

# Hábitos característicos da insônia

Integrantes:

Magno Luiz Gonçalves Melo  
Paulo Henrique S. Dias

RA:133688  
RA:156648



# Sumário

**01**

**Introdução e  
motivação**

**02**

**Conceitos  
Fundamentais**

**03**

**Trabalhos  
relacionados**

**04**

**Objetivo**

**05**

**Metodología  
experimental**

**06**

**Resultados e  
Conclusão**



01

# Introdução e motivação

# Definição de insônia



## Definição

A insônia, é caracterizado como um distúrbio do sono prevalente caracterizado pela dificuldade em adormecer ou manter o sono, que impacta significativamente a vida diária e a saúde geral dos indivíduos.



## Motivação

No Brasil estima-se que 72% dos brasileiros sofrem de insônia, de acordo com um estudo da Fundação Oswaldo Cruz (Fiocruz).

Por ser um tema comum e recorrente surgiu-se a ideia de desenvolver uma IA capaz de predizer se uma pessoa possui insônia ou não

zZ

zZ



zZ



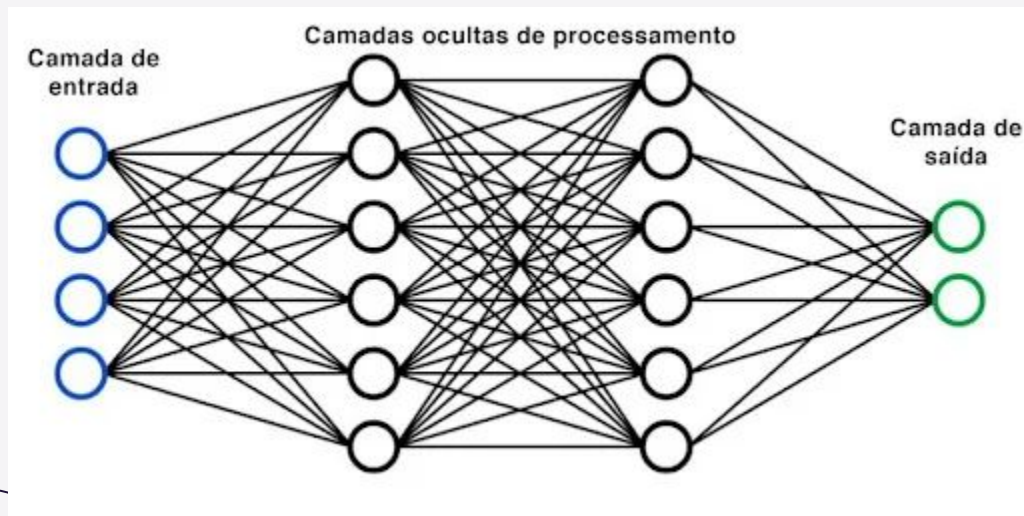


**02**

# **Conceitos Fundamentais**

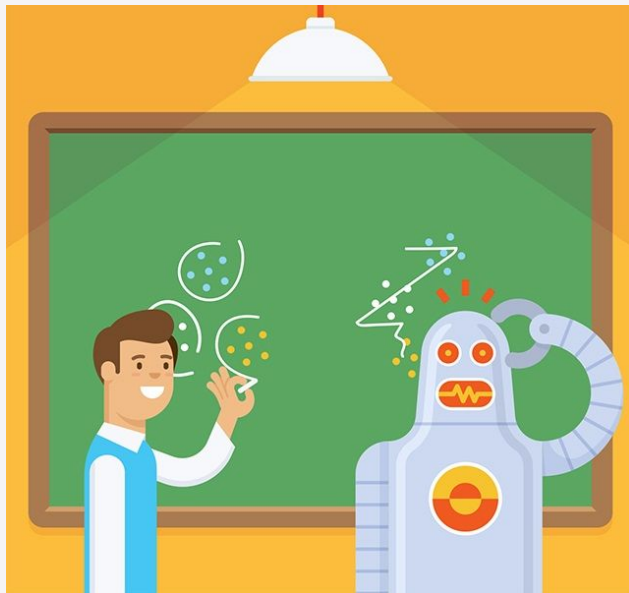
# 1. Redes Neurais

Redes Neurais são modelos computacionais inspirados no cérebro humano, compostos por neurônios artificiais interconectados, que aprendem padrões e relações a partir de dados para realizar tarefas como classificação, regressão e reconhecimento de padrões.



