

Identificação de sentimentos e emoções em textos nas redes sociais

Julian de Campos Teixeira, Luis Gustavo Aguirre Castanho, Luiz Henrique Bonilha Pasquinelli, Vinicius Moreira Tamburu

Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

Graduação em Ciência da Computação – Faculdade de Computação e Informática (FCI)
– Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

{julian.teixeira, luis.castanho, luiz.pasquinelli,
vinicius.tamburu}@mackenzista.com.br

Abstract. *This paper aims to develop a methodology for identifying emotions and feelings present in texts published on social media, using Natural Language Processing and Machine Learning techniques. Emotion detection is crucial for better understanding online interactions, especially in the context of polarized debates, where emotions such as joy, sadness, fear, and hatred have a significant impact. The proposed approach seeks to provide a robust emotional analysis that can be applied in various fields, such as digital marketing, mental health, and conflict mediation.*

Resumo. *Este artigo tem como objetivo desenvolver uma metodologia para identificar emoções e sentimentos presentes em textos publicados em redes sociais, utilizando técnicas de Processamento de Linguagem Natural e Aprendizado de Máquina. A detecção de emoções mostra-se crucial para compreender melhor as interações no meio online, especialmente em contextos de debates polarizados, onde emoções como alegria, tristeza, medo e ódio têm grande impacto. A abordagem proposta visa fornecer uma análise emocional robusta que possa ser utilizada em diversos campos, como marketing digital, saúde mental e mediação de conflitos*

1. Introdução

A crescente utilização das redes sociais como meio para expressão pessoal e debates gera um grande volume de informações, muitas vezes carregadas de emoções e opiniões. Essas manifestações emocionais, em sua maioria, não são estruturadas e são expressas em textos curtos, o que desafia os métodos considerados mais tradicionais de análise de dados (BATBAATAR, 2019). Embora abordagens para análise de sentimentos estejam consolidadas, muitas não capturam a complexidade das emoções de forma eficaz, especialmente em textos em português (KANSAON, 2019). Nesse sentido, técnicas modernas de Processamento de Linguagem Natural (PLN), mineração de texto e aprendizado de máquina têm demonstrado potencial para superar essas limitações (BARBOSA, 2017).

Este projeto tem como objetivo geral desenvolver e aplicar técnicas de Inteligência Artificial e Processamento de Linguagem Natural para analisar e identificar sentimentos e emoções em mensagens de redes sociais escritas em português brasileiro. A partir disso, identificar tendências e percepções da opinião pública sobre temas de interesse coletivo. As emoções selecionadas foram Raiva, Alegria, Medo, Tristeza, Confiança, Amor e Ausente (neutro).

A relevância deste trabalho reside na possibilidade de fornecer insights valiosos para diversas áreas. No marketing digital, por exemplo, a análise emocional pode otimizar estratégias de engajamento. Em políticas públicas, a identificação de emoções pode ajudar a compreender o impacto social de determinadas ações governamentais. Já na saúde mental, compreender as emoções em larga escala pode oferecer suporte à criação de intervenções mais direcionadas (BATBAATAR, 2019; FIGUEIREDO et al, 2018).

2. Descrição do Problema

A dificuldade em identificar automaticamente as emoções presentes em argumentos textuais nas redes sociais impede uma compreensão mais aprofundada do comportamento humano online. A complexidade da linguagem natural, aliada ao uso de ironias, ambiguidades e polarizações, torna o problema desafiador do ponto de vista técnico e semântico.

A análise de emoções em textos enfrenta desafios significativos, como a informalidade e a ambiguidade das expressões, características comuns em textos de redes sociais (BATBAATAR ET AL., 2019). Embora abordagens para análise de sentimentos estejam consolidadas, muitas não capturam a complexidade das emoções de forma eficaz, especialmente em textos em português (KANSAON, BRANDÃO E PINTO, 2019). Nesse sentido, técnicas modernas de Processamento de Linguagem Natural (PLN), mineração de texto e aprendizado de máquina têm demonstrado potencial para superar essas limitações (BARBOSA ET AL., 2017).

3. Ética e responsabilidade no uso da IA para desenvolvimento da solução

A utilização de Inteligência Artificial para identificar emoções em textos exige responsabilidade quanto à privacidade, possibilidade de viés algorítmico e transparência nos critérios utilizados. O modelo deve ser treinado com dados anonimizados e auditáveis, garantindo que decisões automatizadas não reforcem estereótipos ou prejudiquem grupos vulneráveis. Além disso, deve haver supervisão humana e delimitação clara de uso, evitando assim interpretações errôneas ou abusos da ferramenta.

