

Detecção de Comportamentos Depressivos e Suicidas no Twitter por Aprendizado de Máquina usado técnica de PLN

**Quele da Silva Andrade¹, Jeovane dos Santos Santos¹,
Franklin Andrade de Brito¹, Camila Bezerra da Silva¹**

¹Departamento de Ciências Exatas e Tecnológicas - Universidade Federal do Recôncavo da Bahia (UFRB)

Caixa Postal 710 – 44.380-000 – Cruz das Almas – BA - Brasil

queleandrade@aluno.ufrb.edu.br, jeovanessantos@aluno.ufrb.edu.br

frank.engcet@gmail.com, camilabezerra@ufrb.edu.br

Abstract. *This study presents an automated approach for collecting and analyzing textual data from social networks to detect emotions related to depressive behaviors and suicidal ideation. Using text mining and natural language processing, Twitter posts were analyzed and classified into six emotions: joy, disgust, fear, anger, surprise, and sadness. The model achieved excellent results, with AUC values above 0.99 across all categories and high precision in detecting negative emotions such as anger and fear (95.6%). The confusion matrix indicated consistent classification, supporting the use of such technologies as complementary tools for mental health monitoring on digital platforms.*

Keywords *Text mining, Mental health, Emotion detection, Social networks, Natural language processing.*

Resumo. *Este estudo apresenta uma abordagem automatizada para coleta e análise de dados textuais em redes sociais, com foco na detecção de emoções associadas a comportamentos depressivos e ideação suicida. Foram aplicadas técnicas de mineração de texto e processamento de linguagem natural para classificar postagens do Twitter em seis emoções: alegria, desgosto, medo, raiva, surpresa e tristeza. O modelo obteve desempenho excelente, com valores de AUC acima de 0,99 e alta precisão na detecção de emoções negativas como raiva e medo (95,6%). A matriz de confusão revelou classificações coerentes, indicando o potencial do uso dessas tecnologias como ferramentas complementares no monitoramento da saúde mental em plataformas digitais.*

Palavras-Chave *Mineração de texto, Saúde mental, Detecção de emoções, Redes sociais, Processamento de linguagem natural.*

1. Introdução

O Twitter se destaca como um espaço dinâmico de discussão sobre os mais diversos temas, desde acontecimentos cotidianos até questões sociais, políticas e culturais. Nesse ambiente, os usuários expressam espontaneamente opiniões, sentimentos e emoções. Esse comportamento tem motivado pesquisas que aplicam Técnicas de Processamento de Linguagem Natural (PLN) e Aprendizado de Máquina (AM) para analisar automaticamente tais manifestações, com o objetivo de identificar

padrões comportamentais e sinais de sofrimento psíquico, de forma contínua e não intrusiva [Nijhawan et al. 2022]. Estudos recentes, como os de [Mirhatheri et al. 2024], [Sawhney et al. 2021b] e [Salas-Zarate et al. 2024], mostram a eficácia de modelos baseados em PLN e AM na detecção precoce de comportamentos relacionados à depressão e ideação suicida, destacando o papel preditivo das emoções textuais em redes sociais.

Este trabalho tem como principal objetivo desenvolver uma abordagem automatizada para a coleta e análise de dados textuais em redes sociais, com foco na detecção de emoções associadas a comportamentos depressivos e sinais de ideação suicida. A intenção não é a construção imediata de protótipos, mas sim a extração e categorização de sentimentos em postagens do *Twitter*, permitindo identificar padrões emocionais relevantes. Os resultados obtidos poderão subsidiar o desenvolvimento de ferramentas voltadas ao monitoramento e apoio a indivíduos em sofrimento psíquico, ampliando o potencial de aplicação das tecnologias de mineração de texto na área da saúde mental.

Este artigo está organizado em cinco seções, além desta introdução. A seção 2 apresenta a revisão da literatura, incluindo a estratégia de busca, os critérios de seleção e a análise dos estudos. Também são descritos os principais trabalhos encontrados, destacando suas abordagens técnicas, bases de dados utilizadas e resultados obtidos. A seção 3 apresenta a metodologia, *dataset*, métricas e modelos utilizados. A seção 4 apresenta os resultados alcançados neste estudo. Por fim, a seção 5 discute os avanços, limitações e tendências observadas nas pesquisas revisadas, além de sintetizar as contribuições do trabalho e sugerir direções para investigações futuras.

