

Saúde do sono e estilo de vida: Classificação de distúrbio do sono com redes neurais.

Emanuela B. de Lima, Matheus Vinicius F. Pinheiro

Abstract—Sleep is a fundamental physiological state for animals, including humans, essential for memory, health, and energy restoration. Sleep disorders, such as insomnia and sleep apnea, affect physical, cognitive, and social functioning, potentially causing health problems and impacting quality of life. Data collection for this study was conducted through the Kaggle platform. The neural network model used was the Multilayer Perceptron, also known as a multilayer network. The research goal was to develop a neural network model capable of classifying sleep disorders from a set of input parameters. The model achieved the following metrics: accuracy 0.9545, precision 0.9511, sensitivity 0.9518, and F1-Score 0.9514. The model proved to be quite stable and achieved good learning rates.

Keywords—sleep disorder, neural network, Perceptron

I. INTRODUÇÃO

O sono é um estado fisiológico natural e essencial para a maioria dos animais, incluindo os seres humanos. É uma função biológica fundamental na consolidação da memória, na visão binocular, na termorregulação, na conservação e restauração da energia [1]. O mesmo tem um papel crucial na manutenção da saúde, seja ela física ou mental, sendo fundamental para o funcionamento adequado do organismo. É um período de repouso no qual o corpo descansa e se recupera, permitindo que funções vitais sejam restauradas e o cérebro processe informações. Dado a essas importantes funções, as perturbações do sono podem acarretar alterações significativas no funcionamento físico, ocupacional, cognitivo e social do indivíduo [2], além de impactar a qualidade de vida e a saúde em geral, como também comprometer a segurança pública. Dado que pessoas com problemas de sono, têm uma propensão maior a problemas de saúde, podendo apresentar uma expectativa de vida reduzida. As estimativas sobre o índice de acidentes e mortes causados por sonolência ou cansaço variam de 2% a 41% [2].

Como vimos, os distúrbios do sono, também conhecidos como transtornos do sono, são uma classe de doenças ou condições médicas que afetam o padrão natural do sono-vigília. A classificação de distúrbios do sono pode variar dependendo dos critérios e sistemas de classificação utilizados. Ambos são muito predominantes na população geral e podem afetar a saúde do sono, bem como, o estilo de vida, resultando em questões como insônia, apneia do sono e outros distúrbios correlacionados. A insônia é o mais comum, e, quando crônica, geralmente reflete distúrbios psicológicos e comportamentais [3]. É um sintoma que pode ser definido como dificuldade em iniciar e/ou manter o sono, presença de sono não reparador, ou seja, insuficiente para manter uma boa qualidade de alerta e bem-estar físico e mental durante o dia, com o comprometimento consequente do desempenho nas atividades diurnas [4]. Dessa forma, indivíduos que sofrem de insônia costumam mencionar

sentir-se cansados durante o dia, podendo enfrentar desafios para realizar suas tarefas diárias.

Estes episódios de insônia podem estar relacionados a vários fatores, que são eles: expectativas, problemas clínicos, problemas emocionais passageiros e entusiasmos associados a determinados eventos [5]. No entanto, pode tornar-se crônica e provocar muito sofrimento ao longo dos anos. Vale ressaltar que, algumas pessoas apresentam maior tendência à insônia e quando expostas a condições de estresse, doenças ou mudança de hábitos, desenvolvem episódios de insônia. Estes fatores podem se prolongar por muito tempo, dado que o indivíduo tende a associar suas dificuldades de dormir a uma série de comportamentos, o que realimenta a insônia [5].

