

Hábitos característicos da insônia

Estudo baseado em hábitos comportamentais e sensoriais em humanos

Paulo Henrique Silva Dias
Instituto de Tecnologia
Universidade Federal de São Paulo
São José dos Campos - SP, Brasil
phs.dias@unifesp.br

Magno Luiz Gonçalves Melo
Instituto de Tecnologia
Universidade Federal de São Paulo
São José dos Campos - SP, Brasil
magno.melo@unifesp.br

A insônia, um distúrbio do sono prevalente caracterizado pela dificuldade em adormecer ou manter o sono, impacta significativamente a vida diária e a saúde geral dos indivíduos. Este estudo visa desenvolver um sistema de inteligência artificial (IA) capaz de reconhecer hábitos indicativos de insônia. Utilizando um banco de dados abrangente contendo vários padrões de estilo de vida e comportamentais associados à insônia, o sistema de IA analisará esses hábitos para determinar a probabilidade de um indivíduo sofrer de insônia. Ao identificar esses padrões, a IA pode fornecer insights valiosos para o indicar a possibilidade de tal distúrbio, contribuindo, assim, para uma análise precoce da insônia.

Keywords—insônia, IA, lifestyle, hábitos, distúrbio

I. INTRODUÇÃO E MOTIVAÇÃO

Nos dias de hoje é cada vez mais comum encontrar pessoas sofrendo de insônia, um distúrbio do sono que afeta significativamente a qualidade de vida e o bem-estar geral. A rápida rotina em que vivemos, o aumento do estresse e o uso excessivo de dispositivos eletrônicos são alguns dos fatores que contribuem para o aumento dos casos de insônia.

Estudar e compreender os hábitos e características que levam à insônia é crucial para desenvolver métodos eficazes de predição e intervenção desse problema que vem crescendo a quantidade de pessoas que sofrem deste mal, se tornando um problema de saúde pública, onde no Brasil, por exemplo, estima-se que 72% dos brasileiros sofrem de insônia, de acordo com um estudo da Fundação Oswaldo Cruz (Fiocruz).

Com os avanços da tecnologia, torna-se cada vez mais amplo o uso da tecnologia em diagnósticos e no dia a dia como ferramentas para avaliar e tratar de diversas condições fisiológicas adversas, sendo a insônia um desses distúrbios que busca-se criar maneiras de diagnosticar e tratar. Na Inglaterra, por exemplo, o Instituto Nacional de Excelência em Saúde e Cuidados do Reino Unido (NICE) indica o uso de um aplicativo de inteligência artificial como tratamento para a insônia, mostrando que esse problema global, tem agora soluções estudadas de forma ampla pelo mundo.

Dado esse cenário, o desenvolvimento de uma IA que possa prever se uma pessoa possui indicativo desse distúrbio pode contribuir com um diagnóstico precoce da insônia e alertar o indivíduo sobre esta condição.

II. CONCEITOS FUNDAMENTAIS

A. Inteligência Artificial e Aprendizado de Máquina

Definição de Inteligência Artificial (IA), IA refere-se à capacidade de uma máquina de imitar funções cognitivas humanas, como aprendizagem e resolução de problemas. No contexto deste estudo, a IA é utilizada para analisar padrões de dados e fazer previsões.

Aprendizado de Máquina (Machine Learning), o aprendizado de máquina envolve o uso de algoritmos e modelos estatísticos para permitir que os computadores aprendam com e façam previsões ou decisões baseadas em dados. Técnicas de aprendizado supervisionado e não supervisionado podem ser aplicadas para identificar hábitos associados à insônia.

B. Modelos de Predição e Acurácia

No contexto da IA e do aprendizado de máquina, um modelo de predição é um algoritmo treinado para prever um resultado com base em dados de entrada, existindo diferentes tipos de modelos, como regressão linear, redes neurais e árvores de decisão, que podem ser usados para classificar ou determinar um valor dado os dados inseridos.

A eficácia de um modelo de predição é avaliada utilizando métricas como precisão, o método Recall também conhecido como sensibilidade ou taxa de detecção, é a proporção de verdadeiros positivos em relação ao total de positivos reais, bem como se fez uso das métricas precision, f1-score e support.

