

UNIVERSIDADE NOVE DE JULHO – UNINOVE

**PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA E GESTÃO DO
CONHECIMENTO – PPGI**

SABRINNA DELGADO

**INTELIGÊNCIA ARTIFICIAL APLICADA NO MAPEAMENTO DE SINTOMAS E
TRATAMENTO DE DEPRESSÃO, ANSIEDADE E ESTRESSE**

São Paulo

2023

SABRINNA DELGADO

**INTELIGÊNCIA ARTIFICIAL APLICADA NO MAPEAMENTO DE SINTOMAS E
TRATAMENTO DE DEPRESSÃO, ANSIEDADE E ESTRESSE**

Dissertação apresentada ao Programa de Pós-Graduação em Informática e Gestão do Conhecimento da UNINOVE como parte dos requisitos para obtenção do título de Mestre em Informática.

Linha de pesquisa: Sistemas Inteligentes – LP 2

Prof. Orientador: Dr. Sidnei Alves de Araújo

Prof. Coorientador: Dr. Peterson Adriano Belan

São Paulo

2023

Delgado, Sabrina.

Inteligência artificial aplicada no mapeamento de sintomas e tratamento de depressão, ansiedade e estresse. / Sabrina Delgado. 2023.

71 f.

Dissertação (Mestrado) - Universidade Nove de Julho - UNINOVE, São Paulo, 2023.

Orientador (a): Prof. Dr. Sidnei Alves de Araújo.

1. Transtornos mentais. 2. DASS-21. 3. Mapeamento de sintomas. 4. Inteligência artificial. 5. Mineração de dados.

I. Araújo, Sidnei Alves de. II. Título.

CDU 004

PARECER – EXAME DE DEFESA

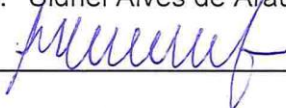
Parecer da Comissão Examinadora designada para o exame de defesa do Programa de Pós-Graduação em Informática e Gestão do Conhecimento a qual se submeteu a aluna Sabrina Delgado.

Tendo examinado o trabalho apresentado para obtenção do título de “Mestre em Informática e Gestão do Conhecimento”, com Dissertação intitulada “INTELIGÊNCIA ARTIFICIAL APLICADA NO MAPEAMENTO DE SINTOMAS E TRATAMENTO DE DEPRESSÃO, ANSIEDADE E ESTRESSE”, a Comissão Examinadora considerou o trabalho:

- Aprovada
 Reprovada com direito a novo exame
- Aprovada condicionalmente
 Reprovada

EXAMINADORES

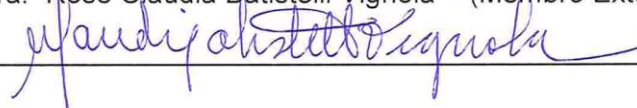
Prof. Dr. Sidnei Alves de Araújo - (Orientador)



Prof. Dr. Peterson Adriano Belan - (Coorientador/Membro Interno)



Profa. Dra. Rose Claudia Batistelli Vignola - (Membro Externo)



Prof. Dr. Renato Jose Sassi - (Membro Interno)



DEDICATÓRIA

Este projeto dedico aos meus orientadores pela sua orientação, parceria, paciência e profissionalismo. Todo o apoio e atenção dedicados a mim foram imprescindíveis para a concretização deste trabalho.

Dedico aos meus pais, meus maiores incentivadores, pela educação que me deram e pelo amor que me dedicaram.

Dedico a Deus, cuja presença me auxilia nas minhas escolhas, abrindo caminhos e me dando confiança frente aos desafios e adversidades.

AGRADECIMENTOS

Nesses anos de mestrado, de muito estudo, esforço e empenho, gostaria de agradecer a algumas pessoas que me acompanharam e foram fundamentais para a realização de mais este sonho. Por isso, expresso aqui, através de palavras sinceras, um pouquinho da importância que elas tiveram, e ainda têm, nesta conquista e a minha sincera gratidão a todas elas. Primeiramente, ao Programa de Pós-Graduação em Informática e Gestão do Conhecimento (PPGI) da Universidade Nove de Julho, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida. Um agradecimento mais que especial para meu orientador, Prof. Dr. Sidnei Alves de Araújo, por todo apoio, ajuda, orientação e zelo em todo decorrer deste trabalho, obrigada por sua dedicação, que o fez, por muitas vezes, deixar de lado seus momentos de descanso para me ajudar e me orientar. E, principalmente, obrigada por sempre ter acreditado e depositado sua confiança em mim ao longo de todos esses anos de trabalho que se iniciaram ainda na graduação. Sem sua orientação, apoio, confiança e amizade, não somente neste trabalho, mas em todo o caminho percorrido até aqui, nada disso seria possível; ao meu coorientador, Prof. Dr. Peterson Adriano Belan, por seus conselhos e orientações; a Dra. Rose Claudia Batistelli Vignola pelo fornecimento dos dados e todo apoio concedido. Não poderia deixar de agradecer aos membros da banca de Qualificação e Defesa de Mestrado, pelos conselhos, sugestões e interesse em contribuir para o desenvolvimento desta pesquisa. Gratidão aos meus pais Luciene e Wagner e ao meu irmão Luccas, pela compreensão, ao serem privados em muitos momentos da minha companhia e atenção, e pelo profundo apoio, me estimulando nos momentos mais difíceis, por desejarem sempre o melhor para mim, pelo esforço que fizeram para que eu pudesse superar cada obstáculo em meu caminho e chegar aqui e, principalmente, pelo amor imenso que vocês têm por mim. Uma gratidão mais que especial para minha mãe que como sempre mergulhou de cabeça comigo nessa jornada, me faltam até palavras para expressar o quanto te agradeço por tudo. À vocês, minha família, sou eternamente grata por tudo que sou, por tudo que consegui conquistar e pela felicidade que tenho. Sem vocês nada disso seria possível! Por fim, o agradecimento mais importante: agradeço a Deus por estar sempre comigo, me guiando, iluminando e me abençoando, por me dar a fé e a força para enfrentar todos os obstáculos, sem nunca desistir.

RESUMO

A cada ano é crescente o número de pessoas no mundo acometidas por transtornos mentais (TM), entre os quais estão a depressão, a ansiedade e o estresse que têm sido os mais comuns e que normalmente estão associados ao estilo de vida moderno. Os dois primeiros TM pertencem ao grupo das principais doenças do século XXI e podem levar a consequências graves, como o suicídio. De acordo com a Organização Mundial de Saúde (OMS), a depressão impacta a rotina de vida de mais de 300 milhões de pessoas, sendo considerada uma das doenças mais importantes do mundo. Além disso, estima-se que 12 bilhões de dias de trabalho são perdidos anualmente no mundo devido à depressão e à ansiedade, impactando em quase um trilhão de dólares na economia global. O tratamento de TM pode incluir, além de medicamentos e psicoterapias, que são essenciais, o emprego de recursos tecnológicos, como a Inteligência Artificial (IA) para indicar terapias e cuidados personalizados. Na literatura existem diversas abordagens de IA aplicadas no contexto de TM, mas é muito comum que elas sejam focadas no auxílio ao diagnóstico. Nesta pesquisa propõe-se um método de IA para mapeamento de sintomas e auxílio ao tratamento de depressão, ansiedade e estresse. Primeiro são aplicadas técnicas de mineração de dados (MD) para geração de regras que, além de mapear os sintomas, representam conhecimentos acerca de uma base contendo dados de 242 pacientes, coletados a partir de um teste denominado DASS-21 (*Depression, Anxiety and Stress Scale*). Em seguida, o conjunto de regras gerado é usado para compor um Sistema de Inferência Fuzzy (SIF) capaz de fazer previsões sobre os TM a partir dos principais sintomas e de alguns dados pessoais do paciente. As altas taxas de acerto nas tarefas de MD (acima de 90%) indicando a existência de padrões consistentes e os resultados obtidos pelo SIF demonstram que o método proposto pode auxiliar os profissionais de saúde na rápida predição de sintomas de depressão, ansiedade e estresse, em triagem ambulatorial e em pronto atendimento. Ele também pode ser útil para uma melhor associação de sintomas, propostas terapêuticas e até mesmo investigações de outras doenças não relacionadas à saúde mental, propiciando diagnósticos e tratamentos diferenciais.

Palavras-chave: Transtornos Mentais, DASS-21, Mapeamento de Sintomas, Inteligência Artificial, Mineração de Dados, Lógica *Fuzzy*.

ABSTRACT

Every year the number of people in the world affected by mental disorders (MD) increases, among which are depression, anxiety and stress that have been more common and that are usually related to the modern lifestyle. The first two belong to the group of the main diseases of the 21st century and can lead to serious consequences, such as suicide. According to the Pan World Health Organization (WHO), depression impacts the daily lives of more than 300 million people, being considered one of the most important diseases in the world. Additionally, an estimated 12 billion workdays are lost annually worldwide due to depression and anxiety, impacting nearly a trillion dollars on the global economy. TM treatment may include, in addition to medication and psychotherapies, which are essential, the use of technological resources, such as Artificial Intelligence (AI) to indicate therapies and personalized care. In the literature, there are several AI approaches applied in the context of MT, but it is very common that they are focused on aiding the diagnosis. This research proposes an AI method for mapping symptoms and helping to treat depression, anxiety and stress. First, data mining (DM) techniques are applied to generate rules that, in addition to mapping the symptoms, represent knowledge about a database containing data from 242 patients, collected from a test called DASS-21 (Depression, Anxiety and StressScale). Then, the generated set of rules is used to compose a Fuzzy Inference System (FIS) capable of making predictions about MDs based on the main symptoms and some personal data of the patient. The high hit rates in the DM tasks (above 90%) indicating the existence of consistent patterns and the results produced by the FIS demonstrate that the proposed method can help health professionals in the rapid prediction of symptoms of depression, anxiety and stress, in outpatient screening and in emergency care. It can also be useful for a better association of symptoms, therapeutic proposals and even investigations of other diseases not related to mental health, providing differential diagnoses and treatments.

Keywords: Mental Disorders, DASS-21, Symptoms Mapping, Artificial Intelligence, Data Mining, Fuzzy Logic.

LISTA DE SIGLAS

AD	Árvores de Decisão
AG	Algoritmo Genético
BDI	<i>Beck Depression Inventory</i>
CAPS	Centros de Atenção Psicossociais
Chatbot	Robô de conversação
CSV	<i>Comma-separated Values</i> (valores Separados por vírgulas)
DASS-21	<i>Depression Anxiety and Stress Scale</i>
DSM-5	<i>Diagnostic and Statistical Manual of Mental Disorders</i>
IA	Inteligência Artificial
IES-R	<i>Impact of Event Scale-Revised</i>
KDD	<i>Knowledge Discovery in Databases</i>
K-NN	<i>K-Nearest Neighbors</i>
LDA	<i>Linear Discriminant Analysis</i>
LNN	<i>Local Nearest Neighbor</i>
MATLAB	<i>MATrix LABoratory</i>
MD	Mineração de Dados
MLP	<i>Multilayer Perceptron</i>
NB	<i>Naïve Bayes</i>
OMS	Organização Mundial de Saúde
OPAS	Organização Pan-Americana de Saúde
PHQ-9	<i>Patient Health Questionnaire</i>
PLN	Processamento de Linguagem Natural
RB	Redes Bayesianas
RBFN	<i>Radial Basis Function Network</i>

RF	<i>Random Forest</i>
RNA	Redes Neurais Artificiais
RNC	Redes Neurais Convolucionais
SIF	Sistema de Inferência <i>Fuzzy</i>
SUS	Sistema Único de Saúde
SVM	<i>Support Vector Machines</i>
TM	Transtornos Mentais
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

LISTA DE ILUSTRAÇÕES

Figura 1 – Etapas do Processo KDD.....	33
Figura 2 – Exemplo de AD	36
Figura 3 – Estrutura de um Sistema de Inferência <i>Fuzzy</i> (FIS).....	39
Figura 4 – Diagrama esquemático do método proposto.....	41
Figura 5 – Fragmento da base de dados em formato CSV	43
Figura 6 – Variáveis de entrada/saída do mecanismo de inferência fuzzy.....	45
Figura 7 – Funções de pertinência associadas à variável idade	45
Figura 8 – Funções de pertinência associadas à variável gênero.....	46
Figura 9 – Funções de pertinência associadas à variável estado civil	46
Figura 10 – Funções de pertinência associadas aos sintomas (S1 a S21)	46
Figura 11 – Funções de pertinência associadas às variáveis de saída.....	47
Figura 12 – Tela do Matlab para predições do SIF	57
Figura 13 – Algumas predições do SIF para depressão	57
Figura 14 – Algumas predições do SIF para ansiedade.....	58
Figura 15 – Algumas predições do SIF para estresse.....	58

LISTA DE TABELAS

Tabela 1 – Divisão dos trabalhos encontrados na literatura a partir das combinações das palavras-chaves	17
Tabela 2 – Respostas do DASS-21 usando pontuação na escala de Likert de 0 a 330	
Tabela 3 – Itens (sintomas) do DASS-21 referentes a cada transtorno	31
Tabela 4 – Níveis de gravidade.....	32
Tabela 5 – Percentuais de Depressão, Ansiedade e Estresse.....	40
Tabela 6 – Matriz de confusão obtida na classificação de dados relativos ao TM “Depressão”.....	49
Tabela 7 – Matriz de confusão obtida na classificação de dados relativos ao TM “Ansiedade”	50
Tabela 8 – Matriz de confusão obtida na classificação de dados relativos ao TM “Estresse”	52
Tabela 9 – Classificação dos atributos dos transtornos do DASS-21	54
Tabela 10 – Carga fatorial vs. ganho de informação dos sintomas para cada TM....	55

SUMÁRIO

1. INTRODUÇÃO	15
1.1 CONTEXTUALIZAÇÃO DO TEMA	15
1.2 TRABALHOS CORRELATOS E LACUNAS DE PESQUISA	16
1.2.1 IA no diagnóstico de transtornos mentais	17
1.2.2 IA no tratamento de transtornos mentais	22
1.2.3 Sumarização da revisão da literatura e lacunas de pesquisa	23
1.3 PERGUNTA DE PESQUISA.....	24
1.4 OBJETIVOS.....	24
1.4.1 Objetivo geral.....	24
1.4.2 Objetivos específicos	24
1.5 JUSTIFICATIVA.....	25
1.6 ESTRUTURA DO TRABALHO.....	26
2. FUNDAMENTAÇÃO TEÓRICA.....	27
2.1 TRANSTORNOS MENTAIS (TM)	27
2.1.1 Escala de depressão, ansiedade e estresse (<i>Depression, Anxiety and Stress Scale – DASS</i>).....	30
2.2 INTELIGÊNCIA ARTIFICIAL (IA)	32
2.3 KNOWLEDGE DISCOVERY IN DATABASES – KDD	33
2.3.1 Etapas do KDD	33
b) Pré-processamento	34
c) Transformação dos dados	34
d) Mineração de Dados (MD).....	34
e) Interpretação/Avaliação	35
2.4 ÁRVORES DE DECISÃO	35

2.5 ALGORITMO APRIORI	37
2.6 LÓGICA FUZZY	38
3. MATERIAL E MÉTODO	40
3.1 BASE DE DADOS	40
3.2 MÉTODO PROPOSTO	41
3.2.1 <i>Aplicação das etapas do KDD para mapeamento dos sintomas</i>	42
3.2.2 <i>Sistema de Inferência Fuzzy (SIF)</i>	44
4. RESULTADOS E DISCUSSÃO	48
4.1 DESCOBERTA DE PADRÕES POR MEIO DE ÁRVORES DE DECISÃO (CLASSIFICAÇÃO)	48
4.1.1 <i>Depressão</i>	48
4.1.2 <i>Ansiedade</i>	50
4.1.3 <i>Estresse</i>	51
4.2 DESCOBERTA DE PADRÕES POR MEIO DO ALGORITMO APRIORI (REGRAS DE ASSOCIAÇÃO)	53
4.3 ANÁLISE DA IMPORTÂNCIA DOS SINTOMAS NA DETERMINAÇÃO DOS GRAUS DOS TM (SELEÇÃO DE ATRIBUTOS)	54
4.4 SISTEMA DE INFERÊNCIA FUZZY (SIF)	56
4.5 DISCUSSÃO DOS RESULTADOS	59
5. CONCLUSÕES E TRABALHOS FUTUROS	62
REFERÊNCIAS	64
APÊNDICE I. AD gerada na classificação dos dados relativos à depressão	69
APÊNDICE II. AD gerada na classificação dos dados relativos à ansiedade	70
APÊNDICE III. AD gerada na classificação dos dados relativos ao estresse	71

1. INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO DO TEMA

Os Transtornos Mentais (TM) estão relacionados à ordem mental ou psicológica de um indivíduo. Entre os TM mais comuns estão a depressão, a ansiedade, o estresse, a síndrome de Burnout e o transtorno bipolar, os quais possuem algum tipo de associação com o estilo de vida moderno (BALDISSERA, 2021). Alguns TM, como a depressão e a ansiedade, pertencem ao grupo das principais doenças do século XXI e podem levar a consequências graves, como o suicídio. Além disso, é comum que uma pessoa que tenha um desses TM possa desencadear o outro, ou ainda desenvolver mais de um deles ao mesmo tempo. Infelizmente muitas pessoas não acreditam que tais TM sejam tão perigosos e, por isso, acabam tendo seu estado de saúde agravado (BALDISSERA, 2021). De acordo com Vignola (2013), o estresse é cada vez mais apontado na literatura como um fator de risco para a depressão e a ansiedade.

