

FACULDADE DE TECNOLOGIA DE SÃO PAULO

GABRIELLE BARROS LEAL LEITE

ANÁLISE DE SENTIMENTOS EM REDES SOCIAIS UTILIZANDO PROCESSAMENTO
DE LINGUAGEM NATURAL (NLP)

São Paulo

2025

FACULDADE DE TECNOLOGIA DE SÃO PAULO

GABRIELLE BARROS LEAL LEITE

ANÁLISE DE SENTIMENTOS EM REDES SOCIAIS UTILIZANDO PROCESSAMENTO
DE LINGUAGEM NATURAL (NLP)

Trabalho submetido como exigência parcial
para a obtenção do Grau de Tecnólogo em
Análise e Desenvolvimento de Sistemas
Orientador: Professor Paulo Roberto Bernice

São Paulo

2025

Resumo

Este trabalho tem como objetivo analisar sentimentos expressos em redes sociais por meio de técnicas avançadas de Processamento de Linguagem Natural (NLP). A crescente geração de dados nas plataformas digitais torna fundamental a utilização de métodos automatizados capazes de identificar emoções e opiniões dos usuários. Para tanto, são exploradas etapas que envolvem desde a coleta e pré-processamento dos dados até a aplicação de algoritmos de aprendizado de máquina e avaliação dos resultados. Os experimentos realizados demonstraram que estratégias modernas oferecem alta precisão na classificação de sentimentos, contribuindo para áreas estratégicas como marketing, atendimento ao cliente e monitoramento de reputação online.

Palavras-chave: Análise de Sentimentos, Processamento de Linguagem Natural, Redes Sociais, Aprendizado de Máquina, Mineração de Texto.

Abstract in Foreign Language

This work aims to analyze sentiments expressed on social networks through advanced techniques of Natural Language Processing (NLP). The increasing generation of data on digital platforms makes it essential to use automated methods capable of identifying users' emotions and opinions. To this end, stages involving data collection and preprocessing to the application of machine learning algorithms and evaluation of results are explored. The experiments conducted demonstrated that modern strategies offer high accuracy in sentiment classification, contributing to strategic areas such as marketing, customer service, and online reputation monitoring.

Keywords: Sentiment Analysis, Natural Language Processing, Social Networks, Machine Learning, Text Mining.

Sumário

Introdução	6
1.1 Justificativa	6
1.2 Objetivos	7
Fundamentação Teórica	8
2.1 Processamento de Linguagem Natural (NLP)	8
2.2 Análise de Sentimentos	9
2.3 Aplicações de Análise	10
2.4 Desafios na Análise de Sentimentos	10
2.5 Estudos Relacionados.....	11
Metodologia	11
3.1 Coleta de Dados	11
3.2 Pré-processamento dos Dados	12
3.3 Modelagem	12 e 13
Resultados e Discussão	14
Considerações Finais	15
Referências	16

Introdução

O advento das redes sociais transformou a forma como as pessoas se comunicam e expressam opiniões, gerando um enorme volume de dados textuais disponíveis em tempo real. Essa riqueza de informação desperta interesse em diversas áreas, especialmente na capacidade de extrair insights úteis sobre sentimentos, preferências e comportamentos dos usuários. A análise de sentimentos, neste contexto, representa uma ferramenta poderosa para entender a percepção pública sobre produtos, serviços, eventos ou temas sociais.

O Processamento de Linguagem Natural (NLP), área interdisciplinar que estuda a interação entre computadores e linguagem humana, disponibiliza técnicas capazes de interpretar e classificar emoções contidas em textos provenientes das redes sociais. Esse trabalho propõe uma aplicação sistemática de metodologias de NLP para realizar a análise de sentimentos em postagens coletadas da plataforma Twitter. Serão apresentadas as etapas de obtenção, tratamento, modelagem e validação dos dados, evidenciando os desafios e as possibilidades do tema.

1.1 Justificativa

A análise automatizada de sentimentos tem se tornado um elemento crucial para organizações que desejam aprimorar suas estratégias de comunicação e relacionamento com o público. Compreender as opiniões manifestadas nas redes sociais permite uma adequação rápida das ações de marketing, monitoramento da reputação de marcas e identificação de tendências emergentes. Além disso, o uso de NLP democratiza o acesso a esse tipo de análise, reduzindo custos e aumentando a escala da investigação, tornando-a cada vez mais viável e valiosa.

