


**IDENTIFICAÇÃO DE DETERMINANTES INSTITUCIONAIS DA QUALIFICAÇÃO  
TÉCNICA NAS IES BRASILEIRAS SOB UMA PERSPECTIVA MATEMÁTICA E  
MÉTODOS DE MACHINE LEARNING**

**IDENTIFICATION OF INSTITUTIONAL DETERMINANTS OF TECHNICAL  
QUALIFICATION IN BRAZILIAN HEIs FROM A MATHEMATICAL PERSPECTIVE  
AND MACHINE LEARNING METHODS**

**IDENTIFICACIÓN DE DETERMINANTES INSTITUCIONALES DE LA CALIFICACIÓN  
TÉCNICA EN LAS IES BRASILEÑAS DESDE UNA PERSPECTIVA MATEMÁTICA Y  
MÉTODOS DE APRENDIZAJE AUTOMÁTICO**

 <https://doi.org/10.56238/arev8n2-117>

**Data de submissão:** 25/01/2026

**Data de publicação:** 25/02/2026

**Tarcila Oliveira Matos Muniz**

Doutoranda em Modelagem Computacional em Ciência e Tecnologia

Instituição: Universidade Estadual de Santa Cruz (UESC)

E-mail: [matostarcila@gmail.com](mailto:matostarcila@gmail.com)

Lattes: <http://lattes.cnpq.br/5834283034031897>

**Gabriel de Mello Loureiro**

Mestre em Modelagem Computacional em Ciência e Tecnologia

Instituição: Universidade Estadual de Santa Cruz (UESC)

E-mail: [gabrielmelloloureiro@gmail.com](mailto:gabrielmelloloureiro@gmail.com)

Lattes: <http://lattes.cnpq.br/8781390666125679>

---

## RESUMO

O objetivo deste estudo foi identificar quais fatores institucionais estão associados à presença de, no mínimo, 30% de servidores técnicos com pós-graduação nas Instituições de Educação Superior (IES) brasileiras. Para isso, utilizaram-se os microdados do Censo da Educação Superior de 2023, disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), considerando variáveis institucionais de natureza administrativa, organizacional e estrutural. A pesquisa adotou a técnica de Árvore de Decisão como método de aprendizado supervisionado, permitindo modelar padrões explicativos e interpretar relações entre as variáveis preditoras e a qualificação técnica institucional. Os resultados indicaram que a categoria administrativa da IES e a quantidade de docentes com doutorado foram os fatores mais relevantes na discriminação das instituições que atingem o limiar estabelecido. Observou-se que instituições não federais de menor porte apresentaram maior proporção relativa de técnicos qualificados, enquanto instituições maiores nem sempre demonstraram o mesmo desempenho proporcional. A avaliação preditiva revelou desempenho moderado do modelo, com acurácia superior ao acaso, porém com sensibilidade reduzida. Conclui-se que, embora a capacidade preditiva seja limitada, a técnica empregada mostrou-se eficaz para gerar insights interpretáveis acerca dos fatores associados à qualificação do corpo técnico, contribuindo para a compreensão estrutural do fenômeno analisado.

**Palavras-chave:** Censo da Educação Superior. Modelagem Preditiva. Machine Learning. Matemática Aplicada.

## ABSTRACT

The objective of this study was to identify which institutional factors are associated with the presence of at least 30% of technical staff holding postgraduate degrees in Brazilian Higher Education Institutions (HEIs). To achieve this, microdata from the 2023 Higher Education Census, provided by the National Institute for Educational Studies and Research Anísio Teixeira (INEP), were used, considering institutional variables of an administrative, organizational, and structural nature. The research adopted the Decision Tree technique as a supervised learning method, allowing the modeling of explanatory patterns and the interpretation of relationships between predictor variables and institutional technical qualification. The results indicated that the administrative category of the HEI and the number of faculty members with doctoral degrees were the most relevant factors in distinguishing institutions that reach the established threshold. It was observed that smaller non-federal institutions presented a higher relative proportion of qualified technical staff, while larger institutions did not always demonstrate the same proportional performance. The predictive evaluation revealed moderate model performance, with accuracy higher than chance, but with reduced sensitivity. It is concluded that, although the predictive capacity is limited, the employed technique proved effective in generating interpretable insights regarding the factors associated with technical staff qualification, contributing to the structural understanding of the analyzed phenomenon.

