

ANÁLISE DE SENTIMENTOS DE *TWEETS* EM RELAÇÃO À ELEIÇÃO PRESIDENCIAL DE 2022 NO BRASIL

Davi de Meneses Silva¹, Adonias Caetano de Oliveira¹, Paulo Cesar de Almeida Junior¹

¹Instituto Federal do Ceará (IFCE) *Campus* Tianguá
CE-187, s/n, Tianguá - CE, 62320-000, Brazil

davisilvameneses@gmail.com, {adonias.oliveira,paulo.almeida}@ifce.edu.br

Abstract. *With technological progress and the rise of social media, the routine use of social media platforms has consolidated, encompassing various activities, including political debates and expression of opinions. Given this scenario, people's opinions on social media play an important role in providing a real-time overview of society's perceptions and feelings, especially in the field of politics. With this in mind, this study analyzes the performance of Artificial Intelligence algorithms in sentiment analysis tasks in relation to the 2022 elections in Brazil. The methodology consisted of collecting tweets about the 2022 election from the old social network Twitter (currently X). The dataset was divided into two subsets, namely training labeled via ChatGPT-3.5 and testing manually labeled. Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors Algorithm (KNN), Random Forest (RF) and BERTimbau Large learning models were evaluated. As result, among the 5 models, BERTimbau Large obtained the best result with 68% accuracy.*

Resumo. *Com o progresso tecnológico e a ascensão das mídias sociais, o uso rotineiro das plataformas de redes sociais se consolidou, abrangendo diversas atividades e entre elas os debates políticos e expressão de opiniões. Diante desse cenário, a opinião das pessoas nas redes sociais desempenha um importante papel para presumir um panorama em tempo real das percepções e sentimentos da sociedade, principalmente no ramo da política. Visando isso, este estudo analisa o desempenho de algoritmos de Inteligência Artificial nas tarefas de análise de sentimentos em relação às eleições de 2022 no Brasil. A metodologia consistiu na coleta de tweets sobre a eleição de 2022 da antiga rede social Twitter (atual X). O conjunto de dados foi dividido em dois subconjuntos, a saber, treinamento rotulado por meio do ChatGPT-3.5 e de teste rotulado manualmente. Foram avaliados os modelos de aprendizagem Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors Algorithm (KNN), Random Forest (RF) e BERTimbau Large. Como resultado, dentre os 5 modelos, o BERTimbau Large obteve o melhor resultado com 68% de acurácia.*

1. Introdução

O ano de 2022 foi ímpar na história do Brasil e das eleições brasileiras. Como descrito por [Perez et al. 2022], o mundo passava pela pandemia da Covid-19 que, em época, já contava com 669 mil mortos somente no Brasil, juntamente com o retorno do então

ex-presidente, Luiz Inácio Lula da Silva, à disputa eleitoral, no qual gerou um conflito entre os três poderes: Executivo, Legislativo e Judiciário. Além disso, foi vivenciado um cenário econômico desfavorável, em que a inflação alta, que afetou diretamente a população, diminuindo o poder de compra, influenciou também o aumento da taxa de juros básicos. Esse período também foi marcado por constantes manifestações, principalmente durante e após o período eleitoral, como as manifestações de 8 de janeiro de 2023, que refletiu a polarização presente no país.

Neste cenário, as diversas opiniões políticas, emitidas nas eleições de 2022, indicaram a importância de compreender o comportamento dos eleitores, bem como os sentimentos que a sociedade vivenciou em relação aos candidatos e também aos conflitos políticos. Segundo [Silva 2022], a Ciência de Dados (CD) visa transformar dados em informação, a análise de sentimentos desempenha um papel fundamental para compreender as emoções humanas. Portanto, através do uso de técnicas avançadas CD, pode-se desvendar padrões de opiniões e emoções que moldam o cenário político, fornecendo percepções valiosas para candidatos, partidos políticos e eleitores.

