UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

RAFAEL LAGE DE OLIVEIRA

Predição de transtornos de saúde mental em redes sociais com o uso de conhecimento extra-linguístico

São Paulo
2023

RAFAEL LAGE DE OLIVEIRA

Predição de transtornos de saúde mental em redes sociais com o uso de conhecimento extra-linguístico

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Metodologia e Técnicas da Computação

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 14 de abril de 2023. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Prof. Dr. Ivandré Paraboni

São Paulo

2023

Ficha catalográfica elaborada pela Biblioteca da Escola de Artes, Ciências e Humanidades,
com os dados inseridos pelo(a) autor(a)
Brenda Fontes Malheiros de Castro CRB 8-7012; Sandra Tokarevicz CRB 8-4936

Oliveira, Rafael Lage de
Predição de transtornos de saúde mental em redes sociais com o uso de conhecimento extra-linguístico / Rafael Lage de Oliveira; orientador, Ivandre Paraboni. -- São Paulo, 2023.
133 p: il.

Dissertacao (Mestrado em Ciencias) - Programa de Pós-Graduação em Sistemas de Informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, 2023.
Versão corrigida

1. Predição de Saúde Mental. 2. Depressão. 3. Ansiedade. 4. Redes Sociais. 5. Interação Social. 6. Homofilia. I. Paraboni, Ivandre, orient. II. Título.

Dissertação de autoria de Rafael Lage de Oliveira, sob o título **"Predição de transtornos de saúde mental em redes sociais com o uso de conhecimento extra-linguístico"**, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovada em 14 de abril de 2023 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Ivandré Paraboni
Universidade de São Paulo
Presidente

Prof. Dr. José de Jesús Pérez Alcázar
Universidade de São Paulo

Prof. Dr. Jó Ueyama
Universidade de São Paulo

Dedico este trabalho à minha querida esposa Raquel. Sua contribuição vai muito além do conhecimento e ensinamentos da área da Psicologia compartilhados a mim e importantes para a compreensão de muitos conceitos neste trabalho. Seu amor, compreensão, confiança, conselhos, orações e incansável apoio foram parte essencial de toda esta jornada, sem os quais nada disso seria possível. Sou grato a Deus por sua vida em minha vida e por toda a dedicação como profissional, esposa, companheira e melhor amiga. Dedico também este trabalho aos meus pais, Amauri e Elizabeth, por todo amor, por serem exemplos a mim e por me ensinarem os caminhos em que eu deveria andar, sobretudo o caminho da fé. De todo o coração, dedico este trabalho a vocês.

Agradecimentos

Agradeço primeiramente a Deus que me sustenta a cada dia e que me deu forças para concluir mais esta etapa em minha vida. Toda honra e toda glória sejam dadas a Ele.

Agradeço ao meu orientador Prof. Dr. Ivandré Paraboni, por toda a liderança e orientação ao longo deste período, as quais foram essenciais para a condução e conclusão deste trabalho. Agradeço também a confiança desde o início e pela oportunidade e privilégio de fazer parte de seu time.

Agradeço aos orientandos do Prof. Dr. Ivandré Paraboni que fizeram parte deste mesmo grupo de trabalho. Obrigado pelas sugestões, trocas de conhecimento e pelas contribuições para que este trabalho fosse aperfeiçoado.

Agradeço a todos os professores que me ensinaram em todo este período e contribuíram para minha formação e para o desenvolvimento deste trabalho.

Agradeço à Universidade de São Paulo por todo o suporte e por me ajudar a ser um profissional e cidadão melhor.

Agradeço a todos os familiares e amigos que me apoiaram, incentivaram, oraram por mim e estiveram comigo nos momentos mais difíceis e também nos momentos de alegria.

A todos estes, meus sinceros e profundos agradecimentos.

Resumo

OLIVEIRA, Rafael Lage de. **Predição de transtornos de saúde mental em redes sociais com o uso de conhecimento extra-linguístico**. 2023. 133 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2023.

Transtornos de saúde mental como depressão e ansiedade têm se tornado cada vez mais comuns em decorrência do ritmo de vida imposto pela sociedade moderna, afetando milhões de pessoas ao redor do mundo. Estes transtornos geram graves consequências comportamentais, refletindo na forma como os indivíduos se expressam e interagem socialmente. Pesquisas recentes nas áreas de Processamento de Língua Natural (PLN) e redes sociais têm utilizado métodos computacionais para estudar o comportamento e explorar o conteúdo gerado por usuários de mídias sociais, como o *Twitter*, em busca de reconhecer padrões que auxiliem a identificação de indivíduos que sofrem com estes tipos de transtornos. Modelos computacionais existentes para predição de transtornos de saúde mental frequentemente utilizam o conteúdo gerado pelos usuários nas redes sociais, principalmente na forma de textos. No entanto, este tipo de informação é esparsa e pode não ser suficiente para o desenvolvimento de modelos robustos. Por outro lado, estudos indicam que atributos extra-linguísticos podem conter conhecimento importante acerca de indivíduos com estes transtornos, como dados demográficos e de atividade, características comportamentais e atributos de rede e relacionamento, conforme sugerem os efeitos da homofilia. Este trabalho apresenta uma pesquisa de nível de mestrado acadêmico cujo objetivo principal é propor e desenvolver modelos computacionais para o reconhecimento de transtornos do tipo depressão e ansiedade a partir de dados do *Twitter* em português, abordando o conhecimento extra-linguístico disponível na rede social.

Palavras-chaves: Predição de Saúde Mental. Depressão. Ansiedade. Redes Sociais. Interação Social. Homofilia.

Abstract

OLIVEIRA, Rafael Lage de. **Predicting mental health disorders in social networks using extra-linguistic knowledge**. 2023. 133 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2023.

Mental health disorders such as depression and anxiety have become increasingly common as a result of the pace of life imposed by modern society, affecting millions of people around the world. These disorders generate serious behavioral consequences, reflecting on the way individuals express themselves and interact socially. Recent research in Natural Language Processing (NLP) and social networks has used computational methods to study behavior and explore content generated by users of social media, such as Twitter, in order to recognize patterns that help identify individuals who suffer from these kind of disorders. Existing computational models for predicting mental health disorders often use user-generated content on social networks, mainly in the form of texts. However, this kind of information is sparse and may not be sufficient for the development of robust models. On the other hand, studies indicate that extra-linguistic attributes may contain important knowledge about individuals with these disorders, such as demographic and activity data, behavioral characteristics and network and relationship attributes, as suggested by the effects of homophilia. This work presents an academic master's-level research that aims to propose and develop computational models for the recognition of depression and anxiety disorders from Twitter data in Portuguese, addressing extra-linguistic knowledge available on the social network.

Keywords: Mental Health Prediction. Depression. Anxiety. Social Networks. Social Interaction. Homophily.

Lista de figuras

Lista de quadros

Lista de tabelas

Lista de abreviaturas e siglas

ACL	*Association for Computational Linguistics*
ACM	*Association for Computing Machinery*
AM	Aprendizado de Máquina
ANOVA	*Analysis of Variance*
BDI	*Beck Depression Inventory*
BERT	*Bidirectional Encoder Representations from Transformers*
BLSTM	*Bidirectional Long Short-Term Memory*
BFS	*Breadth First Search*
BOW	*Bag of Words*
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CAE	*Cross Auto Encoder*
CES-D	*Center for Epidemiological Scale-Depression*
CLPsych	*Computational Linguistics and Clinical Psychology*
CNN	*Convolutional Neural Network*
DFS	*Depth First Search*
FFNN	*Feed-Forward Neural Network*
FGM	*Factor Graph Model*
GCN	*Graph Convolutional Network*
IA	Inteligência Artificial
IEEE	*Institute of Electrical and Electronics Engineers*
kNN	*k-Nearest Neighbors*
LDA	*Latent Dirichlet Allocation*

LIWC	*Linguistic Inquiry and Word Count*
LSTM	*Long Short-Term Memory*
MLM	*Masked Language Model*
MLP	*Multilayer Perceptron*
NLP	*Natural Language Processing*
OMS	Organização Mundial da Saúde
PLN	Processamento de Língua Natural
RBF	*Radial Basis Function*
RNA	Rede Neural Artificial
RNN	*Recurrent Neural Network*
SGD	*Stochastic Gradient Descent*
SVM	*Support Vector Machine*
TEPT	Transtorno do Estresse Pós-traumático
TF-IDF	*Term Frequency-Inverse Document Frequency*
URL	*Uniform Resource Locator*

Sumário

1 Introdução

Estima-se que mais de 300 milhões de pessoas sofram de depressão e transtornos de ansiedade no mundo, o que equivale a aproximadamente 4,5% da população mundial. Consequências mais graves, como o suicídio, podem ter como origem transtornos de saúde mental não tratados adequadamente. Segundo dados da Organização Mundial da Saúde (OMS), depressão é o principal motivo de mortes por suicídio no mundo, com aproximadamente 800 mil casos por ano (WORLD HEALTH ORGANIZATION, 2017). Portanto, um diagnóstico em fases iniciais da doença é essencial para o sucesso do tratamento e a prevenção da ocorrência de consequências extremas.

Para um tratamento adequado a estes tipos de transtornos, é necessário que haja um diagnóstico seguro, realizado por um profissional de saúde especializado. Métodos tradicionais de diagnóstico e avaliação do grau destes transtornos são amplamente utilizados por esses profissionais, como a aplicação de questionários e escalas reconhecidas mundialmente. Contudo, os métodos tradicionais possuem um alto custo de aplicação e são restritos a uma pequena parcela da população, além de frequentemente serem aplicados tardiamente (COPPERSMITH; DREDZE; HARMAN, 2014).

Recentemente, métodos computacionais têm contribuído para a identificação de transtornos de saúde mental a partir de dados extraídos de redes sociais, como *Twitter*, *Facebook* e *Reddit*. Esta abordagem traz benefícios, como identificar comportamentos de risco, prover intervenções precoces e alcançar uma população de difícil acesso (CHANCELLOR; CHOUDHURY, 2020). A identificação de indivíduos com estes tipos de transtornos frequentemente é modelada como um problema computacional de Aprendizado de Máquina (AM) supervisionado que faz uso de um conjunto de dados com características e publicações de usuários de redes sociais. Estes usuários são rotulados com a presença ou ausência de transtornos de saúde mental com base em métodos de validação externa ou autorrelato. A validação externa tipicamente consiste em utilizar questionários, escalas tradicionais ou o parecer de um profissional de saúde especializado para a avaliação da saúde mental dos usuários, enquanto que o método baseado em autorrelato utiliza informações de diagnóstico fornecidas pelos próprios usuários (e.g., “No mês passado, fui diagnosticado com depressão”). Apesar de o método baseado em validação externa ser altamente confiável, possui um custo alto de aplicação e tende a limitar o número de usuários selecionados. Por

outro lado, o método baseado em autorrelato é mais susceptível a imprecisões, no entanto este problema é minimizado pelo alto volume de usuários selecionados (COPPERSMITH *et al.*, 2015). A seleção de usuários baseada em autorrelatos vêm sendo amplamente utilizada em trabalhos recentes (YATES; COHAN; GOHARIAN, 2017; COHAN *et al.*, 2018; LOSADA; CRESTANI; PARAPAR, 2019) e foi também explorada pela presente pesquisa.

Modelos computacionais para a predição de transtornos de saúde mental frequentemente utilizam dois tipos de conhecimento: linguístico e extra-linguístico. O conhecimento linguístico é extraído do conteúdo textual gerado pelos usuários e concentra grande parte das características exploradas por estes modelos, uma vez que é uma das formas mais comuns de expressar sentimentos nas redes sociais. O conhecimento extra-linguístico contempla outros tipos de características que extrapolam o conteúdo textual gerado pelos usuários, como o histórico de atividades na rede social, as interações com outros usuários, dados demográficos e a estrutura da rede de relacionamentos (SHRESTHA; SPEZZANO, 2019). Esses tipos de conhecimento podem ser complementares entre si e em conjunto têm contribuído para o desenvolvimento de modelos mais robustos (LIN *et al.*, 2017; SHRESTHA; SPEZZANO, 2019; SINHA *et al.*, 2019).

No que diz respeito ao uso de conhecimento extra-linguístico, estudos recentes têm explorado características de rede e relacionamento entre usuários de redes sociais para a detecção de transtornos de saúde mental a partir de modelos computacionais, tendo como fundamentação o conceito da homofilia (SINHA *et al.*, 2019). A homofilia é definida no trabalho em McPherson, Smith-Lovin e Cook (2001) como o princípio de que um contato entre pessoas semelhantes ocorre em uma taxa mais alta do que entre pessoas diferentes. Em outras palavras, indivíduos com características semelhantes possuem a tendência de criar conexões sociais entre si, o que também pode ser observado entre pessoas com transtornos de saúde mental (WANG; ZHANG; SUN, 2013).

1.1 Problema de pesquisa

Modelos computacionais existentes para a predição de transtornos de saúde mental em redes sociais digitais frequentemente utilizam conteúdo gerado por usuários, principalmente na forma de textos (WANG *et al.*, 2017; RICARD *et al.*, 2018; SHRESTHA; SERRA;

SPEZZANO, 2020). No entanto, a informação textual disponível em redes sociais é limitada e esparsa, constantemente possui ruídos e pode não ser suficiente para desenvolver um modelo preditivo robusto baseado somente neste tipo de característica (LIN *et al.*, 2017; SINHA *et al.*, 2019), podendo gerar severa perda de informação (CHAI *et al.*, 2019) ou até mesmo o uso de conhecimento irrelevante (ZOGAN *et al.*, 2021). Por outro lado, estudos na área de redes sociais indicam que indivíduos com transtornos de saúde mental preferem formar conexões sociais com indivíduos que possuam transtornos semelhantes, conforme sugerem os efeitos da homofilia (VEDULA; PARTHASARATHY, 2017; GIUNTINI *et al.*, 2021). Com base neste princípio, e no fato de que a interação social é uma das mais importantes características de plataformas de mídias sociais, esses estudos têm passado a considerar este tipo de conhecimento para melhorar a efetividade do reconhecimento de transtornos de saúde mental em redes sociais (LIN *et al.*, 2017; WU *et al.*, 2020). O presente trabalho que explora o conhecimento extra-linguístico disponível em redes sociais e foi desenvolvido para a modalidade de mestrado acadêmico é apresentado a seguir.

1.2 Questões de pesquisa

Este trabalho se propôs a responder as seguintes questões de pesquisa:

Q1. Quais características extra-linguísticas baseadas no comportamento de usuários em redes sociais podem contribuir significativamente para a predição de transtornos de saúde mental?

Q2. Modelos computacionais que combinam diferentes tipos de características comportamentais de usuários de redes sociais possuem resultados significativamente superiores aos modelos de classificação de transtornos de saúde mental baseados apenas em características individuais?

Estas questões de pesquisa foram investigadas comparando-se modelos de classificação de transtornos de saúde mental baseados em atributos extra-linguísticos, sejam eles individuais ou modelos que combinam diferentes fontes de conhecimento, utilizando-se para esse fim métricas tradicionais de AM. Estes modelos foram também comparados com um modelo de *baseline* que considera apenas atributos psicolinguísticos.

1.3 Objetivo

O objetivo deste trabalho foi desenvolver modelos computacionais baseados em técnicas de AM supervisionado para o reconhecimento de transtornos de saúde mental do tipo depressão e ansiedade a partir de dados da rede social *Twitter* em português, contemplando características comportamentais, que envolvam a atividade, estrutura da rede e os relacionamentos entre usuários, de modo a verificar se tais características podem contribuir significativamente para as tarefas de predição propostas.

O presente trabalho toma por base um conjunto de dados multimodal já existente, contendo aproximadamente 38,5 milhões de *tweets* de 24,4 mil usuários, além das listas de conexões destes usuários com aproximadamente 6,6 milhões de amigos e 11,6 milhões de seguidores. Adicionalmente, o conjunto de dados contém a categoria a qual cada usuário pertence (i.e., diagnosticado com depressão, diagnosticado com ansiedade ou grupo de controle) e outros conhecimentos extra-linguísticos, como dados demográficos e comportamentais dos usuários, possibilitando a extração de conteúdo relacionado à estrutura da rede de relacionamentos e à atividade na rede social. As características linguísticas e extra-linguísticas disponibilizadas por este conjunto de dados fornecem importante conhecimento para a construção de modelos preditivos de classificação de transtornos de saúde mental baseados em AM supervisionado.

O restante deste documento está organizado da seguinte forma. O capítulo 2 apresenta os conceitos básicos sobre métodos de representação e modelos computacionais de classificação baseados em técnicas de AM supervisionado. Neste capítulo é ainda apresentado os conceitos básicos sobre o conjunto de dados SetembroBR. O capítulo 3 apresenta a Revisão Bibliográfica Sistemática. O capítulo 4 detalha o método utilizado neste projeto. Nos capítulos 5 e 6 são apresentados os experimentos com classificadores individuais e classificadores combinados, respectivamente. Por fim, o capítulo 7 apresenta as considerações finais desta pesquisa.

2 Conceitos básicos

Este capítulo apresenta os conceitos básicos relacionados a este trabalho e está dividido em cinco seções, onde serão abordados os principais métodos de representação textual (seção 2.1) e de representação baseada em rede (seção 2.2), os principais métodos de AM (seção 2.3), o conceito de comitê de classificadores e o detalhamento dos métodos utilizados neste trabalho (seção 2.4), e, por fim, uma descrição detalhada do conjunto de dados SetembroBR (seção 2.5).

2.1 Representação de dados textuais

Nesta seção serão apresentados os conceitos básicos de alguns dos principais métodos de representação textual, i.e., métodos tradicionais e métodos baseados em modelos pré-treinados de língua.

2.1.1 Representação textual baseada em contagem de palavras

O modelo Saco de Palavras (do inglês, *Bag of Words* - BOW) é um método tradicional e simplificado de representação textual em que cada texto é representado pela contagem de todas as palavras que o compõe. Este modelo foi originalmente proposto em Salton, Wong e Yang (1975) e sugere o uso de um vetor de tamanho igual ao número de palavras únicas do texto, em que cada posição do vetor armazena a frequência da palavra no texto que está sendo representado. Em tarefas de classificação de textos, a frequência de cada palavra é normalmente utilizada como uma característica, o que pode gerar dados muito esparsos devido ao volume elevado de palavras únicas que um texto pode apresentar.

No método utilizado pelo modelo BOW, as palavras possuem a mesma relevância, uma vez que apenas a frequência de cada palavra é computada. Como forma de minimizar este problema, na abordagem TF-IDF (do inglês, *term frequency–inverse document frequency*), pesos são atribuídos a cada palavra de acordo com a importância que cada uma delas possui no texto em comparação a um conjunto de documentos. Em outros termos, a medida TF-IDF considera a frequência das palavras em um documento único, no entanto é ponderada pela frequência destas palavras em um conjunto de documentos, diminuindo a

importância de palavras que são muito comuns a diversos documentos (YUN-TAO; LING; YONG-CHENG, 2005).

2.1.2 Representação textual baseada em aspectos psicolinguísticos

O LIWC (*Linguistic Inquiry and Word Count*) é uma aplicação de análise textual proposta em Pennebaker, Francis e Booth (2001) com o objetivo de fornecer um método efetivo para estudo de aspectos emocionais, cognitivos e estruturais expressados por meio de textos. O método consiste em classificar cada palavra de um documento em uma categoria psicolinguística e calcular o percentual de palavras no texto que corresponde a cada uma destas categorias. Um vetor com os percentuais calculados de tamanho igual ao número de categorias é gerado como resultado da representação textual de um documento.

A estrutura mais importante do LIWC é o dicionário, também chamado de léxico, e originalmente foi desenvolvido para o idioma inglês contendo 74 categorias, sendo 17 relacionadas a dimensões linguísticas padrão (e.g., percentual de pronomes, artigos, preposições), 25 que se referem a processos psicológicos (e.g., afeto, cognição, sentimentos positivos), 10 referentes a relatividade (e.g., tempo, espaço, movimento) e 19 associadas a interesses pessoais (e.g., trabalho, casa, atividades de lazer). Este dicionário possui aproximadamente 2300 palavras ou raízes de palavras, onde cada uma delas pode pertencer a uma ou mais categorias (PENNEBAKER; FRANCIS; BOOTH, 2001).

O pré-processamento dos textos é uma etapa importante do LIWC, podendo influenciar diretamente os resultados finais de modelos baseados neste tipo de representação. Por isso, faz-se necessário a preparação dos textos, como ajustes em palavras com erros ortográficos e palavras de uso inapropriado (PENNEBAKER; FRANCIS; BOOTH, 2001). O LIWC possui versões desenvolvidas para diversos idiomas, incluindo o Português (FILHO; PARDO; ALUISIO, 2013), e frequentemente é utilizado para aplicações de análise de sentimentos em textos e inferências de personalidade.

2.1.3 Representação textual baseada em modelos pré-treinados de língua

Modelos recentes de representação textual têm explorado métodos robustos em diversas aplicações de Processamento de Língua Natural (PLN). Neste contexto, um

dos modelos mais explorados é o BERT (*Bidirectional Encoder Representations from Transformers*), que é um modelo de língua baseado em transformadores, proposto em Devlin *et al.* (2019), e sua arquitetura está baseada em codificadores bidirecionais multicamadas (VASWANI *et al.*, 2017).

A maior contribuição do BERT consiste em demonstrar a importância de um pré-treinamento bidirecional para representação de língua, permitindo que o contexto anterior e posterior a uma palavra ou sentença seja levado em consideração, diferentemente de outros modelos que o fazem em uma única direção. Para atingir tal objetivo, o BERT foi pré-treinado com um Modelo de Língua Mascarado (do inglês, *Masked Language Model* - MLM), que consiste em "mascarar" alguns dos elementos textuais e predizê-los a partir de seu contexto anterior e posterior. Adicionalmente, o BERT foi também pré-treinado na tarefa de "predição da próxima sentença", que consiste em identificar se uma próxima sentença é provável dada uma primeira sentença e seu contexto (DEVLIN *et al.*, 2019). Ambas as tarefas foram originalmente pré-treinadas especificamente para o idioma inglês a partir de córpus de dimensões muito grandes, no entanto existem iniciativas para outros idiomas, incluindo o português (SOUZA; NOGUEIRA; LOTUFO, 2020).

Após a etapa de pré-treinamento, que possui um alto custo computacional e é realizada uma única vez, o passo seguinte consiste em fazer um ajuste fino, ou seja, a partir do modelo pré-treinado, são realizados novos treinamentos com dados rotulados e de dimensões menores para as tarefas específicas de interesse. Essa arquitetura tem gerado resultados que superam outros modelos linguísticos, com importantes contribuições para a área de PLN.

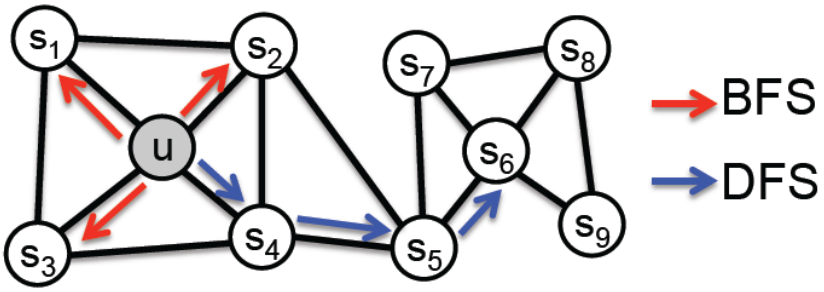
2.2 Representação de redes de relações

O *node2vec* é um algoritmo escalável para o aprendizado de características em redes proposto em Grover e Leskovec (2016), cujo método consiste em aprender representações dos nós por meio de *embeddings* de características, de modo a maximizar a probabilidade de preservar a vizinhança dos nós na rede em um espaço geométrico de baixa dimensão. O algoritmo *node2vec* utiliza métodos semi-supervisionados para o aprendizado de características dos nós em uma rede, buscando solucionar um problema de otimização de máxima probabilidade. Essa abordagem tem como objetivo obter ganhos em relação a métodos supervisionados e não-supervisionados tradicionais que sofrem com o custo da

complexidade de tempo de treinamento e baixa performance computacional e estatística no processo de extração de características, possibilitando a escalabilidade para redes maiores e mais complexas. A estrutura do *node2vec* pode ser dividida em três fases sequenciais: (1) pré-processamento para computar as probabilidades de transição, (2) simulações de passeios aleatórios e (3) otimização utilizando o método Gradiente Descendente Estocástico (do inglês, *Stochastic Gradient Descent* - SGD).

A arquitetura do *node2vec* é uma extensão para redes do modelo *Skip-Gram*, construído originalmente para a representação de palavras em aplicações de PLN, que busca preservar informações do contexto em que estas palavras estão inseridas (MIKOLOV *et al.*, 2013). No entanto, textos possuem uma natureza linear, ou seja, a vizinhança normalmente é definida por meio de palavras consecutivas, enquanto que a vizinhança de nós em redes é mais complexa e não-linear. Para resolver este problema, *node2vec* utiliza uma estratégia de geração de amostras aleatórias de vizinhanças, que explora os dois principais métodos de busca em redes, i.e., Busca em Largura (do inglês, *Breadth-first Sampling* - BFS) e Busca em Profundidade (do inglês, *Depth-first Sampling* - DFS), ilustrados pela figura 1 (GROVER; LESKOVEC, 2016).

Figura 1 – Estratégias de busca BFS e DFS a partir de um nó u



Fonte – Grover e Leskovec (2016)

A estratégia de amostragem de vizinhanças proposta pelo *node2vec* é baseado na teoria matemática do Passeio Aleatório, em que a probabilidade de transição para um nó c_i durante uma simulação é dada pela equação 1. π_{vx} corresponde à probabilidade de transição entre os nós v e x e Z é a constante de normalização (GROVER; LESKOVEC, 2016).

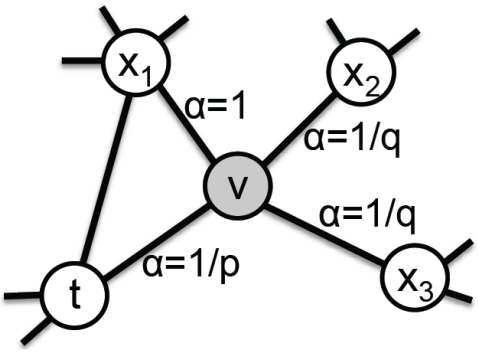
$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{se } (v, x) \in E \\ 0 & \text{caso contrário} \end{cases} \quad (1)$$

A abordagem tradicional do Passeio Aleatório está baseada em pesos estáticos das arestas e não permite que a amostragem dos nós leve em consideração a estrutura da rede e a vizinhança na qual o nó está inserido. No entanto, o método utilizado pelo *node2vec* se baseia no conceito de Passeio Aleatório de segunda ordem, no qual dois parâmetros p e q guiam o passo segundo a probabilidade de transição $\pi_{vx} = \alpha_{pq}(t,x) \cdot w_{vx}$, em que $\alpha_{pq}(t,x)$ é definido conforme a equação 2 e d_{tx} corresponde ao caminho mais curto entre os nós t e x. Os parâmetros p e q permitem que a busca explore os benefícios das estratégias BFS e DFS e os passos sejam dados levando-se em consideração a vizinhança dos nós (GROVER; LESKOVEC, 2016).

$$\alpha_{pq}(t,x) = \begin{cases} \frac{1}{p} & \text{se } d_{tx} = 0 \\ 1 & \text{se } d_{vx} = 1 \\ \frac{1}{q} & \text{se } d_{tx} = 2 \end{cases} \tag{2}$$

A figura 2 ilustra uma simulação do Passeio Aleatório de segunda ordem utilizado pelo *node2vec*, em que o passo é dado do nó t para o v e o próximo passo está sendo avaliado de acordo com a probabilidade de transição ajustada ao parâmetros p e q.

Figura 2 – Ilustração do Passeio Aleatório no *node2vec*



Fonte – Grover e Leskovec (2016)

Por fim, o algoritmo aplica uma função objetivo baseada em grafos que utiliza o método Gradiente Descendente Estocástico com amostragem negativa para a fase de otimização, motivado por trabalhos anteriores na área de PLN. Assim, representações dos nós por meio de *embeddings* podem contribuir para o aprendizado de características ao considerar a vizinhança em que um nó está inserido, de forma semelhante a que o *Skip-Gram* considera palavras em contextos similares como tendo os mesmos significados (GROVER; LESKOVEC, 2016).

2.3 Métodos de aprendizado de máquina

Nesta seção serão apresentados os conceitos básicos de alguns métodos de AM utilizados em modelos de classificação.

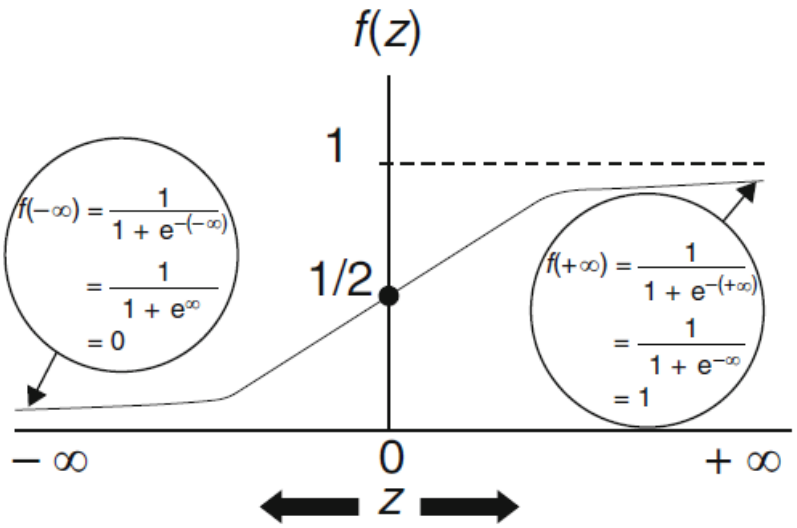
2.3.1 Regressão logística

Regressão logística é um método estatístico utilizado em modelos de classificação e diferencia-se de uma regressão linear por se aplicar a problemas em que o que está sendo modelada é a probabilidade de que um evento aconteça (acomodando valores entre 0 e 1) ou o caso em que a categoria de resposta é particularmente binária (DEMARIS *et al.*, 2013). Este método se baseia em uma função logística que foi originalmente descoberta no ano de 1920 em um estudo do crescimento populacional dos Estados Unidos (PEARL; REED, 1920) e ganhou suporte na área da química quando foi empregada (com algumas variações) para descrever o curso de reações autocatalíticas (CRAMER, 2002; BERTSIMAS; KING, 2017). A função logística é apresentada pela equação 3 no trabalho em Kleinbaum e Klein (2010).

$$f(z) = \frac{1}{1 + e^{-z}} \tag{3}$$

A figura 3 ilustra a curva sigmoide gerada pela função logística. É possível observar que quando o parâmetro z recebe o valor -∞ a função logística *f(z)* é igual a 0 (balão ao lado esquerdo). Por outro lado, quando o parâmetro z recebe o valor ∞ a função logística *f(z)* é igual a 1 (balão ao lado direito). Desta forma, a função logística assegura que o resultado do modelo será sempre valores entre 0 e 1 (KLEINBAUM; KLEIN, 2010).

Figura 3 – Curva sigmoide gerada pela função logística



Fonte – Kleinbaum e Klein (2010)

Para obter o modelo logístico a partir da função logística, z é definido conforme a equação 4, onde a variável X (X_1, X_2,..., X_k) é uma variável independente de interesse e α e β (β_1, β_2,..., β_k) são termos constantes que representam parâmetros desconhecidos e que devem ser estimados com base nos dados obtidos das variáveis independentes (KLEINBAUM; KLEIN, 2010). No contexto de AM, as variáveis independentes são os atributos de interesse que descrevem um objeto.

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \tag{4}$$

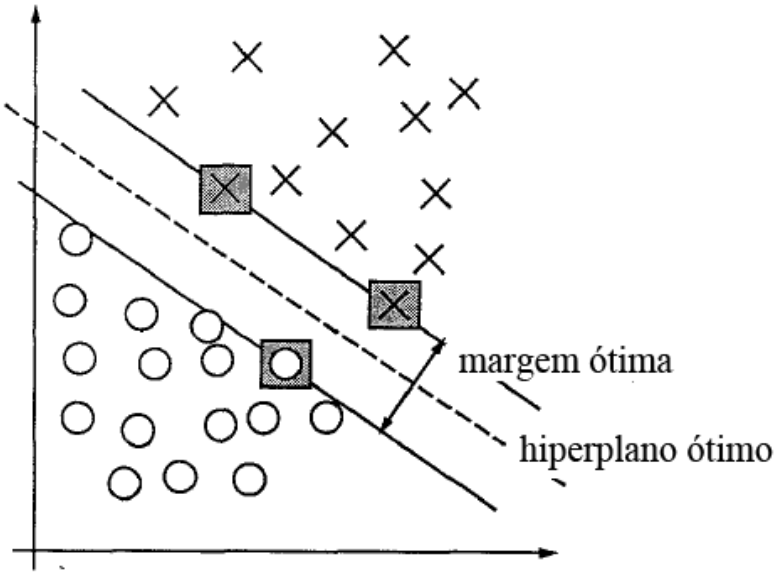
A simplicidade e efetividade de modelos construídos com regressão logística tem contribuído para que este método seja bastante popular e muito utilizado em problemas de AM (BERTSIMAS; KING, 2017).

2.3.2 Máquina de vetores de suporte

Uma Máquina de Vetores de Suporte (SVM) é um algoritmo de AM supervisionado introduzido pelo trabalho em Cortes e Vapnik (1995) que tem como base a teoria do aprendizado estatístico (VAPNIK, 1995) e é utilizado em problemas de regressão e classificação. Esta seção se limitará a a presentar o conceito básico de SVMs em sua formulação original, ou seja, para problemas de classificação binária, embora existam técnicas de generalização para tratar problemas multiclasse.

A ideia básica por trás do algoritmo SVM é encontrar um hiperplano ótimo que separe os dados a partir de um conjunto de treinamento de forma a maximizar a capacidade de generalização, ou seja, de predizer a classe correta para novos dados (NOBLE, 2006). Para este fim, o hiperplano deve ser estabelecido com a máxima distância dos vetores de dados mais próximos de cada uma das classes. Estes vetores de dados são denominados vetores de suporte e a máxima distância é definida como margem do hiperplano (CORTES; VAPNIK, 1995). A figura 4 ilustra o hiperplano ótimo e a margem ótima estabelecidos a partir de vetores de suporte para um problema linearmente separável em um espaço bidimensional de classificação binária.

Figura 4 – Exemplo de um problema linearmente separável em um espaço bidimensional. Os vetores de suporte destacados em quadrados cinzas definem a margem ótima para a separação entre duas classes



Fonte – Adaptado de Cortes e Vapnik (1995)

Encontrar a margem ótima com a máxima separação se resume a um problema de otimização com restrições, ou seja, a maximização da margem com restrições impostas de maneira a assegurar que não haja dados de treinamento entre as margens de separação das classes (LORENA; CARVALHO, 2007). Contudo, a maioria dos problemas reais não são linearmente separáveis, seja pela presença de ruídos nos dados, *outliers*, ou pela própria natureza do problema. Para tratar esta questão, o algoritmo SVM introduz o conceito de margens suaves, que permite que algumas restrições sejam violadas e os dados permaneçam dentro da margem ou mesmo ocorram erros de classificação (LORENA; CARVALHO, 2007). Para lidar com esta "suavização"do problema, um parâmetro C é utilizado como

termo de regularização, controlando a quantidade de dados que podem violar as restrições e o quão distante do hiperplano eles podem permanecer (NOBLE, 2006).

O algoritmo SVM lida com problemas essencialmente não-lineares (em que não é possível traçar um hiperplano para a separação das classes e tampouco ser tratado com margens suaves) mapeando o conjunto de treinamento para um novo espaço de maior dimensão, onde é possível traçar um hiperplano que separe adequadamente os dados (LORENA; CARVALHO, 2007; CORTES; VAPNIK, 1995). Este mapeamento, contudo, pode gerar problemas de sobreajustes, uma vez que o novo espaço de maior dimensão pode ser muito específico aos dados de treinamento. Para lidar com este problema, funções de núcleo são utilizadas nesta tarefa, evitando que dimensões irrelevantes sejam introduzidas (NOBLE, 2006). Os núcleos mais utilizados para problemas que não são linearmente separáveis são os polinomiais, gaussianos ou RBF e os sigmoidais (LORENA; CARVALHO, 2007).

O algoritmo SVM é robusto para problemas com dados de grande dimensão, para categorização de textos e é menos complexo do que Redes Neurais Artificiais (não há mínimos locais na função objetivo, apenas um único mínimo global). Apesar disso, ele é sensível à escolha de parâmetros e não é de fácil interpretação do modelo final (JOACHIMS, 2005; LORENA; CARVALHO, 2007).

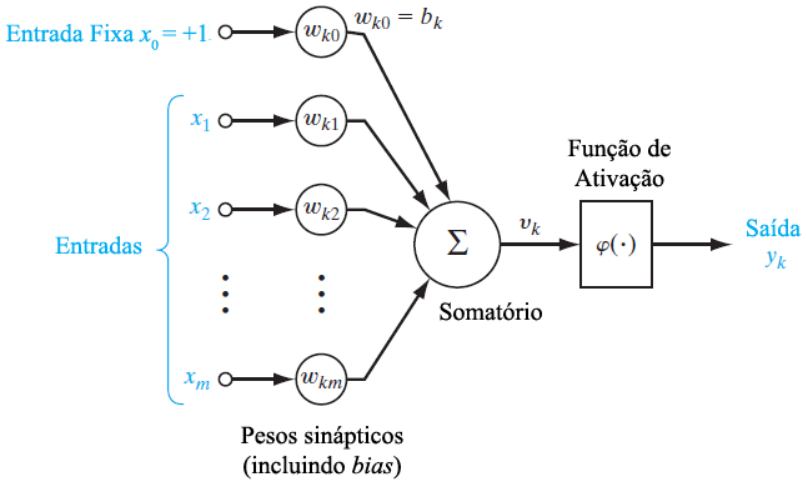
2.3.3 Redes neurais artificiais

Uma Rede Neural Artificial (RNA) é definida em Fausett (2006) como um sistema de processamento de informação que possui características semelhantes ao sistema neural biológico do ser humano, onde a informação é transmitida por meio de sinais entre unidades básicas chamadas neurônios. Existem diferentes tipos de neurônios em RNAs, sendo o *Perceptron* um dos mais comuns em diversos tipos de aplicação.

