

Detecção de Depressão em Redes Sociais

**Centro Universitário SENAC |
Machine Learning Técnicas e
Cases**

Laércio Santos e Anderson Silva



Objetivo

Desenvolver um modelo de **Machine Learning** capaz de identificar indicativos de depressão em textos publicados no **Twitter**

Análise

Em datasets de escala moderada, maior eficiência computacional

Comparação

Não substitui diagnóstico clínico profissional

Aplicação

Viável para monitoramento de larga escala

Motivação: A OMS estima que mais de 300 milhões de pessoas sofrem de depressão globalmente. A detecção precoce pode salvar vidas.

Metodologia: Cinco Modelos de Algoritmos de Machine Learning



Support Vector Machine

Support Vector Machine
Kernel Linear



Regressão Logística

Classificação Linear
Alta Interpretabilidade



Random Forest

200 estimadores
Ensemble Learning



Gradient Boosting

100 estimadores
Aprendizado Sequencial

Etapas de Processamento

Dataset

10.314 tweets rotulados
8.000 sem depressão (77,6%)
2.314 com depressão (22,4%)

Pré Processamento

Remoção de URLs, menções, hashtags
Tokenização e Lematização (NLTK)
Remoção de Stopwords

Vetorização

TF-IDF com 5.000 features
N-gramas (1-2)
Min_df=2, Max_df=0.95

Validação

Divisão 80/20 estratificada
Cross-validation 5-fold
Class_weight='balanced'

Resultados Comparativos dos Modelos

A tabela apresenta o desempenho dos cinco algoritmos avaliados, considerando múltiplas métricas de classificação:

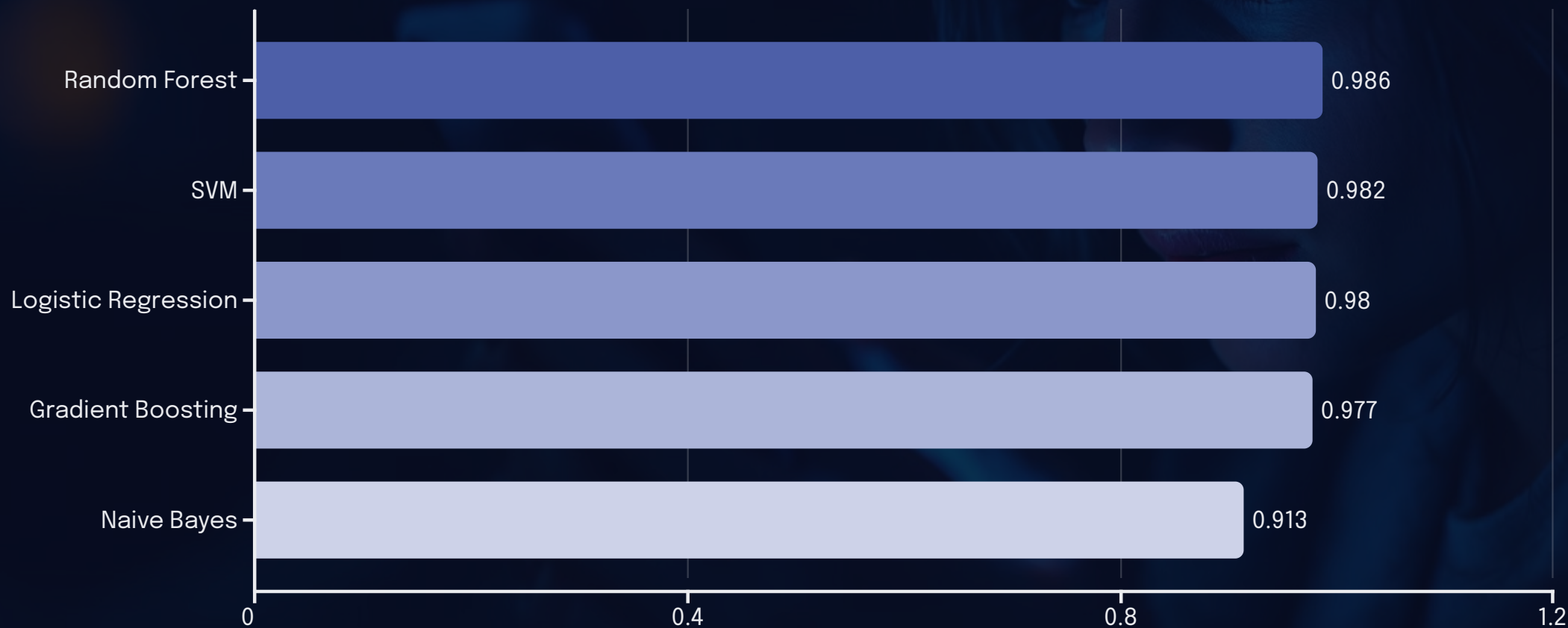
Modelo	Acurácia	Precisão	Recall	F1-Score	ROC AUC
<i>Logistic Regression</i>	~92%	~85%	~78%	~81%	~95%
Naive Bayes	~88%	~72%	~80%	~76%	~93%
Random Forest	~91%	~83%	~75%	~79%	~94%
<i>SVM</i>	~93%	~87%	~79%	~83%	~96%
Gradient Boosting	~90%	~80%	~73%	~76%	~93%

