

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

***DepressionMap: Um Sistema Para Indicadores de Depressão e
Análise de Dados Sociodemográficos***

Dissertação de mestrado

Noirane Getirana de Sá



São Cristóvão – Sergipe

2025

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Noirane Getirana de Sá

***DepressionMap: Um Sistema Para Indicadores de Depressão e
Análise de Dados Sociodemográficos***

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Sergipe como requisito parcial para a obtenção do título de mestre em Ciência da Computação.

Orientador(a): Prof. Dr. Gilton José Ferreira da Silva

São Cristóvão – Sergipe

2025

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL
UNIVERSIDADE FEDERAL DE SERGIPE

S111d Sá, Noirane Getirana de
Depressionmap: um sistema para indicadores de depressão e análise de dados sociodemográficos / Noirane Getirana de Sá ; orientador Gilton José Ferreira da Silva. - São Cristóvão, 2025.
74 f.; il.

Dissertação (mestrado em Ciência da Computação) – Universidade Federal de Sergipe, 2025.

1. Saúde. 2. Depressão mental. 3. Levantamentos demográficos. 4. Aprendizado do computador. I. Silva, José Ferreira da orient. II. Título.

CDU 004:616.89-008.454



UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Ata da Sessão Solene de Defesa da Dissertação do
Curso de Mestrado em Ciência da Computação-UFS.
Candidata: **NOIRANE GETIRANA DE SÁ**

Em 27 dias do mês de janeiro do ano de dois mil e vinte cinco, com início às 16hs, realizou-se na Sala de Seminários do PROCC da Universidade Federal de Sergipe, na Cidade Universitária Prof. José Aloísio de Campos, a Sessão Pública de Defesa de Dissertação de Mestrado do candidata **NOIRANE GETIRANA DE SÁ**, que desenvolveu o trabalho intitulado: **“DepressionMap: Um Sistema Para Indicadores de Depressão e Análise de Dados Sociodemográficos”**, sob a orientação do Prof. Dr. **Gilton José Ferreira da Silva**. A Sessão foi presidida pelo Prof. Dr. **Gilton José Ferreira da Silva** (PROCC/UFS), que após a apresentação da dissertação passou a palavra aos outros membros da Banca Examinadora, a Dra. **Zenith Nara Costa Delabrida** (DPS - UFS) e, em seguida, Dr. **Fábio Gomes Rocha** (PROCC/UFS). Após as discussões, a Banca Examinadora reuniu-se e considerou o mestrando (a) **APROVADA** “(aprovado/reprovado)”. Atendidas as exigências da Instrução Normativa 05/2019/PROCC, do Regimento Interno do PROCC (Resolução 67/2014/CONEPE), e da Resolução nº 04/2021/CONEPE que regulamentam a Apresentação e Defesa de Dissertação, e nada mais havendo a tratar, a Banca Examinadora elaborou esta Ata que será assinada pelos seus membros e pelo mestrando.

Cidade Universitária “Prof. José Aloísio de Campos”, 27 de janeiro de 2025.

Documento assinado digitalmente
gov.br GILTON JOSE FERREIRA DA SILVA
Data: 27/01/2025 22:59:06-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Gilton José Ferreira da Silva
(PROCC/UFS)
Presidente

Documento assinado digitalmente
gov.br ZENITH NARA COSTA DELABRIDA
Data: 28/01/2025 08:45:46-0300
Verifique em <https://validar.iti.gov.br>

Profª. Dra Zenith Nara Costa Delabrida
(DPS - UFS)
Examinador Externo

Documento assinado digitalmente
gov.br FABIO GOMES ROCHA
Data: 10/02/2025 15:47:41-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Fábio Gomes Rocha
(PROCC/UFS)
Examinador Interno

Documento assinado digitalmente
gov.br NOIRANE GETIRANA DE SA
Data: 10/02/2025 14:59:18-0300
Verifique em <https://validar.iti.gov.br>

Noirane Getirana de Sá
Candidata

Agradecimentos

Agradeço primeiramente a Deus por ser meu mestre e guiar meus passos.

Agradeço ao meu pai, José Ginaldo, por ser uma inspiração ao longo da minha vida. A minha mãe, Maria da Glória, minha melhor amiga e a pessoa mais humana que conheço. Obrigada por todas as oportunidades e ensinamentos que vocês dois me proporcionaram.

Ao meu esposo e amigo, Lucas, pela compreensão e apoio em diversos momentos. Por ter sido meu parceiro e ter entendido os momentos de ausência. Meu cachorrinho Rown por deixar os meus dias mais leves.

Aos meus irmãos, Nayane e Francisco, por tornarem a minha infância e adolescência repletas de momentos divertidos e inesquecíveis. Minhas sobrinhas, Clarice e Amélie. Minhas primas e amigas, Cleize e Tacy. Meus tios, José Carlos e Terezinha.

A todos os professores do PROCC/UFS, em especial ao meu orientador, Prof. Dr. Gilton José Ferreira da Silva, pela paciência, energia positiva, dedicação e ensinamentos. Ao Prof. Dr. Daniel Oliveira Dantas, pelo conhecimento agregado no artigo sobre detecção de glaucoma. Agradeço também à Elaine por ser sempre gentil e solícita.

Agradeço à Fundação de Apoio à Pesquisa e à Inovação Tecnológica do Estado de Sergipe (FAPITEC-SE) pela concessão da bolsa, que foi importante para a conclusão deste mestrado.

E a todos aqueles que contribuíram, diretamente ou indiretamente, para que eu pudesse concluir mais uma etapa importante da minha vida.

*Renda-se, como eu me rendi.
Mergulhe no que você não conhece como eu mergulhei.
Não se preocupe em entender, viver ultrapassa qualquer entendimento.
(Clarice Lispector)*

Resumo

Contexto: Segundo a Organização Mundial da Saúde, após a pandemia de COVID-19, ocorreu um aumento da depressão em mais de 25%, sendo a condição mental mais incapacitante do mundo. Um sistema baseado em indicadores de depressão pode auxiliar profissionais de saúde mental na triagem da doença. Essa abordagem facilita a identificação de grupos vulneráveis, o mapeamento das áreas mais afetadas e a possibilidade de um tratamento mais eficiente. **Objetivo:** Desenvolver o *DepressionMap*, um sistema com foco em gerar indicadores relacionados à depressão, contribuindo para soluções inteligentes referentes à doença. **Método:** Esta pesquisa iniciou por meio de um Mapeamento Sistemático da Literatura (MSL), mediante a leitura dos trabalhos relacionados, foram identificadas as lacunas de pesquisa referentes ao tema. Nesse contexto, foi desenvolvido o *DepressionMap*, um sistema que gera indicadores de depressão com base na escala DASS-42 e em variáveis sociodemográficas, além de disponibilizar informações sobre o desempenho de métodos de Aprendizado de Máquina para classificação dos níveis de severidade da doença. Ainda, foi proposta uma estrutura melhorada com a participação de profissionais da saúde mental. **Resultados:** Entre os algoritmos testados, a Máquina de Vetores de Suporte e a Regressão Logística destacaram-se com melhores resultados na classificação dos níveis de depressão, alcançando acurácias de 99,66% e 99,69%, respectivamente. **Conclusão:** O *DepressionMap* é um sistema capaz de gerar indicadores de depressão, contudo foram identificadas limitações relacionadas à base de dados utilizada e à ausência de profissionais especialistas em depressão. Por conta desses fatores, foi proposta uma estrutura melhorada, com maior controle na coleta de dados e com a participação de uma equipe de saúde. Além disso, este estudo apresentou avanços na classificação dos níveis de depressão em comparação com pesquisas relacionadas.

Palavras-chave: Saúde Inteligente, DASS-42, Dados Sociodemográficos, Aprendizado de Máquina, Depressão.

Abstract

Context: According to the World Health Organization, after the COVID-19 pandemic, there was an increase of more than 25% in depression, making it the most disabling mental condition in the world. A system based on depression indicators can assist mental health professionals in screening for the disease. This approach facilitates the identification of vulnerable groups, the mapping of the most affected areas and the possibility of more efficient treatment. **Objective:** To develop *DepressionMap*, a system focused on generating indicators related to depression, contributing to intelligent solutions related to the disease. **Method:** This research began through a Systematic Literature Mapping (SLM), by reading related works, research gaps related to the topic were identified. In this context, *DepressionMap* was developed, a system that generates depression indicators based on the DASS-42 scale and sociodemographic variables, in addition to providing information on the performance of Machine Learning methods for classifying disease severity levels. Furthermore, an improved structure was proposed with the participation of mental health professionals. **Results:** Among the algorithms tested, the Support Vector Machine and Logistic Regression stood out with the best results in classifying depression levels, reaching accuracies of 99.66% and 99.69%, respectively. **Conclusion:** *DepressionMap* is a system capable of generating depression indicators; however, limitations related to the database used and the lack of professionals specialized in depression were identified. Due to these factors, an improved structure was proposed, with greater control over data collection and the participation of a health team. Furthermore, this study presented advances in the classification of depression levels compared to related research.

Keywords: Smart Health, DASS-42, Sociodemographic Data, Machine Learning, Depression.

Lista de ilustrações

Figura 1 – Níveis DASS	23
Figura 2 – Cálculo da Pontuação da Depressão DASS	23
Figura 3 – Características das Cidades Inteligentes	25
Figura 4 – Inteligência Artificial e Subáreas	27
Figura 5 – Etapas da Metodologia	33
Figura 6 – Fluxograma das Etapas da Execução	39
Figura 7 – Quantidade de Trabalhos Encontrados por Base	40
Figura 8 – Quantidade de Trabalhos Selecionados por Base	40
Figura 9 – Técnicas de Inteligência Artificial Utilizadas	44
Figura 10 – Distribuição dos Estudos Selecionados por Ano de Publicação	45
Figura 11 – Mapa de Calor	50
Figura 12 – Gênero e sua Relação com a Depressão	51
Figura 13 – Comparação dos Modelos Acurácia e Medida-F1	56
Figura 14 – <i>DepressionMap</i> x Estrutura Melhorada	60

Lista de quadros

Quadro 1 – Itens do DASS-42	22
Quadro 2 – Questões de Pesquisa	35
Quadro 3 – Campos de Extração	36
Quadro 4 – Bases de Pesquisa e URL	37
Quadro 5 – Critérios de Inclusão e Exclusão	37
Quadro 6 – Lista de Palavras-Chave e Tradução em Inglês	37
Quadro 7 – <i>String</i> de Busca	38
Quadro 8 – Categorização dos Artigos Seleccionados	43
Quadro 9 – Intervalos Válidos Para as Variáveis Sociodemográficas	49
Quadro 10 – Quantidade de Participantes por Região	51
Quadro 11 – Índice da Depressão por Região e País	53

Lista de tabelas

Tabela 1 – Escalas e Seus Respectivos Valores de <i>Cronbach's Alpha</i> (α)	21
Tabela 2 – Bibliotecas, Versão e Endereço Eletrônico	34
Tabela 3 – Distribuição Percentual por Grupo e Nível	52
Tabela 4 – Resultados dos Modelos de Classificação	56
Tabela 5 – Teste de <i>Shapiro-Wilk</i> para Análise da Normalidade	56
Tabela 6 – Teste de <i>Post-Hoc Nemenyi</i>	57
Tabela 7 – Comparação dos Modelos de Aprendizado de Máquina	57

Lista de abreviaturas e siglas

ADABOOST	Adaptive Boosting
BAI	Beck Anxiety Inventory
BDI	Beck Depression Inventory
BLR	Binomial Logistic Regression
CATBOOST	Categorical Boosting
CE	Critérios de Exclusão
CES-D	Center for Epidemiologic Studies Depression Scale
CI	Critérios de Inclusão
CNN	Convolutional Neural Networks
COVID-19	Doença por Coronavírus 2019
CRSD	Carroll Rating Scale for Depression
DAS	Depression, Anxiety, and Stress
DASS	Depression, Anxiety, and Stress Scale
DL	Deep Learning
DT	Decision Tree
FN	Falso Negativo
FP	Falso Positivo
GNB	Gaussian Naive Bayes
GSE	General Self-Efficacy Scale
IA	Inteligência Artificial
IFSSML-DAS	Intelligent Feature Subset Selection with Machine Learning-based DAS
IoMT	Internet of Medical Things
IoT	Internet of Things
KNN	K-Nearest Neighbors

LDA	Linear Discriminant Analysis
LR	Logistic Regression
ML	Machine Learning
MLR	Multinomial Logistic Regression
MSL	Mapeamento Sistemático da Literatura
MTP	Multi-Target Prediction
OMS	Organização Mundial da Saúde
PCA	Principal Component Analysis
RF	Random Forest
RNA	Redes Neurais Artificiais
SMI	Severe Mental Illness
STAI	State Trait Anxiety Inventory
SVM	Support Vector Machine
TI	Tecnologia da Informação
TIC	Tecnologias de Informação e Comunicação
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
WESAD	Wearable Stress and Affect Detection
XGBoost	Extreme Gradient Boosting
Zung SRAS	Zung Self-rating Scale for Anxiety
Zung SRDS	Zung Self-rating Scale for Depression

Sumário

1	Introdução	15
1.1	Justificativa e Motivação	16
1.2	Objetivos	17
1.2.1	Objetivo Geral	17
1.2.2	Objetivos Específicos	17
1.3	Estrutura da Dissertação	18
2	Fundamentação Teórica	19
2.1	Saúde Mental	19
2.1.1	DASS-42	20
2.2	Cidades Inteligentes	23
2.2.1	Saúde Inteligente	24
2.2.2	Promoção da Saúde Mental em Cidades Inteligentes	25
2.3	Aprendizado de Máquina	26
2.3.1	<i>Overfitting</i>	27
2.3.2	Validação Cruzada	27
2.3.3	<i>One-hot Encoding</i>	28
2.3.4	Normalização	28
2.3.5	<i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	28
2.3.6	<i>Logistic Regression (LR)</i>	28
2.3.7	<i>Support Vector Machine (SVM)</i>	29
2.3.8	<i>Random Forest (RF)</i>	29
2.3.9	<i>K-Nearest Neighbors (KNN)</i>	29
2.3.10	<i>Gaussian Naive Bayes (GNB)</i>	30
2.3.11	<i>Decision Tree (DT)</i>	30
2.3.12	<i>Adaptive Boosting (AdaBoost)</i>	30
2.3.13	Métricas de Qualidade	31
3	Materiais e Métodos	32
3.1	Base de Dados	34
3.2	Ferramentas	34
4	Trabalhos Relacionados	35
4.1	Planejamento	35
4.2	Execução	36
4.3	Análise e Divulgação	42

4.3.1	Q1 - Quais Técnicas de IA Foram Utilizadas?	42
4.3.2	Q2 - O Estudo Leva em Consideração a Participação de Profissionais de Saúde para o Diagnóstico da Depressão?	44
4.3.3	Q3 - Qual o Ano de Publicação do Trabalho?	44
4.4	Ameaças à Validade	45
4.5	Discussão Sobre a MSL	46
5	Desenvolvimento do <i>DepressionMap</i>	47
5.1	Preparação do Ambiente	47
5.2	Pré-processamento dos Dados	47
5.3	Análise dos Dados Sociodemográficos	48
5.3.1	Resultados da Análise dos Dados Sociodemográficos	50
5.3.2	Depressão e Região	52
5.3.3	Discussão Sobre a Análise Sociodemográfica	52
5.4	Experimento Classificação dos Níveis de Depressão	54
5.4.1	Objetivo	54
5.4.2	Contexto	54
5.4.3	Formulação do Problema e Hipóteses	54
5.4.4	Variáveis Independentes e Dependentes	54
5.4.5	Execução	55
5.4.6	Validação dos dados	55
5.4.7	Resultados	55
5.4.8	Ameaças à Validade	58
5.4.9	Considerações Sobre o Experimento	58
6	Proposta de uma Estrutura Melhorada para o <i>DepressionMap</i>	59
6.1	Estrutura Melhorada	59
6.1.1	Coleta dos Dados	59
6.1.2	Criação e Disponibilização do <i>DepressionMap</i> Melhorado	61
6.1.3	Atuação da Equipe de Saúde	61
6.2	Ameaças à Validade	61
7	Considerações Finais e Trabalhos Futuros	62
	Referências	64

1

Introdução

As emoções desempenham um papel essencial no comportamento humano (BUCKLEY, 2016). Os estados emocionais negativos podem impactar o processo de cognição (GRAY, 2004), influenciando o raciocínio e o autocontrole (PHAM, 2007). Ainda podem desencadear diversas condições psicopatológicas, tais como: transtornos de humor, ansiedade e dependência de drogas (BARKER, 2018). A saúde mental enfraquecida compromete a qualidade de vida, afetando a produtividade dos indivíduos (EVANS et al., 2006).

Em relação aos estados emocionais negativos que geram impactos nocivos na saúde mental, a depressão se destaca. De acordo com a Organização Mundial da Saúde (OMS), a depressão afeta aproximadamente 350 milhões de pessoas no mundo (JADHAV; BABAR; MAHALLE, 2023), prejudicando indivíduos de diversas faixas etárias. Diante disso, seus efeitos nocivos implicam significativamente na qualidade de vida, comprometendo aspectos como o desempenho no trabalho e os relacionamentos sociais (SHOREY; NG; WONG, 2022; MLAKI et al., 2021). A depressão ocupa o primeiro lugar entre os transtornos mentais mais incapacitantes, representando 39% dos casos, enquanto a ansiedade corresponde a 22% e a esquizofrenia representa 11,7% (World Health Organization et al., 2022).

A saúde e o bem-estar mental são prioridades nas Cidades Inteligentes (MOHANTY; CHOPPALI; KOUKIANOS, 2016), as quais usam a tecnologia para o desenvolvimento de soluções, visando aumentar a qualidade de vida da sociedade (NEIROTTI et al., 2014). As informações são acessadas em tempo real, podendo servir como parâmetro para decisões (JIN et al., 2014). Assim, dados relacionados à saúde mental podem colaborar no bem-estar dos indivíduos (OTEIZA; KELLY; MOONEY, 2021). Diante dos esforços direcionados à melhoria da saúde no contexto de Cidades Inteligentes, surge a área da Saúde Inteligente (*Smart Health*), que usa a tecnologia para desenvolver soluções na área da saúde (SOLANAS et al., 2014).

Por meio de tecnologias como Aprendizado de Máquina (*Machine Learning* - ML) e análise de dados, é possível gerar indicadores de estados emocionais, detectar padrões e entender

os dados, possibilitando o fornecimento das informações aos profissionais de saúde mental, agregando para tratamentos mais eficazes (CHAKRABORTY et al., 2023). Ainda, por meio da telemedicina, que utiliza tecnologias de comunicação para oferecer cuidados médicos à distância, torna-se viável que pacientes em áreas remotas ou com dificuldades de locomoção recebam atendimento médico especializado sem a necessidade de se deslocar até uma unidade de saúde (SHIH et al., 2023).

Compreender as emoções e discernir vários graus de bem-estar é relevante para aprimorar a saúde mental. Esse cenário impulsionou a criação de métodos para auxiliar a mensurar emoções (MCDOWELL, 2006). Como a depressão é um problema significativo, quando se trata de saúde mental, existem diversas iniciativas que fazem uso da tecnologia para ajudar nesse desafio, inclusive fazendo uso de escalas ou questionários. Na Escala de Depressão, Ansiedade e Estresse (*Depression, Anxiety, and Stress Scale* - DASS), são analisados três estados emocionais: Depressão, Ansiedade e Estresse (*Depression, Anxiety and Stress* - DAS). Originalmente, o DASS é composto por 42 itens, também denominado DASS-42. No entanto, existem versões mais curtas, como o DASS-21, que contém apenas 21 itens (SRINATH et al., 2022).

A escala DASS, em conjunto com as técnicas de *Machine Learning*, tem sido amplamente utilizada para gerar indicadores de depressão (BUDIYANTO; SIHOMBING; IM, 2019; SINGH et al., 2022; TAN; AL., 2023). Indicadores podem ajudar a identificar problemas de saúde e fornecer informações úteis para o desenvolvimento de estratégias com foco na melhoria da saúde populacional (GRACE, 1997). A Inteligência Artificial (IA), do inglês *Artificial Intelligence*, é uma ferramenta relevante nesse contexto, auxiliando na saúde mental ao possibilitar soluções que ajudam a prever, detectar e tratar problemas mentais (D'ALFONSO, 2020; MAJUMDER; VEILLEUX, 2021).

Apesar dos obstáculos a serem considerados, as Cidades Inteligentes progridem otimizando e fortalecendo a qualidade dos serviços pela utilização mais inteligente dos recursos urbanos (KULANDAIVEL et al., 2019). Diversos projetos em todo o mundo, iniciativas públicas e empresas estão investigando técnicas e ferramentas para auxiliar a mensurar níveis emocionais (ZHU; SHEN; REN, 2022; ZAKZAK, 2019; MACEDO; BALTATU; SILVA, 2015), sendo assim estruturas de Saúde Inteligente com indicadores de depressão podem ajudar profissionais de saúde mental na triagem e acompanhamento dos pacientes.

1.1 Justificativa e Motivação

A assistência médica inteligente utiliza recursos tecnológicos para acessar informações de forma dinâmica, conectar pessoas com instituições de saúde e ajudar a atender de maneira eficiente às necessidades do setor médico. Esse tipo de abordagem pode apoiar a tomada de decisões da área da saúde (TIAN et al., 2019). Assim sendo, indicadores de saúde mental podem auxiliar na proposta de soluções otimizadas (SAMARTZIS; TALIAS, 2019).

Com o avanço da tecnologia, os sistemas inteligentes têm transformado significativamente a forma como os serviços de saúde são oferecidos. Soluções inteligentes podem facilitar diagnósticos e triagem de doenças para os profissionais da saúde. Dentro desse contexto, os algoritmos de Aprendizado de Máquina têm se mostrado relevantes para aprimorar os sistemas relacionados à saúde. No caso da depressão, essas tecnologias podem ser usadas para realizar previsões ou classificações relacionadas à doença, fornecendo indicadores que podem ajudar os profissionais de saúde a tomar decisões. Essa abordagem não só melhora a eficiência dos serviços de saúde, mas também permite que os tratamentos sejam mais personalizados e direcionados às necessidades específicas dos pacientes (ABDEEN et al., 2022).

Outro aspecto relevante é a consideração das variáveis sociodemográficas e como se relacionam com a depressão. Segundo Link e Phelan (1996), as sociedades desempenham um papel crucial na formação dos padrões de doenças e as variáveis sociodemográficas são determinantes importantes para a compreensão dos resultados de saúde. No caso da depressão, diversos estudos demonstram uma associação entre os fatores sociodemográficos e a doença (JORM, 2000; BUCKMAN; AL., 2021; TAN; AL., 2023).

