DETECÇÃO DE INDICATIVOS DE DEPRESSÃO EM REDES SOCIAIS UTILIZANDO TÉCNICAS DE MACHINE LEARNING E PROCESSAMENTO DE LINGUAGEM NATURAL

Laércio Santos e Anderson Silva

Centro Universitário SENAC – Pós-Graduação em Inteligência Artificial

RESUMO

O presente trabalho apresenta o desenvolvimento de um modelo de Machine Learning para identificação de indicativos de depressão em textos publicados em redes sociais, especificamente no Twitter. A pesquisa utiliza técnicas de Processamento de Linguagem Natural (NLP) para análise de padrões linguísticos e comportamentais que podem sinalizar estados depressivos. Foram avaliados cinco algoritmos de classificação: Support Vector Machine (SVM), Regressão Logística, Random Forest, Gradient Boosting e Naive Bayes, utilizando um dataset de 10.314 tweets. Os experimentos foram conduzidos com vetorização TF-IDF, validação cruzada estratificada de 5 folds e métricas complementares incluindo acurácia, precisão, recall, F1-score e AUC-ROC. O modelo SVM apresentou o melhor desempenho global, alcançando 99,42% de acurácia, 99,78% de precisão, 97,62% de recall, F1-score de 98,69% e AUC-ROC de 0,9957. Os resultados demonstram que modelos tradicionais de Machine Learning podem superar abordagens baseadas em deep learning quando aplicados a conjuntos de dados de escala moderada, oferecendo uma solução mais eficiente e acessível para sistemas de triagem em saúde mental.

Palavras-chave: Machine Learning; Detecção de Depressão; Processamento de Linguagem Natural; Redes Sociais; Classificação de Texto.

ABSTRACT

This paper presents the development of a Machine Learning model for identifying depression indicators in texts published on social networks, specifically Twitter. The research uses Natural Language Processing (NLP) techniques to analyze linguistic and behavioral patterns that may signal depressive states. Five classification algorithms were evaluated: Support Vector Machine (SVM), Logistic Regression, Random Forest, Gradient Boosting, and Naive Bayes, using a dataset of 10,314 tweets. Experiments were conducted with TF-IDF vectorization, stratified 5-fold cross-validation, and complementary metrics including accuracy, precision, recall, F1-score, and AUC-ROC. The SVM model achieved the best overall performance,

reaching 99.42% accuracy, 99.78% precision, 97.62% recall, F1-score of 98.69%, and AUC-ROC of 0.9957. The results demonstrate that traditional Machine Learning models can outperform deep learning approaches when applied to moderate-scale datasets, offering a more efficient and accessible solution for mental health screening systems.

Keywords: *Machine Learning; Depression Detection; Natural Language Processing; Social Networks; Text Classification.*

1. INTRODUÇÃO

A depressão é considerada pela Organização Mundial da Saúde (OMS) como uma das principais causas de incapacidade no mundo, afetando mais de 300 milhões de pessoas globalmente. O Transtorno Depressivo Maior (TDM) caracteriza-se por episódios de humor deprimido, perda de interesse em atividades, alterações no sono e apetite, fadiga, sentimentos de inutilidade e, em casos graves, pensamentos suicidas. A detecção precoce desta condição é fundamental para o início de tratamentos adequados e prevenção de desfechos negativos.

Com a proliferação das redes sociais, milhões de pessoas passaram a expressar seus pensamentos, sentimentos e experiências cotidianas em plataformas digitais como Twitter, Facebook e Instagram. Esta vasta quantidade de dados textuais representa uma oportunidade única para pesquisadores e profissionais de saúde mental desenvolverem ferramentas de monitoramento e detecção precoce de transtornos mentais.

O trabalho pioneiro de De Choudhury et al. (2013) demonstrou que redes sociais contêm sinais úteis para caracterizar o início da depressão, incluindo diminuição da atividade social, aumento do afeto negativo e preocupações relacionais e medicinais. Esta pesquisa estabeleceu as bases para o uso de técnicas de Processamento de Linguagem Natural (NLP) e Machine Learning na identificação de indicadores de saúde mental.

