



VERIFICAÇÃO DE APRENDIZAGEM

ATIVIDADE ATIVA – ANÁLISE VISUAL DO DATASET IRIS

Discente _____

TURMA _____

1. DESCRIÇÃO

Nesta atividade, os você atuará como cientistas de dados em uma empresa de biotecnologia que deseja desenvolver um sistema inteligente para identificar automaticamente espécies de flores com base em suas características físicas. Para isso, a equipe precisa analisar um conjunto de dados reais contendo medidas de três espécies de íris (Setosa, Versicolor e Virginica), extraídas de flores coletadas no campo.

Antes de desenvolver um modelo de machine learning, a equipe deve realizar uma análise visual dos dados para compreender os padrões existentes, identificar atributos relevantes e justificar suas escolhas com base em evidências visuais.

2. SITUAÇÃO-PROBLEMA

A empresa está desenvolvendo um aplicativo de identificação de flores baseado em inteligência artificial, que ajudará biólogos e jardineiros a reconhecer espécies com precisão. Para garantir que o sistema seja confiável, a equipe de cientistas de dados deve analisar os dados coletados no campo, identificar quais características são mais úteis para distinguir as espécies e propor uma solução baseada em aprendizado de máquina.

Como parte da equipe de dados, você recebeu o seguinte desafio:

- ✓ Com base nas visualizações geradas a partir dos dados das flores, quais atributos são mais relevantes para distinguir as espécies?
- ✓ Quais padrões podem ser observados?
- ✓ Que tipo de modelo você recomendaria aplicar nesse cenário?

3. Objetivo Geral:

Permitir que os alunos explorem dados do conjunto Iris e visualizem como diferentes espécies de flores se distribuem em relação a suas características, ajudando-os a perceber padrões e justificar decisões antes de aplicar qualquer algoritmo.

3.1. Objetivos Específicos:

- I. Desenvolver habilidades de visualização e interpretação de dados.
- II. Reconhecer a importância da análise exploratória antes da modelagem.
- III. Estimular o pensamento crítico na escolha de atributos e algoritmos.
- IV. Relacionar a teoria da aprendizagem de máquina com a prática.

4. ATIVIDADE

Requisitos:

- Python instalado (ou Google Colab)
- Bibliotecas: pandas, matplotlib, seaborn

Execução:

- Explorar o dataset usando gráficos
- Analisar padrões entre as espécies
- Responder perguntas interpretando as visualizações

Passo a passo:

I. Carrega o dataset direto da internet

```
url = "https://raw.githubusercontent.com/mwaskom/seaborn-data/master/iris.csv"  
df = pd.read_csv(url)
```

- II. Mostra as 5 primeiras linhas do dataset.
- III. Gere - Gráfico 1 - Relação entre comprimento e largura das pétalas, colorido por espécie.
- IV. Gere - Gráfico 2 - Boxplot para comparar distribuições de sépalas.
- V. Gere - Gráfico 3 - Matriz de correlação.

Perguntas:

- a) Existe algum padrão claro que diferencia as espécies?
- b) Quais atributos parecem ser mais úteis para separar as classes?
- c) Há sobreposição entre as espécies em algum gráfico?
- d) Se você fosse treinar um modelo, quais colunas escolheria como mais relevantes? Por quê?

5. Requisitos da Atividade

Você deverá entregar:

- ✓ Código Python com os gráficos e análises.
- ✓ Respostas das perguntas.
- ✓ Justificativa das decisões baseadas nas visualizações.

OBS:

- ✓ Não será aceito código/projeto copiado de colegas ou da internet ou gerado por IAs.
- ✓ **Não será aceito arquivo zipado ou compactado.**
- ✓ AA deve ser desenvolvido individualmente.
- ✓ *Códigos com estrutura ou lógica idêntica entre diferentes alunos serão desconsiderados.*
- ✓ O relatório deve ser entregue dentro do prazo estipulado.
- ✓ Utilize terminologia adequada e mantenha organização lógica e clareza no código e no relatório.

Explico melhor na primeira web teoria e prática.