Assim como a insônia, a apneia do sono também compromete a qualidade do sono e o bem-estar de um indivíduo. Uma vez que a respiração nesse distúrbio é interrompida ou diminuída durante o sono. Ou seja, quando o fluxo de ar é bloqueado devido ao colapso das vias aéreas. A apneia obstrutiva do sono caracteriza-se pela obstrução da via aérea no nível da garganta durante o sono, levando a uma parada da respiração, que dura em média 20 segundos [5]. O autor ainda afirma que, após esta parada, a pessoa acorda, emitindo um ronco muito barulhento, podendo ocorrer várias vezes durante a noite. Vale enfatizar que, pessoas com apneia obstrutiva do sono podem desenvolver doenças nas artérias, provocadas pelo acúmulo de colesterol nas suas paredes, além de provocar a ocorrência de infarto do miocárdio e acidente vascular cerebral, como também, desenvolver outros tipos de doenças [5].

Sendo assim, dado ao que foi visto, o objetivo do presente estudo é realizar a classificação de distúrbio do sono através da arquitetura de rede, Perceptron Multilayer, especificando quando há presença ou ausência de um distúrbio na pessoa (nenhum, insônia, apneia do sono) uma vez que, ambas são as principais doenças responsáveis por um sono de má qualidade.

II. COLETA DE DADOS

A coleta de dados deste trabalho foi realizada a partir da plataforma Kaggle, em 25 de Julho de 2023. Esta plataforma é destinada a aprendizado de ciências de dados e é também uma comunidade para assuntos relacionados com uma variedade de informações.

O conjunto de dados compreende 374 linhas e 13 colunas, cobrindo uma ampla gama de variáveis relacionadas ao sono e aos hábitos diários. Inclui detalhes como sexo, idade, ocupação, duração do sono, qualidade do sono, nível de atividade física, níveis de estresse, categoria de IMC, pressão arterial, frequência cardíaca, passos diários e presença ou ausência de distúrbios do sono.

Os principais recursos do conjunto de dados envolve as métricas abrangentes do sono, na qual explora a duração, a qualidade e os fatores do sono que influenciam os padrões de sono. Como também, os fatores de estilo de vida, para realizar a análise dos níveis de atividade física, níveis de estresse e as categorias de IMC. Do mesmo modo que a saúde cardiovascular, na qual se examinam as medições de pressão arterial e frequência cardíaca. E por último, a análise dos distúrbios do sono, para identificar a ocorrência de distúrbios do sono, como insônia e apneia do sono.

A Coluna do conjunto de dados é composta pelos seguintes dados: ID da pessoa, no qual é estabelecido um identificador para cada indivíduo. Sexo: O gênero da pessoa (masculino/feminino). Idade: A idade da pessoa em anos. Ocupação: A ocupação ou profissão da pessoa. Duração do sono (horas): O número de horas que a pessoa dorme por dia. Qualidade do Sono (escala: 1-10): Uma classificação subjetiva da qualidade do sono, variando de 1 a 10. Nível de Atividade Física (minutos/dia): O número de minutos que a pessoa se dedica diariamente à atividade física. Nível de Estresse (escala: 1-10): Uma classificação subjetiva do nível de estresse experimentado pela pessoa, variando de 1 a 10. Categoria de IMC: a categoria de IMC da pessoa (por exemplo, abaixo do peso, normal, acima do peso). Pressão arterial (sistólica/diastólica): A medida da pressão arterial da pessoa, indicada como pressão sistólica sobre pressão diastólica. Frequência cardíaca (bpm): A frequência cardíaca em repouso da pessoa em batimentos por minuto. Passos Diários: O número de passos que a pessoa dá por dia e por fim, os distúrbios do sono: Identificar se há presença ou ausência de um distúrbio do sono na pessoa (Nenhum, Insônia, Apneia do Sono).

Os detalhes sobre a coluna de distúrbios do sono é realizado da seguinte maneira: Nenhum: O indivíduo não apresenta nenhum distúrbio específico do sono. Insônia: O indivíduo experimenta dificuldade em adormecer ou permanecer dormindo, levando a um sono inadequado ou de má qualidade. Apneia do sono: O indivíduo sofre de pausas na respiração durante o sono, resultando em padrões de sono interrompidos e riscos potenciais à saúde.