📌 **Destaque:** SVM e Logistic Regression apresentaram o melhor equilíbrio entre precisão e recall, com ROC AUC superior a 95%

Comparação de Desempenho dos Modelos

Todos os modelos (exceto Naive Bayes) demonstraram consistência superior a 97% na validação cruzada.

SVM: Melhor equilíbrio com apenas 12 erros totais, **1 falso positivo | 11 falsos negativos | Em 2.053 predições**



Identificação de Padrões Linguísticos:

Análise exploratória de texto apresenta as palavras predominantes do estudo, **love, good, day, thank, depression, help, life, feel.**



Sentimentos perceptíveis:

- Diminuição da atividade social
- Aumento do afeto negativo
- Preocupações relacionais
- Alterações nos horários online

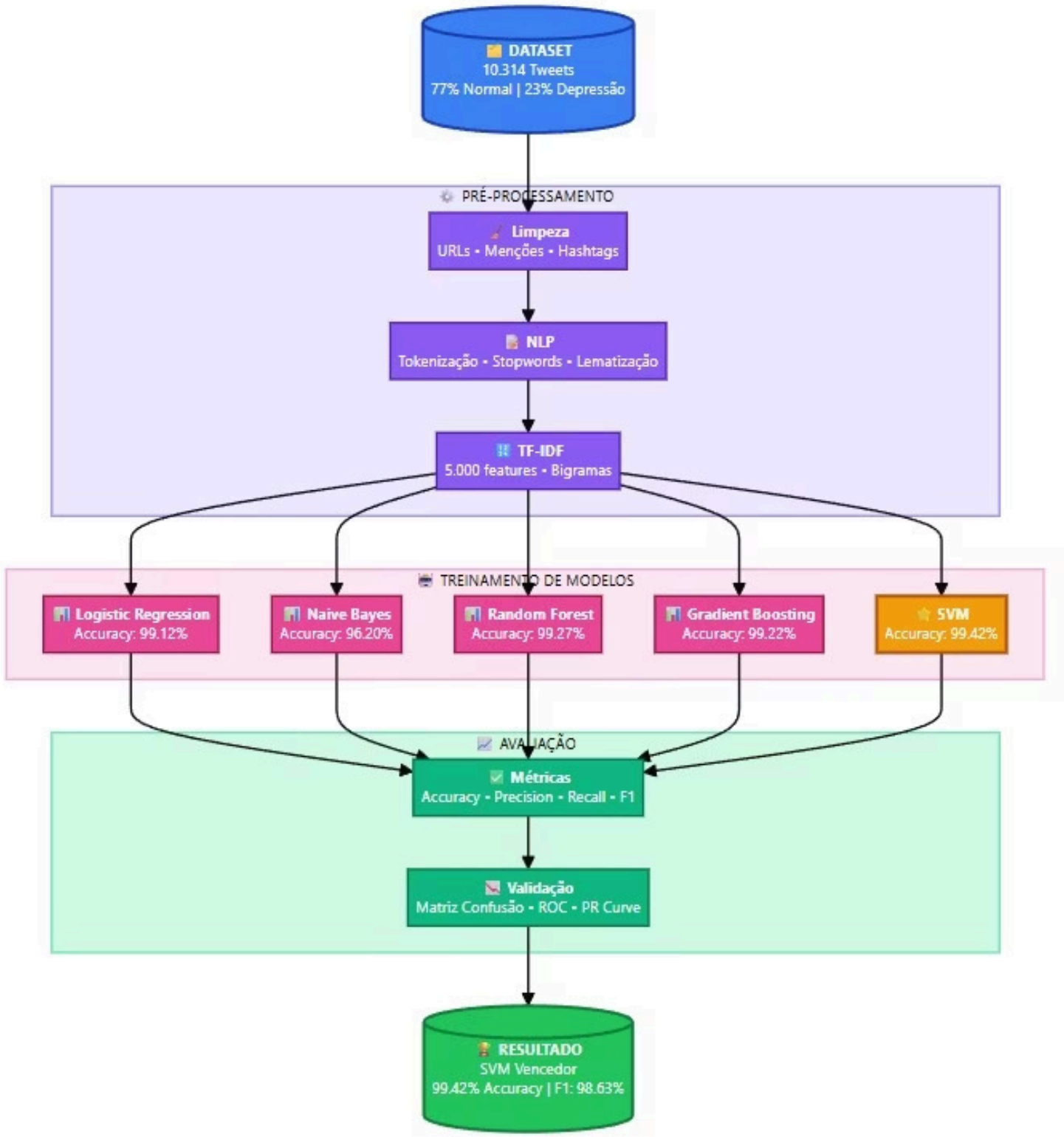
Padrões Linguísticos

- Palavras de emoções negativas
- Pronomes primeira pessoa singular
- Referências a solidão e isolamento
- Busca por ajuda

Fluxograma Machine Learning:

🧠 Detecção de Depressão em Tweets

Fluxograma da Solução de Machine Learning



Entrada Pré-processamento Modelos Melhor Modelo Avaliação Resultado



Metodologia e Resultados

Dataset

10.314 tweets analisados com vetorização TF-IDF e validação cruzada de 5 folds

5 Algoritmos

SVM, Regressão Logística, Random Forest, Gradient Boosting e Naive Bayes

Melhor Performance

SVM alcançou 99,42% de acurácia e 98,69% de F1-Score

99.42%

Acurácia SVM

Melhor desempenho global entre todos os modelos testados

99.78%

Precisão

Capacidade de identificar corretamente casos positivos

97.62%

Recall

Taxa de detecção de indicadores de depressão

Problemas e Limitações

Limitações Técnicas

