
FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

**DO SUBJETIVO AO COMPUTÁVEL: DESEMPENHO DE MODELOS
CLÁSSICOS NA CLASSIFICAÇÃO DE SENTIMENTOS EM
RESENHAS CINEMATOGRÁFICAS**

**FROM THE SUBJECTIVE TO THE COMPUTABLE: PERFORMANCE
OF CLASSICAL MODELS IN SENTIMENT CLASSIFICATION OF
FILM REVIEWS**

Marcos Ribeiro dos Santos Filho^{1*}
Bruno Santos de Lima^{2**}

Resumo

A análise de sentimentos tem se consolidado como uma ferramenta essencial para interpretar opiniões expressas em linguagem natural, especialmente em ambientes digitais. Este trabalho investiga o desempenho de algoritmos clássicos de aprendizado supervisionado na classificação de sentimentos em resenhas cinematográficas, buscando responder se esses modelos são capazes de interpretar adequadamente a polaridade emocional presente em textos opinativos. Utilizou-se o IMDB Dataset of 50K Movie Reviews, amplamente empregado como benchmark em estudos de Processamento de Linguagem Natural, aplicando técnicas de pré-processamento textual e vetorização por meio de TF-IDF. Foram avaliados os algoritmos K-Nearest Neighbors (KNN), Random Forest, Naive Bayes (ComplementNB, MultinomialNB e BernoulliNB) e Linear Support Vector Classifier (Support Vector Machine), em múltiplas configurações e tamanhos de amostra. Os resultados demonstraram que o Support Vector Machine apresentou o melhor desempenho geral, atingindo F1-Score de até 0,8725, enquanto os modelos Naive Bayes se destacaram pelo baixo custo computacional e resultados consistentes. O estudo conclui que algoritmos clássicos são capazes de realizar análise de sentimentos com elevada precisão quando submetidos a adequado pré-processamento e otimização, contribuindo para aplicações como sistemas de recomendação, monitoramento de opinião pública e processamento de grandes volumes de dados textuais.

Palavras-chave: Análise de Sentimentos; Aprendizado de Máquina; Processamento de

^{1*} Aluno do curso superior de Tecnologia em Análise e Desenvolvimento de Sistemas da Faculdade de Tecnologia de Presidente Prudente. E-mail: marcos.santos234@fatec.sp.gov.br

^{2**} Professor orientador Me. em Ciência da Computação, da Faculdade de Tecnologia de Presidente Prudente. E-mail: bruno.lima105@fatec.sp.gov.br.

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

Linguagem Natural; Algoritmos de Classificação.

Abstract

Sentiment analysis has become an essential tool for interpreting opinions expressed in natural language, especially in digital environments. This study investigates the performance of classical supervised learning algorithms in sentiment classification of film reviews, seeking to determine whether these models are capable of adequately interpreting the emotional polarity present in opinionated texts. The IMDB Dataset of 50K Movie Reviews, widely used as a benchmark in Natural Language Processing research, was employed along with preprocessing techniques and TF-IDF vectorization. The evaluated algorithms include K-Nearest Neighbors (KNN), Random Forest, Naive Bayes (ComplementNB, MultinomialNB and BernoulliNB) and Support Vector Machine (SVM), applied across multiple configurations and sample sizes. The results indicate that SVM achieved the best overall performance, reaching an F1-Score of up to 0.8725, while Naive Bayes models stood out for their low computational cost and consistent performance. The study concludes that classical algorithms are capable of performing sentiment analysis with high accuracy when appropriately optimized and combined with robust preprocessing, contributing to applications such as recommendation systems, public opinion monitoring and large-scale text processing.

Keywords: *Sentiment Analysis; Machine Learning; Natural Language Processing; Classification Algorithms.*

1. INTRODUÇÃO

O avanço das técnicas de Inteligência Artificial (IA) e de Processamento de Linguagem Natural (PLN) tem permitido que máquinas interpretem, ainda que parcialmente, as emoções e opiniões humanas expressas em textos. A análise de sentimentos, em particular, busca identificar a polaridade de sentimento em um texto, atribuindo classificação positiva, negativa ou neutra a partir de padrões linguísticos e estatísticos (LOURES, 2024; LUDERMIR et al., 2021). Essa capacidade de traduzir sentimentos em padrões mensuráveis desperta não apenas interesse técnico, mas também uma curiosidade mais ampla: até que ponto é possível compreender o subjetivo por meio do objetivo?

O presente trabalho investiga como os algoritmos clássicos de aprendizado de máquina podem interpretar expressões humanas de gosto e emoção em resenhas de filmes que

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

oferecem um contexto consolidado para análise científica, por dois motivos principais: (i) a existência de bases de dados amplas, rotuladas e padronizadas para críticas cinematográficas, o que favorece a comparabilidade e a reprodutibilidade da pesquisa; e (ii) a maior aplicabilidade interdisciplinar dessas resenhas, frequentemente utilizadas em estudos de mídia, mercado, recomendação de conteúdo e comportamento de audiência digital (PINHO et al., 2024; SIQUEIRA, 2024).

Na literatura, diversos estudos têm explorado a aplicação de algoritmos de aprendizado de máquina à análise de sentimentos em textos. Pinho et al. (2024), por exemplo, aplicam técnicas de IA para classificar fuga ao tema em redações, evidenciando o potencial de modelos clássicos para tarefas de classificação textual. Já Pfitscher (2023) discute o uso do dataset *IMDB* em atividades educacionais de programação, mostrando como algoritmos supervisionados podem ser empregados para análise de sentimentos em críticas de filmes. Esses trabalhos evidenciam um cenário em que, apesar da disseminação de abordagens baseadas em deep learning, os algoritmos clássicos ainda ocupam lugar central nas aplicações de PLN.

Esse contexto conduz à primeira questão de pesquisa que orienta este estudo: *Qual algoritmo de aprendizado de máquina se mostra mais eficaz na identificação de sentimentos em resenhas de filmes?* A literatura apresenta resultados variados, explorando variações de técnicas de pré-processamento e das métricas utilizadas, o que torna relevante a comparação sistemática entre modelos (LUDERMIR et al., 2021; PINHO et al., 2024).

Entretanto, mais do que identificar o “melhor” algoritmo em termos de métricas, interessa compreender se os algoritmos clássicos de classificação supervisionada são capazes de analisar o sentimento humano no contexto das resenhas de filmes, marcado por avaliações explícitas, julgamentos de valor e traços afetivos expressos por humanos. Estudos como os de Loures (2024) e Soares (2023) indicam que modelos como *Naive Bayes*, *Support Vector Machines* e *Random Forest* apresentam bom desempenho em tarefas de análise de opiniões, mas apontam a necessidade de investigações que considerem diferentes tamanhos de conjuntos de dados e ajustes de hiperparâmetros.

