



Previsão do Desempenho dos Alunos com Machine Learning

Projeto que usa notas de Matemática, Leitura e Escrita para prever o resultado final: aprovado, recuperação ou reprovado.



O que é esse projeto?



Análise de Notas

Estudamos as notas dos alunos em três disciplinas principais.



Previsão Inteligente

Usamos algoritmos de ML para prever o resultado final dos estudantes.



Intervenção Precoce

Identificamos quem precisa de ajuda antes da prova final.

Sobre os Dados

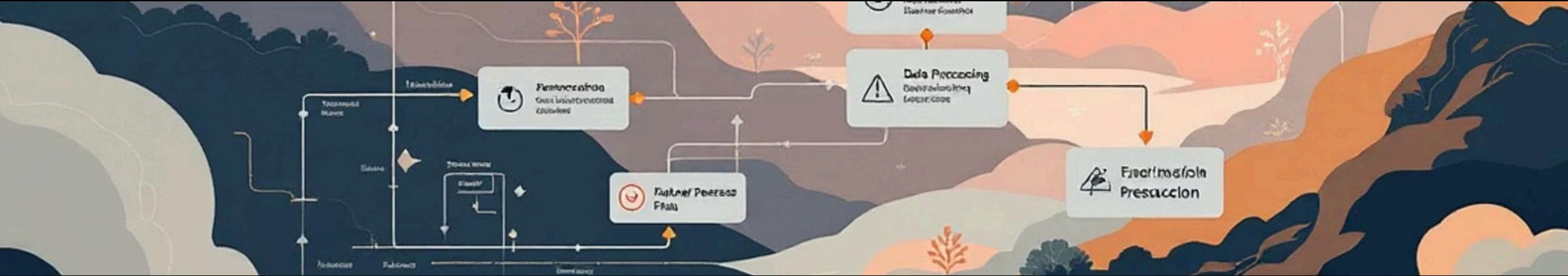
Os dados contêm três notas principais: **Matemática, Leitura e Escrita**.

Criamos uma nova coluna chamada **status_final**, que classifica cada aluno:

- **Aprovado** (média ≥ 70)
- **Recuperação** (média entre 50 e 69)
- **Reprovado** (média < 50)

Organizamos e limpamos esses dados antes de treinar os modelos.





Preparando os Dados

01

Verificação de Qualidade

Conferimos se havia valores faltando nos dados.

02

Padronização

Normalizamos as notas para melhorar o desempenho dos modelos.

03

Divisão Estratégica

70% para treinamento e 30% para teste.

04

Validação

O modelo aprende com uma parte e é avaliado com outra.

Análise Exploratória dos Dados

Distribuição das Notas

Gráficos mostrando como as notas estão distribuídas entre os alunos.

Mapa de Calor

Heatmap revelando quais matérias têm maior correlação.

Comparação de Médias

Diferença entre as médias de cada matéria estudada.

- ☐ **Descoberta importante:** Leitura e escrita apresentam forte correlação, enquanto Matemática tende a ter notas mais baixas.

Modelos de Machine Learning Testados

Comparamos seis algoritmos diferentes para identificar o melhor preditor:

1

Regressão Logística

Modelo linear clássico para classificação.

2

KNN

Classificação baseada em vizinhos próximos.

3

SVM

Máquina de vetores de suporte para separação ótima.

4

Árvore de Decisão

Modelo baseado em regras hierárquicas.

5

Random Forest

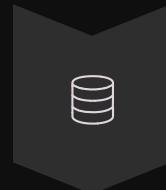
Conjunto de árvores para maior precisão.

6

Naive Bayes

Classificador probabilístico simples.

Como o Modelo Aprende?



Separação

Dividimos notas (X) e resultado final (y).



Normalização

Padronizamos as notas para o mesmo padrão.



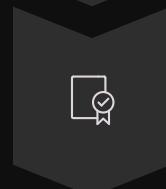
Treinamento

O modelo aprende com exemplos reais.



Previsão

Testamos em novos casos não vistos.



Validação

Comparamos previsões com resultados reais.

Principais Descobertas da Análise



Correlação Forte

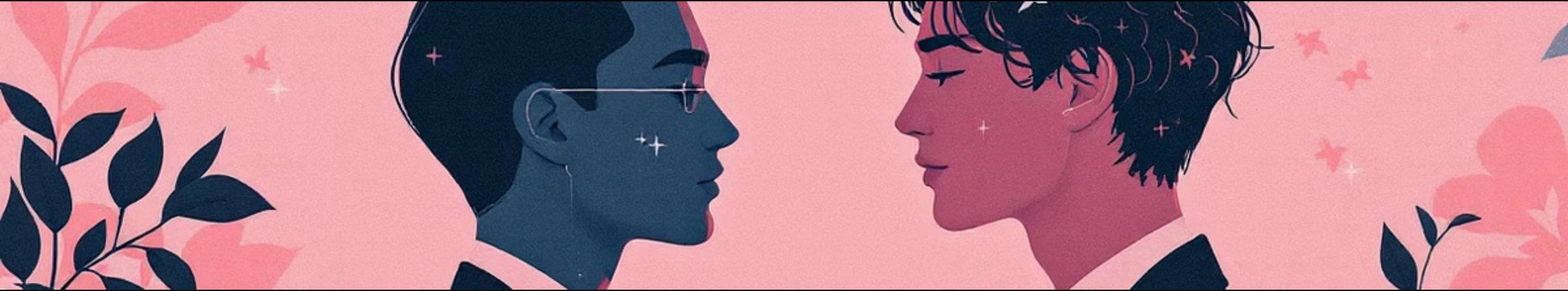
Leitura e escrita apresentam notas muito semelhantes entre os alunos.

Desafio em Matemática

Matemática é, geralmente, a nota mais baixa entre as três disciplinas.

Padrões Identificados

O heatmap revelou relações importantes entre as matérias.



Resultados Finais dos Modelos

1º

SVM

Melhor desempenho geral na classificação.

2º

Random Forest

Excelente acurácia e robustez.



Naive Bayes

Modelo mais simples com mais erros.

A matriz de confusão revelou que SVM e Random Forest tiveram os melhores resultados em acertos, enquanto modelos mais simples apresentaram maior taxa de erro nas previsões.

Conclusão e Próximos Passos

O projeto demonstrou que é possível prever o desempenho dos alunos usando Machine Learning com alta precisão.

Impacto Educacional

Escolas podem identificar alunos em risco antes das provas finais e oferecer suporte direcionado.

Matemática é Crucial

A disciplina tem grande influência no resultado final dos estudantes.

Evolução Contínua

Adicionar mais dados, testar novos modelos e criar um painel visual interativo.

