibilitou a obtenção desse modelo ótimo foi igual a dois.

**Figura 4** – Sincronização de hiperparâmetros da árvore de classificação - *maxdepth* x *acc.test.mean*



```

Console Terminal x Background Jobs x
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

```

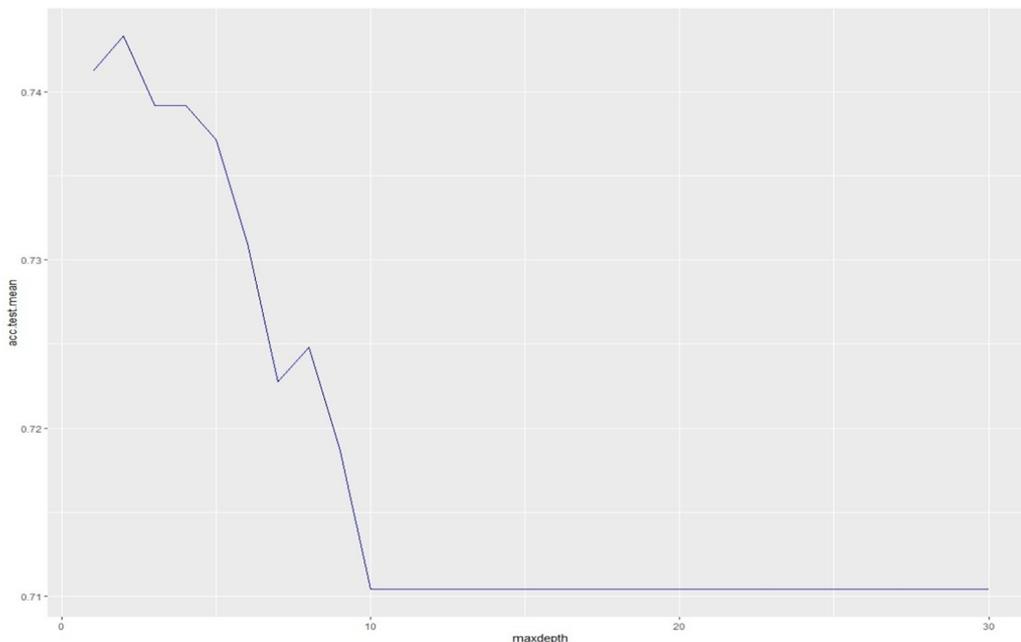
Fonte: os autores.

Nota: A expressão “[Tune-X] número” indica ao usuário quais e quantas combinações de hiperparâmetros já foram testadas até o momento. A expressão “[Tune-Y] número” indica o resultado da iteração do momento. Já “[Tune-X] Result” mostra os parâmetros ótimos e a acurácia média.



A Figura 5 mostra a variação da acurácia média da amostra de validação do modelo em função do maxdepth. É possível notar que a curva da acurácia média diminuiu à medida que o maxdepth aumentou e depois se estabilizou a partir de maxdepth igual a 10. Isso porque, à medida que a árvore se expandiu, menor foi seu poder de generalização, ou seja, menor poder de predição para outros dados que não aqueles pertencentes à amostra de treino.

**Figura 5** – Gráfico *maxdepth x acc.test.mean* (amostra de validação na validação cruzada)



Fonte: os autores.

Realizados esses procedimentos, a acurácia média máxima obtida com esse modelo, agora na amostra de teste, foi calculada em 75,15%. Em seguida, ao aumentar-se a quantidade de hiperparâmetros para quatro, novos parâmetros ótimos foram calculados e um novo resultado de acurácia foi obtido. De acordo com a Tabela 2, o maxdepth de 5, o minbucket de 10; o minsplit de 5; e o cp de 0,01 representam os hiperparâmetros ótimos a serem utilizados no modelo.

**Tabela 2** – Valores testados e o valor ótimo obtido pelo algoritmo

Hiperparâmetro	Intervalo de valores testados	Valor ótimo obtido
<b>Maxdepth</b>	04 a 20	5
<b>Minbucket</b>	01 a 10	10
<b>Minsplit</b>	01 a 10	5
<b>Cp</b>	0,001 a 0,01	0,01

Fonte: os autores.

Na realização desse procedimento (Figura 6), como já exposto, o R processou 17 mil combinações e chegou à maior acurácia média de 72,68% nos diferentes modelos resultantes das combinações (validação cruzada) de hiperparâmetros. Já o tempo de processamento desse comando pelo R ocorreu em, aproximadamente, 14 minutos e 06 segundos.

