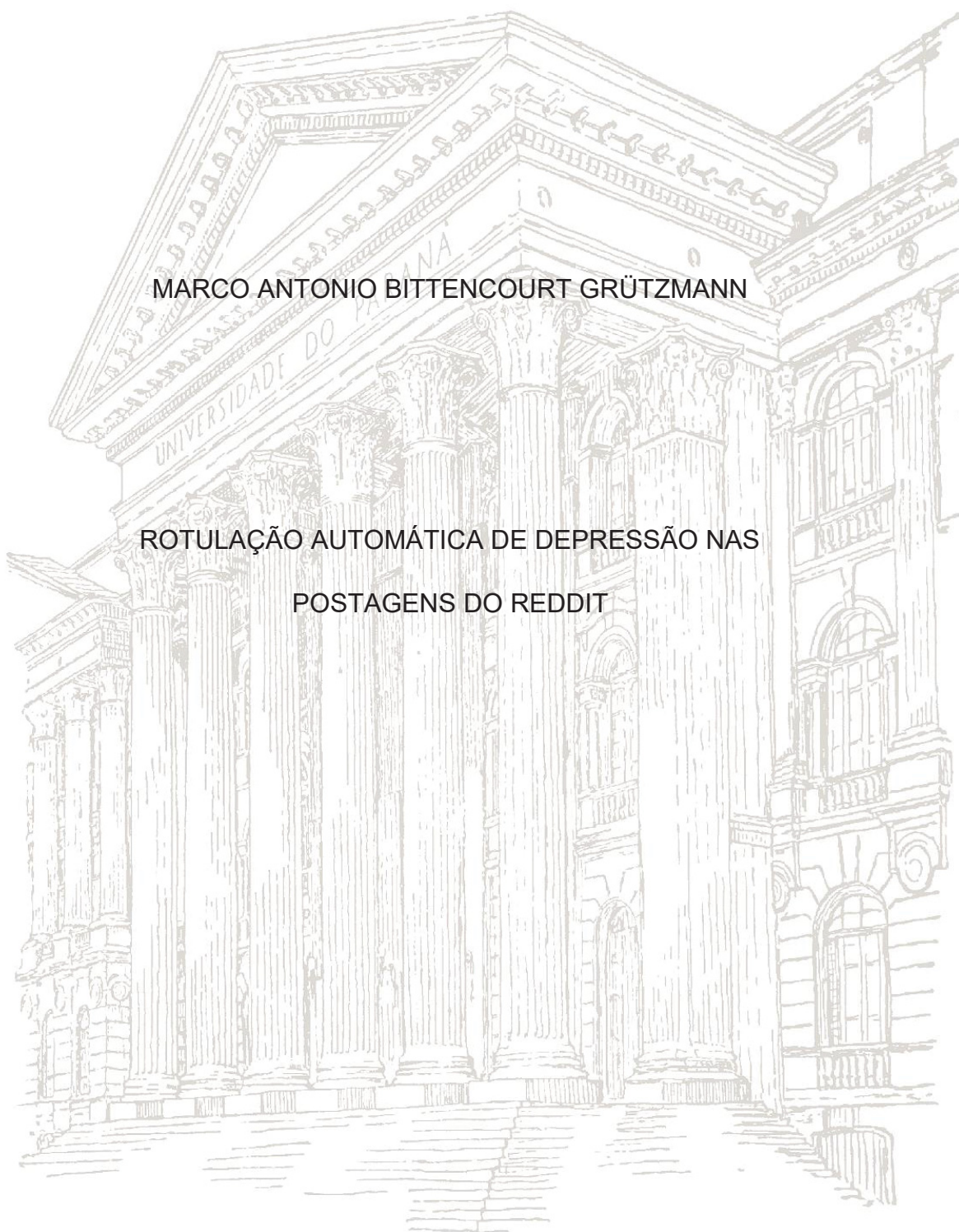


UNIVERSIDADE FEDERAL DO PARANÁ

MARCO ANTONIO BITTENCOURT GRÜTZMANN

ROTULAÇÃO AUTOMÁTICA DE DEPRESSÃO NAS
POSTAGENS DO REDDIT



CURITIBA

2023

MARCO ANTONIO BITTENCOURT GRÜTZMANN

ROTULAÇÃO AUTOMÁTICA DE DEPRESSÃO NAS
POSTAGENS DO REDDIT

Dissertação apresentada ao Programa de Pós-Graduação em Gestão da Informação (PPGGI), Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná (UFPR), como requisito parcial à obtenção do título de Mestre em Gestão da Informação.

Orientadora: Prof^a Dr^a Denise Fukumi Tsunoda.

Coorientadora: Prof^a Dr^a Fernanda Brenneisen Mayer.

CURITIBA

2023

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIAS SOCIAIS APLICADAS

Grützmann, Marco Antonio Bittencourt

Rotulação automática de depressão nas postagens de Reddit /
Marco Antonio Bittencourt Grützmann. – Curitiba, 2023.

1 recurso on-line : PDF.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor
de Ciências Sociais Aplicadas, Programa de Pós-Graduação em
Gestão da Informação.

Orientadora: Profa. Dra. Denise Fukumi Tsunoda.

Coorientadora: Profa. Dra. Fernanda Brenneisen Mayer.

1. Mineração de dados (Computação). 2. Depressão. 3. Redes
sociais. I. Tsunoda, Denise Fukumi. II. Mayer, Fernanda Brenneisen.
III. Universidade Federal do Paraná. Programa de Pós-Graduação
em Gestão da Informação. III. Título.

Bibliotecária: Maria Lidiane Herculano Graciosa CRB-9/2008



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS SOCIAIS E APLICADAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO DA
INFORMAÇÃO - 40001016058P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GESTÃO DA INFORMAÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **MARCO ANTONIO BITTENCOURT GRUTZMANN** intitulada: **ROTULAÇÃO AUTOMÁTICA DE DEPRESSÃO NAS POSTAGENS DO REDDIT**, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua *aprovação* no rito de defesa. A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 30 de Março de 2023.

FERNANDA BRENNEISEN MAYER
Presidente da Banca Examinadora

RONAN ASSUMPCAO SILVA
Avaliador Interno (INSTITUTO FEDERAL DO PARANÁ)

EMERSON CABRERA PARAISO
Avaliador Externo (PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ)

RESUMO

A depressão é a principal causa de incapacidade em todo o mundo, afetando o humor e os sentimentos das pessoas que possuem o transtorno. Os usuários de redes sociais usam as plataformas para obter apoio nesses casos, reconhecer que não estão sozinhos e compartilhar suas experiências. Existe uma crescente área interdisciplinar que utiliza técnicas de mineração de dados para identificar tendências depressivas por meio de redes sociais, porém um dos maiores desafios é a rotulação dos dados e seus atributos. O presente trabalho se propõe a aprofundar sobre quais atributos podem ser considerados para rotulação automática de depressão em postagens em Redes Sociais, utilizando o Reddit como rede social alvo. A metodologia inclui revisão de metodologias e tecnologias, desenho da metodologia geral do trabalho, captura e exploração de dados obtidos por meio da rede social Reddit, pré-processamento e extração de "features", análise estatística descritiva, comparação de desempenho de classificadores para rotulação automática e aferição e discussão dos resultados. A coleta de dados foi realizada utilizando o repositório de dados PushShift e a API pushshift.io Reddit, com dados coletados exclusivamente do subreddit "Desabafos" no período de 01/01/2021 a 30/06/2022. Como resultados, identificado que os atributos metadados, os atributos retirados da rede social em si e o texto transformado em dimensões de atributos (através de técnicas como BoW e TF-IDF) são eficientes na classificação da classe "depressão", utilizando-se do método Random Forest, podendo seu resultado variar dependendo de qual classe será utilizada no contexto de "não depressivo" para treinamento do classificador. Outrossim, avaliado 17 atributos e suas importâncias na classificação, onde os quatro principais são a carga sentimental encontrada no texto, a variância de sentimento, o tamanho do texto em si e quantidade de pessoas citadas nos textos.

Palavras-chave: mineração de dados, depressão, redes sociais, KDD.

ABSTRACT

Depression is the leading cause of disability worldwide, affecting the mood and feelings of people with the disorder. Social media users use the platforms to get support in these cases, recognize that they are not alone and share their experiences. There is a growing interdisciplinary area that uses data mining techniques to try to identify depressive tendencies through social networks, but one of the biggest challenges is the labeling of the data and its attributes. The present work proposes to delve into what attributes can be considered for automatic labeling of Depression in posts on Social Networks, using Reddit as the target social network. The methodology includes review of methodologies and technologies, design of the general methodology of the work, capture and exploration of data obtained through social networks (in this case, Reddit), pre-processing and acceptance of "*features*", descriptive statistics, descriptive analysis, performance comparison of classifiers for automatic labeling and measurement and discussion of results. Data collection was carried out using the PushShift data repository and the pushshift.io API. As a result, we identified that attributes derived from the text, attributes taken from the social network itself and the text transformed into attribute dimensions (through techniques such as BoW and TF-IDF) are efficient in classifying the "Depression" class, using the Random Forest method, allowing its result to vary depending on which class will be used in the context of "non-depressive" for training the classifier. Others evaluate 17 *features* and their importance in the classification, where the four main ones are the sentimental content found in the text, the sentiment variance, the size of the text itself and the number of people mentioned in the texts.

Keywords: Datamining, depression, social media, KDD.

LISTA DE FIGURAS

| | |
|---|----|
| FIGURA 1– PREVALÊNCIA GLOBAL DE DEPRESSÃO, POR IDADE E SEXO | 12 |
| FIGURA 2 - TAXAS DE SUICÍDIO GLOBAIS, POR IDADE..... | 13 |
| FIGURA 3 - FLUXO DE EXEMPLO DE UM USUÁRIO DO REDDIT | 15 |
| FIGURA 4 - VISÃO GERAL DAS ETAPAS QUE COMPÕEM O PROCESSO KDD..... | 25 |
| FIGURA 8 - VISÃO MACRO DO FRAMEWORK PROPOSTO PARA DESENVOLVIMENTO DO PROJETO DE PESQUISA | 33 |
| FIGURA 9 – FLUXO DE TRABALHO DO CONTEXTO DOS CLASSIFICADORES..... | 42 |
| FIGURA 5 – FLUXOGRAMA DETALHANDO PROCESSO DE ELEGIBILIDADE DE TRABALHOS PARA REVISÃO SISTEMÁTICA | 47 |
| FIGURA 6 – RELAÇÃO DE PALAVRAS (EM INGLÊS) RELACIONADAS À SUICÍDIO | 51 |
| FIGURA 7 – GRÁFICOS COMPARATIVOS DOS TRÊS CLASSIFICADORES MAIS UTILIZADOS NOS ESTUDOS COM RESULTADOS DAS MÉTRICAS DE AVALIAÇÃO DE CLASSIFICAÇÃO. | 57 |
| FIGURA 10 - NUVEM DE PALAVRAS CLASSE DEPRESSÃO..... | 80 |
| FIGURA 11 - NUVEM DE PALAVRAS GERAL..... | 81 |

LISTA DE QUADROS

| | |
|--|----|
| QUADRO 1– REDES SOCIAIS MAIS UTILIZADAS PELOS BRASILEIROS E TEMPO MÉDIO DE UTILIZAÇÃO..... | 17 |
| QUADRO 2 – SINTOMAS CONSIDERADOS NO DIAGNÓSTICO DE DEPRESSÃO..... | 23 |
| QUADRO 3 - COLUNAS DO DATASET | 60 |
| QUADRO 4 – EXPLICAÇÃO DE CADA CLASSIFICAÇÃO | 61 |
| QUADRO 5 – QUANTIDADE DE POSTAGENS DE ACORDO COM SUA CLASSIFICAÇÃO..... | 63 |
| QUADRO 6 - EXEMPLO DE UMA POSTAGEM DO REDDIT | 64 |
| QUADRO 7 – EXEMPLO DE SENTENÇAS IDENTIFICADAS PELA API DE SENTIMENTOS | 69 |
| QUADRO 8 – EXEMPLO DE IDENTIFICAÇÃO DE SENTIMENTOS POR SENTENÇA..... | 72 |
| QUADRO 9 – ENTIDADES DISPONÍVEIS NA API GOOGLE CLOUD..... | 75 |
| QUADRO 10 – TOTAL DE REGISTROS E PERCENTUAL POR ENTIDADE . | 76 |
| QUADRO 11 – TOTAL DE REGISTROS E PERCENTUAL POR ENTIDADE APÓS TRATAMENTO..... | 76 |
| QUADRO 12 – COMPARATIVO DE SIMILARIDADE DAS PALAVRAS COM MAIOR FREQUÊNCIA POR CLASSE | 79 |
| QUADRO 13 - RESULTADOS DAS CLASSIFICAÇÕES POR ELEMENTOS TEXTUAIS..... | 82 |
| QUADRO 14 - RESULTADO DA CLASSIFICAÇÃO POR FEATURES | 83 |
| QUADRO 15 - RESULTADO DA CLASSIFICAÇÃO COMBINADA | 84 |
| QUADRO 16–IMPORTÂNCIA DAS FEATURES | 85 |

LISTA DE TABELAS

| | |
|---|----|
| TABELA 1– CRITÉRIOS PARA STRING DE PESQUISA NAS BASES ACADÊMICAS..... | 19 |
| TABELA 2– CRITÉRIOS PARA STRING DE PESQUISA NAS BASES ACADÊMICAS..... | 45 |
| TABELA 3 – DESIGN DE CRITÉRIOS ACEITE OU RECUSA PARA REVISÃO | 46 |

LISTA DE SIGLAS

API - *Application Programming Interface* - Interface de Programação de Aplicação

BoW - Bag of Words - bolsa de palavras

KDD – *Knowledge Discovery in Databases*

RNAs - Rede Neurais Artificiais

SVM - Support Vector Machines

TF-IDF - *Term Frequency – Inverse Document Frequency*

UFPR - Universidade Federal do Paraná

SUMÁRIO

| | | |
|----------|--|-----------|
| 1 | INTRODUÇÃO | 12 |
| 1.1 | CONTEXTUALIZAÇÃO E PROBLEMA..... | 16 |
| 1.2 | OBJETIVOS | 18 |
| 1.2.1 | Objetivo geral | 18 |
| 1.2.2 | Objetivos específicos | 18 |
| 1.3 | JUSTIFICATIVA | 19 |
| 1.3.1 | Justificativa acadêmica | 19 |
| 1.3.2 | Justificativa ao Programa de Pós-graduação em Gestão da Informação | 21 |
| 1.4 | DELIMITAÇÕES DA PESQUISA..... | 21 |
| 1.5 | ESTRUTURA DO DOCUMENTO | 22 |
| 2 | REFERENCIAL TEÓRICO | 23 |
| 2.1 | SAÚDE MENTAL..... | 23 |
| 2.2 | DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS | 24 |
| 2.2.1 | MODELOS DE CLASSIFICAÇÃO | 27 |
| 3 | PERCURSO METODOLÓGICO..... | 32 |
| 3.1 | PERCURSO METODOLÓGICO..... | 32 |
| 3.2 | COMUNIDADE DA REDE SOCIAL REDDIT..... | 34 |
| 3.3 | COLETA DOS DADOS..... | 35 |
| 3.4 | AMOSTRAGEM | 36 |
| 3.5 | PRÉ-PROCESSAMENTO E EXTRAÇÃO DE FEATURES | 36 |
| 3.5.1 | Pré-processamento..... | 37 |
| 3.5.2 | Extração de <i>features</i> | 38 |
| 3.5.3 | Rotulação dos dados..... | 39 |
| 3.6 | ANÁLISE E ESTATÍSTICA DESCRITIVA | 40 |
| 3.6.1 | Variância de Sentimentos | 40 |
| 3.7 | MODELOS DE CLASSIFICAÇÃO..... | 41 |
| 3.8 | AVALIAÇÃO DOS MODELOS DE CLASSIFICAÇÃO | 43 |
| 3.9 | IMPORTÂNCIA DAS FEATURES | 43 |
| 4 | RESULTADO: REVISÃO SISTEMÁTICA..... | 45 |
| 4.1 | METODOLOGIA DA REVISÃO SISTEMÁTICA | 45 |
| 4.2 | RESULTADOS | 47 |
| 4.2.1 | Saúde mental e redes sociais..... | 48 |

| | | |
|----------|---|-----------|
| 4.2.2 | Seleção de amostra e rotulação | 50 |
| 4.2.3 | Técnicas de Tratamento de dados e Classificação..... | 52 |
| 4.2.4 | Aferição de Resultados | 55 |
| 4.2.5 | DISCUSSÃO..... | 57 |
| 4.2.6 | ANÁLISES E CONSIDERAÇÕES | 58 |
| 5 | RESULTADO: ANÁLISE DE DADOS NA BASE REDDIT | 60 |
| 5.1 | PRÉ-PROCESSAMENTO E EXTRAÇÃO DE FEATURES | 64 |
| 5.2 | RESULTADOS CLASSIFICAÇÃO DE TEXTOS..... | 66 |
| 5.2.1 | Análise de sentimentos..... | 66 |
| 5.2.2 | Análise de sentimento geral | 67 |
| 5.2.2 | Total de sentenças | 69 |
| 5.2.3 | Variância de Sentimentos..... | 72 |
| 5.2.3 | Extração de entidades | 74 |
| 5.2.4 | Processamento de texto | 78 |
| 5.2.5 | Modelos de Classificação..... | 82 |
| 6 | CONSIDERAÇÕES FINAIS | 86 |
| | REFERÊNCIAS..... | 90 |

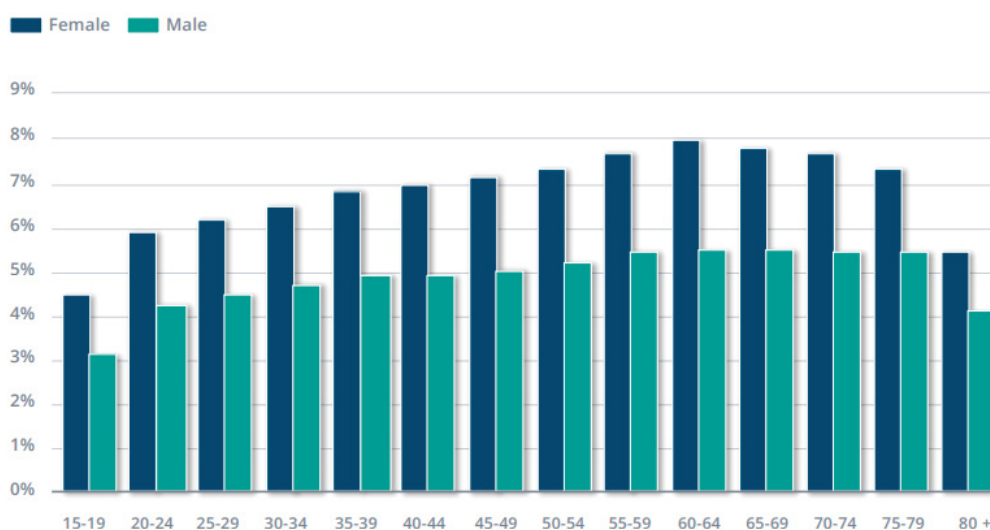
1 INTRODUÇÃO

Os transtornos mentais comuns são classificados em categorias principais: transtornos depressivos e transtornos de ansiedade. Esses distúrbios são altamente prevalentes na população e afetam o humor e/ou os sentimentos das pessoas afetadas, os sintomas variam de leve a grave e podem durar entre alguns dias, meses ou anos. Esses distúrbios são condições de saúde diagnosticáveis e são distintos dos sentimentos de tristeza, estresse ou medo que qualquer pessoa pode experimentar de tempos em tempos em suas vidas (GLOBAL HEALTH ESTIMATES, 2017).

Em nível global, estima-se que mais de 300 milhões de pessoas sofram de depressão, o equivalente a 3,8% da população mundial viva (GLOBAL HEALTH ESTIMATES, 2017).

No Brasil a taxa foi de 10,2% da população com mais de 18 anos em 2019, registrando um aumento de 32% quando comparado com os dados de 2013. As mulheres apresentam diagnóstico de depressão 2,8 vezes mais que os homens (MINISTÉRIO DA MULHER; SECRETARIA NACIONAL DA FAMÍLIA; OBSERVATÓRIO NACIONAL DA FAMÍLIA, 2022), cenário que se reflete mundialmente. Conforme apresentado na Figura 1, independente da faixa etária, as mulheres sempre apresentam maior prevalência de depressão.

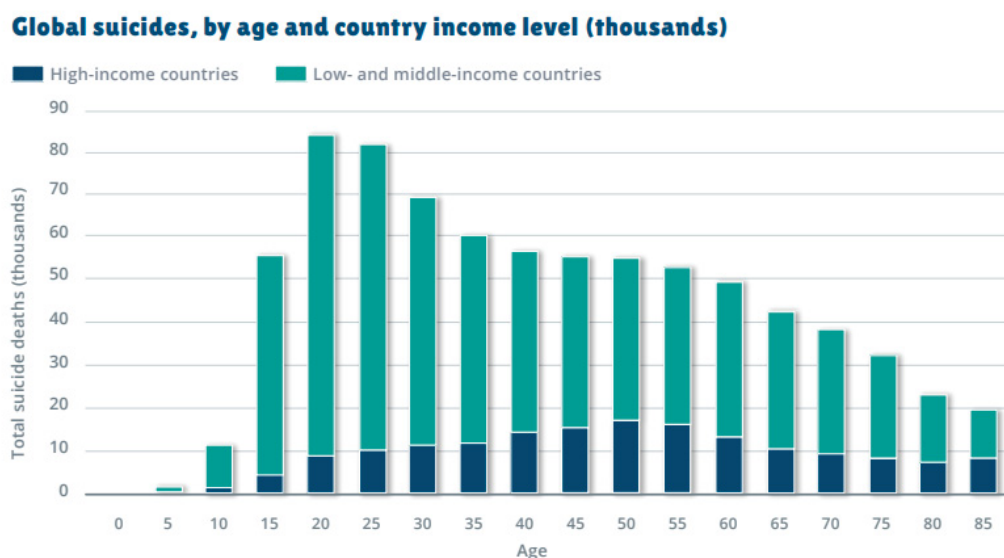
FIGURA 1– PREVALÊNCIA GLOBAL DE DEPRESSÃO, POR IDADE E SEXO



FONTE: Global Health Estimates (2017).