Durante o desenvolvimento do projeto foi utilizado um dataset do Projeto de Pesquisa HEIWA de acesso restrito que apresenta dados pessoais e dados sensíveis, para usá-lo foi necessário que todos os membros do grupo assinassem um termo de responsabilidade se comprometendo integralmente pelo uso adequado dos dados somente para os fins da pesquisa. Para garantir o sigilo dos dados o dataset tem conteúdo totalmente anonimizado, como o dataset é composto de publicações em redes sociais removido o nome da conta do autor da publicação original e quaisquer menções a outras contas no corpo texto foram substituídas pelo termo “@user”.

Durante o desenvolvimento da pesquisa, foi implementada uma etapa de avaliação manual criteriosa dos dados, com foco na verificação e rotulação supervisionada das publicações. Essa intervenção humana não apenas assegura maior qualidade e coerência dos rótulos atribuídos, como também atuou como um cuidado importante contra vieses indesejados que poderiam ser perpetuados por modelos treinados automaticamente. Ao incorporar a análise e decisão humanas nesse processo, buscamos mitigar distorções, respeitando os princípios de justiça, transparência e responsabilidade no uso da Inteligência Artificial.

4. Dataset utilizado

O dataset utilizado pertence ao CIBERDEM, grupo de pesquisa e inovação de professores e estudantes da Universidade Presbiteriana Mackenzie em Ciberdemocracia, consistindo em mais de cinco mil tweets públicos coletados em janeiro de 2023 durante os eventos do dia 8, onde houve uma invasão à Praça dos Três Poderes.

Os dados foram anonimizados para preservar a privacidade dos usuários. Além disso, um recorte inicial de 200 tweets foi feito para a primeira etapa da análise, com o tratamento e rotulação das emoções expressas nas mensagens sendo realizado de maneira manual.

Foram desenvolvidos três modelos, utilizando a linguagem python, dois deles com a aplicação de uma função de pré-processamento, e o outro sem a aplicação do pré-processamento. Os primeiros dois modelos a serem desenvolvidos utilizam o Multinomial Naive Bayes (MNB), para o método de classificação, foram desenvolvidos dois modelos, um para classificar as emoções e outro para classificar os sentimentos. O modelo BERTimbau, foi construído com a aplicação do pré-processamento, porém foi utilizado somente na classificação dos sentimentos.

5. Metodologia

A pesquisa classifica-se como aplicada, de abordagem tanto qualitativa, quanto quantitativa, tendo como objetivo principal compreender e automatizar a identificação de emoções em argumentos expressos nas redes sociais. Para isso, serão utilizados métodos de análise textual e técnicas de Processamento de Linguagem Natural (PLN), combinadas com aprendizado de máquina, para detecção e classificação de emoções em textos.

A metodologia adotada neste trabalho é composta por três grandes etapas: rotulação manual, pré-processamento dos dados e experimentação com diferentes arquiteturas de classificação.

5.1. Rotulação Manual

O processo de classificação das emoções iniciou-se com a avaliação manual dos tweets em estado bruto, sem a aplicação de qualquer pré-processamento. Para garantir a diversidade de interpretações e mitigar vieses individuais, a amostra inicial de 200 tweets foi dividida entre duas duplas de avaliadores, totalizando 100 tweets para cada dupla. Essa estratégia não apenas assegura maior imparcialidade no julgamento das emoções expressas nos textos, como também proporciona uma análise cruzada mais rica, permitindo observar divergências de interpretação entre os avaliadores. Tal abordagem foi essencial considerando que cada tweet pode carregar múltiplas emoções simultaneamente ou mesmo não apresentar nenhuma emoção clara, exigindo, portanto, sensibilidade e atenção ao contexto linguístico e semântico. Em casos de discordância entre os membros da dupla, foi acionado um terceiro avaliador automatizado — o modelo Ollama LLM —, o qual serviu como critério de desempate, conferindo consistência metodológica ao processo de rotulação manual e preparando os dados para as etapas subsequentes de pré-processamento e classificação automatizada. (Dosciatti et al. 2015)

Para a produção do modelo de classificação automática de emoções em textos, foram adotadas múltiplas abordagens, combinando pré-processamento textual e diferentes algoritmos de aprendizado de máquina e aprendizado profundo.

5.2. Análise exploratória das anotações

Após a etapa de classificação, foi realizada uma análise exploratória com o objetivo de identificar potenciais inconsistências na base de dados e antecipar desafios que poderiam impactar o desempenho dos modelos durante o treinamento e a avaliação.