**Figura 6** – Sincronização de hiperparâmetros da árvore de classificação – *maxdepth / CP / minsplit / minbucket x acc.test.mean* (amostra de validação)

```

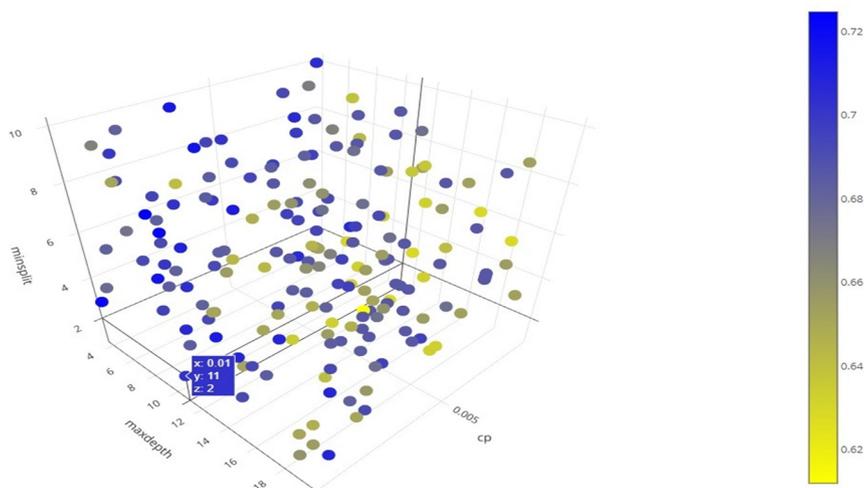
124:1 (Top Level)
Console Terminal Background Jobs
R 4.2.2 ~/04-TCC/02-Algoritmos-TCC/
[Tune-y] 16995: acc.test.mean=0.7206948; time: 0.0 min
[Tune-x] 16996: maxdepth=16; cp=0.01; minsplit=10; minbucket=10
[Tune-y] 16996: acc.test.mean=0.7206948; time: 0.0 min
[Tune-x] 16997: maxdepth=17; cp=0.01; minsplit=10; 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

```

Fonte: os autores.

A maior acurácia média da amostra de validação do procedimento de validação cruzada, conforme exposto na Figura 6, foi de 72,68%, e a que retornou o melhor resultado do modelo, segundo o algoritmo, foi de 73,24%, na amostra de teste. Graficamente, a sincronização dos quatro hiperparâmetros apresentou o resultado exposto na Figura 7. Das 17 mil combinações obtidas, foram utilizadas apenas 200 para representação do gráfico tridimensional. O R permite que, ao se posicionar o cursor em cada ponto, sejam exibidas as coordenadas (cp, maxdepth, minsplit). E a cor de cada ponto representa a acurácia a ser obtida no modelo, de acordo com a barra vertical ao lado direito. Na Figura 7, a seguir, o ponto (0,01, 11, 2), em destaque, é meramente ilustrativo. Na mesma figura, como se nota, é possível plotar até três hiperparâmetros sincronizados nos eixos cartesianos, embora a função `makeParamSet` permita a sincronização de quantidade maior de hiperparâmetros. Nesse caso, os pontos azuis representam as combinações que geram a melhor acurácia, conforme legenda à direita do gráfico. Já os pontos amarelos resultam em menores acurácias.