De acordo com a Organização Pan-Americana da Saúde (OPAS), a depressão impacta a rotina de vida de mais de 300 milhões de pessoas, sendo considerada uma das doenças mais incapacitantes do mundo e, principal candidata ao posto de doença mais comum do mundo até 2030 (OPAS, 2022a).

No primeiro ano da pandemia de COVID-19, a OMS relatou que a prevalência global de ansiedade e depressão aumentou em 25%. Além disso, uma pesquisa realizada pela OMS indicou que a pandemia agravou o estado da saúde mental dos brasileiros e, concluiu que há na população uma elevada proporção de ansiedade (86,5%), seguida por eventos de estresse pós-traumático (45,5%) e, por fim, quadros depressivos graves (16%) (BALDISSERA, 2021). Estima-se que anualmente 12 bilhões de dias de trabalho são perdidos por causa de depressão e ansiedade, custando à economia global quase 1 trilhão de dólares (OPAS, 2022b).

Com o crescimento em ascensão dos TM, o tempo gasto para obtenção de um diagnóstico preciso tem aumentado, o que pode levar à desistência e/ou piora no quadro clínico do paciente. Soma-se a isso a falta de capacitação e ferramentas que permitam aos profissionais de saúde melhor compreensão das reais necessidades dos pacientes e oferecer diagnósticos e tratamentos diferenciais (AVENI, 2020).

Assim, além do uso de medicamentos e psicoterapias, o emprego de recursos tecnológicos vem ganhando cada vez mais espaço no auxílio ao tratamento e diagnóstico de TM. Um desses recursos é a Inteligência Artificial (IA), cuja utilização no campo da medicina tem aumentado significativamente nos últimos anos em virtude da capacidade de auxílio ao diagnóstico, produção de insights a partir de grandes conjuntos de dados clínicos e de propiciar tratamentos personalizados (AVENI, 2020; PINTO-COSTA, 2021). Nesse contexto, são desenvolvidos sistemas computacionais que empregam técnicas de Aprendizagem de Máquina (AM) e de Mineração de Dados (MD) para diagnóstico e tratamento de TM, auxiliando os profissionais de saúde no acompanhamento diário de seus pacientes. Também é possível criar sistemas inteligentes que interpretam os comportamentos de pacientes visando auxiliar em seus tratamentos (MENDES *et al.*, 2020).

Contudo, apesar dos inúmeros trabalhos na literatura propondo aplicações de IA para lidar com TM, a maioria deles é voltada para auxílio ao diagnóstico, deixando uma lacuna no que tange as abordagens para mapeamento de sintomas e tratamento, que possam subsidiar tomada de decisão das equipes multiprofissionais de saúde, principalmente em alguns órgãos públicos com escassez de profissionais. É nesse contexto que se insere a presente pesquisa, que explora o uso de técnicas de MD e de IA para compor um método para mapeamento de sintomas e auxílio ao tratamento de depressão, ansiedade e estresse, a partir de dados coletados com o uso de um teste denominado DASS-21 (*Depression, Anxiety and Stress Scale*).

1.2 TRABALHOS CORRELATOS E LACUNAS DE PESQUISA

Para identificação das lacunas de pesquisa, o primeiro passo foi realizar uma revisão da literatura considerando o período de 2017 a 2022. A pesquisa foi feita nas bases Science Direct, Compendex e Scopus. Após uma filtragem para selecionar os trabalhos mais relevantes e com maiores números de citações, restaram os 26 trabalhos apresentados nas subseções 1.2.1 e 1.2.2, que foram divididos em dois temas definidos pelas combinações das palavras-chave utilizadas na pesquisa (em Português e Inglês), como indicado Tabela 1.

Tabela 1 – Divisão dos trabalhos encontrados na literatura a partir das combinações das palavras-chaves

Tema/Palavras-chaves	Autores e ano de publicação
<p>IA no diagnóstico de transtornos mentais</p> <ul style="list-style-type: none"> • (Aprendizado de Máquina ou Inteligência Artificial) + Diagnóstico + (Doença Mental ou Transtorno Mental) • (Aprendizado de Máquina ou Inteligência Artificial) + DASS-21 + (Doença Mental ou Transtorno Mental) 	<p>Sau <i>et al.</i> (2017), Srividya <i>et al.</i> (2018), Fonseca <i>et al.</i> (2019), Narayanrao e Kumari (2020), Nayan <i>et al.</i> (2022), Vaishnavi <i>et al.</i> (2022), Graham <i>et al.</i> (2019), Bernert <i>et al.</i> (2020), Le Glaz <i>et al.</i> (2021), Altintas <i>et al.</i> (2021), Joshi <i>et al.</i> (2022)</p> <p>Choudhury <i>et al.</i> (2019), Budiyanto <i>et al.</i> (2019), Priya <i>et al.</i> (2020), Kumar <i>et al.</i> (2020), Nizam <i>et al.</i> (2020), Fátima <i>et al.</i> (2021), Palattao <i>et al.</i> (2021), Elahi <i>et al.</i> (2021), Wesabi <i>et al.</i> (2022)</p>
<p>IA no tratamento de transtornos mentais</p> <ul style="list-style-type: none"> • (Aprendizado de Máquina ou Inteligência Artificial) + Tratamento + (Doença Mental ou Transtorno Mental) 	<p>Abreu (2020), Paton e Tiffin (2022), Fulmer <i>et al.</i> (2018), Sangani <i>et al.</i> (2022)</p>

Fonte: Autoria própria.

1.2.1 IA no diagnóstico de transtornos mentais

Os trabalhos de Sau *et al.* (2017), Srividya *et al.* (2018), Fonseca *et al.* (2019), Narayanrao e Kumari (2020), Nayan *et al.* (2022) e Vaishnavi *et al.* (2022) tiveram como foco o desenvolvimento e/ou comparação de algoritmos de aprendizado de máquina (AM) aplicados na identificação e/ou classificação dos graus de TM.

Sau *et al.* (2017) desenvolveram um modelo preditivo para diagnosticar ansiedade e depressão em pacientes idosos a partir de fatores sociodemográficos relacionados à saúde. Para tanto, eles avaliaram dez classificadores com um conjunto de dados de 510 pacientes geriátricos e observaram que maior precisão de predição, de 89%, foi obtida com o classificador *Random Forest* (RF). O modelo baseado em RF foi testado com outro conjunto de dados de 110 pacientes e sua precisão preditiva foi de 91%, com taxa de falso positivo (FP) de 10%.

Srividya *et al.* (2018) propuseram a aplicação de vários algoritmos de AM, como *Support Vector Machines* (SVM), Árvores de Decisão (AD), RF, *Naïve Bayes* (NB), *K-Nearest Neighbor* (KNN) e Regressão Logística (RL) para identificar o estado de saúde mental em um grupo-alvo. Os experimentos conduzidos por eles demonstraram que SVM, KNN e RF tiveram desempenho quase equivalente, e que o uso dos classificadores de forma conjunta melhorou significativamente o desempenho da predição de saúde mental, alcançando 90% de precisão.

Fonseca *et al.* (2019) propuseram o uso de Redes Neurais Artificiais (RNA) para auxiliar no diagnóstico de transtorno bipolar, depressão e esquizofrenia. Para tanto, dois modelos de RNA foram treinados. Para o primeiro modelo empregou-se um banco de dados de distribuição livre da *Stanley Neuropathology Consortium*, o qual consiste de biomarcadores inflamatórios e características da população com diagnósticos de esquizofrenia e transtorno bipolar, além de dados de um grupo controle. Já para o segundo modelo foi utilizado outro banco de dados, com variáveis bioquímicas, características da população e respostas de questionários com diagnósticos de depressão, transtorno bipolar e dados de um grupo controle (sem transtornos). Como resultados eles obtiveram RNAs treinadas com mais de 80% de acurácia nas classificações dos diagnósticos.

Narayanrao e Kumari (2020) analisaram diferentes algoritmos e classificadores de AM, como AD, SVM, NB, RL e KNN, aplicados na identificação do estado de saúde mental em alunos do ensino médio, universitários e profissionais liberais. Os autores também propuseram um mecanismo de aprendizagem profunda para extrair padrões de textos coletados no Twitter, visando detectar se um determinado Twit é depressivo ou não. Contudo, não foram relatados resultados quantitativos dos experimentos conduzidos pelos autores.

Nayan *et al.* (2022) aplicaram os algoritmos RL, RF, SVM, *Linear Discriminant Analysis* (LDA), KNN e NB na predição de doenças mentais entre estudantes universitários da cidade de Dhaka em Bangladesh. Tais algoritmos foram treinados com dados sociodemográficos e de testes comportamentais (incluindo a escala *Patient Health Questionnaire* (PHQ-9) e a escala *Generalized Anxiety Disorder Assessment-7*). Os resultados mostraram que o RF superou outros algoritmos na predição de depressão (acurácia de 89%), enquanto o SVM forneceu o melhor resultado para a predição de ansiedade (91,49% de acurácia).

Vaishnavi *et al.* (2022) compararam cinco algoritmos de AM (RL, K-NN, AD, RF e Stacking) na identificação de problemas de saúde mental. De acordo com os autores, a partir dos resultados obtidos ficou evidente que todos os cinco algoritmos fornecem bons resultados, tendo em vista que a acurácia de todos os classificadores foi superior a 79%. Contudo, verificou-se que as melhores predições foram obtidas com Stacking que apresentou acurácia de 81,75%.

A partir dos trabalhos de Sau *et al.* (2017), Srividya *et al.* (2018), Fonseca *et al.* (2019), Narayanrao e Kumari (2020), Nayan *et al.* (2022) e Vaishnavi *et al.* (2022) pode-se perceber que RF, SVM e AD são técnicas bastantes promissoras para implementação de modelos de predição aplicados ao diagnóstico de TM. Além desses trabalhos, destacam-se também os trabalhos de Graham *et al.* (2019), Bernert *et al.* (2020), Le Glaz *et al.* (2021) e Joshi *et al.* (2022) que embora não tenham proposto ferramentas e/ou modelos para auxílio ao diagnóstico de TM, mas realizaram importantes levantamentos e comparações das técnicas de AM amplamente empregadas nesse contexto.

Graham *et al.* (2019) conduziram um estudo de revisão da literatura sobre uma visão geral da aplicação de IA na área da saúde, incluindo estudos recentes sobre aplicações de IA específicas para a saúde mental. Eles apresentam uma discussão sobre como a IA pode complementar a prática clínica, considerando suas limitações atuais, áreas que necessitam de pesquisas adicionais e as implicações éticas relacionadas às tecnologias de IA.

Bernert *et al.* (2020) realizaram uma revisão sistemática da literatura com foco na aplicação de IA e AM na identificação e na avaliação de comportamentos suicidas,

e concluíram que resultados das pesquisas sugerem altos níveis de precisão de classificação de risco e predição de comportamentos suicidas.

Le Glaz *et al.* (2021) conduziram uma revisão sistemática da literatura com intuito de resumir e caracterizar, em termos metodológicos e técnicos, estudos que empregaram AM e técnicas de Processamento de Linguagem Natural (PLN) para saúde mental. Também foi objetivo do estudo avaliar o uso potencial dos métodos propostos na prática clínica de saúde mental. De acordo com os autores, nos 58 artigos selecionados para o estudo os principais objetivos foram extrair sintomas, classificar a gravidade da doença, comparar a eficácia da terapia e fornecer pistas psicopatológicas. No que diz respeito aos métodos utilizados, pode-observar o uso de PLN para extração de padrões em textos médicos e classificadores baseados em diversos algoritmos de AM para auxílio ao diagnóstico.

Altintas *et al.* (2021) e Joshi *et al.* (2022) analisaram diversos estudos existentes baseados em IA e técnicas de AM que estão sendo utilizadas para detectar a depressão, ansiedade e estresse. Além disso, os autores discutiram diferentes abordagens usadas para detectar emoção e humor em um indivíduo. No estudo de Joshi *et al.* (2022) analisou-se ainda como expressões faciais, imagens, chatbots emocionais e textos em plataformas de mídia social podem ser eficazes na detecção das emoções e, posteriormente, da depressão.

Os trabalhos até aqui descritos abordam o uso de AM no auxílio ao diagnóstico de TM empregando diferentes técnicas e fontes de dados. Estudos que consideram o uso do DASS-21 na obtenção dos dados usados no treinamento dos algoritmos de AM são apresentados a seguir.

Choudhury *et al.* (2019) empregaram os algoritmos RF, SVM e K-NN para predição de depressão em universitários, em seus estágios iniciais, com o objetivo de recomendação a um psiquiatra de modo a evitar incidentes dolorosos como o suicídio. Os dados para a pesquisa foram coletados por meio de uma pesquisa elaborada após consulta a psicólogos, orientadores e professores. Na coleta de dados foram empregadas as escalas DASS-21 e *Beck Depression Inventory-II* (BDI-II). Segundo os autores, o melhor método preditivo foi o RF, com uma acurácia de 75%.

Budiyanto *et al.* (2019) investigaram a possibilidade de mensurar a tendência de depressão e ansiedade por meio de mineração de textos postados no Facebook, usando parâmetros do DASS-21. Para classificação dos padrões extraídos dos textos eles utilizaram o algoritmo de *Naïve Bayes* (NB). Porém, resultados do desempenho do classificador não foram relatados pelos autores.

Priya *et al.* (2020) investigaram a utilização de AM para prever depressão, ansiedade e estresse com base em dados coletados pelo DASS-21. Levando em consideração uma amostra de 348 pacientes em seus experimentos, os pesquisadores avaliaram cinco algoritmos de ML (AD, RF, NB, SVM e K-NN) e chegaram à conclusão de que o RF apresentou os melhores resultados em termos de acurácia.

Kumar *et al.* (2020) investigaram a predição de ansiedade, depressão e estresse aplicando oito algoritmos de AM em dados obtidos com o uso do DASS-42 e do DASS-21. Eles consideraram os seguintes algoritmos: NB, Redes Bayesianas (RB), K-star, *Local Nearest neighbor* (LNN), *Multilayer Perceptron* (MLP), *Radial Basis Function Network* (RBFN), RF e AD. Um algoritmo de classificação híbrido também foi aplicado para a predição de diferentes níveis de gravidade de ansiedade, depressão e estresse. De acordo com os autores, os resultados mostraram que a RBFN teve o melhor desempenho para depressão em ambos os conjuntos de dados. No entanto, o resultado do RF foi de 100% para ansiedade no DASS-21.

Nizam *et al.* (2020) desenvolveram um aplicativo utilizando AM para identificar depressão a partir de dados coletados com o DASS-21. De acordo com os autores, a contribuição deste estudo reside em obter conhecimento sobre a viabilidade da utilização de medidas psicológicas em aplicativos móveis de detecção de depressão, contando com o feedback de especialistas médicos durante o processo de desenvolvimento.

Fátima *et al.* (2021) propuseram um modelo de aprendizado de máquina semi-supervisionado para extrair informações sobre depressão, ansiedade e estresse a partir de texto escrito. Em seus experimentos foi realizado o treinamento para identificar como as sequências de palavras emocionais que se correlacionam com os níveis de depressão, ansiedade e estresse, baseando-se no DASS-21. Esse modelo

foi aplicado a um conjunto de dados clínicos composto por 142 notas de suicídio, e constatou-se que os níveis previstos de depressão e ansiedade correspondiam a diferenças na valência e excitação, conforme esperado em um modelo de afeto.

Palattao *et al.* (2021), Elahi *et al.* (2021) e Wesabi *et al.* (2022) empregaram AM na investigação dos níveis de depressão, ansiedade e estresse durante o isolamento decorrente da COVID-19, também considerando o DASS-21.

Palattao *et al.* (2021) conduziram um estudo com o objetivo determinar os possíveis fatores contribuintes para a depressão, ansiedade e estresse, e o impacto psicológico adverso na população em geral das Filipinas usando abordagens de AM. Os autores consideraram dados de 2119 participantes que responderam a uma pesquisa online, considerando a escala *Impact of Event Scale-Revised* (IES-R) além da DASS-21.

Elhai *et al.* (2021) treinaram vários algoritmos de AM utilizando dados de 908 indivíduos chineses coletados em entre fevereiro e março de 2020, os quais apresentaram pelo menos um diagnóstico de TM. O estudo confirmou a ansiedade com a saúde como o preditor mais importante para a percepção de ameaça de morte por COVID-19.

Já o estudo de Wesabi *et al.* (2022) teve como foco o desenvolvimento de um modelo preditivo baseado em técnicas de AM e de otimização voltado para estudantes universitários. De acordo com os autores, os resultados experimentais do modelo proposto foram promissores em comparação com outros métodos.

1.2.2 IA no tratamento de transtornos mentais

Abreu (2020) exploraram as diferentes utilidades que a Realidade Virtual oferece ao tratamento de TM, bem como as formas em que a IA pode auxiliar no tratamento das doenças mentais. De acordo com os autores, pressupõe-se que a IA na psicologia é benéfica e produtiva, e numa integração disciplinar com a neurociência que estuda o sistema nervoso pode-se aprimorar os estudos nessa área.

Paton e Tiffin (2022) exploram as possibilidades de utilização de abordagens de IA e AM para melhorar a prestação de serviços de saúde mental para crianças e

adolescentes. De acordo com os autores, tais abordagens poderiam ser usadas para estender, em vez de substituir, médicos humanos e ajudar a resolver algumas das escassezes agudas de força de trabalho, de forma a liberar tempo da equipe para se concentrar nos aspectos humanos do atendimento. Os autores concluem seu estudo dizendo que se a IA “é um poderoso "martelo", é vital identificar os pregos certos para que essa tecnologia se traduza em benefícios tangíveis para os serviços de saúde.

