



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUI
CAMPUS FLORIANO
CURSO TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS**

VIVIANE CUNICO DA SILVA

**ANÁLISE COMPARATIVA DE PESQUISAS ELEITORAIS E RESULTADOS
OFICIAIS NAS ELEIÇÕES BRASILEIRAS PARA PRESIDENTE E GOVERNADOR
COM USO DE MACHINE LEARNING**

FLORIANO

2025

VIVIANE CUNICO DA SILVA

ANÁLISE COMPARATIVA DE PESQUISAS ELEITORAIS E RESULTADOS
OFICIAIS NAS ELEIÇÕES BRASILEIRAS PARA PRESIDENTE E
GOVERNADOR COM USO DE MACHINE LEARNING

Trabalho de Conclusão de Curso
(artigo científico) apresentado como
exigência parcial para obtenção do
diploma do Curso de Tecnologia em
Análise e Desenvolvimento de
Sistemas do Instituto Federal de
Educação, Ciência e Tecnologia do
Piauí, Campus Floriano.

Orientador: Prof. Dr. Willamys
Rangel Nunes de Sousa.

FLORIANO

2025

ANÁLISE COMPARATIVA DE PESQUISAS ELEITORAIS E RESULTADOS OFICIAIS NAS ELEIÇÕES BRASILEIRAS PARA PRESIDENTE E GOVERNADOR COM USO DE MACHINE LEARNING

Viviane Cunico da Silva¹
Willamys Rangel Nunes de Sousa²

RESUMO

Este artigo tem como objetivo principal comparar os resultados das pesquisas eleitorais com os dados oficiais das urnas do ano de 2022, utilizando técnicas de Machine Learning, com foco nas eleições para Presidente da República e Governador do estado de Pernambuco. O estudo busca analisar a confiabilidade das pesquisas de intenção de voto e identificar possíveis discrepâncias em relação aos resultados efetivamente registrados, destacando padrões e fatores que possam ter influenciado tais diferenças. A pesquisa adota uma abordagem quantitativa, com base em dados obtidos no portal Poder 360 e no Portal de Dados Abertos do Tribunal Superior Eleitoral (TSE). A metodologia contempla o pré-processamento das bases de dados, a implementação do modelo Random Forest e a avaliação de desempenho por meio das métricas MAE (Erro Absoluto Médio) e RMSE (Raiz do Erro Quadrático Médio). Os resultados evidenciam discrepâncias relevantes entre as fontes, além de apontar a importância de determinadas variáveis no processo de treinamento e previsão do modelo.

Palavras-chave: pesquisas eleitorais; machine learning; desempenho; discrepâncias eleitorais.

ABSTRACT

This article's main objective is to compare the results of electoral polls with official polling data from the 2022 election, using machine learning techniques, focusing on the elections for President of the Republic and Governor of the state of Pernambuco. The study seeks to analyze the reliability of voting intention polls and identify possible discrepancies in relation to the actual results, highlighting patterns and factors that may have influenced such differences. The research adopts a quantitative approach, based on data obtained from the Poder 360 portal and the Open Data Portal of the Superior Electoral Court (TSE). The methodology includes database preprocessing, implementation of the Random Forest model, and performance evaluation using MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) metrics. The results highlight significant discrepancies between the sources and emphasize the crucial importance of certain variables in the model's training and prediction process.

Keywords: electoral polls; machine learning; performance; electoral discrepancies.

Data de aprovação: 14/07/2025

¹ Graduanda em Tecnologia em Análise e Desenvolvimento de Sistemas. Instituto Federal do Piauí Campus Floriano. vicunico6866@gmail.com

² Professor do Ensino Básico, Técnico e Tecnológico. IFPI - Campus Floriano. rangelnunes@ifpi.edu.br.

1 INTRODUÇÃO

No Brasil, a cada dois anos, a população é convocada às urnas para escolher seus representantes em diferentes cargos, definindo os rumos políticos e administrativos do país no próximo ciclo. Durante esse período, as pesquisas eleitorais desempenham um papel crucial, servindo como ferramentas para captar, em momentos específicos, as preferências dos eleitores em relação aos candidatos que disputam uma eleição (Valentim, Seufitelli e Silva, 2023).

Segundo Venturi (1995), a importância dessas pesquisas no processo eleitoral tem crescido substancialmente, uma vez que, conforme apontam Maddalena, Kazama e Savioli (2021), elas permitem não apenas identificar as tendências e preferências da população, mas também ajustar campanhas de acordo com o cenário. Isso inclui estratégias para um eventual segundo turno, a mitigação de riscos de derrota e o reforço da comunicação com grupos estratégicos de eleitores. Entretanto, as frequentes discrepâncias entre os resultados das pesquisas e os dados oficiais divulgados pelo Tribunal Superior Eleitoral (TSE) levantam questionamentos sobre sua precisão, metodologia e impacto na percepção pública do processo democrático (Nogueira, 2012).

Nesse contexto, a análise comparativa entre os resultados das pesquisas eleitorais e os dados oficiais fornecidos pelo TSE torna-se indispensável para avaliar a confiabilidade dessas ferramentas. Compreender as possíveis discrepâncias é fundamental não apenas para aumentar a confiança no sistema eleitoral, mas também para aprimorar as práticas de pesquisa. Além disso, ao incluir uma dimensão geográfica na análise, é possível identificar padrões regionais e sociais que ajudam a explicar as diferenças entre as intenções de voto e os resultados efetivos.

Estudos recentes, como o de Valentim, Seufitelli e Silva (2023), têm abordado essa temática utilizando métodos avançados de aprendizado de máquina, como *K-Nearest Neighbors* (KNN), Regressão Linear e *Random Forest*, destacando o potencial dessas ferramentas para analisar e prever o comportamento eleitoral com maior precisão.

Embora o aprendizado de máquina seja amplamente aplicado em diversas áreas, sua utilização na comparação de dados eleitorais, especialmente com análise geoespacial, ainda é limitada. Este trabalho busca preencher essa lacuna ao treinar

um modelo de *Machine Learning* capaz de correlacionar dados de pesquisas eleitorais com boletins de urna, incorporando uma perspectiva geográfica para identificar padrões e discrepâncias. Com foco nos resultados das eleições de 2022 no Brasil, com foco no estado de Pernambuco para os cargos de presidente e governador, a proposta une ciência política e inteligência artificial.

Métricas como Erro Médio Absoluto e Raiz do Erro Quadrático Médio foram aplicadas para avaliar o desempenho dos modelos, enquanto mapas temáticos e visualizações interativas foram elaborados para destacar padrões e discrepâncias significativas. Essa abordagem busca oferecer uma análise mais aprofundada e contextualizada, contribuindo para o debate sobre a confiabilidade das pesquisas eleitorais e sua relação com o fortalecimento da democracia brasileira.

Este estudo pode oferecer contribuições práticas significativas na análise e exploração de dados ao utilizar modelos de *Machine Learning* para correlacionar pesquisas eleitorais com boletins de urna, possibilitando uma compreensão mais precisa das discrepâncias e padrões existentes, aprimorando as metodologias utilizadas pelos institutos de pesquisa.

2 REFERENCIAL TEÓRICO

Para uma melhor compreensão dos conceitos abordados neste trabalho, esta seção está organizada em tópicos, os quais são: 2.1. Pesquisas Eleitorais, 2.2. Boletins de Urna, 2.3. Comparação de Dados com *Machine Learning*.

2.1 PESQUISAS ELEITORAIS

As eleições são pilares fundamentais da democracia, pois permitem aos cidadãos exercerem o direito de escolha e influenciarem os rumos políticos de uma nação, e de acordo com Valentim, Seufitelli e Silva (2023), essas escolhas impactam diretamente as estratégias de condução do governo eleito, afetando áreas cruciais, com reflexos diretos na vida da população ao longo do mandato.

Por isso, é essencial que cada cidadão analise com cautela os candidatos, considerando que as consequências de suas escolhas nas urnas exercerão uma influência significativa sobre o cenário político futuro. Ademais, é imprescindível refletir sobre a qualidade das informações consumidas pelos eleitores, já que estas podem determinar o voto em um determinado candidato.

Nesse contexto, as pesquisas pré-eleitorais desempenham um papel crucial no processo, pois subsidiam as preferências da população e podem influenciar o voto de um indivíduo, impactando o resultado final das eleições (Venturi, 1995).

Como aponta Trein (2004), em geral, os eleitores utilizam uma quantidade limitada de informações para tomar suas decisões, dedicando pouco tempo e esforço ao processo de análise e definição do voto. As imagens dos candidatos são construídas com base em experiências passadas, conversas com amigos e familiares, propaganda eleitoral e, principalmente, informações da imprensa, adquirindo um peso significativo na decisão final. Essas percepções, muitas vezes superficiais, são formadas sem um conteúdo mais consistente, levando o eleitor a acompanhar o processo eleitoral de forma distanciada e com pouco envolvimento com a política e os políticos.

Diante desse cenário, as pesquisas possuem um papel estratégico tanto na influência do voto individual, como destaca Trein (2004), segundo o qual os eleitores, nos últimos dias da campanha, baseiam suas decisões na intenção de voto da maioria, conforme apontado pelos resultados das pesquisas, quanto para os candidatos.

Nesse sentido, Maddalena, Kazama e Savioli (2021) ressaltam que as pesquisas são ferramentas estratégicas que permitem planejar campanhas e ajustar estratégias conforme os cenários apresentados. Isso inclui, por exemplo, a preparação para um eventual segundo turno ou a identificação precoce de possíveis derrotas, possibilitando ações mais direcionadas e eficazes.

Entretanto, sempre que o Brasil passa por um novo processo eleitoral, as mesmas polêmicas envolvendo as pesquisas voltam à mídia. Uma delas, abordada no trabalho de Nogueira (2012), é a disparidade entre os resultados das pesquisas previamente realizadas e o resultado real do processo eleitoral, levando muitos a questionarem a utilidade de algumas pesquisas para a democracia.