**Figura 7** – Combinações de *maxdepth / minsplit / CP x acc.test.mean* (amostra de validação)



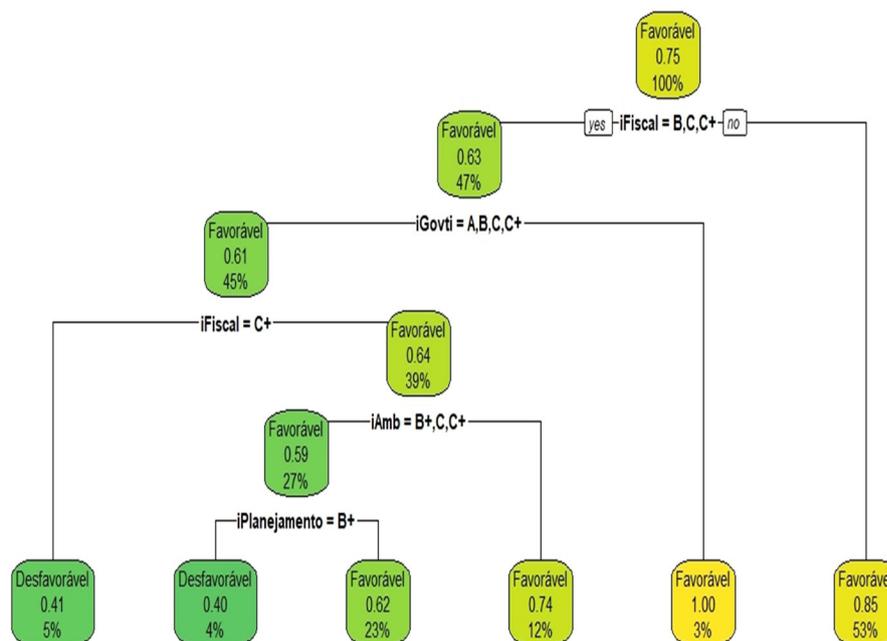
Fonte: os autores.



A vantagem do algoritmo executado na etapa 1 foi a de se estabelecerem e sincronizarem os hiperparâmetros (Figura 3) de maneira prévia ao desenvolvimento do modelo no pacote rpart. Contudo, faz-se o alerta de que o tempo de processamento no comando `tuneparams` foi maior, uma vez que mais hiperparâmetros foram estabelecidos na função `makeParamSet`. Com apenas um hiperparâmetro, foram apenas 30 iterações, procedimento que foi concluído em poucos segundos. Por outro lado, com quatro hiperparâmetros (`maxdepth`, `minbucket`, `minsplit` e `cp`), o tempo de processamento das 17 mil iterações foi de 14 minutos e 06 segundos até sua efetiva conclusão. Outra desvantagem dessa etapa é que a acurácia foi calculada pelo R sem que se variassem os percentuais de ponto de corte. Verificando-se a tabela de acurácias e respectivos pontos de corte (Figura 10), constatou-se que o R adotou 41% como ponto de corte para a previsão de resultados de pareceres favoráveis ou desfavoráveis. Dessa forma, a execução dos comandos da etapa 2 permitiram obter a variação dos valores de acurácia em função dos diferentes pontos de corte, associando-se a essa análise os valores de sensibilidade e especificidade para se encontrar o ponto de corte mais adequado ao modelo.

Em seguida, a partir desses dados obtidos na etapa 1, executou-se a segunda etapa do algoritmo, que foi a geração da árvore de classificação, apresentada na Figura 8.

Figura 8 – Árvore de classificação



Fonte: os autores.

No topo da árvore está o nó-raiz. Abaixo dele, existem dois outros nós. Um deles é um nó terminal e o outro se divide em outros dois subnós. No interior de cada nó está identificada a



variável de resultado predominante daquele nó (favorável ou desfavorável). O primeiro número percentual que aparece em cada nó, logo abaixo da variável de resultado predominante do nó, informa qual é o percentual total da amostra que possui a variável de resultado predominante do modelo (favorável). Já o número abaixo indica o percentual da condição que está expressa na linha horizontal do ramo correspondente. No caso desta pesquisa, por exemplo, verificou-se no nó-raiz que, de 100% da amostra, 75% das cidades receberam parecer favorável. Abaixo, 47% dos municípios receberam a nota “B”, “C”, ou “C+” (vide caixa com o texto “yes”) no iFISCAL – 63% dos 47% que possuíam essas três notas de iFISCAL receberam parecer favorável. Quanto aos municípios que não receberam as notas “B”, “C”, ou “C+”, ou seja, receberam notas “A” ou “B+” no iFISCAL, representaram 53% da amostra, sendo que 85% tiveram suas contas aprovadas. Voltando-se para o lado esquerdo do primeiro ramo, os municípios do que possuem notas “B”, “C”, “C+”, mas que não possuem iGOVTI com notas “A”, “B”, “C” e “C+” receberam a nota “B+” (por exclusão das demais notas). Nessa situação, esse ramo faz a ligação com o nó terminal “Favorável, 1.00, 3%”. Assim, esse nó (iFISCAL “B”, “C”, “C+” e iGOVTI “B+”) representa 3% do total da população de 644 municípios. No entanto, todos os municípios que compõem essa amostra (nó terminal) tiveram as contas aprovadas (1.00 = 100%).