Ademais, no âmbito da CD, é exequível efetuar diversas análises por meio de dados abertos disponibilizados pelo governo, como o projeto "Operação Serenata de Amor"¹, cujo objetivo é fiscalizar gastos públicos e compartilhar informações de maneira acessível a qualquer pessoa [de Jesus Procópio et al. 2023].

Além disso, no processo eleitoral de um país tão diverso e populoso como o Brasil, compreender profundamente os sentimentos e aspirações dos potenciais eleitores é crucial para promover uma representação mais precisa das necessidades da sociedade [Lago et al. 2005], como também aumenta a confiança dos cidadãos, pois segundo [Belchior 2016, p. 35] "os que tendem a confiar mais são também os que estão mais predispostos a estar envolvidos e a apoiar esse mesmo sistema político". Dessa forma, a análise significa bem mais que uma mera identificação dos sentimentos, tornando-se crucial para investigar as causas subjacentes que concebem tais sentimentos. Com isso, é factível traçar um plano de ação eficaz, com políticas alinhadas ao dêmico e que ressoem com às necessidades reais da população.

Nesse sentido, a análise de sentimentos exerce um papel essencial em diversos contextos, incluindo análises políticas [Reis et al. 2015] ao permitir que os candidatos e suas equipes/partidos entendam a dinâmica sociopolítica em que estão inseridos. Dessa forma, ao compreender o que motiva o entusiasmo, frustração, esperança ou a insatisfação dos eleitores, os candidatos conseguem ajustar suas propostas e estratégias de campanhas de uma forma mais precisa, fortalecendo suas conexões com os eleitores e promovendo um processo eleitoral mais representativo e democrático, conforme exposto por [Augusto and da Silva 2014, p. 66], que relata que "o sentimento de falta de representatividade leva o eleitor a não refletir e acreditar que a política partidária não possui qualquer importância para sua vida".

Diante disto, este estudo analisa o desempenho de modelos de aprendizagem em tarefas de análise de sentimentos sobre as eleições de 2022 no Brasil, sobretudo as opiniões acerca das candidaturas presidenciais. Foram coletados *tweets* sobre a eleição de 2022 da antiga rede social *Twitter* (atual X) de modo que o conjunto de dados foi

¹<https://serenata.ai/about/>

dividido em dois subconjuntos: base de treinamento rotulada por meio do *ChatGPT-3.5* e a base de teste rotulada manualmente. Foram avaliados os modelos de aprendizagem *Multilayer Perceptron (MLP)*, *Support Vector Machine (SVM)*, *K-Nearest Neighbors Algorithm (KNN)*, *Random Forest* e *BERTiumbau Large* por meio das métricas de acurácia, precisão, sensibilidade e medida-F1.

2. Fundamentação Teórica

2.1. Eleições de 2022 no Brasil

Para [Motta 2018], as eleições são um assunto recorrente desde o século XIX no mundo inteiro. No Brasil, não é diferente, 2022 foi marcado pela corrida eleitoral, que ocorre a cada quatro anos. Neste período, em particular, notou-se um grande conflito de ideais em relação aos eleitores quanto a escolha de voto para o cargo presidencial. Tal embate ideológico foi mais forte nas redes sociais, sobretudo na antiga rede social *Twitter* (atual X) onde ocorrem inúmeras discussões sobre diversos tópicos e também sendo a plataforma digital mais utilizada pelo governo federal, de acordo com [Ranzani and Caram 2019]. Além disso, consoante com a perspectiva de [Ruediger and Grassi 2018], o Brasil tem sofrido fortes impactos políticos, sociais e também no que tange ao ressurgimento de disputas partidárias.

De acordo com [Rennó 2022], as avaliações positivas do governo do presidente do Brasil de 2022, Jair Messias Bolsonaro, diminuíam à medida que as mortes relacionadas à Covid-19 aumentavam, resultado esse que também teve impacto por sua postura de confronto, o que gerou uma enorme divisão de opiniões políticas. Diante desse cenário, a polarização gerou a homogeneização dos discursos em cada parte dos grupos políticos envolvidos, causando assim o distanciamento destes [Rennó 2022].

