Utilização de Machine Learning para diagnose em estudantes com Transtorno do Espectro Autista a partir de bases de dados públicas

Sara R. A. Leal[1], Suzerlly V. L. Pires[1], Vanderlene C. Rocha[1], Franciely A. de Souza[1], Lucas C. Teixeira[1], Joao F. L. de Oliveira[1], Ticia C. F. Cavalcante[2], Diego M. P. F. Silva[1] e Carlo M. R. da Silva[1]

[1]Programa de Pós-Gradução em Engenharia de Computação (PPGEC) – Universidade de Pernambuco (UPE)

[2]Programa de Pós-Graduação em Educação (PPGEdu) – Universidade Federal de Pernambuco (UFPE)

{sral,svlp,vcr,jflo,cmrs}@ecomp.poli.br

Abstract. *Autism Spectrum Disorder (ASD) is a neurological condition that affects neurodevelopment, communication, and social interaction. It is often underreported, leading to educational challenges due to the lack of appropriate interventions. This study aims to develop a tool that assists educators in diagnosing ASD by utilizing Machine Learning algorithms to detect ASD signs across different ages, based on simple data extracted from three public datasets. These datasets were pre-processed and balanced using the SMOTE technique, and five algorithms were applied: Decision Tree, Random Forest, KNN, Naive Bayes, and Deep Learning. Random Forest stood out for its superior performance, with high accuracy and low error incidence. The results suggest that these models can be effective tools for early ASD screening, offering significant support to educators.*

Resumo. *O Transtorno do Espectro Autista (TEA) é um distúrbio neurológico que afeta o neurodesenvolvimento, a comunicação e a interação social, frequentemente subnotificado, o que resulta em dificuldades educacionais devido à falta de intervenções adequadas. Este estudo visa desenvolver uma ferramenta que possa auxiliar educadores na diagnose do TEA, utilizando algoritmos de Machine Learning para rastrear sinais do TEA em diferentes idades, com base em dados simples extraídos de três bases públicas. Essas bases foram pré-processadas e balanceadas usando a técnica SMOTE, e cinco algoritmos foram aplicados: Decision Tree, Random Forest, KNN, Naive Bayes e Deep Learning. O Random Forest destacou-se pelo desempenho superior, com alta acurácia e baixa incidência de erros. Os resultados sugerem que esses modelos podem ser ferramentas eficazes na triagem precoce de TEA, oferecendo suporte significativo para educadores.*

1. Introdução

O Transtorno do Espectro Autista (TEA) é um distúrbio complexo do desenvolvimento cerebral caracterizado por anormalidades em três domínios, a saber: interação social recíproca, comunicação e repertório comportamental de interesses restritos, repetitivo e estereotipado [Araújo et al. 2021]. O TEA compreende uma condição clinicamente heterogênea, resultando em uma faixa de comportamentos que um determinado individuo dentro do espectro acaba apresentando sintomas e intensidades diferentes.

De acordo com os dados do último censo escolar, disponibilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) [INEP 2021], quase 300

DOI: 10.5753/sbie.2024.241708

mil estudantes com TEA estavam matriculados na educação básica em todo o país no ano de 2021, um aumento de 280%, em comparação com os dados de 2017. O resultado desta crescente é uma busca mais incisiva da educação inclusiva nas escolas. O ambiente escolar é desafiador para estudantes com TEA e eles têm um risco muito maior de desenvolver comportamento de recusa escolar, em comparação com estudantes neurotípicos, principalmente devido à sobrecarga sensorial e social [Lowri 2021, Munkhaugen et al. 2017]. Em indivíduos com TEA adultos, por exemplo, a dificuldade de acompanhar trabalhos escolares, atividades em equipe e demais propostas acadêmicas podem resultar em quadro severo de fadiga, se não tiverem apoio necessário [Overland et al. 2022].

Na maioria dos casos, o transtorno pode ser identificado nos primeiros anos de vida, mas a diversidade de sintomas faz com que algumas crianças passem pela primeira infância sem diagnóstico. A detecção precoce permite um acompanhamento adequado na escola e na família, melhorando a qualidade de vida com terapias focadas em necessidades específicas.

Porém, métodos tradicionais, como avaliações comportamentais e entrevistas clínicas, podem ser subjetivos e dependem da *expertise* do avaliador [Gioia et al. 2021]. Em contraste, os modelos de *Machine Learning* (ML) analisam grandes volumes de dados rapidamente, identificando padrões sutis do TEA, acelerando o diagnóstico e reduzindo erros. Além disso, essas técnicas ampliam o acesso ao diagnóstico, especialmente em áreas com poucos especialistas em TEA.

Assim, a aplicação de ML no diagnose do TEA representa um avanço, contribuindo para intervenções precoces e, consequentemente, qualidade de vida das pessoas afetadas. Essas ferramentas permitem um diagnóstico preliminar mais preciso, o que contribui significativamente para o aprimoramento do aprendizado e do desenvolvimento dos estudantes no ambiente escolar, visto que, os próprios profissionais de educação poderão fazer uma triagem mais rápida e eficiente, possibilitando que as intervenções educacionais sejam implementadas de maneira mais eficaz, além de poder adaptar suas abordagens pedagógicas para atender às necessidades específicas de cada aluno com TEA.

Dessa forma, este trabalho tem como objetivo analisar três bases de dados que propõem fazer uma triagem de pessoas em três faixas etárias distintas. A partir desses dados, serão aplicados modelos de machine learning para comparar as três bases de dados públicas, compostas por perguntas e características simples, com o objetivo de subsidiar a detecção de autismo em crianças e adultos. Tais iniciativas são importantes e válidas desde a educação infantil até o ensino superior [Wang et al. 2011, Brasil. Ministério da Saúde. 2014, Gois et al. 2022].