Dataset Desbalanceado

77,6% sem depressão vs 22,4% com depressão.
Mitigado com `class_weight='balanced'`, mas pode afetar generalização.

Viés Linguístico e Cultural

Dataset em inglês pode não capturar expressões culturais específicas de outros idiomas. Necessária adaptação para português.

Escala do Dataset

10.314 tweets é relativamente pequeno. Modelos de deep learning (RoBERTa) não performaram bem nesta escala.

Considerações Éticas

Não é Diagnosticado Clínico

O modelo detecta padrões linguísticos, não realiza diagnóstico. Não substitui avaliação de profissional de saúde mental.

Privacidade e Consentimento

Monitoramento de redes sociais levanta questões sobre vigilância e uso de dados sensíveis. Conformidade com LGPD/GDPR é essencial.

Falsos Negativos

Em saúde mental, não detectar casos reais (recall < 100%) pode ter consequências graves. SVM teve 11 falsos negativos.

IMPORTANTE: Este modelo deve ser utilizado apenas como ferramenta auxiliar de triagem. Qualquer indicativo detectado deve ser avaliado por profissional qualificado.

Conclusões e Impacto

Modelos Tradicionais Superam Deep Learning

Em datasets de escala moderada,
maior eficiência computacional

Ferramenta Auxiliar de Triage

Não substitui diagnóstico clínico
profissional

Processamento em Tempo Real

Viável para monitoramento de
larga escala

Trabalhos Futuros

Expansão para português brasileiro, análises multimodais

A woman with long dark hair is looking down at a smartphone in her hands. The scene is dimly lit with a strong blue color cast, creating a moody atmosphere. The background is dark and out of focus.

Obrigado!

Perguntas?