Conforme evidenciado nos diagramas de Pareto apresentados na Figura 1, observa-se uma concentração significativa de rótulos em poucas categorias. No caso da classificação por emoções, mais de 70% das anotações estão restritas a apenas duas

emoções predominantes: raiva e tristeza. De maneira similar, na classificação por sentimentos, mais de 70% dos exemplos foram rotulados como negativos.

Essa distribuição altamente desequilibrada já indicava, desde a fase exploratória, possíveis dificuldades no desempenho dos modelos, especialmente na capacidade de prever categorias com baixa representatividade. Tal desequilíbrio impacta diretamente as métricas de avaliação, reduzindo a sensibilidade e a precisão em classes minoritárias.

Figura 1: Distribuição dos sentimentos (à esquerda) e emoções (à direita) nos 200 tweets anotados, em um diagrama de Pareto.



Fonte: Autoria própria.

5.3. Pré-processamento

Inicialmente, os dados passam por um pré-processamento textual, realizado com auxílio da linguagem Python e da biblioteca re (expressões regulares). Essa etapa é essencial para garantir a qualidade dos dados e reduzir ruídos que poderiam impactar negativamente a performance dos modelos. As principais operações incluem:

- Remoção de URLs e links externos, que normalmente não agregam valor semântico relevante ao conteúdo emocional dos tweets;
- Eliminação de menções a usuários (iniciadas por “@”), que tendem a ser elementos contextuais não informativos;
- Exclusão de pontuação e caracteres especiais desnecessários, com o objetivo de simplificar o texto e facilitar sua vetorização futura;
- Padronização para letras minúsculas, a fim de evitar duplicidade na representação das palavras;
- Futuramente, pretende-se também aplicar tokenização, remoção de stopwords, e possivelmente lematização, a depender do modelo utilizado.

Figura 2: Fluxo demonstrando o pré-processamento aplicado sobre os tweets de treino/teste



Fonte: Autoria própria.

5.4. Arquiteturas de classificação

Após o pré-processamento, será conduzida a etapa de experimentação com diferentes modelos de classificação, com o objetivo de comparar seus desempenhos e identificar a arquitetura mais eficiente para a tarefa de detecção de múltiplas emoções. Os modelos considerados são:

Multinomial Naive Bayes (MNB): modelo probabilístico tradicional, amplamente utilizado em tarefas de classificação de texto. É especialmente adequado para bases com alta dimensionalidade e pressupõe independência condicional entre os atributos (KANSAON et al., 2019).

BERTimbau com *fine-tuning* supervisionado: abordagem na qual o modelo BERTimbau, baseado na arquitetura de Transformers ajustado ao português brasileiro, é ajustado diretamente sobre os dados rotulados de emoções. O objetivo é adaptar seus pesos internos de forma supervisionada, otimizando sua capacidade preditiva específica para o domínio da detecção emocional (HAMMES; FREITAS, 2021).

6. Resultados

Como etapa inicial da avaliação dos modelos de classificação multirrótulo, foi testada a combinação do classificador MNB com a estratégia One-vs-Rest, utilizando vetores de contagem simples (CountVectorizer) como representação dos textos. Esse modelo se destaca por sua leveza computacional e simplicidade de implementação, sendo tradicionalmente utilizado como linha de base em tarefas de categorização de texto.

Testamos o modelo com e sem pré-processamento das mensagens. A partir da binarização dos rótulos (permitindo múltiplas emoções por entrada), o modelo foi treinado e avaliado sobre uma divisão de 80% dos dados para treino e 20% para teste. Os rótulos foram tratados com MultiLabelBinarizer, permitindo a associação de cada texto a uma ou mais emoções simultaneamente.

Figura 3: Métricas de desempenho do modelo MNB para emoções, sem pré-processamento aplicado.

MNB para emoções (sem pré-processamento)				
Emoção	Precision	Recall	F1-Score	Tamanho da amostra de teste
ALEGRIA	0.00	0.00	0.00	1
AMOR	0.00	0.00	0.00	0
AUSENTE	0.33	0.14	0.20	7
CONFIANÇA	0.00	0.00	0.00	8
MEDO	0.00	0.00	0.00	4
RAIVA	0.42	0.69	0.52	16
TRISTEZA	0.20	0.09	0.12	11
Accuracy	0.2			

Fonte: Autoria própria.

Figura 4: Métricas de desempenho do modelo MNB para emoções, com pré-processamento aplicado.