Apesar da existência de tratamentos conhecidos e eficazes, mais de 75% das pessoas que vivem em países de baixa renda não recebem tratamento, seja por falta de recursos e/ou pelo estigma social relacionado a transtornos mentais. A depressão, especialmente quando recorrente e de intensidade moderada ou grave traz graves problemas sociais e, em casos graves, a depressão pode levar a ideações suicidas e até mesmo a consumação do fato. Estima-se que mais de 700.000 pessoas morrem por suicídio todos os anos no mundo, sendo está a quarta principal causa de morte entre jovens de 15 a 29 anos no Brasil (MINISTÉRIO DA SAÚDE, 2022; WORLD HEALTH ORGANIZATION, 2017).

FIGURA 2 - TAXAS DE SUICÍDIO GLOBAIS, POR IDADE



FONTE: Global Health Estimates (2017).

Estudos mostram que os usuários de rede social utilizam de rede social para falar sobre sentimentos e compartilhar experiências. (BERRY; BUCCI; LOBBAN, 2017)

De acordo com BUCCI e col. (2019), os usuários das redes sociais foram entrevistados e os principais motivos são:

- apoio disponível quando não podiam sair de casa ou quando havia receio de procurar ajuda pessoalmente;
- conhecer experiência de outras pessoas que já se recuperavam, o que dá esperança de recuperação;
- reconhecer que não estão sozinhos;
- anonimato;
- entre outros.

Estudos mostram que participantes de tratamento de saúde mental em grupo descrevem benefícios semelhantes, sendo os principais benefícios de se estar em grupo é a aprendizagem compartilhada, compartilhar sucesso, desafios e informação, sentimento de pertencimento (ASCHBRENNER; NASLUND; BARTELS, 2016).

Essa pesquisa baseia-se na rede social Reddit. Reddit.com é uma rede social que surgiu em 2005 e se define como:

O Reddit é o lar de milhares de comunidades, conversas intermináveis e conexão humana autêntica. Se você gosta de notícias de última hora, esportes, teorias de fãs de TV ou um fluxo interminável dos animais mais fofos da Internet, há uma comunidade no Reddit para você. Como Funciona o Reddit? Todos os dias, milhões de pessoas em todo o mundo postam, votam e comentam em comunidades organizadas em torno de seus interesses (REDDIT, 2022, sn).

Os dados sobre a quantidade de usuários foram atualizados pelo Reddit em 2021, sendo eles composto por mais de 50 milhões de usuários ativos diariamente em todo o mundo, mais de 100 mil comunidades ativas e mais de 13 bilhões de postagens e comentários. O conteúdo do site é fornecido por seus usuários e a popularidade desse conteúdo também é determinado pela adesão dos usuários.

O registro é simples com poucas restrições e pode ser feito de forma anônima. As postagens de texto são categorizadas como “postagens próprias” e os links podem estar em vários formatos diferentes, incluindo artigos, fotos e vídeos (ANDERSON, 2015; REDDIT, 2022).

No Brasil, em 2021 cerca de 16,6% da população entre 16 até 64 anos utiliza o Reddit mensalmente (“Digital in Brazil: All the Statistics You Need in 2021 — DataReportal – Global Digital Insights”, 2021).

Como forma de auxiliar no diagnóstico preventivo dos transtornos da saúde mental, pesquisadores estão usando comportamento dos usuários e dados linguísticos de redes sociais para prever a presença de transtornos. Desde 2013, pesquisas podem avaliar a presença de depressão, suicídio, distúrbios alimentares, esquizofrenia, entre outros com alta precisão (80–90%) (CHANCELLOR; DE CHOUDHURY, 2020).

Para que seja possível o desenvolvimento para predição de transtornos mentais com base na rede social reddit, se fazem necessários estudos relacionados a rotulação e classificação das postagens dentro da rede social, vinculados aos seus atributos.

Dessa forma, esse trabalho se propõe aprofundar sobre quais atributos podem ser considerados para rotulação automática de depressão em postagens em Redes Sociais.

1.1 CONTEXTUALIZAÇÃO E PROBLEMA

O transtorno depressivo é uma doença com altos níveis de suicídio exigindo cada vez mais um rastreamento precoce (AQUINO, 2019). Além disso, o estigma social da depressão circundante impede que muitos indivíduos afetados procurem assistência profissional adequada. Como resultado, eles se voltam para recursos menos formais, como redes sociais (TADESSE et al., 2019). Objetivando o rastreamento precoce, as redes sociais se destacam como um caminho promissor, pois fazem parte do dia a dia dos brasileiros, sendo uma fonte de conexões, expressões e influências humanas.

Levantamento realizado por *DataReportal* mostrou que o Brasil ocupa o terceiro lugar no mundo em quantidade de horas de utilização de rede social diária, sendo 03h42min, enquanto a média global é 02h25min, bem como 70,3% da população brasileira possui rede social (DATA REPORTAL – GLOBAL DIGITAL INSIGHTS, 2021). As mais utilizadas são:

QUADRO 1– REDES SOCIAIS MAIS UTILIZADAS PELOS BRASILEIROS E TEMPO MÉDIO DE UTILIZAÇÃO

| Rede social | Porcentagem de usuários | Tempo médio mensal de utilização (horas) |
|-------------|-------------------------|--|
| Youtube | 96,4% | 26,3 |
| WhatsApp | 91,7% | 30,3 |
| Facebook | 89,8% | 15,6 |
| Instagram | 86,3% | 14,0 |

FONTE: Adaptado Global digital insights (2021).

Devido à alta utilização e geração de dados, as redes sociais se tornam fontes ricas de estudos de diversas áreas.

Um dos desafios significativos na mineração dos dados dos usuários de redes sociais é lidar com dados complexos, ou seja, texto, *emoticons* e imagens, que se caracterizam como dados não estruturados que somam ao desafio de ser um ambiente dinâmico e sem padrões, por exemplo, os usuários têm frequência de postagem e contextos individuais muito diferentes (GIUNTINI, 2021).

Dentro dos desafios da mineração de dados em redes sociais, com ênfase em desenvolvimento de classificadores, existe a rotulação e contextualização das postagens utilizadas em redes sociais. É comum que seja utilizada técnica de *crowd-sourcing*, que consiste em solicitar que usuários rotulem *datasets* para posterior estudo, porém existem redes sociais que possuem funcionalidades onde a própria comunidade já faz a rotulação de comentários e postagens, de acordo com regras prévias.

O problema em rotulações realizadas pela comunidade é que carecem de validade científica, pois não garantem que são avaliadas por especialistas, além de não terem toda a contextualização e métodos adequados para aferição de saúde mental.

Porém, tal rotulação permite o estudo aprofundado de quais percepções são utilizadas para rotulação e classificação das postagens, além de possibilitar comparativos com demais elementos de outras classes dentro de um mesmo contexto da rede social.

Além do entendimento das características de rotulação, entender o potencial de desenvolvimento de um modelo com característica preditiva de

rotulação, com base nos elementos textuais e derivados das redes sociais, contribui para avanço de modelos para língua portuguesa, entendimento de limitações e possibilidade de avanços metodológicos aprofundados para detecção precoce de indícios de transtornos mentais utilizando como base as redes sociais.

Para o desenvolvimento de um estudo aprofundado, com base em KDD (*knowledge-discovery in databases* – descoberta de conhecimento em base de dados), com foco em mineração em redes sociais, é necessária escolha de uma plataforma com características que permitam a extração de classes, dos seus dados e atributos.

Com base na contextualização do problema, formula-se: **Quais atributos podem ser considerados para rotulação automática de Depressão em postagens em Redes Sociais?**

1.2 OBJETIVOS

Este tópico está dividido em: objetivo geral e específicos.

1.2.1 Objetivo geral

Com o intuito de responder à questão de pesquisa, definiu-se como objetivo geral: desenvolver e avaliar classificadores para rotulação automática de depressão para a comunidade 'Desabafos' da rede social Reddit por meio da identificação de padrões na classificação textual e humana da comunidade.

1.2.2 Objetivos específicos

Derivado do objetivo geral, e para o seu cumprimento, foram definidos os seguintes objetivos específicos:

- realizar revisão sistemática de metodologias e tecnologias utilizadas para resolução de problemas similares;
- coletar e pré-processar os dados para a construção de um conjunto de dados.

- analisar e descrever o conjunto de dados formado, com foco na relação entre as classes da rede social e as dimensões do conjunto de dados, tal como os processos envolvidos na geração das variáveis.
- analisar e comparar os resultados dos métodos de classificação.

1.3 JUSTIFICATIVA

Este tópico foi dividido em: justificativa acadêmica, justificativa ao Programa de Pós-graduação em Gestão da Informação e justificativa à sociedade.

1.3.1 Justificativa acadêmica

Em pesquisa as bases Scopus (Elsevier), PubMed e American Psychologic Association (PsychInfo), em um período de 5 anos, sendo de 01/06/2016 à 01/06/2021, utilizando as *strings* da Tabela 1 como critérios de busca, encontrou-se 89 estudos.

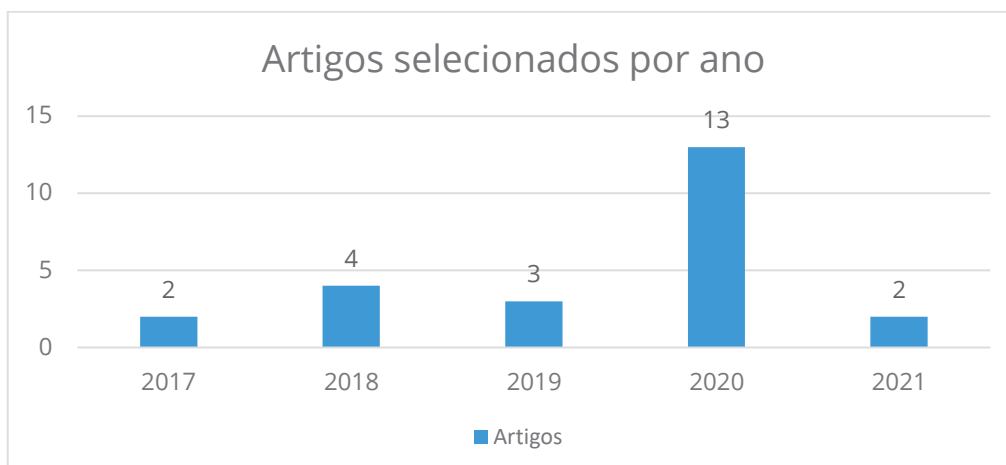
TABELA 1– CRITÉRIOS PARA STRING DE PESQUISA NAS BASES ACADÊMICAS

| Categoria | Palavras-Chave |
|-----------------------------|---|
| Saúde Mental (1) | Mental Health, Mental Disorder, Anxiety, Stress, Suicide, Depression, Psychiatry |
| Redes Sociais (2) | Social Network, Facebook, Instagram, Tiktok, Twitter, Social Media, Reddit |
| Inteligência Artificial (3) | Predictive, Machine Learning, Forecast, AI, Artificial Intelligence, Data Mining, Deep Learning |
| Termos de Busca | (1) AND (2) AND (3) |

FONTE: O Autor (2021).

Após a filtragem dos artigos, foi realizado a leitura e análise de 24 pesquisas (metodologia e resultados descritos no Item 3), a distribuição por data está ilustrada no Gráfico 1.

GRÁFICO 1– TOTAL DE ARTIGOS SELECIONADOS POR ANO DE PUBLICAÇÃO



FONTE: O Autor (2021).

Com base no Gráfico 1, vê-se um aumento de interesse da comunidade acadêmica e científica na exploração dos temas que envolvem redes sociais e saúde mental. Dessa forma, dentro do contexto delimitado no item 3 (revisão sistemática), não foi encontrado pesquisas na língua portuguesa que abordem o tema de saúde mental e redes sociais.

Esse fato acaba dificultando a comunidade em ter referências de técnicas e métodos de mineração de dados que sejam facilmente reprodutíveis para a língua portuguesa, face às características individuais de cada linguagem e contextos brasileiros. Entretanto, a quantidade de trabalhos demonstra interesse acadêmico e científico em evoluir e explorar a área de estudo proposta.

Como contribuições esperadas do trabalho, o desenvolvimento da pesquisa possibilitará a entrega de um artigo referente ao tema objetivo da pesquisa e uma dissertação interdisciplinar sobre o tema proposto.

Será disponibilizado um conjunto de dados sobre o tema, descaracterizando qualquer identificação dos participantes do estudo. Este resultado visa auxiliar futuras análises e trabalhos no meio acadêmico, principalmente pela dificuldade em conseguir materiais similares em português.

As variáveis serão extraídas da rede social Reddit e analisadas para determinação de importância no funcionamento do modelo preditivo, possibilitando maior entendimento de quais variáveis possuem maior correlação com a linguagem depressiva, tendo como base a classificação da comunidade.

Por fim, fomentar a pesquisa Brasileira nessa área de atuação, uma vez que maioria das referências encontradas nesse tema são de outros países e, conseqüentemente outros idiomas.

1.3.2 Justificativa ao Programa de Pós-graduação em Gestão da Informação

Realizando uma busca por dissertações e teses no sistema de bibliotecas da UFPR usando os termos “rede social” e “saúde mental”, tal como “reddit” e também "rede social" e "mineração de dados", não foram encontradas teses nem dissertações em relação aos temas associados.

Com a pesquisa "rede social" e "mineração de dados" encontram-se dois resultados, sendo um com foco em automação de coleta de dados de redes e outro com foco em justificar as redes sociais como fonte de informação.

Já com a busca dos termos “Descoberta de Conhecimento em Base de Dados (KDD)” encontram-se 14 dissertações e 7 Teses. Uma das pesquisas utiliza mineração de dados utilizando redes sociais, porém dentro do contexto de marketing (brand equity). Não foi encontrado nenhum trabalho que envolva o contexto de saúde mental relacionado a redes sociais e mineração de dados.

Uma dissertação na área de KDD com foco em redes sociais e saúde mental é a primeira a ser realizada no PPGGI até o ano de 2022, fato que pode incentivar novas pesquisas na área.

1.4 DELIMITAÇÕES DA PESQUISA

A pesquisa apenas coletará dados da comunidade do Subreddit “Desabafos”, tendo como foco apenas postagens em língua portuguesa.

A rotulação de dados foi realizada de forma colaborativa pela própria comunidade ao longo do tempo, não tendo acompanhamento metodológico ou supervisão de profissionais da saúde mental.

As postagens na comunidade do Reddit são anônimas, isto apresenta uma vantagem, logo tem-se maior tendência de serem verdadeiras e expressarem sentimentos reais, mas esse ponto também limita o entendimento de variáveis individuais dos autores, como gênero, idade, região geográfica, etc.

A pesquisa se limita a variáveis coletadas nos âmbitos citados, não trazendo variáveis que não estejam no escopo da rede social.

1.5 ESTRUTURA DO DOCUMENTO

O documento inicia com a fundamentação teórica, onde são apresentados conceitos das principais técnicas e assuntos utilizados ao longo do trabalho.

Em seguida é detalhada a metodologia do trabalho, que se caracteriza como uma pesquisa exploratória com amostragem por conveniência. É explicado a forma de coleta de dados, armazenamento e processamento para extração de variáveis, complementado pelas métricas de avaliação dos modelos estudados.

Para um melhor entendimento do estado da arte é apresentado uma revisão sistemática, onde realizou-se buscas de estudos que contemplam identificação de transtornos da saúde mental, por meio de aplicação de técnicas de mineração utilizando como fonte de dados as redes sociais.

Então são apresentadas as análises e discussões dos resultados do trabalho, iniciando pelo pré-processamento e extração de *features*, seguindo pela classificação dos textos, que engloba análise de sentimentos, total de sentenças, variância de sentimentos, extração de entidades, processamento de texto e aplicação de modelos de classificação.

Por fim, apresentam-se as considerações finais, com destaque para a continuidade do estudo.

2 REFERENCIAL TEÓRICO

Esta seção apresenta os principais conceitos relacionados à pesquisa realizada, tais como: saúde mental com ênfase para a depressão com suas características e sintomas; descoberta de conhecimento em bases de dados com definições, etapas técnicas e alguns métodos.

2.1 SAÚDE MENTAL

A depressão pode ser considerada uma síndrome que afeta diferentes áreas da vida de uma pessoa, ocorre tanto alterações de humor, como tristeza, irritabilidade, falta de capacidade de sentir prazer, apatia, como também alterações cognitivas, psicomotoras, alterações de sono e apetite (ALBERTO; PORTO, 1999; MAYER, 2017).

Essas alterações são normais em curtos períodos de tempo, mas quando duram mais que duas semanas e prejudicam atividades cotidianas, podem ser considerados sintomas de transtorno depressivo. Para o diagnóstico de depressão são considerados sintomas psicológicos, fisiológicos e comportamentais (ALBERTO; PORTO, 1999; MAYER, 2017).

Quadro 2 – SINTOMAS CONSIDERADOS NO DIAGNÓSTICO DE DEPRESSÃO

| Sintomas Psíquicos | Sintomas Fisiológicos | Evidências Comportamentais |
|--|--|---|
| <ul style="list-style-type: none"> • Humor depressivo, sensação e tristeza, autodesvalorização e sentimento de culpa. • Redução da capacidade de experimentar prazer na maior parte das atividades, antes consideradas como agradáveis. • Fadiga ou sensação de perda de energia. • Diminuição da capacidade de pensar, de se concentrar ou de tomar decisões. | <ul style="list-style-type: none"> • Alterações do sono (insônia e hipersonolência). • Alterações do apetite (perda e aumento do apetite). • Redução do interesse sexual. | <ul style="list-style-type: none"> • Retraimento social. • Crises de choro. • Comportamentos suicidas. • Retardo psicomotor e lentificação generalizada, ou agitação psicomotora. |

FONTE: Adaptada de Mayer (2017) e Porto (1999).

A tendência depressiva, ou tendência à depressão, é convencionalmente estimada por uma série de questionários, itens e entrevistas padronizadas, porém alguns estudos recentes utilizam de análises de texto, imagem e comportamento das pessoas com base no que é comunicado nas redes sociais. Portanto, a tendência depressiva pode ser caracterizada como expressão ou reflexo de alguns dos itens incluídos no inventário de sintomas de depressão (AARØ et al., 2011; KURODA et al., 2017; XIE; JIANG; ZHANG, 2018; YANG et al., 2019).

Tendo em vista que a depressão é um transtorno complexo que utiliza sintomas diversos para seu diagnóstico, bem como o presente trabalho se dispõe a analisar apenas o contexto comportamental inserido na exposição em redes sociais, é adotado o termo “tendência depressiva” como forma de identificar comportamentos correlatos de estudantes de medicina previamente diagnosticados com depressão.

2.2 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

Com a alta utilização e geração de dados nas redes sociais formam-se oportunidades de estudos de diversas áreas, porém os dados são disponibilizados de forma não estruturada, ou seja, sem padrões para fácil consumo computacional. Dentre as metodologias utilizadas para coleta e extração de conhecimento deste tipo de fonte de dados, destaca-se a descoberta de conhecimento em base de dados (do inglês *Knowledge Discovery in Databases - KDD*).

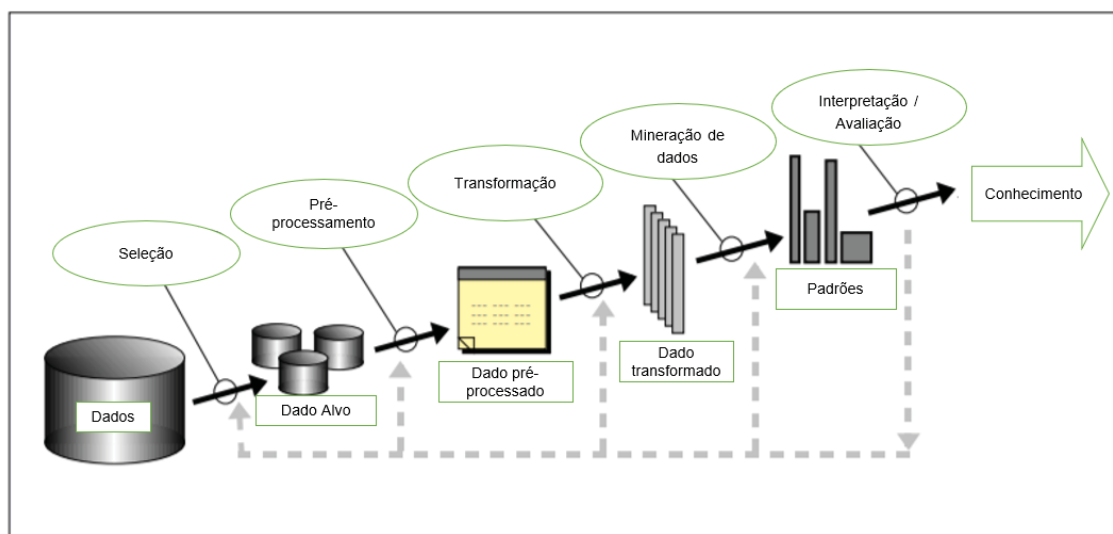
O *KDD* tem como o objetivo o desenvolvimento e utilização de métodos e técnicas computacionais para análise de dados normalmente com alto volume e baixa qualidade, para formatos mais facilmente legíveis, seja através de relatórios e/ ou gráficos ou modelos mais avançados que utilizam mineração de dados para descoberta de padrões e que possibilitam a geração de modelos preditivos (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

O termo descoberta de conhecimento em base de dados foi cunhado no primeiro workshop de *KDD* em 1989 (PIATETSKY-SHAPIRO, 1990), em que os autores, Fayyad e Piatetsky-Shapiro (1996), enfatizaram que o conhecimento é

o produto final de uma descoberta baseada em dados. Esse conceito foi popularizado nas áreas de inteligência artificial e aprendizado de máquina, conhecido em inglês por *machine learning*.

O processo de KDD possui diversas etapas, entre elas: preparação, seleção e limpeza dos dados, mineração, incorporação de conhecimentos prévios, interpretação dos resultados, entre outros (PIATETSKY-SHAPIRO, 1990).

FIGURA 4- VISÃO GERAL DAS ETAPAS QUE COMPÕEM O PROCESSO KDD.



FONTE: Adaptada PIATETSKY-SHAPIRO (1996).