Este artigo está estruturado da seguinte forma: na seção de Conceitos Fundamentais, foi discutido as bases teóricas e estudos relevantes que embasam a utilização de Machine Learning no diagnóstico do TEA. Em seguida, a seção de Método descreve o delineamento da pesquisa, incluindo as bases de dados utilizadas, o pré-processamento dos dados e a configuração dos algoritmos. Na seção de Resultados, foram apresentadas as análises das métricas de desempenho dos modelos aplicados, seguidas pela Discussão, onde foi comparado os resultados obtidos e suas implicações. Por fim, as Considerações Finais sintetizam as principais conclusões do estudo e sugerem direções para futuras pesquisas.

2. Conceitos Fundamentais

A identificação de estudantes com TEA no contexto educacional é uma prática fundamentada na necessidade de fornecer intervenções precoces e adequadas que possam melhorar significativamente o desenvolvimento e a qualidade de vida desses estudantes. Estudos teóricos e empíricos indicam que a detecção precoce de TEA, realizada por meio

de avaliações comportamentais e do uso de ferramentas padronizadas de triagem, permite que educadores e profissionais de saúde desenvolvam Planos Educacionais Individualizados (PEIs) que atendam às necessidades específicas de cada criança [Santos et al. 2021].

A literatura destaca a eficácia de métodos baseados em observações sistemáticas do comportamento, que podem ser definidos como diagnose, uma prática que pode ser realizada por um profissional não clínico, capaz de estabelecer critérios para levantar possíveis suspeitas de um determinado quadro a ser confirmado por um clínico através de um diagnóstico [Brentani et al. 2013]. A diferença entre "diagnóstico" e "diagnose" reside principalmente no uso e na especificidade dos termos. O diagnóstico é amplamente utilizado para se referir ao resultado final da identificação de uma doença ou condição, baseado em uma avaliação clínica. Em contraste, a diagnose é um termo que tende a ser mais associado ao processo de investigação e análise que leva a uma possível suspeita.

O trabalho de [Frota et al. 2019], utilizou o algoritmo de Classification Tree (Árvore de Decisão de Classificação) para construção de um modelo capaz de simplificar um conjunto complexo de decisões e produzir uma estratégia como complemento ao diagnóstico e possível tratamento do TEA, a partir da mesma base de dados pública utilizada neste trabalho. Com uma abordagem parecida com a dos autores citados, [Shinde and Patil 2023] , utilizaram DT e *Random Forest* (RF) como algoritmos de sua pesquisa, para validar o modelo de multi-classificadores para aumentar a precisão na predição de TEA.

Outros estudos, embora não realizem rastreio de TEA, analisaram padrões comportamentais e cognitivos de estudantes diagnosticados para sugerir estratégias de melhoria da qualidade de vida. [Gyori et al. 2018] investigou a eficácia de sistemas automatizados de reconhecimento facial para interpretar expressões emocionais em estudantes autistas. Já [Deng et al. 2021] utilizou um sistema de monitoramento baseado em ML para detectar atenção e estresse em crianças com TEA, considerando as limitações dessas crianças em manter a atenção e gerir o estresse.

Dados biométricos e comportamentais foram analisados a fim de identificar padrões associados aos níveis de atenção e estresse dos participantes. Os modelos SVM (*Support Vector Machine*), RF (*Random Forest*), KNN (*K-Nearest Neighbors*), GBDT (*Gradient Boosting Decision Tree*) e ANN (*Artificial Neural Network*) foram utilizados e comparados na análise dos dados, onde o GBDT obteve a maior precisão (86,67%), para detecção de atenção, e o modelo RF obteve 99% de precisão na detecção de estresse.

Todavia, [Yang et al. 2019] também empregaram o método de validação cruzada (*cross validation*) para a classificação da conectividade funcional entre indivíduos com TEA e indivíduos típicos, demonstrando avanços significativos. Através da aplicação de técnicas de *Deep Learning* (DL), o estudo utilizou quatro classificadores distintos: regressão logística, Ridge, máquina de vetor de suporte linear com penalização e máquina de vetor de suporte com kernel gaussiano. Sob essa abordagem, destaca-se a obtenção de uma acurácia superior a 70%, sendo a técnica Ridge aquela que alcançou o nível de acurácia mais elevado.

Em [Gupta and Hafiz 2022], foram utilizados diferentes métodos de classificação para o diagnóstico de TEA em crianças de 4 a 11 anos. O estudo utilizou uma base de dados pública, separou o conjunto de dados em proporções de 70:30 e propôs testar vários algoritmos como SVM, KNN, DT e análise discriminante linear (Linear Discriminant Analysis - LDA). Uma comparação de várias medidas de desempenho foi feita após a aplicação de cada algoritmo. Observa-se que as árvores de decisão e o SVM apresentam maior acurácia de 99% do que o KNN e o LDA, com 70% e 97%, respectivamente. Similarmente, [Yesilyurt and Diagnosing 2021] aplicaram os oito modelos de *Machine Learning*

mais utilizados em quatro diferentes conjuntos de dados com objetivo de diagnosticar o TEA por meio do rastreio. De acordo com os experimentos, os melhores resultados foram obtidos com C-SVC, um classificador baseado em uma máquina de vetores de suporte.

O trabalho de [Shohieb et al. 2019] por outro lado, realizou coletas de dados de crianças de 16 a 30 meses de idade, a partir do preenchimento de um questionário conhecido por Autism Barta, baseado no padrão M-CHAT, em Bangladesh. O objetivo era considerar na base de dados apenas crianças com residência permanente no país a fim de descobrir características regionais e classificar uma criança com TEA ou não. Classificadores baseados em árvores, como J48, LMT (*Logistic Model Tree*), RF, REPTree (*Reduced Error Pruning Tree*), DS (*Data Science*), foram usados para construir várias árvores de decisão e, em seguida, utilizada a técnica de validação cruzada 10 vezes. O algoritmo J48 foi o melhor classificador do experimento, com 98,44% de acurácia.

3. Método

A pesquisa adotou um delineamento quantitativo, de caráter descritivo e exploratório, que tem como objetivo comparar os algoritmos de *Machine Learning* e avaliar a precisão para prever TEA em crianças e adultos a partir de bases de dados pública. A Figura 1 ilustra as etapas do processo, que consiste em analisar e tratar os dados (pré-processamento), dividir as bases de dados em dados de treinamento (que serão aplicados ao modelo) e dados de teste (que serão aplicados ao algoritmo) e extrair o resultado da classificação.

