

## **Comparação entre Modelos para Análise de Sentimentos no Contexto de Avaliações de Filmes**

Rodrigo Franciozi Rodrigues da Silva<sup>1\*</sup>; Adâmara Santos Gonçalves Felício<sup>2</sup>

<sup>1</sup> Analista de dados. Avenida Getúlio Vargas, 405 – Baeta Neves; 09751-250 São Bernardo do Campo, São Paulo, Brasil

<sup>2</sup> PECEGE. Doutora em Desenvolvimento Econômico, UNICAMP. Orientadora MBA Data Science & Analytics. – Rua Alexandre Herculano, 120 – Vila Monteiro; CEP:13418-445 Piracicaba, São Paulo, Brasil.

\*autor correspondente: rodrigofranciozi@outlook.com

## **Comparação entre Modelos para Análise de Sentimentos no Contexto de Avaliações de Filmes**

### **Resumo**

Este estudo explorou a capacidade de modelos de aprendizado de máquina na classificação de sentimentos em textos de comentário de filmes. O objetivo foi comparar a eficácia de modelos conhecidos como clássicos e de aprendizado profundo, especificamente o Naïve Bayes e “Bidirectional Encoder Representations from Transformers” [BERT] ao classificar sentimentos contidos em comentários de acordo com as polaridades positiva, negativa e neutra. Utilizou-se um conjunto de dados de avaliações oriundas da plataforma do “Internet Movie Database” [IMDB], aplicando uma metodologia que envolveu a preparação dos dados, implementação dos modelos, treinamento e avaliação de sua precisão e eficiência. Os resultados demonstraram que o BERT com sua capacidade de compreender o contexto e as nuances da linguagem, superou em 8% o Naive Bayes em acurácia e apresentou uma melhora nas classificações de sentimentos para as classes positivas e negativas com relação as métricas precisão, revocação e pontuação F1, mesmo em condições de recursos computacionais limitados, embora este último, tenha se destacado por sua rapidez e simplicidade. Os resultados alcançados reforçam a importância do emprego de técnicas de aprendizado profundo para análise de sentimentos, evidenciando a complexidade e as respectivas importâncias do contexto e da subjetividade inerentes aos textos analisados.

**Palavras-chave:** PLN; aprendizado profundo; mineração de opinião; inteligência artificial; classificação de texto.

## **Comparison Between Models for Sentiment Analysis in the Context of Movie Reviews**

### **Abstract**

This study explored the capability of machine learning models in classifying sentiments in movie review texts. The goal was to compare the effectiveness of models known as classic and deep learning, specifically Naïve Bayes and “Bidirectional Encoder Representations from Transformers” [BERT], in classifying sentiments contained in comments according to positive, negative, and neutral polarities. A dataset of reviews from the “Internet Movie Database” [IMDB] platform was used, applying a methodology that involved data preparation, model implementation, training, and evaluation of their accuracy and efficiency. The results showed that BERT, with its ability to understand context and the nuances of language, outperformed naïve Bayes by 8% in accuracy and presented an improvement in sentiment classifications for positive and negative classes with respect to precision, recall, and F1 score metrics, even when possessing limited computational resources, although the latter stood out for its speed and simplicity. The study reinforces the importance of employing deep learning techniques for sentiment analysis, highlighting the complexity and respective importances of context and subjectivity inherent in the texts analyzed.

**Keywords:** NLP; deep learning; opinion mining; artificial intelligence; text classification

## Introdução

Nos últimos anos a indústria audiovisual cresceu consideravelmente e junto a esse crescimento advém o interesse de agradar ao público para que obras de sucesso de bilheteria sejam criadas. Uma maneira de se verificar tal sucesso é por meio da checagem geral de notas e pela leitura de comentários sobre críticos especializados e o público geral que já assistiram determinado conteúdo (Yassine, 2019).

Quando observado o conjunto dos comentários produzidos sobre o assunto, identificamos plataformas online que trabalham no registro das críticas e avaliações por meio do texto escrito, como o “Rotten Tomatoes”, IMDB e “Metacritic”. Esses sites contêm uma biblioteca vasta de registros que funcionam como bases de dados, para que pesquisadores consigam avaliar o engajamento e potencial sucesso de produções cinematográficas (Harsh *et al.*, 2023).

As redes sociais têm se tornado plataformas cada vez mais relevantes para a expressão de sentimentos e ideias, especialmente no que diz respeito a filmes e séries. Segundo Faisal e Ubaid (2023), muitos indivíduos utilizam esses espaços virtuais para compartilhar suas avaliações sobre obras cinematográficas, influenciando assim as decisões de outros consumidores em relação ao investimento de tempo e dinheiro em determinadas produções.