Além disso, o rápido avanço de modelos de aprendizado profundo levanta outra importante questão: *o desempenho dos algoritmos clássicos permanece eficiente frente às novas abordagens?* Ludermir et al. (2021) argumentam que, embora redes neurais profundas

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

ofereçam ganhos expressivos em alguns domínios, algoritmos clássicos continuam competitivos quando bem parametrizados e aliados a representações adequadas dos dados, especialmente pelo menor custo computacional e maior interpretabilidade. Essa discussão reforça a importância de reavaliar o papel desses modelos em cenários atuais de análise de sentimentos.

Dessa forma, o foco deste estudo passou a avaliar se os algoritmos clássicos de classificação supervisionada conseguem analisar o sentimento humano no contexto de resenhas de filmes, considerando desempenho, estabilidade e custo computacional. Assim, o estudo busca responder a principal questão de pesquisa: “*Os algoritmos clássicos de classificação conseguem analisar o sentimento humano no contexto de resenhas de filmes?*”, e de modo complementar às questões de pesquisa: “*Qual algoritmo apresenta o melhor desempenho nesse cenário?*” e “*O desempenho desses algoritmos ainda pode ser considerado eficiente diante de modelos mais robustos?*”.

Para tanto, empregou-se o *IMDB Dataset of 50K Movie Reviews*, um dos *benchmarks* mais utilizados em tarefas de análise de sentimentos, composto por 50 mil críticas de filmes igualmente distribuídas entre as classes positiva e negativa. Estudos como: *Sentiment Analysis on IMDB Review Dataset (2023)*, reforçam a centralidade dessa base em experimentos comparativos de modelos de aprendizado de máquina, o que contribui para a validação externa dos resultados obtidos neste trabalho.

Neste contexto, o presente estudo propõe uma análise experimental e quantitativa de diferentes algoritmos — *K-Nearest Neighbors (KNN)*, *Support Vector Machines (SVM)*, *Random Forest* e *Naive Bayes (ComplementNB, MultinomialNB, BernoulliNB)* — aplicados à vetorização de textos via *TF-IDF*, em diferentes tamanhos de dataset e condições de otimização de hiperparâmetros. Busca-se, portanto, identificar qual modelo oferece o melhor equilíbrio entre desempenho, estabilidade e eficiência computacional na tarefa de reconhecimento de sentimentos humanos em resenhas cinematográficas.

Este trabalho, assim, reflete o esforço de articular embasamento teórico e investigação empírica, aproximando linguagem, emoção e tecnologia. Ao avaliar o desempenho de algoritmos clássicos em um corpus amplamente utilizado, pretende-se contribuir para o debate sobre os limites e as potencialidades da IA na interpretação de opiniões humanas em contextos reais de uso.

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

Este artigo está organizado como segue: Na seção 2 é apresentada a fundamentação teórica acerca de Aprendizado de Máquina e principais algoritmos clássicos para classificação de sentimentos. Na seção 3 é descrita a metodologia empregada. Na seção 4 são apresentados e discutidos os resultados do estudo. Por fim, na seção 5 são destacadas as considerações finais.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Aprendizado de máquina

O aprendizado de máquina (Machine Learning – ML) é um dos pilares da Inteligência Artificial contemporânea e consiste no desenvolvimento de modelos capazes de identificar padrões, aprender comportamentos e tomar decisões com base em dados. Paixão e Campos (2022) descrevem o ML como a construção de estruturas matemáticas e estatísticas que permitem ao sistema aperfeiçoar seu desempenho à medida que recebe novas informações, aproximando-se do conceito de aprendizado humano na perspectiva computacional.

Barroso (2024) ressalta que o avanço do poder computacional e o aumento do volume de dados digitais impulsionaram a aplicação do ML em tarefas complexas, como análise textual, vigilância epidemiológica, detecção de anomalias e apoio à tomada de decisões em diferentes setores. Menezes (2024) destaca que essas aplicações podem ser organizadas em técnicas supervisionadas, não supervisionadas e semi-supervisionadas, escolhidas conforme a disponibilidade de rótulos e o objetivo da modelagem. Em problemas binários, como a classificação de sentimentos em positivo ou negativo, o aprendizado supervisionado é a abordagem mais adequada.

No contexto brasileiro, o uso de ML tem crescido em áreas como educação, saúde, economia e governança. Rocha (2025) aponta seu emprego na previsão de desempenho estudantil, enquanto Schaab et al. (2024) demonstram resultados expressivos na classificação de sintomas psicológicos entre universitários, evidenciando a capacidade do ML de lidar com dados subjetivos e complexos. Esses cenários são semelhantes ao da análise de sentimentos, em que emoções e opiniões são modeladas a partir de textos.

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

Além disso, o ML é capaz de lidar com alta dimensionalidade, característica dos dados textuais vetorizados, e enfrenta desafios como overfitting, underfitting, balanceamento de classes e regularização (PAIXÃO; CAMPOS, 2022; MENEZES, 2024). Dessa forma, o aprendizado de máquina fornece a base teórica e computacional para este estudo, oferecendo instrumentos para representar textos opinativos, organizar informações emocionais e reconhecer padrões afetivos em resenhas de filmes.

2.2 Aprendizado supervisionado e problemas de classificação

O aprendizado supervisionado é uma das abordagens mais consolidadas do ML, pois utiliza dados rotulados para ensinar o modelo a reconhecer padrões e realizar previsões futuras. Segundo Paixão e Campos (2022), o treinamento ocorre a partir de pares de entrada-saída, permitindo que o algoritmo aprenda a relação entre características dos dados e suas classes e generalize esse conhecimento para novos exemplos.

Barroso (2024) observa que avanços recentes tornaram essas técnicas centrais em decisões automatizadas de alto impacto, como análises judiciais, políticas públicas e interpretação de grandes volumes de dados textuais. Menezes (2024) destaca que o aprendizado supervisionado é especialmente útil quando a tarefa envolve categorizar dados em classes bem definidas, como diagnóstico médico, análise de crédito, detecção de comportamentos anômalos e classificação de textos, desde que haja bases estruturadas e rotuladas.

Estudos brasileiros ilustram esse movimento. Gomes e Balaniuk (2024) mostraram que técnicas de classificação identificam empresas de fachada e fraudes fiscais com alta precisão, enquanto Rossati et al. (2025) aplicaram classificadores na detecção de pragas agrícolas em imagens multiespectrais. Schaab et al. (2024) evidenciam ainda o potencial dessas técnicas na identificação de padrões subjetivos e emocionais em dados psicológicos, aproximando-se do cenário da análise de sentimentos.

Em problemas textuais, desafios como alta dimensionalidade, ruído e desbalanceamento de classes tornam indispensáveis métricas como *Acurácia*, *Precisão*, *Revocação* e *F1-Score* (MENEZES, 2024). No contexto deste trabalho, o aprendizado supervisionado constitui o núcleo metodológico para treinar modelos capazes de distinguir a

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

polaridade emocional de resenhas de filmes e avaliar se algoritmos clássicos de classificação conseguem interpretar emoções humanas expressas em linguagem natural.

