UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
PRÓ-REITORIA DE GRADUAÇÃO
DEPARTAMENTO DE CIÊNCIAS EXATAS E TECNOLOGIA DA INFORMAÇÃO
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

WESLEY DOS SANTOS SILVA

USO DE TÉCNICAS DE *MACHINE LEARNING* PARA DETECÇÃO DE TRANSTORNOS DEPRESSIVOS EM ESTUDANTES UNIVERSITÁRIOS

ANGICOS-RN
2025

WESLEY DOS SANTOS SILVA

USO DE TÉCNICAS DE *MACHINE LEARNING* PARA DETECÇÃO DE TRANSTORNOS DEPRESSIVOS EM ESTUDANTES UNIVERSITÁRIOS

Trabalho de Conclusão de Curso apresentado à Universidade Federal Rural do Semi-Árido como pré-requisito para cumprimento do componente curricular Trabalho de Conclusão de Curso II.

Orientadora: Profa. Dra. Samara Martins Nascimento Gonçalves.

ANGICOS-RN

2025

S586u Silva, Wesley dos Santos.
Uso de técnicas de machine learning para detecção de transtornos depressivos em estudantes universitários / Wesley dos Santos Silva. - 2025.
59 f. : il.

Orientadora: Samara Martins Nascimento Gonçalves.
Monografia (graduação) - Universidade Federal Rural do Semi-árido, Curso de Sistemas de Informação, 2025.

1. Machine learning. 2. Estudantes. 3. Saúde mental. I. Gonçalves, Samara Martins Nascimento, orient. II. Título.

Ficha catalográfica elaborada por sistema gerador automáto em conformidade com AACR2 e os dados fornecidos pelo) autor(a).
Biblioteca Campus Angicos
Bibliotecário: Helder Romero Maia Duarte
CRB: 15/673

WESLEY DOS SANTOS SILVA

USO DE TÉCNICAS DE *MACHINE LEARNING* PARA DETECÇÃO DE TRANSTORNOS DEPRESSIVOS EM ESTUDANTES UNIVERSITÁRIOS

Trabalho de Conclusão de Curso apresentado à Universidade Federal Rural do Semi-Árido como pré-requisito para cumprimento do componente curricular Trabalho de Conclusão de Curso II.

Defendido em: 10/07/2025

BANCA EXAMINADORA

_____

Samara Martins Nascimento Gonçalves, Profa. Dra. (UFERSA)

Presidente

_____

Joêmia Leilane Gomes de Medeiros, Profa. Dra. (UFERSA)

Membro Examinador

_____

Thatiana Cunha Navarro Diniz, Profa. Dra. (UFERSA)

Membro Examinador

AGRADECIMENTOS

Em primeiro lugar, a Deus, por sempre me dar forças para superar obstáculos e sabedoria para atingir os meus objetivos.

Aos meus pais, que desde sempre se fizeram presentes nas horas difíceis durante a minha jornada acadêmica, me motivando e ajudando no que fosse necessário.

À minha família em geral, pelo apoio e pela assistência, que foram fundamentais para a conclusão deste trabalho.

Aos amigos, que fiz durante o percurso da graduação, pelos momentos de descontração e fraternidade.

À minha orientadora, Profa. Dra. Samara Martins Nascimento, pelo auxílio na condução desse estudo com dedicação e amizade.

RESUMO

A saúde mental é um tema de crescente preocupação, especialmente devido aos estigmas sociais que ainda a cercam. Os alarmantes índices de suicídio entre pessoas com problemas mentais reforçam a urgência de abordagens eficazes. A saúde mental, que engloba o bem-estar psicológico e social, é fortemente influenciada pelo ambiente acadêmico, onde desafios como a alta carga horária, o grande número de disciplinas e a pressão por desempenho podem exacerbar transtornos, como o depressivo. Nesse contexto, técnicas de *Machine Learning* apresentam-se como ferramentas promissoras para auxiliar na detecção precoce de transtornos depressivos em estudantes universitários. Neste sentido, este trabalho investiga a aplicação de modelos para predizer a probabilidade de transtornos depressivos nesse grupo, visando fornecer *insights* sobre padrões emocionais e comportamentais que possam auxiliar profissionais de saúde e educadores, para intervir de forma mais rápida e direcionada. O objetivo é demonstrar como a *Machine Learning* pode apoiar a identificação de estudantes em risco, complementando, e não substituindo, a avaliação clínica.

Palavras-chave: *Machine Learning*; estudantes; saúde mental.

ABSTRACT

Mental health is a topic of growing concern, especially due to the social stigmas that still surround it. The alarming suicide rates among people with mental health problems reinforce the urgency of effective approaches. Mental health, which encompasses psychological and social well-being, is strongly influenced by the academic environment, where challenges such as high workload, large number of courses and pressure to perform can exacerbate disorders, such as depressive. In this context, Machine Learning techniques appear to be a promising tool to aid in the early detection of depressive disorders in university students. This work investigates the application of models to predict the probability of depressive disorders in this group, aiming to provide insights into emotional and behavioral patterns that can help health professionals and educators to intervene more quickly and in a more targeted manner. The objective is to demonstrate how Machine Learning can support the identification of students at risk, complementing, and not replacing, clinical assessment.

Keywords: machine learning; students; mental health.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

WHO	World Health Organization
HDRS	Hamilton Depression Rating Scale
PCA	Principal Component Analysis
SVM	Support Vector Machine
RF	Random Forest
MLP	Multilayer Perceptron
NB	Naive Bayes
DT	Decision Tree
BMI	Body Mass Index
PHQ	Patient Health Questionnaire
GAD	Generalized Anxiety Disorder
HIV	Human Immunodeficiency Virus
MS	Ministério da Saúde
ML	Machine Learning
IA	Inteligência Artificial
SLR	Sparse Logistic Regression
CDRS	Cornell Dysthymia Rating Scale
RBF	Radial Basis Function
NaN	Not a Number
VP	Verdadeiro Positivo
VN	Verdadeiro Negativo
FP	Falso Positivo
FN	Falso Negativo
API	Application Programming Interface
APA	American Psychiatric Association

SUMÁRIO

1 INTRODUÇÃO

A saúde mental tem sido um campo de pesquisa contínuo, revelando diversas implicações em diferentes períodos da história (SCULL, 2015; QUEVEDO, 2019). Apesar disso, o tema permanece relevante para estudos e um alerta constante para a população global. A *World Health Organization* (WHO) (2023) define saúde mental como o bem-estar do indivíduo, abrangendo sua capacidade de lidar com adversidades cotidianas e contribuir para seu ciclo social. A WHO também ressalta que os fatores pessoais, sociais e socioeconômicos exercem influência significativa sobre saúde mental, enquanto o Ministério da Saúde (MS) destaca que características biopsicossociais também contribuem na manutenção do bem-estar (BRASIL, 2019a).

O transtorno depressivo é caracterizado pela falta de interesse na execução de atividades do cotidiano e pelo desânimo, que transcende fronteiras geográficas (SOUZA, 2023; MARTINS, 2022b). Dados da *American Psychiatric Association* (APA) (2023) apontam que 3,8% da população mundial vivencia a algum tipo de transtorno depressivo e que cerca de 280 milhões de indivíduos convivem com esse transtorno, sendo 50% mais prevalente em mulheres. A APA destaca, em seu Manual Diagnóstico e Estatístico de Transtornos Mentais (DMS-5-TR), os diferentes tipos de transtornos depressivos, como: o transtorno disruptivo da desregulação do humor, o transtorno depressivo maior e o transtorno depressivo persistente. O Brasil se destaca como o país com o maior número de pessoas acometidas por transtornos depressivos entre as nações da América Latina, com aproximadamente 15,5 % de prevalência entre a população (MARTINS, 2022b; BRASIL, 2019b).

Ao longo das últimas décadas, a limitação dos sistemas computacionais tradicionais e, posteriormente, dos sistemas especialistas em resolver problemas complexos naturalmente solucionados por humanos levou à necessidade de automatização e à busca por extrair conhecimento sem intervenção humana. Esse cenário impulsionou o surgimento do *Machine Learning* (ML), sendo um ramo da Inteligência Artificial (IA) fundamentado em estatística (FACELI *et al.*, 2021; HUYEN, 2024).

O ML capacita computadores a aprenderem por conta própria a partir de padrões em dados, guiados pelo princípio da inferência indutiva. Essa capacidade se manifesta em três categorias principais de tarefas: Aprendizado Supervisionado, Não Supervisionado e por Reforço (ALMEIDA, CARVALHO E MENINO, 2020; GÉRON, 2019). Técnicas de ML são amplamente utilizadas na área da saúde para o diagnóstico de doenças (BEZERRA, ALMEIDA, 2024).

Uma pesquisa realizada por Trigueiro *et al.* (2023) apontou que os estudantes universitários são mais propensos a desenvolverem transtornos depressivos. Os resultados da pesquisa indicaram que 4,57% dos estudantes apresentavam sintomas graves de algum tipo de transtorno depressivo enquanto 20,51% apresentavam sintomas moderados. Com isso, surge uma oportunidade de trabalho voltada para o uso de técnicas de ML, com o intuito de investigar a saúde mental de estudantes universitários. É nesse sentido que surge uma questão norteadora para esta pesquisa: "*É possível utilizar técnicas de ML para detecção de transtorno depressivo em estudantes universitários?*".

1.1 Justificativa e Problemática

O suicídio é uma ocorrência multifacetada que pode afetar indivíduos de diferentes características demográficas e sociais. Segundo o MS, esse fenômeno ocupa o quarto lugar entre as causas que mais provocam mortes entre pessoas de 15 e 29 anos de idade do mundo todo (BRASIL, 2022). A WHO (2021) aponta que, anualmente, cerca de 800 mil casos de suicídio são registrados, ultrapassando doenças como *Human Immunodeficiency Virus* (HIV) e mortes causadas por violência. No Brasil, segundo o MS, são documentados aproximadamente 14 mil casos de suicídio por ano, com uma média de 38 suicídios diariamente (BRASIL, 2022).

Com o intuito de conter o alto número de suicídios entre os países americanos, a WHO (2021) introduziu a iniciativa *Live life*, que tem como uma de suas abordagens a identificação de pessoas com pensamentos e comportamentos suicidas, para apoiá-las continuamente. No Brasil, destaca-se a campanha Setembro Amarelo, cujo intuito é reforçar o alerta para a população sobre a necessidade de prevenção contra o suicídio e consequentemente, promover a conscientização acerca da saúde mental (WEBER, 2023).

A instituição americana Chegg (2023), especializada em tecnologia educacional, realizou um estudo com estudantes universitários de 15 diferentes países ao redor do mundo, onde foi possível identificar três fatores preocupantes em relação à saúde mental de alunos do ensino superior, sendo eles: esgotamento mental, sentimentos diários de ansiedade e sono insuficiente, com percentuais de 46%, 54% e 59%, respectivamente (CHEGG, 2023). Diante desse cenário, este trabalho surge com o intuito de analisar é possível aplicar técnicas de ML na detecção precoce de transtornos depressivos, permitindo que futuras intervenções sejam realizadas de forma mais eficaz.

1.2 Objetivos

1.2.1 Objetivo Geral

O objetivo geral deste trabalho é investigar a aplicação de técnicas de ML para auxiliar na detecção de transtornos depressivos em estudantes universitários.

1.2.2 Objetivos Específicos

Para atingir o propósito geral desta pesquisa, foram estabelecidos os seguintes objetivos específicos:

- Explorar os principais fatores preditivos relacionados a transtornos depressivos em estudantes universitários;
- Adquirir e pré-processar o conjunto de dados de saúde mental de universitários;
- Implementar modelos de ML a fim de auxiliar na detecção de transtornos depressivos;
- Avaliar os resultados de desempenho dos modelos de ML implementados;
- Analisar os resultados obtidos para propor recomendações que possam apoiar a saúde mental dos estudantes.

1.3 Organização do Trabalho

O presente trabalho está estruturado, além da introdução, da seguinte maneira: no Capítulo 2, é exposta a fundamentação teórica, com o objetivo de destacar essa área de estudo e apresentar trabalhos relacionados a esta pesquisa; no Capítulo 3, é descrita a metodologia que adotada; no Capítulo 4, os resultados dos experimentos são revelados e discutidos; e por fim, no Capítulo 5, são apresentadas as considerações finais e sugestões de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados alguns dos principais conceitos referentes à área de estudo investigada neste trabalho, a citar: Saúde Mental, Técnicas de *Machine Learning*, com seus tipos de aprendizado e os Trabalhos Relacionados.

2.1 Saúde Mental

Durante séculos, a ideia por trás do significado de saúde mental assumiu diferentes interpretações, e por um longo período, foi associada à noção de insanidade (SCULL, 2015). Entre os antigos hebreus, acreditava-se que transtornos mentais tinham origem sobrenatural, sendo resultante da desobediência às leis divinas. Essa transgressão não apenas acarretava graves consequências para o indivíduo, mas também o colocava sob a influência de forças malignas, interpretadas como um castigo divino (SCULL, 2015; QUEVEDO, 2019). Em contraste com os hebreus, os gregos antigos defendiam uma perspectiva naturalista sobre o conceito, atribuindo a ausência de saúde mental a desequilíbrios do corpo, sem associá-los a intervenções divinas (SCULL, 2015).

A WHO (2020) define que a saúde mental está diretamente relacionada ao bem-estar. Nesse contexto, Maggi (2006) esclarece que o termo bem-estar, nos dicionários latino-americanos, é formado pela palavra “bem”, que expressa um alto grau, e “estar”, que denota viver. Segundo o autor, durante o período renascentista, esse termo foi utilizado inicialmente para se referir à satisfação de necessidades físicas, como alimentação e descanso, e, posteriormente, às necessidades materiais.

Reforçando a importância desse conceito, a WHO (2020) destaca que o estado de bem-estar permite ao indivíduo desenvolver e executar suas atividades diárias de maneira eficaz, além de contribuir para a comunidade à qual pertence. A organização ainda enfatiza que fatores pessoais, sociais e socioeconômicos exercem influência significativa sobre a saúde mental. Complementando essa perspectiva, o MS ressalta que a saúde mental possui características biopsicossociais, e que esses aspectos contribuem para a construção e manutenção do bem-estar do indivíduo (BRASIL, 2019a). A atenção voltada à saúde mental evidencia a importância de identificar e compreender os transtornos mentais que podem comprometer o bem-estar e a qualidade de vida dos indivíduos.

Além das definições anteriores, a APA (2023) destaca o transtorno mental como uma condição caracterizada por um desequilíbrio nos processos mentais, dificuldades no controle emocional e comportamentos atípicos. Segundo dados da entidade, cerca de 3,8% da população mundial vivencia algum transtorno depressivo. A organização estima que cerca de

280 milhões de indivíduos convivem com esse transtorno, sendo ele 50% mais prevalente entre mulheres do que entre homens.

O transtorno depressivo, é um termo utilizado para indicar um quadro clínico caracterizado pela falta de interesse na realização das atividades do cotidiano e pelo desânimo, conforme considera Souza (2023). Nessa perspectiva, Martins (2022b) destaca que o transtorno depressivo transcende fronteiras geográficas, podendo acometer indivíduos em diferentes contextos socioculturais. Somada a estas contribuições, a APA (2023) em seu trabalho DSM-5-TR, define os diferentes tipos de transtornos depressivos, destacando, entre eles, o transtorno disruptivo da desregulação do humor, o transtorno depressivo maior, o transtorno depressivo persistente, o transtorno disfórico pré-menstrual, o transtorno depressivo induzido por substância/medicamento, transtorno depressivo devido a outra condição médica, outro transtorno depressivo especificado e transtorno depressivo não especificado. De acordo com a organização, fatores como temperamento, ambiente e genética podem influenciar o desenvolvimento desses transtornos, cujas principais características são apresentadas no Quadro 1.