A conexão entre neurônios do tipo *Perceptron* possui um peso que é multiplicado ao sinal de entrada, e a soma destas multiplicações alimenta uma função de ativação que irá calcular o sinal de saída. A figura 5 ilustra o modelo de um neurônio que recebe x_1, x_2,...x_m como sinais de entrada, aplica os pesos w_{k1}, w_{k2},...w_{km} nestes sinais, e gera um sinal de saída y_k após aplicar uma função de ativação $\varphi(\cdot)$ no somatório dos sinais de entrada multiplicados por seus pesos correspondentes. Esse modelo pode conter também

um *bias* (b_k), que tem o efeito de aumentar ou diminuir o sinal de entrada na função de ativação (HAYKIN, 2009).

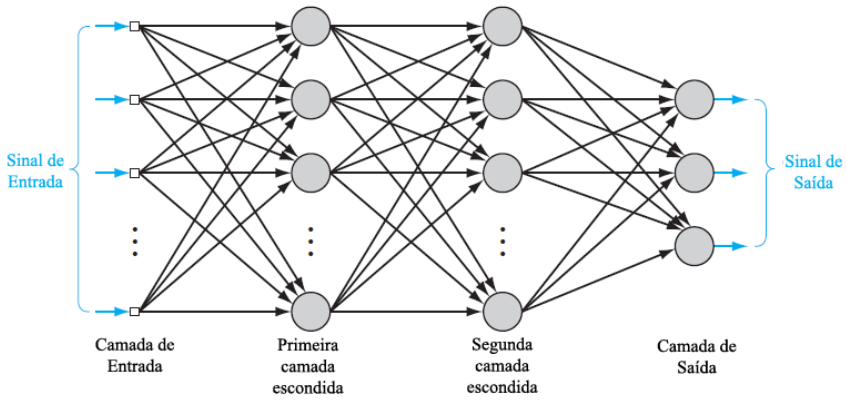
Figura 5 – Modelo de um neurônio *Perceptron*



Fonte – Adaptado de Haykin (2009)

Neurônios do tipo *Perceptron*, com função de ativação linear, são capazes de resolver apenas problemas linearmente separáveis. Para lidar com essa limitação, uma nova estrutura de rede neural conhecida como *Multilayer Perceptron* (MLP) é proposta. Uma MLP é basicamente composta por uma camada de entrada com neurônios sensoriais, uma ou mais camadas escondidas que fazem processamento de sinal a partir de funções de ativação não-lineares, e uma camada de saída que processa os sinais a partir de funções de ativação lineares ou não-lineares (HAYKIN, 2009). A figura 6 ilustra a estrutura de uma MLP com duas camadas escondidas

Figura 6 – Arquitetura de uma MLP com duas camadas escondidas



Fonte – Adaptado de Haykin (2009)

Em uma rede neural de múltiplas camadas de aprendizado supervisionado, o treinamento se torna uma tarefa mais custosa. Um método popular utilizado para o treinamento de MLPs é a utilização do algoritmo de retropropagação (do inglês, *backpropagation*), que consiste em propagar um sinal de erro que é produzido como resultado da comparação da saída da rede com a resposta desejada. Esse sinal é transmitido no sentido oposto ao original da rede, permitindo que ajustes nos pesos das conexões sejam realizados e consequentemente a rede execute a tarefa de aprendizado a partir dos novos dados de entrada.

2.3.4 Redes neurais de aprendizado profundo

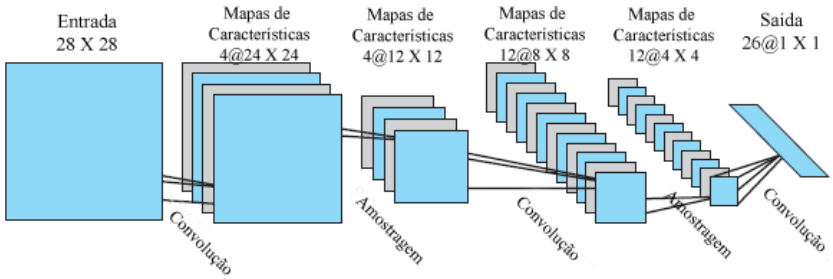
Nesta seção serão apresentadas as principais arquiteturas de redes neurais baseadas em aprendizado profundo utilizadas em problemas de AM.

Redes neurais convolucionais

Redes Neurais Convolucionais (do inglês, *Convolutional Neural Network* - CNN) são apresentadas em Haykin (2009) como um tipo especial de *Perceptron* multicamadas, desenhadas especificamente para reconhecer formas bidimensionais com um alto grau de invariância a diversos tipos de distorção. Esse tipo de rede tem como inspiração a neurobiologia, mais especificamente a organização do córtex visual dos animais, e é muito explorado em aplicações de processamento de imagens e vídeo, além de outros usos menos frequentes, como processamento de voz e língua natural.

A estrutura de uma CNN é composta por camadas de convolução, mapeamento de características e camada de amostragem. As camadas de convoluções funcionam como filtros que buscam as características mais relevantes. A tarefa de mapeamento de características permite simplificar as informações na saída da camada convolucional, mapeando as características de forma condensada. Por fim, a camada de amostragem reduz o número de características e outras formas de distorção (HAYKIN, 2009). A figura 7 apresenta a arquitetura de uma CNN em um exemplo de processamento de imagem.

Figura 7 – Rede Convolucional em processamento de imagem



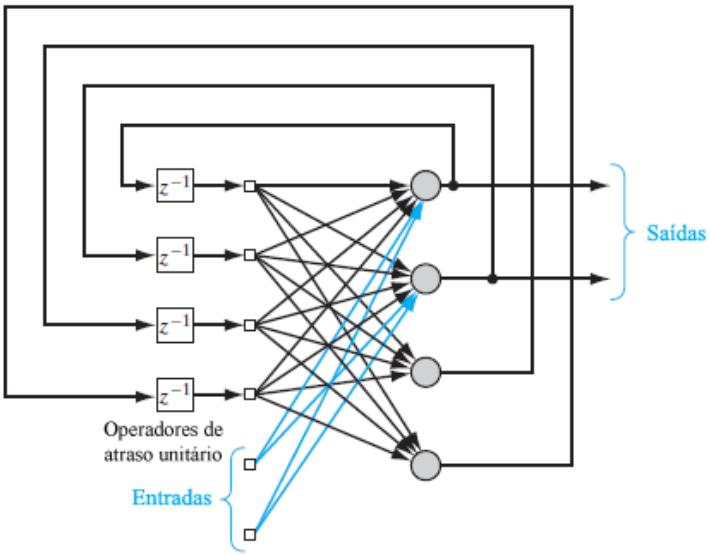
Fonte – Adaptado de Haykin (2009)

Os pesos das camadas de uma CNN são aprendidos a partir do treinamento, reduzindo o esforço de implementação de filtros, sendo uma das principais vantagens em comparação a outras redes neurais tradicionais.

Redes neurais recorrentes

A arquitetura mais simples de uma rede neural é conhecida como *Feed-Forward Neural Network* (FFNN), caracterizada por não haver ciclos entre as conexões dos neurônios. De acordo com o trabalho em Haykin (2009), Redes Neurais Recorrentes (do inglês, *Recurrent Neural Network* - RNN) se diferenciam de FFNNs no aspecto de haver pelo menos um ciclo de retroalimentação. Estes ciclos podem retroalimentar o próprio neurônio, ou seja, a saída de um neurônio é utilizada como entrada do mesmo neurônio, ou alimentar outros neurônios de uma mesma camada. A figura 8 apresenta uma RNN com neurônios escondidos e retroalimentação.

Figura 8 – Rede Neural Recorrente com neurônios escondidos



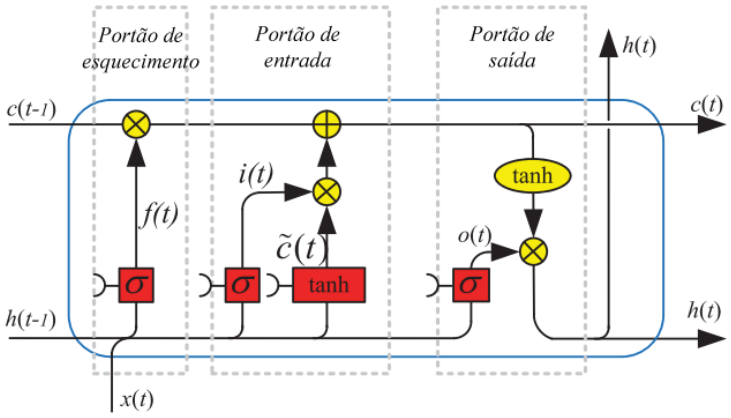
Fonte – Adaptado de Haykin (2009)

A presença de ciclos de retroalimentação traz um profundo impacto na capacidade de aprendizado da rede, permitindo que a dependência dos ciclos simule uma estrutura de memória de rede. O uso deste tipo de estrutura é favorável em aplicações relacionadas a sequências, listas e outras aplicações, como processamento de textos e imagens.

Um tipo específico de RNN denominado *Long Short-Term Memory* (LSTM) foi proposto originalmente em Hochreiter e Schmidhuber (1997) e suas versões mais sofisticadas frequentemente são utilizadas em aplicações de PLN por conseguirem processar cadeias longas de texto. As redes LSTM foram propostas com o objetivo de lidar com dependências de longo prazo e resolvem o problema de dissipação do gradiente, em que a atualização dos pesos praticamente não traz mais efeitos no treinamento da rede. Esse problema ocorre devido à multiplicação por valores menores a cada iteração, gerando um custo elevado de processamento, dada a necessidade de muitas iterações nas camadas mais iniciais para um ajuste relevante (HAYKIN, 2009).

A estrutura de uma rede LSTM é composta por blocos de memórias denominadas células, que armazenam as memórias de curto e longo prazo, e portões que controlam o fluxo de informações e o estado das células. O portão do esquecimento remove as informações que não são mais úteis no estado da célula, enquanto que o portão de entrada determina quais informações de curto prazo serão adicionadas ao estado da célula. Por fim, o portão de saída extrai as informações úteis do estado da célula, determinando o valor para o próximo estado (YU *et al.*, 2019). A figura 9 apresenta a arquitetura de uma rede LSTM.

Figura 9 – Arquitetura de uma rede LSTM



Fonte – Adaptado de Yu *et al.* (2019)

2.4 Comitê de classificadores

Comitê de classificadores é um conjunto de classificadores cujas decisões individuais são combinadas de alguma forma para classificar novos exemplos (DIETTERICH, 2000). Este conceito é também encontrado na literatura por outros nomes, como sistemas de classificação múltipla, mistura de especialistas, combinação de classificadores múltiplos e *ensembles* (POLIKAR, 2006). A ideia do comitê de classificadores é construir um modelo de predição que combina a força de uma coleção de modelos base mais simples (HASTIE *et al.*, 2009).

O conceito de comitê de classificadores foi inicialmente explorado por trabalhos como em Dasarathy e Sheela (1979), que discute o particionamento do espaço de características por meio do uso de dois ou mais classificadores. Já o trabalho em Hansen e Salamon (1990) demonstra que a performance de generalização de uma rede neural pode ser melhorada por meio da combinação de várias redes neurais artificiais. Por fim, o trabalho em Schapire (1990) prova que um classificador mais forte pode ser gerado pela combinação de classificadores fracos por meio do método de *boosting* (POLIKAR, 2006).

Dado este contexto, o trabalho em Lima (2004) cita como forte motivação para o uso de um comitê de classificadores o fato de que diferentes propostas de solução podem explorar diferentes aspectos relevantes de um problema, enquanto que muitas vezes uma única proposta de solução não é capaz de explorar todos os aspectos relevantes

simultaneamente. Nas subseções a seguir serão apresentados alguns dos principais métodos de combinação utilizados em comitês de classificadores.

2.4.1 Comitê de classificadores baseado em votação

O método de votação é uma das abordagens mais comuns para a combinação de classificadores e foi introduzida pelo trabalho em Hansen e Salamon (1990), que utilizou os resultados de modelos individuais baseados em redes neurais como entrada para a decisão final de classificação por votação. Este método consiste basicamente em um esquema consensual de decisão da classificação final baseado em votação majoritária, ou seja, o resultado apresentado pela maioria dos classificadores individuais será definido como classe final. O trabalho em Lima (2004) apresenta algumas possíveis variações deste método:

- **Unanimidade**: uma observação x pertence a uma classe k se, e somente se, todos os classificadores tomarem a mesma decisão
- **Unanimidade modificada**: uma observação x pertence a uma classe k se uma grande parcela dos classificadores tomarem esta decisão, sendo a margem de discordância um parâmetro a ser definido
- **Majoritário (ponderado ou não-ponderado)**: uma observação x pertence a uma classe k se mais que a metade dos classificadores tomarem esta decisão. A diferença entre este item e o anterior se dá apenas na margem de discordância aceitável
- **Pluralidade limitada**: uma observação x pertence a uma classe k se o número de classificadores que a rotulam desta maneira é maior que o número de classificadores que a rotulam em qualquer outra classe. Em um problema com apenas duas classes, seria equivalente ao voto majoritário

O método de votação pode utilizar os rótulos ou as probabilidades de cada classe geradas pelos classificadores individuais para predizer a classe final. Além disso, nos casos em que há um empate entre duas ou mais classes majoritárias, alguma estratégia deve ser adotada, como, por exemplo, atribuir a decisão final à classe mais frequente no conjunto de treinamento.

Ainda segundo o trabalho em Lima (2004), o método de votação possui a vantagem de ser simples, não requerendo qualquer conhecimento prévio do comportamento dos

classificadores individuais, necessitando apenas a contagem do número de classificadores e suas saídas individuais para a decisão final de classificação. Por outro lado, os pesos da decisão de todos os classificadores são iguais, independentemente do desempenho geral de cada classificador, o que pode piorar o desempenho do comitê, caso existam classificadores com desempenho global ruim.

2.4.2 Comitê de classificadores baseado em *stacking*

O conceito de *stacking* (ou *stacking generalization*) para a combinação de classificadores foi originado no trabalho em Wolpert (1992) e se refere a um esquema de alimentação de informação a partir de um conjunto de classificadores para um outro classificador que fará a predição final. Este método foi criado com o objetivo de reduzir a taxa de erro de um ou mais classificadores individuais, e, por isso, é recomendado que estes classificadores consigam abordar aspectos diferentes entre si.

O método de *stacking* pode ser separado em duas etapas: preparação dos dados (primeiro nível) e combinação dos dados (segundo nível). No primeiro nível, classificadores individuais de diferentes tipos (por exemplo, que façam uso de algoritmos de classificação diferentes, utilizem espaços dimensionais de características distintos, possuam como entrada amostras do conjunto de treinamento diferentes, etc) geram suas predições (que podem ser os rótulos ou as probabilidades das classes), que serão utilizadas como entrada para um classificador do próximo nível. Este segundo nível explora o conceito de meta-aprendizado (que substitui a etapa de votação descrita na subseção anterior), em que um meta-classificador fará a combinação das predições dos classificadores do primeiro nível e gerará a classificação final para todas as instâncias de entrada (LIMA, 2004).

Modelos que utilizam meta-aprendizado, como *stacking*, exigem aprendizado adicional para o combinador e por isso consomem mais tempo de treinamento (MAIMON; ROKACH, 2008). Por outro lado, quando construído da forma adequada, estes modelos costumam gerar bons resultados (tão bons quanto o melhor classificador individual) uma vez que reduz a taxa de erro dos classificadores do primeiro nível.

2.5 Conjunto de dados SetembroBR

O córpus SetembroBR é uma coletânea de dados de usuários do *Twitter* contendo conhecimento linguístico extraído do conteúdo textual de postagens (*tweets*), dados demográficos e informações relacionadas ao comportamento destes usuários na rede social, como a estrutura de suas redes de relacionamento e dados de atividades. Este recurso linguístico-computacional é composto de *tweets* em português com o foco em predição de transtornos de saúde mental do tipo depressão e ansiedade. O projeto inicial de construção do córpus foi apresentado em Santos, Funabashi e Paraboni (2020) e nas próximas seções será detalhada a versão final, conforme apresentado em Santos, Oliveira e Paraboni (2023).

2.5.1 Visão geral do córpus

O córpus SetembroBR contém dados de usuários diagnosticados com depressão e ansiedade selecionados a partir de autorrelatos em que há indicação explícita do momento do diagnóstico (e.g., "Mês passado o psiquiatra me diagnosticou com depressão") e que indicam um suporte clínico, ou seja, relacionados a diagnósticos por especialistas da área da saúde e/ou tratamentos que envolvem principalmente o uso de medicação específica para estes tipos de transtornos. O conteúdo textual disponível no córpus é, no entanto, restrito aos *tweets* dos usuários (i.e., excluindo mensagens escritas por outros usuários, conhecido como *retweets*) anteriores ao evento do diagnóstico, selecionados a partir de uma inspeção manual. O quadro 1 ilustra por meio de um marcador [end] a porção de *tweets* disponível no córpus após inspeção manual que considerou a postagem de diagnóstico identificada pelo marcador [msg] como ponto referencial.

O método de seleção baseada em autorrelatos permite que apenas usuários da classe positiva (diagnosticado) sejam identificados, uma vez que a ausência de relatos sobre eventos relacionados a estes transtornos não implica necessariamente a falta de um diagnóstico. Nestes casos, utiliza-se um grupo de controle como classe negativa de proporções superiores às da classe positiva. A classe de controle do córpus SetembroBR consiste em usuários selecionados aleatoriamente, que não relataram explicitamente um evento de diagnóstico, ou seja, que não atendem os mesmos critérios de seleção de usuários diagnosticados. Além disso, a seleção inicial de usuários de controle considerou um mínimo

Quadro 1 – *Timeline* de um usuário com marcador [end] indicando o término da porção de dados a ser considerada na predição de depressão, em que todos os *tweets* abaixo do ponto [end] são descartados

Data	**Marcador**	**Texto**
Mon March 25		Deixei meu celular em casa e agora a bateria está morta.
Tue March 26		Eu assisti esse filme duas vezes no ano passado.
Thu March 28	[end]	tão feliz que finalmente comprei meus óculos novos LOL
Mon April 1		Vou dormir agora. Amanhã é um grande dia.
Wed April 3		@usuário você nunca me contou isso.
Mon April 8		Pensando em ligar para ela de novo hoje à noite...
Fri May 3		Oiiii! como você está?
Sun May 5	[msg]	Mês passado o psiquiatra me diagnosticou com depressão :(

Fonte – Santos, Oliveira e Paraboni (2023)

de 1000 *tweets* escritos em português (posteriormente, na construção da versão final do córpus, as *timelines* destes usuários foram truncadas para ficarem compatíveis com usuários diagnosticados) e usuários que não possuíssem mais do que 10 mil seguidores ou amigos.

O córpus é composto por usuários diagnosticados pareados com sete usuários de controle aleatórios de mesmo gênero estimado, mesmo número de *tweets* e intervalo de postagem aproximado (no limite de 2 meses a mais ou a menos em relação à data de postagem mediana para evitar grandes discrepâncias de vocabulário). A relação de sete usuários de controle para um usuário diagnosticado é sugerido pelo trabalho em Coppersmith *et al.* (2015) e seguido pela tarefa compartilhada no eRisk-2018 (LOSADA; CRESTANI; PARAPAR, 2018). A tabela 1 apresenta estatísticas descritivas dos subconjuntos de depressão e ansiedade. As colunas C/D mostram as relações de número de usuários de Controle/ Diagnosticado, e têm como objetivo ilustrar o balanceamento de cada conjunto de dados (SANTOS; OLIVEIRA; PARABONI, 2023).

Tabela 1 – Estatísticas descritivas do córpus SetembroBR

	Depressão				**Ansiedade**			
Estatística	**Diagnosticado**	**Controle**	**Total**	**C/D**	**Diagnosticado**	**Controle**	**Total**	**C/D**
Todos os usuários	1684	11788	13472	7,0	2219	15533	17752	7,0
Feminino	76,7%	76,7%	76,7%	7,0	78,8%	78,8%	78,8%	7,0
Tweets (milhões)	2,43	16,99	19,42	7,0	3,43	23,98	27,41	7,0
Palavras (milhões)	29,32	201,94	231,26	6,9	42,24	281,51	323,75	6,7
Tweets/usuários	1441	1441	1441	1,0	1543	1543	1543	1,0
Palavras/tweets	12,08	11,88	11,91	1,0	12,33	11,74	11,81	1,0
Dias (média)	492	553	546	1,1	524	540	538	1,0
Dias (máximo)	3712	4165	4165	1,1	4088	4211	4211	1,0

Fonte – Adaptado de Santos, Oliveira e Paraboni (2023)

A partir das estatísticas apresentadas é possível observar que as proporções entre as classes estão próximas a sete para o gênero feminino e para o número total de usuários,

tweets e palavras, conforme sugerido em trabalhos anteriores. As estatísticas também revelam a quantidade de dias médio e máximo entre a mensagem mais antiga do usuário e a mais recente (anterior ao autorrelato de diagnóstico ou tratamento), indicando que é possível explorar conhecimento que auxilie na predição de depressão e ansiedade com ampla antecedência (em média, 508 dias) (SANTOS; OLIVEIRA; PARABONI, 2023).

O córpus foi originalmente dividido em conjuntos de treinamento e teste, objetivando o uso em modelos de classificação baseados em técnicas de AM Supervisionado, como proposto neste trabalho. A divisão foi realizada aleatoriamente, na proporção de 80% para treino e 20% para teste, de forma a não haver intersecção entre os conjuntos (ou seja, cada usuário é integralmente parte do treinamento ou teste, mas nunca de ambos). Além disso, o balanceamento entre as classes presente no conjunto original foi preservado, ou seja, para cada usuário da classe diagnosticado, existem sete contrapartidas da classe controle em ambos os conjuntos de treino e teste. A tabela 2 apresenta a quantidade de usuários de cada classe nos conjuntos de treino e teste.

Tabela 2 – Quantidade de usuários após divisão do córpus em treino e teste

	Depressão			**Ansiedade**		
	Diagnosticado	**Controle**	**Total**	**Diagnosticado**	**Controle**	**Total**
Treino	1347	9429	10776	1775	12425	14200
Teste	337	2359	2696	444	3108	3552
Total	1684	11788	13472	2219	15533	17752

Fonte – Rafael Lage de Oliveira, 2023

2.5.2 Conhecimento extra-linguístico

Além dos dados textuais, representados pela união de todas as postagens dos usuários em suas *timelines*, o córpus SetembroBR contém dados não-textuais que expressam atributos dos usuários e seus comportamentos na rede social. Os atributos dos usuários são representados pelo gênero (masculino ou feminino) estimado para cada usuário, que corresponde à categoria de característica demográfica do conjunto de dados. A porção comportamental do córpus é composta por dados relacionados à estrutura da rede, interações com outros indivíduos da rede e atividade dos usuários, e estas características serão detalhadas ao longo desta seção. O quadro 2 sumariza as características extra-linguísticas presentes no córpus SetembroBR.

Quadro 2 – Características extra-linguísticas do córpus SetembroBR

Categoria	**Característica**
Rede	Lista de amigos Lista de seguidores Número de amigos Número de seguidores
Interação	Lista de menções Lista de top menções Número de menções
Atividade	Número de atualizações de *status* Lista de data/hora dos *tweets* do usuário Quantidade total de *tweets* Quantidade de palavras Quantidade de *tweets* a cada hora do dia Quantidade de *tweets* postados entre 21h e 6h
Demográfico	Gênero (masculino e feminino)

Fonte – Rafael Lage de Oliveira, 2023

Os amigos representam conexões que são seguidos pelos usuários a classificar do córpus, sendo suficiente uma ação por parte do seguidor para a formação de uma conexão. De forma semelhante, no sentido oposto, os seguidores representam conexões de rede que seguem usuários a classificar do córpus. As menções caracterizam outros indivíduos da rede social mencionados em postagens por um usuário do córpus, precedido pelo símbolo "@". A lista de top menções corresponde aos 20 indivíduos mais mencionados por usuários diagnosticados ou de controle em postagens relacionadas a depressão ou ansiedade. Estas listas foram originalmente complementadas por usuários comuns em caso de necessidade até que atingissem o total de 20 menções, sendo colocado um separador para a identificação das duas partes da lista. As listas foram originalmente ordenadas no córpus de forma decrescente por usuários mais mencionados em postagens sobre depressão ou ansiedade, ou seja, o primeiro usuário da lista corresponde àquele que mais foi mencionado sobre estes temas. No entanto, não há a informação no córpus sobre a quantidade de menções para cada uma destas conexões.

A tabela 3 apresenta estatísticas comportamentais que envolvem as conexões dos usuários a classificar do córpus e que estão relacionadas tanto à estrutura da rede de relacionamentos quanto às interações por meio de menções. Estas estatísticas correspondem às médias dos números de amigos, seguidores e menções identificadas originalmente nas listas de cada usuário na data de incorporação deste ao córpus. É possível notar que a proporção entre as classes para o número de amigos, seguidores e menções é próxima a

1, sugerindo que explorar o atributo de quantidade de forma independente pode não ser eficaz para a diferenciação das duas classes.

Tabela 3 – Médias de amigos, seguidores e menções no córpus SetembroBR

	Depressão				Ansiedade			
Média	**Diagnosticado**	**Controle**	**Total**	**C/D**	**Diagnosticado**	**Controle**	**Total**	**C/D**
Amigos	659	710	704	1,1	678	729	722	1,0
Seguidores	777	945	924	1,2	810	975	954	1,2
Menções	125	122	122	1,0	115	114	114	1,0

Fonte – Adaptado de Santos, Oliveira e Paraboni (2023)

Os usuários a classificar do córpus foram anonimizados, assim como suas listas de amigos, seguidores e menções, de modo a garantir a confidencialidade dos dados. No entanto, identificadores coincidentes nas listas de relacionamento de dois usuários indicam que estes possuem amigos ou seguidores em comum ou mencionam uma mesma conexão em suas respectivas postagens. Desta forma, é possível identificar o número de usuários únicos entre todas as listas de usuários, tanto para o subconjunto de depressão quanto para o subconjunto de ansiedade, o qual está sumarizado na tabela 4 para cada uma das redes.

Tabela 4 – Total de usuários únicos que formam conexões de rede

	Depressão	**Ansiedade**
Amigos	3.678.662	4.448.835
Seguidores	6.236.404	7.829.049
Menções	910.180	1.070.567
Top Menções	16.651	20.191

Fonte – Rafael Lage de Oliveira, 2023

Com relação aos atributos de atividade dos usuários, foram coletados o número de atualizações de *status* do perfil do usuário e dados temporais representados pela data e horário de cada *tweet*, quantidade de *tweets* postados a cada hora do dia e, mais especificamente, *tweets* postados entre 21h e 6h, podendo contribuir para avaliação do "índice de insônia" (SHRESTHA; SERRA; SPEZZANO, 2020; TSUGAWA *et al.*, 2015; CHOUDHURY *et al.*, 2013). A característica *hour_18*, por exemplo, representa a quantidade de postagens feitas por um usuário no horário entre 18 e 19 horas. A tabela 5 exemplifica um comportamento mais noturno do usuário *D_1*, que realizou postagens no período da madrugada, enquanto que o usuário *CD_1* revela um comportamento mais diurno, com postagens no período da tarde.

Tabela 5 – Características de horário de postagens

Usuário	**hour_00**	**hour_01**	**hour_02**	**hour_15**	**hour_16**	**hour_17**
D_1	0	10	5	0	0	0
CD_1	0	0	0	4	7	2

Fonte – Rafael Lage de Oliveira, 2023

Adicionalmente, foram computadas características que estão diretamente relacionadas ao conteúdo textual mas que expressam um comportamento dos usuários na forma de atividade de postagem, como o número de palavras e a quantidade de *tweets* válidos de cada usuário, que permitem derivar outros tipos de características de atividade (e.g., a frequência de postagem diária ou semanal, a média de palavras por *tweet*, etc).

3 Revisão bibliográfica

Este capítulo apresenta uma revisão bibliográfica sistemática sobre métodos e técnicas utilizados para a predição de transtornos de saúde mental em redes sociais a partir de modelos computacionais. Revisões semelhantes foram identificadas em trabalhos anteriores, como em Wongkoblap, Vadillo e Curcin (2017), que identificou 48 artigos a partir de uma revisão sistemática de publicações feitas entre 2010 e 2017. O trabalho realizado em Chancellor e Choudhury (2020) identificou 75 trabalhos, publicados entre 2013 e 2018, que abordam os transtornos de saúde mental mais comuns e as sintomatologias relacionadas, como estresse, instabilidade emocional e automutilação.

Diferentemente das revisões sistemáticas anteriores realizadas sobre este tema de pesquisa, este trabalho possui um enfoque em atributos sociais e comportamentais, que envolvem principalmente a estrutura da rede de relacionamentos e as interações entre usuários nas redes sociais, de forma a identificar quais características dessa natureza são exploradas para o problema em questão e quais os métodos e técnicas são utilizados para a extração de tais atributos e para a tarefa de classificação. Os estudos considerados foram selecionados conforme metodologia descrita na seção 3.1 e as respostas para as questões de pesquisa que esta revisão sistemática visa responder são apresentadas na seção 3.2. Os resultados detalhados são descritos na seção 3.3. Trabalhos que foram identificados após a revisão sistemática serão apresentados na seção 3.4.

3.1 Metodologia da revisão sistemática

Inicialmente, na etapa de planejamento, foram definidas as principais questões de pesquisa que esta revisão sistemática se propõe responder. São elas:

A1. Quais são os principais transtornos de saúde mental explorados na literatura para a classificação automática de usuários em redes sociais?

A2. Em quais idiomas e em quais plataformas de redes sociais digitais os trabalhos mais recentes buscam classificar usuários com transtornos de saúde mental?

A3. Quais são as características relacionadas à atividade, comportamento, interação e rede dos usuários extraídas das redes sociais que podem contribuir para a identificação de usuários com transtornos de saúde mental?

Ainda na etapa de planejamento, foram definidos os tipos de trabalhos e a seleção de fontes onde estes seriam buscados. Estudos em inglês e português que utilizem técnicas de AM, PLN ou análises estatísticas para a classificação de transtornos de saúde mental em plataformas de redes sociais foram definidos como população principal de trabalhos a serem selecionados. Além disso, trabalhos que tratem de homofilia em redes sociais também foram considerados. Alguns trabalhos com estas características, como em Santos, Funabashi e Paraboni (2020), Tsugawa *et al.* (2015) e Vedula e Parthasarathy (2017), foram definidos como controle para a validação da qualidade da busca, após uma análise exploratória inicial. Para a seleção das fontes, foram considerados repositórios de artigos especializados nas áreas de estudo em questão, incluindo eventos e periódicos, preferencialmente com melhores classificações segundo critérios da CAPES. Para artigos, as fontes de busca incluem o SCOPUS e bibliotecas digitais da ACM, IEEE e ACL Anthology, enquanto que para teses e dissertações foram consideradas a Biblioteca Digital da Universidade de São Paulo e o Catálogo de Teses e Dissertações da CAPES.

A condução da revisão sistemática foi realizada a partir da busca de termos específicos nos campos de Título, Resumo e Palavras-chave das fontes definidas na etapa de planejamento. Duas categorias de termos foram definidas para as buscas: *Saúde Mental* e *Rede Social*. Além de considerar palavras-chave genéricas relacionadas a estas categorias, foram incluídos também os termos específicos mais comuns associados ao tema de pesquisa. O quadro 3 mostra as palavras-chave e a estratégia utilizada para a busca de trabalhos na literatura. Para as plataformas que permitem a busca por periódicos ou conferências de diversas áreas de aplicação, como o SCOPUS, foram filtrados apenas aqueles que possuam a área da *computação* como área de interesse.

Quadro 3 – Palavras-chave para busca de trabalhos em fontes especializadas

Categoria	**Palavras-chave**
Saúde Mental (1)	*mental OR depression*
Rede Social (2)	*social media OR social network*
Estratégia de busca	(1) *AND* (2)

Fonte – Rafael Lage de Oliveira, 2023

Para a seleção dos trabalhos na etapa de condução, foram aplicados critérios de inclusão e exclusão definidos na etapa de planejamento da revisão sistemática. Para a inclusão de trabalhos, foram considerados aqueles que tratem especificamente da identificação de transtornos de saúde mental em redes sociais a partir de métodos computacionais

ou estatísticos e também trabalhos que tratem sobre homofilia em redes sociais. Como critérios iniciais de exclusão, foram desconsiderados trabalhos que não possuíssem resumo, que não tratassem especificamente da identificação de transtornos de saúde mental em redes sociais a partir de métodos computacionais ou estatísticos e que foram publicados em um período anterior ao ano de 2015. Além disso, foram excluídos os trabalhos que não apresentassem na seção resumo os detalhes dos métodos e técnicas utilizados, que não estivessem relacionados aos principais transtornos de saúde mental e suas sintomatologias associadas, e que não utilizassem técnicas computacionais, como AM ou métodos estatísticos, com exceção dos trabalhos que abordem a homofilia em redes sociais.

As buscas pelas palavras-chave nas fontes selecionadas foram realizadas em junho de 2020 e resultaram na seleção de 1026 trabalhos, dentre os quais não foram identificadas teses ou dissertações relevantes para o trabalho atual. Foram incluídos ainda 5 trabalhos apresentados no *workshop CLPsych*, que não foram retornados nas buscas e que são considerados relevantes para o tema em questão. Após a aplicação dos critérios de inclusão e exclusão, foram obtidos 239 trabalhos, conforme quadro 4.

Quadro 4 – Resultado da busca de trabalhos e aplicação dos critérios de inclusão e exclusão

Fonte	**Data da busca**	**Trabalhos retornados**	**Excluídos**	**Incluídos**
IEEE	13/06/2020	200	136	64
ACM	13/06/2020	158	126	32
SCOPUS	18/06/2020	629	515	114
ACL	22/06/2020	39	15	24
CLPsych	22/06/2020	5	0	5

Fonte – Rafael Lage de Oliveira, 2023

Para a avaliação da qualidade dos estudos selecionados após a aplicação dos critérios de inclusão e exclusão, foram aplicados critérios de qualidade definidos na etapa de planejamento da revisão sistemática. A aplicação dos critérios de qualidade consiste em verificar se alguns conceitos considerados relevantes para o trabalho atual são encontrados nos estudos selecionados, atribuindo pontuações após esta verificação. Em relação às características de usuários em redes sociais, foram verificados se os trabalhos abordam características de rede, homofilia ou interação entre usuários das redes sociais, além de características de comportamento ou atividade dos usuários. Com respeito aos transtornos de saúde mental, foram verificados se os trabalhos consideram a depressão ou ansiedade como tema central. Além disso, receberam uma pontuação maior os trabalhos que abordem as principais redes sociais encontradas na literatura: *Facebook*, *Twitter*, *Reddit* e *Instagram*.

Por fim, foram considerados mais relevantes os trabalhos que utilizam técnicas de AM. Os trabalhos considerados menos relevantes, segundo cada um dos critérios de qualidade estabelecidos, obtiveram pontuação 0 para o critério em questão, enquanto que os trabalhos mais relevantes obtiveram pontuação 1. O critério relacionado à presença de características de rede, homofilia ou interação entre usuários, recebeu pontuação 2 para os trabalhos mais relevantes, devido à importância dada a este critério no contexto do trabalho atual. O quadro 5 mostra os critérios de qualidade utilizados e a pontuação associada a cada um dos critérios.

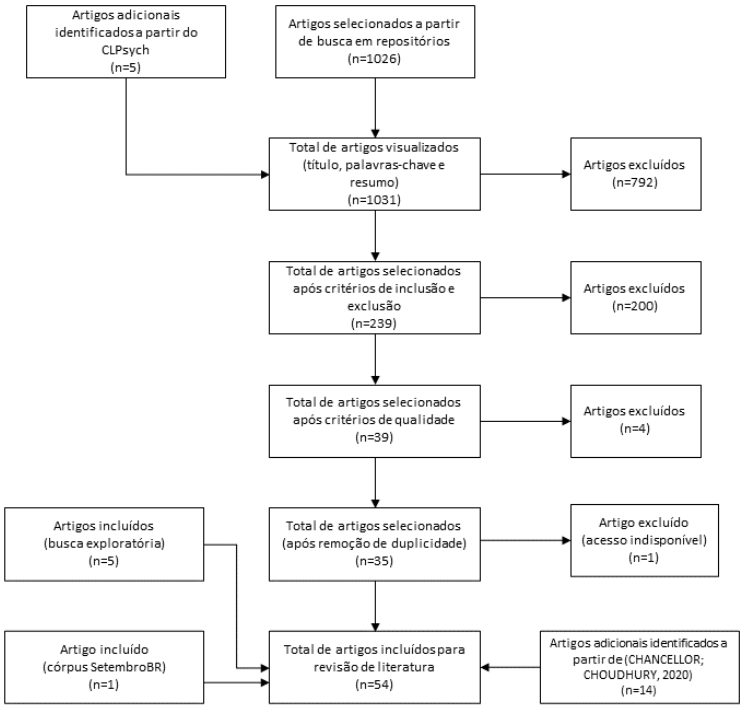
Quadro 5 – Critérios de qualidade e pontuação

Critério de Qualidade	**Pontuação**
Rede, homofilia ou interação entre usuários	[0;2]
Comportamento ou atividade	[0;1]
Depressão ou Ansiedade	[0;1]
Facebook, *Twitter*, *Reddit*, *Instagram*	[0;1]
Aprendizado de Máquina	[0;1]

Fonte – Rafael Lage de Oliveira, 2023

Após a aplicação dos critérios de qualidade, foram selecionados 39 trabalhos, sendo que quatro destes estavam duplicados e um estava indisponível para acesso, os quais foram excluídos. Foram adicionados ainda à seleção final um artigo relacionado ao córpus SetembroBR, objeto de estudo deste trabalho, cinco artigos identificados como relevantes a partir de uma busca exploratória inicial, e 14 artigos citados pela revisão sistemática em Chancellor e Choudhury (2020). Estes últimos artigos foram considerados relevantes para o trabalho atual por abordarem características de rede, homofilia ou interação entre usuários, porém não foram encontrados na busca realizada por terem sido publicados no período anterior a 2015 ou em veículos não considerados pelas plataformas utilizadas. A figura 10 ilustra o processo de revisão sistemática por meio do PRISMA (Principais Itens para Relatar Revisões Sistemáticas e Meta-análises) (LIBERATI *et al.*, 2009), totalizando 54 trabalhos selecionados.

Figura 10 – PRISMA - Principais Itens para Relatar Revisões Sistemáticas e Meta-análises



Fonte – Rafael Lage de Oliveira, 2023

3.2 Síntese dos dados

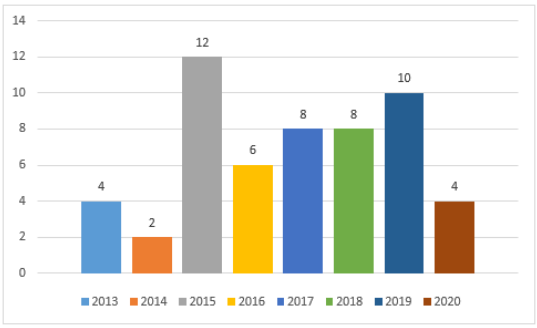
Esta seção apresenta os resultados sintetizados dos trabalhos selecionados na etapa de condução da revisão sistemática, conforme descrito na seção 3.1.