2.3 Evolução, finalidade e importância dos algoritmos de classificação

Os algoritmos de classificação acompanham a evolução da estatística aplicada e da ciência de dados, consolidando-se como uma das formas mais tradicionais de modelagem preditiva. Paixão e Campos (2022) apontam que o aumento da disponibilidade de dados estruturados e não estruturados exigiu técnicas capazes de automatizar a categorização de informações em larga escala.

Barroso (2024) ressalta que esses algoritmos passaram a ocupar papel estratégico em áreas que demandam decisões rápidas e precisas, como sistemas de recomendação, diagnósticos assistidos, monitoramento de riscos e interpretação de linguagem natural. Sua finalidade é identificar padrões que permitam distinguir grupos com base em atributos relevantes, inclusive em ambientes com milhares de variáveis, como textos vetorizados (MENEZES, 2024).

Aplicações nacionais reforçam essa relevância. Schaab et al. (2024) utilizaram classificadores na detecção de sintomas psicológicos; Rossati et al. (2025) mostraram sua eficácia em dados multiespectrais na agricultura; e Gomes e Balaniuk (2024) destacaram seu uso na detecção de fraudes fiscais. No campo textual, a classificação permite transformar expressões linguísticas em categorias emocionais, como positivo ou negativo, organizando grandes volumes de avaliações, resenhas e comentários online (FERREIRA; FREITAS, 2020; ROCHA, 2025).

Assim, a evolução e a importância dos algoritmos de classificação justificam sua escolha neste estudo, uma vez que possibilitam investigar, de forma sistemática, se aspectos afetivos da linguagem podem ser representados computacionalmente em resenhas cinematográficas.

2.4 Algoritmos de classificação utilizados

Os algoritmos de classificação empregados neste trabalho representam abordagens consolidadas na literatura de aprendizado de máquina. Paixão e Campos (2022) destacam que

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

tais modelos foram desenvolvidos para aprender relações entre variáveis com base em exemplos rotulados, enquanto Barroso (2024) enfatiza sua difusão em setores como saúde, justiça, finanças e administração pública. Em comum, esses algoritmos buscam construir funções capazes de atribuir rótulos a novas instâncias, mesmo em cenários com alta dimensionalidade, ruído e desbalanceamento (MENEZES, 2024).

No contexto da análise textual, Rocha (2025) e Ferreira e Freitas (2020) salientam que a combinação entre vetorização de textos e classificadores supervisionados é fundamental para interpretar grandes volumes de dados opinativos. A seguir, são apresentados, de forma sucinta, os principais algoritmos utilizados neste estudo.

2.4.2 K-Nearest Neighbors

O K-Nearest Neighbors (KNN) é um algoritmo baseado em similaridade, que classifica novas instâncias considerando as classes dos k vizinhos mais próximos no espaço de atributos. Menezes (2024) o caracteriza como um método não paramétrico, em que o esforço computacional se concentra na etapa de predição. No Brasil, tem sido utilizado em áreas como detecção de fraudes e reconhecimento de padrões (GOMES; BALANIUK, 2024).

Apesar de sua simplicidade, o KNN é sensível à alta dimensionalidade, situação em que a maldição da dimensionalidade torna as distâncias menos informativas e prejudica o desempenho, especialmente em textos representados por TF-IDF (MENEZES, 2024). Ainda assim, o algoritmo é útil como modelo de referência em experimentos comparativos.

2.4.3 Support Vector Machines

As SVMs são amplamente reconhecidas pela capacidade de separar classes por meio de hiperplanos que maximizam a margem entre elas, favorecendo a generalização (MENEZES, 2024). Em dados textuais, versões lineares como o SVM são particularmente adequadas, pois lidam bem com matrizes esparsas e de alta dimensionalidade (ROCHA, 2025).

Barroso (2024) e Schaab et al. (2024) destacam o bom desempenho do SVM em problemas que envolvem dados complexos e subjetivos, como sinais emocionais e comportamentais. Na análise de sentimentos, a combinação entre TF-IDF e SVM permite que

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

o modelo identifique padrões linguísticos associados à polaridade textual (FERREIRA; FREITAS, 2020), justificando sua inclusão como principal algoritmo deste estudo.

2.4.4 Random Forest

O Random Forest é um método de ensemble que combina múltiplas árvores de decisão construídas a partir de amostras aleatórias de dados e atributos. Essa estratégia reduz o sobreajuste e aumenta a estabilidade do modelo (MENEZES, 2024; PAIXÃO; CAMPOS, 2022). Estudos nacionais evidenciam sua versatilidade em contextos diversos, como detecção de pragas agrícolas em imagens multiespectrais (ROSSATI et al., 2025) e identificação de fraudes fiscais (GOMES; BALANIUK, 2024).

Embora apresente bom desempenho em tarefas de classificação, o Random Forest tende a ser mais custoso computacionalmente, especialmente em bases muito grandes, razão pela qual, neste trabalho, é utilizado como modelo de comparação com algoritmos lineares e probabilísticos.

2.4.5 Naive Bayes

Os classificadores Naive Bayes formam uma família de modelos probabilísticos baseados no Teorema de Bayes, que assumem independência condicional entre atributos. Menezes (2024) destaca sua simplicidade, baixo custo computacional e boa capacidade de generalização. Em dados textuais, esse tipo de modelo é tradicionalmente empregado na análise de sentimentos, uma vez que consegue explorar padrões de frequência de termos associados à polaridade (FERREIRA; FREITAS, 2020).

No cenário brasileiro, Rocha (2025) e Schaab et al. (2024) relatam resultados consistentes do Naive Bayes em tarefas de classificação educacional e de saúde mental. As variações MultinomialNB, BernoulliNB e ComplementNB são adequadas a diferentes formas de representação textual e lidam de maneira eficiente com bases balanceadas e de grande porte, o que justifica sua adoção neste estudo como abordagem probabilística de referência.

2.5 Processamento de Linguagem Natural e Representação Textual

O Processamento de Linguagem Natural (PLN) integra linguística, ciência da computação e inteligência artificial com o objetivo de permitir que máquinas processem e

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

analisem textos escritos ou falados. Silva e Moreira (2020) explicam que o PLN transforma a linguagem humana, subjetiva e contextual, em estruturas passíveis de tratamento matemático, viabilizando tarefas como classificação de textos, sumarização automática e análise de opiniões.

Com a intensificação das interações digitais, resenhas de filmes, avaliações de produtos e comentários em redes sociais tornaram-se fontes importantes para estudos computacionais sobre linguagem (ROCHA, 2025). Para que algoritmos de aprendizado de máquina possam interpretar esses textos, é necessário convertê-los em vetores numéricos, etapa conhecida como representação vetorial. Ferreira e Freitas (2020) destacam que a qualidade dessa representação influencia diretamente o desempenho dos classificadores, pois determina o quanto nuances semânticas e emocionais são preservadas.