2. Revisão da literatura

A revisão da literatura foi realizada exclusivamente na base Scopus sobre o período de 2020 a 2025, escolhida por sua abrangência e indexação de periódicos de relevância científica nas áreas essenciais do estudo. A busca focou na detecção de emoções em textos do Twitter, com ênfase em comportamentos depressivos e ideação suicida, aplicando técnicas de PLN e AM. A seguir, apresenta-se a string de busca completa:

```
TITLE-ABS-KEY (( "Twitter") AND ( "suicidal ideation"OR "depressive behavior") AND ( "emotion detection") AND ( "natural language processing"OR "NLP") AND ( "machine learning") AND ( "mental health"))
```

Como resultado, foram identificados 18 artigos no total. Desses, 7 estavam disponíveis em acesso aberto. A partir de uma análise baseada em critérios de relevância, qualidade metodológica, uso de *datasets* públicos e aplicabilidade dos métodos, foi realizada uma triagem aprofundada, resultando em um conjunto final de 6 artigos selecionados para análise detalhada.

A Tabela 1 detalha as principais características dos artigos selecionados, incluindo ano de publicação, conjunto de dados utilizados, métricas de avaliação, modelos empregados e técnicas associadas.

Tabela 1. Classificação e Características dos Artigos Utilizados na Fundamentação da Literatura

Referência	Ano	Conjunto de Dados	Métodos aplicados	Métricas de Avaliação	Técnicas Utilizadas
[Mirhatheri et al. 2024]	2024	Twitter e Reddit	LSTM (Long Short-Term Memory), TCN (Temporal Convolutional Networks)	Acurácia, recall e F1-score.	Autoatenção, Análise Estatística, Análise Semântica
[Luu e Follmann 2023]	2023	Twitter	TextBlob, AFINN, SentimentR para análise de sentimento	Polaridade, subjetividade e correlação estatística (coeficiente de Pearson)	Correlação estatística (Pearson), Pré-processamento de texto (remoção de stopwords, links, normalização), Geração de nuvem de palavras
[Salas-Zarate et al. 2024]	2024	Twitter	Classificação multiclasses, Correlação entre emoções e sintomas de depressão	Precisão, recall, correlação entre emoções e sintomas de depressão	Extração de sentimentos com NLP, Análise semântica e estatística, Modelo baseado em PHQ-9
[Alhuzali et al. 2022]	2022	Twitter	SenticNet 6 para análise de sentimentos, SpanEmo para reconhecimento de emoções, CTM (Combined Topic Modeling) para modelagem de tópicos	Análise de sentimentos (polaridade), reconhecimento de emoções e modelagem de tópicos	Geolocalização de tweets, Tokenização e Normalização, Extração de características, Classificação de sentimentos e emoções
[Sawhney et al. 2021b]	2021	Twitter	Graph Networks (GNN), Hawkes Process	Precisão, recall, F1-score e AUC-ROC	Hyperbolic Graph Convolutions, HEAT (Hawkes Temporal Emotion Aggregation), Embeddings BERT
[Sawhney et al. 2021a]	2021	Twitter	LSTM sensível ao tempo (TSE-LSTM), Convoluçãoes adaptativas para fases emocionais (PAC)	Precisão, recall, F1-score e AUC-ROC	Plutchik Transformer para reconhecimento de emoções, Modelagem de progressão emocional, Atenção temporal

Entre os trabalhos identificados, destaca-se o modelo AL-BTCN proposto por [Mirhatheri et al. 2024], que integra redes neurais LSTM e TCN com mecanismos de autoatenção para aprimorar a detecção de padrões linguísticos associados à ideação suicida. Avaliado com dados do Twitter e Reddit, o modelo obteve desempenho superior a 94% em métricas como acurácia e F1-score, oferecendo subsídios relevantes para estratégias de monitoramento digital. Na mesma linha, [Sawhney et al. 2021a] introduzem o modelo PHASE, que combina Plutchik Transformer, LSTM temporal e convoluções adaptativas para analisar a progressão emocional de usuários ao longo do tempo, identificando fases críticas que antecedem postagens que indicam risco suicida e ressaltando a importância do fator temporal.