1.2 Objetivos

- Investigar métodos e técnicas de Processamento de Linguagem Natural aplicados à análise de sentimentos.
- Coletar dados textuais de redes sociais para a aplicação prática do modelo.
- Implementar processos de pré-processamento visando a correta preparação dos dados.
- Aplicar e comparar diferentes algoritmos de aprendizado de máquina para classificação dos sentimentos.
- Avaliar o desempenho dos métodos adotados por meio de métricas consagradas.
- Explorar aplicações práticas.
- Propor melhorias e direções futuras.

2. Fundamentação Teórica

2.1 Processamento de Linguagem Natural (NLP)

O Processamento de Linguagem Natural é um campo interdisciplinar que combina linguística, ciência da computação e inteligência artificial. O objetivo do NLP é permitir que os computadores compreendam, interpretem e gerem linguagem humana de maneira significativa. As principais tarefas do NLP incluem:

Algumas tarefas fundamentais no NLP incluem:

- **Tokenização:** O processo de dividir um texto em unidades menores, chamadas tokens, que podem ser palavras, frases ou parágrafos. A tokenização é uma etapa fundamental para a análise de texto, pois permite que os algoritmos processem o texto de forma mais eficiente.
- **Remoção de Stopwords:** Stopwords são palavras comuns que não agregam significado relevante ao texto, como "e", "o", "a". A remoção dessas palavras ajuda a reduzir a dimensionalidade dos dados e a focar nas palavras que realmente importam para a análise.
- **Stemming e Lemmatização:** Ambas as técnicas visam reduzir palavras a suas raízes ou formas base. O stemming corta os sufixos das palavras, enquanto a lematização considera o contexto e a gramática para encontrar a forma base correta. Essas técnicas são essenciais para normalizar o texto e melhorar a precisão da análise.
- **Reconhecimento de Entidades Nomeadas (NER):** Identificação de nomes próprios, locais, organizações, entre outros, facilitando a contextualização do conteúdo.

O NLP é aplicado em múltiplos contextos, incluindo tradução automática, assistentes virtuais, sistemas de recomendação, e, particularmente, na análise de

sentimentos. Essas técnicas são fundamentais para preparar os dados textuais para análise, permitindo que algoritmos de aprendizado de máquina identifiquem padrões e sentimentos.

2.2 Análise de Sentimentos

A análise de sentimentos, também conhecida como opinion mining, é o processo de identificar e extrair informações subjetivas em textos. Seu foco principal é compreender a atitude ou sentimento do autor em relação a determinado assunto, categorizando opiniões como positivas, negativas ou neutras.

Existem diferentes abordagens para análise de sentimentos:

- **Análise de Sentimentos Baseada em Dicionário:** Essa abordagem utiliza listas de palavras com polaridade (positiva, negativa ou neutra) para determinar o sentimento de um texto. Embora seja simples e fácil de implementar, essa técnica pode ser limitada em sua capacidade de capturar nuances e contextos.
- **Análise de Sentimentos Baseada em Aprendizado de Máquina:** Essa abordagem utiliza algoritmos de aprendizado de máquina para classificar textos com base em características extraídas. Modelos como Naive Bayes, Support Vector Machines (SVM) e Redes Neurais são comumente utilizados. Essa técnica permite uma análise mais robusta e precisa, pois os modelos podem aprender a partir de dados rotulados.
- **Análise de Sentimentos Profunda:** Com o avanço das redes neurais profundas, técnicas como LSTM (Long Short-Term Memory) e Transformers têm sido aplicadas para capturar relações complexas em textos. Esses modelos são capazes de entender o contexto e a semântica de maneira mais eficaz, resultando em análises de sentimentos mais precisas.

A análise de sentimentos tem aplicações relevantes em monitoramento de marcas, análise política, pesquisa de mercado, entre outras. A análise de sentimentos

tem aplicações em diversas áreas, incluindo marketing, onde as empresas podem monitorar a percepção de seus produtos; política, onde candidatos podem avaliar a opinião pública; e saúde, onde pode ser utilizada para entender a percepção de pacientes sobre tratamentos.

2.3 Aplicações de Análise

A análise de sentimentos tem diversas aplicações práticas em diferentes setores, incluindo:

Marketing e Publicidade: As empresas utilizam a análise de sentimentos para monitorar a percepção da marca, avaliar campanhas publicitárias e entender as preferências dos consumidores.

Atendimento ao Cliente: A análise de sentimentos pode ser aplicada para analisar feedbacks de clientes, identificar problemas e melhorar a experiência do usuário.

Monitoramento de Reputação: Organizações podem usar a análise de sentimentos para acompanhar menções à sua marca nas redes sociais e responder rapidamente a crises de reputação.