Entre as técnicas de representação, o Term Frequency–Inverse Document Frequency (TF-IDF) se destaca por ponderar a importância dos termos considerando sua frequência no documento e raridade no corpus (SILVA; MOREIRA, 2020). Em análise de sentimentos, essa estratégia é útil para destacar palavras associadas à polaridade emocional, reduzindo o peso de termos muito comuns. Estudos brasileiros indicam que o TF-IDF apresenta desempenho superior a representações puramente baseadas em contagem e mantém a interpretabilidade dos modelos, permitindo identificar quais termos contribuem para as decisões do algoritmo (FERREIRA; FREITAS, 2020; ROCHA, 2025).

Técnicas de pré-processamento, como remoção de pontuação, normalização e remoção de stopwords, complementam esse processo ao reduzir ruídos e padronizar a linguagem (SILVA; MOREIRA, 2020). No contexto deste trabalho, o uso combinado de PLN, pré-processamento e TF-IDF é fundamental para representar de forma consistente as resenhas cinematográficas, possibilitando que os algoritmos clássicos de classificação avaliem, com precisão, a polaridade dos sentimentos expressos nos textos.

3. METODOLOGIA

3.1 Objetivo geral

O presente estudo tem como objetivo avaliar se algoritmos clássicos de aprendizado de máquina são capazes de analisar o sentimento humano em resenhas de filmes,

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

identificando aquele que apresenta melhor equilíbrio entre desempenho preditivo e custo computacional.

3.2 Objetivos específicos

Visando atingir o objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- Comparar o desempenho de diferentes algoritmos clássicos de aprendizado supervisionado na tarefa de classificação de sentimentos em resenhas cinematográficas;
- Analisar o comportamento desses modelos em diferentes escalas de dados;
- Avaliar o impacto do ajuste de hiperparâmetros nas métricas de desempenho.

3.3 Procedimentos metodológicos

A pesquisa caracteriza-se como experimental e quantitativa, uma vez que envolve a manipulação de variáveis (como o tipo de algoritmo, o tamanho dos dados e os hiperparâmetros) e a mensuração objetiva do desempenho dos modelos por meio de métricas.

Segundo Souza et al. (2021), estudos experimentais permitem controlar fatores e observar efeitos sobre variáveis dependentes, enquanto para Pinho et al. (2024) a abordagem quantitativa baseia-se na análise numérica de dados e no uso de métricas estatísticas para comparar métodos. O processo metodológico conduzido por meio de experimento controlado é ilustrado na Figura 1:

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

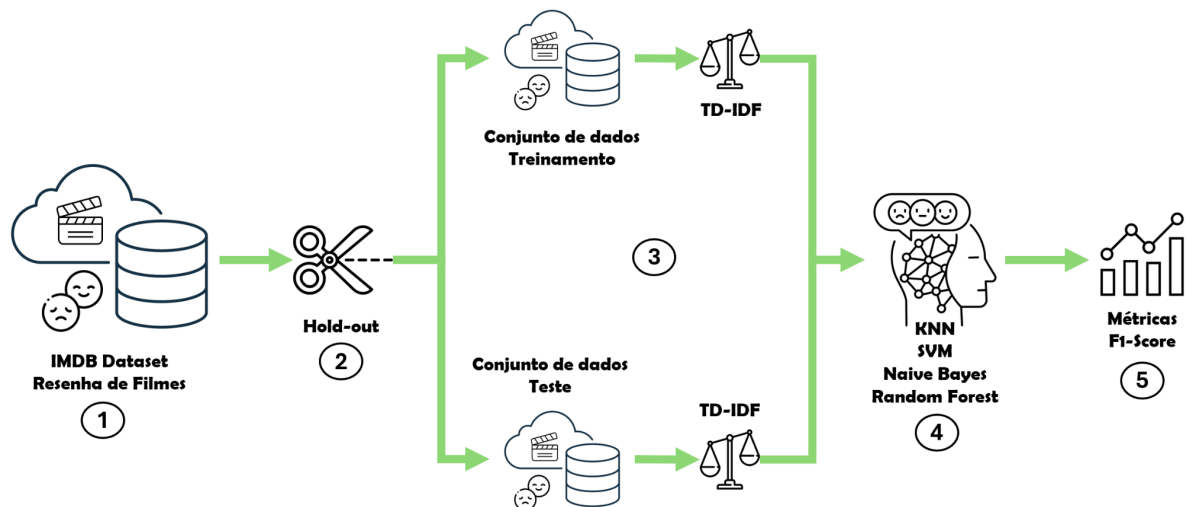


Figura 1 - Esquema do experimento controlado de avaliação de algoritmos clássicos de aprendizado de máquina

Fonte: elaborada pelo autor.

3.3.1 Conjunto de dados

O conjunto de dados utilizado na aplicação do experimento é composto por resenhas de filmes classificadas por humanos em polaridades de sentimentos positivos e negativos (Figura 1 - Etapa 1). O conjunto de dados está disponibilizado publicamente na plataforma *Kaggle* no repositório *IMDB Dataset of 50K Movie Reviews - Large Movie Review Dataset* - <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>. Este conjunto de dados em larga escala possui 50.000 avaliações de filmes e fornece o contexto adequado para análise de sentimentos em língua inglesa aplicadas ao domínio cinematográfico.

A escolha do conjunto de dados *IMDB Dataset of 50K Movie Reviews* justifica-se pela relevância e pela ampla utilização dessa base em pesquisas recentes sobre análise de sentimentos aplicada a textos de opinião. Como o conjunto possui 50.000 resenhas de filmes publicadas por usuários da plataforma *Internet Movie Database (IMDB)*, igualmente distribuídas entre as classes positiva e negativa, assegura equilíbrio estatístico e representatividade das amostras utilizadas.

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

As resenhas presentes na base refletem opiniões humanas, redigidas de forma espontânea e em linguagem natural, oferecendo um contexto realista para avaliar a capacidade dos algoritmos clássicos de classificação em interpretar e modelar aspectos subjetivos do sentimento humano.

De acordo com Soares (2023) e Loures (2024), o *IMDB* tem sido amplamente empregado como base de referência (*benchmark*) em estudos contemporâneos de Processamento de Linguagem Natural (PLN), por conter textos ricos em nuances emocionais e diversidade semântica. Um estudo de 2023 intitulado de *Sentiment Analysis on IMDB Review Dataset* reforça essa importância ao demonstrar que essa base continua sendo utilizada em pesquisas que comparam algoritmos de aprendizado de máquina e técnicas de otimização para análise de sentimentos.