Conforme dados divulgados pelo [TSE 2022], no primeiro turno, Luiz Inácio Lula da Silva e Jair Bolsonaro disputaram a liderança, em que ambos terminaram com 48,43% e 43,20% respectivamente, seguidos por Simone Tebet, com 4,16% e Ciro Gomes, com 3,04%. Dessa forma e como previsto, ao final do primeiro turno, Lula e Bolsonaro continuaram na disputa. Durante o período de espera até a votação do 2º turno, a população demonstrou um empenho mais intenso nas redes sociais, gerando diversas discussões e trocas de opiniões de caráter político. Ao final das votações Lula venceu Bolsonaro com 50,9%, contra 49,10%, de votos válidos, sendo eleito Presidente do Brasil pela 3ª vez [TSE 2022].

2.2. Classificadores

O *Random Forest (RF)* é um modelo que possui, como diferencial, uma combinação de diferentes modelos — árvores de decisão — para se obter um único resultado. No caso desse classificador, são usadas diversas árvores de decisão, construídas com base em dois passos, sendo eles: seleção de amostras feitas com o método *bootstrap*, utilizado para a seleção de amostras de treino; e seleção de variáveis para cada nó da árvore, escolhendo-se essas variáveis de forma aleatória [Pedregosa et al. 2011]. As vantagens desse algoritmo é sua usabilidade e por usar o sistema de votação de múltiplas árvores.

O *Multilayer Perceptron* de acordo com [Popescu et al. 2009] é o mais conhecido e mais usado algoritmo de rede neural, ele possui uma ou mais camadas ocultas e utiliza o *backpropagation* para atualizar os pesos da rede conforme as entradas fornecidas, possui

bom desempenho em problemas não linearmente separáveis com o aumento de camadas ocultas, que também pode aumentar a sua complexidade.

A Máquina de Vetores de Suporte (*Support Vector Machine*, SVM) é uma técnica de aprendizado supervisionado, útil para reconhecer padrões sutis em conjuntos de dados complexos [Pavlidis et al. 2004]. Ele também possui diversas aplicações, sendo utilizados em tarefas de classificação de textos, reconhecimento de imagens e dígitos. Uma de suas vantagens é a classificação de dados não regulares, já que ele consegue se sair bem em classificações não lineares, porém em *datasets* maiores o custo computacional aumenta muito.

Segundo [Andrade 2018, p. 97], o classificador *K-Nearest Neighbors* (KNN) consiste em processo de aprendizado em que a classe da amostra é calculada através da maior frequência dos k vizinhos mais próximos à instância analisada. A classificação baseada em vizinhos é um tipo de aprendizagem fundamentada em instâncias ou aprendizagem não generalizante. Dessa forma, ela não tenta construir um modelo interno geral, mas simplesmente armazenar instâncias dos dados de treinamento [Pedregosa et al. 2011]. Sua vantagem está na sua implementação e simplicidade e em obter bons resultados na maioria dos problemas, porém para cada nova amostra exige um grande custo computacional.

As variantes *BERTimbau* [Souza et al. 2020] são modelos de BERT (*Bidirectional Encoder Representations from Transformers*) pré-treinados e treinados na língua portuguesa que alcançam desempenhos de última geração em três tarefas *downstream* de processamento de linguagem natural: reconhecimento de entidade nomeada, similaridade textual de frase e reconhecimento de implicação textual. As variantes *BERT-Base* e *BERT-Large Cased* foram treinadas no BrWaC (Web brasileira como Corpus), um grande corpus português, neste trabalho será utilizado o modelo *BERTimbau Large Cased*.