O transtorno disruptivo da desregulação do humor é marcado pela irritabilidade persistente intensa, com maior prevalência entre 6 e 18 anos. Essa agitação se configura em dois quadros clínicos, sendo o primeiro os episódios recorrentes de ataques de fúria desencadeados em virtude do estado de frustração, resultando em ações verbais ou comportamentais, como a autoagressão. O segundo está relacionado com o humor predominantemente zangado, com o passar do tempo (APA, 2023).

O transtorno depressivo maior tem como principal característica a perda de motivação para realizar atividades diárias por pelo menos duas semanas seguidas. Embora possa surgir em qualquer fase da vida, indivíduos em período de puberdade são particularmente mais vulneráveis devido a mudanças hormonais (APA, 2023). Esse transtorno frequentemente se manifesta por meio de alterações no apetite ou peso, no sono e nos movimentos corporais. Além de baixa disposição, sentimento de culpa e a sensação de inferioridade, as capacidades cognitivas como raciocínio e concentração também são afetadas. Outro aspecto preocupante é a presença constante de pensamentos de morte ou de origem suicida (APA, 2023).

O transtorno depressivo persistente é caracterizado como um humor depressivo com duração de no mínimo dois anos, para adultos, e um ano para crianças e adolescentes, esse transtorno se instala de forma gradual e discreta. Quando presente desde a infância ou adolescência, tende a se integrar à personalidade do indivíduo (APA, 2023).

O transtorno disfórico pré-menstrual caracteriza-se por labilidade do humor, irritabilidade, disforia e ansiedade que ocorrem recorrentemente na semana anterior à menstruação, diminuindo após seu início. Podem surgir sintomas comportamentais e físicos (APA, 2023).

O transtorno depressivo induzido por substância/medicamento manifesta-se por sintomas depressivos que se desenvolvem durante ou logo após a intoxicação ou abstinência de uma substância (APA, 2023).

O transtorno depressivo devido a outra condição médica envolve um humor deprimido proeminente ou acentuada diminuição de interesse/prazer, que é uma consequência fisiopatológica direta de outra condição médica (APA, 2023).

O outro transtorno depressivo especificado é aplicado quando os sintomas depressivos estão presentes de forma clinicamente significativa, mas não preenchem todos os critérios de duração ou quantidade de sintomas exigidos para um diagnóstico de transtorno depressivo principal. Essa classificação é utilizada quando o profissional consegue identificar e registrar o motivo pelo qual o quadro não se encaixa totalmente nas categorias estabelecidas (APA, 2023).

O transtorno depressivo não especificado, por sua vez, é utilizado quando há sintomas depressivos relevantes, porém não há informações suficientes para especificar uma categoria diagnóstica precisa, sendo comum em contextos de atendimentos emergenciais ou quando a avaliação clínica ainda está em fase inicial, impossibilitando o detalhamento necessário para enquadrar o quadro em outra classificação mais específica (APA, 2023).

Quadro 1 – Diferenciação dos Transtornos Depressivos do DSM-5-TR

Transtorno	**Humor Dominante**	**Duração Mínima**	**Início Comum**
Disruptivo da Desregulação do Humor	Irritabilidade persistente e raiva	12 meses	Antes dos 10 anos
Depressivo Maior	Tristeza profunda, perda de interesse/prazer, pensamentos suicidas	2 semanas	Qualquer idade, pico na puberdade
Persistente	Humor depressivo crônico	2 anos para adultos (1 ano para crianças/adolescente)	Gradual
Disfórico Pré-menstrual	Labilidade, irritabilidade, ansiedade	Semana pré-menstrual	Após a menarca

Induzido por Substância/Medicamento	Humor depressivo ou diminuição de interesse/prazer	Durante/logo após uso de substância/medicamento	Enquanto usa ou durante a abstinência
Devido a Outra Condição Médica	Desânimo ou perda de motivação	Persistente, consequência fisiológica direta da condição médica	Varia (ex: agudo pós-AVC, precoce na doença de Parkinson)
Outro Transtorno Depressivo Especificado / Especificado	Humor deprimido, irritável, anedonia, mas não satisfaz todos os critérios de outros transtornos depressivos	Variável	Variável

Fonte: Autoria própria.

Dessa forma, este estudo se concentra na identificação do transtorno depressivo persistente. Nesse contexto, é relevante destacar que Martins (2022b) ressalta que o Brasil detém o maior índice de pessoas diagnosticadas com algum transtorno depressivo entre os países da América Latina. Corroborando essa perspectiva, dados de um estudo epidemiológico do MS indicam que aproximadamente 15,5% da população brasileira pode desenvolver esse transtorno ao longo da vida (BRASIL, 2019b). Como forma de enfrentamento a esse cenário, iniciativas como a campanha Setembro Amarelo foram desenvolvidas. Segundo Oliveira (2020), esse projeto, implementado no Brasil em 2015, tem como objetivo fornecer à população informações essenciais para a conscientização e prevenção de casos de suicídio.

No ambiente universitário, são recorrentes os relatos preocupantes relacionados à saúde mental dos estudantes. Trigueiro *et al.* (2023) argumentam que estudantes universitários são mais propensos a desenvolverem transtorno depressivo devido à exposição frente a situações de estresse. Nesse sentido, realizaram um estudo com 545 alunos de uma universidade privada do Ceará, com o objetivo de investigar elementos que influenciam o bem-estar individual. Os autores descobriram que 4,57% e 20,51% da quantidade total de universitários apresentavam sintomas graves e moderados de algum transtorno depressivo, respectivamente.

2.2 Técnicas de *Machine Learning*

De acordo com Faceli *et al.* (2021), ao longo das últimas décadas, grande parte dos sistemas computacionais, para resolver um determinado problema, implementava uma série

ordenada de passos com a finalidade de encontrar uma solução. Segundo os autores, embora essa abordagem fosse eficaz para lidar com determinados problemas, ela ainda não era suficiente para resolver desafios que são naturalmente solucionados por humanos, como o reconhecimento de um indivíduo por meio da identificação de suas características físicas, bem como a combinação de experiências anteriores com os conhecimentos adquiridos por meio da educação para embasamento de decisões conscientes. Nesse sentido, em meados da década de 1970, surgiram os sistemas especialistas (ou sistemas baseados em conhecimento), com o objetivo de fornecer soluções para problemas reais. Essa abordagem, baseava-se no conhecimento de um especialista de determinado domínio que então era implementado em um programa de computador por meio de regras lógicas (FACELI *et al.*, 2021).

Segundo Faceli *et al.* (2021), a estratégia de sistemas especialistas apresentava limitações que iam desde a influência da perspectiva individual do especialista até a falta de colaboração no desenvolvimento do sistema, muitas vezes motivada pelo receio de desligamento de sua função. Os autores destacam que o aumento gradual na complexidade dos problemas a serem processados, aliado à elevada escalabilidade dos dados e à velocidade em que os dados chegam aos usuários finais, impõe novos desafios ao desenvolvimento de soluções computacionais eficientes. Nesse contexto, surgiu a necessidade de automatização desses problemas, com o objetivo de extrair conhecimento sem a dependência de intervenção humana, introduzindo os primeiros conceitos relacionados ao ML (FACELI *et al.*, 2021; HUYEN, 2024).

De acordo com Almeida, Carvalho e Menino (2020), o ML é um ramo da IA constituído por técnicas desenvolvidas com base na estatística, que fornece a possibilidade de automatização de tarefas de acordo com a identificação de estruturas padronizadas nos conjuntos de dados ou por meio de um histórico de aprendizado. Já para Géron (2019), o ML se trata de um conhecimento e prática relacionado à codificação de computadores para que estes consigam ter a capacidade de autoaprendizado por meio de padrões extraídos de um conjunto de dados. Nesse contexto, conforme explicitado por Faceli *et al.* (2021), o processo de aprendizado dos modelos de ML é pautado pelo princípio da inferência indutiva, que confere a capacidade de inferir um padrão ou resultado genérico a partir de dados ou casos particulares.

Bezerra e Almeida (2024) destacam que, no setor da saúde, as técnicas de ML são vastamente aplicadas para enfrentar desafios complexos. Essas aplicações abrangem desde a predição e o auxílio no diagnóstico de diversas patologias até a otimização de planos de tratamento e a personalização de terapias. Os resultados obtidos por esses métodos, segundo

os autores, têm se mostrado altamente promissores no reconhecimento acurado de sintomas e na identificação de padrões clinicamente relevantes a partir de grandes volumes de dados. Tais capacidades são cruciais para a extração de *insights* valiosos, que, por sua vez, possibilitam o auxílio eficaz na classificação e triagem de pacientes, contribuindo para uma tomada de decisão mais rápida e para a implementação de intervenções mais direcionadas.

Segundo Almeida, Carvalho e Menino (2020), o campo das técnicas de ML é vasto e diversificado, abrangendo uma ampla gama de metodologias e estratégias para a resolução de problemas complexos. Essa amplitude metodológica se organiza fundamentalmente em três principais categorias, que refletem os distintos paradigmas de aprendizado e as diferentes formas de interação com os dados para a extração de inteligência. Essas categorias são: Aprendizado Supervisionado, Aprendizado não Supervisionado e Aprendizado por Reforço. Tal classificação, que serve como um pilar conceitual para a compreensão das diversas aplicações e fundamentos do ML. Nesse sentido, a Figura 1 mostra uma representação dos tipos de aprendizado e fornece um panorama inicial sobre as abordagens que moldam o desenvolvimento de sistemas inteligentes. Diante do exposto, é válido conhecer um pouco mais cada aprendizado, por esse motivo, eles serão elencados a seguir.

Figura 1 - Tipos de aprendizado em *Machine Learning*

Fonte: Almeida, Carvalho, Menino (2020).

2.2.1 Aprendizado Supervisionado

De acordo com Faceli *et al.* (2021), o aprendizado supervisionado é classificado como uma tarefa preditiva, cujo intuito é realizar estimativas por meio de um conjunto de

dados rotulados, que, segundo Almeida, Carvalho, Menino (2020), correspondem às soluções que o modelo deve aprender e replicar. Em tarefas de aprendizado supervisionado, o classificador ou preditor recebe dados de entrada acompanhados de suas respectivas saídas, com o objetivo de inferir a relação subjacente entre elas (ALMEIDA, CARVALHO, MENINO, 2020).

A rotulagem dos dados demanda a expertise de um profissional especialista na área do problema da qual se deseja trabalhar. Esse profissional detém o conhecimento necessário para discernir entre previsões verdadeiras e falsas, geradas a partir de um modelo de ML. Essa estratégia pode ser dividida em dois tipos de tarefas, que diferem com base no significado do rótulo: na classificação, os rótulos são discretos, enquanto na regressão, os rótulos são contínuos (FACELI *et al.*, 2021).

As tarefas de classificação caracterizam-se pela busca de uma fronteira de decisão, responsável por fazer a distinção entre os objetos. Para isso, utilizam rótulos discretos, que são valores finitos e qualitativos, resultando na atribuição do atributo alvo a uma classe específica (ALMEIDA, CARVALHO, MENINO, 2020).

No tocante às tarefas de regressão, a principal característica é a presença de rótulos numéricos contínuos, que consistem em um conjunto infinito de valores. O objetivo é estimar uma função que se aproxime dos valores da função original do problema, resultando em um atributo alvo quantitativo (FACELI *et al.*, 2021).

2.2.2 Aprendizado Não Supervisionado

O aprendizado não supervisionado faz parte das tarefas descritivas, isto é, objetiva encontrar similaridades entre os dados sem que haja a rotulagem dos mesmos (FACELI *et al.*, 2021). Técnicas como agrupamento e redução de dimensionalidade podem ser computadas nessa estratégia. Segundo Almeida, Carvalho e Menino (2020), o aprendizado não supervisionado é comumente empregado para revelar padrões ocultos e identificar anomalias em um conjunto de dados.

Em tarefas de agrupamento, o modelo de ML recebe como entrada, dados não rotulados e, a partir deles, cria grupos de registros com base nas similaridades entre esses objetos. Já em tarefas de redução de dimensionalidade, ocorre a diminuição do número de atributos em um conjunto de dados (ALMEIDA, CARVALHO, MENINO, 2020). A aplicação dessa estratégia pode otimizar o desempenho da técnica utilizada, ao reduzir a complexidade computacional.

A redução de dimensionalidade pode ser efetuada de duas maneiras: em agregação,

quando acontece a substituição de atributos antigos por novos, tendo como principal método o *Principal Component Analysis* (PCA); e em seleção de atributos, quando uma parcela dos atributos é retida e o restante é eliminado (FACELI *et al*., 2021).

2.2.3 Aprendizado Por Reforço

Esse tipo de aprendizado visa simular o comportamento humano na resolução de problemas específicos por meio de tentativa e erro. Ele se baseia em três áreas de pesquisa: psicologia do conhecimento humano, programação dinâmica e aprendizagem de diferença temporal (SHAKYA, PILLAI, CHAKRABARTY, 2023). Conforme Almeida, Carvalho e Menino (2020), essa abordagem capacita um agente a tomar decisões em cenários autênticos e a solucionar problemas potenciais, com esse processo sendo avaliado durante seu treinamento.

Ris-Ala (2023) define que os elementos do aprendizado por reforço são: o agente (modelo de ML), responsável pela tomada de decisão; o ambiente, que é contexto no qual o agente opera; a ação, que consiste no comportamento do agente; o estado, que representa a circunstância atual do agente; e a recompensa, um retorno (positivo ou negativo) concedido mediante o estado do agente. É a partir dessas recompensas que a política do agente é atualizada (ALMEIDA,CARVALHO,MENINO,2020;SHAKYA,PILLAI, CHAKRABARTY, 2023).

Segundo Almeida (2024), uma política em aprendizado por reforço define o conjunto de estados que o agente pode assumir, com base nas ações resultantes de suas interações com o ambiente externo. Conforme Géron (2019) afirma, essa política permite que o modelo determine a ação apropriada conforme as circunstâncias, visando otimizar a obtenção de recompensas.

2.3 Trabalhos Relacionados

Cruz, Flores e Quispe (2023) desenvolveram um modelo de ML com o classificador *Naive Bayes* (NB), para prever o nível de depressão entre estudantes universitários, e obtiveram resultados promissores sobre a doença. Para isso, utilizaram amostras de dados de 519 estudantes pertencentes à *Universidad Nacional del Altiplano*. Como instrumento de análise, os autores adotaram o *Hamilton Depression Rating Scale* (HDRS), que corresponde a uma escala para determinar níveis de depressão proposto pela *National Institute of Mental Health* nos Estados Unidos.

O HDRS consiste nos seguintes níveis: estado normal, depressão leve, depressão moderada, depressão grave e depressão muito grave; e todos eles foram utilizados para a

rotulagem dos dados. O modelo proposto pelos pesquisadores demonstrou desempenho eficaz na detecção de casos positivos, em especial, aos níveis grave e moderado, com valores da métrica de *recall* de 72,41% e 57,14% respectivamente, indicando potencial para identificação de pessoas com níveis altos de depressão. A técnica criada por Cruz, Flores e Quispe (2023) obteve uma acurácia de 78,03%, mostrando ser uma ótima escolha para o diagnóstico precoce da doença em um ambiente acadêmico.