Essa tendência de compartilhamento de opiniões e avaliações não apenas molda a percepção do público, mas também serve como um valioso recurso para os produtores de conteúdo. Conforme apontado por Haoran (2023), os produtores podem aproveitar essas informações disponíveis nas redes sociais para obter as reações do público, facilitando a identificação de potenciais sucessos ou fracassos, permitindo-lhes adaptar e aprimorar suas obras cinematográficas.

A análise de sentimentos, conhecida também como mineração de opinião, consiste em uma técnica que avalia a polaridade dos sentimentos, emoções e atitudes de determinado indivíduo de acordo com um objeto de interesse podendo ele ser um filme, produto, serviço e organização (Azilawati *et al.*, 2019). Essa tecnologia se baseia em inteligência artificial que usa o processamento de linguagem natural [PLN] para transformar textos não estruturados em dados normalizados e adequados para análises de algoritmos de aprendizado de máquina, que por sua vez auxiliam na identificação de “insights” e tomada de decisão a respeito de um determinado tópico de interesse (Faisal e Ubaid, 2023). De acordo com Yassine (2019), até pouco tempo essas análises de comentários eram feitas de forma

bastante manual, mas com o advento de técnicas de aprendizado de máquina como a análise de sentimentos, essa realidade acabou sendo modificada.

A importância do emprego de técnicas de PLN na análise de sentimentos de filmes, reflete tanto na automatização do processo de coleta e interpretação de um grande volume de dados, como também oferece uma visão sistematizada das percepções sobre determinados espectadores. Ao utilizar PLN, pesquisadores e profissionais podem extrair nuances que outrora seriam difíceis de se identificar manualmente, facilitando assim uma compreensão mais abrangente a respeito das reações do público (Yankang Su e Zbigniew j. Kabala, 2023).

A distinção entre modelos clássicos ou convencionais de aprendizado de máquina e os tidos como profundos, marcaram uma revolução na capacidade de análise e processamento de dados complexos. Enquanto os modelos clássicos dependem de técnicas estatísticas e abordagens preditivas baseadas em características previamente definidas, os modelos de aprendizado profundo aprendem automaticamente essas características a partir dos dados, com uma arquitetura denominada rede neural que imita características do cérebro humano (Yann Lecun *et al.*, 2015).

Destarte, este trabalho busca investigar como modelos de análise de sentimentos podem efetivamente distinguir opiniões positivas, negativas e neutras, com o intuito principal de se avaliar e comparar a eficácia de modelos convencionais, como o Naïve Bayes, frente a abordagem de aprendizado profundo baseada em redes neurais, representada pelo BERT.

## **Material e Métodos**

### **Base de dados**

A base de dados foi obtida por meio da aplicação de uma técnica de raspagem de dados (“web scrapping”) dentro da plataforma do IMDB<sup>1</sup>. A técnica de “Web Scrapping” consiste na extração de dados da internet de maneira programática, transformando-os em uma base de dados estruturada e permitindo uma coleta de grandes volumes de maneira automatizada minimizando potenciais erros (Mine Çetinkaya-Rundel e Mine Dogucu, 2021).

A lista de filmes foi selecionada de maneira aleatória levando em consideração os anos de lançamento, no período entre 2022 e 2023, para que as classificações fossem realizadas em comentários avaliados recentemente, mantendo apenas comentários em inglês devido ao

---

<sup>1</sup> Disponível em: [www.imdb.com](http://www.imdb.com)

site ser de origem norte américa e a natureza desses comentários estar majoritariamente na língua nativa. Após a obtenção e estruturação dos dados, técnicas de limpeza e transformação como a remoção de espaços em branco, emoticons, números e caracteres especiais, assim como a padronização de palavras para letras minúsculas e remoção de palavras vazias (“stop-words”) foram aplicadas. As palavras vazias representam uma lista de palavras que tipicamente aparecem com uma certa frequência na construção de frases, mas que não denotam uma significância relevante ao usuário final Faisal e Ubaid (2023). O fluxograma que implementa as etapas de transformação está descrito na Figura 1.