Em cada etapa aplicam-se métodos e ferramentas computacionais que evoluem constantemente, ressaltando a natureza interdisciplinar do KDD. Áreas como Aprendizado de Máquina, Reconhecimento de Padrões, Estatística e Inteligência Artificial (conhecido como IA) tem recebido destaque (AZEVEDO; SANTOS, 2008; FAYYAD; HAUSSLER; STOLORZ, 1996; KAYACIK; ZINCIR-HEYWOOD; HEYWOOD, 2005; STEINER et al., 2006).

Alguns dos elementos essenciais e poderosos comumente utilizados em KDD são técnicas de aprendizado de máquina, principalmente nas etapas de pré-processamento e mineração de dados.

Conforme definição de Monard e Baranauskas (2003), Aprendizado de Máquina é:

uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas através da solução bem sucedida de problemas anteriores. Os diversos sistemas de aprendizado de máquina possuem características particulares e comuns que possibilitam sua classificação quanto à linguagem de descrição, modo, paradigma e forma de aprendizado utilizado[...].(MONARD; BARANAUSKAS, 2003, p. 14).

As técnicas de aprendizado de máquina são classificadas em técnicas de aprendizado supervisionado, não-supervisionado, semi-supervisionado e de aprendizagem por reforço.

As técnicas de aprendizado supervisionado utilizam ganho de conhecimento prévio e atual através de rotulação de dados para prever eventos. A abordagem inicial parte do processo de treinamento com um conjunto de dados, e através do aprendizado de máquina, desenvolver função de inferência para prever valores de saída para dados não rotulados(SARAVANAN; SUJATHA, 2019).

Técnicas de aprendizado não-supervisionado são empregadas quando os dados de treinamento não são classificados nem rotulados. O sistema não prediz a saída, mas grava observações de conjunto de dados para encontrar padrões ocultos e não rotulados(SARAVANAN; SUJATHA, 2019).

Técnicas semi-supervisionadas situam-se entre técnicas supervisionadas e não supervisionadas onde usam dados rotulados e não rotulados para o processo de treinamento(SARAVANAN; SUJATHA, 2019).

As técnicas de reforço de aprendizado interagem com o ambiente por ações e localizam erros ou recompensas, isto permite que os sistemas e programas identifiquem o comportamento ideal em um contexto específico para aumentar o seu desempenho(SARAVANAN; SUJATHA, 2019).

Existem diversas técnicas de aprendizagem de máquina, compostas por diversos algoritmos, onde a utilização de cada uma dependerá do problema a ser resolvido.

2.2.1 MODELOS DE CLASSIFICAÇÃO

Os algoritmos de análise preditiva são derivados principalmente de métodos estatísticos tradicionais ou de técnicas contemporâneas de aprendizado de máquina.

Os métodos estatísticos que tiveram o maior impacto na evolução da análise preditiva e da mineração de dados incluem análise discriminante, regressão linear e regressão logística. As técnicas de aprendizado de máquina mais populares usadas em vários projetos bem-sucedidos de análise preditiva incluem árvores de decisão, vizinhos mais próximos, redes neurais artificiais e máquinas de vetores de suporte. Todas essas técnicas de aprendizado de máquina podem lidar com problemas de classificação e previsão do tipo regressão. Frequentemente, são aplicadas em problemas complexos de predição onde outras técnicas mais tradicionais não são capazes de produzir resultados satisfatórios (DELEN, 2020).

No presente trabalho foram utilizados quatro métodos de classificação, sendo eles: Random Forest, Naïve Bayes, Regressão Logística e SVM. Os métodos são abordados nas subseções abaixo.

2.2.1.1 Random Forest

Random Forest (florestas aleatórias) são um tipo de ensemble method (método de conjunto). Um ensemble method é um processo no qual vários modelos são ajustados e os resultados são combinados para previsões mais fortes. Embora isso forneça ótimas previsões, a inferência e a explicabilidade geralmente são limitadas. (LANDER P. JARED, 2017)

Random Forest são compostas por um número de árvores de decisão onde os preditores e observações incluídos são escolhidos aleatoriamente. O nome vem da construção aleatória de árvores para formar uma floresta. (LANDER P. JARED, 2017)

A Random Forest oferece suporte à importância das variáveis (*feature importance*), mas não há uma árvore de decisão única pela qual é possível percorrer, apenas inspecionar árvores isoladas do conjunto. (HARRISON; SAFARI, 2019)

O pseudo código do Random Forest pode ser dividido em duas etapas, sendo a primeira a criação da Random Forest e a segunda o pseudo código para realizar previsões. As duas etapas do pseudo código são apresentadas abaixo (LI et al., 2018):

1. Dado um conjunto de treinamento com n amostras e m características:
 - Defina o número de árvores no modelo (número de estimadores).
 - Defina o número máximo de características a serem consideradas em cada divisão (*max_features*).
 - Para cada árvore no modelo:
 - i. Amostra aleatoriamente o conjunto de treinamento com reposição (*bootstrap*).
 - ii. Crie uma árvore de decisão utilizando o conjunto de treinamento amostrado.
 - iii. Para cada nó na árvore:
 1. Escolha um subconjunto de características aleatoriamente (de acordo com *max_features*).
 2. Divida o nó com base em uma métrica de impureza (por exemplo, Gini ou entropia).
 3. Repita os passos anteriores para cada nó filho até atingir um critério de parada (profundidade máxima da árvore ou número mínimo de amostras em um nó).
 - iv. Armazene todas as árvores criadas.
2. Para fazer uma previsão para uma nova amostra:
 - Para cada árvore no modelo:
 - i. Passe a amostra pela árvore seguindo as divisões até chegar a uma folha.
 - ii. Registre a classe ou o valor da folha alcançada.
 - Calcule a classe ou o valor final com base nas previsões de todas as árvores (por exemplo, moda para classificação ou média para regressão).

2.2.1.2 Naïve Bayes

Naïve Bayes é um método simples de classificação baseado em probabilidade derivado do conhecido teorema de Bayes. O método exige que as variáveis de saída tenham valores nominais/categóricos. Embora as variáveis de entrada possam ser uma mistura de tipos numéricos e nominais, a variável de saída numérica precisa ser discretizada por meio de algum tipo de método de categorização antes de poder ser usada no classificador Naïve Bayes. (DELEN, 2020)

A parte “ingênua” do nome Naïve Bayes (Naïve, do inglês, significa “Ingenuo”) vem da suposição forte e um tanto irreal de independência entre as variáveis de entrada. Simplificando, o classificador Naïve Bayes assume que as variáveis de entrada não dependem umas das outras, e a presença (ou ausência) de uma variável particular na mistura dos preditores não tem nada a ver com a presença ou ausência de quaisquer outras variáveis.(DELEN, 2020)

Naïve Bayes é frequentemente associado a aplicativos NLP (Natural Language Processing), como reconhecimento de spam ou análise de sentimentos. (COOK, 2017)

O exemplo abaixo ilustra o fluxo geral do algoritmo do Naïve Bayes (GARG, 2013):

1. Dado um conjunto de treinamento com n amostras e m características:
 - Calcule a probabilidade a priori de cada classe no conjunto de treinamento;
 - Para cada característica e cada classe:
 - i. Calcule a probabilidade de ocorrência da característica dado que a amostra pertence à classe.
 - Armazene as probabilidades calculadas.

2. Para fazer uma previsão para uma nova amostra:
 - Calcule a probabilidade a posteriori de cada classe para a nova amostra, utilizando as probabilidades armazenadas;

- Selecione a classe com a maior probabilidade a posteriori como a classe prevista para a nova amostra.

2.2.1.3 Regressão Logística

Regressão logística é um algoritmo de classificação baseado em probabilidade muito popular, estatisticamente sólido, que emprega aprendizado supervisionado.

Foi desenvolvido na década de 1940 como um complemento à regressão linear e aos métodos de análise discriminante linear. Tem sido amplamente utilizado em várias disciplinas, incluindo os campos das ciências médicas e sociais.

A regressão logística é semelhante à regressão linear na medida em que também visa regredir a uma função matemática que explica a relação entre a variável de resposta e as variáveis explicativas, usando uma amostra de observações passadas (dados de treinamento).

Ela difere da regressão linear em um ponto importante: sua saída (variável de resposta) é uma classe em oposição a uma variável numérica. Ou seja, enquanto a regressão linear é usada para estimar uma variável numérica contínua, a regressão logística é usada para classificar uma variável categórica.

Embora a forma original de regressão logística tenha sido desenvolvida para uma variável de saída binária (por exemplo, 1/0, sim/não, aprovado/reprovado, aceitar/rejeitar), a versão modificada atual é capaz de prever variáveis de saída de múltiplas classes (isto é, regressão logística multinomial) (DELEN, 2020).

2.2.1.4 SVM

SVM é um aprendizado supervisionado não probabilístico modelo que pertence à família dos classificadores binários lineares. Ele basicamente constrói um hiperplano que atua como um descritor de classe entre dois conjuntos de dados de forma otimizada.

O SVM se concentra nos vetores de treinamento que estão exatamente na fronteira dos descritores de classe que são chamados de vetores de suporte.

Durante a fase de teste, apenas os vetores de suporte são considerados para decidir sobre a classe do vetor de teste (SURIYA PRAKASH et al., 2012).

Embora algumas pessoas caracterizem as SVMs como um caso especial de RNAs – Rede Neurais Artificiais, a maioria as reconhece como duas técnicas concorrentes de aprendizado de máquina com qualidades diferentes.

Historicamente, o desenvolvimento de RNAs seguiu um caminho heurístico, com aplicações e extensa experimentação precedendo a teoria. Em contraste, o desenvolvimento de SVMs envolveu primeiro uma sólida teoria de aprendizado estatístico, seguida de implementação e experimentos.

Uma vantagem significativa das SVMs é que, embora as RNAs possam sofrer de múltiplos mínimos locais, as soluções para SVMs são globais e únicas. Além disto, os SVMs têm interpretações geométricas simples e fornecem soluções amplas.

A razão pela qual as SVMs geralmente superam as RNAs na prática é que elas lidam com sucesso com o problema de “overfitting”, que é um grande problema com as RNAs.(DELEN, 2020)

Apresentados os principais conceitos, a próxima seção apresenta os encaminhamentos metodológicos adotados.

3 PERCURSO METODOLÓGICO

A metodologia utilizada para esta pesquisa é do tipo qualitativa exploratória. Segundo Moreira (2011), caracteriza-se pela busca de explicações interpretativas, levando em consideração aspectos sociais do contexto pesquisado, através de uma realidade social construída.

Para o autor, o pesquisador assume um papel de grande participação no estudo, estando imerso neste, cabendo a ele realizar observações, tomar notas, buscar significados, levantar hipóteses e então, após a constituição de todos os dados, realizar uma análise reflexiva.

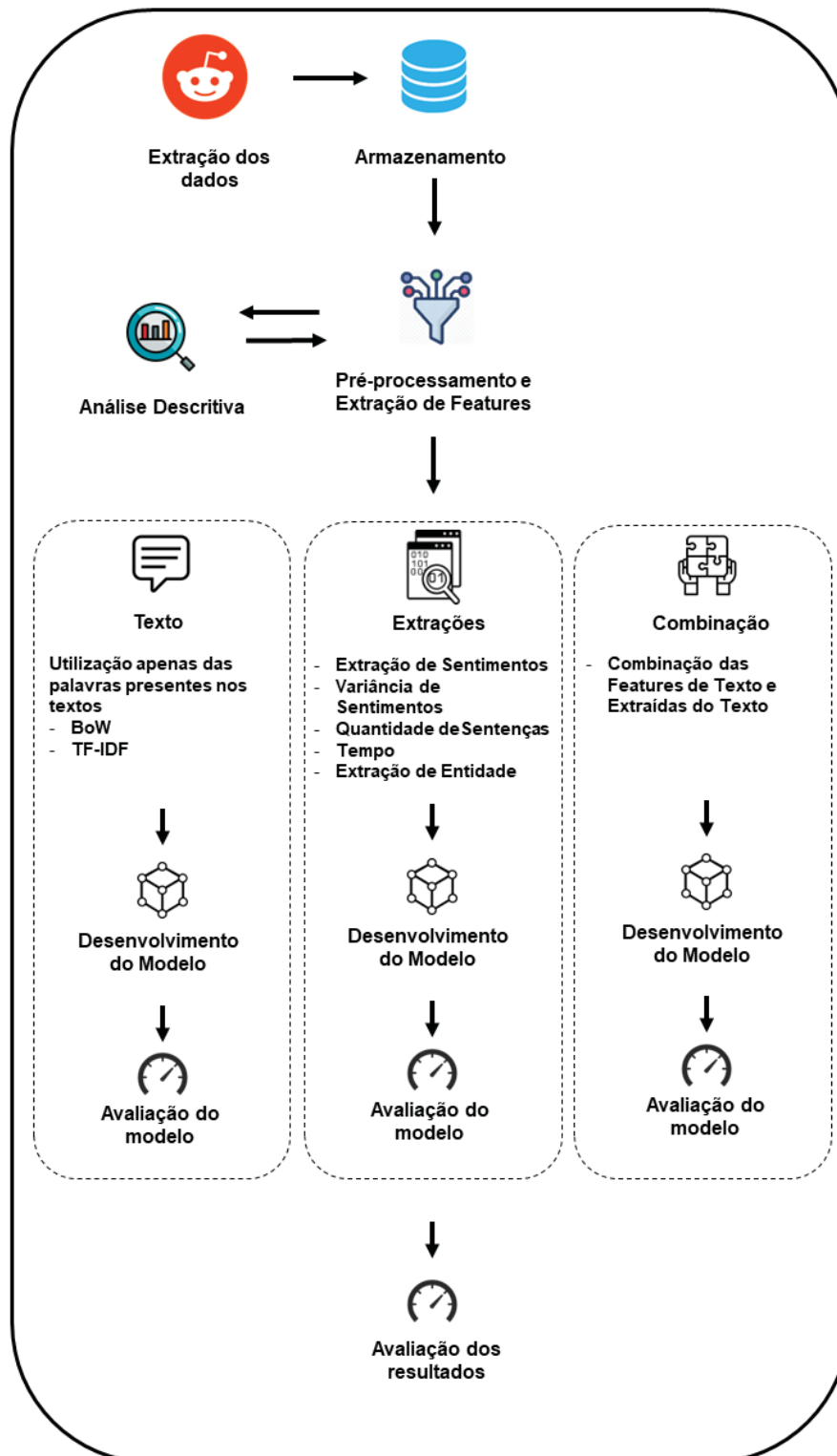
3.1 PERCURSO METODOLÓGICO

A metodologia deste trabalho está dividida nas seguintes etapas:

1. Revisão sistemática das metodologias e tecnologias utilizadas para resolução do problema alvo;
2. Desenho da metodologia geral do trabalho;
3. Captura e exploração dos dados obtidos por meio das redes sociais;
4. Pré-processamento e extração de “*features*” dos dados;
5. Análise e estatística descritiva para maior entendimento dos dados e “*features*” extraídas;
6. Comparativo de performance de classificadores para rotulação automática;
7. Aferição e discussão dos resultados

As etapas descritas apresentadas e detalhadas na Figura 8. A etapa de extração de dados foi realizada utilizando a api do repositório de dados do Reddit, a PushShift, utilizando a linguagem de programação Python. Como parâmetros, os dados foram coletados da comunidade ‘/r/ Desabafos’ considerando o período de 01/01/2021 até 30/06/2022. Os dados foram armazenados em um banco de dados local, estando prontos para as etapas de pré-processamento e análise descritiva.

FIGURA 5 - VISÃO MACRO DO FRAMEWORK PROPOSTO PARA DESENVOLVIMENTO DO PROJETO DE PESQUISA



FONTE: O Autor (2022).

O pré-processamento foi organizado em duas etapas, a etapa de limpeza dos dados e a etapa de extração de features. A etapa de limpeza dos dados compreende um conjunto de técnicas foi aplicado visando maior qualidade dos dados para a análise descritiva e implementação dos algoritmos de classificação. A segunda etapa trata-se de da extração de features, onde técnicas de extração de metadados foram implementados no dataset, enriquecendo as dimensões disponíveis para a etapa de classificação.

A etapa de análise descritiva do dataset foi realizada utilizando técnicas de estatística descritiva, possibilitando maior entendimento dos dados coletados e gerados após processo de extração de features. Essa etapa ocorreu em paralelo ao pré-processamento, pois além de facilitar o entendimento do conjunto de dados, evidenciou a necessidade de limpeza dos dados e também oportunidades na criação dos metadados.

As etapas de classificações das dimensões foram segmentadas em três fases. A primeira fase contemplou a implementação de classificadores considerando apenas a dimensão do texto das postagens na comunidade (após pré-processamento), a segunda fase compreendeu a implementação de classificadores com as dimensões de metadados. Por fim, a terceira etapa foi implementada considerando as dimensões do texto e metadados. Todas as etapas foram submetidas para 4 algoritmos, sendo eles: Naive Bayes, Random Forest, Regressão Logística e SVM. A avaliação dos algoritmos se deu através da métrica de acurácia.

Após resultados, é realizada uma leitura dos mesmos e possíveis hipóteses, de forma complementar também é exposto o resultado da importância das features do resultado de maior destaque, para entendimento de quais dimensões tiveram maior impacto nas classificações.

Na sequência cada etapa é explicada com maior detalhamento.

3.2 COMUNIDADE DA REDE SOCIAL REDDIT

A escolha da comunidade (que são denominadas como 'subreddit') '/r/ Desabafos' (sendo o '/r/' o prefixo padrão de uma comunidade) na rede social Reddit como foco do presente trabalho foi determinado pelos seguintes fatores:

1. Relevância da comunidade dentro da rede social, que se traduz pela quantidade de inscritos dentro da comunidade
2. Postagens devem ser em português brasileiro
3. Comunidade deve ter classificação manual das postagens
4. Classificação das postagens devem considerar tema da pesquisa (saúde mental e depressão)

Em pesquisa realizada na rede social Reddit no mês de Maio de 2022, considerando os fatores acima, a comunidade ‘/r/ Desabafos’ se destacou por possuir cerca de 470 mil usuários, suas postagens são em português e os temas envolvem saúde mental. Sua relevância é comprovada quando identificado que menos de 0,5% das comunidades possuem mais que 300 mil inscritos (como referência, na data da pesquisa o Reddit contava com cerca de 326 mil comunidades).

Em perspectiva com comunidades Brasileiras, a maior comunidade brasileira no Reddit é a ‘/r/brasil’, possuindo cerca de 1,5 milhões de usuários, porém seu conteúdo é generalista, englobando todos temas que possuem referência o Brasil.

3.3 COLETA DOS DADOS

A coleta de dados da rede social Reddit foi realizada utilizando o repositório de dados PushShift, que possui todo o histórico de dados de postagens do Reddit. A escolha da rede social Reddit como foco do presente trabalho se deve após a revisão sistemática e exploração dos dados presente nas redes sociais. O fato de maior impacto foi a facilidade de extração dos dados em conjunto com a anonimização dos usuários.

A API pushshift.io Reddit foi projetada e criada pela equipe de modificação /r/datasets para ajudar a fornecer funcionalidade aprimorada e recursos de pesquisa para pesquisar comentários e envios do Reddit (BAUMGARTNER, 2017).

O PushShift possui APIs (application programming interfaces) públicas para coleta de dados, sendo esse o método principal para coleta das postagens.

Os dados foram coletados exclusivamente utilizando o Subreddit /r/ Desabafos, no período de 01/01/2021 até 30/06/2022.

Para intermediação da API com o banco de dados foi utilizado a linguagem de programação Python utilizando a biblioteca *requests*. Os dados foram armazenados em um banco de dados *Mysql* local para facilitar a modelagem para a fase de pré-processamento.

O Dataset está disponível para consulta através do link público https://github.com/MarcoGrutzmann/Dataset_Reddit_Desabafos/wiki. Nota-se que o dataset possui as dimensões de metadados que serão citados no decorrer do trabalho para facilitar a análise, utilização e reprodução do presente trabalho.

3.4 AMOSTRAGEM

Devido às características do trabalho, a amostragem será do tipo por conveniência, uma amostragem não probabilística que seleciona unidades amostrais mais acessíveis e usualmente não é representativa da população alvo. Deve ser utilizada em um estudo exploratório inicial, abrindo campo para uma amostragem probabilística posterior.

Por ser não-probabilística, essa amostragem não permite a realização de inferências. (FARIAS, A. A.; SOARES, J. F.; CÉSAR, C. C. Introdução à Estatística. 2.ed. Rio de Janeiro: LTC, 2008. 3 LOESCH, C. Probabilidade e Estatística. Rio de Janeiro: LTC, 2012).

3.5 PRÉ-PROCESSAMENTO E EXTRAÇÃO DE FEATURES

Para maior compreensão do conjunto de dados e preparação para implementação dos modelos de classificação, serão utilizadas técnicas de pré-processamento para extração das *features*.

3.5.1 Pré-processamento

O conjunto de dados foi submetido em um processo de limpeza dos dados, composto pelas seguintes etapas:

- exclusão de postagens com quantidade de caracteres inferior a 15 caracteres, realizado no banco de dados em SQL (PostgreSQL®) utilizando filtro com a função *length*;
- concatenação dos títulos das postagens com o corpo do texto (devido a característica da rede social), realizado no banco de dados em SQL (PostgreSQL®) utilizando a função *concat*;
- remoção de stop words utilizando a biblioteca nltk em python;
- remoção de pontuação utilizando a biblioteca nltk em python;
- substituição de caracteres especiais utilizando a biblioteca unidecode em python;
- normalização de letras maiúsculas, tornando todo o texto com letras minúsculas), realizado no banco de dados em SQL (PostgreSQL®) utilizando a função *concat*;
- implementação da técnica de lematização para redução de dimensionalidade das palavras, utilizando a biblioteca *SpaCy* em python carregando como parâmetro a coleção *'pt_core_news_sm'*.

Na etapa de utilização dos classificadores também foi implementada a técnica de Tokenização das palavras, utilizando a biblioteca *sklearn* em python (*sklearn.feature_extraction.text*).

A filtragem dos textos pela quantidade de caracteres e a junção do título com o corpo do texto, foram realizadas utilizando a linguagem SQL dentro do próprio banco de dados (PostgreSQL®).