3.2.1 Visão geral dos resultados

A figura 11 mostra os anos de publicação dos trabalhos selecionados. É possível observar uma concentração baixa de trabalhos publicados em 2013 e 2014 devido à seleção criteriosa realizada para publicações anteriores a 2015, além de ser um período onde o tema de pesquisa estava em fase inicial de exploração (CHANCELLOR; CHOUDHURY, 2020). Em contrapartida, é possível verificar que existe uma concentração maior de publicações no ano de 2015 em comparação aos demais anos, possivelmente por influência do *II Workshop on Computational Linguistics and Clinical Psychology (CLPsych)*, realizado em 2015, cuja tarefa principal do evento era classificar usuários do *Twitter* que tenham o diagnóstico de depressão ou Transtorno do Estresse Pós-Traumático (TEPT). Entre 2016 e 2019 (último

ano completo anterior à data de execução da revisão sistemática), nota-se uma tendência de aumento de publicações, sugerindo que o tema deste trabalho está ainda em ascensão.

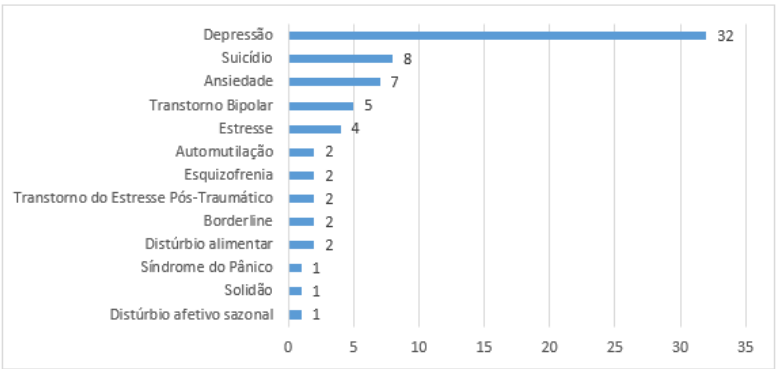
Figura 11 – Volume de trabalhos selecionados por ano de publicação



Fonte – Rafael Lage de Oliveira, 2023

A figura 12 ilustra os transtornos de saúde mental abordados pelos trabalhos selecionados e também as sintomatologias associadas, como suicídio, automutilação, estresse e solidão. É possível observar que depressão concentra a maior parte dos trabalhos, conforme observado também no trabalho em Chancellor e Choudhury (2020). Este resultado é consistente com as estimativas da OMS que indicam que a depressão será a doença mais comum até o ano de 2030, além de ser o principal motivo de mortes por suicídio no mundo (WORLD HEALTH ORGANIZATION, 2017).

Figura 12 – Transtornos de saúde mental e sintomatologias associadas

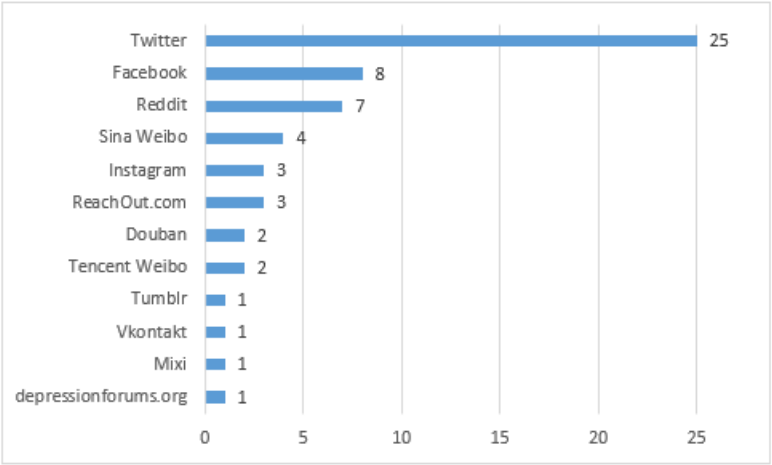


Fonte – Rafael Lage de Oliveira, 2023

As figuras 13 e 14 mostram, respectivamente, as plataformas de rede social e os idiomas abordados pelos trabalhos selecionados. Assim como descrito pelo trabalho em Chancellor e Choudhury (2020), o *Twitter* é a plataforma de rede social mais adotada para estudos de classificação de transtornos de saúde mental, embora não seja a mídia social mais popular. Segundo o trabalho em Chai *et al.* (2019), a maioria dos estudos fazem predições apenas utilizando o conteúdos das postagens, o que pode explicar o uso

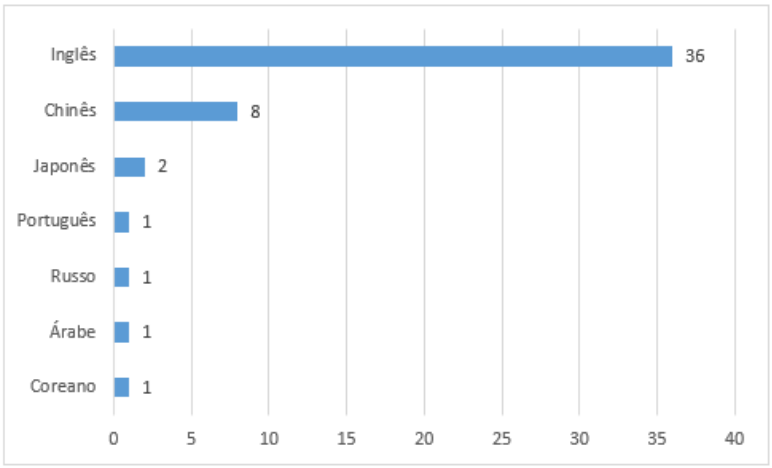
massivo do *Twitter*, uma vez que a natureza da plataforma privilegia o compartilhamento de conteúdo, principalmente na forma de textos. É possível observar também que o inglês é o idioma mais explorado nos trabalhos selecionados, embora haja um crescimento de publicações que exploram o idioma chinês.

Figura 13 – Plataformas de rede social nos trabalhos selecionados



Fonte – Rafael Lage de Oliveira, 2023

Figura 14 – Idioma nos trabalhos selecionados



Fonte – Rafael Lage de Oliveira, 2023

3.2.2 Características de comportamento em redes sociais

Como resultado da revisão sistemática, foram identificados trabalhos que exploram características comportamentais, de rede e relacionamento entre usuários de redes sociais para a predição de transtornos de saúde mental, quase sempre em conjunto com carac-

terísticas de conteúdo produzido por esses usuários, principalmente na forma de textos. Por exemplo, o trabalho em Sinha *et al.* (2019) destaca que muitas vezes a informação na forma de texto pode ser esparsa, conter ruídos e faz-se necessário explorar outras características, como metadados e interações entre usuários, dada a complexidade do problema a ser modelado.

A nomenclatura para designar atributos que envolvem o comportamento dos indivíduos em redes sociais ainda não possui um padrão bem definido na literatura. Os trabalhos selecionados por esta revisão sistemática utilizam termos mais genéricos, como atributos sociais, metadados ou metavariáveis, engajamento social, padrões de comportamento, comportamento social, atributos de interação e características de padrões de vida, além de termos mais específicos, como estrutura de rede, propriedades diáticas e características de rede egocêntrica. Neste trabalho, será adotada a mesma nomenclatura utilizada em Chancellor e Choudhury (2020), que se refere a essas características como atributos de comportamento e os divide em quatro grupos: (1) atividade, (2) rede, (3) interação e (4) específica do domínio. Atributos de atividade são definidos por ações realizadas pelos indivíduos na rede social, normalmente relacionadas às características principais da plataforma (e.g., atividade de postagem) ou a estatísticas de utilização (e.g., frequência de acesso), e que não possuem algum tipo de interação direta com outros usuários. Atributos de rede estão relacionados à estrutura de redes ou grafos que pode ser construída a partir de diversos tipos de conexões, como relações de amizade e seguidores, menções a outros usuários, respostas a postagens e *retweets*. Atributos de interação são caracterizados por possuir algum tipo de relação com outros usuários da rede, seja uma relação direta, como a quantidade de menções recebidas de um amigo, ou indireta, como o número de comunidades a que um usuário pertence. Por fim, atributos específicos de domínio expressam características que não podem ser extraídas da maior parte das plataformas de rede social, limitando-se a apenas algumas em particular, devido à sua natureza e propriedades específicas. O quadro 6 apresenta um resumo das características de comportamento identificadas nos trabalhos selecionados por esta revisão sistemática.

Um dos primeiros esforços para identificar indivíduos com transtornos de saúde mental em redes sociais de forma automática é encontrado no trabalho em Choudhury *et al.* (2013), onde são explorados atributos de conteúdo, como o estilo linguístico das postagens, o uso de linguagem específica, como antidepressivos e termos que expressam sentimento, e também atributos de comportamento do usuário para a detecção de depressão. Dentre

Quadro 6 – Características de comportamento nos trabalhos selecionados

Categoria	Sub-categoria	Característica
(1) Atividade	Postagem	Frequência de postagem
		Horário da postagem (índice de insônia)
		Tipo do conteúdo da postagem
		Tamanho da postagem
	Diversos	Frequência de acesso à rede social em dias
		Horas dispendidas diariamente na rede social
		Nº de páginas curtidas
		Nº de tags de localização geográfica
		Nº de eventos de interesse
(2) Rede	Estrutura	Nº de nós
		Nº de arestas
		Nº de componentes conectados
		Diâmetro da rede
		Densidade da rede
		Tamanho da rede em até dois níveis
		Excentricidade
		Assortatividade
		Coeficiente de agrupamento
		Coeficiente de propagação
		Média de tamanho do caminho mais curto
		Laços fortes e fracos
		Fração de laços negativos
		PageRank
		Embeddings de nós
(3) Interação	Seguidores e amigos	Nº de seguidores
		Nº de seguidores em comum
		Nº de amigos
		Nº de requisições de amizade pendentes
		Laços unidirecionais e bidirecionais
	Grupos e comunidades	Nº de grupos/comunidades a que um usuário pertence
		Nº de grupos como administrador
	Curtidas e favoritos	Nº de curtidas a perfis de usuários
		Nº de curtidas a postagens
		Diferença de votos (atribuições positivas vs negativas)
		Nº de marcação de usuários como favoritos
		Nº de marcação de postagens como favoritas
	Menções	Nº de menções por usuário
		Nº de automenções
		Nº de usuários únicos mencionados
		Nº de menções frequentes
		Proporção de postagens contendo menções
		Nº de menções recebidas de um amigo
	Respostas e comentários	Nº de publicações sem respostas
		Conteúdo dos comentários de amigos
		Origem dos comentários (se feito por moderador ou amigo)
		Tempo até o primeiro comentário a uma postagem
		Tempo desde o último comentário
		Tamanho da resposta
		Nº de comentários recebidos por uma postagem
		Nº de respostas recebidas de um amigo
		Frequência de comentários recebidos
		Proporção de postagem do tipo “resposta”
	Retweets	Fração de *retweets*
		Nº de *retweets* de uma postagem
		Nº de *retweets* recebidos de um amigo
	Diversos	Entropia
		Reciprocidade
		Taxa de prestígio
(4) Específica de domínio	Diversos	Compartilhamento de tópicos de tendência
		Marcação de outros usuários em fotos
		Fração de postagens sem legendas
		Investimento e intensidade no *Instagram*

Fonte – Rafael Lage de Oliveira, 2023

os atributos de atividade, destaca-se o volume de postagens por dia e o horário de cada postagem. Muitos atributos de interação entre usuários foram extraídos, como a proporção de postagens do tipo "resposta", a fração de *retweets*, o número de seguidores e amigos, a reciprocidade, que mede a quantidade de vezes que dois usuários trocam respostas a uma postagem inicial, e a taxa de prestígio, que mede o número de respostas recebidas por um usuário em comparação ao número de respostas recebidas por outro usuário da rede de relacionamento, onde ambos possuem um histórico de interação através de respostas a postagens. Com relação aos atributos de rede, foram exploradas as propriedades de uma rede baseada nas respostas entre usuários, como densidade do grafo, coeficiente de agrupamento, tamanho da vizinhança em até dois níveis e número de componentes conectados.

Segundo o trabalho desenvolvido em Park *et al.* (2013), o estudo em questão foi a primeira tentativa de identificar uma associação entre características sociais no *Facebook* e sintomas depressivos de usuários. Neste estudo, foram consideradas características relacionadas à atividade dos usuários, como o número de páginas curtidas por um usuário, o número de marcações do tipo *tags* de localização física e o número de eventos no qual um usuário possui interesse em participação. Foram exploradas também características de interação, como o número de grupos a que um usuário pertence e a quantidade de grupos em que ele é administrador, o número de amigos na rede social e a quantidade de requisições de amizade pendentes.

Com o objetivo de detectar depressão em usuários de redes sociais, o trabalho em Coppersmith, Dredze e Harman (2014) utilizou um método semiautomático para identificar usuários diagnosticados a partir dos textos. Apesar do enfoque em características de conteúdo textuais, o estudo também extraiu informações de comportamento relacionadas à atividade dos usuários e suas interações na rede. Para os atributos relacionados à atividade, foram utilizados o horário e a frequência das postagens, enquanto que para os atributos de interação foram explorados as menções e seus diversos usos, como a proporção de postagens contendo menções, o volume total de menções por usuário, a quantidade de automenções, o número de usuários únicos mencionados e o número de usuários mencionados pelo menos três vezes. Semelhantemente, o trabalho em Lin *et al.* (2014) combinou características de conteúdo com atributos de comportamento a partir de uma rede neural convolucional para a detecção automática de estresse, sintomatologia frequentemente relacionada a transtornos de saúde mental. Foram consideradas características de atividade, como o

horário de postagem e o tipo do conteúdo (por exemplo, imagens, postagens originais, informações de compartilhamento, etc), além de atributos de interação, como a quantidade de comentários, compartilhamentos e curtidas de uma postagem, e também as menções, respostas e compartilhamentos de postagens de usuários da lista de amigos na rede social.

O trabalho em Masuda, Kurahashi e Onari (2013) demonstrou a importância da homofilia como característica para a ideação suicida, sendo um dos pioneiros a analisar este conceito como forma de contágio em redes sociais. Para sustentar essas análises, foram consideradas outras variáveis de comportamento, como o número de comunidades a que um indivíduo faz parte, o número de amigos na rede social e o coeficiente de agrupamento local. De semelhante forma, o trabalho em Wang, Zhang e Sun (2013) explorou o princípio da homofilia em usuários de redes sociais diagnosticados com depressão. Para este fim, foram consideradas características relacionadas à força dos laços e às interações entre usuários, como o número de seguidores, número de menções e comentários entre eles, tempo do último comentário e número de seguidores em comum. Além disso, um coeficiente de propagação foi utilizado como forma de medir a influência dos sintomas de depressão em outros membros da rede de relacionamento, o que está diretamente relacionado ao princípio da homofilia.

Durante o período de 2015 a 2020, outros trabalhos foram publicados no âmbito de predição de transtornos de saúde mental em redes sociais, explorando características comportamentais dos usuários e seus relacionamentos na rede. Um resumo geral destes trabalhos será abordado a seguir.

Em termos de características relacionadas à atividade do usuário na rede social, alguns trabalhos optaram por utilizar a frequência e o volume de postagens (SINHA *et al.*, 2019; CACHEDA *et al.*, 2019; FRAGA; SILVA; MURAI, 2018; VEDULA; PARTHASARATHY, 2017; WANG *et al.*, 2017; SHEN *et al.*, 2017; SARAVIA *et al.*, 2016; CHANG; SARAVIA; CHEN, 2016; CHOUDHURY *et al.*, 2016; XU; ZHANG, 2016a; PREOŢIUC-PIETRO *et al.*, 2015; PARK *et al.*, 2015; TSUGAWA *et al.*, 2015; CHOUDHURY, 2015; MCMANUS *et al.*, 2015; WU *et al.*, 2020; DUTTA; CHOUDHURY, 2020), o horário de postagem ou índice de insônia (CACHEDA *et al.*, 2019; WONGKOBLAP; VADILLO; CURCIN, 2018; LIN *et al.*, 2017; SHEN *et al.*, 2017; PARK *et al.*, 2015; TSUGAWA *et al.*, 2015; MCMANUS *et al.*, 2015; WU *et al.*, 2020), o tipo do conteúdo (por exemplo, imagens, postagens originais, informações de compartilhamento, etc) (LIN *et al.*, 2017; WANG *et al.*, 2017), o tamanho da postagem (CHOUDHURY *et al.*, 2016; KUMAR *et al.*,

2015), a quantidade de localizações geográficas (LARSEN *et al.*, 2015) e o tempo gasto na rede social, seja em dias ativos (XU; ZHANG, 2016a) ou horas dispendidas diariamente (LUP; TRUB; ROSENTHAL, 2015).

A diversidade de opções de interação em redes sociais permite que haja a exploração de uma grande variedade de características deste tipo nos trabalhos selecionados. Muitas destas características estão relacionadas a comentários e respostas a outras postagens, menções a outros usuários, citações a outras postagens e republicações. Estas características podem ser expressas a partir da frequência, quantidade e tamanho dos comentários, tempo até a primeira resposta a uma publicação, reciprocidade, ou seja, o número de vezes em que usuários trocam respostas a partir de uma postagem, quantidade de menções únicas ou menções frequentes, origem dos comentários (por exemplo, se feito por um moderador de grupo ou amigo), taxa de prestígio (ou seja, a relação entre o número de respostas feitas para dois usuários distintos), ou até mesmo quantidade de publicações sem comentários (KAKULAPATI; REDDY, 2020; SINHA *et al.*, 2019; CHAI *et al.*, 2019; FRAGA; SILVA; MURAI, 2018; RICARD *et al.*, 2018; ALVAREZ-MON *et al.*, 2018; SOLDAINI *et al.*, 2018; VEDULA; PARTHASARATHY, 2017; LIN *et al.*, 2017; WANG *et al.*, 2017; AKAY; DRAGOMIR; ERLANDSSON, 2016; SARAVIA *et al.*, 2016; CHANG; SARAVIA; CHEN, 2016; CHOUDHURY *et al.*, 2016; KUMAR *et al.*, 2015; PARK *et al.*, 2015; TSUGAWA *et al.*, 2015; HUSSAIN *et al.*, 2015; ZHAO; JIA; FENG, 2015; WU *et al.*, 2020; DUTTA; CHOUDHURY, 2020). Outra característica de interação bastante comum é o relacionamento através dos laços de amizade na forma de seguidores e amigos, ou seja, laços bidirecionais, quando essa relação é mútua entre usuários, ou unidirecional, quando parte de apenas um dos lados (LOWE-CALVERLEY; GRIEVE; PADGETT, 2019; WONGKOBLAP; VADILLO; CURCIN, 2018; DIJK; TREUR, 2018; WANG *et al.*, 2017; SHEN *et al.*, 2017; PREOŢIUC-PIETRO *et al.*, 2015; TSUGAWA *et al.*, 2015; SEMENOV *et al.*, 2015; LUP; TRUB; ROSENTHAL, 2015; HUSSAIN *et al.*, 2015; MCMANUS *et al.*, 2015; WU *et al.*, 2020; DUTTA; MA; CHOUDHURY, 2018). Outras formas de interação envolvem o número de atribuições positivas a perfis de usuários, páginas de conteúdo e publicações, a diferença de votos, ou seja, a diferença entre atribuições positivas e negativas (WONGKOBLAP; VADILLO; CURCIN, 2018; RICARD *et al.*, 2018; CHOUDHURY *et al.*, 2016; KUMAR *et al.*, 2015; PARK *et al.*, 2015; HUSSAIN *et al.*, 2015; CHOUDHURY, 2015; ZHAO; JIA; FENG, 2015; LIN *et al.*, 2014; WU *et al.*, 2020), as participações em grupos ou comunidades (WONGKOBLAP; VADILLO; CURCIN, 2018; SEMENOV *et al.*,

2015), a marcação de usuários ou publicações como favoritos (PREOŢIUC-PIETRO *et al.*, 2015) a entropia, ou seja, a medida da diversidade nas interações (WANG *et al.*, 2017) e, por fim, o conteúdo dos comentários de usuários da rede de relacionamento (LIN *et al.*, 2017; SOLDAINI *et al.*, 2018; DUTTA; MA; CHOUDHURY, 2018).

Para a extração de características de rede, muitos trabalhos utilizaram a estrutura de grafos para modelar os relacionamentos de usuários e as interações entre eles. Estas redes foram criadas a partir de diferentes formas de ligação entre os usuários, tendo como principais as menções, respostas, citações, republicações e laços de amizade entre estes usuários e seus seguidores ou amigos. Dentre as características de rede identificadas, destaca-se o número de nós, número de arestas, densidade, diâmetro da rede, excentricidade, assortatividade, coeficiente de agrupamento local, média de tamanho do caminho mais curto, laços fortes e fracos, coeficiente de propagação, fração de laços negativos, *PageRank* e *embeddings* de nós (SINHA *et al.*, 2019; SHRESTHA; SPEZZANO, 2019; CHAI *et al.*, 2019; FRAGA; SILVA; MURAI, 2018; DIJK; TREUR, 2018; VEDULA; PARTHASARATHY, 2017; LIN *et al.*, 2017; WANG *et al.*, 2017; AKAY; DRAGOMIR; ERLANDSSON, 2016; XU; ZHANG, 2016a; PARK *et al.*, 2015; SEMENOV *et al.*, 2015; XU; ZHANG, 2016b; DUTTA; CHOUDHURY, 2020).

Alguns trabalhos utilizaram características específicas de domínio relacionadas ao comportamento, que não necessariamente podem ser facilmente aplicadas a todas as plataformas de redes sociais, como, por exemplo, o compartilhamento de tópicos de tendência (ALMOUZINI; KHEMAKHEM; ALAGEEL, 2019; WANG *et al.*, 2017), a marcação de outros usuários em fotos (WONGKOBLAP; VADILLO; CURCIN, 2018) e a fração de postagens sem legendas (RICARD *et al.*, 2018). O trabalho em Lowe-Calverley, Grieve e Padgett (2019) utiliza ainda outras características específicas, como o investimento e intensidade no *Instagram*, onde são medidos o tempo gasto na rede social e as principais estratégias para maximizar respostas a publicações.

O princípio da homofilia também foi explorado como característica e como forma de entender as relações nas redes sociais entre usuários que possuem comportamentos semelhantes. O trabalho em Dijk e Treur (2018) explora a homofilia no *Twitter* a partir de um modelo de rede adaptativo temporal-causal, com foco no contágio de atividade física como forma de prevenção a transtornos como depressão. O trabalho em Sinha *et al.* (2019) analisou o grafo social a partir de uma Rede Neural Convolucional baseada em grafo para entender a comunicação e propagação de conteúdo suicida no *Twitter*. Padrões de homofilia

foram analisados no trabalho em Wang *et al.* (2017) a partir de comunidades de usuários com transtornos alimentares no *Twitter*. Foram identificados que estes usuários possuem significantes padrões de mistura assortativa no comportamento na rede, como preferência de postagens e uso de linguagem, além de tenderem a se relacionar com usuários similares em termos de peso corporal. A homofilia também foi explorada como característica a partir da fração de amigos que participam das mesmas comunidades nas redes sociais, principalmente relacionadas a depressão e suicídio (SEMENOV *et al.*, 2015).

3.3 Resultados detalhados

Nesta seção são apresentados os principais estudos que têm como proposta a utilização de características comportamentais para a classificação automática de usuários com transtornos de saúde mental em redes sociais a partir de técnicas de AM supervisionado.

3.3.1 Detecção de usuários depressivos em fóruns online

No estudo em Shrestha e Spezzano (2019), o problema de detecção de usuários com depressão é tratado a partir do fórum online australiano denominado *Rechout.com*, disponível gratuitamente para acesso a qualquer usuário. Conforme citado pelos autores, um fórum online é diferente de uma rede social comum, uma vez que o foco é a busca por ajuda e suporte a partir de tópicos específicos, principalmente relacionados à saúde mental. O problema deste trabalho foi formulado a partir de uma tarefa de classificação binária e foram analisadas características relacionadas ao estilo linguístico das postagens dos usuários, características de rede e atividade de postagem.

O conjunto de dados foi extraído a partir da tarefa organizada pelo *CLPsych* de 2017 e contém postagens anotadas manualmente por especialistas de acordo com categorias relacionadas à urgência de atenção por parte dos moderadores e o motivo relacionado. Uma vez que o problema neste estudo é orientado a usuário, uma nova anotação foi realizada, classificando os usuários em diagnosticados com depressão e usuários de controle, a partir das categorias específicas identificadas no conjunto de dados original. O conjunto de dados resultante possui 65 usuários classificados como depressivos e 94 usuários de controle. Com o objetivo de aumentar o conjunto de treinamento, utilizou-se o método

de classificação *k-vizinhos mais próximos* (do inglês, k-Nearest Neighbors – kNN), onde os usuários que possuíam uma classe desconhecida passaram a ser classificados como depressivos ou controle, aumentando o conjunto de dados total para 1716 usuários.

Para a extração de características de conteúdo na forma de texto, foi utilizada a Consulta Linguística e Contagem de Palavras (do inglês, Linguistic Inquiry and Word Count - LIWC), que é uma ferramenta de análise textual para a classificação das palavras em categorias sob diversas perspectivas, sobretudo linguísticas e psicológicas. Quatro principais categorias foram consideradas neste trabalho, que incluem linguística, pontuação, psicológica e síntese. Para a redução de sobreajustes, foram selecionadas apenas as 20 características mais relevantes de um total de 93 características disponibilizadas pelo LIWC.

Para a extração das características de rede, foi construída uma rede baseada em respostas a postagens, ou seja, "quem responde a quem". A modelagem desta rede foi realizada a partir de uma estrutura de grafos do tipo G=V, E, onde V é o conjunto de nós que representam os usuários e E é o conjunto de arestas que representam a interação na forma de respostas a comentários entre usuários. Foram utilizadas quatro características de rede que incluem *PageRank*, reciprocidade, coeficiente de agrupamento e *embeddings* de nós (*node2vec*). *PageRank* é um algoritmo que explora a importância de um nó na rede por meio de associações que este possui. A reciprocidade é uma métrica que permite verificar a taxa de conexões mútuas entre dois nós em relação ao total de interações entre eles, ou seja, quanto maior a reciprocidade, mais interações em ambas as direções existem entre os dois nós. O coeficiente de agrupamento mede a probabilidade de que vizinhos de um nó estejam conectados entre si, formando assim um grupo local. *node2vec* corresponde a uma técnica para gerar representações vetoriais de um nó em uma rede a partir do mapeamento destes nós em um espaço geométrico de alta dimensão, permitindo que características relativas ao contexto em que o nó está inserido sejam exploradas.

Para a análise de características de atividade de postagem, foi extraído o índice de insônia, que mede o período do dia em que estas postagens ocorrem para um determinado usuário. Esta característica foi identificada em Choudhury *et al.* (2013) como sendo relevante, indicando que usuários com depressão utilizam o período noturno com maior frequência para a criação de postagens.

Os experimentos foram realizados a partir de um classificador baseado em regressão logística, com a utilização da regularização L2 para mitigar os sobreajustes, e o uso de

pesos nas classes, devido ao desbalanceamento. Os testes foram realizados com validação cruzada (*3-fold*) e os resultados de classificação foram avaliados a partir das métricas de precisão média e medida-F.

Os experimentos mostraram que utilizando apenas as características linguísticas extraídas a partir do LIWC, o resultado obtido para o classificador foi de 71% de precisão média. Considerando apenas as características *PageRank*, reciprocidade e coeficiente de agrupamento, o resultado do classificador foi de 60% de precisão média. O melhor resultado de classificação foi gerado a partir da combinação das características extraídas do LIWC com as características de rede *PageRank*, reciprocidade e coeficiente de agrupamento, alcançando 78% de precisão média. Adicionando *embeddings* de nós ao subconjunto de características anterior, os resultados atingem 67% de precisão média, reduzindo o desempenho do classificador, possivelmente pela introdução de ruído ao conjunto de dados. Com relação ao índice de insônia, não foi observada uma atividade noturna mais intensa, possivelmente pela natureza do fórum online que possui o objetivo principal de prover ajuda e, portanto, este atributo não foi considerado na classificação.

O trabalho em Yates, Cohan e Goharian (2017) foi utilizado como base de comparação, obtendo o resultado de 50% de medida-F a partir de um classificador baseado apenas em características textuais, utilizando o método de aprendizado profundo. O modelo do presente trabalho obteve 63% de medida-F no mesmo conjunto de dados, utilizando características linguísticas e baseadas em rede, demonstrando a relevância de tais características para o problema de classificação proposto.

O trabalho em Shrestha e Spezzano (2019) demonstrou o potencial de características de rede em conjunto com atributos linguísticos para a classificação de usuários com depressão em fóruns online, superando inclusive modelos que utilizam técnicas mais complexas para a extração e classificação a partir de características textuais. A reciprocidade foi a característica de rede mais importante para usuários de controle, enquanto que o coeficiente de agrupamento local foi a característica de rede mais relevante para usuários com depressão. Com estas análises, pode-se concluir que usuários com depressão possuem a tendência de responder menos nas redes sociais, enquanto que seus vizinhos estão mais interconectados do que os vizinhos de usuários sem depressão.

3.3.2 Sinais emocionais e linguísticos de depressão a partir de mídias sociais

O trabalho em Vedula e Parthasarathy (2017) procurou identificar se usuários do *Twitter* podem revelar sinais relevantes, tanto emocionais quanto linguísticos, para a detecção de depressão. Para isto, foram examinados um grupo de usuários clinicamente diagnosticados com depressão, um grupo de controle, e os laços de relacionamento de ambos os grupos na rede social. Para a identificação do grupo de diagnosticados, foram buscados usuários com postagens no *Twitter* contendo termos frequentemente associados à depressão, definidos a partir de um dicionário da área da saúde, além da menção explícita do uso de antidepressivos, cujos nomes foram obtidos por meio de um especialista. Para a identificação do grupo de controle, foi realizada uma busca aleatória, excluindo os usuários com sobreposição da rede de amigos em relação aos usuários diagnosticados, com exceção de perfis mais populares, como os de celebridades, onde normalmente não há uma interação mútua. Como resultado, foram identificados 50 usuários diagnosticados com depressão e 100 usuários de controle.

A partir deste conjunto de usuários, foram analisadas características de comportamento na rede social e que estão relacionadas à participação e atividade de rede, interação com a rede de relacionamento, conteúdo linguístico e emoção expressa na forma de texto. O objetivo destas análises foi entender o comportamento dos usuários e construir um modelo preditivo para a identificação de usuários com depressão em redes sociais. Adicionalmente, o trabalho também se propôs a analisar a contribuição dos relacionamentos para a emoção expressa por indivíduos com depressão em redes sociais, seja a partir de uma relação direta (primeiro nível) ou indireta (segundo nível), ou seja, amigos dos amigos.

Ao analisar a participação e atividade de rede dos usuários, foram identificados comportamentos que condizem com estudos das áreas de psicologia e sociologia. Observou-se um comportamento noturno mais frequente em usuários com depressão em comparação a usuários de controle. Com relação às interações na forma de repostagens e menções a outros usuários, observou-se uma tendência menos frequente por parte de usuários com depressão a interagirem desta forma em comparação a usuários de controle, com um percentual de repostagens bastante inferior na classe positiva. Usuários com depressão também foram identificados com uma baixa entropia regional, ou seja, indivíduos desta classe estão mais próximos fisicamente em suas redes do que indivíduos da classe de controle.

Com relação às características de rede, observou-se que usuários com depressão possuem redes menores, tanto nos laços de primeiro nível quanto nos laços de segundo nível. Além disso, a rede destes usuários é mais densa e consequentemente formada por grupos de usuários semelhantes, o que pôde ser também observado por meio do coeficiente de agrupamento maior, conforme sugere o efeito da homofilia. Analisou-se também a posição relativa dos usuários na estrutura da rede, a partir de uma representação vetorial de todas as postagens de um usuário e a similaridade com outros usuários da rede. Observou-se que a excentricidade de usuários com depressão é significativamente maior, refletindo na tendência destes usuários de serem menos centrais em suas redes de relacionamento.

Com o objetivo de verificar o engajamento dos usuários na rede, foram analisadas as interação por meio de repostas, repostagens e menções. Foi possível observar que grande parte dos usuários com depressão possui um engajamento menor e sofre pouca influência de outros usuários da rede de relacionamento, sugerindo um isolamento maior por parte destes usuários. Por outro lado, foram identificados alguns usuários com depressão que reagem com maior frequência a estes tipos de interação, exercendo alguma influência na rede, enquanto que outros também recebem um suporte maior por meio de respostas e menções.

Para a análise de conteúdo linguístico, foi explorado o uso de pronomes em postagens dos dois grupos de usuários. Notoriamente, usuários com depressão fizeram uso mais intenso de pronomes que indicam a si próprios de uma forma individual (e.g., “eu”, “meu” e “a mim mesmo”) ou que possuem uma conotação negativa (e.g., “nenhum” e “ninguém”). Os resultados destas análises sugerem que usuários diagnosticados possuem uma tendência a se expressar de uma forma mais isolada, sem normalmente incluir outras pessoas em suas postagens.

Ainda com relação ao conteúdo, as postagens foram submetidas a uma ferramenta chamada *SentiStrengt* (THELWALL; BUCKLEY; PALTOGLOU, 2012), que detecta sentimentos positivos e negativos em textos. Para os usuários com depressão, o sentimento expresso nas postagens foi identificado como predominantemente negativo e não houve diferenciação entre os dias da semana. Em contrapartida, usuários de controle foram identificados com uma tendência geral mais positiva, com exceção dos primeiros dias da semana, em que foram expressos sentimentos mais negativos. As postagens dos usuários da rede de relacionamentos de ambas as classes possuem a tendência de serem mais positivas. Essa tendência sugere que usuários com depressão são pouco influenciados pelo sentimento

de usuários da rede, permanecendo mais isolados, enquanto que usuários de controle são mais facilmente influenciados por seus relacionamentos da rede.

Para a classificação, foi construído um modelo preditivo a partir do algoritmo Árvore de Decisão com Aumento de Gradiente e utilizada validação cruzada (*5-fold*). O classificador utilizou as características de média de emoção expressa em texto, coeficiente de agrupamento relacionado à rede de primeiro nível, estilo linguístico (uso dos pronomes e negatividade), número de postagens, média do número de menções e repostagens recebidas, e entropia regional da rede. Como resultado, o classificador alcançou 90% de medida-F, sugerindo que as interações nas redes sociais atreladas a características de conteúdo e estrutura de rede podem contribuir significativamente para identificar usuários com depressão em mídias sociais.

3.3.3 Caracterização de transtornos de ansiedade com redes interacionais e sociais online

O trabalho em Dutta e Choudhury (2020) procurou responder se a estrutura de rede, as interações entre usuários e os comportamentos sociais em redes sociais online poderiam fornecer sinais de que um indivíduo sofre de transtornos de ansiedade. Para isso, este trabalho explorou características de interação e rede extraídas do *Twitter* em conjunto com características de conteúdo textuais e construiu um classificador binário baseado em AM supervisionado para a detecção de usuários com transtornos de ansiedade.

A extração de dados foi realizada a partir da plataforma *Twitter*, com dados de usuários entre 2013 e 2017. O conjunto de dados foi construído a partir de dados de usuários classificados como diagnosticados com transtornos de ansiedade e usuários de controle. Para a extração de usuários com ansiedade (classe positiva), foram identificados perfis que informaram ter um diagnóstico deste transtorno por meio de autorrelato em uma postagem pública. Além disso, dois especialistas fizeram a revisão das postagens para validar se os autorrelatos se tratavam de diagnósticos genuínos. Para a extração de usuários de controle (classe negativa), foi utilizado o método de busca aleatória, selecionando usuários que nunca mencionaram qualquer tipo de diagnóstico de ansiedade ou frases relacionadas a este transtorno. Para ambas as classes, foram removidos os usuários com menos de 20 ou mais de 10 mil seguidores e amigos e também aqueles com menos de 10 postagens no geral. Além do conteúdo gerado por estes usuários, também foram extraídos os dados de

rede e interação com outros usuários para as duas classes. Estas redes foram construídas a partir de conexões bidirecionais, ou seja, entre usuários que são seguidores e seguidos ao mesmo tempo entre eles. Para os dados de interação social, foram consideradas as interações na forma de respostas a postagens de outros usuários e postagens que fazem menções a outros usuários. Como resultado da aplicação destes métodos, foram extraídos 200 usuários diagnosticados com ansiedade e 200 usuários de controle, além dos usuários pertencentes à rede de relacionamento, tanto de primeiro quanto de segundo nível.

Para a extração das características dos usuários, foram explorados atributos de engajamento social, rede, interações e comportamento social no *Twitter*. Os atributos de engajamento social utilizados foram o número de postagens por dia, a proporção diária de postagens do tipo resposta, a fração de repostagens, e a fração de citações a outras postagens por dia. Além disso, foram consideradas também a fração de respostas, repostagens e menções a amigos, diferenciando, assim, a força dos laços entre aqueles que possuem um relacionamento bidirecional de amizade para aqueles que não possuem tal laço. Os atributos de rede foram explorados a partir de uma rede construída por meio de conexões do tipo seguidor e seguido em até dois níveis de relacionamento, e foram divididos em propriedades do nó, propriedades diádicas e propriedades de rede. Os atributos de propriedades do nó extraídos foram o número de seguidores, número de seguidos e número de amigos (representando um laço bidirecional). Os atributos de propriedade diádicas considerados foram a reciprocidade e a taxa de prestígio, que mostra a relação entre o número de respostas entre dois usuários que possuem laços bidirecionais. Com relação aos atributos de propriedades de rede, foram considerados a centralidade, a densidade do grafo, o coeficiente de agrupamento, o tamanho da vizinhança em dois níveis, a taxa de vizinhos em comum entre dois usuários com relação ao total de vizinhos de ambos, o número de componentes conectados de um usuário e a média do tamanho destes componentes. Para a extração dos atributos de interação, foi construída uma rede baseada nas respostas a postagens entre os usuários. Estes atributos foram divididos em duas categorias que são as propriedades de rede "sem sinal", onde foram consideradas seis das características utilizadas também na rede do tipo seguidor e seguido, e também propriedades de rede "com sinal", que considera a polaridade das interações em positivo e negativo, de acordo com a aplicação da técnica de análise de sentimentos. Para estas últimas, foram consideradas a fração de laços negativos, a média do grau de nós que possuem laços negativos com um usuário, e a fração de postagens da categoria *família* que expressam sentimentos

negativos. Por fim, em relação aos atributos de comportamento social, foram extraídas características de conteúdo textuais baseados no LIWC, sendo quatro delas relacionadas ao estado emocional dos usuários (i.e., afeto positivo, afeto negativo, nervosismo e tristeza) e outras 22 categorias relacionadas ao estilo linguístico (e.g., artigos, verbos auxiliares, conjunções, prenomes pessoais, etc.).