Gil, Kim e Min (2022) investigaram modelos preditivos do ML para o diagnóstico de riscos de depressão em estudantes coreanos, tendo como elementos de identificação componentes relacionados à influência da família e aspectos do próprio indivíduo. Com esse objetivo, utilizaram em seu estudo uma amostra de dados composta por 513 indivíduos, sendo 171 referentes a tríades formadas por mãe, pai e filho, representando o núcleo familiar do estudante. Esses dados foram utilizados nos classificadores *Support Vector Machine* (SVM), *Random Forest* (RF) e no regressor *Sparse Logistic Regression* (SLR). Os autores relataram que tiveram dificuldades em seu trabalho devido à alta dimensionalidade dos dados, que ultrapassava o quantitativo de exemplos, o que acabava afetando as estimativas feitas por técnicas regressoras e, consequentemente, dificultava a interpretação dos resultados. Por outro lado, reportaram que as técnicas SVM e RF não foram impactadas pelo grande número de atributos, considerando que são capazes de lidar com grandes quantidades de variáveis, sendo o RF o algoritmo de melhor desempenho em relação aos outros dois, com acurácia de 86,27%. O estudo também mostrou que a depressão materna, as doenças respiratórias e o câncer paterno, elementos relacionados à família, bem como a coesão familiar, o apego evitativo-medroso, o neuroticismo, a conscienciosidade e a autoavaliação da saúde mental, aspectos associados ao estudante, foram considerados fatores importantes que contribuem para a predição de casos de depressão em estudantes universitários.

Sandhu *et al.* (2022) desenvolveram um modelo de ML baseado no algoritmo *k-Nearest Neighbors* (KNN) para prever níveis de depressão e ansiedade em estudantes universitários no Paquistão. O estudo utilizou dados de 1.366 participantes, coletados por meio de questionários validados clinicamente, incluindo a *Hospital Anxiety and Depression Scale* (HADS), que categorizou os estados mentais em normal, *borderline* (limítrofe) e anormal (depressão/ansiedade clinicamente relevantes). Os autores destacaram que 22% dos estudantes apresentaram níveis anormais de depressão, enquanto 42,5% foram classificados com ansiedade anormal, reforçando a prevalência desses distúrbios no ambiente acadêmico. Para a classificação, os pesquisadores empregaram múltiplas variações do KNN (*weighted, cosine, cubic*), com e sem a técnica de PCA para redução de dimensionalidade. O modelo

KNN ponderado (*weighted* KNN) obteve os melhores resultados, com 76,6% de acurácia na detecção de depressão quando combinado com a técnica de PCA.

Upadhyay, Mohapatra e Singh (2024) propuseram um modelo baseado em ML para a avaliação precoce do Transtorno Depressivo Persistente (TDP), utilizando uma abordagem de combinação de classificadores. Para o estudo, foram coletados dados comportamentais de 137 estudantes universitários por meio de questionários baseados em escalas clínicas como a HDRS e a *Cornell Dysthymia Rating Scale* (CDRS). Os autores aplicaram a técnica *SelectKBest* para extração das melhores características e compararam diferentes algoritmos de classificação, incluindo KNN, *Multilayer Perceptron* (MLP), *XGBoost* e SVM com diferentes *kernels*, como *Linear*, *Sigmoid* e *Radial Basis Function* (RBF), além de variações como o *LinearSVC*. Os resultados indicaram que a abordagem baseada em SVM proposta apresentou desempenho superior em relação às demais, alcançando uma acurácia de 89,4%.

Xiaocheng (2023) investigou a aplicação do algoritmo de *Decision Tree* (DT) na avaliação da saúde mental de estudantes universitários. O estudo utilizou um conjunto de dados com 1.574 registros e adotou o algoritmo C4.5 para construir um modelo de classificação voltado à detecção de sintomas psicológicos, com ênfase em dimensões como ansiedade, estresse e relacionamentos interpessoais. A construção do modelo seguiu uma abordagem recursiva de divisão dos dados e incorporou técnicas de poda pós-processamento para reduzir o sobreajuste e simplificar a estrutura da árvore. Os resultados mostraram alta acurácia nas previsões, onde o modelo apresentou taxas de acerto superiores a 91% em todas as categorias avaliadas. O estudo concluiu que a técnica de DT é eficaz na identificação de fatores psicológicos relevantes e pode auxiliar psicólogos e instituições educacionais na detecção precoce de transtornos mentais entre estudantes.

Recapitulando os trabalhos relacionados a esta pesquisa, o Quadro 2 foi criado. Nele, é possível ver um resumo geral de todas as técnicas consideradas pelos autores, assim como o resultado do melhor classificador testado.

Quadro 2 – Comparativo dos Trabalhos Relacionados

Trabalho	**Usou dados de Universitários**	**Modelos**	**Melhor Modelo**
Cruz, Flores e Quispe (2023)	Sim	NB	NB

Gil, Kim e Min (2022)	Sim	SVM, RF, SLR	RF
Sandhu *et al.* (2022)	Sim	KNN	KNN
Upadhyay, Mohapatra e Singh (2024)	Sim	KNN, MLP, *XGBoost,* SVM	SVM
Xiaocheng (2023)	Sim	DT	DT

Fonte: Autoria própria.

3 METODOLOGIA DE PESQUISA

A pesquisa realizada neste trabalho se constitui de natureza aplicada, descritiva, com abordagem quantitativa, fundamentada na comparação de modelos de ML para detecção de transtornos depressivos em estudantes universitários. A pesquisa aplicada, segundo Vieira, Leite e Kuhn (2023), permite o estabelecimento de investigações com o intuito de contribuir para a solução de problemas, que possam existir no contexto do cotidiano ao qual o investigador está inserido por meio da aplicação prática do conhecimento elaborado.

A pesquisa descritiva tem como principal intuito relatar fatos e características de uma realidade específica a ser analisada, podendo ser conduzida mediante diferentes estratégias para a coleta de dados, com ou sem padronização, e sem interferência por parte do pesquisador (ALMEIDA, 2021; SAMPAIO, 2022). A abordagem quantitativa de pesquisa caracteriza-se pelo uso de dados numéricos para analisar e interpretar informações de forma objetiva e mensurável (ALMEIDA, 2021).

3.1 Proposta Metodológica

A proposta metodológica para a comparação de modelos de ML na detecção de transtornos depressivos foi estabelecida com base nas etapas de Amante e Morgado (2001), adaptadas para os objetivos deste estudo. Esse processo foi dividido nas seguintes fases: Concepção, Planejamento e Implementação.

A fase de concepção tem como objetivo definir os aspectos iniciais do estudo, incluindo o público-alvo, a ideia principal e a definição dos conteúdos a serem tratados. Já a fase de planejamento se configura com o uso de técnicas de ML na obtenção de resultados de treinamento, utilizando o conjunto de dados *depression_anxiety_data*, disponível na plataforma *Kaggle*[1], que é formado por 783 registros de estudantes da Universidade de *Lahore*.

Os atributos presentes no conjunto de dados escolhido incluem variáveis demográficas e sociais, como idade e gênero, além de indicadores de saúde, como o *Body Mass Index* (BMI) e sua classificação na *World Health Organization* (WHO). Além disso, são fornecidos dados detalhados sobre a saúde mental dos participantes, como os *scores* de *Patient Health Questionnaire* (PHQ) e *Generalized Anxiety Disorder* (GAD), que avaliam a gravidade de sintomas relacionados à depressão e ansiedade, respectivamente, além de informações sobre diagnósticos, tratamentos prévios, severidade dos sintomas e pensamentos suicidas. Tais características são relevantes para a criação de classificadores que possam identificar padrões significativos, permitindo a análise e auxílio na detecção de transtornos

[1] Disponível em: https://www.kaggle.com/code/geovaniwoll/machine-learningproject. Acesso em: 25 maio 2025.

depressivos em estudantes universitários.

Por fim, a fase de implementação constitui-se na execução e avaliação comparativa dos modelos de ML sobre o conjunto de dados previamente preparado. Para isso, foram selecionados os melhores modelos reportados pelos trabalhos relacionados (Quadro 2), a saber: KNN, SVM, NB, DT e RF; e seus resultados serão mostrados no próximo Capítulo.

3.2 Modelos de *Machine Learning* utilizados

Cada modelo considerado tem sua particularidade. Especificamente, o KNN é um algoritmo de aprendizado supervisionado que se baseia no princípio de que amostras com características semelhantes tendem a pertencer à mesma classe. Sua operação central funciona calculando a distância entre a amostra a ser classificada e todas as amostras do conjunto de treinamento, sendo a distância Euclidiana a métrica mais comumente utilizada (FACELI *et al.*, 2021; GÉRON, 2019).

Já o modelo SVM tem como objetivo principal encontrar um hiperplano que separe os dados de classes distintas, de forma linear ou não linear. Isso pode ser realizado mediante a criação de uma fronteira de decisão (HARRISON, 2019). E o NB trata-se de um algoritmo probabilístico, de aprendizado supervisionado, que é baseado no teorema de Bayes, com a premissa de que as características são independentes entre si, dado o valor de uma classe alvo (FACELI *et al.*, 2021).

Quanto ao DT, trata-se de um modelo de aprendizado supervisionado, que permite a interpretação visual do conhecimento extraído. Baseia-se na estratégia de dividir para conquistar, onde o problema original é recursivamente dividido em subproblemas mais simples (FACELI *et al.*, 2021). Para isso, os dados são particionados em subconjuntos por meio de regras baseadas em características, utilizando critérios de impureza, como o coeficiente de Gini e a Entropia (GÉRON, 2019). Por fim, o RF destaca-se como um algoritmo baseado em conjunto (*ensemble*), utilizado principalmente em tarefas de classificação e regressão (GÉRON, 2019). Ele funciona por meio da construção de múltiplas DT independentes, cujos resultados são combinados para produzir uma predição mais robusta e precisa (FACELI *et al.*, 2021).

4 RESULTADOS

Neste capítulo, são apresentados os resultados detalhados da pesquisa, englobando a

caraterização do conjunto de dados, as características selecionadas e a aplicação dos modelos de ML, especificamente, SVM, KNN, NB, DT e RF. Os resultados expostos representam o desfecho da metodologia descrita no Capítulo 3, abrangendo o desempenho desses algoritmos nos cenários de experimentação definidos.

4.1 Tecnologias e ferramentas

Para a condução dos experimentos, utilizou-se a linguagem de programação *Python* na versão 3.11.9, e, como plataforma, optou-se pelo *Anaconda Navigator* na versão 2.6.4, que simplifica o gerenciamento de pacotes e ambientes virtuais. Dentro da plataforma, o *Jupyter Lab* na versão 4.2.5, foi adotado como ambiente principal para a escrita e execução dos códigos relacionados aos modelos. O *Jupyter Lab* fornece uma interface interativa e versátil, permitindo a combinação de códigos, visualizações e textos explicativos em um único ambiente, o que facilitou a exploração dos dados e a documentação do processo.

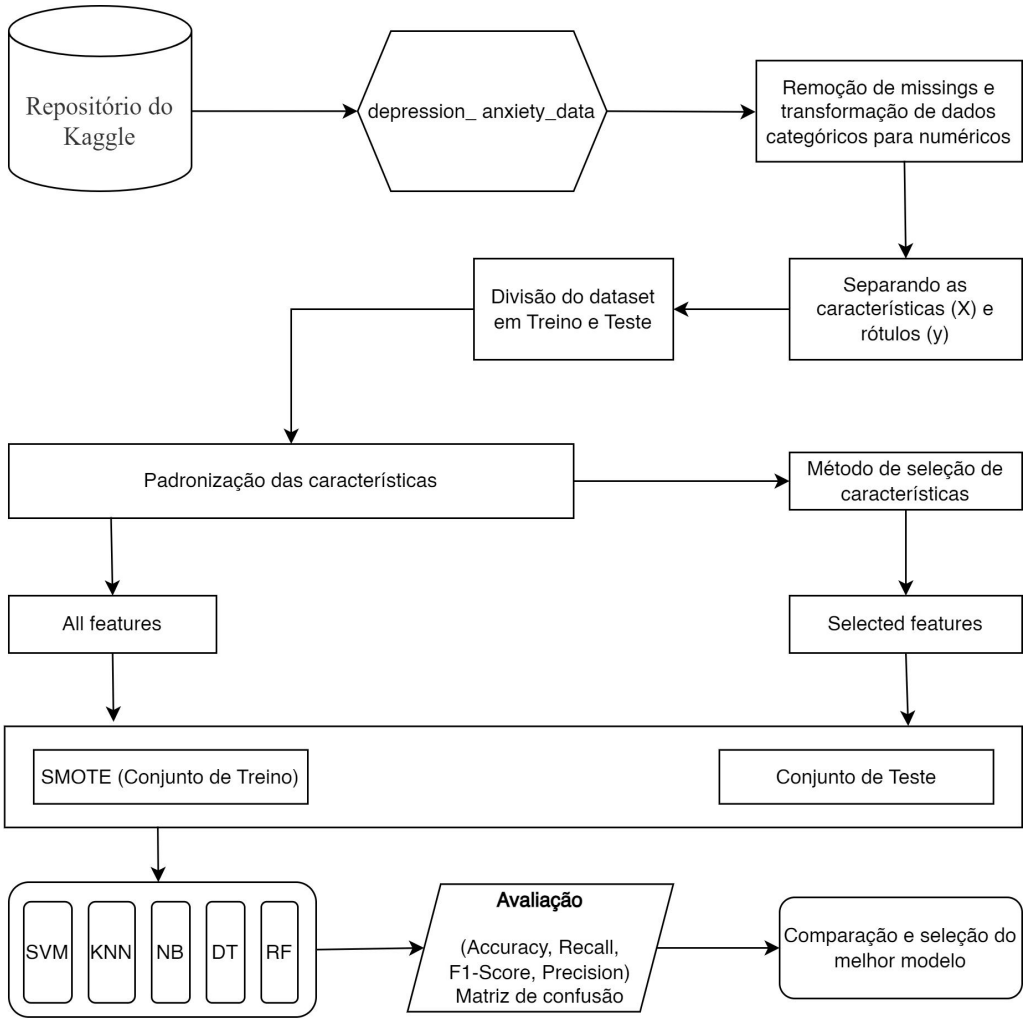
Para o controle de versão dos códigos, foi utilizado o *Git* em conjunto com o *GitHub*. O *Git* permitiu o rastreamento de alterações nos códigos, enquanto o *GitHub* serviu como repositório remoto, garantindo a segurança e a acessibilidade do projeto. Essa combinação foi fundamental para manter um histórico organizado das modificações e para facilitar a revisão e o compartilhamento do trabalho.

As bibliotecas utilizadas variaram conforme o modelo de ML aplicado, mas algumas foram comuns a todos os métodos, como: *pandas*, *numpy*, *seaborn, collections* e *matplotlib*, que são amplamente utilizadas para manipulação de dados, análise exploratória e visualização. Além disso, foram empregadas bibliotecas como *Scikit-learn* (para implementação e análise dos algoritmos), *imblearn* (para balanceamento de classes, com *Synthetic Minority Over-sampling Technique* (SMOTE)) e *graphviz* (para visualização de árvores de decisão). A escolha dessas bibliotecas foi guiada por critérios como eficiência, facilidade de uso e compatibilidade com os modelos propostos.

4.2 Configuração dos experimentos

A Figura 2 ilustra o planejamento que guiou a execução dos experimentos deste trabalho. O ponto de partida foi a aquisição dos dados da plataforma *Kaggle*, especificamente o conjunto de dados *depression_anxiety_data*. Essa base de dados, crucial para o treinamento e a avaliação dos modelos propostos, é composta por 19 características iniciais. A fim de otimizar a *performance* dos modelos e garantir a qualidade dos dados, todas essas características foram submetidas a uma estratégia de pré-processamento.

Figura 2 - Planejamento para Implementação dos Modelos



Fonte: Autoria própria.

Inicialmente, o conjunto de dados adquirido continha 783 instâncias. No entanto, ao realizar-se uma análise sobre os dados e iniciar a fase de limpeza, foi necessário remover 18 instâncias, restando 765. Além disso, foram identificados 8 valores ausentes representados como "*Not Available*" na variável *who_bmi*, os quais foram removidos por meio de uma filtragem, restando 757 instâncias no conjunto de dados.

No conjunto de treinamento, a variável-alvo (*y_train*) apresentava um desequilíbrio significativo: a Classe "0" contava com 484 instâncias, enquanto a Classe "1" possuía apenas 45. Após a aplicação da técnica SMOTE, ambas as classes foram equilibradas, totalizando 484 instâncias para cada uma.