Fulmer *et al.* (2018), Wang *et al.* (2020) e Sangani *et al.* (2022) conduziram estudos com foco no uso de *chatbots* (robôs de conversação) como alternativa para o tratamento de TM.

O estudo de Fulmer *et al.* (2018) teve como objetivo avaliar a viabilidade e a eficácia do uso de um *chatbot*, que eles denominaram “IA psicológica integrativa”, para reduzir os sintomas autoidentificados de depressão e ansiedade em estudantes universitários recrutados em 15 universidades nos Estados Unidos. De acordo com os autores, os resultados indicaram uma redução significativa nos sintomas de ansiedade dos indivíduos que usaram a ferramenta de IA.

Wang *et al.* (2020) propuseram um *chatbot* para monitorar e avaliar o estado mental de mulheres perinatais. Eles empregaram algoritmo de AM supervisionado para analisar 31 características de 223 amostras e treinar um modelo para determinar o índice de ansiedade, depressão e hipomania de mulheres perinatais. Em adição, escalas de testes psicológicos foram usadas para auxiliar na avaliação e fazer sugestões de tratamento para ajudar na melhora da saúde mental das pacientes.

Por fim, Sangani *et al.* (2022) fizeram uma revisão de literatura e apresentaram uma arquitetura de *chatbot* para detectar e auxiliar na recuperação de indivíduos com depressão, estresse e ansiedade. De acordo com os autores, o *chatbot* tem como objetivo ajudar as pessoas recomendando sessões de terapia adequadas aos seus níveis de depressão.

1.2.3 Sumarização da revisão da literatura e lacunas de pesquisa

Como pode ser visto nas subseções anteriores, boa parte dos trabalhos da literatura têm foco no auxílio ao diagnóstico usando técnicas como RF, SVM, AD, NB, RL e RNA, sendo RF a técnica que apresenta os melhores resultados. Vale ressaltar

que essa técnica é baseada em um conjunto de ADs que compõem a “floresta”. Daí o seu nome “floresta aleatória” (*Random Forest* – RF). Destaca-se ainda a pouca utilização de técnicas capazes de lidar com dados ruidosos ou imprecisos, como a lógica fuzzy.

Observa-se ainda na literatura uma carência de pesquisas que visam a aplicação de IA na construção de sistemas computacionais para mapeamento de sintomas e que levem em conta dados dos pacientes para auxiliar no tratamento de TM. Tais sistemas podem ser de grande interesse dos profissionais de saúde, visto que podem proporcionar uma melhor compreensão dos padrões associados aos sintomas de TM permitindo tratamentos e diagnósticos que levam em conta as particularidades de cada paciente.

1.3 PERGUNTA DE PESQUISA

Com base nas lacunas encontradas na análise dos trabalhos da literatura apresentados na seção anterior, esta pesquisa se propõe a responder a seguinte questão de pesquisa: como desenvolver um método baseado em IA para auxiliar os profissionais de saúde no mapeamento e predição de sintomas de depressão, ansiedade e estresse na triagem ambulatorial e/ou pronto atendimento, a partir de dados coletados com o uso do DASS-21?

1.4 OBJETIVOS

1.4.1 Objetivo geral

Desenvolvimento de um método que combina Mineração de Dados (MD) e Lógica Fuzzy para mapeamento de sintomas e auxílio ao tratamento de depressão, ansiedade e estresse.

1.4.2 Objetivos específicos

- Tratar (aplicando as etapas do KDD) a base de dados coletados com o uso do DASS-21 visando adequá-la para aplicação dos algoritmos de MD.

- Aplicar Árvores de Decisão (AD) e o algoritmo Apriori, que consistem em técnicas de MD, para geração de regras que representam conhecimentos sobre a base de dados dos pacientes, permitindo mapear os sintomas dos TM.
- Aplicar o conjunto de regras geradas pelas AD na composição de um Sistema de Inferência Fuzzy (SIF) capaz de fazer predições sobre os TM a partir dos principais sintomas e de alguns dados pessoais, que pode ser empregado na triagem de pacientes.

1.5 JUSTIFICATIVA

De acordo com a Organização Pan-Americana da Saúde (OPAS), depressão impacta a rotina de vida de mais de 300 milhões de pessoas, sendo considerada uma das doenças mais importantes do mundo (OPAS, 2022a).

Além disso, estima-se que 12 bilhões de dias de trabalho são perdidos anualmente no mundo devido à depressão e à ansiedade, impactando em quase um trilhão de dólares na economia global. O tratamento de TM pode incluir, além de medicamentos e psicoterapias, que são essenciais, o emprego de recursos tecnológicos, como a Inteligência Artificial (IA).

Na pandemia de COVID-19, a prevalência global de ansiedade e depressão, aumentou em 25%, de acordo com um resumo científico divulgado pela (OPAS, 2022c). Como consequência disto, quando essas pessoas são infectadas, elas são mais propensas a sofrer hospitalização, doença grave e morte em comparação com pessoas sem transtornos mentais. Pessoas com TM mais graves, como psicoses, e jovens com TM, estão particularmente em risco (OPAS, 2022c).

De acordo com a OPAS, apesar dos dados alarmantes sobre as doenças e transtornos mentais, atualmente muitas pessoas continuam sem os cuidados e o apoio que precisam para tratar suas condições de saúde mental pré-existentes e recém-desenvolvidas (OPAS, 2022d).

Conforme instrui a OPAS (2022e), o compromisso com a saúde mental precisa ser acompanhado por um aumento global de investimento. A situação ressalta uma escassez global crônica de recursos de saúde mental que continua. O mais recente Atlas de Saúde Mental da OMS mostrou que, em 2020, governos em todo o mundo

gastaram em média pouco mais de 2% de seus orçamentos de saúde mental e muitos países de baixa renda relataram ter menos de um profissional de saúde mental por cada 100 mil habitantes.

Foram divulgados em dezembro de 2019, pelo Departamento de Informática do SUS (DATASUS), dados sobre os CAPS (Centros de Atenção Psicossocial), além da quantidade de psicólogos e psiquiatras no País na rede pública e privada. O Nexo Jornal apresentou uma série de infográficos sobre os números disponibilizados, dentre eles o fato de que 300 cidades não possuem psicólogos e quase 3 mil estão sem psiquiatras. São cerca de 4,1 psicólogos para cada 10 mil habitantes no Brasil e 1,1 psiquiatra também para cada 10 mil habitantes (LOCH, 2020).

Os relatos sobre problemas tão atuais evidenciam a importância da busca por alternativas que contribuam para amenizar esse cenário tão alarmante, sendo uma dessas alternativas o uso de IA para geração de conhecimentos para auxiliar os profissionais de saúde na obtenção de diagnósticos e tratamentos diferenciais.

Assim, o emprego de MD pode auxiliar as equipes multiprofissionais de saúde na rápida predição de sintomas de depressão, ansiedade e estresse, tanto em consultas clínicas quanto não clínicas. Não obstante, o SIF proposto pode ser útil para triagem ambulatorial e pronto atendimento, bem como em entrevistas cujas queixas não sejam claras, visando melhor associação de sintomas às propostas terapêuticas.

1.6 ESTRUTURA DO TRABALHO

Este trabalho está estruturado como segue: O capítulo 1 apresenta a contextualização, os trabalhos correlatos encontrados na literatura, as lacunas e a pergunta de pesquisa, os objetivos e, por fim, a justificativa da pesquisa. Em seguida, no capítulo 2 é apresentada a fundamentação teórica, sendo descritos conceitos sobre transtornos mentais, inteligência artificial, mineração de dados e lógica *fuzzy*, necessários ao entendimento do método proposto. O capítulo 3 contempla a caracterização da pesquisa, a base de dados empregada e o método proposto. Em seguida, no capítulo 4 são apresentados os resultados obtidos com o método proposto e a discussão dos resultados. Por fim, no capítulo 5 são apresentadas as conclusões, principais contribuições científicas e tecnológicas, e as sugestões para trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os conceitos teóricos necessários para o entendimento do método proposto neste trabalho. Primeiro são apresentadas algumas definições sobre transtornos mentais (seção 2.1). Em seguida são apresentados o DASS-21 – *Depression, Anxiety and Stress Scale* (seção 2.2), conceitos sobre IA (seção 2.3), *Knowledge Discovery in Database* – KDD e Mineração de Dados (seção 2.4), Árvores de Decisão (seção 2.5), algoritmo Apriori (seção 2.6) e Lógica Fuzzy (seção 2.7).

2.1 TRANSTORNOS MENTAIS (TM)

Os transtornos mentais (TM) estão relacionados à ordem mental ou psicológica. Eles incluem qualquer quadro que possa comprometer a vida pessoal, familiar, social e profissional de um paciente, influenciando inclusive a forma como ele enxerga a si próprio, as pessoas e situações ao seu redor (OPAS, 2022d).

As transformações radicais pelas quais a sociedade vem passando nas últimas décadas, especialmente por conta dos avanços tecnológicos, da globalização e dos novos meios de comunicação, estão levando a uma série de problemas comportamentais, físicos e emocionais. Em meio a tantas disponibilidades de comunicação, como as redes sociais, TVs, o acesso a tantas informações faz com que as pessoas vivam sob intensa pressão, seja no trabalho ou na vida pessoal. Esse contexto é propício para o surgimento de diversos TM, como depressão, ansiedade e estresse, entre outros (EXAL, 2019).

Conhecer a diferença entre doença, síndrome e transtorno é importante para que o paciente e as pessoas que convivem com algum deles possam lidar melhor com o quadro clínico apresentado, aumentando suas qualidades de vida e contribuindo para um restabelecimento mais rápido (MED PREV, 2023).

Sintomas são alterações no corpo, que são percebidas pelo próprio paciente, como dor de cabeça, tontura, cansaço, entre outras sensações. Eles são subjetivos, pois não é possível medi-los, embora jamais devam ser descartados ou ignorados (OPAS, 2022d).

No entanto, a doença é qualquer “ausência de saúde” acompanhada por alterações do estado de equilíbrio de uma pessoa em relação ao meio ambiente, o termo “doença” engloba o prejuízo das funções da psique, de um órgão ou do organismo, o que dá origem a sintomas e sinais característicos (MED PREV, 2023).

Já a síndrome é definida como uma reunião de sintomas e sinais associados a mais de uma causa, diferente do que acontece em uma doença, a sintomatologia das síndromes é inespecífica (MED PREV, 2023).

Assim as doenças têm uma razão conhecida e definida por trás de sua manifestação clínica, enquanto as síndromes são quadros que podem ter diversas origens. Dessa forma, alguns pacientes diagnosticados com síndromes podem nunca chegar a um veredicto definitivo sobre a causa de seus sinais e sintomas (MED PREV, 2023).

Os TM não apresentam uma única causa definida, podendo ser resultado de aspectos biológicos (como o déficit ou o excesso de produção de um neurotransmissor) e psicológicos (a forma como o paciente se comporta e interage com o ambiente). Algumas vezes, eles também são chamados de “distúrbios” (OPAS, 2022d).

Quando se fala em alterações de ordem mental, prefere-se o termo “transtorno mental” ao invés de “doença mental” ou “doença psíquica” porque o paciente raramente apresenta todos os sintomas e sinais que caracterizariam uma doença. Além disso, a escolha desse termo colabora para a mitigação do preconceito que ainda existe em torno das alterações mentais (OPAS, 2022d).

Entre os diversos TM existentes estão a depressão, a ansiedade, o estresse, o transtorno afetivo bipolar, a esquizofrenia e outras psicoses, demência, deficiência intelectual e transtornos de desenvolvimento, incluindo o autismo. Os três primeiros, considerados nesta pesquisa, são definidos a seguir:

- a) Depressão: pode ser caracterizada pelo sentimento de tristeza, perda de interesse ou prazer, sentimento de culpa ou baixa autoestima, alterações no sono e no apetite, cansaço e falta de concentração. Pessoas que sofrem com essa condição podem também apresentar múltiplas queixas físicas sem causas

aparentes. Os quadros de depressão podem persistir por um longo período ou de forma recorrente, e prejudicam substancialmente a capacidade das pessoas nas atividades profissionais e/ou acadêmicas, bem como a capacidade de lidar com o dia a dia. Quadros mais graves de depressão podem levar ao suicídio (OPAS, 2022a).

- b) Ansiedade: pode ser causada por preocupações excessivas, a ponto de serem desproporcionais aos motivos. A ansiedade pode ser percebida quando a pessoa apresenta dificuldades para relaxar, mostra-se muito nervosa, se sente bastante pressionada pelos compromissos e se cobra muito. Quadros mais graves de ansiedade podem levar a crises de ansiedade que acarretam sintomas como respiração ofegante, taquicardia, e dificuldades para se auto localizar no espaço e no tempo (EXAL, 2019).
- c) Estresse: pode ser proveniente do trabalho desgastante, excesso de expectativas sociais e preocupações com o futuro. Algumas causas comuns são: problemas financeiros, dificuldades no trabalho e dificuldades para lidar com familiares e amigos. Estresse persistente por um longo período de tempo pode levar a sintomas como dificuldade de concentração, alterações de memória, mau humor, ansiedade e até mesmo depressão (CABRAL *et al.*, 2016; EXAL, 2019). O estresse também pode ser causado pela ansiedade e pela depressão em virtude de mudança brusca no estilo de vida e de exposição a um determinado ambiente, que provoquem algum tipo de angústia (CABRAL *et al.*, 2016).

Como pode ser observado, depressão, ansiedade e estresse são TM que possuem relações entre si, sendo que o estresse é cada vez mais apontado na literatura como um fator de risco para depressão e ansiedade (VIGNOLA, 2013). Segundo Mello *et al.* (2007) depressão, ansiedade e estresse transitam, entre o afeto negativo, desconforto emocional e alterações fisiológicas: na ansiedade, há uma hiperativação fisiológica; na depressão, há uma baixa afetividade positiva, compreendida com definhamento do hipocampo; e nas atividades da hipófise, submetendo os organismos a cargas hormonais exageradas. Embora em alguns casos esses TM possam se confundir ou até mesmo se manifestarem ao mesmo

tempo, eles têm causas, sintomas e tratamentos específicos e, obviamente, precisam de acompanhamento profissional (MEDLEY, 2022).

2.1.1 Escala de depressão, ansiedade e estresse (*Depression, Anxiety and Stress Scale – DASS*)

Lovibond e Lovibond (2004) desenvolveram um instrumento para identificação de sintomas associados à depressão, ansiedade e estresse que pode ser empregado em estudos clínicos, auxiliar em diagnósticos, e para indicação de tratamento adequado levando em conta as especificidades de cada paciente. Trata-se da Escala de Depressão, Ansiedade e Estresse (*Depression,–Anxiety and Stress Scale – DASS*), que consiste em um conjunto de 42 itens divididos em três partes (escalas) com 14 itens (questões) cada (VIGNOLA, 2013). Os itens são referentes aos sintomas vivenciados pela pessoa na semana anterior à aplicação do instrumento, e são quantificados numa escala de Likert de 0 a 3, como mostra a Tabela 2 a seguir:

Tabela 2 – Respostas do DASS-21 usando pontuação na escala de Likert de 0 a 3

0	não se aplica a mim
1	aplicado a mim em algum grau, ou algumas vezes
2	se aplica a mim em um grau considerável ou em boa parte do tempo
3	se aplica a mim muito ou na maioria das vezes

Fonte: Adaptado de Vignola (2013).

O DASS-21 é uma versão reduzida do DASS-42, tem a mesma estrutura, mas requer apenas a metade do tempo para sua aplicação, uma vez que consiste em apenas 21 itens (sintomas), sendo 7 itens atribuídos a cada uma das escalas de Estresse, Ansiedade e Depressão. Tal instrumento, cujas questões estão apresentadas na Tabela 3, foi adaptado e validado para o português do Brasil por Vignola (2013). De acordo com a autora, os resultados da aplicação da versão em português do DASS-21 revelaram propriedades que comprovam a sua qualidade e capacidade de avaliar os estados emocionais associados a cada TM.

Tabela 3 – Itens (sintomas) do DASS-21 referentes a cada transtorno

Depressão	Ansiedade	Estresse
<p>S3. Não consegui vivenciar nenhum sentimento positivo.</p> <p>S5. Achei difícil ter iniciativa para fazer as coisas.</p> <p>S10. Senti que não tinha nada a desejar.</p> <p>S13. Senti-me deprimido e sem ânimo.</p> <p>S16. Não consegui me entusiasmar com nada.</p> <p>S17. Senti que não tinha valor como pessoa.</p> <p>S21. Senti que a vida não tinha sentido.</p>	<p>S2. Senti minha boca seca.</p> <p>S4. Tive dificuldade em respirar em alguns momentos. (ex. ofegante, sem ar, sem esforço físico).</p> <p>S7. Senti tremores (ex. nas mãos).</p> <p>S9. Preocupei-me com situações em que pudesse entrar em pânico e fosse ridículo.</p> <p>S15. Senti que ia entrar em pânico.</p> <p>S19. Percebi meu coração alterado mesmo não tendo feito nenhum esforço físico.</p> <p>S20. Senti medo sem motivo.</p>	<p>S1. Achei difícil me acalmar.</p> <p>S6. Tive tendência a reagir de forma exagerada às situações.</p> <p>S8. Senti que estava sempre nervoso.</p> <p>S11. Senti-me agitado.</p> <p>S12. Achei difícil relaxar.</p> <p>S14. Fui intolerante com coisas que me impedissem de continuar o que eu estava fazendo.</p> <p>S18. Senti que estava um pouco emotiva / sensível demais.</p>

Fonte: Adaptado de Vignola (2013).