A imagem 1 ilustra a distribuição de distúrbio entre os gêneros, com 49,5% representando o feminino e 50,5% representando o masculino.

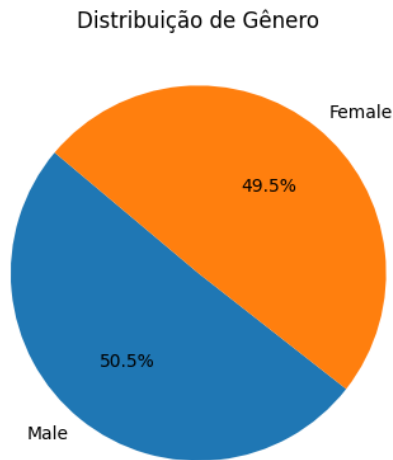


Imagem 1: Distribuição de Gênero
Autores (2023)

Conforme o gráfico, pode-se perceber que o gênero masculino possui um índice maior de distúrbio, comparado ao feminino.

A imagem 2 nos mostra a ocupação por gênero. Nela, podemos identificar que a profissão de enfermagem é predominantemente composta por mulheres e apresenta um índice de distúrbios mais elevado em comparação com outras profissões. Por outro lado, ocupações como médico e advogado exibem uma taxa de distúrbios mais alta entre homens do que entre mulheres. Da mesma forma, as profissões de professor e contador demonstram uma taxa superior entre as mulheres em comparação com os homens. Por fim, observa-se que as funções de gerente e representante de vendas apresentam índices mais baixos em comparação com as outras profissões, sendo que a ocupação de gerente possui a menor taxa e é composta exclusivamente por mulheres.

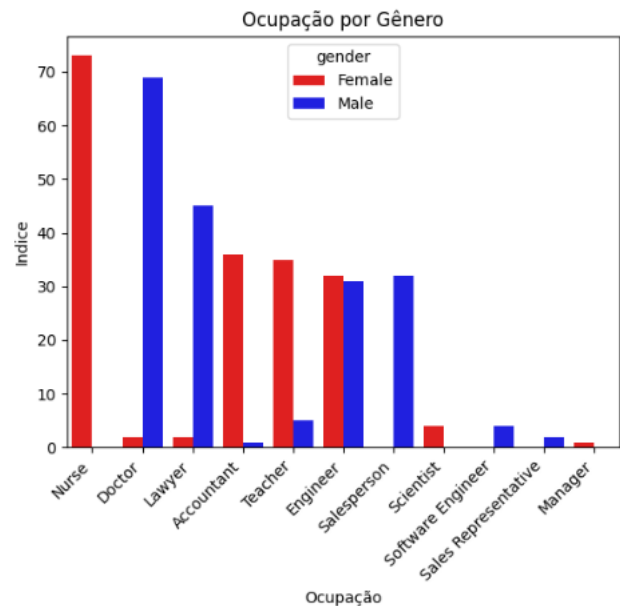


Imagem 2: Ocupação por Gênero
Autores (2023)