Dessa forma, a adoção do *IMDB Dataset of 50K Movie Reviews* neste estudo possibilita uma investigação atualizada e consistente sobre se os algoritmos clássicos de aprendizado supervisionado: *Naive Bayes*, *SVMs*, *Random Forest* e *K-Nearest Neighbors*, são capazes de analisar o sentimento humano no contexto das resenhas cinematográficas, oferecendo subsídios empíricos para compreender as potencialidades e limitações dessas abordagens em dados textuais reais.

3.3.2 Pré-processamento dos dados:

Aplicou-se a técnica Hold-out (Figura 1 - Etapa 2) para dividir o conjunto de dados original em conjunto de treinamento (80 %) e conjunto de teste (20 %). Os rótulos das classes foram codificados por meio de *LabelEncoder*. Os textos foram vetorizados utilizando *CountVectorizer* e posteriormente *TF-IDF (Term Frequency–Inverse Document Frequency)*, que transformam o conteúdo textual em representações numéricas (Figura 1 - Etapa 3).

3.3.4 Algoritmos de Classificação e Otimização de hiperparâmetros:

Os seguintes algoritmos clássicos de classificação foram utilizados para análise de sentimentos em resenhas cinematográficas: *Support Vector Machine (SVM)*; *Naive Bayes*, considerando as variantes *Complement*, *Multinomial* e *Bernoulli*; *K-Nearest Neighbors (KNN)*; *Random Forest* (Figura 1 - Etapa 4). A seguir são apresentadas as configurações dos

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

hiperparâmetros considerando as otimizações obtidas de cada algoritmo de classificação de sentimentos:

- **Support Vector Machine:** os parâmetros $max_iter=100000$, $dual=False$ e $tol=1e-4$ foram ajustados para garantir convergência, eliminando avisos de não convergência observados nas execuções anteriores.
- **Naive Bayes:** o parâmetro $alpha=0.5$ foi utilizado para suavização, proporcionando maior estabilidade das probabilidades.
- **KNN:** foi empregada a métrica de distância cosseno, buscando mitigar o impacto da alta dimensionalidade dos vetores *TF-IDF*.
- **Random Forest:** ajustes no número de estimadores e profundidade máxima foram realizados, com o objetivo de equilibrar desempenho e tempo de execução.

3.3.5 Métricas de Verificação de Aprendizagem e Análise dos resultados:

Os algoritmos de Aprendizado de Máquina precisam ser avaliados com o objetivo de verificar sua eficácia de treinamento para classificar sentimentos em resenhas de filmes. Conforme dos Santos (2025), a métrica *F1-Score* (Equação 1) é o principal mecanismo de verificação de aprendizagem. O *F1-Score*, considera em sua média harmônica as métricas de Precisão (Equação 2) e Revocação (Equação 3), também utilizadas no experimento controlado em conjunto com a métrica de Acurácia (Equação 4).

As métricas de Precisão, Revocação e Acurácia são calculadas a partir da comparação entre a classificação manual e a classificação automática. Para isso, utilizam-se os quantificadores Verdadeiros Positivos (VP), Falsos Positivos (FP), Verdadeiros Negativos (VN) e Falsos Negativos (FN). A obtenção de uma alta precisão indica que o algoritmo de classificação apresenta alta confiabilidade, pois raramente classifica um sentimento incorretamente como positivo ou negativo. Já uma alta revocação indica que o algoritmo é eficaz em capturar todas as resenhas que realmente pertencem a uma determinada classe, ou seja, evita perder resenhas positivas classificando-as como negativas (ou vice-versa). Por fim, uma acurácia elevada indica que o algoritmo consegue acertar a polaridade dos sentimentos tanto positivos quanto negativos de forma geral, refletindo um bom desempenho global.

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

$$F1Score = 2 \cdot \frac{Precisão \cdot Revocação}{Precisão + Revocação} \quad (\text{Equação 1})$$

$$Precisão = \frac{VP}{VP + FP} \quad (\text{Equação 2})$$

$$Revocação = \frac{VP}{VP + FN} \quad (\text{Equação 3})$$

$$Acurácia = \frac{VP + VN}{VP + FN + VN + FP} \quad (\text{Equação 4})$$

Os algoritmos foram submetidos a um mesmo processo de avaliação (Figura 1 - Etapa 5), implementado por meio de uma função denominada `eval_model()`. Essa função calculou as métricas Acurácia, Precisão, Revocação e *F1-Score*, utilizando média binary (duas classes). Foram testados diferentes tamanhos de dataset (10k, 26k, 50k e completo), repetindo-se os experimentos após ajustes nos hiperparâmetros.

As métricas obtidas foram registradas e organizadas em uma tabela comparativa de *F1-Score*, o que permitiu avaliar o comportamento dos modelos em diferentes tamanhos de dataset e sob condições de otimização. A análise concentrou-se em responder à questão central: os algoritmos clássicos de classificação conseguem analisar o sentimento humano em resenhas de filmes? Para tanto, considerou-se o desempenho, a escalabilidade e a eficiência computacional dos modelos. Os procedimentos adotados asseguraram a reprodutibilidade dos experimentos e forneceram subsídios para a discussão crítica dos resultados, apresentada na seção subsequente.

3.3.6 Configuração Experimental e Reprodutibilidade:

Os experimentos foram realizados no ambiente de desenvolvimento *Visual Studio Code (VS Code)*, utilizando a linguagem *Python 3.10*, escolhida por sua elevada adoção em pesquisa científica e por dispor de bibliotecas sólidas para análise de dados e aprendizado de máquina. Conforme Rodrigues (2025), o *Python* consolida-se como uma das principais linguagens para Inteligência Artificial devido à facilidade de integração entre suas ferramentas.

Para o desenvolvimento dos experimentos foram empregadas as bibliotecas *Pandas*, *Numpy*, *scikit-learn* e *tqdm*. Conforme Ludermir et al. (2021), tais bibliotecas oferecem funcionalidades eficientes para a manipulação de dados, implementação de algoritmos de

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

aprendizado de máquina. Visando à reprodutibilidade deste estudo, a implementação dos algoritmos de aprendizado de máquina clássicos, bem como o ambiente do experimento está disponível no repositório público vinculado a este trabalho no *GitHub*: <https://github.com/marcosrsfilho/tcc-analise-sentimentos-imdb>

4. RESULTADOS E DISCUSSÕES

Os experimentos realizados permitiram avaliar o desempenho de diferentes algoritmos clássicos de aprendizado de máquina na tarefa de análise de sentimentos em resenhas de filmes, utilizando o conjunto de dados *IMDB Dataset of 50K Movie Reviews*. As métricas analisadas foram Acurácia, Precisão, Revocação e F1-Score, sendo esta última adotada como principal medida de desempenho por representar o equilíbrio entre precisão e abrangência das classificações (FERREIRA, 2023).