As técnicas de remoção de stop words, remoção de pontuação, substituição de caracteres especiais e normalização de letras minúsculas foram realizadas utilizando a linguagem Python, com apoio da biblioteca NLTK (Natural Language Toolkit) e unidecode.

Para a implementação da técnica de lematização, utilizou-se a linguagem Python com biblioteca SpaCy.

3.5.2 Extração de *features*

Com base nos aprendizados da revisão da literatura, em conjunto com os dados disponíveis extraídos do Reddit, utilizou-se uma combinação de modelos pré-treinados do Google Cloud Platafom (discorridos abaixo) e construção de atributos para a composição das dimensões do conjunto de dados.

Dimensões geradas utilizando modelos pré-treinados:

- identificação de sentimentos: extração de sentimento dos textos será realizada com o auxílio da API de processamento de linguagem natural do Google Cloud Platform que possui suporte ao idioma português. O retorno da API é o Score de sentimento do texto de -1 a 1, sendo menor que 0, negativo, e maior que zero, positivo.
- desenvolvimento da variação de sentimentos no texto: a API utilizada acima também consegue classificar o sentimento específico dos trechos do texto, isso possibilita entendermos a variação dos scores de sentimento do documento como outra dimensão de *feature* em si, esse processo é descrito com detalhes na seção de análise de resultados.
- extração de entidades: a extração de sentimento dos textos será realizada com o auxílio da API de processamento de linguagem natural do Google Cloud Platform que possui suporte ao idioma português. O retorno da API são as entidades presentes no texto, como menção a pessoas, eventos, datas, etc. Isso possibilita entendermos mais a fundo o contexto do texto. A presença ou ausência de determinadas entidades, em conjunto com a quantidade de menções de cada entidade, podem ter relevância na classificação das classes.

O Google Cloud API é uma plataforma mantida pela empresa Google, seus serviços são pagos, porém os serviços utilizados neste trabalho se

enquadram dentro das camadas gratuitas do serviço de processamento de linguagem natural, não sendo necessário direcionar recursos financeiros para extração dos atributos descritos acima.

Dimensões extraídas pelo texto e horário de postagem:

- total de sentenças por postagem: contagem do total das sentenças (frases) dentro do escopo do texto.
- categorização do horário das postagens: categorização por turno de acordo com o horário da postagem, sendo “manhã”, “tarde”, “noite” e “madrugada”.

Também foram utilizadas técnicas de *features* textuais, como bag of words e TF-IDF (citadas na revisão sistemática). Para a criação do TF-IDF utilizado como parâmetro de máximo de *features* de 4.000 (parâmetro em python: *max_features*).

3.5.3 Rotulação dos dados

O conjunto de dados foi rotulado pela própria comunidade do Reddit, utilizando a funcionalidade de adição de rótulos (marcações) vinculadas a postagem. Esses rótulos são fixas e criadas pelos administradores da comunidade, utilizando como base a maior quantidade de temas discutidos dentro da comunidade. O conteúdo de cada rótulo e seus exemplos são abordados no quadro 4 da seção 4 deste documento.

O processo para inclusão de rótulos pode ser realizado por meio de dois fluxos. O primeiro fluxo é realizado pelo próprio usuário responsável pela criação da sua postagem realizar a marcação de acordo com os rótulos disponíveis.

O segundo fluxo pode ser realizado pelos moderadores da comunidade, onde podem editar a tag colocada previamente pelo usuário, ou inserir o rótulo caso o usuário não tenha inserido previamente.

3.6 ANÁLISE E ESTATÍSTICA DESCRITIVA

Com o objetivo de aprofundar nas dimensões do conjunto de dados, tal como entender a relação das dimensões com as suas classes, realizou-se a análise dos dados utilizando como base a estatística descritiva.

A estatística descritiva tem como objetivo a descrição dos dados, sejam eles de uma amostra ou de uma população. Pode incluir (FERREIRA, 2005):

- verificação da representatividade ou da falta de dados;
- ordenação dos dados;
- organização dos dados em tabela;
- criação de gráficos com os dados;
- calcular valores de sumário, tais como médias;
- obter relações funcionais entre variáveis.

3.6.1 Variância de Sentimentos

Devido a feature “Variância de Sentimentos” ser desenvolvida utilizando um Método estatístico, será descrito neste tópico a sua construção.

Na teoria da probabilidade e na estatística, a variância de uma variável aleatória ou processo estocástico é uma medida da sua dispersão estatística, indicando, de forma geral, o quão distante os seus valores se encontram do valor esperado. (RUNGER, 2002)

A definição da variância se dá pela fórmula:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Onde:

σ^2 : variância

x_i : valor analisado

\bar{x} : média aritmética do conjunto

n : número de dados do conjunto

Neste trabalho a implementação da variância foi realizada utilizando métodos nativos do banco de dados PostgreSQL®, descritos no tópico de pré-processamento de dados.

3.7 MODELOS DE CLASSIFICAÇÃO

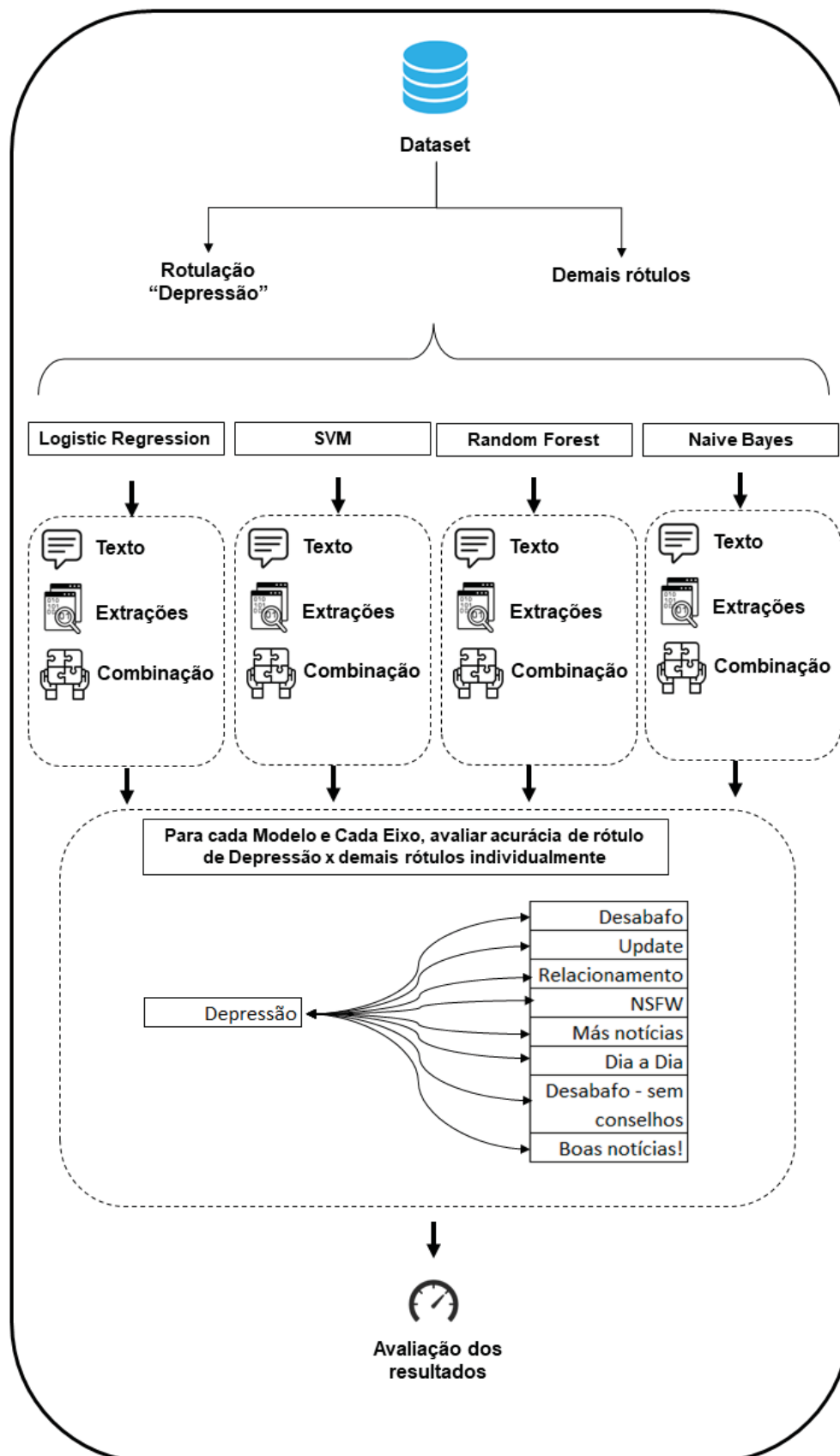
Com o texto devidamente tratado e processado, em conjunto com o processamento e geração de *features* derivadas de técnicas de mineração de dados, desenvolvido os modelos de classificação conforme fluxo de trabalho presente na Figura 9.

Segmentado o *dataset* entre as *features* extraídas do texto e informações adicionais da comunidade, como o horário da postagem, e as *features* puramente textuais. Mesmo dependentes, o experimento será para verificar a acurácia das técnicas de classificação de três formas:

- 1) classificação apenas pelos elementos textuais;
- 2) classificação apenas pelas *features* extraídas do texto e adicionais da extração, aos quais são chamados de metadados;
- 3) classificação combinada das opções anteriores.

Para desenvolvimento do classificador, considerando que o foco é depressão, cada classificador será implementado com a classe alvo “depressão” presente, com a classe comparativa como “não depressão”. Exemplo: “depressão” e “desabafo” assumindo a notação de “não depressão”. Dessa forma é possível comparar a performance dos classificadores e qual classe faz mais sentido em ser a notação de “não depressão”. Para os rótulos que não são ‘depressão’, a figura 9 os cita como ‘demais rótulos’ para exemplificar e facilitar o visual da figura.

FIGURA 6 – FLUXO DE TRABALHO DO CONTEXTO DOS CLASSIFICADORES



FONTE: O Autor (2022).

Os algoritmos selecionados para utilização são o Random Forest, Naïve Bayes, Regressão Logística e SVM, sendo esses os que tiveram prevalência em trabalhos semelhantes encontrados na etapa de revisão da literatura.

3.8 AVALIAÇÃO DOS MODELOS DE CLASSIFICAÇÃO

Para treinamento e validação do modelo, será utilizada a técnica de “tem-fold cross-validation”, que consiste em quebrar o dataset em partes menores e validando cada novo treinamento entre as partes menores (KOHAVI, 1995).

Para todos os modelos, a métricas de acurácia foram acompanhadas como sinal de qualidade do modelo. A acurácia consiste na divisão do número de classificações corretas pelo número total de classificações.

3.9 IMPORTÂNCIA DAS FEATURES

As *features* (ou características) são variáveis utilizadas para treinar modelos de Machine Learning, como o Random Forest. A importância das *features* refere-se à sua contribuição na tomada de decisão do modelo.

No algoritmo Random Forest, a importância das features é geralmente avaliada usando a medida chamada "Importância de Gini" ou "Gini Importance". Essa medida é baseada no índice de impureza de Gini, que é uma medida de quão misturados estão os dados de diferentes classes em um nó de decisão. A Importância de Gini mede o quanto cada feature contribui para reduzir o índice de impureza de Gini médio nos nós do modelo. (BREIMAN, 2001)

O fluxo de cálculo da importância das *features* é dado através dos passos (ALSAGRI; YKHLEF, 2020):

- 1) Calcular a importância dos nós n_j para cada nó j para cada árvore de decisão, usando a seguinte equação:

$$n_j = W_j C_j - W_{left(j)} C_{left(j)} - W_{right(j)} C_{right(j)}$$

Onde W_j é o nó j probabilidade de atingimento e C_j é o nó da impureza de Gini. O mesmo é para o nó à direita e a esquerda do nó filho de j .

- 2) A importância de cada feature (F) na árvore é calculado usando a equação 2, onde m é o total de nós:

$$F_{(j)} = \frac{n_j}{\sum_{i=1}^m n_i}$$

- 3) A importância de cada feature na Random Forest (coleção de k árvores) é calculada usando a seguinte equação:

$$(i) = \frac{\sum_{j=1}^m F_{(j)}}{k}$$

Para implementação no presente trabalho, foi utilizada a biblioteca sklearn com a linguagem de programação python que já possui os métodos de cálculo da feature importance utilizando Random Forest.

4 RESULTADO: REVISÃO SISTEMÁTICA

Para desenvolvimento da revisão sistemática, utilizou-se a metodologia Prisma. “A recomendação PRISMA consiste em um *checklist* com 27 itens e um fluxograma de quatro etapas. O objetivo do PRISMA é ajudar os autores a melhorarem o relato de revisões sistemáticas e meta-análises.” (GALVÃO; PANSANI; HARRAD, 2015)

O objetivo da revisão é identificar e analisar os métodos e técnicas existentes para identificação de tendências depressivas em redes sociais utilizando técnica de mineração de dados.

4.1 METODOLOGIA DA REVISÃO SISTEMÁTICA

O critério da seleção de bases acadêmicas deu-se através da compatibilidade da pesquisa e dos acessos da instituição vinculado, se restringindo à Scopus (Elsevier), PubMed e American Psychologic Association (PsychInfo), em um período de 5 anos, sendo de 01/06/2016 à 01/06/2021.

Como critérios de busca, pesquisou-se nos títulos e palavras-chaves das mencionadas bases. Foi desconsiderado o resumo (ou Abstract) devido à alta densidade de artigos não relacionados com o tema aparecem, fugindo do escopo desse trabalho.

Para resgate dos artigos para a revisão, realizaram-se buscas utilizando um *string* de Busca (com adaptações para cada revista) no idioma Inglês, conforme tabela abaixo:

TABELA 2– CRITÉRIOS PARA STRING DE PESQUISA NAS BASES ACADÊMICAS

| Categoria | Palavras-Chave |
|-----------------------------|---|
| Saúde Mental (1) | Mental Health, Mental Disorder, Anxiety, Stress, Suicide, Depression, Psychiatry |
| Redes Sociais (2) | Social Network, Facebook, Instagram, Tiktok, Twitter, Social Media, Reddit |
| Inteligência Artificial (3) | Predictive, Machine Learning, Forecast, AI, Artificial Intelligence, Data Mining, Deep Learning |
| Termos de Busca | (1) AND (2) AND (3) |

FONTE: O Autor (2021).

A lógica “ou” (OR) foi utilizada entre as palavras-chave de uma mesma categoria. A lógica “e” (AND) foi utilizada entre as três categorias garantindo que exista a combinação de pelo menos uma palavra-chave de cada categoria. Como resultado, obteve-se um total de 89 artigos.

Os critérios de inclusão e exclusão de artigos são citados a seguir:

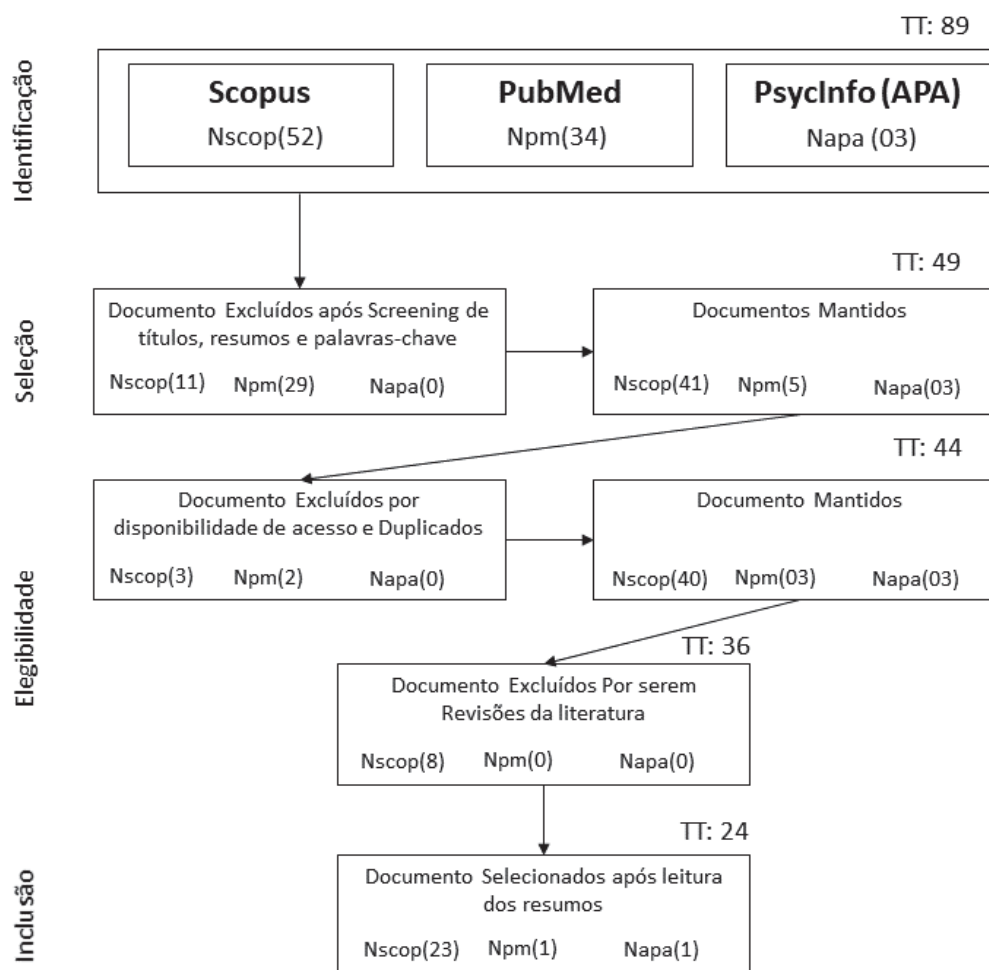
TABELA 3 – DESIGN DE CRITÉRIOS ACEITE OU RECUSA PARA REVISÃO

| Critérios Inclusão | Critérios Exclusão |
|---|--|
| Incluídos trabalhos recentes (publicados a partir de 2015) que já possuam aprovação pela comunidade científica. | Excluídos trabalhos que aferem a saúde mental por meios físicos |
| Incluídos trabalhos cujo foco é predição de transtornos mentais pela utilização das redes sociais | Excluídos trabalhos que apresentam avaliações sem apresentar o método utilizado. |
| Incluídos trabalhos que estejam aberto para consulta | Excluídos trabalhos cujo foco é o impacto da utilização de redes sociais na saúde mental |
| | Excluídos trabalhos que não utilizam técnicas de IA/Data Mining ou correlatos |
| | Excluídos trabalhos cujo foco não é aferição da saúde mental |
| | Excluídos artigos que utilizam fontes complementares à rede social para aferição dos transtornos mentais do objeto de estudo |
| | Excluídos trabalhos que não estão abertos para consulta |
| | Excluídos trabalhos que são revisões de literatura |

FONTE: O Autor (2021).

Após aplicação dos critérios de Elegibilidade nos 89 artigos extraídos das bases, seguindo a recomendação PRISMA, empregou-se os critérios de inclusão e criou-se o fluxograma disposto abaixo demonstrado a metodologia de seleção.

FIGURA 7– FLUXOGRAMA DETALHANDO PROCESSO DE ELEGIBILIDADE DE TRABALHOS PARA REVISÃO SISTEMÁTICA



FONTE: O Autor (2021).

Destaca-se que 8 artigos de revisão (sistemática, crítica e bibliométrica) foram excluídos e, ao final dos filtros, obteve-se 24 artigos para análise e maior detalhamento.

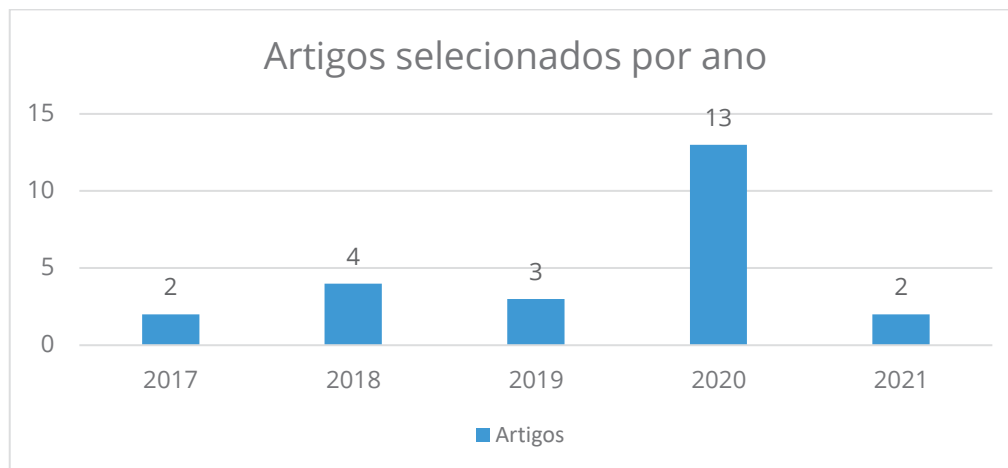
4.2 RESULTADOS

Neste tópico são apresentados os resultados derivados do método relatado.

A distribuição por ano de publicação dos artigos selecionados tem predominância no ano de 2020, sendo 13 artigos (52%), seguido de 2018 com 4 artigos (16%), 2019 com 3 artigos (12%) e 2017 e 2021 ambos com 2

artigos cada (8% e 8%, respectivamente). A distribuição pode ser acompanhada através do **Erro! Fonte de referência não encontrada.**

GRÁFICO 2 – TOTAL DE ARTIGOS SELECIONADOS POR ANO DE PUBLICAÇÃO



FONTE: O Autor (2021).

A revisão foi segmentada em 4 grandes eixos de aprofundamento: saúde mental e redes sociais, seleção de amostra e rotulação, técnicas de classificação e predição e aferição de resultados.