**Keywords:** Higher Education Census. Predictive Modeling. Machine Learning. Applied Mathematics.

## RESUMEN

El objetivo del presente estudio fue identificar los factores institucionales asociados a la presencia de, como mínimo, un 30% de personal técnico con estudios de posgrado en las Instituciones de Educación Superior (IES) brasileñas. Para ello, se utilizaron los microdatos del Censo de la Educación Superior de 2023, disponibilizados por el Instituto Nacional de Estudios e Investigaciones Educativas Anísio Teixeira (INEP), considerando variables institucionales de carácter administrativo, organizativo y estructural. La investigación adoptó la técnica de Árbol de Decisión como método de aprendizaje supervisado, lo que permitió modelar patrones explicativos e interpretar las relaciones entre las variables predictoras y la cualificación técnica institucional. Los resultados evidenciaron que la categoría administrativa de la IES y la cantidad de docentes con título de doctorado constituyeron los factores más relevantes en la discriminación de las instituciones que alcanzan el umbral establecido. Se observó que las instituciones no federales de menor tamaño presentaron una mayor proporción relativa de personal técnico cualificado, mientras que las instituciones de mayor tamaño no siempre demostraron el mismo desempeño proporcional. La evaluación predictiva reveló un desempeño moderado del modelo, con una exactitud superior al azar, aunque con sensibilidad reducida. Se concluye que, si bien la capacidad predictiva es limitada, la técnica empleada demostró ser eficaz para generar resultados interpretables acerca de los factores asociados a la cualificación del personal técnico, contribuyendo a la comprensión estructural del fenómeno analizado.

**Palabras clave:** Censo de Educación Superior. Modelado Predictivo. Aprendizaje Automático. Matemáticas Aplicadas.

## 1 INTRODUÇÃO

Diante das transformações estruturais observadas no sistema de educação superior brasileiro, torna-se cada vez mais relevante compreender os fatores institucionais associados à qualificação dos recursos humanos que compõem as Instituições de Educação Superior (IES). Nesse contexto, a qualificação do corpo técnico-administrativo assume papel estratégico, uma vez que contribui diretamente para a eficiência dos processos organizacionais, para o suporte às atividades acadêmicas e para a consolidação de políticas institucionais voltadas à qualidade educacional.

Assim, este estudo busca responder à seguinte questão norteadora: quais fatores estão associados à presença de, no mínimo, 30% de servidores técnicos com pós-graduação em uma instituição de educação superior? A definição do limiar de 30% fundamenta-se na adoção de um valor intermediário considerado estatisticamente e institucionalmente razoável, evitando extremos que poderiam limitar a análise — seja por representar um percentual muito elevado, restringindo o número de instituições analisáveis, seja por representar um percentual muito baixo, reduzindo a capacidade discriminatória do indicador.

Adicionalmente, a escolha desse patamar encontra respaldo indireto em parâmetros utilizados por órgãos de avaliação da pós-graduação brasileira. A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por exemplo, adota referenciais percentuais relacionados à titulação docente em processos avaliativos de programas de pós-graduação, nos quais, em determinados contextos, a presença mínima de aproximadamente 30% de docentes doutores podem ser considerada critério relevante para enquadramento em faixas de qualidade, especialmente em regiões com menor densidade acadêmica, conforme diretrizes normativas vigentes.

Sob a perspectiva institucional, a qualificação do corpo técnico também está associada à valorização profissional e à melhoria dos indicadores organizacionais. Servidores técnicos que agregam, ao longo da carreira, cursos de pós-graduação lato sensu e/ou stricto sensu frequentemente apresentam progressão funcional e salarial, o que, por sua vez, tende a elevar gradualmente o nível médio de qualificação institucional. Dessa forma, o estabelecimento de um patamar mínimo de técnicos qualificados pode ser interpretado como um indicador indireto de maturidade organizacional, bem como da capacidade institucional de promover políticas contínuas de formação e desenvolvimento profissional.