Na Tabela 1 é sumarizado o desempenho consolidado dos algoritmos em diferentes tamanhos de amostras e configurações de otimização.

Tabela 1 – Comparação final do desempenho dos algoritmos (F1-Score)

Dataset Configuração	KNN	Support Vector Machine	Complement NB	MultinomialNB	Bernoulli NB	Random Forest
10k	0,6386	0,8420	0,8219	0,8220	0,8386	0,8491
26k	0,6533	0,8632	0,8353	0,8353	0,8465	0,8552
50k	0,6630	0,8725	0,8400	0,8400	0,8492	0,8600
Dataset completo (Linear padrão)	0,6623	0,8721	0,8460	0,8460	0,8549	0,8620
Dataset completo (Linear otimizado)	0,6524	0,8725	0,8459	0,8459	0,8553	0,8523

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

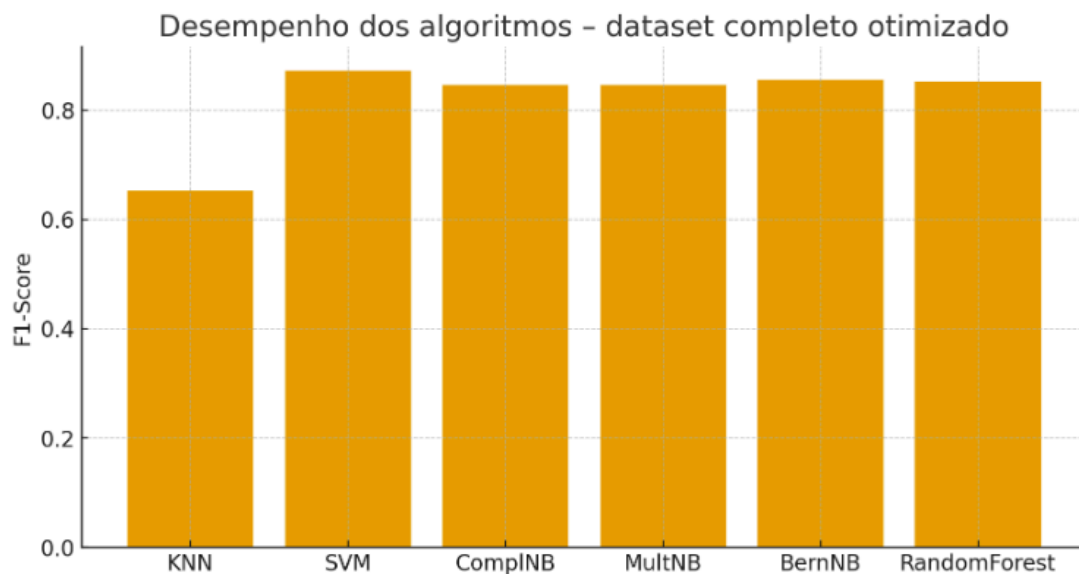


Figura 2 - Desempenho dos algoritmos clássicos em resenhas cinematográficas considerando a métrica F1-Score.

Fonte: elaborada pelo autor.

4.2 Análise dos resultados

Observa-se que o modelo *SVM* apresentou o melhor desempenho geral em todos os cenários testados, alcançando *F1-Score* de 0,8725 no dataset completo otimizado (Figura 2). Esse resultado corrobora o que aponta Ludermir et al. (2021) ao afirmar que modelos lineares tendem a apresentar desempenho superior em tarefas de classificação de texto de alta dimensionalidade, devido à sua capacidade de generalização e separação eficiente de classes em espaços vetoriais.

O *Random Forest* apresentou desempenho competitivo (*F1-Score* entre 0,849 e 0,862), porém com custo computacional elevado, o que restringe sua aplicação em sistemas de tempo real ou de grande volume de dados. Conforme Ferreira (2023), esse tipo de algoritmo, por combinar múltiplas árvores de decisão, tende a oferecer boa precisão, mas em detrimento da eficiência temporal.

Os modelos baseados em *Naive Bayes* (*ComplementNB*, *MultinomialNB* e *BernoulliNB*) apresentaram desempenho médio entre 0,83 e 0,85, com tempo de execução consideravelmente inferior, o que os torna adequados para sistemas de baixa latência. Esse

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

comportamento confirma a observação de Soares (2023) de que o *Naive Bayes* é um modelo estatístico de baixo custo computacional e que se mostra eficiente em cenários textuais balanceados.

O modelo *K-Nearest Neighbors* (KNN) obteve os piores resultados, com *F1-Score* entre 0,63 e 0,66, mantendo desempenho instável mesmo após o uso da média distância do cosseno. Esse comportamento é esperado, já que, conforme Loures (2024), algoritmos baseados em vizinhança sofrem degradação de desempenho em espaços vetoriais de alta dimensionalidade, como os gerados pelo método TF-IDF.

Além disso, os resultados confirmam o efeito positivo do aumento do tamanho do dataset, visto que todos os modelos apresentaram ganhos graduais de desempenho à medida que o número de amostras cresceu. O *SVM*, em particular, mostrou-se mais sensível à otimização de hiperparâmetros, destacando-se após o ajuste dos parâmetros *max_iter*, *dual* e *tol*. Essa melhoria ilustra a importância da calibração dos algoritmos supervisionados para garantir convergência e estabilidade numérica (PINHO et al., 2024).

4.3 Discussão teórica e implicações

Os resultados obtidos reforçam que os algoritmos clássicos de aprendizado supervisionado ainda apresentam desempenho expressivo em tarefas de análise de sentimentos, mesmo frente a abordagens mais recentes baseadas em redes neurais profundas. Estudos como *Sentiment Analysis on IMDB Review Dataset* (2023) e Loures (2024) também identificam que, quando otimizados, modelos lineares e probabilísticos podem atingir métricas superiores a 0,85 em bases de resenhas cinematográficas.

Do ponto de vista metodológico, os achados demonstram que a análise de sentimentos pode ser adequadamente modelada por algoritmos tradicionais, desde que aplicados a um processo robusto de vetorização e pré-processamento textual. Isso confirma o pressuposto de Quintanilha et al. (2023), segundo o qual a escolha do modelo deve ser orientada não apenas pela complexidade do algoritmo, mas pela qualidade do pré-processamento e pela natureza dos dados.

Dessa forma, o estudo responde à questão central proposta “os algoritmos clássicos de classificação conseguem analisar o sentimento humano no contexto de resenhas de

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

filmes?” de maneira afirmativa: sim, desde que os modelos sejam devidamente otimizados e aplicados em bases bem estruturadas, como o *IMDB*. O desempenho consistente do *SVM* e dos modelos *Naive Bayes* demonstra que a interpretação computacional de emoções expressas em linguagem natural é possível, ainda que limitada pela ausência de contexto pragmático e de compreensão semântica profunda, desafios que, conforme Ludermir et al. (2021), continuam sendo objeto de pesquisa no campo do PLN.