Complementando essas abordagens, [Salas-Zarate et al. 2024] propõem o sistema Mental-Health, voltado à classificação de níveis de depressão com base em emoções extraídas de tweets, utilizando o questionário PHQ-9 como referência. O modelo obteve alta precisão, sobretudo na identificação de casos moderados e severos, reforçando

o potencial do Twitter como fonte complementar no rastreamento da saúde mental. Além dos estudos focados em ideação suicida e depressão, pesquisas como as de [Luu e Follmann 2023] e [Alhuzali et al. 2022] exploram o impacto de eventos externos, como a pandemia de COVID-19, na expressão emocional *on-line*. O primeiro analisou variações no sentimento público conforme o avanço da doença nos EUA, enquanto o segundo mapeou mais de 500 mil tweets no Reino Unido, aplicando modelagem de tópicos e emoções para acompanhar mudanças regionais e temporais no discurso público.

Ainda nesse contexto, [Sawhney et al. 2021b] propõem um modelo baseado em Graph Neural Networks que incorpora o histórico emocional e as interações sociais dos usuários, oferecendo uma avaliação mais abrangente do risco de suicídio ao considerar fatores sociais e temporais. Essas investigações demonstram o crescente interesse e o potencial das técnicas de PLN e AM na detecção automática de estados emocionais em redes sociais, ressaltando a importância de integrar componentes linguísticos, temporais e sociais na identificação precoce e escalável de sinais de sofrimento psíquico.

3. Materiais e Métodos

Nesta seção, apresentamos os dados utilizados, a metodologia de mineração de texto, os modelos de classificação aplicados, a coleta via API do Twitter e as métricas usadas para avaliar os resultados.

3.1. Dataset

O conjunto de dados utilizado para o treinamento foi rotulado manualmente, totalizando 1.718 exemplos textuais classificadas em seis das sete emoções básicas propostas por Ekman: alegria, tristeza, raiva, medo, nojo e surpresa. A emoção “desprezo” foi excluída devido à baixa frequência e à dificuldade de rotulagem confiável. As emoções consideradas são reconhecidas como universais e biologicamente fundamentadas, o que justifica sua aplicação em tarefas de classificação emocional em PLN [Ekman 1992, Ekman 1999].

A distribuição das classes no conjunto de treinamento foi: medo (556), surpresa (264), raiva (251), tristeza (251), alegria (249) e nojo (147), evidenciando um desbalanceamento que pode afetar o desempenho de modelos supervisionados. Já para avaliação, foi utilizada uma base de teste com 340 frases do Twitter, visando a representar um cenário textual ruidoso e não estruturado. A classificação resultou em: medo (164), surpresa (46), raiva (47), tristeza (58), alegria (20) e nojo (5). O desbalanceamento observado em ambos os conjuntos pode comprometer a capacidade discriminativa do modelo, especialmente em classes minoritárias. Para contornar esse problema, foram utilizadas métricas robustas à desproporcionalidade, como F1-score, AUC (Área sob a Curva), precisão e recall.

3.2. Mineração de texto

A mineração de texto é uma área que visa a extrair informações relevantes e padrões significativos a partir de grandes volumes de dados textuais não estruturados. No contexto deste trabalho, a mineração de texto foi aplicada a postagens do Twitter com o objetivo de identificar sinais de comportamentos depressivos ou suicidas. Esse processo envolve múltiplas etapas, desde a coleta dos textos até a análise dos resultados produzidos por modelos de aprendizado de máquina.

Uma das bases da mineração de texto é o Processamento de Linguagem Natural, um conjunto de técnicas computacionais que permite a interpretação e manipulação automática de linguagem humana. O PLN desempenha papel essencial na limpeza, normalização e preparação dos dados textuais antes de sua utilização em modelos preditivos.

A Figura 1 apresenta o fluxograma com todas as etapas do processo de mineração de texto adotado neste trabalho.