Por fim, a falta de interação direta com profissionais de saúde também é um aspecto importante, pois a tecnologia não substitui o acompanhamento de profissionais da saúde, que realizam uma análise detalhada e personalizada dos pacientes (VASCONCELLOS-SILVA; CASTIEL, 2009).

1.2 Objetivos

Nesta seção serão apresentados o objetivo geral e os objetivos específicos do trabalho.

1.2.1 Objetivo Geral

Desenvolver o *DepressionMap*, um sistema que apresenta indicadores de depressão e sua relação com variáveis sociodemográficas, além de disponibilizar dados sobre a classificação de níveis de depressão com técnicas de Aprendizado de Máquina.

1.2.2 Objetivos Específicos

Com o intuito de atingir o objetivo geral, este estudo tem os seguintes objetivos específicos:

- Realizar o Mapeamento Sistemático da Literatura (MSL): análise exploratória para selecionar modelos relevantes ao tema;
- Realizar o pré-processamento do *dataset*: limpeza dos dados e cálculo dos indicadores dos níveis de depressão;

- Implementar os indicadores de depressão e suas relações com variáveis sociodemográficas;
- Realizar um experimento visando à análise dos métodos de Aprendizado de Máquina na classificação dos níveis de depressão;
- Modelar a proposta de uma estrutura melhorada para o *DepressionMap*, com participação de profissionais da saúde mental.

1.3 Estrutura da Dissertação

Para facilitar a compreensão e a navegação, este documento segue uma estrutura organizada em capítulos da seguinte forma:

- Capítulo 1 - Introdução: É apresentado o contexto e os objetivos deste trabalho;
- Capítulo 2 - Fundamentação Teórica: São explicados os conceitos necessários para compreender a proposta deste estudo;
- Capítulo 3 - Materiais e Métodos: É apresentada a metodologia empregada para o desenvolvimento do *DepressionMap*, bem como o *dataset* e as ferramentas utilizadas;
- Capítulo 4 - Trabalhos Relacionados: São abordados os estudos relevantes para o tema;
- Capítulo 5 - Desenvolvimento do *DepressionMap*: Fornece uma descrição detalhada do desenvolvimento da ferramenta *DepressionMap*;
- Capítulo 6 - Proposta de uma estrutura melhorada para o *DepressionMap*: Proposta de uma estrutura com a participação de profissionais de saúde;
- Capítulo 7 - Considerações Finais e Trabalhos Futuros: Expõe considerações finais sobre o estudo, principais contribuições e trabalhos futuros.

2

Fundamentação Teórica

Neste capítulo, são expostos os conceitos teóricos relevantes para a realização deste estudo. A seguir, serão apresentados conceitos referentes a: Saúde Mental, Cidades Inteligentes e Aprendizado de Máquina.

2.1 Saúde Mental

O conceito de saúde mental tem origem antiga que se estende até filósofos como Platão e Aristóteles ([SEESKIN, 2008](#)). Saúde mental é um estado de bem-estar no qual o indivíduo reconhece suas próprias habilidades, podendo lidar com o estresse da vida, trabalhar de forma produtiva e ser capaz de dar uma contribuição à sua comunidade ([ORGANIZATION et al., 2004](#)). A preservação da saúde mental contribui para a saúde do ser humano, influenciando o desenvolvimento social e econômico ([HERRMAN; JANÉ-LLOPIS, 2012](#)). Fatores sociais como baixa renda, moradia insegura, educação limitada, desemprego e trabalho de alta demanda, podem impactar a saúde mental, assim como aspectos genéticos e tabagismo ([FISHER; BAUM, 2010](#); [LOUIS, 2022](#); [FIRTH et al., 2020](#)).

Além disso, doenças crônicas podem ter um impacto significativo na saúde mental de um indivíduo ([LANDO et al., 2006](#)). A doença mental grave inclui esquizofrenia, transtorno bipolar e depressão grave ([UHER, 2014](#)). Uma Doença Mental Grave (*Severe Mental Illness* - SMI) está associada ao aumento de doenças nutricionais e metabólicas, doenças cardiovasculares, doenças virais, doenças do trato respiratório, disfunção sexual, complicações na gravidez e, possivelmente a problemas relacionados à obesidade ([HERT et al., 2011](#)).

As emoções negativas, tais como tristeza, medo e raiva, influenciam diferentes maneiras de pensar e reagir ([BERNBAU; FUJITA; PFENNIG, 1995](#)). Estados emocionais negativos têm sido associados ao aumento de cortisol circulante e à baixa da função imunológica, o que pode aumentar a suscetibilidade à infecção ([CHESNEY et al., 2005](#)). As emoções negativas estão

associadas a estados emocionais como: depressão, ansiedade e estresse (KEDAR; ROKADE, 2019; LOVIBOND, 1995). A depressão é um dos transtornos mentais mais comuns, impactando tanto o bem-estar físico quanto emocional de quem a enfrenta (VIMALESWARAN; RATNAYAKE, 2021).

A Higiene Mental trata da promoção da saúde mental, bem como a prevenção e tratamento precoce de neuroses, psicoses, transtornos de caráter e personalidade e outros tipos de desajustes em adultos e crianças (MARFATIA, 1972). Com a popularização do termo Higiene Mental, surgiu em 1909 nos Estados Unidos a Comissão Nacional de Higiene Mental. O movimento se internacionalizou, o que gerou associações em países como: França, África do Sul, Itália e Hungria. Essas associações posteriormente deram origem à Federação Mundial de Saúde Mental, com foco principal no cuidado de pessoas com transtornos mentais (BERTOLOTE, 2008).

A promoção da saúde implica em modos de vida saudáveis e na criação de condições de vida e ambientes propícios à saúde, juntamente com a prevenção e o tratamento de doenças mentais (HERRMAN; JANÉ-LLOPIS, 2012), visando instruir indivíduos a assumirem maior controle sobre o seu bem-estar, englobando iniciativas que têm como objetivo viabilizar efeitos positivos na saúde mental (HODGSON; ABBASI; CLARKSON, 1996).

A atividade física pode ser um fator protetor da saúde mental, destacando a recomendação de 150 minutos de atividade moderada por semana para reduzir o risco de doenças mentais. Sono adequado pode ajudar na prevenção de transtornos mentais (FIRTH et al., 2020). Técnicas e abordagens psicológicas, como: intervenções baseadas em atenção plena e terapia cognitiva-comportamental, podem ajudar indivíduos que sofrem com estresse, ansiedade e depressão (MELNYK et al., 2020; SUNDQUIST et al., 2015).

2.1.1 DASS-42

Entender e avaliar a experiência humana são objetivos relevantes para a sociedade. Diferentes facetas da saúde, como deficiência, doenças crônicas e saúde mental, são minuciosamente estudadas por pesquisadores e profissionais de saúde (BERGER et al., 2016). A busca pelo bem-estar resultou na criação de métodos para capturar reações emocionais cotidianas, resultando em avaliações de satisfação com a vida. Questionários e escalas, nas avaliações psiquiátricas, são amplamente utilizados (MCDOWELL, 2006), sendo que uma escala emocional negativa pode ser uma estrutura que analisa aspectos como: tristeza, medo, raiva e vergonha (SMITS et al., 2002). Escalas podem ter sua confiabilidade avaliada pelo *Cronbach's Alpha* (TABER, 2017; PETERSON, 1994). A confiabilidade é o grau em que as medidas estão livres de erros, produzindo resultados consistentes (PETER, 1979).

A medida *Cronbach's Alpha* foi utilizada no estudo Crawford et al. (2011), que compara diversas escalas de adoecimento mental, como: Inventário de Ansiedade de Beck (*Beck Anxiety Inventory* - BAI), Inventário de Depressão de Beck (*Beck Depression Inventory* - BDI), Escala de

Avaliação de Depressão de Carroll (*Carroll Rating Scale for Depression* - CRSD), Escala de Depressão do Centro de Estudos Epidemiológicos (*Center for Epidemiologic Studies Depression Scale* - CES-D), DASS-42, DASS-21, Escala de Autoavaliação de Zung para Ansiedade (*Zung Self-rating Scale for Anxiety* - Zung SRAS), Escala de Autoavaliação de Zung para Depressão (*Zung Self-rating Scale for Depression* - Zung SRDS), e Inventário de Ansiedade Traço-Estado (*State Trait Anxiety Inventory* - STAI).

A **Tabela 1** apresenta o comparativo entre as escalas, onde o DASS-42 e o DASS-21, na subescala depressão, apresentam resultados consistentes em relação ao *Cronbach's Alpha*, com valores de 0,95 e 0,90 respectivamente. A escala STAI é a única que faz uma diferenciação entre a ansiedade, onde ansiedade-estado reflete como um indivíduo se sente atualmente, e ansiedade-traço reflete como um indivíduo geralmente se sente (BERNSTEIN; EVELAND, 1982).

Tabela 1 – Escalas e Seus Respectivos Valores de *Cronbach's Alpha* (α)

Escala	Subescala	Cronbach's Alpha (α)
BDI	Depressão	0.89
BAI	Ansiedade	0.90
CRSD	Depressão	0.90
CES-D	Depressão	0.90
DASS-42	Depressão	0.95
DASS-42	Ansiedade	0.88
DASS-42	Estresse	0.94
DASS-21	Depressão	0.90
DASS-21	Ansiedade	0.79
DASS-21	Estresse	0.89
STAI-X	Ansiedade-estado	0.94
STAI-X	Ansiedade-traço	0.91
STAI-Y	Ansiedade-estado	0.94
STAI-Y	Ansiedade-traço	0.94
Zung SRDS	Depressão	0.86
Zung SRAS	Ansiedade	0.83

Fonte: Adaptado da fonte, 2024.

O DASS foi escrito originalmente em inglês (PELEGRINO et al., 2012), as traduções para diversos idiomas podem ser encontradas na plataforma da *University of New South Wales*¹. A única versão que tem tradução oficial para o português do Brasil é DASS-21 (VIGNOLA; TUCCI, 2014). No **Quadro 1** são expostos os 42 itens do DASS-42 em uma tradução para o português feita pelos autores.

O DASS-42 pode ser utilizado para gerar indicadores relacionados somente a depressão (FADHEL; MAHARANI, 2024; PARIDA; AL., 2023; JAMSHIDI; AL., 2020). A subescala de depressão DASS-42 consiste em 14 itens que avaliam sintomas como: sentir-se deprimido;

¹ <<https://dass.psy.unsw.edu.au/DASS%20Translations.htm>>

Quadro 1 – Itens do DASS-42

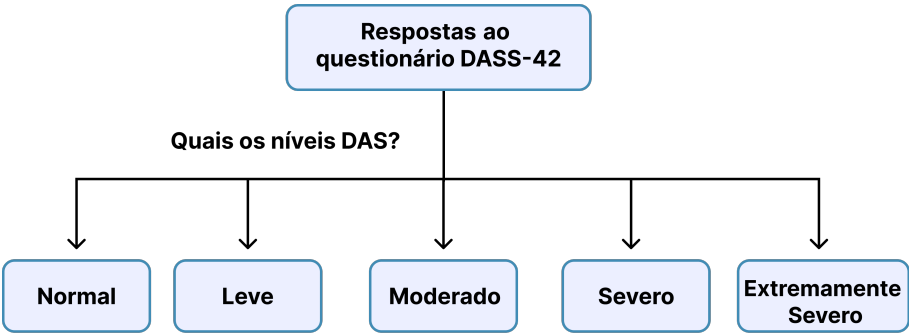
Q	Descrição
Q1	Eu me peguei ficando chateado com coisas triviais.
Q2	Eu senti a minha boca seca.
Q3	Não consegui sentir nenhum sentimento positivo.
Q4	Tive dificuldade para respirar (por exemplo, respiração excessivamente rápida, falta de ar na ausência de esforço físico).
Q5	Eu simplesmente não conseguia começar.
Q6	Tive tendência a reagir exageradamente às situações.
Q7	Tive uma sensação de tremores (por exemplo, sem forças nas pernas).
Q8	Achei difícil relaxar.
Q9	Eu me vi em situações que me deixaram tão ansioso que fiquei mais aliviado quando elas terminaram.
Q10	Eu senti que não tinha o que esperar do futuro.
Q11	Eu me percebi ficando chateado com bastante facilidade.
Q12	Senti que estava usando muita energia nervosa.
Q13	Eu me senti triste e deprimido.
Q14	Fiquei impaciente quando me atrasavam de alguma forma (por exemplo, elevadores, semáforos, ficar esperando).
Q15	Tive uma sensação de desmaio.
Q16	Senti que havia perdido o interesse em quase tudo.
Q17	Eu sentia que não valia muito como pessoa.
Q18	Eu me senti sensível.
Q19	Eu transpirava bastante (por exemplo, mãos suadas) na ausência de altas temperaturas ou esforço físico.
Q20	Fiquei assustado sem nenhum motivo aparente.
Q21	Eu sentia que a vida não valia a pena.
Q22	Achei difícil ir diminuindo a intensidade.
Q23	Tive dificuldade em engolir.
Q24	Eu não conseguia tirar nenhum prazer das coisas que fazia.
Q25	Eu senti alterações do meu coração na ausência de esforço físico
Q26	Eu me sentia desanimado e triste.
Q27	Percebi que estava muito irritado.
Q28	Eu senti que estava em pânico.
Q29	Achei difícil me acalmar depois que algo me aborreceu.
Q30	Eu temia enfrentar tarefas simples, mas desconhecidas.
Q31	Não consegui me entusiasmar com nada.
Q32	Achei difícil tolerar interrupções enquanto estava fazendo algo.
Q33	Eu estava em um estado de tensão nervosa.
Q34	Eu me sentia muito inútil.
Q35	Eu era intolerante com qualquer coisa que me impedisse de continuar a executar o que estava fazendo.
Q36	Fiquei aterrorizado.
Q37	Não consigo ver nada no futuro que me dê esperança.
Q38	Eu sentia que a vida não tinha sentido.
Q39	Eu fiquei agitado.
Q40	Fiquei preocupado com situações em que eu poderia entrar em pânico e fazer papel de bobo.
Q41	Senti tremores (por exemplo, nas mãos).
Q42	Achei difícil ter iniciativa para fazer as coisas.

sentir-se melancólico; sentir-se autodepreciativo; sentir-se pessimista sobre o futuro; sentir-se incapaz de sentir prazer ou satisfação; sentir que a vida não tem significado ou valor; e sentir-se incapaz de se interessar ou se envolver (BEYER et al., 2014). Os entrevistados classificam a extensão em que experimentaram esses sintomas na semana anterior (POORAVARI et al., 2017), cada item é classificado em uma escala de 4 pontos, com as seguintes opções:

- 0 – Não se aplica a mim de forma alguma;
- 1 – Aplica-se a mim em algum grau, ou algumas vezes;
- 2 – Aplica-se a mim em um grau considerável, ou boa parte do tempo;
- 3 – Aplica-se a mim muito, ou a maior parte do tempo;

Para calcular a pontuação dos indicadores de depressão, é necessário somar os valores das respostas de 14 questões específicas: 3, 5, 10, 13, 16, 17, 21, 24, 26, 31, 34, 37, 38 e 42. A pontuação pode ser classificada em níveis de severidade: Normal (0-9), Leve (10-13), Moderado (14-20), Severo (21-27) e Extremamente Severo (28+) (SRINATH et al., 2022). Para deixar mais claro, a Figura 1 demonstra os possíveis níveis de depressão segundo o DASS-42, enquanto a Figura 2 mostra exemplos dos cálculos das pontuações e dos respectivos níveis de depressão.

Figura 1 – Níveis DASS



Fonte: Produção dos autores, 2024.

Figura 2 – Cálculo da Pontuação da Depressão DASS

Q3	Q5	Q10	Q13	Q16	Q17	Q21	Q24	Q26	Q31	Q34	Q37	Q38	Q42	PONTUAÇÃO	NÍVEL
2	3	1	1	1	3	3	0	2	2	3	3	2	2	28	EXTREMAMENTE SEVERO
3	0	0	0	2	2	2	1	3	1	2	2	0	1	19	MODERADO
3	2	1	0	0	2	2	1	1	0	0	1	0	0	13	LEVE

Fonte: Produção dos autores, 2024.

2.2 Cidades Inteligentes

O conceito de Cidades Inteligentes tem se mostrado relevante na literatura científica nas últimas décadas (ALBINO; BERARDI; DANGELICO, 2015). Cidades Inteligentes são áreas

urbanas altamente adaptáveis e inovadoras, com uma infraestrutura digital robusta, o que as capacitam a evoluir continuamente (HOLLANDS, 2020), fazem uso de tecnologias avançadas de comunicação para oferecer serviços eficientes tanto para a administração urbana quanto para os cidadãos (ZANELLA et al., 2014).

As Cidades Inteligentes surgiram para integrar os elementos essenciais da produção urbana contemporânea, destacando o papel das Tecnologias de Informação e Comunicação (TIC) na melhoria do perfil competitivo das cidades (CARAGLIU; BO; NIJKAMP, 2011), podendo fazer uso da tecnologia para auxiliar em desafios sociais: coleta de lixo controlada pela *internet*, detectar áreas propensas ao crime e alertas precoces a deslizamentos de terra (GUPTA, 2023). Iniciativas inteligentes não são somente mudanças tecnológicas, mas também investimentos em recursos humanos, assim a TIC é complementar ao capital humano (NEIROTTI et al., 2014).

As cidades estão se tornando inteligentes e permitindo monitorar, compreender e analisar, almejando a eficiência, a equidade e a qualidade de vida dos seus cidadãos (BATTY et al., 2012). Na literatura é possível encontrar diversos projetos para melhoria das Cidades Inteligentes, por exemplo, na sustentabilidade ambiental: Benedict, McMahon et al. (2002) e Esty e Winston (2009), saúde: Allam e Jones (2020) e Ramaswami et al. (2016), trânsito: Djahel et al. (2014) e Barba et al. (2012) e segurança: Elmaghraby e Losavio (2014).

A maioria das propostas de Cidades Inteligentes abrange principalmente quatro características fundamentais: sustentabilidade, qualidade de vida, urbanização e inteligência. A sustentabilidade busca aprimorar aspectos como: infraestrutura e governança, questões ambientais, sociais, econômicas e de saúde. Melhorar a qualidade de vida envolve o bem-estar mental e financeiro dos habitantes. Urbanização considera aspectos tecnológicos, econômicos e de infraestrutura. A inteligência visa melhorar aspectos ambientais, sociais e econômicos dos seus habitantes (MOHANTY; CHOPPALI; KOUIGIANOS, 2016). Essas informações são demonstradas na Figura 3.

2.2.1 Saúde Inteligente

O propósito da Saúde Inteligente é disponibilizar cuidados médicos aos pacientes de forma acessível, a qualquer momento e em qualquer lugar. Para isso, os sistemas de monitoramento da Saúde Inteligente utilizam, predominantemente, conexões por redes sem fio (BUTT et al., 2019), sendo possível realizar o monitoramento ágil de pacientes com condições de saúde anormais. Por meio de um sistema de assistência médica inteligente, é possível acompanhar essas condições em tempo real e oferecer soluções rápidas aos pacientes (ANANTH; SATHYA; MOHAN, 2019).

A abordagem da Saúde Inteligente traz melhorias significativas ao integrar acessibilidade, qualidade, redução de custos e inteligência ao sistema de saúde. Ela permite que pacientes, independentemente de sua localização, acessem serviços, equipamentos remotos e especialistas médicos, enquanto profissionais podem monitorar continuamente as condições dos pacientes.

Figura 3 – Características das Cidades Inteligentes



Fonte: Produção dos autores, 2024.

Além disso, melhora a qualidade do atendimento ao reduzir o tempo de resposta de emergências por meio de monitoramento contínuo e algoritmos avançados (ABDEEN et al., 2021).

Na Saúde Inteligente, o foco é ajudar em desafios como o aumento de doenças crônicas, promovendo o monitoramento contínuo da saúde e a detecção precoce de problemas. Tecnologias como prontuários eletrônicos e dispositivos da Internet das Coisas (*Internet of Things* - IoT), para medir frequência cardíaca e glicose, são componentes que podem auxiliar em soluções inteligentes na área da saúde (DIMITROV, 2016). Como apresentado anteriormente, o bem-estar mental é uma das prioridades nas Cidades Inteligentes, existindo diversos trabalhos no ramo da Saúde Inteligente que abordam a saúde mental, inclusive a depressão (DOGRUCU et al., 2020; KIM; CHUNG, 2017; BOUKHECHBA et al., 2018).

2.2.2 Promoção da Saúde Mental em Cidades Inteligentes

O crescimento das cidades mudou o estilo de vida das pessoas e essas alterações no ambiente urbano têm um impacto na segurança, saúde e outros aspectos dos residentes (CHEN et al., 2024). As Cidades Inteligentes contribuem para a diminuição de problemas urbanos (SHAHAZI et al., 2018), utilizando tecnologias como: IA, Aprendizado de Máquina e Internet das Coisas. Diversos dados são coletados e trabalhados para gerar *insights* que podem ajudar a melhorar a qualidade de vida, desenvolvimento socioeconômico e saúde (THOMAS et al., 2022), onde a saúde mental é uma das maiores preocupações (CHAKRABORTY et al., 2023; MOHANTY; CHOPPALI; KOUGIANOS, 2016).

A aplicação de IA no setor de saúde é revolucionária (BAJAJ; MAHESH, 2024), na área

da saúde mental tem-se um interesse na adoção da tecnologia para aprimorar processos, desde o diagnóstico, monitoramento e tratamento (ANTONIOU; PAPADAKIS; BARYANNIS, 2022). Com o Aprendizado de Máquina, é possível processar um enorme volume de dados, criando reações automatizadas, em tempo real, e quando apropriado fornecer análises (SORNLERTLAM-VANICH et al., 2023). Os sistemas automatizados de diagnóstico auxiliam os profissionais da área de saúde no processo de tomada de decisão (MODY; MODY, 2019; MENDIS et al., 2022).

Existem diversos estudos envolvendo IA aplicada à saúde mental em Cidades Inteligentes: em Yadav, Jin e Stevano (2019) foi desenvolvida uma tecnologia para auxiliar na saúde mental, fazendo uso de realidade aumentada, Aprendizado de Máquina e sensores da Internet das Coisas Médicas (*Internet of Medical Things* - IoMT); o estudo de Wang (2023) concentra-se no uso do Aprendizado de Máquina para analisar as mudanças psicológicas e a adaptabilidade de estudantes universitários; em Li et al. (2019) foi gerado um modelo de Aprendizado de Máquina para avaliação de crises psicológicas.