C. Definição de insônia

De acordo com Summers MO et al. em Recent Developments in the classification, evaluation and treatment of insomnia. (2006); 130(1):276-8, a insônia é definida como uma dificuldade para iniciar o sono ou para se manter dormindo, quando pode haver uma diminuição total ou parcial da quantidade e/ou da qualidade do sono. Pode ser classificada em inicial, intermediária ou final, e, quanto à duração, em transitória (< 1 mês), de curto tempo (1 – 6 meses) ou crônica (> 6 meses). Também pode ser primária ou secundária a algum fator conhecido.

D. Redes Neurais

As redes neurais, modelos computacionais inspirados pelo funcionamento do cérebro humano, são amplamente empregadas nos campos da inteligência artificial e do aprendizado de máquina. Essas estruturas consistem em uma série de unidades de processamento interconectadas, denominadas neurônios artificiais, que se organizam em camadas. Cada neurônio artificial recebe um conjunto de entradas, processa essas informações e gera uma saída que é transmitida aos neurônios da camada subsequente.

O treinamento dessas redes é realizado por meio de dados, ajustando seus parâmetros internos através de algoritmos de otimização, com o objetivo de minimizar o erro entre as previsões da rede e os resultados esperados. Esse processo permite que as redes neurais aprendam padrões e executem tarefas complexas, como reconhecimento de imagens, processamento de linguagem natural e previsão de séries temporais. A habilidade das redes neurais em modelar relações não lineares e generalizar a partir dos dados as transforma em ferramentas extremamente poderosas para a resolução de uma ampla gama de problemas.

E. Aprendizado Supervisionado

Aprendizagem supervisionada, também conhecida como aprendizado de máquina supervisionado, é uma subcategoria de aprendizado de máquina e inteligência artificial. É definido pelo uso de conjuntos de dados rotulados para treinar algoritmos que classificam dados ou preveem resultados com precisão.

De forma simplificada, à medida que os dados de entrada são inseridos no modelo, ele ajusta seus pesos até que o modelo tenha sido ajustado adequadamente, o que ocorre como parte do processo de validação cruzada. O aprendizado supervisionado ajuda a resolver diversos problemas do mundo real em escala, como classificar o spam em uma pasta separada da sua caixa de entrada. Ele pode ser usado para criar modelos de aprendizado de máquina altamente precisos.

F. Keras

Keras é uma biblioteca de código aberto em Python para criar e treinar modelos de aprendizado profundo de forma simples e intuitiva.

Ela fornece uma interface simples para construir redes neurais, facilitando a experimentação com diferentes camadas, otimizadores e funções de perda. Keras é altamente modular e se integra ao TensorFlow, permitindo a execução eficiente de modelos em CPUs, GPUs e TPUs. Popular entre pesquisadores e desenvolvedores, Keras é usado para prototipagem rápida e desenvolvimento de soluções de inteligência artificial.

G. Random Forest

Florestas aleatórias ou florestas de decisão aleatórias é um método de aprendizado de conjunto para classificação, regressão e outras tarefas que opera construindo uma infinidade de árvores de decisão no momento do treinamento. Para tarefas de classificação, a

saída da floresta aleatória é a classe selecionada pela maioria das árvores. Para tarefas de regressão, a média ou previsão média das árvores individuais é retornada. Florestas de decisão aleatórias corrigem o hábito das árvores de decisão de overfitting ao seu conjunto de treinamento.

H. Gradient Boosting

O algoritmo Gradient Boosting é uma técnica Boosting, incluída dentro do grupo de classificadores Ensemble. Esses classificadores são métodos que utilizam uma combinação de resultados de preditores fracos, com o objetivo de produzir um melhor modelo preditivo. Os preditores fracos são modelos que, quando utilizados individualmente possuem uma acurácia abaixo do esperado.

III. TRABALHOS RELACIONADOS

Os estudos sobre a insônia são abrangentes dentro da área da medicina, onde busca-se entender melhor os agentes que causam a insônia, quanto os efeitos que a insônia causa no indivíduo. Nesse projeto especificamente, visamos entender características causadoras da insônia que nos ajudem a identificar se um indivíduo possivelmente sofre de tal distúrbio.

Como estudos semelhantes podemos citar o trabalho de Alexander A. Huang e Samuel Y. Huang com o projeto Use of machine learning to identify risk factors for insomnia, onde foi avaliado o modelo de aprendizado de máquina XGBoost famoso na literatura por sua utilização para previsões no âmbito dos autocuidados.

