

UNIVERSIDADE PRESBITERIANA MACKENZIE

Faculdade de Computação e Informática
Tecnologia em Ciência de Dados
Projeto Aplicado 2

Aplicando Conhecimento 2

Análise de Sentimentos em Redes Sociais
Grupo Projeto Aplicado 18

Allana Silva Santana	10424950
Cristiano Prado do Carmo	10720249
Deborah Silvério Alves Morales	10728563
Maria Fernanda Salles Vasconcellos	10424791

Sumário

Seção 1 - Apresentação	3
Título do Trabalho	3
Membros do Grupo	3
Objetivo do Projeto	3
Contexto de Estudo	3
Descrição da Origem	3
Descrição do Dataset	4
Repositório do Projeto	4
Seção 2 – Premissas do Projeto	5
Empresa	5
Problema de Estudo	5
Objetivo	7
Metas Específicas	7
Seção 3 – Cronograma do Projeto	8
Seção 4 - Bibliografia	9
Referências	9

Seção 1 - Apresentação

Título do Trabalho

Análise de sentimentos em redes sociais.

Membros do Grupo

Aluno	RA
Allana Silva Santana	10424950
Cristiano Prado do Carmo	10720249
Deborah Silvério Alves Morales	10728563
Maria Fernanda Salles Vasconcellos	10424791

Objetivo do Projeto

Desenvolver um modelo de análise de sentimentos capaz de classificar automaticamente tweets em categorias positivas, negativas ou neutras, demonstrando a aplicação de técnicas de PLN e aprendizado de máquina.

Contexto de Estudo

As redes sociais se tornaram um espaço central de expressão de opiniões sobre produtos, marcas, política e eventos sociais. A Análise de Sentimentos é uma área do Processamento de Linguagem Natural (PLN) que busca identificar automaticamente a polaridade de textos - geralmente classificados como positivos, negativos ou neutros.

Esse tipo de análise enfrenta desafios, como textos curtos, linguagem informal, gírias e ironias, mas é amplamente aplicado em marketing, política e atendimento ao cliente. *Datasets* públicos, como o Twitter US Airline Sentiment, permitem desenvolver modelos que vão de técnicas estatísticas tradicionais até abordagens modernas de *deep learning* (ex.: BERT).

Assim, a análise de sentimentos em redes sociais é uma ferramenta estratégica para entender percepções, apoiar decisões e antecipar tendências na sociedade.

Descrição da Origem

A origem dos dados é o Kaggle, que é um repositório de milhares de datasets, contendo os mais variados temas.

Especificamente, o dataset escolhido para este trabalho é o **Twitter US Airline Sentiment**¹, que é um conjunto de dados composto por 15.000 tweets postados por viajantes, em fevereiro de 2015, expressando seus sentimentos.

¹ <https://www.kaggle.com/datasets/crowdfunder/twitter-airline-sentiment>

Descrição do Dataset

O dataset escolhido é formado por um único arquivo em formato “.csv” nomeado como *tweets.csv*, e detalhado a seguir.

O arquivo contém 15.000 tweets que foram devidamente classificados (0 – negativo; 2 – neutro; 4 – positivo) e que podem ser usados para detectar o sentimento expresso em cada um deles.

De forma geral, as variáveis que compõem o dataset, e que usaremos em nosso trabalho, são as seguintes:

Variável	Descrição	Tipo de dado
tweet_id	Id do tweet	numeric
airline_sentiment	Sentimento a respeito da companhia aérea	string
airline_sentiment_confidence	Confiança a respeito do sentimento	numeric
Negativereason'	Motivo negativo	string
negativereason_confidence	Confiança a respeito do motivo negativo	numeric

Repositório do Projeto

O endereço do repositório do projeto contendo toda a documentação produzida, datasets utilizados e códigos-fontes criados é:

https://github.com/httpsdebs/Projeto_Aplicado_II/tree/main

Seção 2 – Premissas do Projeto

Empresa

O X, anteriormente conhecido como Twitter, é uma das principais plataformas de redes sociais do mundo, lançada em 2006 e atualmente pertencente à X Corp. A plataforma permite que usuários publiquem mensagens curtas chamadas tweets, com limite de caracteres, podendo incluir links, imagens, vídeos, hashtags e menções a outros usuários.

O X tornou-se um espaço central para a expressão de opiniões, debates públicos, compartilhamento de notícias e tendências, sendo amplamente utilizado por indivíduos, empresas, jornalistas e autoridades. A característica de publicar textos curtos e instantâneos faz do X uma fonte valiosa para análises de sentimentos e mineração de opiniões, pois os usuários frequentemente manifestam percepções sobre produtos, serviços, eventos sociais e políticos.

No contexto deste projeto, os tweets representam dados textuais que serão utilizados para treinar modelos de Classificação de Sentimentos, permitindo a análise automática da polaridade (positiva, negativa ou neutra) das postagens. A plataforma é escolhida devido à abundância de dados disponíveis publicamente, à diversidade de usuários e à relevância de seu conteúdo para estudos de comportamento, marketing e comunicação digital.

Problema de Estudo

As redes sociais se tornaram espaços centrais de expressão de opiniões sobre produtos, marcas, política e eventos sociais. Entre essas plataformas, o X — anteriormente conhecido como Twitter — se destaca como um ambiente de grande circulação de informações e debates públicos. Lançado em 2006 e atualmente pertencente à X Corp, o X permite que usuários publiquem mensagens curtas chamadas tweets, com a possibilidade de incluir links, imagens, vídeos, hashtags e menções.