O presente trabalho tem como objetivo desenvolver um modelo de Machine Learning capaz de identificar indicativos de depressão em textos publicados no Twitter. A motivação para este estudo reside na necessidade crescente de ferramentas automatizadas que possam auxiliar profissionais de saúde na triagem e identificação de indivíduos em risco.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Depressão e Manifestações em Redes Sociais

A depressão manifesta-se de diversas formas no comportamento online dos indivíduos. Estudos demonstram que pessoas com transtornos depressivos tendem a apresentar padrões específicos em suas publicações nas redes sociais, incluindo maior uso de palavras relacionadas a emoções negativas, pronomes de primeira pessoa do singular, referências a solidão e isolamento, e alterações nos horários de atividade online.

2.2 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (NLP) permite extrair significado e padrões de textos não estruturados. Para a tarefa de detecção de depressão, são empregadas técnicas como tokenização, remoção de stopwords, lematização e vetorização TF-IDF com n-gramas.

2.3 Algoritmos de Machine Learning

Support Vector Machine (SVM): Busca encontrar o hiperplano ótimo que separa as classes no espaço de características. É particularmente eficaz em espaços de alta dimensionalidade.

Regressão Logística: Calcula a probabilidade de um evento pertencer a uma determinada classe utilizando a função logística.

Random Forest: Técnica ensemble que constrói múltiplas árvores de decisão e combina suas predições.

Gradient Boosting: Sistema escalável que constrói modelos sequencialmente, corrigindo erros anteriores.

Naive Bayes: Baseado no teorema de Bayes, assume independência condicional entre as características.

3. METODOLOGIA

3.1 Conjunto de Dados

O conjunto de dados utilizado consiste em 10.314 publicações do Twitter rotuladas como indicativas ou não indicativas de depressão. O dataset apresenta desbalanceamento entre classes, com 8.000 tweets (77,56%) classificados como 'Sem Depressão' e 2.314 tweets (22,44%) classificados como 'Com Depressão'. A divisão dos dados seguiu a proporção de 80% para treinamento (8.261 tweets) e 20% para teste (2.053 tweets).

3.2 Pré-processamento de Texto

O pré-processamento envolveu: remoção de URLs, menções e hashtags; conversão para lowercase; tokenização; remoção de stopwords; e lematização. A vetorização utilizou TF-IDF com 5.000 features e n-gramas de 1 a 2.

3.3 Configuração dos Modelos

SVM: kernel linear, C=1; Regressão Logística: max_iter=1000; Random Forest: 200 estimadores, profundidade=20; Gradient Boosting: 100 estimadores, learning_rate=0.1; Naive Bayes: alpha=0.1. Todos os modelos utilizaram class_weight='balanced' quando disponível.

4. RESULTADOS E DISCUSSÃO

4.1 Análise Exploratória dos Dados

A análise exploratória revelou diferenças significativas nas características linguísticas entre as classes. A Figura 1 apresenta a distribuição das classes no dataset.