Outro projeto com o mesmo intuito é o de pesquisadores da Daffodil International University, de Bangladesh, onde eles compararam diversos tipos de algoritmos de inteligência artificial na busca de melhores resultados na previsão de insônia crônica em pacientes.

Por fim, vale citar o trabalho Automated classification of cyclic alternating pattern sleep phases in healthy and sleep-disordered subjects using convolutional neural network onde usando de redes neurais convolucionais eles criaram um algoritmo para identificar diferentes tipos de problemas relacionados ao sono usando da inteligência artificial para identificar e reconhecer os padrões através do sono do indivíduo em questão.

IV. OBJETIVO

O objetivo deste projeto, visa, desenvolver um algoritmo de inteligência artificial com o intuito de identificar se um indivíduo sofre ou não de insônia através de características físicas do mesmo, visando um diagnóstico preciso e rápido e que venha a superar a acurácia de outros algoritmos já existentes e com o mesmo objetivo, isto é, deduzir se uma pessoa sofre do distúrbio de insônia.

V. METODOLOGIA EXPERIMENTAL

Neste trabalho, buscamos em um base de dados públicas informações sobre características físicas de indivíduos com e sem insônia, onde realizaremos tarefas variadas de mineração de dados com objetivo de criar uma base apropriada para uso neste projeto, o refinamento da base irá contribuir para a proficiência do modelo e sua interpretação. E de modo a tangibilizar os acertos um aprendizado supervisionado será utilizado.

Com tais dados, passaremos por todos os passos relacionados a limpeza e tratamento dos dados para obtermos um cenário único, porém com a qualidade necessária para que possamos, usando algoritmos de inteligência artificial, mais especificamente do tipo de redes neurais convolutivas para criarmos um modelo capaz de prever se uma pessoa sofre ou não com insônia, com apoio do Keras. Métodos de validação da precisão dos resultados serão aplicados.

A. Domínio da Aplicação

A base de dados consiste em um csv de aproximadamente 400 rows e leva em consideração os seguintes atributos/características: Gênero, idade, profissão, duração do sono, qualidade do sono, nível de atividade física, nível de stress, BMI Category/IMC(classificação: acima do peso, normal, obeso), pressão sanguínea, batimentos cardíacos, passos diários e por fim o apontamento se pessoa possui insônia(classificação: Apneia do sono, insônia, sem distúrbio de sono).

Abaixo temos alguns gráficos que podem ilustrar e facilitar o contexto quantitativo de algumas variáveis através de relações diversas.