3. Trabalhos Relacionados

[Joyce and Deng 2017] analisaram o sentimento expresso pelos *tweets* relacionados à eleição presidencial dos Estados Unidos da América no ano de 2016 e compararam os sentimentos coletados com os dados das pesquisas eleitorais. O estudo combinou análise de sentimento baseada em *lexicon* e algoritmo *Naive Bayes* para classificar os *tweets*. Além disso, foram utilizados *hashtags* específicas dos candidatos a fim de automatizar parte do processo de rotulagem dos *tweets*. A análise de sentimento baseada em *lexicon* teve uma correlação positiva alta para os *tweets* relacionados a Donald Trump, e uma negativa para os *tweets* relacionados a Hillary Clinton. A análise de sentimento baseada em aprendizagem de máquina teve uma correlação positiva alta para os *tweets* relacionados a Donald Trump, mas não superou a análise de *lexicon* quando comparada aos dados de pesquisa.

[Córdova Sáenz and Becker 2021] avaliaram os modelos BERT a classificação de posturas políticas expressas na rede social *Twitter* (atual X) em relação à vacinação contra a Covid-19 no contexto brasileiro. O modelo BERT foi pré-treinado em uma larga quantidade de dados e, em seguida, foi ajustado (*fine-tuned*) para a tarefa específica de classificação de posturas. Foram utilizados três conjuntos de dados contendo *tweets* em português do Brasil e três modelos BERT com suporte para essa língua. Foram alcançados desempenhos muito semelhantes entre os três modelos BERT, quando treinados em cada um dos três conjuntos de dados, com medidas F1 variando de 0,722 (problema de três

posturas) a 0,859 (problema de duas posturas).

[Gonçalves et al. 2022] analisaram *tweets* relacionadas à Covid-19 tendo em vista identificar as emoções predominantes dos usuários. O objetivo final foi compreender os comportamentos e emoções dos usuários da rede social *Twitter* (atual X) durante a pandemia, com intuito de obter evidências úteis para situações semelhantes no futuro. Foram utilizados diversos classificadores, incluindo RF, KNN, *Passive Aggressive*, *Naive Bayes*, *Gradient Boosting*, XGB, MLP, *Logistic Regression* e SVC. No estudo de [Gonçalves et al. 2022], foram obtidos resultados satisfatórios para algumas combinações de emoções. A combinação de Alegria/Tristeza apresentou acurácia acima de 70%.

[Silva 2023] utilizou técnicas de mineração de dados textuais para analisar os comentários oriundos do *Twitter* em relação aos candidatos da eleição presidencial de 2022 no Brasil. O objetivo principal foi verificar se o desempenho dos candidatos estava relacionado com a sua popularidade nas redes sociais. Os resultados indicaram semelhanças entre a popularidade dos candidatos na rede social e sua intenção de voto em pesquisas eleitorais. No entanto, não foram encontradas associações entre a popularidade de um determinado candidato e seu desempenho nas eleições.

O diferencial deste estudo em relação aos trabalhos relacionados é identificação dos sentimentos positivos e negativos em tweets publicados durante eleições brasileiras de 2022, como também por investigar classificadores capazes de alcançar resultados satisfatórios na tarefa de classificação proposta utilizando uma base de treino rotulada unicamente por um modelo de inteligência artificial ChatGPT-3.5.

4. Materiais e Métodos

A metodologia utilizada neste trabalho tem como característica a pesquisa exploratória-descritiva quantitativa. Além disso, este estudo também está relacionado com a aplicação de métodos e técnicas de CD para os procedimentos de extração de dados e de Inteligência Artificial (IA). O processo metodológico foi organizado em quatro etapas descritas nas próximas subseções.

4.1. Extração de Tweets

Para obter os *tweets* dispostos nesse trabalho, foi utilizada a API do *Twitter*, por meio de um portal de desenvolvimento, no qual foi necessário vincular uma conta de usuário com as chaves de acesso gerada pelo próprio portal. A partir disto, foi escrito um *script* utilizando a linguagem *python* em conjunto da biblioteca *tweepy*. O código foi executado através do *Google Colab* e as consultas da API eram compostas por determinados parâmetros, sendo eles: os dados de autenticação com a aplicação do *Twitter*, nome da aplicação e palavra-chave da consulta.