Nesse sentido, diante do cenário apresentado, o objetivo geral deste trabalho consiste em realizar uma busca sistemática por um conjunto de fatores institucionais que, quando combinados, possam oferecer explicabilidade para as questões relacionadas à qualificação do corpo técnico no setor de educação superior. Busca-se, portanto, identificar padrões estruturais e organizacionais que

estejam associados ao alcance do limiar estabelecido.

Para a consecução desse objetivo, será utilizada como base empírica o banco de Microdados do Censo da Educação Superior de 2023, disponibilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). A análise será conduzida por meio da aplicação da técnica de Árvore de Decisão, no contexto do aprendizado supervisionado, permitindo modelar relações entre variáveis institucionais e interpretar padrões explicativos associados ao fenômeno investigado. A escolha dessa técnica justifica-se por sua capacidade de gerar modelos interpretáveis, facilitando a compreensão das interações entre fatores institucionais e contribuindo para análises aplicadas ao campo da gestão educacional.

Dessa forma, espera-se que os resultados obtidos contribuam para o aprofundamento da compreensão sobre a estrutura e a dinâmica da qualificação técnica nas IES brasileiras, além de subsidiar discussões relacionadas ao planejamento institucional, formulação de políticas educacionais e tomada de decisão baseada em evidências.

## **2 REFERENCIAL TEÓRICO**

A qualificação dos recursos humanos constitui elemento central para o desenvolvimento institucional e para a melhoria da qualidade da educação superior. Segundo Dias Sobrinho (2005), a qualidade universitária deve ser compreendida de forma multidimensional, envolvendo aspectos acadêmicos, administrativos e estruturais. Nesse contexto, a formação continuada e a qualificação do corpo técnico-administrativo assumem papel estratégico na consolidação de práticas institucionais eficientes.

A gestão do capital humano, conforme argumentam Becker (1993) e Schultz (1961), está diretamente relacionada ao conceito de capital humano, entendido como o conjunto de conhecimentos, habilidades e competências incorporadas aos indivíduos, capazes de gerar produtividade e desenvolvimento organizacional. No ambiente das Instituições de Educação Superior (IES), a qualificação técnica contribui para maior eficiência administrativa, melhor suporte às atividades acadêmicas e fortalecimento institucional.

No cenário brasileiro, os dados educacionais produzidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira desempenham papel fundamental na formulação de políticas públicas e na avaliação institucional. O Censo da Educação Superior constitui, portanto, importante base empírica para análises estruturais e diagnósticos institucionais.

O avanço das técnicas de aprendizado de máquina tem ampliado as possibilidades de análise de grandes bases de dados educacionais. De acordo com Mitchell (1997), aprendizado de máquina

consiste na capacidade de um sistema computacional aprimorar seu desempenho em determinada tarefa por meio da experiência. No contexto supervisionado, os algoritmos aprendem padrões a partir de dados previamente rotulados.

As Árvores de Decisão destacam-se entre os métodos supervisionados por sua interpretabilidade e capacidade de modelar relações não lineares. O trabalho seminal de Breiman et al. (1984), ao desenvolver o método CART (Classification and Regression Trees), estabeleceu os fundamentos teóricos da técnica, baseada na divisão recursiva do espaço amostral com o objetivo de maximizar a pureza dos nós, geralmente medida pelo índice de Gini ou entropia.

Segundo James et al. (2013), uma das principais vantagens das Árvores de Decisão reside em sua transparência interpretativa, permitindo compreender como cada variável influencia a classificação final. Diferentemente de modelos mais complexos, como redes neurais, as árvores facilitam a extração de regras decisórias explícitas, aspecto particularmente relevante quando o objetivo do estudo não é apenas prever, mas também explicar fenômenos.

No âmbito educacional, a aplicação de técnicas de mineração de dados tem se mostrado relevante para identificar padrões institucionais e apoiar decisões estratégicas (REZENDE; MARCACINI; MOURA, 2011).