Após a execução do método *SelectKBest,* que emprega critérios estatísticos para avaliar e ranquear a relevância dos atributos, as melhores características foram: *phq_score,*

depressiveness, depression_treatment, anxiety_diagnosis e anxiety_treatment. Tais atributos, ao demonstrarem a mais forte correlação com a variável-alvo, consolidaram-se como elementos essenciais para a construção dos modelos preditivos. Essa seleção de fatores é crucial, pois permite que os algoritmos de ML concentrem seu aprendizado nas informações mais impactantes, otimizando o processo de reconhecimento de padrões e, consequentemente, aprimorando significativamente a capacidade de auxílio preditivo na detecção de transtornos depressivos em estudantes universitários.

4.3 Pré-processamento dos Dados

Das 19 características existentes, uma delas foi removida por ser um identificador (*id*), cuja presença não traz contribuição para o desempenho dos modelos. A instrução para essa remoção está representada na Figura 3. O comando de exclusão da característica é composto pelos seguintes elementos: *df_data*, que corresponde a variável que contém todo o conjunto de dados; *drop*, que corresponde a um método do *pandas* capaz de remover linhas ou colunas da variável *df_data*; *"id"*, que diz respeito ao nome da coluna passado como argumento ao parâmetro *label* do método; *axis = 1*, que determina o eixo para a remoção (0 para linhas e 1 para colunas), onde a operação é realizada apenas na coluna especificada da variável *df_data*; e *inplace = True*, que corresponde ao parâmetro que define se a alteração será aplicada diretamente na variável *df_data* (quando o valor *True*) ou se será criada uma cópia com a alteração (quando o valor é *False*).

Figura 3 - Remoção da coluna *"id"* do conjunto de dados

```
df_data.drop('id', axis = 1, inplace = True)
```

Fonte: Autoria própria.

Outra etapa de análise do conjunto de dados consistiu na identificação de valores ausentes (representados por *NaN*) e inconsistências (dados com "*Not Available"*). O uso de valores nulos, como destacado por Géron (2019), é problemático na execução de modelos de ML, pois afeta negativamente o treinamento. Para mitigar esse problema, a instrução para remover esses valores em todo o conjunto de dados foi aplicada, conforme detalhado na Figura 4. Além disso, na Figura 5, é ilustrado a instrução utilizada para a filtragem dos dados com o objetivo de eliminar dados inconsistentes.

Figura 4 - Remoção de linhas com valores ausentes do conjunto de dados

```
df_data.dropna(inplace = True)
```

Fonte: Autoria própria.

Figura 5 - Removendo linhas com *'Not Available'*

```
df_data= df_data[df_data['who_bmi'] != 'Not Available']
```

Fonte: Autoria própria.

Para adequar o conjunto de dados aos modelos de ML, que em sua maioria exigem entradas numéricas, foram realizadas transformações nas variáveis categóricas. Para isso, a técnica de *One-Hot Encoding* foi aplicada às variáveis *who_bmi*, *gender*, *depression_severity* e *anxiety_severity*, utilizando a função *get_dummies* da biblioteca *pandas*. Essa função converte cada categoria dessas variáveis em colunas binárias, onde 1 representa a presença da categoria e 0 sua ausência (HARRISON, 2019). Essa abordagem foi escolhida para garantir que as relações ordinais ou nominais entre as categorias fossem preservadas de forma adequada, sem impor ponderações arbitrárias ou suposições sobre distâncias entre os valores. As transformações foram realizadas mediante as instruções da Figura 6, sendo composta pelos seguintes elementos: *df*, que representa a variável que armazena o conjunto de dados completo; *pd.get_dummies*, uma função do *pandas* empregada para implementar a técnica de *One-Hot Encoding*; *data = df_data*, que indica o *DataFrame* de origem que será transformado; *columns* = [*'who_bmi'*, *'gender'*, *'depression_severity'*, *'anxiety_severity'*], que especifica as colunas categóricas a serem convertidas em variáveis *dummy*; e *dtype = int*, que assegura que as colunas geradas sejam codificadas diretamente como valores inteiros (1 e 0), em vez de booleanos (*True* e *False*). Além disso, uma lista denominada “colunas” é criada, contendo as características *depressiveness*, *suicidal*, *depression_treatment*, *anxiousness*, *anxiety_diagnosis*, *anxiety_treatment*, *sleepiness*, *depression_diagnosis*. Por fim, um *loop for* percorre cada item dessa lista, utilizando o método *.astype(int),* também da biblioteca *pandas,* para converter os valores booleanos (*True* e *False*) em inteiros (1 e 0).

Figura 6 - Aplicação de *One-Hot Encoding*

```
#One-hot-encoding
Df = pd.get_dummies(data = df_data, columns = ['who_bmi',
'gender','depression_severity', 'anxiety_severity'], dtype = int)
#Mudança das características do tipo bool para int
colunas = ['depressiveness','suicidal', 'depression_treatment', 'anxiousness',
'anxiety_diagnosis', 'anxiety_treatment', 'sleepiness',
'depression_diagnosis']
for coluna in colunas:
```

Fonte: Autoria própria.

4.4 Separação e Padronização dos Dados

Logo após os ajustes relacionados à limpeza, o conjunto de dados foi dividido em duas partes: as variáveis independentes (X) e a variável dependente (y). Para isso, a característica *depression_diagnosis*, que representa o diagnóstico de depressão, foi considerada como atributo alvo da predição, e, por esse motivo, foi separada do restante do *dataframe*, sendo atribuído à variável y, conforme mostrado na Figura 7.

Figura 7 - Separação das características e rótulos

```
X = df.drop(columns=['depression_diagnosis'])
y = df['depression_diagnosis']
```

Fonte: Autoria própria.

A variável X armazena todas as características do conjunto de dados, exceto o atributo alvo, isto é, *depression_diagnosis*. Para isso, utiliza-se o método *drop*, conforme aplicado anteriormente, removendo, dessa vez, a coluna especificada no parâmetro *columns* por meio de indexação (Figura 7). De modo semelhante, na instrução seguinte, a variável y recebe apenas os rótulos da classe, que são definidos da seguinte forma: Classe “0” para representar a ausência de depressão (valores iguais a 0) e Classe “1” para indicar a presença de depressão (valores iguais a 1).

Para garantir que os modelos sejam treinados e avaliados, os dados foram divididos em conjuntos de treinamento e teste, conforme a instrução mostrada na Figura 8, utilizando a função *train_test_split*. Sendo assim, as variáveis independentes (X) e a variável dependente (y) foram particionadas em dois grupos: um para treinamento (*X_train* e *y_train*) e outro para

teste (*X_test* e *y_test*).

Figura 8 - Divisão dos dados em conjuntos de treinamento e teste

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
       test_size=0.3, random_state=42, stratify=y)
```

Fonte: Autoria própria.

A proporção definida para o conjunto de teste foi de 30% (*test_size=0.3*), enquanto os 70% restantes foram destinados ao treinamento. Além disso, o parâmetro *random_state=42*, que foi utilizado para garantir a reprodutibilidade da divisão, assegurando que os mesmos dados sejam selecionados sempre que o código for executado. Por fim, o parâmetro *stratify=y* foi incluído para preservar a proporção original da variável dependente (y) em ambos os conjuntos, o que é especialmente importante em casos de desbalanceamento de classes, garantindo que a distribuição de *depression_diagnosis* seja mantida tanto no treinamento quanto no teste.

Após a divisão dos dados em conjuntos de treinamento e teste, foi realizada a normalização das variáveis numéricas, que apresentavam diferentes escalas (Figura 9). Para isso, utilizou-se o *StandardScaler*, uma técnica que padroniza as características, transformando-as para uma distribuição com média zero e desvio padrão igual a um (HARRISON, 2019).

Figura 9 - Padronização de características numéricas

```
scaler = StandardScaler()
#Selecionando apenas as características com diferentes escalas
cols_to_scale = ['age', 'bmi', 'phq_score', 'gad_score', 'epworth_score']
X_train[cols_to_scale] = scaler.fit_transform(X_train[cols_to_scale])
X_test[cols_to_scale] = scaler.transform(X_test[cols_to_scale])
```

Fonte: Autoria própria.

Conforme mostrado na Figura 9, as colunas selecionadas para essa transformação foram: *age*, *bmi*, *phq_score*, *gad_score* e *epworth_score*, pois essas variáveis possuíam escalas distintas: idade (*age*) em anos, índice de massa corporal (*bmi*) em kg/m², e os *scores* dos questionários (*phq_score*, *gad_score* e *epworth_score*) em escalas variadas. A

normalização foi aplicada ao conjunto de treinamento (*X_train*) utilizando o método *fit_transform*, que ajusta o *scaler* aos dados e já realiza a transformação. Para o conjunto de teste (*X_test*), foi utilizado apenas o método *transform*, garantindo que a mesma escala aprendida no treinamento seja aplicada, evitando vazamento de dados. As demais variáveis não passaram por essa transformação, pois já estavam em escalas compatíveis ou eram categóricas previamente codificadas, não necessitando de ajustes adicionais.

4.5 Cenários de Experimentação e Balanceamento de Classes

Ao realizar a padronização dos dados, foram definidos dois cenários para os experimentos com os modelos: o Cenário 1, que corresponde ao treinamento dos modelos com todas as características do conjunto de dados, e o Cenário 2, que diz respeito ao treinamento dos modelos com apenas as melhores características.

Para ambos os cenários, foi realizado um balanceamento dos dados devido ao significativo desequilíbrio na variável-alvo de treino (*y_train*). Para lidar com essa disparidade, aplicou-se a técnica SMOTE, que gera amostras sintéticas da classe minoritária, equilibrando a distribuição dos rótulos. A técnica foi aplicada utilizando o parâmetro *random_state=42* para garantir a reprodutibilidade dos resultados. A função *fit_resample* foi utilizada para ajustar o SMOTE aos dados de treinamento com todas as características e apenas as melhores características (*X_train_selected*, *y_train*), gerando um novo conjunto balanceado (Figura 10).

Figura 10 - Balanceamento das classes com SMOTE

```
smote = SMOTE(random_state=42)
X_train_balanced, y_train_balanced = smote.fit_resample(X_train, y_train)
print(f"Distribuição das classes após SMOTE: {Counter(y_train_balanced)}")
```

Fonte: Autoria própria.

É importante destacar que o balanceamento foi realizado apenas no conjunto de treinamento, evitando qualquer vazamento de dados para o conjunto de teste, o que poderia comprometer a avaliação dos modelos. Após a aplicação do SMOTE, a distribuição das classes foi verificada por meio da função *Counter*, que exibe a quantidade de exemplos em cada classe, permitindo confirmar que o desbalanceamento foi corrigido.

4.6 Seleção de Características

A respeito do Cenário 1, após a aplicação do SMOTE, seguiu-se com treinamento dos modelos. Já em relação ao Cenário 2, logo após a transformação dos dados, foi aplicado a estratégia de seleção de características por meio do método *SelectKBest*, disponível na biblioteca *Scikit-learn*. Essa técnica tem como objetivo avaliar a relevância de cada atributo em relação à variável-alvo, utilizando critérios estatísticos para selecionar os k atributos mais significativos[2]. No contexto deste estudo, o hiperparâmetro *score_func* foi configurado como *f_classif*, que é uma função estatística baseada em ANOVA, enquanto o valor de k foi definido como 5 (Figura 11).

Figura 11 - Seleção de características

```
selector = SelectKBest(score_func = f_classif, k=5)
X_train_selected = selector.fit_transform(X_train, y_train)
X_test_selected = selector.transform(X_test)
view_features = X.columns[selector.get_support()]
print("Características selecionadas:\n", view_features)
```

Fonte: Autoria própria.

Conforme mostrado na Figura 11, o método *get_support* foi utilizado para identificar e exibir essas características, que foram posteriormente exploradas nas etapas seguintes do estudo. Somente após a seleção das melhores características, foi aplicada a técnica SMOTE no conjunto de treinamento, seguida pelo treinamento dos modelos. A motivação por trás dessa abordagem é garantir que o balanceamento das classes seja realizado com base nas variáveis mais relevantes, evitando que ruídos ou atributos menos significativos influenciem na geração de exemplos sintéticos.

4.7 Avaliação dos Modelos

Após o treinamento dos modelos, procedeu-se à avaliação individual de cada um, comparando seu desempenho nos dois cenários propostos. Para isso, utilizou-se a acurácia média da validação cruzada, que divide os dados em múltiplos subconjuntos (*folds*) para treinar e testar o modelo em diferentes combinações, fornecendo uma estimativa robusta e

[2] Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html. Acesso em: 25 maio 2025.

confiável do desempenho e reduzindo o risco de *overfitting*, que ocorre quando um modelo perde a capacidade de generalização para dados novos, devido ao ajuste excessivo aos dados de treinamento (FACELI *et al.*, 2021). Além disso, a matriz de confusão foi analisada para entender o comportamento dos modelos em relação às classes negativa (Classe “0”) e positiva (Classe “1”), mostrando a quantidade de previsões corretas e incorretas, permitindo identificar possíveis dificuldades na classificação de cada classe (Figura 12).

Figura 12 - Matriz de confusão e Relatório de Classificação