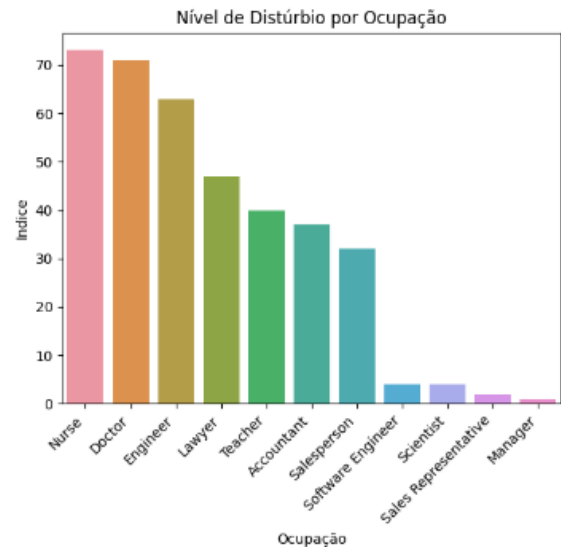
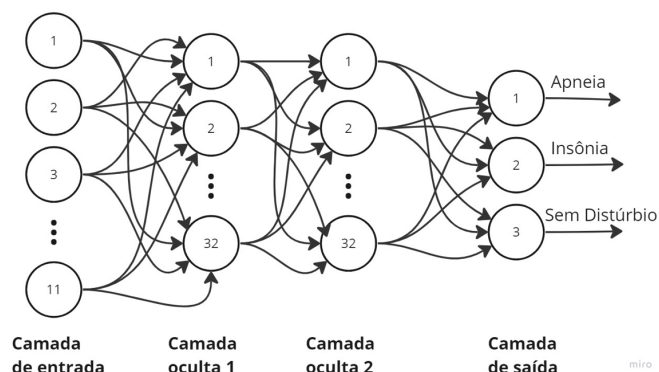


Imagem 3: Nível de Distúrbio por Ocupação
Autores (2023)

A imagem 3 ilustra os níveis de distúrbios associados a diferentes profissões. O gráfico revela que a enfermagem é a profissão com a maior taxa de perturbação, excedendo os 70%. Além disso, em comparação com outras ocupações, a função de gerente exibe uma taxa significativamente menor, com aproximadamente 1%.

Com base nos dados, é possível perceber o quanto o estilo de vida pode impactar a saúde das pessoas, especialmente no que se refere ao sono, que é o foco da pesquisa. A imagem 3 reflete que profissões que envolvem níveis mais elevados de estresse têm uma maior probabilidade de causar distúrbios nos indivíduos, como é o caso de enfermeiros e médicos, que acabam sofrendo as consequências de suas rotinas de trabalho cansativas.

representa de forma diagramática o modelo da rede construído.



IMC por Distúrbio

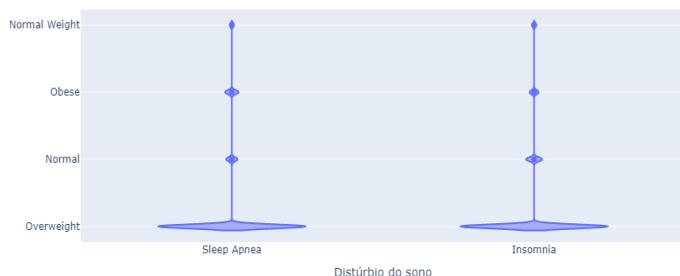


Imagem 4: Distribuição por IMC
Autores (2023)

Segundo a imagem 4 é possível perceber que pessoas com sobrepeso têm maior propensão a apresentarem algum tipo de distúrbio do sono.

Com base nas informações coletadas, foram realizadas as classificações dos distúrbios do sono, a partir dos dados de entrada, nos quais informam qual distúrbio a pessoa apresenta. Foi utilizado o modelo de rede neural, Perceptron Multilayer, também conhecido como rede de multicamadas. Foi fundamental a conversão dos valores de gênero visto que, a base de dados apresentava elementos do tipo caracter, como ocupação, valores IMC e pressão arterial. Para realizar este procedimento, foi utilizado a função `astype` que consegue converter tipos de dados para um tipo especial. Como parâmetro da função, foi utilizado a diretiva "category", responsável por categorizar os dados e convertê-los para números.