Hastie, Tibshirani e Friedman (2009), apontam que as árvores são capazes de capturar relações não lineares e interações entre variáveis sem a necessidade de pressupostos paramétricos rígidos, o que amplia sua aplicabilidade em contextos nos quais a estrutura dos dados é complexa ou heterogênea. No campo educacional, tais propriedades permitem explorar padrões institucionais de forma estruturada e compreensível.

Dessa forma, a utilização da Árvore de Decisão neste estudo fundamenta-se tanto em sua robustez metodológica quanto em sua capacidade explicativa. Ao modelar a relação entre características institucionais e a proporção de técnicos com pós-graduação, a técnica possibilita identificar fatores discriminantes, hierarquizar variáveis relevantes e explicitar regras interpretáveis, contribuindo para uma compreensão mais aprofundada do fenômeno investigado.

### **3 METODOLOGIA**

O presente estudo caracteriza-se como uma pesquisa quantitativa, de natureza aplicada e com abordagem explicativa, fundamentada na análise de dados secundários provenientes do Censo da Educação Superior de 2023, disponibilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. A escolha dessa base justifica-se por sua abrangência nacional e pelo detalhamento das variáveis institucionais, permitindo análises estruturais consistentes.

Inicialmente, realizou-se o pré-processamento dos dados, etapa essencial em estudos de modelagem preditiva (JAMES et al., 2013). Foram selecionadas variáveis institucionais consideradas teoricamente relevantes, incluindo categoria administrativa, organização acadêmica, localização (capital ou interior), quantidade total de técnicos e número de docentes com doutorado.

Em seguida, construiu-se a variável de interesse (variável alvo), definida como binária: instituições com proporção igual ou superior a 30% de técnicos com pós-graduação (especialização, mestrado ou doutorado) foram classificadas como 1; as demais, como 0. Esse limiar foi estabelecido com base em critérios de razoabilidade institucional e comparabilidade com parâmetros de qualificação utilizados em avaliações acadêmicas.

Para garantir consistência analítica, foram excluídas observações com valores ausentes nas variáveis selecionadas e removidos casos com total de técnicos igual a zero, evitando distorções percentuais. Posteriormente, os dados foram particionados em conjunto de treinamento (70%) e teste (30%), utilizando procedimento estratificado para manter a proporção das classes, prática recomendada na literatura de aprendizado supervisionado (KUHN; JOHNSON, 2013).