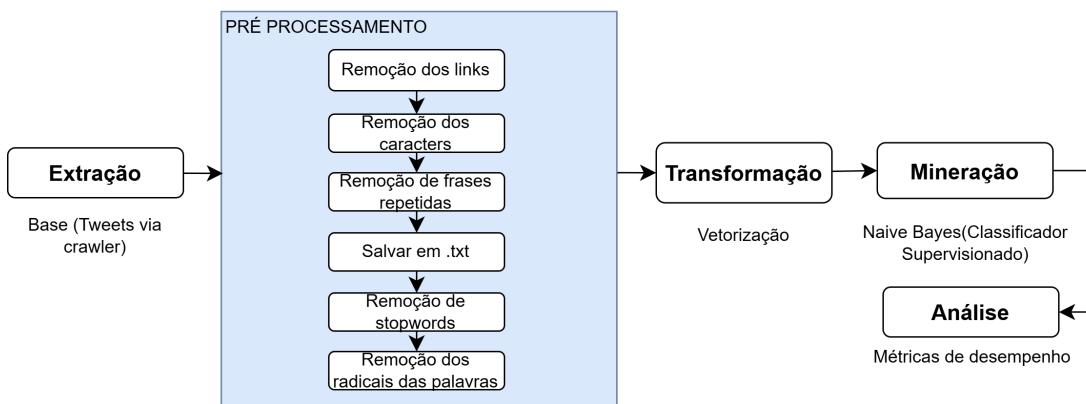


Figura 1. Metodologia de Mineração de Texto Aplicado ao Twitter

O processo inicia-se pela extração, em que tweets são extraídos por meio de um *crawler* utilizando a API do Twitter, com base em palavras-chave e *hashtags* associadas a sentimentos ou transtornos emocionais.

Em seguida, realiza-se o pré-processamento, etapa crítica que envolve a limpeza dos dados textuais. Nesta fase são aplicadas as seguintes operações: remoção de links e caracteres especiais, eliminação de frases repetidas, os textos processados são salvos em arquivos texto, e após ocorre a remoção de *stopwords* e de radicais das palavras.

A transformação consiste na vetorização dos textos, convertendo o conteúdo textual em representações numéricas. Já a etapa de Mineração envolve a aplicação de um modelo supervisionado de classificação, neste caso, o algoritmo Naive Bayes. Este classificador probabilístico é baseado no Teorema de Bayes e assume independência entre os atributos, sendo amplamente utilizado para tarefas de classificação de texto por sua simplicidade e bom desempenho.

Por fim, na análise, são utilizadas métricas de avaliação para mensurar o desempenho do classificador, como F1-score, precision, recall, curva-ROC e Matriz de Confusão.

3.3. Modelo utilizado

Foi utilizado o classificador Naive Bayes, baseado no Teorema de Bayes e amplamente aplicado em machine learning por sua simplicidade e eficiência. O algoritmo tem bom desempenho em tarefas de classificação com muitos dados e alta dimensionalidade, mesmo com conjuntos reduzidos de treino [Becker 2019], sendo comum em classificação de textos. Ele estima, via probabilidade, a chance de um elemento pertencer a uma classe específica, conforme apresentado por [Ray 2016]:

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)} \quad (1)$$

Equação 1: Fórmula do Teorema de Bayes.

Onde, segundo [Ray 2016], as incógnitas da equação acima são:

- $P(c | x)$: é a probabilidade posterior da classe (c , alvo) dado o preditor (x , atributos).
- $P(c)$: é a probabilidade original da classe.
- $P(x | c)$: é a probabilidade do preditor dada a classe.
- $P(x)$: é a probabilidade marginal do preditor.

3.4. API

Para a coleta de dados deste estudo, foi utilizada a API do *Twitter* por meio da biblioteca *Tweepy*. O acesso à API requer a criação de uma conta de desenvolvedor na plataforma *Twitter*, seguida pela criação de um aplicativo, onde são geradas as seguintes credenciais:

- *Access token* (Token de acesso)
- *Access token secret* (Token secreto de acesso)
- *Get api key* (Obter chave da api)
- *Get api secret key* (Obter chave secreta da api)

Essas credenciais permitem autenticar a aplicação, concedendo acesso ao conteúdo público do *Twitter*. O algoritmo desenvolvido utiliza essas informações para efetuar o login automático na API e realizar buscas por hashtags relacionadas a sentimentos, como #triste, #feliz, #raiva, entre outras.