III. MODELO DE APRENDIZADO DE MÁQUINA

No desenvolvimento da pesquisa, utilizamos a rede Perceptron Multilayer (também conhecida como rede multicamada). No nosso modelo, implementamos quatro camadas de neurônios: uma de entrada, uma de saída e outras duas camadas internas para processamento dos dados. O objetivo da rede era determinar, a partir de um conjunto de entradas, se um indivíduo possuía ou não algum distúrbio de sono. Nosso modelo é capaz de categorizar três cenários distintos: sem distúrbio, apneia e insônia. A imagem 4

Imagem 5: Rede Perceptron
Autores (2023)

IV. RESULTADOS

A rede Perceptron utilizada na pesquisa foi construída com duas camadas internas de processamento, cada uma contendo 32 neurônios. A função de ativação escolhida foi a sigmoide, que mapeia os dados de entrada para valores dentro do intervalo de 0 a 1. Essa função é particularmente útil para resolver problemas de classificação binária. Na camada de saída, foram utilizados 3 neurônios, levando em consideração as categorias de classificação do modelo. Nessa camada, optou-se pela utilização da função de ativação Softmax, que possui como característica principal, a capacidade de lidar com classificações em várias classes.

Foi necessário realizar alguns procedimentos na base de dados para realizar o processo de classificação, uma vez que ela continha vários elementos do tipo caracter, e as funções de aprendizagem e análise que seriam aplicadas demandam o uso de valores numéricos, foi necessário efetuar uma conversão desses valores. Foram convertidos os valores de gênero, ocupação, valores IMC e pressão arterial. Para realizar o procedimento foi utilizado a função `astype` que consegue converter tipos de dados para um tipo especificado, como parâmetro da função foi utilizado a diretiva "category", responsável por categorizar os dados e converter para números, é muito eficiente quando se pretende realizar operações numéricas entre categoria de dados.

Após a análise da base de dados, tornou-se evidente um desequilíbrio nas categorias. A classe "sem distúrbio" possuía um total de 219 amostras, enquanto as classes "apneia" e "insônia" contavam com 78 e 77 amostras, respectivamente. Para solucionar esse desafio, recorreu à função SMOTE. Essa função desempenha o papel de equilibrar a base de dados, de modo a igualar o número de amostras em todas as classes. Dessa forma, evita-se que o modelo seja influenciado de forma tendenciosa devido a uma predominância excessiva de exemplos de uma classe específica.

Foi utilizado o método `train_test_split`, o qual tem a responsabilidade de dividir a base de dados em conjuntos destinados ao treinamento e ao teste. Como argumento desse método, foi fornecida a base de dados que já havia sido normalizada. Além disso, como instrução para o método, estabeleceu-se que 80% dos elementos seriam alocados para o conjunto de treinamento, enquanto os 20% restantes seriam destinados ao conjunto de testes.

Foi utilizada também uma técnica de regularização, conhecida por prevenir o overfitting em redes neurais. O overfitting ocorre quando um modelo se ajusta excessivamente aos dados de treinamento, não sendo capaz de generalizar para novos dados. A função escolhida foi o Dropout, que funciona desativando aleatoriamente um número específico de neurônios da rede durante o treinamento. Em nosso modelo, o parâmetro foi configurado como 0.1, o que significa que 10% dos neurônios da rede serão desativados aleatoriamente. Essa escolha se deve ao fato de a base de dados não ser muito extensa.

Para medir a precisão do modelo construído, foram utilizadas as seguintes métricas aplicando o módulo `sklearn.metrics`. Sendo elas acurácia, precisão, sensibilidade e f1-score.

A acurácia é capaz de medir o percentual de classificações assertivas do modelo, levando em consideração a seguinte fórmula: $(tp + tn) / (tp + tn + fp + fn)$ [8]. Em nosso caso, como o modelo trabalha com valores no intervalo de 0 a 1, quanto mais próxima de 1 mais acurácia ele possui. A precisão mede a porcentagem de predições positivas que estavam corretas entre todos os exemplos previstos como positivo, calculado dado a fórmula: $tp / (tp + fp)$ [8]. A sensibilidade mede a porcentagem de valores positivos classificados corretamente entre todos os exemplos reais pertencentes a classe positiva, calculado de acordo com a fórmula: $tp / (tp + fn)$ [8]. E a F1-score é a média harmônica entre a precisão e revocação (recall), calculado da seguinte forma:

$$\begin{aligned} \text{pre} &= tp / (tp + fp) \\ \text{rec} &= tp / (tp + fn) \\ (2 * \text{pre} * \text{rec}) / (\text{pre} + \text{rec}) &[8]. \end{aligned}$$