MNB para emoções (com pré-processamento)				
Emoção	Precision	Recall	F1-Score	Tamanho da amostra de teste
ALEGRIA	0.00	0.00	0.00	1
AMOR	0.00	0.00	0.00	0
AUSENTE	1.00	0.14	0.25	7
CONFIANÇA	0.00	0.00	0.00	8
MEDO	0.00	0.00	0.00	4
RAIVA	0.42	0.69	0.52	16
TRISTEZA	0.50	0.09	0.15	11
Accuracy	0.225			

Fonte: Autoria própria.

Figura 5: Métricas de desempenho do modelo MNB para sentimento, sem pré-processamento aplicado.

MNB para sentimentos (sem pré-processamento)				
Sentimento	Precision	Recall	F1-Score	Tamanho da amostra de teste
NEGATIVO	0.76	0.93	0.84	30
NEUTRO	0.33	0.14	0.20	7
POSITIVO	0.00	0.00	0.00	3
Accuracy	0.675			

Fonte: Autoria própria.

Figura 6: Métricas de desempenho do modelo MNB para sentimento, com pré-processamento aplicado.

MNB para sentimentos (com pré-processamento)				
Sentimento	Precision	Recall	F1-Score	Tamanho da amostra de teste
NEGATIVO	0.76	0.93	0.84	30
NEUTRO	1.00	0.14	0.25	7
POSITIVO	0.00	0.00	0.00	3
Accuracy	0.725			

Fonte: Autoria própria.

O desempenho do modelo, de forma geral, revelou limitações esperadas para essa abordagem. Na classificação de emoções, a acurácia obtida foi de 20% (Figura 3) sem pré-processamento e 22,5% com pré-processamento (Figura 4), indicando uma dificuldade do modelo em acertar todas as emoções, especialmente aquelas com menor frequência no dataset, mas evidenciando a vantagem de utilizar o pré-processamento. Na classificação de sentimentos, a acurácia apresentada foi maior, atingindo 67,5% sem pré-processamento (Figura 5), e 72,5% com pré-processamento (Figura 6). Esse resultado se deve, em parte, à menor quantidade de classes e à estrutura mais simples da tarefa. No entanto, ainda é possível observar o impacto do desequilíbrio dos rótulos, especialmente na previsão da classe positivo. Apesar dos resultados limitados, esse experimento foi fundamental para estabelecer um ponto de referência, permitindo comparações futuras com modelos mais sofisticados.

Como continuação da etapa de experimentação, foi avaliado o desempenho de um modelo baseado no BERTimbau padrão, uma versão do BERT pré-treinado para o português. Nessa configuração, a arquitetura foi mantida conforme disponibilizada originalmente, sem ajuste dos pesos internos do modelo. A tarefa considerada foi a de classificação de sentimentos em três categorias mutuamente exclusivas (positivo, neutro e negativo), utilizando a função de ativação softmax na saída e a função de perda cross-entropy, adequadas para problemas de classificação multiclasse. O pré-processamento textual, a divisão dos dados (80% para treino e 20% para teste) e a tokenização com o BertTokenizer da versão cased foram mantidos consistentes com os experimentos anteriores.

Figura 7: Métricas de desempenho do modelo Bertimbau padrão, para sentimento, com pré-processamento aplicado.

Bertimbau para sentimentos (com pré-processamento)				
Sentimento	Precision	Recall	F1-Score	Tamanho da amostra de teste
NEGATIVO	0.77	1.00	0.87	30
NEUTRO	1.00	0.14	0.25	7
POSITIVO	0.00	0.00	0.00	3
Accuracy	0.775			

Fonte: Autoria própria.

Nesta configuração, o modelo alcançou uma acurácia de 77,5% (Figura 7), superando levemente o desempenho obtido com o modelo Naive Bayes (MNB). Ainda assim, observou-se que a classe positivo apresentou desempenho inferior em relação às demais, reflexo direto de sua menor frequência no conjunto de dados.

De maneira geral, os testes indicaram avanços limitados na acurácia dos modelos, principalmente devido às limitações estruturais da base de dados, como número reduzido de exemplos, desequilíbrio entre classes e múltiplos rótulos por tweet.