Essa configuração da árvore, vale ressaltar, deve ser entendida sob o conceito de entropia e de ganho de informação. Géron (2021) e Lauretto (2010) asseveram que entropia é o grau de incerteza que existe em cada nó. Assim, se em um determinado nó existem amostras de municípios representados pelas duas decisões distintas – favorável e desfavorável –, então há, nesse nó, determinado grau de entropia. Os nós que possuem amostras representadas somente por municípios com parecer de contas exclusivamente “favorável” ou “desfavorável” não possuem entropia. Como se pode ver na Figura 8, no primeiro nó à esquerda, houve 63% de decisões favoráveis, residindo nessa situação um grau de entropia. Por outro lado, no caso do nó terminal que vem logo abaixo, com 100% das decisões “favoráveis”, denomina-se que a entropia é zero, nula ou não existente. Esse conceito é importante porque mostra na árvore quais variáveis mais contribuem para que o modelo faça previsões corretas. Assim, variáveis que geram menor entropia possuem o que se denomina “ganho de informação”, ou seja, possuem maior poder preditivo. Por isso, variáveis que geram maiores ganhos de informação estão mais próximas do nó-raiz. No caso desse estudo, o iFISCAL é a variável que permite o maior ganho de informação no modelo. Logo em seguida, o iGOV é a variável que mais gera ganho de informação, e assim sucessivamente.

A partir dessas premissas, a árvore obtida pode ser descrita por extenso na forma a seguir discriminada. Ressalva-se que as expressões “município(s)”, “prefeitura(s)” e “órgão(s)”, “amostra(gem)”, quando se referirem aos dados analisados, são consideradas sinônimas. Isto posto, constataram-se os seguintes resultados:

- a. 75% dos municípios tiveram suas contas aprovadas;
- b. 47% das prefeituras possuíam iFISCAL com notas “B”, “C” e “C+” e, destes, 63% tiveram suas contas com parecer “favorável”;
- c. 53% dos órgãos possuíam iFISCAL com notas “A” ou “B+”, de modo que 85% tiveram suas contas julgadas “favoráveis”;



- d. 45% da amostra possuem iGOVTI com notas “A”, “B”, “C”, “C+” e iFISCAL com notas “B”, “C” e “C+”;
- e. 5% dos municípios tiveram nota “C” no iFISCAL e 41% deles tiveram as contas consideradas “favoráveis”. Lembra-se que estes municípios estão contidos nas notas de iGOVTI mencionadas na alínea d;
- f. 39% dos órgãos receberam a nota “B” e “C” no iFISCAL, e 64% deles tiveram as contas aprovadas (“favorável”) pelo TC pesquisado;
- g. das prefeituras citadas na alínea f, 27% possuíam iAMB com as notas “B+”, “C”, “C+”, e 59% destes tiveram as contas consideradas “favoráveis” pela Corte de Contas;
- h. em relação à alínea g, 4% alcançaram a nota “B+” no iPLAN, sendo que 40% tiveram as contas consideradas “favoráveis”;
- i. daqueles municípios que tinham outras notas no iPLAN, o que representou 23% da amostra, 62% tiveram as contas referendadas pelo TC;
- j. 12% dos órgãos tiveram notas “A” ou “B” no iAMB, sendo que 74% tiveram as contas consideradas “favoráveis”;
- k. por fim, como já dito anteriormente, 3% das prefeituras possuíam iFISCAL com notas “B”, “C” e “C+” e, ao mesmo tempo, iGOVTI com nota “B+”. Esse foi o único caso em que a entropia foi zero.

Além dos resultados em cada um dos nós e ramos – considerando-se que o nó terminal apresentou seis folhas –, obtiveram-se os valores dos critérios necessários à formulação das respostas do modelo, ou seja, os possíveis pontos de corte. Essas folhas ou nós terminais foram apresentados também por extenso no R, conforme transcrito na Figura 9. É possível observar que o próprio R destaca as folhas ou nós terminais, os quais estão com asterisco.