Durante os meses de agosto e setembro do ano de 2022, a coleta de *tweets* que continham os nomes dos quatro primeiros candidatos à presidência como palavras-chaves da busca. Os quatro candidatos foram aqueles com maiores intenções de votos, a saber: Lula (intenção de votos entre 40% e 45%), Jair Bolsonaro (entre 33% e 36%), Ciro Gomes (entre 5% e 9%) e Simone Tebet (entre 2% e 6%).

4.2. Rotulação Manual e Automática dos Textos

Foi realizada uma seleção de 1300 *tweets*, dos quais 1000 seriam utilizados como base de treinamento e o restante como teste. Estas 1000 amostras supramencionadas foram usa-

das para rotulação automática utilizando a API da *OpenAI*, empresa responsável pelo chatGPT, através de um *script* em *Python*, que solicitava ao modelo que realizasse a identificação entre as polaridades positiva e negativa para cada *tweet*. Os demais 300 *tweets* foram rotulados manualmente por três pessoas de faixa etária 23-26 anos, sendo eles: um graduado em Ciência da Computação e Pedagogia, um estudante de Finanças e um graduado em História, cada pessoa realizou a rotulação de 100 amostras com a ciência de manter a abordagem imparcial sobre os textos, a fim de formar a base de teste deste trabalho.

4.3. Processamento de textos

Segundo [de Ávila et al. 2022], o processamento de textos é uma etapa fundamental do Processamento de Linguagem Natural (PLN) que irá preparar o texto para a extração automatizada. Desta forma, foi desenvolvido um *script* na linguagem *Python* em conjunto da biblioteca *Natural Language Toolkit (NLTK)*, utilizada para realizar a remoção de informações desnecessárias contidas nos textos, tais como *stopwords*, *links*, nomes de usuários e a remoção dos nomes dos candidatos a presidência para evitar que os algoritmos associem os nomes dos candidatos a qualquer sentimento. Ainda, foi aplicado uma transformação de todos os caracteres em letras minúsculas. Após este processo foi realizada a construção da base dados em um arquivo de formato *.xlsx* para realizar a “tokenização” e aplicação posterior nos classificadores.

4.4. Experimentos Computacionais

Para a realização dos experimentos, foram desenvolvidos dois *scripts* na plataforma *Google Colaboratory*. O primeiro *script* foi utilizado somente para a realização dos testes com o *BERTimbau*, com fim de classificação, utilizando o corpus em português do *BERTimbau*. O segundo *script* foi utilizado para os demais classificadores em conjunto, sendo eles: RF, SVM, MLP e KNN. Além disso, foram utilizadas duas abordagens de *Embeddings* de palavras da biblioteca *Word2Vec* e *BERTimbau*, ambos em português.