2.2 BOLETINS DE URNA

A Lei nº 9.100, de 1995, teve relevância significativa para as eleições municipais de 1996, introduzindo importantes mudanças no processo eleitoral. Uma das inovações mais destacadas foi a adoção das urnas eletrônicas para a votação, representando um avanço tecnológico significativo na contagem e registro dos votos

(Brasil, 1995).

Segundo Paz (2023), um dos aspectos mais relevantes desta lei foi a introdução da impressão dos boletins de urna³, proporcionando maior segurança e transparência ao processo eleitoral. A apuração das urnas com votação por cédulas por meio de sistema eletrônico representou um aprimoramento na eficiência e rapidez na contagem dos resultados. A impressão desses boletins de urna diretamente nas seções eleitorais teve o mérito de tornar os resultados mais acessíveis ao público local, promovendo a divulgação imediata e transparente dos votos. A realização de auditorias nos sistemas eleitorais e nas urnas eletrônicas evidenciou o compromisso das autoridades com a lisura do processo, proporcionando uma camada adicional de segurança e confiabilidade.

2.3 COMPARAÇÃO DE DADOS COM *MACHINE LEARNING*

A vasta quantidade de dados disponíveis tem sido amplamente explorada pela aprendizagem de máquina (*machine learning*), que destaca-se pela capacidade de processar grandes volumes de informações para realizar previsões e análises de tendências (Rocha, 2019).

Dentro do contexto eleitoral brasileiro, Valentim, Seufitelli e Silva (2023) analisaram o uso de modelos preditivos nas eleições presidenciais de 2022. O estudo utilizou dados de pesquisas eleitorais de fontes como Poder360⁴ e Polling Data⁵ e demonstrou que modelos de aprendizado supervisionado superaram as previsões dos institutos tradicionais. Métodos como *Random Forest* (Breiman, 2001), *KNeighbors Regressor* (Hastie, Tibshirani e Friedman, 2008) e *Regressão Linear* (Darlington, 1985) apresentaram alto nível de precisão, especialmente nas projeções do segundo turno, destacando o potencial dessas tecnologias para gerar previsões detalhadas sobre o comportamento eleitoral.

O algoritmo *k-nearest neighbors* (KNN) é amplamente empregado em tarefas

³Assegurando a ampla transparência no processo, o Tribunal Superior Eleitoral (TSE) sempre disponibilizou para consulta pública os boletins de urna alguns dias após a votação. Com uma inovação em 2022, mais de 472 mil boletins de urna foram disponibilizados em tempo real. Disponível em: <<https://www.tse.jus.br/comunicacao/noticias/2022/Outubro/boletins-de-urna-do-1o-turno-ja-estao-disponiveis-para-consulta-saiba-como-acessar>> Acesso em: 25 nov. 2024.

⁴ Base dos Dados. *Pesquisas Eleitorais Poder360*. Disponível em: <<https://basedosdados.org/>>. Acesso em: 1 dez. 2024.

⁵PollingData. *PollingData Pesquisas de Opinião*. Disponível em: <<https://www.pollingdata.com.br>>. Acesso em: 1 dez. 2024.

de reconhecimento de padrões, abrangendo tanto problemas de classificação quanto de regressão. Ele baseia-se em métricas de distância ou similaridade para identificar os k pontos mais próximos dentro do espaço das amostras de treinamento, caracterizando-se como uma abordagem não paramétrica adequada para problemas que requerem modelos flexíveis, especialmente quando envolvem relações não lineares (Hastie, Tibshirani e Friedman, 2008). Durante o processo de treinamento, o modelo apenas armazena os preditores e suas respectivas respostas, utilizando essas informações para realizar previsões com base nas amostras mais próximas da nova observação.

Dois fatores críticos no KNN são a definição do número de vizinhos (k) e a escolha da métrica de distância. O valor de k atua como um hiperparâmetro que controla o equilíbrio entre viés e variância: valores baixos (como 1) tornam o modelo mais suscetível a ruídos, enquanto valores altos podem levar a um viés elevado. A otimização desse parâmetro é frequentemente realizada por validação cruzada (*k-fold*) (Raschka, 2017; Hastie; Tibshirani; Friedman, 2008).

Já o modelo de regressão linear, uma abordagem clássica, representa um bom ponto de partida para a compreensão de algoritmos de Machine Learning, pois muitas das ferramentas atuais podem ser vistas como generalizações ou extensões deste modelo (James *et al.*, 2014). Como explicam Kutner, Nachtsheim e Neter (2004) e McCullagh e Nelder (1989), esses modelos são utilizados para prever respostas contínuas e oferecem uma descrição interpretável das relações entre variáveis preditoras e a resposta de interesse. Nesse contexto, o objetivo é “aprender” os pesos (ou parâmetros) da equação que melhor descreve a relação entre os preditores e a variável resposta, para então utilizá-los na predição dessa resposta em novas observações (Hastie; Tibshirani; Friedman, 2008).

O *Random Forest* expande as Árvores de Classificação e Regressão (CART) ao combinar diversas árvores construídas a partir de amostras aleatórias do conjunto de treinamento. Esse processo cria um modelo mais robusto e preciso, que utiliza a média ponderada das árvores para prever valores contínuos (Segal, 2004; Hastie; Tibshirani; Friedman, 2008). Cada árvore é treinada com um subconjunto de dados, garantindo maior estabilidade e redução do risco de superajuste.

E a avaliação da performance desses modelos preditivos é essencial para validar a eficácia das abordagens utilizadas. Entre as métricas mais aplicadas destacam-se o Erro Absoluto Médio (MAE) e a Raiz do Erro Quadrático Médio

(RMSE). Segundo Willmott e Matsuura (2005), o MAE mede a média dos erros absolutos entre os valores previstos e observados, sendo útil por fornecer uma interpretação direta da magnitude média dos erros.

Já o RMSE penaliza mais severamente os grandes erros, pois eleva ao quadrado as diferenças antes de calcular a média e a raiz. Isso o torna adequado para contextos em que desvios maiores devem ter maior impacto na métrica final (Chai; Draxler, 2014). Em estudos eleitorais, essas métricas são fundamentais para aferir a precisão dos modelos de *Machine Learning*, especialmente quando comparados aos resultados oficiais das urnas, conforme destacado por Valentim, Seufitelli e Silva (2023).

3 PROCEDIMENTOS METODOLÓGICOS

3.1 DESIGN DA PESQUISA

Este estudo utilizou uma abordagem quantitativa, com modelos de *Machine Learning* para análise dos dados. O objetivo foi explicativo, já que envolveu comparações, interpretação e plotagem dos resultados obtidos (Saunders, Lewis e Thornhill, 2016).

Em relação ao método de coleta de dados, tratou-se de um estudo de monitoramento que acompanhou dados eleitorais para analisar, de forma comparativa, as pesquisas de intenção de voto com os boletins de urna, levando em conta a distribuição geográfica dos resultados.

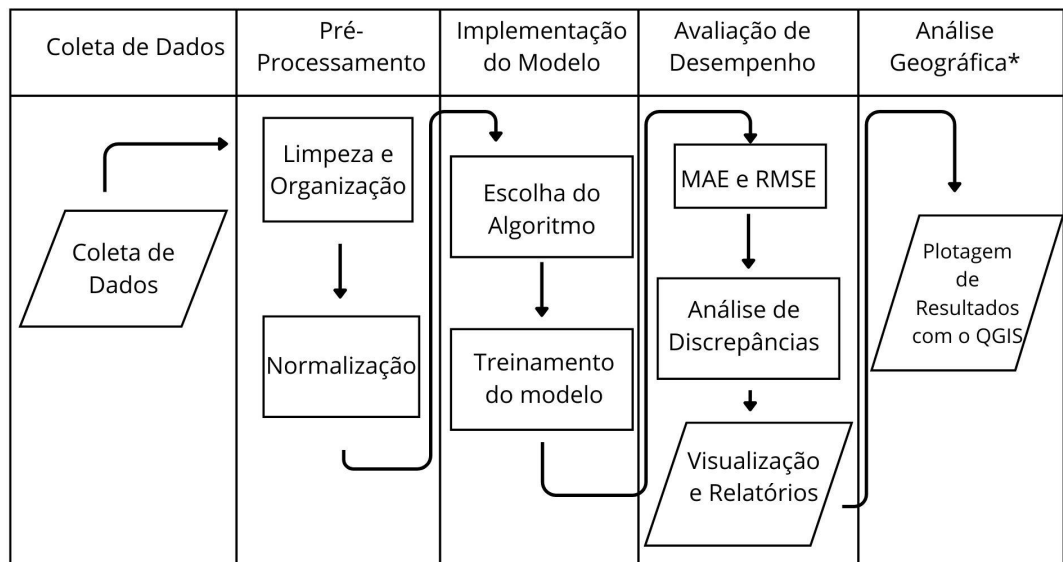
A análise focou nas eleições do ano de 2022, com destaque para os cargos de governador e presidente, do estado de Pernambuco, investigando padrões de divergência ou semelhança entre as pesquisas eleitorais e os resultados oficiais.

Este estudo foi classificado como *ex post facto*, pois observou eventos já ocorridos, sem manipulação de variáveis. Além disso, foi um estudo de natureza transversal, pois a extração dos dados das pesquisas eleitorais e dos boletins de urna foi realizada em um único período de tempo (Gil, 2002).

3.2 PERCURSO METODOLÓGICO

A metodologia adotada neste estudo foi estruturada em quatro etapas principais, além de uma etapa adicional voltada para uma futura implementação, conforme representado na Figura 1.

Figura 1 - Fluxograma do Percurso Metodológico



***Etapa voltada para uma implementação futura**

Fonte: Autoria Própria (2024)

3.3 COLETA DE DADOS

Para a execução da pesquisa, os dados foram coletados de fontes públicas, abrangendo os dados das pesquisas eleitorais do ano 2022, para os cargos de governador e presidente no estado de Pernambuco, por meio do portal Poder 360.

Além destes, foram utilizados os dados dos boletins de urnas, fornecidos pelo Tribunal Superior Eleitoral (TSE), por meio do Portal de Dados Abertos.