A seguir, apresentamos o fluxograma do algoritmo, na Figura 2, que ilustra cada etapa da conexão com a API e da coleta dos dados:

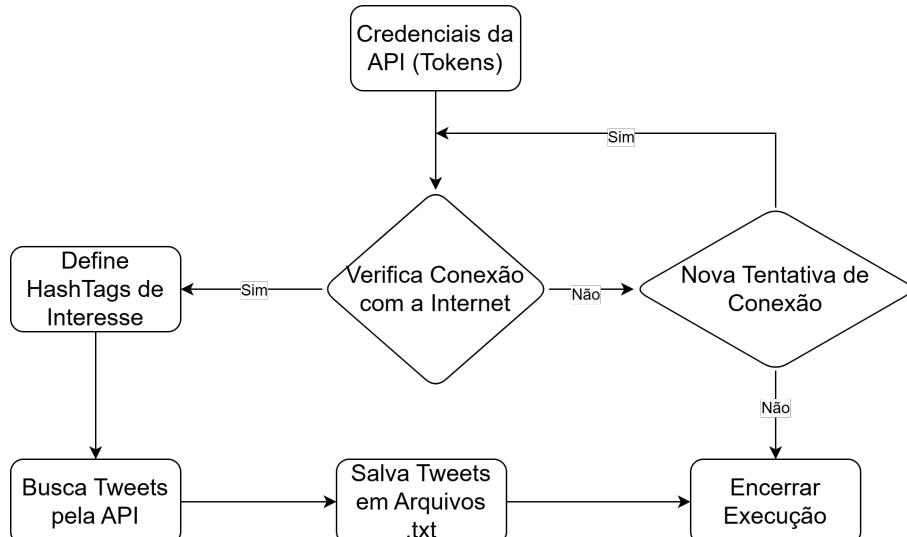


Figura 2. Fluxograma do algoritmo de busca

O fluxograma apresentado na Figura 2 ilustra o processo automatizado de coleta de dados via API do *Twitter*. A execução inicia com a autenticação, por meio das Credenciais da API (*Tokens*). Em seguida, o sistema verifica se há conexão com a internet (Verifica Conexão com a Internet). Caso positivo, o algoritmo prossegue para a etapa de definição de termos de busca (Define *Hashtags* de Interesse), responsáveis por orientar a coleta de publicações. Com os critérios estabelecidos, os dados são obtidos (Busca *Tweets* pela API) e imediatamente salvos localmente (Salva *Tweets* em arquivo texto). Se a conexão falhar, o sistema entra em modo de recuperação (Nova Tentativa de Conexão) e, após um número limitado de tentativas, finaliza a execução de forma segura (Encerrar Execução). Esse fluxo garante robustez ao processo, permitindo retomadas automáticas em caso de falhas de rede.

3.5. Métricas

A avaliação de desempenho do modelo é uma etapa fundamental para verificar sua eficácia na tarefa de classificação. Como o problema tratado neste trabalho apresenta dados desbalanceados, a métrica de acurácia torna-se pouco confiável. Dessa forma, utilizamos métricas mais robustas, descritas a seguir:

- **Precision (Precisão):** mede a proporção de classificações positivas corretas (TP), ajudando a reduzir falsos positivos (FP).
- **Recall (Sensibilidade):** avalia a capacidade do modelo de reconhecer corretamente exemplos positivos, sendo crucial para evitar falsos negativos (FN) em contextos críticos, como a detecção de depressão.
- **F1-Score:** equilibra precisão e recall por meio de uma média harmônica, essencial em conjuntos de dados desbalanceados.
- **Curva ROC e AUC:** A Curva ROC representa graficamente a relação entre a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR) para diferentes limiares. A AUC quantifica esse desempenho: quanto mais próxima de 1, melhor a capacidade do modelo em distinguir as classes.

