



APRENDIZADO DE MÁQUINA

DECISION TREE X RANDOM FOREST

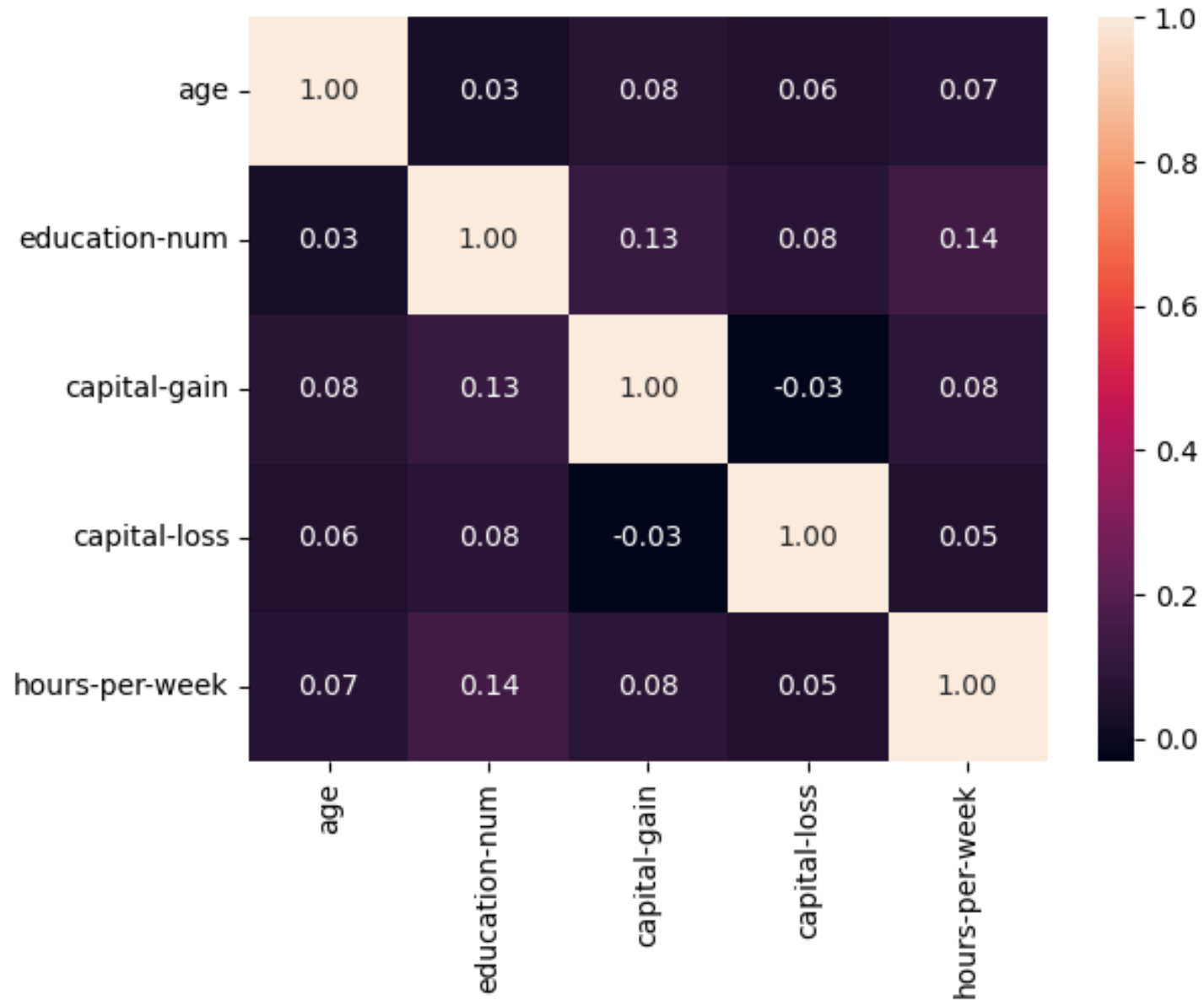
INFORMAÇÕES SOBRE O DATASET

- O conjunto de dados "**Adult**" contém informações sobre:
 - Renda, educação, idade, sexo, etnia, trabalho, estado civil, ganho capital;
 - Perda capital, horas de trabalho por semana, nacionalidade, relacionamento familiar;
- O conjunto de dados contém **48.842** entradas com um total de **15** colunas representando diferentes atributos das pessoas, conforme a tabela abaixo;
- O conjunto de dados foi dividido previamente em dois datasets, o dataset de de treino **adult.data** que contém **32.560** entradas, e o dataset de testes **adult.test** que contém **16.282** entradas.
- Os dados foram coletados em 1994 e a variável alvo, **Renda** (income) é dividida em duas classes:
 - "**Alta Renda**" ($> \$50K$) e "**Baixa Renda**" ($\leq \$50K$).