3.4 PRÉ-PROCESSAMENTO DE DADOS

A preparação dos dados envolveu diversas etapas essenciais para garantir a consistência e a precisão das análises. Entre essas etapas, destacam-se a limpeza, organização e transformação dos dados.

Os dados foram estruturados de forma a assegurar a compatibilidade entre as diferentes fontes, permitindo comparações válidas. Para isso, aplicaram-se técnicas de limpeza que incluíram a remoção de ruídos, outliers e duplicatas, bem como o

preenchimento de valores ausentes e a correção de inconsistências.

Além disso, foi realizada a normalização dos dados para padronizar os formatos provenientes das pesquisas eleitorais e dos boletins de urna. Essa padronização foi fundamental para permitir uma comparação consistente entre os conjuntos de dados e para reduzir a possibilidade de erros durante a implementação do algoritmo.

3.5 IMPLEMENTAÇÃO DO MODELO

Foi escolhido o modelo que melhor se adequa aos dados disponíveis e à abordagem proposta no estudo — neste caso, o modelo Random Forest — considerando tanto sua capacidade de interpretação quanto sua aplicabilidade ao contexto das eleições de 2022 para os cargos analisados.

Após essa escolha, o modelo foi treinado utilizando os dados das pesquisas eleitorais como variáveis independentes (features) e os resultados dos boletins de urna como variável dependente (target). Para garantir a robustez do modelo e reduzir o risco de overfitting, foi empregada a técnica de validação cruzada (cross-validation).

3.6 AVALIAÇÃO DE DESEMPENHO

As métricas de *Erro Médio Absoluto* (Willmott; Matsuura, 2005) e *Raiz do Erro Quadrático Médio* (Chai; Draxler, 2014) foram utilizadas para avaliar a performance do modelo na comparação entre os dados das pesquisas eleitorais e os boletins de urna.

Com base nos resultados obtidos, foi realizada uma análise detalhada das discrepâncias identificadas entre as intenções de voto captadas pelas pesquisas e os resultados reais das eleições. Para complementar essa análise, foram gerados relatórios detalhados acompanhados de visualizações informativas, destacando as comparações e evidenciando as discrepâncias que foram mais significativas.

4 ANÁLISE DE RESULTADOS

Nesta seção, são apresentados os resultados da análise comparativa entre os

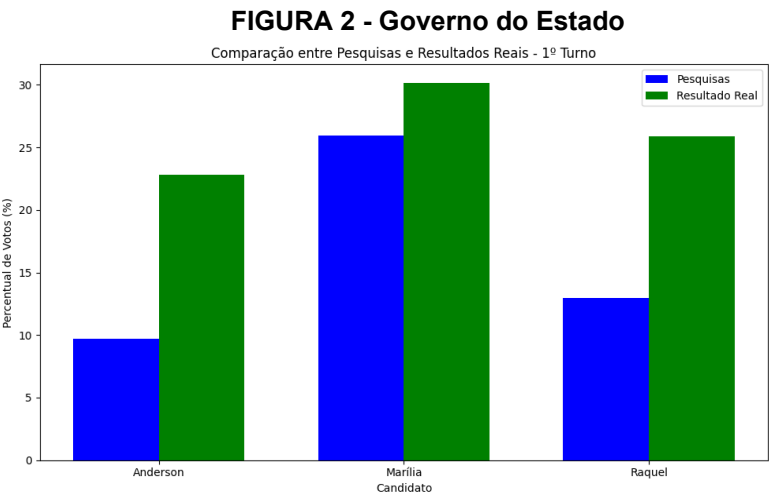
dados das pesquisas eleitorais e os boletins de urna referentes às eleições de 2022, com foco no estado de Pernambuco, nos cargos de Governador e Presidente. A metodologia adotada teve como objetivo identificar discrepâncias entre as estimativas pré-eleitorais e os resultados oficiais, além de avaliar o desempenho dos modelos de *Machine Learning* utilizados na predição dos resultados.

Para viabilizar a análise, foram realizadas as etapas de pré-processamento, treinamento e validação dos algoritmos, conforme descrito na seção metodológica. A performance dos modelos foi mensurada por meio de métricas quantitativas, com destaque para o *Erro Quadrático Médio (MSE)* e a *Raiz do Erro Quadrático Médio (RMSE)*, indicadores que permitiram avaliar a precisão das previsões em relação aos dados reais.

Os resultados estão organizados por gráficos comparativos, tabelas estatísticas e análises interpretativas, de forma a evidenciar os principais achados e padrões observados ao longo do estudo.

4.1 GRÁFICOS COMPARATIVOS POR CARGO

A Figura 2 apresentada a seguir compara as projeções de intenção de voto obtidas nas pesquisas eleitorais com os resultados oficiais do primeiro turno para o cargo de Governador, destacando o desempenho dos candidatos Anderson Ferreira, Marília Arraes e Raquel Lyra.



Fonte : Elaborado pela autora (2025)

A Figura 2 apresenta o gráfico comparativo entre os percentuais de votos indicados pelas pesquisas eleitorais e os resultados reais obtidos no primeiro turno

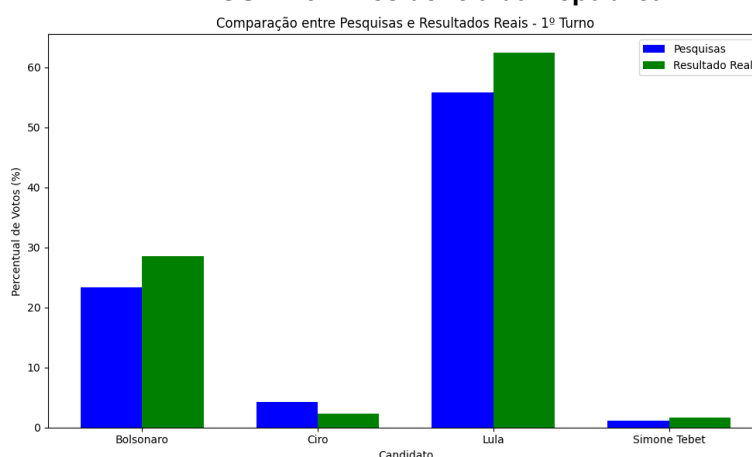
para o cargo de Governador em Pernambuco. Observa-se uma variação significativa entre as projeções das pesquisas e o desempenho efetivo dos candidatos nas urnas.

Para o candidato Anderson, as pesquisas indicaram aproximadamente 10% dos votos, enquanto o resultado real foi de cerca de 22.5%. Essa diferença representa uma subestimação considerável do desempenho do candidato pelas pesquisas. A candidata Marília, por sua vez, teve suas intenções de voto projetadas em aproximadamente 26% pelas pesquisas, com um resultado real próximo de 30%, indicando uma diferença menor, mas ainda presente. Já a candidata Raquel, que obteve cerca de 26% dos votos reais, foi subestimada pelas pesquisas, que a apontavam com aproximadamente 13% dos votos.

As discrepâncias encontradas revelam limitações das pesquisas em captar com precisão a intenção de voto. Os dados indicam que os candidatos Anderson e Raquel tiveram desempenhos reais significativamente superiores às projeções, evidenciando desafios na acurácia das estimativas, que serão explorados em maior profundidade nas próximas subseções.

Prosseguindo com a análise, a figura 3 a seguir apresenta o gráfico que ilustra a comparação entre os dados das pesquisas de intenção de voto e os resultados apurados no primeiro turno para o cargo de Presidência da República, com destaque para o desempenho dos quatro principais candidatos: Luiz Inácio Lula da Silva, Jair Bolsonaro, Ciro Gomes e Simone Tebet.

FIGURA 3 - Presidência da República



Fonte: Elaborado pela autora (2025)

Neste gráfico apresentado é ilustrada a comparação entre os percentuais de votos indicados pelas pesquisas eleitorais e os resultados reais obtidos no primeiro turno para o cargo de Presidente em Pernambuco. Assim como na análise para

governador, observa-se que as pesquisas apresentaram variações em relação aos resultados finais.

Para o candidato Jair Bolsonaro, observou-se uma subestimação por parte das pesquisas, que indicavam cerca de 23,5% das intenções de voto, enquanto o resultado real alcançou aproximadamente 28,5%. No caso de Ciro Gomes, houve uma superestimação: as pesquisas projetavam 4,5%, mas o candidato obteve apenas 2% dos votos válidos. Luiz Inácio Lula da Silva, candidato mais votado no estado, teve sua intenção de voto estimada em 56%, sendo que o resultado oficial foi de aproximadamente 62%, caracterizando também uma subestimação, embora mantendo a liderança apontada pelas pesquisas. Já a candidata Simone Tebet foi projetada com 0,8% das intenções de voto, mas obteve um desempenho real em torno de 1,5%, indicando mais uma subestimação. Esses resultados evidenciam variações importantes entre as projeções pré-eleitorais e os resultados finais.

Esses resultados reforçam a complexidade da previsão eleitoral, mesmo para cargos de maior visibilidade como o de Presidente. As discrepâncias observadas, tanto de subestimação quanto de superestimação, indicam a necessidade de uma análise aprofundada dos fatores que podem influenciar a precisão das pesquisas, como a metodologia utilizada, o perfil dos entrevistados e o período de realização da pesquisa em relação à data da eleição.

4.2 ANÁLISE DO ERRO DAS PESQUISAS E DOS INSTITUTOS POR CARGO

A tabela 1 apresenta, para cada candidato, os valores estimados pelas pesquisas eleitorais e os resultados oficiais apurados nas urnas para governador, acompanhados dos respectivos cálculos do erro absoluto e do erro relativo, com o objetivo de quantificar a acurácia das projeções em relação aos dados reais.