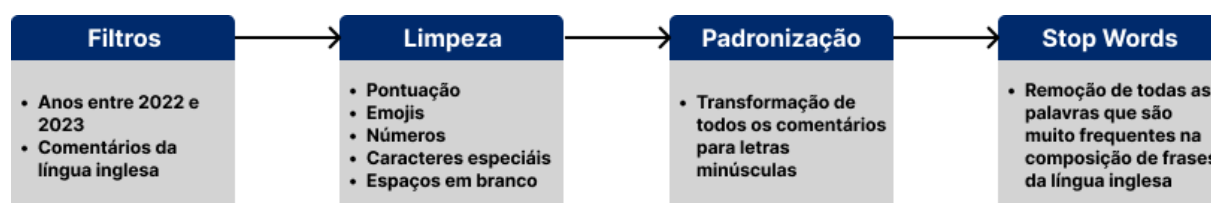


Figura 1. Fluxo de limpeza e tratamento de dados  
Fonte: Dados originais da pesquisa

Ao final de todas as etapas, obteve-se um  $n$  amostral de 14.480 comentários para a definição das polaridades juntamente com o treino e teste dos modelos.

## Metodologia

Foi necessário a implementação de um processo de rotulagem de comentários, para que esses pudessem ser classificados de acordo com a polaridade (positivo, negativo e neutro) presente no texto. Essa se torna uma das grandes dificuldades, devido a subjetividade presente na tarefa de classificação, assim como o tempo necessário para que a rotulagem seja feita. Para facilitar o processo de rotulagem, foram implementadas duas técnicas de aprendizado não supervisionado com a finalidade de comparação. Tais técnicas consistem na implementação de dicionários léxicos contendo palavras que já possuem sentimentos atrelados e expressos de uma maneira quantitativa, podendo ser números entre -1 e 1 onde -1 representa uma polaridade mais negativa, 0 neutra e 1 positiva (Fabrício *et al.*, 2015).

As técnicas escolhidas foram o “Opinion Lexicon”, composto por uma lista de aproximadamente 13.000 palavras incluindo gírias e abreviações na língua inglesa, que foi desenvolvido a partir de textos de avaliações de sites de compra e o “Valence Aware Dictionary and Sentiment Reasoner” [VADER] que é uma composição de outros dicionários já bem estabelecidos, com o adendo de emoticons, acrônimos e gírias (Fabrício *et al.*, 2015).

Ambas as saídas dos dicionários são valores que podem ser -1, 0 e 1, representando as polaridades. Esses valores são somados para cada uma das avaliações e classificados de acordo com os resultados, sendo negativos para somas menores do que 0, positivos para maiores que 0 e neutras iguais a zero.

Dois modelos de aprendizado de máquina, a título de comparação foram treinados e testados, avaliando métricas como Acurácia, Precisão, Revocação e Teste F1 a fim de escolher aquele que melhor se ajustou a tarefa de predição das classificações de polaridade.

A métrica de precisão, consiste nas previsões corretas de uma classe com relação a todas as previsões feitas para a mesma classe, a revocação representa as previsões corretas de uma classe com relação ao total de casos que realmente pertencem a essa classe e a pontuação F1 é uma média harmônica entre a precisão e revocação, fornecendo um valor que equilibra ambas as métricas. Essa etapa foi conduzida não apenas para determinar o melhor modelo em termos de capacidade preditiva, mas também para aprimorar a capacidade de análise de sentimentos e fornecer insights mais profundos sobre a recepção de conteúdos cinematográficos ao público.

Devido aos rótulos estarem desbalanceados, a técnica de “oversampling” envolvendo repetição foi aplicada com o auxílio do método “resample”, tendo como base a classificação positivo, que contém o maior número de exemplos. A técnica de “oversampling” é utilizada quando se tem dados insuficientes, fazendo com que haja uma expansão da base em relação as amostras com menor volume de informações, gerando mais dados ao invés de deletar os já existentes. Caso essa técnica não seja aplicada, pode existir um impacto no treinamento dos modelos justamente por ferir a capacidade de generalização buscada, tendo em vista que com menos exemplos fica mais complicado do modelo conseguir identificar nuances (Nassera *et al.*, 2023).

Para ambos os modelos foi realizado um particionamento de 80:20 entre os dados alocados para treinamento e teste respectivamente. Essa abordagem segue o princípio de Pareto<sup>2</sup> onde 20% dos fatores estão representados nos 80% outros fatores (Rosie *et al.*, 2014).

O primeiro modelo treinado foi o Naïve Bayes. Esse modelo é derivado do teorema de Bayes, onde se calcula a probabilidade de um evento acontecer com base em probabilidades

---

<sup>2</sup> Há 100 anos atrás, o economista Italiano Wilfred Pareto fez a observação de que 80% de toda a riqueza estava concentrada em 20% da população, o que mais tarde a partir do desenvolvimento de uma fórmula matemática acabou descrevendo esse fenômeno de distribuição desigual presente em seu país. Essa formulação ficou conhecida como distribuição de Pareto e mais tarde, em 1940 como regra de Pareto devido a generalização por parte de J.M Juran.

condicionais, podendo ser aplicado no contexto de análise de sentimentos que leva em consideração a probabilidade de um texto pertencer a determinada classificação (Murtadha *et al.*, 2022).