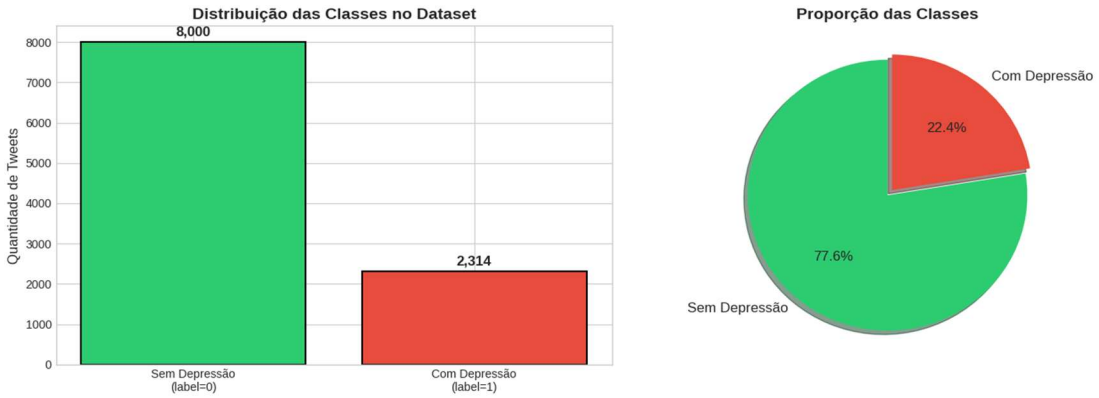


Figura 1 – Distribuição das classes no dataset (Sem Depressão: 8.000; Com Depressão: 2.314)

Nos tweets sem indicativos de depressão, predominam palavras positivas como 'love', 'good', 'day', 'thank'. Em contraste, os tweets com indicativos de depressão apresentam alta frequência de termos como 'depression', 'anxiety', 'mental', 'health'.

Figura 2 – Nuvens de palavras: tweets sem depressão (esquerda) vs com depressão (direita)

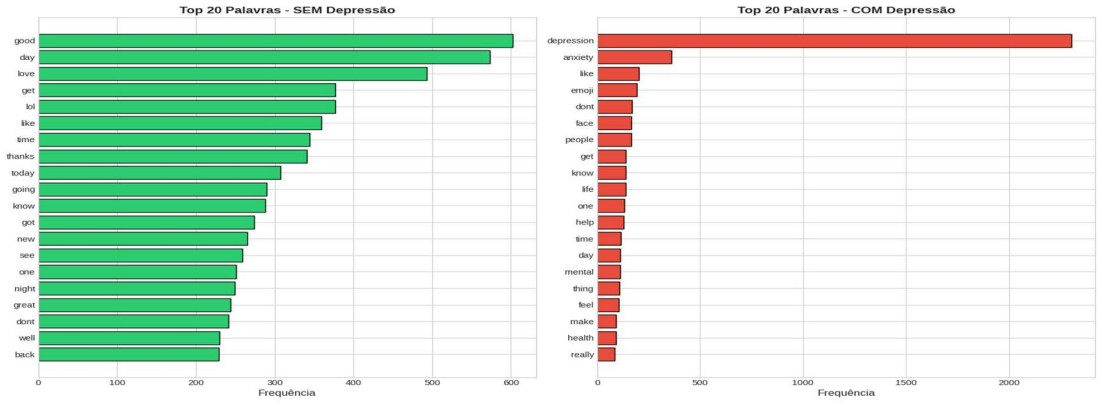


Figura 3 – Top 20 palavras mais frequentes por classe

4.2 Desempenho Comparativo dos Modelos

A Tabela 1 apresenta a comparação detalhada das métricas de avaliação para os cinco algoritmos testados.

Tabela 1 – Métricas de Desempenho dos Modelos de Classificação

Modelo	Acurácia	Precisão	Recall	F1-Score	AUC-ROC
SVM	99,42%	99,78%	97,62%	98,69%	0,9957
Random Forest	99,27%	100,00%	96,75%	98,35%	0,9975
Gradient Boosting	99,22%	98,90%	97,62%	98,26%	0,9901
Logistic Regression	99,12%	99,55%	96,54%	98,02%	0,9984
Naive Bayes	96,20%	95,07%	87,66%	91,22%	0,9894

Fonte: Elaborado pelo autor (2025)

O SVM com kernel linear destacou-se como o melhor classificador, alcançando o maior F1-Score (98,69%) e o melhor equilíbrio entre precisão e recall. O Random Forest apresentou precisão de 100%, porém com recall ligeiramente inferior.