Figura 9 – Representação da árvore por extenso no 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) *

```

Fonte: os autores.

Com efeito, cada um dos intervalos foi utilizado na previsão do modelo, por meio da comparação dos dados reais do IEG-M com a árvore gerada, definindo-se as probabilidades de que o parecer fosse “favorável” ou “desfavorável”. Com base nesses intervalos, definiram-se os 6 pontos de corte (coluna “Valores de corte estabelecidos” – Tabela 3).

**Tabela 3** – Pontos de corte do modelo

Folha	Intervalo de probabilidades	Valores de corte estabelecidos
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

Fonte: os autores.

Quanto ao desempenho do modelo, a observação desses pontos de corte (Figura 10) permitiu a visualização de duas situações distintas, as quais demandam uma investigação dos tomadores de decisão que os utilizarão. O primeiro deles é o ponto de corte de 73%, o qual apresentou uma acurácia de 71,97%, sensibilidade (previsão correta de julgamento “desfavorável”) de 61,53% e especificidade (previsão correta de julgamento “favorável”) de 75,42%. Já para o ponto de corte de 84%, a acurácia foi menor (63,69%), mas a sensibilidade alcançada foi maior (76,92%) e a especificidade foi de 59,32%. Desta forma, é possível notar que, embora o ponto de corte “73%” tenha uma acurácia mais elevada, sua sensibilidade mostra um poder preditivo de parecer de contas desfavoráveis menor.

Figura 10 – Acurácia, Sensibilidade e Especificidade das amostras de treino e teste

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

Fonte: os autores.

Nota: Os nomes das colunas com final “tr” se referem à amostra de treino, e as de final “tes” se referem à amostra de teste.

Na Figura 11, está representada a matriz de confusão no ponto de corte de 73% e, na Figura 12, está a matriz de confusão para um ponto de corte de 84%.



**Figura 11** – Matriz de confusão de treino e de teste com ponto de corte de 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
    
```

Fonte: os autores.

