

CRONOGRAMA E EMENTA

Aula 1 – 08/11

Conceitos Gerais de Machine Learning

- Apresentações
- Grade Curricular
- Objetivos
- Método de Ensino
- Critérios de Avaliação
- Referências Bibliográficas
- Introdução a Machine Learning
- Tipos de aprendizado (supervisionado, não supervisionado)
- Análise Exploratória, Pré-processamento de dados e engenharia de atributos (construção de pipelines)
- Divisão em treino e teste
- Desbalanceamento dos dados
- Overfitting e Underfitting
- Validação Cruzada e Métricas de Avaliação

Aula 2 – 29/11

Modelos Lineares e Árvores

- Regressão Linear (+ SVM Linear)
- Regressão Logística
- Métricas de avaliação para regressão
- Árvores de decisão

Aula 3 – 06/12

Ensemble e Métodos de Avaliação/Interpretabilidade

- Random Forest
- Bagging, Boosting e principais algoritmos
- Métodos mais robustos de avaliação de modelos de classificação (classification report, matriz de confusão, curvas ROC-AUC, AUC-PR)
- Validação cruzada e redução de overfitting
- Tuning de hiperparâmetros – Grid Search, Random Search, Optuna e Ray Tune
- Interpretabilidade com SHAP Values e Lime

Aula 4 – 13/12

Modelos Baseados em Distância e Clusterização

- KNN e SVM
- K-means e DBScan
- Identificação de anomalias com clusterização
- Métodos de escolha de K
- Apresentação do Trabalho Final

COMPOSIÇÃO DA NOTA

- Atividade em aula: 1 ponto (3 atividades)
- Atividade para casa: 1 ponto (3 atividades)
- Trabalho Final: 3 pontos
- Participação: 1 ponto

TRABALHO FINAL

- O trabalho final será em grupo e o grupo deverá escolher um dataset voltado para resolução de problemas de negócio (churn, fraude, crédito, inadimplência, etc).
- O grupo deverá fazer uma apresentação com duas abordagens: técnica e de negócio.
- A entrega deverá conter um notebook com todo o código para a resolução do problema comentado e organizado, e os slides da apresentação.

O notebook deve conter:

- Análise Exploratória com principais insights
- Processamento dos Dados
- Modelagem e Avaliação dos modelos

A apresentação deve conter:

- O que o grupo considerar relevante para ser apresentado para um público técnico **E** para um público focado no resultado do negócio (não técnico).