2.3 Aprendizado de Máquina

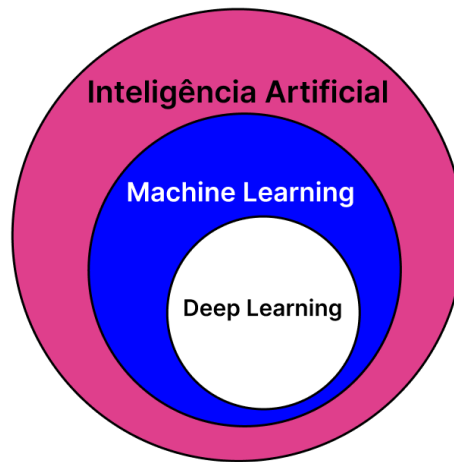
Antes de abordar o Aprendizado de Máquina, é interessante descrever o campo da computação onde ele está contido: a Inteligência Artificial, essa que é a automação do processo cognitivo (ABBASS, 2021). A IA tem impacto na economia, melhora padrões de vida e promove interação inteligente no sistema humano-máquina-ambiente (WANG, 2022).

O Aprendizado de Máquina é baseado em algoritmos treinados para a tomada de decisões que aprendem com os dados analisados (JORDAN; MITCHELL, 2015). O Aprendizado Profundo, do inglês *Deep Learning* (DL), é um subgrupo do Aprendizado de Máquina baseado em algoritmos que usam Redes Neurais Artificiais (RNA) organizadas em várias camadas para interpretar e tirar conclusões das informações (BELLINI et al., 2022). A relação entre Inteligência Artificial, Aprendizado de Máquina e Aprendizado Profundo é representada na Figura 4.

Alan Turing, em 1950, reconheceu a importância de desenvolver programas capazes de aprender (MUGGLETON, 2014). Humanos aprendem e se adaptam e a Aprendizagem de Máquina busca replicar isso para computadores (JAPKOWICZ; SHAH, 2011). Esse processo resulta em tomadas de decisão mais assertivas em diversas áreas, como saúde, educação, finanças, segurança pública e estratégias de mercado (JORDAN; MITCHELL, 2015). A máquina aprende por si mesma, absorvendo conhecimento da interação com dados, dispensando uma programação direta. Com o tempo, esses sistemas discernem os padrões e tendências subjacentes nos dados. Uma vez treinados, conseguem antecipar com precisão eventos futuros (SHAHRIAR et al., 2020).

Machine Learning é dividido em três principais categorias: supervisionado, que usa dados rotulados para treinamento; não supervisionado, onde não há orientação e a inteligência encontra padrões em dados não rotulados; e a Aprendizagem por Reforço, onde um agente interage com um ambiente para maximizar recompensas a longo prazo (ULLAH et al., 2020).

Figura 4 – Inteligência Artificial e Subáreas



Fonte: Produção dos autores, 2024.

O Aprendizado Supervisionado é uma técnica comum em problemas de classificação, onde o objetivo é treinar um modelo para prever uma saída com base em entradas fornecidas. O processo de aprendizagem pode ser dividido em duas etapas: treinamento e teste. No treinamento, o modelo aprende a partir de dados rotulados, comparando suas projeções com as saídas corretas e ajustando-se para melhorar os resultados. No teste, o modelo aplica o que aprendeu para prever os dados (NASTESKI, 2017).

2.3.1 *Overfitting*

Overfitting é um problema relevante quando se trata de Aprendizado de Máquina, ocorrendo quando um algoritmo se ajusta tão bem ao conjunto de dados de treinamento que acaba aprendendo ruídos e propriedades irrelevantes dos dados de treinamento. Como resultado, o desempenho do algoritmo de aprendizagem diminui no conjunto de dados de teste (POURTAHERI; ZAHIRI, 2016). O *overfitting* pode ser detectado por uma estimativa da precisão do modelo no teste em comparação com sua precisão nos dados de treinamento (KHOSHGOFTAAR; ALLEN; DENG, 2001).

2.3.2 Validação Cruzada

A Validação Cruzada é uma técnica de avaliação de modelos de Aprendizado de Máquina usada para testar a capacidade de generalização de um modelo e evitar *overfitting*. A ideia principal da Validação Cruzada é dividir o conjunto de dados disponível em várias partes (ou dobras), de forma que cada parte seja usada tanto para treinamento quanto para teste em diferentes iterações. Na Validação Cruzada, *10-Fold Cross Validation*, os dados são divididos em 10 partes, e o modelo é treinado e testado em diferentes combinações dessas partes. Em cada ciclo, uma

parte é usada para testar o modelo, enquanto as outras 9 partes servem para o treinamento. Esse processo é repetido até que cada parte seja usada como conjunto de teste (WEN; ARIS, 2022).

2.3.3 *One-hot Encoding*

A técnica *One-hot encoding* é usada para transformar variáveis categóricas em dados numéricos, podendo ser usada em modelos de Aprendizado de Máquina. Quando uma variável categórica possui várias categorias distintas, o *One-hot encoding* cria uma nova coluna para cada categoria. Para cada registro, a coluna correspondente à categoria presente recebe o valor 1, e as outras colunas recebem 0 (POTDAR; PARDAWALA; PAI, 2017).

2.3.4 Normalização

A normalização padroniza os valores numéricos em uma escala comum, evitando que valores de recursos numéricos maiores possam dominar os valores de recursos numéricos menores. Isso ajuda a melhorar o desempenho dos algoritmos de classificação (SINGH; SINGH, 2020). Existem diversas técnicas de normalização, dentre elas, a técnica normalização *Min-Max*, que realiza uma transformação linear dos dados originais para um intervalo pré-definido, ajustando os dados dentro de uma faixa específica, preservando o relacionamento entre os dados originais (PATRO, 2015).

2.3.5 *Synthetic Minority Over-sampling Technique (SMOTE)*

Synthetic Minority Over-sampling Technique, é uma técnica que envolve criar exemplos sintéticos de classes minoritárias para aumentar sua quantidade no conjunto de dados. Essa técnica ajuda a equilibrar conjuntos de dados desequilibrados, permitindo que o classificador tenha mais exemplos da classe minoritária para aprender (CHAWLA et al., 2002).

2.3.6 *Logistic Regression (LR)*

A Regressão Logística (*Logistic Regression* - LR) é uma técnica de Aprendizado de Máquina Supervisionada, normalmente utilizada em problemas de classificação, quando a variável alvo é categórica. O objetivo é prever a probabilidade de um exemplo pertencer a uma classe, mapeando uma função que relaciona as características do conjunto de dados aos resultados. (BISONG; BISONG, 2019). Segundo Srinath et al. (2022), os principais tipos são:

a) Regressão Logística Binomial (*Binomial Logistic Regression* - BLR): Existem apenas dois possíveis tipos de variáveis de resposta disponíveis, como falso ou verdadeiro, 0 ou 1. A modelagem é baseada na função de distribuição de probabilidade binomial, que calcula a probabilidade de um exemplo pertencer a uma dessas classes.

b) Regressão Logística Multinomial (*Multinomial Logistic Regression* - MLR): Quando há mais de dois tipos de resposta possíveis, como 0, 1 ou mais, o problema passa a ser de classificação multiclasse. Para aplicar a *Logistic Regression* nesse contexto, o problema multiclasse é dividido em vários subproblemas de classificação binária. Em cada subproblema, o modelo de regressão logística padrão é configurado para distinguir entre duas classes. No caso de conjuntos de dados com múltiplas classes, utiliza-se a regressão logística multinomial.

2.3.7 *Support Vector Machine* (SVM)

A Máquina de Vetores de Suporte (*Support Vector Machine* - SVM) é um método de Aprendizado de Máquina baseado na teoria de aprendizado estatístico, que pode ser usado para classificar dados (YANG; LI; YANG, 2015). A técnica SVM, inicialmente introduzida por Vapnik (1999), é utilizada para reconhecimento de padrões (CERVANTES et al., 2020).

Na configuração mais simples, um SVM gera, ao receber um conjunto de dados de treinamento, um classificador que encontra um hiperplano de margem máxima que separa os dados (GRØNLUND; KAMMA; LARSEN, 2020). O SVM em sua versão original requer um algoritmo de aprendizado supervisionado baseado em amostras rotuladas limitadas para aprender as regras e estendê-las a amostras desconhecidas sem *tags* (DEQIAN, 2011).

2.3.8 *Random Forest* (RF)

O conceito principal da Floresta Aleatória (*Random Forest* - RF) é combinar várias árvores de decisão geradas aleatoriamente para melhorar a precisão do modelo preditivo. Em vez de otimizar um único preditor, como em uma árvore de decisão tradicional, a Floresta Aleatória utiliza um conjunto de preditores, o que aumenta a capacidade de exploração do espaço de possíveis soluções, melhorando o desempenho do modelo (GENUER et al., 2020). Um algoritmo flexível com uma ampla gama de aplicações e que desempenha bem em grande volume de conjuntos de dados, com alta dimensionalidade e valores ausentes, podendo ser utilizado em problemas de classificação e regressão (ZHU, 2020; IZQUIERDO-VERDIGUIER; ZURITA-MILLA, 2020).

2.3.9 *K-Nearest Neighbors* (KNN)

Destaca-se como um dos algoritmos mais conhecidos em reconhecimento de padrões para classificação supervisionada não paramétrica. Sua popularidade se deve à simplicidade conceitual, baseada na comparação de distâncias entre elementos (CALVO-ZARAGOZA; VALERO-MAS; RICO-JUAN, 2015). O K-Vizinhos Mais Próximos (*K-Nearest Neighbors* - KNN) funciona comparando amostras desconhecidas com amostras de treinamento já conhecidas.

É baseado em instâncias que classifica uma amostra com base em seus *k* vizinhos mais próximos. A proximidade é determinada por uma métrica de distância, podendo ser utilizada a

distância euclidiana para essa finalidade. Técnicas como redução de dimensionalidade ajudam a melhorar seu desempenho (CUNNINGHAM; DELANY, 2021).

2.3.10 *Gaussian Naive Bayes* (GNB)

O *Gaussian Naive Bayes* (GNB) é um algoritmo de classificação simples que prevê a qual classe uma amostra pertence, com base em suas características. O GNB assume que as características são independentes entre si e seguem uma distribuição gaussiana. O algoritmo calcula a probabilidade de cada classe para a amostra usando o Teorema de *Bayes*. Em seguida, o GNB atribui à amostra o rótulo da classe com a maior probabilidade posterior, ou seja, a classe mais provável (KAMEL; ABDULAH; AL-TUWAIJARI, 2019). O algoritmo é altamente escalável. Entretanto, uma fraqueza do GNB é a suposição de independência entre preditores, dificilmente válida na realidade (AMPOMAH et al., 2021).

2.3.11 *Decision Tree* (DT)

O algoritmo Árvore de Decisão (*Decision Tree* - DT) é um algoritmo simples que organiza os dados em subgrupos homogêneos (THOMAS; VIJAYARAGHAVAN; EMMANUEL, 2019). Ainda, possui um nó raiz, ramos e nós folhas. O nó raiz é o nó superior da árvore e o ponto de partida para todos os outros nós. Cada nó representa um atributo (característica), cada ramo simboliza uma decisão (regra) e cada folha indica o resultado (PATEL; PRAJAPATI, 2018).

O algoritmo seleciona, em cada nó, o atributo que melhor divide o conjunto de dados em subconjuntos mais homogêneos, direcionando-os a uma categoria específica. Sua estrutura é hierárquica, organizada por condições que começam na raiz da árvore e avançam até as folhas. Cada folha representa um rótulo de classe, definido pelas condições atendidas ao longo do percurso (ROMERO et al., 2008).

2.3.12 *Adaptive Boosting* (AdaBoost)

AdaBoost é uma técnica de Aprendizado de Máquina projetada para melhorar o desempenho de classificadores fracos, combinando-os em um classificador forte (SHEN; LI, 2010). Os pesos são atribuídos a todas as instâncias, sendo maiores para aquelas situações difíceis, de modo que o modelo foque mais nelas (SUDHARSON et al., 2021).

O algoritmo combina várias hipóteses fracas para criar um modelo preditivo robusto e altamente preciso. O foco central do *AdaBoost* é ajustar o peso de cada hipótese fraca com base em sua performance, dando mais importância a exemplos difíceis de classificar corretamente. O algoritmo tradicional trabalha com previsões binárias, mas em Schapire (2013) o conceito é expandido, onde ocorrem melhorias no uso de valores de confiança nas previsões das hipóteses, permitindo que cada uma não apenas forneça um rótulo, mas também uma medida de confiança na sua precisão.

2.3.13 Métricas de Qualidade

As métricas de qualidade são relevantes para avaliar o desempenho de modelos de Aprendizado de Máquina. Métricas comuns para avaliação de problemas de classificação incluem: acurácia (*accuracy*), precisão (*precision*), sensibilidade (*recall*) e medida-F1 (*F1-score*) (VAKILI; GHAMSARI; REZAEI, 2020). Os termos utilizados nessas métricas são apresentados a seguir:

- VP (Verdadeiro Positivo): O modelo acertou ao prever a classe positiva.
- VN (Verdadeiro Negativo): O modelo acertou ao prever a classe negativa.
- FP (Falso Positivo): O modelo errou, prevendo positivo quando era negativo.
- FN (Falso Negativo): O modelo errou, prevendo negativo quando era positivo.

Por meio da acurácia ou *accuracy*, é possível identificar o percentual de instâncias que foram classificadas de forma correta.

$$\text{acurácia} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FN} + \text{VN} + \text{FP}}$$

Quanto à sensibilidade ou *recall*, a métrica mede a capacidade do modelo de identificar corretamente as instâncias da classe positiva.

$$\text{sensibilidade} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

Em relação à precisão ou *precision*, a métrica indica a proporção de previsões positivas corretas em relação ao total de previsões feitas como positivas.

$$\text{precisão} = \frac{\text{VP}}{\text{VP} + \text{FP}}$$

A medida-F1 ou *F1-score* combina duas métricas, sendo a média harmônica entre a precisão e a sensibilidade.

$$\text{medida-F1} = \frac{2 \times \text{precisão} \times \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}}$$

3

Materiais e Métodos

A depressão é um grande problema social e soluções em Saúde Inteligente que utilizam Aprendizado de Máquina podem viabilizar ferramentas úteis para gerar indicadores e consequentemente auxiliar no tratamento da doença. Ainda, o DASS é uma ferramenta confiável que pode auxiliar na mensuração da depressão ([CRAWFORD et al., 2011](#)).

Nesse contexto, foi realizado um Mapeamento Sistemático da Literatura (MSL), para encontrar trabalhos relacionados a soluções inteligentes que fizessem uso de Aprendizado de Máquina em conjunto com a escala DASS, onde dois estudos tiveram destaque: [Srinath et al. \(2022\)](#) e [Singh et al. \(2022\)](#), esses que fizeram uso do mesmo *dataset* apresentado na [Seção 3.1](#), com respostas ao DASS-42 e informações sociodemográficas.

Mediante a leitura dos artigos, foram identificadas lacunas relevantes, com possibilidade de gerar avanços no desenvolvimento de Sistemas Inteligentes para depressão:

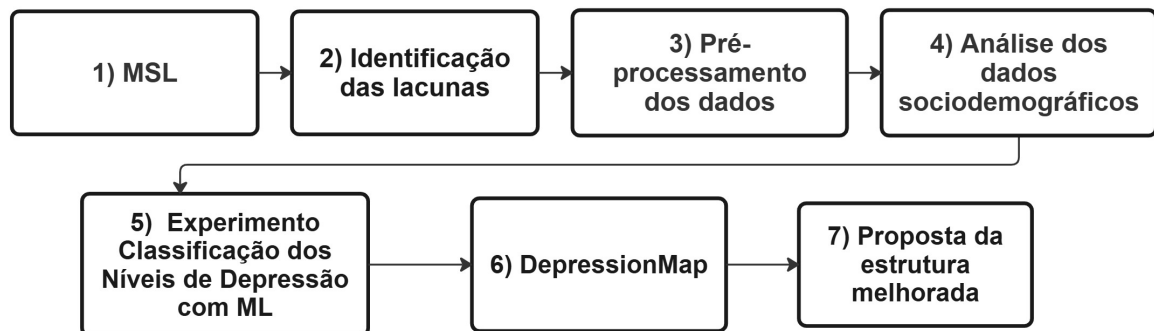
- 1) Explorar a relação entre níveis de depressão e dados sociodemográficos;
- 2) Analisar o desempenho dos algoritmos de Aprendizado de Máquina na classificação dos níveis de severidade da depressão com um pré-processamento diferenciado.
- 3) Propor uma estrutura melhorada com participação de profissionais da saúde mental.

Para a implementação do sistema, ocorreu um pré-processamento dos dados, no qual foi realizada a limpeza dos dados e o cálculo dos níveis de depressão. Posteriormente, foi realizada a análise dos dados sociodemográficos, onde ocorreu um levantamento das informações sociodemográficas e das suas relações com os níveis de depressão, nessa etapa os resultados foram apresentados por meio de um mapa de calor, além de gráficos e tabelas.

Ainda foi realizada, por meio de um experimento, a classificação dos níveis de depressão em cinco categorias: normal, leve, moderado, severo e extremamente severo, o procedimento foi realizado utilizando técnicas de Aprendizado de Máquina. Os resultados da análise dos dados sociodemográficos e da classificação foram disponibilizados *online* por meio do *DepressionMap*.

Por fim, foi elaborada a proposta de uma estrutura melhorada do *DepressionMap*, com a participação de profissionais de saúde. A nova abordagem proposta inclui coleta de dados com informações pessoais, retorno sobre os indicadores de depressão e um painel para auxiliar a equipe médica na triagem e acompanhamento dos indivíduos depressivos. Na [Figura 5](#), são apresentadas as etapas da metodologia.

Figura 5 – Etapas da Metodologia



Fonte: Produção dos autores, 2024.

Em relação à classificação da pesquisa, quanto à natureza, pode ser classificada como aplicada, pois produziu conhecimento com o objetivo de aplicar seus resultados na solução de problemas práticos. No que se refere à abordagem dos dados, é considerada quantitativa, pois os dados foram analisados de forma estatística. Em relação ao objetivo da pesquisa, pode ser considerada de caráter exploratório e descritivo. A pesquisa exploratória visa fornecer uma maior compreensão do problema, ajudando a torná-lo mais claro. Enquanto a pesquisa descritiva tem como objetivo retratar os acontecimentos e especificações de uma realidade específica ([GERHARDT; SILVEIRA, 2009](#)). O procedimento adotado é experimental, pois envolve identificar um objeto de estudo, selecionar as variáveis que podem influenciá-lo, estabelecer métodos para controlar essas variáveis e observar os efeitos que elas causam no objeto analisado ([GIL, 2002](#)).

Segundo a Resolução nº 510, de 07 de abril de 2016 ([Conselho Nacional de Saúde, 2016](#)), não serão registradas nem avaliadas pelo sistema CEP/CONEP do Conselho Nacional de Saúde pesquisa com bancos de dados, cujas informações são agregadas, sem possibilidade de identificação individual. A presente pesquisa utilizou um *dataset* com dados secundários anônimos, sem possibilidade de identificação individual, disponibilizados publicamente, como detalhado na [Seção 3.1](#).

3.1 Base de Dados

Para o desenvolvimento do *DepressionMap* foi utilizado um *dataset* disponibilizado publicamente pela *Openpsychometrics*¹, possuindo respostas anônimas de 39.775 participantes. Esses dados foram coletados entre 2017 e 2019.

No total, o conjunto de dados possui 172 colunas, contendo respostas ao DASS-42, itens do Inventário de Personalidade de Dez Itens, do inglês *Ten Item Personality Inventory*, perguntas sobre definições de palavras (algumas fictícias, usadas para verificar a validade das respostas do participante). Além disso, o conjunto de dados inclui dados sociodemográficos, como: escolaridade, em que tipo de área morava quando era criança, gênero, se o inglês é a língua nativa, idade, mão dominante, religião, orientação sexual, raça, situação eleitoral, estado civil, tamanho da família e ocupação.

Ainda, foram registrados automaticamente pelo sistema, dados como: tempo dedicado à introdução da pesquisa, tempo para responder ao DASS-42, tempo para concluir o restante da pesquisa, país de conexão, tamanho da tela e origem do acesso (categorizada como página inicial, *Google* ou outras fontes).

3.2 Ferramentas

O desenvolvimento do *DepressionMap* ocorreu na plataforma *Google Colab*², usando a linguagem *python*, as bibliotecas utilizadas para a programação do sistema, bem como as versões e endereços eletrônicos são apresentados abaixo na [Tabela 2](#).

Tabela 2 – Bibliotecas, Versão e Endereço Eletrônico

Biblioteca	Versão	Endereço eletrônico
<i>NumPy</i>	1.26.4	< https://numpy.org/news/#releases >
<i>Pandas</i>	2.2.2	< https://pandas.pydata.org/ >
<i>Streamlit</i>	1.41.1	< https://streamlit.io/ >
<i>Matplotlib</i>	3.9.0	< https://matplotlib.org/ >
<i>Plotly</i>	5.22.0	< https://pypi.org/project/plotly/ >
<i>GeoJSON</i>	3.1.0	< https://pypi.org/project/geojson/ >
<i>Folium</i>	0.17.0	< https://python-visualization.github.io/folium/ >
<i>Scikit-learn</i>	1.6.0	< https://scikit-learn.org/ >
<i>Imbalanced-learn</i>	0.13.0	< https://imbalanced-learn.org/ >
<i>Autorank</i>	1.2.1	< https://github.com/sherbold/autorank >

Fonte: Produção dos autores, 2024.

¹ <<https://openpsychometrics.org/>>

² <<https://colab.research.google.com/>>

4

Trabalhos Relacionados

Esta etapa descreve uma investigação direcionada à busca de trabalhos relevantes relacionados ao tema em questão. Assim, conduzindo uma pesquisa exploratória e utilizando uma abordagem de Mapeamento Sistemático da Literatura (MSL). A pesquisa foi conduzida em março de 2024. Durante a execução dessas fases, foi utilizada a plataforma *online Parsifal* (versão 2.2) ¹. O presente trabalho adotou a diretriz apresentada no relatório [Kitchenham \(2004\)](#), que foi criada para atender às demandas e particularidades dos problemas de pesquisa em Engenharia de Software. Os procedimentos abrangem as três fases críticas de um mapeamento sistemático: planejamento, execução e, por fim, análise e divulgação.

4.1 Planejamento

Essa fase destaca a importância de formular as perguntas certas para a área de interesse ([BRERETON et al., 2007](#)). Após a esquematização do planejamento, o protocolo será utilizado para guiar a condução do mapeamento, com o principal propósito de responder às questões de pesquisa conforme detalhadas no [Quadro 2](#).

Quadro 2 – Questões de Pesquisa

ID da Questão	Questões de Pesquisa
Q1	Quais técnicas de IA foram utilizadas?
Q2	O estudo leva em consideração a participação de profissionais de saúde para o diagnóstico da depressão?
Q3	Qual o ano de publicação do trabalho?

Fonte: Produção dos autores, 2024.

Com o objetivo de fornecer respostas às questões de pesquisa, foram estabelecidos campos a serem extraídos, conforme demonstrado no [Quadro 3](#). Após a leitura dos artigos, foram obtidos

¹ <https://www.parsif.al>

os dados correspondentes de cada estudo analisado.