A técnica de vetorização denominada “term frequency-inverse document frequency” [TF-IDF] foi aplicada permitindo a conversão de dados textuais em numéricos. Essa estatística reflete a importância das palavras dentro de um documento e é definida pelo número de vezes que determinada palavra aparece, em comparação com o número de documentos que contenham a palavra. Entenda documentos como sendo cada comentário presente na base de dados (Christine *et al.*, 2023).

Por se tratar de uma classificação multi-classes, ou seja, contendo mais do que duas categorias a serem classificadas, foi aplicado o Naïve Bayes multinomial que é especificamente adequado para variáveis que apresentam contagens de palavras e que normalmente são obtidas usando técnicas de bolsa de palavras (“bag of words”) ou TF-IDF. Esse modelo assume como base que as palavras pertencentes a uma determinada classe são independentes entre si de acordo com determinada classificação, seguindo uma distribuição multinomial. Dessa forma calcula-se a probabilidade de um documento pertencer a essa classe, de acordo com a frequência de contagem das palavras naquele documento (Christine *et al.*, 2023).

O segundo modelo treinado foi o BERT, que consiste em um modelo de rede neural desenvolvido pela Google em 2018 com o intuito de compreender o contexto de frases, auxiliando máquinas no processo de linguagem natural com o emprego da arquitetura de transformadores e transferência de aprendizado. BERT é o primeiro modelo a atingir o estado da arte com relação a performance para atividades em nível de sentença e “token” (Jacob *et al.* 2018).

Uma versão pré-treinada do BERT foi utilizada, mas com uma adaptação referente ao volume de classes previstas que passou de duas para três. Devido a essa mudança nas classes apenas a estrutura definida pela rede neural foi aproveitada, havendo a necessidade da realização de um retreino com os dados presentes no estudo. A técnica de “tokenização” foi aplicada para configurar corretamente os dados da camada de entrada da rede neural, habilitando assim a possibilidade de treinamento do modelo.

A função de perda escolhida foi a entropia cruzada que é universalmente utilizada em trabalhos que tangem processos de classificação em redes neurais. Além disso, essa função se torna mais favorável em termos de otimização quando se trabalha com classificações de mais de duas classes Like e Mikhail (2021). O otimizador mais comum e o escolhido para atualização dos pesos das redes foi o “Adaptative Moment Estimation with Weight Decay” [AdamW] que consiste em uma versão modificada da otimização estocástica de primeira

ordem Adam. Ambos se diferem devido ao ADAMw omitir uma das etapas de correção de viés auxiliando no processo de evitar o sobre ajuste na etapa de treinamento (Tianyi *et al.*, 2021).

Devido a limitações técnicas presentes no processo de treinamento, foi considerado um tamanho de lote de 32 e um total de 3 épocas. Como a função de perda continuou apresentando uma redução ao invés de estagnação, isso permitiu uma margem para a continuação do treinamento considerando um volume maior de épocas, devido ao fato dos algoritmos de aprendizado profundo terem como objetivo principal a minimização das funções de perda entre as predições dadas pelo modelo e os valores reais pertencentes a base de dados de treinamento (Alexander *et al.*, 2023).

Todas as etapas foram executadas por meio da linguagem Python, em particular com o uso das bibliotecas “selenium”, “requests” e “random” para extração de dados da internet e aplicação de aleatoriedades, “nltk”, “pandas” e “re” para limpeza, tratamento e pré-processamento dos dados, “sklearn” e “transformers” para implementação de modelos de aprendizado de máquina e “matplotlib” e “seaborn” para a criação de gráficos e visualizações.

## **Resultados e Discussão**

### **Resultados**

A Figura 2 exibe as quatro classificações realizadas para ambos os dicionários léxicos, sendo consideradas padrão aquelas contendo apenas a implementação da técnica, e customizada que envolveram uma remoção de palavras extras que aparecem com muita frequência em ambas as avaliações, sem trazer um significado relevante.

O dicionário “Opinion Lexicon” customizado foi escolhido como rótulo oficial das avaliações, devido a apresentar uma sensibilidade maior em suas classificações, principalmente levando em consideração os comentários de natureza neutra.