TABELA 1 - Análise do Erro nas Estimativas de Voto por Candidato - Governo do Estado

candidato	percentual_pesquisa	percentual_real	erro_absoluto	erro_relativo
Anderson	9.707621	22.812718	13.105097	57.446452
Marília	25.971652	30.127153	4.155502	13.793210
Raquel	12.945718	25.870814	12.925096	49.960143

Fonte: Elaborado pela Autora (2025)

A tabela 1 mostra a comparação entre as projeções das pesquisas eleitorais e os resultados oficiais por candidato ao Governo de Pernambuco (Anderson Ferreira, Marília Arraes e Raquel Lyra), evidenciando os respectivos erros absolutos e relativos. Observa-se que todos os candidatos obtiveram votações reais superiores às estimativas das pesquisas, indicando uma tendência generalizada de subestimação.

O candidato Anderson Ferreira foi o mais subestimado, com projeção de 9,7% frente a um resultado real de 22,8%, o que corresponde a um erro absoluto de 13,1 pontos percentuais e um erro relativo de 57,4%. Raquel Lyra também foi significativamente subestimada, com um erro absoluto de 12,9% e erro relativo de quase 50%. Marília Arraes apresentou a menor discrepância entre os dados projetados e reais, com erro absoluto de 4,2% e relativo de 13,8%.

Essas discrepâncias indicam possíveis limitações metodológicas nas pesquisas, como amostragem não representativa, mudanças de intenção de voto em períodos próximos à eleição e o impacto do chamado “voto útil” — a migração de votos para candidatos mais competitivos na reta final da campanha. Os resultados reforçam a necessidade de aprimoramento nos métodos de coleta e análise dos dados eleitorais, a fim de aumentar a acurácia das projeções.

Complementando a análise, a Tabela 2 apresenta as métricas de desempenho agregadas por candidato, com base em indicadores como o Erro Médio Absoluto (MAE) e a Raiz do Erro Quadrático Médio (RMSE), que permitem avaliar de forma quantitativa a acurácia das estimativas fornecidas pelas pesquisas.

TABELA 2 - MAE E RMSE - Governo do Estado

Erro Médio Absoluto (MAE)	Raiz do Erro Quadrático Médio (RMSE)
10.06%	10.89%

Fonte: Elaborado pela Autora (2025)

Os indicadores mostram que as pesquisas apresentaram uma margem significativa de imprecisão, com o MAE de 10.06% E RMSE de 10.89%, que confirmam a tendência de subestimação observada nos dados individuais dos candidatos, indicando que os erros não se limitaram a casos isolados, mas foram sistêmicos. O fato da Raiz do Erro Quadrático Médio (RMSE) ser ligeiramente superior ao MAE sugere a presença de outliers com erros particularmente grandes

(como os casos de Anderson e Raquel), O que reforça a necessidade de revisão metodológica nas pesquisas, especialmente na captação de eleitores indecisos ou de última hora.

Na sequência, a Tabela 3 apresenta, para cada candidato à Presidência da República, os valores estimados pelas pesquisas eleitorais e os resultados oficiais apurados nas urnas, acompanhados dos respectivos cálculos de erro absoluto e erro relativo, com o objetivo de mensurar a acurácia das projeções em relação aos dados reais.

TABELA 3 - Análise do Erro nas Estimativas de Voto por Candidato – Presidência da República

candidato	percentual_pesquisa	percentual_real	erro_absoluto	erro_relativo
Bolsonaro	23.362619	27.601873	5.239254	18.317872
Ciro	4.240320	2.280082	1.960238	85.972245
Lula	55.818508	62.402540	6.584031	10.550903
Simone	1.091730	1.693555	0.601824	35.536168

Fonte: Elaborado pela Autora (2025)

A tabela 3 evidencia que os candidatos Lula e Bolsonaro foram moderadamente subestimados pelas pesquisas, com erros absolutos de 6,58% e 5,24%, respectivamente. Lula apresentou o menor erro relativo (10,55%), indicando maior acurácia na estimativa de intenções de voto para o candidato líder. O caso mais crítico foi o de **Ciro Gomes**: apesar do erro absoluto moderado (1,96%), seu erro relativo alcançou 85,97%, sugerindo que as pesquisas superestimaram significativamente sua performance, possivelmente por não captarem adequadamente sua rejeição ou uma migração tardia de votos. Já **Simone Tebet** obteve uma votação 57% superior à projetada, o que pode refletir uma mobilização tardia de um segmento específico do eleitorado não detectado nas pesquisas.

Dando complemento a análise, a Tabela 4 apresenta as métricas de desempenho agregadas por candidato, com base em indicadores como o *Erro Médio Absoluto (MAE)* e a *Raiz do Erro Quadrático Médio (RMSE)*, que permitem avaliar de forma quantitativa a acurácia das estimativas fornecidas pelas pesquisas.

TABELA 4 - MAE e RMSE

Erro Médio Absoluto (MAE)	Raiz do Erro Quadrático Médio (RMSE)
3.60%	4.33%

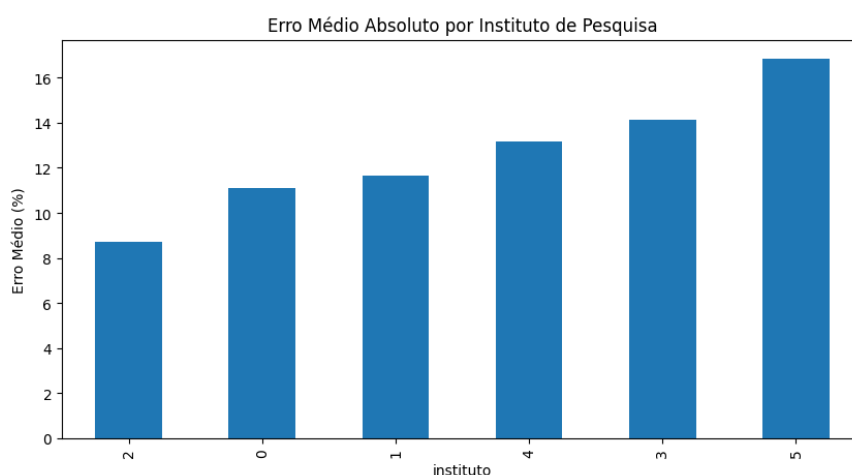
Fonte: Elaborado pela autora (2025)

O *Erro Médio Absoluto (MAE)* indica que, em média, as pesquisas divergiram dos resultados reais em aproximadamente 3,6 pontos percentuais por candidato, um valor considerado aceitável em contextos eleitorais, embora com ressalvas. Já a *Raiz do Erro Quadrático Médio (RMSE)*, de 4,33%, revela a influência de outliers, como os casos de *Ciro Gomes* e *Simone Tebet*, cujos desvios significativos elevaram a média dos erros.

Pode-se concluir que, embora as pesquisas tenham captado adequadamente a tendência geral do pleito, elas apresentaram maior imprecisão em cenários de baixa intenção de voto, como no caso de *Ciro* e *Simone*. A aparente precisão global representada pelo MAE pode ocultar erros relevantes em candidatos com baixa adesão, reforçando a necessidade de aprimoramentos metodológicos.

Seguindo com a análise, a Figura 4 apresenta o gráfico da comparação do *Erro Médio Absoluto (MAE)* entre os diferentes institutos de pesquisa, identificados por números de 0 a 5, considerando as estimativas para o cargo de Governador do Estado.

FIGURA 4 - Gráfico do MAE (Erro Médio Absoluto) por Instituto de Pesquisa - Governo do Estado



Fonte: Elaborado pela Autora (2025)

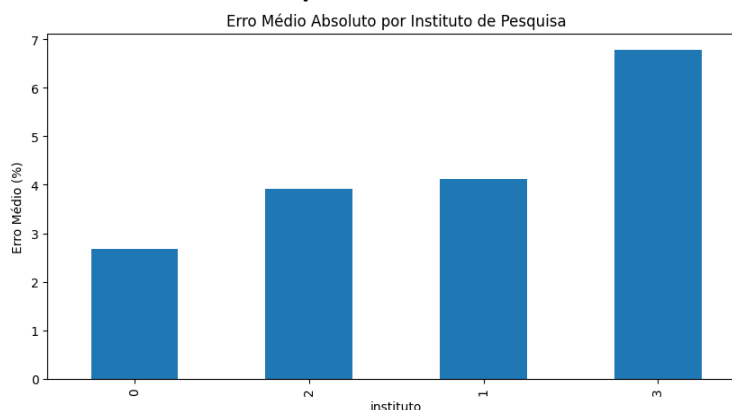
Neste gráfico podemos observar que o instituto 2 foi o mais preciso, com erro médio de aproximadamente 8%, seguido pelos institutos 0 e 1, que também

apresentaram desempenho razoável, com erros em torno de 10% a 12%. Por outro lado, os institutos 4,3 e 5 apresentaram os maiores erros, sendo o instituto 5 com o pior desempenho, com erro acima de 16%.

Essa variação mostra as diferenças significativas na qualidade das estimativas fornecidas, que pode estar relacionado a metodologias distintas, amostragens inadequadas ou interpretações incorretas dos dados. Esse tipo de análise é importante para avaliar a confiabilidade dos institutos e serve como base para decisões futuras quanto à credibilidade das pesquisas.

Seguindo adiante com os resultados, a Figura 5 apresenta o gráfico da comparação do *Erro Médio Absoluto (MAE)* entre os diferentes institutos de pesquisa, identificados de 0 a 3, considerando as estimativas para o cargo de Presidência da República.

