

# FIAP - Faculdade de Informática e Administração Paulista

---

## Cap 3 – (IR ALÉM) Implementando Algoritmos de Machine Learning com Scikit-learn (Seeds)

---

Nome do grupo

Grupo 4 - DeepThinkers

Integrantes

- André Pessoa Gaidzakian – RM567877
- Erick Prados Pereira – RM566833
- Guilherme Ferreira Santos – RM568523
- Viviane de Castro – RM567367



Professores:

Tutora

- Sabrina Otoni

Coordenador

- André Godoi Chiovato



Descrição

Em cooperativas agrícolas de pequeno porte, a classificação de grãos é realizada manualmente por especialistas, processo sujeito a erros e demorado. Este projeto aplica a metodologia **CRISP-DM** para automatizar a classificação de variedades de trigo (Kama, Rosa, Canadian) com base no “Seeds Dataset” (UCI).

A solução foi desenvolvida inteiramente em um **Jupyter Notebook**, cobrindo:

1. **Business Understanding**: Definição do problema de classificação.
2. **Data Understanding**: Análise exploratória (EDA) com estatísticas e visualizações.
3. **Data Preparation**: Limpeza, imputação e padronização dos dados.
4. **Modeling**: Treinamento de algoritmos (KNN, SVM, Random Forest, Naive Bayes).
5. **Evaluation**: Comparação de métricas e otimização via GridSearchCV.



Estrutura de pastas

Dentre os arquivos e pastas presentes na raiz do projeto, definem-se:

- **assets**: aqui estão os arquivos relacionados a elementos não-estruturados deste repositório, como imagens.

- **config:** Posicione aqui arquivos de configuração que são usados para definir parâmetros e ajustes do projeto.
- **document:** aqui estão todos os documentos do projeto que as atividades poderão pedir. Na subpasta "other", adicione documentos complementares e menos importantes.
- **scripts:** Posicione aqui scripts auxiliares para tarefas específicas do seu projeto.
- **src:** Todo o código fonte criado para o desenvolvimento do projeto ao longo das 7 fases.
- **README.md:** arquivo que serve como guia e explicação geral sobre o projeto (o mesmo que você está lendo agora).

## Como executar o código

### Pré-requisitos

- Python 3.10+
- Jupyter Notebook ou Google Colab

### Instalação Local

1. Clone o repositório e acesse a pasta:

```
git clone <seu-repo>
cd <seu-repo>
```

2. Crie um ambiente virtual (recomendado):

```
# Windows
python -m venv .venv
.venv\Scripts\activate

# Linux/Mac
python3 -m venv .venv
source .venv/bin/activate
```

3. Instale as dependências:

```
pip install pandas==2.2.2 numpy==2.1.1 scikit-learn==1.4.2 matplotlib==3.9.0
seaborn==0.13.2 joblib==1.4.2 notebook
```

### Execução

1. Inicie o Jupyter Notebook:

```
jupyter notebook
```

2. Navegue até a pasta [document/](#) e abra o arquivo [FASE\\_04\\_CTWP\\_Cap3.ipynb](#).
3. Execute todas as células sequencialmente para reproduzir a análise, o treinamento e a geração dos resultados.



## Histórico de lançamentos

- 0.1.0 - 23/11/2025



## Licença

[MODELO GIT FIAP](#) por [Fiap](#) está licenciado sobre [Attribution 4.0 International](#).

Este projeto utiliza o [Seeds Dataset](#) (CC BY 4.0).