Quadro 3 – Campos de Extração

Campos de extração
Técnicas de IA utilizadas
Profissionais de saúde
Ano de Publicação

Fonte: Produção dos autores, 2024.

Os critérios para escolha das bases científicas foram os seguintes: as bases precisavam estar disponíveis online, oferecer conteúdo das áreas relacionadas ao tema: Ciência da Computação e Saúde, mais especificamente Psicologia. Após estabelecer esses parâmetros, foram identificadas quatro bases científicas, apresentadas no [Quadro 4](#). Essas foram: **IEEE Xplore**, **Scopus**, **PubMed** e **LILACS**.

O [Quadro 5](#) apresenta os Critérios de Inclusão (CI) e os Critérios de Exclusão (CE) definidos para a seleção dos estudos na pesquisa. Os Critérios de Inclusão consideram estudos primários, disponíveis integralmente em bases de pesquisa, e escritos em português ou inglês, estudos com foco em soluções tecnológicas baseadas no DASS para o tratamento da depressão e artigos publicados entre 2017 e 2024, visando selecionar estudos recentes, respeitando assim os avanços tecnológicos. Já os Critérios de Exclusão eliminam estudos duplicados, revisões sistemáticas, publicações fora do período estabelecido, documentos irrelevantes ao tema e aqueles indisponíveis integralmente nas bases de dados analisadas.

Com base nas palavras-chave que remetem ao tema proposto, apresentadas no [Quadro 6](#), foi formulada uma *string* de busca, utilizada como um padrão para conduzir a pesquisa de maneira uniforme em todas as bases de dados. Essa sequência específica está detalhada na [Quadro 7](#).

4.2 Execução

A fase de execução do mapeamento sistemático de [Kitchenham \(2004\)](#) tem o objetivo de buscar e sintetizar os trabalhos relevantes ao tema, sendo estabelecida em 5 etapas, representadas no fluxograma da [Figura 6](#), onde o processo começa com a etapa de identificação, uma busca com a *string* foi conduzida em quatro bases de pesquisa, ocorrendo o retorno de 105 estudos; Scopus (33), IEEE Xplore (29), PubMed (38) e Lilacs (5), conforme apresentado na [Figura 7](#).

Em seguida, na etapa de seleção, foram aplicados Critérios de Inclusão e Exclusão com base em títulos e resumos para eliminar estudos que não atendiam aos requisitos pré-definidos. Foram eliminados 7 estudos duplicados e 79 foram excluídos pelos outros Critérios de Exclusão, restando 19 artigos. Essa quantidade se manteve depois da etapa de avaliação, onde foi realizada a leitura completa dos estudos. A [Figura 8](#) mostra a distribuição dos artigos selecionados em relação às bases de pesquisa.

Quadro 4 – Bases de Pesquisa e URL

Palavras-Chave	Termos em inglês
IEEE Xplore	https://ieeexplore.ieee.org/Xplore/home.jsp
Scopus	https://www.scopus.com/home.uri
PubMed	https://pubmed.ncbi.nlm.nih.gov/
LILACS	https://lilacs.bvsalud.org/en/

Fonte: Produção dos autores, 2024.

Quadro 5 – Critérios de Inclusão e Exclusão

Critérios de Inclusão	Critérios de Exclusão
CI1 - Estudos que apresentem soluções tecnológicas para auxiliar no tratamento da depressão com o DASS	CE1- Estudos duplicados
CI2 - Estudos primários	CE2 - Estudos secundários publicados como mapeamento sistemática
CI3 - Escritos em português ou inglês	CE3 - Estudos que não estejam escritos em inglês ou português
CI4 - Estudos publicados entre 2017 a 2024	CE4 - Estudos que não tenham relevância para o projeto
CI5 - Estudos disponíveis integralmente nas bases de pesquisas	CE5 - Estudos publicados fora da faixa de 2017 e 2024
	CE6 - Estudos que não estejam disponíveis integralmente em bases de pesquisa

Fonte: Produção dos autores, 2024.

Quadro 6 – Lista de Palavras-Chave e Tradução em Inglês

Palavras-Chave	Tradução em inglês
Cidade Inteligente	Smart City
DASS	DASS
Escala de Depressão, Ansiedade e Estresse	Depression, Anxiety, and Stress Scale
Aprendizado de Máquina	Machine Learning
DASS-42	DASS-42

Fonte: Produção dos autores, 2024.

Após a seleção dos 19 estudos, foi realizada a avaliação dos estudos, por meio de uma leitura completa. A extração dos dados e sínteses também foram realizadas. Assim, cada questão de pesquisa do [Quadro 2](#) foi abordada e solucionada com base nas informações obtidas. A seguir, os resumos dos artigos:

A pesquisa de [Brown et al. \(2024\)](#) mostrou que a ligação entre os jogadores e seus avatares pode fornecer informações sobre o risco de depressão. Foram utilizados dados de 565 indivíduos. O algoritmo da Floresta Aleatória foi o método mais eficaz, enquanto a imersão de avatar foi um preditor de treinamento relacionado à depressão.

Quadro 7 – String de Busca

<i>String</i>
(“DASS” OR “DASS-42” OR “DEPRESSION ANXIETY STRESS SCALES”) AND (“MACHINE LEARNING” OR “SMART CITY”)

Fonte: Produção dos autores, 2024.

O trabalho [Azmi, Fathima e Wasid \(2023\)](#) investiga a detecção de emoções por meio da análise da caligrafia em tarefas diárias, explorando a relação entre estados emocionais e a escrita. Utilizando dados de 129 usuários, um modelo de Floresta Aleatória é desenvolvido para prever estados emocionais com base nas características da escrita.

No estudo de [Priya, Garg e Tigga \(2020\)](#), utilizou-se Aprendizado de Máquina para prever ansiedade, depressão e estresse em diferentes níveis de gravidade, analisando dados coletados de pessoas empregadas e desempregadas. Os dados foram coletados fazendo uso do DASS-21. Foram utilizadas cinco técnicas de classificação diferentes: Árvore de Decisão, Floresta Aleatória, *Naive Bayes*, SVM e KNN.

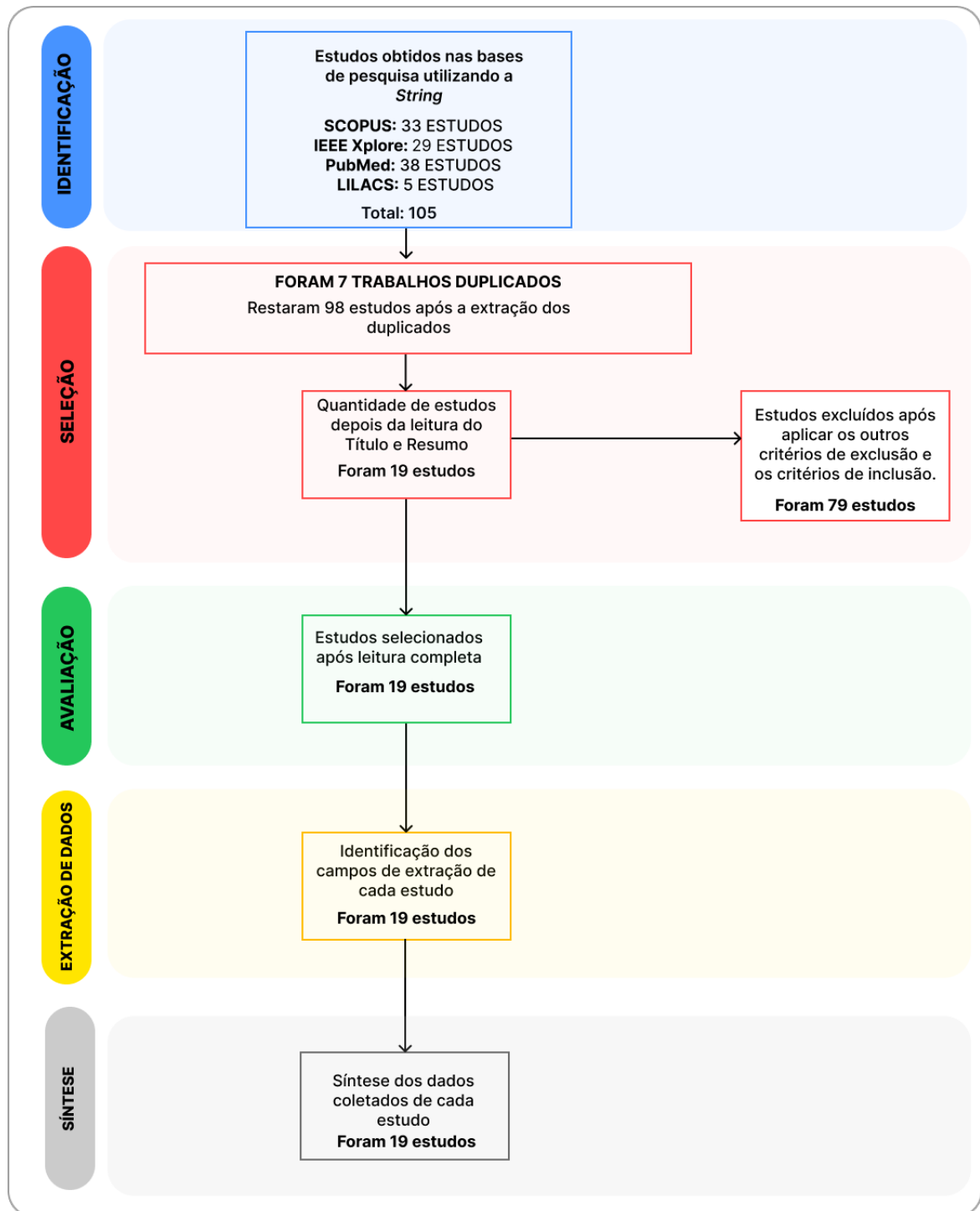
No artigo de [Maskat et al. \(2021\)](#), aborda-se a Previsão Multi-alvo (*Multi-Target Prediction* - MTP), uma técnica voltada para a predição simultânea de múltiplos resultados. Essa abordagem tem aplicação em diversos campos, como previsão do tempo, interações em redes sociais e ecossistemas. O estudo propõe a aplicação do MTP para antecipar os níveis de depressão, ansiedade e estresse, utilizando dados da escala DASS-21 versão Malásia. Ao utilizar algoritmos de Aprendizado de Máquina já estabelecidos para prever a gravidade de cada categoria (ou seja, depressão, ansiedade e estresse), os resultados do estudo revelam uma abordagem mais eficaz ao incorporar a matriz de custos como um elemento secundário no contexto do MTP.

O estudo [Sun et al. \(2022\)](#) empregou dados do DASS-42 de 31.715 participantes para criar modelos de Aprendizado de Máquina para prever o estado de depressão. Foi identificado que apenas sete itens da escala são suficientes para atingir pelo menos 90% de acurácia, utilizando a técnica de generalização empilhada multimodelo. O modelo foi integrado em uma ferramenta online para Avaliação Rápida de Depressão, oferecendo autoavaliações práticas e rápidas, com duração aproximada de um minuto.

O estudo de [Al-Wesabi et al. \(2022\)](#) visa compreender o DAS com foco em estudantes universitários. Ele apresenta o modelo de Seleção Inteligente de Subconjuntos de Recursos com DAS Baseado em Aprendizado de Máquina (*Intelligent Feature Subset Selection with Machine Learning-based DAS* - IFSSML-DAS). O modelo analisa o DAS utilizando um conjunto de dados DASS-21, sugerindo a necessidade de intervenções personalizadas para abordar estas preocupações de saúde mental, a fim de superar a crise da Doença por Coronavírus 2019 (COVID-19).

Já no estudo de [Singh et al. \(2022\)](#) foi apresentada uma comparação entre diversos métodos

Figura 6 – Fluxograma das Etapas da Execução

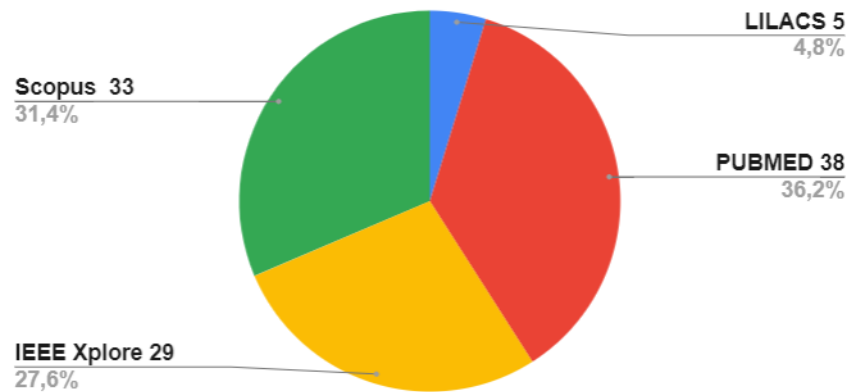


Fonte: Produção dos autores, 2024.

de Aprendizado de Máquina na tarefa de classificação de estados emocionais negativos, utilizando dados relacionados ao DASS-42 e informações sociodemográficas. A Floresta Aleatória, Árvore de Decisão, SVM, *AdaBoost*, *Categorical Boosting (CatBoost)* e *XGBoost* foram usados para classificar os níveis de Depressão, Ansiedade e Estresse.

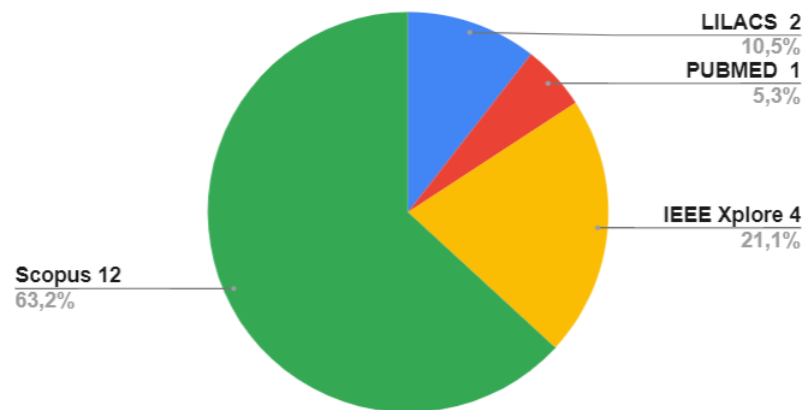
O estudo de [Swain e Patra \(2022\)](#) apresenta uma comparação entre diversos métodos de

Figura 7 – Quantidade de Trabalhos Encontrados por Base



Fonte: Produção dos autores, 2024.

Figura 8 – Quantidade de Trabalhos Seleccionados por Base



Fonte: Produção dos autores, 2024.

Aprendizado de Máquina, utilizando a escala DASS-21 e informações demográficas. No artigo, observou-se que o método de SVM foi o método mais eficaz para a classificação dos estados emocionais.

Este estudo [Budiyanto, Sihombing e IM \(2019\)](#) tem como objetivo explorar o impacto das interações nas redes sociais na detecção precoce de sinais de depressão e ansiedade. Utilizando técnicas de mineração de texto e Aprendizado de Máquina, especificamente o modelo *Naive Bayes*, busca identificar tendências e padrões que possam indicar a presença de depressão e ansiedade nas interações *online*.

Este estudo [Fatima et al. \(2021\)](#) apresenta o DASentimental, um modelo de Aprendizado de Máquina para detectar depressão, ansiedade e estresse de textos escritos. Foi utilizada a Escala de Depressão, Ansiedade e Estresse (DASS-21) como base, envolvendo amostra de informações de 200 indivíduos. Ao incorporar ciência de redes cognitivas, o DASentimental analisa redes semânticas, utilizando a distância entre palavras-chave e sua centralidade na rede para prever os níveis de depressão, ansiedade e estresse.

Neste estudo [Ghorpade-Aher et al. \(2023\)](#) foram analisados: ansiedade, estresse e depressão em adultos. Com 39.775 respostas coletadas entre 2017 e 2019, foram explorados os estados emocionais usando algoritmos de Aprendizado de Máquina. A técnica SVM teve um desempenho máximo preditivo de 96%. O modelo foi melhorado após a utilização de algoritmos de DL.

Neste trabalho [Jha, Abirami e Kumar \(2022\)](#), alguns algoritmos de Aprendizado de Máquina foram usados para prever a prevalência de problemas psicológicos, como depressão e ansiedade, usando DASS-42 e o *dataset* chamado de Detecção de Estresse e Afeto Vestível (*Wearable Stress and Affect Detection* - WESAD). Os pesquisadores do artigo chegaram à conclusão de que o trabalho teve um resultado significativo em termos de precisão quando comparado a outros trabalhos do gênero.

Este estudo [Kumar, Chandra et al. \(2023\)](#) aborda a identificação precoce de problemas de saúde mental. Comparando as percepções e níveis de ansiedade, estresse e depressão entre praticantes de *Rajyoga* e não praticantes. Utiliza o DASS-21. Além disso, utiliza algoritmos de Aprendizado de Máquina para prever DAS. O SVM apresentou os maiores resultados.

Este estudo [Jalandra et al. \(2022\)](#) investigou o estresse, a ansiedade e a depressão entre profissionais de saúde. O estudo envolveu 156 participantes em vários ambientes de atendimento à COVID-19 em hospitais indianos. Ele utilizou o DASS-21 e um algoritmo de Aprendizado de Máquina, *CatBoost*. Fatores como tempo pessoal limitado, idade, experiência de trabalho externa e ser médico residente impactaram significativamente os níveis de estresse.

No artigo de [Sun, Luo e Lee \(2022\)](#), foi demonstrado como o DASS-42 pode realizar a avaliação de ansiedade em adultos, envolvendo coleta de dados, seleção de recursos e treinamento de modelo para prever pontuações totais com precisão com um subconjunto mínimo de itens. A abordagem gera uma ferramenta baseada na web para uso em avaliações futuras.

Este estudo [Srinath et al. \(2022\)](#) visa analisar o desempenho de algoritmos de Aprendizagem de Máquina na previsão dos níveis de gravidade de depressão, ansiedade e estresse (DAS). Utilizando DASS-42 para mensurar esses níveis, o objetivo é aprimorar e comparar o desempenho de dois algoritmos de ML, o SVM e a LR, com foco na acurácia da classificação. O SVM atinge acurácias de 97,35%, 97,49% e 97,20% para depressão, ansiedade e estresse, enquanto o LR atinge acurácias de 98,15%, 98,05% e 98,45%. A análise revelou um desempenho superior do LR em termos de acurácia em comparação ao SVM.

O objetivo do estudo [Singh e Kumar \(2021\)](#) é empregar técnicas de Aprendizado de Máquina e Inteligência Artificial (SVM, Árvores de Decisão, Floresta Aleatória, *Naive Bayes* e KNN) para identificar sinais de ansiedade e depressão utilizando respostas de questionários, especificamente o DASS-21. O trabalho busca analisar o desempenho de diferentes algoritmos de classificação na identificação desses sintomas emocionais, explorando as soluções do uso dessas ferramentas para auxiliar na detecção precoce de ansiedade e depressão a partir das respostas

fornecidas pelos usuários nos questionários.

O objetivo deste estudo [Mendis et al. \(2022\)](#) é desenvolver um aplicativo móvel, promovendo a autoavaliação dos níveis de estresse, ansiedade e depressão (DAS). Por meio de técnicas de Aprendizado de Máquina e processamento de imagens, o aplicativo de busca oferece orientação e aumenta a conscientização local sobre saúde mental. Utilizando escalas como Escala de Auto-Eficácia Geral (*General Self-Efficacy Scale* - GSE) e DASS 21, o aplicativo identifica e mensura problemas de saúde mental.

O artigo de [Pandey, Khurana e Misra \(2021\)](#) apresenta os resultados preliminares da análise de uma pesquisa pública realizada na Índia, avaliando os efeitos psicológicos durante a segunda onda da pandemia de COVID-19. Uma pesquisa foi elaborada para categorizar a população com base em diversos dados socioeconômicos. Os participantes responderam ao questionário DASS-21 para avaliar níveis de ansiedade, depressão e estresse. Foram utilizadas as técnicas de Aprendizado de Máquina: MLR e K-Means.

A leitura dos artigos permitiu a extração dos dados de interesse, apresentados no [Quadro 8](#). Assim, torna-se relevante uma breve explicação sobre cada uma das colunas do quadro:

- **ID:** Identificação do artigo;
- **Título:** Título do artigo;
- **Referência:** Referência do artigo;
- **Base:** Base de pesquisa onde o artigo foi encontrado;
- **Técnica de IA:** Técnicas de Inteligência Artificial utilizadas no estudo;
- **ES:** Participação de especialistas da área da saúde (“S” para estudos com participação, “N” para estudos sem participação).

4.3 Análise e Divulgação

Após a extração dos dados dos artigos selecionados, tornou-se possível responder às questões de pesquisa.