FIGURA 5 - Gráfico do MAE (Erro Médio Absoluto) por Instituto de Pesquisa - Presidência da República



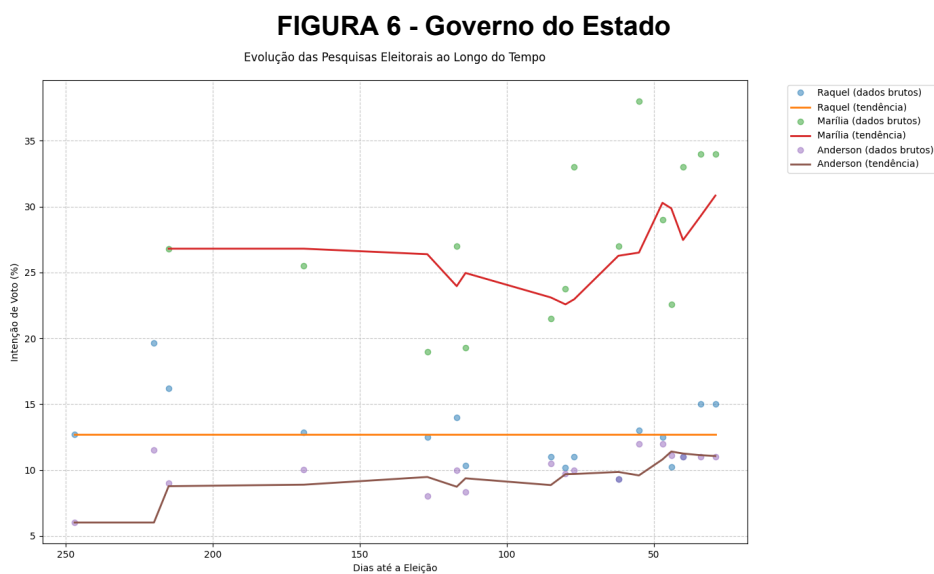
Fonte: Elaborado pela Autora (2025)

O gráfico mostrou que o instituto 0 foi o mais preciso com erro médio de aproximadamente 2,5%, seguido pelos institutos 2 e 1, com erros em torno de 4%, o que indica que esses institutos tiveram boa proximidade entre as previsões realizadas e os resultados reais. Já o instituto 3 foi o que apresentou o maior erro médio, próximo de 6,5%, indicando uma menor precisão nas estimativas.

Essa diferença mostra que, mesmo dentro de um mesmo contexto, a qualidade das previsões variou significativamente, o que destaca a importância de avaliar a confiabilidade de cada um com base nos seus acertos.

4.3 EVOLUÇÃO DAS PESQUISAS ELEITORAIS AO LONGO DO TEMPO POR CARGO

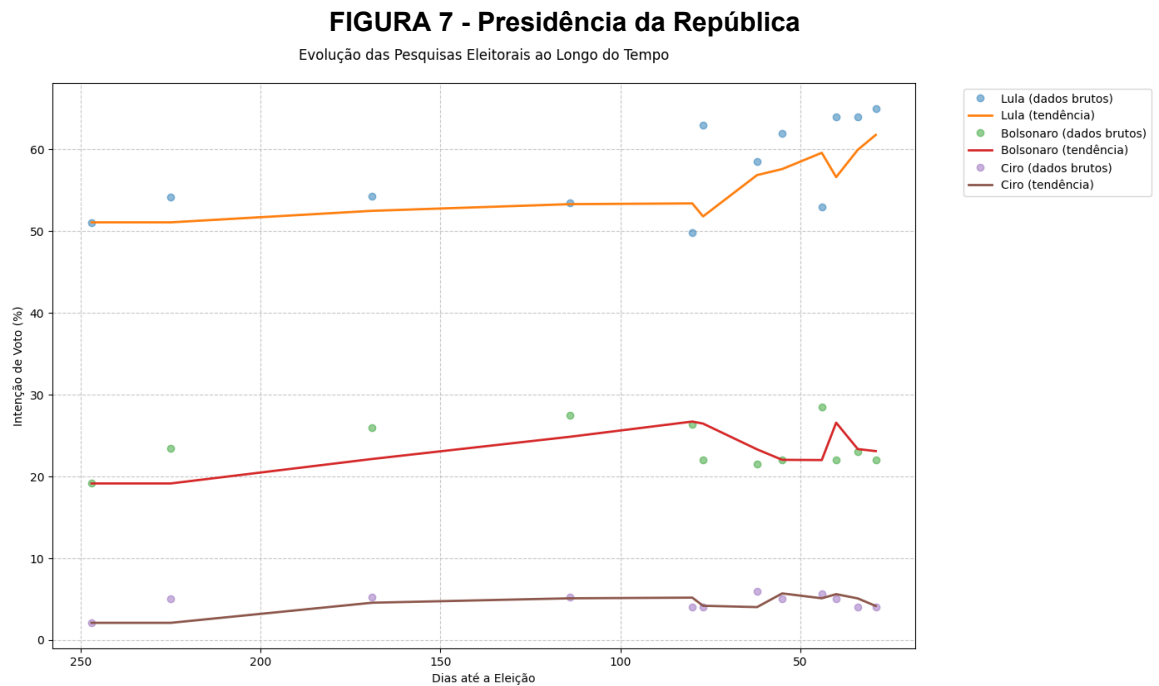
Figura 6 a seguir apresenta o gráfico com a evolução das intenções de voto ao longo do tempo para os principais candidatos ao cargo de Governador do Estado. A análise contempla os dados brutos das pesquisas eleitorais, bem como as tendências observadas ao longo do período.



Fonte: Elaborado pela Autora (2025)

O gráfico apresenta a evolução da intenção de voto dos candidatos Raquel, Marília e Anderson ao longo do tempo. Observa-se que Marília, apesar de algumas oscilações, teve uma trajetória de crescimento contínuo, especialmente nos períodos mais próximos à data da eleição. Raquel manteve-se estável durante toda a campanha, com variações mínimas nas intenções de voto, indicando um eleitorado consolidado. Já Anderson apresentou um crescimento lento e constante, partindo de um patamar mais baixo. A análise sugere que Marília ganhou força ao longo da campanha, Raquel sustentou sua base de apoio e Anderson teve uma ascensão discreta.

Dando prosseguimento à análise, a figura 7 apresenta o gráfico com a evolução das pesquisas de intenções de voto ao longo do tempo para os principais candidatos ao cargo de Presidente. A análise contempla os dados brutos das pesquisas eleitorais, bem como as tendências observadas ao longo do período.



Fonte: Elaborado pela Autora (2025)

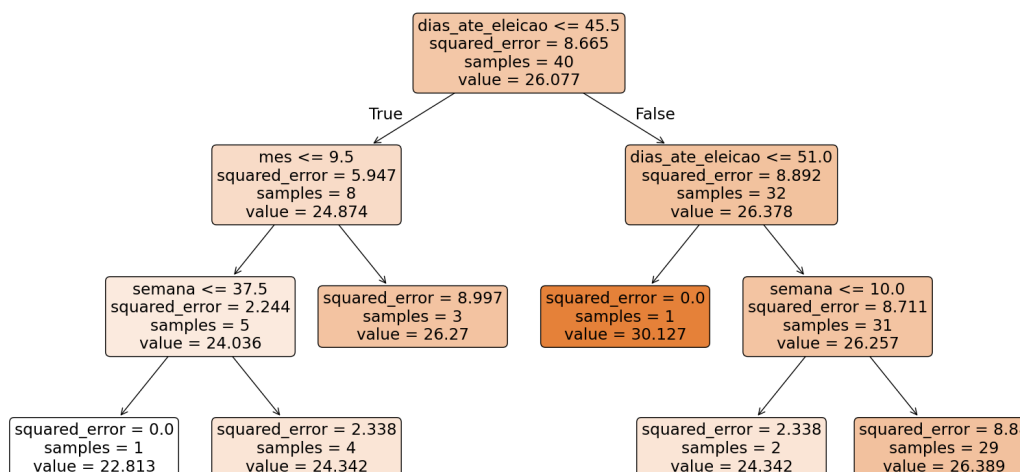
O gráfico mostra a evolução da intenção de voto para os candidatos Lula, Bolsonaro e Ciro ao longo do tempo. Lula manteve uma liderança consistente, com tendência de crescimento mais acentuado nos últimos 50 dias antes da eleição. Bolsonaro apresentou uma trajetória de oscilação, com crescimento moderado até certo ponto, seguido de uma leve queda na reta final. Já Ciro manteve-se estável com uma leve tendência de alta no início, mas sem grandes avanços, permanecendo com baixo percentual de intenção de votos. A análise revela um cenário de consolidação para Lula, estabilidade com flutuações para Bolsonaro e estagnação para Ciro.

4.4 IMPLEMENTAÇÃO DA ÁRVORE DE DECISÃO POR CARGO E AVALIAÇÃO (MAE E RMSE)

A Figura 8 a seguir mostra a Árvore de Decisão para prever Resultados Eleitorais para o cargo de Governador. O modelo representa de forma visual o processo de tomada de decisão do algoritmo, mostrando como os dados são divididos em diferentes ramos de acordo com as variáveis relevantes.

FIGURA 8 - Governo do Estado

Árvore de Decisão para Prever Resultados Eleitorais



Fonte: Elaborado pela Autora (2025)

A árvore de decisão apresentada tem como objetivo prever resultados eleitorais com base no tempo restante até a eleição. A variável mais importante para a divisão inicial é o número de dias até a eleição, com um ponto de corte em 45,5 dias. Quando restam poucos dias ($\leq 45,5$), outros fatores como o mês e a semana se tornam relevantes, resultando em previsões de intenção de voto próximas a 24-26%. Quando faltam mais dias ($> 45,5$), há subdivisões adicionais com base em faixas de dias e semanas, revelando que quanto mais distante da eleição, maior a variabilidade nas previsões (com valores entre 24,3% e 30,1%). A árvore demonstra como o tempo influencia diretamente na estabilidade e precisão das previsões eleitorais.

Seguindo com a análise, a tabela 5 apresenta as métricas de desempenho em relação à Árvore de Decisão dentro do contexto do Governo do Estado.