Alguns classificadores binários baseados em aprendizado supervisionado foram construídos para a predição de transtornos de ansiedade em usuários do *Twitter* por meio de diferentes algoritmos de classificação, como Processo Gaussiano, Árvore de Decisão, Floresta Aleatória, Perceptron Multicamadas, Estímulo Adaptativo (Adaboost), Naive Bayes Gaussiano, Regressão Logística e SVM. Cada um desses classificadores foi treinado a partir de cada uma das categorias de atributos definidas neste trabalho. Além disso, um classificador adicional foi treinado, utilizando todas as características em um espaço dimensional reduzido. O conjunto de dados foi dividido em treinamento (80%) e teste (20%) e a validação foi feita por meio de validação cruzada (*5-fold*). Os melhores resultados foram obtidos pelo classificador construído com SVM e todas as características, atingindo 79% de acurácia.

Apesar de individualmente a categoria de atributos de comportamento social contribuir para os melhores resultados (73% de acurácia), é possível observar que os atributos de rede e relacionamento em conjunto com atributos extraídos a partir do conteúdo textual podem contribuir significativamente para a classificação de transtornos de ansiedade em usuários de plataformas de rede social online. A análise dos resultados também permitiu observações importantes, como o fato de que, apesar de redes sociais como o *Twitter* serem formadas na grande maioria por laços fracos (i.e., relações unidirecionais), indivíduos foram mais afetados pelas interações dos laços fortes (i.e., relações bidirecionais), resultando em um risco de ansiedade ainda maior quando as interações foram consideradas negativas.

3.3.4 Detecção de estresse baseado em interações sociais em redes sociais

O trabalho em Lin *et al.* (2017) explorou dados públicos em redes sociais online para a detecção de estresse psicológico em usuários destas plataformas. Apesar de o estresse ser comum a grande parte da população, sua forma crônica e excessiva pode levar a

danos mentais e físicos e inclusive estar relacionado a depressão, insônia e, em casos mais graves, levar ao suicídio. Este trabalho considerou atributos a nível de postagens, ou seja, relacionados ao conteúdo gerado por usuários nas redes sociais, e também atributos a nível de usuário, explorando seus comportamentos, atividades e interações com outros usuários durante períodos semanais.

Para a condução dos experimentos, quatro conjuntos de dados foram considerados, em que metodologias diferentes foram aplicadas para a extração dos dados. O primeiro conjunto (DB1) foi extraído da plataforma *Sina Weibo* a partir de autorrelatos de usuários por meio de postagens, tanto para a classe positiva (e.g., "Eu me sinto estressado esta semana") como para a classe negativa (e.g., "Eu me sinto relaxado"). Ao final, foram extraídos 23304 usuários, sendo 11074 da classe estressados e 12230 da classe não-estressados. O segundo conjunto de dados (DB2) foi extraído a partir do conjunto DB1, do qual foi selecionada uma amostra de usuários que compartilharam os resultados da escala de estresse psicológico PSTR. Para este conjunto de dados, foram selecionados 98 usuários da classe estressados e 112 usuários da classe não-estressados. Os dois últimos conjuntos de dados (DB3 e DB4) foram construídos para teste do método proposto e foram extraídos das plataformas *Tencent Weibo* (DB3) e *Twitter* (DB4), seguindo a mesma metodologia baseada em autorrelatos utilizada para a extração dos conjuntos DB1 e DB2. Para o conjunto DB3, foram extraídos 7845 usuários da classe estressados e 8239 usuários da classe não-estressados, enquanto que para o conjunto DB4, foram considerados 4905 usuários da classe estressados e 4018 usuários da classe não-estressados.

Para a seleção das características potencialmente relevantes para a diferenciação das duas classes de usuários, foram considerados conjuntos de atributos a nível de postagem e a nível de usuário. As características a nível de postagem foram selecionadas a partir do conteúdo linguístico, conteúdo visual e atributos sociais. Os atributos linguísticos foram extraídos do dicionário LIWC chinês e contemplaram características relacionadas a análise de sentimentos, como o número de palavras com emoções positiva e negativa, e o grau de emoção expresso em advérbios e palavras associadas. Além disso, foram extraídos o número de *emoticons* positivos e negativos, e as marcas de pontuação associadas a palavras que expressam emoção. Para os atributos visuais, foram extraídas características de processamento de imagens e atributos relacionados à cor, como saturação, brilho, cores quentes ou frias, cores claras ou opacas e cores dominantes. Para os atributos sociais, foram

extraídas as quantidades de comentários, repostagens e atribuições positivas relacionadas a cada postagem.

As características a nível de usuário foram categorizadas em comportamento de postagem e interação social, e foram extraídas a partir de uma lista de postagens dos usuários em períodos de uma semana, considerando tanto o conteúdo destas postagens como as interações sociais relacionadas. As características de comportamento de postagem extraídas foram o engajamento social (número de menções a outros usuários, repostagens e respostas de outros usuários), o horário das postagens, o tipo da postagem (imagens, postagens de autoria do próprio usuário, postagens com dúvidas ou busca por ajuda e postagens contendo *links* a outras páginas) e o estilo linguístico (baseado em 10 categorias do LIWC, e.g., pronomes pessoais, amigos, família, dinheiro, religião, etc). As características de interação social selecionadas foram o estilo do conteúdo gerado na relação com outros usuários (número de palavras em 10 categorias do LIWC e número de *emoticons* positivos e negativos), a influência social (número de vizinhos da classe estressados, número de laços fortes e de laços fracos, quantidade de seguidores e quantidades de fãs) e a estrutura social, representando a distribuição das conexões da rede de amizade dos usuários.

A arquitetura do classificador proposto por este trabalho consiste em um modelo híbrido composto por Grafo-fator (do inglês, *Factor Graph Model* - FGM) para a modelagem das interações, combinado com uma Rede Neural Convolucional (do inglês, *Convolutional Neural Network* - CNN) com codificadores automáticos (do inglês, *Cross Auto Encoder* - CAE), que possui bom desempenho na modelagem de características extraídas de múltiplas modalidades (e.g., texto, imagens e atributos de relacionamento). A CNN foi aplicada para modelar um usuário a partir de uma série de postagens em um período de uma semana, devido a sua alta capacidade de aprendizagem de atributos locais. Os codificadores automáticos foram treinados por meio do método de Gradiente Descendente Estocástico e permitem o aprendizado das correlações entre as diferentes modalidades de atributos. O Grafo-fator foi utilizado para incorporar as interações sociais e o conteúdo das postagens por meio da fatoração de uma probabilidade global a partir de funções fatores locais. Foram utilizadas três tipos de funções fatores que consideram fatores de atributos (correlação entre o estado de estresse de um usuário em um determinado momento e seus atributos), fatores dinâmicos (a correlação entre o estado de estresse de um usuário em um momento t e $t + 1$) e fatores sociais (correlação entre o estado de estresse de dois usuários em um determinado momento) para gerar uma função objetivo que calcula a probabilidade do

modelo. O processo de aprendizagem consiste em estimar os parâmetros de configuração do modelo a partir do conjunto de dados e maximizar a função objetivo.

Para a geração de resultados base de comparação, foram construídos classificadores para a detecção de estresse psicológico a nível de usuário que foram treinados no conjunto DB1, utilizando todas as características selecionadas, e testados com validação cruzada (*5-fold*) e 10 corridas experimentais aleatórias. Os métodos de classificação utilizados foram Regressão Logística (F1-escore: 83%), SVM com núcleo RBF (F1-escore: 80,82%), Floresta Aleatória (F1-escore: 84,18%), Árvore de Decisão com Aumento de Gradiente (F1-escore: 84,43%) e CNN com codificadores automáticos (F1-escore: 89%). O modelo proposto (FGM + CNN) foi também treinado e testado no conjunto DB1, sendo avaliado por meio da combinação de diferentes grupos de características, alcançando o melhor desempenho com todos os atributos (F1-escore: 93,40%). Outros experimentos foram também realizados nos conjuntos de dados DB2, DB3 e DB4, em que o modelo proposto foi testado com todas as características e utilizando quatro camadas na CNN, alcançando resultados satisfatórios de F1-escore para todos os conjuntos de dados (DB2: 87,85%, DB3: 88,32% e DB4: 82,24%), demonstrando a universalidade da aplicação do modelo.

Ao longo do trabalho, foram realizados testes com os diferentes grupos de características para validar a efetividade de cada um deles. Utilizando apenas os atributos a nível de usuário, verificou-se que o desempenho foi bastante inferior à utilização somente dos atributos a nível de postagem, revelando a importância das características de conteúdo. Porém, combinando os dois grupos de características, foi possível observar um aumento de 10% a 20% nos resultados, indicando que os atributos a nível de usuário são complementares aos atributos a nível de postagem. Os resultados com os atributos de interação social a nível de usuário revelaram que estes atributos são bastante efetivos para a detecção de estresse. De forma geral, os atributos de conteúdo, por serem a parte necessária da postagem, podem alcançar resultados bastante elevados, porém, utilizando todos os atributos, os resultados foram superiores àqueles que utilizam apenas o conteúdo das postagens.

Análises complementares foram realizadas para entender como o estresse é desenvolvido e como pode afetar outros usuários na rede social. Foi identificado que o conteúdo das postagens de cada uma das classes é nitidamente diferente, principalmente no uso de categorias de palavras. A análise da estrutura social das duas classes por meio de conexões na forma de seguidores e amigos indicou que usuários da classe estressados

tende a ser menos complexa e menos conectada em comparação à classe não-estressados, confirmando resultados já identificados em pesquisas da área de Psicologia. Além disso, foram encontradas evidências de que os usuários possuem a tendência de mudar seus comportamentos de acordo com os comportamentos dos amigos na rede social, sugerindo uma forte influência social. Por fim, a força dos laços foi identificada como um fator relevante na influência social, sugerindo que aquelas conexões em que exite mais interação tendem a influenciar mais o comportamento dos usuários nos laços de amizade.

3.3.5 Detecção e caracterização de comunidades de transtornos alimentares em mídias sociais

O objetivo principal do trabalho em Wang *et al.* (2017) foi analisar as diferenças entre usuários do *Twitter* em relação aos padrões de comportamento que possam estar ligados a transtornos alimentares, resultando no desenvolvimento de um modelo preditivo para a classificação binária destes usuários referente a apresentação destes transtornos. Além disso, este trabalho buscou explorar a homofilia em comunidades de indivíduos com transtorno alimentar no *Twitter*, caracterizando as redes e identificando semelhanças de comportamento e preferências de conteúdo. Bulimia nervosa e anorexia nervosa são dois dos mais comuns transtornos alimentares e possuem como principais sintomas a ansiedade e depressão, além de comportamentos e respostas emocionais relativos a alimentação e ganho de peso.

A extração do conjunto de dados ocorreu de forma a adotar estratégias diferentes para a identificação de cada uma das classes. Para a extração da classe de usuários com transtornos alimentares, foram identificados indivíduos que relataram ter sido diagnosticados com este transtorno na descrição de seus perfis na rede social. Este relato foi baseado em palavras semanticamente semelhantes a "transtorno alimentar", extraídas a partir de um dicionário de palavras e frases online. Além disso, estes usuários deveriam ter informações biológicas pessoais, principalmente relacionadas ao peso, para a diferenciação de perfis que não fazem parte do público alvo, como terapeutas, organizações, institutos, etc. A amostra de usuários da classe positiva foi ainda expandida por meio do método de amostragem "bola de neve", que considerou os seguidores e amigos destes usuários na rede social. Para a extração de usuários de controle, foram utilizadas duas estratégias para a construção de conjuntos de dados apartados. A primeira estratégia correspondeu

a uma seleção aleatória de usuários que possuem postagens no idioma inglês e que são ativos na plataforma. A segunda estratégia considerou 14 populares artistas musicais para a identificação de usuários jovens e também seus nomes próprios para a identificação de usuários do sexo feminino, compondo o conjunto de dados "jovens". Em ambas as estratégias, foram também extraídas as redes de seguidores e amigos. Como resultado, foram identificados 3380 usuários da classe positiva, 30684 usuários de controle do tipo aleatório e 37983 usuários de controle do tipo jovens.

As características dos usuários selecionados foram divididas em atributos sociais, padrões de comportamento e propriedades psicométricas. Foram consideradas 6 características sociais relacionadas ao engajamento na rede social por meio do volume geral de seguidores, amigos e postagens, e também relacionadas à atividade do usuário, por meio da média normalizada de seguidores, amigos e postagens por dia. Foram também consideradas 11 características de padrões de comportamento, relacionadas à preferência de postagem e a diversidade na interação. As características de preferência de postagem foram extraídas a partir do tipo de publicação (postagens, repostagens ou citações a postagens de outros usuários), interação com usuários (menções e respostas), postagens relacionadas a tópicos públicos e, por fim, o compartilhamento de *links* externos. As características de interação foram basicamente representadas pela entropia, que mede a diversidade de interesses que um usuário possui (e.g., preferência por seguir uma variedade grande ou pequena de tópicos ou a preferência na interação com poucos ou muitos usuários). As características psicométricas estão diretamente relacionadas ao conteúdo das postagens na forma de textos e foram analisadas por meio do léxico LIWC, onde foram extraías 80 categorias que correspondem a emoções, estilo linguístico, uso de pronomes, dentre outros. Após análises comparativas entre as classes baseadas nestas características, verificou-se que usuários com transtornos alimentares possuem um número menor de seguidores, amigos e postagens. Em contrapartida, estes usuários são relativamente ativos ao se considerar um pequeno período. Observou-se também que esta classe de usuários possui menor interação com outros usuários, normalmente seguem uma menor quantidade de tópicos públicos e a diversidade nas interações é baixa em comparação ao grupo de controle.

Para a etapa de classificação, foram consideradas 97 características dos usuários que foram utilizadas para treinar classificadores baseados em diversos algoritmos, como Naive Bayes, k-vizinhos mais próximos e SVM com núcleos linear, RBF, sigmoide e polinomial. A classificação foi realizada de forma binária com aplicação apartada para

cada par de conjunto de dados (diagnosticados com transtorno alimentar, usuários de controle do tipo aleatório e usuários de controle do tipo jovens). Os testes foram realizados utilizando validação cruzada (*5-fold*) e foram aplicados pesos de classes para ajuste do desbalanceamento. Como métricas de avaliação, foram consideradas a acurácia, precisão, revocação e medida-F. Os melhores resultados foram obtidos a partir do classificador com SVM linear, atingindo valores acima de 97% para todas as métricas utilizadas, tanto na comparação entre a classe de diagnosticados com a classe de controle aleatória, quanto entre a classe de diagnosticados com a classe de controle do tipo jovens. Os altos resultados foram explicados principalmente pela eficácia do método de seleção da classe de diagnosticados e pelo volume de usuários em cada classe, relativamente maior que em trabalhos anteriores.

Além das características utilizadas na construção dos classificadores, foram analisados também o comportamento dos usuários em sua rede de relacionamento. Para a extração destas características, foram consideradas quatro formas de modelagem de rede, sendo elas por meio de seguidores e amigos, repostagens, respostas e menções a outros usuários. As principais características de rede analisadas foram o número de nós, o número de arestas, a densidade, a média do tamanho do caminho mais curto entre nós, o número de nós fracamente conectados, a fração de nós fracamente conectados do tipo gigante, o coeficiente de agrupamento, a reciprocidade e a assortatividade. Algumas análises destas características mostraram que usuários com transtornos alimentares possuem uma densidade maior de laços sociais e formam comunidades mais firmemente conectadas. Além disso, verificou-se que estes usuários na rede do tipo seguidores e amigos possuem uma assortatividade mais baixa, ou seja, usuários com muitos seguidores normalmente não são seguidos por usuários que também possuem muitos seguidores. Este mesmo resultado foi encontrado para as redes criadas a partir de respostas e menções, porém, na rede formada por repostagens, a assortatividade foi maior, ou seja, usuários que fazem muitas republicações possuem a tendência de utilizarem postagens de outros usuários que também fazem muitas repostagens. A assortatividade está fortemente relacionada ao princípio da homofilia, que foi observado em diversas características analisadas, como as preferências das postagens, uso da linguagem, além de tendência de relacionamento entre usuários com pesos similares.

3.3.6 #suicida - uma abordagem multifacetada para identificar e explorar a ideação suicida no *Twitter*

O estudo abordado em Sinha *et al.* (2019) teve como foco identificar postagens com ideação suicida a partir do *Twitter*. Para isso, foram construídos modelos baseados em conteúdo textual, atividade histórica de postagens e características de rede, que contribuíram para as análises do comportamento de usuários com tendência suicida e para a classificação das postagens de acordo com o risco de pertencer a este tipo de ideação. O contexto que envolve o suicídio pode ser entendido por meio de conteúdo e comportamentos que envolvem pensamentos, intenções, seja da completude do ato ou mesmo em tentativas de execução e, na maioria dos casos, pode ser considerada uma sintomatologia decorrente de outros transtornos de saúde mental, como depressão e ansiedade.

Para a identificação de postagens relacionadas à ideação suicida e criação de um conjunto de dados, foi construído um léxico com 248 frases relacionadas ao assunto, que foi utilizado para extrair o conteúdo do *Twitter* em dezembro de 2018. Três especialistas fizeram anotações manuais, verificando quais dessas postagens realmente possuíam uma ideação suicida (classe positiva), resultando na geração de um conjunto de dados com 3984 postagens da classe positiva e 30322 da classe negativa (sem ideação suicida). A partir desta seleção de postagens, foram identificados 32558 usuários únicos responsáveis por estas publicações, dos quais foi também extraído o histórico de até 3200 postagens na rede social.

Com o objetivo de criar representações dos usuários, de seus comportamentos de postagens e de suas interações na rede, foram construídos três modelos que melhor representam essas informações extraídas do conjunto de dados. Ao final, um comitê de classificadores foi construído no intuito de explorar as diferentes abordagens de cada um dos modelos implementados, com a hipótese de gerar resultados de classificação ainda melhores que cada um dos modelos individuais.

Para a extração de características textuais, cinco modelos foram construídos e comparados com o objetivo de selecionar o modelo com melhor desempenho para o comitê de classificadores. O primeiro modelo utilizou Regressão Logística e foi treinado a partir de características n-gramas, variando de 1 a 4 caracteres. O segundo modelo foi construído a partir de uma CNN com 100 filtros em cada camada, seguidas por uma camada máxima de agrupamento global, utilizando o *GloVe* para a representação das palavras, que é um

modelo para a representação distribuída de palavras a partir de vetores globais. O terceiro modelo utilizou o *GloVe* para a representação das palavras e a arquitetura LSTM, que é uma rede neural recorrente baseada em memória de longo prazo. O quarto modelo foi construído a partir de uma C-LSTM, que utiliza uma CNN para capturar características locais de frases e uma RNN para capturar semânticas de frases globais e temporais. Por fim, o quinto modelo se baseou em uma arquitetura LSTM bidirecional (BLSTM) que permite capturar o contexto tanto à direita, como à esquerda de uma palavra. Este modelo também utilizou uma camada de atenção, que fornece um contexto útil a partir de uma palavra em questão.

A modelagem da atividade histórica de postagens foi também realizada por meio de um modelo BLSTM com camada de atenção. A representação de cada postagem histórica foi obtida a partir do vetor latente da penúltima camada do modelo. Além disso, foram atribuídos pesos temporais, onde a importância dada a uma postagem histórica varia inversamente com a distância temporal da postagem em questão. A atividade histórica foi representada a partir de um vetor de características e um modelo de Regressão Logística foi treinado para aprender tais atributos.

Para a modelagem de interações entre usuários, quatro diferentes grafos foram construídos baseando-se em interações com outros usuários dos tipos citações ao conteúdo, menções, respostas a postagens e a combinação destas três formas de interação. Os grafos foram construídos com a mesma quantidade de nós que representam os usuários, porém o número de arestas variou conforme o tipo de interação. Para a representação dos usuários e suas características, foi utilizada a arquitetura de uma rede convolucional baseada em Grafo (do inglês, Graph Convolutional Networks - GCN). Essa arquitetura permite a extração efetiva de informações de contexto, dados linguísticos e estruturais dos usuários, a partir de aprendizado semi-supervisionado. Um grafo estendido foi criado contendo nós que representam tanto os usuários como suas postagens. Para os nós que representam postagens, foram extraídas características TF-IDF a partir de unigramas e bigramas, e foram selecionadas as 1000 melhores características a partir do método ANOVA. Para os nós que representam usuários, o vetor de características foi criado a partir da média dos vetores de características dos nós de postagens correspondentes.

Os três modelos foram treinados separadamente e a combinação desses modelos foi explorada por meio de um comitê de classificadores do tipo *stacking*. Para estes experimentos, foi utilizada validação cruzada estratificada (*10-fold*) e as métricas de

avaliação consideradas foram medida-F, precisão, revocação e área sob a curva de precisão-revocação. Os três modelos obtiveram resultados significativos, com 92,26% de medida-F para o modelo textual (BLSTM e camada de atenção), 89,53% de medida-F para o modelo de atividade histórica de postagens considerando pesos para a natureza temporal das postagens e 90,29% de medida-F para o modelo de interações em grafos. A combinação dos modelos por meio de um comitê de classificadores resultou em 92,33% de medida-F para a combinação dos modelos textuais e histórico de atividades e 92,76% para a combinação dos três modelos. Analisando os resultados, foi possível observar que o modelo textual obteve os melhores resultados de forma individual, destacando a importância deste conteúdo para o problema de classificação em questão. Porém, o aumento de desempenho ao combinar as características de texto, histórico de atividade e interação, sugere que estes tipos de características possuem um potencial elevado para contribuir com a detecção de ideação suicida em redes sociais online.

3.3.7 Reconhecimento de depressão a partir de atividade no *Twitter*

O trabalho em Tsugawa *et al.* (2015) utilizou dados extraídos entre 2013 e 2014 de usuários do *Twitter* para a construção de um classificador capaz de identificar usuários com depressão, sendo um dos primeiros trabalhos a explorar mais enfaticamente características de atividade dos usuários e outras relacionadas às interações na rede. O estudo utilizou como base comparativa o trabalho em Choudhury *et al.* (2013), considerado um dos primeiros a explorar a classificação automática de usuários com depressão em redes sociais. Diferentemente da maior parte dos trabalhos relacionados que utilizam um conjunto de dados no idioma inglês, este trabalho utilizou dados no idioma japonês, demonstrando a efetividade do método na aplicação em diferentes contextos.

A identificação dos usuários para a coleta de dados foi feita a partir do método de contribuição colaborativa (do inglês, *crowdsoursing*), em que dois questionários foram disponibilizados em uma página web para serem respondidos por usuários do *Twitter*. Estes questionários correspondem ao CES-D e BDI, ambos amplamente utilizados na área de saúde para avaliar o grau de depressão. Após a aplicação dos questionários, 81 indivíduos foram identificados com depressão ativa, enquanto que 128 não foram identificados com o mesmo transtorno, totalizando 209 indivíduos. Adicionalmente, foram coletados até 3200 postagens desses usuários no *Twitter* e também a lista de seguidores e amigos.

Para a construção de um classificador de depressão, foram extraídas características relacionadas ao conteúdo gerado pelos usuários, suas atividades na rede social e interação com outros usuários. A frequência de palavras em uma postagem foi extraída como característica de conteúdo a partir do método Saco de Palavras (do inglês, *Bag of Words* - BOW), após a remoção de alguns termos mais comuns, como partículas, verbos auxiliares, adjuntos adnominais e símbolos visuais ou expressões pouco frequentes entre os usuários. Foi utilizado também o modelo LDA para a identificação dos tópicos das postagens e um dicionário afetivo de palavras para a extração da taxa de palavras positivas e negativas.

No contexto de atributos de atividade do usuário, foram extraídos o horário e frequência das postagens, a média do número de palavras das postagens e a taxa de postagens contendo localizadores uniformes de recurso (do inglês, *Uniform Resource Locator* - URL). Para a extração de características relacionadas à interação dos usuários, foram consideradas a taxa de repostagens, a taxa de menções a outros usuários e o número de seguidores e amigos na rede de relacionamentos.

O classificador foi construído a partir de um modelo com SVM e núcleo RBF para classificação binária de depressão em usuários do *Twitter*. A avaliação dos resultados foi feita por meio de validação cruzada (*10-fold*) e foram avaliados os resultados a partir das métricas de precisão, revocação, medida-F e acurácia. Foram realizados testes considerando grupos de características para avaliar a relevância de cada grupo no processo de classificação. Considerando apenas as características de conteúdo, os resultados demonstraram uma acurácia entre 61% e 64% na detecção de usuários com depressão. Os melhores resultados que consideram apenas estas características foram obtidos a partir da limitação a 2000 palavras, uma vez que dimensões muito altas podem gerar sobreajustes, e da limitação a 10 tópicos gerados pelo modelo LDA, uma vez que um valor muito alto para esta característica pode dificultar a captação dos tópicos dominantes. Incluindo as características de atividade e interação, foi observado um aumento da acurácia para 66%. O melhor resultado foi obtido após uma análise do período de dados que deve ser considerado na classificação. Observou-se que considerando dois meses de dados, a acurácia gerada pelo modelo foi de 69%, sugerindo que períodos muito longos de dados podem gerar resultados piores, uma vez que podem conter dados desatualizados e não refletir o estado atual de depressão do usuário.

Este trabalho apresentou análises relevantes sobre características de conteúdo, atividade e interação, capazes de contribuir para a classificação de usuários com depressão

no *Twitter*. Diferenças significantes foram identificadas entre as duas classes para as características analisadas, principalmente na taxa de palavras negativas, frequência de postagem, taxa de repostagem, taxa de postagens contendo URLs e número de seguidores e amigos. Além disso, foi observado que há desafios na utilização de características de conteúdo linguístico, uma vez que as características específicas do idioma e questões culturais podem afetar os modelos. Embora existam estes desafios, o modelo de tópicos se mostrou promissor para este tipo de problema, explorando o conteúdo gerado pelos usuários. Os resultados obtidos sugerem que atributos comportamentais atrelados a características de conteúdo podem contribuir significativamente para o reconhecimento de depressão em redes sociais.

3.4 Trabalhos identificados após a revisão sistemática

Esta seção apresenta trabalhos relacionados a detecção de transtornos de saúde mental a partir de dados de mídias sociais que fazem uso de atributos comportamentais e que foram identificados após a condução da revisão sistemática descrita nas seções anteriores deste capítulo. Estes trabalhos foram selecionados de forma exploratória, não contemplando, portanto, a totalidade de trabalhos sobre o assunto no período em questão.

O trabalho em Ruch (2020) combinou dados textuais (*embeddings* de palavras) com dados de estrutura de rede (*embeddings* de nós) por meio da concatenação de características para predizer ideação suicida a partir da mídia social *Reddit*. O conteúdo textual foi extraído de postagens na língua inglesa, enquanto que os dados de rede foram extraídos de um grafo que correlaciona os autores, suas postagens e a comunidade na qual estas foram submetidas. Como resultados, foi destacada a importância da integração de dados linguísticos com dados comportamentais para a tarefa de predição, principalmente com o uso de *embeddings* que permite evitar o uso de dados esparsos, diminuindo a complexidade computacional do modelo.

O trabalho em Mendu *et al.* (2020) apresentou um *framework* para a extração de características que exploram as relações e comportamentos de usuários do *Facebook* por meio de mensagens de texto privadas e a detecção de ansiedade e solidão nestes usuários. Foram extraídas características relacionadas ao conteúdo textual das mensagens na língua inglesa do tipo psicolinguísticas (LIWC) e semânticas (LDA e TF-IDF), e outras relacionadas ao comportamento dos usuários, como temporais (horário e latência da

mensagens) e estrutura de rede (tamanho da rede e peso das arestas). Estas características foram combinadas por meio do método de *stacking* de forma a construir um modelo multimodal. Os resultados apresentados corroboraram os resultados de trabalhos prévios da literatura, que indicaram diferenças nos comportamentos relacionais, temporais e uso do vocabulário dos usuários de mídias sociais diagnosticados com os transtornos em questão.

O trabalho em Bi, Li e Wang (2021) explorou a possibilidade de usar dados do *Sina Microblog* como uma ferramenta para detecção de depressão entre os usuários. Para esta tarefa, foram extraídas características relacionadas ao perfil do usuário (gênero, classe do usuário e assinatura pessoal), engajamento social (número de seguidores, amigos, média do número de curtidas, repostagens e comentários recebidos pelo usuário) e ao conteúdo textual das postagens (aspectos psicolinguísticos do LIWC e termos contidos em um léxico relacionados à depressão). As características foram concatenadas e utilizadas para o treinamento de classificadores baseados em SVM, Regressão Logística e Floresta Aleatória. Os resultados dos experimentos demonstraram a habilidade do modelo em identificar usuários com depressão na mídia social.

O trabalho em Ghosh e Anwar (2021) explorou o problema de detecção de depressão a partir do comportamento de postagens de usuários do *Twitter*. Para isso, foi proposto um método baseado em aprendizado profundo (rede LSTM usando Swish como função de ativação) para estimar a intensidade da depressão, e não apenas a presença ou ausência do transtorno, como ocorre na maior parte dos trabalhos encontrados na literatura. O modelo foi construído com a representação de um usuário por meio de um vetor de 527 características, incluindo sentimentos (aspectos psicolinguísticos), tópicos (LDA), comportamento (número de *tweets*, *retweets* e comentários), linguísticas no idioma inglês (n-gramas) e características de usuários disponibilizadas pela plataforma do *Twitter*. O modelo proposto superou outros modelos de *baseline* para a tarefa de estimativa da intensidade de depressão. Além disso, uma adaptação do método para o problema de classificação binária superou outros modelos existentes em 2% de acurácia.

O trabalho em Zogan *et al.* (2022) propôs um modelo explicável de detecção automática de usuários com depressão em mídias sociais. Este modelo combinou dados linguísticos e comportamentais de usuários do *Twitter* extraídos dos *tweets* e de suas atividades de postagem. Todas as características comportamentais foram extraídas a partir de contagens, como de suas redes de relacionamento (e.g., número de seguidores, amigos, favoritos, etc) e atividade de postagem (e.g., número de *tweets*, número de *retweets*, tamanho

total das postagens, quantidade de *tweets* a cada hora do dia, etc). Estas características comportamentais foram combinadas com características extraídas do conteúdo textual na língua inglesa (*embeddings* de palavras, tópicos a partir do LDA, número de *emoticons*, etc) para criar um modelo multimodal. Este modelo obteve excelentes resultados, superando outros modelos de *baseline* que fazem uso de métodos robustos.

O trabalho em Cheng e Chen (2022) utilizou o *Instagram* como plataforma de estudo para a detecção de depressão a partir de um conjunto de dados de usuários que foram rotulados como depressivos ou não-depressivos de acordo com seus autorrelatos. Um modelo multimodal foi desenvolvido para esta tarefa, fazendo uso de características textuais na língua chinesa, temporais e de imagem. Para a extração das características textuais e de imagem, foram utilizados modelos pré-treinados (BERT e CNN, respectivamente), enquanto que as características temporais foram extraídas do horário das postagens (estação do ano, dia da semana, período do dia e intervalo entre as postagens). As características foram concatenadas e utilizadas em um modelo baseado em LSTM com mecanismo de atenção para a detecção de depressão. Os resultados foram superiores aos de modelos apresentados em trabalhos prévios, superando, inclusive, outros modelos de referência em plataformas como o *Twitter*.

3.5 Considerações

Os resultados da revisão sistemática indicam que a detecção de transtornos de saúde mental a partir de redes sociais digitais é uma área de pesquisa em ascensão, com os primeiros estudos que utilizam modelos computacionais baseados em aprendizado de máquina supervisionado identificados em 2013. A partir de 2015, houve um aumento de publicações nesta área, possivelmente influenciado por tarefas organizadas por eventos como o *CLPsych*, cujo objetivo é propor o uso de métodos computacionais para melhor entender o uso da língua e comportamento de indivíduos frente a transtornos de saúde mental. Com base no estudo realizado, retomamos a seguir as questões A1, A2 e A3 enunciadas na seção 3.1.

Com relação à questão A1 - *Quais são os principais transtornos de saúde mental explorados na literatura?* - observamos que a depressão foi identificada como transtorno mais explorado em trabalhos nesta área de pesquisa, enquanto que o suicídio foi a sintomatologia

encontrada mais vezes. A ansiedade aparece também como transtorno bastante explorado nos trabalhos selecionados, muitas vezes associada a outros transtornos como depressão. Estes resultados são condizentes com relatórios da OMS que indicam estes problemas como muito comuns até 2030.

Em referência à questão A2 - *Em quais idiomas e em quais plataformas de redes sociais digitais os principais trabalhos se baseiam?* - diversas plataformas de redes sociais digitais foram identificadas pela revisão sistemática, sendo o *Twitter* a plataforma mais explorada, possivelmente porque favorece o compartilhamento massivo de conteúdo dos usuários, principalmente da forma de textos, de onde são frequentemente extraídas características relacionadas aos transtornos de saúde mental. O principal idioma identificado nestes trabalhos é o inglês, seguido pelo chinês, enquanto que outros idiomas são ainda pouco explorados.

Com relação à questão A3 - *Quais são as características relacionadas à atividade, comportamento, interação e rede dos usuários exploradas?* - foram identificados trabalhos que exploram características comportamentais dos usuários, classificadas em quatro tipos: atividade, rede, interação e específica de domínio. As características de atividade mais exploradas estão relacionadas à geração de conteúdo, como a frequência e horário da postagem, mas também podem caracterizar outros tipos de atividades, como a frequência de acesso à rede social, por exemplo. As características de rede foram extraídas a partir da estrutura da rede modelada por meio de diversos tipos de conexões, como seguidores, amigos, menções a outros usuários e respostas a postagens. Alguns trabalhos exploram redes de até dois níveis de conexões (e.g., amigos dos amigos de um usuário), porém geralmente há nestes casos uma limitação no tamanho do conjunto de dados, uma vez que redes mais complexas exigem recursos computacionais mais avançados. A maior diversidade de características de comportamento identificadas nos trabalhos selecionados são do tipo interação, pois permitem explorar diferentes formas de relacionamento, seja por meio dos laços de amizade (e.g., seguidores, amigos, grupos e comunidades) ou por ações que envolvem outros usuários (e.g., curtidas, menções, respostas e *retweets*). Por fim, características específicas de domínio também foram identificadas, porém em um volume reduzido de trabalhos, visto que este tipo de atributo limita a universalidade da aplicação do método para redes sociais em geral.

Os resultados da revisão sistemática indicam que atributos comportamentais são complementares às características textuais, sendo que geralmente os melhores resultados de

classificação são obtidos por modelos que utilizam estes dois grupos em conjunto. Dentre os principais trabalhos identificados pela revisão sistemática, a maior parte utiliza modelos textuais menos complexos, como LIWC e BOW, uma vez que o foco está nos atributos comportamentais, principalmente aqueles relacionados a rede e interação dos usuários. Uma vez que atributos comportamentais e textuais são complementares, faz-se necessário explorar modelos baseados em texto mais robustos, principalmente aqueles fundamentados em redes neurais e aprendizado profundo. Além disso, grande parte das características comportamentais são baseadas em extrações mais simples (e.g., baseadas em contagem), e menos a partir de representações mais robustas, como a partir de modelos de grafos e *embeddings* de nós (e.g., *node2vec*). Por fim, os métodos de classificação utilizados pelos principais trabalhos identificados na revisão sistemática são em grande parte baseados em algoritmos tradicionais de AM supervisionado, como Regressão Logística, SVM, Naive Bayes e Árvores de Decisão. Neste contexto, ainda existem oportunidades para a exploração de métodos mais robustos de representação das características textuais e comportamentais e também outros métodos de classificação.

4 Método

Com base na revisão bibliográfica apresentada no capítulo 3, observamos que modelos computacionais existentes para a predição de transtornos de saúde mental em redes sociais digitais frequentemente utilizam conteúdo gerado por usuários, principalmente na forma de textos (WANG *et al.*, 2017; RICARD *et al.*, 2018; SHRESTHA; SERRA; SPEZZANO, 2020). No entanto, a informação textual disponível em redes sociais é limitada e esparsa, constantemente possui ruídos e pode não ser suficiente para desenvolver um modelo preditivo robusto baseado somente neste tipo de característica (LIN *et al.*, 2017; SINHA *et al.*, 2019), podendo gerar severa perda de informação (CHAI *et al.*, 2019) ou até mesmo o uso de conhecimento irrelevante (ZOGAN *et al.*, 2021). Por outro lado, estudos na área de redes sociais indicam que indivíduos com transtornos de saúde mental preferem formar conexões sociais com indivíduos que possuam transtornos semelhantes, conforme sugerem os efeitos da homofilia (VEDULA; PARTHASARATHY, 2017; GIUNTINI *et al.*, 2021). Com base neste princípio, e no fato de que a interação social é uma das mais importantes características de plataformas de mídias sociais, esses estudos têm passado a considerar este tipo de conhecimento para melhorar a efetividade do reconhecimento de transtornos de saúde mental em redes sociais (LIN *et al.*, 2017; WU *et al.*, 2020).

Os resultados do trabalho em Shrestha e Spezzano (2019), por exemplo, mostram o potencial de características baseadas em rede para identificar usuários com depressão em fóruns online, enquanto que o estudo em Sinha *et al.* (2019) demonstra que a força do uso de características como o engajamento na rede social, o histórico de atividade do usuário e as redes de homofilia, está na habilidade de olhar além de características linguísticas e fazer predições baseadas nos aspectos comportamentais e interacionais.

Apesar da potencial relevância de características comportamentais para a detecção de transtornos de saúde mental em redes sociais digitais, poucos estudos têm pesquisado sobre laços sociais, atividades e interações entre usuários para o desenvolvimento de modelos preditivos (WANG *et al.*, 2017; SHEN *et al.*, 2017). De acordo com o trabalho em Chai *et al.* (2019), por exemplo, o uso dessas fontes de informação permanece praticamente inexplorado na literatura. Além disso, a maioria dos experimentos são conduzidos em conjuntos de dados muito pequenos, o que torna muito difícil de justificar a robustez e a generalidade dos resultados em larga escala (SHEN *et al.*, 2017; WANG *et al.*, 2020).

Estas observações motivaram o desenvolvimento de um estudo para a predição de transtornos de saúde mental em redes sociais digitais que explora o comportamento dos usuários e suas conexões. Com foco no conhecimento extra-linguístico, este estudo priorizou as características de estrutura de rede, interação e atividade dos usuários e suas conexões. Para esse fim, retomamos a seguir as questões de pesquisa e objetivos enunciados no capítulo 1, e acrescentamos detalhes adicionais deste projeto.