4.3.1 Q1 - Quais Técnicas de IA Foram Utilizadas?

Na questão de pesquisa sobre as técnicas de IA utilizadas, os resultados indicam uma variedade de algoritmos adotados nos estudos revisados. Destacou-se a Máquina de Vetores de Suporte (SVM) com o maior número de ocorrências, totalizando 10 estudos que a utilizaram. Em seguida: métodos como a Floresta Aleatória (RF) com 8 estudos; Árvore de Decisão (DT) com

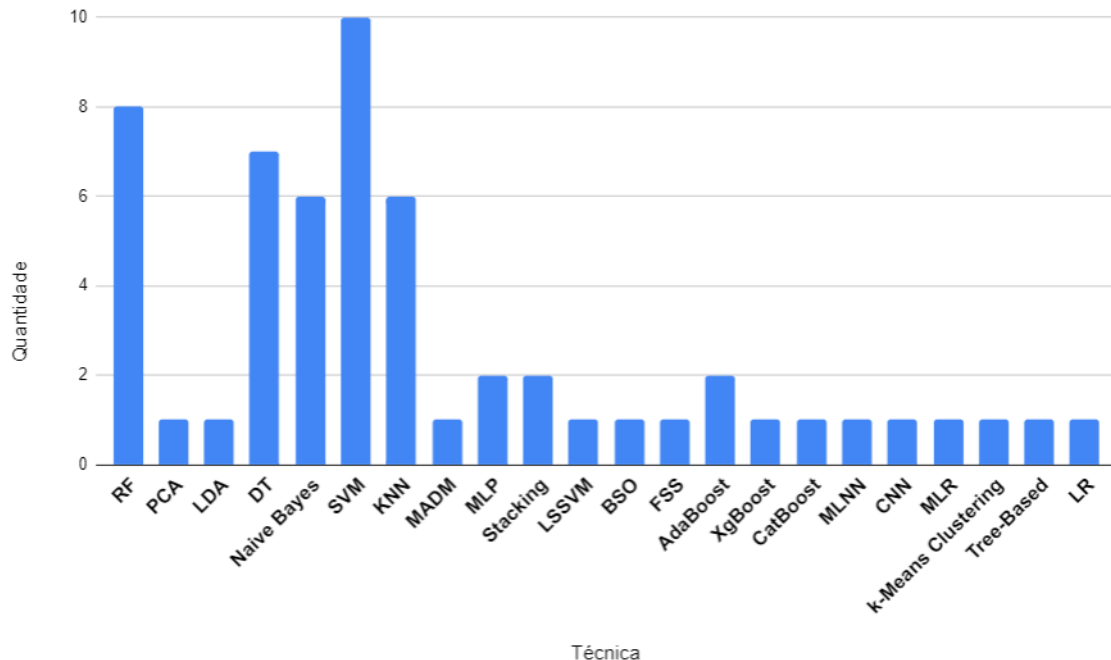
Quadro 8 – Categorização dos Artigos Seleccionados

ID	Título	Referência	Base	Técnica de IA	ES
A1	Translating the user-avatar bond into depression risk: A preliminary machine learning study	(BROWN et al., 2024)	Scopus	Random Forest (RFT) / Naive Bayes / Support Vector Machine (SVM) / K-Nearest Neighbors (KNN)	N
A2	Unveiling Emotions through Handwriting: A Data Analysis Approach	(AZMI; FATHIMA; WASID, 2023)	Scopus	Random Forest Model / Principal Component Analysis (PCA) / Linear Discriminant Analysis (LDA)	N
A3	Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms	(PRIYA; GARG; TIGGA, 2020)	Scopus	Decision Tree (DT) / Random Forest (RFT) / Naive Bayes / Support Vector Machine (SVM) / K-Nearest Neighbors (KNN)	N
A4	Utilising Target Adjacency Information for Multi-target Prediction	(MASKAT et al., 2021)	Scopus	MADM"(Multitarget Adapted Decision Methods) / MLP (Multilayer Perceptron) / MLNN"(Multilabel Neural Network)	N
A5	A novel machine learning approach to shorten depression risk assessment for convenient uses	(SUN et al., 2022)	Scopus	Stacked Generalization Ensemble (Stacking)	N
A6	Machine learning based depression, anxiety, and stress predictive model during COVID-19 crisis	(AL-WESABI et al., 2022)	Scopus	Least Square Support Vector Machine (LSSVM) / Beetle Swarm Optimization (BSO) / Feature Selection Strategy (FSS)	N
A7	Comparative Analysis of Machine Learning Models to Predict Depression, Anxiety and Stress	(SINGH et al., 2022)	Scopus	Random Forest / Decision Tree / Support Vector Machine (SVM) / AdaBoost / CatBoost / Extreme Gradient Boosting (XGBoost)	N
A8	Analysis of Depression, Anxiety, and Stress Chaos Among Children and Adolescents Using Machine Learning Algorithms	(SWAIN; PATRA, 2022)	Scopus	Support Vector Machines (SVM) / K-Nearest Neighbors (KNN) / Naive Bayes / Decision Trees	N
A9	Depression and anxiety detection through the Closed-Loop method using DASS-21	(BUDIYANTO; SIHOMBING; IM, 2019)	Scopus	Naive Bayes	N
A10	Dasentimental: Detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning	(FATIMA et al., 2021)	Scopus	Redes Neurais Perceptron Multicamadas (MLP)	S
A11	DASS-21 Based Psychometric Prediction Using Advanced Machine Learning Techniques	(GHORPADE-AHER et al., 2023)	Scopus	Support Vector Machine (SVM) / AdaBoost	N
A12	Predictive Model for Depression and Anxiety Using Machine Learning Algorithms	(JHA; ABIRAMI; KUMAR, 2022)	Socopus	K-Nearest Neighbors: KNN / Tree-Based Methods	N
A13	Prediction and comparison of psychological health during COVID-19 among Indian population and Rajyoga meditators using machine learning algorithms	(KUMAR; CHANDRA et al., 2023)	PubMed	Decision Tree / Naive Bayes / Random Forest / Support Vector Machine (SVM) / K-Nearest Neighbors (KNN)	N
A14	Emotional distress among health professionals involved in care of inpatients with COVID-19: a survey based cross-sectional study.	(JALANDRA et al., 2022)	LILACS	CatBoost	N
A15	A Novel Approach for Developing Efficient and Convenient Short Assessments to Approximate a Long Assessment	(SUN; LUO; LEE, 2022)	LILACS	Support Vector Machine (SVM) / Random Forest (RF) / Stacked Generalization Ensemble (Ensemble)	N
A16	Prediction of Depression, Anxiety and Stress Levels Using Dass-42	(SRINATH et al., 2022)	IEEE	Support Vector Machine (SVM) / Regressão Logística (LR)	N
A17	Identification of Anxiety and Depression Using DASS-21 Questionnaire and Machine Learning	(SINGH; KUMAR, 2021)	IEEE	Support Vector Machine (SVM) / Decision Trees / Random Forest / Naive Bayes / K-Nearest Neighbors (KNN)	N
A18	Mobile Application for Mental Health Using Machine Learning	(MENDIS et al., 2022)	IEEE	Decision Trees / Random Forest / CNN - Convolutional Neural Networks	S
A19	Towards Understanding the Psychological Effects of the COVID-19 Pandemic on the Indian Population	(PANDEY; KHURANA; MISRA, 2021)	IEEE	Multinomial Logistic Regression (MLR) / K-Means Clustering	N

Fonte: Produção dos autores, 2024.

7; *Naive Bayes* e KNN com 6 estudos cada um. Outras técnicas que tiveram resultados menos expressivos, tiveram seus resultados apresentados no gráfico da [Figura 9](#).

Figura 9 – Técnicas de Inteligência Artificial Utilizadas



Fonte: Produção dos autores, 2024.

4.3.2 Q2 - O Estudo Leva em Consideração a Participação de Profissionais de Saúde para o Diagnóstico da Depressão?

A maioria dos estudos, um total de 17, não considerou a validação médica do diagnóstico, limitando-se a indicadores automáticos baseados em respostas a questionários ou informações retiradas de textos, em conjunto com técnicas de Inteligência Artificial.

O artigo [Fatima et al. \(2021\)](#), que faz análise de palavras e sentimentos para identificar estados emocionais DAS, faz uma breve menção, nos trabalhos futuros, sobre como a tecnologia criada pode auxiliar na elaboração de relatórios médicos. Dos artigos analisados, o [Mendis et al. \(2022\)](#) foi o único que propôs uma estrutura completa, avaliando desde a triagem dos pacientes até a possibilidade de assistência de profissionais na área da saúde mental.

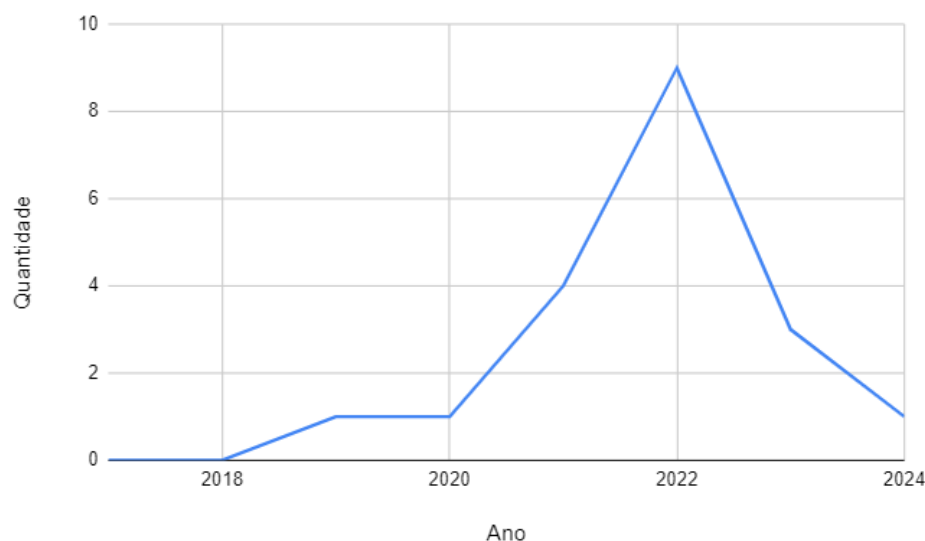
4.3.3 Q3 - Qual o Ano de Publicação do Trabalho?

Para responder à terceira questão de pesquisa, foi criado o gráfico apresentado na [Figura 10](#), contendo a distribuição dos estudos selecionados por ano de publicação, fornecendo uma visão do volume de pesquisas ao longo do tempo. Não houve estudos publicados em 2017 e 2018, ocorrendo um crescimento discreto em 2019 e 2020, sugerindo um início de interesse sobre o tema. Pode-se notar um aumento significativo em 2021 e 2022, onde foram identificados

4 e 9 artigos respectivamente. No ano de 2023, 3 estudos foram selecionados e 2024 com apenas um artigo.

O crescente interesse na saúde mental durante a pandemia de COVID-19 é evidenciado pela análise dos artigos relacionados à mensuração emocional por meio do DASS. A busca por estudos utilizando a DASS revelou um aumento significativo, especialmente em 2020, 2021 e 2022. Esses resultados sugerem uma preocupação em entender o impacto emocional da pandemia, refletindo a busca por uma compreensão mais profunda dos efeitos psicológicos desencadeados por esse evento global. Foi verificado ao analisar os estudos que vários deles fazem referência à pandemia, como: [Al-Wesabi et al. \(2022\)](#), [Kumar, Chandra et al. \(2023\)](#), [Jalandra et al. \(2022\)](#) e [Pandey, Khurana e Misra \(2021\)](#).

Figura 10 – Distribuição dos Estudos Selecionados por Ano de Publicação



Fonte: Produção dos autores, 2024.

4.4 Ameaças à Validade

Validade de construção: a elaboração da *string* de busca utilizada para coletar os dados pode não abranger todos os detalhes do tema em questão.

Validade Externa: O Mapeamento Sistemático da Literatura foi realizado utilizando quatro plataformas: IEEE Xplore, Scopus, PubMed e Lilacs. Isso significa que alguns estudos relevantes, presentes em outras bases de dados não consideradas, podem não ter sido incluídos. A exclusão de determinados estudos pode ter impacto nos resultados apresentados na MSL.

4.5 Discussão Sobre a MSL

Neste capítulo, foram apresentados os procedimentos realizados no Mapeamento Sistemático da Literatura visando encontrar estudos relevantes ao tema. Esse conteúdo proporcionou uma base conceitual para o desenvolvimento desta dissertação. Além disso, por intermédio da leitura dos artigos relacionados ao tema, foi possível avaliar lacunas que podem ser trabalhadas visando avanços relacionados ao contexto proposto.

Os estudos [Srinath et al. \(2022\)](#) e [Singh et al. \(2022\)](#) fizeram uso do mesmo *dataset*, detalhado na [Seção 3.1](#), com respostas ao questionário DASS-42 e informações sociodemográficas, para realizar a classificação dos indicadores DAS, utilizando algoritmos de Aprendizado de Máquina. Os estudos abordaram poucas etapas de pré-processamento dos dados e não realizaram análises sociodemográficas, surgindo a oportunidade de classificar os níveis de depressão com uma abordagem diferente de pré-processamento e realizar uma análise sociodemográfica dos dados e como se relacionam com os níveis de depressão.

O estudo de [Sun, Luo e Lee \(2022\)](#) utilizou uma fração de um *dataset* com respostas ao questionário DASS-42 e informações sociodemográficas para classificar a ansiedade. A leitura do estudo foi importante para entender como técnicas como: conversão de variáveis categóricas, normalização dos dados e o tratamento de classes desbalanceadas podem ser relevantes para problemas de classificação relacionados à saúde mental que fazem uso de base de dados.

Outro fator relevante da MSL foi identificar como os estudos têm uma preocupação maior em entender as estruturas envolvendo o DASS e o Aprendizado de Máquina, somente no contexto computacional, algo que pode ser questionado, uma vez que o DASS não foi projetado para substituir uma avaliação psicológica ([PARKITNY; MCAULEY, 2010](#)), mas sim para gerar indicadores úteis que podem ajudar no diagnóstico ([MAKARA-STUDZIŃSKA; AL., 2022](#)). Assim, gera-se a oportunidade de pensar em estruturas mais completas que usem os aparatos tecnológicos em conjunto com profissionais de saúde.

5

Desenvolvimento do *DepressionMap*

Ao acessar o *DepressionMap*, um menu de navegação é exibido com duas opções: Página Inicial e Inteligência Artificial. Na Página Inicial, é exibido um mapa de calor que permite visualizar a relação entre os países e seus níveis de depressão. Além disso, por meio de uma caixa de seleção, é possível escolher o país, uma vez selecionada a opção desejada, gráficos e dados que mostram a relação entre a depressão e variáveis sociodemográficas são exibidos. Na página Inteligência Artificial, é apresentado um gráfico comparativo relacionado à acurácia e à medida-F1 de diversas técnicas de Aprendizado de Máquina na tarefa de classificação dos indicadores dos níveis de depressão. Um vídeo demonstrando como a ferramenta funciona pode ser assistido em <https://bit.ly/415vLfQ>. Além disso, é possível acessar o *DepressionMap* por meio do link <https://depression-map.streamlit.app/>.

A seguir, serão apresentados detalhes importantes para a elaboração do *DepressionMap*: preparação do ambiente, pré-processamento dos dados, análise dos dados sociodemográficos e um experimento sobre a classificação dos níveis de depressão fazendo uso de técnicas de Aprendizado de Máquina.

5.1 Preparação do Ambiente

O ambiente do *Google Colab* foi preparado para o desenvolvimento do *DepressionMap* com a instalação das bibliotecas detalhadas em [Tabela 2](#).

5.2 Pré-processamento dos Dados

O *dataset* utilizado foi detalhado na [Seção 3.1](#). Dado que o presente estudo tem como foco a depressão e variáveis sociodemográficas, o conjunto de dados foi preparado levando em consideração esses aspectos. Fatores como país, idade, gênero, estado civil, nível educacional,

status de imigrante e renda são frequentemente reconhecidos como determinantes importantes na compreensão da depressão (AKHTAR-DANESH; LANDEEN, 2007; LIM et al., 2018). Assim, foram selecionadas e trabalhadas 19 colunas para investigar informações relacionadas à depressão e às variáveis sociodemográficas.

Ainda nesse contexto, as colunas escolhidas foram: as 14 questões do DASS-42 na subescala da depressão e 5 colunas relacionadas às informações sociodemográficas: país, idade, gênero, estado civil e educação. *Status* de imigrante e renda não foram incluídos, pois não havia dados disponíveis sobre esses aspectos no conjunto de dados utilizado. Alguns procedimentos adicionais foram necessários no pré-processamento, esses que serão detalhados a seguir.

No conjunto de dados, as pontuações para respostas relacionadas ao DASS-42 variaram de 1 a 4, como mencionado anteriormente na Seção 2.1.1, as respostas devem variar de 0 a 3. Como essa divergência poderia impactar os cálculos dos indicadores, foi necessário reduzir 1 nos valores das questões correspondentes ao DASS-42. Esse procedimento também foi realizado no artigo de Singh et al. (2022). Depois desse processo, foi feito o cálculo dos indicadores dos níveis de depressão.

Ao tentar implementar o mapa de calor com os indicadores de depressão, foi notado que alguns países tinham uma representatividade muito baixa, por exemplo: alguns países tinham apenas um, ou dois representantes, e a média do país ficava resumida a uma amostra muito pequena. Diante desse problema, foi adotada a teoria do limite central, onde considera que pelo menos 30 elementos são necessários para garantir boas estimativas (DAVIS; CAMPBELL; WHITEFORD, 2018; PRIEBE; AL., 2019). Assim, foram removidas do conjunto de dados as linhas correspondentes a países com menos de 30 entrevistados.

Foram excluídas as linhas que continham pelo menos uma coluna com valor ausente. Também foram aplicados filtros para garantir que as informações estivessem nas faixas válidas para cada coluna. Essa verificação foi necessária, pois foi observado que alguns casos apresentavam números inválidos. No Quadro 9, são expostas às colunas e seus respectivos valores válidos. Os intervalos de gênero, estado civil e nível educacional estavam detalhados nas instruções do *dataset*. A escolha da faixa de 18 a 100 foi para que o foco fosse em adultos, assim como em Sun, Luo e Lee (2022).

Após a limpeza dos dados, 14.848 linhas foram eliminadas, restando 24.927 entrevistas. Esses dados foram utilizados tanto para análise dos dados sociodemográficos como para a classificação de níveis de severidade da depressão usando ML.

5.3 Análise dos Dados Sociodemográficos

Nesta seção, serão abordados detalhes sobre o procedimento realizado para execução da análise sociodemográfica dos dados. Assim, torna-se relevante entender que índices médios

Quadro 9 – Intervalos Válidos Para as Variáveis Sociodemográficas

Coluna	Intervalo Válido	Descrição dos Valores
Gênero	1 a 3	1 = Masculino, 2 = Feminino, 3 = Outro
Estado Civil	1 a 3	1 = Nunca Casado, 2 = Casado, 3 = Anteriormente Casado
Nível Educacional	1 a 4	1 = Menos que o Ensino Médio, 2 = Ensino Médio, 3 = Graduação, 4 = Pós-Graduação
Idade	18 a 100	Valores numéricos entre 18 e 100 anos

Fonte: Produção dos autores, 2024.

baseados na escala DASS para analisar estados emocionais já foram utilizados em diversos estudos (BINTARI; SUDIBYO; KARIMAH, 2021; NG et al., 2007), incluindo mapas de calor associados a países (TAN; AL., 2023). No presente estudo, as taxas de depressão foram agrupadas pelos grupos de interesse, calculando-se a média aritmética simples deles. Segue a fórmula utilizada:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Onde:

- $\bar{x}$: é a média aritmética simples do grupo;
- $x_1, x_2, \dots, x_n$: representam os indicadores dos níveis de depressão de cada participante no grupo desejado;
- n : é o número total de participantes no grupo desejado.

Quanto à criação do mapa de calor apresentado na Figura 11, o desenvolvimento ocorreu fazendo uso da biblioteca *Folium* e também do arquivo *countries.geo.json*¹. Em relação à legenda, os países em tom cinza escuro tiveram amostras insuficientes, ou seja, menos de 30 participantes. Tons mais claros, como amarelo claro, indicam taxas médias de depressão mais baixas, enquanto tons mais intensos, como o vermelho escuro, indicam taxas médias de depressão mais altas. A divisão das regiões foi embasada em Tan e al. (2023), um estudo com contexto mundial que fez uso do questionário DASS, e adotou 6 regiões principais: Ásia, América do Norte, Europa, Austrália/Nova Zelândia, América do Sul e África. Detalhes relacionados às quantidades das regiões e seus respectivos países são demonstrados no Quadro 10, no total foram selecionados 34 países.

No *DepressionMap*, o usuário pode selecionar um país específico, ou todos os países, para que gráficos com informações sociodemográficas relacionados à depressão sejam listados. Os gráficos pizza mostram as proporções de participantes de cada categoria da variável

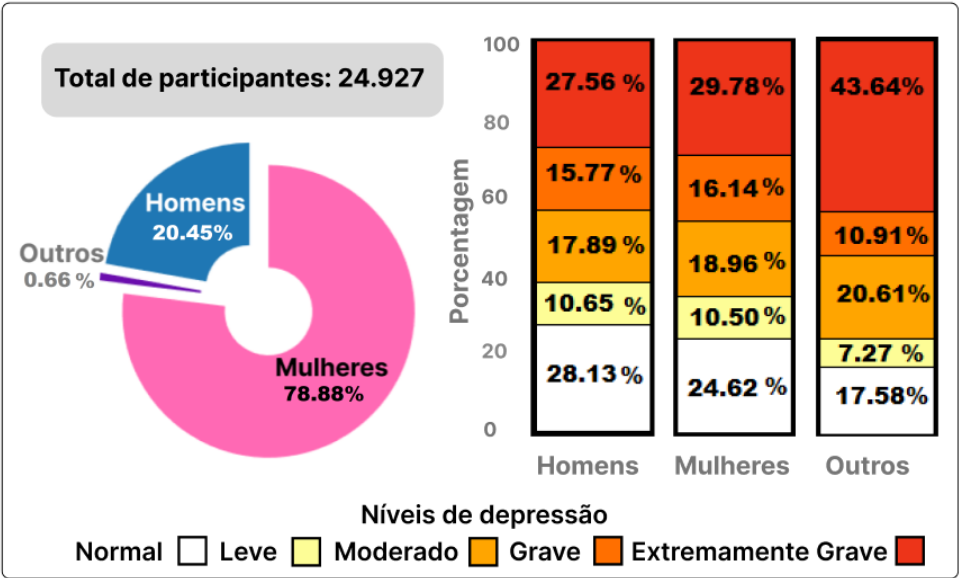
¹ <<https://raw.githubusercontent.com/johan/world.geo.json/master/countries.geo.json>>

Quadro 10 – Quantidade de Participantes por Região

Região (Total de Participantes)	Países (Participantes)
África (107)	Egito (55); África do Sul (52)
Ásia (19.180)	Malásia (17.252); Indonésia (642); Filipinas (472); Índia (331); Singapura (123); Brunei (97); Japão (74); Turquia (54); Paquistão (53); Arábia Saudita (43); Hong Kong (39)
América do Norte (3.813)	Estados Unidos (3.277), Canadá (428); México (75); Jamaica (33)
Europa (1.322)	Reino Unido (517); Alemanha (134); França (125); Polônia (85); Itália (72); Romênia (71); Países Baixos (59) Espanha (43); Irlanda (39); Rússia (39); Grécia (37); Sérvia (35); Portugal (34); Finlândia (32)
Austrália/Nova Zelândia (400)	Austrália (278); Nova Zelândia (122)
América do Sul (105)	Brasil (105)

Fonte: Produção dos autores, 2024.

Figura 12 – Gênero e sua Relação com a Depressão



Fonte: Produção dos autores, 2024.

do “gênero feminino”, com 19.664 participantes no total. O “gênero masculino” e a categoria “Outro” apresentam participações menores. A média do índice de depressão foi mais alta para o grupo “Outro” (2.56), em comparação com o grupo Feminino (2.16) e o Masculino (2.04).