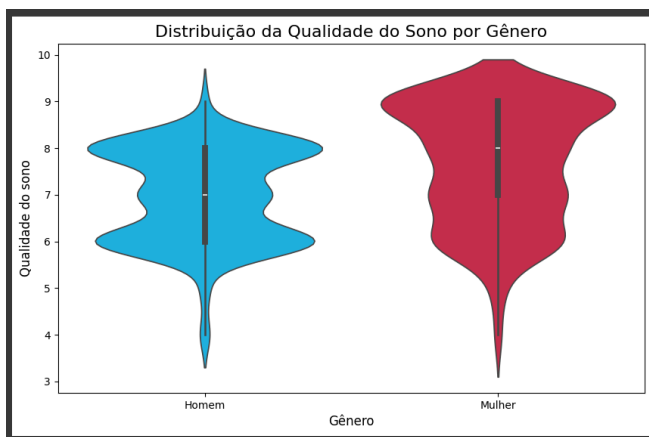


Fig 1. Relação entre gêneros e qualidade do sono

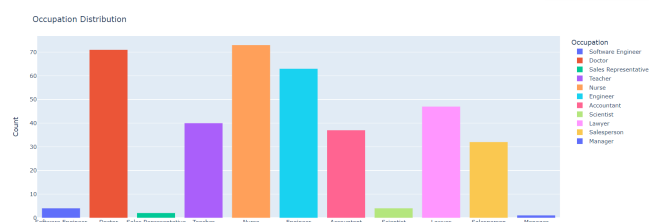


Fig 2. Proporção de profissões na base de dados

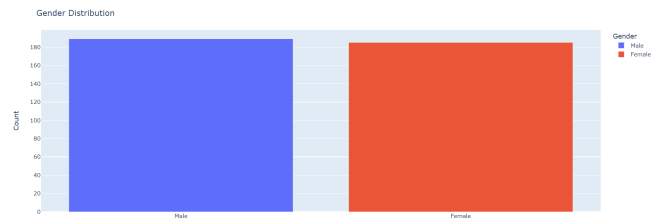


Fig 3. Proporção de homens e mulheres na base de dados

B. Pré-processamento

Primeiramente usamos de uma abordagem gráfica para visualização do comportamento de alguns dos atributos da nossa base de dados, a fim de uma melhor compreensão da mesma. A partir disso, traçamos uma estratégia para o tratamento dos dados. Como todos os modelos usados são baseados em pesos, partimos para 2 tipos de tratamentos com técnicas diferentes a partir da base inicial.

As principais atividades dessa etapa foram a limpeza dos dados, transformações e renomeações para a normalização dos dados.

Na primeira abordagem utilizamos transformações categóricas diretamente em números inteiros, para a segunda optamos por uma abordagem com técnicas como o hot-encoding onde ele transforma categorias em elementos vetoriais para cada categoria de um atributo. Dessa forma evitamos que uma errônea hierarquia seja interpretada pelo algoritmo, com isso evitando atribuir relevância indevida às diferentes profissões (campo *Occupation*), em gênero (campo *Gender*), no IMC (campo *BMI Category*) e por fim no campo de classificação Sleep Disorder.

A pressão sanguínea, representada pelo campo *Blood Pressure*, de forma simples, foi transformado dividindo em pressão sistólica(*BloodPressure_Upper*) e pressão diastólica(*BloodPressure_Lower*), e convertido para inteiro.

Todos os campos numéricos foram convertidos para o formato int para garantir uniformidade. E o campo *PersonID* foi removido por ser um índice que não influencia na análise.

C. Reconhecimento de Padrões

Para este projeto, foi realizado uma tarefa de classificação, onde o objetivo é categorizar o conjunto de dados de entrada dentro das classes disponíveis, onde para este experimento, consiste em definir se o indivíduo possui insônia, apneia do sono ou não possui problemas do sono, sendo essas as classes onde iremos categorizar as entradas. Não somente isso, para este trabalho, iremos fazer uma comparação com algoritmos já utilizados em outros projetos para este mesmo cenário, assim teremos a comparação entre algoritmos como random forest, gradient boosting (onde esse é notavelmente já conhecido dentro da área da saúde pelos excelentes resultados), logistic regression e um algoritmo de neural network.

Para realizar tais experimentos, utilizaremos a linguagem python, munida de seus frameworks como pandas, aplicação da biblioteca matplotlib para visibilidade gráfica e seaborn de forma complementar, bibliotecas do scikit-learn tanto para criação dos modelos quanto para validação das métricas no passo de pós-processamento, sendo o principal framework deste projeto, pois usaremos o mesmo tanto para o protocolo de validação, onde inicialmente usaremos modelos hold-out com separação de 80-20 para as bases de treino e teste e seguiremos com medidas de avaliação para classificação como as contidas no scikit-learn para classificadores, sendo elas a F1-score, recall, precision e o suporte usado, além claro da pontuação presente nos modelos do scikit learn para cada um dos algoritmos citados.

O uso da estratégia de validação cruzada K-Folds junto a modelos de rede neurais utilizando a Keras API foi utilizado em determinado momento para diversificar ainda mais os tipos de modelos aplicáveis na análise de classificação e aumentar a qualidade dos resultados encontrados.

D. Pós-processamento

Utilizando os classificadores mencionados anteriormente, realizamos uma comparação dos algoritmos, analisando seus resultados finais em relação aos demais. Avaliamos o desempenho de cada algoritmo, determinando qual deles foi superior em cada método empregado. Também examinamos os scores de cada modelo individualmente. No caso do uso dos K-Folds, incluímos os cálculos de média, bem como os desvios de acurácia e regressão linear.

Por fim, utilizamos gráficos de barra para ilustrar visualmente o desempenho dos algoritmos. Com base nessas análises, obtivemos um veredito claro sobre o desempenho dos algoritmos e avaliamos sua qualidade.

Em seguida temos os gráficos com os resultados obtidos em cada estratégia, que será discutida nos resultados.