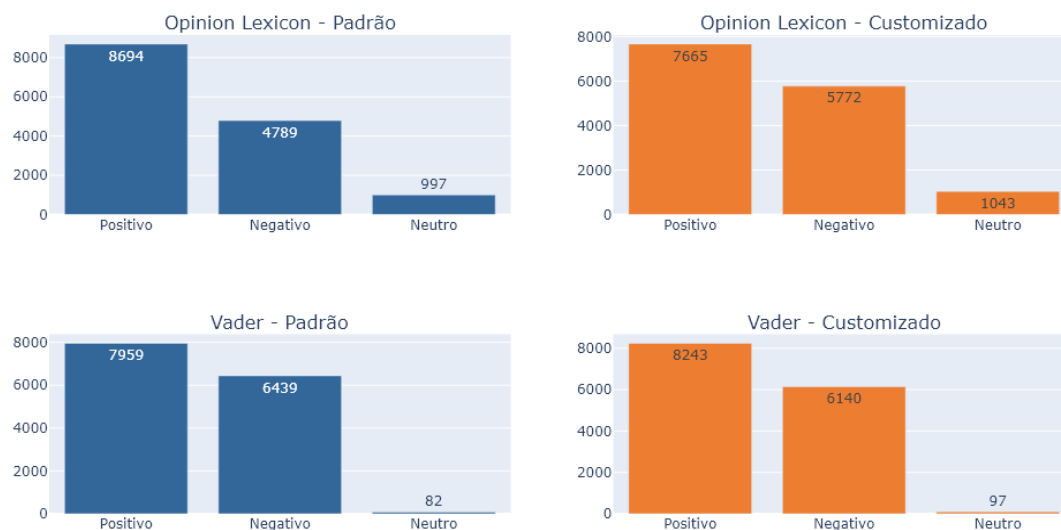


Figura 2. Classificações das polaridades por meio dos dicionários léxicos  
Fonte: Resultados originais da pesquisa

Utilizou-se uma matriz de confusão multiclasse para quantificar a precisão dos modelos de classificação propostos. Tal matriz, proporciona uma visão clara de seus comportamentos em todas as classes, em que as colunas representam as previsões do modelo e as linhas os valores reais. Cada célula mostra o número de amostras real que foram previstas na categoria correspondente e está representada na Figura 3.

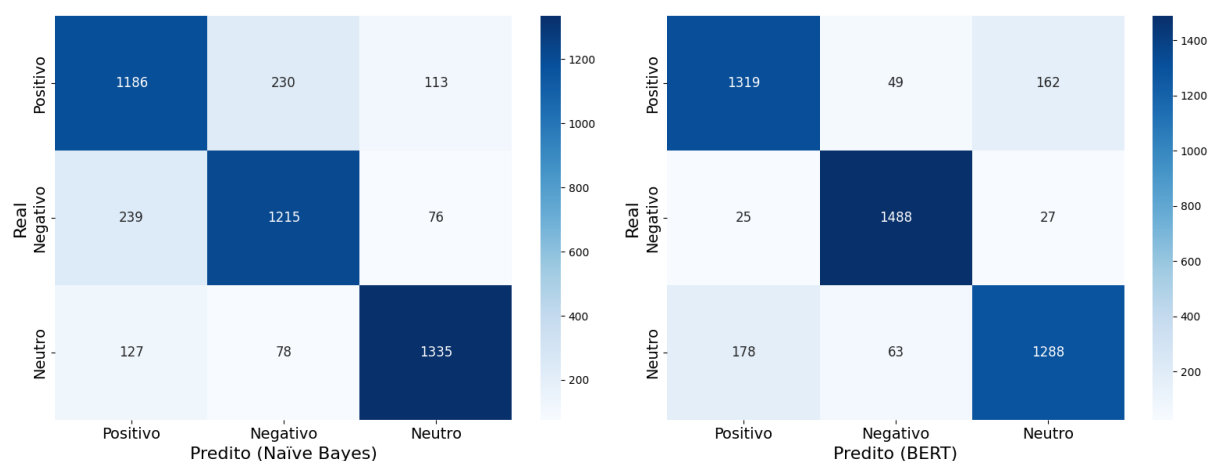


Figura 3. Comparação entre as matrizes de confusão dos modelos Naive Bayes e BERT  
Fonte: Resultados originais da pesquisa

O modelo Naïve Bayes alcançou desempenho comparativamente melhor na classificação dos sentimentos de categoria neutra, o que não se aplicou para as duas outras categorias, sugerindo uma maior dificuldade em diferenciar as classes positivo a negativo. Por outro lado, o modelo BERT apresentou um desempenho superior, com quantidades significativamente maiores de predições corretas e menos confusões entre as duas categorias, como evidenciado pelos valores elevados na diagonal principal.

No geral, a taxa de acurácia total para ambos os modelos foi de 81% para o Naïve Bayes e 89% para o BERT, o que denota uma eficácia maior nas classificações por parte do modelo de rede neural sendo consistente com a natureza do algoritmo que, como um modelo de linguagem textual profundo, é capaz de compreender com maior capacidade a complexidade e os matizes do texto.