```
matriz_conf = confusion_matrix(y_test, y_pred)
sns.heatmap(matriz_conf, annot=True, fmt='d', cmap='Blues',
            xticklabels=['0', '1'],
            yticklabels=['0', '1'])
plt.title('Matriz de Confusão')
plt.xlabel('Valor Predito')
plt.ylabel('Valor Real');
print('Relatório do modelo\n\n', classification_report(y_test, y_pred))
```

Fonte: Autoria própria.

A Figura 12 apresenta a implementação da função *classification_report*, que gera um relatório detalhado com métricas como acurácia, *precision*, *recall* e *F1-score*. Esse relatório permite uma análise completa do desempenho dos modelos, considerando todas as características e as melhores características (*X_test_selected*) do conjunto de dados.

A acurácia mede a proporção de previsões corretas, a *precision* indica a precisão das previsões positivas, o *recall* avalia a capacidade de identificar corretamente os casos positivos, e o *F1-score* representa o equilíbrio entre *precision* e *recall* usando a média harmônica, sendo especialmente útil em cenários onde as classes são desbalanceadas, conforme discutido por Harrison (2019).

A avaliação detalhada é essencial para compreender as nuances do desempenho dos modelos em cada cenário, garantindo uma análise crítica e embasada para a escolha do modelo mais adequado ao problema em questão. A seguir, são explanados os resultados de cada modelo, para posteriormente serem comparados com base em seu desempenho nos dois cenários propostos. Com os resultados alcançados, pretende-se compreender qual modelo teve melhor eficácia para auxiliar na detecção da transtornos depressivos em estudantes universitários.

4.7.1 SVM

Os experimentos realizados com o modelo SVM foram conduzidos utilizando a configuração padrão da classe SVC, conforme implementado na biblioteca *Scikit-learn*. Todo o código construído para a validação deste modelo está disponível no *Github*[3]. Nesse sentido, a única exceção foi a mudança do *kernel*, que foi ajustado para os tipos RBF, *Linear* e *Polynomial*, com o objetivo de avaliar o desempenho dos modelos nos dois cenários do estudo (ou seja, todas as características e as melhores características), por meio da comparação entre eles de seus desempenhos individuais. Nesse cenário, o modelo do SVM foi instanciado a partir da classe SVC (*Scikit-learn*) e os diferentes *kernels* foram passados como argumentos; assim como o valor do *random_state*, que foi padronizado em 42, com o objetivo de controlar a semente de aleatoriedade, possibilitando a reprodutibilidade dos experimentos.

Os modelos desenvolvidos foram submetidos ao processo de validação cruzada, essencial para avaliar seu comportamento em diferentes partições do conjunto de dados e garantir uma estimativa mais robusta de sua capacidade de generalização. Subsequentemente, cada modelo foi treinado em dois cenários distintos: um utilizando todas as características disponíveis e outro empregando apenas as melhores características selecionadas, conforme detalhado na metodologia. Após o treinamento, as predições foram geradas sobre o conjunto de teste (dados não vistos), e a avaliação abrangente do desempenho de cada modelo foi conduzida. Esta avaliação detalhada foi realizada por meio da análise da matriz de confusão e do relatório de classificação, ferramentas cruciais que permitem uma compreensão aprofundada dos acertos e erros do modelo em relação a cada classe, oferecendo *insights* sobre sua *performance* preditiva.

Para a comparação entre *kernels* RBF, *Linear* e *Polynomial*, os modelos foram avaliados em termos de *recall*, *precision e F1-score*, sendo esta última a métrica determinante para a escolha do melhor *kernel*, por representar a média harmônica entre *precision* e *recall*, além de atribuir maior peso a valores mais baixos, o que contribui para uma análise mais justa e equilibrada (GÉRON, 2019). Desse modo, como estratégia para a comparação entre os *kernels*, foi extraída a macro *avg* de cada critério de avaliação, fornecida pelo *classification report* com base nos valores da Classe “0” e da Classe “1”, sendo os resultados mostrados nas Tabelas 1 e 2.

Conforme visto na Tabela 1, o *kernel Linear* destacou-se em relação aos outros *kernels* no Cenário 1, devido ao alcance de resultados mais consistentes nas métricas de

[3] https://github.com/WesleydosSantos/machine-learning-tecnicas/blob/main/Modelos/svm.ipynb.

avaliação do modelo. Além disso, o *kernel* RBF e o *Polynomial* alcançaram resultados aproximados, indicando um desempenho mediano mas inferior ao *Linear*.

Tabela 1 – Comparativo dos *kernels* utilizando todas as características

Kernel	*Precision*	*Recall*	*F1-score*
Linear	85%	78%	81%
RBF	79%	77%	78%
Poly	71%	78%	74%

Fonte: Autoria própria.

Tabela 2 – Comparativo dos *kernels* utilizando as melhores características

Kernel	*Precision*	*Recall*	*F1-score*
Linear	80%	87%	83%
RBF	83%	88%	85%
Poly	85%	88%	86%

Fonte: Autoria própria.

Com relação ao Cenário 2, o *kernel Linear* apresentou um valor de *precision* inferior em comparação ao Cenário 1. Contudo, observou-se uma melhora nas métricas de *recall* e *F1-score*. Apesar dessas melhorias, o desempenho do *kernel Linear* não foi suficiente para superar os *kernels* RBF e *Polynomial* que também alcançaram valores mais consistentes em contraste ao primeiro cenário (Tabela 2). O *kernel Polynomial* destacou-se em comparação aos outros dois *kernels* (RBF e *Linear*), apresentando resultados mais sólidos nas métricas de avaliação dos modelos.

Com base na análise comparativa entre os três *kernels* (*Linear*, RBF e *Polynomial*), no primeiro e no segundo cenário, foi identificado o *kernel Linear* como sendo o mais promissor, devido ao desempenho consistente em relação à métrica *F1-score* (média harmônica entre *recall* e *precision*) para ambos os cenários, evidenciando sua capacidade de proporcionar resultados estáveis, mesmo com variações nas características, garantindo um bom equilíbrio entre desempenho e simplicidade do modelo. Com isso, o *kernel* foi utilizado para a análise em ambos os cenários, considerando os resultados individuais por classe (Classes “0” e “1”).

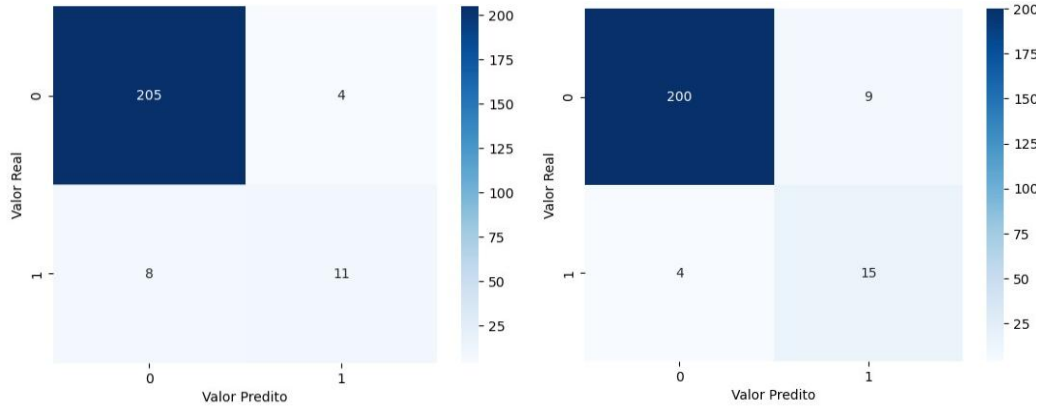
Outra análise investigada foi a visualização da matriz de confusão, conforme apresentado na Figura 13(a), que mostra o modelo alcançando alta precisão na classificação

de instâncias pertencentes à classe majoritária (Classe “0”). O número de verdadeiros negativos (VN) foi de 205, evidenciando que o modelo teve um bom desempenho na identificação de casos negativos. Por outro lado, o número de verdadeiros positivos (VP) foi de 11, um valor menor em comparação à classe negativa, mas condizente com a distribuição desbalanceada das classes. Em relação aos falsos negativos (FN), o modelo classificou incorretamente 8 casos da Classe “0” como pertencentes à Classe “1”. Já os falsos positivos (FP) totalizaram 4, indicando que o modelo classificou 4 instâncias como positivas, quando na verdade pertenciam à classe negativa.

Esses resultados reforçam a tendência do modelo de classificar com maior precisão a classe majoritária, enquanto enfrenta desafios na distinção correta da classe minoritária, refletindo o impacto do desbalanceamento entre as classes (rótulos), mas propondo uma avaliação real do modelo. Nesse contexto, uma avaliação aprofundada se torna essencial para mensurar realisticamente a capacidade preditiva do modelo

Figura 13 - Matriz de confusão do modelo SVM

(a) com todas as características. (b) com as melhores características.



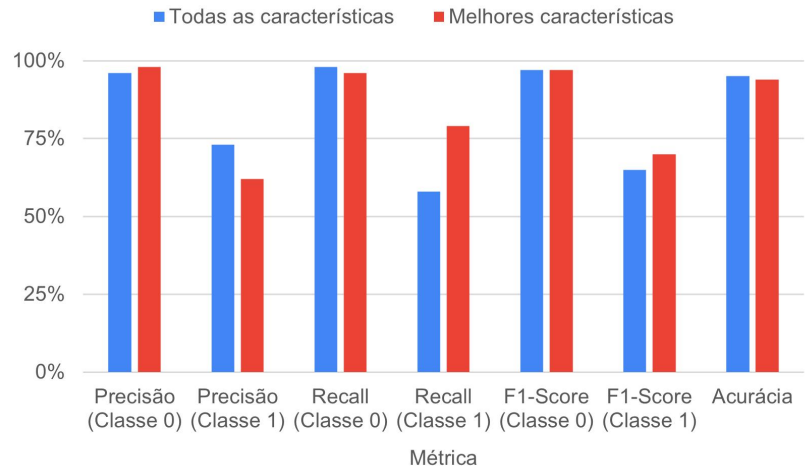
Fonte: Autoria própria.

Em relação ao primeiro cenário, o modelo SVM atingiu 95% de acurácia, o que indica que o modelo apresenta um bom desempenho geral, mantendo uma taxa de acerto consistente em diferentes subconjuntos do *dataset* e demonstrando capacidade de generalização nas previsões. Em relação às demais métricas, o modelo apresentou desempenho consistente na Classe “0”, atingindo 96% de *precision*, 98% de *recall* e 97% de *F1-score*. Por outro lado, para a Classe “1”, os resultados foram inferiores, com 73% de *precision*, 58% de *recall* e 65% de *F1-score*. Esses resultados evidenciam a preferência do modelo em classificar corretamente instâncias pertencentes à classe majoritária (Classe “0”).

Entretanto, mesmo apresentando um baixo desempenho para Classe “1”, o modelo conseguiu distinguir corretamente os exemplos desse rótulo.

Os resultados da matriz de confusão, considerando as melhores características, apresentada na Figura 13(b), revelaram que o modelo preservou um alto índice de acertos na classificação da classe majoritária (Classe “0”). Observou-se uma leve diminuição no número de VN, que passou de 205 para 200, reforçando a capacidade do modelo de identificar corretamente os casos negativos. Em contrapartida, os VP tiveram um pequeno incremento, subindo de 11 para 15, valor que, embora ainda modesto em relação à classe majoritária, está alinhado com a distribuição desbalanceada das classes. Quanto aos FN, houve uma redução, com o modelo classificando incorretamente 4 instâncias da Classe “0” como pertencentes à Classe “1”, contra 8 no cenário anterior. Por outro lado, os FP sofreram um leve aumento, de 4 para 9, mostrando que o modelo cometeu menos erros ao classificar instâncias negativas como positivas. Esses resultados destacam que, embora o modelo continue a priorizar a classe majoritária, houve avanços na redução de erros, especialmente nos FP, o que sugere uma maior precisão geral. No entanto, o desbalanceamento das classes ainda representa um desafio, especialmente para a classificação correta da classe minoritária.

Figura 14 - Resultados do desempenho com o SVM (*kernel Linear*)



Fonte: Autoria própria.

Ainda em relação aos resultados alcançados nos experimentos realizados para o segundo cenário, observou-se que o modelo manteve um desempenho consistente nas métricas de avaliação. No entanto, apresentou alguns valores inferiores em comparação ao primeiro cenário, nas duas classes, conforme ilustra a Figura 14. Nesse contexto, ao

considerar a métrica de acurácia, o modelo atingiu 94%, evidenciando uma leve redução na taxa geral de acertos, em comparação ao cenário anterior. Observando a Figura 14, percebe-se que o modelo manteve um desempenho sólido na classificação de instâncias pertencentes à Classe “0”, com métricas de *precision* (98%), *recall* (96%) e *F1-score* (97%) atingindo valores elevados. Por outro lado, em relação à Classe “1”, o modelo obteve um resultado inferior na métrica de *precision* (62%), em comparação com o primeiro cenário. Contudo, apresentou melhorias significativas nas métricas de *recall* (79%) e *F1-score* (70%). Esses achados indicam que o modelo ainda tende a classificar com maior precisão os exemplos da Classe “0”, refletindo um viés em direção à classe majoritária. No entanto, a possibilidade de acertos para a Classe “1” aumentou, sugerindo um avanço na capacidade do modelo em lidar com o desequilíbrio entre as classes.

4.7.2 KNN

Os experimentos com o modelo KNN foram conduzidos utilizando a configuração padrão da classe *KNeighborsClassifier*, conforme implementado na biblioteca *Scikit-learn*. Nesse sentido, a única alteração realizada foi a variação do hiperparâmetro k, que corresponde ao número de vizinhos e foi testado em um intervalo de números para avaliar seu desempenho nos dois cenários do estudo (isto é, todas as características e as melhores características).

Conforme código mostrado na Figura 15, foi definido um intervalo de números (1 a 10) por meio da função *range*() da linguagem *Python*. Essa função criou uma sequência de valores obedecendo aos argumentos de intervalo especificados (*start, stop*, *step*). Em seguida, foi criada uma lista que armazena os valores referentes à acurácia do modelo e, para cada iteração, o valor atual de k foi utilizado para configurar o modelo. Essa estrutura permitiu testar todos os valores de k de forma organizada.

Figura 15 - Implementação do modelo KNN

```
# testando diferentes valores de k
k_values = range(1, 11)
accuracies = []
for k in k_values:
       knn = KNeighborsClassifier(n_neighbors=k,)
       knn.fit(X_train_balanced, y_train_balanced)
       y_pred = knn.predict(X_test)
       accuracies.append(accuracy_score(y_test, y_pred))
best_k = k_values[accuracies.index(max(accuracies))]
print(f"Melhor valor de k: {best_k}")
```

Fonte: Autoria própria.

Cada valor de k teve uma instância do modelo de KNN e foi usada para fazer previsões no conjunto de teste, por meio da função *predict()*. Nesse sentido, os resultados dessas previsões foram avaliados mediante o uso da função *accuracy_score* (*Scikit-learn*). Os resultados apontaram que o melhor valor de k, considerando todas as características e apenas as melhores características, foi 1 (Figura 15). Isso indica que, entre os valores testados no intervalo de 1 a 10, o modelo obteve a maior acurácia utilizando 1 vizinho mais próximo, em ambos os cenários.

Com base nas descobertas elencadas, um novo treinamento foi realizado considerando esse valor. Para isso, dois novos modelos do KNN (*knn_best*) foram instanciados com as configurações padrão da classe *KNeighborsClassifier*. O hiperparâmetro *n_neighbors* foi ajustado para o valor definido como 1 para o modelo treinado com todas as características e para o modelo treinado apenas com as melhores características.

Antes do treinamento, os dois modelos foram validados por meio de validação cruzada para avaliar sua capacidade de generalização em diferentes subconjuntos dos dados. Seus resultados (acurácia média e o desvio padrão) foram revelados por meio do método *print()*, conforme apresenta a Figura 16. Em seguida, os modelos foram treinados com o conjunto de treinamento.

Figura 16 - Validação cruzada e treinamento do KNN