Na coleta de dados, as respostas são quantificadas por meio de valores numéricos de 0 a 3, e os escores são então calculados para cada uma das três escalas (depressão, ansiedade e estresse) somando-se os valores atribuídos aos itens pertencentes a elas. Então, os escores finais são rotulados de acordo com o nível de gravidade, que pode ser Normal, Mínimo, Moderado, Grave e Muito Grave, conforme indicado na Tabela 4.

Tabela 4 – Níveis de gravidade

Níveis	Depressão	Ansiedade	Estresse
Normal	0-9	0-7	0-14
Leve	10-13	8-9	15-18
Moderado	14-20	10-14	19-25
Grave	21-27	15-19	26–32
Muito Grave	28 +	20 +	33 +

Fonte: Adaptado de Vignola (2013).

De acordo com Vignola (2013), o DASS-21 é um instrumento de rápida aplicação, sendo o seu resultado de simples aferição e podendo ser obtido imediatamente após a sua aplicação.

2.2 INTELIGÊNCIA ARTIFICIAL (IA)

A criação de sistemas computacionais com capacidade de simular ou imitar comportamentos inteligentes e relevantes aos seres humanos pode ser classificada como uma tecnologia de inteligência artificial (IA), mesmo que suas ações se dêem de forma diferente dos seres humanos (BHAVE *et al.*, 2020).

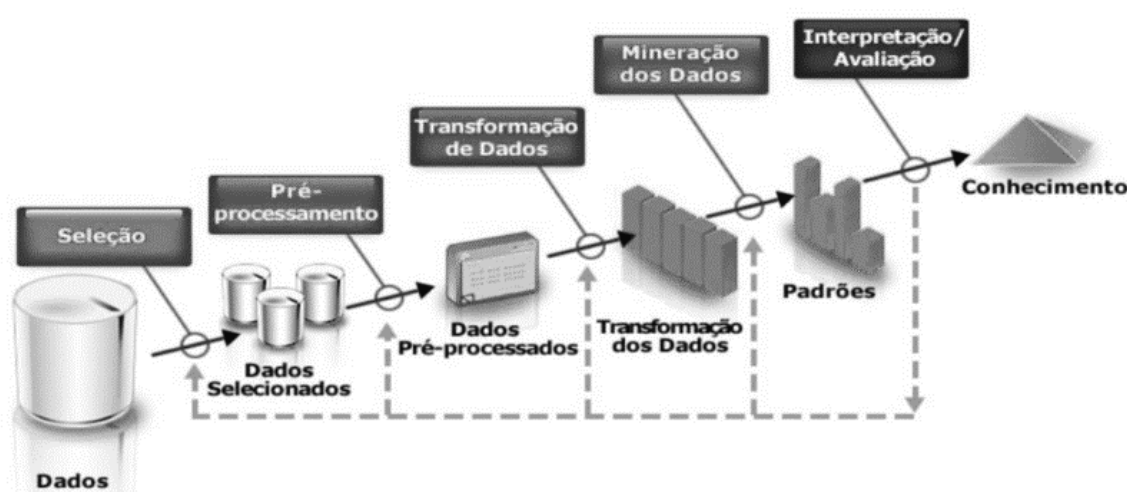
De acordo Com Russell e Norvig (1995) IA "é a arte de criar máquinas que realizam funções que exigiriam inteligência se fossem executadas por humanos". Essa definição destaca a ideia de que a IA busca replicar em máquinas a capacidade de executar tarefas que normalmente requerem algum nível de inteligência humana. Isso pode incluir tarefas como raciocínio, aprendizado, resolução de problemas, percepção visual, processamento de linguagem natural e muito mais.

Entre as técnicas de IA mais utilizadas no desenvolvimento de sistemas inteligentes estão as RNAs, as Redes Neurais Convolucionais (RNC), SVM, Lógica Fuzzy e AD.

2.3 KNOWLEDGE DISCOVERY IN DATABASES – KDD

O KDD, de acordo com Fayyad *et al.* (1996), é um processo não trivial de descoberta de conhecimento útil, válido, novo e compreensível a partir de grandes conjuntos de dados. Conforme ilustrado na Figura 1, o KDD contempla todas as etapas para a descoberta do conhecimento a partir da base de dados analisada, as quais estão descritas na subseção 2.3.1, sendo a Mineração de Dados uma das etapas do processo.

Figura 1 – Etapas do Processo KDD



Fonte: Adaptado de SILVA NETO *et al.* (2010).

2.3.1 Etapas do KDD

a) Seleção

A etapa de seleção consiste em selecionar o(s) conjunto(s) de dados que será(ão) analisado(s). As fontes de dados podem ser bancos de dados, planilhas ou data warehouses contendo dados em diversos formatos (estruturados, semiestruturados e não-estruturados). Esta é uma etapa crucial para identificar e coletar dados relevantes para a análise posterior e definir critérios de seleção de dados para filtrar informações irrelevantes e evitar a inclusão de dados incorretos ou incompletos. Durante essa etapa, também é considerada a representatividade e qualidade dos dados selecionados para garantir que os resultados da análise sejam confiáveis. Além disso, é fundamental avaliar a capacidade de processamento dos dados selecionados para que possam ser manipulados adequadamente pelos algoritmos de mineração de dados utilizados na etapa seguinte. Por fim, documentar

todas as decisões tomadas durante a seleção de dados para garantir a transparência e reprodutibilidade do processo (MURTHY, 1998).

b) Pré-processamento

A etapa de pré-processamento de dados do KDD é responsável por preparar os dados selecionados para a análise posterior. Nessa etapa, são realizadas diversas atividades, como limpeza de dados para remover ruídos e dados inconsistentes ou anomalias (outliers). O objetivo geral do pré-processamento de dados é melhorar a qualidade dos dados para que possam ser analisados por algoritmos de mineração de dados com maior precisão e eficiência (WITTEN; FRANK, 2005).

c) Transformação dos dados

A etapa de transformação de dados do KDD é responsável por aplicar técnicas de manipulação e modificação dos dados pré-processados para que possam ser analisados por algoritmos de mineração de dados. Nessa etapa, são realizadas diversas atividades, como discretização de dados contínuos, agregação de dados, normalização de dados (para garantir que todos os dados tenham a mesma escala e distribuição), combinação de dados e geração de novas características. O objetivo principal da transformação de dados é criar uma representação mais útil e significativa dos dados para a análise. Essa etapa é crucial para obter resultados precisos e confiáveis na etapa de mineração de dados, pois a qualidade dos dados transformados influencia diretamente a eficácia dos algoritmos de análise (GOLDSCHMIDT; PASSOS, 2005).

d) Mineração de Dados (MD)

O conceito de MD surgiu na década de 1990 advindo da evolução da tecnologia de banco de dados que possibilitou às organizações realizarem coleta e armazenamento contínuo de grandes volumes de dados sobre clientes, fornecedores, produtos e serviços de dados que estavam sendo subutilizados (FAYYAD *et al.*, 1996; GONÇALVES, 2011; DOMINGUES *et al.*, 2022).

A MD pode ser entendida como o processo de exploração e análise de dados por meio do uso de algoritmos de IA e de estatística, com o objetivo de descoberta de

padrões e regras que forneçam compreensão mais apurada das informações que estes conjuntos de dados “escondem” (FAYYAD *et al.*, 1996; BHAT, 2019).

As tarefas mais comuns empregadas em MD são: classificação, regras de associação e agrupamento. A classificação tem a finalidade de estudar o conjunto de dados de entrada e compor a descrição das propriedades desses dados para cada classe (FAYYAD *et al.*, 1996).

As regras de associação apresentam modelos descobertos em forma de regras SE...ENTÃO alta competência (BHAT, 2019). Em geral, os algoritmos aplicados no emprego desta tarefa criam um número de regras razoavelmente grande, deixando a cabo do utilizador fazer a seleção das regras que lhe sejam mais relevantes para dada aplicação. Por fim, o agrupamento tem o propósito de agrupar os dados investigados em dois ou mais grupos, utilizando alguma medida para demonstrar a semelhança entre os padrões conforme suas particularidades.

As técnicas e algoritmos de IA mais comumente empregados em MD são AD, RNA, Algoritmos Genéticos (AG), Lógica Fuzzy, *K-means*, SVM e Redes Bayesianas (BHAT, 2019). Nesta pesquisa foram empregadas as técnicas AD e Lógica Fuzzy, descritas nas seções a seguir, em virtude do funcionamento e resultados gerados por elas serem de fácil interpretação por pessoas de qualquer área do conhecimento (JUNG *et al.*, 2017).

e) Interpretação/Avaliação

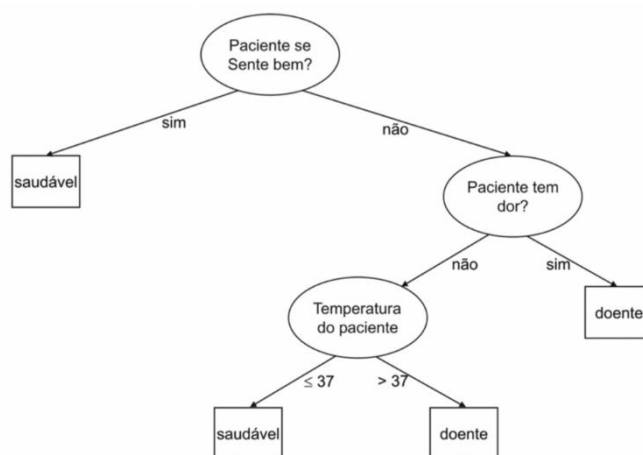
O objetivo principal da interpretação de dados é obter insights e conhecimentos a partir dos dados analisados visando solucionar problemas complexos. Além disso, é fundamental elaborar avaliações sobre os padrões encontrados na etapa anterior, por meio de análises de testes e validações usando medidas estatísticas como taxa de acerto, acurácia, precisão e índice Kappa, entre outras (FAYYAD *et al.*, 1996).

2.4 ÁRVORES DE DECISÃO

De acordo com Mitchell (2010), dentre os algoritmos de aprendizado supervisionado, a AD pode ser considerado um dos mais simples e descrito como uma coleção de regras do tipo SE...ENTÃO . Por causa disso, as AD têm sido largamente

empregadas em tarefas de classificação, como sendo um caminho eficaz na construção de classificadores que predizem classes com base nos valores dos atributos que expõem os modelos. Desta forma, elas podem ser empregues em diversas aplicações como diagnóstico médico (exemplificado na Figura 2), análise de risco e análise de crédito, entre outros exemplos.

Figura 2 – Exemplo de AD



Fonte: Monard e Baranauskas (2003).

A chave para o sucesso de um algoritmo AD é como gerar a árvore, ou seja, como escolher os atributos mais significativos para gerar as regras e quais regras podem ser descartadas da árvore (MITCHELL, 2010). Um bom conceito está na geração da AD baseado na relevância das suas particularidades, ou seja, dos seus atributos. Desta maneira, o atributo mais significativo será o que está na raiz da árvore. Sendo assim, um menor número de regras poderá ser empregado para a solução de um determinado problema. De acordo com De Araújo *et al.* (2018), em fator de que as AD são construídas com baseada na importância de cada atributo, nem sempre elas se valem de todos os atributos que refletem um padrão para geração do conjunto de regras, trazendo o benefício de reduzir o tempo computacional nas tarefas de classificação.

Entre os algoritmos de AD estão ID3 (QUINLAN, 1986; MITCHELL, 2010), C4.5 (QUINLAN, 1993) e CART (BRAMER, 2007), sendo o primeiro dos mais básicos. O algoritmo C4.5, empregado neste trabalho, constrói AD a partir de um conjunto de

dados da mesma forma que o algoritmo ID3, usando o conceito de entropia e ganho de informação para definir a importância dos atributos, conforme apresentado nas equações 1 e 2. Em cada nó da árvore, o algoritmo C4.5 escolhe o atributo que melhor particiona o conjunto de amostras em subconjuntos, tendendo a uma categoria ou outra. O atributo com maior ganho de informação normalizado é escolhido para tomar a decisão (QUINLAN, 1993).

Para realizar o cálculo do ganho de informação (Equação 2), primeiro obtém-se a entropia (Equação 1). A entropia de um conjunto S pode ser definida como sendo o grau de pureza desse conjunto, que pode ter n classes distintas.

$$Entropia(S) = \sum_{i=1}^n -P_i \log_2 P_i \quad (1)$$

P_i é a proporção de dados em S que pertencem à classe i

$$Gain(S, A) = Entropia(S) - \sum_{v \text{ valores}(A)} \frac{S_v}{S} Entropia(S_v) \quad (2)$$

v é um elemento dos valores que o atributo A pode assumir e S_v é o subconjunto de S formado pelos dados em que $A = v$.

2.5 ALGORITMO APRIORI

O algoritmo Apriori, sugerido por Agrawal e Srikant (1994), é um dos mais antigos algoritmos amplamente conhecidos para mineração de regras de associação em grandes conjuntos de dados. O nome do algoritmo é derivado do fato de que ele utiliza o conhecimento prévio das propriedades frequentes do conjunto de itens. Ele identifica todos os conjuntos de itens frequentes, conhecidos como conjuntos frequentes de itens.

A primeira etapa do algoritmo Apriori, que pode ser encontrado em Agrawal e Srikant (1994), envolve a contagem das ocorrências dos itens para determinar os conjuntos frequentes de itens de tamanho único. As etapas subsequentes compõem duas fases. Primeiramente, os conjuntos frequentes de itens encontrados na primeira etapa são empregados para gerar conjuntos candidatos de itens. Posteriormente, uma

nova busca é realizada no conjunto de dados, contabilizando o suporte de cada candidato.

Considerando o conjunto de dados hipotético {(alfa, beta, épsilon); (alfa, beta, teta); (alfa, beta, épsilon); (alfa, beta, teta)}, no qual cada par de parênteses simboliza uma linha, as seguintes regras de associação poderiam ser determinadas a partir deste conjunto de dados: a) 100% dos conjuntos com alfa também contêm beta, b) 50% dos conjuntos com alfa, beta também têm épsilon, e c) 50% dos conjuntos com alfa, beta também têm teta.

Alguns conceitos normalmente associados à geração de regras de associação são confiança e suporte. A confiança mede o quão provável é que um item B seja adquirido dado que o item A está presente. É calculada dividindo o suporte da regra (a frequência com que a regra ocorre) pelo suporte do item antecedente (a frequência com que o item antecedente ocorre). Já o suporte mede a frequência com que um item ou conjunto de itens aparece no conjunto de transações analisado. O suporte é calculado dividindo o número de transações que contêm o item ou conjunto de itens pelo número total de transações.

2.6 LÓGICA FUZZY

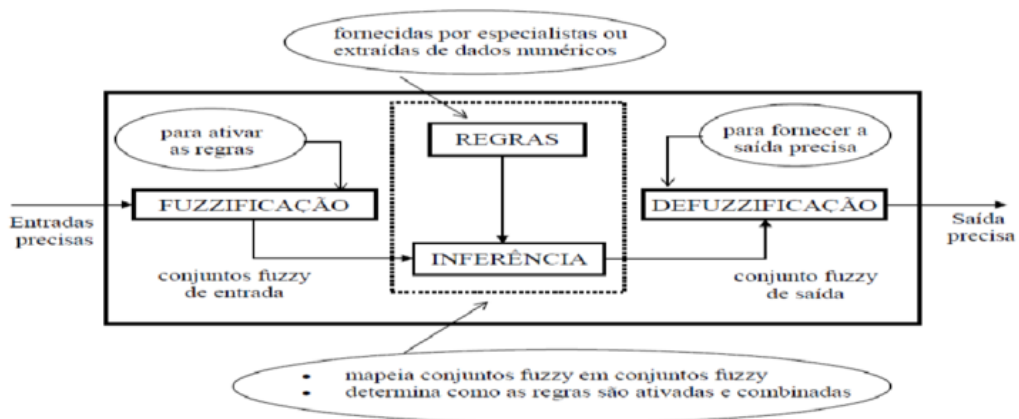
A palavra *Fuzzy* em inglês pode ter diversos significados, que mudam dependendo do contexto. “Nebuloso” e “difuso” são algumas das traduções mais comuns para *Fuzzy* (MARRO, 2010), mas o sentido fundamental desse adjetivo está sempre relacionado a algo vago, indistinto e incerto.

A lógica *Fuzzy*, também conhecida como lógica nebulosa, é considerada uma forma de lógica que lida com modelos de raciocínios imprecisos ou aproximados (ZADEH, 1973). Essa lógica permite trabalhar com problemas de decisão que não são facilmente representados em modelos matemáticos convencionais. Um controle *Fuzzy* procura “imitar” um agente humano por meio de uma descrição descritiva e experimental de um processo específico (ZADEH, 1973).

No campo da Inteligência Artificial, a lógica *Fuzzy* é considerada um mecanismo fundamental para representar de forma apropriada o conhecimento e o raciocínio, devido à sua habilidade de trabalhar com incertezas, raciocínio

aproximado, termos vagos e ambíguos, que são maneiras pelas quais as pessoas expressam seus pensamentos. Por isso, os sistemas *Fuzzy* possuem "capacidade de "raciocinar" levando em conta perspectivas relacionadas à incerteza e aos processos práticos, tornando-os mais semelhantes a um comportamento "humano" (MARRO, 2010), conforme mostrado na Figura 3.