4.4 Análise da matriz de confusão e desempenho por classe

Para aprofundar a interpretação dos resultados, foi gerada a matriz de confusão do modelo *SVM* na configuração final (dataset completo otimizado). Essa matriz permite identificar como o modelo se comporta ao classificar cada polaridade, indicando separadamente acertos e erros por classe.

Os resultados mostram que, entre as resenhas realmente negativas, 6.357 foram corretamente classificadas como negativas e 1.143 foram classificadas incorretamente como positivas. Entre as resenhas realmente positivas, 6.419 foram classificadas corretamente como positivas e 1.081 foram classificadas incorretamente como negativas. Com base nesses valores, é possível calcular as métricas por classe:

- Precisão da classe negativa: aproximadamente 0,855
- Revocação da classe negativa: aproximadamente 0,848
- Precisão da classe positiva: aproximadamente 0,849
- Revocação da classe positiva: aproximadamente 0,856
- Acurácia global: aproximadamente 0,852

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

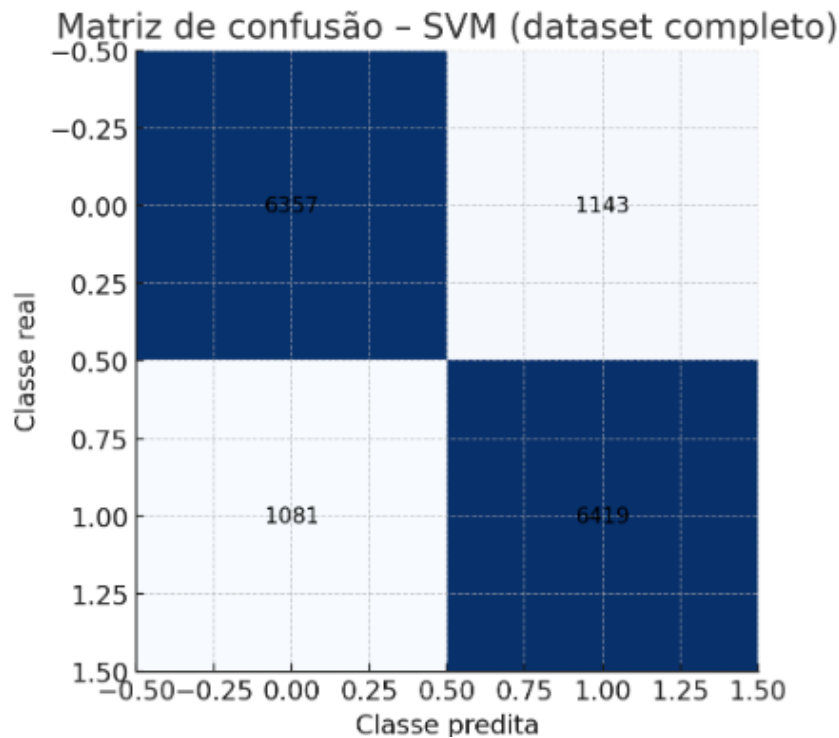


Figura 3 - Matriz de Confusão considerando o Algoritmo de Classificação SVM

Fonte: elaborada pelo autor.

Esses resultados mostram que o *SVM* apresentou desempenho equilibrado entre as duas polaridades, com leve vantagem na revocação da classe positiva e leve vantagem na precisão da classe negativa. Assim, o modelo acerta um pouco mais quando a resenha é realmente positiva (revocação positiva maior) e comete menos falsos positivos quando prediz que uma resenha é negativa (precisão negativa maior). Isso indica que o classificador linear tem boa capacidade de identificar tanto avaliações positivas quanto negativas, sem enviesamento acentuado em favor de apenas um dos polos emocionais.

Os achados quantitativos obtidos a partir da matriz de confusão (Figura 3) complementam as análises anteriores ao demonstrar, de forma mais detalhada, como o melhor modelo do estudo, o *SVM*, se comporta individualmente para cada polaridade. Além de apresentar o maior *F1-Score* geral, o modelo mostrou equilíbrio entre precisão e revocação nas duas classes, o que reforça sua adequação para tarefas de análise de sentimentos em bases textuais de grande escala.

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE**5. CONSIDERAÇÕES FINAIS**

Os experimentos realizados neste estudo permitiram responder à questão central proposta “os algoritmos clássicos de classificação conseguem analisar o sentimento humano no contexto de resenhas de filmes?” de forma afirmativa e fundamentada. Os resultados demonstraram que os modelos de aprendizado supervisionado tradicionais continuam sendo ferramentas eficazes para a análise de sentimentos, desde que adequadamente configurados e aplicados a bases de dados representativas.

Entre os modelos testados, o *SVM* destacou-se como o mais consistente, alcançando *F1-Score* de aproximadamente 0,87 após otimização dos parâmetros de convergência. Tal desempenho reflete a capacidade dos modelos lineares de lidar com alta dimensionalidade e vetorização textual, características típicas de dados provenientes de linguagem natural. O *Random Forest* apresentou métricas próximas, mas com custo computacional elevado, enquanto os classificadores *Naive Bayes* obtiveram um equilíbrio notável entre precisão e eficiência, tornando-se uma alternativa viável em contextos de baixa infraestrutura. O *KNN*, por sua vez, apresentou o desempenho menos satisfatório, corroborando a literatura que o aponta como menos adequado para espaços vetoriais densos, como o gerado pelo *TF-IDF*.

De maneira geral, os resultados indicam que modelos clássicos de aprendizado supervisionado permanecem relevantes em cenários de análise de sentimentos, especialmente quando associados a estratégias robustas de pré-processamento e vetorização. Essa constatação reforça o que Ludermit et al. (2021) e Loures (2024) defendem: que a eficiência de um modelo não depende unicamente de sua complexidade estrutural, mas da coerência entre os dados, a técnica de representação e os parâmetros utilizados.

O presente estudo também evidencia a importância de compreender a dimensão humana dos dados textuais. As resenhas do *IMDB* expressam emoções, julgamentos e subjetividades, o que desafia os modelos matemáticos a traduzir sentimentos em estruturas numéricas. Nesse sentido, ainda que os algoritmos tenham demonstrado competência técnica, há limites intrínsecos à interpretação computacional do afeto humano, aspecto que abre espaço para novas investigações que integrem abordagens híbridas entre ML e redes neurais profundas.

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

Como perspectivas futuras, recomendamos: a) expandir o estudo para outras bases de dados multilíngues, a fim de avaliar a generalização dos modelos; b) testar abordagens de *Deep Learning* (como *BERT* ou *LSTM*) e compará-las aos algoritmos clássicos; c) explorar a combinação entre técnicas de aprendizado supervisionado e análise semântica contextual, buscando aprimorar a capacidade das máquinas de reconhecer nuances emocionais e ironias linguísticas, que são dialetos mais complexos da nossa língua.

