



Treinamento CIS - 1º Período (Classificação)

1. Conteúdos:

- a. Introdução a classificação;
 - i. Conceitos básicos e aplicações;
 - ii. Classificação binária;
 - iii. Classificação multiclasse;
 - iv. Métricas de avaliação (acurácia, precisão, recall, F1-score, AUC-ROC).
- b. Algoritmos de classificação;
 - i. Árvore de decisão;
 - ii. Naive Bayes;
 - iii. Support Vector Machines (SVM);
- c. Validação e ajuste de modelos;
 - i. Validação cruzada;
 - ii. Matriz de confusão;
 - iii. Ensemble Learning;
 - iv. Ajuste de hiperparâmetros.

2. Materiais:

- a. [DataCamp](#) - Introdução a classificação;
- b. [Vídeo explicando](#) - Métricas de avaliação;
- c. [Vídeo explicando](#) - Árvore de decisão;
- d. [Vídeo explicando](#) - Naive Bayes;
- e. [Vídeo explicando](#) - Support Vector Machines;
- f. [Vídeo explicando](#) - Validação cruzada;
- g. [Vídeo explicando](#) / [Aula](#) - Random forest;
- h. [Artigo explicando](#) - Ensemble Learning.



DESAFIO: Classificação de Renda — Dataset: [Adult \(Census Income\)](#)

Descrição:

O conjunto de dados “Adult” (Census Income) reúne informações do censo dos EUA de 1994, com 48.842 registros e 14 atributos demográficos e socioeconômicos. Seu objetivo é construir um modelo de **classificação binária** para prever se a renda anual de um indivíduo excede US\$ 50.000.

Atributos

1. age (idade)
2. hours-per-week (horas trabalhadas por semana)
3. capital-gain (ganho de capital)
4. capital-loss (perda de capital)
5. fnlwgt (peso da amostra)
6. workclass (classe de trabalho)
7. education (nível de escolaridade)
8. marital-status (estado civil)
9. occupation (ocupação)
10. relationship (grau de parentesco)
11. race (raça)
12. sex (sexo)
13. native-country (país de origem)
14. income (“≤50K” ou “>50K”) (**Variável alvo**)

TAREFAS

1. **Comparação de Modelos (OBRIGATÓRIA)**
 - 1.1. Treinar dois classificadores: Árvore de Decisão e Random Forest
 - 1.2. Avaliar cada modelo usando:
 - Acurácia
 - Precisão
 - Recall
 - F1-score
 - 1.3. Gerar:
 - Matriz de confusão
 - Gráfico de importância de atributos
2. **Análise Exploratória dos Dados (OPCIONAL)**
 - Examinar distribuições (idade, horas, ganhos, etc.)
 - Investigar correlações e influência de ‘education’ e ‘occupation’
3. **Validação Cruzada (OPCIONAL)**
 - Aplicar e comparar: hold-out, K-fold e Stratified K-fold
 - Discutir impacto na variância e no viés dos modelos
4. **Redução de Dimensionalidade (OPCIONAL)**
 - Codificar e normalizar variáveis
 - Aplicar PCA e avaliar efeito no desempenho dos classificadores