Figura 3 – Estrutura de um Sistema de Inferência *Fuzzy* (FIS)



Fonte: Rodrigues *et al.* (2010)

McNeil e Thro (1994) relacionam algumas características de sistemas nos quais a aplicação da Lógica *Fuzzy* é benéfica:

- a) Sistemas complexos que são difíceis ou impossíveis de modelar;
- b) Sistemas controlados por especialistas (humanos);
- c) Sistemas com entradas e saídas complexas e contínuas;
- d) Sistemas que se utilizam da observação humana como entradas ou como base para regras. Sistemas que são naturalmente "vagos", como os que envolvem ciências sociais e comportamentais, cuja descrição é extremamente complexa.

3. MATERIAL E MÉTODO

Neste capítulo são apresentados os materiais e métodos empregados na descoberta de padrões associados aos TM depressão, ansiedade e estresse, levando em conta os sintomas elencados no questionário DASS-21. Primeiro é apresentada a base de dados na seção 3.1 e, em seguida, o método proposto na seção 3.2.

3.1 BASE DE DADOS

Uma vez que não foram encontradas bases de dados públicas aderentes ao foco da pesquisa, a busca se concentrou na literatura. Com isso, foi localizado o trabalho de Vignola (2013) abordando a adaptação e a validação do DASS-21 no Brasil. Após contato com a autora, a base de dados foi gentilmente cedida por ela para realização desta pesquisa, que é a primeira envolvendo a aplicação de IA nos dados.

A base é composta por dados de 242 indivíduos com idade mínima de 18 e máxima de 75 anos; com maior frequência de mulheres (62%), sendo 154 pacientes casados, 62 solteiros, 15 divorciados e 5 viúvos. Cada um dos 242 registros contempla 24 atributos: estado civil, idade, sexo e as vinte e uma questões do questionário DASS-21. Com relação à severidade dos transtornos, 40% dos indivíduos apresentaram algum tipo de depressão, sendo que 17% deles mostraram estar com níveis graves ou muito graves; 39% apresentaram algum grau de ansiedade, sendo que 22% pertenciam aos graus grave ou muito grave; e 45% apresentaram algum nível de estresse, com 22% em níveis grave ou muito grave (VIGNOLA, 2013). A Tabela 5 apresenta os percentuais de ansiedade, depressão e estresse da amostra considerada na pesquisa de Vignola (2013).

Tabela 5 – Percentuais de Depressão, Ansiedade e Estresse

Níveis DASS	Depressão	Ansiedade	Estresse
MÍNIMO / NORMAL	60%	62%	55%
LEVE	9%	6%	11%
MODERADO	14%	11%	12%
GRAVE	5%	8%	16%
MUITO GRAVE	12%	14%	6%

Fonte: Adaptado de Vignola (2013).

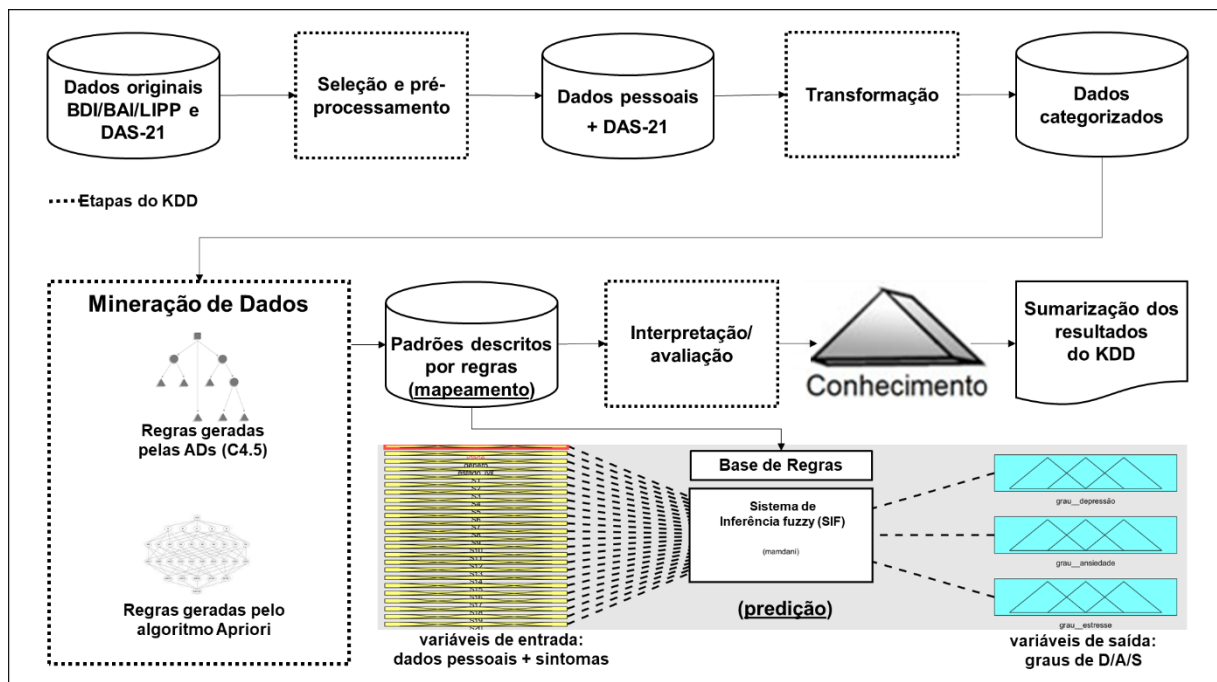
3.2 MÉTODO PROPOSTO

O método proposto neste trabalho, cujo diagrama de funcionamento é ilustrado na Figura 4, baseia-se na descoberta de padrões em dados coletados com o uso do DAS-21 para auxiliar no mapeamento de sintomas e tratamento de depressão, ansiedade e estresse.

Foram realizados experimentos envolvendo as tarefas de classificação, geração de regras de associação e seleção de atributos empregando, respectivamente, AD, algoritmo Apriori e algoritmo de ranking baseado na informação de ganho (Equação 2). Uma vez que o DASS-21 associa 7 sintomas a cada TM, foram criadas três ADs, uma para cada TM (depressão, ansiedade e estresse).

Para realizar os experimentos, o Software WEKA – *Waikato Environment for Knowledge Analysis* (HALL *et al.*, 2009) foi empregado. Trata-se de uma ferramenta que fornece diversos algoritmos de estatística e IA, amplamente utilizada em tarefas de mineração de dados em diversos campos científicos. É válido ressaltar que o algoritmo de AD C4.5, empregado nesta pesquisa, é chamado de J48 no WEKA.

Figura 4 – Diagrama esquemático do método proposto



Fonte: Autoria própria.

A classificação com AD teve como objetivo investigar a existência de padrões nos perfis de pacientes. As regras de associação têm a finalidade de descobrir correlações/associações nos dados, a fim de derivar novos conhecimentos do conjunto de dados analisados.

Já os experimentos envolvendo a seleção de atributos usando a informação de ganho foram realizados com intuito de averiguar a importância dos sintomas na determinação do grau associado a cada TM, o que também pode contribuir no auxílio ao tratamento. Os resultados obtidos nesses experimentos foram comparados com os estudos estatísticos feitos por (VIGNOLA, 2013).

A lógica *fuzzy* foi adotada na segunda etapa do método visando a composição de um sistema de inferência capaz de capturar informações vagas, geralmente descritas por regras em linguagem natural, como as fornecidas por AD, e convertê-las em dados numéricos expressando conceitos ou termos linguísticos (ARAUJO *et. al.*, 2019).

3.2.1 Aplicação das etapas do KDD para mapeamento dos sintomas

Seleção de dados: A base original está armazenada em uma planilha Excel contendo diversos tipos de dados, entre os quais dados pessoais que não identificam o paciente (idade, estado civil e sexo), valores associados aos sintomas de depressão, ansiedade e estresse e mensurações dos graus desses TM de acordo com o DASS-21, os inventários de BECK para Depressão e Ansiedade (CUNHA, 2011) e o Teste LIPP – Inventário de Sintomas de Estresse, para avaliação de Estresse (LIPP, 2005).

O primeiro passo foi compor uma nova planilha Excel contendo apenas os dados selecionados para pesquisa, quais sejam: dados pessoais dos pacientes, as respostas deles para cada um dos 21 sintomas e os graus de depressão, estresse e ansiedade calculados com base no DASS-21 (nulo, mínimo, moderado, grave e muito grave). Em resumo, cada linha da planilha composta para os experimentos contempla 27 atributos de um paciente (3 dados pessoais + 21 valores associados aos sintomas + 3 graus).

Pré-processamento: Nesta etapa foram feitas exclusões dos registros de cinco pacientes em virtude de dados faltantes, restando os dados de 237 indivíduos. Além

disso, foi excluído o atributo profissão, em virtude da falta de padronização nas respostas.

Transformação: Para utilização da base de dados foi necessário categorizar o atributo idade, como segue: “18 a 25 anos”, “26 a 35 anos”, “36 a 45 anos”, “46 a 55 anos” e “acima de 55 anos”. Após isso, os dados foram salvos no em um arquivo no formato CSV (valores separados por vírgula), para posterior importação e análise com o WEKA. Um fragmento do arquivo CSV é ilustrado na Figura 5.

Figura 5 – Fragmento da base de dados em formato CSV

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	
	Idade_cat	Est_civil	Sexo	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	Diag_Dep_DASS21	Diag_Ans_DASS21	Diag_Str_DASS21	
1	50-59	casado	F	1	2	3	0	0	2	0	1	1	2	2	1	2	0	0	0	0	1	0	0	0	0	MODERADO	NORMAL	MINIMO
2	50-59	casado	M	1	1	3	1	2	1	0	2	1	1	2	2	3	1	2	2	2	3	3	2	3	MUITO GRAVE	MUITO GRAVE	MODERADO	
3	18-29	casado	F	2	1	0	0	0	0	2	2	2	1	2	3	3	3	2	3	0	3	1	3	2	MODERADO	MUITO GRAVE	GRAVE	
4	50-59	casado	M	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	NORMAL	NORMAL	NORMAL	
5	18-29	solteiro	M	3	2	0	0	1	3	3	2	2	1	2	2	0	1	1	0	1	2	1	0	0	NORMAL	GRAVE	GRAVE	
6	18-29	solteiro	F	1	2	0	1	1	1	1	1	0	0	0	2	1	2	0	0	0	1	0	0	0	NORMAL	MINIMO	MINIMO	
7	18-29	solteiro	M	1	1	2	1	0	3	2	2	2	1	0	3	2	1	2	3	1	2	0	2	1	MODERADO	MUITO GRAVE	MODERADO	
8	18-29	solteiro	F	1	0	1	0	1	3	1	2	1	1	1	0	2	0	0	0	0	0	0	0	1	MINIMO	NORMAL	NORMAL	
9	18-29	solteiro	F	2	0	2	2	3	2	2	2	3	2	2	2	2	2	2	2	2	3	2	2	3	MUITO GRAVE	MUITO GRAVE	GRAVE	
10	18-29	solteiro	F	3	0	0	2	2	1	1	2	0	3	3	3	3	2	3	3	2	3	0	0	3	MUITO GRAVE	MODERADO	MUITO GRAVE	
11	30-39	casado	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NORMAL	NORMAL	NORMAL
12	30-39	divorciado	M	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	NORMAL	NORMAL	NORMAL
13	18-29	solteiro	F	2	1	3	0	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	3	MUITO GRAVE	MUITO GRAVE	MUITO GRAVE
14	18-29	solteiro	F	3	1	3	1	1	3	1	2	1	3	3	3	3	2	1	3	3	3	1	2	3	MUITO GRAVE	GRAVE	MUITO GRAVE	
15	18-29	solteiro	F	1	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	NORMAL	NORMAL	NORMAL

Fonte: Autoria própria.

Mineração de dados: Nesta etapa foram realizados experimentos com o emprego de AD e o algoritmo Apriori, os quais estão apresentados nas seções 4.1 e 4.2. Também foram realizados experimentos (descritos na seção 4.3) envolvendo a seleção/ranqueamento de atributos com base no ganho de informação, nos quais pode-se verificar a importância de cada sintoma para identificação dos graus associados aos TM.

No caso das AD, utilizou-se o algoritmo C4.5, que representa as relações entre as variáveis no formato de uma árvore, a partir da qual podem ser lidas as regras (cada ramificação da AD descreve uma regra). O algoritmo C4.5 foi aplicado com os seguintes parâmetros: número mínimo de instâncias por nó igual a 2 e fator de

confiança mínima utilizada nas podas igual a 0,25. Os experimentos de classificação foram feitos utilizando todo conjunto de dados, com e sem poda da AD. Como os resultados foram semelhantes, foram escolhidas as árvores usando poda.

As regras de associação foram geradas pelo algoritmo Apriori, com base no fator de confiança (70%), que é uma forma simples de avaliar a precisão de uma regra obtida a partir dos dados de treinamento.

Tanto as regras geradas pelas AD quanto pelo algoritmo Apriori representam padrões (associações e correlações) descobertos a partir da base de dados e são usadas para compor a base de regras do SIF apresentado na seção 4.3.

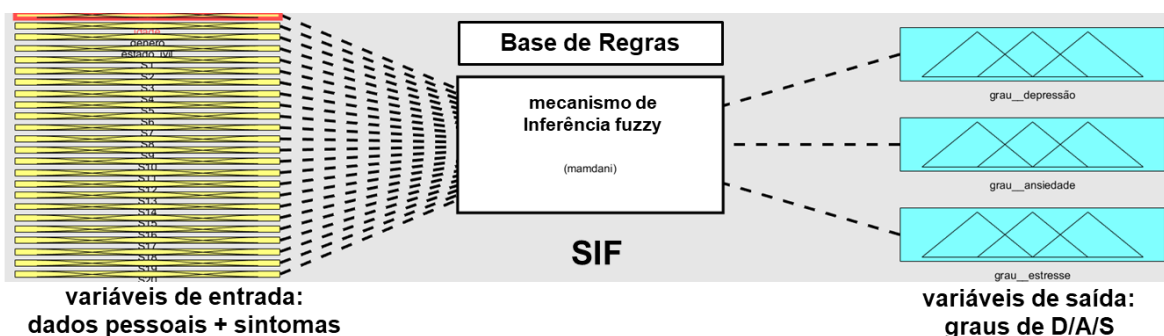
Interpretação/avaliação: A interpretação dos resultados produzidos pelas AD e pelas Regras de Associação foi feita a partir das regras do tipo SE...ENTÃO geradas. Elas foram analisadas e, quando possível agrupadas, produzindo padrões generalizados e fáceis de serem lidos pelos profissionais da saúde. Em adição, os padrões produzidos pelas AD foram avaliados com base na taxa de acerto (sensibilidade) na classificação dos dados, enquanto as regras de associação foram avaliadas com base no fator de confiança (explicado na seção 2.5). Quanto maior a taxa de acertos ou do fator de confiança, maior é a consistência dos padrões descritos pelas regras.

3.2.2 Sistema de Inferência Fuzzy (SIF)

O SIF proposto neste trabalho, cuja arquitetura é ilustrada na Figura 6, foi desenvolvido com o auxílio da *toolbox fuzzy* do ambiente MatLab¹, e sua base de regras foi composta pelas regras geradas pelas ADs.

¹ <https://www.mathworks.com/products/matlab.html>

Figura 6 – Variáveis de entrada/saída do mecanismo de inferência fuzzy

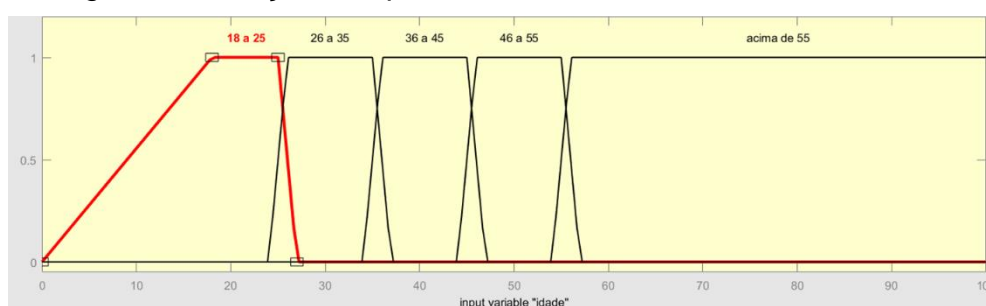


Fonte: Autoria própria.

O SIF foi projetado com 24 variáveis de entrada: dados pessoais do paciente (idade, gênero e estado civil) + 21 sintomas do DASS-21 e 3 variáveis de saída: grau_depressão, grau_estresse e grau_ansiedade. Em outras palavras, o objetivo do SIF é prever os graus de depressão, estresse e ansiedade a partir dos dados de entrada.