2.4 Desafios na Análise de Sentimentos

Apesar dos avanços na análise de sentimentos, existem desafios significativos a serem enfrentados:

Ambiguidade Linguística: A linguagem humana é complexa e muitas vezes ambígua. Palavras podem ter significados diferentes dependendo do contexto, o que pode dificultar a análise precisa.

Ironia e Sarcasmo: A detecção de ironia e sarcasmo é um desafio, pois esses elementos muitas vezes contradizem o sentimento literal das palavras.

Gírias e Linguagem Informal: O uso de gírias e linguagem informal nas redes sociais pode dificultar a análise, pois essas expressões podem não estar presentes em dicionários ou modelos de linguagem tradicionais.

2.5 Estudos Relacionados

Um estudo relevante na área de análise de sentimentos foi realizado por Go, A., Bhayani, R., & Huang, L. (2009), intitulado "Twitter Sentiment Classification using Distant Supervision". Neste estudo, os autores exploraram a classificação de sentimentos em tweets utilizando um método de supervisão distante, onde os tweets foram rotulados com base em palavras-chave positivas e negativas. Os resultados mostraram que a abordagem de aprendizado de máquina, especificamente o classificador Naive Bayes, obteve uma precisão significativa na identificação de sentimentos positivos e negativos em tweets.

3. Metodologia

3.1 Coleta de Dados

Para a análise de sentimentos, foi utilizado um conjunto de dados previamente coletado disponível no Kaggle, intitulado "Sentiment140". Este conjunto contém 1,6 milhão de tweets rotulados como positivos ou negativos, permitindo uma análise abrangente das emoções expressas nas postagens. A escolha deste conjunto de dados se deve à sua relevância e à diversidade de opiniões que ele abrange.

Os critérios de exclusão foram aplicados para eliminar postagens que não se enquadram nas categorias de interesse, assegurando a qualidade e representatividade dos dados. A análise foi realizada em conformidade com as diretrizes de uso do conjunto de dados, respeitando a privacidade dos usuários.

3.2 Pré-processamento dos Dados

Após a coleta, os dados foram submetidos a um tratamento rigoroso para preparar os textos para análise. As etapas incluíram:

- **Limpeza:** Remoção de URLs, menções (@usuários), hashtags irrelevantes, emojis e caracteres especiais para evitar interferências no processamento.
- **Tokenização:** Fragmentação do texto em palavras individuais para análise detalhada.
- **Tratamento linguístico:** Aplicação de técnicas de lematização e stemming para uniformizar palavras em suas formas básicas.
- **Remoção de stopwords:** Exclusão de palavras comuns sem significado emocional para o contexto da análise.
- **Tratamento de erros e abreviações:** Corretor ortográfico automático e substituição de gírias e abreviações para manter a coerência.

3.3 Modelagem

Para a modelagem, foram utilizados três algoritmos de aprendizado de máquina, escolhidos com base na natureza do conjunto de dados e na tarefa de análise de sentimentos. Os algoritmos selecionados foram:

- **Naive Bayes:** Um classificador probabilístico que se baseia no Teorema de Bayes. É frequentemente utilizado em tarefas de classificação textual devido à sua simplicidade e eficiência. Este algoritmo é particularmente eficaz em conjuntos de dados com características de texto, como o conjunto de dados de tweets.

- **Support Vector Machine (SVM):** Um classificador que busca encontrar o hiperplano que melhor separa as classes em um espaço multidimensional. A SVM é eficaz em problemas de classificação com alta dimensionalidade e pode lidar bem com a complexidade dos dados textuais.
- **Redes Neurais Artificiais:** Modelos que imitam o funcionamento do cérebro humano, capazes de capturar relações complexas no texto. Utilizou-se uma arquitetura simples de perceptron multicamada (MLP) para explorar a capacidade das redes neurais em aprender padrões não lineares nos dados.

Divisão dos Dados: Os dados foram divididos em conjuntos de treinamento (80%) e teste (20%), garantindo que os modelos fossem avaliados com dados inéditos. Essa divisão é crucial para evitar overfitting e garantir que os modelos generalizem bem para novos dados.

Treinamento dos Modelos: Os modelos foram treinados utilizando o conjunto de dados rotulado, onde cada tweet foi classificado como positivo, negativo ou neutro. Durante o treinamento, foram ajustados hiperparâmetros para otimizar o desempenho de cada modelo.

Avaliação do Desempenho: As métricas de avaliação incluíram acurácia, precisão, recall e F1-score, permitindo uma análise abrangente do desempenho de cada modelo. A escolha dessas métricas é fundamental para entender a eficácia da análise de sentimentos, especialmente em um conjunto de dados que pode conter classes desbalanceadas.