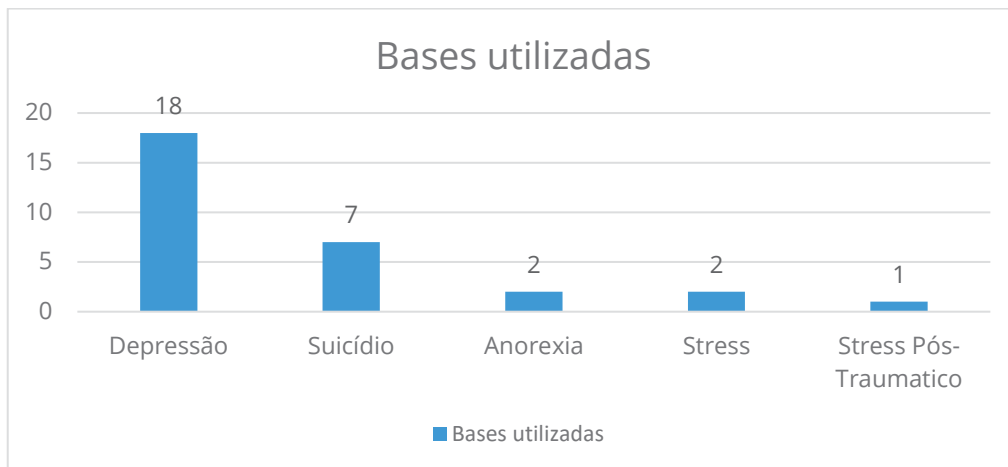
4.2.1 Saúde mental e redes sociais

Dentre os trabalhos selecionados, o transtorno com maior quantidade de estudo é a depressão, contendo 18 trabalhos (72%) que procuram desenvolver modelos para sua predição. Dentro dos trabalhos de depressão, os que apresentam como foco exclusivo à depressão são 56% (14), porém 4 trabalhos (16%) possuem focos mistos, sendo 2 (8%) trabalhos relacionados à depressão e suicídio e 2 (8%) trabalhos relacionados à depressão e anorexia. Além dos trabalhos mistos citados, ainda há 1 (4%) trabalho misto de stress e suicídio. Apenas 1 (4%) trabalho foca em stress pós-traumático e 1 (4%) focado em stress.

Os demais trabalhos (7) possuem como foco o desenvolvimento de modelos relacionados a suicídio.

A distribuição está ilustrada no Gráfico 3.

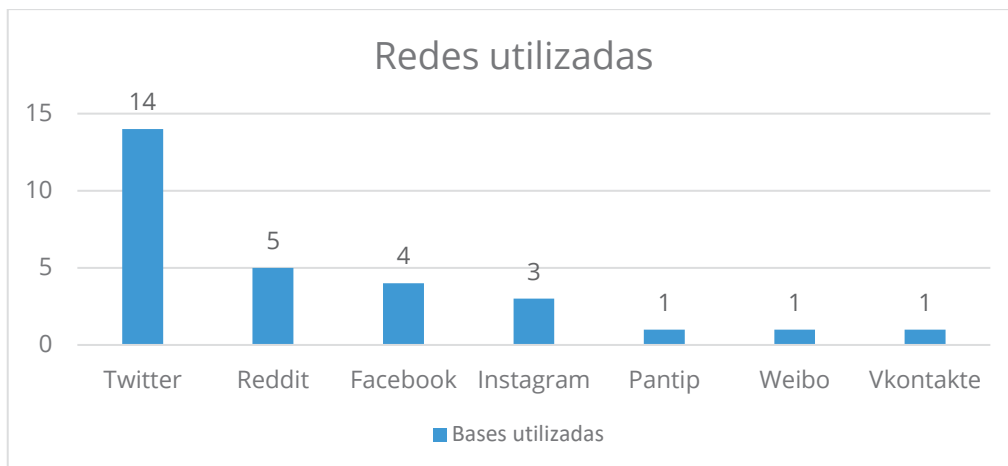
GRÁFICO 3 – DISTRIBUIÇÃO DOS TRANSTORNOS MENTAIS ESTUDADOS NOS TRABALHOS



FONTE: O Autor (2021).

Com relação as redes sociais estudadas, observa-se a predominância do Twitter como base para estudos (14 ou 56%), destes, 11 utilizam apenas o Twitter, outros 3 combinam o estudo com o Twitter e outras redes sociais.

GRÁFICO 4 – REDES SOCIAIS UTILIZADAS COMO BASE DOS ARTIGOS



FONTE: O Autor (2021).

Twitter difere das demais redes sociais por ser uma rede onde cada postagem do usuário deve ter no máximo 140 caracteres. Percebe-se que a ampla utilização da rede acontece pela facilidade na extração de dados através

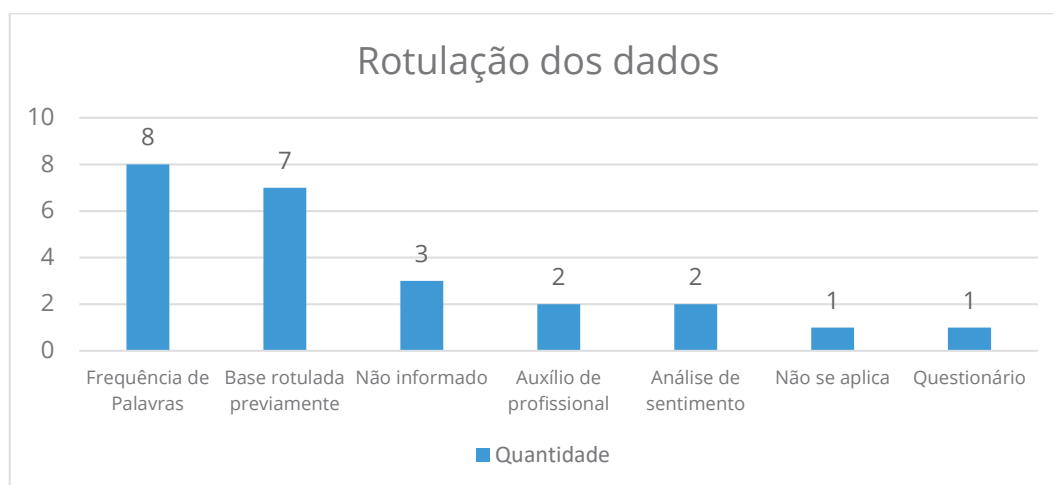
de criação de queries complexas pela sua API(GOVINDASAMY; PALANICHAMY, 2021).

Na sequência tem-se Reddit em 5 trabalhos (20%), Facebook em 4 (16%) e Instagram em 3 (12%). Pantip (Rede Social Indiana), Weibo (Rede Social Asiática) e Vkontakte (Rede Social Russa) aparecem cada um em 1 trabalho, totalizando os 3 em 12% dos trabalhos analisados.

4.2.2 Seleção de amostra e rotulação

Pela característica da rotulação e construção dos modelos, não foi observado nos trabalhos analisados a presença de fundamentações estatísticas no que se refere à escolha amostral dos objetos de estudo.

GRÁFICO 5 – TÉCNICAS UTILIZADAS PARA ROTULAÇÃO DOS DADOS



FONTE: O Autor (2021).

A técnica de rotulação com maior presença nos estudos foi a de frequência ou presença de palavras-chave que possuem relação aos transtornos de saúde mental, totalizando 8 estudos (33%). Como exemplo do trabalho de Samah J. Fodeh e colaboradores (2019), foi utilizado como referência os termos apresentados na imagem abaixo para classificação de “Alto Risco para Suicídio” todos os Tweets (nomes das postagens realizadas na rede social Twitter) que continham os termos (em inglês) apresentados na **Erro! Fonte de referência não encontrada..**

FIGURA 8 – RELAÇÃO DE PALAVRAS (EM INGLÊS) RELACIONADAS À SUICÍDIO

| | | | |
|------------|-----------------|------------------------|----------------|
| Abused | Prozac | Cut | Panic Disorder |
| Depressed | Pills | Bully | Social Anxiety |
| Hopeless | Suicide Abused | Bullied | Fight Dad |
| Worthless | Suicide Pain | Suicide Gun | Fight Mom |
| Empty | Suicide Tried | Suicide Shoot | Fight Parents |
| Anxious | Suicide Mom | Schizophrenia | Fight Sister |
| Sleeping | Suicide Sister | Anorexia | Fight Brother |
| Irritable | Suicide Brother | Bulimia | Argue Dad |
| Restless | Suicide Friend | OCD | Argue Mom |
| Alcohol | Suicide Thought | Bipolar | Argue Parents |
| Sertraline | Suicide Kill | PTSD | Impulsive |
| Zoloft | Suicide Think | Borderline Personality | Suicide Before |

FONTE: Fodeh et al (p. 942, 2019).

Doravante, 29% (7) utilizaram bases já rotuladas em outros trabalhos, ou que também foram rotuladas por CrowdSourcing. O Crowd sourcing (em português, contribuição colaborativa ou colaboração coletiva), dentro do contexto de rotulação, é um formato onde um grupo de pessoas colaboram para rotular conjuntos de dados sobre um determinado tema, de forma aberta a todos os participantes. Esse formato é presente dentro da rede social Reddit, onde 3 trabalhos utilizaram essa fonte de dados.

Infelizmente 3 (13%) trabalhos não apresentaram método de rotulação das bases antes de implementar os algoritmos.

Apenas 2 (8%) trabalhos contaram com auxílio de equipe especializada para supervisão ou rotulação, e 2 trabalhos (8%) utilizaram análise de sentimentos de textos para rotulação. Nesse caso, utilizam bibliotecas como o TextBlob, da linguagem de programação Python para classificar o sentimento do texto com o score de -1 até 1. Caso o score seja acima de 0, classifica-se como sentimento positivo, quando é abaixo de 0, é negativo. Desta forma, utilizam o

score negativo como rotulação “Depressivo” e positivo como “Não depressivo”(GOVINDASAMY; PALANICHAMY, 2021).

Por fim, 1 trabalho (4%) se propõe a apresentar um framework de classificação de saúde mental, não se aplicando a necessidade de rotulação, e 1 trabalho (4%) utiliza de questionário de saúde mental para identificar usuários com tendências depressivas.

De todos os trabalhos, apenas 3 (13%) explicitaram que solicitaram o consentimento dos usuários para coleta dos dados, e apenas 1 (4%) expõe que o trabalho teve sua aprovação por um comitê de ética.

4.2.3 Técnicas de Tratamento de dados e Classificação

4.2.3.1 Pré-processamento

Devido as redes sociais possuírem dados não estruturados, se faz necessário processo de limpeza de dados dentro das etapas de pré-processamento. Os processos mais recorrentes identificados e um breve resumo individual estão destacados abaixo:

- remoção de stop words (9 trabalhos utilizaram esta técnica): a remoção de stopwords consiste na aplicação de algoritmos que identificam palavras que são irrelevantes para o conjunto do resultado, como ‘as’, ‘e’, ‘o’, etc. Lembrando que para cada idioma as stopwords mudam e a relevância delas para o conjunto de dados depende do contexto da análise;
- remoção de caracteres especiais (8 trabalhos utilizaram esta técnica): a remoção de caracteres especiais consiste na aplicação de algoritmos que identificam e removem caracteres especiais do texto, como “ç” (“cedilha”), “” (acento) etc;
- remoção de emojis (2 trabalhos utilizaram esta técnica): a remoção de emojis consiste na aplicação de algoritmos que identificam e removem emojis do texto, que são abreviações da escrita que demonstram sentimentos pontuais, como “=)” (sorriso invertido), “=/” (rosto triste) etc;

- remoção de pontuação (7 trabalhos utilizaram esta técnica): a remoção de pontuação consiste na aplicação de algoritmos que identificam e removem a pontuação do texto, como “.” (ponto), “,” (vírgula) etc;
- tokenização (9 trabalhos utilizaram esta técnica): processo para segmentar o texto em frases ou palavras que são chamados de “tokens”;
- stemming (6 trabalhos utilizaram esta técnica): consiste na aplicação de técnicas que reduzem cada palavra à sua raiz, como por exemplo, a palavra “amigos” se aplicado a técnica de Stemming, seria transformada em “amig”, reduzindo a sua dimensionalidade;
- spell check (1 trabalho utilizou esta técnica): do inglês “verificador ortográfico”, consiste em utilizar técnicas para identificar e corrigir palavras escritas de maneira incorreta, exemplo: “traduaco” (seria corrigido para “tradução”);
- lemmatization (5 trabalhos utilizaram esta técnica): similar à técnica de stemming, a lematização reduz cada palavra em sua raiz, porém diferente do stemming, o resultado sempre será em uma palavra existente. Como por exemplo, se aplicarmos a técnica de lematização para a palavra “amigos”, o resultado será “amigo”;
- tradução (1 trabalho utilizou esta técnica): consiste em traduzir palavras em um texto para um idioma desejado;
- vetorização (8 trabalhos utilizaram esta técnica): consiste em representar um texto em um vetor de termos.

Para os processamentos envolvendo textos, a extração de *features* utilizando técnicas de frequências de palavras esteve presente em 13 trabalhos, sendo as técnicas mais citadas a TF-IDF e Bag of Words (BoW).

- TF-IDF: Essencialmente, o TF-IDF funciona por determinar a frequência relativa de palavras em um específico documento em comparação com a proporção inversa dessa palavra sobre todo o corpus do documento. Intuitivamente, este cálculo determina o quão

relevante uma determinada palavra é em um documento particular (RAMOS, 2003);

- BoW: Neste modelo, um texto é representado como uma coleção não ordenada de suas palavras, desconsiderando a gramática e até mesmo a ordem das palavras. No caso de classificação de texto, uma palavra em um documento é atribuída a um peso de acordo com sua frequência no documento e frequência entre diferentes documentos. Palavras junto com seus pesos formam uma “bolsa de palavras” (Bag Of Words) (GEORGE; JOSEPH, 2014).

Classificação de sentimentos em textos também esteve presente em etapas de pré-processamento, conforme citado anteriormente, em dois trabalhos. A análise de sentimento ou mineração de opinião é o tratamento computacional de opiniões, sentimentos e subjetividade de texto. A análise de sentimento é um processamento de linguagem natural e tarefa de extração de informação que visa obter os sentimentos do escritor expressa em comentários positivos ou negativos, perguntas e pedidos, analisando um grande número de documentos (KHAIRNAR; KINIKAR, 2013).

4.2.3.2 Classificação

Em todos os trabalhos houve a preocupação dos pesquisadores em ter parâmetros de referência para os resultados dos classificadores, desta forma, utilizaram mais de um algoritmo por trabalho para avaliar os resultados. Os classificadores utilizados bem como a frequência de utilização, estão apresentados no Gráfico 6.

GRÁFICO 6 – TÉCNICAS DE CLASSIFICAÇÃO UTILIZADAS NOS TRABALHOS



FONTE: O Autor (2021).

4.2.4 Aferição de Resultados

A aferição de resultados dos trabalhos concentra-se na eficiência dos classificadores com base nas suas classificações, sendo as métricas de aferição dos classificadores apresentadas como performance dos trabalhos. A composição das métricas utiliza os conceitos:

- verdadeiros negativos (VN): classificação foi falsa e o valor real era falso, sendo uma classificação correta;
- falso positivo (FP): classificação foi verdadeira, porém o valor real era falso, sendo uma classificação incorreta;
- falso negativo (FN): classificação foi falsa, porém o valor real era verdadeiro, sendo uma classificação incorreta;

- verdadeiro positivo (VP): classificação foi verdadeira e o valor real era verdadeiro, sendo uma classificação correta.

As métricas finais acompanhadas nos artigos foram, em sua maioria, as dispostas abaixo:

- acurácia (*accuracy*): número de classificações corretas / número total de classificações;
- precisão (*precision*): Total de verdadeiros positivos / total de verdadeiros positivos + falsos positivos;
- *recall*: verdadeiros positivos / verdadeiros positivos + falsos negativos;
- F-measure: Também conhecido como F-score, é calculado através da fórmula: $2 \times \text{recall} \times \text{precision} / (\text{recall} + \text{precision})$.

Comparando os três classificadores com maior frequência de utilização nos estudos selecionados em conjunto com as quatro métricas citadas acima, normalizados em uma escala de 0 a 100 e desconsiderando as diferenças conceituais, foi identificado que, em média, o algoritmo Naïve Bayes teve acurácia superior de 6 p.p em comparação com o algoritmo Random Forest, porém todos possuem resultados, em média, de $81 \pm 3,3$.

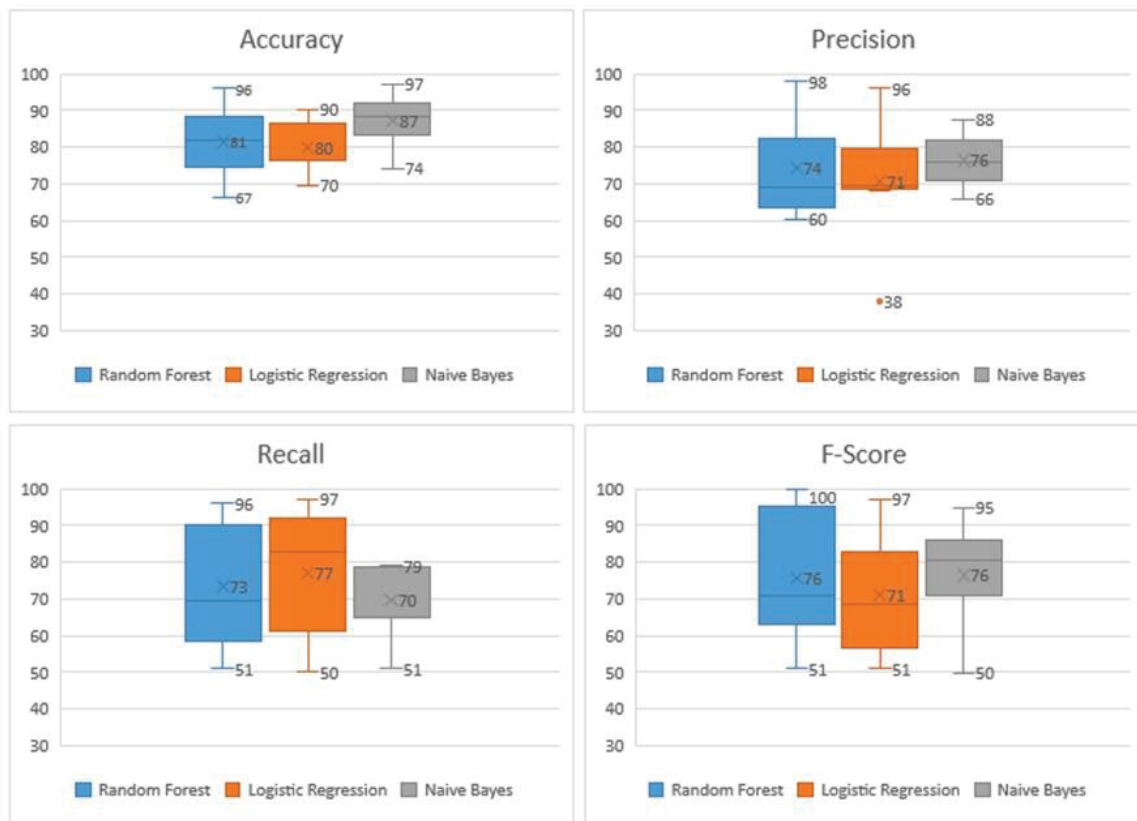
Em termos de precisão, os algoritmos possuem maior variância de resultados, sendo que a regressão logística possui um resultado baixo de 38 pontos, sendo sua média 71. O Naïve Bayes também supera o Random Forest em 2 p.p em média, mesmo o Random Forest possuindo maior máximo (98 pontos).

A métrica Recall destaca positivamente o algoritmo de Regressão Logística, que, em média, demonstrou resultados de 77 pontos, 4 pontos acima da segunda média (Random Forest).

Por fim, a métrica F-score tem a mesma média do algoritmo Random Forest e Naïve Bayes, sendo 76 p.p.

Os resultados estão disponíveis nos gráficos da **Erro! Fonte de referência não encontrada..**

FIGURA 9 – GRÁFICOS COMPARATIVOS DOS TRÊS CLASSIFICADORES MAIS UTILIZADOS NOS ESTUDOS COM RESULTADOS DAS MÉTRICAS DE AVALIAÇÃO DE CLASSIFICAÇÃO.



FONTE: O autor, 2021.

4.2.5 DISCUSSÃO

Observa-se aumento na pesquisa e exploração de algoritmos para identificação de transtornos da saúde mental, principalmente de Depressão e Suicídio. As pesquisas possuem maior ênfase na aplicação da tecnologia em si (arquitetura de machine learning e deep learning). Por ser uma área experimental

com diversas possibilidades de obtenção dos resultados desejados, não existe padronização de metodologia e, em alguns casos, não são apresentados a completude da metodologia que possibilite a reprodutividade dos trabalhos.

Duas pesquisas discutem nos resultados do seu trabalho, quais variáveis tiveram maior relevância nos modelos, além de associá-las à comportamentos dos grupos rotulados e não rotulados. Porém, na maioria dos trabalhos, não existe a preocupação de entender quais variáveis possuem maior correlação com os transtornos de saúde mental analisados, diminuindo potencial de extração de insumos para profissionais da área.

A maioria dos estudos apenas foca em análises textuais simples, não utilizando todas possíveis variáveis que podem ser extraídas das redes sociais, como variáveis de imagens, comportamento e interações entre usuários e entre usuários com a plataforma.

Existe preocupação em relação à validade da rotulação dos trabalhos cujo processo de rotulação não possui metodologia alinhada com profissionais da área, ou que depende exclusivamente de análise de frequência de palavras que não são derivadas de áreas da saúde.

4.2.6 ANÁLISES

Esta revisão sistemática teve como objetivo identificar e analisar os métodos e técnicas existentes para identificação de transtornos mentais por meio das redes sociais, com ênfase em depressão e suicídio, utilizando técnicas de mineração de dados. Os estudos, em sua maioria, utilizam de rotulação de dados sem apoio de áreas especializadas, com maior foco em testar o resultado dos algoritmos de classificação em determinados conjuntos de dados, sem contextualização e planejamento de inserção dos trabalhos em situações que possam ser utilizados por profissionais da área. Os pontos destaque da etapa são:

- as redes sociais mais utilizadas para identificação de transtornos da saúde mental são, respectivamente, o Twitter, Reddit, Facebook, Instagram, Pantip, Weibo e Vkontakte.

- as variáveis extraídas de maior utilização para identificação de transtornos da saúde mental foram de textos, emojis e imagens.
- a rotulação dos datasets para classificação utilizaram, em sua maioria, frequência de palavras relacionadas à depressão ou demais transtornos, seguido de estudos com bases já rotuladas previamente, estudos com rotulação e auxílio de profissionais, estudos apenas com análise de sentimentos e através de questionários. Também há estudos que não informaram a origem da rotulação.
- as principais técnicas utilizadas para tarefas de identificação e classificação foram Random Forest, Regressão Logística, Naïve Bayes, SVM (Support Vector Machines). Em muitos estudos também se combinaram técnicas para análises textuais, sendo os principais o TF_IDF e Bag of Words.