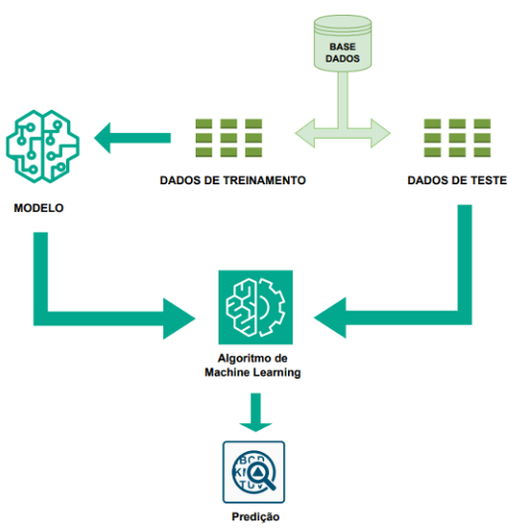

Figura 1. Fluxograma do Modelo Proposto

Para realizar o experimento proposto pelo estudo, foram obtidas três bases de dados distintas, cada uma direcionada a diferentes faixas etárias no contexto de triagem do TEA. A primeira base de dados[1] foi adquirida do *UCI Machine Learning Repository*[2], contendo dados de triagem para crianças (4-12 anos). A segunda base[3] foi extraída do *Kaggle*[4] e aborda a triagem de autismo em adultos (18+ anos), oferecendo uma perspectiva sobre a identificação do transtorno em uma população mais madura. Por fim, a terceira base[5],

[1]https://archive.ics.uci.edu/dataset/419/autistic+spectrum+disorder+screening+data+for+children
[2]https://archive.ics.uci.edu/
[3]https://www.kaggle.com/datasets/andrewmvd/autism-screening-on-adults
[4]kaggle.com
[5]https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers

também proveniente do *Kaggle*, foca na triagem de autismo para crianças pequenas, complementando o conjunto de dados com informações sobre a detecção precoce em uma fase ainda mais inicial da vida (0-3 anos).

Embora sejam bases distintas, elas compartilham características e classes bastante semelhantes. Essas bases foram escolhidas devido à simplicidade de seus dados e à semelhança entre as classes, o que facilita a comparação entre diferentes faixas etárias no contexto da triagem do TEA. Cada base inclui informações sobre 10 perguntas relacionadas a características comuns em pessoas com TEA, além de dados básicos como idade, etnia, entre outros, que serão descritos detalhadamente posteriormente. Antes da aplicação dos algoritmos, os dados foram submetidos a um rigoroso pré-processamento. Esta etapa incluiu a limpeza dos dados, removendo entradas inconsistentes, duplicadas ou incompletas, e a normalização, ajustando as variáveis para uma escala comum. Para tratar o desbalanceamento das classes em todas as bases, aplicamos a técnica SMOTE (*Synthetic Minority Oversampling Technique*). Isso assegura uma representação equilibrada das classes. Essa técnica é fundamental para melhorar a análise dos dados, proporcionando melhores resultados nas métricas utilizadas, que neste estudo são acurácia, precisão, *recall* e *F1 score*.

Após realizar o pré-processamento dos dados, descrito na subseção 4.2, foram escolhidos cinco algoritmos de *Machine Learning* para realização dos experimentos: *Decision Tree* (DT), *RF*, *KNN*, *Naive Bayes* (NB) e *Deep Learning* (DL) a Para realizar uma análise mais precisa e conseguir examinar as bases de dados utilizando mais algoritmos. Com o propósito de identificar os hiperparâmetros mais adequados visando aprimorar a acurácia do modelo, empregou-se a técnica de otimização conhecida como *Grid Search*, que opera mediante uma busca exaustiva em uma grade pré definida de parâmetros.

O *Grid Search* realizará todas as combinações possíveis dentro do espaço de hiperparâmetros , a fim de selecionar a configuração que oferece o melhor desempenho, medido pela métrica de maior acurácia, para o modelo em questão. Esta técnica foi aplicada em todos os algoritmos utilizados, exceto no modelo de DL. Devido ao grande número de hiperparâmetros envolvidos em DL, como o número de camadas, neurônios por camada, taxa de aprendizado, tamanho do batch, número de épocas, função de ativação, dropout rate e otimizadores, o Grid Search tornaria a análise excessivamente demorada.

Para a análise e aplicação dos algoritimos nas bases, os dados foram divididos em conjuntos de treinamento (70%) e teste (30%). A fim de ter uma análise mais acurada, cada algoritimo foi rodado 30 vezes e com os resultados feito uma média artimética.

4. Experimentos

Para dar início ao processo, o qual envolve um experimento que faz uso de *Machine Learning* foram utilizados os seguintes recursos: a linguagem de programação *Python*, e as *bibliotecas NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, TensorFlow/Keras, Google Colab e Imbalanced-learn*.

4.1. Bases de dados

A base de dados utilizada disponibilizada por Fadi Fayez Thabtah no repositório UCI Machine Learning (Base de dados I), trás dados a respeito de crianças e é composta por informações sobre 292 crianças, com 21 atributos relacionado a características e dez atributos com variáveis comportamentais, cujas respostas eram armazenadas como dados binários (0 ou 1). Embora sejam provenientes de repositórios diferentes, as bases de dados possuem atributos semelhantes, variando poucas classes entre elas. A base de dados rela-

cionada a adultos (Base de Dados II) compartilha as mesmas classes que a base de dados I, totalizando 704 linhas e 22 colunas. A descrição dos atributos pode ser vista na Tabela 1.