## 2. Aprendizado Supervisionado

É um tipo de aprendizado de máquina onde um modelo é treinado usando um conjunto de dados rotulado, aprendendo a mapear entradas para saídas corretas com base em exemplos fornecidos, para depois realizar previsões ou classificações em novos dados.





**03**

## **Trabalhos relacionados**



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.

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.





04

## Objetivo

# O que queremos fazer?

O objetivo deste projeto, visa, desenvolver um algoritmo de inteligência artificial com o intuito de identificar se um indivíduo sofre ou não de insônia (classificador), 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.



**05**

# **Metodología experimental**

# Sobre a base de dados

## 1- Domínio da Aplicação

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
2	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None
3	2	Male	28	Doctor	6.2	6	60	60	Normal	125/80	75	10000	None
4	3	Male	28	Doctor	6.2	6	60	60	Normal	125/80	75	10000	None
5	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
6	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
7	6	Male	28	Software Engineer	5.9	4	30	8	Obese	140/90	85	3000	Insomnia
8	7	Male	29	Teacher	6.3	6	40	7	Obese	140/90	82	3500	Insomnia
9	8	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
10	9	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
11	10	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
12	11	Male	29	Doctor	6.1	6	30	8	Normal	120/80	70	8000	None
13	12	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
14	13	Male	29	Doctor	6.1	6	30	8	Normal	120/80	70	8000	None
15	14	Male	29	Doctor		6	30	8	Normal	120/80	70	8000	None
16	15	Male	29	Doctor		6	30	8	Normal	120/80	70	8000	None
17	16	Male	29	Doctor		6	30	8	Normal	120/80	70	8000	None
18	17	Female	29	Nurse	6.5	5	40	7	Normal	132/87	80	4000	Sleep Apnea
19	18	Male	29	Doctor		6	30	8	Normal	120/80	70	8000	Sleep Apnea
20	19	Female	29	Nurse	6.5	5	40	7	Normal	132/87	80	4000	Insomnia
21	20	Male	30	Doctor	7.6	7	75	6	Normal	120/80	70	8000	None
22	21	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None
23	22	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None
24	23	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None
25	24	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None
26	25	Male	30	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
27	26	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None
28	27	Male	30	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
29	28	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None
30	29	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None
31	30	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None
32	31	Female	30	Nurse	6.4	5	35	7	Normal	130/86	78	4100	Sleep Apnea
33	32	Female	30	Nurse	6.4	5	35	7	Normal	130/86	78	4100	Insomnia
34	33	Female	31	Nurse	7.9	8	75	4	Normal	117/76	69	6800	None
35	34	Male	31	Doctor	6.1	6	30	8	Normal	125/80	72	5000	None

Rows: ~400

Class.:3

Total de  
atributos: 13

fonte:Kaggle

# Limpeza e transformação dos dados

## 2- Pré-processamento

A partir da base de dados inicial, criamos duas bases com tratamentos distintos.

### Tratamento 1:

- Existência de campos null devem ser preenchidos
- Despadronização dos valores da coluna BMI category(IMC)
- Transformação do campo Blood Pressure
- Campos numéricos com diferentes tipos
- Remoção do campo id
- Padronização dos campos com valores categóricos (Profissão,IMC,Gênero)
- Neste tratamento, transformamos valores categóricos usando uma tabela de conversão

# Limpeza e transformação dos dados

## 2- Pré-processamento

A partir da base de dados inicial, criamos duas bases com tratamentos distintos.

### Tratamento 2:

- Existência de campos null devem ser preenchidos
- Despadronização dos valores da coluna BMI category(IMC)
- Transformação do campo Blood Pressure
- Campos numéricos com diferentes tipos
- Remoção do campo id
- Padronização dos campos com valores categóricos (Profissão,IMC,Gênero)
- Hot-encoding para campos categóricos

## Hot encoding

É uma técnica comum para transformar atributos categóricos em uma representação numérica. Ele cria colunas binárias (0 ou 1/true or false) para cada categoria de um atributo. Dessa forma evitamos que uma errônea hierarquia seja interpretada pelo algoritmo.