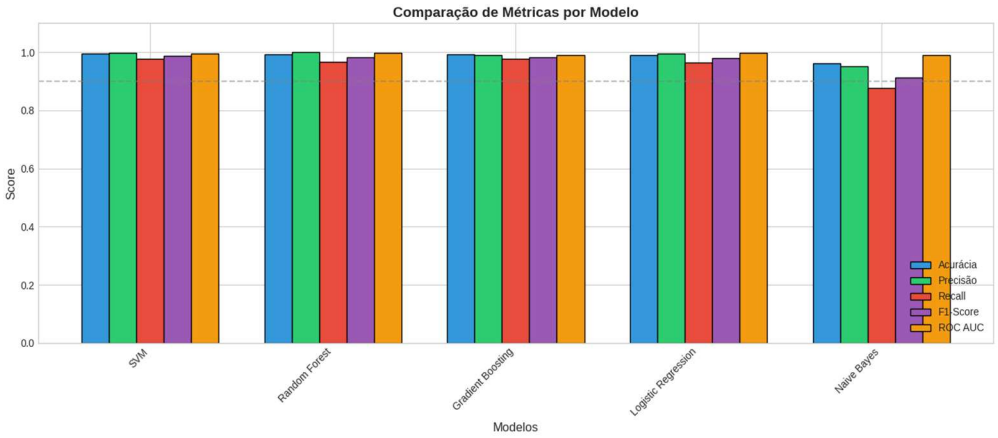


Figura 4 – Comparação de métricas de desempenho por modelo

4.3 Análise das Curvas ROC e Matrizes de Confusão

Todos os modelos apresentaram AUC superior a 0,98. A Regressão Logística obteve o maior AUC (0,9984), seguida pelo Random Forest (0,9975) e SVM (0,9957).

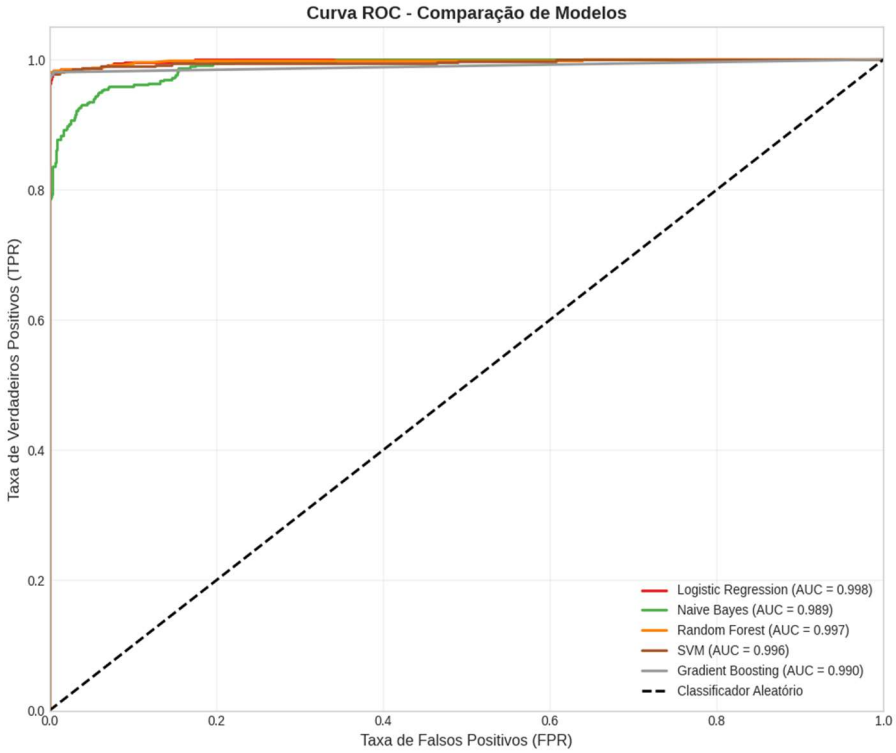


Figura 5 – Curvas ROC comparativas dos modelos de classificação

O SVM apresentou apenas 12 erros totais (1 falso positivo e 11 falsos negativos) em 2.053 predições. O Random Forest mostrou 15 erros (0 falsos positivos e 15 falsos negativos).