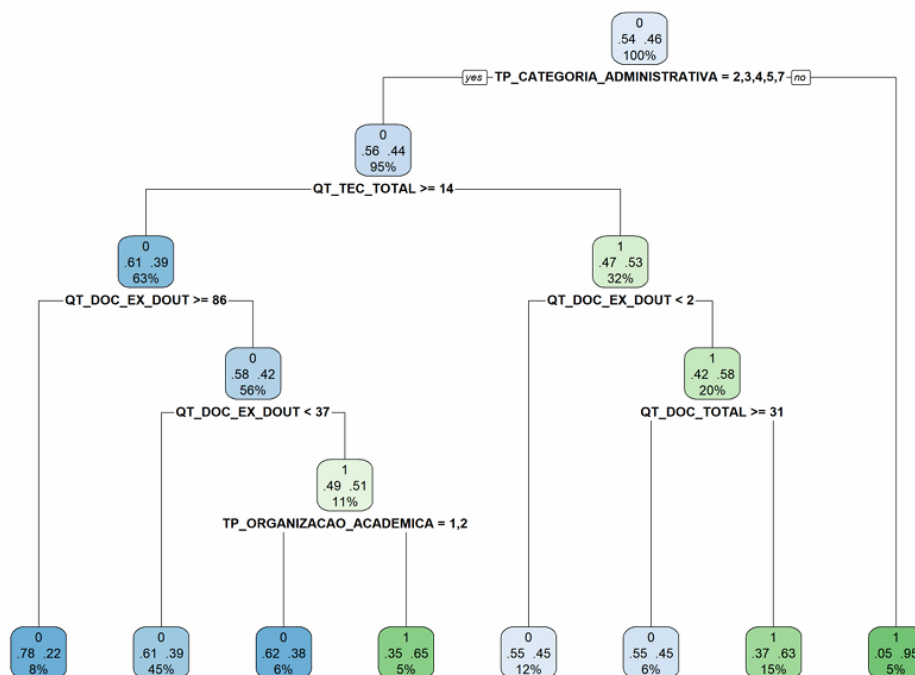
### 3.1 MODELAGEM COM ÁRVORE DE DECISÃO

A modelagem inicial foi realizada por meio do algoritmo CART (Classification and Regression Trees), conforme proposto por Breiman et al. (1984). O método baseia-se na divisão recursiva do conjunto de dados em subconjuntos mais homogêneos, buscando maximizar a pureza dos nós por meio da redução da impureza de Gini.

Foram definidos parâmetros de controle visando equilibrar interpretabilidade e prevenção de sobreajuste (overfitting), como profundidade máxima da árvore e número mínimo de observações por nó terminal. O modelo foi ajustado utilizando o conjunto de treinamento e posteriormente aplicado ao conjunto de teste.

A representação gráfica da árvore gerada é apresentada na Figura 1, na qual se observa a hierarquização das variáveis segundo sua capacidade discriminatória. O nó raiz evidencia a primeira divisão realizada pelo algoritmo, indicando a variável com maior poder de separação entre as classes.

Figura 1. Árvore de Decisão ajustada para classificação das IES quanto à proporção de técnicos com pós-graduação.



Fonte: Elaborada pelos próprios autores.

A análise estrutural da árvore permitiu identificar padrões relevantes, como a influência da categoria administrativa e da quantidade de docentes com doutorado na probabilidade de a instituição atingir o limiar estabelecido.

### 3.2 AVALIAÇÃO DO MODELO

A avaliação do desempenho foi realizada por meio da matriz de confusão, apresentada na Figura 2, que permite examinar acertos e erros de classificação, incluindo verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos.

Figura 2. Matriz de Confusão do modelo de Árvore de Decisão.

```

Confusion Matrix and Statistics

      Reference
Prediction 0  1
0      349 249
1       66 109

      Accuracy : 0.5925
      95% CI : (0.5569, 0.6274)
      No Information Rate : 0.5369
      P-Value [Acc > NIR] : 0.001049

      Kappa : 0.1507

McNemar's Test P-Value : < 2.2e-16

      Sensitivity : 0.3045
      Specificity : 0.8410
      Pos Pred Value : 0.6229
      Neg Pred Value : 0.5836
      Prevalence : 0.4631
      Detection Rate : 0.1410
      Detection Prevalence : 0.2264
      Balanced Accuracy : 0.5727

      'Positive' Class : 1
    
```

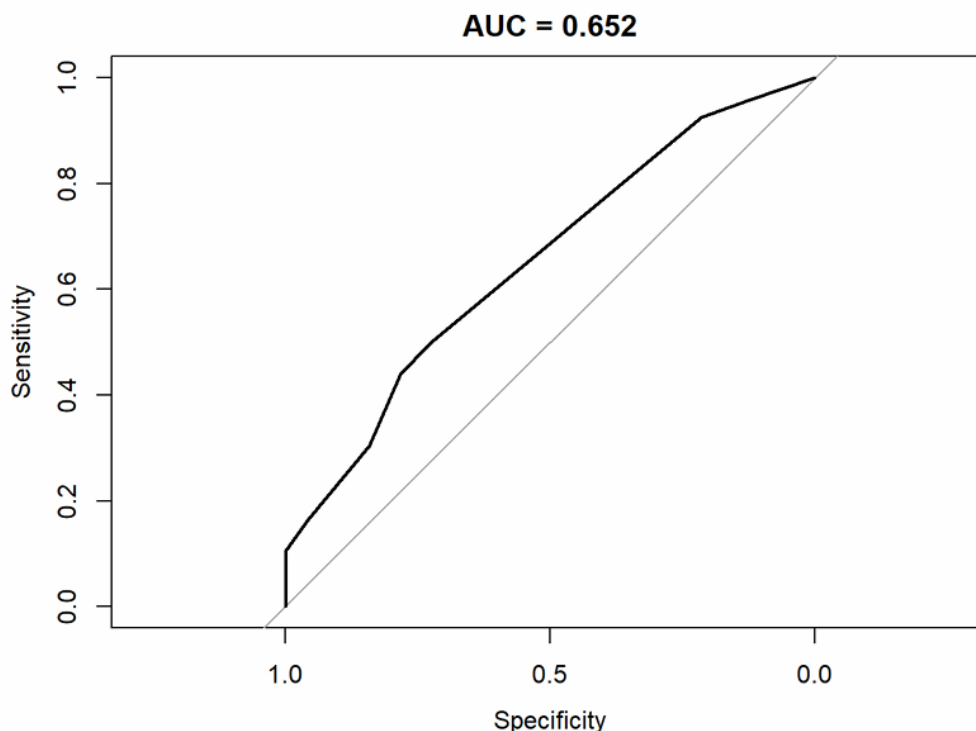
Fonte: Elaborada pelos próprios autores.

