



Sleep Research

PROJETO EM CIÊNCIA DE DADOS

ARTHUR HENRIQUE CHAVES OLIVEIRA

JÚLIO LAGE ALEXANDRE

RHUDSON DOUGLAS MOTA SAMPAIO

SARA ALVES MARTINS

SOPHIA NEVES ALVIM OTTONI



PUC MINAS
PRAÇA DA
LIBERDADE

Tópicos

INTRODUÇÃO AO PROBLEMA

OBJETIVOS DO PROJETO

PREPARAÇÃO DOS DADOS

DESCRIÇÃO DOS DADOS

MODELAGEM: MODELOS 1 E 2

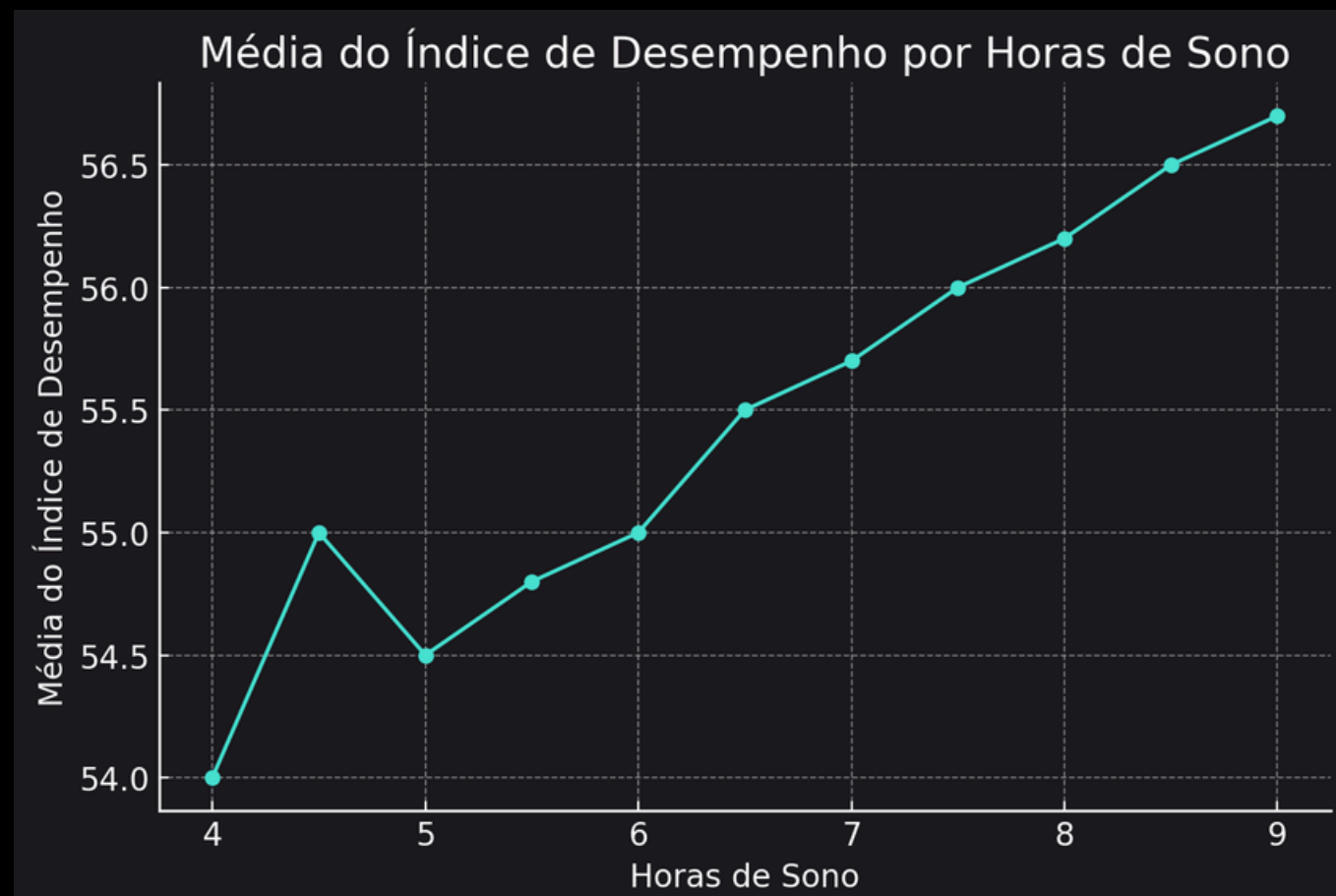
AMEAÇAS A VALIDADE

RESULTADOS

CONCLUSÕES

Introdução ao problema

[Voltar para o slide de temas](#) 