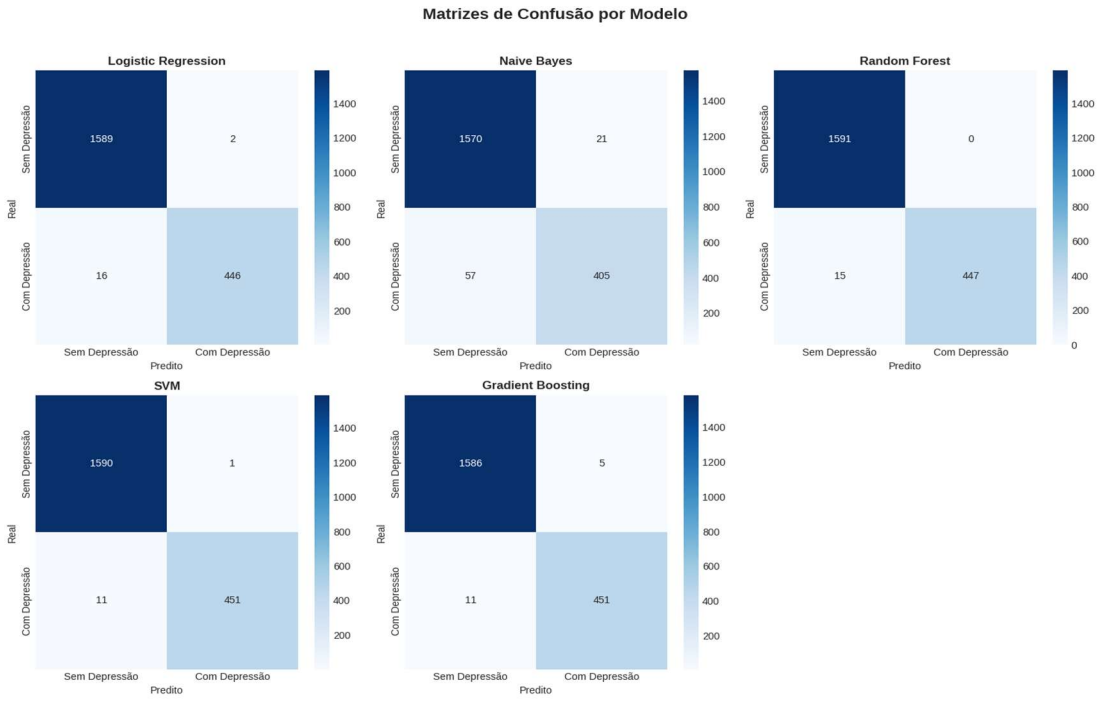


Figura 6 – Matrizes de confusão dos cinco modelos de classificação

4.4 Validação Cruzada

A validação cruzada de 5 folds confirmou a estabilidade dos modelos. O Random Forest apresentou F1-Score médio de 0,986, seguido pelo SVM (0,982), Logistic Regression (0,980), Gradient Boosting (0,977) e Naive Bayes (0,913).