Apesar da abundância de informações, as organizações enfrentam o desafio de monitorar e compreender rapidamente a percepção do público em relação a produtos, serviços ou eventos. A análise manual desses dados é inviável devido ao grande volume de postagens e à diversidade de opiniões expressas. Esse problema destaca a necessidade de ferramentas automatizadas capazes de classificar o sentimento dos usuários de forma eficiente, fornecendo insights estratégicos para empresas, instituições e tomadores de decisão.

Essa característica de textos curtos e instantâneos torna o X uma fonte valiosa para análises de comportamento e opinião, oferecendo dados textuais ricos para estudos de Processamento de Linguagem Natural (PLN) e Ciência de Dados. A plataforma é amplamente utilizada por indivíduos, empresas,

jornalistas e autoridades, sendo relevante para monitoramento de tendências, análise de marketing e avaliação da percepção pública.

Objetivo

Desenvolver um modelo de análise de sentimentos aplicado a tweets, capaz de classificar automaticamente postagens em categorias de polaridade (positiva, negativa ou neutra), de modo a demonstrar como técnicas de Processamento de Linguagem Natural (PLN) e aprendizado de máquina podem ser utilizadas para compreender percepções expressas em redes sociais.

Metas Específicas

Para alcançar o objetivo proposto, este trabalho será conduzido por meio de etapas organizadas em metas específicas. Essas metas visam orientar o processo analítico de forma estruturada, assegurando que cada fase — desde a exploração do dataset até a avaliação dos modelos — contribua para a obtenção de resultados consistentes e relevantes. Assim, pretende-se:

Selecionar e compreender o dataset (Twitter Airline Sentiment), verificando sua estrutura, variáveis e distribuição de classes de sentimentos.

- **Realizar o pré-processamento textual**, incluindo limpeza de caracteres especiais, links, hashtags, menções e normalização das palavras.
- **Aplicar técnicas de representação textual**, como *Bag of Words*, *TF-IDF* e *embeddings*, preparando os dados para modelagem.
- **Treinar e comparar diferentes algoritmos de classificação**, como Naive Bayes, Regressão Logística, Random Forest e Redes Neurais, avaliando o desempenho em termos de acurácia, precisão, recall e F1-score.
- **Visualizar e interpretar os resultados obtidos**, destacando a distribuição dos sentimentos nos dados e discutindo as limitações do modelo.
- **Apresentar conclusões sobre a viabilidade do uso da análise de sentimentos em redes sociais**, ressaltando sua relevância para empresas e organizações na tomada de decisões.

Seção 3 – Cronograma do Projeto

Etapa		Atividade	Início	Término	Responsável
1	Kick-Off				
	1.1	Definição do grupo de trabalho	19/08/2025	22/08/2025	Todos
	1.2	Definição das premissas do projeto	25/08/2025	02/09/2025	Todos
2	Definições do Produto Analítico				
	2.1	Seleção de bibliotecas da linguagem Python a serem usadas no projeto	08/09/2025	10/09/2025	Todos
	2.2	Definição do repositório no GitHub	11/09/2025	12/09/2025	Todos
	2.3	Análise exploratória dos dados do dataset escolhido	12/09/2025	15/09/2025	Todos
	2.4	Tratamento da base para preparação e treinamento	16/09/2025	23/09/2025	Cristiano Prado/Deborah Morales
	2.5	Descrição da base teórica e dos métodos analíticos que serão empregados	19/09/2025	25/09/2025	Deborah Morales/Allana Silva
	2.6	Definição do cálculo da acurácia	26/09/2025	02/10/2025	Allana Silva/Maria Fernanda
3	Apresentação de Produtos e Storytelling				
	3.1	Consolidação dos resultados do método analítico escolhido	05/10/2025	08/10/2025	Cristiano Prado/Allana Silva
	3.2	Aplicação das medidas de acurácia	05/10/2025	08/10/2025	Cristiano Prado/Allana Silva
	3.3	Descrição dos resultados preliminares - produto gerado	09/10/2025	16/10/2025	Maria Fernanda
	3.4	Esboço do Storytelling	17/10/2025	23/10/2025	Deborah Morales/ Maria Fernanda
4	Apresentação de Resultados (Entrega do projeto)				
	4.1	Entrega do relatório técnico do projeto	26/10/2025	21/11/2025	Deborah Morales/Cristiano Prado
	4.2	Reunião de gravação do vídeo	16/11/2025	18/11/2025	Todos
	4.3	Apresentação do Storytelling	26/10/2025	21/11/2025	Todos
	4.4	Atualização final do repositório no GitHub	18/11/2025	19/11/2025	Cristiano Prado/Maria Fernanda
	4.5	Entrega de vídeo do Storytelling, em um canal do YouTube	26/10/2025	21/11/2025	Todos

Seção 4 - Bibliografia

Referências

SANTANA, Allana Silva; CARMO, Cristiano Prado do; MORALES, Deborah Silvério Alves; *Projeto Aplicado 2*. Disponível em: <https://github.com/httpsdebs/Projeto_Aplicado_II>. Acesso em: 03 set. 2025.

CROWD-FLOWER. *Twitter US Airline Sentiment* [conjunto de dados]. Kaggle. Disponível em: <<https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment>>. Acesso em: 3 set. 2025.

GO, Alec; BHAYANI, Richa; HUANG, Lei. Twitter sentiment classification using distant supervision. *CS224N Project Report*, Stanford, v. 1, p. 12, 2009.