...	Occupation_Lawyer	Occupation_Manager	Occupation_Nurse	Occupation_Sales Representative	Occupation_Salesperson	Occupation_Scientist
...	0.0	0.0	0.0	0.0	0.0	0.0
...	0.0	0.0	0.0	0.0	0.0	0.0
...	0.0	0.0	0.0	0.0	0.0	0.0
...	0.0	0.0	0.0	1.0	0.0	0.0
...	0.0	0.0	0.0	1.0	0.0	0.0



# Reconhecimento de Padrões

## 3- Reconhecimento de Padrões

Neste projeto, vamos realizar uma tarefa de classificação para determinar se um indivíduo possui insônia, apneia do sono, ou não apresenta problemas de sono. Para isso, compararemos diferentes algoritmos utilizados em cenários similares, incluindo Random Forest, Gradient Boosting (reconhecido na área da saúde por seus excelentes resultados), Regressão Logística e duas Redes Neurais.

Usamos Python com frameworks como Pandas para manipulação de dados, Matplotlib e Seaborn para visualização, e Scikit-Learn para criação e validação dos modelos. O Scikit-Learn será o principal framework, aplicado tanto no protocolo de validação quanto na avaliação de métricas. Iniciaremos com o método hold-out, separando 80% dos dados para treino e 20% para teste, e usaremos métricas como F1-score, recall, precision e suporte para avaliar o desempenho dos classificadores.

# Sobre a base de dados

## 4- Pós-processamento

Para a validação dos modelos, como utilizamos dois tipos de tratamento inicial distintos, nossa etapa de pós-processamento contou com 4 validações diferentes, sendo 2 para cada tipo de validação cruzada que escolhemos.

### 1) Hold-out

Dividimos as 2 bases tratadas para o teste utilizando a validação Hold-out dividindo a base e o teste em proporção de 80-20 para cada um dos testes dos 5 algoritmos

### 2) K-Folds

Dividimos as 2 bases tratadas em 10-folds para cada um dos testes dos 5 algoritmos

# Sobre a base de dados

## 4- Pós-processamento

Validação	Base	Modelo	Precisão/Acurácia
Hold-Out	Primeira Base	Regressão Logística	0.9067
		Random Forest	0.9467
		Gradient Boosting	0.9467
		Rede Neural	0.9467
	Segunda Base	Regressão Logística	0.9200
		Random Forest	0.9467
		Gradient Boosting	0.9467
		Rede Neural	0.9200

<b>K-Fold</b>	Primeira Base	Regressão Logística	0.8721
		Random Forest	0.7728
		Gradient Boosting	0.7522
		Rede Neural	0.8617
	Segunda Base	Regressão Logística	0.8400
		Random Forest	0.8030
		Gradient Boosting	0.7655
		Rede Neural	0.8668
<b>K-Fold (Keras)</b>	Primeira Base	Rede Neural (Keras)	0.9094
	Segunda Base	Rede Neural (Keras)	>0.90

06

## Conclusão



# Conclusão

Para cenários reais onde a distribuição dos dados é semelhante à divisão utilizada na validação Hold-Out, os melhores modelos são o XGBoost e a Floresta Aleatória, que apresentaram resultados imbatíveis.

No entanto, como a distribuição real dos dados é desconhecida, modelos menos especializados, como a rede neural Keras, mostram melhor desempenho geral devido à sua maior flexibilidade e capacidade de generalização. Assim, o modelo Keras teve uma vantagem significativa, alcançando uma média de acurácia comparável aos melhores modelos, sendo considerado o vencedor da competição.

# Referências

<https://www.gov.br/saude/pt-br/assuntos/noticias/2023/marco/voce-ja-teve-insonia-saiba-que-72-dos-brasileiros-sofrem-com-alteracoes-no-sono>  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10096447/>  
[https://link.springer.com/chapter/10.1007/978-3-030-81462-5\\_31](https://link.springer.com/chapter/10.1007/978-3-030-81462-5_31)

# Obrigad o pela atenção!

