



Avaliação AV3 – Parte 1 – Atividade 2

Importação de bibliotecas essenciais:

Importa o “Pandas” para manipulação de dados, ferramentas de divisão de conjunto de dados e os classificadores do “scikit-learn”.

Leitura e pré-processamento dos dados:

- Carrega o arquivo CSV e usa “LabelEncoder” para converter variáveis categóricas (como “DeptoAtuacao” e “Salario”) em valores numéricos.

Divisão dos dados:

Separa o conjunto de dados em X (features) e Y (variável alvo) e, em seguida, divide em conjuntos de treino e teste.

Treinamento e avaliação de modelos:

- Usa três algoritmos de classificação: **Regressão Logística**, **Árvore de Decisão** e **K-Nearest Neighbors (K-NN)**.
- Calcula e armazena a acurácia e a matriz de confusão para cada modelo.

Saída dos resultados: Imprime a acurácia e a matriz de confusão de cada modelo para comparação.

```
Model: Logistic Regression
Accuracy: 0.76
Confusion Matrix:
[[3158, 270], [818, 254]]

Model: Decision Tree
Accuracy: 0.98
Confusion Matrix:
[[3367, 61], [42, 1030]]

Model: K-Nearest Neighbors
Accuracy: 0.93
Confusion Matrix:
[[3208, 220], [115, 957]]
```

Regressão Logística

- **Precisão:** 0.76 (76%)

Matriz de Confusão

- **Interpretação:**
 - **3158:** Número de verdadeiros negativos (funcionários corretamente previstos como não saindo da empresa).
 - **270:** Número de falsos positivos (funcionários previstos como saindo, mas que realmente não saíram).
 - **818:** Número de falsos negativos (funcionários previstos como não saindo, mas que realmente saíram).
 - **254:** Número de verdadeiros positivos (funcionários corretamente previstos como saindo da empresa).

Árvore de Decisão

- **Precisão:** 0.98 (98%)

Matriz de Confusão:

- **Interpretação:**
 - **3367:** Número de verdadeiros negativos.(funcionários corretamente previstos como não saindo da empresa).
 - **61:** Número de falsos positivos.(funcionários previstos como saindo, mas que realmente não saíram).
 - **42:** Número de falsos negativos.(funcionários previstos como não saindo, mas que realmente saíram).
 - **1030:** Número de verdadeiros positivos.(funcionários corretamente previstos como saindo da empresa).

K-Nearest Neighbors (K-NN)

- **Precisão:** 0.93 (93%)

Matriz de Confusão:

- **Interpretação:**
 - **3208:** Número de verdadeiros negativos.(funcionários corretamente previstos como não saindo da empresa).
 - **220:** Número de falsos positivos.(funcionários previstos como saindo, mas que realmente não saíram).
 - **115:** Número de falsos negativos.(funcionários previstos como não saindo, mas que realmente saíram).

- **957**: Número de verdadeiros positivos.(funcionários corretamente previstos como saindo da empresa).

Conclusão: Para esse problema específico de predição de turnover, a árvore de decisão é a escolha mais eficaz devido à sua alta acurácia e à baixa quantidade de erros de classificação.