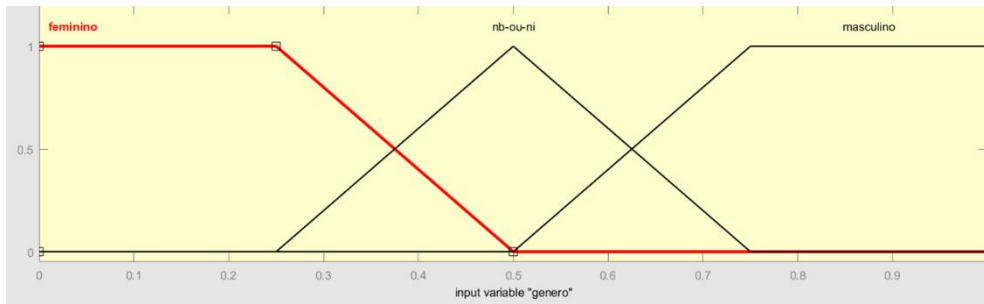
Os 21 sintomas foram igualmente modelados com as seguintes funções de pertinência: nulo, baixo, médio e alto, correspondentes aos valores de 0 a 3 determinados pelo DASS-21. Analogamente, as 3 variáveis de saída foram modeladas com as funções de pertinência: normal (0); mínimo (1); moderado (2); grave (3) e muito grave (4), correspondentes aos graus de depressão, ansiedade e estresse também determinados pelo DASS-21. As figuras 7 a 11 ilustram, respectivamente, as funções de pertinência associadas às variáveis idade, gênero, estado civil, sintomas (S1 a S21) e ao grau dos TM (grau_depressão, grau_ansiedade e grau_estresse).

Figura 7 – Funções de pertinência associadas à variável idade



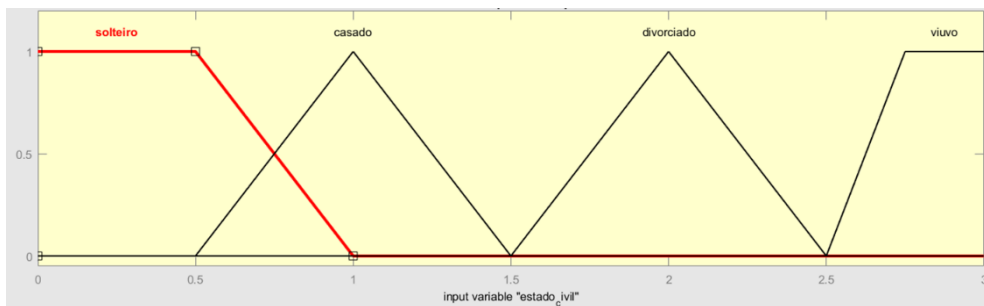
Fonte: Autoria própria.

Figura 8 – Funções de pertinência associadas à variável gênero



Fonte: Autoria própria.

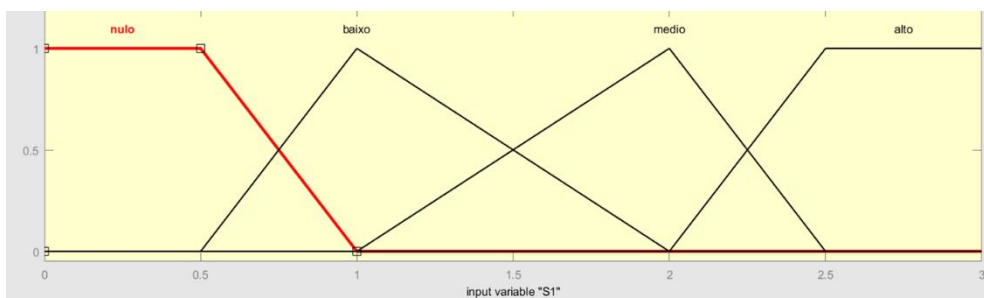
Figura 9 – Funções de pertinência associadas à variável estado civil



Fonte: Autoria própria.

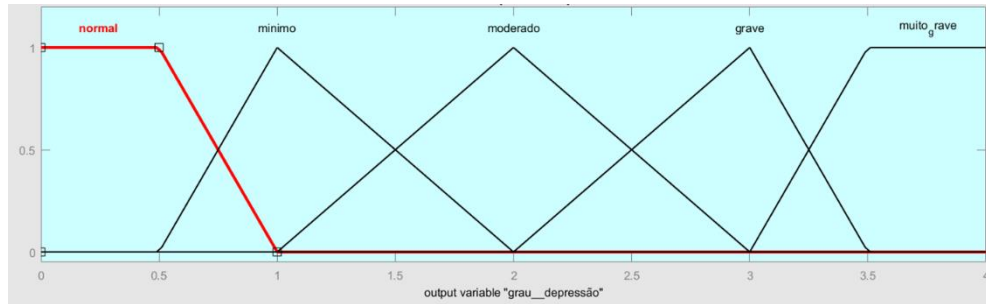
As variáveis idade e estado civil contemplam funções de pertinência correspondentes às categorias dos dados originais ou transformados, conforme descrito nas seções 3.1 e 3.2, ou seja: idade – “18 a 25 anos”, “26 a 35 anos”, “36 a 45 anos”, “46 a 55 anos” e “acima de 55 anos”; estado civil – solteiro, casado, divorciado e viúvo. Já a variável gênero é uma adaptação da variável original sexo, para representar feminino, não-binário e masculino. O fato dessas variáveis não serem naturalmente nebulosas não invalida a aplicabilidade do SIF.

Figura 10 – Funções de pertinência associadas aos sintomas (S1 a S21)



Fonte: Autoria própria.

Figura 11 – Funções de pertinência associadas às variáveis de saída



Fonte: Autoria própria.

A base de regras do SIF contempla as 81 regras obtidas a partir das 3 ADs geradas pelo algoritmo C4.5 para descrever os padrões associados à depressão, ansiedade e estresse por meio de um valor numérico de 0 a 4, como indicado anteriormente. São regras que possuem um encadeamento **SE... ENTÃO...** como no exemplo a seguir: **SE** S16 = Baixo **E** S17= Médio **ENTÃO** grau de depressão = **Moderado**. Por fim, para o processo de defuzzyficação o método empregado foi o de Mamdani. A avaliação do SIF foi feita aplicando-o para predizer os graus dos TM para os indivíduos representados nos dados coletados e comparando as predições com os resultados reais.

4. RESULTADOS E DISCUSSÃO

Este capítulo contempla os resultados da descoberta de padrões por meio das AD (seção 4.1) e pelo algoritmo Apriori (seção 4.2), os resultados alcançados na seleção/ranqueamento dos sintomas (seção 4.3) e por fim, os resultados obtidos com o Sistema de Inferência Fuzzy (SIF) proposto.

4.1 DESCOBERTA DE PADRÕES POR MEIO DE ÁRVORES DE DECISÃO (CLASSIFICAÇÃO)

Nas subseções 4.1.1 a 4.1.3 são apresentadas as análises feitas para cada fator motivacional por meio de AD gerada pelo algoritmo C4.5 (J48 no WEKA), mostrando o seu desempenho com base no número de instâncias classificadas corretamente (taxa de acerto). Foram criadas três AD, apresentadas nos Apêndices I a III, sendo uma para cada TM, quais sejam: depressão, ansiedade e estresse

4.1.1 Depressão

A AD mostrada no Apêndice I, correspondente ao TM depressão, classificou corretamente 228 instâncias obtendo uma taxa de acerto de 94,2149%. Foram 14 instâncias classificadas incorretamente com a taxa de erro de 5,7851%. A partir da matriz de confusão (Tabela 6) pode-se observar o seguinte: para os pacientes que não apresentaram depressão (normal) foram 144 acertos e 2 erros; para os que apresentaram um grau mínimo foram 17 acertos e 4 erros; nos casos moderados houve 32 acertos e 3 erros; para os casos graves de depressão foram 8 acertos e 4 erros; nos casos muito graves foram 27 acertos e apenas 1 erro.

Tabela 6 – Matriz de confusão obtida na classificação de dados relativos ao TM “Depressão”

Classe real					Classe predita
a	b	c	d	E	
144	2	0	0	0	a = Normal
3	17	1	0	0	b = Mínimo
0	3	32	0	0	c = Moderado
0	1	2	8	1	d = Grave
0	0	0	1	27	e = Muito Grave

Fonte: Autoria própria com base na matriz de confusão gerada pelo software Weka

A partir da AD pode-se ainda derivar os seguintes conhecimentos:

- Pacientes que conseguem se entusiasmar com facilidade, sentem que têm algo a desejar e sentem-se animados(as) foram classificados(as) sem nenhum grau de depressão (precisão da regra = 97,7942%, 133 de 136 casos possíveis);
- Pacientes que quase sempre estão entusiasmados(as), que enxergam seu valor como pessoa, sentem que a vida faz sentido, sentem que têm algo a desejar, mas que em geral se sentem deprimidos(as) e sem ânimo foram classificados(as) com grau mínimo de depressão (precisão da regra = 73,6843%, 14 de 19 casos possíveis);
- Pacientes que quase sempre estão entusiasmados(as), mas que às vezes sentem-se deprimidos(as) e sem ânimo, com dificuldades em desejar algo e com dificuldade de enxergar seu valor como pessoa, foram classificados(as) com grau moderado de depressão (precisão da regra = 100%, 10 de 10 casos possíveis);
- Pacientes com alguma facilidade de se entusiasmar, mas que na maioria das vezes não tem desejos, em geral não enxergam seu valor como pessoa e quase sempre estão deprimidos(as) e sem ânimo, foram classificados(as) com grau grave de depressão (precisão da regra = 100%, 2 de 2 casos possíveis);;

- e) Pacientes que em geral não se entusiasmam com nada e sentem que não tem nada a desejar foram classificados(as) com grau muito grave de depressão (precisão da regra = 100%, 17 de 17 casos possíveis).

4.1.2 Ansiedade

A AD mostrada no Apêndice II, correspondente ao TM ansiedade, classificou corretamente 225 instâncias obtendo uma taxa de acerto de 92.9752%. Foram 17 instâncias classificadas incorretamente com a taxa de erro de 7.0248%. A partir da matriz de confusão (Tabela 7) pode-se observar o seguinte: para os pacientes que não apresentaram depressão (normal) foram 145 acertos e 4 erros; para os que apresentaram um grau mínimo foram 11 acertos e 3 erros; nos casos moderados houve 22 acertos e 4 erros; para os casos graves de depressão foram 14 acertos e 5 erros; já nos casos muito graves foram 33 acertos e apenas 1 erro.

Tabela 7 – Matriz de confusão obtida na classificação de dados relativos ao TM “Ansiedade”

Classe real					Classe predita
a	b	c	d	E	
145	2	2	0	0	a = Normal
1	11	2	0	0	b = Mínimo
1	1	22	1	1	c = Moderado
0	2	3	14	0	d = Grave
0	0	1	0	33	e = Muito Grave

Fonte: Autoria própria com base na matriz de confusão gerada pelo software Weka.

A partir da AD pode-se ainda derivar os seguintes conhecimentos:

- a) Pacientes não sentem o coração alterado na ausência de esforço físico, em geral não se preocupam com a possibilidade de entrar em pânico e passar vergonha e não sentem a boca seca, foram classificados(as) sem ansiedade (precisão da regra = 98.5186%, 133 de 135 casos possíveis);

- b) Pacientes que em geral não sentem o coração alterado na ausência de esforço físico, mas que têm alguma preocupação com a possibilidade de entrar em pânico e às vezes apresentam dificuldade de respirar mesmo na ausência de esforço físico (sentem-se ofegantes), foram classificados(as) com grau mínimo de ansiedade (precisão da regra = 76.2516%, 5 de 7 casos possíveis);
- c) Pacientes que em geral sentem o coração alterado na ausência de esforço físico e tremores no corpo em algumas situações, foram classificados(as) com grau moderado de ansiedade (precisão da regra = 75%, 6 de 8 casos possíveis);
- d) Pacientes que algumas vezes sentem o coração alterado na ausência de esforço físico, que às vezes se preocupam com a possibilidade de entrar em pânico e passar vergonha, e apresentam dificuldade de respirar mesmo na ausência de esforço físico (sentem-se ofegantes), foram classificados(as) com grau grave de ansiedade (precisão da regra = 100%, 3 de 3 casos possíveis);
- e) Pacientes que sentem o coração alterado na ausência de esforço físico, que se preocupam com a possibilidade de entrar em pânico e passar vergonha e sentem medo sem motivo algum, foram classificados(as) com grau muito grave de ansiedade (precisão da regra = 100%, 21 de 21 casos possíveis).

4.1.3 Estresse

E por fim, a AD corresponde ao TM estresse (Apêndice III) classificou corretamente 225 instâncias obtendo uma taxa de acerto de 92.9752%. Foram 17 instâncias classificadas incorretamente com a taxa de erro de 7.0248%. A partir da matriz de confusão (Tabela 8) pode-se observar o seguinte: para os pacientes que não apresentaram depressão (normal) foram 132 acertos e 2 erros; para os que apresentaram um grau mínimo foram 22 acertos e 5 erros; nos casos moderados houve 27 acertos e 2 erros; para os casos graves de depressão foram 23 acertos e 6 erros; para os casos muito graves foram 21 acertos e 2 erros.

Tabela 8 – Matriz de confusão obtida na classificação de dados relativos ao TM “Estresse”

Classe real					Classe
a	b	c	d	E	Predita
132	2	0	0	0	a = Normal
3	22	2	0	0	b = Mínimo
0	2	27	0	0	c = Moderado
1	0	3	23	2	d = Grave
0	0	0	2	21	e = Muito Grave

Fonte: Autoria própria com base na matriz de confusão gerada pelo software Weka

A partir da AD pode-se ainda derivar os seguintes conhecimentos:

- a) Pacientes que dificilmente estão nervosos(as), que têm facilidade de se acalmar, na maioria do tempo são tolerantes com coisas que os(as) impedem de continuar o que estavam fazendo e que na maior parte do tempo não são agitados, foram classificados(as) sem estresse (precisão da regra = 97,5%, 117 de 120 casos possíveis);
- b) Pacientes que às vezes estão nervosos(as), se sentem agitados, reagem de forma exagerada às situações e não conseguem se acalmar foram classificados(as) com grau mínimo de estresse (precisão da regra = 88.8889%, 8 de 9 casos possíveis);
- c) Pacientes que às vezes estão nervosos(as) e não conseguem se acalmar, tendem a reagir de forma exagerada às situações, na maioria das vezes estão emotivos(as) ou sensíveis demais e não conseguem relaxar, foram classificados(as) com grau moderado de estresse (precisão da regra = 80%, 8 de 10 casos possíveis);
- d) Pacientes que na maior parte do tempo estão nervosos(as) e agitados, tendem a reagir de forma exagerada às situações e na maioria das vezes não conseguem se acalmar, foram classificados(as) com grau grave de estresse (precisão da regra = 83.3333%, 10 de 12 casos possíveis);

- e) Pacientes que geralmente estão nervosos(as) e agitados e que na maioria do tempo são intolerantes com coisas que os(as) impedem de continuar o que estavam fazendo, foram classificados(as) com grau muito grave de estresse (precisão da regra = 90.4762%, 19 de 21 casos possíveis).

4.2 DESCOBERTA DE PADRÕES POR MEIO DO ALGORITMO APRIORI (REGRAS DE ASSOCIAÇÃO)

Nesses experimentos, para cada um dos TM elencados a seguir sintetizou-se o conjunto de regras com grau de confiabilidade de no mínimo 70%.

Depressão:

- 76% dos pacientes homens casados não têm depressão;
- 72% dos pacientes com faixa etária de 30 a 39 anos não têm depressão e 71% são casados.

Ansiedade:

- 82% dos homens casados não tem ansiedade;
- 79% dos homens não têm ansiedade;
- 74% dos pacientes com faixa etária entre 50 e 59 anos e que estão casados não tem ansiedade;
- 71% dos pacientes com faixa etária entre 50 e 59 não tem ansiedade;
- 70% dos pacientes com faixa etária entre 30 e 39 anos não têm ansiedade.

Estresse:

- 70% dos homens casados não têm estresse.

Com o número de pacientes sem TM serem maior, as regras de maior confiabilidade são para aqueles que não têm esses três transtornos mentais. Também se conclui que as regras criadas pelo algoritmo Apriori são complementares às das AD.

4.3 ANÁLISE DA IMPORTÂNCIA DOS SINTOMAS NA DETERMINAÇÃO DOS GRAUS DOS TM (SELEÇÃO DE ATRIBUTOS)

Nesses experimentos o objetivo foi verificar a importância dos sintomas dos TM e comparar com os resultados da análise fatorial obtidos no estudo de Vignola (2013). Para tanto, empregou-se o ganho de informação (Equação 2) de cada um dos 21 sintomas para cada TM. Os resultados enfatizando os 7 sintomas de maior importância (peso) para cada TM estão apresentados na Tabela 9.

Tabela 9 – Classificação dos atributos dos transtornos do DASS-21

Depressão		Ansiedade		Estresse	
<i>Sintoma</i>	<i>Ganho</i>	<i>Sintoma</i>	<i>Ganho de informação</i>	<i>Sintoma</i>	<i>Ganho de informação</i>
S16	0.827	S15	0.611	S8	0.635
S13	0.744	S9	0.54	S11	0.617
S10	0.732	S20	0.525	S14	0.611
S21	0.625	S19	0.468	S1	0.609
S17	0.615	S18	0.425	S12	0.586
S18	0.508	S16	0.368	S18	0.578
S3	0.5	S4	0.361	S6	0.542
S14	0.406	S21	0.36	S16	0.54
S8	0.4	S8	0.353	S13	0.464
S5	0.392	S13	0.346	S10	0.42
S15	0.359	S7	0.338	S17	0.415
S11	0.342	S17	0.309	S15	0.381
S12	0.338	S12	0.308	S21	0.336
S1	0.326	S1	0.298	S3	0.314
S20	0.308	S14	0.296	S9	0.285
S19	0.295	S11	0.294	S5	0.266
S4	0.262	S2	0.274	S20	0.261
S9	0.26	S10	0.272	S7	0.238
S6	0.252	S3	0.234	S19	0.237
S7	0.182	S6	0.229	S2	0.223
S2	0.139	S5	0.227	S4	0.2

Fonte: Autoria própria.