No gráfico da Figura 4, correspondente ao Naïve Bayes, o modelo apresentou um equilíbrio entre todas as métricas para as diferentes classes de sentimentos, indicando uma consistência em seu desempenho. Essa consistência é importante, pois sugere que o modelo não favoreceu indevidamente nenhuma classe em particular durante o processo de classificação. A precisão foi de 76% para as classes positivas, e destacou-se em 88% para as neutras mostrando uma maior acurácia neste grupo. A revocação indicou uma recuperação das instâncias relevantes de maneira eficiente sendo 78% para as positivas, 79% para as negativas e 87% para as neutras. Os fatores F1 foram consistentes, com 77% e 80% para as positivas e negativas e 87% para neutras, revelando um equilíbrio entre precisão e revocação.

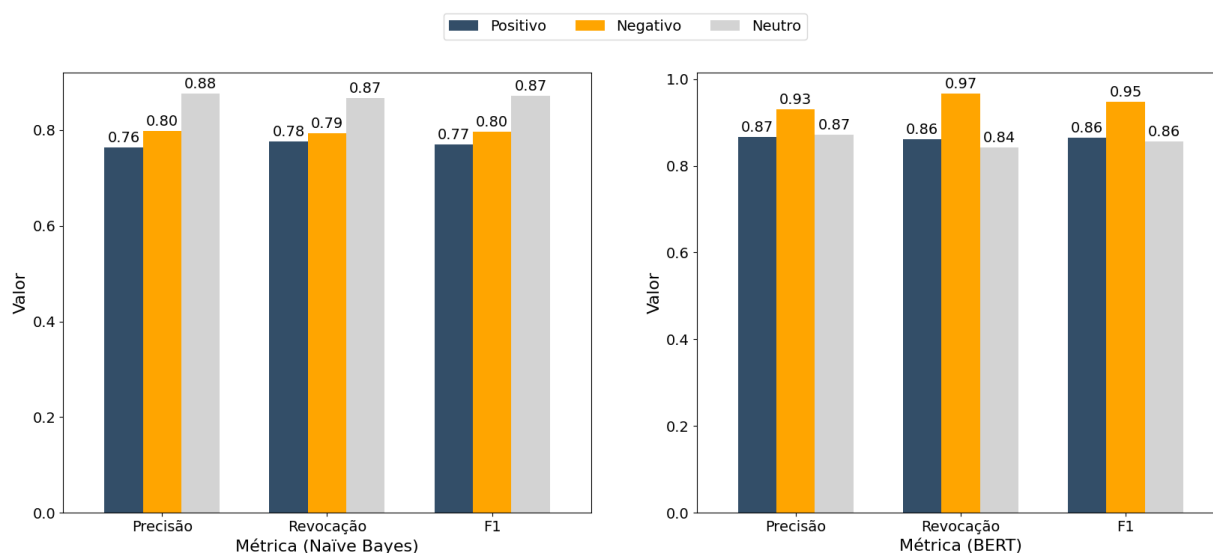


Figura 4. Comparativo de métricas de avaliação para modelos de classificação Naïve Bayes e BERT

Fonte: Resultados originais da pesquisa

Embora no cenário geral o BERT tenha um melhor desempenho do que o Naïve Bayes, o modelo mais simples exibiu uma capacidade notável nas classificações dos sentimentos neutros, o que é relevante em contextos em que essa distinção é crucial.

Por outro lado, o modelo de rede neural evidenciou uma performance superior, com uma predominância particularmente expressiva para as métricas revocação e pontuação F1, levando em consideração os sentimentos das classes positiva e negativa. O BERT com sua arquitetura avançada baseada em redes neurais e que se distingue por entender o contexto das palavras em uma frase, demonstrou uma consistência que contribui para uma taxa menor de falsos negativos e uma e uma melhor identificação global dos sentimentos.

A precisão para as avaliações negativas alcançou 93% e a revocação 97%, demonstrando alta confiabilidade e capacidade de identificação correta das avaliações positivas. O escore F1 foi de 95%, indicando um excelente balanceamento entre precisão e revocação. Para as avaliações positivas, a precisão e a revocação foram de 87% e 86%, respectivamente, com um escore F1 de 86%, evidenciando consistência. As métricas para as avaliações neutras apresentaram uma leve redução quando comparadas com as positivas especificamente para a revocação, mas ainda assim mostraram um desempenho sólido.