TABELA 5 - MAE E RMSE DA ÁRVORE DE DECISÃO - Governo do Estado

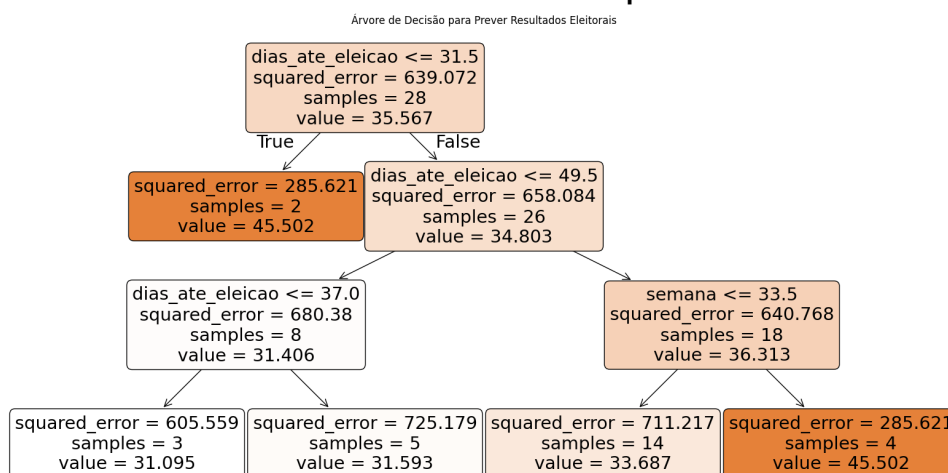
Erro Médio Absoluto (MAE)	Raiz do Erro Quadrático Médio (RMSE)
4.61%	5.00%

Fonte: Elaborado pela Autora (2025)

O fato de o *RMSE* estar próximo ao *MAE* mostra que não houve muitos erros extremos (*outliers*), o que é positivo. Esses valores indicam que o modelo está capturando bem os padrões nos dados e realizando previsões com boa precisão, especialmente considerando a complexidade de fenômenos eleitorais, que costumam ser influenciados por diversas variáveis externas e imprevisíveis.

Seguindo adiante com a análise, a Figura 9 a seguir mostra a Árvore de Decisão para prever Resultados Eleitorais para o cargo de Presidente.

FIGURA 9 - Presidência da República



Fonte: Elaborado pela Autora (2025)

A árvore de decisão apresentada tem como objetivo prever resultados eleitorais com base principalmente na variável "dias até a eleição". Ela inicia dividindo os dados em dois grandes grupos: eleições com 31,5 dias ou menos de antecedência e aquelas com mais de 31,5 dias. Observa-se que, quanto mais próxima a eleição, maior tende a ser a intenção de voto (com valor médio de 45,502), embora essa previsão ainda apresenta certa variabilidade (erro quadrático). Para os casos em que faltam mais de 31,5 dias, a árvore realiza novas divisões, considerando novamente os dias até a eleição e também a semana do levantamento. Mesmo com essas divisões adicionais, os erros quadráticos permanecem elevados, o que indica que há variações consideráveis nos dados e que apenas o tempo até a eleição pode não ser suficiente para explicar completamente os resultados.

Complementando a análise, a tabela 6 apresenta as métricas de desempenho em relação à Árvore de Decisão dentro do contexto da Presidência.

TABELA 6 - MAE E RMSE DA ÁRVORE DE DECISÃO - Presidência

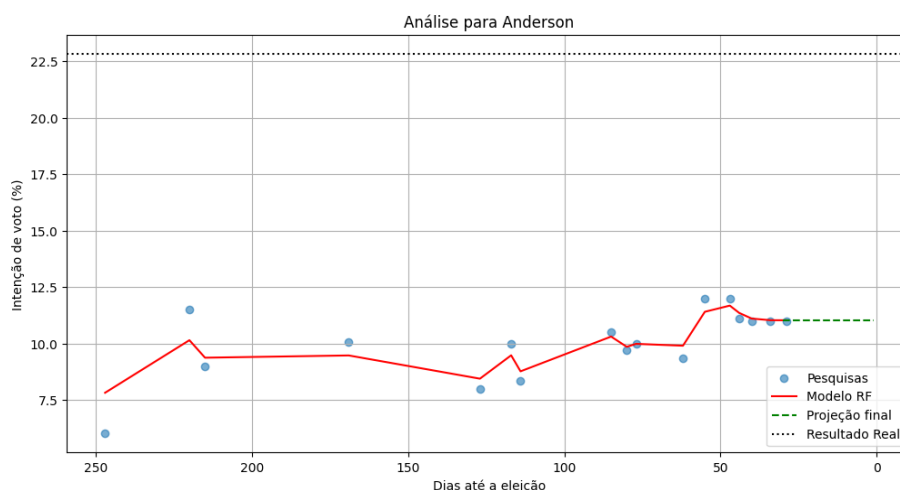
Erro Médio Absoluto (MAE)	Raiz do Erro Quadrático Médio (RMSE)
22.42%	28.89%

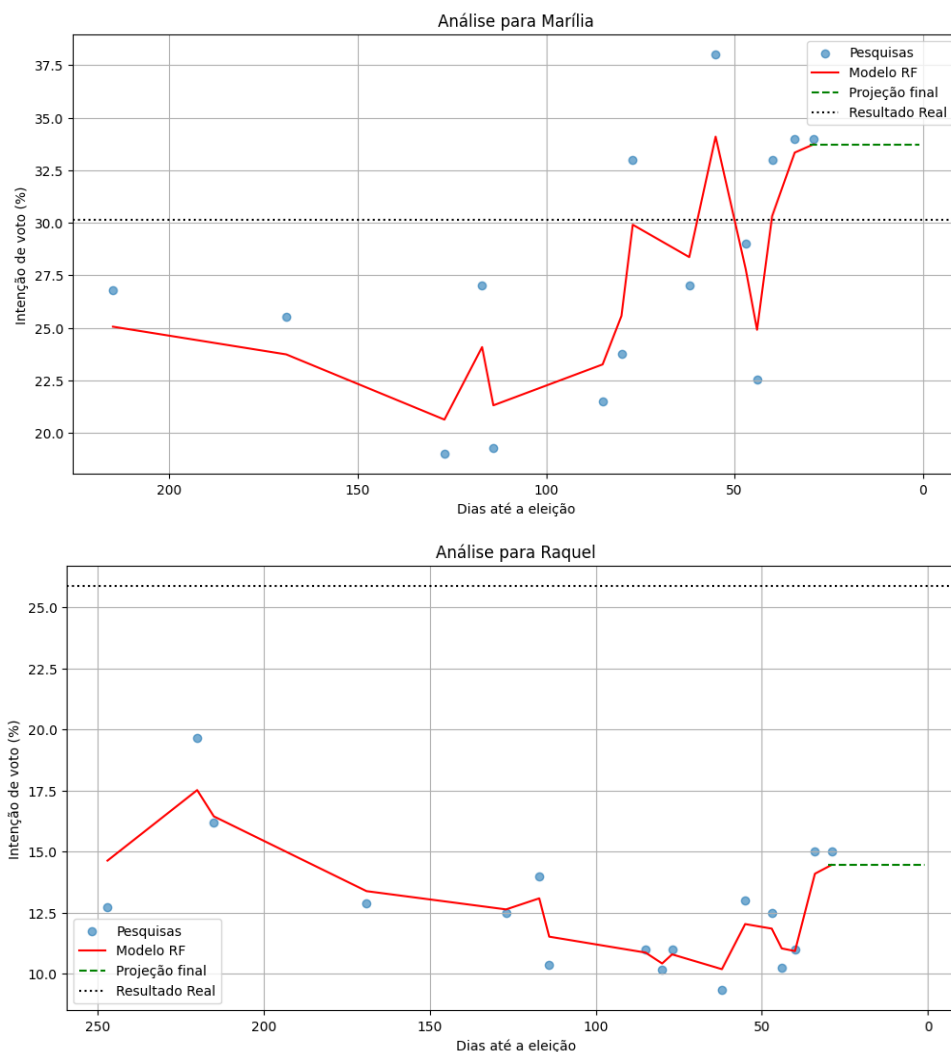
Fonte: Elaborada pela Autora (2025)

A segunda árvore de decisão teve desempenho fraco, com *MAE* de 22,42% e *RMSE* de 28,89%, indicando erros altos nas previsões. Isso sugere que o modelo não conseguiu generalizar bem os dados e pode estar cometendo erros frequentes e graves, possivelmente por sobreajuste ou escolhas inadequadas de variáveis ou estrutura da árvore.

4.5 ANÁLISE SEPARADA POR CANDIDATO DE CADA CARGO

A Figura 11 a seguir apresenta três gráficos que analisam, separadamente por candidato, a evolução das intenções de voto ao longo do tempo durante as eleições de 2022 para o cargo de Governador. Em cada gráfico, os pontos azuis representam os dados coletados nas pesquisas eleitorais, enquanto a linha vermelha indica a previsão contínua gerada pelo modelo *Random Forest*. A linha verde tracejada mostra a projeção final do modelo nos dias que antecederam a eleição. Por fim, a linha preta pontilhada marca o resultado real obtido nas urnas por cada candidato, permitindo uma comparação visual entre os dados previstos e os resultados efetivos.

FIGURA 11 - Governo do Estado



Fonte: Elaborado pela autora (2025)

Os gráficos analisados evidenciam que o modelo *Random Forest* apresentou desempenhos distintos na previsão da intenção de voto para os diferentes candidatos.

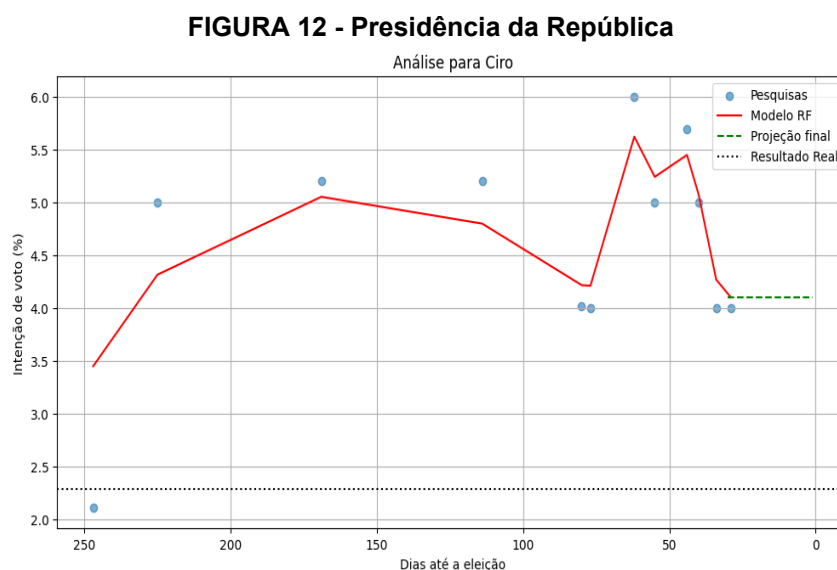
Para o candidato Anderson, a projeção gerada pelo modelo manteve-se relativamente estável ao longo do tempo, porém subestimou o resultado final, que se revelou significativamente superior à estimativa. Apesar das pesquisas indicarem oscilações crescentes, o modelo não conseguiu captar de forma satisfatória essa tendência de alta. Dessa forma, conclui-se que o modelo adotou uma postura conservadora, resultando na subestimação do desempenho final do candidato.