Tabela 1. Significado dos atributos da base de dados I e II - Tabela I

Variável	Descrição
A1_Score	Alta percepção de ruídos
A2_Score	Maior concentração da visão em relação a pequenos detalhes
A3_Score	Facilidade de comunicação
A4_Score	Facilidade em fazer várias tarefas ao mesmo tempo
A5_Score	Dificuldade em manter conversa com seus colegas
A6_Score	Facilidade em manter conversas informais
A7_Score	Dificuldade de percepção de intenção e sentimento em histórias
A8_Score	Dificuldade de brincar com a imaginação, fugindo da realidade
A9_Score	Facilidade em reconhecer sentimentos por meio de expressões faciais alheias
A10_Score	Dificuldade em fazer novas amizades
Age	Idade
Gender	Gênero
Ethnicity	Etnia
Jundice	Icterícia
Autism	Autismo
Country_of_res	País de residência
Used_app before	Aplicativo usado antes
Result	Resultado
Age_desc	Descrição de idade
Relation	Relação
Class/ASD	Classe/TEA

A base de dados de crianças pequenas (Base de Dados III) apresenta algumas diferenças nos atributos em comparação com as outras bases. Os dez atributos comportamentais permanecem os mesmos, assim como os atributos *'Age Mons'* (relacionado à idade em meses), *'Sex'* (equivalente à classe *'gender'* das outras bases), *'Ethnicity'* (relacionado à etnia), *'Jaundice'* (relacionado à icterícia) e *'Class/ASD Traits'* (relacionado a Classe/TEA). Alguns dos atributos são exclusivos dessa base, como pode ser visto na Tabela 2. Ao todo, esta base tem um total de 1054 linhas e 19 colunas.

Tabela 2. Atributos exclusivos da base de dados III

Variável	Descrição
Qchat-10-Score	relacionado as 10 questões sobre variáveis comportamentais (Menor ou igual a 3 traços sem TEA; maior que 3 traços de TEA)
Family mem with ASD	Familiar com histórico de TEA
Who completed the test	Quem está concluindo o teste

4.2. Pré-processamento dos dados

O pré-processamento dos dados é uma etapa inicial para experimentos que envolvem *Machine Learning* e consiste em analisar os dados para detectar valores nulos ou inconsistentes, a fim de remover informações desnecessárias ou que possam prejudicar o resultado do experimento. Inicialmente, constatou-se que não há ocorrência de dados nulos na Base de Dados III. Os atributos categóricos com valores *'no'* (Não) e *'yes'* (Sim) foram substituídos por valores binários 0 e 1, respectivamente. Essa substituição foi aplicada nas classes *'Jaundice', 'Family_mem_with_ASD' e 'Class/ASD Traits'*. Para a classe *'Sex'*, os valores foram codificados como 0 para 'f' (feminino) e 1 para 'm' (masculino) em todas as bases.

Durante o pré-processamento, na base I, foi identificada a necessidade de substituir um dos valores na coluna *'AGE'* (Idade) devido à presença de um espaço em branco. Uma vez que esta coluna se refere a idades, optou-se por calcular a média aritmética dos valores

existentes e, posteriormente, efetuou-se a substituição pelo valor resultante, que foi estabelecido como sendo 6, o que também ocorreu na base de dados II, sendo este, substituído pelo valor 29 (média aritmética das idades)

Na preparação dos dados das bases I e II, os atributos da base *autism*, *used_app_before* e *Class/ASD* foram alterados para seguir a mesma convenção (Não: 0; Sim: 1). Os atributos, *id*, *age_desc*, *result* e *ethnicity* foram removidos por não serem relevantes para este estudo. Após o pré-processamento, o conjunto de dados resultante possui as seguintes características: a base de dados I contém 292 linhas e 18 colunas, enquanto a base de dados II possui 704 linhas e 18 colunas. Na base de dados III, foram removidos os atributos *"Case No"* (O índice) , *"Who completed the test"* e *"Qchat 10 Score"*. Este último foi excluído por fornecer diretamente o resultado do rastreamento de TEA, comprometendo assim a viabilidade do rastreamento. Ao final do processamento, a base de dados III teve, ao todo, 1054 linhas e 15 colunas.

Nestas bases de dados, a classe de saída era *"Class/TEA"*. Após a análise, constatou-se que essa classe estava desbalanceada. Portanto, foi necessário empregar técnicas de balanceamento dos dados. A técnica escolhida foi o SMOTE, que identifica a classe minoritária e gera novos exemplos sintéticos. O SMOTE seleciona aleatoriamente um ou mais dos k-vizinhos mais próximos da classe minoritária e cria novos pontos de dados no espaço entre o exemplo original e os vizinhos por meio de interpolação. Essa interpolação é feita ao escolher um ponto ao longo da linha que conecta o exemplo original a um dos seus vizinhos. O processo é repetido até que a classe minoritária seja suficientemente ampliada para equilibrar a distribuição das classes.

4.3. Hiperparâmetros

Os hiperparâmetros são parâmetros configuráveis de um algoritmo de Machine Learning que precisam ser definidos antes do processo de treinamento. Eles não são aprendidos a partir dos dados, mas são essenciais para guiar a aprendizagem do modelo. A seleção adequada dos hiperparâmetros é crucial para otimizar o desempenho dos algoritmos, e diversas técnicas podem ser utilizadas para encontrar as melhores combinações. Portanto, com base nos resultados observados com *Grid Search*, os hiperparâmetros foram ajustados de forma a otimizar o desempenho e evitar problemas como *overfitting*. A seguir, detalhamos os valores específicos de cada hiperparâmetro conforme configurados nos experimentos.

Para a DT, foram configurados os seguintes hiperparâmetros: *max_depth* foi definido como 10, limitando a profundidade máxima da árvore para evitar que ela se torne complexa e reduza o overfitting. O *min_samples_split* foi definido como 2, determinando o número mínimo de amostras necessárias para dividir um nó. O *min_samples_leaf* foi definido como 1, especificando o número mínimo de amostras que um nó folha deve ter. Por fim, o *criterion* foi configurado como "gini", utilizado para medir a qualidade de uma divisão.