A seguir, apresentam-se as fórmulas matemáticas associadas a cada uma dessas métricas:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1-score} = \frac{2 \cdot (\text{Recall} \cdot \text{Precision})}{\text{Recall} + \text{Precision}} \quad (4)$$

$$\text{AUC (Área sob a Curva ROC)} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}) \quad (5)$$

4. Resultados

A seguir, apresentam-se a análise dos resultados e, posteriormente, a discussão da revisão da literatura, contextualizando os achados do estudo em relação às pesquisas existentes.

4.1. Análise dos resultados

A Tabela 2 apresenta as métricas de desempenho do modelo na classificação automática de emoções em postagens de redes sociais. Os resultados demonstram alta capacidade discriminativa, com valores de AUC superiores a 0.99 em todas as classes emocionais, o que é relevante em contextos de saúde mental.

Tabela 2. Tabela de resultados das métricas de desempenho.

Classes	Precision	Recall	F1-Score	AUC
Alegria	0.80	0.96	0.87	0.99
Desgosto	0.99	0.59	0.74	1.00
Medo	0.91	0.97	0.94	0.99
Raiva	0.90	0.88	0.89	0.99
Surpresa	0.98	0.86	0.92	0.99
Tristeza	0.92	0.94	0.93	1.00

As classes *Tristeza* e *Desgosto* destacam-se com AUC de 1.00, indicando que o modelo diferencia essas emoções com altíssima precisão. No caso da tristeza (Precision: 0.92, Recall: 0.94), esse desempenho é especialmente promissor para aplicações voltadas à detecção de quadros depressivos.

A classe *Alegria* apresentou o segundo menor F1-score (0.87), impactada por uma precisão relativamente baixa (0.80), apesar do alto recall (0.96). Isso indica que o modelo identifica a maioria das postagens alegres, mas comete erros ao rotular como alegria textos que pertencem a outras emoções. Essa imprecisão pode estar associada à ambiguidade semântica de expressões alegres em redes sociais, muitas vezes marcadas por ironia, sarcasmo ou sobreposição com outras emoções positivas.



Figura 3. Matriz de Confusão

A classe *Raiva* obteve o melhor desempenho absoluto na Matriz de Confusão (Figura 3), com 561 classificações corretas em 566 exemplos ($\approx 95,6\%$). *Medo* e

Desgosto também alcançaram altas taxas de acerto (95,6% e 93,6%, respectivamente), evidenciando a capacidade do modelo de captar expressões linguísticas associadas a emoções negativas importantes para identificar sinais de sofrimento psíquico.

O principal ponto de confusão ocorre entre *Raiva* e *Surpresa*, com 17 casos mal classificados. Essa sobreposição pode indicar semelhanças linguísticas em contextos de surpresa negativa. Já a classe *Alegria* apresenta dispersão em sua classificação: foi confundida com medo (15), raiva (14) e tristeza (5), o que reforça a hipótese de ambiguidade semântica em postagens alegres.

Embora *Tristeza* tenha uma taxa de acerto satisfatória ($222/251 \approx 88,4\%$), apresenta confusões principalmente com medo (17) e raiva (8), o que pode refletir a complexidade emocional de posts associados a depressão. Ainda assim, o modelo mostra estabilidade frente à variação no número de exemplos por classe, mantendo desempenho robusto mesmo nas classes com menos exemplos, como *Medo* e *Desgosto*.

A capacidade do modelo de identificar emoções como raiva, medo e tristeza com alta precisão é fundamental para detectar precocemente sinais de sofrimento psicológico. No entanto, as confusões entre emoções negativas e a ambiguidade na alegria representam desafios para uma caracterização emocional mais precisa, especialmente em contextos de diagnóstico e prevenção em saúde mental.