DECISION TREE E RANDOM FOREST

- Metodologia:
 - Critério: Gini;
 - Profundidade máxima da árvore: 10;
 - Random State: 0;
 - K-Fold: 6;
 - As linhas que possuíam dados faltantes foram eliminadas;
 - As linhas duplicadas foram eliminadas;
 - Os dados categóricos foram transformados em numéricos;
 - Utilizou-se o Scikit-Learn.

Matriz de Correlação



DECISION TREE - METRICS

	Precision	Recall	F1-Score	Quantidade
False	87,06%	88,29%	87,67%	23028
True	62,38%	59,68%	61,00%	7492

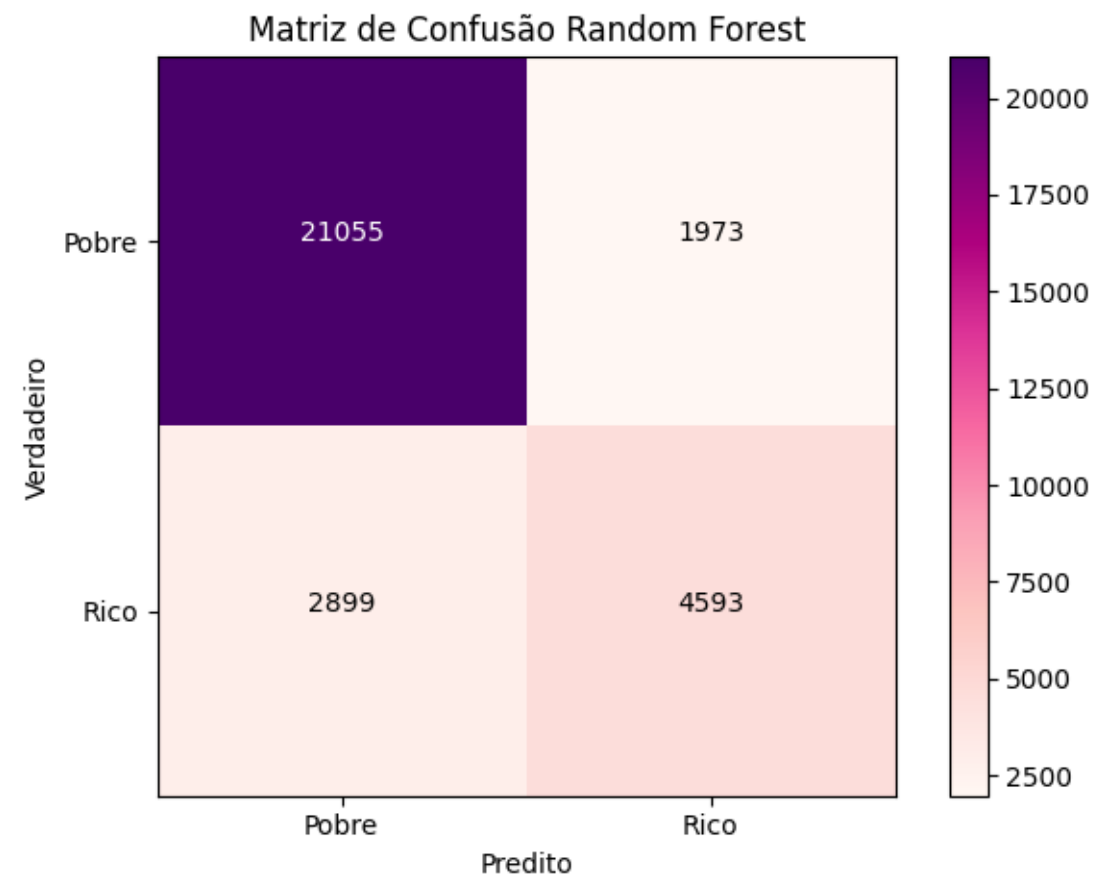
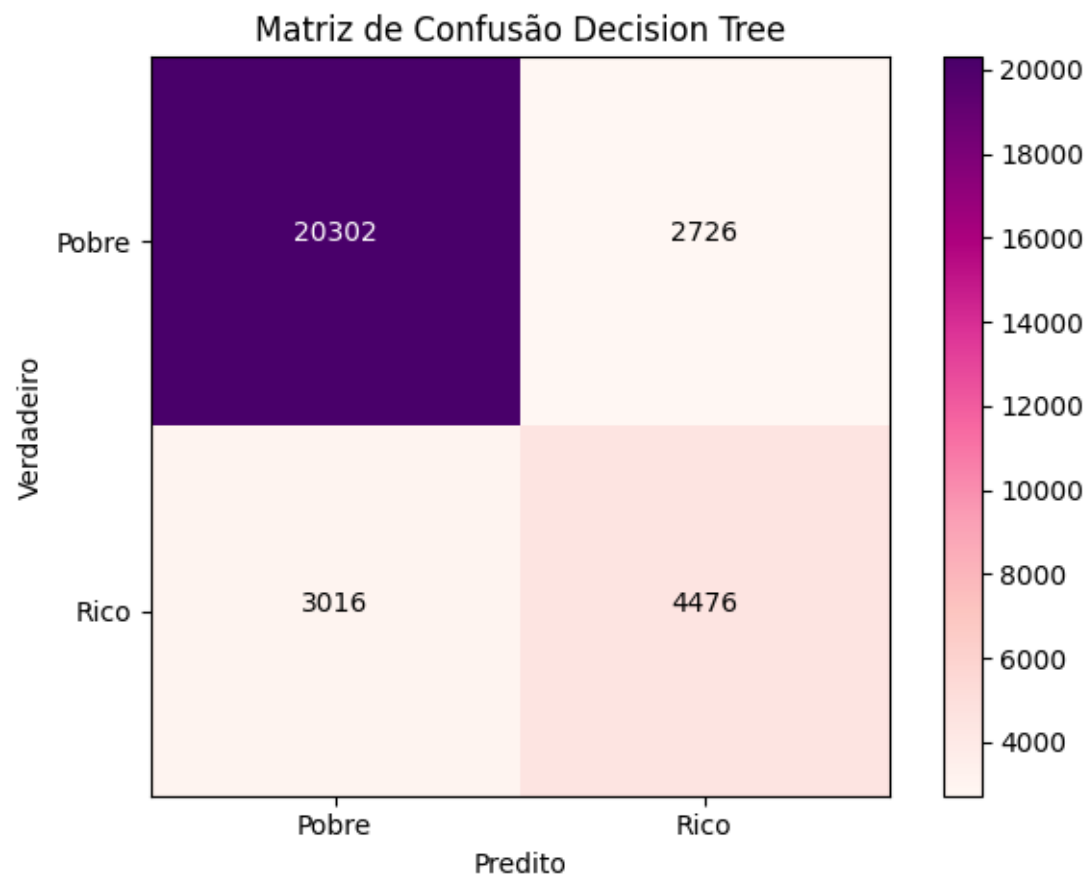
Métricas	Percentual	Quantidade Respondentes
Accuracy	81,27%	30520
ROC	76,99%	-

RANDOM FOREST - METRICS

	Precision	Recall	F1-Score	Quantidade Resp.
False	87,90%	91,43%	89,63%	23028
True	69,95%	61,31%	65,34%	7492

Métricas	Percentual	Quantidade Respondentes
Accuracy	84,04%	30520
ROC	88,52%	-

COMPARAÇÃO DE MATRIZES DE CONFUSÃO



CONSIDERAÇÕES FINAIS

- O **Random Forest** superou a única **Árvore de Decisão** em todas as métricas;
 - Torna-se claro ao comparar visualizando as matrizes de confusões.
- Com essa experiência é importante sempre notarmos a importância desses modelos de ensemble;



OBRIGADO

- Marcelo Victor Sousa Lima
- Danilo Rodrigues Barbosa