Foram analisadas métricas como acurácia, sensibilidade, especificidade e acurácia balanceada. Observou-se desempenho moderado do modelo, com maior capacidade de identificar instituições que não atingem o limiar de 30% (alta especificidade), mas menor sensibilidade na identificação das instituições qualificadas.

Adicionalmente, utilizou-se a curva ROC para avaliar a capacidade discriminatória independente do limiar de classificação, conforme recomendado por Fawcett (2006). A Figura 3 apresenta a curva ROC e o respectivo valor da área sob a curva (AUC), indicador sintético da performance do classificador.



Figura 3. Curva ROC e área sob a curva (AUC) para o modelo de Árvore de Decisão.



Fonte: Elaborada pelos próprios autores.

O valor da AUC indicou capacidade discriminatória superior ao acaso, porém distante de níveis considerados elevados, reforçando o caráter exploratório-explicativo do modelo.

#### 4 DISCUSSÃO E RESULTADOS OBTIDOS

A aplicação da Árvore de Decisão (CART) permitiu identificar a hierarquia das variáveis associadas à probabilidade de as Instituições de Educação Superior (IES) atingirem o limiar de 30% de técnicos com pós-graduação. Conforme ilustrado na Figura 1, a variável TP\_CATEGORIA\_ADMINISTRATIVA constituiu o nó raiz do modelo, evidenciando seu maior poder discriminatório. Esse resultado indica que a natureza administrativa da instituição representa o principal fator estrutural associado ao padrão de qualificação técnica.

Na sequência, a variável QT\_TEC\_TOTAL (porte institucional) foi utilizada como critério de segmentação, sugerindo efeito de escala na estrutura organizacional. Instituições de maior porte apresentaram comportamento distinto das de menor porte, indicando possível associação entre complexidade institucional e qualificação do quadro técnico.

As divisões subsequentes envolveram principalmente variáveis relacionadas ao corpo docente, como QT\_DOC\_EX\_DOUT e QT\_DOC\_TOTAL. Observou-se que instituições com maior número de docentes doutores apresentaram probabilidade significativamente mais elevada de

pertencer à classe 1. Em determinados nós terminais, a probabilidade ultrapassou 90%, indicando elevada pureza de classificação.

A matriz de confusão (Figura 2) evidenciou desempenho moderado do modelo. Verificou-se maior especificidade do que sensibilidade, indicando melhor desempenho na identificação das instituições que não atingem o limiar estabelecido. A curva ROC (Figura 3) confirmou capacidade discriminatória intermediária, com AUC superior ao acaso, porém distante de níveis considerados elevados.

Apesar da interpretabilidade do modelo — característica central das árvores de decisão (BREIMAN et al., 1984) — observou-se limitação preditiva, especialmente quanto à identificação da classe positiva. Conforme discutem HASTIE, TIBSHIRANI e FRIEDMAN (2009), árvores individuais tendem a apresentar alta variância e instabilidade estrutural.