Durante o treinamento da base de dados, foram realizadas um total de 200 épocas, representando cada passagem completa pela base de dados. Nesse cenário, foram obtidas as seguintes métricas de avaliação conforme a tabela abaixo:

Métricas	Resultados
Acurácia	0.9545454545454546
Precisão	0.9511347397485892
Sensibilidade	0.9518831499963575
F1-Score	0.9514081483165366

Tabela 1: Métricas de avaliação

Autores (2023)

De acordo com a tabela 1, pode-se perceber que as métricas de avaliação foram boas, dado que ambas obtiveram valores semelhantes e um resultado de mais de 90%.

A imagem 6 da representação gráfica, ilustra de maneira clara a eficácia do modelo ao comparar os desempenhos entre a base de treinamento e a base de teste. Foi utilizado como função perda, o método Entropia Cruzada, na qual avalia a divergência entre duas distribuições de probabilidade [9]. Esta função, é empregada em aprendizado de máquina e aprendizado profundo para medir o quão bem um modelo de previsão está desempenhando uma tarefa de classificação.

Conforme a imagem 6, é evidente que o modelo conseguiu aprender e adquirir um nível significativo de conhecimento para realizar a classificação a partir de qualquer dado inserido.

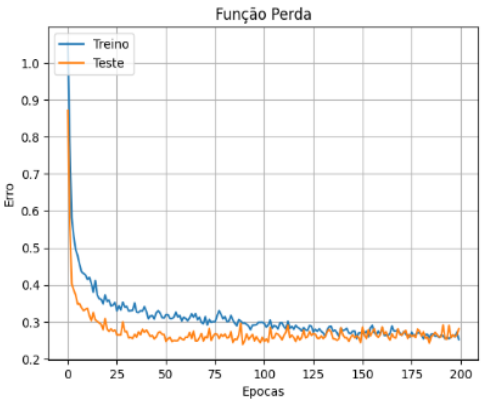


Imagem 6: Função perda
Autores (2023)

A matriz de confusão é uma ferramenta utilizada em aprendizado de máquina e análise de classificação para avaliar o desempenho de um modelo em relação a um conjunto de dados rotulados em um cenário de aprendizado supervisionado [8].

Um classificador binário pode ter quatro resultados de classificação: TP (True Positives, ou Verdadeiros Positivos), TN (True Negatives, ou Verdadeiros Negativos), FP (False Positives, ou Falso-Positivos) e FN (False Negatives, ou Falso-Negativos). As duas primeiras classificações são corretas [8]. Em que, verdadeiro positivo (TP) representa os casos em que o modelo previu corretamente como positiva uma classe que era realmente positiva. Verdadeiro negativo (TN) representa os casos em que o modelo previu corretamente como negativa uma classe que era realmente negativa. Falso positivo (FP) representa os casos em que o modelo previu incorretamente uma classe como positiva quando na verdade ela é negativa. também conhecida como “tipo 1” e falso negativo (FN) representa os casos em que o modelo previu corretamente uma classe como negativa quando na verdade ela é positiva. também conhecida como “tipo 2”.