Os resultados obtidos nesta etapa são utilizados para a condução do experimento detalhado na próxima seção, com a adoção de uma das redes sociais mais pesquisadas (Reddit), variáveis, rotulação e técnicas.

5 RESULTADO: ANÁLISE DE DADOS NA BASE REDDIT

O conjunto de dados (dataset) extraído dos dados do Subreddit (de 01/01/2021 até 30/06/2022), ainda sem tratamento e adição das demais *features*, possui 23.438 linhas, contendo as seguintes colunas:

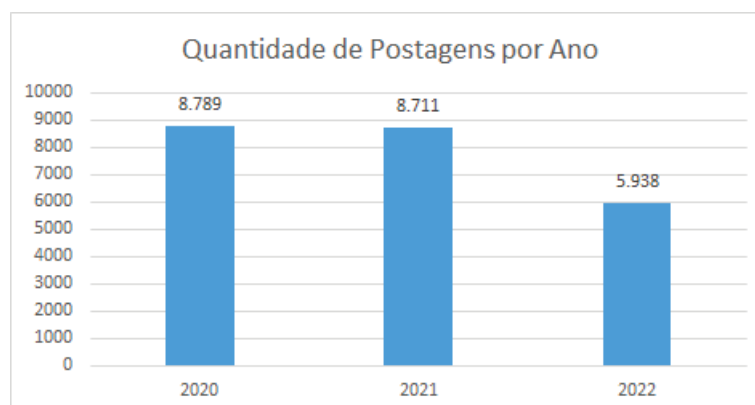
QUADRO 3 - COLUNAS DO DATASET

| Coluna | Descrição |
|-----------------|---|
| Author | Nome do Autor do Reddit |
| Title | Título do post |
| url | url para a postagem dentro do Reddit |
| Selftext | Conteúdo do corpo textual do post |
| Link_flair_text | Classificação do conteúdo presente dentro do post |
| Created_date | Data de criação do post |

FONTE: O Autor (2022).

A quantidade de postagens por ano se mantém próxima nos anos de 2020 e 2021, sendo 8.789 e 8.711, respectivamente. Em 2022, houve extração de 5.938 postagens no período, conforme GRÁFICO 7.

GRÁFICO 7 – QUANTIDADE DE POSTAGEM POR ANO NA COMUNIDADE DESABAFOS REDDIT



FONTE: O Autor (2022).

O campo com o nome "Link_flair_text" representa a rotulação/classificação dos textos, realizados de forma manual pela própria comunidade do Subreddit. Conforme citado anteriormente, as classificações são realizadas de forma empírica, sem utilização de método específico. Porém cada

rotulação possui seu significado que representa de forma macro o texto classificado.

O quadro 4 contém a explicação de cada classe disponível no conjunto de dados, seguido de um exemplo.

QUADRO 4 – EXPLICAÇÃO DE CADA CLASSIFICAÇÃO

| Classificação | Explicação | Exemplo |
|-------------------|--|---|
| Desabafo | A Classificação Desabafo se encontra em um relato de sentimentos e pensamentos íntimos, não tendo padrão se é relacionado a uma situação específica ou uma reflexão ampla. | como o clima de final de ano me atormenta. não tem explicação na minha cabeça, isso vem desde criança, eu fico muito melancólico. [...] não sei se isso se deve pelo fato de eu também odiar clima de despedida, e também não saber lidar com isso desde pequeno, e essa sensação de algo estar indo me deixe assim. [...] |
| Não classificados | São postagens que não foram classificadas pela comunidade. | em 3 meses, li 6 livros!! é isso aí mesmo. nada muito relevante! [...] |
| Dia a Dia | Relatos que refletem o cotidiano. | a vida passa muito rápido. foi essa a conclusão que tive às 02:49, bebendo uma cerveja, ouvindo jazz e apreciando o céu noturno enquanto a cidade está em completo silêncio. |
| Relacionamento | Relatos que envolvem relacionamentos interpessoais, com maior tendência para relacionamentos amorosos | sinto falta da paixão tenho uma vida romântica com uma frequência relativamente normal. não sou de ficar trocando de par romântico toda hora, e nem de ficar muito tempo na seca. de vez em quando conheço alguém novo, saio com essa pessoa por um tempo, desfruto da companhia e |

| | | |
|---------------------------|---|--|
| | | da eventual intimidade, mas é sempre meio morno emocionalmente, e parece que fica estagnado depois de pouco tempo. faz muito tempo que eu não desejo alguém.[...] |
| Boas notícias! | Relatos para expressar boas notícias | desabafar aqui no reddit tem me ajudado muito! estou começando a me sentir melhor e menos ansiosa, ontem e hoje sai de casa pra fazer coisas comuns do cotidiano como comprar pão e ir pra academia [...] |
| Desabafos - sem conselhos | Desabafos que não possuem a opção de comentários e interações pela comunidade, apenas n62ntuitoto de desabafar. | valorizem a sua vida a vida pertence a cada um de vocês. valorizem antes que seja tarde. [...] |
| Depressão | Relatos que envolvem temas de depressão em específico | minha depressão ta voltando no meio de 2019 eu comecei a ter minhas primeiras crises de pânico. eu tava no cursinho na época e provavelmente era por causa de pressão de la. mas enfim, fiz tratamento com antidepressivos e passei na facul tudo parecia ta ótimo [...] |
| Más notícias | Relatos para comentar sobre notícias negativas | acho que to com covid fui no médico e ele falou para eu fazer um teste dia 15, eu to com dor de garganta, febre e cansaço. to com medo que aconteça alguma coisa comigo e deixar os meus pais sozinhos. |
| NSFW | Relatos com comentários de conteúdos adultos. NSFW " "No Suitable For Wo"k" | é errado julgar fetiche? [...] |

| | | |
|--------|--|--|
| Update | Relatos que são atualizações de temas anteriores | há algumas semanas postei aqui sobre a situação do meu ex[...] |
|--------|--|--|

FONTE: O Autor (2022).

Por serem classificações empíricas, conforme citado nas limitações da pesquisa, as classificações dependem dos vieses e percepções individuais de quem está classificando, inclusive existe certa sobreposição conceitual entre as classificações das postagens.

Contando a quantidade de postagens por classificação das mesmas, constrói-se o quadro abaixo:

QUADRO 5 – QUANTIDADE DE POSTAGENS DE ACORDO COM SUA CLASSIFICAÇÃO

| Classificação | Quantidade | Percentual |
|-------------------------|-------------------|-------------------|
| Desabafo | 7926 | 39,62% |
| Não classificados | 5062 | 25,30% |
| Dia a Dia | 2105 | 10,52% |
| Relacionamento | 1604 | 8,02% |
| Desabafo -sem conselhos | 836 | 4,18% |
| Boas notícias! | 817 | 4,08% |
| Depressão | 689 | 3,44% |
| Más notícias | 599 | 2,99% |
| NSFW | 193 | 0,96% |
| Update | 175 | 0,87% |

FONTE: O Autor (2022).

A maior classificação é sobre “desabafo”, além de ser um termo de classificação genérico, traz o nome da comunidade que é objeto de análise deste trabalho. Quanto mais específico a classificação se torna, menor é o percentual de representatividade dentro da Base. Depressão, que é o foco do trabalho, representa 3,44% das classificações. Não foi escopo do trabalho neste momento lidar com possíveis desbalanceamentos entre as classes descritas acima.

5.1 PRÉ-PROCESSAMENTO E EXTRAÇÃO DE FEATURES

Como característica da rede social Reddit, o título da publicação muitas vezes faz parte da postagem em si, por exemplo:

QUADRO 6 - EXEMPLO DE UMA POSTAGEM DO REDDIT

| Título | Corpo do Texto |
|----------------------------|---|
| A vida passa muito rápido. | Foi essa a conclusão que tive as 02:49, bebendo uma cerveja, ouvindo jazz e apreciando o céu noturno enquanto a cidade está em completo silêncio. |

FONTE: O Autor (2022).

Devido a essa característica, o texto que será analisado é composto pela concatenação do título com o corpo do text. Utilizado a função do MySQL para realizar essa tarefa (função “*concat*”).

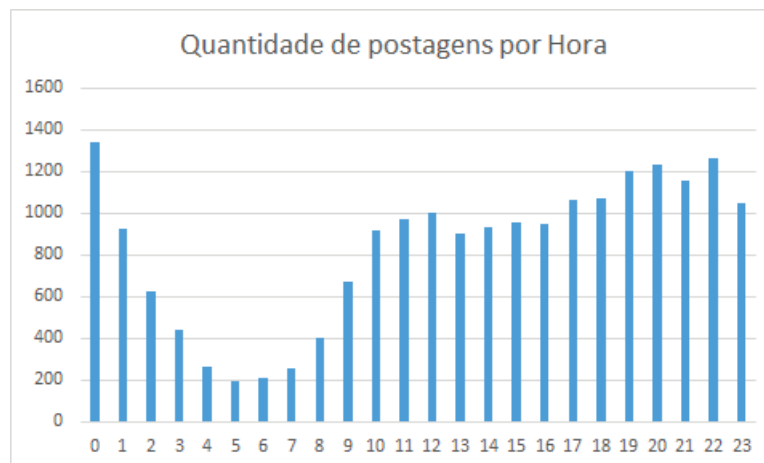
Na sequência, será utilizada a função “*length*” para identificar a quantidade de caracteres presentes dentro do texto. Postagens extremamente curtas (menos de 15 caracteres) foram removidas (3.432 postagens), após o filtro restou 20.006 postagens.

Utilizando o campo “*created_date*”, que refere-se a data de criação da postagem, extraiu-se o período de postagem por ser um campo “*datetime*” (campo que possui além da data, horário da postagem).

Observando as postagens por hora do dia, evidencia-se um pico de postagens a meia-noite (horário zero), porém com tendência de queda até as 6 horas, conforme GRÁFICO 8.

Na sequência verifica-se uma tendência de crescimento que se estabiliza entre 12 horas e 18 horas. Entre as 19 e 23 horas observa-se outro pico de postagens com certa estabilidade, sendo o período que maior se destaca na quantidade de postagens.

GRÁFICO 8 - QUANTIDADE DE POSTAGENS NA COMUNIDADE DESABAFOS DO REDDIT



FONTE: O Autor (2022).

Para preparar o conjunto de dados para as próximas etapas, o campo de hora da postagem será categorizado. A categorização do turno foi realizada utilizando o seguinte pseudo-código em SQL:

CASO

QUANDO horário da postagem ENTÃO “manhã”

QUANDO horário da postagem ENTRE 13 E 18 ENTÃO “tarde”

QUANDO horário da postagem ENTRE 19 E 24 ENTÃO “noite”

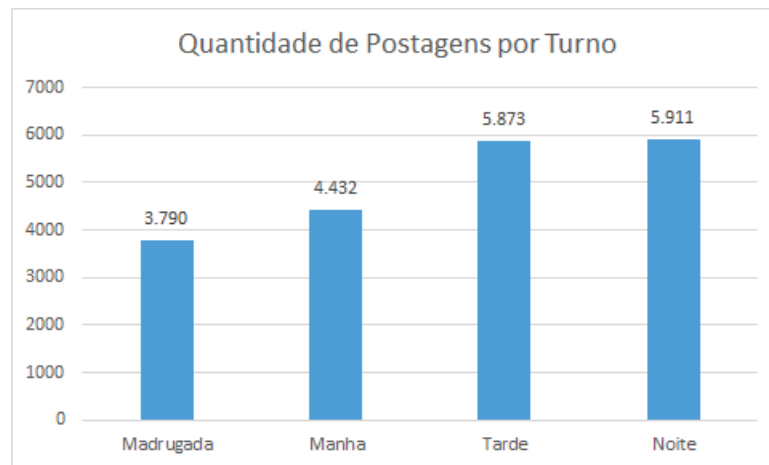
SENÃO “madrugada”

Através dessa classificação dos horários, gera-se a classificação de quatro dimensões:

- 1) Postagens no período da manhã (entre 6h a 12h e 59 min)
- 2) Postagens no período da tarde (entre 13h a 18h e 59 min)
- 3) Postagens no período da noite (entre 19h a 24h e 59 min)
- 4) Postagens no período da madrugada (entre 01h a 05h e 59 min)

A maior quantidade de postagens foi realizada no período da noite (5.911 postagens), porém no período da tarde foram realizadas 5.873, sendo uma pequena diferença do período da noite. Por fim, tiveram 5.532 postagens no período da manhã e 3.790 postagens no período da madrugada.

GRÁFICO 9 - QUANTIDADE DE POSTAGENS POR TURNO



FONTE: O Autor (2022).

5.2 RESULTADOS CLASSIFICAÇÃO DE TEXTOS

Neste tópico serão abordados os resultados das classificações de texto, contendo as suas análises e por fim as implementações dos modelos de classificação considerando a classificação de depressão.

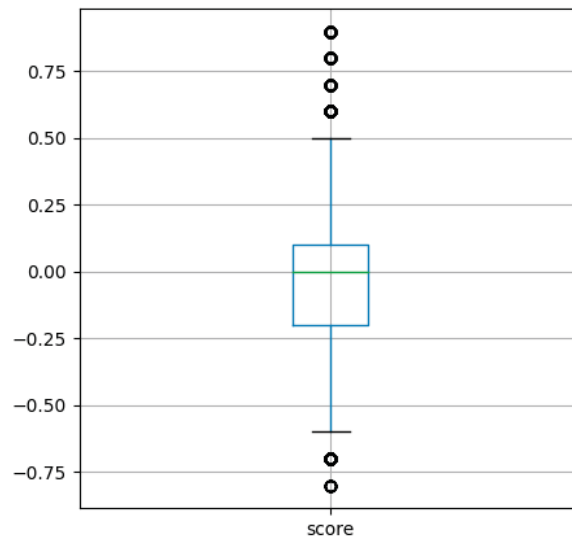
5.2.1 Análise de sentimentos

Através do fluxo de classificação pela API de Sentimentos do Google Cloud, os textos presentes do conjunto de dados foram classificados pelo teor sentimental presentes em cada texto. Como retorno da API, além do sentimento do documento, também houve o retorno da quantidade de sentenças (frases) presentes em cada documento e o sentimento presente em cada sentença. Para extrair o máximo de *features* úteis além do sentimento geral, considerado o total de sentenças e criado a métrica de variância de sentimentos, ambos serão abordados abaixo.

5.2.2 Análise de sentimento geral

Identificado que a média da classificação de sentimentos para todo o conjunto de dados é $-0,05 \pm 0,26$, ou seja, grande parte das postagens possui característica neutra com leve tendência para sentimentos negativos.

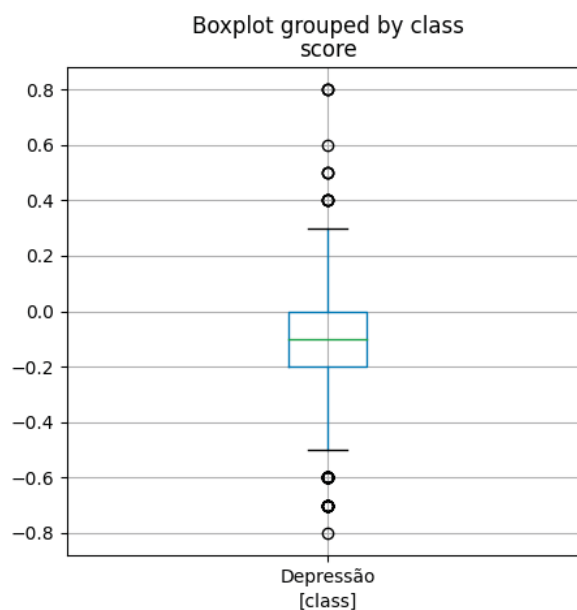
GRÁFICO 10 – VARIACÃO DO ÍNDICE DE SENTIMENTO PRESENTE NOS TEXTOS



FONTE: O Autor (2022).

Quando a classe “depressão” é isolada, vê-se que a média fica em $-0,13 \pm 0,23$, demonstrando maior carga negativa com menor desvio.

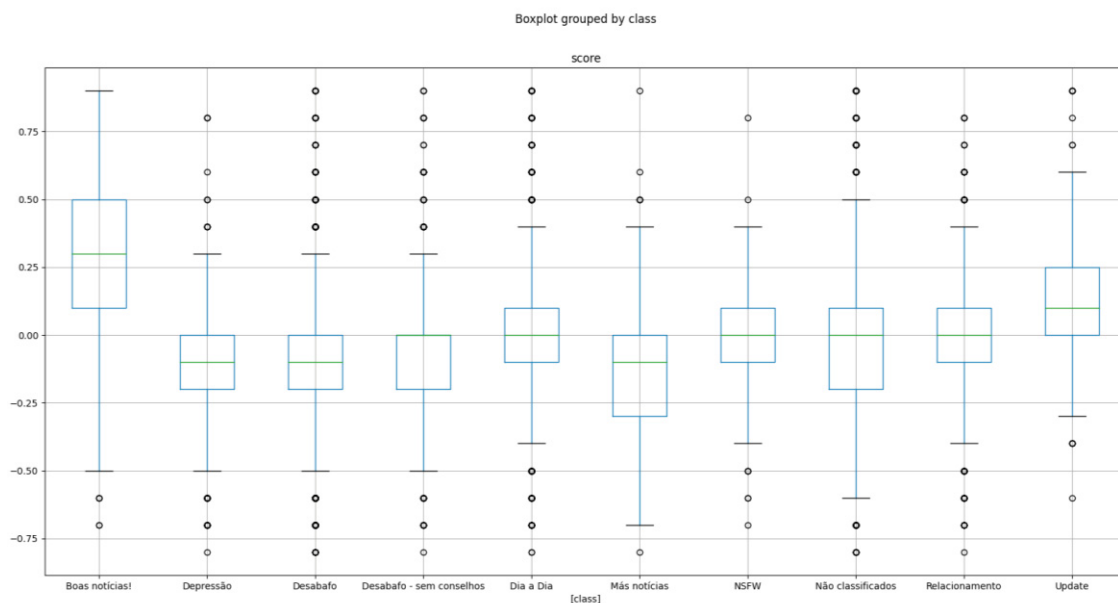
GRÁFICO 11 – VARIACÃO DOS ÍNDICES DE SENTIMENTOS PRESENTE NA CLASSE DEPRESSÃO



FONTE: O Autor (2022).

Quando observado o Gráfico 11, a quebra da variação de sentimentos pela classificação dos textos da comunidade do Reddit, é possível entender algumas características de separação de cada classe pelo teor sentimental dos textos.

GRÁFICO 12 - VARIAÇÃO DOS ÍNDICES DE SENTIMENTOS POR CLASSE



FONTE: O Autor (2022).

A classe “boas notícias!” evidentemente possui um teor sentimental mais positivo que demais classes (média de $0,30 \pm 0,10$). As classes “depressão”, “desabafo” e “desabafo – sem conselhos” possuem distribuição de sentimentos com alta similaridade.

As classes “dia a dia”, “NSFW”, “não classificado” e “relacionamento” possui tendência de neutralidade de sentimentos, sendo o “não classificado” com maior amplitude de sentimentos. Isso se deve possivelmente por essa classe representar vários assuntos sem classificações prévias. A classe “Update”, que representa a atualização de postagens, possui média positiva (0,15). Destaca-se que todas as classificações possuem exceções (pontos de outliers).

5.2.2 Total de sentenças

As sentenças representam frases presentes dentro do corpo do texto. Um texto pode ter uma ou mais sentenças presentes.

Conforme exemplo:

“sinto falta da paixão tenho uma vida romântica com uma frequência relativamente normal. não sou de ficar trocando de par romântico toda hora, e nem de ficar muito tempo na seca. de vez em quando conheço alguém novo, saio com essa pessoa por um tempo, desfruto da companhia e da eventual intimidade, mas é sempre meio morno emocionalmente, e parece que fica estagnado depois de pouco tempo. faz muito tempo que eu não desejo alguém.eu quero me apaixonar por alguém antes mesmo de poder ter um encontro romântico com essa pessoa. quero conhecer alguém de forma orgânica, alguém que me faça sonhar acordado de novo. não quero mais investir tempo, dinheiro e esforço pra ver se eventualmente algum sentimento aflora. é injusto comigo e também com a pessoa que estiver junto, que merece alguém que goste dela.”

O identificador de sentenças identifica cada frase e a separa, com o seguinte output:

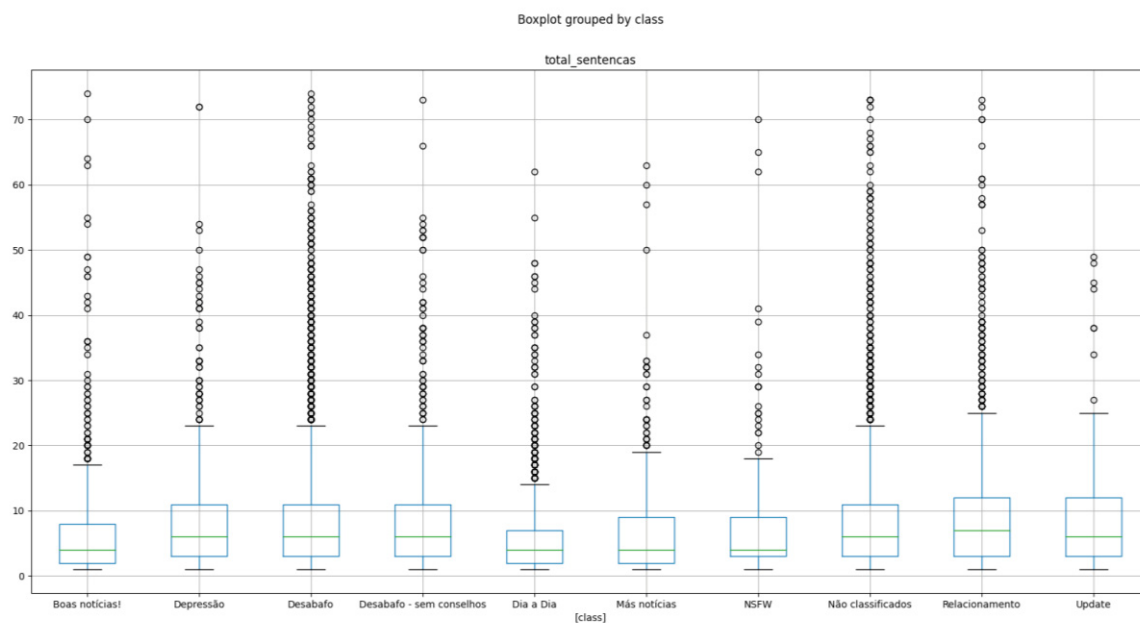
QUADRO 7 – EXEMPLO DE SENTENÇAS IDENTIFICADAS PELA API DE SENTIMENTOS

| Número | Sentença |
|--------|---|
| 1 | Sinto falta da paixão Tenho uma vida romântica com uma frequência relativamente normal. |
| 2 | Não sou de ficar trocando de par romântico toda hora, e nem de ficar muito tempo na seca. |
| 3 | De vez em quando conheço alguém novo, saio com essa pessoa por um tempo, desfruto da companhia e da eventual intimidade, mas é sempre mei“ "mor"o" emocionalmente, e parece que fica estagnado depois de pouco tempo. |
| 4 | Faz muito tempo que eu não desejo alguém. |
| 5 | Eu quero me apaixonar por alguém antes mesmo de poder ter um encontro romântico com essa pessoa. |
| 6 | Quero conhecer alguém de forma orgânica, alguém que me faça sonhar acordado de novo. |
| 7 | Não quero mais investir tempo, dinheiro e esforço pra ver se eventualmente algum sentimento aflora. |
| 8 | É injusto comigo e também com a pessoa que estiver junto, que merece alguém que goste dela. |

FONTE: O Autor (2022).