```
# instanciando o modelo com o valor de k com maior desempenho em accuracy
knn_best = KNeighborsClassifier(n_neighbors=best_k)
cv_scores = cross_val_score(knn_best, X_train_balanced, y_train_balanced, cv
= 10, scoring='accuracy')
print(f"Acurácia média da validação cruzada: {cv_scores.mean() * 100:.2f}%")
print(f"Desvio padrão da acurácia: {cv_scores.std() * 100:.2f}%")
knn_best.fit(X_train_balanced, y_train_balanced)
```

Fonte: Autoria própria.

Após o treinamento, foi realizada a avaliação dos modelos. Para isso, as duas instâncias previamente treinadas (*knn_best*) foram usadas para realizar predições no conjunto de teste. O primeiro modelo considerou todas as características (Figura 17) e valor de k igual a 1, enquanto o segundo modelo levou em consideração apenas as melhores características (*X_test_selected*), com o valor de k também igual a 1, mediante o da função *predict()* da classe *KNeighborsClassifier*.

Figura 17 - Avaliação dos modelos KNN

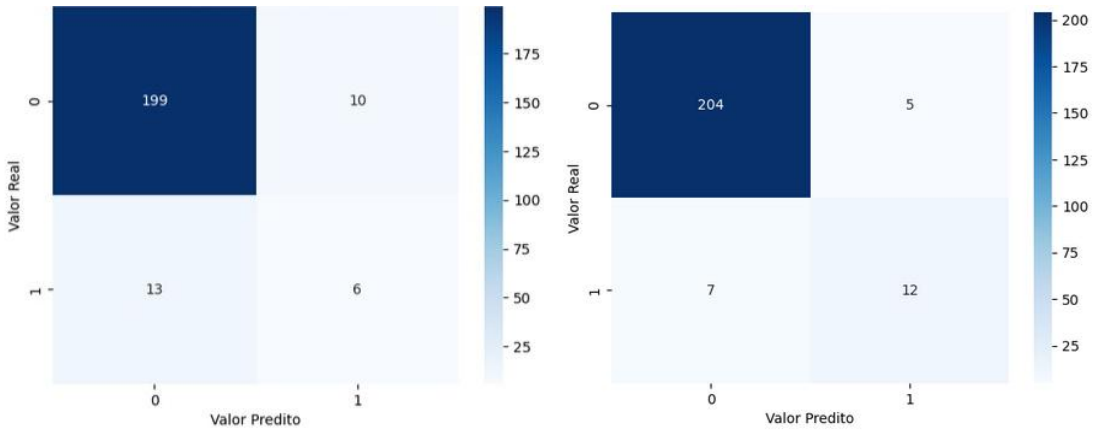
```
y_pred = knn_best.predict(X_test)
matriz_conf = confusion_matrix(y_test, y_pred)
sns.heatmap(matriz_conf, annot=True, fmt='d', cmap='Blues',
            xticklabels=['0', '1'],
            yticklabels=['0', '1'])
plt.title('Matriz de Confusão')
plt.xlabel('Valor Predito')
plt.ylabel('Valor Real');
print('Relatório do modelo\n\n', classification_report(y_test,
y pred))
```

Fonte: Autoria própria.

A avaliação do modelo ocorreu com base nos resultados da matriz de confusão e do relatório de classificação. Os resultados alcançados indicam uma alta precisão na identificação das instâncias da classe majoritária (Classe “0”), conforme mostrado na Figura 18(a). O número de VN foi de 199, demonstrando que o modelo obteve um bom desempenho na classificação de casos negativos da doença. Por sua vez, o número de VP foi de 6, um resultado que reflete a assimetria entre as classes. Com relação aos FN, o modelo apresentou erros na classificação, onde 13 casos foram preditos como pertencentes à Classe “0” mas que na verdade, pertenciam à Classe “1”. Por outro lado, o número de FP foi de 10. Isso reflete que o modelo classificou 10 instâncias como positivas, embora elas pertencessem à classe negativa.

Figura 18 - Matriz de confusão do modelo KNN

(a) com todas as características. (b) com as melhores características.

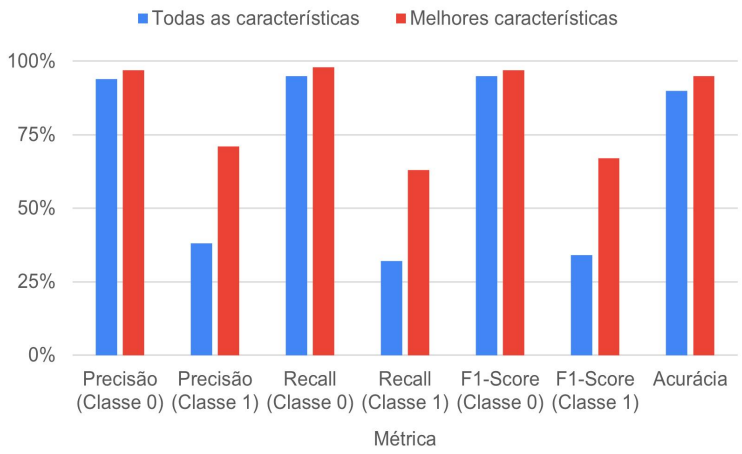


Fonte: Autoria própria.

Considerando todas as características, a acurácia obtida foi 90%. Em relação à Classe "0", o modelo apresentou resultados consistentes, alcançando 94% na métrica de *precision*. Além disso, as métricas de *recall* e *F1-score* também foram robustas, com 95% para ambas. No entanto, para a Classe "1", o modelo apresentou dificuldades para classificar as instâncias corretamente, apresentando *precision* de 38%, *recall* de 32%, e um *F1-score* de 34%. Esse desempenho evidencia uma dificuldade significativa do modelo em classificar corretamente a classe minoritária.

Ao considerar as melhores características, o modelo manteve um elevado desempenho na identificação correta das instâncias da classe majoritária (Classe "0"), conforme visto na Figura 18(b). O número de VN aumentou em relação ao cenário anterior (204), refletindo uma melhoria na identificação de casos negativos da doença. Por outro lado, o número de VP aumentou, de 6 para 12, indicando que o modelo sofreu um leve incremento de eficácia, a respeito da classificação correta das instâncias pertencentes à classe positiva. Em contrapartida, o número de FN diminuiu, com o modelo classificando incorretamente 7 instâncias como pertencentes à classe negativa, quando na verdade pertenciam à classe positiva, em contraste ao cenário anterior, com 13 casos. De modo semelhante, o número de FP também caiu de forma significativa, em comparação ao cenário anterior, de 10 para 5, refletindo que o modelo errou menos na identificação de casos negativos. O modelo obteve resultados superiores relacionados à acurácia, do cenário com as melhores características, conforme apresenta a Figura 19, cujo valor da acurácia chegou a 95%.

Figura 19 - Resultados do desempenho com o KNN



Fonte: Autoria própria.

Os valores alcançados indicam que o modelo manteve um comportamento consistente, com um leve incremento na taxa de acerto geral nas previsões. A respeito da Classe “0”, o modelo manteve o mesmo resultado na métrica *precision* (97%), enquanto obteve uma melhoria em relação ao *recall* (98%) e *F1-score* (97%) em comparação ao cenário anterior. De modo semelhante, para à Classe “1”, o modelo alcançou uma melhora significativa em contraste com o Cenário 1. Com isso, a técnica atingiu 71% em *precision*, 63% em *recall* e 67% em *F1-score*. Esses resultados refletem a importância da estratégia de seleção de características para obtenção de modelos mais eficazes e consistentes para tarefas que exigem classificação.

4.7.3 DT

Os testes com o modelo de DT foram conduzidos empregando as configurações padrão da classe *DecisionTreeClassifier*, conforme implementado na biblioteca *Scikit-learn*. Essa abordagem permitiu a avaliação do desempenho do algoritmo sem ajustes personalizados, garantindo a reprodutibilidade e a comparação justa com outras técnicas implementadas neste estudo. A implementação do modelo pode ser vista no Github[4] do trabalho.

Para obter resultados relacionados ao desempenho do modelo, foi realizada uma validação cruzada com 10 *folds,* que permitiu estimar a acurácia média e o desvio padrão do modelo, fornecendo uma visão geral da consistência e precisão. Em seguida, o modelo foi treinado com todas as características e melhores características.

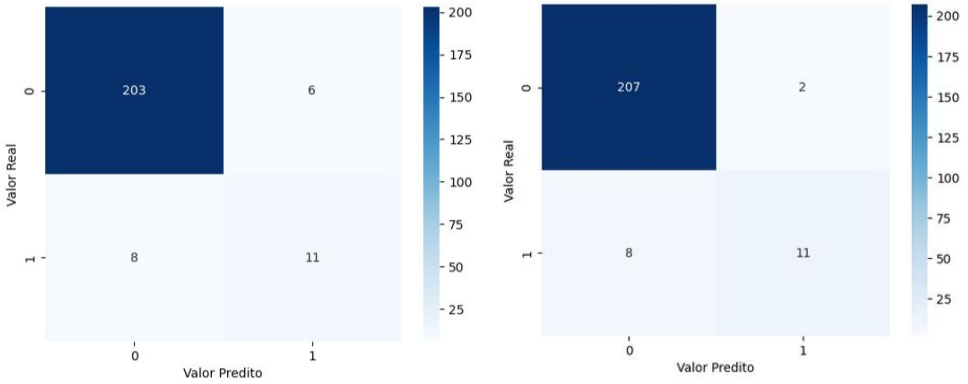
Após o treinamento, o modelo foi utilizado para fazer predições no conjunto de teste com todas as características e melhores características, permitindo uma avaliação dos resultados por meio da matriz de confusão e do relatório de classificação. Nesse sentido, conforme ilustrado na Figura 20(a), os resultados da matriz de confusão do modelo treinado com todas as características indicam um desempenho significativo na identificação das instâncias da classe majoritária (Classe “0”). O número de VN foi de 203, demonstrando que o modelo classificou corretamente a maioria dos casos negativos. Por outro lado, o número de VP foi de 11, refletindo a dificuldade do modelo em identificar corretamente os casos da classe minoritária (Classe “1”). Em relação aos FN, o modelo apresentou 8 erros, onde instâncias da Classe “1” foram erroneamente classificadas como pertencentes à Classe “0”. Já os FP totalizaram 6, indicando que o modelo classificou incorretamente algumas instâncias da Classe “0” como pertencentes à Classe “1”. Esses resultados evidenciam a assimetria entre as

[4]https://github.com/WesleydosSantos/machine-learning-tecnicas/blob/main/Modelos/decision_tree.ipynb.

classes e destacam os desafios enfrentados pelo modelo na classificação de casos da classe minoritária.

Figura 20 - Matriz de confusão do modelo DT

(a) com todas as características. (b) com as melhores características.



Fonte: Autoria própria.

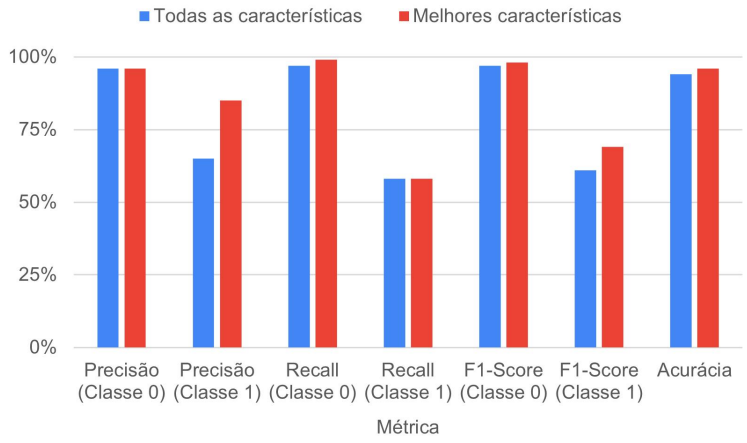
Os resultados da matriz de confusão do modelo treinado com as melhores características, apresentados na Figura 20(b), mostram que o número de VN aumentou para 207, indicando uma melhora na classificação correta dos casos negativos (Classe “0”). Por outro lado, o número de VP manteve-se similar ao modelo anterior, com 11 predições corretas, demonstrando que a seleção de características não impactou significativamente a identificação da classe minoritária (Classe “1”). De modo similar, com relação à FN, o modelo apresentou o mesmo desempenho, permanecendo com 8 instâncias classificadas incorretamente. Por sua vez, o número de FP diminuiu, atingindo 2, contra 6 do cenário anterior, sugerindo um pequeno incremento na eficácia do modelo ao classificar instâncias da Classe “0”.

Ainda em relação ao cenário em que o modelo foi treinado com todas as características, a acurácia atingiu 94%, indicando que o modelo manteve um desempenho consistente e generalizável em diferentes subconjuntos do *dataset*. O modelo apresentou um desempenho robusto na classificação da Classe “0”, com 96% de *precision*, 97% de *recall* e *F1-score*, respectivamente, demonstrando eficácia na identificação correta das instâncias da classe majoritária. Por outro lado, para a Classe “1”, os resultados foram inferiores, com 65% de *precision*, 58% de *recall* e 61% de *F1-score*. Esses valores evidenciam a dificuldade do modelo em classificar corretamente as instâncias da classe minoritária, refletindo a assimetria entre as classes. Apesar de suas limitações na detecção da classe minoritária, o modelo demonstrou capacidade de identificar corretamente parte das instâncias da Classe “1”, o que

significa que, mesmo com eficiência reduzida, ele não negligencia totalmente a classe de interesse.

Observou-se uma melhoria discreta no desempenho do modelo com as melhores características, em comparação ao cenário anterior. Desse modo, o modelo atingiu 96% de acurácia, mantendo um resultado sólido. É possível observar um desempenho ainda mais robusto na classificação da Classe “0”, com 96% de *precision*, 99% de *recall* e 98% de *F1-score*, demonstrando uma melhora na identificação correta das instâncias da classe majoritária em comparação ao modelo que utilizou todas as características (Figura 21). Para a Classe “1”, os resultados também melhoraram, com 85% de *precision*, 58% de *recall* e 69% de *F1-score*. Embora o *recall* tenha permanecido inalterado, o aumento na *precision* e no *F1-score* é um indicativo crucial da otimização do modelo com as melhores características. Tal melhoria na *precision* demonstra que as predições para a classe minoritária se tornaram mais confiáveis, resultando em uma redução desejável de FP. Consequentemente, a elevação do *F1-score* valida a maior confiabilidade das predições para a classe de interesse, aprimorando a capacidade do modelo em auxiliar na detecção de transtornos depressivos.

Figura 21 - Resultados do desempenho com o DT



Fonte: Autoria própria.

4.7.4 RF

Para o modelo de RF, os experimentos foram executados com base nas configurações padrão da classe *RandomForestClassifier*, disponível na biblioteca *Scikit-learn*. Essa escolha permitiu avaliar o desempenho do algoritmo em sua forma original, sem modificações ou otimizações adicionais, assegurando a consistência dos resultados e facilitando a comparação com os outros modelos utilizados no estudo. O código completo da implementação está

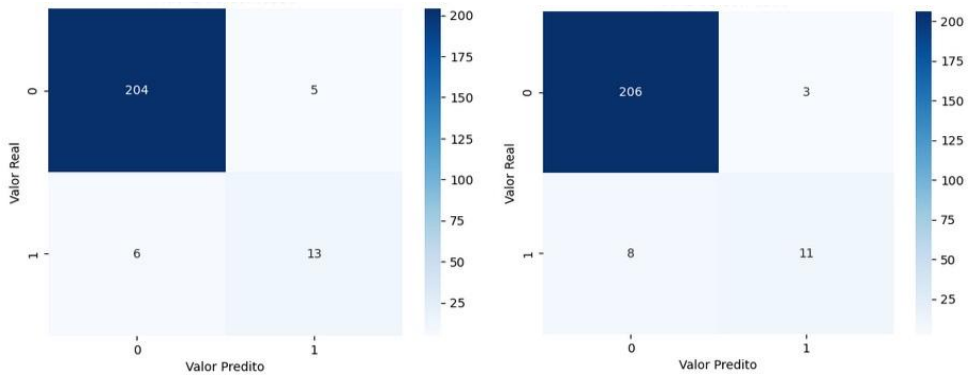
acessível no repositório do *Github*[5].

De modo semelhante às técnicas anteriores, realizou-se uma validação cruzada, possibilitando a obtenção da taxa de acerto geral e o desvio padrão do modelo, fornecendo um panorama quanto à robustez do modelo nos diferentes subconjuntos da base de dados. Na sequência, efetuou-se o treinamento do modelo, considerando todas as características e as melhores características. Nesse sentido, logo após o treinamento, o modelo foi usado para realizar estimativas no conjunto de testes, tornando possível a análise dos resultados com base na matriz de confusão e no relatório de classificação do modelo.

A análise da matriz de confusão do modelo treinado com todas as características revela um desempenho consistente na classificação das instâncias da classe majoritária (Classe “0”), conforme ilustrado na Figura 22(a). Nesse contexto, o número de VN foi de 204, indicando uma precisão na identificação correta de casos negativos. Por outro lado, o número de VP foi de 13, refletindo um desempenho moderado na classificação de amostras da classe minoritária (Classe “1”). Quanto aos FN, o modelo classificou erroneamente 6 exemplos como pertencentes à classe positiva (Classe “1”), quando, na verdade, eram da classe negativa (Classe “0”). De modo semelhante, o número de FP foi de 5, o que significa que o modelo classificou incorretamente 5 casos negativos como positivos, dificuldade do modelo em lidar com a classe minoritária, apesar de seu bom desempenho na identificação da classe majoritária.

Figura 22 - Matriz de confusão do modelo RF

(a) com todas as características. (b) com as melhores características.



Fonte: Autoria própria.

Conforme ilustrado na Figura 22(b), ao considerar as melhores características, o

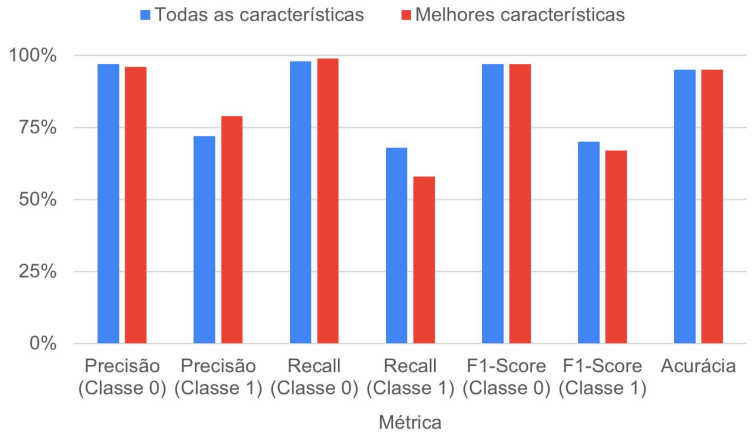
[5] https://github.com/WesleydosSantos/machine-learning-tecnicas/blob/main/Modelos/random_forest.ipynb.

modelo manteve um elevado desempenho na identificação correta das instâncias da classe majoritária (Classe “0”). O número de VN foi de 206, evidenciando uma melhoria na identificação de casos negativos em relação ao cenário anterior. Por outro lado, o número de VP diminuiu, de 13 para 11, indicando uma leve redução na capacidade do modelo de classificar corretamente as instâncias da classe positiva (Classe “1”). Além disso, o número de FN aumentou para 8, em contraste com o cenário anterior, refletindo que o modelo cometeu mais erros ao classificar instâncias da classe positiva como pertencentes à classe negativa. Por fim, o número de FP caiu para 3, em relação ao cenário anterior, mostrando que o modelo errou menos na identificação de casos negativos.