4.1 Questões de pesquisa

Este trabalho se propôs a responder as seguintes questões de pesquisa:

Q1. Quais características extra-linguísticas baseadas no comportamento de usuários em redes sociais podem contribuir significativamente para a predição de transtornos de saúde mental?

Q2. Modelos computacionais que combinam diferentes tipos de características comportamentais de usuários de redes sociais possuem resultados significativamente superiores aos modelos de classificação de transtornos de saúde mental baseados apenas em características individuais?

Estas questões de pesquisa foram investigadas comparando-se modelos de predição de transtornos de saúde mental baseados em atributos extra-linguísticos, sejam eles individuais ou modelos que combinam diferentes fontes de conhecimento, utilizando-se para esse fim métricas tradicionais de AM. Estes modelos foram também comparados com um modelo de *baseline* que considera apenas atributos psicolinguísticos.

4.2 Objetivo

O objetivo deste trabalho foi desenvolver modelos computacionais baseados em técnicas de AM supervisionado para o reconhecimento de transtornos de saúde mental do tipo depressão e ansiedade a partir de dados da rede social *Twitter* em português, contemplando características comportamentais, que envolvam a atividade, estrutura da rede e os relacionamentos entre usuários, de modo a verificar se tais características podem contribuir significativamente para as tarefas de predição propostas.

4.3 Justificativa

Uma vez comprovado que modelos computacionais baseados em conhecimento extra-linguístico de usuários de redes sociais digitais podem contribuir significativamente para a predição de transtornos de saúde mental, então, futuramente, aplicações deste tipo poderão auxiliar no diagnóstico destes transtornos em indivíduos que utilizam plataformas de mídias sociais, possibilitando que ações preventivas sejam tomadas de forma mais eficaz e tratamentos adequados sejam aplicados nos estágios mais iniciais.

4.4 Metodologia

Este trabalho foi conduzido por meio da execução de atividades correspondentes ao desenvolvimento de modelos computacionais baseados em técnicas de AM para o reconhecimento de transtornos de saúde mental e classificação de usuários em redes sociais. Estes modelos foram baseados no córpus SetembroBR discutido na seção 2.5.

Inicialmente, foram realizados experimentos com classificadores individuais que fazem uso de características extra-linguísticas baseadas no comportamento de usuários para a predição de depressão e transtornos de ansiedade em redes sociais, desenvolvidos com o objetivo de responder a questão *Q1* apresentada na seção 4.1. Isso é descrito no capítulo 5. A seguir, foram desenvolvidos experimentos com modelos que combinam características e predições dos classificadores individuais com o objetivo de responder a questão *Q2* apresentada na seção 4.1. Estes experimentos são descritos no capítulo 6.

4.5 Avaliação

Este trabalho utiliza como conjunto de dados o córpus SetembroBR apresentado em Santos, Oliveira e Paraboni (2023) e detalhado na seção 2.5. O projeto inicial de construção do córpus foi apresentado em Santos, Funabashi e Paraboni (2020) e a versão final detalhada em Santos, Oliveira e Paraboni (2023). Para a avaliação de qualidade deste trabalho, foi utilizada a divisão em treinamento (80%) e teste (20%) originalmente disponibilizada pelo córpus, descrito na seção 2.5. Foram utilizadas métricas tradicionais de AM (precisão, revocação e medida F1) para a avaliação dos resultados para cada classe

e foram construídas matrizes de confusão que permitem a visualização do desempenho dos classificadores de forma mais detalhada, com a classe controle representada pelo valor 0 e a classe diagnosticado representada pelo valor 1. Além disso, a significância estatística entre pares de modelos foi avaliada por meio do teste de McNemar (MCNEMAR, 1947). Todos os experimentos conduzidos neste estudo utilizaram os métodos de avaliação descritos nesta seção.

5 Experimentos com uso de classificadores individuais

Este capítulo apresenta os experimentos realizados no contexto de predição de transtornos de saúde mental a partir de dados de usuários de redes sociais digitais que visam a responder a questão *Q1* apresentada na seção 4.1: *Quais características extra-linguísticas baseadas no comportamento de usuários em redes sociais podem contribuir significativamente para a predição de transtornos de saúde mental?* Para este fim, foram construídos modelos básicos que exploram características de rede, interação e atividade de forma individualizada, com o objetivo principal de investigar a contribuição de cada conjunto de características nas tarefas de predição de depressão e transtornos de ansiedade em usuários do *Twitter* e identificar qual modelo é mais efetivo para cada uma destas tarefas. Além disso, estes experimentos tiveram como objetivo secundário comparar dois métodos de extração de características baseados em rede: (1) seleção univariada e (2) *embeddings* de nós por meio do *node2vec*. Apesar de o foco destes experimentos ser a comparação de modelos baseados em características extra-linguísticas, não foram encontrados na literatura trabalhos correlatos que forneçam modelos comparáveis em termos de volume de dados, características de mesmo tipo, rede social de mesma natureza, e tarefa de predição semelhante. Assim, um modelo que faz uso de características linguísticas extraídas da versão em português do LIWC foi construído para ser utilizado como *baseline* aos demais classificadores. As seções a seguir apresentam o detalhamento da construção destes modelos.

5.1 Modelos desenvolvidos

Nesta seção serão apresentados os modelos [1] individuais que fazem uso de características extra-linguísticas baseadas no comportamento de usuários para a predição de depressão e transtornos de ansiedade em redes sociais, desenvolvidos com o objetivo de responder a questão *Q1* apresentada na seção 4.1.

Estes modelos tiveram como princípio explorar as três principais categorias de características comportamentais disponíveis no córpus SetembroBR, i.e., rede, interação e atividade. As características de rede foram extraídas por meio da estrutura de redes baseadas em amigos e seguidores, enquanto que as características de interação foram extraídas a

[1] ⟨https://github.com/rlagedo/ExtraLinguistic_SetembroBR⟩

partir das menções a outros indivíduos que formam uma rede de interação. Neste contexto, dado um usuário X (diagnosticado ou de controle) existente no córpus, será utilizado o termo "conexão" para designar um amigo, seguidor ou uma menção a outro indivíduo feita por X, e será reservado o termo "usuário" para designar apenas os indivíduos que se deseja classificar, ou seja, os indivíduos cujas *timelines* são representadas no córpus. A atividade dos usuários a classificar foi modelada por meio de características temporais relacionadas ao horário de suas postagens na rede social. Sete modelos foram construídos por meio de redes de relações, sendo três (*amigos.univ.reglog*, *seguidores.univ.reglog* e *menções.univ.reglog*) com o uso de seleção univariada como método de extração de características e quatro (*amigos.n2v.svm*, *seguidores.n2v.svm*, *menções.n2v.svm* e *topmenções.n2v.svm*) com o uso do *node2vec*. O modelo *horario.svm* foi construído com base em características de atividade de postagens no tempo. Estes modelos foram comparados com o *baseline liwc.reglog* construído por meio de aspectos psicolinguísticos extraídos do LIWC em português. O quadro 7 apresenta os modelos individuais construídos no contexto destes experimentos e que serão detalhados nas seções a seguir.

Quadro 7 – Lista de modelos individuais desenvolvidos

Modelo	**Características**			**Método**
	Categoria	**Representação**	**Método de Extração**	
amigos.univ.reglog	Estrutura de rede	Matriz de adjacências	Seleção Univariada	Regressão Logística
seguidores.univ.reglog	Estrutura de rede	Matriz de adjacências	Seleção Univariada	Regressão Logística
menções.univ.reglog	Rede de interação	Matriz de adjacências	Seleção Univariada	Regressão Logística
amigos.n2v.svm	Estrutura de rede	*Embeddings* de nós	*node2vec*	SVM (núcleo RBF)
seguidores.n2v.svm	Estrutura de rede	*Embeddings* de nós	*node2vec*	SVM (núcleo RBF)
menções.n2v.svm	Rede de interação	*Embeddings* de nós	*node2vec*	SVM (núcleo RBF)
topmenções.n2v.svm	Rede de interação	*Embeddings* de nós	*node2vec*	SVM (núcleo RBF)
horario.svm	Atividade	Contagem de postagens	Seleção de subconjunto	SVM (núcleo linear)
liwc.reglog	Linguística	Contagem de palavras	LIWC (categorias psicolinguísticas)	Regressão Logística

Fonte – Rafael Lage de Oliveira, 2023

5.1.1 Modelos de redes de relações baseados em seleção univariada

Nesta seção serão apresentados os modelos *amigos.univ.reglog*, *seguidores.univ.reglog* e *menções.univ.reglog*, construídos por meio de redes de relações com base em seleção univariada como método de extração de características.

Para os modelos de estrutura de rede e de interação, foram exploradas as conexões entre os usuários a classificar do córpus e suas listas de amigos, seguidores e menções. Estas redes foram modeladas a partir de grafos não-dirigidos e a representação das conexões foi feita por meio de matrizes de adjacências, onde o valor 1 representa a existência de

uma conexão entre dois indivíduos e o valor 0 representa a ausência de uma conexão. A modelagem destas redes e conexões foram realizadas de forma independente para amigos seguidores e menções, sendo, portanto, gerados três grafos e três matrizes de adjacências. O quadro 8 exemplifica a representação destas conexões na forma de uma matriz de adjacências para a rede de amigos.

Quadro 8 – Conexões de amizade a partir de uma matriz de adjacências

	Amigo 1	**Amigo 2**	**Amigo 3**	**Amigo 4**
Usuário a classificar 1	1	0	0	0
Usuário a classificar 2	0	1	1	0
Usuário a classificar 3	1	0	0	1

Fonte – Rafael Lage de Oliveira, 2023

Uma conexão de rede representa uma característica na modelagem do problema para o experimento em questão. O espaço dimensional de características é, portanto, definido pela união de todas as listas de conexões dos usuários a classificar do córpus. No entanto, uma vez que a lista de conexões únicas é extensa (e.g., aproximadamente um milhão de menções em ambos os conjuntos de depressão e ansiedade), a representação adotada gera dados esparsos, resultando na necessidade de aplicação de métodos de seleção de características. Neste contexto, optou-se por reduzir o conjunto de características de duas formas: selecionando as conexões mais frequentes e realizando seleção univariada sobre as conexões remanescentes. Estes passos serão discutidos individualmente a seguir.

Para a seleção de conexões mais frequentes, foram analisados os números de conexões para cada classe em ordem decrescente. Os quadros 9, 10, 11, 12, 13 e 14 resumem estas análises, apresentando um *ranking* de conexões mais frequentes para as classes diagnosticado e controle (amigos nos quadros 9 e 10, seguidores nos quadros 11 e 12 e menções nos quadros 13 e 14). Por exemplo, conforme quadro 9, a conexão de amizade mais frequente do córpus (na posição 1º) ocorre 578 vezes com usuários da classe diagnosticado para a tarefa de predição de depressão. De forma semelhante, conforme quadro 10, a conexão de amizade mais frequente com usuários da classe controle ocorre 3984 vezes, também para a tarefa de predição de depressão.

Observou-se que as conexões na parte inferior do *ranking* são extremamente esparsas, e possivelmente pouco relevantes para uso como características de AM. Neste contexto e, em virtude do tamanho dos grafos, optou-se por uma seleção inicial de 15 mil conexões mais frequentes para cada classe (diagnosticado e controle). Para cada uma das redes, a

Quadro 9 – Ranking de conexões mais frequentes da rede de amigos para a classe diagnosticado

Posição da conexão	Usuários Diagnosticados	
	Depressão	Ansiedade
1º	578	815
2º	550	767
3º	514	729
4º	371	554
5º	370	498
6º	292	473
7º	270	393
8º	252	379
9º	248	367
10º	244	365
100º	110	154
500º	45	63
1000º	30	40
3500º	13	18
5000º	11	14
10000º	8	10
15000º	6	8
20000º	5	7
25000º	5	6

Quadro 10 – Ranking de conexões mais frequentes da rede de amigos para a classe de controle

Posição da conexão	Usuários de Controle	
	Depressão	Ansiedade
1º	3984	5146
2º	3904	5037
3º	3773	4904
4º	2504	3201
5º	2339	3018
6º	2338	2990
7º	2308	2984
8º	2161	2796
9º	2089	2707
10º	2067	2651
100º	807	1037
500º	339	435
1000º	222	286
3500º	91	119
5000º	70	92
10000º	43	57
15000º	33	44
20000º	28	37
25000º	24	32

Fonte – Rafael Lage de Oliveira, 2023

lista de conexões selecionadas da classe diagnosticado foi unificada à lista de conexões da classe controle, formando uma lista única de 30 mil conexões. Após a eliminação de duplicidades de conexões formadas com as duas classes, esta seleção resultou em redes com tamanhos significativamente menores, conforme dados das tabelas 6 e 7 para depressão e ansiedade, respectivamente. Estas tabelas apresentam a quantidade de nós de cada uma das redes (total de usuários a classificar e suas conexões), a quantidade de arestas (relações entre todos os nós), e, por fim, o número total de conexões.

Tabela 6 – Dados descritivos das redes originais em comparação às redes formadas pelas 15 mil conexões mais frequentes por classe para a tarefa de predição de depressão

Rede	Original			Seleção de conexões mais frequentes		
	Nós	Arestas	Conexões	Nós	Arestas	Conexões
Amigos	3.692.134	9.338.030	3.678.662	34.860	1.719.519	21.388
Seguidores	6.249.876	11.775.705	6.236.404	36.391	814.883	22.919
Menções	923.652	1.649.334	910.180	37.392	463.329	23.920

Fonte – Rafael Lage de Oliveira, 2023

Uma vez que os dados permaneceram esparsos após a aplicação do método de seleção inicial, um segundo método de seleção de características foi introduzido. O método

Quadro 11 – Ranking de conexões mais frequentes da rede de seguidores para a classe diagnosticado

Posição da conexão	Usuários Diagnosticados	
	Depressão	Ansiedade
1º	101	167
2º	93	142
3º	67	108
4º	64	103
5º	61	85
6º	53	73
7º	50	72
8º	49	71
9º	49	67
10º	44	60
100º	24	34
500º	15	21
1000º	13	17
3500º	9	11
5000º	8	10
10000º	6	8
15000º	5	6
20000º	4	6
25000º	4	5

Quadro 12 – Ranking de conexões mais frequentes da rede de seguidores para a classe de controle

Posição da conexão	Usuários de Controle	
	Depressão	Ansiedade
1º	1106	1559
2º	764	1112
3º	640	867
4º	580	785
5º	463	632
6º	444	618
7º	408	575
8º	402	573
9º	373	551
10º	373	517
100º	151	211
500º	84	117
1000º	66	90
3500º	42	56
5000º	36	49
10000º	28	37
15000º	23	31
20000º	21	28
25000º	19	25

Fonte – Rafael Lage de Oliveira, 2023

Tabela 7 – Dados descritivos das redes originais em comparação às redes formadas pelas 15 mil conexões mais frequentes por classe para a tarefa de predição de ansiedade

Rede	Original			Seleção de conexões mais frequentes		
	Nós	Arestas	Conexões	Nós	Arestas	Conexões
Amigos	4.466.587	12.403.709	4.448.835	38.547	2.251.127	20.795
Seguidores	7.846.801	16.160.129	7.829.049	39.570	1.085.014	21.818
Menções	1.088.319	2.033.469	1.070.567	40.958	563.263	23.206

Fonte – Rafael Lage de Oliveira, 2023

de seleção univariada baseado em ANOVA e medida F1 como função de avaliação foi utilizado para tal tarefa, o qual selecionou as K características mais relevantes do conjunto. Para a definição do K, foram realizados testes com valores a partir do número máximo de conexões disponíveis em cada uma das redes, onde, em cada execução de teste, o número de conexões era decrescido em 500 unidades. As métricas de avaliação dos resultados para cada execução de teste serão apresentadas em forma de gráficos no apêndice A para as tarefas de predição de depressão e transtornos de ansiedade.

A partir dos resultados gerados, foram selecionados os valores de K que apresentaram os melhores resultados de F1 gerados por um classificador baseado em Regressão Logística.

Quadro 13 – Ranking de conexões mais frequentes da rede de menções para a classe diagnosticado

Posição da conexão	Usuários Diagnosticados	
	Depressão	Ansiedade
1º	337	473
2º	261	322
3º	208	264
4º	190	231
5º	152	219
6º	152	202
7º	150	200
8º	127	195
9º	118	186
10º	114	159
100º	36	47
500º	13	17
1000º	8	10
3500º	4	4
5000º	3	3
10000º	2	2
15000º	2	2
20000º	1	2
25000º	1	1

Quadro 14 – Ranking de conexões mais frequentes da rede de menções para a classe de controle

Posição da conexão	Usuários de Controle	
	Depressão	Ansiedade
1º	2061	2659
2º	1785	2197
3º	1376	1776
4º	1228	1519
5º	911	1299
6º	877	1156
7º	875	1130
8º	862	1015
9º	752	1007
10º	739	952
100º	270	326
500º	106	129
1000º	64	79
3500º	23	27
5000º	17	21
10000º	10	12
15000º	7	9
20000º	6	7
25000º	5	6

Fonte – Rafael Lage de Oliveira, 2023

A tabela 8 apresenta as K principais características de cada modelo nos conjuntos de depressão e ansiedade.

Tabela 8 – K principais características selecionadas pelo método de seleção univariada

Modelo	K principais características	
	Depressão	Ansiedade
amigos.univ.reglog	14500	17000
seguidores.univ.reglog	13000	21000
menções.univ.reglog	19500	10500

Fonte – Rafael Lage de Oliveira, 2023

5.1.2 Modelos de redes de relações baseados em *node2vec*

Nesta seção serão apresentados os modelos *amigos.n2v.svm*, *seguidores.n2v.svm*, *menções.n2v.svm* e *topmenções.n2v.svm* construídos por meio de redes de relações com base no *node2vec* (GROVER; LESKOVEC, 2016) como método de extração de características.

A representação de conexões em matrizes de adjacências e seleção univariada para extração das principais características (utilizadas nos experimentos descritos na seção 5.1.1) não prioriza a natureza complexa e não-linear de redes de conexões e interação, além de trazer uma complexidade adicional para o modelo por gerar vetor de características de espaço dimensional elevado. Como forma de tratar estes problemas, preservar a vizinhança dos nós e alcançar um número maior de relações, foram realizados experimentos utilizando representações de redes por meio de *embeddings* de nós com base no *node2vec*. Neste tipo de arquitetura, cada usuário a classificar do córpus é representado por um vetor de características de mais baixo espaço dimensional em relação ao vetor original que considera todas as conexões da rede.

As redes de amigos, seguidores e menções foram inicialmente modeladas a partir dos grafos não-dirigidos originais e completos utilizados também nos modelos descritos na seção 5.1.1. Apesar de a utilização de *embeddings* de nós possibilitar o acesso a um número maior de conexões de rede, não houve neste projeto recursos computacionais suficientes para gerar representações das redes completas para os modelos *amigos.n2v.svm*, *seguidores.n2v.svm* e *menções.n2v.svm*. Ao contrário da seleção de características inicial utilizada nos modelos descritos na seção 5.1.1 que selecionou as 15 mil conexões mais frequentes de cada classe, estes três modelos baseados em *node2vec* utilizaram um mecanismo de poda, em que foram eliminadas da rede as conexões menos frequentes com usuários diagnosticados e de controle. Como exemplo, foram eliminadas da rede de amigos do conjunto de depressão todas as conexões que possuem menos de dez ligações com usuários diagnosticados e de controle. A definição destes valores foi baseada nos melhores resultados de classificação dos modelos, onde variou-se o número mínimo de conexões entre 10 e 30 para as redes de amigos e seguidores e entre uma e duas menções a conexões da rede. A tabela 9 indica os números de conexões definidos para a poda na modelagem com *node2vec*, enquanto que as tabelas 10 e 11 apresentam a estrutura das redes após a poda de amigos, seguidores e menções em comparação às redes originais.

Tabela 9 – Número de conexões para poda nos modelos baseados em *node2vec*

Modelo	Número de conexões para poda	
	Depressão	Ansiedade
amigos.n2v.svm	10	30
seguidores.n2v.svm	10	15
menções.n2v.svm	2	2

Fonte – Rafael Lage de Oliveira, 2023

Tabela 10 – Estrutura das redes após poda de conexões menos frequentes para a tarefa de predição de depressão

Rede	Original			Poda de conexões menos frequentes		
	Nós	Arestas	Conexões	Nós	Arestas	Conexões
Amigos	3.692.134	9.338.030	3.678.662	116.138	3.133.850	102.666
Seguidores	6.249.876	11.775.705	6.236.404	110.000	1.941.540	96.528
Menções	923.652	1.649.334	910.180	184.652	889.113	171.180

Fonte – Rafael Lage de Oliveira, 2023

Tabela 11 – Estrutura das redes após poda de conexões menos frequentes para a tarefa de predição de ansiedade

Rede	Original			Poda de conexões menos frequentes		
	Nós	Arestas	Conexões	Nós	Arestas	Conexões
Amigos	4.466.587	12.403.709	4.448.835	45.833	2.556.652	28.081
Seguidores	7.846.801	16.160.129	7.829.049	93.421	2.280.330	75.669
Menções	1.088.319	2.033.469	1.070.567	232.052	1.150.693	214.300

Fonte – Rafael Lage de Oliveira, 2023

Variações dos parâmetros do *node2vec* foram também experimentadas com o objetivo de encontrar os melhores resultados de classificação, partindo dos valores padrão da implementação utilizada. As dimensões de *embeddings* foram variadas entre os valores 128 (valor padrão) e 64. O tamanho do passo, ou seja, o número de nós em cada passo aleatório, foi variado entre 80 (valor padrão), 30 e 15. O número de passos por nó foi variado entre 10 (valor padrão) e 20. Os melhores resultados foram computados para os parâmetros de dimensões de *embeddings* igual a 64, tamanho do passo igual a 30 (amigos) ou 80 (seguidores e menções) e o número de passos igual a 10.

Conforme será discutido na seção 5.3, os modelos *menções.univ.reglog* e *menções.n2v.svm* apresentaram desempenho superior aos demais modelos que utilizam os mesmos métodos de extração de características, com destaque para o modelo baseado em *node2vec*. Por esse motivo, optou-se por explorar a lista de top menções disponibilizada no córpus SetembroBR, detalhada na seção 2.5, para a construção do modelo *topmenções.n2v.svm*. A rede de top menções foi modelada a partir de grafos não-dirigidos e a representação das conexões por meio do *node2vec* com o uso de todas as conexões disponíveis, uma vez que

trata-se de uma rede significativamente menor que as demais, com 16651 conexões de rede. Além disso, com o objetivo de capturar a ordenação original de conexões mais mencionadas, foram atribuídos pesos simples às arestas no grafo, em que o indivíduo mais mencionado recebe o peso 20, o segundo na sequência recebe o peso 19 e assim sucessivamente. Estes pesos, contudo, não são proporcionais à frequência de menções que cada indivíduo recebeu, servindo apenas para modelar a noção de prioridade. Foram realizados experimentos com as listas completas, que incluem indivíduos comuns (ou seja, que não fazem parte da lista de conexões mais mencionadas), e com a remoção destes indivíduos, e também com a aplicação de pesos para preservar a ordenação original e sem esta aplicação. Os melhores resultados foram obtidos para as redes em que foram excluídos os usuários comuns e nas quais foram aplicados os pesos de ordenação. Estes resultados serão apresentados na seção 5.3.

5.1.3 Modelo temporal

O modelo *horario.svm* foi construído por meio da seleção de características que expressam a atividade dos usuários a classificar do córpus ao longo do tempo na rede social. Foram extraídas 24 características (*hour_00* a *hour_23*) presentes no córpus SetembroBR que indicam a quantidade de postagens feitas a cada hora do dia, conforme descrito na seção 2.5. Estas características foram utilizadas sem qualquer tipo de pré-processamento, i.e., na forma original em que foram disponibilizadas no conjunto de dados.

5.1.4 Modelo linguístico de *baseline*

Embora o uso de características linguísticas não seja o foco deste trabalho, um modelo baseado em conteúdo textual foi definido como *baseline* para investigação da questão *Q1* e comparação aos demais modelos individuais desenvolvidos. A versão do LIWC para o português (FILHO; PARDO; ALUISIO, 2013) foi utilizada para a construção do modelo *liwc.reglog* que faz uso de características psicolinguísticas extraídas a partir da porção textual do córpus SetembroBR, tanto para o conjunto de dados de depressão quanto para o conjunto de ansiedade. Para cada usuário a classificar do córpus, foi criado um único documento, resultado da concatenação de todos os seus *tweets*. As características LIWC foram então computadas a partir da criação de um vetor de 64 posições correspondentes

às categorias LIWC e, para cada palavra do documento, foram incrementadas todas as categorias às quais a palavra pertence. Ao final do documento, estas contagens são divididas pela quantidade de palavras do mesmo, criando assim uma proporção de palavras de cada categoria no intervalo 0 e 1. Para a seleção final, foram consideradas todas as categorias disponíveis nesta versão do LIWC, resultando em 64 características psicolinguísticas relacionadas diretamente ao conteúdo textual gerado pelos usuários a classificar do córpus. O modelo *liwc.reglog*, que faz uso destas 64 características, foi construído como modelo de *baseline* para comparação aos modelos que fazem uso de características extra-linguísticas descritos nas seções 5.1.1, 5.1.2 e 5.1.3.

5.2 Procedimento

Os modelos *amigos.univ.reglog*, *seguidores.univ.reglog* e *menções.univ.reglog*, baseados em seleção univariada para a extração das características, foram construídos com o método de Regressão Logística por se tratar de um técnica pouco complexa em termos de parametrização e recursos de processamento. Uma vez que o foco do experimento é avaliar a relevância das características extraídas para a predição de transtornos de saúde mental, este algoritmo atende aos objetivos iniciais propostos. A baixa complexidade na parametrização e em recursos de processamento são aspectos relevantes no contexto de um espaço dimensional alto de características, como é o caso dos experimentos com seleção univariada. Como principais parâmetros do algoritmo, foram considerados o mecanismo de regularização L2, o peso de classe do tipo balanceado e o número máximo de 400 iterações.

Os modelos *amigos.n2v.svm*, *seguidores.n2v.svm*, *menções.n2v.svm* e *top-menções.n2v.svm*, que utilizam *node2vec* para a representação das redes, foram construídos com base no método de SVM. Foram ainda realizados alguns experimentos com métodos baseados em redes neurais simples (MLP), porém os resultados obtidos não foram satisfatórios e foram descartados. A utilização de SVM nestes modelos se tornou viável computacionalmente, uma vez que o espaço dimensional de características foi reduzido com o uso *node2vec* em comparação aos modelos que utilizaram seleção univariada. Como principais parâmetros do algoritmo de SVM, foram considerados o parâmetro de regularização igual a 1, núcleo RBF e foram aplicados pesos de classes para ajuste do desbalanceamento.

O modelo *horario.svm*, que faz uso de características temporais de horário de postagem foi construído com o método de SVM, com parâmetro de regularização igual a 1, núcleo linear e foram aplicados pesos de classes para ajuste do desbalanceamento. Foram também realizados testes com o núcleo RBF em que foram gerados resultados inferiores ao modelo com núcleo linear e, portanto, estes resultados foram desconsiderados.

O modelo de *baseline liwc.reglog*, que faz uso de características linguísticas, foi construído com o método de Regressão Logística. Como principais parâmetros, foram considerados o mecanismo de regularização L2, o peso de classe do tipo balanceado e o número máximo de 400 iterações.

Para a avaliação de qualidade dos modelos, foi utilizada a divisão em treinamento (80%) e teste (20%) originalmente disponibilizada pelo córpus, descrito na seção 2.5. Foram utilizadas as métricas precisão, revocação e medida F1 para a avaliação dos resultados para cada classe e foram construídas matrizes de confusão que permitem a visualização do desempenho dos classificadores de forma mais detalhada. Além disso, a significância estatística entre pares de modelos, quando pertinente, foi avaliada por meio do teste de McNemar (MCNEMAR, 1947).

5.3 Resultados

Esta seção apresenta os resultados dos experimentos que utilizam modelos individuais para a predição de depressão e ansiedade. De forma complementar, os resultados destes classificadores serão apresentados no apêndice B na forma de matrizes de confusão e os testes completos de significância estatística realizados serão apresentados no apêndice D.

5.3.1 Predição de depressão

A tabela 12 apresenta os resultados dos modelos individuais apresentados na seção 5.1 para a tarefa de predição de depressão. O melhor resultado de F1 para a classe diagnosticado está em destaque.

Dentre todos os modelos individuais desenvolvidos para a tarefa de predição de depressão, os melhores resultados foram obtidos para o modelo *menções.n2v.svm*. A diferença entre o modelo *menções.n2v.svm* e o modelo *amigos.n2v.svm*, que obteve o

Tabela 12 – Resultados dos modelos individuais para a tarefa de predição de depressão

Modelo	Diagnosticado			Controle		
	Precisão	**Revocação**	**F1**	**Precisão**	**Revocação**	**F1**
amigos.univ.reglog	25,34%	43,92%	32,14%	91,05%	81,52%	86,02%
seguidores.univ.reglog	21,60%	60,24%	31,79%	92,37%	68,76%	78,83%
menções.univ.reglog	35,62%	32,34%	33,90%	90,46%	91,65%	91,05%
amigos.n2v.svm	22,90%	72,70%	34,83%	94,34%	65,03%	76,99%
seguidores.n2v.svm	22,81%	69,44%	34,34%	93,83%	66,43%	77,79%
menções.n2v.svm	27,47%	70,92%	**39,60%**	94,63%	73,25%	82,58%
topmenções.n2v.svm	20,67%	54,90%	30,03%	91,56%	69,90%	79,28%
horario.svm	16,08%	40,36%	22,99%	89,14%	69,90%	78,36%
liwc.reglog	18,52%	67,06%	29,03%	92,48%	57,86%	71,19%

Fonte – Rafael Lage de Oliveira, 2023

segundo melhor desempenho, é estatisticamente significante (χ =45,891; p<0,001). Este resultado oferece suporte à questão de pesquisa *Q1*, que visa a identificar as características extra-linguísticas que possam contribuir significativamente para a predição de transtornos de saúde mental em usuários de redes sociais, e evidencia que as interações por meio de menções a outras conexões de rede são altamente preditivas de depressão.

Os resultados apresentados evidenciam ainda que os modelos que utilizam os métodos de *embeddings* de nós por meio do *node2vec* foram superiores aos modelos semelhantes que fazem uso de seleção univariada para a extração de características baseadas em rede. A diferença entre os modelos *amigos.univ.reglog* e *amigos.n2v.svm* é estatisticamente significante (χ =103,776; p<0,001), assim como entre os modelos *menções.univ.reglog* e *menções.n2v.svm* (χ =122,412; p<0,001). A diferença entre os modelos *seguidores.univ.reglog* e *seguidores.n2v.svm* não é estatisticamente significante.

5.3.2 Predição de transtornos de ansiedade

A tabela 13 apresenta os resultados dos modelos individuais apresentados na seção 5.1 para a tarefa de predição de ansiedade. O melhor resultado de F1 para a classe diagnosticado está em destaque.

Assim como ocorreu para a tarefa de predição de depressão, os melhores resultados dos modelos individuais para a tarefa de predição de ansiedade foram obtidos para o modelo *menções.n2v.svm*. A diferença entre o modelo *menções.n2v.svm* e o modelo *amigos.n2v.svm*, que obteve o segundo melhor desempenho, é estatisticamente significante (χ =133,500; p<0,001). Este resultado sugere que as interações por meio de menções a outras conexões

Tabela 13 – Resultados dos modelos individuais para a tarefa de predição de ansiedade

Modelo	**Diagnosticado**			**Controle**		
	Precisão	**Revocação**	**F1**	**Precisão**	**Revocação**	**F1**
amigos.univ.reglog	23,26%	42,79%	30,13%	90,71%	79,83%	84,92%
seguidores.univ.reglog	20,00%	49,55%	28,50%	90,86%	71,69%	80,14%
menções.univ.reglog	30,29%	30,63%	30,46%	90,07%	89,93%	90,00%
amigos.n2v.svm	18,88%	75,00%	30,16%	93,79%	53,96%	68,50%
seguidores.n2v.svm	19,30%	68,69%	30,14%	92,95%	58,98%	72,17%
menções.n2v.svm	22,67%	67,34%	**33,92%**	93,51%	67,18%	78,19%
topmenções.n2v.svm	19,47%	51,58%	28,27%	90,95%	69,53%	78,81%
horario.svm	17,95%	54,05%	26,95%	90,79%	64,70%	75,56%
liwc.reglog	18,25%	66,67%	28,65%	92,33%	57,34%	70,74%

Fonte – Rafael Lage de Oliveira, 2023

de rede são altamente preditivas de transtornos de ansiedade, oferecendo suporte à questão de pesquisa *Q1*, que visa a identificar as características extra-linguísticas que possam contribuir significativamente para a predição de transtornos de saúde mental em usuários de redes sociais.

Os resultados apresentados evidenciam ainda que os modelos que utilizam os métodos de *embeddings* de nós por meio do *node2vec* foram superiores aos modelos semelhantes que fazem uso de seleção univariada para a extração de características baseadas em rede. A diferença entre os modelos *amigos.univ.reglog* e *amigos.n2v.svm* é estatisticamente significante (χ =335,335; p<0,001), assim como entre os modelos *seguidores.univ.reglog* e *seguidores.n2v.svm* (χ =79,834; p<0,001) e entre os modelos *menções.univ.reglog* e *menções.n2v.svm* (χ =269,023; p<0,001).

5.4 Considerações

Neste capítulo foram apresentados os experimentos realizados no contexto de predição de transtornos de saúde mental a partir de dados de usuários de redes sociais que utilizam modelos individuais para as tarefas de predição de depressão e transtornos de ansiedade. Estes modelos fizeram uso de características extra-linguísticas baseadas no comportamento dos usuários nas redes sociais com o objetivo de verificar se estas características podem contribuir individualmente e significativamente para a predição de transtornos de saúde mental, conforme questão *Q1* apresentada na seção 4.1.

Os resultados dos experimentos sugerem que as interações por meio de menções a outras conexões de rede são altamente efetivas para a predição de depressão e ansiedade, superando inclusive modelos mais simples baseados exclusivamente em recursos linguísticos. Estes resultados indicam ainda que é possível identificar usuários diagnosticados com estes transtornos com uma acurácia relativamente alta sem recorrer ao conteúdo textual.

Além disso, os resultados dos modelos construídos por meio de características de estrutura de rede e de relacionamento indicam que o método de extração de características baseado em *node2vec* foi mais efetivo para as tarefas de predição de depressão e ansiedade quando comparados ao método de seleção univariada, possivelmente beneficiado pelo acesso a um número maior de conexões de rede e ao contexto em que estas conexões estão inseridas.

6 Experimentos com uso de classificadores combinados

Este capítulo apresenta os experimentos realizados no contexto de predição de transtornos de saúde mental a partir de dados de usuários de redes sociais digitais que visam a responder a questão *Q2* apresentada na seção 4.1: *Modelos computacionais que combinam diferentes tipos de características comportamentais de usuários de redes sociais possuem resultados significativamente superiores aos modelos de classificação de transtornos de saúde mental baseados apenas em características individuais?* Para este fim, foram construídos modelos que combinam as características extra-linguísticas e predições dos modelos individuais apresentados na seção 5.1 de forma a verificar se estes modelos combinados possuem um desempenho preditivo significativamente superior aos modelos individuais para as tarefas de predição de depressão e transtornos de ansiedade em usuários do *Twitter*. Para a combinação das características e predições, foram utilizados métodos de concatenação dos vetores de características e comitês de classificadores. As seções a seguir apresentam o detalhamento da construção destes modelos.

6.1 Modelos desenvolvidos

Nesta seção serão apresentados os modelos[1] que combinam características e predições dos modelos individuais descritos na seção 5.1 para as tarefas de predição de depressão e ansiedade desenvolvidos com o objetivo de responder a questão *Q2* apresentada na seção 4.1.

A construção destes modelos teve como princípio explorar a combinação de características extra-linguísticas de diferentes naturezas (i.e., rede, interação e atividade) e, para isso, foram utilizados três métodos de combinação: concatenação de características, voto majoritário e *stacking*. Para a combinação de características e predições dos modelos individuais, foram considerados os modelos de estrutura de rede e de interação baseados em *node2vec* em detrimento aos modelos baseados em seleção univariada, uma vez que estes apresentaram resultados superiores, conforme descrito na seção 5.3. As combinações de características e estratégias computacionais consideradas deram origem aos modelos relacionados no quadro 15. Estes modelos serão discutidos em mais detalhes nas próximas seções.

[1] ⟨https://github.com/rlagedo/ExtraLinguistic_SetembroBR⟩

Quadro 15 – Lista de modelos desenvolvidos que fazem uso de classificadores combinados (A=Amigos, S=Seguidores, M=Menções, T=Top Menções, H=Horário)

Modelo	**Método de combinação**	**Características**
concat.ASM	Concatenação de características	A; S; M
concat.ASMT		A; S; M; T
concat.ASMH		A; S; M; H
concat.ASMTH		A; S; M; T; H
vot.ASM	Voto majoritário	A; S; M
vot.ASMT		A; S; M; T
vot.ASMH		A; S; M; H
vot.ASMTH		A; S; M; T; H
stack.ASM	*Stacking*	A; S; M
stack.ASMT		A;S; M; T
stack.ASMH		A; S; M; H
stack.ASMTH		A; S; M; T; H

Fonte – Rafael Lage de Oliveira, 2023

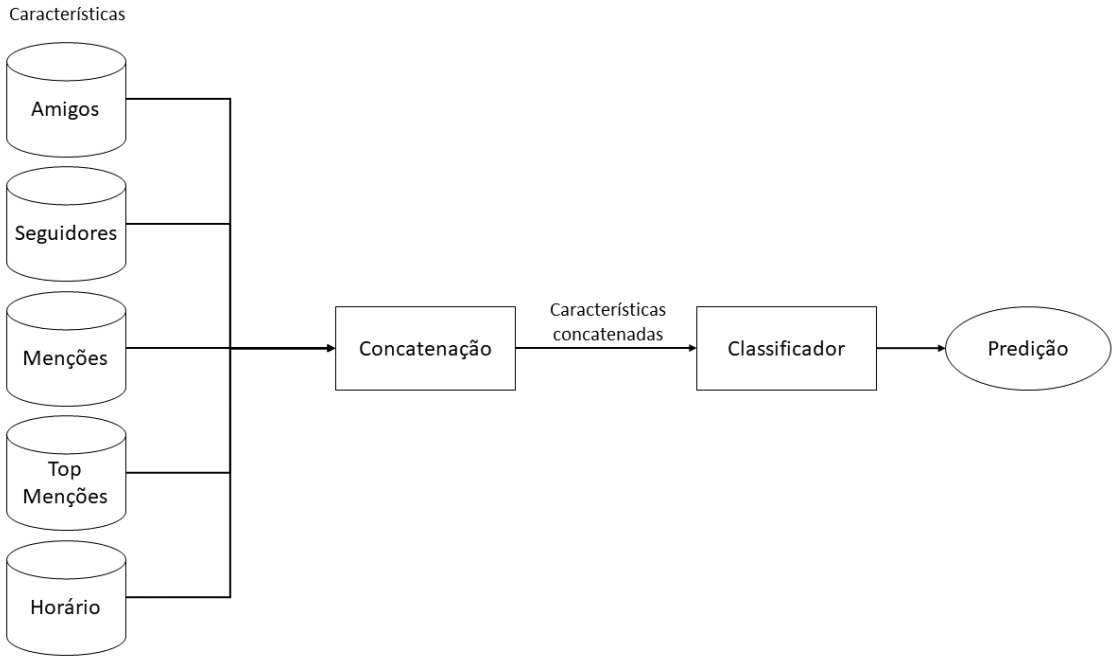
6.1.1 Modelos de predição baseados em concatenação de características

Nesta seção serão apresentados os modelos *concat.ASM*, *concat.ASMT*, *concat.ASMH* e *concat.ASMTH* que fazem uso da concatenação de características extra-linguísticas disponíveis no córpus SetembroBR e exploradas pelos modelos individuais descritos na seção 5.1.