Diante disso, fez-se necessário empregar o método Random Forest (BREIMAN, 2001), técnica ensemble baseada na construção de múltiplas árvores via amostragem bootstrap e seleção aleatória de variáveis a cada divisão. O modelo final resulta da agregação por votação majoritária.

Os resultados do Random Forest (Figura 4 – matriz de confusão do modelo ensemble) indicaram melhora marginal nas métricas, com aumento discreto da sensibilidade e da acurácia global.

Figura 4. Matriz de Confusão do método Random Forest.  
 Confusion Matrix and Statistics

```

Reference
Prediction  0  1
0  295 171
1  120 187

Accuracy : 0.6235
95% CI : (0.5883, 0.6578)
No Information Rate : 0.5369
P-Value [Acc > NIR] : 6.886e-07

Kappa : 0.2355

McNemar's Test P-Value : 0.003378

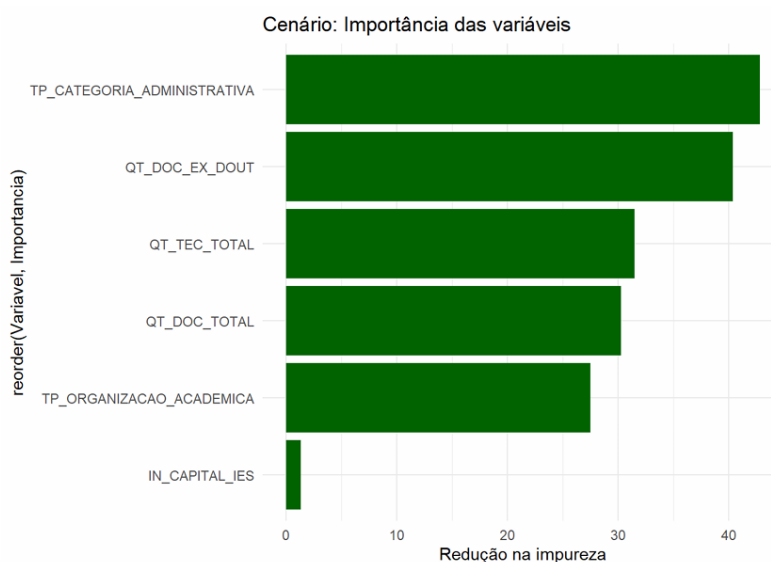
Sensitivity : 0.5223
Specificity : 0.7108
Pos Pred Value : 0.6091
Neg Pred Value : 0.6330
Prevalence : 0.4631
Detection Rate : 0.2419
Detection Prevalence : 0.3972
Balanced Accuracy : 0.6166

'Positive' Class : 1
    
```

Fonte: Elaborada pelos próprios autores.

A Figura 5 demonstrou manutenção da centralidade da categoria administrativa e da densidade de doutores, reforçando a consistência dos achados.

Figura 5. Importância das variáveis



Fonte: Elaborada pelos próprios autores.

Assim, embora o Random Forest tenha proporcionado maior robustez estatística e redução da variância do modelo, os ganhos preditivos não foram substancialmente elevados, sugerindo que parte da limitação decorre da própria natureza das variáveis disponíveis.

## 5 CONSIDERAÇÕES FINAIS

Os resultados obtidos neste estudo evidenciam contribuições relevantes tanto no plano metodológico quanto no campo da gestão da educação superior. Ao integrar técnicas de aprendizado de máquina com dados oficiais do Censo da Educação Superior, a pesquisa avançou na aplicação de métodos quantitativos explicativos ao contexto educacional brasileiro, área ainda marcada por predominância de análises descritivas.

A modelagem por Árvore de Decisão permitiu identificar, de maneira hierarquizada e interpretável, os principais fatores associados à qualificação técnica nas Instituições de Educação Superior (IES). A evidência de que a categoria administrativa constitui o principal critério de segmentação demonstra que a estrutura jurídica e organizacional das instituições influencia diretamente a composição e o desenvolvimento do capital humano técnico. Esse achado reforça a importância de considerar características institucionais estruturais em análises sobre qualidade e qualificação profissional no ensino superior.