1. Validação usando Hold-out

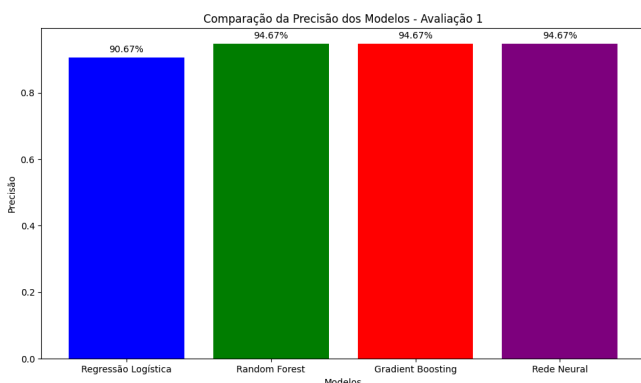


Fig 4. Gráfico de barras sobre o resultado de score de cada modelo na primeira avaliação

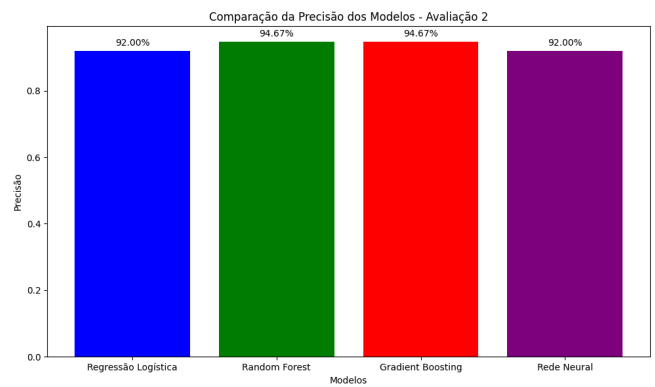


Fig 5. Gráfico de barras sobre o resultado de score de cada modelo na segunda avaliação

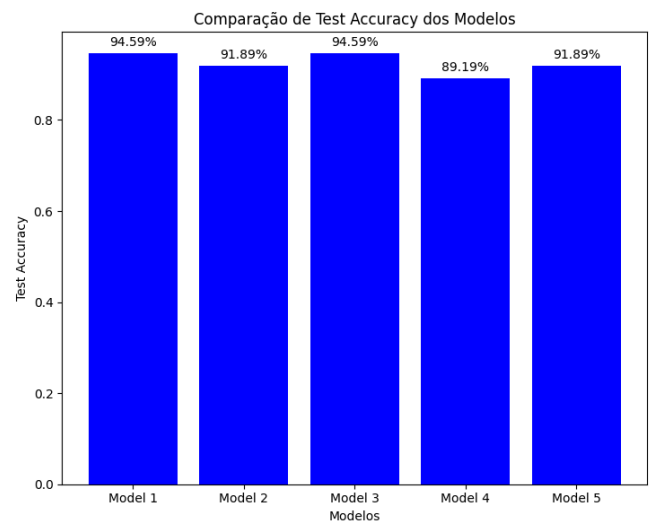


Fig 6. Gráfico de barras com a precisão de cinco modelos com parâmetros distintos aplicados usando Keras sem k-folds, na primeira avaliação.

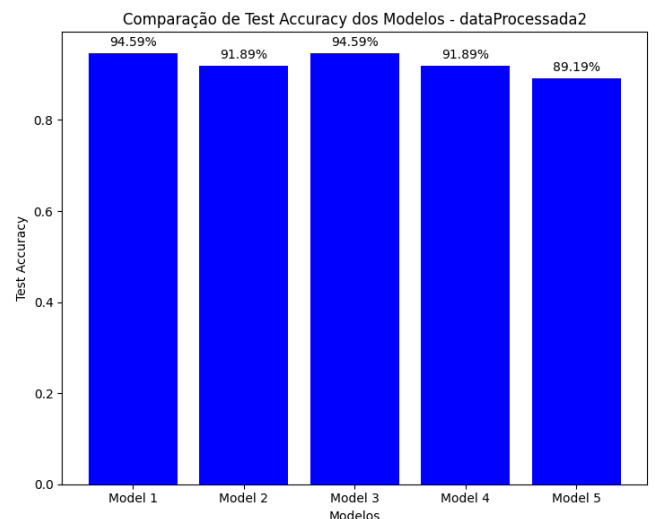


Fig 7. Gráfico de barras com a precisão de cinco modelos com parâmetros distintos aplicados usando Keras sem k-folds, na segunda avaliação