A concatenação de características é comum em aplicações de AM que visam a agregar diferentes tipos de conhecimentos em tarefas de classificação (CHENG; CHEN, 2022; SAWHNEY *et al.*, 2021; BI; LI; WANG, 2021; GHOSH; ANWAR, 2021) e consiste em mapear diferentes vetores de características em um único vetor de mais alto espaço dimensional. Essa abordagem permite que o conhecimento combinado contribua para o reconhecimento de padrões que não seria possível apenas com conjuntos individuais de características. A figura 15 ilustra a combinação das características amigos (64 dimensões), seguidores (64 dimensões), menções (64 dimensões), top menções (64 dimensões) e horário (24 dimensões) em um único vetor de 4*64 + 24 = 280 dimensões.

A estratégia de combinação de características adotada neste trabalho priorizou os melhores resultados dos modelos individuais apresentados na seção 5.3. O modelo *concat.ASM* foi construído a partir da concatenação dos três tipos de características cujos modelos individuais tiveram os melhores resultados de F1 para a classe diagnosticado, i.e., amigos, seguidores e menções, explorando a combinação de características de estrutura de rede e de interação. Os demais modelos desenvolvidos com base no método de concatenação

Figura 15 – Combinação de características por meio de concatenação



Fonte – Rafael Lage de Oliveira, 2023

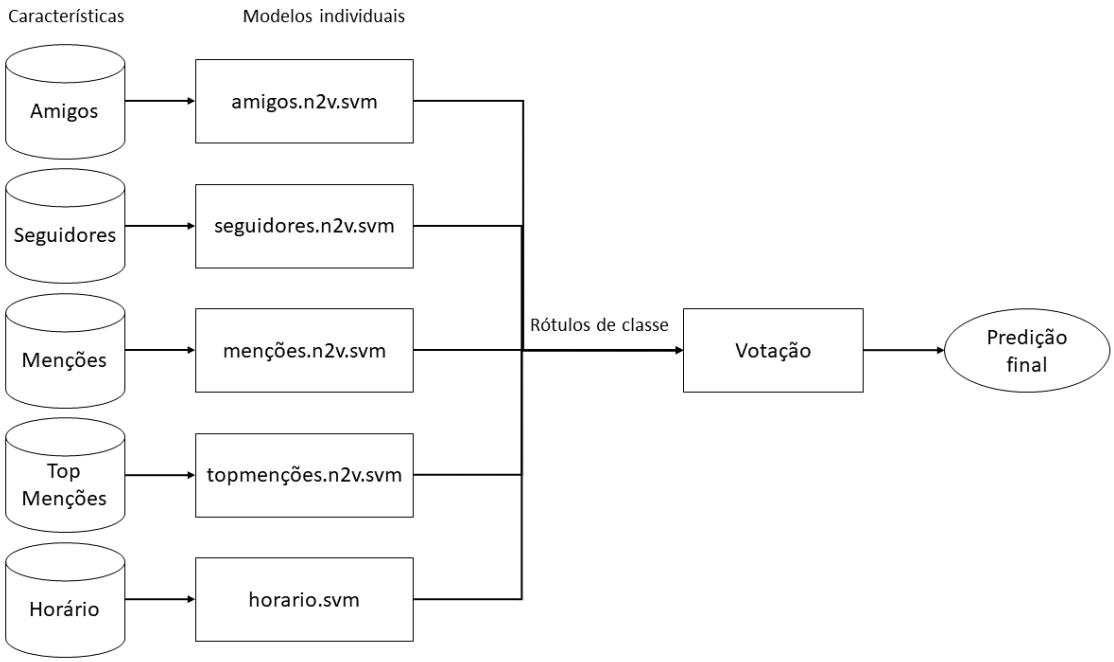
de características buscaram agregar novos conhecimentos ao modelo *concat.ASM* com o objetivo de avaliar a contribuição destes para a tarefa de predição. O modelo *concat.ASMT* uniu as top menções aos amigos, seguidores e menções, preservando a combinação de características de estrutura de rede e redes de interação do modelo *concat.ASM*. O modelo *concat.ASMH* combinou o conhecimento temporal às características de estrutura de rede e de interação com a inclusão do horário de postagem. Por fim, o modelo *concat.ASMTH* concatenou todas as características extra-linguísticas exploradas pelos modelos individuais, i.e., amigos, seguidores, menções, top menções e horário de postagem.

6.1.2 Modelos de combinação de classificadores baseados em votação

Nesta seção serão apresentados os modelos *vot.ASM*, *vot.ASMT*, *vot.ASMH* e *vot.ASMTH* que fazem uso do método de votação para a predição de depressão e ansiedade em redes sociais.

O método de votação faz parte da estratégia de combinar diferentes modelos por meio de um comitê de classificadores, conforme detalhado na seção 2.4.1. A implementação realizada neste trabalho considera os rótulos de classes preditos pelos modelos individuais descritos na seção 5.1 e prediz a classe final de um usuário do córpus por meio de uma votação por maioria, ou seja, a classe mais frequente dentre todos os classificadores individuais considerados será escolhida como classe final do usuário. Por exemplo, se dois classificadores predizem a classe diagnosticado para um determinado usuário e um terceiro classificador prediz a classe controle, a classe final para este usuário, por maioria, será diagnosticado. Em casos em que há empate de votos, a classe negativa (controle) será considerada como classe final pois ela é, por definição, majoritária no córpus. A figura 16 ilustra a combinação de classificadores por meio do método de votação.

Figura 16 – Combinação de classificadores por meio de votação



Fonte – Rafael Lage de Oliveira, 2023

Assim como nos modelos de classificação baseados em concatenação de características, a estratégia de combinação utilizada priorizou os melhores resultados de F1

para a classe diagnosticado dos modelos individuais apresentados na seção 5.3. O modelo *vot.ASM* utilizou os rótulos de classe preditos pelos modelos *amigos.n2v.svm*, *seguidores.n2v.svm* e *menções.n2v.svm* para selecionar, por voto de maioria, a classe final dos usuários do córpus, combinando classificadores baseados em estrutura de rede e interação. Estes classificadores foram ainda combinados no modelo *vot.ASMT*, onde foram consideradas também as predições do modelo *topmenções.n2v.svm* baseado em redes de interação. O modelo *vot.ASMH*, por sua vez, combinou as predições dos modelos *amigos.n2v.svm*, *seguidores.n2v.svm*, *menções.n2v.svm* e *horario.svm*, que consideram, além de características de estrutura de rede e de interação, a atividade de postagens no tempo dos usuários a classificar do córpus. Por fim, o modelo *vot.ASMTH* combinou os rótulos atribuídos pelos modelos *amigos.n2v.svm*, *seguidores.n2v.svm*, *menções.n2v.svm*, *topmenções.n2v.svm* e *horario.svm*, incorporando todos os tipos de conhecimento extra-linguísticos explorados pelos modelos individuais descritos na seção 5.1.

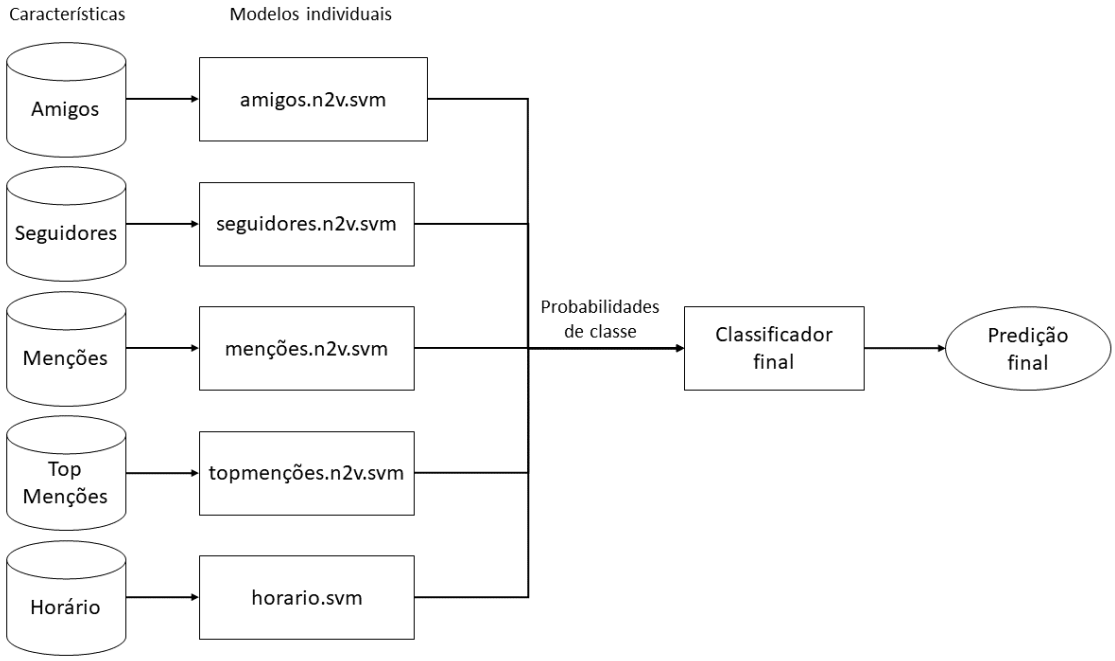
6.1.3 Modelos de combinação de classificadores baseados em *stacking*

Nesta seção serão apresentados os modelos *stack.ASM*, *stack.ASMT*, *stack.ASMH* e *stack.ASMTH* que fazem uso do método de *stacking* para a predição de depressão e ansiedade em redes sociais.

Conforme descrito na seção 2.4.2, o *stacking* é um método de comitê de classificadores que busca combinar as predições de dois ou mais modelos diferentes e utilizá-las como entrada para um classificador final. O objetivo é que este classificador aprenda partes diferentes do problema por meio dos modelos individuais e alcance um desempenho superior na classificação final. A implementação utilizada neste trabalho considera as probabilidades de classe geradas pelos modelos individuais descritos na seção 5.1 como entrada para um novo classificador que fará as predições finais para cada usuário a classificar do córpus. A figura 17 ilustra a combinação de classificadores por meio de *stacking*.

O desenvolvimento dos modelos baseados em *stacking* seguiu a mesma estratégia adotada para o desenvolvimento dos modelos baseados em votação descritos na seção 6.1.2, combinando inicialmente os modelos com melhores resultados de F1 para a classe diagnosticado. Os três classificadores individuais com melhor desempenho (*amigos.n2v.svm*, *seguidores.n2v.svm* e *menções.n2v.svm*) foram combinados na construção do modelo

Figura 17 – Combinação de classificadores por meio de *stacking*



Fonte – Rafael Lage de Oliveira, 2023

stack.ASM, extraindo o conhecimento gerado pela estrutura da rede e pelas interações entre as conexões. Estes classificadores foram ainda combinados no modelo *stack.ASMT*, onde foram consideradas também as predições do modelo *topmenções.n2v.svm*, baseado em redes de interação. O modelo *stack.ASMH*, por sua vez, combinou as predições dos modelos *amigos.n2v.svm*, *seguidores.n2v.svm*, *menções.n2v.svm* e *horario.svm*, que consideram, além da estrutura de rede e das interações entre as conexões, a atividade de postagens no tempo dos usuários a classificar do córpus. Por fim, o modelo *stack.ASMTH* combinou as probabilidades de classe atribuídas pelos modelos *amigos.n2v.svm*, *seguidores.n2v.svm*, *menções.n2v.svm*, *topmenções.n2v.svm* e *horario.svm*, incorporando todos os tipos de conhecimento extra-linguísticos explorados pelos modelos individuais descritos na seção 5.1.

6.2 Procedimento

Os modelos baseados em concatenação de características foram construídos com o método de SVM, seguindo o mesmo padrão dos modelos individuais que tiveram melhor desempenho, conforme apresentado na seção 5.3. Foram considerados o parâmetro de regularização igual a 1, núcleo RBF e foram aplicados pesos de classes para ajuste do

desbalanceamento. Os modelos baseados em um comitê de classificadores que fazem uso de votação não utilizam um algoritmo de AM específico, visto que o método explora os algoritmos individuais de cada modelo de entrada e faz a predição final por meio de uma decisão final por voto majoritário. Os modelos baseados em um comitê de classificadores que fazem uso de *stacking* também exploram os algoritmos particulares de cada modelo individual de entrada, contudo, há também um classificador final que é treinado a partir das probabilidades de classe geradas por cada um destes modelos individuais. Este classificador final foi construído com o método de SVM, considerando o parâmetro de regularização igual a 1, núcleo RBF e foram aplicados pesos de classes para ajuste do desbalanceamento.

Assim como nos experimentos do capítulo anterior, para a avaliação de qualidade dos modelos, foi utilizada a divisão em treinamento (80%) e teste (20%) originalmente disponibilizada pelo córpus e descrito na seção 2.5. Foram utilizadas as métricas precisão, revocação e medida F1 para a avaliação dos resultados para cada classe e foram construídas matrizes de confusão que permitem a visualização do desempenho dos classificadores de forma mais detalhada. Além disso, a significância estatística entre pares de modelos, quando pertinente, foi avaliada por meio do teste de McNemar (MCNEMAR, 1947).

6.3 Resultados

Esta seção apresenta os resultados dos experimentos que utilizam modelos combinados para as tarefas de predição de depressão e transtornos de ansiedade. Os resultados dos modelos individuais apresentados na seção 5.3 serão também reproduzidos nesta seção para facilitar a comparação entre todos os modelos desenvolvidos. De forma complementar, os resultados destes classificadores combinados serão apresentados no apêndice C na forma de matrizes de confusão e os testes completos de significância estatística realizados serão apresentados no apêndice D.

6.3.1 Predição de depressão

A tabela 14 apresenta os resultados dos modelos que fazem uso de classificadores combinados para a tarefa de predição de depressão (parte superior da tabela) em conjunto com os resultados dos modelos individuais apresentados na seção 5.3.1 (parte inferior

da tabela). Esta tabela representa os resultados finais dos experimentos realizados neste trabalho para a tarefa de predição de depressão, em que são comparados os principais modelos individuais e combinados desenvolvidos. Os modelos foram ordenados pelos melhores resultados para a medida F1 da classe diagnosticado. O melhor resultado para esta métrica aparece em destaque.

Tabela 14 – Resultados finais para a tarefa de predição de depressão, com os modelos combinados na parte superior e os modelos individuais na parte inferior

Modelo	**Método**	**Diagnosticado**			**Controle**		
		Precisão	**Revocação**	**F1**	**Precisão**	**Revocação**	**F1**
stack.ASMT	SVM (núcleo RBF)	29,45%	69,73%	**41,41%**	94,63%	76,13%	84,38%
stack.ASMTH	SVM (núcleo RBF)	29,05%	69,73%	41,01%	94,59%	75,67%	84,08%
stack.ASM	SVM (núcleo RBF)	28,73%	70,33%	40,79%	94,66%	75,07%	83,74%
concat.ASMT	SVM (núcleo RBF)	28,26%	72,70%	40,70%	94,97%	73,63%	82,95%
concat.ASM	SVM (núcleo RBF)	27,32%	74,18%	39,94%	95,12%	71,81%	81,84%
stack.ASMH	SVM (núcleo RBF)	28,32%	67,06%	39,82%	94,15%	75,75%	83,96%
vot.ASMT	voto por maioria	29,50%	60,83%	39,73%	93,40%	79,23%	85,73%
vot.ASMTH	voto por maioria	29,46%	53,41%	37,97%	92,47%	81,73%	86,77%
vot.ASMH	voto por maioria	35,66%	39,47%	37,46%	91,22%	89,83%	90,52%
vot.ASM	voto por maioria	24,65%	74,18%	37,01%	94,83%	67,61%	78,94%
concat.ASMH	SVM (núcleo RBF)	50,00%	0,30%	0,59%	87,53%	99,96%	93,33%
concat.ASMTH	SVM (núcleo RBF)	25,00%	0,30%	0,59%	87,52%	99,87%	93,29%
menções.n2v.svm	SVM (núcleo RBF)	27,47%	70,92%	39,60%	94,63%	73,25%	82,58%
amigos.n2v.svm	SVM (núcleo RBF)	22,90%	72,70%	34,83%	94,34%	65,03%	76,99%
seguidores.n2v.svm	SVM (núcleo RBF)	22,81%	69,44%	34,34%	93,83%	66,43%	77,79%
topmenções.n2v.svm	SVM (núcleo RBF)	20,67%	54,90%	30,03%	91,56%	69,90%	79,28%
liwc.reglog	Regressão Logística	18,52%	67,06%	29,03%	92,48%	57,86%	71,19%
horario.svm	SVM (núcleo linear)	16,08%	40,36%	22,99%	89,14%	69,90%	78,36%

Fonte – Rafael Lage de Oliveira, 2023

É possível observar que *stacking* e concatenação de características foram as estratégias de combinação que apresentaram os melhores resultados para a tarefa de predição de depressão, com destaque ao método de *stacking*. A diferença entre estas estratégias é estatisticamente significante (χ =7,361; p<0,01), o que pode ser observado com a comparação dos melhores modelos de cada estratégia, i.e., os modelos *stack.ASMT* e *concat.ASMT*. A combinação do melhor modelo *stack.ASMT* com um modelo baseado apenas em aspectos psicolinguísticos do LIWC não gerou resultados significativamente superiores, ainda que com melhor medida F1 para a classe diagnosticado, sugerindo que é possível predizer depressão sem acessar o conteúdo textual dos usuários a classificar.

Os dados apresentados evidenciam também que a combinação de classificadores é geralmente superior a modelos individuais para a tarefa de predição de depressão. A diferença entre o modelo *stack.ASMT* e o melhor modelo individual *menções.n2v.svm* é estatisticamente significante (χ =12,403; p<0,001). Este resultado oferece suporte à questão de pesquisa *Q2*, que visa a identificar se modelos que combinam características

extra-linguísticas podem fornecer resultados superiores a modelos de classificação baseados apenas em características individuais.

6.3.2 Predição de transtornos de ansiedade

A tabela 15 apresenta os resultados dos modelos que fazem uso de classificadores combinados para a tarefa de predição de ansiedade (parte superior da tabela) em conjunto com os resultados dos modelos individuais apresentados na seção 5.3.2 (parte inferior da tabela). Esta tabela representa os resultados finais dos experimentos realizados neste trabalho para a tarefa de predição de ansiedade, em que são comparados os principais modelos individuais e combinados desenvolvidos. Os modelos foram ordenados pelos melhores resultados para a medida F1 da classe diagnosticado. O melhor resultado para esta métrica aparece em destaque.

Tabela 15 – Resultados finais para a tarefa de predição de ansiedade, com os modelos combinados na parte superior e os modelos individuais na parte inferior

Modelo	**Método**	**Diagnosticado**			**Controle**		
		Precisão	**Revocação**	**F1**	**Precisão**	**Revocação**	**F1**
concat.ASMT	SVM (núcleo RBF)	24,24%	73,42%	**36,44%**	94,65%	67,21%	78,61%
stack.ASMTH	SVM (núcleo RBF)	24,51%	67,34%	35,94%	93,78%	70,37%	80,40%
stack.ASMT	SVM (núcleo RBF)	24,47%	67,34%	35,89%	93,78%	70,30%	80,36%
stack.ASMH	SVM (núcleo RBF)	23,95%	67,12%	35,31%	93,67%	69,56%	79,84%
stack.ASM	SVM (núcleo RBF)	23,44%	65,32%	34,50%	93,35%	69,53%	79,70%
vot.ASMT	voto por maioria	24,06%	60,81%	34,48%	92,84%	72,59%	81,47%
concat.ASM	SVM (núcleo RBF)	22,44%	72,52%	34,27%	94,24%	64,19%	76,36%
vot.ASMTH	voto por maioria	25,23%	50,23%	33,58%	91,72%	78,73%	84,73%
vot.ASM	voto por maioria	20,40%	72,75%	31,87%	93,85%	59,46%	72,80%
vot.ASMH	voto por maioria	26,94%	38,29%	31,63%	90,62%	85,17%	87,81%
concat.ASMTH	SVM (núcleo RBF)	40,00%	0,45%	0,89%	87,54%	99,90%	93,31%
concat.ASMH	SVM (núcleo RBF)	0,00%	0,00%	0,00%	87,50%	99,97%	93,32%
menções.n2v.svm	SVM (núcleo RBF)	22,67%	67,34%	33,92%	93,51%	67,18%	78,19%
amigos.n2v.svm	SVM (núcleo RBF)	18,88%	75,00%	30,16%	93,79%	53,96%	68,50%
seguidores.n2v.svm	SVM (núcleo RBF)	19,30%	68,69%	30,14%	92,95%	58,98%	72,17%
liwc.reglog	Regressão Logística	18,25%	66,67%	28,65%	92,33%	57,34%	70,74%
topmenções.n2v.svm	SVM (núcleo RBF)	19,47%	51,58%	28,27%	90,95%	69,53%	78,81%
horario.svm	SVM (núcleo linear)	17,95%	54,05%	26,95%	90,79%	64,70%	75,56%

Fonte – Rafael Lage de Oliveira, 2023

É possível observar que concatenação de características e *stacking* foram as estratégias de combinação que apresentaram os melhores resultados para a tarefa de predição de ansiedade, com destaque ao método de concatenação de características. A diferença entre estas estratégias é estatisticamente significante (χ =12,099; p<0,001), o que pode ser observado com a comparação dos melhores modelos de cada estratégia, i.e., os modelos *concat.ASMT* e *stack.ASMTH*. A combinação do melhor modelo *concat.ASMT* com um

modelo baseado apenas em aspectos psicolinguísticos do LIWC não gerou resultados significativamente superiores, sugerindo que é possível predizer transtornos de ansiedade sem acessar o conteúdo textual dos usuários a classificar.

Os dados apresentados evidenciam também que os melhores resultados foram gerados pelo modelo *concat.ASMT*, sendo a concatenação de características a melhor estratégia geral para a tarefa de predição de ansiedade. Contudo, a diferença entre o modelo *concat.ASMT* e o melhor modelo individual *menções.n2v.svm* não é estatisticamente significante, sugerindo que o conteúdo de interação representado pelas menções a outras conexões de rede é suficiente para a tarefa de predição de ansiedade. Este resultado oferece suporte à questão de pesquisa *Q2*, que visa a identificar se modelos que combinam características extra-linguísticas podem fornecer resultados superiores a modelos de classificação baseados apenas em características individuais.

6.4 Considerações

Neste capítulo, foram apresentados os experimentos realizados no contexto de predição de transtornos de saúde mental a partir de dados de usuários de redes sociais que utilizam modelos combinados para as tarefas de predição de depressão e transtornos de ansiedade. Estes modelos exploraram a combinação de características extra-linguísticas de diferentes naturezas (i.e., rede, interação e atividade) por meio de métodos de concatenação de vetores de características e comitês de classificadores com o objetivo de verificar se a combinação de características/classificadores pode contribuir de forma mais significativa do que as características individuais para as tarefas de predição de depressão e transtornos de ansiedade em redes sociais, conforme questão *Q2* apresentada na seção 4.1.

Os resultados dos experimentos sugerem que a combinação de múltiplas fontes de informação (principalmente aquelas provenientes da estrutura de rede e de interações entre conexões) são altamente efetivas para a predição de transtornos de saúde mental em redes sociais digitais. A combinação de amigos, seguidores, menções e top menções se mostrou a mais efetiva para as tarefas de predição, com os melhores resultados gerados pelo método de *stacking* para o problema de depressão e pelo método de concatenação de características para o problema de transtornos de ansiedade. Estes resultados sugerem ainda que a combinação de características e classificadores melhoram a robustez de modelos de predição de depressão e ansiedade em comparação aos modelos individuais.

Além disso, o conhecimento temporal (representado pelo horário de postagem) foi, de forma geral, responsável por resultados inferiores de modelos de combinação em comparação a modelos que não utilizam esta característica, sugerindo que este tipo de conhecimento não se mostrou eficaz da forma como foi empregado. Por fim, a inclusão de características psicolinguísticas aos melhores modelos que combinam características extra-linguísticas teve pouco impacto em seus desempenhos, sugerindo que o conhecimento comportamental de usuários nas redes sociais podem contribuir por si só para as tarefas de predição de depressão e transtornos de ansiedade.

7 Considerações finais

Este trabalho apresentou modelos computacionais de predição de depressão e transtornos de ansiedade que exploram o conhecimento extra-linguístico, com foco em características comportamentais de usuários e de suas conexões em redes sociais, e que contribuíram para a área da computação, nos campos da Inteligência Artificial (IA) e Processamento de Língua Natural (PLN). As principais contribuições obtidas são as seguintes:

- Modelos individuais baseados em redes de relações e características temporais
- Modelos de predição baseados em concatenação de características extra-linguísticas
- Modelos baseados em comitês de classificadores com o uso dos métodos de votação e *stacking*
- Resultados de referência para as tarefas de predição de depressão e transtornos de ansiedade no córpus SetembroBR

Os resultados apresentados neste trabalho sugerem que as interações em redes sociais (representadas pelas menções a outras conexões de rede) são altamente efetivas para a predição de depressão e ansiedade, superando inclusive modelos mais simples baseados exclusivamente em recursos linguísticos. Por outro lado, características temporais (representadas pelo horário de postagem) não tiveram resultados expressivos, sugerindo que este tipo de conhecimento não se mostrou eficaz da forma como foi empregado. Além disso, a combinação de múltiplas fontes de informação (principalmente aquelas provenientes da estrutura de rede e de interações entre conexões, como amigos, seguidores e menções) por meio de estratégias simples (concatenação de características e comitês de classificadores) foram capazes de melhorar os resultados em comparação ao uso do conhecimento individual, sugerindo que arquiteturas mais sofisticadas podem superar os resultados apresentados. Os resultados indicam ainda que é possível identificar usuários diagnosticados com transtornos de saúde mental com uma acurácia relativamente alta sem recorrer ao conteúdo textual.

Apesar dos resultados de modo geral positivos, este trabalho apresenta também algumas limitações que podem ser tratadas em trabalhos futuros. Este trabalho é baseado integralmente no córpus SetembroBR e, portanto, limita-se a dados do *Twitter* em português e demais características do córpus discutidas na seção 2.5, não constituindo necessariamente uma amostra representativa dos transtornos de saúde mental na população

em geral ou mesmo em redes sociais. Apesar do enfoque deste trabalho ser em uma única plataforma e estar sujeito às limitações do córpus, os conceitos teóricos e a abordagem dos modelos computacionais se aplicam potencialmente a redes sociais em geral.

Observa-se ainda que este trabalho foi limitado ao desenvolvimento de modelos computacionais de predição binária de usuários em redes sociais a partir de técnicas de AM supervisionado para a detecção de transtornos de saúde mental. Por se tratar de um trabalho com enfoque computacional, aspectos psicossociais não foram tratados com a mesma profundidade. Por fim, este trabalho não teve como objetivo explorar exaustivamente os diversos algoritmos e técnicas de AM, mas, sim, estudar métodos que considerem conhecimento extra-linguístico, como características comportamentais, de rede e relacionamento em redes sociais.

O presente trabalho deixa algumas oportunidades de melhorias e trabalhos futuros. Primeiramente, o desenvolvimento de modelos textuais mais robustos baseados em BERT (DEVLIN *et al.*, 2019). Segundo, a oportunidade de explorar outras características extra-linguísticas disponíveis no córpus SetembroBR, como atributos demográficos, frequência de interação, padrões de tempo e atividade expressa por meio das postagens. Terceiro, explorar métodos mais sofisticados de AM, principalmente aqueles baseados em aprendizado profundo. Por fim, a combinação de modelos e características baseados em conhecimento extra-linguístico com conhecimento linguístico extraído do conteúdo textual, para entender como multimodalidades interagem e como obter resultados ótimos dessas combinações.

Referências

AKAY, A.; DRAGOMIR, A.; ERLANDSSON, B. Assessing antidepressants using intelligent data monitoring and mining of online fora. *IEEE Journal of Biomedical and Health Informatics*, v. 20, n. 4, p. 977–986, 2016. Citado 2 vezes nas páginas 54 e 55.

ALMOUZINI, S.; KHEMAKHEM, M.; ALAGEEL, A. Detecting arabic depressed users from twitter data. In: . [S.l.]: Elsevier B.V., 2019. v. 163, p. 257–265. ISSN 18770509. Conference of 16th International Learning and Technology Conference, L and T 2019 ; Conference Date: 30 January 2019 Through 31 January 2019; Conference Code:157350. Citado na página 55.

ALVAREZ-MON, M.; BARCO, A. A. D.; LAHERA, G.; QUINTERO, J.; FERRE, F.; PEREIRA-SANCHEZ, V.; ORTUñO, F.; ALVAREZ-MON, M. Increasing interest of mass communication media and the general public in the distribution of tweets about mental disorders: Observational study. *Journal of medical Internet research*, NLM (Medline), v. 20, n. 5, p. e205, 2018. ISSN 14388871. Citado na página 54.

BERTSIMAS, D.; KING, A. Logistic regression: From art to science. *Statistical Science*, JSTOR, p. 367–384, 2017. Citado 2 vezes nas páginas 26 e 27.

BI, Y.; LI, B.; WANG, H. Detecting depression on sina microblog using depressing domain lexicon. In: *2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. [S.l.: s.n.], 2021. p. 965–970. Citado 2 vezes nas páginas 75 e 98.

CACHEDA, F.; FERNANDEZ, D.; NOVOA, F.; CARNEIRO, V. Early detection of depression: Social network analysis and random forest techniques. *Journal of Medical Internet Research*, Journal of Medical Internet Research, v. 21, n. 6, 2019. ISSN 14388871. Citado na página 53.

CHAI, Y.; WU, F.; SUN, R.; ZHANG, Z.; BAO, J.; MA, R.; PENG, Q.; WU, D.; WAN, Y.; LI, K. Predicting future alleviation of mental illness in social media: An empathy-based social network perspective. In: *2019 IEEE Intl Conf on Parallel Distributed Processing with Applications, Big Data Cloud Computing, Sustainable Computing Communications, Social Computing Networking (ISPA/BDCloud/SocialCom/SustainCom)*. [S.l.: s.n.], 2019. p. 1564–1571. Citado 5 vezes nas páginas 19, 48, 54, 55 e 79.

CHANCELLOR, S.; CHOUDHURY, M. D. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine*, Nature Publishing Group, v. 3, n. 1, p. 1–11, 2020. Citado 6 vezes nas páginas 17, 43, 46, 47, 48 e 50.

CHANG, C.-H.; SARAVIA, E.; CHEN, Y.-S. Subconscious crowdsourcing: A feasible data collection mechanism for mental disorder detection on social media. In: *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. [S.l.]: IEEE Press, 2016. (ASONAM '16), p. 374–379. ISBN 9781509028467. Citado 2 vezes nas páginas 53 e 54.

CHENG, J. C.; CHEN, A. L. P. Multimodal time-aware attention networks for depression detection. *Journal of Intelligent Information Systems*, v. 59, n. 2, p. 319 – 339, 2022. Citado 2 vezes nas páginas 76 e 98.

CHOUDHURY, M. D. Anorexia on tumblr: A characterization study. In: *Proceedings of the 5th International Conference on Digital Health 2015*. New York, NY, USA: Association for Computing Machinery, 2015. (DH '15), p. 43–50. ISBN 9781450334921. Disponível em: ⟨https://doi.org/10.1145/2750511.2750515⟩. Citado 2 vezes nas páginas 53 e 54.

CHOUDHURY, M. D.; GAMON, M.; COUNTS, S.; HORVITZ, E. Predicting depression via social media. In: *Seventh international AAAI conference on weblogs and social media*. [S.l.: s.n.], 2013. Citado 4 vezes nas páginas 41, 50, 57 e 72.

CHOUDHURY, M. D.; KICIMAN, E.; DREDZE, M.; COPPERSMITH, G.; KUMAR, M. Discovering shifts to suicidal ideation from mental health content in social media. In: . [S.l.]: Association for Computing Machinery, 2016. p. 2098–2110. ISBN 9781450333627. Conference of 34th Annual Conference on Human Factors in Computing Systems, CHI 2016 ; Conference Date: 7 May 2016 Through 12 May 2016; Conference Code:121621. Citado 2 vezes nas páginas 53 e 54.

COHAN, A.; DESMET, B.; YATES, A.; SOLDAINI, L.; MACAVANEY, S.; GOHARIAN, N. SMHD: A large-scale resource for exploring online language usage for multiple mental health conditions. *CoRR*, abs/1806.05258, 2018. Disponível em: ⟨http://arxiv.org/abs/1806.05258⟩. Citado na página 18.

COPPERSMITH, G.; DREDZE, M.; HARMAN, C. Quantifying mental health signals in twitter. In: *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Baltimore, Maryland, USA: Association for Computational Linguistics, 2014. p. 51–60. Disponível em: ⟨https://www.aclweb.org/anthology/W14-3207⟩. Citado 2 vezes nas páginas 17 e 52.

COPPERSMITH, G.; DREDZE, M.; HARMAN, C.; HOLLINGSHEAD, K.; MITCHELL, M. CLPsych 2015 shared task: Depression and PTSD on Twitter. In: *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Denver, Colorado: Association for Computational Linguistics, 2015. p. 31–39. Disponível em: ⟨https://aclanthology.org/W15-1204⟩. Citado 2 vezes nas páginas 18 e 38.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, p. 273–297, 1995. Citado 3 vezes nas páginas 27, 28 e 29.

CRAMER, J. S. The origins of logistic regression. Tinbergen Institute Working Paper, 2002. Citado na página 26.

DASARATHY, B. V.; SHEELA, B. V. A composite classifier system design: Concepts and methodology. *Proceedings of the IEEE*, IEEE, v. 67, n. 5, p. 708–713, 1979. Citado na página 34.

DEMARIS, A.; SELMAN, S. H.; DEMARIS, A.; SELMAN, S. H. Logistic regression. *Converting Data into Evidence: A Statistics Primer for the Medical Practitioner*, Springer, p. 115–136, 2013. Citado na página 26.

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. Citado 2 vezes nas páginas 23 e 109.

DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1*. [S.l.], 2000. p. 1–15. Citado na página 34.

DIJK, M. van; TREUR, J. Physical activity contagion and homophily in an adaptive social network model. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, v. 11055 LNAI, p. 87–98, 2018. ISSN 03029743. Conference of 10th International Conference on Computational Collective Intelligence, ICCCI 2018 ; Conference Date: 5 September 2018 Through 7 September 2018; Conference Code:217559. Citado 2 vezes nas páginas 54 e 55.

DUTTA, S.; CHOUDHURY, M. D. Characterizing anxiety disorders with online social and interactional networks. In: SPRINGER. *International Conference on Human-Computer Interaction*. [S.l.], 2020. p. 249–264. Citado 4 vezes nas páginas 53, 54, 55 e 61.

DUTTA, S.; MA, J.; CHOUDHURY, M. D. Measuring the impact of anxiety on online social interactions. In: *Proceedings of the International AAAI Conference on Web and Social Media*. [S.l.: s.n.], 2018. v. 12, n. 1. Citado 2 vezes nas páginas 54 e 55.

FAUSETT, L. V. *Fundamentals of neural networks: architectures, algorithms and applications*. [S.l.]: Pearson Education India, 2006. Citado na página 29.

FILHO, P. P. B.; PARDO, T. A. S.; ALUISIO, S. M. An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In: *Brazilian Symposium in Information and Human Language Technology - STIL*. [S.l.]: SBC, 2013. Citado 2 vezes nas páginas 22 e 91.

FRAGA, B. S.; SILVA, A. P. C. da; MURAI, F. Online social networks in health care: A study of mental disorders on reddit. In: *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. [S.l.: s.n.], 2018. p. 568–573. Citado 3 vezes nas páginas 53, 54 e 55.

GHOSH, S.; ANWAR, T. Depression intensity estimation via social media: A deep learning approach. *IEEE Transactions on Computational Social Systems*, v. 8, n. 6, p. 1465 – 1474, 2021. Citado 2 vezes nas páginas 75 e 98.

GIUNTINI, F. T.; MORAES, K. L. de; CAZZOLATO, M. T.; KIRCHNER, L. de F.; REIS, M. de J. D. D.; TRAINA, A. J. M.; CAMPBELL, A. T.; UEYAMA, J. Modeling and assessing the temporal behavior of emotional and depressive user interactions on social networks. *IEEE Access*, v. 9, p. 93182–93194, 2021. Citado 2 vezes nas páginas 19 e 79.

GROVER, A.; LESKOVEC, J. Node2vec: Scalable feature learning for networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 855–864. ISBN 9781450342322. Disponível em: ⟨https://doi.org/10.1145/2939672.2939754⟩. Citado 4 vezes nas páginas 23, 24, 25 e 89.

HANSEN, L. K.; SALAMON, P. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 12, n. 10, p. 993–1001, 1990. Citado 2 vezes nas páginas 34 e 35.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H.; FRIEDMAN, J. H. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer, 2009. v. 2. Citado na página 34.

HAYKIN, S. *Neural networks and learning machines, 3/E*. [S.l.]: Pearson Education India, 2009. Citado 4 vezes nas páginas 30, 31, 32 e 33.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 33.

HUSSAIN, J.; ALI, M.; BILAL, H.; AFZAL, M.; AHMAD, H.; BANOS, O.; LEE, S. Sns based predictive model for depression. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, v. 9102, p. 349–354, 2015. ISSN 03029743. Conference of 13th International Conference on Smart Homes and Health Telematics, ICOST 2015 ; Conference Date: 10 June 2015 Through 12 June 2015; Conference Code:156989. Citado na página 54.