Utilizando a técnica citada, identificado que a Média de Sentenças presente no conjunto de dados é de 8,6 sentenças \pm 10,36.

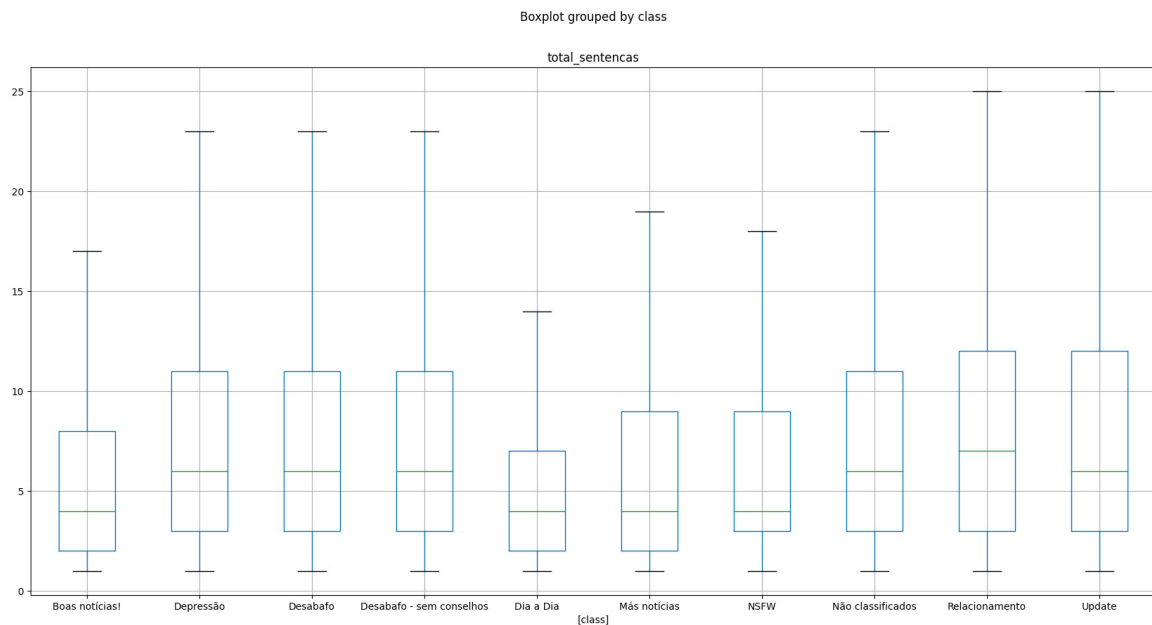
GRÁFICO 13 – VARIAÇÃO DA QUANTIDADE DE SENTENÇAS POR CLASSE



FONTE: O Autor (2022).

Ao gerar a visualização com a quebra por classificação, presente no gráfico 13, entende-se alta variação da quantidade de sentenças em todas as classes, evidenciado pela expressiva quantidade de outliers.

GRÁFICO 14 - VARIAÇÃO DA QUANTIDADE DE SENTENÇAS POR CLASSE COM REMOÇÃO DE OUTLIERS



FONTE: O Autor (2022).

Analisando a quantidade de sentenças por classificação, utilizando o gráfico 14 que possui supressão dos outliers, observa-se que a classe “Relacionamento” possui a maior média de sentenças, e também a maior amplitude. A classe “update” também tem amplitude similar ao de “Relacionamento”, porém com média inferior. As classes “depressão”, “desabafo”, “desabafo – sem conselhos” e “não classificados” possui tanto a média de sentenças similar quanto a amplitude.

As postagens com menor tendência de quantidade de sentenças são pertencentes as classes “boas notícias!”, “dia a dia”, “más notícias” e “NSFW”, sendo o menor a classe “dia a dia”.

5.2.3 Variância de Sentimentos

A avaliação de sentimento através da API do Google Cloud considera a média dos sentimentos das sentenças presentes dentro do texto, em conjunto com o peso de cada sentimento, para avaliar o sentimento geral do texto. Para complementar essa visão, foi elaborado uma nova métrica denominada “variância de sentimentos” que consiste em calcular a variância dos sentimentos das sentenças presentes no texto, conforme exemplo abaixo:

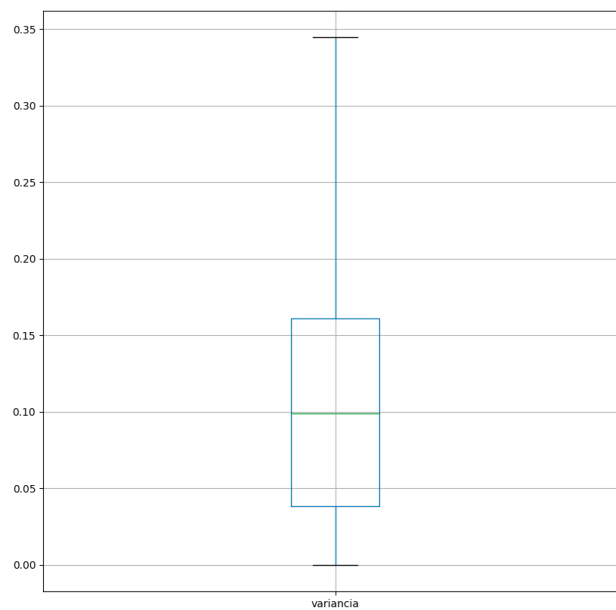
QUADRO 8 – EXEMPLO DE IDENTIFICAÇÃO DE SENTIMENTOS POR SENTENÇA

| Número | Sentença | Sentimentos |
|--------|---|-------------|
| 1 | Sinto falta da paixão Tenho uma vida romântica com uma frequência relativamente normal. | 0,00 |
| 2 | Não sou de ficar trocando de par romântico toda hora, e nem de ficar muito tempo na seca. | 0,10 |
| 3 | De vez em quando conheço alguém novo, saio com essa pessoa por um tempo, desfruto da companhia e da eventual intimidade, mas é sempre mei“ "mor"o" emocionalmente, e parece que fica estagnado depois de pouco tempo. | -0,20 |
| 4 | Faz muito tempo que eu não desejo alguém. | 0,20 |
| 5 | Eu quero me apaixonar por alguém antes mesmo de poder ter um encontro romântico com essa pessoa. | 0,10 |
| 6 | Quero conhecer alguém de forma orgânica, alguém que me faça sonhar acordado de novo. | 0,40 |
| 7 | Não quero mais investir tempo, dinheiro e esforço pra ver se eventualmente algum sentimento aflora. | - 0,50 |
| 8 | É injusto comigo e também com a pessoa que estiver junto, que merece alguém que goste dela. | - 0,50 |

FONTE: O Autor (2022).

Para o documento acima, o score geral de sentimento foi 0,00 (neutro), porém ao se calcular a variância, tem-se o valor de 0,09, ou seja, mesmo o documento tendo o sentimento neutro, ele possui variação de sentimentos com baixa variação total (considerando a média de variância de 0,10).

GRÁFICO 15 – ÍNDICE DE VARIAÇÃO DA MÉTRICA VARIÂNCIA DE SENTIMENTOS



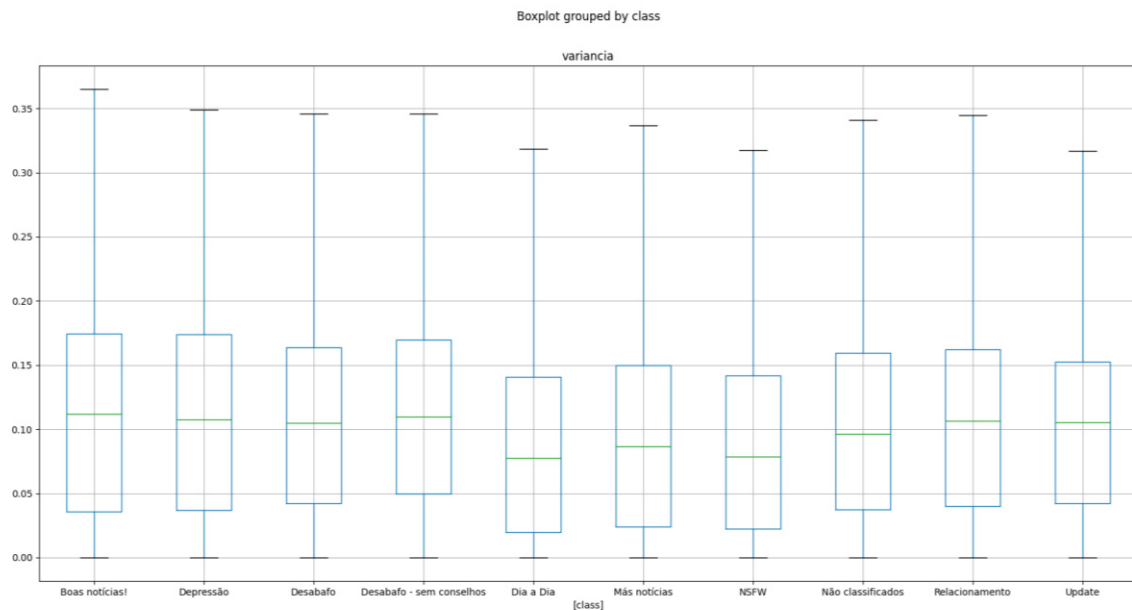
FONTE: O Autor (2022).

A variância média de sentimentos das sentenças do conjunto de dados analisado é $0,10 \pm 0,08$, o mínimo é 0 e o máximo de variância é 0,64.

Agrupando a variância de sentimentos por classe, entende-se que mesmo com similaridade entre as médias entre as classes, todas são distintas.

A Classe “depressão” possui média de variância de sentimentos de $0,11 \pm 0,09$, sendo um pouco acima da média geral, sendo a segunda menor média, ficando abaixo apenas da classe “boas notícias!”.

GRÁFICO 16 - ÍNDICE DE VARIAÇÃO DA MÉTRICA VARIÂNCIA DE SENTIMENTOS POR CLASSE



FONTE: O Autor (2022).

De acordo com o gráfico 16, as classes “dia a dia”, “más notícias” e “NSFW” possuem as menores médias de variância de sentimentos, isso se deve por serem as classes que também possuem menores quantidades de sentenças, ou seja, possuem menor espaço para variação sentimental nos textos.

5.2.3 Extração de entidades

A extração de entidades disponíveis através da API da Google Cloud consiste em identificar as entidades presentes dentro de um texto em específico.

É identificado todas as entidades presentes e classificados nas seguintes dimensões, conforme detalhado no Quadro 9.

QUADRO 9 – ENTIDADES DISPONÍVEIS NA API GOOGLE CLOUD

| ENTIDADE | DESCRIÇÃO | EXEMPLOS DO CONJUNTO DE DADOS |
|---------------|--|--|
| OTHER | Outros tipos de entidades. Consegue identificar que é uma entidade, porém não consegue classificá-la | aspecto, vontade, vida, falta |
| PERSON | Identificação de uma pessoa no texto, pode ser através do nome ou de títulos | psicóloga, marido, sogros, Mark, Henrique |
| NUMBER | Números | um, dois, 1, 2 |
| EVENT | Identificação de um evento | aniversário, casamento, churrasco, festa |
| LOCATION | Identificação de um local ou uma localização | praia, casa, igreja, academia |
| CONSUMER_GOOD | Identificação de um Produto de consumo | comida, carro, brinquedo, móveis |
| ORGANIZATION | Identificação de uma organização | whatsapp, instagram, reddit, faculdade |
| WORK_OF_ART | Identificação de uma Obra de arte | música, tatuagem, fotos, poesia |
| DATE | Identificação de uma Data | 2022, Dezembro, 30 de março |
| PRICE | Identificação de Preços | 100 euros, R\$ 20 |
| ADDRESS | Identificação de Endereços | Rua Universal Feliz, 31 (exemplo fictício) |
| PHONE_NUMBER | Identificação de números de telefone | - |

FONTE: O Autor (2022).

Nesta etapa, foi identificado e extraído o total de 774.941 entidades presentes nos textos do conjunto de dados analisado, conforme detalhado no Quadro 10.

A classificação “other” acaba sendo atribuída quando é reconhecido como uma entidade, mas não tem classificação dentro do modelo pré-treinado. a entidade “number” refere-se à números presentes dentro do texto. tanto o “other” quanto o “number” serão removidos das próximas análise (e para o modelo) por não terem relevância dentro da proposta da pesquisa.

QUADRO 10 – TOTAL DE REGISTROS E PERCENTUAL POR ENTIDADE

| ENTIDADE | REGISTROS | % |
|---------------|-----------|--------|
| OTHER | 422860 | 54,57% |
| PERSON | 120514 | 15,55% |
| NUMBER | 116277 | 15,00% |
| EVENT | 42990 | 5,55% |
| LOCATION | 27555 | 3,56% |
| CONSUMER_GOOD | 19696 | 2,54% |
| ORGANIZATION | 13826 | 1,78% |
| WORK_OF_ART | 8892 | 1,15% |
| DATE | 1660 | 0,21% |
| PRICE | 646 | 0,08% |
| ADDRESS | 17 | 0,00% |
| PHONE_NUMBER | 8 | 0,00% |

FONTE: O Autor (2022).

A presença da entidade “person” (pessoa) é bem alta dentro dos textos do conjunto de dados, representando 51,11% de todas as entidades identificadas. Isso implica em que independente da classe, grande parte dos relatos presentes na comunidade envolvem pessoas terceiras do autor.

QUADRO 11 – TOTAL DE REGISTROS E PERCENTUAL POR ENTIDADE APÓS TRATAMENTO

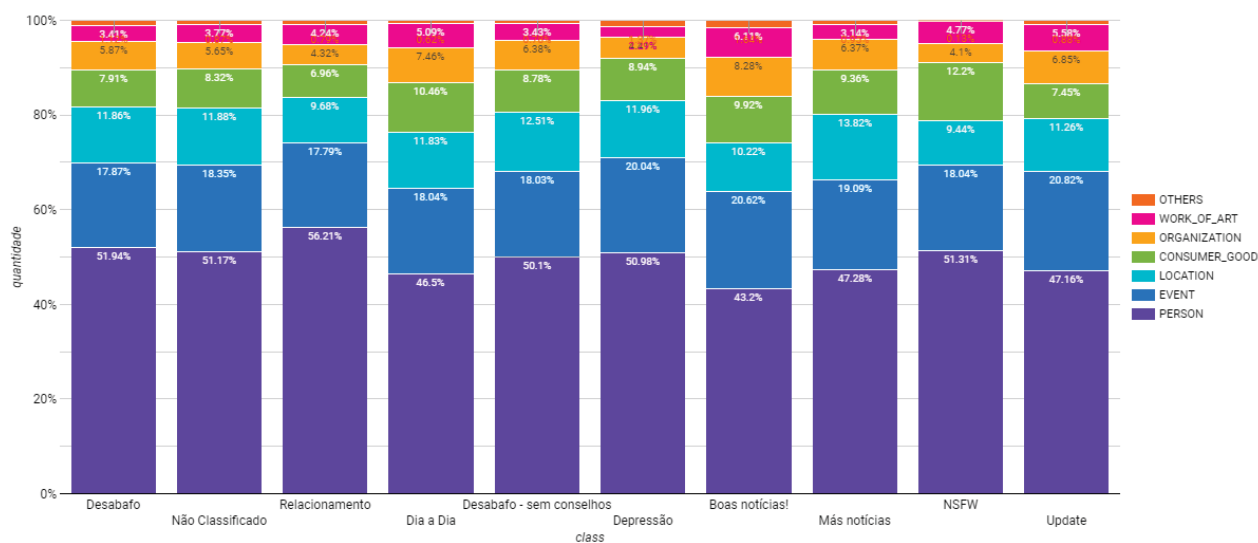
| ENTIDADE | REGISTROS | % |
|---------------|-----------|--------|
| PERSON | 120514 | 51,11% |
| EVENT | 42990 | 18,23% |
| LOCATION | 27555 | 11,69% |
| CONSUMER_GOOD | 19696 | 8,35% |
| ORGANIZATION | 13826 | 5,86% |
| WORK_OF_ART | 8892 | 3,77% |
| DATE | 1660 | 0,70% |
| PRICE | 646 | 0,27% |

| | | |
|--------------|----|-------|
| ADDRESS | 17 | 0,01% |
| PHONE_NUMBER | 8 | 0,00% |

FONTE: O Autor (2022).

A segunda maior entidade é “event” (eventos) seguida de “location” (localização), que são entidades que podem ser relacionadas dentro do relato. As entidades “consumer_good” (produto de consumo), “organization” (organização), “work_of_art” (obra de arte) somam 17,98%, e as entidades com menor quantidade presente no conjunto de dados são “date” (data), “price” (preço), “address”(endereço) e “phone_number” (número de telefone) somam menos de 1%. Agrupando a quantidade de entidades identificadas pelas classes presentes do conjunto de dados, gerado o gráfico abaixo.

GRÁFICO 17 – TOTAL DE ENTIDADES IDENTIFICADAS POR CLASSE



FONTE: O Autor (2022).

A classe “relacionamento” é a que apresenta maior presença da entidade “pessoa”, o que faz sentido por ser a classe que representa relatos de relações interpessoais. As classes “depressão”, “boas notícias!” e “update” apresentam presença superior à 20% da entidade “evento”.

A entidade “localização” tem maior presença em “más notícias” e “desabafo – sem conselhos”. Já a entidade “produtos de consumo” tem destaque evidente dentro da classe “NSFW”.

De forma geral, a quantidade de cada entidade presente em cada classe não se distância muito da média, apenas nos casos citados anteriormente.

5.2.4 Processamento de texto

O processamento de texto está em alinhamento ao descrito na seção de encaminhamentos metodológicos e prevê a aplicação de cada uma das próximas etapas em cada texto:

- remoção de pontuação;
- remoção de *stopwords*;
- substituição da acentuação das palavras;
- normalização das palavras em letras minúsculas.

Para a classificação de textos também foi aplicado o método de lematização. A verificação das palavras mais frequentes foi realizada antes da aplicação do lematização, pois não é necessária a redução de dimensionalidade nessa etapa, além de podermos visualizar as palavras da forma que foram escritas.

Comparando as palavras mais frequentes de cada classe com a classe Depressão, observado que existe alta similaridade nas palavras mais frequentes, com média de 79,2% \pm 4,3. As classes “desabafo”, “desabafo – sem conselhos” e “NSFW” possuem mais de 80% de similaridade nas 50 palavras mais frequentes, sendo o eixo “desabafo” com 88% de similaridade.

QUADRO 12 – COMPARATIVO DE SIMILARIDADE DAS PALAVRAS COM MAIOR FREQUÊNCIA POR CLASSE

| Classe | X Depressão | % Comparativo |
|--------------------------|-------------|---------------|
| Desabafo | 44 | 88% |
| Desabafo - sem conselhos | 42 | 84% |
| NSFW | 40 | 80% |
| Relacionamento | 39 | 78% |
| Dia a Dia | 39 | 78% |
| Update | 38 | 76% |
| Más notícias | 38 | 76% |
| Boas notícias! | 37 | 74% |

FONTE: O Autor (2023).

A classe com menor similaridade é “boas notícias!”, com 74%. Pela característica da classe “depressão” ser predominantemente de sentimentos negativos, o maior distanciamento ser da classe “boas notícias!”, que possui sentimento predominantemente positivos, é coerente.

Porém, nota-se que a linguagem dominante da comunidade possui certa homogeneidade. Isolando as palavras mais frequentes do eixo depressão, gera-se a figura abaixo:

FIGURA 11 - NUVEM DE PALAVRAS GERAL



FONTE: O Autor (2023).

A maior diferença é a palavra “sempre” e “até”, sendo a primeira uma ênfase de recorrência de uma situação e a outra uma preposição que expressa um limite posterior de tempo.

5.2.5 Modelos de Classificação

5.2.5.1 Classificação por elementos textuais

Para o comparativo amplo das classificações, inserido as acurácias dos classificadores (em porcentagens) comparado com as classes no quadro 13.

QUADRO 13 - RESULTADOS DAS CLASSIFICAÇÕES POR ELEMENTOS TEXTUAIS

| | Logistic Regression | RandomForest | SVM | Naïve Bayes |
|--------------------------|---------------------|--------------|-------|-------------|
| Desabafo | 91,35 | 92,30 | 91,58 | 91,41 |
| Update | 76,30 | 79,19 | 82,08 | 76,30 |
| Relacionamento | 87,80 | 87,36 | 91,94 | 72,55 |
| NSFW | 77,97 | 79,66 | 87,57 | 76,84 |
| Más notícias | 77,52 | 73,26 | 81,01 | 56,59 |
| Dia a Dia | 85,69 | 84,26 | 88,55 | 77,82 |
| Desabafo - sem conselhos | 90,60 | 93,29 | 91,28 | 90,60 |
| Boas notícias! | 85,76 | 82,12 | 88,41 | 86,09 |

FONTE: O Autor (2023).