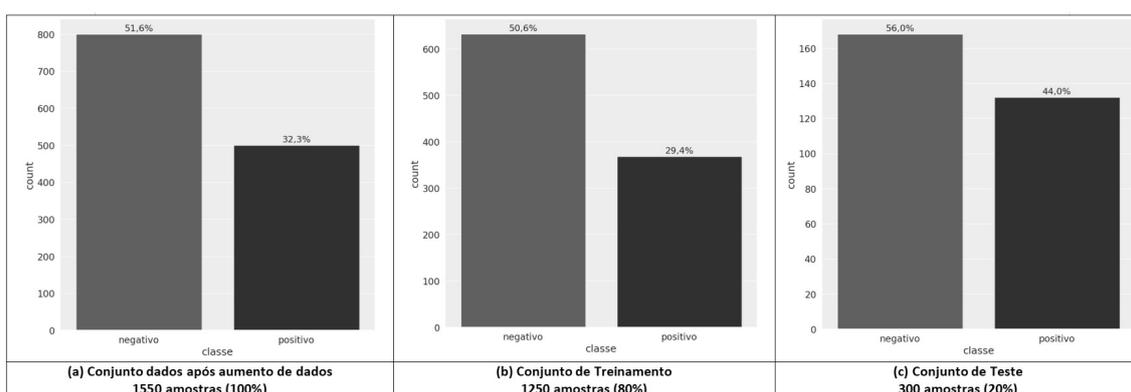


Figura 1. Distribuição das amostras entre as classes positivo e negativo.

Para o conjunto de teste foram selecionados os 300 *tweets* rotulados manualmente. Para o restante do conjunto de dados (1000 amostras) foi aplicada a técnica de geração de amostras *Data augmentation* baseado no *BERTimbau Large* mediante *nlpaug library* [Ma 2019]. Neste sentido, foram obtidas 250 amostras a mais totalizando 1250 amostras. Portanto, o conjunto de dados foi distribuído em 1150 (74%) amostras para treinamento,

100 (6%) amostras para validação e 300 (20%) amostras para teste. A Figura ?? apresenta a distribuição das amostras entre as classes.

Em seguida, para realizar os experimentos com *BERTimbau Large*, neste estudo considerou-se os melhores resultados de épocas 3 e 4, bem como, com taxas de aprendizado de $2e-6$ e $3e-5$, na qual, para cada época, foi realizado teste com as duas taxas de aprendizado.

4.5. Métricas de avaliação

Neste estudo, foram avaliados os modelos a partir da acurácia, precisão, sensibilidade, *Medida-F1* e especificidade. As métricas consideram os verdadeiros positivos (VP) como a quantidade no qual o classificador prevê corretamente uma amostra como classe positiva; os verdadeiros negativos (VN) como a quantidade de padrões classificados corretamente como classe negativa; os falsos positivos (FP) como o total de previsões na qual o classificador prediz incorretamente um padrão de classe negativa como positiva; e os falsos negativos (FN) como total de exemplos da classe positiva onde o classificador prevê incorretamente como negativa.

A acurácia considera, dentre todas as entradas de teste, o percentual de previsões corretas dos sentimentos de cada tweet. Essa medida pode ser obtida a partir da equação $acuracia = \frac{VP+VN}{VP+VN+FP+FN}$.

A equação $precisao = \frac{VP}{VP+FP}$ estabelece o indicador de precisão, sendo este o indicador responsável por determinar a quantidade de tweets da classe sentimento positivo classificados corretamente (Verdeiro Positivo) em relação à soma dos verdadeiros positivos com falso positivos.

A sensibilidade ou *recall* indica a proporção de tweets corretamente classificados como positivos sobre o total de tweets realmente positivos. Pode ser calculado conforme a equação $sensibilidade = \frac{VP}{VP+FN}$.

A medida-F1 calcula a média harmônica entre precisão e sensibilidade e pode ser determinada a partir da equação $medida-F1 = \frac{2 * precisao * sensibilidade}{precisao + sensibilidade}$.

5. Análise dos Resultados

Diante a realização dos experimentos computacionais utilizando o modelo *BERTimbau Large*, com as diferentes épocas e taxas de aprendizado, foram obtidos os resultados apresentados na Tabela 1.

Observa-se que, nos experimentos com *BERTimbau Large*, a maior taxa de aprendizado experimentada gerou melhores resultados de acurácia. A classe negativa, que possuía a maior quantidade de amostras, mantinha as maiores taxas de precisão. O melhor resultado em relação à acurácia, portanto, foi utilizando quatro épocas com a taxa de aprendizado $3e-5$. Entretanto, a precisão para classe positiva foi muito baixa, indicando muitos falsos positivos. Ademais, na métrica de Medida-F1 para classe negativa observa-se sempre valores acima de 70% e para classe positiva a melhor média harmônica foi de 52% que em comparação aos classificadores tradicionais obtiveram valores abaixo de 50%