No caso do RF, os hiperparâmetros configurados foram: *n_estimators* definido como 100, representando o número de árvores na floresta, onde um valor mais alto geralmente melhora a performance, mas também aumenta o tempo de processamento. O *max_features* foi configurado como "auto", definindo o número de características a serem consideradas ao procurar a melhor divisão. O *bootstrap* foi configurado como True, indicando que as amostras devem ser retiradas com reposição. Além disso, o *oob_score* foi definido como True, permitindo a utilização de amostras fora da bolsa para avaliar a precisão do modelo.

Para o KNN, os hiperparâmetros configurados foram: *n_neighbors* definido como 5, especificando o número de vizinhos a serem considerados para classificar um ponto. O *weights* foi configurado como "uniform", determinando que todos os vizinhos têm o mesmo

peso. O *algorithm* foi definido como ”auto”, especificando que o algoritmo usado para encontrar os vizinhos mais próximos será selecionado automaticamente pelo Scikit-learn. O *p* foi configurado como 2, definindo a potência do parâmetro de distância de Minkowski, onde $p = 2$ é equivalente à distância euclidiana.

Para o Gaussian Naive Bayes, o principal hiperparâmetro configurado foi *var_smoothing*, definido como 1×10^{-9}. Este valor adiciona uma pequena quantidade de variação a cada característica para evitar divisões por zero, crucial para estabilizar a estimativa da variância das características e garantir que o modelo funcione corretamente, mesmo com dados que têm variação muito baixa.

No contexto de DL, os hiperparâmetros configurados incluíram: o *número de camadas* e *neurônios por camada*, que foram definidos como 3 camadas com 64, 32 e 16 neurônios, respectivamente. A *taxa de aprendizado* foi configurada como 0.001, definindo o tamanho dos passos que o modelo dá na direção da minimização da função de perda durante o treinamento. O *batch size* foi definido como 32, e o *número de épocas* foi configurado como 50, determinando quantas vezes o modelo passa pelos dados. As *funções de ativação* utilizadas foram ReLU para as camadas ocultas e Sigmoid para a camada de saída. A *dropout rate* foi configurada como 0.5, ajudando a prevenir overfitting ao desligar aleatoriamente uma fração dos neurônios durante o treinamento. Por fim, o *otimizador* escolhido foi Adam, ajustando os pesos do modelo baseado no gradiente da função de perda.

5. Resultados

Para compreender melhor as relações entre as variáveis do conjunto de dados, foi realizada uma análise de correlação. A matriz de correlação resultante fornece uma visão detalhada das inter-relações entre as diferentes variáveis. Em particular, foi focado na identificação das variáveis que possuem correlações significativas com a classe de saída *“Class/ASD”*, que indica a presença de autismo. A análise da matriz de correlação da base de dados I revela que as variáveis de pontuação dos atributos (especialmente *“A5_Score”* e *“A9_Score”*) têm correlações moderadas a altas com *“Class/ASD”*, sugerindo que são importantes para a previsão do TEA, isso também ocorre na matriz de correlação da base II.

Nas três bases, as classes *“Age”* (idade), *“Gender”* (gênero), *“jundice”* (icterícia), *“austim”* (autismo), *“contry_of_res”* (país de residência) e *“relation”* tem baixas correlações com a classe de saída, sugerindo que estes pontos podem não ser um fator importante na determinação das outras variáveis. Quando é analisado a matriz de correlação, é possivel notar que as variáveis de pontuação dos atributos (especialmente “A5”, “A6”, “A7” e “A9”) têm correlações altas com *“Class/ASD Traits”*, o que pode ser inferido que elas tem grande relevância para o rastreio de autismo.

As métricas utilizadas para a análise dos dados foram a acurácia, precisão, *F1-score* e recall, cada métrica foi iniciada 30 vezes e retirado a média arimética entre elas. Deve-se considerar a interpretação das métricas de performance. A acurácia, embora seja uma métrica útil, pode ser enganosa em conjuntos de dados desbalanceados. Outras métricas como precisão, *recall* e *F1-score* oferecem uma visão mais completa da performance dos modelos, especialmente em contextos onde o custo de falsos positivos e falsos negativos é diferente. Portanto, uma avaliação holística e crítica dos resultados é necessária para garantir a validade das conclusões tiradas a partir dos experimentos.

A Tabela 3 apresenta os resultados obtidos após a aplicação de cinco algoritmos de ML em três diferentes bases de dados: DT, RF, KNN, NB e DL. Cada algoritmo foi avaliado com base em quatro métricas principais: acurácia, precisão, *recall* e *F1-Score*. A acurácia

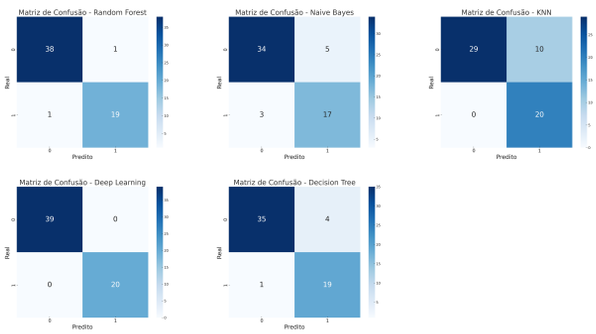

Figura 2. Matrizes de Confusão para Base de Dados I

mede a proporção de predições corretas. A precisão reflete a proporção de verdadeiros positivos entre os positivos preditos. O *recall* indica a capacidade do algoritmo de identificar corretamente as instâncias positivas, e o *F1-Score* é a média harmônica da precisão e do *recall*. Essas métricas são essenciais para avaliar o desempenho dos modelos, oferecendo uma visão abrangente de sua eficácia em diferentes cenários.