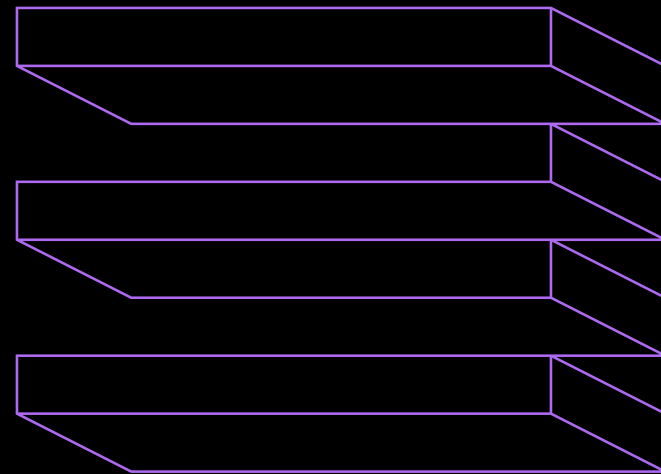
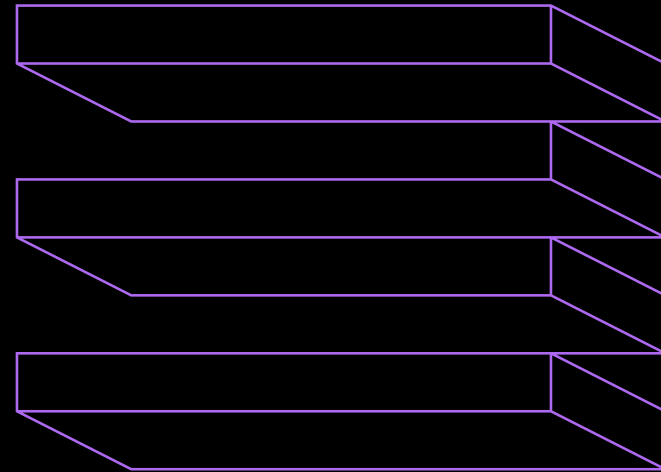


- **Desafios Globais:** Saúde e bem-estar exigem colaboração global, apesar dos avanços científicos e tecnológicos.
- **Qualidade do Sono:** Relação forte com a saúde humana, influenciando o desempenho físico e mental.
- **Objetivos da ONU:** Em 2015, destacaram a saúde e bem-estar como um dos 17 objetivos globais.
- **Evidências:** Estudos mostram impacto significativo do sono na saúde, apesar de não haver meta específica sobre sono.
- **Estatísticas:** Maus hábitos de sono afetam negativamente o desempenho em atividades físicas e mentais.

Objetivos



- Investigar a correlação entre a qualidade do sono e o desempenho acadêmico de estudantes universitários.
- Examinar como noites mal dormidas influenciam negativamente a produtividade em sala de aula.
- Analisar a necessidade de promover uma boa higiene do sono entre os estudantes.



Preparação dos dados

- Seleção dos atributos dos datasets
 - Tratamento dos valores faltantes ou omissos
 - Tratamento dos valores inconsistentes
 - Conversão de dados
-

Descrição dos dados

[Voltar para o slide de temas](#) 

Características dos dados	
Origem dos dados	Datasets: Student Stress Factors Student Performance
Tamanho da amostra	Student Stress Factors: 520 registros Student Performance: 10.000 registros
Tipos de dados coletados	Números inteiros, booleanos e reais.