Em síntese, este trabalho reforça que, mesmo diante dos avanços recentes da Inteligência Artificial, os algoritmos clássicos de classificação continuam a oferecer excelente custo-benefício, estabilidade e interpretabilidade em tarefas de análise de sentimentos, contribuindo para aplicações práticas em áreas como marketing, mídia, educação e psicologia.

REFERÊNCIAS

BARROSO, L. R. Inteligência artificial, promessas, riscos e regulação. *Revista Direito e Práxis*, v. 15, n. 1, 2024. Disponível em: <https://www.scielo.br/j/rdp/a/n89PjvWXTdthJKwb6TtYXy/>

DOS SANTOS, G. C. M.; RODRIGUES, Y. P.; DE LIMA, B. S. Análise de Sentimentos em Dívida Técnica Auto-Admitida. *Revista Alomorfia*, v. 9, n. 1, 2025. Disponível em: <https://www.alomorfia.com.br/index.php/alomorfia/issue/view/33/24>

FERREIRA, J. S. P. Elementos básicos de aprendizagem de máquina. *InterMaths: Revista de Matemática e Educação Matemática*, v. 1, n. 1, 2023. Disponível em: <https://periodicos2.uesb.br/intermaths/article/download/15813/9308>

FERREIRA, W. A.; FREITAS, D. C. Análise de sentimentos em avaliações de produtos utilizando métodos de aprendizado de máquina. *Revista de Sistemas e Computação*, 2020. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/rsc/article/view/51388>

GOMES, G. S. L.; BALANIUK, R. Identificando evasão fiscal em empresas de fachada e em créditos ilegais de ICMS. *Revista de Administração Pública*, v. 58, n. 2, 2024. Disponível em: <https://www.scielo.br/j/rap/a/sy68bRbXvnfDVXxFCqPQGck/>

LOURES, S. A. Análise de sentimentos: como a inteligência artificial pode aprimorar a compreensão de emoções em textos. *Revista Contemporânea de Educação, Cultura e Comunicação*, v. 9, n. 2, 2024. Disponível em: <https://ojs.revistacontemporanea.com/ojs/index.php/home/article/download/6887/4912/20047>

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

LUDERMIR, T. B. et al. Inteligência artificial e aprendizado de máquina: estado da arte. Educação e Pesquisa, v. 47, n. 2, 2021. Disponível em: <https://www.scielo.br/j/ea/a/wXBdv8yHBV9xHz8qG5RCgZd/?lang=pt>

MENEZES, R. P. B. Aprendizado de máquina aplicado a QSAR. Química Nova, v. 47, n. 3, 2024. Disponível em: <https://www.scielo.br/j/qn/a/dhgmXqHrNSFY4Tb7P4kfvkd/>

PAIXÃO, G. M. M.; CAMPOS, B. Machine learning na medicina, revisão e aplicabilidade. Arquivos Brasileiros de Cardiologia, v. 119, n. 5, 2022. Disponível em: <https://www.scielo.br/j/abc/a/WMgVngCLbYfJrkmC65VFCkp/>

PFITSCHER, R. J. Análise de sentimentos em turmas de programação com uso de dataset de resenhas de filmes do IMDb. SBIE – Simpósio Brasileiro de Informática na Educação, 2023. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/download/26758/26577/>

PINHO, F. A. et al. Aplicação de técnicas de inteligência artificial para classificação de fuga ao tema em redações. Educação em Revista, v. 40, n. 2, 2024. Disponível em: <https://www.scielo.br/j/edur/a/YPPVRRKGdLdFbXTQXMmZgCc/>

QUINTANILHA, R. F. et al. Estudo de ferramentas para ciência de dados aplicadas a aprendizado de máquina: utilização do Python e bibliotecas Pandas, Matplotlib e Scikit-Learn. XXV Jornada de Iniciação Científica do CTI Renato Archer, 2023. Disponível em: <https://www.gov.br/cti/pt-br/publicacoes/producao-cientifica/jicc/xxv-jicc-2023/pdf/jicc-2023-paper-29.pdf>

ROCHA, L. A. Machine learning approaches in educational environments, a systematic literature review. Texto Livre, v. 18, n. 1, 2025. Disponível em: <https://www.scielo.br/j/tl/a/dTZ7XFwLzHWWby9kn5Cs6wP/>

RODRIGUES, J. P. Análise numérica e visualização de osciladores via Python. Revista Brasileira de Ensino de Física, v. 47, n. 1, 2025. Disponível em: <https://www.scielo.br/j/rbef/a/x5zFrtGhCYyK79mSFCLjrN/>

ROSSATI, K. F. et al. Planococcus citri in coffee trees by supervised classification using multispectral images. Ciência Rural, 2025. Disponível em: <https://www.scielo.br/j/cr/a/m9MxQtX979bzNVk9jYMKX6d/>

SCHAAB, B. L. et al. Desempenho de modelos de aprendizagem de máquina na detecção de depressão, ansiedade e estresse entre estudantes. Cadernos de Saúde Pública, v. 40, n. 4, 2024. Disponível em: <https://www.scielo.br/j/csp/a/J9PYjQCZDb7fdzQHY68nGZP/>

SIQUEIRA, V. X. Dataset anotado de sentimentos a partir de comentários em português. Data Science Workshop (DSW), 2024. Disponível em: <https://sol.sbc.org.br/index.php/dsw/article/download/30616/30419/>

FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

Agradecimentos

Aos meus pais, pela persistência. Mesmo diante das dificuldades, nunca deixaram de me apoiar, acreditaram na minha formação e se sacrificaram para que eu chegasse até aqui.

Ao meu orientador Bruno, por ajustar minhas expectativas com clareza e por sustentar o processo quando pensei em desistir. Pelo apoio constante, paciência e respeito absoluto.

Ao coordenador Rodrigo, exemplo de profissional e de ser humano. Acompanha cada aluno de perto, trata todos com humanidade, respeito e transparência. Celebra nossas conquistas e abre caminhos para novas oportunidades. Um coordenador verdadeiramente presente.

Ao meu tech lead Henrique, pelo apoio, incentivo e orientação. Por ter guiado meu desenvolvimento com dedicação e respeito, sendo um pilar na minha trajetória profissional. Se hoje posso dizer que sou desenvolvedor, muito se deve ao seu apoio.

À minha amiga Flaviana, minha principal base neste processo. Sem ela, este trabalho não existiria. Agradeço não só pela parceria acadêmica, mas pelo apoio emocional, amizade, companhia e orientação. Sua presença fez toda a diferença.

E à vida, por ter me conduzido até aqui. Apesar dos obstáculos, sou grato pela pessoa que me tornei, pela trajetória construída com esforço e resiliência e pela oportunidade de seguir em frente com respeito e gratidão pela existência.