4.2. Análise da Revisão da Literatura

Este estudo avança na análise emocional em redes sociais por meio de PLN e aprendizado de máquina, complementando abordagens como AL-BTCN [Mirhatheri et al. 2024] e PHASE [Sawhney et al. 2021a], que focam em padrões temporais para identificar risco suicida. Diferentemente dessas propostas, nosso modelo destaca-se pela capacidade de classificar múltiplas emoções com precisão, atingindo AUC de 1.00 para tristeza e desgosto, além de manter desempenho consistente em outras classes conforme F1-score, precision, recall e curva-ROC. No entanto, a classificação da alegria enfrentou desafios semânticos, com erros que sugerem sobreposição com outras emoções positivas e nuances de ironia e sarcasmo.

Comparado ao Mental-Health [Salas-Zarate et al. 2024], que identifica níveis de depressão via tweets, nosso modelo se mostrou robusto diante da variação interclasse, garantindo estabilidade mesmo em emoções menos representadas, como medo e desgosto. Para avaliar seu desempenho, utilizamos métricas como F1-score, precision, recall, curva-ROC e Matriz de Confusão, essenciais para uma triagem emocional eficaz no monitoramento da saúde mental digital.

5. Conclusão e Trabalhos futuros

Este estudo apresentou uma abordagem baseada em mineração de texto para detecção de emoções em redes sociais, com foco em sinais de sofrimento psíquico. O modelo proposto obteve alta precisão, com valores de AUC superiores a 0.99, destacando sua capacidade discriminativa, especialmente para emoções negativas como raiva e medo (95,6% de acerto). A análise revelou padrões coerentes com a complexidade das manifestações emocionais digitais.

Apesar dos resultados promissores, desafios como ironia e sarcasmo requerem técnicas mais avançadas para melhorar a classificação. A análise longitudinal das

emoções pode identificar tendências associadas a transtornos depressivos, enquanto características multimodais (imagens e vídeos) podem enriquecer a compreensão emocional dos usuários.

A validação clínica é essencial para garantir aplicabilidade em contextos reais, viabilizando estudos colaborativos com profissionais da saúde. No longo prazo, espera-se desenvolver ferramentas de intervenção, como sistemas de alerta e plataformas de apoio. A adaptação do modelo para diferentes contextos culturais e linguísticos ampliará sua aplicabilidade, promovendo um impacto positivo na prevenção e tratamento de transtornos mentais.

Referências

- Alhuzali, H., Zhang, T., e Ananiadou, S. (2022). Emotions and topics expressed on twitter during the covid-19 pandemic in the united kingdom: Comparative geolocation and text mining analysis. *Journal of Medical Internet Research*, 24(10):e40323.
- Becker, L. (2019). Algoritmo de classificação naive bayes. Acesso em: 5 jun. 2025.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4):169–200.
- Ekman, P. (1999). Basic emotions. In Dalgleish, T. e Power, M., editors, *Handbook of Cognition and Emotion*, pages 45–60. John Wiley Sons.
- Luu, T. J. P. e Follmann, R. (2023). The relationship between sentiment score and covid-19 cases in the united states. *Journal of Information Science*, 49(6):1615–1630.
- Mirhatheri, S. L., Greco, S., e Shahbazian, R. (2024). A self-attention tcn-based model for suicidal ideation detection from social media posts. *Expert Systems with Applications*, 255:124855.
- Nijhawan, T., Attigeri, G., e Ananthakrishna, T. (2022). Stress detection using natural language processing and machine learning over social interactions. *Journal of Big Data*, 9(1):33.
- Ray, S. (2016). 6 passos fáceis para aprender o algoritmo naive bayes (com o código em python). Tradução de: *6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python)*. Acesso em: 5 jun. 2025.
- Salas-Zarate, R., Alor-Hernandez, G., Paredes-Valverde, M. A., del Pilar Salas-Zárate, M., Bustos-López, M., e Sánchez-Cervantes, J. L. (2024). Mental-health: An nlp-based system for detecting depression levels through user comments on twitter (x). *Mathematics*, 12(1926).
- Sawhney, R., Joshi, H., Flek, L., e Shah, R. R. (2021a). Phase: Learning emotional phase-aware representations for suicide ideation detection on social media. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2415–2428. Association for Computational Linguistics.
- Sawhney, R., Joshi, H., Shah, R. R., e Flek, L. (2021b). Suicide ideation detection via social and temporal user representations using hyperbolic learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2176–2190. Association for Computational Linguistics.