## **Discussão**

No estudo conduzido por Huay *et al.* (2021), análise de sentimentos foi aplicada nas linhas aéreas da Malásia com o objetivo de comparar resultados entre seis modelos diferentes. A pesquisa mostrou que as técnicas de aprendizado profundo obtiveram performances mais elevadas em relação as técnicas tidas como clássicas. O exemplo de destaque foi o BERT que obteve uma acurácia de 86,0% quando comparado com valores entre 66,5% e 77,0% dos outros modelos implementados, já o Naïve Bayes atingiu a marca dos 70,4% sendo considerado o melhor em termos de acurácia entre os modelos clássicos.

Outro estudo feito por Gheeta e Karthika (2021) e centrado nas avaliações de consumidores, categorizou comentários em positivos e negativos com o auxílio de análise de sentimentos e avaliou a performance de quatro modelos distintos. O BERT obteve uma performance melhor em todas as métricas de avaliação escolhidas (acurácias, precisão, revocação e pontuação F1), demonstrando a potencial relevância da aplicação desse potente modelo em problemas de análise de sentimentos.

Uma versão já pré-treinada do modelo BERT explorada com o intuito de classificar textos em sueco obteve uma excelente performance devido a seu vasto treino anterior em dados dessa natureza, permitindo assim uma facilidade em identificar complexidades

presentes na linguagem indicando uma menor necessidade de dados adicionais na condução de tarefas específicas (Daniel, 2020).

A aplicação de um modelo híbrido combinando uma rede convolucional e o modelo BERT para classificação de comentários de filmes extraídos da plataforma do IMDB, superou uma série de modelos selecionados. De acordo com Pawan *et al.*(2024), as métricas registradas pelo Naïve Bayes como exemplo de modelo utilizado foram de 86,8%, 87,4%, 86,3% e 86,2% para acurácia, precisão, revocação e pontuação F1 respectivamente, quando comparadas com 95,9%, 96,8%, 95,1% e 95,9% geradas pelo modelo híbrido, denotam uma importância de se misturarem estados da arte atuais na obtenção de melhorias em performance.

A comparação entre os modelos desse estudo revela uma tendência consistente com a descoberta das pesquisas anteriores, que também indicaram um potencial superior dos modelos de aprendizado profundo, como o BERT. Isso corrobora para um consenso emergente na comunidade científica sobre a robustez e a precisão atreladas ao aprendizado profundo, com o intuito de capturar a complexidade e a riqueza de expressões humanas em textos. Ao seguir a tendência, o estudo não apenas valida essas observações anteriores, como também contribui para o crescente corpo de evidências que aponta o BERT como uma ferramenta valiosa e eficaz na compreensão de sentimentos expressos nos mais diversos âmbitos.

Uma das limitações encontradas na realização do trabalho foi a limitação de recursos computacionais disponíveis para treinamento do modelo BERT, o que impôs restrições tanto com relação ao tamanho do conjunto de dados que pode ser processado, quanto em relação a complexidade dos experimentos realizados. Um segundo fator a se mencionar foi a natureza dos comentários ter sido apenas atrelada a língua inglesa, o que não necessariamente captou nuances culturais e linguísticas presentes em avaliações de filmes de outras regiões do mundo. Por fim, houve restrição monetária, referente a inviabilidade de compra da permissão de acesso aos dados por meio do serviço de “API” [Application Programming Interface] provido pela plataforma IMDB, o que potencialmente impôs limites na velocidade e volume de dados coletados.

## **Conclusão(ões) ou Considerações Finais**

Nesse estudo, foi proposto uma comparação entre um modelo clássico e de aprendizado profundo para análise de sentimentos de comentários de filmes, com ênfase nos modelos Naïve Bayes e BERT. Os resultados demonstraram que, apesar da simplicidade e rapidez do Naïve Bayes, o BERT superou significativamente em precisão, beneficiando-se da

capacidade de compreender melhor as nuances da linguagem. Este achado reforça a importância da implementação do aprendizado profundo em tarefas de processamento de linguagem natural, especialmente na análise de sentimentos onde o contexto e a subjetividade dos textos são fundamentais.

Para futuros trabalhos, propõe-se a exploração de modelos híbridos, visando otimizar tanto a precisão quanto a eficiência computacional. A investigação sobre o desempenho desses modelos em bases de dados desbalanceadas, bem como a aplicação de técnicas de processamento de linguagem natural mais avançadas, como a análise de sentimentos de tempo real e a exploração da performance considerando outros modelos de linguagem, também são direções promissoras para a continuidade do trabalho.

## **Agradecimento**

Gostaria de agradecer a todos os colegas que me auxiliaram no refinamento da escolha do tema para esse trabalho, a minha família pelo suporte e especialmente a professora Adâmara por todo trabalho de revisão, orientação e apoio durante esse período de escrita e desenvolvimento.

