Depressão e Suicídio nas Redes Sociais: Uma Abordagem Supervisionada Para Classificação de Textos

Keterly Geovana Gouveia Silva
Instituto de Geociências e Ciências Exatas
Universidade Estadual Paulista (UNESP)
Rio Claro - SP, Brasil
keterly.silva@unesp.br

Fabricio Aparecido Breve
Instituto de Geociências e Ciências Exatas
Universidade Estadual Paulista (UNESP)
Rio Claro - SP, Brasil
fabricio.breve@unesp.br

***Resumo*—A depressão é um transtorno mental frequentemente associado ao suicídio, o ato de tirar a própria vida. A detecção precoce do risco de suicídio e depressão é essencial para a redução do número de vítimas. Este estudo aplica Processamento de Linguagem Natural e Aprendizado de Máquina para identificar risco de suicídio e depressão, analisando conjunto de dados com e sem pré-processamento e extraindo características com o modelo BERT. A classificação de suicídio com Regressão Logística em textos com pré-processamento obteve acurácia de 90,42% e a remoção do pré-processamento resultou em 92,67%.**

Área: Inteligência Computacional

I. Introdução

A depressão é um transtorno mental que afeta a mente e altera como o indivíduo manifesta suas emoções [1]. Em 2015, as estatísticas mostraram que a depressão já atingia cerca de 322 milhões de pessoas no mundo todo. No mesmo ano, estimou-se que 788.000 pessoas cometeram suicídio, o ato intencional de causar sua própria morte [2], [3].

Em uma análise feita por [4] no Instagram de uma jovem que cometeu suicídio, é possível compreender que os textos podem indicar o sentimento daquele que o escreve. A análise indica que o conteúdo das publicações possibilitaram a identificação de sentimentos como tristeza e culpa.

Diante disso, o objetivo desta pesquisa é aplicar técnicas de Processamento de Linguagem Natural (PLN) e Aprendizado de Máquina para a detecção de potencial depressão e risco de suicídio em textos extraídos de redes sociais, representados por *Word Embeddings*. O estudo também analisa de que modo os cenários de pré-processamento podem influenciar nos resultados obtidos.

II. Conceitos e Técnicas

PLN é uma área relacionada à inteligência artificial e à linguística computacional e busca dar às máquinas a habilidade de compreender a língua falada ou escrita pelos seres humanos, língua esta usada no dia a dia para a comunicação [5]. Para que textos possam ser utilizados em modelos de classificação, por exemplo, o pré-processamento dos dados textuais se torna uma das primeiras etapas realizadas.

Pré-processar os dados pode incluir a remoção de *stopwords* e a representação de textos por meio de *word embeddings*. *Word Embedding* é uma forma de capturar o significado de cada palavra e transformar textos em números reais, de modo que essa transformação é representada na forma de vetores n-dimensionais [6]. A criação de *embeddings* pode ser feita de modo estático, com frameworks como *word2vec*, ou de modo contextual com modelos como BERT [7].

III. Metodologia de Desenvolvimento

Este trabalho propõe a classificação de um conjunto para suicídio e um para depressão, sendo eles representados por *word embeddings* geradas pelo modelo BERT. Para a classificação, os modelos supervisionados SVM, RandomForestClassifier e LogisticRegression foram construídos e testados com o auxílio da biblioteca scikit-learn.

A. Conjunto de Dados

O primeiro conjunto, Mental Health Twitter, foi escolhido para a detecção da depressão. Ele é utilizado no trabalho de [8], contém 20mil tweets e está disponível na plataforma Kaggle. O segundo conjunto, Suicidal Data, é um conjunto disponibilizado pelo trabalho de [9] para a predição de ideação suicida, com 5.121 mil exemplos negativos e 3.998 mil positivos.

B. Pré-processamento e geração de embeddings

Os conjuntos foram representados por *Word Embeddings* contextuais geradas pelo modelo BERT, com dimensão fixa de 768. Para fins comparativos, as *embeddings* foram geradas em cenários diferentes. O primeiro cenário aplicou técnicas de pré-processamento, realizando conversão de emojis, tokenização, remoção de caracteres especiais, *stopwords* e stemização. O segundo, no entanto, não realizou nenhuma manipulação inicial. Os conjuntos foram divididos em 80% dos dados para treino e 20% para teste. Esta forma de divisão, conhecida como holdout, foi executada 50 vezes para cada um dos algoritmos.

XIII WORKSHOP DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DA UNESP (WPPGCC 2024), Unesp, 26 e 27 de novembro de 2024

IV. Resultados Preliminares

Esta seção apresenta os resultados experimentais para os cenários com e sem utilização de pré-processamento.

A. Conjuntos pré-processados

As tabelas I, II e III exibem a média dos resultados obtidos após 50 execuções. Para o conjunto de depressão, Regressão Logística e Random Forest podem ser comparados com o trabalho de [8], enquanto a aplicação da SVM para o suicídio pode ser comparada com o trabalho de [9].