Em relação aos grupos de idade, o grupo 18-34 teve uma maior participação, com 22.418 participantes. Ao comparar os diferentes grupos de idade, observa-se que o grupo mais jovem, “18-34 anos” (2.19), apresenta a maior média de depressão, seguida pelo “35-54 anos” (1.70) e, por fim, o grupo “55+” (1.66). Assim, pessoas mais jovens tiveram os maiores níveis de

Tabela 3 – Distribuição Percentual por Grupo e Nível

Grupo	Total	Nível 0 %	Nível 1 %	Nível 2 %	Nível 3 %	Nível 4 %	Média
Gênero Masculino	5098	28.13	10.65	17.89	15.77	27.56	2.04
Gênero Feminino	19664	24.62	10.50	18.96	16.14	29.78	2.16
Gênero Outro	165	17.58	7.27	20.61	10.91	43.64	2.56
Idade 18-34	22418	23.84	10.42	19.07	16.46	30.21	2.19
Idade 35-54	2083	37.78	11.47	16.18	12.48	22.08	1.70
Idade 55+	426	40.38	10.56	15.02	10.80	23.24	1.66
Nunca Casado	21087	23.14	10.23	18.95	16.47	31.20	2.22
Casado	3202	38.48	12.34	18.11	13.12	17.96	1.60
Anteriormente Casado	638	30.25	10.50	15.52	15.99	27.74	2.00
Menos que o Ensino Médio	83	26.51	14.46	15.66	9.64	33.73	2.10
Ensino Médio	6510	21.43	10.28	18.77	17.47	32.06	2.28
Graduação	13886	24.98	10.45	19.28	15.91	29.38	2.14
Pós-Graduação	4448	31.88	10.97	17.15	14.43	25.56	1.91

Fonte: Produção dos autores, 2024.

depressão.

Quanto à variável estado civil, o grupo “Nunca Casado” apresentou a maior representatividade, com 21.087 participantes. Quanto às médias, o grupo “Casado” apresentou o menor indicativo de depressão, com uma média de 1.60, enquanto o grupo “Anteriormente Casado” apresentou uma média de 2.00 . O grupo que apresentou o maior nível foi o “Nunca Casado” com 2.22 de média.

A nível da escolaridade, o grupo “Graduação” apresentou a maior representatividade, com 13.886 participantes. O grupo “Ensino Médio” (2.28), apresentou a maior média. Em seguida, os grupos: “Graduação” (2.14) e “Menos que o Ensino Médio (2.10)”. O grupo com Pós-Graduação (1.91) apresentou a menor média.

5.3.2 Depressão e Região

A comparação foi feita entre as seis áreas determinadas em Tan e al. (2023), abrangendo os 34 países selecionados. A análise das médias do índice de depressão entre as regiões, em ordem decrescente, mostra que a África apresenta a maior média, com 2,36, seguida pela América do Sul, com uma média de 2,20. As regiões da Ásia, América do Norte e Europa têm médias iguais, de 2,18. Por fim, a Austrália/Nova Zelândia apresenta a média mais baixa, com 2,16. Os detalhes são apresentados no Quadro 11.

5.3.3 Discussão Sobre a Análise Sociodemográfica

É relevante destacar que diversos estudos investigam a relação entre depressão e variáveis sociodemográficas (REISNER; AL., 2016; VARMA; AL., 2021; BJELLAND; AL., 2008; TAN; AL., 2023). A mensuração dos níveis de depressão envolve múltiplos fatores, o que dificulta a formulação de conclusões absolutas.

Quadro 11 – Índice da Depressão por Região e País

Região (Média)	Países (Média)
África (2,36)	África do Sul (2,13); Egito (2,58)
Ásia (2,18)	Malásia (2,12); Indonésia (2,32); Filipinas (2,46); Índia (2,23); Singapura (2,08); Brunei (2,18); Japão (1,84); Turquia (2,39); Paquistão (2,32); Arábia Saudita (1,91); Hong Kong (2,13)
América do Norte (2,18)	Estados Unidos (2,09); Canadá (2,12); México (2,25); Jamaica (2,24)
Europa (2,18)	Reino Unido (2,16); Alemanha (2,26); França (2,28); Polónia (2,69); Itália (2,50); Romênia (2,01); Países Baixos (2,24); Finlândia (2,34); Irlanda (2,26); Espanha (1,81); Grécia (2,03); Sérvia (1,46); Portugal (2,12); Rússia (2,31)
Austrália/Nova Zelândia (2,16)	Austrália (2,11); Nova Zelândia (2,21)
América do Sul (2,20)	Brasil (2,20)

Fonte: Produção dos autores, 2024.

Ainda nesse sentido, embora uma série de estudos indique que as mulheres são mais propensas à depressão do que os homens (KUEHNER, 2017; WILHELM, 2001; PARKER; BROTHIE, 2010), uma pesquisa conduzida na Noruega, utilizando a Escala Hospitalar de Ansiedade e Depressão (HADS), revelou que os homens apresentavam uma maior probabilidade de desenvolver depressão em comparação às mulheres (NORTVEDT; RIISE; SANNE, 2006). Dessa forma, a relação entre depressão e variáveis sociodemográficas pode variar a depender das características do público analisado.

O objetivo central da análise sociodemográfica conduzida nesta dissertação não foi identificar quais grupos são mais afetados pela depressão, uma vez que os dados utilizados foram obtidos a partir de um *dataset* público, o que impossibilitou o controle da qualidade da coleta. Além disso, o conjunto de dados apresentou limitações, como o desbalanceamento das classes. Por exemplo, a maioria dos participantes era da Ásia, enquanto a América do Sul apresentou poucos indivíduos, trazendo ameaças à validade em relação à representatividade das regiões ou países.

Ainda, o processo de pré-processamento dos dados foi fundamental para a realização da análise sociodemográfica, uma vez que com os dados faltantes não seria possível gerar os indicadores. Mas, a remoção dos dados pode ter impactado nos resultados obtidos. Diante de todos aspectos apresentados, o foco principal não foi os resultados alcançados, mas sim o desenvolvimento de um sistema capaz de gerar indicadores relacionados à depressão e a variáveis sociodemográficas.

5.4 Experimento Classificação dos Níveis de Depressão

Nesta seção, serão abordados detalhes sobre o procedimento realizado no experimento, que envolveu a classificação dos níveis de depressão por meio de técnicas de Aprendizado de Máquina.

5.4.1 Objetivo

Este experimento teve como objetivo classificar os níveis de depressão com o uso de algoritmos de Aprendizado de Máquina. Como os níveis DASS-42 podem ser classificados em categorias (normal, leve, moderado, severo e extremamente severo), o problema foi tratado como uma tarefa de classificação supervisionada. O objetivo principal foi atribuir corretamente cada conjunto de entrada a uma dessas classes.

5.4.2 Contexto

Os dados utilizados foram obtidos a partir das respostas de 24.927 participantes, conforme descrito na seção 5.2. Para a classificação, foram analisados os seguintes algoritmos: Máquina de Vetores de Suporte, Floresta Aleatória, Regressão Logística, K-Vizinhos Mais Próximos, *Gaussian Naive Bayes*, Árvore de Decisão e *AdaBoost*. A eficácia dos modelos foi avaliada por meio das métricas de acurácia, precisão, sensibilidade e medida-F1. Apenas os modelos que alcançaram pelo menos 95% de acurácia foram incluídos na análise estatística, cujo objetivo foi comparar o desempenho dos algoritmos e identificar quais apresentam maior eficácia.

5.4.3 Formulação do Problema e Hipóteses

O experimento busca responder à seguinte questão de pesquisa:

No contexto da classificação dos níveis de depressão, entre os algoritmos analisados que atingem a meta de 95% de acurácia, quais apresentam melhor eficácia?

Para responder à questão de pesquisa, as seguintes hipóteses foram elaboradas:

- H_0 : Os algoritmos (1,2...n) têm a mesma eficácia.
- H_1 : Os algoritmos (1,2...n) têm eficácias diferentes.

5.4.4 Variáveis Independentes e Dependentes

Variáveis independentes: São as variáveis do *dataset*, respostas ao DASS-42 e informações sociodemográficas. Como também os algoritmos selecionados.

Variáveis dependentes: Classificações, das quais podem ser derivadas: acurácia, sensibilidade, precisão e medida-F1.

5.4.5 Execução

Antes do treinamento dos modelos, as seguintes técnicas foram aplicadas: *One-Hot Encoding*, com o objetivo de converter variáveis categóricas em um formato numérico; técnica Min-Max, para a normalização dos dados; e a técnica *SMOTE*, visando melhorias para o balanceamento das classes. Os algoritmos selecionados foram utilizados para classificar os níveis de severidade da depressão. As 24.927 linhas foram divididas da seguinte forma: 80% dos dados dedicados ao treinamento e 20% ao teste. Os algoritmos não sofreram ajustes nos parâmetros.

O treinamento foi realizado fazendo uso da técnica *10-Fold Cross Validation*, onde os dados foram divididos em 10 partes, e o modelo foi treinado e testado 10 vezes. A cada iteração, uma parte diferente dos dados foi usada para teste, e as outras 9 partes foram usadas para treinamento. Para cada métrica avaliada, foi calculada a média das 10 iterações. O modelo foi previamente treinado e os resultados apresentados no *DepressionMap*.

5.4.6 Validação dos dados

Para a análise, interpretação e validação dos dados, foram utilizados quatro testes estatísticos: *Shapiro-Wilk* (SHAPIRO; WILK, 1965), para verificar a normalidade dos dados; *Bartlett* (BARTLETT, 1937), para avaliar a homogeneidade das variâncias; *Friedman* (FRIEDMAN, 1937), para comparar o desempenho dos algoritmos; e o *Post-Hoc Nemenyi* (NEMENYI, 1963), para identificar diferenças significativas entre as eficácias das técnicas. O fluxo para escolha dos métodos estatísticos foi feito com o auxílio da ferramenta *Autorank*, desenvolvida por Herbold (2020).

5.4.7 Resultados

Os algoritmos Regressão Logística, SVM e Floresta Aleatória tiveram os valores mais elevados em termos de acurácia, com 0.9969, 0.9966 e 0.9592, respectivamente. Os demais algoritmos não alcançaram a meta de 95% de acurácia. O K-Vizinhos Mais Próximos obteve 0.9187, enquanto a Árvore de Decisão apresentou 0.8463. O *Gaussian Naive Bayes* teve um resultado de 0.6854, o *AdaBoost* registrou o valor mais baixo entre os modelos analisados, com 0.4936. Esses resultados foram expostos na Tabela 4 e Figura 13.

Embora a Regressão Logística tenha alcançado a acurácia mais alta, a escolha do melhor algoritmo entre os que alcançaram a meta (acurácia $\geq 95\%$) requer evidências estatísticas suficientemente conclusivas. Para as análises estatísticas as métricas acurácia e medida-F1, que harmoniza a precisão e a sensibilidade, foram utilizadas. Foi adotado um nível de significância de 0.05 para todo experimento.

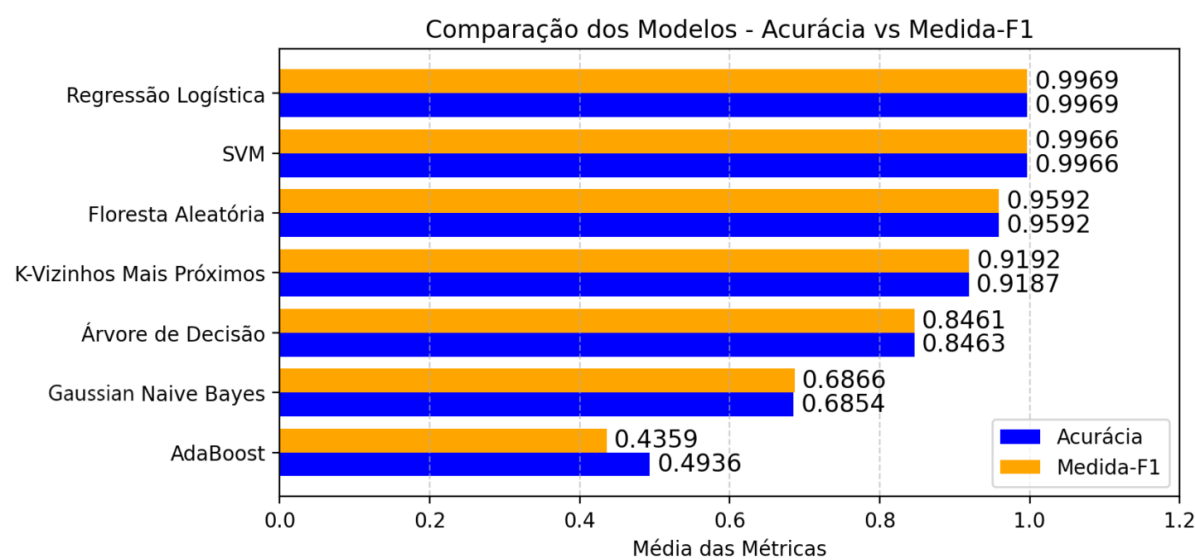
Inicialmente o teste de *Shapiro-Wilk* foi aplicado para verificar a normalidade da distribuição dos dados, com resultados apresentados em Tabela 5. Como todos os valores foram

Tabela 4 – Resultados dos Modelos de Classificação

Modelo	Acurácia	Precisão	Sensibilidade	Medida-F1
Regressão Logística	0.9969	0.9969	0.9969	0.9969
SVM	0.9966	0.9966	0.9966	0.9966
Floresta Aleatória	0.9592	0.9594	0.9592	0.9592
K-Vizinhos Mais Próximos	0.9187	0.9231	0.9187	0.9192
Árvore de Decisão	0.8463	0.8462	0.8463	0.8461
Gaussian Naive Bayes	0.6854	0.7350	0.6854	0.6866
AdaBoost	0.4936	0.5946	0.4936	0.4359

Fonte: Produção dos autores, 2024.

Figura 13 – Comparação dos Modelos Acurácia e Medida-F1



Fonte: Produção dos autores, 2024.

superiores ao nível de significância previsto, conclui-se que as distribuições das métricas Acurácia e Medida-F1 seguem uma distribuição normal.

Tabela 5 – Teste de *Shapiro-Wilk* para Análise da Normalidade

Algoritmo	Acurácia	Medida-F1
Regressão Logística	0.2667	0.2697
SVM	0.9038	0.9040
Floresta Aleatória	0.4894	0.4721

Fonte: Produção dos autores, 2024.

No entanto, o teste de *Bartlett* indicou que as variâncias não são homogêneas, pois os *p-values* de 0,00176 para acurácia e 0,00172 para medida-F1 foram menores que o nível de significância (*p-value* < 0.05). Como os dados seguem uma distribuição normal, mas as variâncias não são homogêneas, optou-se pelo teste de *Friedman* para comparar o desempenho dos algoritmos.

Ao realizar o teste de *Friedman*, os resultados obtidos apresentaram um *p-values* de 0.00036 para acurácia e 0.00050 para medida-F1, ambos inferiores ao nível de significância (*p-value* < 0.05). Isso indica que pelo menos um dos modelos apresentou um desempenho significativamente diferente dos demais, o que é suficiente para rejeitar a hipótese nula (H_0), que assume que os algoritmos (1, 2, ..., n) possuem a mesma eficácia.

O teste de *Friedman* não permite identificar quais algoritmos se diferenciam entre si. Para isso, foi realizado o teste *Post-Hoc de Nemenyi*, que avalia comparações par a par entre os modelos, com resultados apresentados na [Tabela 6](#). Os testes entre Floresta Aleatória e SVM, assim como entre Floresta Aleatória e Regressão Logística, apresentaram *p-values* abaixo de 0.05, indicando diferenças estatisticamente significativas entre esses modelos.

Apesar disso, ao analisar a relação entre SVM e Regressão Logística, os *p-values* obtidos foram superiores a 0.05 tanto para acurácia quanto para Medida-F1, o que indica que as diferenças entre esses algoritmos não são estatisticamente relevantes. Isso sugere que ambos os algoritmos apresentam desempenhos equivalentes e foram os modelos que mais se destacaram em termos de eficácia na classificação dos níveis de depressão.

Tabela 6 – Teste de *Post-Hoc Nemenyi*

Algoritmo 1	Algoritmo 2	Acurácia	Medida-F1
Floresta Aleatória	SVM	0.007165	0.004965
Floresta Aleatória	Regressão Logística	0.000658	0.001010
SVM	Regressão Logística	0.780467	0.895638

Fonte: Produção dos autores, 2024.

A [Tabela 7](#) apresenta um comparativo, relativo a métrica acurácia, entre os resultados obtidos neste estudo e estudos relacionados. As técnicas Regressão Logística, SVM, Floresta Aleatória, K-Vizinhos Mais Próximos, Árvore de Decisão e *Gaussian Naive Bayes* demonstraram acurácias superiores no presente estudo em relação aos estudos relacionados. Enquanto, a técnica *AdaBoost* apresentou desempenho inferior em termos de acurácia.

Tabela 7 – Comparação dos Modelos de Aprendizado de Máquina

Modelo	Neste Estudo	(SRINATH et al., 2022)	(SINGH et al., 2022)
Regressão Logística	99,69	98,15	-
SVM	99,66	97,35	94,00
Floresta Aleatória	95,92	84,97	85,00
K-Vizinhos Mais Próximos	91,87	75,14	73,00
Árvore de Decisão	84,63	-	75,00
Gaussian Naive Bayes	68,54	-	33,00
AdaBoost	49,36	-	80,00

Fonte: Produção dos autores, 2024

5.4.8 Ameaças à Validade

Para validar um experimento, é preciso considerar os fatores que podem impactar o resultado final. Nesta seção, discutem-se as possíveis ameaças identificadas durante a execução do experimento.

Validação interna: Os dados eliminados no pré-processamento podem gerar viés nos resultados.

Validade externa: Uma limitação do *dataset* consiste em que a maioria dos participantes é de asiáticos, mulheres e jovens, o que resultou em dados demográficos desequilibrados, visando melhorias nesse aspecto, foi utilizada a técnica *SMOTE*. Ainda, não se pode garantir que não haja respostas incorretas no *dataset*, o que pode impactar diretamente os resultados da pesquisa.

5.4.9 Considerações Sobre o Experimento

Foram aplicadas as técnicas de *One-Hot Encoding*, para converter variáveis categóricas em formato numérico; a técnica *Min-Max*, para normalização dos dados; e *SMOTE*, com o objetivo de melhorar o balanceamento das classes. Por intermédio dessas técnicas, foi possível trabalhar os dados de forma a torná-los mais adequados para os modelos.

Os algoritmos analisados, como: Regressão Logística, Máquina de Vetores de Suporte, Floresta Aleatória, KNN e Árvore de Decisão, alcançaram resultados sólidos na classificação. No entanto, as técnicas *Gaussian Naive Naves* e *AdaBoost* tiveram acurácias abaixo de 70%, sendo necessária uma análise mais detalhada para verificar o motivo do rendimento baixo, podendo ser testadas situações diferenciadas de ajustes de parâmetros ou ainda utilizar maior quantidade de dados.

Os resultados indicam que a Regressão Logística e o SVM tiveram desempenhos estatisticamente equivalentes e superiores aos demais modelos, com acurácia de 0.9969 e 0.9966, respectivamente. Isso evidencia a alta eficácia de ambos na classificação de níveis de depressão.

Os estudos relacionados [Srinath et al. \(2022\)](#) e [Singh et al. \(2022\)](#) abordaram poucas etapas de pré-processamento, e realizaram uma abordagem de classificação para os três estados emocionais: estresse, ansiedade e depressão. O presente estudo focou em uma abordagem direcionada somente à depressão, com mais etapas de pré-processamento e menos dados. As técnicas analisadas atingiram melhores resultados em termos de acurácia do que nos estudos relacionados, com exceção da técnica *AdaBoost*, que apresentou desempenho inferior.

6

Proposta de uma Estrutura Melhorada para o *DepressionMap*

Segundo [Hoda \(2021\)](#), a eficácia da tecnologia não pode ser plenamente compreendida sem considerar as pessoas envolvidas. Além disso, um fator que merece atenção é que o DASS não deve ser considerado um diagnóstico, mas sim um indicador ou ferramenta de triagem, sendo importante a presença de profissionais da saúde para acompanhamento e diagnóstico ([YAHAYA et al., 2018](#)). Diante desse contexto, foi elaborada a proposta de uma estrutura melhorada para o *DepressionMap*.

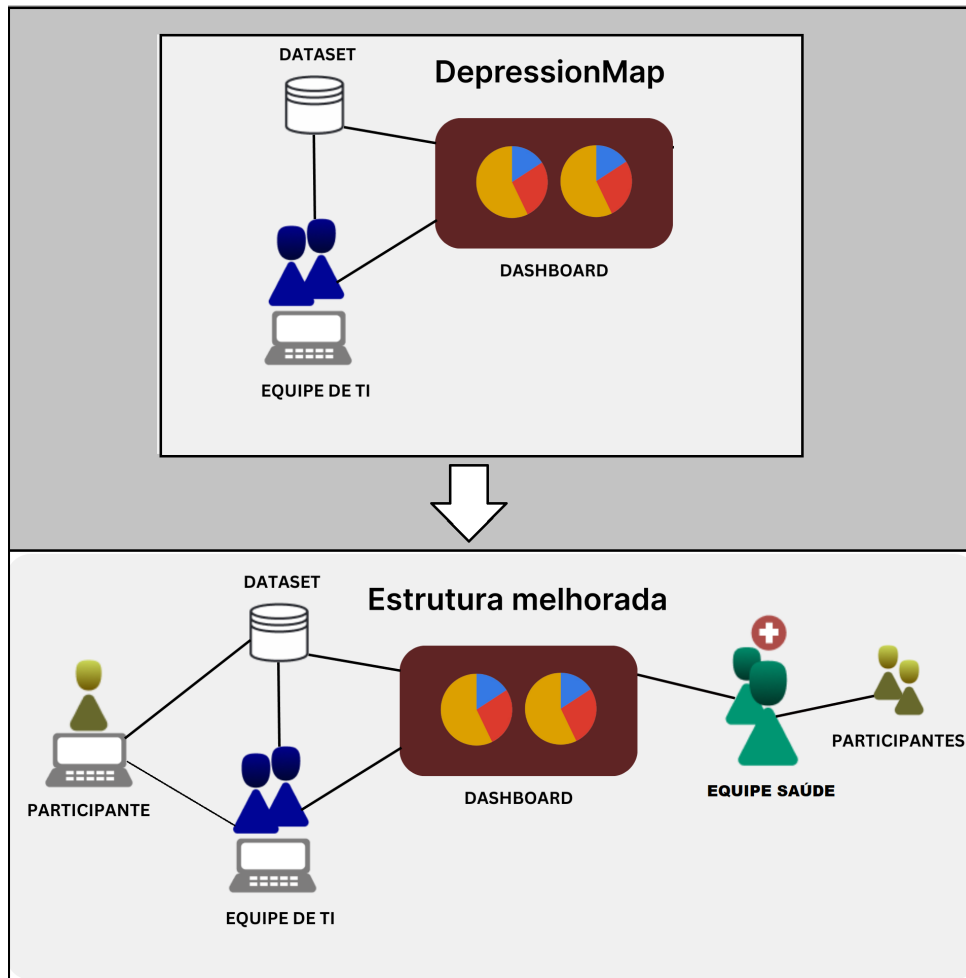
6.1 Estrutura Melhorada

O *DepressionMap* foi desenvolvido dentro de um contexto computacional, não incluindo a participação de profissionais de saúde mental e apresentando limitações relacionadas ao *dataset* utilizado. Por conta desses aspectos, surgiu a necessidade de realizar a proposta conceitual de uma estrutura melhorada, com participação de profissionais da saúde e controle na coleta dos dados.