Quando se trata de elementos textuais, verificou-se que os classificadores apresentam alta acurácia quando são utilizadas as classes “Desabafo” e “Desabafo – sem conselhos”.

De forma geral, a média dos classificadores com a classe “Desabafo” sendo comparativa com “Depressão” teve média de acurácia de $91,66 \pm 0,38$, já com “Desabafo – sem conselhos” teve $91,44 \pm 1,10$. Já o menor índice foi com a classe “Más notícias”, tendo média de acurácia de $72,10 \pm 9,36$, mantendo também uma alta dispersão entre as acurácias dos classificadores. Naïve Bayes, nesse contexto, teve seu valor de acurácia igual a 56,59.

De forma individual, a maior acurácia é a do método Random Forest, com 93,29 utilizando a classe “desabafo – sem conselhos”. O método SVM teve seu pico em 91,58 com a classe “SVM”. Já o método de regressão logística chegou em 91,35 com “desabafo”, e por fim o método Naïve Bayes teve sua melhor performance também com a classe desabafo, com 91,41. A pior performance desse experimento foi do classificador Naïve Bayes, que teve

acurácia de 56,59 comparando com a classe “más notícias”, curiosamente essa classe também teve baixa performance nos classificadores Random Forest e SVM.

5.2.5.2 Classificação por *features*

Para o comparativo amplo das classificações, inserido as acurácias dos classificadores comparado com as classes no quadro 14:

QUADRO 14 - RESULTADO DA CLASSIFICAÇÃO POR FEATURES

| | Logistic Regression | RandomForest | SVM | Naïve Bayes |
|--------------------------|---------------------|--------------|-------|-------------|
| Desabafo | 91,53 | 91,76 | 91,47 | 91,47 |
| Update | 79,77 | 84,97 | 78,61 | 76,30 |
| Relacionamento | 73,20 | 77,12 | 71,68 | 71,68 |
| NSFW | 77,40 | 79,10 | 76,84 | 76,84 |
| Más notícias | 60,08 | 60,47 | 53,1 | 48,45 |
| Dia a Dia | 78,53 | 80,14 | 77,46 | 77,46 |
| Desabafo - sem conselhos | 89,93 | 93,29 | 90,60 | 90,60 |
| Boas notícias! | 85,43 | 87,42 | 87,42 | 59,93 |

FONTE: O Autor (2023).

Os resultados das classificações das *features* derivadas e extraídas dos textos teve resultado semelhante das classificações por elementos textuais, sendo as classes comparativas com “Depressão”, “Desabafo” e “Desabafo – sem conselhos” com médias de acurácia de $91,56 \pm 0,12$ e $91,11 \pm 1,29$, respectivamente.

Porém, o comparativo entre os resultados menos favoráveis dos classificadores, a classe “Más notícias” teve o pior desempenho de acurácia, tanto comparando com as classificações de elementos textuais do item 5.2.3.1 quando por ela própria, tendo média de $55,53 \pm 5,03$.

Dos métodos utilizados, a maior acurácia é do método Random Forest, com o mesmo score da classificação apenas por texto, 93,29, provida pela utilização da classe “Desabafo – sem conselhos”. O método de Regressão Logística teve seu pico de acurácia com a utilização da classe “Desabafo”, com

91,53. Já os métodos SVM e Naïve Bayes tiveram o mesmo score de acurácia também com a classe “Desabafo”, sendo 91,47.

O método Naïve Bayes novamente foi o que teve menor vale de performance, alcançando 48,45 com a utilização da classe “Más notícias”.

Interessante notar que para a criação de um classificador nas condições utilizadas nesse trabalho, tem-se, de forma geral, possibilidade de escolher utilizar apenas *features* extraídas do texto para ter uma acurácia com performance positiva.

A classificação através das *features* derivadas e extraídas do texto apresentam algumas vantagens na hora do aprendizado de máquina, como por exemplo, por ter uma quantidade mais enxuta de dimensões, o processo de aprendizado dos algoritmos é mais rápido (por ter menor gasto computacional), e conseqüentemente, também a sua classificação, mas o tempo do pré-processamento é superior pela quantidade de métodos utilizados (extração de sentimentos, entidades e demais dimensões).

5.2.5.3 Classificação combinada

A maior acurácia dos três experimentos é a classificação combinada, utilizando o classificador Random Forest e tendo como base de “Não Depressão” a classe “Desabafo – sem conselhos”. Com acurácia de 93,49, sendo apenas 0,10 acima do mesmo método em ambas classificações de “apenas *features*” e “apenas texto”.

QUADRO 15 - RESULTADO DA CLASSIFICAÇÃO COMBINADA

| | Logistic Regression | Random Forest | SVM | Naïve Bayes |
|--------------------------|---------------------|---------------|-------|-------------|
| Desabafo | 91,53 | 92,34 | 91,53 | 90,37 |
| Update | 83,24 | 78,61 | 87,86 | 76,30 |
| Relacionamento | 87,58 | 86,71 | 91,29 | 87,36 |
| NSFW | 81,36 | 80,23 | 85,88 | 80,79 |
| Más notícias | 77,52 | 74,03 | 81,01 | 72,87 |
| Dia a Dia | 84,79 | 85,33 | 88,37 | 84,79 |
| Desabafo - sem conselhos | 89,93 | 93,49 | 91,95 | 89,26 |
| Boas notícias! | 87,42 | 85,76 | 88,41 | 84,44 |

FONTE: O Autor (2023).

A utilização dos métodos BoW e TD-IDF geram uma quantidade ampla de *features* extraídas do texto (conforme mencionado no tópico 3.4.2 deste documento), a implementação desse método dificulta identificarmos a importância de cada variável dentro da classificação do conjunto de dados. Para entender a importância e relevância das *features*, realizou-se a implementação do modelo no formato anterior (apenas *features* externas ao texto), onde o resultado é apresentado no Quadro 16.

QUADRO 16–IMPORTÂNCIA DAS FEATURES

| Posição | Feature | Importância |
|---------|----------------------|-------------|
| 1 | sentimentos | 16,00% |
| 2 | varianciaSentimentos | 15,00% |
| 3 | totalsentencas | 14,00% |
| 4 | PERSON | 12,00% |
| 5 | EVENT | 8,00% |
| 6 | LOCATION | 7,00% |
| 7 | CONSUMER_GOOD | 7,00% |
| 8 | ORGANIZATION | 5,00% |
| 9 | WORK_OF_ART | 5,00% |
| 10 | manha | 3,00% |
| 11 | tarde | 3,00% |
| 12 | noite | 3,00% |
| 13 | DATE | 2,00% |
| 14 | madrugada | 2,00% |
| 15 | PRICE | 0,00% |
| 16 | ADDRESS | 0,00% |
| 17 | PHONE_NUMBER | 0,00% |

FONTE: O Autor (2022).

Na leitura da tabela de importância das *features*, um score maior significa que a *feature* possui um amplo efeito no modelo que está sendo usado para prever uma variável. No contexto deste documento, representa qual *feature* tem maior importância para a classificação da classe depressão.

Dessa forma identifica-se que a carga sentimental presente no texto (*feature* sentimentos), seguida da variação de sentimentos (*feature* varianciaSentimentos), tamanho do texto (*feature* totalsentencas) e presença de citação de pessoas (*feature* PERSON), são de alta relevância para classificação dos textos.

6 CONSIDERAÇÕES FINAIS

A avaliação de saúde mental possui total dependência de um profissional de saúde especializado, que em contato com o seu paciente, pode determinar se possui algum transtorno da saúde mental.

Esse ponto se faz importante para lembrar que a rotulação derivada das postagens da comunidade do Reddit não possui profundidade para determinar se o autor da postagem possui de fato o transtorno depressivo em si, mas auxilia a entendermos quais as motivações derivadas da percepção do texto fazem a rotulação acontecer, e também possibilitam testes e experimentos para desenvolvimentos de modelos de classificação utilizando os elementos textuais e disponíveis na rede social do estudo.

O objetivo específico de **realizar uma revisão sistemática de metodologias e tecnologias utilizadas para resolução de problemas similares**, foi realizado na sessão 3, dentro dos limites temporais descritos na mesma, onde evidencia-se o aumento do estudo dentro da área de interesse, porém ainda com carência de estudos em português.

Além disto, pontua-se sobre a validade de utilização dos modelos que, mesmo tendo como foco a rotulação de depressão, não possuem metodologia com validação de especialistas da área. Derivado da revisão sistemática, foi definido os classificadores que compuseram a tarefa de classificação do trabalho, sendo eles os métodos Random Forest, Naïve Bayes, Regressão Logística e SVM.

Evidencia-se o crescimento desta área multidisciplinar que trará benefícios para nossa sociedade por meio da criação de ferramentas e análises para potencializar os insumos dos profissionais da área.

O objetivo específico que propõe **coletar e pré-processar os dados para a construção de um conjunto de dados**, foi apresentado na sessão 4.2 e evidenciado no decorrer da sessão 5. Conforme descrito na sessão 4, os dados foram coletados através de API e pré-processados com diferentes métodos no decorrer da sessão 5.

O objetivo específico de **analisar e descrever o conjunto de dados formado, com foco na relação entre as classes da rede social e as dimensões do conjunto de dados**, tal como os processos envolvidos na

geração das variáveis, teve sua evolução dentro da sessão 5. Para esse objetivo evidencia-se a descoberta de conhecimento gerada pela estatística descritiva realizada, combinada com os métodos de pré-processamento, para entendimento das dinâmicas textuais das classificações realizadas pela comunidade do Reddit.

O objetivo específico de **analisar e comparar os resultados dos métodos de classificação utilizados e discutir a sua aplicabilidade**, foi descrito e evidenciado dentro da sessão 5.2.3.

Finalmente, quanto ao objetivo geral de **desenvolver e avaliar classificadores para rotulação automática de depressão para a comunidade ‘Desabafos’ da rede social Reddit pela identificação de padrões na classificação textual e humana da comunidade**, foram implementados os classificadores Random Forest, Naïve Bayes, Regressão Logística e SVM, de acordo com a metodologia proposta no item 4.6 e avaliados conforme item 4.7.

Observou-se que os classificadores usados possuem alta acurácia nos contextos do experimento (classificação apenas com texto, classificação atributos sem o texto e classificação combinada), porém existe variação de acurácia de acordo com qual classe é utilizada como “não depressão”.

A maior acurácia foi alcançada com o classificador Random Forest no contexto de classificação combinado, com a classe “Depressão” de variável alvo e a classe “Desabafo – sem conselhos” como variável dependente, no entanto mesmo em um contexto que não se faz presente a classificação combinada, o Random Forest alcança os maiores valores de acurácia, se adaptando bem ao problema. Devido ao Random Forest ser um algoritmo mais robusto (conforme descrito no item 3.6.1), teve resultados positivo comparado com algoritmos menos robustos (como o Naïve Bayes).

Em relação à questão de pesquisa **“Quais atributos podem ser considerados para rotulação automática de Depressão em postagens em Redes Sociais?”** afirma-se que dentro do escopo limitado dessa pesquisa, os atributos derivados do texto, os atributos retirados da rede social em si e o texto transformado em dimensões de atributos (através de técnicas como BoW e TF-IDF) são eficientes na classificação da classe “Depressão”, utilizando-se do

método Random Forest, podendo seu resultado variar dependendo de qual classe será utilizada no contexto de “não depressivo” para treinamento do classificador. Outrossim, observando o quadro 16 (importância das *features*), os atributos “price”, “address” e “phone number” não possui influência na classificação do Random Forest, podendo serem retiradas sem gerar resultados inferiores aos alcançados.

De forma geral, a pesquisa auxilia na construção de literatura em português sobre o tema de rotulação de saúde mental derivada de redes sociais, além de incentivar a combinação de estatística descritiva, análise de dados, mineração de sentimentos e utilização de métodos de classificação para resolução do problema, sendo que em cada etapa existem aprendizados e maior entendimento da base de dados.

Com base nas etapas do trabalho, têm-se evidências dos vieses utilizados pela classificação humana dentro das classes disponíveis na comunidade estudada, auxiliando pesquisas que buscam entender questões culturais dos vieses humanos inseridos em elementos textuais.

O fluxo de trabalho criado para realização do trabalho (macro e específico para classificação) contribui para fundamentação de experimentos de classificação com dados providos por redes sociais, podendo ser evoluído para pipelines de classificação automática, desde que o retorno da classificação seja conectado com as redes sociais. Além disso, pode ser utilizado como comparativo com demais frameworks criados para objetivos semelhantes.

Em relação à aplicabilidade do modelo, é possível utilizá-lo para rotulação automática da comunidade, com alta acurácia. No entanto é enfatizado que o trabalho pode evoluir no contexto de aferição da saúde mental, desde que tenha auxílio especializado e aprofundamento metodológico.

Na execução do trabalho, percebeu-se a escassez de literatura sobre o tema em português, dessa forma, o presente trabalho contribui para o desenvolvimento da pesquisa no Brasil e em português.

Identificadas oportunidades de evolução metodológica, futuras pesquisas podem contar com a identificação dos usuários e apoio de especialistas em saúde mental, para que os classificadores possam ter aplicação no apoio da identificação de tendências depressivas.

Também existem oportunidades para criação de modelos com maior profundidade na utilização dos elementos textuais, como detecção de ironia, além da utilização de outras redes sociais que possuam formatos diferentes de atributos, como imagem, som, etc.

Por fim, espera-se que as evoluções das pesquisas permitam formar um cenário futuro em que a inteligência artificial auxilie na identificação precoce de tendências depressivas dos seus usuários, possibilitando a atuação dos especialistas de forma preventiva no cuidado da saúde mental da população.

REFERÊNCIAS

- AARØ, L. E. et al. Depressive episodes and depressive tendencies among a sample of adults in Kielce, south-eastern Poland. **Annals of Agricultural and Environmental Medicine**, v. 18, n. 2, p. 273–278, 2011.
- ALBERTO, J.; PORTO, D. Conceito e diagnóstico. **Brazilian Journal of Psychiatry**, v. 21, n. suppl 1, p. 06–11, maio 1999.
- ALSAGRI, H.; YKHLEF, M. Quantifying feature importance for detecting depression using Random Forest. **IJACSA) International Journal of Advanced Computer Science and Applications**, v. 11, n. 5, 2020.
- ANDERSON, K. E. Ask me anything: what is Reddit? **Library Hi Tech News**, v. 32, n. 5, p. 8–11, 6 jul. 2015.
- AQUINO, D. R. **Sintomas de depressão em universitários de medicina**. Disponível em: <<http://pepsic.bvsalud.org/pdf/bapp/v39n96/v39n96a09.pdf>>. Acesso em: 21 dez. 2021.
- ASCHBRENNER, K. A.; NASLUND, J. A.; BARTELS, S. J. A mixed methods study of peer-to-peer support in a group-based lifestyle intervention for adults with serious mental illness. **Psychiatric rehabilitation journal**, v. 39, n. 4, p. 328, 1 dez. 2016.
- AZEVEDO, A.; SANTOS, M. KDD, SEMMA and CRISP-DM: a parallel overview. **IADIS European Conf. Data Mining**, 2008.
- BAUMGARTNER, J. **Pushshift Reddit API documentation**. Disponível em: <https://medium.com/@jason_82699/pushshift-reddit-api-md-c2d70745c270>. Acesso em: 25 dez. 2022.
- BREIMAN, L. Random Forests. jan. 2001.
- CHANCELLOR, S.; DE CHOUDHURY, M. **Methods in predictive techniques for mental health status on social media: a critical review**. **Digital Medicine** Nature Research, , 1 dez. 2020.
- COOK, Darren. **Practical machine learning with H2O**: powerful, scalable techniques for Deep Learning and AI. **Sebastopol**: O'Reilly Media. 1st. ed. 2017.
- DATAREPORTAL – GLOBAL DIGITAL INSIGHTS. **Digital in Brazil: All the Statistics You Need in 2021**. Disponível em: <<https://datareportal.com/reports/digital-2021-brazil?rq=Brazil>>. Acesso em: 21 dez. 2021.
- DELEN, D. **Predictive analytics : data mining, machine learning and data science for practitioners**. 2nd ed ed. [s.l.: s.n.].
- Digital in Brazil: All the Statistics You Need in 2021 — DataReportal – Global Digital Insights**. Disponível em: <<https://datareportal.com/reports/digital-2021-brazil>>. Acesso em: 25 dez. 2022.
- FAYYAD, U.; HAUSSLER, D.; STOLORZ, P. **KDD for science data analysis: issues and examples**. 1996. Disponível em: <<https://dl.acm.org/doi/10.5555/3001460.3001471>>. Acesso em: 23 dez. 2021

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, p. 37–37, 15 mar. 1996.

FERREIRA, P. L. **Estatística descritiva e inferencial**: breves notas. 2005. Disponível em: <<https://estudogeral.uc.pt/handle/10316/9961>>. Acesso em: 19 jun 2023.

GALVÃO, T. F.; PANSANI, T. DE S. A.; HARRAD, D. Principais itens para relatar Revisões sistemáticas e Meta-análises: A recomendação PRISMA. **Epidemiologia e Serviços de Saúde**, v. 24, n. 2, p. 335–342, jun. 2015.

GARG, B. Design and development of Naïve Bayes classifier. **Dissertação**. 2013. Disponível em: <<https://library.ndsu.edu/ir/handle/10365/23048>>. Acesso em: 19 jun. 2023.

GEORGE, S.; JOSEPH, S. Text classification by augmenting Bag of Words (BOW): representation with co-occurrence feature. **Journal of Computer Engineering**, n. 1, p. 34–38, 2014.

GIUNTINI, F. T. n approach to the sequential evaluation of emotional behaviors of depressive users on social networks in groups and individually. 2021.

GLOBAL HEALTH ESTIMATES. **Depression and other common mental disorders global health estimates**. Geneva, 2017. Disponível em: <<https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf?sequence=1>>. Acesso em: 9 out. 2021

GOVINDASAMY, K. A. L.; PALANICHAMY, N. Depression detection using machine learning techniques on twitter data. **Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021**, p. 960–966, 6 maio 2021.

HARRISON, M.; SAFARI, AN O. M. COMPANY. **Machine learning pocket reference**. Revised Ed ed. [s.l.: s.n.].

KAYACIK, H. G.; ZINCIR-HEYWOOD, A. N.; HEYWOOD, M. selecting features for intrusion detection: a feature relevance analysis. **KDD 99**., 2005.

KHAIRNAR, J.; KINIKAR, M. machine learning algorithms for opinion mining and sentiment classification. **International Journal of Scientific and Research Publications**, v. 3, n. 6, 2013.

KOHAVI, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. **IJCAI**, 1995.

KURODA, Y. et al. Occurrence of depressive tendency and associated social factors among elderly persons forced by the Great East Japan Earthquake and nuclear disaster to live as long-term evacuees: a prospective cohort study. **BMJ open**, v. 7, n. 9, 1 set. 2017.

LANDER P. JARED. **R for everyone**: advanced analytics and graphics | **Amazon.com.br**. 2nd ed ed. [s.l.: s.n.].

LI, X. et al. A multi-dimensional context-aware recommendation approach based on improved random forest algorithm. **IEEE Access**, v. 6, p. 45071–45085, 13 ago. 2018.

MAYER, F. B. **A prevalência de sintomas de depressão e ansiedade entre os estudantes de medicina**: um estudo multicêntrico no Brasil. São Paulo: Universidade de São Paulo, 4 ago. 2017.

MINISTÉRIO DA MULHER, DA F. E DOS D. H.; SECRETARIA NACIONAL DA FAMÍLIA; OBSERVATÓRIO NACIONAL DA FAMÍLIA. **Boletim - Fatos e Números - Saúde Mental**. Brasília: [s.n.]. . Acesso em: 25 dez. 2022.

MINISTÉRIO DA SAÚDE. **Anualmente, mais de 700 mil pessoas cometem suicídio, segundo OMS — Português (Brasil)**. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/noticias/2022/setembro/anualmente-mais-de-700-mil-pessoas-cometem-suicidio-segundo-oms>>. Acesso em: 27 nov. 2022.

MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos sobre aprendizado de máquina. Sistemas inteligentes fundamentos e aplicações**. 1. ed. Barueri-SP: Manole Ltda, 2003. v. 1

PIATETSKY-SHAPIRO, G. Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop. **AI Magazine**, v. 11, n. 4, p. 68–68, 31 dez. 1990.

RAMOS, J. Using TF-IDF to Determine Word Relevance in Document Queries. v. 242, 2003.

REDDIT. **Reddit - Explore o que quiser**. Disponível em: <<https://www.reddit.com/>>. Acesso em: 26 nov. 2022.

RUNGER, G. C. ; M. D. C. **Applied statistics and probability for engineers**. . 3rd ed ed. [s.l: s.n.].

SARAVANAN, R.; SUJATHA, P. A State of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. **Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018**, p. 945–949, 7 mar. 2019.

STEINER, M. T. A. et al. Abordagem de um problema médico por meio do processo de KDD com ênfase à análise exploratória dos dados. **Gestão & Produção**, v. 13, n. 2, p. 325–337, 2006.

SURIYA PRAKASH, J. et al. Multi class support vector machines classifier for machine vision application. 2012 International Conference on Machine Vision and Image Processing, MVIP 2012. **Anais...2012**.

TADESSE, M. M. et al. Detection of depression-related posts in Reddit social media forum. **IEEE Access**, v. 7, p. 44883–44893, 2019.

WORLD HEALTH ORGANIZATION. **Suicídio**. Disponível em: <https://www-who-int.translate.google/news-room/fact-sheets/detail/suicide?_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt-BR&_x_tr_pto=sc>. Acesso em: 26 nov. 2022.

XIE, H.; JIANG, D.; ZHANG, D. Individuals with depressive tendencies experience difficulty in forgetting negative material: two mechanisms revealed by ERP data in the directed forgetting paradigm. **Scientific Reports 2018 8:1**, v. 8, n. 1, p. 1–14, 18 jan. 2018.

YANG, C. et al. **Depression tendency screening use text based emotional analysis technique**. p. 32035, 2019.