Tabela 3. Métricas de desempenho dos modelos para as três bases de dados

Base de Dados	Modelo	Acurácia	Precisão		Recall		F1 Score	
			Classe 0	Classe 1	Classe 0	Classe 1	Classe 0	Classe 1
I	DT	0.92	0.97	0.84	0.91	0.94	0.93	0.88
	RF	0.94	0.98	0.88	0.93	0.96	0.95	0.92
	KNN	0.83	1.00	0.67	0.74	1.00	0.85	0.80
	NB	0.86	0.92	0.77	0.87	0.85	0.89	0.81
	DL	0.98	1.00	0.96	0.98	1.00	0.99	0.98
II	DT	0.91	0.98	0.76	0.90	0.95	0.94	0.85
	RF	0.99	0.99	0.97	0.99	0.98	0.99	0.97
	KNN	0.91	1.00	0.75	0.89	1.00	0.94	0.86
	NB	0.99	0.99	0.97	0.99	0.97	0.99	0.97
	DL	1.00	1.00	0.99	1.00	1.00	1.00	1.00
III	DT	0.92	0.84	0.97	0.93	0.91	0.88	0.94
	RF	0.97	0.95	0.99	0.97	0.98	0.96	0.98
	KNN	0.97	0.91	1.00	1.00	0.95	0.95	0.97
	NB	0.96	1.00	0.94	0.87	1.00	0.93	0.97
	DL	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Para uma análise mais aprofundada, também foi utilizada a matriz de confusão. Esta última é particularmente importante em questões de saúde, pois os índices de falsos positivos e, principalmente, falsos negativos, têm grande impacto. Um falso negativo significa que uma criança que deveria ter recebido um diagnóstico não terá as intervenções educacionais e de saúde necessárias para seu desenvolvimento. Cada base de dados gerou cinco matrizes de confusão, sendo elas ilustradas nas Figuras 2, 3, 4.

5.1. Discussão Geral dos Resultados

Nesta seção, será analisada a viabilidade da aplicação de técnicas de machine learning para o rastreamento de sinais de TEA nas três bases de dados provenientes de repositórios públicos. Além disso, serão comparadas as performances dos modelos em cada uma das bases por meio das métricas extraídas, discutindo os pontos fortes e limitações.

Os resultados obtidos através da aplicação dos cinco algoritmos de ML nas três bases de dados fornecem insights importantes sobre a aplicabilidade de cada técnica em diferentes

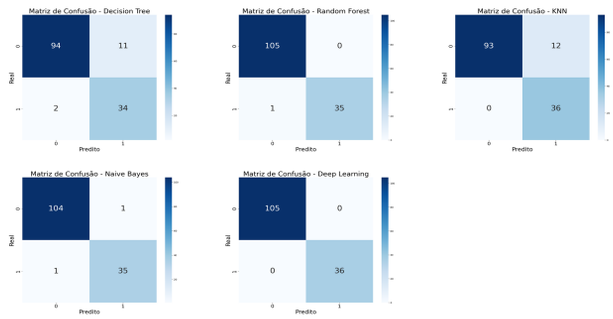

Figura 3. Matrizes de Confusão para Base de Dados II

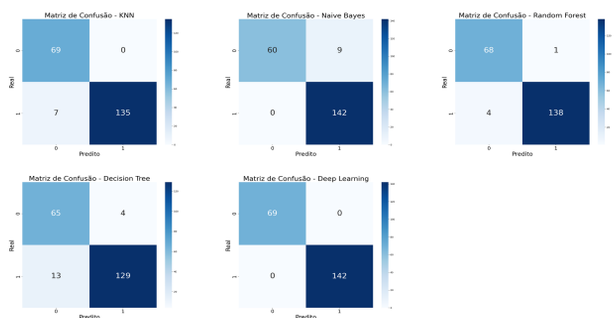

Figura 4. Matrizes de Confusão para Base de Dados III

contextos. A Tabela 3 revela variações significativas nas métricas de desempenho, refletindo a capacidade de cada algoritmo de lidar com as características específicas dos dados. A classe 0 refere-se à ausência de TEA, enquanto a classe 1 refere-se à presença de TEA.

Em relação ao algoritmo RF, a acurácia foi alta nas três bases de dados, com um desempenho ligeiramente superior nas Bases de Dados I e III (92%). Na Base de Dados II, a precisão para a Classe 1 foi de 76%, enquanto a Base de Dados III se destacou com uma taxa de 97%. Ao analisar a matriz de confusão, observamos a ocorrência de 4 falsos positivos e 13 falsos negativos na Base de Dados III.

O número elevado de falsos negativos é preocupante, pois indica que 13 pessoas deixariam de receber a intervenção necessária para melhorar seu desenvolvimento. No entanto, apesar desse alto índice, a Base de Dados III apresentou um desempenho percentual maior, pois o número de dados dessa base é superior ao das outras bases. Considerando todas as métricas analisadas, a Base de Dados III obteve um desempenho melhor em uma análise geral. O algoritmo DT demonstrou um melhor desempenho no *recall* e no *F1-Score* na Base de Dados III, indicando maior assertividade e viabilidade para uso prático. Estes resultados reforçam a capacidade do algoritmo de identificar corretamente os casos positivos e negativos, apesar dos desafios apresentados pelos falsos negativos.

O algoritmo RF apresentou um desempenho notável em todas as bases de dados, com destaque para a Base de Dados II, onde alcançou uma acurácia de 99%. Além disso, os outros valores de desempenho também foram superiores nesta base. Embora o algoritmo tenha se destacado na Base de Dados II, ele também apresentou boa performance nas demais métricas nas outras bases de dados. Em relação à matriz de confusão, o RF demonstrou baixos índices de falsos positivos e falsos negativos em todas as bases de dados. Isso reforça sua eficácia na detecção de sinais de TEA em diferentes faixas etárias.

Na Base de Dados I, o modelo obteve uma acurácia de 94%, com apenas 1 falso pos-

itivo e 1 falso negativo, indicando um equilíbrio sólido entre precisão e *recall*. Na Base de Dados III, a acurácia foi de 97%, com 4 falsos negativos e 1 falso positivo, o que novamente demonstra sua capacidade de minimizar erros de classificação.