A [Figura 14](#) apresenta a estrutura atual do *DepressionMap*, com os componentes envolvidos para sua elaboração, onde o procedimento envolveu apenas três elementos principais: o *dataset*, com as informações dos participantes, os autores do presente trabalho, com conhecimento em Tecnologia da Informação (TI), e a ferramenta desenvolvida, o *DepressionMap*. Ainda é apresentada na figura a estrutura melhorada, que envolve desde a coleta de dados até o acompanhamento de profissionais da saúde a pessoas depressivas. A seguir serão apresentados mais detalhes.

6.1.1 Coleta dos Dados

Inicialmente ocorre a coleta de dados por meio de um *site* disponibilizado *online*, onde o participante será convidado a responder o questionário, assinando um termo, autorizando sua

Figura 14 – *DepressionMap* x Estrutura Melhorada

Fonte: Produção dos autores, 2024.

participação e envio das suas informações para a equipe de saúde responsável pelo projeto. O indivíduo fornecerá as informações preenchendo o DASS-42 na subescala de depressão e dados pessoais, como: nome, telefone, idade, renda, localidade, gênero, estado civil e nível educacional. Uma rotina desenvolvida pela equipe de TI processa as respostas, verificando se a pesquisa foi preenchida de forma adequada. Bem como é realizado o cálculo dos níveis de depressão de acordo com o DASS-42.

Depois do processamento das informações, o participante recebe um *feedback*, contendo o nível de depressão, com uma nota explicativa informando que o resultado serve apenas como um indicador inicial, e que a equipe entrará em contato caso o participante tenha apresentado algum nível de depressão. Os dados coletados, incluindo as respostas ao questionário e o nível de depressão calculado, são armazenados de forma segura no *dataset*. Para melhorar as questões de segurança e privacidade, técnicas como criptografia de dados e controle de acesso serão utilizadas, garantindo que somente pessoas autorizadas possam visualizar ou manipular essas informações.

6.1.2 Criação e Disponibilização do *DepressionMap* Melhorado

Um painel é implementado pela equipe de TI para facilitar o trabalho da equipe de saúde, composta por profissionais especialistas em saúde mental. No *dashboard* são disponibilizadas operações relacionadas ao cadastro dos usuários, como inserir, consultar, atualizar e deletar, permitindo que a equipe de saúde tenha acesso aos dados dos participantes. Ainda será disponibilizada uma lista de espera das pessoas que aguardam atendimento.

Quanto aos indicadores sociodemográficos relacionados à depressão, um dos recursos oferecidos é um mapa de calor, que apresenta visualmente a prevalência de depressão por região, utilizando abordagens semelhantes ao *DepressionMap*. Esse mapa pode ser ajustado pela equipe de TI para mostrar informações em diferentes áreas, exemplo: bairros, municípios ou estados, ajudando a identificar regiões com maior incidência de casos de depressão. Também serão disponibilizados gráficos e tabelas que trazem informações sobre a depressão e variáveis sociodemográficas. Ainda, o sistema poderá proporcionar análises realizadas com utilização de Aprendizado de Máquina, para identificar grupos e regiões com maior risco de depressão.

6.1.3 Atuação da Equipe de Saúde

Com base nos dados apresentados no *dashboard*, a equipe de saúde pode realizar análises e tomar decisões. Quanto ao acompanhamento, os profissionais de saúde podem contatar a lista de espera de atendimento, agendar e realizar consultas (seja *online* ou presencial). Dessa forma, ocorre a finalização da proposta da estrutura melhorada.

6.2 Ameaças à Validade

Por ser uma sugestão desenvolvida em um contexto teórico, pode apresentar ameaças à validade relacionadas à sua capacidade de considerar adequadamente as condições reais de uso, possuindo o risco de não refletir as situações práticas em que será aplicada.

Como também, a estrutura foi proposta de forma superficial, sem o devido aprofundamento nas funcionalidades do sistema e sem a consulta a profissionais de saúde. Isso dificulta a verificação das funcionalidades essenciais para garantir que o modelo atenda de maneira eficaz às necessidades de diagnóstico e ofereça suporte adequado na prática clínica. Por fim, as implicações éticas, como a proteção da privacidade dos pacientes, e o respeito a normas inerentes às áreas da saúde.

7

Considerações Finais e Trabalhos Futuros

Este trabalho desenvolveu o *DepressionMap*, uma ferramenta focada na geração de indicadores relacionados à depressão, com o objetivo de contribuir para soluções inteligentes no campo da saúde mental. Para a implementação da ferramenta, foram necessários diversos objetivos específicos. A seguir, apresenta-se uma conclusão sucinta sobre cada um deles:

Por meio da realização do objetivo específico **realizar o MSL**, foi possível identificar estudos relevantes ao tema, os quais serviram de referências para a escolha do conjunto de dados e para alguns procedimentos realizados na elaboração do *DepressionMap*. Um fator relevante foi o entendimento de que a maioria dos estudos relacionados ao tema é explorado exclusivamente em um contexto computacional, com apenas uma minoria considerando o trabalho conjunto entre a tecnologia e os profissionais de saúde.

Quanto à execução do objetivo específico **realizar o pré-processamento do conjunto de dados**, foi feita uma análise criteriosa, dividida em várias fases, pois não seria possível realizar análises com dados faltantes ou inconsistentes. Muitas das decisões foram baseadas na literatura relacionada, com a intenção de garantir que as escolhas fossem pertinentes.

Em relação ao objetivo específico **implementar os indicadores de depressão e suas correlações com variáveis sociodemográficas**, por intermédio da análise dos dados, foi possível disponibilizar o mapa de calor e diversos gráficos expondo informações sobre a relação das variáveis sociodemográficas e dos níveis de depressão. As limitações do *dataset* e seu impacto nos resultados foram aspectos relevantes na análise sociodemográfica.

Mediante a execução do objetivo específico **realizar um experimento visando à análise dos métodos de Aprendizado de Máquina na classificação dos níveis de depressão**, foi possível, por meio dos dados pré-processados e de técnicas de Aprendizado de Máquina, classificar os níveis de depressão. A Regressão Logística e o SVM obtiveram respectivamente acurácias de 99.69 e 99.66, sendo estatisticamente equivalentes e os algoritmos mais eficazes na classificação de níveis de depressão.

Em relação ao objetivo específico **modelar a proposta de uma estrutura melhorada para o *DepressionMap***, foi elaborada uma proposta conceitual de uma estrutura melhorada, considerando aprimorações do presente estudo, como a falta de controle na coleta de dados. Além disso, levou-se em conta a participação de profissionais da saúde, com o objetivo de contribuir para o aumento de estudos que fazem a interseção entre computação e saúde mental.

Em relação às contribuições, foram realizadas pesquisas relacionadas à saúde mental, Cidades Inteligentes e Inteligência Artificial, como apresentadas abaixo:

- O artigo: Relevância de Escalas Emocionais no Aprendizado de Máquina e em Cidades Inteligentes. Submetido e aguardando resultado da revista Abakós;
- O artigo: Saúde Mental e Inteligência Artificial, Identificação de Temas e Técnicas Relevantes: um Mapeamento Sistemático. Submetido e aguardando resultado da Revista Tecnologia e Sociedade;
- O artigo: *Accessible Smart Health: Framework for Depression Indicators*. Submetido ao Simpósio Brasileiro de Sistemas de Informação 2025 (SBSI/2025).

Uma sugestão de trabalho futuro seria envolver a utilização do *dashboard* para análise de dados sociodemográficos em um contexto real, integrando a colaboração de profissionais especialistas em saúde mental. Nesse cenário, seria possível realizar entrevistas com pacientes, acompanhar sua evolução e cruzar informações sobre suas condições sociodemográficas, a fim de entender melhor os fatores que influenciam o quadro de depressão.

A proposta de estrutura aprimorada exposta no Capítulo 6 seguiu uma linha sucinta. Na prática, seria necessária uma investigação minuciosa envolvendo profissionais de tecnologia e da saúde. No nível tecnológico, seria importante entender os detalhes do sistema, bem como a participação da equipe médica na definição das funcionalidades da estrutura para garantir que o sistema atenda às necessidades clínicas.

Referências

ABBASS, H. What is artificial intelligence? *IEEE Transactions on Artificial Intelligence*, IEEE, v. 2, n. 2, p. 94–95, 2021. Citado na página 26.

ABDEEN, M. et al. Improving the performance of ambulance emergency service using smart health systems. In: IEEE. *2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. [S.l.], 2021. p. 205–209. Citado na página 25.

ABDEEN, M. A. R. et al. Smart health systems components, challenges, and opportunities. *IEEE Canadian Journal of Electrical and Computer Engineering*, v. 45, n. 4, p. 436–441, 2022. Citado na página 17.

AKHTAR-DANESH, N.; LANDEEN, J. Relation between depression and sociodemographic factors. *International Journal of Mental Health Systems*, v. 1, p. 1–9, 2007. Citado na página 48.

AL-WESABI, F. N. et al. Machine learning based depression, anxiety, and stress predictive model during covid-19 crisis. *CMC-Comput. Mat. Contin.*, p. 5803–5820, 2022. Citado 3 vezes nas páginas 38, 43 e 45.

ALBINO, V.; BERARDI, U.; DANGELICO, R. M. Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of urban technology*, Taylor & Francis, v. 22, n. 1, p. 3–21, 2015. Citado na página 23.

ALLAM, Z.; JONES, D. S. On the coronavirus (covid-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (ai) to benefit urban health monitoring and management. In: MDPI. *Healthcare*. [S.l.], 2020. v. 8, n. 1, p. 46. Citado na página 24.

AMPOMAH, E. K. et al. Stock market prediction with gaussian naïve bayes machine learning algorithm. *Informatica*, v. 45, n. 2, 2021. Citado na página 30.

ANANTH, S.; SATHYA, P.; MOHAN, P. M. Smart health monitoring system through iot. In: *2019 International Conference on Communication and Signal Processing (ICCSP)*. [S.l.: s.n.], 2019. p. 0968–0970. Citado na página 24.

ANTONIOU, G.; PAPADAKIS, E.; BARYANNIS, G. Mental health diagnosis: a case for explainable artificial intelligence. *International Journal on Artificial Intelligence Tools*, World Scientific, v. 31, n. 03, p. 2241003, 2022. Citado na página 26.

AZMI, M.; FATHIMA, S. L.; WASID, M. Unveiling emotions through handwriting: A data analysis approach. In: IEEE. *2023 12th International Conference on Advanced Computing (ICoAC)*. [S.l.], 2023. p. 1–6. Citado 2 vezes nas páginas 38 e 43.

BAJAJ, S.; MAHESH, R. Converged avenues: depression and alzheimer’s disease–shared pathophysiology and novel therapeutics. *Molecular Biology Reports*, Springer, v. 51, n. 1, p. 225, 2024. Citado na página 25.

BARBA, C. T. et al. Smart city for vanets using warning messages, traffic statistics and intelligent traffic lights. In: IEEE. *2012 IEEE intelligent vehicles symposium*. [S.l.], 2012. p. 902–907. Citado na página 24.

- BARKER, D. J. Ultrasonic vocalizations as an index of positive emotional state. v. 25, p. 253–260, 2018. Citado na página 15.
- BARTLETT, M. S. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, The Royal Society London, v. 160, n. 901, p. 268–282, 1937. Citado na página 55.
- BATTY, M. et al. Smart cities of the future. *The European Physical Journal Special Topics*, Springer, v. 214, p. 481–518, 2012. Citado na página 24.
- BELLINI, V. et al. Understanding basic principles of artificial intelligence: A practical guide for intensivists. *Acta Bio Medica: Atenei Parmensis*, Mattioli 1885, v. 93, n. 5, 2022. Citado na página 26.
- BENEDICT, M. A.; MCMAHON, E. T. et al. Green infrastructure: smart conservation for the 21st century. *Renewable resources journal*, RENEWABLE NATURAL RESOURCES FOUNDATION, v. 20, n. 3, p. 12–17, 2002. Citado na página 24.
- BERGER, N. et al. Harmonising summary measures of population health using global survey instruments. *J Epidemiol Community Health*, BMJ Publishing Group Ltd, v. 70, n. 10, p. 1039–1044, 2016. Citado na página 20.
- BERNBAU, H.; FUJITA, F.; PFENNIG, J. Consistency, specificity, and correlates of negative emotions. *Journal of Personality and Social Psychology*, American Psychological Association, v. 68, n. 2, p. 342, 1995. Citado na página 19.
- BERNSTEIN, I.; EVELAND, D. C. State vs trait anxiety: A case study in confirmatory factor analysis. *Personality and Individual Differences*, v. 3, p. 361–372, 1982. Citado na página 21.
- BERTOLETE, J. The roots of the concept of mental health. *World psychiatry*, World Psychiatric Association, v. 7, n. 2, p. 113, 2008. Citado na página 20.
- BEYER, K. M. et al. Exposure to neighborhood green space and mental health: evidence from the survey of the health of wisconsin. *International journal of environmental research and public health*, Multidisciplinary Digital Publishing Institute, v. 11, n. 3, p. 3453–3472, 2014. Citado na página 23.
- BINTARI, D. C.; SUDIBYO, D. A.; KARIMAH, A. Correlation between depression level and headache severity: A study among medical students during the covid-19 pandemic. *Narra J*, Narra Sains Indonesia, v. 1, n. 3, 2021. Citado na página 49.
- BISONG, E.; BISONG, E. Logistic regression. *Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners*, Springer, p. 243–250, 2019. Citado na página 28.
- BJELLAND, I.; AL. et. Does a higher educational level protect against anxiety and depression? the hunt study. *Social Science Medicine*, v. 66, n. 6, p. 1334–1345, 2008. Citado na página 52.
- BOUKHECHBA, M. et al. Demonicsalmon: Monitoring mental health and social interactions of college students using smartphones. *Smart Health*, Elsevier, v. 9, p. 192–203, 2018. Citado na página 25.

- BRERETON, P. et al. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and software*, Elsevier, v. 80, n. 4, p. 571–583, 2007. Citado na página 35.
- BROWN, T. et al. Translating the user-avatar bond into depression risk: A preliminary machine learning study. *Journal of Psychiatric Research*, Elsevier, v. 170, p. 328–339, 2024. Citado 2 vezes nas páginas 37 e 43.
- BUCKLEY, R. Aww: The emotion of perceiving cuteness. *Frontiers in Psychology*, v. 7, 2016. Citado na página 15.
- BUCKMAN, J. E.; AL. et. Role of age, gender and marital status in prognosis for adults with depression: An individual patient data meta-analysis. *Epidemiology and Psychiatric Sciences*, v. 30, p. e42, 2021. Citado na página 17.
- BUDIYANTO, S.; SIHOMBING, H. C.; IM, F. R. Depression and anxiety detection through the closed-loop method using dass-21. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, v. 17, n. 4, p. 2087–2097, 2019. Citado 3 vezes nas páginas 16, 40 e 43.
- BUTT, S. A. et al. Iot smart health security threats. In: *2019 19th International Conference on Computational Science and Its Applications (ICCSA)*. [S.l.: s.n.], 2019. p. 26–31. Citado na página 24.
- CALVO-ZARAGOZA, J.; VALERO-MAS, J. J.; RICO-JUAN, J. R. Improving knn multi-label classification in prototype selection scenarios using class proposals. *Pattern Recognition*, Elsevier, v. 48, n. 5, p. 1608–1622, 2015. Citado na página 29.
- CARAGLIU, A.; BO, C. D.; NIJKAMP, P. Smart cities in europe. *Journal of urban technology*, Taylor & Francis, v. 18, n. 2, p. 65–82, 2011. Citado na página 24.
- CERVANTES, J. et al. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, Elsevier, v. 408, p. 189–215, 2020. Citado na página 29.
- CHAKRABORTY, A. et al. A framework of intelligent mental health monitoring in smart cities and societies. *IETE Journal of Research*, Taylor & Francis, p. 1–14, 2023. Citado 2 vezes nas páginas 16 e 25.
- CHAWLA, N. V. et al. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, v. 16, p. 321–357, 2002. Citado na página 28.
- CHEN, F. et al. Optimization strategy of community planning for environmental health and public health in smart city under multi-objectives. *Frontiers in Public Health*, Frontiers Media SA, v. 12, p. 1347122, 2024. Citado na página 25.
- CHESNEY, M. et al. Positive emotions: exploring the other hemisphere in behavioral medicine. *International Journal of Behavioral Medicine*, v. 12, p. 50–58, 2005. Citado na página 19.
- Conselho Nacional de Saúde. *Resolução nº 510, de 7 de abril de 2016*. 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Disponível em: <<https://www.gov.br/conselho-nacional-de-saude/pt-br/aceso-a-informacao/legislacao/resolucoes/2016/resolucao-no-510.pdf>>. Citado na página 33.

CRAWFORD, J. et al. Percentile norms and accompanying interval estimates from an Australian general adult population sample for self-report mood scales (bai, bdi, crsd, ces-d, dass, dass-21, stai-x, stai-y, srds, and sras). *Australian Psychologist*, Taylor & Francis, v. 46, n. 1, p. 3–14, 2011. Citado 2 vezes nas páginas 20 e 32.

CUNNINGHAM, P.; DELANY, S. J. K-nearest neighbour classifiers-a tutorial. *ACM computing surveys (CSUR)*, ACM New York, NY, USA, v. 54, n. 6, p. 1–25, 2021. Citado na página 30.

DAVIS, E. M.; CAMPBELL, M. A.; WHITEFORD, C. Bullying victimization in non-heterosexual university students. *Journal of Gay & Lesbian Social Services*, Taylor & Francis, v. 30, n. 3, p. 299–313, 2018. Citado na página 48.

DEQIAN, X. The research on newly improved bound semi-supervised support vector machine learning algorithm. In: SPRINGER. *International Conference on Information and Management Engineering*. [S.l.], 2011. p. 383–390. Citado na página 29.

DIMITROV, D. V. Medical internet of things and big data in healthcare. *Healthcare informatics research*, Korean Society of Medical Informatics, v. 22, n. 3, p. 156–163, 2016. Citado na página 25.

DJAHEL, S. et al. A communications-oriented perspective on traffic management systems for smart cities: Challenges and innovative approaches. *IEEE Communications Surveys & Tutorials*, IEEE, v. 17, n. 1, p. 125–151, 2014. Citado na página 24.

DOGRUCU, A. et al. Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data. *Smart Health*, Elsevier, v. 17, p. 100118, 2020. Citado na página 25.

D'ALFONSO, S. Ai in mental health. *Current opinion in psychology*, Elsevier, v. 36, p. 112–117, 2020. Citado na página 16.

ELMAGHRABY, A. S.; LOSAVIO, M. M. Cyber security challenges in smart cities: Safety, security and privacy. *Journal of advanced research*, Elsevier, v. 5, n. 4, p. 491–497, 2014. Citado na página 24.

ESTY, D. C.; WINSTON, A. *Green to gold: How smart companies use environmental strategy to innovate, create value, and build competitive advantage*. [S.l.]: John Wiley & Sons, 2009. Citado na página 24.

EVANS, S. et al. The impact of mental illness on quality of life: A comparison of severe mental illness, common mental disorder and healthy population samples. *Quality of Life Research*, v. 16, p. 17–29, 2006. Acesso em: 23 maio 2024. Disponível em: <<https://doi.org/10.1007/s11136-006-9002-6>>. Citado na página 15.

FADHEL, M.; MAHARANI, W. Depression detection of users in social media x using indobertweet. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, v. 8, n. 2, p. 885–891, 2024. Citado na página 21.

FATIMA, A. et al. Dasentimental: Detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning. *Big Data and Cognitive Computing*, Multidisciplinary Digital Publishing Institute, v. 5, n. 4, p. 77, 2021. Citado 3 vezes nas páginas 40, 43 e 44.

- FIRTH, J. et al. A meta-review of “lifestyle psychiatry”: the role of exercise, smoking, diet and sleep in the prevention and treatment of mental disorders. *World psychiatry*, Wiley Online Library, v. 19, n. 3, p. 360–380, 2020. Citado 2 vezes nas páginas 19 e 20.
- FISHER, M.; BAUM, F. The social determinants of mental health: implications for research and health promotion. *Australian & New Zealand Journal of Psychiatry*, Sage Publications Sage UK: London, England, v. 44, n. 12, p. 1057–1063, 2010. Citado na página 19.
- FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, Taylor & Francis, v. 32, n. 200, p. 675–701, 1937. Citado na página 55.
- GENUER, R. et al. *Random forests*. [S.l.]: Springer, 2020. Citado na página 29.
- GERHARDT, T. E.; SILVEIRA, D. T. *Métodos de pesquisa*. [S.l.]: Plageder, 2009. Citado na página 33.
- GHORPADE-AHER, J. et al. Dass-21 based psychometric prediction using advanced machine learning techniques. *Journal of Advances in Information Technology*, v. 14, n. 3, 2023. Citado 2 vezes nas páginas 41 e 43.
- GIL, A. C. *Como elaborar projetos de pesquisa*. [S.l.]: Editora Atlas SA, 2002. Citado na página 33.
- GRACE, T. W. Health problems of college students. Taylor & Francis, 1997. Citado na página 16.
- GRAY, J. Integration of emotion and cognitive control. *Current Directions in Psychological Science*, v. 13, p. 46 – 48, 2004. Citado na página 15.
- GRØNLUND, A.; KAMMA, L.; LARSEN, K. G. Near-tight margin-based generalization bounds for support vector machines. In: PMLR. *International Conference on Machine Learning*. [S.l.], 2020. p. 3779–3788. Citado na página 29.
- GUPTA, R. Intelligent technology, systems support, and smart cities. In: *IoT Based Smart Applications*. [S.l.]: Springer, 2023. p. 295–307. Citado na página 24.
- HERBOLD, S. Autorank: A python package for automated ranking of classifiers. *Journal of Open Source Software*, v. 5, n. 48, p. 2173, 2020. Citado na página 55.
- HERRMAN, H.; JANÉ-LLOPIS, E. The status of mental health promotion. *Public Health Reviews*, Springer, v. 34, p. 1–21, 2012. Citado 2 vezes nas páginas 19 e 20.
- HERT, M. et al. Physical illness in patients with severe mental disorders. i. prevalence, impact of medications and disparities in health care. *World Psychiatry*, v. 10, 2011. Citado na página 19.
- HODA, R. Socio-technical grounded theory for software engineering. *IEEE Transactions on Software Engineering*, IEEE, v. 48, n. 10, p. 3808–3832, 2021. Citado na página 59.
- HODGSON, R.; ABBASI, T.; CLARKSON, J. Effective mental health promotion: a literature review. *Health Education Journal*, Sage Publications Sage CA: Thousand Oaks, CA, v. 55, n. 1, p. 55–74, 1996. Citado na página 20.