Tabela I
Resultados - Mental Health Twitter pré-processado

Métricas	[8] LR	Atual LR	[8] RF	Atual RF	SVM RBF	SVM Linear
Acurácia	0,56	0,6809	**0,71**	0,6798	0,6876	0,6810
F1	0,56	0,6847	**0,72**	0,6875	0,7020	0,6892

Tabela II
Comparativo entre trabalhos - Suicidal Data pré processado

Métrica	[9] Word2vec SVM	[9] Doc2vec SVM	[9] TF-IDF SVM	BERT SVM RBF	BERT SVM Linear
Acurácia	0,7958	0,7399	0,8494	**0,8988**	0,8987

Tabela III
Resultados - Suicidal Data pré processado

Métricas	LR	RF	SVM RBF	SVM Linear
Acurácia	**0,9042**	0,8882	0,8988	0,8987
F1	**0,8880**	0,8634	0,8767	0,8823

A pesquisa atual atingiu os melhores resultados na classificação com Regressão Logística e SVM quando comparada aos trabalhos relacionados. Além disso, em uma análise geral, nota-se que o modelo Random Forest apresentou desempenho inferior aos outros classificadores nos dois conjuntos. O conjunto de suicídio, comparado ao de depressão, obteve os melhores resultados, chegando a acurácia máxima de 0,9267 com a RL.

B. Conjuntos não processados

Para o cenário sem pré-processamento, as tabelas IV e V exibem os resultados, agora sem a comparação com trabalhos relacionados.

Tabela IV
Resultados - Mental Health Twitter não processado

Métricas	LR	RF	SVM RBF	SVM Linear
Acurácia	0,8056	0,7828	**0,8115**	0,8067
F1	0,8045	0,7835	**0,8119**	0,8060

Com a ausência das técnicas de pré-processamento, houve melhora dos resultados nos dois conjuntos, sendo o maior impacto observado no conjunto Mental Health Twitter. A classificação da depressão com LR, que obteve acurácia de 0,6809 no cenário pré-processado, atingiu 0,8056 no cenário atual, um aumento de 0,1247.

Tabela V
Resultados - Suicidal Data não processado

Métricas	LR	RF	SVM RBF	SVM Linear
Acurácia	**0,9267**	0,9050	0,9237	0,9182
F1	**0,9153**	0,8844	0,9093	0,9061

V. Considerações Finais

O objetivo de aplicar técnicas de PLN e Aprendizado de Máquina para a detecção de depressão e suicídio foi atingido através da utilização do modelo BERT para a extração de características e da aplicação de modelos supervisionados. Os modelos SVM, Random Forest e Regressão Logística foram escolhidos e testados em conjuntos contendo postagens em língua inglesa.

Os resultados evidenciaram o impacto das etapas de pré-processamento no resultado das classificações. Com o uso do modelo BERT, algumas técnicas demonstraram efeito negativo na geração dos *embeddings* contextuais. Como próximo passo, a pesquisa buscará explorar os impactos da redução de dimensionalidade e seleção de características na classificação supervisionada, além de avaliar e comparar o uso do aprendizado semi-supervisionado na detecção de depressão e risco de suicídio.

Referências

[1] K. F. L. Vieira and M. d. P. d. L. Coutinho, "Representações sociais da depressão e do suicídio elaboradas por estudantes de psicologia," *Psicologia: Ciência e Profissão*, vol. 28, no. 4, p. 714–727, 2008. [Online]. Available: https://doi.org/10.1590/S1414-98932008000400005

[2] World Health Organization, *Depression and Other Common Mental Disorders: Global Health Estimates*. Geneva: World Health Organization, 2017.

[3] F. d. O. Barbosa, P. C. M. Macedo, and R. M. C. d. Silveira, "Depressao e suicídio," *Revista da SBPH*, vol. 14, pp. 233 – 243, 2011. [Online]. Available: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-08582011000100013&nrm=iso

[4] A. D. Rios, "Suicídio e redes sociais: uma análise comportamental das postagens de uma jovem no instagram," Dissertação, Universidade Católica de Brasília, 2021, dissertação (Programa Stricto Sensu em Psicologia) - Universidade Católica de Brasília, Brasília, 2021. [Online]. Available: https://bdtd.ucb.br:8443/jspui/handle/tede/2841

[5] H. M. Caseli and M. G. V. Nunes, Eds., *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*, 2nd ed. BPLN, 2024. [Online]. Available: https://brasileiraspln.com/livro-pln/2a-edicao/

[6] A. Neelima and S. Mehrotra, "A comprehensive review on word embedding techniques," in *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*, 2023, pp. 538–543.

[7] W. Sarakul and A. T. Rutherford, "Contextualized vs. static word embeddings for word-based analysis of opposing opinions," in *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2023, pp. 95–100.

[8] M. A. Abbas, K. Munir, A. Raza, N. A. Samee, M. M. Jamjoom, and Z. Ullah, "Novel transformer based contextualized embedding and probabilistic features for depression detection from social media," *IEEE Access*, vol. 12, pp. 54 087–54 100, 2024.

[9] K. Yatapala and B. Kumara, "Detection of suicide ideation in twitter using ann," in *2021 6th International Conference on Information Technology Research (ICITR)*, 2021, pp. 1–5.