O algoritmo KNN apresentou um desempenho variado entre as bases de dados utilizadas, destacando-se mais na Base de Dados III, onde obteve métricas elevadas. Na Base de Dados II, o KNN teve uma acurácia de 91% e registrou 12 ocorrências de falsos positivos. Embora a acurácia tenha melhorado, a presença de um número significativo de falsos positivos indica que o modelo tende a classificar algumas instâncias negativas reais como positivas, o que é uma limitação importante. Na Base de Dados I, a acurácia do KNN foi relativamente baixa (83%) em comparação com as outras duas bases de dados. A precisão para a Classe 1 foi de 67%, e o recall para a Classe 0 foi de 74%, valores que, em comparação com outras bases, são baixos. A matriz de confusão revelou um total de 10 falsos positivos e nenhum falso negativo. Embora a ausência de falsos negativos seja um aspecto positivo, o número relativamente alto de falsos positivos reduz a confiabilidade do modelo.

De maneira análoga, o NB apresentou um desempenho variável nas três bases de dados analisadas. Na Base de Dados I, o NB alcançou uma acurácia de 86%. A matriz de confusão revelou um total de 3 falsos negativos e 5 falsos positivos. A precisão para a Classe 1 foi de 77%, enquanto o recall para a Classe 0 foi de 87%. Embora esses índices sejam razoáveis, o número de falsos positivos e negativos pode ser preocupante para cenários educacionais e de saúde, onde a precisão é crucial. Na Base de Dados III, o NB obteve uma acurácia de 96%. A matriz de confusão revelou um total de 9 falsos positivos. Embora a acurácia e o recall sejam elevados, o quantitativo de falsos positivos ainda representa um desafio, pois pode levar a diagnósticos incorretos que impactam negativamente o tratamento e a intervenção precoce.

Em contraponto, na Base de Dados II, o NB teve um desempenho muito forte, com uma acurácia de 99%. A matriz de confusão mostrou apenas 1 falso negativo e 1 falso positivo. O modelo apresentou precisão e recall muito elevados, resultando em F1 Scores altos para ambas as classes. Neste caso, o algoritmo demonstrou ser uma ferramenta eficaz no rastreio, mostrando um excelente equilíbrio entre sensibilidade e especificidade.

O DL apresentou resultados notavelmente altos em todas as bases de dados analisadas, mas alguns índices extremamente elevados, especialmente na Base de Dados III, requerem uma análise mais crítica. Na Base de Dados III, todas as métricas analisadas atingiram 100% em todos os casos, o que levanta suspeitas sobre a possibilidade de overfitting. De maneira similar, na Base de Dados II, o modelo alcançou uma acurácia de 100%, com todas as métricas também perfeitas, exceto a precisão da Classe 1, que foi de 99%.

Essa perfeição nos resultados indica que o modelo pode estar ajustado de forma muito específica aos dados de treinamento, o que comprometeria sua capacidade de generalização para novos conjuntos de dados. Overfitting é uma preocupação significativa em modelos de DL, pois pode levar a um desempenho enganadoramente alto nos dados de treinamento, mas pobre em dados não vistos.

Diferente das outras bases, a Base de Dados I apresentou uma acurácia de 98%, com a matriz de confusão revelando 0 falsos positivos e 0 falsos negativos. Esse desempenho resultou em precisão e recall de 100% para ambas as classes, além de altos valores de *F1-Score* (0.99 e 0.98). Embora o desempenho seja excelente, ele é mais realista e sugere uma menor possibilidade de *overfitting* em comparação com as Bases de Dados II e III.

De maneira geral, os algoritmos foram capazes de realizar ótimas previsões e apresentaram uma alta taxa de precisão na diagnose de TEA em estudantes de diferentes faixas

etárias. O algoritmo RF demonstrou-se consistentemente eficaz em todas as bases de dados analisadas, destacando-se por sua alta acurácia e baixos índices de erros de classificação.

As três bases de dados eram compostas por características simples de estudantes de diferentes faixas etárias e provenientes de diversos países. Os algoritmos alcançaram altos índices de assertividade, o que demonstra seu potencial para serem utilizados por professores, pedagogos e profissionais da educação na diagnose de sinais de TEA. Isso possibilita a implementação das intervenções necessárias e tratamentos adequados, visando melhorar o desenvolvimento dos estudantes tanto no campo escolar quanto no social.

5.2. Ameaças da Avaliação

Embora os resultados sejam promissores, é essencial considerar possíveis ameaças à validade que podem impactar a interpretação dos resultados. O viés de amostragem é uma ameaça significativa; se a base de dados utilizada não for representativa da população-alvo, os modelos podem não generalizar bem para novos dados. Além disso, a qualidade dos dados é crucial, pois dados incompletos ou incorretos podem levar a conclusões errôneas sobre a performance dos modelos. Outra ameaça é o *overfitting*, especialmente com técnicas de otimização como o *Grid Search*, que podem fazer com que o modelo se ajuste excessivamente aos dados específicos utilizados, comprometendo sua capacidade de generalização. É importante avaliar os modelos com conjuntos de dados independentes para validar sua robustez. A análise da matriz de confusão é crucial para entender a distribuição de falsos positivos e falsos negativos, e seu impacto na prática clínica, como diagnósticos incorretos que podem afetar negativamente o desenvolvimento das crianças.

Além disso, a variabilidade dos dados entre diferentes bases de dados pode influenciar os resultados dos modelos de ML. Realizar uma análise comparativa das performances dos modelos em diferentes bases de dados é fundamental para garantir a validade e a generalização dos resultados. O uso de uma base de dados de adultos juntamente com bases de dados de crianças permite avaliar a generalização dos modelos em diferentes faixas etárias, aumentando a utilidade prática dos modelos ao verificar se um modelo eficaz para adultos pode ser aplicado em crianças e vice-versa.

6. Agradecimentos

Este trabalho foi apoiado pela Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), através do Processo IBPG-2308-1.03/22.

7. Conclusão

Este estudo comparou diversos modelos de machine learning, como *DT*, *RF*, *KNN*, *NB* e *DL*, para aprimorar a detecção precoce do TEA em diferentes faixas etárias utilizando três bases de dados públicas. Os resultados mostraram que o *RF* teve um desempenho superior em termos de acurácia, precisão, *recall* e *F1-score*, demonstrando robustez e eficácia consistentes nas três bases de dados.