Considerando todas as características, o modelo obteve 95% de acurácia, indicando um alto desempenho relacionado à taxa de acerto geral. Além disso, em relação à Classe “0”, o modelo apresentou resultados sólidos, com *precision* de 97%, indicando eficácia na identificação de casos realmente negativos. Além disso, obteve um *recall* de 98%, sinalizando uma alta capacidade de distinguir instâncias da classe negativa (Classe “0”) e um *F1-score* de 97%, evidenciando equilíbrio alto entre *precision* e *recall*. Por outro lado, para a Classe “1”, o modelo apresentou resultados mais modestos. A *precision* foi de 72%, indicando uma precisão razoável na identificação de casos positivos. O *recall* foi de 68%, mostrando uma capacidade limitada de identificação de instâncias da classe minoritária. O *F1-score* ficou em 70%, sugerindo que, embora o modelo tenha um desempenho aceitável, ainda enfrenta desafios na classificação correta da classe minoritária. Esses resultados reforçam a assimetria entre as classes, com o modelo apresentando um desempenho superior na classificação da classe majoritária (Classe “0”) em comparação com a classe minoritária (Classe “1”).

Ao utilizar as melhores características, o modelo manteve um desempenho consistente, alcançando 95% de acurácia, valor semelhante ao observado no cenário anterior, conforme ilustrado na Figura 23. Em relação à Classe “0”, observou-se um desempenho levemente inferior na métrica *precision*, que caiu para 96% em relação ao cenário anterior. Por outro lado, houve um pequeno aumento no *recall*, que subiu de 98% para 99%, indicando uma maior eficácia na identificação de casos da classe majoritária. Além disso, o modelo manteve um *F1-score* de 97%, reflexo da alta média harmônica entre *precision* e *recall*. No que diz respeito à Classe “1”, o modelo apresentou um aumento na *precision*, que passou de 72% para 79% em relação ao cenário anterior. No entanto, as métricas de *recall* e *F1-score* tiveram desempenhos inferiores, atingindo 58% e 67%, respectivamente. Isso sugere uma maior dificuldade em distinguir instâncias da classe positiva (Classe “1”).

Figura 23 - Resultados do desempenho com o RF



Fonte: Autoria própria.

4.7.5 NB

Para o modelo NB, a avaliação foi conduzida utilizando a implementação padrão da classe *GaussianNB* da biblioteca *Scikit-learn*. Essa abordagem permitiu observar o funcionamento do algoritmo de forma espontânea, preservando suas propriedades estatísticas essenciais, sem ajustes manuais nos parâmetros. Isso assegurou a confiabilidade dos resultados obtidos e possibilitou uma comparação direta com as outras técnicas analisadas no estudo. O código do modelo desenvolvido pode ser acessado publicamente por meio do repositório *GitHub*[6] associado a este estudo.

Assim como nas demais técnicas, a avaliação do modelo incluiu a aplicação da validação cruzada, permitindo estimar a taxa de acerto geral e analisar sua variação entre os subconjuntos do *dataset*. Em seguida, o treinamento foi realizado considerando tanto o conjunto completo de atributos quanto a versão reduzida, composta pelas melhores características. Após isso, o modelo foi submetido à etapa de predição e viabilizada a avaliação dos resultados, por meio da matriz de confusão e do relatório de classificação, possibilitando uma análise detalhada do desempenho do modelo.

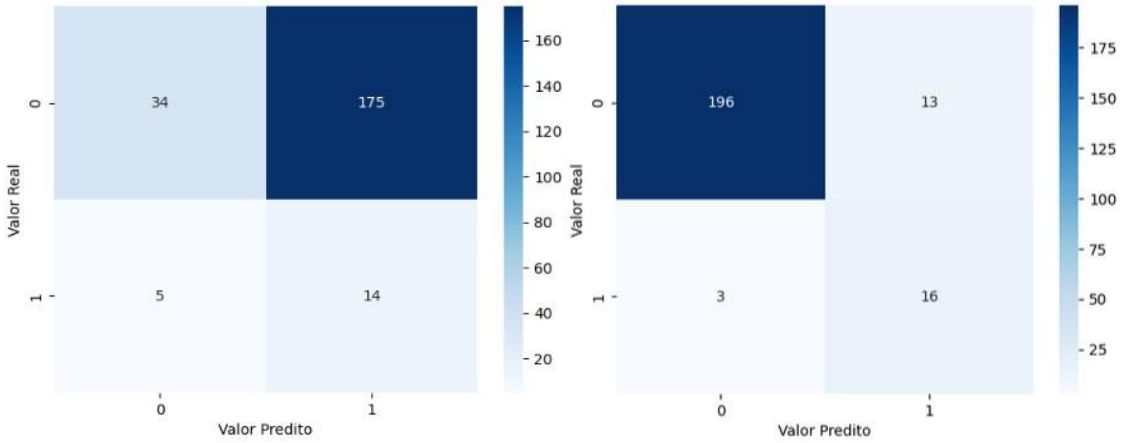
Os resultados do modelo treinado com todas as características indicam uma queda significativa na identificação correta das instâncias pertencentes à classe negativa (Classe “0”), conforme ilustrado na Figura 24(a). O número de VN foi de 34, um desempenho inferior em comparação aos demais modelos avaliados neste estudo. Por outro lado, o número de VP foi de 14, sugerindo um desempenho equilibrado na classificação entre instâncias das classes

[6]https://github.com/WesleydosSantos/machine-learning-tecnicas/blob/main/Modelos/naive_bayes.ipynb.

positiva e negativa. No entanto, o modelo apresentou 5 FN, ou seja, classificou erroneamente instâncias da Classe “1” como pertencentes à Classe “0”. Além disso, o número de FP foi consideravelmente alto, totalizando 175, o que indica uma tendência do modelo em classificar equivocadamente muitas instâncias da Classe “0” como pertencentes à Classe “1”. Esses resultados evidenciam uma limitação do modelo na diferenciação entre as classes, especialmente na identificação da Classe “0”. Esse comportamento pode estar associado à distribuição dos dados e à suposição de independência entre as variáveis, característica do algoritmo.

Figura 24 - Matriz de confusão do modelo NB

(a) com todas as características. (b) com as melhores características.



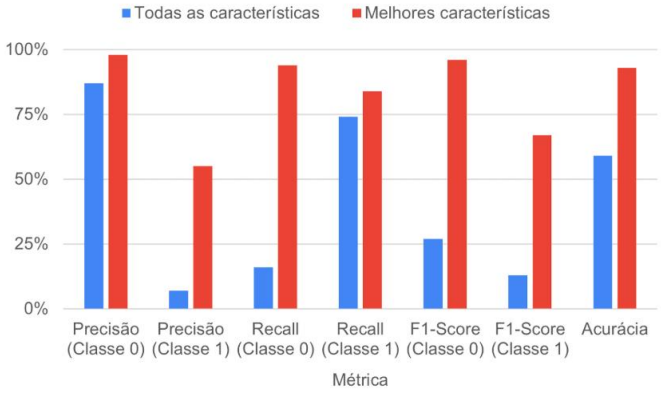
Fonte: Autoria própria.

Os resultados do modelo treinado com as melhores características indicam uma melhora na classificação das instâncias em comparação ao cenário anterior, conforme ilustrado na Figura 24(b). O número de VN aumentou para 196, um valor substancialmente superior aos 34 observados no modelo treinado com todas as características, evidenciando uma capacidade aprimorada do algoritmo em identificar corretamente casos pertencentes à classe negativa (Classe “0”). Quanto ao número de VP, observou-se um leve incremento para 16, sugerindo uma maior eficiência na classificação correta de instâncias pertencentes à Classe “1”. Além disso, o número de FN caiu para 3, o que indica uma maior eficácia na identificação de casos positivos (Classe “1”). Por fim, verificou-se uma queda significativa no número de FP em relação ao cenário anterior, com o modelo classificando incorretamente 13 instâncias como pertencentes à classe positiva, quando, na verdade, pertenciam à classe negativa.

Ao considerar todas as características, o modelo atingiu 59% de acurácia, sinalizando um desempenho geral baixo. Na Classe “0”, o modelo atingiu 87% de *precision*, indicando que a maioria das previsões negativas foi correta. No entanto, o *recall* foi de apenas 16%, evidenciando uma baixa capacidade na distinção entre as classes. Como resultado, o *F1-score* foi de 27%, devido ao desequilíbrio entre as métricas *precision* e *recall*. Em relação à Classe “1”, o modelo obteve 7% de *precision*, refletindo um desempenho limitado na classificação correta das instâncias dessa classe. Por outro lado, o *recall* foi de 74%, indicando uma boa capacidade de identificar a maioria dos casos positivos (Classe “1”). Por sua vez, em relação à métrica de *F1-score*, o algoritmo obteve 13%, em razão da assimetria entre os resultados das métricas *precision* e *recall*.

O desempenho do modelo treinado com as melhores características apresentou melhorias significativas em relação ao modelo que utilizou todas as características, conforme ilustrado na Figura 25. Nesse contexto, a acurácia atingiu 93%, evidenciando uma maior capacidade de generalização. Em relação à Classe “0”, a métrica *precision* manteve sua robustez, com 98% dos casos classificados corretamente como pertencentes à categoria. Além disso, o *recall* obteve 94%, evidenciando um aumento na capacidade do modelo em distinguir casos dessa classe. Consequentemente, o *F1-score* chegou a 96%, impulsionado pelo aumento das métricas *precision* e *recall*. Quanto à Classe “1”, a *precision* atingiu 55%, reforçando um desempenho superior em comparação ao modelo treinado com todas as características. Por outro lado, o *recall* manteve um resultado robusto (84%), evidenciando uma melhora na distinção de instâncias pertencentes à essa classe. Por fim, o *F1-score* registrou 67%, refletindo os avanços nas métricas *precision* e *recall.*

Figura 25 - Resultados do desempenho com o NB



Fonte: Autoria própria.

4.8 Discussão dos Resultados

Os experimentos deste trabalho demonstram a eficácia dos modelos de ML na predição de transtornos depressivos em estudantes universitários, utilizando o conjunto de dados de treinamento. Os resultados alcançados destacam o impacto positivo da seleção de características no desempenho dos classificadores, permitindo comparar o uso de todas as características e apenas as melhores características (usando o *SelectKBest*). Esse teste comparativo evidenciou melhorias significativas em métricas como *precision*, *recall*, *F1-score* e acurácia dos modelos em geral, usando as melhores características, o que reforça a importância dessa análise desde a etapa de pré-processamento dos dados. Os resultados agrupados são mostrados na Tabela 3, com o intuito de viabilizar a análise comparativa dos experimentos realizados.

Tabela 3 - Comparação das métricas alcançadas

Modelo	**Todas as características**			**Melhores características**		
	Precision (%)	*Recall* (%)	*F1-score* (%)	*Precision* (%)	*Recall* (%)	*F1-score* (%)
NB	47	45	20	77	89	81
KNN	66	63	64	84	80	82
SVM	85	78	81	80	87	83
RF	85	83	84	87	78	82
DT	80	78	79	90	78	83

Fonte: Autoria própria.

O modelo NB apresentou uma melhora expressiva em todas as métricas quando utilizou as melhores características: a *precision* subiu de 47% para 77%, o *recall* de 45% para 89%, e o *F1-score* de 20% para 81%. Já o KNN mostrou um equilíbrio entre as métricas, com *precision* aumentando de 66% para 84%, e o *recall* de 63% para 80%, resultando em um *F1-score* superior (82% contra 64%). O SVM, por sua vez, manteve um desempenho alto em ambos os cenários, mas com uma mudança interessante: a *precision* diminuiu de 85% para 80%, enquanto o *recall* aumentou de 78% para 87%, refletindo um *F1-score* ligeiramente melhor (83% contra 81%). Os modelos RF e DT apresentaram comportamentos distintos. O RF manteve alta *precision* (85% para 87%), mas registrou uma redução no *recall* (83% para 78%). Já o DT destacou-se ao alcançar a maior *precision* entre todos os modelos (90%), com

as melhores características, embora com um *recall* estável em 78% em ambos os cenários. Ambos os modelos mantiveram um *F1-score* elevado: 82% (RF) e 83% (DT) com as melhores características, contra 84% e 79% no cenário com todas as características.

Acerca da análise da acurácia, a Tabela 4 mostra uma visualização comparativa entre os resultados alcançados com todas as características, as melhores características e os trabalhos relacionados a esta pesquisa. Percebe-se que o modelo NB apresentou um salto expressivo de desempenho após a seleção das melhores características, com a acurácia subindo de 59% para 93%, superando os resultados reportados por Cruz, Flores e Quispe (2023), que alcançaram 78% (Tabela 4).

Tabela 4 - Comparação entre Acurácias

Modelo	**Acurácia (Todas as características)**	**Acurácia (Melhores as características)**	**Estudos anteriores (Acurácia)**
NB	59%	93%	Cruz, Flores e Quispe (2023): 78%
KNN	90%	95%	Sandhu *et al.* (2022): 76%
SVM	95%	94%	Gil, Kim e Min (2022): 80% Upadhyay, Mohapatra e Singh (2024): 89%
RF	95%	95%	Gil, Kim e Min (2022): 86%
DT	95%	96%	Xiaocheng (2023): 91%

Fonte: Autoria própria.

Conforme mostrado na Tabela 4, os modelos KNN e SVM mantiveram desempenhos elevados, com acurácias acima de 90% em ambos os cenários, demonstrando robustez. O KNN, em particular, atingiu 95% com as melhores características, superando os 76,6%

relatados por Sandhu *et al.* (2022). Já o SVM, embora tenha apresentado uma leve queda na acurácia após a seleção (de 95% para 94%), manteve-se competitivo, alinhando-se com estudos como os de Gil, Kim e Min (2022), que reportaram 80% e Upadhyay, Mohapatra e Singh (2024), que registraram 89%.

Os modelos baseados em árvores de decisão (DT e RF) destacaram-se pela consistência, com acurácias próximas ou superiores a 95%. O DT, em especial, alcançou 96% com as melhores características, desempenho superior ao de Xiaocheng (2023) que registrou aproximadamente 91%. O RF manteve-se estável, com 95% em ambos os cenários, apresentando desempenho comparável ao de Gil, Kim e Min (2022), que obtiveram 86%.

Conforme os resultados ilustrados nas Tabelas 3 e 4, o DT foi o modelo que apresentou o melhor desempenho geral na predição de transtornos depressivos em estudantes universitários. Além de alcançar a maior acurácia entre os modelos testados, demonstrou consistência e equilíbrio nas demais métricas de avaliação, mostrando-se uma alternativa promissora para aplicações práticas nesse contexto.

Apesar dos resultados interessantes, é importante reconhecer que uma limitação desta pesquisa se relaciona com o tamanho reduzido do conjunto de dados utilizado. Para superá-la, propõe-se, em estudos futuros, a ampliação da base de dados.

5 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

A saúde mental continua sendo um dos maiores desafios globais, com impacto significativo sobre a qualidade de vida e bem-estar dos indivíduos. No ambiente universitário, esse problema torna-se ainda mais preocupante, uma vez que estudantes são frequentemente expostos a fatores estressantes, como pressão acadêmica e incertezas profissionais.