**Figura 12** – Matriz de confusão de treino e de teste com ponto de corte de 84%

```

c_treino      Desfavoravel Favoravel
Desfavoravel      82     128
Favoravel         42     235
> tabtes

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

Fonte: os autores.

Em seguida, gerou-se uma tabela para comparar os valores previstos, os observados e as respectivas probabilidades (Figura 13) para cada município analisado.

**Figura 13** – Tabela comparativa: resultados “observados” x “previsão” x probabilidades de cada ocorrência de acordo com o modelo

	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

Fonte: os autores.

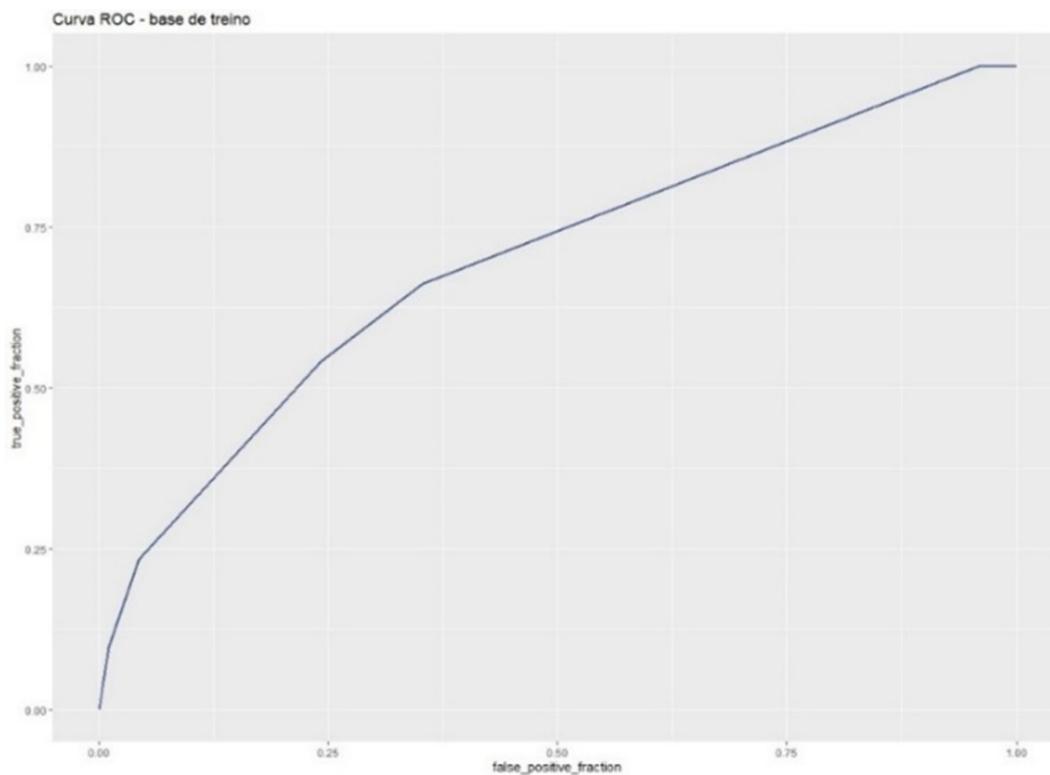


Portanto, para uma análise adequada, seria preciso gerar a matriz de confusão dos dois pontos de corte (73% e 84%) – conforme já realizado nas Figuras 11 e 12. Para o ponto de corte de 84%, na amostra de treino, dos 124 municípios com parecer desfavorável, 42 foram equivocadamente classificados como parecer favorável (33,87% de erro). E, da amostra de teste, foram 9 contas julgadas na classe favorável de um total de 39 contas desfavoráveis (23,08%). Com a redução do ponto de corte na previsão de 84% para 73%, dos 124 que foram treinados no modelo, 57 foram classificados erroneamente como “favoráveis”, ou seja, 15 unidades a mais que a amostra de teste no ponto de “84%”. Já na amostra de teste, o erro aumentou de 9 para 15 prefeituras. Com efeito, são 21 erros de previsão a mais. Dos 644 municípios, isso representa apenas 3,73% da população em estudo. Mas se forem levados em consideração apenas os órgãos que receberam parecer desfavorável pela Corte – em um total de 157 –, essa taxa sobe para 13,38%. Por isso, a escolha do ponto de corte é uma decisão que deve ser tomada pelos responsáveis pelo planejamento das auditorias. O uso de outras fontes de informação pode contribuir para o incremento do poder preditivo do modelo ou para a mitigação de erros e, assim, possibilitar a melhor tomada de decisão do ponto de vista institucional.

Em relação à curva ROC, dois resultados foram obtidos. Na base de treino, a área obtida foi de 0,6937 e, na base de teste, de 0,7106. Esse valor independe do ponto de corte do modelo já estabelecido, conforme se verifica nos cálculos. Na presente pesquisa, não se examinaram diferentes formulações de curva ROC. Um aspecto importante da curva ROC, conforme se verifica nas Figuras 11, 12, 14 e 15, é que os pontos de corte “73%” e “84%” representam os pontos de inflexão das curvas plotadas no plano cartesiano. Os pares ordenados (1-especificidade, sensibilidade) das curvas de treino e teste refletem o contido na Figura 10 e nas matrizes de confusão das Figuras 11 e 12. Os pares (0,40, 0,76), do ponto de corte de “84%”, mostram um ponto mais elevado no eixo das ordenadas do que o ponto (0,24, 0,61). Desta forma, é possível visualizar graficamente a maior sensibilidade ou poder preditivo do evento, ou seja, maior desempenho na detecção de contas rejeitadas. Segundo os estudos de Cristiano (2017), quando o teste diagnóstico utiliza um ponto do plano cartesiano no meio do espaço da curva ROC, os verdadeiros positivos são ligeiramente superiores aos falsos positivos, acarretando o que se denomina de critério “estrito” de escolha do teste diagnóstico. Demais disso, observou-se nesta pesquisa que a curva ROC na amostra de teste (0,7106) possui área ligeiramente superior à área da amostra de treino (0,6937).

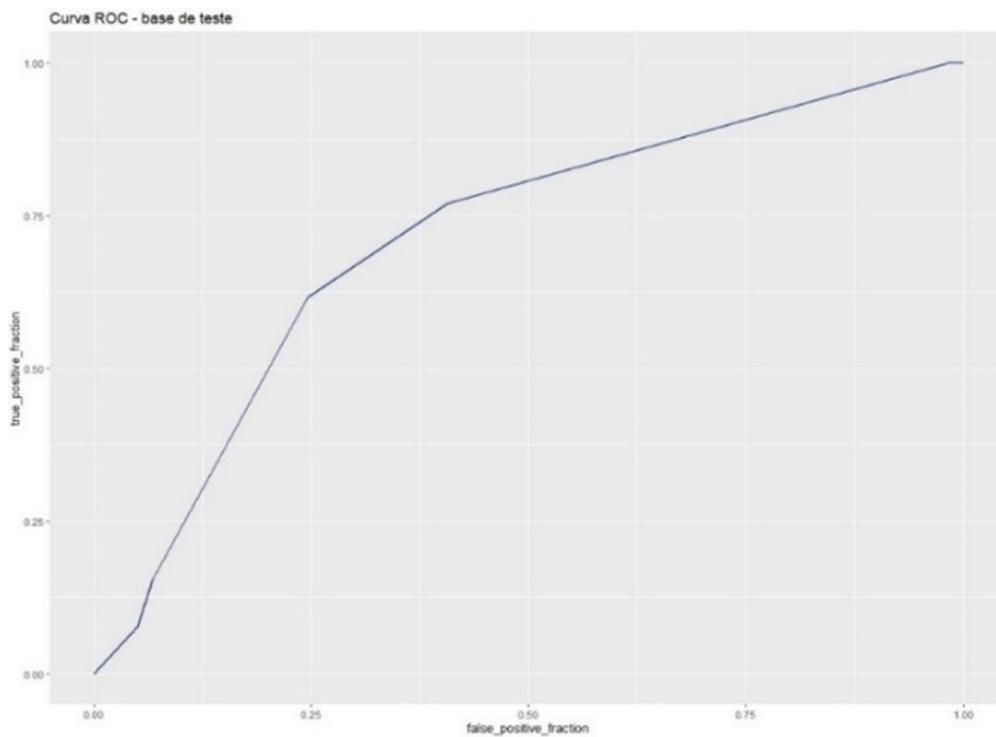


**Figura 14 – Curva ROC – Base de treino**



Fonte: os autores.

**Figura 15 – Curva ROC – Base de teste**



Fonte: os autores.