A imagem 7 retrata como ficou a matriz de confusão do modelo:

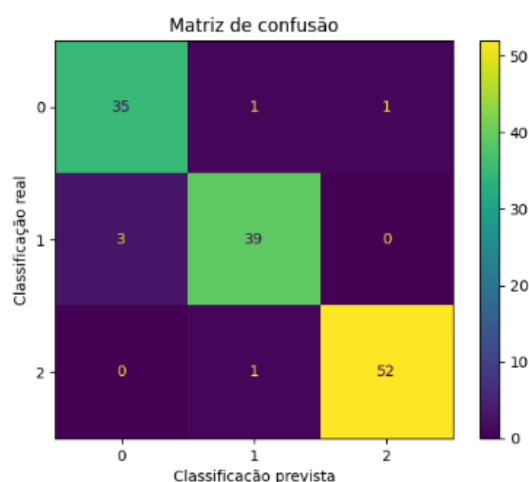


Imagem 7: Matriz de confusão

Autores (2023)

A mesma nos permite analisar de forma rápida o desempenho do sistema. Os valores que compõem a matriz são obtidos fornecendo os segmentos do conjunto de teste ao método de classificação e comparando sua predição com a classe correta de cada segmento.

Diante disso, de acordo com a imagem 6, podemos perceber que a matriz de confusão apresentou bons resultados, visto que os valores presentes na diagonal, que representam os acertos do modelo, são superiores aos demais valores da matriz, que representam as classificações erradas feitas pelo modelo.

V. CONCLUSÃO

Neste trabalho, realizou-se a classificação de distúrbios do sono, identificando em qual categoria cada pessoa se encontrava (Nenhum, Insônia, Apneia do Sono). A coleta de dados foi conduzida através da plataforma Kaggle. Com base nisso, concluiu-se que a profissão de enfermagem apresentava um índice de distúrbio significativamente maior em comparação com outras profissões, ultrapassando os 70%. Em contraste, a função de gerente demonstrou uma taxa consideravelmente mais baixa, em torno de 1%. Com relação aos dados é possível perceber o quanto o estilo de vida pode impactar na saúde das pessoas.

Na pesquisa, utilizou-se uma rede Perceptron composta por duas camadas internas de processamento, juntamente com as funções de ativação sigmoid e softmax. Devido ao desequilíbrio nas categorias após a análise da base de dados, aplicou-se a função SMOTE. Em seguida, utilizou-se o método `train_test_split` para dividir a base de dados em conjuntos, alocando 80% dos elementos para o conjunto de treinamento e reservando os 20% restantes para o conjunto de testes.

Durante o treinamento da base de dados, foram realizadas um total de 200 épocas, resultando em métricas de avaliação que consistentemente superaram os 90%, um valor aceitável para um bom nível de aprendizado. Da mesma forma, a matriz de confusão também apresentou resultados satisfatórios, uma vez que os valores na diagonal, que representam as classificações corretas, foram consistentemente superiores aos demais valores, indicando um baixo nível de desvio nas classificações.

Portanto, pode-se concluir que, a partir de qualquer dado inserido, o modelo conseguiu aprender e adquirir um nível significativo de conhecimento para realizar a classificação de forma precisa.

REFERÊNCIAS

- [1] Reimão, R. Sono: estudo abrangente (2a. ed.). São Paulo: Atheneu, 1996.
- [2] Müller, M. Impacto dos transtornos do sono. Campinas. Estudos de Psicologia, 2007.
- [3] Vgontzas, A. N. & Kales, A. Sleep and its disorders, 1999, *Annu Rev Med* 50,387-400.
- [4] Sociedade Brasileira de Sono. I Consenso Brasileiro de Insônia, 2003, *Hypnos - Journal of Clinical and Experimental Sleep Research* 4 (Supl 2),9/18.
- [5] Ministério da Saúde. Biblioteca Virtual em Saúde, Dicas em Saúde: Distúrbios do sono. Brasília, 2013 [acesso em: 01 jul. 2023], Disponível em: http://bvsms.saude.gov.br/bvs/dicas/256_disturbios_sono.html
- [6] Haykin, S. Neural networks. A comprehensive foundation, 1994.
- [7] Haykin, S. Neural networks and Learning Machines, 2009.
- [8] Harrison, M. Machine Learning: Guia de Referência Rápida, Trabalhando com Dados Estruturados em Python. Novatec, 2019.
- [9] Murphy, K.P. Machine Learning: A Probabilistic Perspective, 2012.