Diante desse cenário e da complexidade dos desafios atuais, técnicas de ML têm se mostrado promissoras ao possibilitar a análise automatizada de grandes volumes de dados, favorecendo a identificação de padrões relevantes que podem ser utilizados tanto na detecção de doenças quanto no apoio à atuação de equipes médicas. Estudos têm investigado o uso de técnicas de ML no contexto da saúde mental, com ênfase na identificação de padrões associados à transtornos depressivos. Essas pesquisas evidenciam e reforçam o potencial desses algoritmos para auxiliar em diagnósticos precoces e apoiar estratégias de intervenção.

Alinhado aos trabalhos elencados anteriormente, a presente pesquisa realizou um comparativo entre os principais modelos supervisionados: SVM, RF, KNN, NB e DT. O objetivo foi avaliar o desempenho desses algoritmos na detecção de transtornos depressivos em estudantes universitários. Os resultados demonstraram um desempenho promissor dos modelos, notadamente influenciado pela aplicação do *SelectKBest* como estratégia de seleção de atributos. Quando empregando as características selecionadas (*phq_score, depressiveness, depression_treatment, anxiety_diagnosis e anxiety_treatment*), o modelo DT alcançou uma acurácia de 96% na predição da variável-alvo, evidenciando sua alta confiabilidade e o seu grande potencial para auxiliar na detecção da transtornos depressivos nesse público.

Acerca dos trabalhos futuros, destaca-se a realização de novos experimentos considerando outros conjuntos de dados, preferencialmente, contextualizados à realidade brasileira, e, mais especificamente, de estudantes universitários da região semiárida do Nordeste do Brasil. Adicionalmente, propõe-se o desenvolvimento de uma *Application Programming Interface* (API) para integrar o modelo de DT, treinado com as melhores características, dado que este modelo teve o melhor desempenho. A criação de uma API possibilitará a implementação de futuras aplicações na área de saúde, destinadas a apoiar a detecção precoce de transtornos depressivos entre os estudantes universitários.

REFERÊNCIAS

ALMEIDA, Adriano; CARVALHO, Felipe; MENINO, Felipe. Introdução ao Machine Learning. Brasil: **Grupo Dataat**, 2020. *E-book.* Disponível em: https://dataat.github.io/introducao-ao-machine-learning/. Acesso em: 29 set. 2024.

ALMEIDA, Ítalo D.'Artagnan. **Metodologia do trabalho científico**. Recife: Editora Universitária da UFPE, 2021. *E-book*. Disponível em: https://repositorio.ufpe.br/handle/123456789/49435. Acesso em: 29 set. 2024.

ALMEIDA, Renato França de. Pesquisa bibliográfica sobre aprendizado de máquina aplicado à condução veicular autônoma: Uma revisão. *In*: SEVEN INTERNATIONAL MULTIDISCIPLINARY CONGRESS, 4., 2024, [*s. l.*]. **Anais** [...]. [*S. l.*]: Seven Publicações, 2024. Disponível em: https://doi.org/10.56238/sevenVImulti2024-075. Acesso em: 29 set. 2024.

AMANTE, Lúcia; MORGADO, Lina. Metodologia de concepção e desenvolvimento de aplicações educativas: o caso de materiais hipermedia. **Discursos**, Lisboa, v.3, p. 27 - 43, jun. 2001. Disponível em: http://hdl.handle.net/10400.2/4348. Acesso em: 25 set. 2024.

APA, AMERICAN PSYCHIATRIC ASSOCIATION. **Manual diagnóstico e estatístico de transtornos mentais**: DSM-5-TR. Tradução de Daniel Vieira, Marcos Viola Cardoso e Sandra Maria Mallmann da Rosa. 5.ed. Porto Alegre: Artmed, 2023.

ARAÚJO, Beatriz Evangelista de. Revisão Narrativa Acerca do Conceito de Ansiedade em Psicologia. **Revista Científica Gênero na Amazônia**, [Pará], n. 2, p. 59-70, jul./dez. 2022. Disponível em: https://periodicos.ufpa.br/index.php/generoamazonia/article/view/13484. Acesso em: 29 set. 2024.

BEZERRA, Jackson Henrique da Silva; ALMEIDA, Fabrício Moraes de. Desenvolvimento de modelos preditivos com machine learning: análise de dados para saúde de gestantes e puérperas. **InterSciencePlace**, [*s. l.*], v. 19, n.16, p. 304-335, jan./dez. 2024. Disponível em: https://www.interscienceplace.org/index.php/isp/article/view/763. Acesso em: 29 set. 2024.

BRACHER-SMITH, Matthew; CRAWFORD, Karen; ESCOTT-PRICE, Valentina. Machine learning for genetic prediction of psychiatric disorders: a systematic review. **Molecular Psychiatry**, [*s. l.*], v.26, n.1, p. 70-79, Jan. 2021. Disponível em: https://www.nature.com/articles/s41380-020-0825-2. Acesso em: 29 set. 2024.

BRASIL. Ministério da Saúde. **Saúde mental**. [*S. l.*]: Ministério da Saúde, [2019?a]. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/s/saude-mental. Acesso em: 29 set. 2024.

BRASIL. Ministério da Saúde. **Depressão**. [*S. l.*]: Ministério da Saúde, [2019?b]. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/depressao. Acesso em: 12 set. 2024.

BRASIL. Ministério da Saúde. **Setembro amarelo:** precisamos falar sobre a saúde mental: A campanha deste mês é marcada pela conscientização sobre a prevenção ao suicídio. [*S. l.*]: Ministério da Saúde, 23 set. 2022. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-brasil/eu-quero-me-exercitar/noticias/2022/setembro-amarelo-precisamos-

falar-sobre-a-saude-mental. Acesso em: 29 set. 2024.

BHATNAGAR, Shaurya; AGARWAL, Jyoti; SHARMA, Ojasvi Rajeev. Detection and classification of anxiety in university students through the application of machine learning. **Procedia Computer Science**, [*s. l.*], v. 218, p. 1542-1550, Jan. 2023. Disponível em: https://doi.org/10.1016/j.procs.2023.01.132. Acesso em: 29 set. 2024.

CHEGG. **Global Student Survey 2023**. Santa Clara: [*s. n*.], 2023. *E-book*. Disponível em: https://www.chegg.org/global-student-survey-2023. Acesso em: 29 set. 2024.

CRUZ, Fred Torres; FLORES, Evelyn Eliana Coaquira; QUISPE, Sebastian Jarom Condori. Prediction of depression status in college students using a Naive Bayes classifier based machine learning model. [*S. l. : s. n.*], July. 2023. Disponível em: https://arxiv.org/abs/2307.14371. Acesso em: 25 set. 2024. *In press*.

CAMPOS, João Heli de; STORRODUMOF, Pamela de Souza; CAVALCANTI, Noemi Borgas de Góes. Visagismo, fisiognomia e análise facial fundamentada no cruzamento de ferramentas diagnósticas. **Simmetria Orofacial Harmonization in Science**, [*s. l.*], v. 1, n. 1, p. 08-19, 2019. Disponível em: https://pt.scribd.com/document/683103505/96-Visagismo-Fisiognomia-e-Analise-Facial. Acesso em: 27 set. 2024.

OLIVEIRA, Milena Edite Casé de *et al*. Série temporal do suicídio no Brasil: o que mudou após o Setembro Amarelo? **Revista Eletrônica Acervo Saúde**, [*s. l.*], n. 48, e3191, mai. 2020. Disponível em: https://acervomais.com.br/index.php/saude/article/view/3191. Acesso em: 27 set. 2024.

FACELI, Katti *et al*. **Inteligência artificial:** uma abordagem de aprendizado de máquina. 2. ed. Rio de Janeiro: LTC, 2021.

GÉRON, Aurélien. **Mãos à obra:** aprendizado de máquina com Scikit-Learn & TensorFlow. Rio de Janeiro: Alta Books, 2019.

GIL, Minji; KIM, Suk-Sun; MIN, Eun Jeong. Machine learning models for predicting risk of depression in Korean college students: identifying family and individual factors. **Frontiers in Public Health**, [Lausanne], v.10, p. 1023010. Nov. 2022. Disponível em: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.1023010/full. Acesso em: 27 set. 2024.

HARRISON, Matt. **Machine Learning–Guia de referência rápida**: trabalhando com dados estruturados em Python. São Paulo: Novatec Editora, 2019.

HUYEN, Chip. **Projetando sistemas de Machine Learning:** processo interativo para aplicações prontas para produção. Rio de Janeiro: Editora Alta Books, 2024.

IPARRAGUIRRE-VILLANUEVA, Orlando *et al*. Machine Learning Models to Classify and Predict Depression in College Students. **International Journal of Interactive Mobile Technologies (iJIM)**, [*s. l.*], v. 18, n. 14, p. 148-163, Aug. 2024. Disponível em: https://online-journals.org/index.php/i-jim/article/view/48669. Acesso em: 09 abr. 2025.

JAYATILAKE, Senerath Mudalige Don Alexis Chinthaka; GANEGODA, Gamage Upeksha. Involvement of machine learning tools in healthcare decision making. **Journal of healthcare engineering**, [*s. l.*], v. 2021, n. 1, p. 6679512, Jan. 2021. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1155/2021/6679512. Acesso em: 20 set. 2024.

LEAHY, Robert L. **Livre de ansiedade**. Porto Alegre: Grupo A, 2010.

MAGGI, Bruno. Bem-estar. **Laboreal**, [Porto], v.2, n.1, p.13864, 2006. Disponível em: https://journals.openedition.org/laboreal/13864#quotation. Acesso em: 20 set. 2024.

MARTINS, Fran. **Transtornos de ansiedade podem estar relacionados a fatores genéticos**: Causas ambientais também são capazes de desencadear o problema. [*S. l*]: Ministério da Saúde, 21 set. 2022a. Disponível em: https://www.gov.br/saude/pt-br/assuntos/noticias/2022/setembro/transtornos-de-ansiedade-podem-estar-relacionados-a-fatores-geneticos. Acesso em: 13 set. 2024.

MARTINS, Fran. **Na América Latina, Brasil é o país com maior prevalência de depressão**: Segundo a Organização Pan-Americana da Saúde, essa é a principal causa de incapacidade em todo o mundo. [*S. l*]: Ministério da Saúde, 22 set. 2022b. Disponível em: https://www.gov.br/saude/pt-br/assuntos/noticias/2022/setembro/na-america-latina-brasil-e-o-pais-com-maior-prevalencia-de-depressao. Acesso em: 28 set. 2024.

MITCHELL, Tom M. **Machine Learning**. 1. st. ed. [*S. l.*]: McGraw-Hill, 1997.

QUEVEDO, João; NARDI, Antonio Egidio; SILVA, Antônio Geraldo da. **Depressão:** teoria e clínica. 2. ed. Porto Alegre: Artmed Editora, 2019.

RIS-ALA, Rafael. **Fundamentos de Aprendizagem por Reforço**. 1. ed. Rafael Ris-Ala, 2023.

SAMPAIO, Tuane Bazanella. **Metodologia da pesquisa**. 1 ed. Santa Maria: UFSM, CTE, UAB, 2022. E-*book.* Disponível em: https://repositorio.ufsm.br/handle/1/26138. Acesso em: 28 set. 2024.

SANDHU, Ejaz *et al.* Compute Depression and Anxiety among Students in Pakistan, using Machine Learning. **Lahore Garrison University Research Journal of Computer Science and Information Technology**, [*s. l.*], v. 6, n. 04, Dec. 2022. Disponível em: https://lgurjcsit.lgu.edu.pk/index.php/lgurjcsit/article/view/402. Acesso em: 25 maio 2025.

SCULL, Andrew. **Madness in Civilization:** A Cultural History of Insanity, from the Bible to Freud, from the Madhouse to Modern Medicine. Princeton: Princeton University Press, 2015. *E-book.* Disponível em: https://press.princeton.edu/books/ebook/9781400865710/madness-in-civilization-0. Acesso em: 02 out. 2024.

SHAKYA, Ashish Kumar; PILLAI, Gopinatha; CHAKRABARTY, Sohom. Reinforcement learning algorithms: A brief survey. **Expert Systems with Applications**, [*s. l.*], v. 231, p. 120495, Nov. 2023.

SOUZA, Isabel C. Weiss de; KOZASA, Elisa H. **Saúde mental:** desafios contemporâneos. 1 ed. Barueri: Editora Manole, 2023.

TRIGUEIRO, Emilia Suitberta de Oliveira; CALDAS, Germana Freire Rocha; SILVA, João Marcos Ferreira de Lima. Saúde mental e sofrimento psíquico em estudantes universitários. **Educação em Perspectiva**, Viçosa, v. 12, p. e021021, dez. 2023. Disponível em: https://periodicos.ufv.br/educacaoemperspectiva/article/view/9675. Acesso em: 28 set. 2024.

UPADHYAY, Devesh Kumar; MOHAPATRA, Subrajeet; SINGH, Niraj Kumar. An early assessment of Persistent Depression Disorder using machine learning algorithm. **Multimedia Tools and Applications**, [*s. l.*], v. 83, n.16, p. 49149-49171, Oct. 2024. Disponível em: https://link.springer.com/article/10.1007/s11042-023-17369-4. Acesso em: 30 abr. 2025.

VIEIRA, Josimar de Aparecido; LEITE, Amanda Regina; KUHN, Adele Stein. Perspectivas da Produção de Pesquisa Aplicada, Inovação e Desenvolvimento Científico e Tecnológico nos Institutos Federais. **Revista Valore**, [*s. l.*], v. 8, p. e-8024, mar. 2023. Disponível em: https://revistavalore.emnuvens.com.br/valore/article/view/1344. Acesso em: Acesso em: 28 set. 2024.

WEBER, César Augusto Trinta. A internacionalização da campanha Setembro Amarelo®. **Debates em Psiquiatria**, Rio de Janeiro, v. 13, p. 1-6, set. 2023. Disponível em: https://revistardp.org.br/revista/article/view/1050. Acesso em: 29 set. 2024.

WORLD HEALTH ORGANIZATION. **Depressive disorder (depression)**. [Genebra]: WHO, 2023. Disponível em: https://www.who.int/news-room/fact-sheets/detail/depression. Acesso em: 29 maio 2025.

WORLD HEALTH ORGANIZATION. **Mental Health**. [Genebra]: WHO, [2020?]. Disponível em: https://www.who.int/health-topics/mental-health#tab=tab_1. Acesso em: 28 set. 2024.

WORLD HEALTH ORGANIZATION. **Live life:** an implementation guide for suicide prevention in countries. Genebra, 2021. *E-book.* Disponível em: https://www.who.int/publications/i/item/9789240026629. Acesso em: 30 abr. 2025.

XIAOCHENG, He. Application of Decision Tree Algorithm in College Students' Mental Health Evaluation. *In*: IEEE INTERNATIONAL CONFERENCE ON INTEGRATED CIRCUITS AND COMMUNICATION SYSTEMS (ICICACS), 1., 2023, Raichur. **Electronic proceedings** [...]. Raichur: IEEE, 2023. p. 10099654. Disponível em: https://ieeexplore.ieee.org/document/10099654. Acesso em: 30 abr. 2025.