Adicionalmente, a relevância das variáveis relacionadas à densidade acadêmica, especialmente o número de docentes doutores, evidencia que ambientes com maior capital intelectual tendem a apresentar maior probabilidade de qualificação técnica elevada. Tal resultado sugere que a qualificação do corpo técnico não ocorre de forma isolada, mas está inserida em um ecossistema institucional mais amplo, no qual o capital humano docente e a cultura acadêmica exercem influência significativa.

A aplicação do método Random Forest representou etapa fundamental para validação e robustez dos resultados. Embora o ganho preditivo tenha sido moderado, a consistência das variáveis mais importantes entre os dois modelos fortalece a confiabilidade das conclusões. A utilização combinada de um modelo altamente interpretável (CART) e um modelo ensemble mais robusto (Random Forest) demonstra maturidade metodológica do estudo e amplia sua credibilidade científica.

Do ponto de vista das políticas institucionais, os achados oferecem subsídios relevantes para gestores educacionais. A identificação de padrões estruturais associados à qualificação técnica pode orientar estratégias de desenvolvimento institucional, planejamento de recursos humanos e formulação de políticas de capacitação profissional. O estudo contribui, portanto, não apenas para o avanço teórico-metodológico, mas também para a prática da gestão educacional baseada em evidências.

Além disso, ao utilizar dados públicos nacionais, a pesquisa reforça o potencial do Censo da Educação Superior como fonte estratégica para análises preditivas e diagnósticos institucionais, incentivando o uso mais sofisticado dessas bases em investigações futuras.

Os resultados obtidos indicam caminhos relevantes para aprofundamento da investigação. Primeiramente, recomenda-se ampliar o conjunto de variáveis explicativas, incorporando indicadores financeiros, dados orçamentários e métricas de investimento em capacitação técnica, a fim de capturar dimensões institucionais não observadas neste estudo.

Como proposta metodológica, sugere-se testar modelos de maior complexidade não linear, como Gradient Boosting Machines (FRIEDMAN, 2001) e XGBoost, os quais tendem a apresentar melhor desempenho em problemas de classificação com interações complexas.

Além disso, recomenda-se explorar abordagens de validação cruzada k-fold, conforme sugerem James et al. (2013), para avaliação mais robusta da capacidade de generalização dos modelos.

Outra possibilidade consiste na análise temporal longitudinal, utilizando séries históricas do Censo da Educação Superior, o que permitiria investigar tendências na evolução da qualificação técnica ao longo dos anos.

Em síntese, o presente estudo estabelece uma base metodológica sólida para investigações futuras e contribui para consolidar a aplicação de técnicas de aprendizado de máquina no campo da avaliação e gestão da educação superior, reforçando a importância de análises estruturais baseadas em evidências para a formulação de políticas educacionais mais eficazes.

## REFERÊNCIAS

- BECKER, Gary S. Human capital: a theoretical and empirical analysis, with special reference to education. 3. ed. Chicago: University of Chicago Press, 1993.
- BREIMAN, Leo. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.
- BREIMAN, Leo; FRIEDMAN, Jerome; OLSHEN, Richard; STONE, Charles. Classification and regression trees. New York: Routledge, 1984.
- DIAS SOBRINHO, José. Avaliação da educação superior: democratização, qualidade e crise. Campinas: Autores Associados, 2005.
- FAWCETT, Tom. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006.
- HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. The elements of statistical learning. 2. ed. New York: Springer, 2009.
- JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. An introduction to statistical learning. New York: Springer, 2013.
- KUHN, Max; JOHNSON, Kjell. Applied predictive modeling. New York: Springer, 2013.
- MITCHELL, Tom M. Machine learning. New York: McGraw-Hill, 1997.
- REZENDE, Solange Oliveira; MARCACINI, Ricardo; MOURA, Maria Fernanda. Mineração de dados. In: REZENDE, Solange Oliveira (org.). *Sistemas inteligentes: fundamentos e aplicações*. Barueri: Manole, 2011.
- SCHULTZ, Theodore W. Investment in human capital. *The American Economic Review*, v. 51, n. 1, p. 1–17, 1961.