Uma imagem contendo Logotipo

Descrição gerada automaticamente

|  |  |  |  |
| --- | --- | --- | --- |
| **Curso:** | Ciência de Dados e Inteligência Artificial | **Aluno:** | Juliano França da Mata |
| **Modular:** | 5º | **RA:** | 23200190 |
| **Disciplina:** | Projeto Integrador V A | **Data Entrega:** | 15/04/2025 |

**Introdução**

A linguagem de programação Python, lançada em 1991, destaca-se por permitir o desenvolvimento de soluções com menos linhas de código em comparação a outras linguagens. Sua simplicidade, legibilidade e vasta comunidade contribuíram para sua ampla adoção em diversas áreas da tecnologia. Atualmente, o Python está integrado em praticamente todas as novas tecnologias, sendo utilizado em aplicações web, mobile, data science, machine learning, blockchain, entre outras.

Desde 2009, o Python tornou-se a linguagem padrão do curso de Ciência da Computação do Massachusetts Institute of Technology (MIT), o que reforça sua importância no meio acadêmico e científico. Além disso, foi eleito "Linguagem do Ano" pelo índice TIOBE em diferentes ocasiões: 2007, 2010, 2018 e 2020, demonstrando sua constante evolução e relevância no cenário da programação contemporânea (TIOBE, 2020).

O presente projeto integrador tem como base um tutorial publicado no site Data Flair, que aborda o pré-processamento e visualização de dados com Python. A primeira etapa do estudo contempla técnicas como normalização, padronização, transformação e binarização de dados, utilizando os pacotes Pandas e Scikit-learn. Já a segunda etapa do tutorial explora a visualização de dados por meio de histogramas e gráficos de densidade, com o uso das bibliotecas NumPy e Matplotlib.

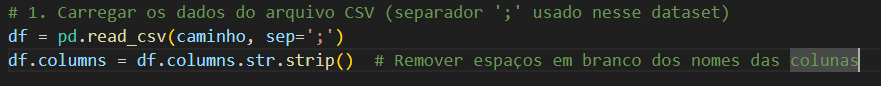
A base de dados utilizada refere-se à análise de vinhos portugueses e está disponível no repositório UCI Machine Learning Repository, na versão winequality-red.csv, a qual também se encontra nos arquivos da disciplina na plataforma Canvas. Como parte da proposta do projeto, será elaborado um relatório técnico no formato ABNT, contendo os programas desenvolvidos e os gráficos gerados a partir das operações realizadas sobre os dados.

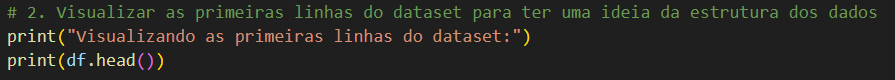
Desenvolvimento

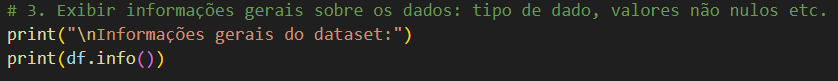
Texto

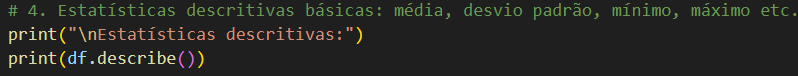
O conteúdo gerado por IA pode estar incorreto.











Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Gráfico, Aplicativo, Tabela

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Gráfico, Histograma

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Gráfico, Gráfico de caixa estreita

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Gráfico, Gráfico de caixa estreita

O conteúdo gerado por IA pode estar incorreto.

Texto

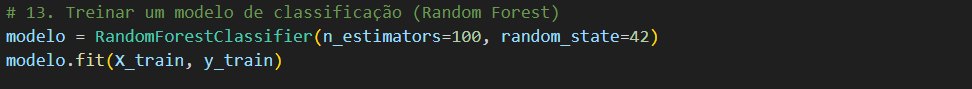
O conteúdo gerado por IA pode estar incorreto.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.



Uma imagem contendo Texto

O conteúdo gerado por IA pode estar incorreto.

# Conclusão:

# A análise inicial forneceu uma visão geral do conjunto de dados, incluindo a estrutura e estatísticas básicas.

# A matriz de correlação identificou variáveis com maior influência na qualidade do vinho.

# A distribuição da acidez e da qualidade revelou padrões importantes.

# Boxplots ajudaram a detectar outliers que podem impactar modelos futuros.

# A relação entre teor alcoólico e qualidade mostrou uma tendência clara de maior qualidade com mais álcool.

#

# Também foi aplicado um modelo de Random Forest para prever a qualidade do vinho com base nas variáveis do dataset.

# Neste caso, a qualidade foi agrupada em três categorias (baixa, média e alta) para facilitar a predição.

# As métricas de avaliação ajudam a entender a performance e precisão do modelo com essa nova abordagem.

#

# Próximas etapas recomendadas:

# - Otimização de hiperparâmetros do modelo (com GridSearchCV, por exemplo)

# - Teste com outros algoritmos como SVM, KNN, ou redes neurais

# - Análise de importância das variáveis para compreender quais características mais influenciam na qualidade