Nesta tabela os sintomas são mostrados de acordo com a importância deles no diagnóstico de cada transtorno, e ao lado são apontados os seus pesos. A comparação entre a carga fatorial empregada em Vignola (2013) e o ganho de informação apresentou algumas diferenças que são indicadas na Tabela 10.

Tabela 10 – Carga fatorial vs. ganho de informação dos sintomas para cada TM

	Sintomas	Carga Fatorial	Sintomas	Ganho
Depressão	S16	0,747	S16	0.827
	S17	0,743	S13	0,744
	S10	0,728	S10	0.732
	S13	0,707	S21	0.625
	S21	0,657	S17	0.615
	S3	0,622	S18	0.508
	S5	0,494	S3	0.5
Ansiedade	S20	0,730	S15	0.611
	S9	0,688	S9	0.54
	S19	0,605	S20	0.525
	S2	0,586	S19	0.468
	S15	0,585	S18	0.425
	S7	0,585	S16	0.368
	S4	0,560	S4	0.361
Estresse	S18	0,497	S8	0.635
	S6	0,749	S11	0.617
	S8	0,675	S14	0.611
	S12	0,671	S1	0.609
	S11	0,665	S12	0.586
	S1	0,657	S18	0.578
	S14	0,653	S6	0.542

Conforme pode ser visto na Tabela 9, a maioria dos sintomas se mantiveram relacionados aos mesmos TM independentemente da medida utilizada (carga fatorial ou ganho de informação).

No TM depressão os atributos S16 e S10 permaneceram na mesma posição, enquanto S17, S13, S21, S3 trocaram de posição de acordo com a medida utilizada. Contudo, houve uma troca do sintoma do S5 pelo S18 quando de ganho de informação foi empregado.

No caso do TM ansiedade, somente os sintomas S9 e S4 se mantiveram nas mesmas posições, enquanto os sintomas S20, S19 e S15 ocuparam diferentes posições. Neste TM houve duas trocas: S2 e S7 foram substituídos pelos atributos S18 e S16. E no TM estresse todos os atributos mudaram de posição, mas não houve substituições.

Essas diferenças nas classificações de atributos podem ter como causas a forma de cálculo empregada para associação dos sintomas aos três TM. Não obstante, de acordo com Vignola (2013) embora os vinte e um sintomas estejam divididos em três grupos (depressão, ansiedade e estresse), eles não nem sempre estão associados a apenas um transtorno. Eles podem se referir a dois ou até mesmo aos três transtornos com “pesos” diferentes. Neste sentido, investigações mais aprofundadas devem ser realizadas para que esta hipótese possa ser mitigada.

4.4 SISTEMA DE INFERÊNCIA FUZZY (SIF)

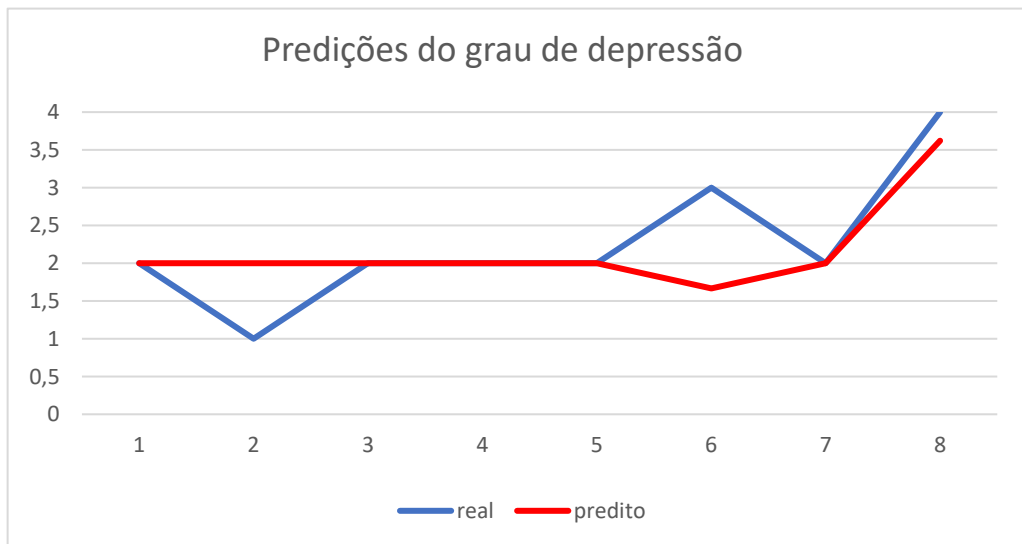
A avaliação do SIF pode ser feita comparando-se os graus dos TM preditos por ele (linha vermelha) com os resultados reais (linha azul). Em virtude da quantidade de dados disponíveis, foram escolhidos aleatoriamente alguns casos da base que foram submetidos aos testes de predição. Na Figura 12 é ilustrada a tela do Matlab por meio da qual as predições são feitas, enquanto nas figuras 13 a 15 são ilustrados resultados do SIF para algumas predições dos graus de depressão, ansiedade e estresse.

Figura 12 – Tela do Matlab para predições do SIF



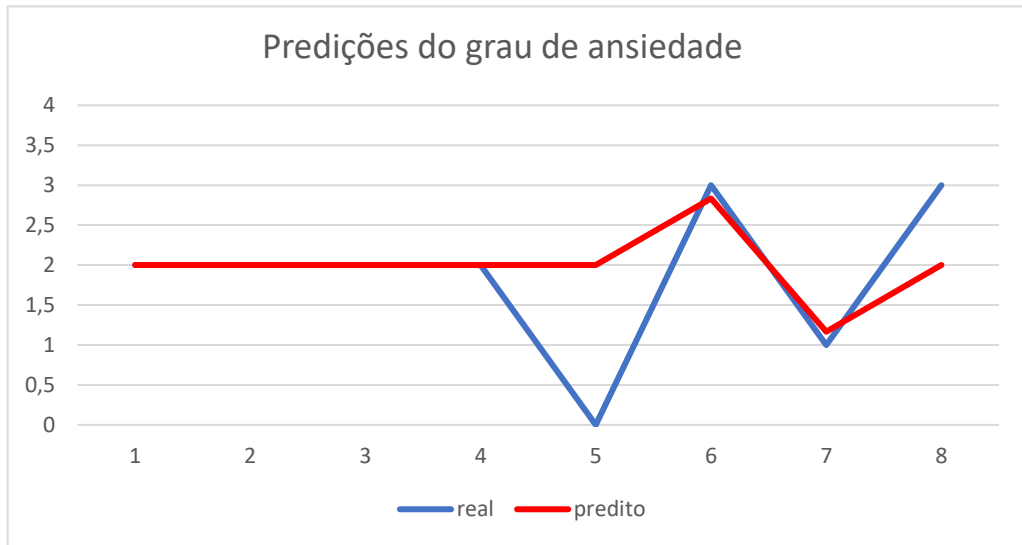
Fonte: Autoria Própria.

Figura 13 – Algumas predições do SIF para depressão



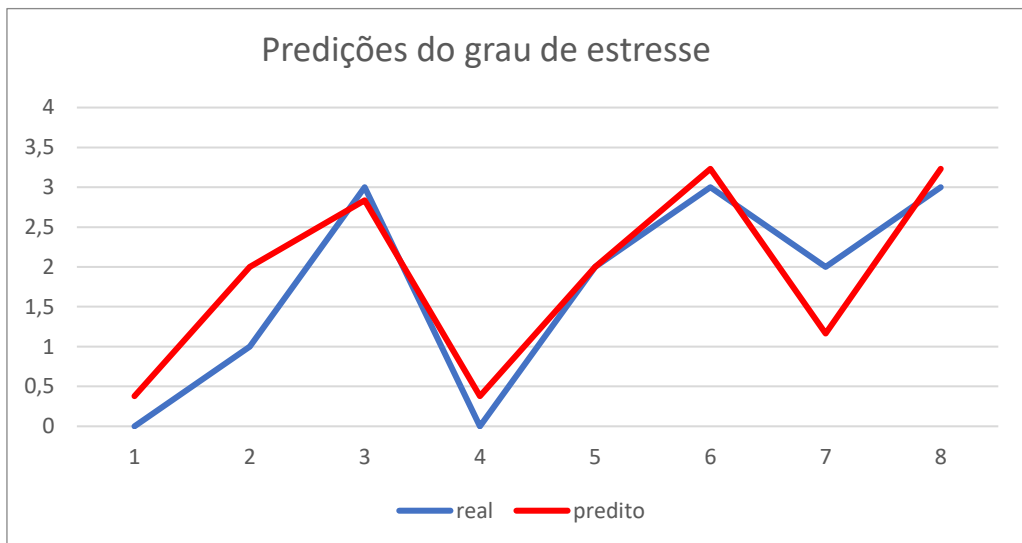
Fonte: Autoria Própria.

Figura 14 – Algumas predições do SIF para ansiedade



Fonte: Autoria Própria.

Figura 15 – Algumas predições do SIF para estresse



Fonte: Autoria Própria.

Como pode ser observado nos exemplos ilustrados nas figuras 13 a 15, os melhores resultados foram obtidos nas predições dos graus de estresse, seguidos dos resultados para as predições dos graus de ansiedade. Isso pode estar relacionado ao número de regras relativas à cada TM, quais sejam: 19, 26 e 29 para depressão, estresse e ansiedade, respectivamente. Em outras palavras, o que se pode observar é que quanto maior o número de regras relativas à um TM, melhor foi resultado obtido

nas predições para esse TM. Neste caso, o fato do algoritmo C4.5 ter sido parametrizado para usar podas nas ADs gerando menores quantidades de regras pode ter impactado negativamente o desempenho do SIF.

Outro ponto que merece destaque é que no caso da AD gerada para depressão os dados pessoais não foram levados em conta, provavelmente por terem sido considerados pouco relevantes para definição dos graus dessa TM. Isso também pode ter sido causado pela parametrização do algoritmo que forçou a poda da AD.

Apesar dos pontos aqui levantados, de um modo geral pode-se dizer que o SIF obteve um desempenho satisfatório. Contudo, cabe uma análise mais aprofundada visando torná-lo ainda mais efetivo.

4.5 DISCUSSÃO DOS RESULTADOS

Com relação aos experimentos envolvendo MD (seções 4.1 e 4.2), tanto as regras geradas pelas ADs quanto pelo algoritmo Apriori, a partir dos dados coletados com o uso do DAS-21, permitiram identificar padrões e correlações que levam a insights significativos, permitindo um mapeamento de como os sintomas impactam nos graus de depressão, ansiedade e estresse, para diferentes grupos de pacientes. Por exemplo: **SE S16 = Baixo E S17= Médio ENTÃO grau de depressão = Moderado**; os resultados obtidos revelaram que a presença de determinados sintomas, combinada com certas características pessoais, é um forte indicativo de depressão. Uma das principais vantagens das ADs é a sua interpretabilidade. As regras geradas por uma AD podem ser facilmente compreendidas e interpretadas pelos profissionais de saúde, contribuindo nos processos de tratamento e diagnóstico.

Destaca-se também a importância dos experimentos relativos à seleção de atributos ou sintomas (seção 4.3). Além de corroborar o estudo de Vignola (2013), por meio da análise da importância relativa dos diferentes atributos, foi possível identificar os sintomas mais relevantes e informativos para o mapeamento dos TM considerados, o que pode contribuir para otimização do processo de triagem e diagnóstico, economizando tempo e recursos da equipe multiprofissional de saúde.

Os resultados apresentados nas seções 4.1 a 4.3 corroboram o fato que a MD tem se mostrado uma ferramenta valiosa na área da saúde, permitindo analisar

grandes volumes de informações e descobrir insights significativos, como também demonstrado nos trabalhos de Fernando *et al.* (2022) e Khanna *et al.* (2023). No contexto do mapeamento de sintomas de depressão, ansiedade e estresse, a MD mostrou ser um importante instrumento na identificação e compreensão desses TM.

Como foi evidenciado na presente pesquisa, um dos principais benefícios da MD no tratamento de TM é a capacidade de identificar padrões e correlações em dados clínicos. Ao analisar dados pessoais, históricos médicos, resultados de exames e outros registros clínicos, algoritmos de MD podem descobrir associações entre variáveis, permitindo uma melhor compreensão dos transtornos mentais e aprimorando a precisão dos diagnósticos.

Além disso, a MD pode ser útil na identificação de subgrupos de pacientes com características semelhantes. Isso pode ajudar os profissionais de saúde na personalização do tratamento com base em características específicas dos pacientes, levando em consideração fatores como idade, sexo, estado civil, sintomas, histórico médico, resposta a terapias anteriores e outros dados relevantes. A abordagem personalizada pode levar a intervenções mais eficazes e resultados de tratamento melhores. Não obstante, a MD também pode ser aplicada no monitoramento contínuo do progresso do tratamento. Ao analisar dados coletados ao longo do tempo, como respostas a questionários, resultados de testes psicológicos ou registros de sintomas, algoritmos de MD poderiam ser utilizados para identificar tendências e padrões que indicam a eficácia ou ineficácia de um determinado tratamento. Isso permite aos profissionais de saúde fazerem ajustes nas terapias e abordagens, oferecendo um cuidado mais personalizado e adaptado às necessidades individuais dos pacientes.

Com relação ao SIF proposto, um exemplo de aplicação é na triagem de pacientes. Os dados pessoais e os sintomas relatados pelos pacientes podem ser avaliados e associados a conjuntos fuzzy que representam diferentes graus de pertinência a cada TM. Com base em regras fuzzy, originadas das regras geradas pelas ADs, o SIF é capaz de determinar a probabilidade de cada transtorno com base nos dados e sintomas relatados. Ressalta-se que, diferentemente dos métodos de aprendizagem de máquina propostos na literatura para auxílio ao diagnóstico de depressão, ansiedade e estresse com o uso do DAS-21, como os de Nizam *et al.* (2020), Altintas *et al.* (2021), Elhai *et al.* (2021), Sangani *et al.* (2021) e Wesabi *et al.*

(2022), o SIF proposto leva em conta dados pessoais do paciente permitindo uma avaliação mais abrangente e precisa dos sintomas, considerando a variação e a subjetividade frequentemente presentes tanto nos sintomas quanto nos TM.

Além disso, o SIF pode ser aplicado para personalizar as intervenções terapêuticas, com base nas características individuais de cada paciente. Por exemplo, nos Centros de Atenção Psicossocial (CAPS) – que oferecem atendimento a pessoas com transtornos mentais e que nem sempre possuem quantidade de profissionais de saúde proporcional às necessidades do atendimento da população, um SIF como o proposto neste trabalho poderia ser útil para: i) auxiliar na avaliação do estado emocional de um paciente visando encaminhá-lo para o profissional adequado dentro do CAPS; ii) identificar indícios de risco e alertar os profissionais de saúde para que intervenções precoces sejam realizadas; ou iii) detectar precocemente sinais de agravamento das condições mentais. Esse tipo de ferramenta pode ser especialmente útil para casos de pacientes que têm dificuldade em expressar seus sentimentos ou que apresentam comportamentos que indicam um possível agravamento do quadro.

Este trabalho também contribui demonstrando como abordagens baseadas em IA podem desempenhar um papel relevante no tratamento de transtornos mentais, fornecendo insights valiosos, identificando padrões e ajudando na personalização do cuidado, como indicado nos trabalhos realizados por Fulmer *et al.* (2018), Oliveira *et al.* (2020), Le Glaz *et al.* (2021) e Paton e Tiffin (2022). Com a utilização adequada, tais abordagens podem contribuir para uma melhor compreensão dos TM, auxiliar no diagnóstico e proporcionar tratamentos mais eficazes aos pacientes.

Contudo, é importante destacar que a aplicação de IA no contexto da saúde levanta questões éticas e de privacidade. A confidencialidade e proteção dos dados dos pacientes são de extrema importância, e medidas adequadas de segurança devem ser implementadas para garantir a privacidade e anonimato das informações pessoais. Além disso, como pontuado em Le Glaz *et al.* (2021) e Paton e Tiffin (2022), é fundamental que a interpretação dos resultados produzidos por abordagens de IA seja sempre realizada por profissionais qualificados, que levem em consideração o contexto clínico e as particularidades de cada paciente. Também é fundamental que os profissionais de saúde estejam envolvidos no desenvolvimento e supervisão dessas tecnologias para assegurar que elas sejam eficazes e seguras.

5. CONCLUSÕES E TRABALHOS FUTUROS

Nesta pesquisa foi proposto um método de IA combinando mineração de dados (MD) e lógica fuzzy para mapeamento de sintomas e auxílio ao tratamento de depressão, ansiedade e estresse.

Os experimentos com o método, acompanhados pela psicóloga que cedeu os dados para esta pesquisa, evidenciaram a importância das técnicas de MD na identificação de padrões e correlações que levaram a insights significativos, permitindo um mapeamento de como os sintomas impactam nos graus de depressão, ansiedade e estresse, para diferentes grupos de pacientes. Já o SIF, baseado em regras fuzzy geradas pelas ADs, é capaz de determinar o grau de cada transtorno com base nos dados do paciente e nos sintomas relatados. O fato do SIF proposto levar em conta os dados pessoais do paciente, diferentemente dos trabalhos apresentados na literatura que consideram apenas os sintomas, faz com que ele seja mais robusto às variações e subjetividades frequentemente presentes tanto nos sintomas quanto nos TM.

