



Pandas

Semana 6 / Atividade 2

Nesta atividade, você vai aprender Pandas, a biblioteca padrão para manipulação e análise de dados tabulares em Python, usando Series e DataFrame. A ideia é ganhar fluência em carregar dados, filtrar, agrregar, ordenar e gerar análises rápidas com poucas linhas de código.

Por que estudar Pandas se o foco é visão computacional?

Em visão computacional, você vai lidar o tempo todo com logs de treino (loss, acc, auc), previsões, rótulos e resultados por época/fold. O Pandas facilita:

- Consolidar resultados em tabela;
- Calcular médias/desvios (ex: validação cruzada);
- Comparar arquiteturas;
- Gerar relatórios e exportar tudo.

Objetivos

Ao final deste estudo dirigido, você deverá:

- Carregar dados (CSV) em DataFrames.
- Inspecionar dados (`head`, `info`, `describe`) e tratar colunas.
- Filtrar e selecionar linhas/colunas.
- Criar colunas derivadas (ex: acerto/erro).

- Calcular TP, TN, FP, FN e métricas (accuracy, precision, recall, specificity).
- Fazer rankings e consultas (top-k, menores/maiores probabilidades).

Atividades

Acompanhar as aulas e codar junto

Assista à playlist, faça exemplos no seu editor e salve no diretório src/07-pandas:

<https://www.youtube.com/playlist?list=PL3ZslI15yo2pfkf7EGNR14xTwe-wZ2bNX>

Exercícios Práticos

Ex01.py - Análise exploratória do CSV de classificação

Arquivo: classification_results_trial_0001.csv disponível no diretório de arquivos compartilhados do onboarding.

Exemplo dos dados presentes no csv:

```
image_path,real_class,predicted_class,prob_benign,prob_malign
image_001.jpg,malign,malign,0.0314291856867343,0.968570814313266
image_002.jpg,benign,benign,0.63641041126378,0.36358958873622
image_003.jpg,malign,malign,0.314355981076327,0.685644018923673
image_004.jpg,benign,benign,0.508570691164703,0.491429308835297
```

Realize a seguinte análise exploratória utilizando o Pandas:

1. Quantas imagens são “benign” e “malign” em `real_class`?
2. Identifique em quais imagens o modelo errou a predição.
3. Verifique se o modelo estava confiante mesmo quando errou.
4. Calcule:
 - a. TP (real malign, previsto malign)
 - b. TN (real benign, previsto benign)
 - c. FP (real benign, previsto malign)

d. FN (real malign, previsto benign)

5. Calcule:

a. Acurácia: $(TP+TN)/(TP+TN+FP+FN)$

b. Precisão: $TP/(TP+FP)$

c. Recall: $TP/(TP+FN)$

d. Especificidade: $TN/(TN+FP)$

6. Encontre as 5 imagens **benign** com menor **prob_benign**. O que isso pode indicar?

7. Encontre as 5 imagens **malign** com maior **prob_benign**. O que isso pode indicar?

Entrega da Atividade

Gerar um arquivo PDF que contenha:

- Seu nome.
- O link do seu repositório no Github.
- O código de todas as atividades desse roteiro de estudos (código das videoaulas, reflexões e exercícios)

Enviar o arquivo PDF na tarefa S6.A2. Análise Exploratória Pandas.