Modelo 1: KNN (K-Nearest Neighbors)

DESCRIÇÃO:

- UTILIZA A PROXIMIDADE ENTRE PONTOS DE DADOS PARA CLASSIFICAÇÃO.
- PARÂMETRO PRINCIPAL: NÚMERO DE VIZINHOS (K).

CONFIGURAÇÕES E PARÂMETROS:

- INICIAL: $K = 1$
- AJUSTE: $K = 3$
- MÉTRICA DE DISTÂNCIA: EUCLIDIANA

PROCESSO DE AJUSTE:

- VALIDAÇÃO CRUZADA: -5 FOLD
- CRITÉRIO DE ESCOLHA: MENOR ERRO DE CLASSIFICAÇÃO

Antes do ajuste de vizinhos:

Tabela Student Performance:

Acurácia: 98%

	precision	recall	f1-score	support
1	1.00	1.00	1.00	10
2	1.00	1.00	1.00	13
3	0.95	0.98	0.97	43
4	0.98	1.00	0.99	45
5	1.00	0.89	0.94	19
accuracy			0.98	130
macro avg	0.99	0.97	0.98	130
weighted avg	0.98	0.98	0.98	130

Tabela Student Stress Factors:

Acurácia: 84%

	precision	recall	f1-score	support
1	0.72	0.71	0.71	51
2	0.84	0.84	0.84	463
3	0.83	0.84	0.84	686
4	0.85	0.85	0.85	583
5	0.85	0.82	0.84	217
accuracy			0.84	2000
macro avg	0.82	0.81	0.81	2000
weighted avg	0.84	0.84	0.84	2000

Depois do ajuste de vizinhos:

Tabela Student Performance:

Acurácia: 98%

	precision	recall	f1-score	support
1	1.00	1.00	1.00	10
2	1.00	1.00	1.00	13
3	0.95	0.98	0.97	43
4	0.98	1.00	0.99	45
5	1.00	0.89	0.94	19
accuracy			0.98	130
macro avg	0.99	0.97	0.98	130
weighted avg	0.98	0.98	0.98	130

Tabela Student Stress Factors:

Acurácia: 89%

	precision	recall	f1-score	support
1	1.00	0.12	0.21	51
2	0.86	0.91	0.88	463
3	0.89	0.93	0.91	686
4	0.88	0.93	0.90	583
5	0.96	0.76	0.85	217
accuracy			0.89	2000
macro avg	0.92	0.73	0.75	2000
weighted avg	0.89	0.89	0.88	2000

Modelo 2: Árvore de decisão

DESCRIÇÃO:

- DIVIDE OS DADOS EM RAMOS BASEADOS EM CONDIÇÕES.
- UTILIZA O CRITÉRIO DE GINI OU ENTROPIA PARA DIVISÃO.

CONFIGURAÇÕES E PARÂMETROS:

- CRITÉRIO DE DIVISÃO: GINI
- PROFUNDIDADE MÁXIMA: 5
- TAMANHO MÍNIMO DO NÓ: 10

PROCESSO DE AJUSTE:

- VALIDAÇÃO CRUZADA: 10-FOLD
- CRITÉRIO DE ESCOLHA: MENOR ERRO DE CLASSIFICAÇÃO

Resultados do modelo 2

Tabela Student Performance:

Acurácia: 98%

```
print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
1	1.00	1.00	1.00	10
2	1.00	1.00	1.00	13
3	0.95	0.98	0.97	43
4	0.98	1.00	0.99	45
5	1.00	0.89	0.94	19
accuracy			0.98	130
macro avg	0.99	0.97	0.98	130
weighted avg	0.98	0.98	0.98	130

Tabela Student Stress Factors:

Acurácia: 98%

```
y_pred = clf.predict(x_test)
print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
1	1.00	1.00	1.00	10
2	1.00	1.00	1.00	13
3	0.95	0.98	0.97	43
4	0.98	1.00	0.99	45
5	1.00	0.89	0.94	19
accuracy			0.98	130
macro avg	0.99	0.97	0.98	130
weighted avg	0.98	0.98	0.98	130

Ameaças à validade

[Voltar para o slide de temas](#) 

Interna:

- Viés na escolha do número de vizinhos (n_neighbors) no KNN
- Importância da validação criteriosa para n_neighbors
- Mitigação de viés com cross-validation
- Análise da predominância de um atributo dominante
- Importância da escolha do critério de divisão
- Impacto da métrica de distância na performance
- Impacto do critério de divisão ['gini' vs 'entropy'] na árvore
- Sensibilidade da feature_importances_ à seleção de características

Externa:

- Limitação na representatividade dos dados de treinamento
- Generalização inadequada para diferentes sistemas educacionais e culturais
- Necessidade de validação cruzada extensa e cuidadosa
- Impacto negativo de dados sintéticos e subjetivos
- Influência da escolha e disponibilidade das features na validade dos resultados

Resultados

IMPACTO DA QUALIDADE DO SONO NO DESEMPENHO ACADÊMICO:

CARACTERÍSTICAS-CHAVE IDENTIFICADAS:

- FATORES DE ESTRESSE SÃO PREDITORES IMPORTANTES DA QUALIDADE DO SONO.
- ÁRVORE DE DECISÃO IDENTIFICOU ESSAS CARACTERÍSTICAS, DIRECIONANDO INTERVENÇÕES ESPECÍFICAS.

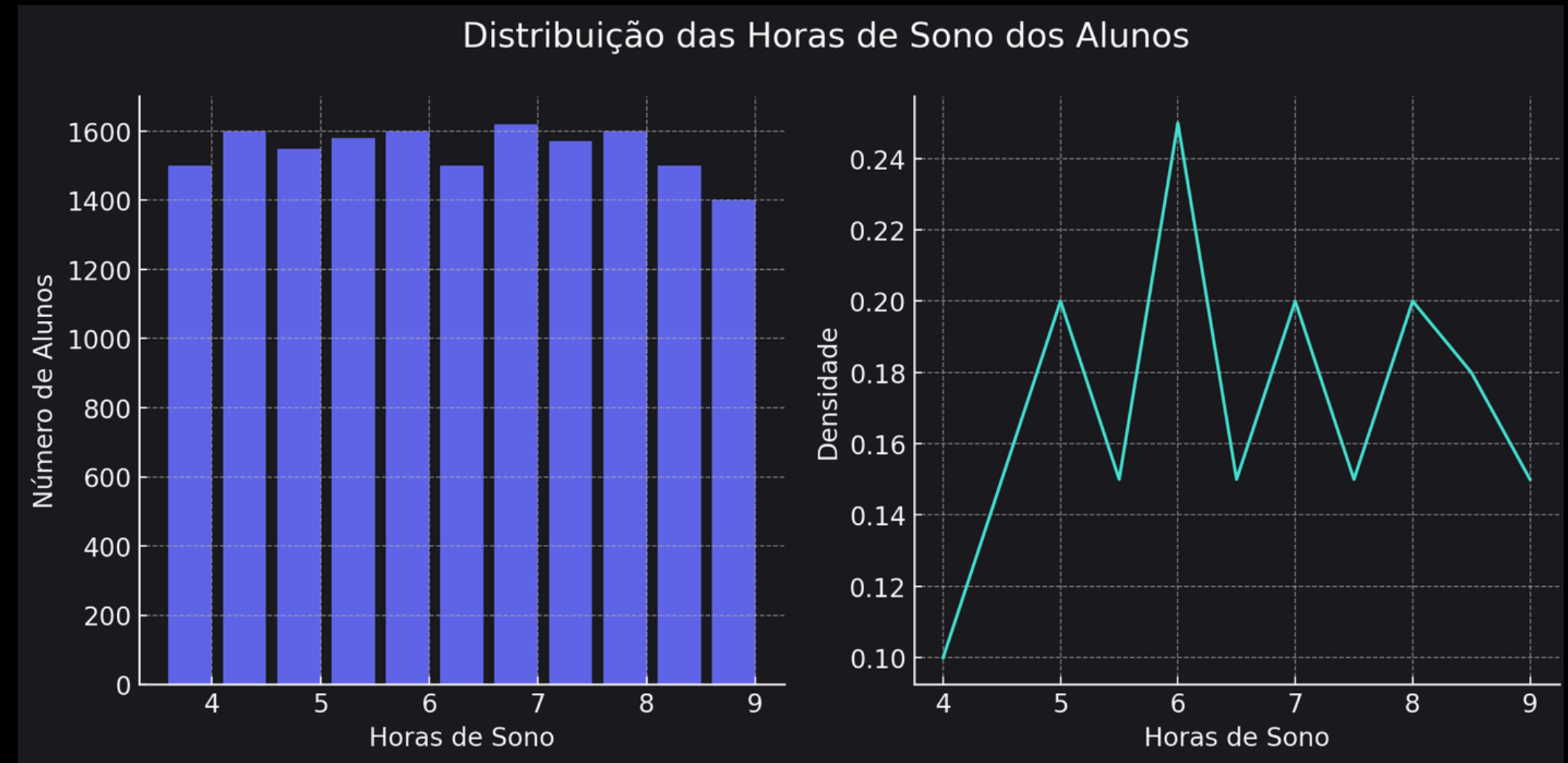
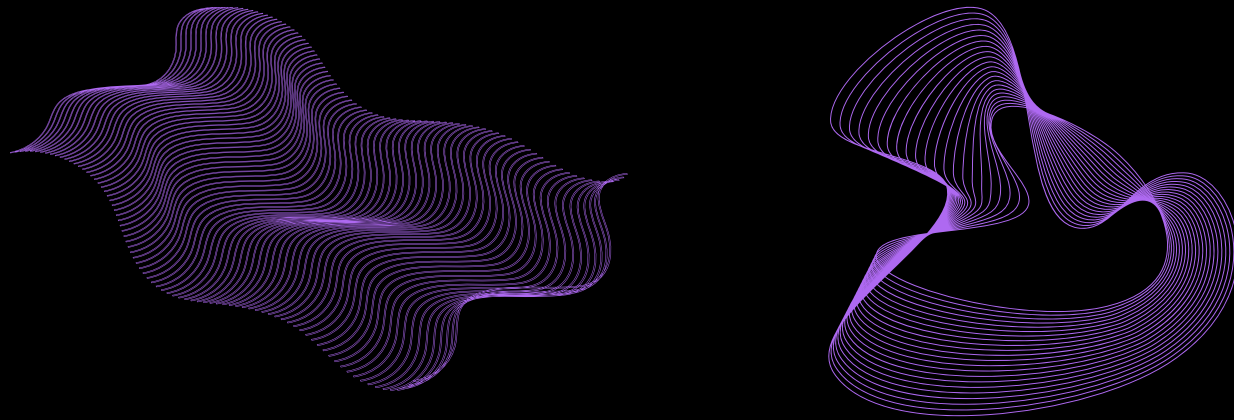
CORRELAÇÕES FORTES:

- CORRELAÇÃO SIGNIFICATIVA ENTRE A QUALIDADE DO SONO E O DESEMPENHO ACADÊMICO.
- MELHOR SONO CORRELACIONA COM MELHOR DESEMPENHO ACADÊMICO.

INFLUÊNCIA DAS NOITES MAL DORMIDAS:

- NOITES MAL DORMIDAS AFETAM NEGATIVAMENTE A PRODUTIVIDADE EM SALA DE AULA.
- ESTUDANTES COM SONO INADEQUADO TÊM NOTAS MAIS BAIXAS E MENOR CONCENTRAÇÃO.

Conclusão



OS RESULTADOS DESTE ESTUDO REFORÇAM A IMPORTÂNCIA DA QUALIDADE DO SONO NO DESEMPENHO ACADÊMICO E SUGEREM QUE INTERVENÇÕES FOCADAS EM MELHORAR OS HÁBITOS DE SONO PODEM TER UM IMPACTO SIGNIFICATIVO. A UTILIZAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA COMO KNN E ÁRVORE DE DECISÃO PROVOU SER EFICAZ PARA ANALISAR E INTERPRETAR OS DADOS, OFERECENDO INSIGHTS VALIOSOS PARA A APLICAÇÃO EM CONTEXTOS EDUCACIONAIS E DE SAÚDE PÚBLICA.