JOACHIMS, T. Text categorization with support vector machines: Learning with many relevant features. In: SPRINGER. *Machine Learning: ECML-98: 10th European Conference on Machine Learning Chemnitz, Germany, April 21–23, 1998 Proceedings*. [S.l.], 2005. p. 137–142. Citado na página 29.

KAKULAPATI, V.; REDDY, S. M. Lexical analysis and mathematical modelling for analysing depression detection of social media reviews. *Lecture Notes in Electrical Engineering*, Springer, v. 643, p. 85–93, 2020. ISSN 18761100. Citado na página 54.

KLEINBAUM, D. G.; KLEIN, M. *Logistic regression*. [S.l.]: Springer, 2010. 1–39 p. Citado 2 vezes nas páginas 26 e 27.

KUMAR, M.; DREDZE, M.; COPPERSMITH, G.; CHOUDHURY, M. D. Detecting changes in suicide content manifested in social media following celebrity suicides. In: *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. New York, NY, USA: Association for Computing Machinery, 2015. (HT '15), p. 85–94. ISBN 9781450333955. Disponível em: ⟨https://doi.org/10.1145/2700171.2791026⟩. Citado 2 vezes nas páginas 53 e 54.

LARSEN, M. E.; BOONSTRA, T. W.; BATTERHAM, P. J.; O'DEA, B.; PARIS, C.; CHRISTENSEN, H. We feel: Mapping emotion on twitter. *IEEE Journal of Biomedical and Health Informatics*, v. 19, n. 4, p. 1246–1252, 2015. Citado na página 54.

LIBERATI, A.; ALTMAN, D. G.; TETZLAFF, J.; MULROW, C.; GØTZSCHE, P. C.; IOANNIDIS, J. P.; CLARKE, M.; DEVEREAUX, P. J.; KLEIJNEN, J.; MOHER, D. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of internal medicine*, American College of Physicians, v. 151, n. 4, p. W–65, 2009. Citado na página 46.

LIMA, C. A. de M. *Comitê de Máquinas: uma abordagem unificada empregando máquinas de vetores-suporte*. Tese (Doutorado) — Universidade Estadual de Campinas, 2004. Citado 3 vezes nas páginas 34, 35 e 36.

LIN, H.; JIA, J.; GUO, Q.; XUE, Y.; LI, Q.; HUANG, J.; CAI, L.; FENG, L. User-level psychological stress detection from social media using deep neural network. In: *Proceedings of the 22nd ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, 2014. (MM '14), p. 507–516. ISBN 9781450330633. Disponível em: ⟨https://doi.org/10.1145/2647868.2654945⟩. Citado 2 vezes nas páginas 52 e 54.

LIN, H.; JIA, J.; QIU, J.; ZHANG, Y.; SHEN, G.; XIE, L.; TANG, J.; FENG, L.; CHUA, T. Detecting stress based on social interactions in social networks. *IEEE Transactions on Knowledge & Data Engineering*, IEEE Computer Society, Los Alamitos, CA, USA, v. 29, n. 09, p. 1820–1833, sep 2017. ISSN 1558-2191. Citado 7 vezes nas páginas 18, 19, 53, 54, 55, 63 e 79.

LORENA, A. C.; CARVALHO, A. C. D. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Citado 2 vezes nas páginas 28 e 29.

LOSADA, D. E.; CRESTANI, F.; PARAPAR, J. Overview of eRisk: Early Risk Prediction on the Internet. In: *Lecture Notes in Computer Science vol 11018*. Cham: Springer, 2018. p. 343–361. Citado na página 38.

LOSADA, D. E.; CRESTANI, F.; PARAPAR, J. Overview of erisk 2019 early risk prediction on the internet. In: CRESTANI, F.; BRASCHLER, M.; SAVOY, J.; RAUBER, A.; MÜLLER, H.; LOSADA, D. E.; BÜRKI, G. H.; CAPPELLATO, L.; FERRO, N. (Ed.). *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Cham: Springer International Publishing, 2019. p. 340–357. ISBN 978-3-030-28577-7. Citado na página 18.

LOWE-CALVERLEY, E.; GRIEVE, R.; PADGETT, C. A risky investment? examining the outcomes of emotional investment in instagram. *Telematics and Informatics*, Elsevier Ltd, v. 45, 2019. ISSN 07365853. Citado 2 vezes nas páginas 54 e 55.

LUP, K.; TRUB, L.; ROSENTHAL, L. Instagram #instasad?: Exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. *Cyberpsychology, Behavior, and Social Networking*, Mary Ann Liebert Inc., v. 18, n. 5, p. 247–252, 2015. ISSN 21522715. Citado na página 54.

MAIMON, O. Z.; ROKACH, L. *Data mining with decision trees: theory and applications*. [S.l.]: World scientific, 2008. Citado na página 36.

MASUDA, N.; KURAHASHI, I.; ONARI, H. Suicide ideation of individuals in online social networks. *PloS one*, Public Library of Science, v. 8, n. 4, p. e62262, 2013. Citado na página 53.

MCMANUS, K.; MALLORY, E. K.; GOLDFEDER, R. L.; HAYNES, W. A.; TATUM, J. D. Mining twitter data to improve detection of schizophrenia. *AMIA Joint Summits on Translational Science proceedings*, v. 2015, p. 122–126, 2015. ISSN 2153-4063. Citado 2 vezes nas páginas 53 e 54.

MCNEMAR, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, v. 12, n. 2, p. 153—157, June 1947. ISSN 0033-3123. Disponível em: ⟨https://doi.org/10.1007/bf02295996⟩. Citado 4 vezes nas páginas 82, 93, 103 e 132.

MCPHERSON, M.; SMITH-LOVIN, L.; COOK, J. M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, v. 27, n. 1, p. 415–444, 2001. Disponível em: ⟨https://doi.org/10.1146/annurev.soc.27.1.415⟩. Citado na página 18.

MENDU, S.; BAGLIONE, A.; BAEE, S.; WU, C.; NG, B.; SHAKED, A.; CLORE, G.; BOUKHECHBA, M.; BARNES, L. A framework for understanding the relationship between social media discourse and mental health. *Proc. ACM Hum.-Comput. Interact.*, Association for Computing Machinery, New York, NY, USA, v. 4, n. CSCW2, oct 2020. Disponível em: ⟨https://doi.org/10.1145/3415215⟩. Citado na página 74.

MIKOLOV, T.; SUTSKEVER, I.; CHEN, K.; CORRADO, G. S.; DEAN, J. Distributed representations of words and phrases and their compositionality. In: BURGES, C. J. C.; BOTTOU, L.; WELLING, M.; GHAHRAMANI, Z.; WEINBERGER, K. Q. (Ed.). *Advances in Neural Information Processing Systems 26*. [S.l.]: Curran Associates, Inc., 2013. p. 3111–3119. Citado na página 24.

NOBLE, W. S. What is a support vector machine? *Nature biotechnology*, Nature Publishing Group UK London, v. 24, n. 12, p. 1565–1567, 2006. Citado 2 vezes nas páginas 28 e 29.

PARK, S.; KIM, I.; LEE, S. W.; YOO, J.; JEONG, B.; CHA, M. Manifestation of depression and loneliness on social networks: A case study of young adults on facebook. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY, USA: Association for Computing Machinery, 2015. (CSCW '15), p. 557–570. ISBN 9781450329224. Disponível em: ⟨https://doi.org/10.1145/2675133.2675139⟩. Citado 3 vezes nas páginas 53, 54 e 55.

PARK, S.; LEE, S. W.; KWAK, J.; CHA, M.; JEONG, B. Activities on facebook reveal the depressive state of users. *J Med Internet Res*, v. 15, n. 10, p. e217, Oct 2013. ISSN 14388871. Disponível em: ⟨https://doi.org/10.2196/jmir.2718⟩. Citado na página 52.

PEARL, R.; REED, L. J. On the rate of growth of the population of the united states since 1790 and its mathematical representation. *Proceedings of the national academy of sciences*, National Acad Sciences, v. 6, n. 6, p. 275–288, 1920. Citado na página 26.

PENNEBAKER, J. W.; FRANCIS, M. E.; BOOTH, R. J. *Inquiry and Word Count: LIWC*. Mahwah, NJ: Lawrence Erlbaum, 2001. Citado na página 22.

POLIKAR, R. Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, IEEE, v. 6, n. 3, p. 21–45, 2006. Citado na página 34.

PREOŢIUC-PIETRO, D.; SAP, M.; SCHWARTZ, H. A.; UNGAR, L. Mental illness detection at the world well-being project for the CLPsych 2015 shared task. In: *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Denver, Colorado: Association for Computational Linguistics, 2015. p. 40–45. Disponível em: ⟨https://www.aclweb.org/anthology/W15-1205⟩. Citado 3 vezes nas páginas 53, 54 e 55.

RICARD, B.; MARSCH, L.; CROSIER, B.; HASSANPOUR, S. Exploring the utility of community-generated social media content for detecting depression: An analytical study on instagram. *Journal of Medical Internet Research*, Journal of Medical Internet Research, v. 20, n. 12, 2018. ISSN 14388871. Citado 5 vezes nas páginas 18, 19, 54, 55 e 79.

RUCH, A. Can x2vec save lives? integrating graph and language embeddings for automatic mental health classification. *Journal of Physics: Complexity*, IOP Publishing, v. 1, n. 3, p. 035005, 2020. Citado na página 74.

SALTON, G.; WONG, A.; YANG, C. S. A vector space model for automatic indexing. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 18, n. 11, p. 613–620, nov. 1975. ISSN 0001-0782. Disponível em: ⟨https://doi.org/10.1145/361219.361220⟩. Citado na página 21.

SANTOS, W.; FUNABASHI, A.; PARABONI, I. Searching Brazilian Twitter for signs of mental health issues. In: *Proceedings of The 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, 2020. p. 6111–6117. ISBN 979-10-95546-34-4. Disponível em: ⟨https://www.aclweb.org/anthology/2020.lrec-1.750⟩. Citado 3 vezes nas páginas 37, 44 e 81.

SANTOS, W. R. dos; OLIVEIRA, R. L. de; PARABONI, I. SetembroBR: a social media corpus for depression and anxiety disorder prediction. *Language Resources and Evaluation*, 2023. Citado 5 vezes nas páginas 37, 38, 39, 41 e 81.

SARAVIA, E.; CHANG, C.; LORENZO, R. J. D.; CHEN, Y. Midas: Mental illness detection and analysis via social media. In: *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. [S.l.: s.n.], 2016. p. 1418–1421. Citado 2 vezes nas páginas 53 e 54.

SAWHNEY, R.; JOSHI, H.; SHAH, R. R.; FLEK, L. Suicide ideation detection via social and temporal user representations using hyperbolic learning. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. [S.l.]: Association for Computational Linguistics, 2021. p. 2176–2190. Citado na página 98.

SCHAPIRE, R. E. The strength of weak learnability. *Machine learning*, Springer, v. 5, p. 197–227, 1990. Citado na página 34.

SEMENOV, A.; NATEKIN, A.; NIKOLENKO, S.; UPRAVITELEV, P.; TROFIMOV, M.; KHARCHENKO, M. Discerning depression propensity among participants of suicide and depression-related groups of vk.com. *Communications in Computer and Information Science*, Springer Verlag, v. 542, p. 24–35, 2015. ISSN 18650929. Conference of 4th International Conference on Analysis of Images, Social Networks and Texts, AIST 2015 ; Conference Date: 9 April 2015 Through 11 April 2015; Conference Code:158559. Citado 3 vezes nas páginas 54, 55 e 56.

SHEN, G.; JIA, J.; NIE, L.; FENG, F.; ZHANG, C.; HU, T.; CHUA, T.-S.; ZHU, W. Depression detection via harvesting social media: A multimodal dictionary learning solution. In: *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. [S.l.]: AAAI Press, 2017. (IJCAI'17), p. 3838–3844. ISBN 9780999241103. Citado 3 vezes nas páginas 53, 54 e 79.

SHRESTHA, A.; SERRA, E.; SPEZZANO, F. Multi-modal social and psycho-linguistic embedding via recurrent neural networks to identify depressed users in online forums. *Network Modeling Analysis in Health Informatics and Bioinformatics*, Springer, v. 9, n. 1, p. 1–11, 2020. Citado 4 vezes nas páginas 18, 19, 41 e 79.

SHRESTHA, A.; SPEZZANO, F. Detecting depressed users in online forums. In: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. New York, NY, USA: Association for Computing Machinery, 2019. (ASONAM '19), p. 945–951. ISBN 9781450368681. Disponível em: ⟨https://doi.org/10.1145/3341161.3343511⟩. Citado 5 vezes nas páginas 18, 55, 56, 58 e 79.

SINHA, P. P.; MISHRA, R.; SAWHNEY, R.; MAHATA, D.; SHAH, R. R.; LIU, H. #suicidal - a multipronged approach to identify and explore suicidal ideation in twitter. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2019. (CIKM '19), p. 941–950. ISBN 9781450369763. Disponível em: ⟨https://doi.org/10.1145/3357384.3358060⟩. Citado 8 vezes nas páginas 18, 19, 50, 53, 54, 55, 70 e 79.

SOLDAINI, L.; WALSH, T.; COHAN, A.; HAN, J.; GOHARIAN, N. Helping or hurting? predicting changes in users' risk of self-harm through online community interactions. In: *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. New Orleans, LA: Association for Computational Linguistics, 2018. p. 194–203. Disponível em: ⟨https://www.aclweb.org/anthology/W18-0621⟩. Citado 2 vezes nas páginas 54 e 55.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. Bertimbau: Pretrained bert models for brazilian portuguese. In: CERRI, R.; PRATI, R. C. (Ed.). *Intelligent Systems*. Cham: Springer International Publishing, 2020. p. 403–417. Citado na página 23.

THELWALL, M.; BUCKLEY, K.; PALTOGLOU, G. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, Wiley Online Library, v. 63, n. 1, p. 163–173, 2012. Citado na página 60.

TSUGAWA, S.; KIKUCHI, Y.; KISHINO, F.; NAKAJIMA, K.; ITOH, Y.; OHSAKI, H. Recognizing depression from twitter activity. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2015. (CHI '15), p. 3187–3196. ISBN 9781450331456. Disponível em: ⟨https://doi.org/10.1145/2702123.2702280⟩. Citado 5 vezes nas páginas 41, 44, 53, 54 e 72.

VAPNIK, V. N. *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag, 1995. ISBN 0387945598. Citado na página 27.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. Attention is all you need. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2017. p. 5998–6008. Citado na página 23.

VEDULA, N.; PARTHASARATHY, S. Emotional and linguistic cues of depression from social media. In: *Proceedings of the 2017 International Conference on Digital Health*. New York, NY, USA: Association for Computing Machinery, 2017. (DH '17), p. 127–136. ISBN

9781450352499. Disponível em: ⟨https://doi.org/10.1145/3079452.3079465⟩. Citado 7 vezes nas páginas 19, 44, 53, 54, 55, 59 e 79.

WANG, T.; BREDE, M.; IANNI, A.; MENTZAKIS, E. Detecting and characterizing eating-disorder communities on social media. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2017. (WSDM '17), p. 91–100. ISBN 9781450346757. Disponível em: ⟨https://doi.org/10.1145/3018661.3018706⟩. Citado 8 vezes nas páginas 18, 19, 53, 54, 55, 56, 67 e 79.

WANG, X.; ZHANG, C.; SUN, L. An improved model for depression detection in micro-blog social network. In: *2013 IEEE 13th International Conference on Data Mining Workshops*. [S.l.: s.n.], 2013. p. 80–87. Citado 2 vezes nas páginas 18 e 53.

WANG, Y.; WANG, Z.; LI, C.; ZHANG, Y.; WANG, H. A multitask deep learning approach for user depression detection on sina weibo. *arXiv preprint arXiv:2008.11708*, 2020. Citado na página 79.

WOLPERT, D. H. Stacked generalization. *Neural networks*, Elsevier, v. 5, n. 2, p. 241–259, 1992. Citado na página 36.

WONGKOBLAP, A.; VADILLO, M.; CURCIN, V. Researching mental health disorders in the era of social media: Systematic review. *Journal of Medical Internet Research*, Journal of Medical Internet Research, v. 19, n. 6, 2017. ISSN 14388871. Citado na página 43.

WONGKOBLAP, A.; VADILLO, M. A.; CURCIN, V. A multilevel predictive model for detecting social network users with depression. In: *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. [S.l.: s.n.], 2018. p. 130–135. Citado 3 vezes nas páginas 53, 54 e 55.

WORLD HEALTH ORGANIZATION. *Depression and other common mental disorders: global health estimates*. [S.l.], 2017. Citado 2 vezes nas páginas 17 e 48.

WU, M. Y.; SHEN, C.-Y.; WANG, E. T.; CHEN, A. L. A deep architecture for depression detection using posting, behavior, and living environment data. *Journal of Intelligent Information Systems*, Springer, v. 54, n. 2, p. 225–244, 2020. Citado 4 vezes nas páginas 19, 53, 54 e 79.

XU, R.; ZHANG, Q. Social dynamics of the online health communities for mental health. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, v. 9545, p. 267–277, 2016. ISSN 03029743. Conference of International Conference for Smart Health, ICSH 2015 ; Conference Date: 17 November 2015 Through 18 November 2015; Conference Code:163109. Citado 3 vezes nas páginas 53, 54 e 55.

XU, R.; ZHANG, Q. Understanding online health groups for depression: Social network and linguistic perspectives. *Journal of Medical Internet Research*, JMIR Publications Inc., v. 18, n. 3, 2016. ISSN 14388871. Disponível em: ⟨https://www.scopus.com/inward/record.uri?eid=2-s2.0-84962109790&doi=10.2196%2fjmir.5042&partnerID=40&md5=43bf3ac83a3817b0d933c76f091246f3⟩. Citado na página 55.

YATES, A.; COHAN, A.; GOHARIAN, N. Depression and self-harm risk assessment in online forums. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, 2017. p. 2968–2978. Disponível em: ⟨https://www.aclweb.org/anthology/D17-1322⟩. Citado 2 vezes nas páginas 18 e 58.

YU, Y.; SI, X.; HU, C.; ZHANG, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, v. 31, n. 7, p. 1235–1270, 07 2019. ISSN 0899-7667. Disponível em: ⟨https://doi.org/10.1162/neco_a_01199⟩. Citado 2 vezes nas páginas 33 e 34.

YUN-TAO, Z.; LING, G.; YONG-CHENG, W. An improved tf-idf approach for text classification. *Journal of Zhejiang University-Science A*, Springer, v. 6, n. 1, p. 49–55, 2005. Citado na página 22.

ZHAO, L.; JIA, J.; FENG, L. Teenagers' stress detection based on time-sensitive micro-blog comment/response actions. In: DILLON, T. (Ed.). *Artificial Intelligence in Theory and Practice IV*. Cham: Springer International Publishing, 2015. p. 26–36. ISBN 978-3-319-25261-2. Citado na página 54.

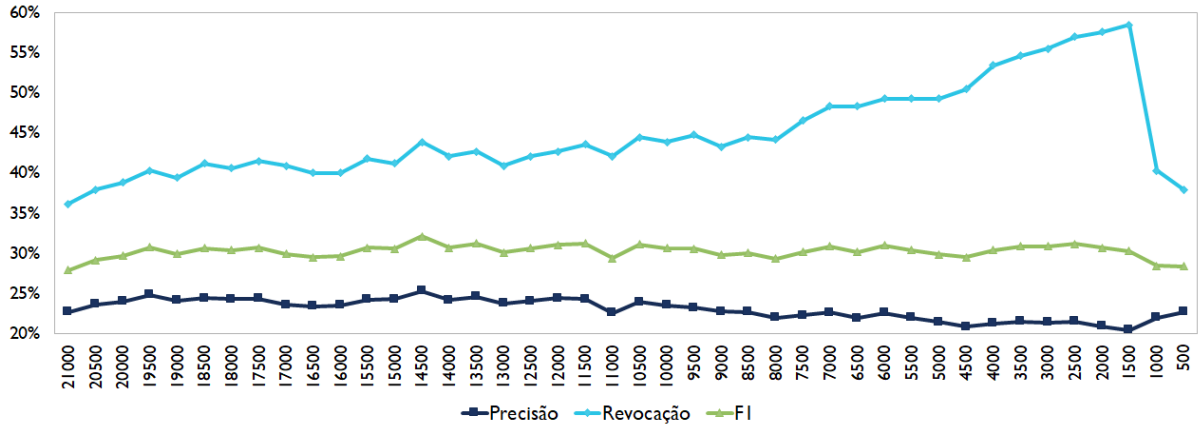
ZOGAN, H.; RAZZAK, I.; JAMEEL, S.; XU, G. Depressionnet: A novel summarization boosted deep framework for depression detection on social media. *arXiv preprint arXiv:2105.10878*, 2021. Citado 2 vezes nas páginas 19 e 79.

ZOGAN, H.; RAZZAK, I.; WANG, X.; JAMEEL, S.; XU, G. Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. *World Wide Web*, Springer, v. 25, n. 1, p. 281–304, 2022. Citado na página 75.

Apêndice A – Definição das K características mais relevantes no método de seleção univariada

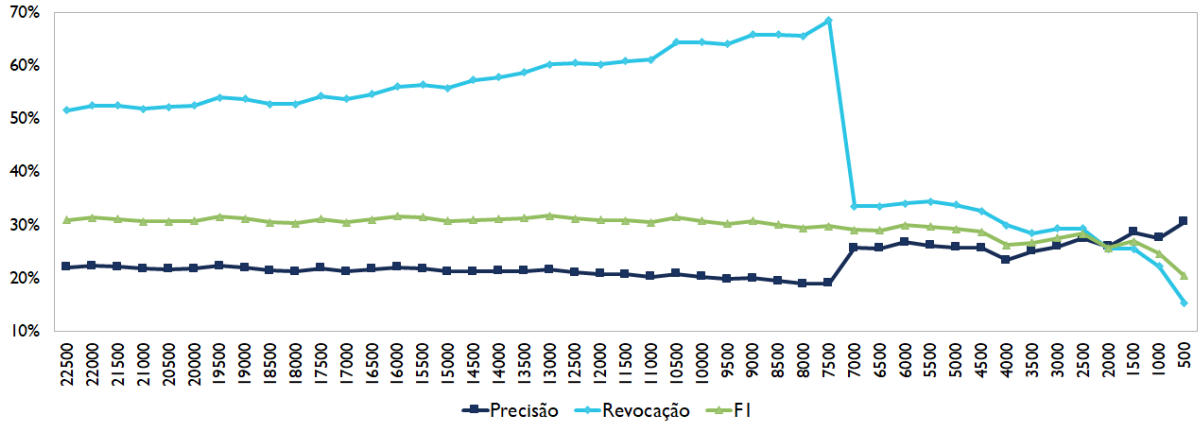
Conforme discutido na seção 5.1.1, para a seleção das K características mais relevantes pelo método de seleção univariada no experimento com uso de classificadores individuais, foram realizados testes com valores a partir do número máximo de conexões disponíveis em cada uma das redes (amigos, seguidores e menções), onde, em cada execução de teste, o número de características era decrescido em 500 unidades. As métricas de avaliação dos resultados gerados por um classificador baseado em Regressão Logística para cada execução de teste estão ilustradas nas figuras 18, 19 e 20 para o conjunto de depressão e nas figuras 21, 22 e 23 para o conjunto de ansiedade.

Figura 18 – Seleção univariada para o modelo *amigos.univ.reglog* na tarefa de predição de depressão



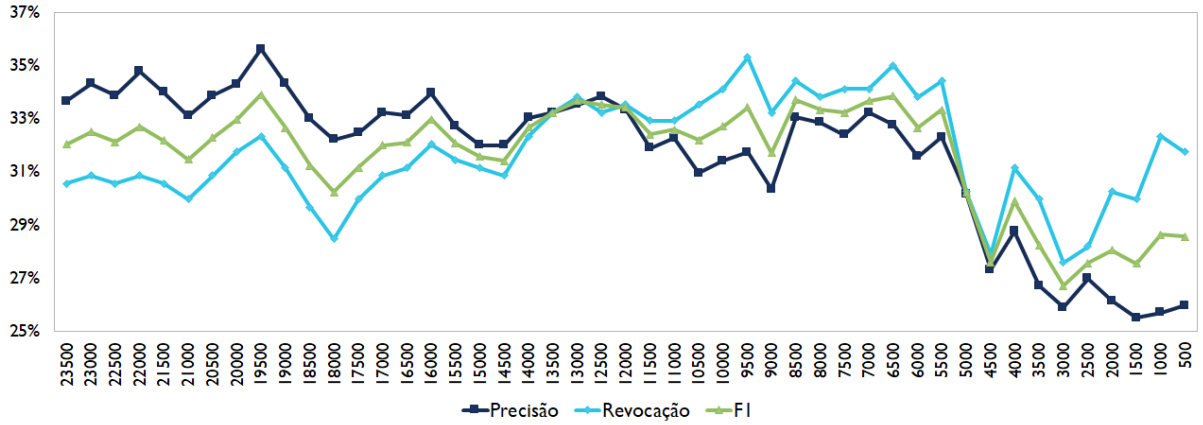
Fonte – Rafael Lage de Oliveira, 2023

Figura 19 – Seleção univariada para o modelo *seguidores.univ.reglog* na tarefa de predição de depressão



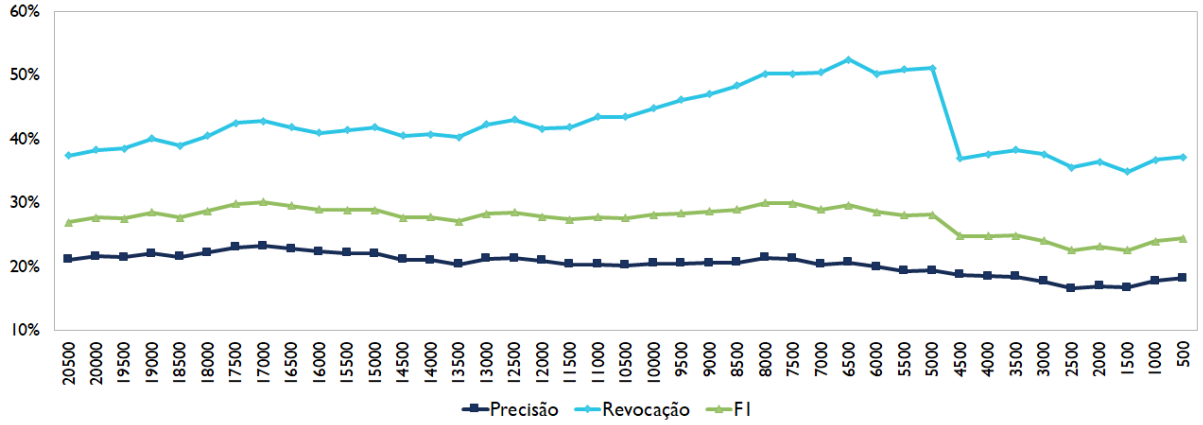
Fonte – Rafael Lage de Oliveira, 2023

Figura 20 – Seleção univariada para o modelo *menções.univ.reglog* na tarefa de predição de depressão



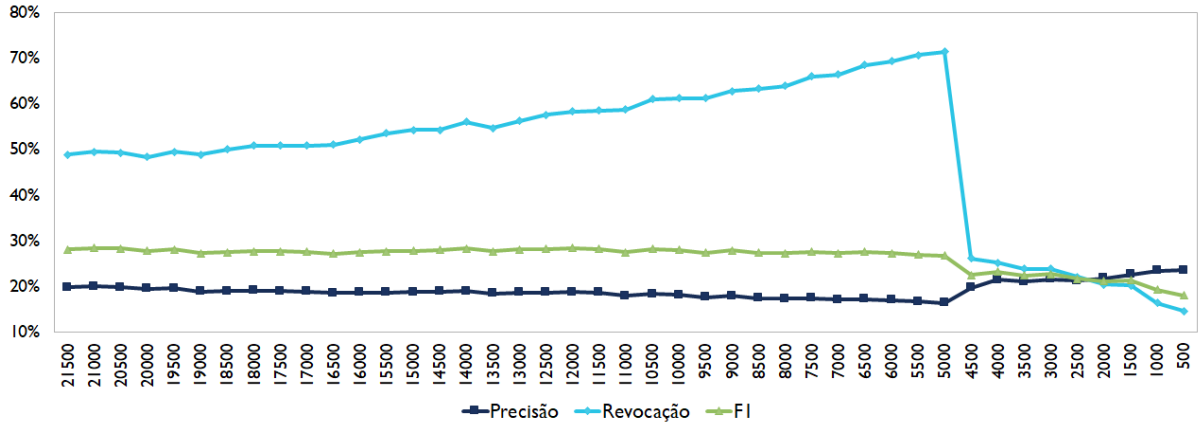
Fonte – Rafael Lage de Oliveira, 2023

Figura 21 – Seleção univariada para o modelo *amigos.univ.reglog* na tarefa de predição de ansiedade



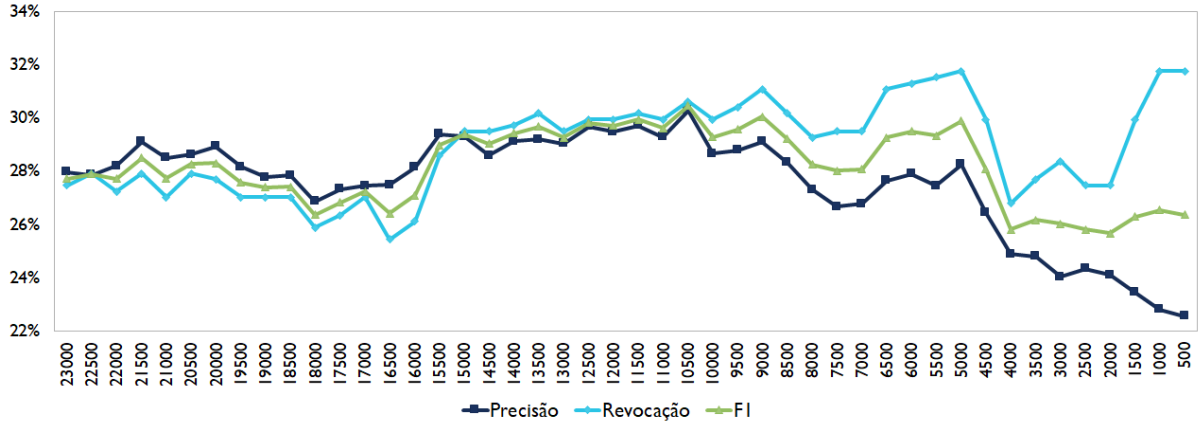
Fonte – Rafael Lage de Oliveira, 2023

Figura 22 – Seleção univariada para o modelo *seguidores.univ.reglog* na tarefa de predição de ansiedade



Fonte – Rafael Lage de Oliveira, 2023

Figura 23 – Seleção univariada para o modelo *menções.univ.reglog* na tarefa de predição de ansiedade

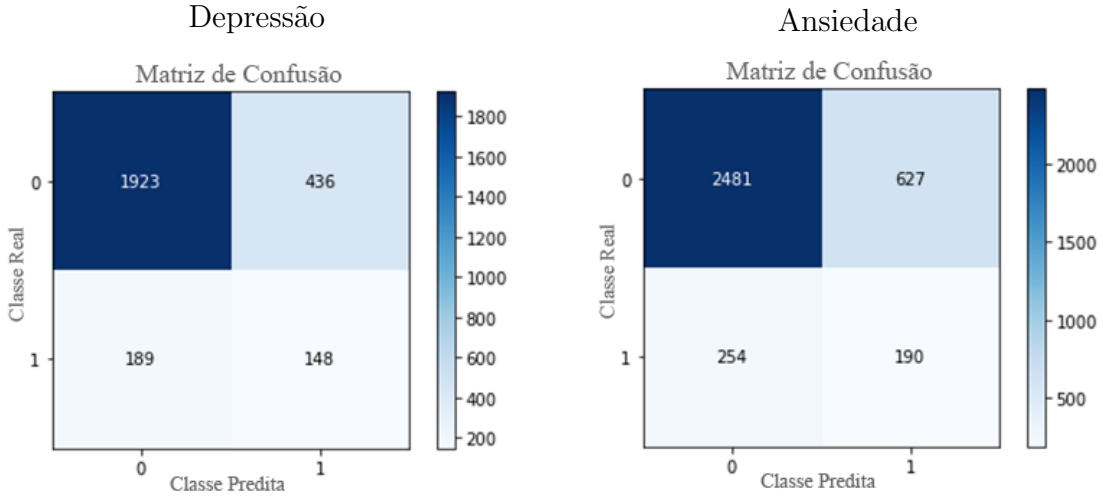


Fonte – Rafael Lage de Oliveira, 2023

Apêndice B – Matrizes de confusão dos classificadores individuais

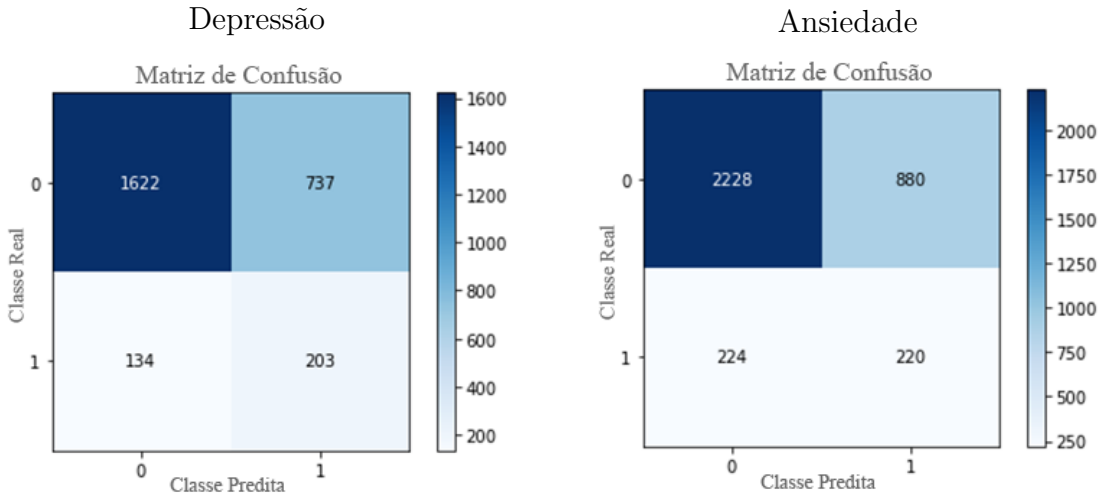
As figuras 24, 25 e 26 mostram as matrizes de confusão referente aos modelos baseados em seleção univariada para as tarefas de predição de depressão e ansiedade. As figuras 27, 28, 29 e 30 mostram as matrizes de confusão referentes aos modelos baseados em *node2vec* para as mesmas tarefas de predição. Por fim, as figuras 31 e 32 mostram as matrizes geradas para os modelos baseados em atividade de postagens no tempo e características linguísticas, respectivamente.