No desenvolvimento da pesquisa sobre sensibilidade e especificidade na curva ROC, Cristiano (2017) desenvolveu dois aspectos de avaliação do modelo proposto. Um deles foi a procura pela melhor curva ROC e o outro representou a busca pelo melhor ponto de corte. Esse último aspecto foi também adotado no presente estudo. Para tanto, é preciso lembrar da função constitucional dos TCs, que é avaliar a gestão dos chefes do Poder Executivo. Com isso, uma avaliação equivocada de que o município receberia parecer favorável poderia induzir a Corte a deixar de intensificar os esforços fiscalizatórios sobre uma determinada prefeitura sob má gestão. Assim, embora existam outras técnicas de planejamento de acompanhamento dos órgãos, a premissa do modelo deve ser a identificação do maior número de cidades com problemas de execução das políticas públicas. Por isso, a abordagem acerca da obtenção do melhor ponto de corte foi a solução adotada.

Quanto às pesquisas predecessoras, o modelo desenvolvido nesta pesquisa corrobora a análise de Macieira (2016), visto que as receitas possuem correlação negativa com o parecer desfavorável. Isso significa que, quanto maior a receita, menor a probabilidade de o município ter suas contas rejeitadas. Dessa forma, como no IEG-M as receitas são captadas pelo iFISCAL – variável que demonstrou o maior ganho de informação, estando, portanto, posicionada logo após o nó-raiz –, observa-se que há convergência com as conclusões daquele autor. Na pesquisa que comparou o IEG-M com os pareceres de contas obtidos pelos municípios potiguares, embora não se tenha identificado correlação entre o IEG-M e o pareceres desfavoráveis e não tenha sido possível concluir pela correlação positiva entre o IEG-M e o parecer de contas favorável, Rodrigues (2022) demonstrou que as cidades que possuíam superávit fiscal no período avaliado apresentaram melhores resultados no iEDUC e no IEG-M Geral, se comparados com municípios que possuíam déficit nessa dimensão. Outra importante conclusão desse estudo é que o volume de receitas não foi relevante para o resultado do IEG-M Final. A autora assinalou, contudo, que o fato de que 156 dos 158 municípios tiveram as contas rejeitadas no exercício fiscal examinado constituiu-se em fator limitador da pesquisa. Nesse contexto, a posição do iFISCAL como a variável de maior efeito preditivo no modelo alinha-se aos achados de Macieira (2016) e Rodrigues (2022) no tocante ao papel do superávit nos resultados do IEG-M. No modelo obtido nesta pesquisa, o iEDUC e iSAUDE – dimensões atinentes a duas políticas públicas prioritárias e especificamente normatizadas pela Constituição Federal – não constaram da árvore de classificação ótima. Contudo, ao se efetuar o procedimento de alternar o maxdepth para 6 e 7, respectivamente, essas dimensões passaram a compor as novas árvores obtidas. Nesse contexto, destaca-se que as pesquisas retrocitadas não apontaram correlação entre cada uma das sete dimensões do IEG-M (isoladamente) e os pareceres de contas, o que é corroborado pelo fato de essas duas dimensões não terem se mostrado nos níveis mais próximos do nó-raiz, justamente por não possuírem elevados ganhos de informação. Por isso, o aumento do maxdepth, a título de simulação do modelo, embora não sirva para fins preditivos, pode ajudar a compreender o comportamento das possíveis variáveis preditoras e a aperfeiçoar as dimensões do IEG-M.

#### 4. CONSIDERAÇÕES FINAIS

Os resultados do estudo demonstraram que foi possível o desenvolvimento de um modelo de previsão do resultado do julgamento das prestações de contas dos prefeitos com base



nas notas do IEG-M dos respectivos municípios, o que de fato ocorreu com a construção de uma árvore de classificação. E, entre as dimensões do IEG-M, a que mais apresentou ganho de informação ou poder preditivo foi a do iFISCAL. Outro aspecto verificado neste trabalho é que uma boa acurácia do modelo, por si só, pode não atender aos objetivos da organização. No presente caso, umas das principais missões do TC analisado era encontrar municípios com problemas de gestão. Por isso, a variável sensibilidade, ao medir municípios com chances de ter seu julgamento corretamente medido como desfavorável, tornou-se um aspecto preponderante na escolha do ponto de corte. Ademais, a escolha dos hiperparâmetros, notadamente o maxdepth (profundidade da árvore) foi relevante, de modo a permitir uma capacidade de generalização do modelo. O procedimento de sincronização foi um dos meios adequados para realizar essa tarefa.