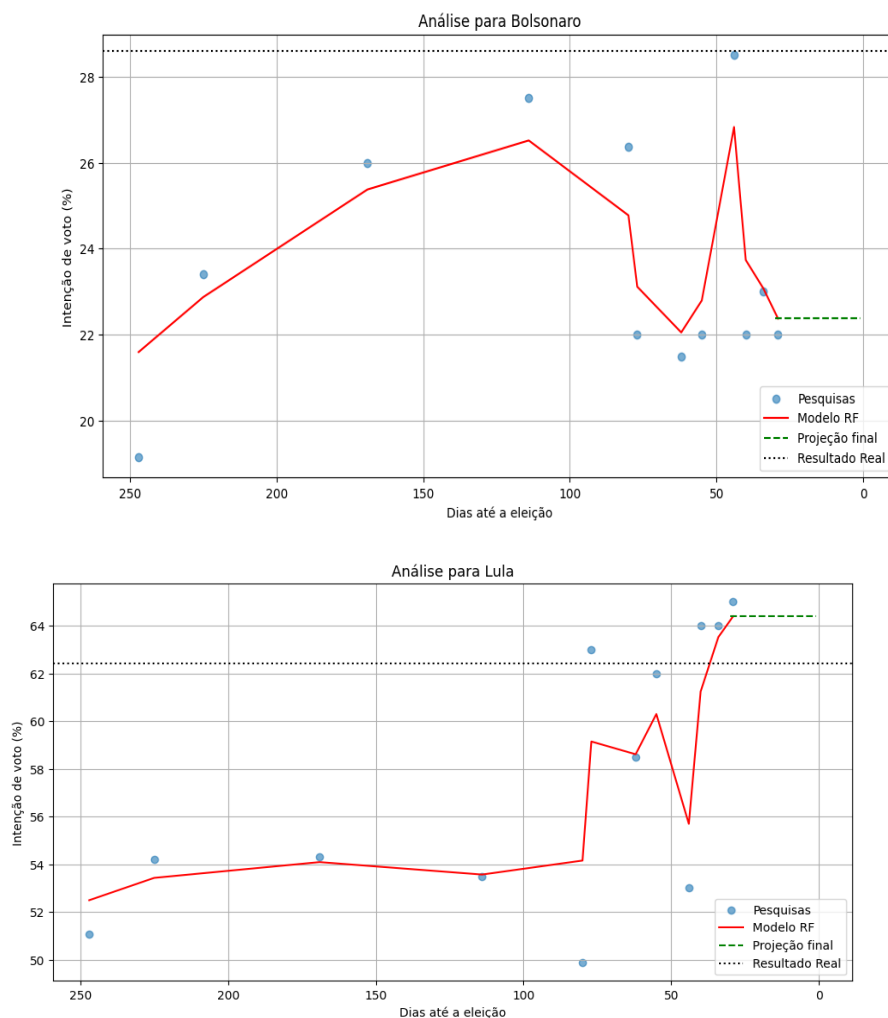
No caso da candidata Marília, a curva gerada pelo modelo oscilou consideravelmente, em tentativa de acompanhar as variações observadas nas pesquisas. Contudo, a projeção final superou o resultado real, indicando uma

superestimação. As fortes flutuações nas intenções de voto, especialmente nos últimos 60 dias, podem ter comprometido a estabilidade do modelo. Assim, conclui-se que o modelo apresentou alta variabilidade e uma tendência de superestimar o desempenho da candidata.

Em relação à candidata Raquel, o modelo mostrou-se moderadamente coerente com os dados das pesquisas, ainda que tenha subestimado o resultado final. A projeção indicava uma tendência de estabilidade ao final da série temporal, porém não acompanhou a elevação real registrada nos últimos dias da campanha. Concluindo que o modelo teve um desempenho relativamente equilibrado, mas falhou ao estimar com precisão o resultado final, subestimando-o.

Dando prosseguimento à análise, a Figura 12 a seguir apresenta três gráficos que analisam, separadamente por candidato, a evolução das intenções de voto ao longo do tempo durante as eleições de 2022 para o cargo de presidente. Em cada gráfico, os pontos azuis representam os dados coletados nas pesquisas eleitorais, enquanto a linha vermelha indica a previsão contínua gerada pelo modelo *Random Forest*. A linha verde tracejada mostra a projeção final do modelo nos dias que antecederam a eleição. Por fim, a linha preta pontilhada marca o resultado real obtido nas urnas por cada candidato, permitindo uma comparação visual entre os dados previstos e os resultados efetivos.





Fonte: Elaborado pela Autora (2025)

A análise dos gráficos mostra que o modelo de *Random Forest* teve precisões distintas para cada candidato.

Para o candidato Ciro Gomes, observou-se uma tendência oscilante nas pesquisas de intenção de voto, seguida por uma leve queda nas semanas finais da campanha. No entanto, o modelo *Random Forest* não foi capaz de capturar essa redução no apoio, o que resultou em uma projeção final de 4,1%, valor significativamente superior ao resultado real de 2,2%. Esse desvio evidencia um erro expressivo do modelo, possivelmente relacionado à dificuldade em refletir a migração de votos motivada pelo voto útil, fenômeno que ganhou força nos momentos finais do processo eleitoral.

Para Bolsonaro, as pesquisas apresentaram uma tendência relativamente estável, variando entre 22% e 26%, com picos de crescimento próximos à data da

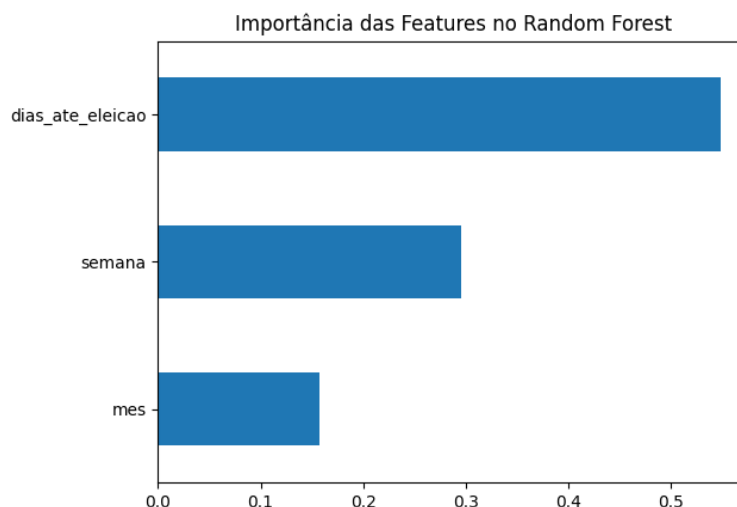
eleição. O modelo *Random Forest* seguiu essa tendência geral, mas não conseguiu captar de forma satisfatória o aumento mais acentuado observado nos últimos dias. A projeção final do modelo foi de aproximadamente 22,5%, valor inferior ao resultado real, que foi de 28,5%. Esse desvio indica um erro moderado na estimativa, possivelmente atribuído à limitação do modelo em detectar mudanças abruptas de opinião ou à ocorrência de uma reconfiguração no comportamento do eleitorado nas semanas finais da campanha.

No caso do candidato Lula, o modelo apresentou um bom desempenho. A tendência observada nas pesquisas manteve-se forte e crescente ao longo de toda a campanha, e o modelo foi capaz de capturar com relativa precisão essa curva ascendente. A projeção final estimada foi de aproximadamente 64%, valor bastante próximo ao resultado real, de 62,5%, ainda que com leve superestimação. Diante disso, conclui-se que o modelo cometeu erros pequenos, evidenciando maior eficácia na previsão para esse candidato — possivelmente devido à menor volatilidade nas intenções de voto e à baixa incidência do chamado voto útil em sua base eleitoral.

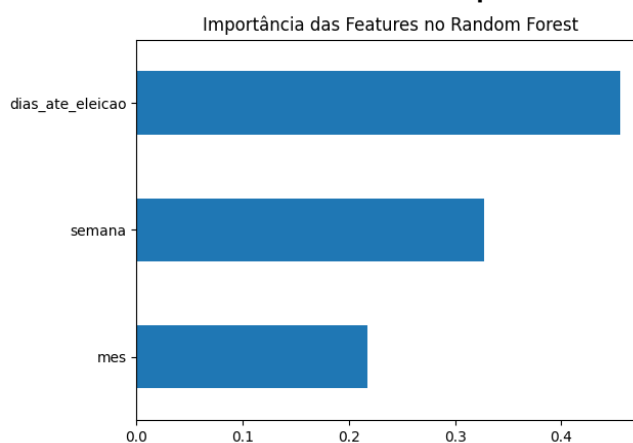
4.6 IMPORTÂNCIA DAS FEATURES PARA O MODELO RANDOM FOREST

A análise da importância das variáveis (features) no modelo Random Forest é essencial para compreender quais fatores exercem maior influência nas previsões realizadas. Esse tipo de avaliação contribui significativamente para a interpretação do modelo, pois evidencia os atributos que mais impactaram o processo de tomada de decisão. Além disso, o conhecimento sobre a relevância das variáveis permite uma redução da dimensionalidade dos dados, viabilizando a exclusão de atributos com baixa contribuição para o desempenho do modelo. Essa prática pode, inclusive, reduzir o risco de sobreajuste (*overfitting*), tornando o modelo mais robusto, eficiente e com maior capacidade de generalização.