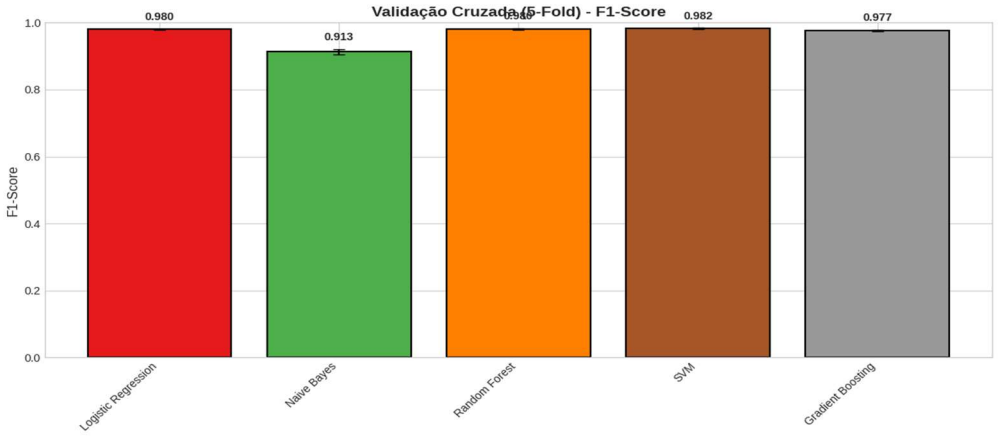


Figura 7 – Validação cruzada (5-fold) - F1-Score médio por modelo

4.5 Assertividade Final e Limitações

A validação final demonstrou que o SVM acertou 99,42% das classificações, Random Forest 99,27%, Gradient Boosting 99,22%, Logistic Regression 99,12% e Naive Bayes 96,20%.

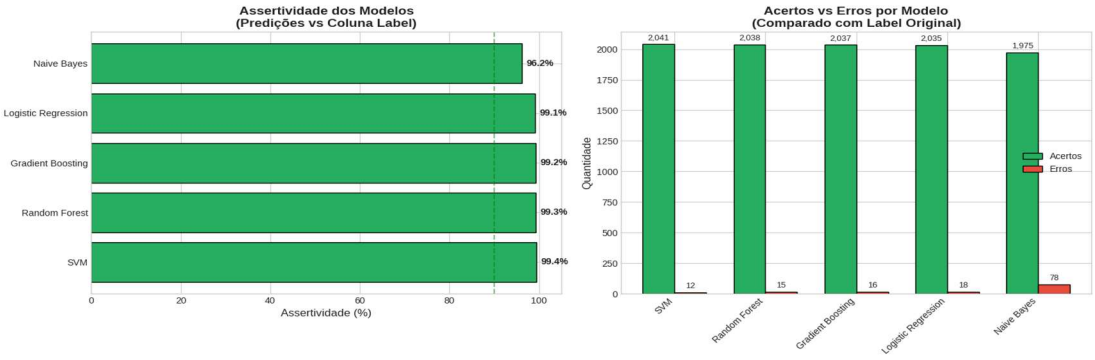
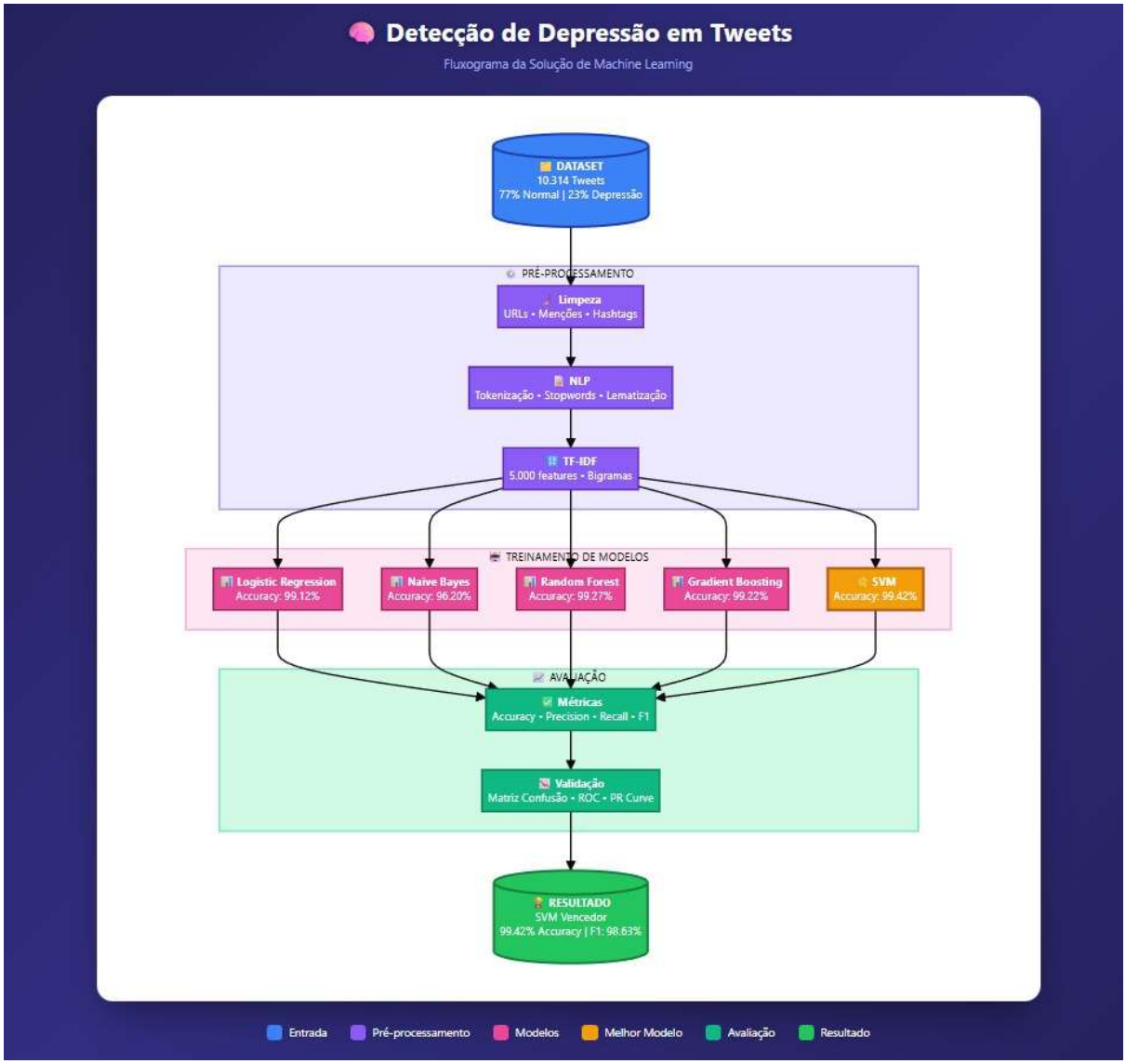