2. Validação usando K-Folds

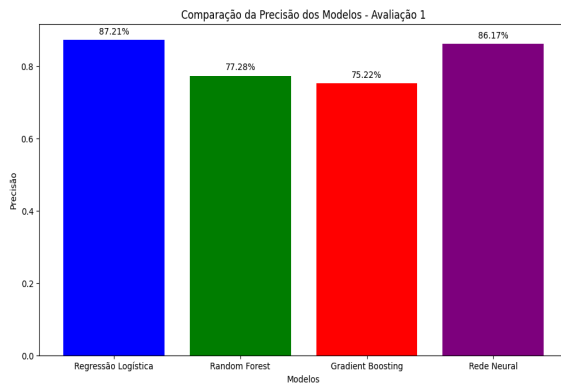


Fig 8. Gráfico de barras da média de precisão obtida para cada um dos modelos usando K-folds validation para o segundo tratamento dos dados

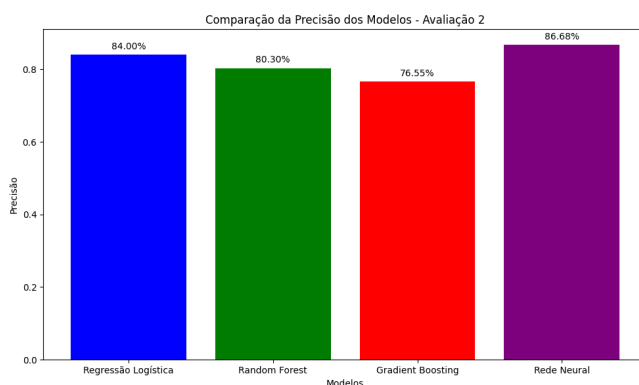


Fig 9. Gráfico de barras da média de precisão obtida para cada um dos modelos usando K-folds validation para o segundo tratamento dos dados

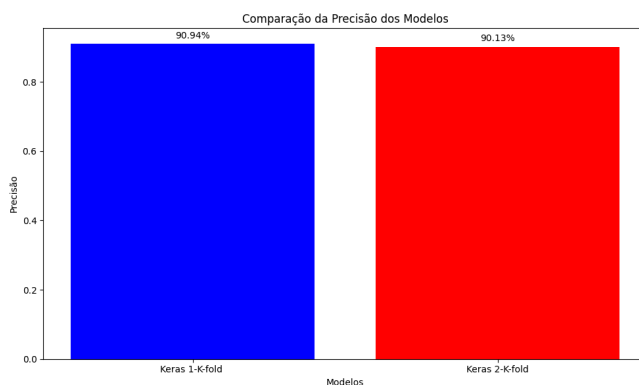


Fig 10. Gráfico de barras da média de precisão obtida para o Keras em cada um dos tratamentos de dados testado.

VI. RESULTADOS

Nossos testes foram feitos através de duas formas de validação distintas: Hold-Out e através de K-Fold validation. Dentro destes 2 tipos de validação encontram-se também duas bases diferentes de treino e teste, onde para cada uma foram aplicadas diferentes formas de tratamento dos dados.

Começando então a avaliação através dos testes em Hold-Out. Para a primeira base temos que a maior precisão foi de 0.9466666666666667 obtida por três modelos, os modelos comparados eram o modelo XGBoost, Regressão Logística, floresta aleatória e a rede neural, sendo que as duas últimas possuem os mesmos resultados dentro da base avaliada.

Resultados das precisões individuais foram Regressão Logística 0.9066666666666666, Random Forest 0.9466666666666667, Gradient Boosting 0.9466666666666667, Rede Neural 0.9466666666666667.

Na segunda base temos os mesmos modelos avaliados, e a maior precisão atingida foi de 0.9466666666666667, onde o modelo XGBoost e floresta aleatória recebem as mesmas notas, porém com a rede atingindo um valor menor de sólidos 0.92. Ressaltando que para ambos os casos usa-se os mesmos parâmetros nos 2 testes para os modelos e uma base de dados devidamente tratada.

Resultados das precisões individuais foram Regressão Logística 0.92, Random Forest 0.9466666666666667, Gradient Boosting 0.9466666666666667, Rede Neural 0.92.