HOLLANDS, R. G. Will the real smart city please stand up?: Intelligent, progressive or entrepreneurial? In: *The Routledge companion to smart cities*. [S.l.]: Routledge, 2020. p. 179–199. Citado na página 24.

IZQUIERDO-VERDIGUIER, E.; ZURITA-MILLA, R. An evaluation of guided regularized random forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 88, p. 102051, 2020. Citado na página 29.

JADHAV, G.; BABAR, S.; MAHALLE, P. A survey: Performance-aware depression detection. In: *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*. [S.l.]: IEEE, 2023. p. 1242–1249. Citado na página 15.

JALANDRA, R. N. et al. Emotional distress among health professionals involved in care of inpatients with covid-19: a survey based cross-sectional study. *Advances in Respiratory Medicine*, 2022. Citado 3 vezes nas páginas 41, 43 e 45.

JAMSHIDI, J.; AL. et. Diverse phenotypic measurements of wellbeing: Heritability, temporal stability and the variance explained by polygenic scores. *Genes, Brain and Behavior*, v. 19, n. 8, p. e12694, 2020. Citado na página 21.

JAPKOWICZ, N.; SHAH, M. *Evaluating learning algorithms: a classification perspective*. [S.l.]: Cambridge University Press, 2011. Citado na página 26.

JHA, A.; ABIRAMI, M.; KUMAR, V. Predictive model for depression and anxiety using machine learning algorithms. In: SPRINGER. *International Conference on Deep Sciences for Computing and Communications*. [S.l.], 2022. p. 133–147. Citado 2 vezes nas páginas 41 e 43.

JIN, J. et al. An information framework for creating a smart city through internet of things. *IEEE Internet of Things journal*, IEEE, v. 1, n. 2, p. 112–121, 2014. Citado na página 15.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, American Association for the Advancement of Science, v. 349, n. 6245, p. 255–260, 2015. Citado na página 26.

JORM, A. F. Does old age reduce the risk of anxiety and depression? a review of epidemiological studies across the adult life span. *Psychological Medicine*, v. 30, p. 11 – 22, 2000. Citado na página 17.

KAMEL, H.; ABDULAH, D.; AL-TUWAIJARI, J. M. Cancer classification using gaussian naive bayes algorithm. In: *2019 International Engineering Conference (IEC)*. [S.l.: s.n.], 2019. p. 165–170. Citado na página 30.

KEDAR, S.; ROKADE, S. Recognition of emotional state based on handwriting analysis and psychological assessment. *Int. J. Eng. Adv. Technol*, v. 8, n. 6, p. 4395–4402, 2019. Citado na página 20.

KHOSHGOFTAAR, T.; ALLEN, E.; DENG, J. Controlling overfitting in software quality models: experiments with regression trees and classification. In: *Proceedings Seventh International Software Metrics Symposium*. [S.l.: s.n.], 2001. p. 190–198. Citado na página 27.

KIM, J.-C.; CHUNG, K. Depression index service using knowledge based crowdsourcing in smart health. *Wireless Personal Communications*, Springer, v. 93, p. 255–268, 2017. Citado na página 25.

- KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele University, Citeseer*, v. 33, n. 2004, p. 1–26, 2004. Citado 2 vezes nas páginas 35 e 36.
- KUEHNER, C. Why is depression more common among women than among men? *The Lancet Psychiatry*, v. 4, n. 2, p. 146–158, 2017. Citado na página 53.
- KULANDAIVEL, R. et al. Intelligent data delivery approach for smart cities using road side units. *IEEE Access*, v. 7, p. 139462–139474, 2019. Citado na página 16.
- KUMAR, P.; CHANDRA, S. et al. Prediction and comparison of psychological health during covid-19 among indian population and rajyoga meditators using machine learning algorithms. *Procedia computer science*, Elsevier, v. 218, p. 697–705, 2023. Citado 3 vezes nas páginas 41, 43 e 45.
- LANDO, J. et al. A logic model for the integration of mental health into chronic disease prevention and health promotion. *Preventing Chronic Disease*, Centers for Disease Control and Prevention, v. 3, n. 2, 2006. Citado na página 19.
- LI, H. et al. Construction of psychological crisis assessment model based on machine learning. In: IEEE. *2019 IEEE International Smart Cities Conference (ISC2)*. [S.l.], 2019. p. 547–550. Citado na página 26.
- LIM, G. et al. Prevalence of depression in the community from 30 countries between 1994 and 2014. *Scientific Reports*, v. 8, p. 2861, 2018. Citado na página 48.
- LINK, B. G.; PHELAN, J. C. Understanding sociodemographic differences in health—the role of fundamental social causes. *American journal of public health*, American Public Health Association, v. 86, n. 4, p. 471–473, 1996. Citado na página 17.
- LOUIS, R. The global socioeconomic impact of mental health. *SocioEconomic Challenges (SEC)*, v. 6, n. 2, 2022. Citado na página 19.
- LOVIBOND. Manual for the depression anxiety stress scales. In: *Manual for the Depression Anxiety Stress Scales*. [S.l.: s.n.], 1995. Citado na página 20.
- MACEDO, A. E. G.; BALTATU, O. C.; SILVA, B. A. da. Development of a monitoring system for evaluation of the mental health program in the city of santarém, state of pará, brazil. In: IEEE. *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)*. [S.l.], 2015. p. 1–6. Citado na página 16.
- MAJUMDER, A. J. A.; VEILLEUX, C. B. Smart health and cybersecurity in the era of artificial intelligence. In: *Computer-mediated communication*. [S.l.]: IntechOpen, 2021. p. 59. Citado na página 16.
- MAKARA-STUDZIŃSKA, M.; AL. et. Confirmatory factor analysis of three versions of the depression anxiety stress scale (dass-42, dass-21, and dass-12) in polish adults. *Frontiers in psychiatry*, Frontiers Media SA, v. 12, p. 770532, 2022. Citado na página 46.
- MARFATIA, J. Psychiatric services for children in india. *Child Psychiatry and Human Development*, Springer, v. 2, n. 3, p. 145–149, 1972. Citado na página 20.
- MASKAT, R. et al. Utilising target adjacency information for multi-target prediction. *Indonesian Journal of Electrical Engineering and Informatics (IJEI)*, v. 9, n. 4, p. 929–942, 2021. Citado 2 vezes nas páginas 38 e 43.

- MCDOWELL, I. *Measuring health: a guide to rating scales and questionnaires*. [S.l.]: Oxford university press, 2006. Citado 2 vezes nas páginas 16 e 20.
- MELNYK, B. et al. Interventions to improve mental health, well-being, physical health, and lifestyle behaviors in physicians and nurses: A systematic review. *American Journal of Health Promotion*, v. 34, p. 929 – 941, 2020. Citado na página 20.
- MENDIS, E. et al. Mobile application for mental health using machine learning. In: IEEE. *2022 4th International Conference on Advancements in Computing (ICAC)*. [S.l.], 2022. p. 387–392. Citado 4 vezes nas páginas 26, 42, 43 e 44.
- MLAKI, D. A. et al. Prevalence and associated factors of depression among older adults in rural tanzania. *International journal of geriatric psychiatry*, Wiley Online Library, v. 36, n. 10, p. 1559–1566, 2021. Citado na página 15.
- MODY, V.; MODY, V. Mental health monitoring system using artificial intelligence: A review. In: IEEE. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*. [S.l.], 2019. p. 1–6. Citado na página 26.
- MOHANTY, S. P.; CHOPPALI, U.; KOUGIANOS, E. Everything you wanted to know about smart cities: The internet of things is the backbone. *IEEE Consumer Electronics Magazine*, IEEE, v. 5, n. 3, p. 60–70, 2016. Citado 3 vezes nas páginas 15, 24 e 25.
- MUGGLETON, S. Alan turing and the development of artificial intelligence. *AI communications*, IOS Press, v. 27, n. 1, p. 3–10, 2014. Citado na página 26.
- NASTESKI, V. An overview of the supervised machine learning methods. *Horizons. b*, v. 4, n. 51-62, p. 56, 2017. Citado na página 27.
- NEIROTTI, P. et al. Current trends in smart city initiatives: Some stylised facts. *Cities*, Elsevier, v. 38, p. 25–36, 2014. Citado 2 vezes nas páginas 15 e 24.
- NEMENYI, P. B. *Distribution-free multiple comparisons*. [S.l.]: Princeton University, 1963. Citado na página 55.
- NG, F. et al. The validity of the 21-item version of the depression anxiety stress scales as a routine clinical outcome measure. *Acta neuropsychiatrica*, Cambridge University Press, v. 19, n. 5, p. 304–310, 2007. Citado na página 49.
- NORTVEDT, M. W.; RIISE, T.; SANNE, B. Are men more depressed than women in norway? validity of the hospital anxiety and depression scale. *Journal of psychosomatic research*, Elsevier, v. 60, n. 2, p. 195–198, 2006. Citado na página 53.
- ORGANIZATION, W. H. et al. *Promoting mental health: Concepts, emerging evidence, practice: Summary report*. [S.l.]: World Health Organization, 2004. Citado na página 19.
- OTEIZA, T. P.; KELLY, L.; MOONEY, P. Using contextual cues in understanding urban mental well-being. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Copernicus GmbH, v. 46, p. 85–89, 2021. Citado na página 15.
- PANDEY, D.; KHURANA, P.; MISRA, A. Towards understanding the psychological effects of the covid-19 pandemic on the indian population. In: IEEE. *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. [S.l.], 2021. Citado 3 vezes nas páginas 42, 43 e 45.

- PARIDA, D.; AL. et. Prevalence and correlates of depression among school-going adolescents in the urban schools of central india: A cross-sectional study. *Cureus*, v. 15, n. 8, 2023. Citado na página 21.
- PARKER, G.; BROTHIE, H. Gender differences in depression. *International Review of Psychiatry*, v. 22, n. 5, p. 429–436, 2010. Citado na página 53.
- PARKITNY, L.; MCAULEY, J. The depression anxiety stress scale (dass). *J Physiother*, v. 56, n. 3, p. 204, 2010. Citado na página 46.
- PATEL, H. H.; PRAJAPATI, P. Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering*, v. 6, n. 10, p. 74–78, 2018. Citado na página 30.
- PATRO, S. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*, 2015. Citado na página 28.
- PELEGRINO, F. M. et al. Cross-cultural adaptation and psychometric properties of the brazilian–portuguese version of the duke anticoagulation satisfaction scale. *Journal of clinical nursing*, Wiley Online Library, v. 21, n. 17-18, p. 2509–2517, 2012. Citado na página 21.
- PETER, J. P. Reliability: A review of psychometric basics and recent marketing practices. *Journal of marketing research*, SAGE Publications Sage CA: Los Angeles, CA, v. 16, n. 1, p. 6–17, 1979. Citado na página 20.
- PETERSON, R. A meta-analysis of cronbach’s coefficient alpha. *Journal of Consumer Research*, v. 21, p. 381–391, 1994. Citado na página 20.
- PHAM, M. T. Emotion and rationality: A critical review and interpretation of empirical evidence. *Review of General Psychology*, v. 11, p. 155 – 178, 2007. Citado na página 15.
- POORAVARI, M. et al. Confirmatory factor analysis of persian version of depression, anxiety and stress (dass-42): Non-clinical sample. *Razavi International Journal of Medicine*, Mashhad Razavi Hospital, v. 5, n. 4, 2017. Citado na página 23.
- POTDAR, K.; PARDAWALA, T. S.; PAI, C. D. A comparative study of categorical variable encoding techniques for neural network classifiers. *International journal of computer applications*, v. 175, n. 4, p. 7–9, 2017. Citado na página 28.
- POURTAHERI, Z. K.; ZAHIRI, S. H. Ensemble classifiers with improved overfitting. In: *2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*. [S.l.: s.n.], 2016. p. 93–97. Citado na página 27.
- PRIEBE, S.; AL. et. Resource-oriented interventions for patients with severe mental illnesses in low-and middle-income countries: trials in bosnia-herzegovina, colombia and uganda. *BMC Psychiatry*, v. 19, p. 1–10, 2019. Citado na página 48.
- PRIYA, A.; GARG, S.; TIGGA, N. P. Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*, Elsevier, v. 167, p. 1258–1267, 2020. Citado 2 vezes nas páginas 38 e 43.
- RAMASWAMI, A. et al. Meta-principles for developing smart, sustainable, and healthy cities. *Science*, American Association for the Advancement of Science, v. 352, n. 6288, p. 940–943, 2016. Citado na página 24.

- REISNER, S. L.; AL. et. Social epidemiology of depression and anxiety by gender identity. *Journal of Adolescent Health*, v. 59, n. 2, p. 203–208, 2016. Citado na página 52.
- ROMERO, C. et al. Data mining algorithms to classify students. In: *Educational data mining 2008*. [S.l.: s.n.], 2008. Citado na página 30.
- SAMARTZIS, L.; TALIAS, M. Assessing and improving the quality in mental health services. *International Journal of Environmental Research and Public Health*, v. 17, 2019. Citado na página 16.
- SCHAPIRE, R. E. Explaining adaboost. In: *Empirical inference: festschrift in honor of vladimir N. Vapnik*. [S.l.]: Springer, 2013. p. 37–52. Citado na página 30.
- SEESKIN, K. Plato and the origin of mental health. *International Journal of Law and Psychiatry*, Elsevier, v. 31, n. 6, p. 487–494, 2008. Citado na página 19.
- SHAHBAZI, M. et al. A review on importance of smart city indexes for citizens' health case study: Esfahan city. *Global Journal of Health Science*, Canadian Center of Science and Education, v. 10, n. 2, p. 1–70, 2018. Citado na página 25.
- SHAHRIAR, S. et al. Machine learning approaches for ev charging behavior: A review. *IEEE Access*, IEEE, v. 8, p. 168980–168993, 2020. Citado na página 26.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, Oxford University Press, v. 52, n. 3-4, p. 591–611, 1965. Citado na página 55.
- SHEN, L.; LI, C. Water body extraction from landsat etm+ imagery using adaboost algorithm. In: *2010 18th International Conference on Geoinformatics*. [S.l.]: IEEE, 2010. p. 1–4. Citado na página 30.
- SHIH, Y.-H. et al. The effects of treatment via telemedicine interventions for patients with depression on depressive symptoms and quality of life: a systematic review and meta-analysis. *Annals of medicine*, Taylor & Francis, v. 55, n. 1, p. 1092–1101, 2023. Citado na página 16.
- SHOREY, S.; NG, E. D.; WONG, C. H. Global prevalence of depression and elevated depressive symptoms among adolescents: A systematic review and meta-analysis. *British Journal of Clinical Psychology*, Wiley Online Library, v. 61, n. 2, p. 287–305, 2022. Citado na página 15.
- SINGH, A.; KUMAR, D. Identification of anxiety and depression using dass-21 questionnaire and machine learning. In: IEEE. *2021 First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT)*. [S.l.], 2021. p. 69–74. Citado 2 vezes nas páginas 41 e 43.
- SINGH, D.; SINGH, B. Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, Elsevier, v. 97, p. 105524, 2020. Citado na página 28.
- SINGH, S. et al. Comparative analysis of machine learning models to predict depression, anxiety and stress. In: IEEE. *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*. [S.l.], 2022. p. 1199–1203. Citado 8 vezes nas páginas 16, 32, 38, 43, 46, 48, 57 e 58.
- SMITS, D. et al. The structure of negative emotion scales: generalization over contexts and comprehensiveness. *European Journal of Personality*, v. 16, p. 127 – 141, 2002. Citado na página 20.

- SOLANAS, A. et al. Smart health: A context-aware health paradigm within smart cities. *IEEE Communications Magazine*, IEEE, v. 52, n. 8, p. 74–81, 2014. Citado na página 15.
- SORNLERLAMVANICH, V. et al. Data analytics and aggregation platform for comprehensive city-scale ai modeling. *Information Modelling and Knowledge Bases XXXIV; IOS Press: Hamburg, Germany*, p. 92–109, 2023. Citado na página 26.
- SRINATH, K. et al. Prediction of depression, anxiety and stress levels using dass-42. In: IEEE. *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*. [S.l.], 2022. p. 1–6. Citado 9 vezes nas páginas 16, 23, 28, 32, 41, 43, 46, 57 e 58.
- SUDHARSON, D. et al. Performance evaluation of improved adaboost framework in randomized phases through stumps. In: *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*. [S.l.: s.n.], 2021. p. 1–6. Citado na página 30.
- SUN, Y. H. et al. A novel machine learning approach to shorten depression risk assessment for convenient uses. *Journal of Affective Disorders*, Elsevier, v. 312, p. 275–291, 2022. Citado 2 vezes nas páginas 38 e 43.
- SUN, Y. H.; LUO, H.; LEE, K. A novel approach for developing efficient and convenient short assessments to approximate a long assessment. *Behavior Research Methods*, Springer, v. 54, n. 6, p. 2802–2828, 2022. Citado 4 vezes nas páginas 41, 43, 46 e 48.
- SUNDQUIST, J. et al. Mindfulness group therapy in primary care patients with depression, anxiety and stress and adjustment disorders: randomised controlled trial. *The British journal of psychiatry : the journal of mental science*, v. 206 2, p. 128–35, 2015. Citado na página 20.
- SWAIN, S.; PATRA, M. R. Analysis of depression, anxiety, and stress chaos among children and adolescents using machine learning algorithms. In: SPRINGER. *International Conference on Innovations in Intelligent Computing and Communications*. [S.l.], 2022. p. 261–272. Citado 2 vezes nas páginas 39 e 43.
- TABER, K. The use of cronbach’s alpha when developing and reporting research instruments in science education. *Research in Science Education*, v. 48, p. 1273 – 1296, 2017. Citado na página 20.
- TAN, Y. Q.; AL. et. Psychological health of surgeons in a time of covid-19: a global survey. *Annals of Surgery*, v. 277, n. 1, p. 50–56, 2023. Citado 4 vezes nas páginas 16, 17, 49 e 52.
- THOMAS, J. J. et al. (Ed.). *Smart Cities and Machine Learning in Urban Health*. Hershey, PA: IGI Global, 2022. Citado na página 25.
- THOMAS, T.; VIJAYARAGHAVAN, A.; EMMANUEL, S. Applications of decision trees. In: *Advanced Computing and Systems for Security*. [S.l.]: Springer, 2019. p. 1–20. Citado na página 30.
- TIAN, S. et al. Smart healthcare: making medical care more intelligent. *Global Health Journal*, v. 3, n. 3, p. 62–65, 2019. ISSN 2414-6447. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2414644719300508>>. Citado na página 16.
- UHER, R. Gene–environment interactions in severe mental illness. *Frontiers in Psychiatry*, v. 5, 2014. Citado na página 19.

- ULLAH, Z. et al. Applications of artificial intelligence and machine learning in smart cities. *Computer Communications*, Elsevier, v. 154, p. 313–323, 2020. Citado na página 26.
- VAKILI, M.; GHAMSARI, M.; REZAEI, M. Performance analysis and comparison of machine and deep learning algorithms for iot data classification. *arXiv preprint arXiv:2001.09636*, 2020. Citado na página 31.
- VAPNIK, V. *The nature of statistical learning theory*. [S.l.]: Springer science & business media, 1999. Citado na página 29.
- VARMA, P.; AL. et. Younger people are more vulnerable to stress, anxiety and depression during covid-19 pandemic: A global cross-sectional survey. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, v. 109, p. 110236, 2021. Citado na página 52.
- VASCONCELLOS-SILVA, P. R.; CASTIEL, L. D. As novas tecnologias de autocuidado e os riscos do autodiagnóstico pela internet. *Revista Panamericana de Salud Pública*, SciELO Public Health, v. 26, n. 2, p. 172–175, 2009. Citado na página 17.
- VIGNOLA, R. C. B.; TUCCI, A. M. Adaptation and validation of the depression, anxiety and stress scale (dass) to brazilian portuguese. *Journal of affective disorders*, v. 155, p. 104–9, 2014. Citado na página 21.
- VIMALESWARAN, B.; RATNAYAKE, G. S. E-therapy improvement monitoring platform for depression using facial emotion detection of youth. In: IEEE. *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. [S.l.], 2021. p. 950–955. Citado na página 20.
- WANG, H. The application of interactive artificial intelligence in the intelligent integration of infants playing teaching aids. In: IEEE. *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*. [S.l.], 2022. p. 30–33. Citado na página 26.
- WANG, J. Modeling and analysis of psychological change and adaptability of college students based on machine learning as an infrastructure to a smart city. *Journal of Testing and Evaluation*, ASTM International, v. 51, n. 3, p. 1650–1660, 2023. Citado na página 26.
- WEN, T. Y.; ARIS, S. A. M. Hybrid approach of eeg stress level classification using k-means clustering and support vector machine. *IEEE Access*, IEEE, v. 10, p. 18370–18379, 2022. Citado na página 28.
- WILHELM, K. A. Men and depression. *Australian Family Physician*, v. 30, n. 1, p. 102–105, 2001. Citado na página 53.
- World Health Organization et al. World mental health report: transforming mental health for all. World Health Organization, 2022. Citado na página 15.
- YADAV, N.; JIN, Y.; STEVANO, L. J. Ar-iomt mental health rehabilitation applications for smart cities. In: IEEE. *2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT)*. [S.l.], 2019. p. 166–170. Citado na página 26.
- YAHAYA, S. N. et al. Prevalence and associated factors of stress, anxiety and depression among emergency medical officers in malaysian hospitals. *World Journal of Emergency Medicine*, v. 9, n. 3, p. 178, 2018. Citado na página 59.

- YANG, Y.; LI, J.; YANG, Y. The research of the fast svm classifier method. In: IEEE. *2015 12th international computer conference on wavelet active media technology and information processing (ICCWAMTIP)*. [S.l.], 2015. p. 121–124. Citado na página [29](#).
- ZAKZAK, L. Citizen-centric smart city development: The case of smart dubai’s “happiness agenda”. In: *Proceedings of the 20th Annual International Conference on Digital Government Research*. [S.l.: s.n.], 2019. p. 141–147. Citado na página [16](#).
- ZANELLA, A. et al. Internet of things for smart cities. *IEEE Internet of Things journal*, Ieee, v. 1, n. 1, p. 22–32, 2014. Citado na página [24](#).
- ZHU, H.; SHEN, L.; REN, Y. How can smart city shape a happier life? the mechanism for developing a happiness driven smart city. *Sustainable cities and society*, Elsevier, v. 80, p. 103791, 2022. Citado na página [16](#).
- ZHU, T. Analysis on the applicability of the random forest. In: IOP PUBLISHING. *Journal of Physics: Conference Series*. [S.l.], 2020. v. 1607, n. 1, p. 012123. Citado na página [29](#).