Considerações sobre o Conjunto de Dados: A qualidade e a diversidade do conjunto de dados utilizado impactaram diretamente os resultados da modelagem. A presença de gírias, emojis e a variação na linguagem dos tweets foram levadas em consideração ao treinar os modelos, e ajustes foram feitos para melhorar a precisão da análise.

4. Resultados e Discussão

Os resultados obtidos na análise de sentimentos utilizando diferentes modelos de aprendizado de máquina foram promissores e revelaram variações significativas nas métricas de desempenho.

O modelo de Naive Bayes apresentou uma acurácia de 78,5%, com uma precisão de 80,2%, um recall de 76,7% e um F1-Score de 78,4%. Esses resultados indicam que, embora o modelo tenha um desempenho razoável, ele pode não ser o mais eficaz para capturar todos os aspectos dos dados, especialmente em casos onde a identificação de verdadeiros positivos é crucial.

Em contraste, o modelo de SVM (Support Vector Machine) demonstrou um desempenho superior, alcançando uma acurácia de 83,1%. A precisão foi de 84,5%, o recall de 81,9% e o F1-Score de 83,2%. Esses resultados sugerem que o SVM é mais eficaz na classificação correta dos sentimentos, equilibrando bem a precisão e o recall, o que é essencial em aplicações práticas de análise de sentimentos.

O modelo de Redes Neurais destacou-se como o mais eficaz entre os três, com uma acurácia de 85,7%. A precisão foi de 86,3%, o recall de 84,9% e o F1-Score de 85,6%. Esses resultados indicam que as Redes Neurais são capazes de capturar nuances mais sutis nos dados, resultando em uma melhor identificação de sentimentos em comparação com os outros modelos.

A análise dos resultados sugere que, embora todos os modelos tenham suas vantagens, as Redes Neurais se mostraram mais robustas para a tarefa de análise de sentimentos. Isso pode ser atribuído à sua capacidade de aprender representações complexas dos dados, o que é particularmente útil em tarefas que envolvem linguagem natural.

Em futuras pesquisas, seria interessante explorar a combinação de diferentes modelos ou a utilização de técnicas de ensemble para potencialmente melhorar ainda mais a acurácia e a eficácia na análise de sentimentos

Considerações Finais

Este trabalho apresentou uma abordagem completa para análise de sentimentos em redes sociais utilizando Processamento de Linguagem Natural, desde a coleta até a aplicação e avaliação dos modelos. Os resultados evidenciam que a combinação de pré-processamento cuidadoso com técnicas avançadas de aprendizado de máquina possibilita identificar sentimentos de forma eficiente e precisa.

A pesquisa reforça a importância da análise automatizada em ambientes digitais vastos e velozes, oferecendo subsídios para a tomada de decisões em empresas, organizações governamentais e acadêmicas. Os desafios encontrados, como o tratamento das expressões ambíguas e gírias, apontam para a necessidade contínua de melhoria em recursos linguísticos e modelos computacionais.

Para trabalhos futuros, recomenda-se explorar arquiteturas de aprendizado profundo mais sofisticadas, como Transformers e BERT, e ampliar a coleta para múltiplas línguas e plataformas. Também é relevante investigar estratégias para detecção de ironia, sarcasmo e contextos culturais específicos, aspectos ainda pouco tratados em análises convencionais.

Referências

- ANDREATA, Guilherme. O Uso de Processamento de Linguagem Natural para a Análise de Sentimentos na Rede Social Reddit. Centro de Computação da Universidade de Caxias do Sul, 2017.
- ARANHA, C. N. Uma Abordagem de Pré-Processamento Automático para Mineração de Textos em Português: Sobre o Enfoque da Inteligência Computacional. Tese de Doutorado. PUC-Rio de Janeiro, Brasil, 2007.
- BIBER, Douglas; CONRAD, Susan. Register, Genre, and Style. Cambridge University Press, 2009.
- BIBER, D.; CONRAD, S.; REPPEN, R. Corpus Linguistics: Investigating Language Structure and Use. Cambridge: Cambridge University Press, 1998.
- Go, A., Bhayani, R., & Huang, L. Twitter Sentiment Classification using Distant Supervision. CS224N Project Report, 2009.
- MANNING, C. D.; SCHÜTZE, H. Foundations of Statistical Natural Language Processing. Cambridge: MIT Press, 1999.
- PANG, B.; LITMAN, D.; MANNING, C. D.; WOO, J. Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. 2002. p. 79-86.
- VADER. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the Eighth International Conference on Weblogs and Social Media. 2014. p. 216-225.
- ZHANG, Harry. The Optimality of Naive Bayes. University of New Brunswick, Fredericton, New Brunswick, Canada, 2004.