Para a validação via K-Fold usando a métrica da acurácia média, o modelo feito a partir de 3 camadas fully connected usando a api do Keras, teve resultados superiores, com uma precisão de 0.9093883335590363 para o primeiro tratamento de dados e com mais de 90% de precisão também para o segundo tratamento, sendo seguido pelo modelo de rede, para ambos os tratamentos usados na bases.

Médias dos modelos usando K-folds e a mesma base aplicada ao primeiro tratamento de dados, foram Regressão Logística 0.8721194879089618, Random Forest 0.7728307254623045, Gradient Boosting 0.7522048364153627, Rede Neural 0.8616642958748223. Médias dos modelos usando K-folds e a mesma base aplicada no segundo tratamento da base, foram Regressão Logística 0.8400426742532006, Random Forest 0.8029871977240399, Gradient Boosting 0.7655049786628734, Rede Neural 0.8667852062588907.

Assim, entendemos que para cenários da vida real onde a distribuição dos dados seja a mesma distribuição que encontramos na divisão da base para os modelos validados via Hold-Out, o melhor modelo a ser utilizado é o XGBoost e a floresta aleatória, onde não foi possível superar seu resultado. Porém na vida real, como não conhecemos a real distribuição dos dados, modelos menos especializados tendem a apresentarem um melhor desempenho pois seu resultado é menos especialista na base de dados e abrange dados cuja distribuição não seja equivalente a distribuição obtida no treino. Assim, consideramos que obtivemos um resultado de maior valor pela maior flexibilidade entre as entradas aceitas ao utilizarmos o Keras e uma boa média de taxa de acerto em geral.

Portanto temos que o modelo Keras de rede neural tem sim uma determinada vantagem sobre os demais

modelos e obteve a mesma média de acurácia, sendo o vencedor da competição.

VII. CONCLUSÃO

Entregamos ao final do projeto um modelo de aprendizado de máquina, capaz de identificar um indivíduo com insônia através de determinadas características físicas, com alta acurácia, superando a acurácia de outros projetos relacionados e apresentamos os números obtidos em teste para comprovar tais resultados de acurácia.

VIII. REFERENCIAS

"Use of machine learning to identify risk factors for insomnia". Home Page. Acesso.
2024-07-11. Disponível:
<https://doi.org/10.1371/journal.pone.0282622>

Você já teve insônia? Saiba que 72% dos brasileiros sofrem com alterações no sono. Disponível em:
<<https://www.gov.br/saude/pt-br/assuntos/noticias/2023/marco/voce-ja-teve-insomnia-saiba-que-72-dos-brasileiros-sofrem-com-alteracoes-no-sono>>.

DIGITAL, O.; BARROS, M. Médicos do Reino Unido indicam aplicativo de IA para tratar insônia. Disponível em:
<<https://olhardigital.com.br/2022/05/26/medicina-e-saude/medicos-do-reino-unido-indicam-aplicativo-de-ia-para-tratar-insomnia/#:~:text=A%20recomenda%C3%A7%C3%A3o%20significa%20que%20o>>. Acesso em: 11 jul. 2024.

M. M. Islam, A. K. M. Masum, S. Abujar and S. A. Hossain, "Prediction of chronic Insomnia using Machine Learning Techniques," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225570.
keywords: {Classification algorithms;Machine learning algorithms;Prediction algorithms;Data models;Sensitivity;Depression;Logistics;Insomnia;Machine-learning;Sleep-disorder;Logistic regression},

Shruti Murarka, Aditya Wadichar, Ankit Bhurane, Manish Sharma, U. Rajendra Acharya, Automated classification of cyclic alternating pattern sleep phases in healthy and sleep-disordered subjects using convolutional neural network, *Computers in Biology and Medicine*, Volume 146, 2022, 105594, ISSN 0010-4825, <https://doi.org/10.1016/j.combiomed.2022.105594>.

(<https://www.sciencedirect.com/science/article/pii/S0010482522003869>)

O que é aprendizado supervisionado? | IBM.
Disponível em:
<<https://www.ibm.com/br-pt/topics/supervised-learning>>. Acesso em: 11 jul. 2024.

SILVA, J. Uma breve introdução ao algoritmo de Machine Learning Gradient Boosting utilizando a biblioteca.... Disponível em:
<<https://medium.com/equals-lab/uma-breve-introdu%C3%A7%C3%A3o-ao-algoritmo-de-machine-learning-gradient-boosting-utilizando-a-biblioteca-311285783099>>.