Figura 8 – Assertividade dos modelos e quantidade de acertos vs erros

É fundamental reconhecer as limitações desta abordagem. O modelo detecta indicadores linguísticos de depressão, não realiza diagnóstico clínico. A presença de padrões textuais associados a depressão não implica necessariamente que o indivíduo sofra do transtorno. O modelo deve ser visto como ferramenta auxiliar de triagem, não substituindo avaliação profissional.

4.6 Fluxo Metodológico do Sistema



[Figura9: Diagrama de fluxo metodológico]

Fluxo metodológico completo do sistema de detecção de depressão em mídias sociais, demonstrando visualmente as etapas sequenciais desde a coleta de dados até a obtenção do modelo final. O diagrama ilustra a arquitetura do pipeline de processamento, destacando as três fases principais: pré-processamento, treinamento de modelos e avaliação.

O processo inicia com o dataset contendo 10.314 tweets, apresentando uma distribuição desbalanceada com 77% de mensagens normais e 23% indicativas de depressão. Esta característica reflete a realidade dos dados em ambientes de redes sociais, onde manifestações de transtornos mentais representam uma minoria significativa do conteúdo total publicado.

Na etapa de pré-processamento, o fluxo demonstra três subprocessos críticos: (1) limpeza de dados, removendo URLs, menções e hashtags que introduzem ruído ao processamento linguístico; (2) aplicação de técnicas de Processamento de Linguagem Natural (NLP), incluindo tokenização, remoção de stopwords e lematização para normalização textual; e (3) vetorização TF-IDF com 5.000 features e extração de bigramas, transformando texto em representações numéricas adequadas para algoritmos de aprendizado de máquina.