## **Referências**

Alexander Demidovskij, Artyom Turgaryov, Aleksei Trutnev, Marina Kazyulina, Igor Salnikov, Stanislav Pavlov. Lightweight and Elegant Data Reduction Strategies for Training Acceleration of Convolutional Neural Networks. MDPI Journal, 2023.

Alok Singh Chauhan, Ankit Kumar, Arun Pratap Srivastava, Manoj Kumar, Pawan Kumar Mall, Swapnita Srivastava, Vipul Narayan. Self-Attentive CNN+BERT: An Approach for Analysis of Sentiment on Movie Reviews Using Word Embedding. International Journal of Intelligent Systems and Applications in Engineering, 2024.

Azilawati Azizan, Masurah Mohamad, Mohammad Nasir Abdullah, Nurkhairizan Khairudin, Nurul Najwa SK Abdul Jamal. Lexicon-based sentiment analysis for movie review tweets. 1st International Conference on Artificial Intelligence and Data Sciences, 2019.

Abayomi Bello, Man-Fai Leung, Sin-Chun Ng. A BERT Framework to Sentiment Analysis of Tweets. Sensors, 2023.

Chi Wee Tan, Huay Wen Kang, Kah Kien Chye, Ong Zi Yuan. The Science of Emotion: Malaysian Airlines Sentiment Analysis using BERT Approach. International Conference on Digital Transformation and Applications, 2021.

Christine Dewi, Rung-Ching Chen, Henochi Juli Cristiano, Francesco Cauteruccio. Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database. Vietnam Journal of Computer Science, 2023.

Daniel Holmer. Context matters: Classifying Swedish texts using BERT's deep bidirectional word embeddings. Department of Computer and Information Science, 2020.

D. Karthika Renuka, M.P. Geetha. Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. International Journal of Intelligent Networks, 2021.

Fabício Bevenuto, Filipe Ribeiro, Matheus Araujo. Métodos para Análises de Sentimentos em Mídias Sociais. Curso rápido em uma conferência de webmedia, 2015.

Faisal Kevin Alkindy, Ubaid Mohamed Dahir. Utilizing Machine Learning for Sentiment Analysis of IMDB Movie Review Data. International Journal of Engineering Trends and Technology, 2023.

Haoran Li. Sentiment Analysis on Internet Movie Database (IMDb) Movie Review Dataset: Hyperparameters Tuning for Naïve Bayes Model. Department of Material Sciences and Engineering, 2023.

Harsh Sharma, Prakash Rokade, Reena Gunjan, Satyajit Pangaonkar. Sentimental Analysis of Movie Reviews Using Machine Learning. ITM Web of Conferences, 2023.

Hota HS, Sharma DK, Verma N. Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad. Data Science for COVID-19, 2021.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Google AI Language, 2018.

Like Hui, Mikhail Belkin. Evaluation Of Neural Architectures Trained with Square Loss vs Cross-Entropy in Classification Taks. International Conference on Learning Representations, 2021.

Mine Çetinkaya-Rundel, Mine Dogucu. Web Scraping in the Statistics and Data Science Curriculum: Challenges and Opportunities, Journal of Statistics and Data Science Education, 2021.

Murtadha B. Rissan, Rehab F. Hassan. Naïve-Bayes family for sentiment analysis during COVID-19 pandemic and classification tweets. Department of Computer science, University of Technology, Baghdad, Iraq, 2022.

Nassera Habbat, Hicham Nouri, Houda Anoun, Larbi Hassouni. Sentiment analysis of imbalanced datasets using BERT and ensemble stacking for deep learning. Research Laboratory on New Economy and Development, 2023.

Rosie Dunfornd, Quanrong Su, Ekraj Tamang, Abigail Wintour. The Pareto Principle. The Plymouth Student Scientist, 2024.

Tianyi Zhang, Felix Wu, Arzoo Katiyar, Kilian Q. Weinberger, Yoav Artzi. Revisiting Few-Sample BERT Fine-Tuning. International Conference on Learning Representations, 2021.

Yann Lecun, Yoshua Bengio, Geoffrey Hinton. Deep Learning. Nature, 2015.

Yankang Su, Zbigniew j. Kabala. Public perception of ChatGPT and Transformer Learning for Tweets Sentiment Analysis Using Wolfram Mathematica. MDPI Journal, 2023.

Yassine Rodani. Movie Sentiment Analysis: A Multinomial Naive Bayes-Based Approach for Assessing User and Critic Opinions. University of Haute-Alsace, FR, 2019.