Contudo, é preciso ressaltar que, apesar de o IEG-M representar um importante componente do atual formato de julgamento das contas pelos TCs brasileiros, ele não é o único aspecto levado em consideração para essa tarefa. Existem outros parâmetros – objetivos e subjetivos – para avaliar a gestão das prefeituras e dos gestores de modo geral. Por isso, propõe-se que novos trabalhos possam aprimorar o assunto por meio do desenvolvimento de outras análises ou modelos. A utilização do método de validação cruzada com podas, por meio do comando printcp do pacote rpart, ou a utilização de modelos de regressão são exemplos dessas outras vertentes. Sugere-se também a replicação desta pesquisa em outros estados e exercícios fiscais, com vistas a aferir se as árvores ótimas obtidas diferem da presente, e as causas dessas possíveis diferenças. Tendo em vista que o IEG-M é aferido meses antes de serem julgadas as contas dos municípios – e, por vezes, superando-se até mais do que um ano de intervalo processual, se considerados os períodos recursais até o trânsito em julgado –, a vantagem de se adotar esse tipo de modelo preditivo é que ele pode contribuir para o planejamento dos trabalhos de fiscalização dos exercícios subsequentes.

## REFERÊNCIAS

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 providências. Manaus: TCE/AM, 2019. Available at: [https://www2.tce.am.gov.br/?page\\_id=20964#68-130-2019](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](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://TheComprehensiveRArchiveNetwork.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 grid 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; JUDD, 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://TheComprehensiveRArchiveNetwork.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](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](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](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](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://TheComprehensiveRArchiveNetwork.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://TheComprehensiveRArchiveNetwork.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://TheComprehensiveRArchiveNetwork.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](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://congressosp.fipecafi.org/anais/16Usplnternational/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://congressosp.fipecafi.org/anais/17Usplnternational/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://TheComprehensiveRArchiveNetwork.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://TheComprehensiveRArchiveNetwork.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](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://TheComprehensiveRArchiveNetwork.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 TCE-SP. **TCE-SP**, 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](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](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](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://TheComprehensiveRArchiveNetwork.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://TheComprehensiveRArchiveNetwork.r-project.org/web/packages/rpart/index.html>. Consulted on: 23 jan. 2023.

TRIVIÑOS, Augusto Nivaldo 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://TheComprehensiveRArchiveNetwork.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://TheComprehensiveRArchiveNetwork.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 Comprehensve 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.

WILLEMANN, Marianna Montebello. **O desenho institucional dos Tribunais de Contas e sua vocação para a tutela da accountability 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](https://www2.dbd.puc-rio.br/pergamum/tesesabertas/1221597_2016_completo.pdf). Consulted on: 6 nov. 2022.

---

Os conceitos e interpretações emitidos nos trabalhos assinados são de exclusiva responsabilidade de seus autores.