Conforme a Tabela 2, os resultados dos classificadores tradicionais indicam uma melhora significativa quando utilizado o *BERTimbau Large* nos *embeddings* em

Tabela 1. Resultados *BERTimbau Large*

Épocas	Taxa de Aprendizado	Classe	Acurácia	Precisão	Sensibilidade	Medida-F1
3	2e-6	Negativo	0.63	0.85	0.63	0.72
		Positivo		0.36	0.65	0.46
3	3e-5	Negativo	0.67	0.89	0.65	0.75
		Positivo		0.40	0.74	0.52
4	2e-6	Negativo	0.62	0.86	0.61	0.72
		Positivo		0.30	0.63	0.41
4	3e-5	Negativo	0.68	0.92	0.65	0.76
		Positivo		0.37	0.78	0.50

Tabela 2. Resultados dos modelos de aprendizagem tradicionais

Model	Embedding	Train acc	Test acc
SVM	Word2Vec	0.784	0.58
	<i>BERTimbau Large</i>	0.734	0.653
Multilayer Perceptron	Word2Vec	0.634	0.61
	<i>BERTimbau Large</i>	0.813	0.623
Random Forest	Word2Vec	0.995	0.63
	<i>BERTimbau Large</i>	0.995	0.68
K-NN	Word2Vec	0.773	0.53
	<i>BERTimbau Large</i>	0.787	0.56

comparação à técnica *Word2Vec*. Com *BERTimbau Large* como *embeddings*, os quatro classificadores tradicionais obtiveram melhores resultados, dos quais o RF foi o melhor modelo com uma acurácia de 68%, seguido pelos modelos SVM e MLP com acurácias de 65% e 62%, respectivamente.

Tabela 3. *Random Forest* com *BERTimbau Large Embedding*

Classe	Precisão	Sensibilidade	Medida-F1
Negativo	0.62	0.85	0.72
Positivo	0.65	0.35	0.45

Entretanto, o RF quando comparado com o Modelo *BERTimbau Large* nas demais métricas, fica evidenciado que a sensibilidade para a classe positiva foi bem abaixo dos resultados, conforme pode ser visto nas Tabelas 1 e 3. A partir desses dados, pode-se concluir que o modelo RF com *BERTimbau Large Embedding* obteve desempenho mais balanceado na predição de sentimentos negativos (medida-F1 = 72%).

6. Conclusão

Este estudo analisou o desempenho de modelos de IA na classificação de sentimentos positivos e negativos dos brasileiros acerca das eleições brasileiras de 2022. Os modelos com *BERTimbau Large Embedding* obtiveram melhores resultados, em comparação ao *Word2Vec*. O *BERTimbau Large* e RF apresentaram o melhor desempenho com 68% de acurácia, enquanto o SVM teve 65% de acurácia. Esta pesquisa também pôde verificar

que mesmo utilizando um modelo como chatGPT para automatizar a rotulação de dados de treino os resultados em casos reais foi satisfatório.

Para trabalhos futuros é recomendável, incluir mais métodos de análises de sentimentos, bem como classificar mais *tweets*, para aumentar o tamanho da base de dados. Realizar validação cruzada usando otimizador de hiperparâmetros para encontrar o melhor modelo. Além disso, pode ser realizada uma triagem rigorosa de perfis para verificar se o perfil é falso ou robotizado, e também uma triagem de perfis de menores de 16 anos, já que não possuem poder de voto no Brasil, assim, poderemos ter um resultado mais específico e com mais credibilidade para estudos nessa temática. Também é necessário pesquisar uma solução para criar um melhor resultado, diminuindo a taxa de falsos positivos.