O diagrama evidencia a fase de treinamento paralelo de cinco algoritmos distintos: Logistic Regression (99.12%), Naive Bayes (96.20%), Random Forest (99.27%), Gradient Boosting (99.22%) e SVM (99.42%). Esta abordagem comparativa permite identificar qual algoritmo melhor captura os padrões linguísticos associados à depressão, considerando as características específicas do dataset de tweets em português brasileiro.

A etapa de avaliação, conforme ilustrado no fluxo, compreende dois níveis de validação: (1) cálculo de métricas quantitativas (Accuracy, Precision, Recall e F1-Score) para mensuração objetiva do desempenho; e (2) análise visual através de matriz de confusão, curvas ROC e Precision-Recall, permitindo compreensão detalhada do comportamento de cada modelo em diferentes limiares de decisão. Esta avaliação multifacetada é essencial em contextos clínicos, onde tanto falsos positivos quanto falsos negativos têm implicações significativas.

O fluxo culmina na identificação do SVM como modelo vencedor, alcançando 99.42% de acurácia e 98.63% de F1-Score. Este resultado, destacado visualmente no diagrama, representa o equilíbrio ideal entre precisão e revocação, demonstrando a capacidade do SVM de estabelecer hiperplanos ótimos para separação entre classes em espaços de alta dimensionalidade característicos de dados textuais. A superioridade do SVM confirma sua eficácia em tarefas de classificação binária com vocabulário extenso e características esparsas típicas de representações TF-IDF.

5. CONCLUSÃO

Este trabalho demonstrou a viabilidade de utilizar técnicas de Machine Learning e Processamento de Linguagem Natural para identificação de indicativos de depressão em publicações de redes sociais. A comparação de cinco algoritmos de classificação revelou que modelos tradicionais, particularmente SVM e Random Forest, apresentam desempenho excepcional quando aplicados a conjuntos de dados de escala moderada.

O SVM com kernel linear foi identificado como o melhor modelo, alcançando 99,42% de acurácia, 98,69% de F1-Score e AUC-ROC de 0,9957. A análise das características mais relevantes confirmou achados da literatura sobre marcadores linguísticos de depressão, incluindo aumento de afeto negativo, referências a ansiedade e saúde mental, e expressões de busca por ajuda.

A eficiência computacional dos modelos tradicionais representa uma vantagem significativa para aplicações práticas, permitindo processamento em tempo real e implantação em ambientes com recursos limitados. Esta característica é especialmente relevante para sistemas de monitoramento de larga escala.

Como trabalhos futuros, sugerimos a expansão do conjunto de dados para incluir textos em português brasileiro, a incorporação de análises multimodais, e a integração com sistemas de apoio psicológico. Em síntese, este estudo contribui para o campo da saúde mental digital ao demonstrar que soluções de Machine Learning eficientes e acessíveis podem auxiliar na detecção precoce de indicadores de depressão.

REFERÊNCIAS BIBLIOGRÁFICAS

BREIMAN, L. Random Forests. Machine Learning, v. 45, n. 1, p. 5-32, 2001.

CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In: PROCEEDINGS OF THE 22ND ACM SIGKDD, 2016. p. 785-794.

CORTES, C.; VAPNIK, V. Support-Vector Networks. Machine Learning, v. 20, n. 3, p. 273-297, 1995.

DE CHOUDHURY, M. et al. Predicting Depression via Social Media. In: ICWSM, 2013.

HOSMER, D. W.; LEMESHOW, S. Applied Logistic Regression. 2. ed. New York: Wiley, 2000.

LIU, Y. et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692, 2019.

MURPHY, K. P. Machine Learning: A Probabilistic Perspective. Cambridge: MIT Press, 2012.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: WHO, 2017.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. JMLR 12, pp. 2825-2830, 2011.