Figura 24 – Matrizes de confusão do modelo *amigos.univ.reglog*



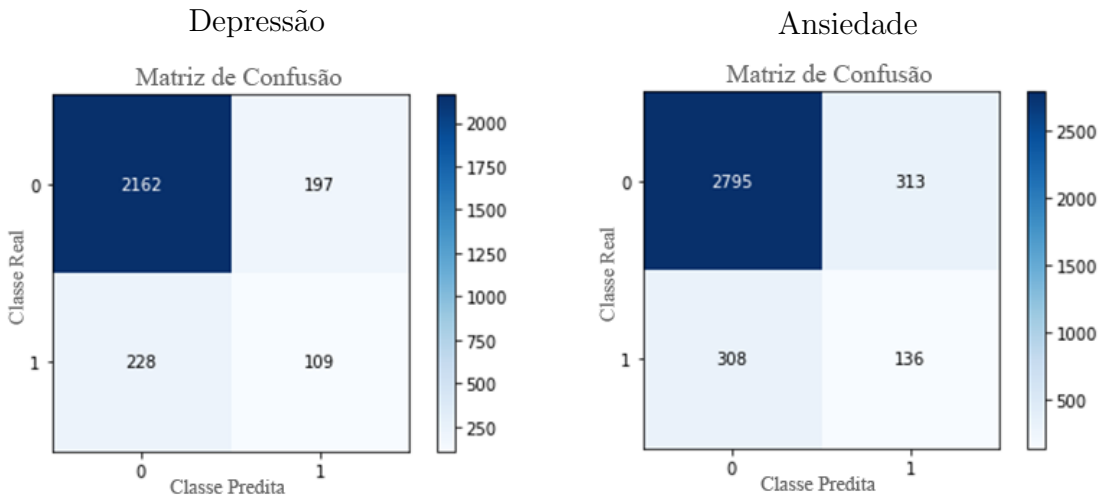
Fonte – Rafael Lage de Oliveira, 2023

Figura 25 – Matrizes de confusão do modelo *seguidores.univ.reglog*



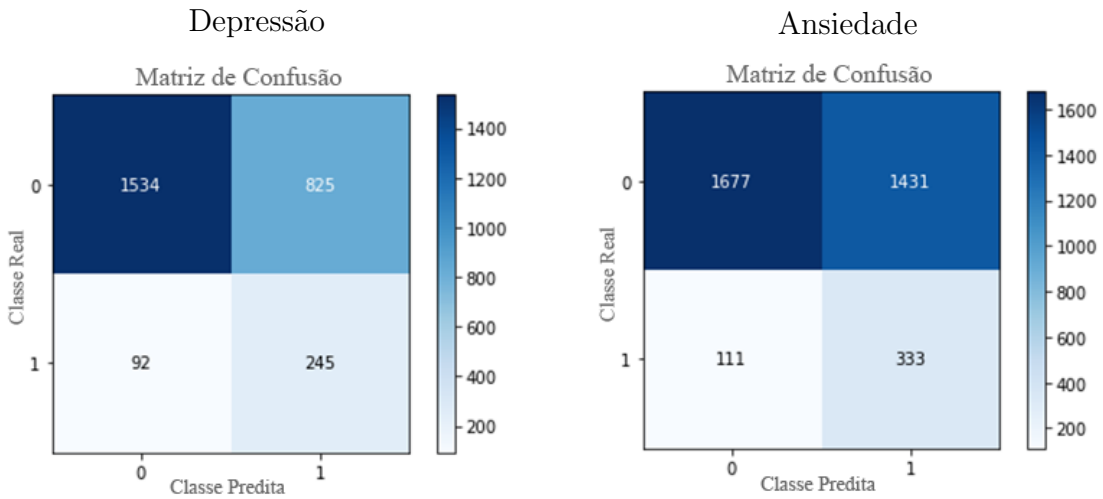
Fonte – Rafael Lage de Oliveira, 2023

Figura 26 – Matrizes de confusão do modelo *menções.univ.reglog*



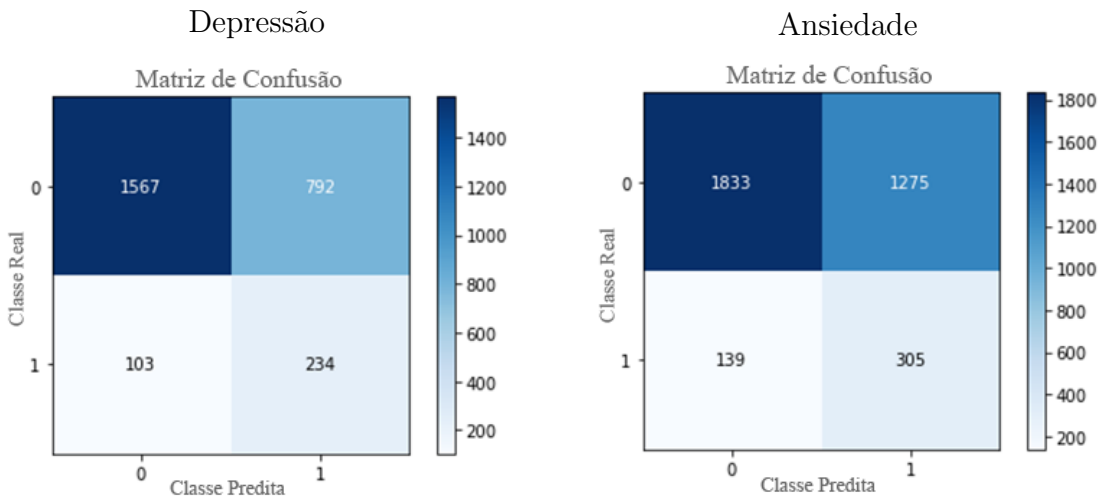
Fonte – Rafael Lage de Oliveira, 2023

Figura 27 – Matrizes de confusão do modelo *amigos.n2v.svm*



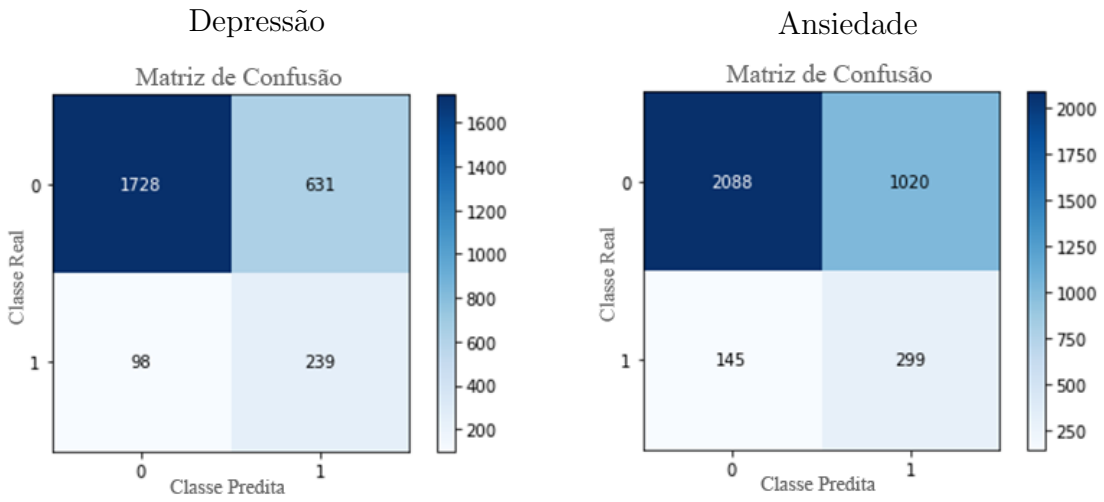
Fonte – Rafael Lage de Oliveira, 2023

Figura 28 – Matrizes de confusão do modelo *seguidores.n2v.svm*



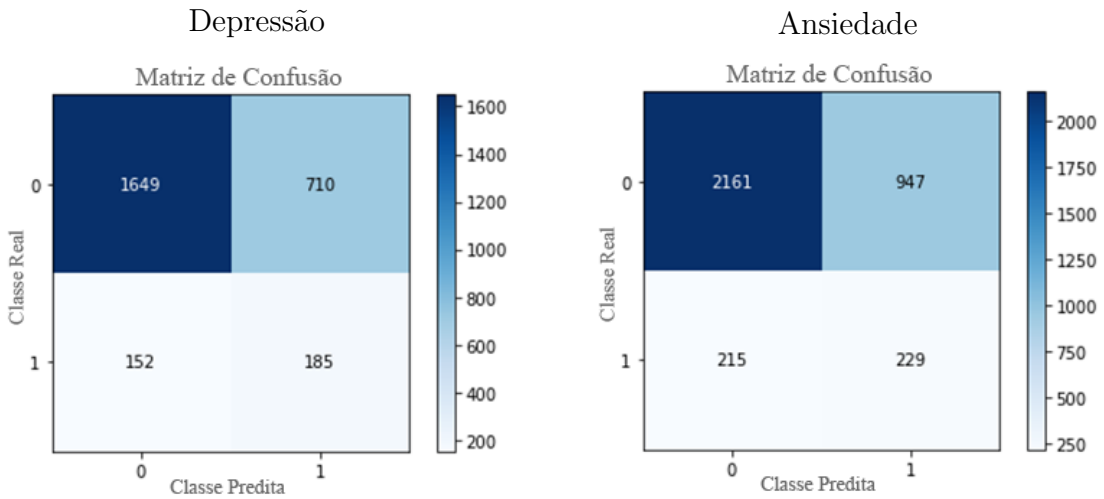
Fonte – Rafael Lage de Oliveira, 2023

Figura 29 – Matrizes de confusão do modelo *menções.n2v.svm*



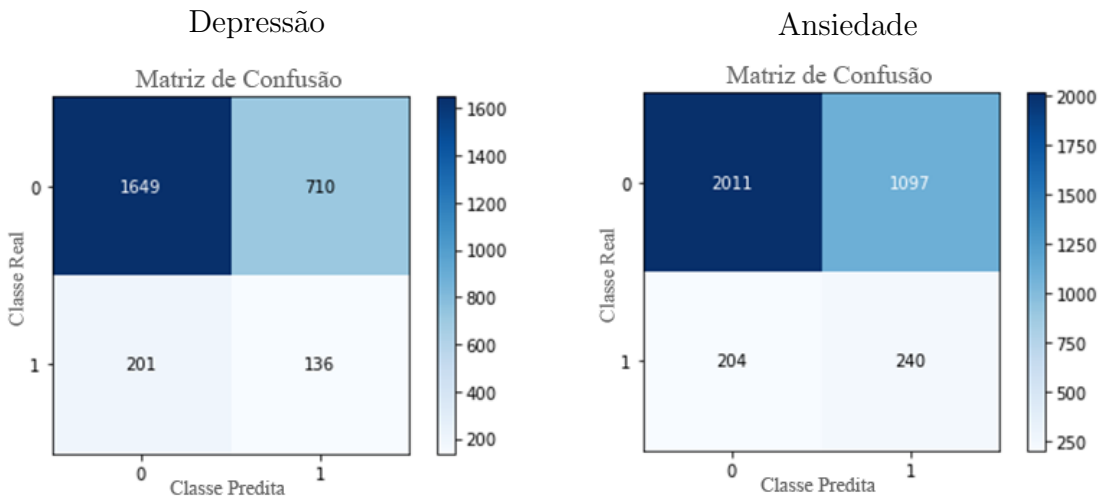
Fonte – Rafael Lage de Oliveira, 2023

Figura 30 – Matrizes de confusão do modelo *topmenções.n2v.svm*



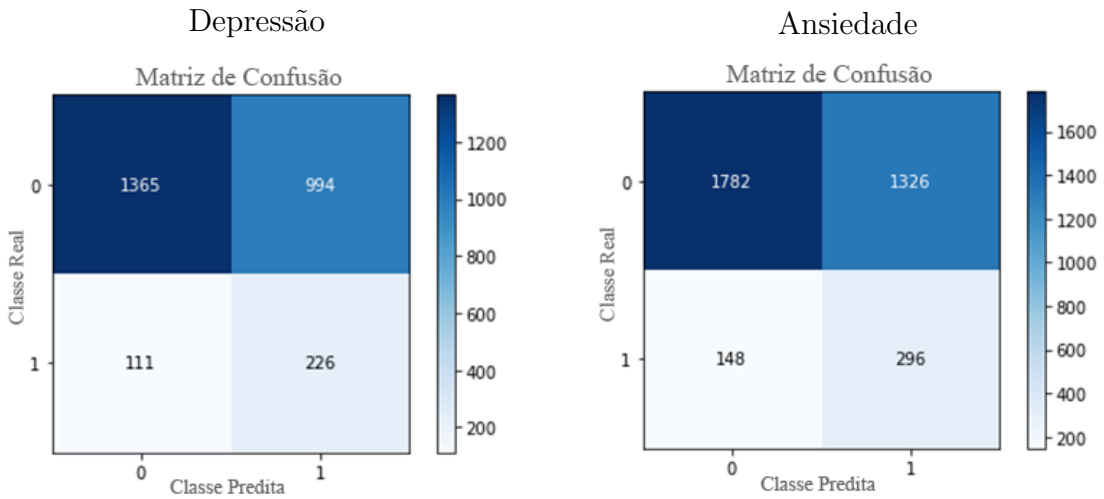
Fonte – Rafael Lage de Oliveira, 2023

Figura 31 – Matrizes de confusão do modelo *horario.svm*



Fonte – Rafael Lage de Oliveira, 2023

Figura 32 – Matrizes de confusão do modelo *liwc.reglog*

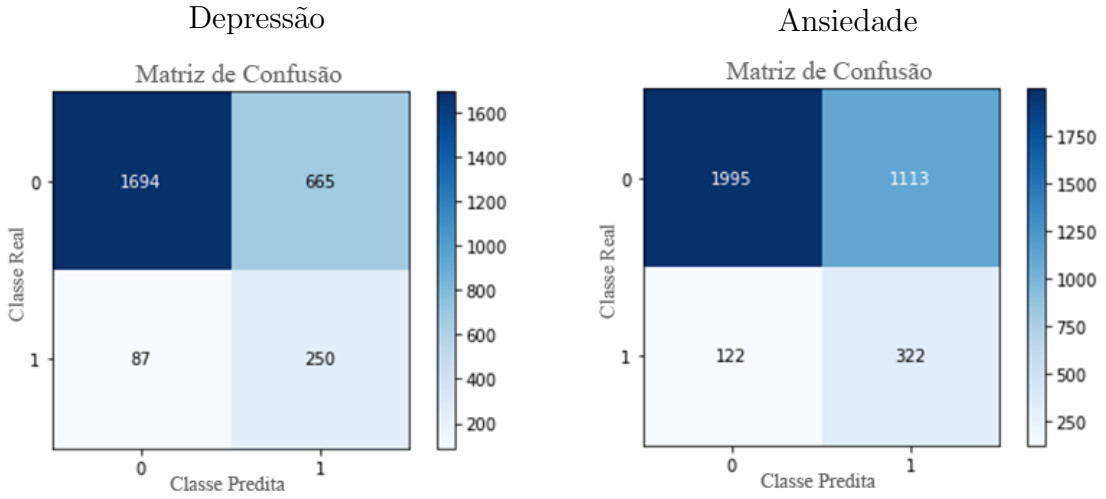


Fonte – Rafael Lage de Oliveira, 2023

Apêndice C – Matrizes de confusão dos classificadores combinados

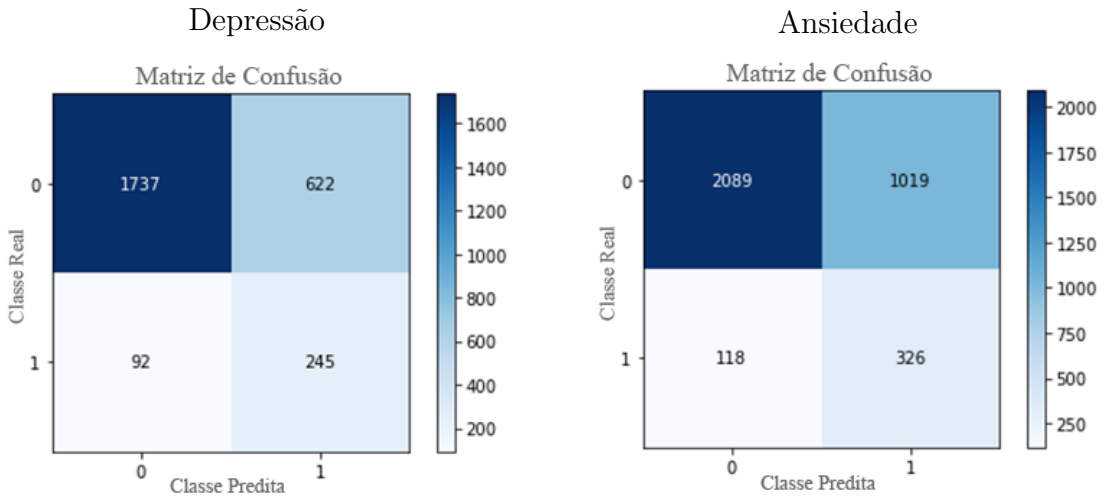
As figuras 33, 34, 35 e 36 mostram as matrizes de confusão referentes aos modelos baseados em concatenação de características para as tarefas de predição de depressão e ansiedade. As figuras 37, 38, 39 e 40 mostram as matrizes de confusão referentes aos modelos baseados em votação. Por fim, as figuras 41, 42, 43 e 44 mostram as matrizes geradas para os modelos baseados em *stacking*.

Figura 33 – Matrizes de confusão do modelo *concat.ASM*



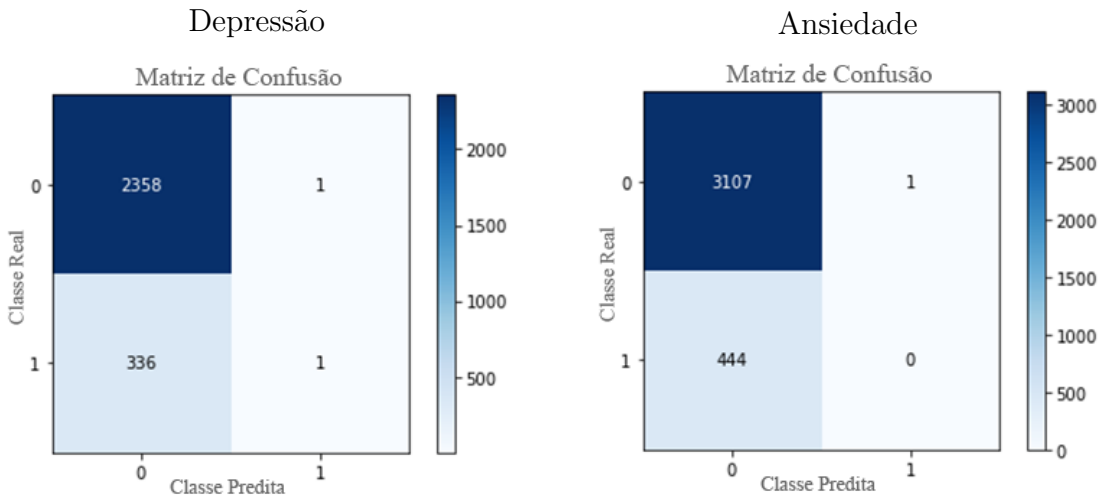
Fonte – Rafael Lage de Oliveira, 2023

Figura 34 – Matrizes de confusão do modelo *concat.ASMT*



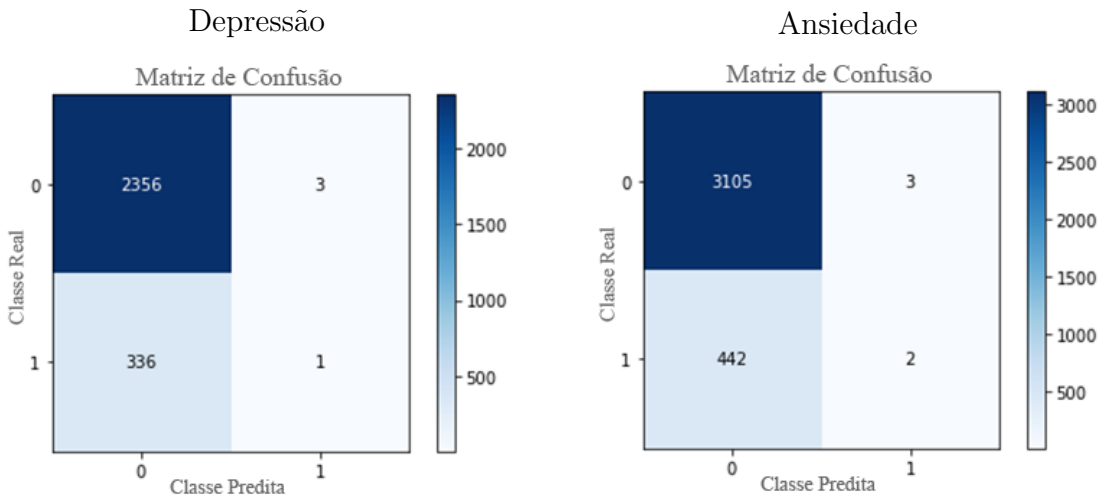
Fonte – Rafael Lage de Oliveira, 2023

Figura 35 – Matrizes de confusão do modelo *concat.ASMH*



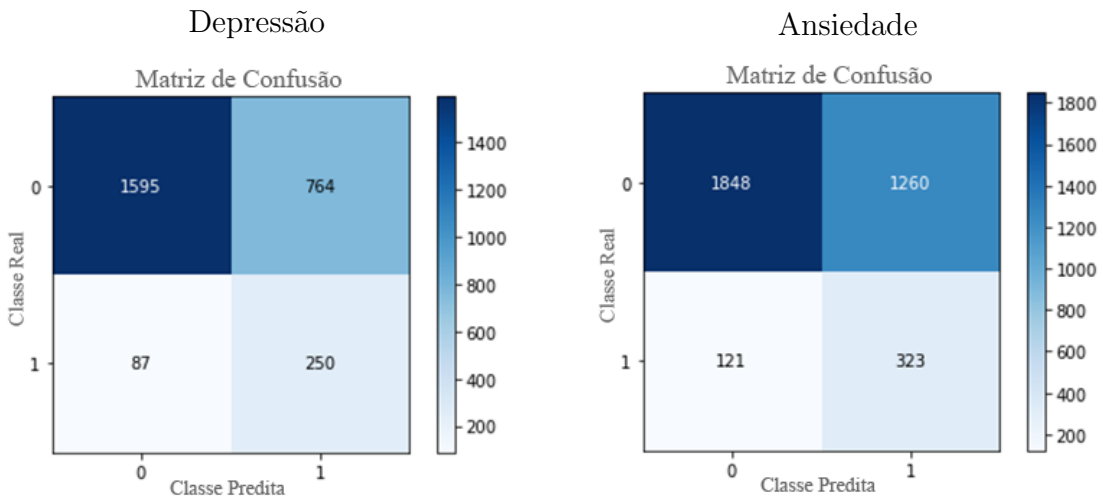
Fonte – Rafael Lage de Oliveira, 2023

Figura 36 – Matrizes de confusão do modelo *concat.ASMTH*



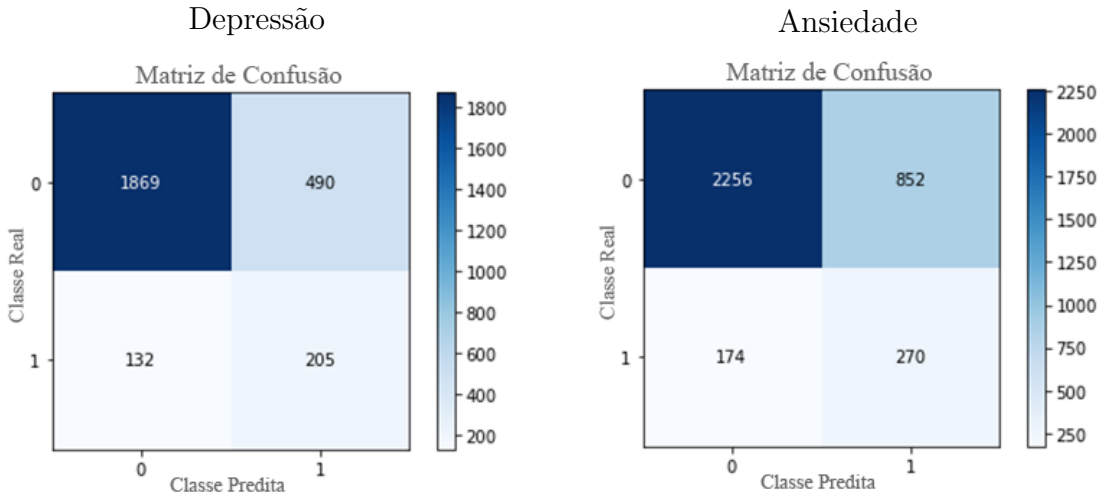
Fonte – Rafael Lage de Oliveira, 2023

Figura 37 – Matrizes de confusão do modelo *vot.ASM*



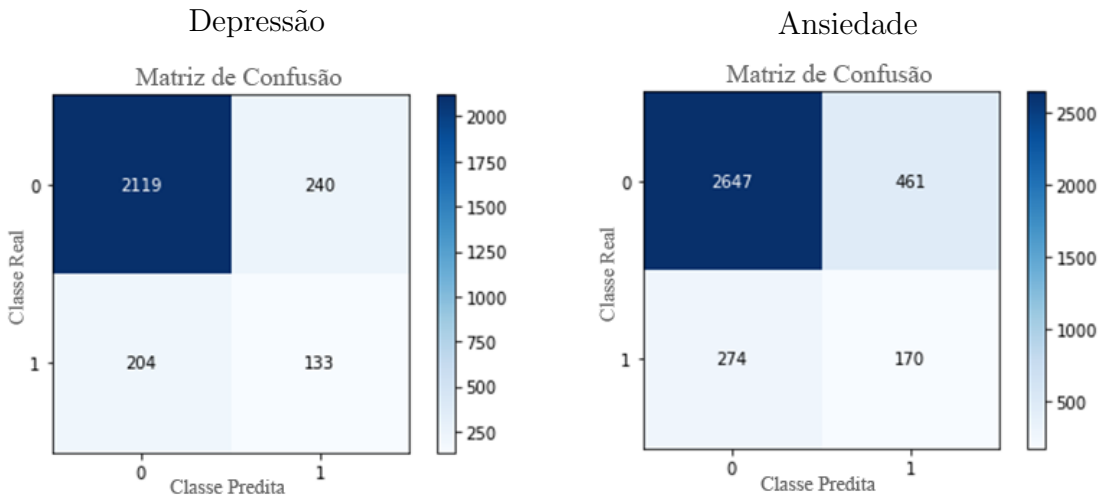
Fonte – Rafael Lage de Oliveira, 2023

Figura 38 – Matrizes de confusão do modelo *vot.ASMT*



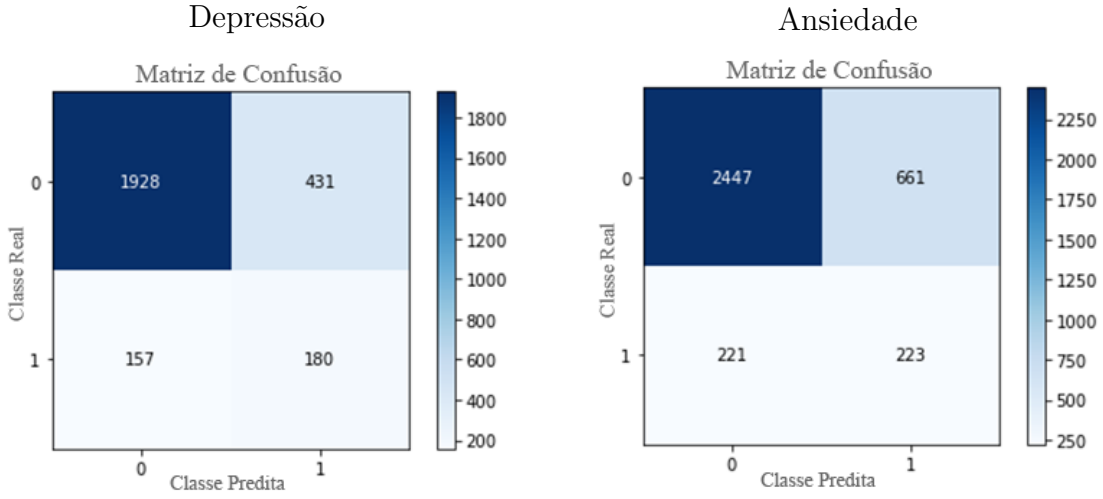
Fonte – Rafael Lage de Oliveira, 2023

Figura 39 – Matrizes de confusão do modelo *vot.ASMH*



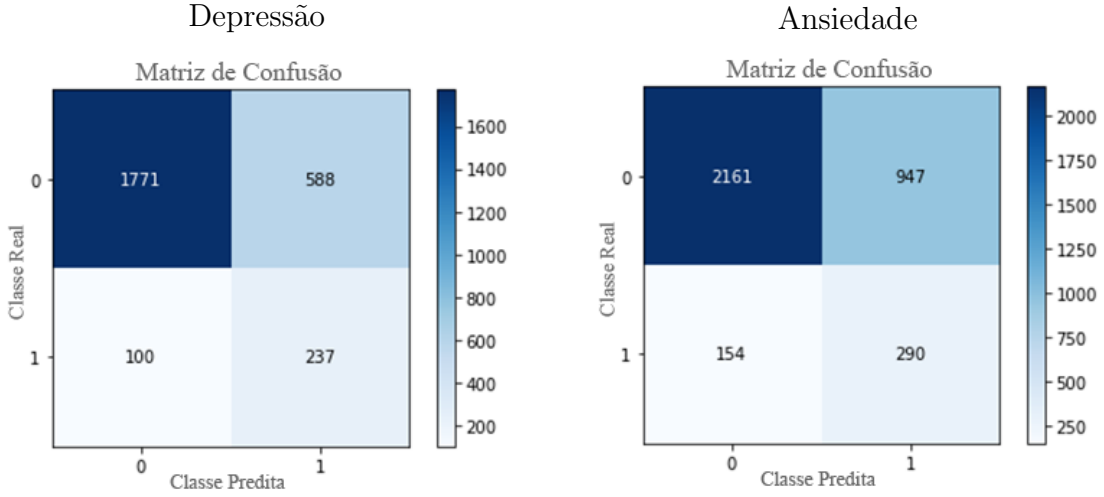
Fonte – Rafael Lage de Oliveira, 2023

Figura 40 – Matrizes de confusão do modelo *vot.ASMTH*



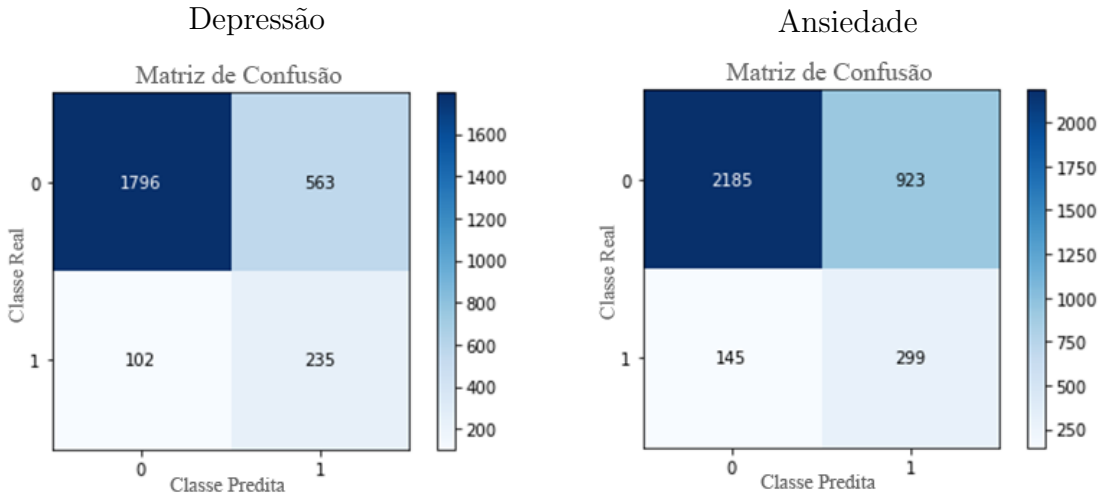
Fonte – Rafael Lage de Oliveira, 2023

Figura 41 – Matrizes de confusão do modelo *stack.ASM*



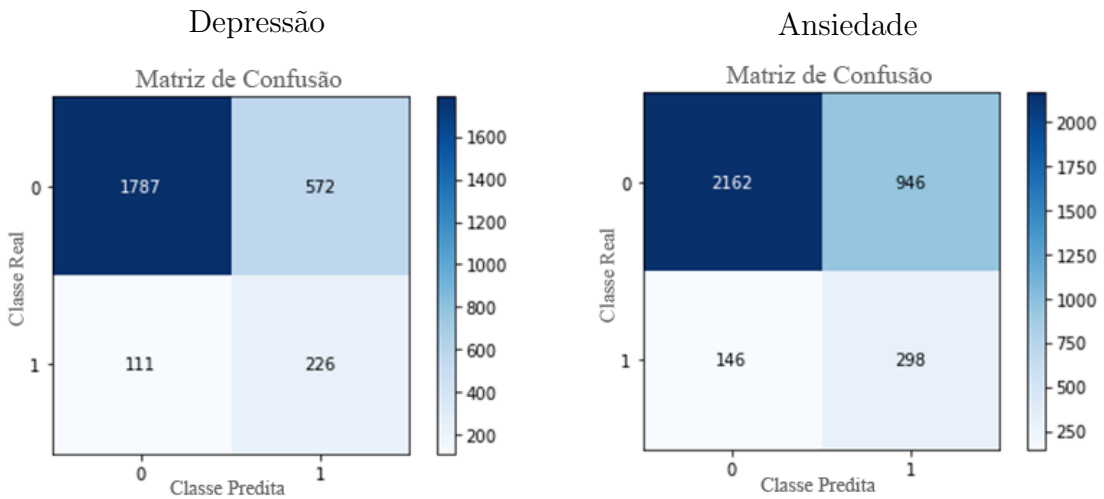
Fonte – Rafael Lage de Oliveira, 2023

Figura 42 – Matrizes de confusão do modelo *stack.ASMT*



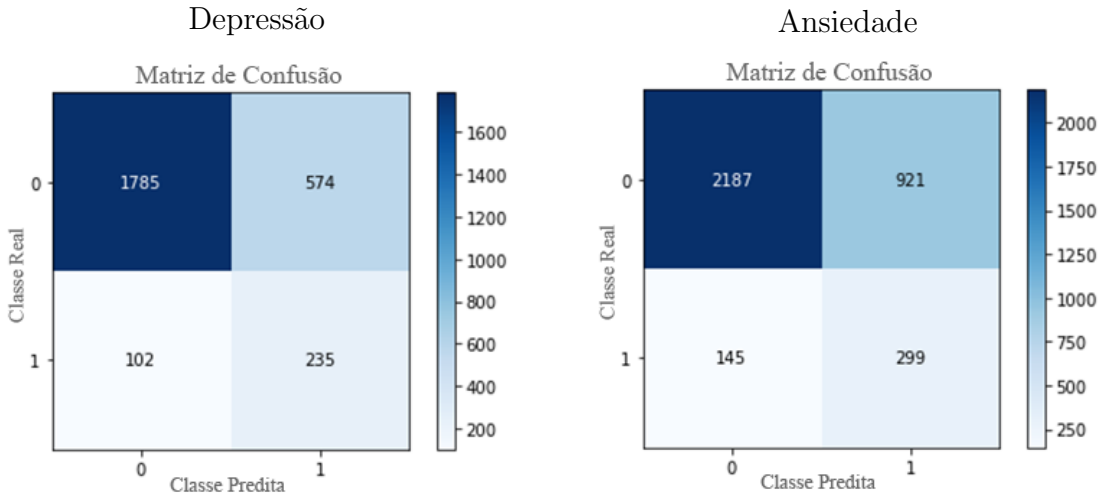
Fonte – Rafael Lage de Oliveira, 2023

Figura 43 – Matrizes de confusão do modelo *stack.ASMH*



Fonte – Rafael Lage de Oliveira, 2023

Figura 44 – Matrizes de confusão do modelo *stack.ASMTH*



Fonte – Rafael Lage de Oliveira, 2023

Apêndice D – Testes de significância estatística

As tabelas 16 e 17 apresentam os resultados dos testes de significância estatística de McNemar (MCNEMAR, 1947) entre pares de modelos para as tarefas de predição de depressão e ansiedade, respectivamente. Para cada comparação entre pares de modelos, foram registrados os resultados para a classe diagnosticado e global (que considera as classes diagnosticado e controle). Em todos os testes, o Modelo 1 obteve resultados superiores ao Modelo 2 para a medida F1.

Tabela 16 – Testes de significância estatística para a tarefa de predição de depressão

Comparação	**Modelo 1**	**Modelo 2**	**Teste McNemar**	
			Diagnosticado	**Global**
Individual x Individual	menções.n2v.svm	amigos.n2v.svm	Sem significância estatística	χ =45,891; p<0,001
	menções.n2v.svm	seguidores.n2v.svm	Sem significância estatística	χ =34,815; p<0,001
	menções.n2v.svm	topmenções.n2v.svm	χ =22,294; p<0,001	χ =17,512; p<0,001
	menções.n2v.svm	horario.svm	χ =58,123; p<0,001	χ =29,621; p<0,001
Combinação x Combinação	stack.ASMT	concat.ASMT	Sem significância estatística	χ =7,361; p<0,01
	stack.ASMT	vot.ASMT	χ =8,000; p<0,001	χ =6,759; p<0,01
	concat.ASMT	vot.ASMT	χ =8,000; p<0,001	χ =21,565; p<0,001
Combinação x Individual	stack.ASMT	menções.n2v.svm	Sem significância estatística	χ =12,403; p<0,001
	stack.ASMT	amigos.n2v.svm	Sem significância estatística	χ =120,692; p<0,001
	stack.ASMT	seguidores.n2v.svm	Sem significância estatística	χ =96,755; p<0,001
	stack.ASMT	topmenções.n2v.svm	χ =20,008; p<0,001	χ =46,792; p<0,001
	stack.ASMT	horario.svm	χ =51,358; p<0,001	χ =56,627; p<0,001
node2vec x Seleção univariada	amigos.n2v.svm	amigos.univ.reglog	χ =14,000; p<0,001	χ =103,776; p<0,001
	seguidores.n2v.svm	seguidores.univ.reglog	χ =8,257; p<0,01	Sem significância estatística
	menções..n2v.svm	menções.univ.reglog	χ =9,000; p<0,001	χ =122,412; p<0,001
Individual x Baseline	menções.n2v.svm	liwc.reglog	Sem significância estatística	χ =129,014; p<0,001
	amigos.n2v.svm	liwc.reglog	Sem significância estatística	χ =36,275; p<0,001
	seguidores.n2v.svm	liwc.reglog	Sem significância estatística	χ =43,507; p<0,001
	topmenções.n2v.svm	liwc.reglog	χ =9,467; p<0,01	χ =46,443; p<0,001
	liwc.reglog	horario.svm	χ =42,133; p<0,001	χ =27,839; p<0,001
Combinação x Baseline	stack.ASMT	liwc.reglog	Sem significância estatística	χ =183,195; p<0,001
	concat.ASMT	liwc.reglog	Sem significância estatística	χ =146,110; p<0,001
	vot.ASMT	liwc.reglog	Sem significância estatística	χ =217,736; p<0,001

Fonte – Rafael Lage de Oliveira, 2023

Tabela 17 – Testes de significância estatística para a tarefa de predição de ansiedade

Comparação	**Modelo 1**	**Modelo 2**	**Teste McNemar**	
			Diagnosticado	**Global**
Individual x Individual	menções.n2v.svm	amigos.n2v.svm	χ =8,377; p<0,01	χ =133,500; p<0,001
	menções.n2v.svm	seguidores.n2v.svm	Sem significância estatística	χ =54,093; p<0,001
	menções.n2v.svm	topmenções.n2v.svm	χ =17,654; p<0,001	Sem significância estatística
	menções.n2v.svm	horario.svm	χ =7,758; p<0,01	χ =11,743; p<0,001
Combinação x Combinação	concat.ASMT	stack.ASMTH	χ =17,000; p<0,001	χ =12,099; p<0,001
	concat.ASMT	stack.ASMT	χ =9,000; p<0,01	χ =10,074; p<0,01
	concat.ASMT	vot.ASMT	χ =7,000; p<0,001	χ =23,048; p<0,01
	stack.ASMTH	vot.ASMT	χ =10,000; p<0,001	Sem significância estatística
Combinação x Individual	concat.ASMT	menções.n2v.svm	χ =7,116; p<0,01	Sem significância estatística
	concat.ASMT	amigos.n2v.svm	Sem significância estatística	χ =203,258; p<0,001
	concat.ASMT	seguidores.n2v.svm	Sem significância estatística	χ =96,793; p<0,001
	concat.ASMT	topmenções.n2v.svm	χ =24,000; p<0,001	Sem significância estatística
	concat.ASMT	horario.svm	χ =16,129; p<0,001	χ =18,684; p<0,001
node2vec x Seleção univariada	amigos.n2v.svm	amigos.univ.reglog	χ =21,000; p<0,001	χ =335,335; p<0,001
	seguidores.n2v.svm	seguidores.univ.reglog	χ =35,805; p<0,001	χ =79,834; p<0,001
	menções.n2v.svm	menções.univ.reglog	χ =16,000; p<0,001	χ =269,023; p<0,001
Individual x Baseline	menções.n2v.svm	liwc.reglog	Sem significância estatística	χ =66,759; p<0,001
	amigos.n2v.svm	liwc.reglog	Sem significância estatística	Sem significância estatística
	seguidores.n2v.svm	liwc.reglog	Sem significância estatística	Sem significância estatística
	liwc.reglog	topmenções.n2v.svm	χ =20,238; p<0,001	χ =55,651; p<0,001
	liwc.reglog	horario.svm	χ =11,456; p<0,001	χ =16,052; p<0,001
Combinação x Baseline	concat.ASMT	liwc.reglog	Sem significância estatística	χ =78,893; p<0,001
	stack.ASMTH	liwc.reglog	Sem significância estatística	χ =113,770; p<0,001
	vot.ASMT	liwc.reglog	χ =7,758; p<0,01	χ =137,990; p<0,001

Fonte – Rafael Lage de Oliveira, 2023