Referências

- Andrade, V. S. (2018). Reconhecimento de texturas utilizando padrão binário local e classificador knn.
- Augusto, D. C. and da Silva, M. (2014). Apontamentos sobre política partidária e eleitores no município de guarapuava/pr: Considerações a partir de uma geografia eleitoral qualitativa. *REVISTA PERCURSO*, 6(2):55–76.
- Belchior, A. M. (2016). *Confiança nas instituições políticas*. Fundação Francisco Manuel dos Santos, Lisboa.
- Córdova Sáenz, C. A. and Becker, K. (2021). Assessing the use of attention weights to interpret bert-based stance classification. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pages 194–201, online.
- de Ávila, P. V. M. L., de Brito, D. M., Santos, D. M., and Ferreira, E. d. A. M. (2022). Processamento de linguagem natural (pln) para automatização da checagem de conformidade: uma investigação do pré-processamento de um código regulatório urbanístico brasileiro. *ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO*, 19:1–12.
- de Jesus Procópio, J. V., Lopes, L., and Feitosa, W. R. (2023). Geoanalytics aplicado ao conceito de cidades inteligentes. *Revista de Inovação e Tecnologia-RIT*, 13(1):109–117.
- Gonçalves, G. F., Rocha, A. A. d. A., and Paes, A. (2022). Analisando as emoções dos tweets relacionadas à covid-19 no rio de janeiro. In *Anais do VI Workshop de Computação Urbana*, pages 210–223, Fortaleza. SBC.
- Joyce, B. and Deng, J. (2017). Sentiment analysis of tweets for the 2016 us presidential election. In *2017 ieee mit undergraduate research technology conference (urtc)*, pages 1–4, Cambridge, MA, USA. IEEE.
- Lago, I. C. et al. (2005). O significado do voto em eleições municipais: análise dos processos de decisão de voto em eleições para prefeito em itajaí/sc.
- Ma, E. (2019). Nlp augmentation. <https://github.com/makcedward/nlpaug>.
- Motta, K. S. d. (2018). *Eleições no Brasil do Oitocentos: entre a inclusão e a exclusão da patuleia na cidadela política (1822-1881)*. PhD thesis, Tese (Doutorado em História). Universidade Federal do Espírito Santo.

- Pavlidis, P., Wapinski, I., and Noble, W. S. (2004). Support vector machine classification on the web. *Bioinformatics*, 20(4):586–587.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Perez, R. T., Barbosa, L. G. C., de Oliveira, M. T., and da Silva Vaccari, G. (2022). Comentários sobre a diáde ‘voto ideológico’/‘voto econômico’ nas eleições presidenciais de 2022 no brasil. *Revista InterAção*, 13(2).
- Popescu, M.-C., Balas, V. E., Perescu-Popescu, L., and Mastorakis, N. (2009). Multi-layer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7):579–588.
- Ranzani, L. H. and Caram, N. R. (2019). A migração da comunicação política para o ambiente digital: Twitter de jair bolsonaro no segundo turno das eleições presidenciais de 2018. *Revista Multiplicidade*, 9(9).
- Reis, J., Gonçalves, P., Araújo, M., Pereira, A. C., and Benevenuto, F. (2015). Uma abordagem multilíngue para análise de sentimentos. In *Anais do IV Brazilian Workshop on Social Network Analysis and Mining*, Recife. SBC.
- Rennó, L. (2022). Bolsonarismo e as eleições de 2022. *Estudos Avançados*, 36:147–163.
- Ruediger, M. A. and Grassi, A. (2018). Redes sociais nas eleições 2018.
- Silva, L. A. N. d. (2022). Ciência de dados como método de transformação de dados em informação.
- Silva, S. M. B. (2023). Análise de sentimentos expressos no twitter em relação aos candidatos da eleição presidencial de 2022.
- Souza, F., Nogueira, R., and Lotufo, R. (2020). Bertimbau: Pretrained bert models for brazilian portuguese. In Cerri, R. and Prati, R. C., editors, *Intelligent Systems*, pages 403–417, Cham. Springer International Publishing.
- TSE, T. (2022). Resultados 2022.