A análise das métricas revelaram que o *Random Forest* (RF) apresenta alta acurácia, juntamente com baixas taxas de falsos negativos e falsos positivos, mostrando-se eficaz no diagnóstico precoce do TEA. No entanto, o estudo reconhece certas limitações, como viés de amostragem e o risco de *overfitting*, ressaltando a importância de pesquisas futuras para explorar novos modelos e técnicas de otimização. Ademais, é recomendada a utilização de outras bases de dados para obter novas perspectivas e permitir comparações com os resultados desta pesquisa. Em síntese, o estudo evidencia o potencial significativo do RF na detecção precoce do TEA, com a possibilidade de transformar a prática clínica por meio de diagnósticos mais rápidos e precisos, além de promover intervenções mais eficazes.

Referências

[Araújo et al. 2021] Araújo, P. H., dos Santos, V. A., and Borges, I. C. (2021). O autismo e a inclusão na educação infantil: estudo e revisão. *Brazilian Journal of Development, DOI: https://doi.org/10.34117/bjdv7n2-563*.

[Brasil. Ministério da Saúde. 2014] Brasil. Ministério da Saúde. (2014). *Diretrizes de Atenção à Reabilitação da Pessoa com Transtornos do Espectro do Autismo (TEA)*. Ministério da Saúde, Brasília.

[Brentani et al. 2013] Brentani, H., Paula, C. S. d., Bordini, D., Rolim, D., Sato, F., Portolese, J., Pacifico, M. C., and McCracken, J. T. (2013). Autism spectrum disorders: an overview on diagnosis and treatment. *Brazilian Journal of Psychiatry*, 35:S62–S72.

[Deng et al. 2021] Deng, L., Rattadilok, P., and Xiong, R. (2021). A machine learning-based monitoring system for attention and stress detection for children with autism spectrum disorders. In *Proceedings of the International Conference on Intelligent Medicine and Health*. ACM.

[Frota et al. 2019] Frota, M., Vilela, M., Hericles, S., Aguiar, G., Renoir, P., Nunes, R., Gomes, D., and Cavalcante, I. (2019). Aplicação de Árvore de decisão para auxílio ao diagnóstico do transtorno do espectro autista. In *Anais da VII Escola Regional de Computação Aplicada à Saúde*. SBC. [Acesso em: 14 ago. 2023].

[Gioia et al. 2021] Gioia, P. S., Barbieri, L., Guilhardi, C., Sarilho, C. A., Vargas, D. K., de Carvalho, D. C. B., Costa, M. M., and Keiner, S. A. (2021). Protocolo de avaliação e intervenção precoces de sinais de risco de autismo: comparando grupos de alto e baixo risco. *SciELO Preprints*.

[Gois et al. 2022] Gois, T., Cordeiro, A. A. d. A., Pernambuco, L., and Queiroga, B. (2022). Risk identification for autistic spectrum disorder in preschool children: Design and validation of a screening instrument. *SciELO Preprints*.

[Gupta and Hafiz 2022] Gupta, K. N. and Hafiz, G. (2022). Accurate estimate of autism spectrum disorder in children utilizing several machine learning techniques. In *14th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE.

[Gyori et al. 2018] Gyori, M. et al. (2018). Automated vs human recognition of emotional facial expressions of high-functioning children with autism in a diagnostic-technological context: Explorations via a bottom-up approach. In *Lecture Notes in Computer Science*, pages 466–473. Springer International Publishing.

[INEP 2021] INEP (2021). Caderno de conceitos e orientações do censo escolar 2021. *Disponível em: https://bit.ly/3SEkkWO Data do último acesso: 25/05/2022*.

[Lowri 2021] Lowri, C. (2021). Issues in persistent non attendance at school of autistic pupils and recommendations following the reintegration of 11 autistic pupils. *Good Autism Practice (GAP)*, 22:12–20.

[Munkhaugen et al. 2017] Munkhaugen, E., Gjevik, E., Pripp, A., Sponheim, E., and Diseth, T. (2017). School refusal behaviour: are children and adolescents with autism spectrum disorder at a higher risk? *Research in Autism Spectrum Disorders*, 41-42:31–38.

[Overland et al. 2022] Overland, E., Hauge, A. L., Orm, S., Pellicano, E., Oie, M. G., Skogli, E. W., and Andersen, P. N. (2022). Exploring life with autism: Quality of life, daily functioning and compensatory strategies from childhood to emerging adulthood: A qualitative study protocol. *Frontiers in Psychiatry*, 13.

[Santos et al. 2021] Santos, J. O. L., Sadim, G. P. T., Schmidt, C., and de S. Matos, M. A. (2021). O atendimento educacional especializado para os educandos com autismo na rede municipal de manaus-am. *Revista Brasileira de Estudos Pedagógicos (RBEP), DOI: https://doi.org/10.24109/2176-6681.rbep.102.i260.4150*.

[Shinde and Patil 2023] Shinde, A. V. and Patil, D. D. (2023). A multi-classifier-based recommender system for early autism spectrum disorder detection using machine learning. *Healthcare Analytics*, 4(100211):100211.

[Shohieb et al. 2019] Shohieb, S. M. et al. (2019). Early detection of autism by extracting features: A case study in bangladesh. In *International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. IEEE.

[Wang et al. 2011] Wang, H., Sandall, S. R., Davis, C. A., and McIntosh, K. E. (2011). Social skills assessment in young children with autism: a comparison evaluation of the ssrs and pkbs. *Journal of Autism and Developmental Disorders*, 41(11):1487–1495.

[Yang et al. 2019] Yang, X., Islam, M. S., and Khaled, A. M. A. (2019). Functional connectivity magnetic resonance imaging classification of autism spectrum disorder using the multisite abide dataset. In *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE.

[Yesilyurt and Diagnosing 2021] Yesilyurt, T. H. and Diagnosing, S. (2021). Diagnosing autism spectrum disorder using machine learning techniques. In *6th International Conference on Computer Science and Engineering (UBMK)*. IEEE.