O método de IA proposto nesta pesquisa pode auxiliar na rápida predição de sintomas de depressão, ansiedade e estresse, tanto em consultas clínicas quanto não clínicas, podendo ser útil para profissionais de saúde especialistas ou não em saúde mental, no contexto de triagem ambulatorial, pronto atendimento, bem como em entrevista abrangente cujas queixas não sejam claras, visando melhor associação de sintomas, propostas terapêuticas e até mesmo na investigação de outras doenças orgânicas, não associadas à saúde mental, otimizando e facilitando possíveis diagnósticos diferenciais. Contudo, para que o método possa ser aplicado na prática ele ainda precisa ser incorporado em um software com uma interface adequada aos usuários da área de saúde.

Ressalta-se que ferramentas de IA para auxiliar no diagnóstico e tratamento da depressão, ansiedade e estresse podem contribuir com os Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU), especialmente no que tange a ODS 3 (Saúde e Bem-Estar), uma vez que visam fornecer suporte contínuo e personalizado para indivíduos que enfrentam problemas de saúde mental, facilitando o acesso a recursos terapêuticos e identificando sinais

precoces de problemas. Obviamente, é imprescindível que a interpretação dos resultados produzidos por tais ferramentas seja realizada por profissionais de saúde qualificados, que levem em consideração o contexto clínico e as particularidades de cada paciente.

Em trabalhos futuros pretende-se ampliar as análises no âmbito da MD, principalmente para gerar ADs mais detalhadas (sem poda) com o objetivo de robustecer o SIF. Também neste sentido, uma alternativa é empregar na composição da base do SIF as regras geradas pelo algoritmo Apriori. Além disso, pretende produzir um software com uma interface apropriada aos profissionais de saúde e testar a sua utilização no campo prático, visando principalmente a otimização dos processos de triagem e tratamento das TM consideradas nesta pesquisa.

REFERÊNCIAS

ABREU, R. F. de. Inteligência Artificial e Doenças Mentais. **RECISATEC – Revista Científica Saúde e Tecnologia** - ISSN 2763-8405, v. 1, n. 5, p. 1538, 2021.

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. *In: Proc. 20th int. conf. very large data bases, VLDB*, vol. 1215, p. 487-499, 1994.

ALTINTAŞ, E.; AKSU, Z. U.; DEMIR, Z. G. Machine learning techniques for anxiety disorder. n. 31, p. 365-374. **Avrupa Bilim ve Teknoloji Dergisi**, 2021.

ARAUJO, A. C. de; SANTOS, J. A. N. dos; LIMA, G. B. A.; NICHIOKA, J.; GAVIÃO, L. O. Avaliação de segurança em empresas da construção civil: uma aplicação da lógica Fuzzy. **Revista de Gestão dos Países de Língua Portuguesa**, Rio de Janeiro, RJ, v. 18, n. 1, p. 33-56, 2019.

AVENI, A. Tendências da Inovação na Economia da Saúde: Um Possível Cenário Para Empreendedores. **Revista Coleta Científica**. v. 4, n. 7, p. 71-87, 2020.

BALDISSERA, O. Promoção da saúde mental: um dos maiores desafios do século 21. Website **PÓS PUCPR DIGITAL**, 15 mar. 2021. Disponível em: <https://posdigital.pucpr.br/blog/principais-problemas-de-saude-mental>, Acesso em: 03/09/2023.

BHAT, P.; MALAGANVE, P.; HEGDE, P. A new framework for social media content mining and knowledge discovery. *In: International Journal of Computer Applications*. 182(36), 17-20, 2019.

BHAVE, D. P.; TEO, L. H.; DALAL, R. S. Privacy at work: A review and a research agenda for a contested terrain. *In: Journal of Management*. 46(1), p. 127– 164, 2020.

BRAMER, M. Principles of data mining. **Springer**, London, 2007.

BERNERT, R. A.; HILBERG, A. M.; MELIA, R.; KIM, J. P.; SHAH, N. H.; ABNOUSI, F. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. *In: International journal of environmental research and public health*. 17(16), 5929. 2020.

BUDIYANTO, S.; SIHOMBING, H.C.; RAHAYU, F. I. M. Depression and anxiety detection through the Closed-Loop method using DASS-21. *In: Telkomnika Journal*, Vol.17, No.4, p.2087-2097. 2019.

CABRAL, J. C. C.; VELEDA, G. W.; MAZZOLENI, M.; COLARES, E. P.; NEIVA SILVA, L.; NEVES, V. T. DAS. Stress and Cognitive Reserve as independent factors of neuropsychological performance in healthy elderly. **Cien Saude Colet.** 21 (11): p. 3499–3508, 2016.

CHOUDHURY, A. A.; KHAN, M. R. H.; NAHIM, N. Z.; TULON, S. R.; ISLAM, S.; CHAKRABARTY, A. Predicting depression in Bangladeshi undergraduates using machine learning. **IEEE Region 10 Symposium (TENSYMP)**. p. 789- 794. IEEE, 2019.

DE ARAÚJO, S. A.; DE BARROS, D. F.; DA SILVA, E. M.; CARDOSO, M. V. Applying computational intelligence techniques to improve the decision making of business game players. v. 23, n. 18, p. 8753-8763, **Soft Computing**. 2019.

DOMINGUES, A. M.; DELGADO, S.; BISSACO, M. A. S.; ARAÚJO, S. A. Computational Intelligence in Serious Games: A Case Study to Identify Patterns in a Game for Children with Learning Disabilities. *In: International Journal of Computer Applications*, v. 184, p. 40-44. São Paulo, 2022.

ELHAI, J. D.; YANG, H.; MCKAY, D.; ASMUNDSON, G. J.; MONTAG, C. Modeling anxiety and fear of COVID-19 using machine learning in a sample of Chinese adults: associations with psychopathology, sociodemographic, and exposure variables. **Anxiety, Stress, & Coping**, 34(2), p. 130-144. 2021.

EXCELÊNCIA EM ALIMENTAÇÃO (EXAL). Conheça as principais doenças do século 21 e como evitá-las. **Website da Exal**. 2019. Disponível em: <https://exal.com.br/conheca-as-principais-doencas-do-seculo-21-e-como-evita-las/#>. Acesso em: 03/09/2023.

FÁTIMA, A., LI, Y.; HILLS, TT; STELLA, M. Dasentimental: Detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning. **Big Data and Cognitive Computing**, v. 5, n. 4, p. 77, 2021.

FAYYAD, U; PIATETSKY-SHAPIRO G; SMYTH, P; UTHURUSAMY, R. Advances in knowledge discovery and data mining. American Association for Artificial Intelligence. **Menlo Park: MIT Press**, 1996.

FERNANDO, J. G.; JAIN, S. K. Predictive Data Mining in Clinical Medicine: Current Issues and Guidelines. *Technoarete Transactions on Intelligent Data Mining and Knowledge Discovery*. 2(3), 13-17, 2022.

FONSECA, M. B. Classification of Bipolar Disorder, Schizophrenia and Depression Using Artificial Neural Networks. **Tese de doutorado (Catholic University of Pelotas)**. Pelotas, 2019.

FULMER, R.; JOERIN, A.; GENTILE, B.; LAKERINK, L.; RAUWS, M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. **JMIR mental health**, v. 5, n. 4, p. e9782, 2018.

GONÇALVES, E. C. Data Mining com a Ferramenta Weka. Rio de Janeiro: **Escola Nacional de Ciências Estatísticas (IBGE/ENCE)**, 2011.

GRAHAM, S., DEPP, C., LEE, E.E.; NEBEKER, C.; TU, X.; KIM, H. C.; JESTE, D. V. Artificial intelligence for mental health and mental illnesses: an overview. **Current Psychiatry Reports**, v. 21, p. 1-18, 2019.

HALL, M.; FRANK, E.; HOLMES G.; PFAHRINGER, B.; REUTEMANN P.; WITTEN, Ian H. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v. 11, n. 1, p. 10-18, 2009.

JOSHI, MANJU & KANOONGO, NEHAL. Depression detection using emotional artificial intelligence and machine learning: A closer review. **Materials Today: Proceedings**. 58. 10.1016/j.matpr.2022.01.467. 2022.

JUNG, J.; SONG, H.; KIM, Y.; IM, H.; OH, S. Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. **Computers in Human Behavior**, 71, 291–298. 2017.

KHANNA, D.; JINDAL, N.; SINGH, H.; RANA, P. S. Applications and Challenges in Healthcare Big Data: A Strategic Review. *Current Medical Imaging*, 19(1), 27-36, 2023.

KUMAR, P.; GARG, S.; GARG, A. Assessment of anxiety, depression and stress using machine learning models. **Procedia Computer Science**, 171, p. 1989-1998. 2020.

LE GLAZ, A.; HARALAMBOUS, Y.; KIM-DUFOR, D.H.; LENCA, P.; BILLOT, R.; TAYLOR, R.; MARSH, J.; DEVYLDER J.; WALTER, M.; BERROUIGUET, S.; LEMEY, C. Machine learning and natural language processing in mental health: systematic review. **Journal of Medical Internet Research**, 23(5), e15708.

LOCH, A. Governo deve estimular a ida de psicólogos para o interior do País Atualidades. **Jornal da USP no Ar. Ed.1. Rádio USP**. São Paulo, 2020.

LOVIBOND, S.H.; LOVIBOND, P.F. Manual for the Depression Anxiety Stress Scales, 4nd. **Sydney: Psychology Foundation**, 2004.

MARRO, A. A.; SOUZA, A. D. C.; CAVALCANTE, E. D. S.; BEZERRA, G. S.; NUNES, R. O. Lógica fuzzy: conceitos e aplicações. **Natal: Universidade Federal do Rio Grande do Norte (UFRN)**, p. 2, 2010.

MED PREV. Qual a diferença entre doença, síndrome, sintoma e transtorno? Website **Med Prev**. Disponível em: <https://medprev.online/blog/doencas/qual-a-diferenca-entre-doenca-sindrome-sintoma-e-transtorno/>. Acesso em: 02/06/2023.

MELLO, A. F.; JURUENA, M. F.; PARIANTE, C. M.; TYRKA, A. R.; PRICE, L. H.; CARPENTER, L. L.; DEL PORTO, J. A. Depressão e estresse: existe um endofenótipo? *In: Brazilian Journal of Psychiatry*, 29, s13-s18. 2007.

MITCHELL, T. M. Machine Learning. New York: **McGraw-Hill**, 2010.

NARAYANRAO, P. V.; KUMARI, P. L. S. Analysis of machine learning algorithms for predicting depression. In: 2020 IEEE International Conference on Computer Science, Engineering and Applications (ICCSEA), 2020. p. 1-4.

NAYAN, M. I. H.; UDDIN, M. S. G.; HOSSAIN, M. I.; ALAM, M. M.; ZINNIA, M. A.; HAQ, I., ... & METHUN, M. I. H. Comparison of the performance of machine learning-

based algorithms for predicting depression and anxiety among University Students in Bangladesh: A result of the first wave of the COVID-19 pandemic. **Asian Journal of Social Health and Behavior**, 5(2), 75. 2022.

NIZAM, D. N. M.; AHMAD, N. H.; BAHARUM, A.; SANI, Z. H. A.; SALIMUN, C.; ABAS, R. B. Developing a Depression Detector Application by integrating the DASS-21 measures: A Preliminary Study on University Students. *In: International Journal of Advanced Trends in Computer Science and Engineering*, 9(1.4). 2020.

OLIVEIRA, L. M. de; FERNANDES JUNIOR, L. C. C. Aplicabilidade da inteligência artificial na psiquiatria: uma revisão de ensaios clínicos. **Debates em Psiquiatria**, Rio de Janeiro, v. 10, n. 1, p. 14–25, 2020. DOI: 10.25118/2236-918X-10-1-2.

OPAS – Organização Pan-Americana da Saúde. **Depressão**. Disponível em: [https://www.paho.org/pt/topicos/depressao#:~:text=Em%20todo%20o%20mundo%20C%20estima,s%C3%A3o%20mais%20afetadas%20que%20homens](https://www.paho.org/pt/topicos/depressao#:~:text=Em%20todo%20o%20mundo%20C%20estima,s%C3%A3o%20mais%20afetadas%20que%20homens.). Acesso em: 03/11/2022a.

OPAS – Organização Pan-Americana da Saúde. **OMS e OIT fazem chamado para novas medidas de enfrentamento das questões de saúde mental no trabalho**. 28 Set. 2022. Disponível em: <https://www.paho.org/pt/noticias/28-9-2022-oms-e-oit-fazem-chamado-para-novas-medidas-enfrentamento-das-questoes-saude>. Acesso em: 03/11/2022b.

OPAS – Organização Pan-Americana da Saúde. **Pandemia de COVID-19 desencadeia aumento de 25% na prevalência de ansiedade e depressão em todo o mundo**. 02 Mar 2022. Disponível em: <https://www.paho.org/pt/noticias/2-3-2022-pandemia-covid-19-desencadeia-aumento-25-na-prevalencia-ansiedade-e-depressao-em>. Acesso em: 03/11/2022c.

OPAS – Organização Pan-Americana da Saúde. **Transtornos Mentais**. Disponível em: <https://www.paho.org/pt/topicos/transtornos-mentais>. Acesso em: 22/11/2022d.

OPAS – Organização Pan-Americana da Saúde. **Relatório da OMS destaca déficit global de investimentos em saúde mental**. 8 Out. 2021. Disponível em: <https://www.paho.org/pt/noticias/8-10-2021-relatorio-da-oms-destaca-deficit-global-investimentos-em-saude-mental>. Acesso em: 22/11/2022e.

PALATTAO, C. A. V.; SOLANO, G. A.; TEE, C. A.; TEE, M. L. Determining factors contributing to the psychological impact of the COVID-19 Pandemic using machine learning. **International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)**. p. 219-224. IEEE, 2021.

PATON, L. W.; TIFFIN, P. A. Technology Matters: Machine learning approaches to personalised child and adolescent mental health care. *Child and Adolescent Mental Health*, v. 27, n. 3, p. 307-308, 2022.

PINTO-COSTA, I. C.; BARBOSA COSTA, A. C.; ROCHA SILVA ASSUNÇÃO, M.; LOPES CHAVES, G.; BRESSAN, V. R.; LOPES, E. D. C. Consequências da pandemia

do novo coronavírus (covid-19) na saúde mental da população mundial: scoping review. **Revista Científica de Enfermagem-RECIEN**, 11(36). 2021.

PRIYA, A.; GARG, S.; TIGGA, N.P. Predicting anxiety, depression and stress in modern life using machine learning algorithms. **Procedia Computer Science**, 167: 1258-1267. 2020.

QUINLAN, J. R. C4.5: programs for machine learning. **San Francisco: Morgan Kaufmann Publ. Inc.**, 1993.

QUINLAN, J. R. Induction of decision trees. **Machine Learning**, 1(1):81-106.1986.

RODRIGUES, W. O. P.; WOLF, R.; VIEIRA, D. A. H.; dos REIS NETO, J. F.; de SOUZA, C. C. Mensuração da satisfação do cliente: uma comparação entre lógica fuzzy e regressão linear. **Congresso Internacional de Administração**. 2010.

RUSSELL, S.; NORVIG, P. A modern, agent-oriented approach to introductory artificial intelligence. **Acm Sigart Bulletin**, v. 6, n. 2, p. 24-26, 1995.

SANGANI, D. P.; NANDANKAR, P. J.; ANEKAR, D.; PANDA, S.; NOOR, S. S.. Psychotherapist Chatbot – a Literature Review. **EasyChair**, 2022.

SAU, A.; BHAKTA, I. Predicting anxiety and depression in elderly patients using machine learning technology. **Healthcare Technology Letters**, 4(6), 238- 243. 2017.

SRIVIDYA, M.; MOHANAVALLI, S.; BHALAJI, N. Behavioral modeling for mental health using machine learning algorithms. *In: Journal of medical systems*, v. 42, n. 5, p. 1-12, 2018.

VAISHNAVI, K.; KAMATH, U. N.; RAO, B. A.; REDDY, N. S. Predicting mental health illness using machine learning algorithms. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012021). IOP Publishing, 2022.

VIGNOLA, R.C.B.; TUCCI, A.M. Vignola, R. C. B. (2013). Escala de depressão, ansiedade e estresse (DASS): adaptação e validação para o português do Brasil. Dissertação de Mestrado, Programa de Pós-Graduação Interdisciplinar em Ciências da Saúde, Universidade Federal de São Paulo, 2013.

WANG, R.; WANG, J.; LIAO, Y.; WANG, J. Supervised machine learning chatbots for perinatal mental healthcare. *In: International Conference on Intelligent Computing and Human-Computer Interaction* (ICHCI 2020), 378-383, 2020.

WESABI, F. N., ALSOLAI, H., HILAL, A. M., HAMZA, M. A., AL DUHAYYIM, M., EGM, N. Machine Learning Based Depression, Anxiety, and Stress Predictive Model During COVID-19 Crisis. *CMC-Comput. Mat. Contin.*, p. 5803-5820, 2022.

ZADEH, L. A. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Transactions on System*, **Man and Cybernetics**, Vol.3, p. 2844, 1973.

APÊNDICE I. AD gerada na classificação dos dados relativos à depressão



APÊNDICE III. AD gerada na classificação dos dados relativos ao estresse