Essa compreensão é especialmente relevante em estudos eleitorais, nos quais é fundamental identificar os fatores que mais influenciaram as projeções de voto e compará-los com os resultados reais. Com isso, as Figuras 11 e 12 apresentam os gráficos que mostram a importância das features no Random Forest dentro do contexto de cada cargo.

FIGURA 13 - Governo do Estado

Fonte: Elaborado pela Autora (2025)

FIGURA 14 - Presidência da República

Fonte: Elaborado pela Autora (2025)

O modelo utilizou três variáveis principais para prever as intenções de voto: *dias_ate_eleicao*, *semana* e *mes*. A análise de importância das features revelou tanto no gráfico para governo do estado quanto para presidente que a variável *dias_ate_eleicao* foi a mais relevante, indicando que a proximidade da data da eleição tem um impacto significativo na precisão das pesquisas, já as variáveis *semana* e *mes* apresentaram importância menor, mesmo havendo uma pequena diferença entre os gráficos, o que sugere que períodos mais amplos têm influência limitada na previsão dos votos, o que mostra que as pesquisas eleitorais são mais sensíveis a eventos de curto prazo, como debates e notícias recentes do que a tendências de longo prazo. Prosseguindo com a análise, a tabela 7 abaixo mostra o *Erro Médio Absoluto (MAE)* e a *Raiz do Erro Quadrático Médio (RMSE)* dentro do

contexto da análise de dados para o Governo do Estado.

TABELA 7 - MAE E RMSE - Random Forest : Governo do Estado

Erro Médio Absoluto (MAE)	Raiz do Erro Quadrático Médio (RMSE)
4.04	4.49

Fonte: Elaborado pela Autora (2025)

Esses resultados de *MAE* e *RMSE* indicam que o modelo apresentou um desempenho razoável de forma geral. No entanto, a diferença entre os dois indicadores sugere a ocorrência de variações mais acentuadas em determinados momentos, evidenciando a presença de erros pontuais mais expressivos nas previsões. Seguindo com a análise a tabela 8 abaixo mostra o *Erro Médio Absoluto (MAE)* e a *Raiz do Erro Quadrático Médio (RMSE)* dentro do contexto da análise de dados para a Presidência da República.

TABELA 8 MAE E RMSE - Random Forest : Presidência

Erro Médio Absoluto (MAE)	Raiz do Erro Quadrático Médio (RMSE)
22.53	29.18

Fonte: Elaborado pela autora (2025)

Os resultados de desempenho do modelo Random Forest, com um *Erro Médio Absoluto (MAE)* de 22,53 e uma *Raiz do Erro Quadrático Médio (RMSE)* de 29,18, indicam uma considerável dificuldade na previsão precisa dos resultados eleitorais para a Presidência da República. A diferença significativa entre os dois indicadores reforça a presença de erros extremos nas estimativas, possivelmente decorrentes da alta volatilidade do eleitorado, da influência de fatores externos de última hora ou da complexidade inerente ao cenário político analisado.

5 CONSIDERAÇÕES FINAIS

Este Trabalho de Conclusão de Curso teve como objetivo realizar uma análise comparativa entre os dados de pesquisas de intenção de voto e os resultados oficiais das eleições de 2022 para os cargos de Governador de Pernambuco e Presidente da República. Para isso, foi adotada uma metodologia baseada no pré-processamento dos dados, aplicação do modelo de Machine Learning Random

Forest e avaliação de desempenho por meio das métricas *MAE (Erro Absoluto Médio)* e *RMSE (Raiz do Erro Quadrático Médio)*.

Os resultados obtidos evidenciaram divergências significativas entre os dados das pesquisas eleitorais e os resultados das urnas. No pleito para o governo de Pernambuco, observou-se uma tendência de subestimação generalizada dos candidatos. Já na disputa presidencial, identificaram-se tanto subestimação quanto superestimação, refletindo a complexidade e as dinâmicas distintas de cada eleição. A análise da importância das variáveis (*features*) indicou que o fator *dias_ate_eleicao* teve maior relevância na performance do modelo, sugerindo que a proximidade do pleito é determinante para a consolidação das intenções de voto.

A principal contribuição deste estudo reside na demonstração prática de que a aplicação de técnicas de *Machine Learning*, como o modelo *Random Forest*, pode aprimorar de forma significativa a análise e a previsão de resultados eleitorais. A comparação entre pesquisas e resultados oficiais revelou que os algoritmos preditivos não apenas identificam discrepâncias relevantes, mas também detectam padrões ocultos que dificilmente seriam percebidos por métodos tradicionais.

Reconhecem-se, no entanto, as limitações desta pesquisa. A análise esteve condicionada à disponibilidade e à qualidade dos dados coletados, tanto das pesquisas quanto dos boletins de urna, o que pode impactar diretamente na precisão dos modelos desenvolvidos. Diferenças metodológicas entre os institutos de pesquisa, amostragens pouco representativas e potenciais vieses regionais são fatores que podem influenciar os resultados.

Diante desses pontos, sugerem-se direções para estudos futuros, como a ampliação da análise para outros estados e o Distrito Federal, bem como a incorporação de um conjunto mais amplo de variáveis — incluindo aspectos demográficos e indicadores de engajamento digital — visando ao desenvolvimento de modelos ainda mais robustos. Além disso, propõe-se como perspectiva futura a geração de visualizações geográficas detalhadas por meio de ferramentas como o *QGIS*. Por fim, a investigação de fenômenos como o “voto útil” e a migração de votos na reta final da campanha, com base nos modelos aqui aplicados, apresenta-se como uma linha de pesquisa promissora para aprofundar a compreensão do comportamento eleitoral no Brasil.

REFERÊNCIAS

- BRASIL. **Lei nº 9.100, de 29 de setembro de 1995. Institui normas para as eleições, altera a Lei nº 8.212, de 24 de julho de 1991, e dá outras providências.** *Diário Oficial da União*, Brasília, 29 set. 1995. Disponível em: <https://www.planalto.gov.br>. Acesso em: 1 dez. 2024.
- BREIMAN, Leo. *Random forests. Machine Learning*, v. 45, n. 1, p. 5–32, 2001. DOI: 10.1023/A:1010933404324. Acesso em: 1 dez. 2024.
- CHAI, T.; DRAXLER, R. R. *Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development*, v. 7, p. 1247–1250, 2014. Acesso em: 1 dez. 2024.
- DARLINGTON, R. B.; WILKINSON, L. G. *Linear Regression*. 1. ed. New York: Wiley, 1985. ISBN 978-0471055669. Acesso em: 1 dez. 2024.
- GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.
- HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. New York: Springer, 2009. ISBN 978-0-387-84858-7. Acesso em: 1 dez. 2024.
- JAMES, G. et al. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer, 2014. Acesso em: 1 dez. 2024.
- JUNCKES, Ivan Jairo; JÚNIOR, Wilson Silva; SILVA, José Maria; SILVA, Edson Armando. **Representação espacial de dados eleitorais no Brasil: críticas e possibilidades**. *Revista de Sociologia e Política*, 2021. Acesso em: 1 dez. 2024.
- KUTNER, M. H.; NACHTSHEIM, C. J.; NETTER, J. B. *Applied Linear Regression Models*. 4. ed. New York: McGraw-Hill, 2004. Acesso em: 1 dez. 2024.
- MADDALENA, Giovanni Lemos; KAZAMA, Matheus Yudy; SAVIOLI, Vivian Monaco. **Pesquisa de Intenção de Voto em Redes Sociais**. 2021. Trabalho de Conclusão de Curso (Graduação em Engenharia de Controle e Automação) – Centro Universitário do Instituto Mauá de Tecnologia, São Caetano do Sul, 2021. Acesso em: 1 dez. 2024.
- MCCULLAGH, P.; NELDER, J. A. *Generalized Linear Models*. 2. ed. Boca Raton: CRC Press, 1989. Acesso em: 1 dez. 2024.
- NOGUEIRA, Érico Ferrari. **As Pesquisas Eleitorais como Condicionantes do Jogo Democrático**. Dissertação (Mestrado em Ciência Política) – Universidade Federal de Minas Gerais, Belo Horizonte, 2012. Acesso em: 1 dez. 2024.
- PAZ, Ana Cristina Ferreira da. **Voto eletrônico nas eleições democráticas gerais e municipais no Brasil**. *Revista Brasileira de Ensino e Aprendizagem*, 2023. Acesso em: 1 dez. 2024.
- RASCHKA, S. *Python Machine Learning*. Birmingham: Packt Publishing, 2017.

Acesso em: 1 dez. 2024.

SAUNDERS, M.; LEWIS, P.; THORNHILL, A. *Research Methods for Business Students*. 7. ed. Harlow: Pearson, 2016. Acesso em: 1 dez. 2024.

SEGAL, M. R. *Machine Learning: A Practical Approach*. ResearchGate, 2004. Acesso em: 1 dez. 2024.

TREIN, Sérgio Roberto. **O duplo papel das pesquisas eleitorais: eleitora de quem está na frente, adversária de quem está atrás**. In: V Encontro dos Núcleos de Pesquisa da Intercom, Universidade do Vale do Rio dos Sinos (UNISINOS), Universidade de Caxias do Sul (UCS), 2004. Acesso em: 1 dez. 2024.

VALENTIM, Gabriel P.; SEUFITELLI, Danilo P.; SILVA, Carlos A. **Eleições Presidenciais de 2022: Predições com Dados de Pesquisas Eleitorais do Poder360 e PollingData**. Trabalho de Conclusão de Curso – Instituto Federal de Minas Gerais, 2023. Acesso em: 1 dez. 2024.

VENTURI, Gustavo. **Pesquisas pré-eleitorais: legitimidade, influência e contribuições à cidadania**. Centro de Estudos de Opinião Pública (CESOP) – Universidade Estadual de Campinas (UNICAMP), 1995. Acesso em: 8 dez. 2024.

WILLMOTT, C. J.; MATSUURA, K. *Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance*. *Climate Research*, v. 30, p. 79–82, 2005. Acesso em: 9 jul. 2025.