7. Links

Github: <https://github.com/ViniTamburu26/Inteligencia-Artificial>

Youtube: <https://youtu.be/mxqNiozouvg>

8. Referências

- BATBAATAR, Erdenebileg; LI, Meijing; RYU, Keun Ho. Semantic-Emotion Neural Network for Emotion Recognition From Text. *IEEE Access*, v. 7, p. 111866–111878, 2019. Disponível em: <https://ieeexplore.ieee.org/document/8794541>. Acesso em: 15 setembro 2024.
- BARBOSA, Jardeson et al. Introdução ao Processamento de Linguagem Natural usando Python. In: ESCOLA REGIONAL DE INFORMÁTICA DO PIAUÍ – ERIPI, 3., 2017. Anais [...]. Teresina: Universidade Federal do Piauí, 2017. v. 1, n. 1, p. 336–360. Disponível em: https://www.facom.ufu.br/~wendelmelo/terceiros/tutorial_nltk.pdf. Acesso em: 14 setembro 2024.
- CARVALHO, Isabel Correia de et al. Analysing Customer-Support Trends in Social Networks through Dialogue Flow Discovery. In: *IBER_SPEECH 2024*, 2024. p. 56–60. DOI: 10.21437/IberSPEECH.2024-12.
- DOSCIATTI, L. F.; FERREIRA, K. S.; PARAÍSO, E. C. Construção e anotação de um corpus de notícias para análise de sentimentos. In: *SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM NATURAL*, 11., 2015, São Carlos. Anais [...]. São Carlos: Sociedade Brasileira de Computação, 2015. p. 1–6. Disponível em: <https://aclanthology.org/W15-5616.pdf>. Acesso em: 16 setembro 2024.
- ESBALTAR, Thyerre. Desenvolvimento de uma aplicação que analisará dados do Twitter. 2021. 33 f. Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Software) – Centro Universitário de Anápolis – UniEVANGÉLICA, Anápolis, 2021. Disponível em: <https://rincon061.org/bitstream/aee/19632/1/TCC1%20AliThyerre%206778.pdf>. Acesso em: 12 setembro 2024.
- FIGUEIREDO, Elaine; CATINI, Rita; MENDES, Leonardo. Mineração de textos: análise de sentimento em redes sociais – revisão sistemática. In: *WORKSHOP DE COMPUTAÇÃO APLICADA*, 5., 2018, Campo Limpo Paulista. Anais [...]. Campo Limpo Paulista: Centro Universitário Campo Limpo Paulista, 2018. p. 1–6. Disponível em: https://www.cc.faccamp.br/anaisdowcf/edicoes_antteriores/wcf2018/arquivos/04/paper_04.pdf. Acesso em: 13 setembro 2024.
- GUIMARÃES, L. M. S.; MEIRELES, M. R. G.; ALMEIDA, P. E. M. de. Avaliação das etapas de pré-processamento e de treinamento em algoritmos de classificação de textos no contexto da recuperação da informação. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 24, n. 1, p. 169–190, 2019. DOI: <https://doi.org/10.1590/1981-5344/3505>.

- HAMMES, Luiz Otávio Alves; FREITAS, Larissa Astrogildo de. Utilizando BERTimbau para a Classificação de Emoções em Português. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (STIL), 2021, Porto Alegre. Anais [...]. Porto Alegre: SBC, 2021. DOI: <https://doi.org/10.5753/stil.2021.17784>. Disponível em: https://www.researchgate.net/publication/356825341_Utilizando_BERTimbau_para_a_Classificacao_de_Emocoes_em_Portugues. Acesso em: 17 maio 2025.
- KANSAON, Daniel; BRANDÃO, Michele; PINTO, Saulo. Analysis of classification algorithms for emotion detection in Brazilian Portuguese tweets. iSys: Revista Brasileira de Sistemas de Informação, v. 12, n. 3, p. 116–138, 2019. Disponível em: <https://journals-sol.sbc.org.br/index.php/isys/article/view/600>. Acesso em: 14 setembro 2024.
- LUDERMIR, Teresa Bernarda. Inteligência artificial e aprendizado de máquina: estado atual e tendências. Estudos Avançados, São Paulo, v. 35, n. 101, p. 85–94, 2021. Disponível em: <https://www.scielo.br/j/ea/a/wXBdv8yHBV9xHz8qG5RCgZd>. Acesso em: 14 setembro 2024.
- PALOMINO, M. A.; AIDER, F. Evaluating the effectiveness of text pre-processing in sentiment analysis. Applied Sciences, Basel, v. 12, n. 8765, 2022. DOI: <https://doi.org/10.3390/app12178765>. Acesso em: 15 maio 2025.
- SCHAEFER, R.; STEDE, M. Argument mining on Twitter: a survey. Information Technology, v. 63, n. 1, p. 45–58, 2021.
- ŠNAJDER, Jan. Social media argumentation mining: The quest for deliberateness in raucousness, 2016.