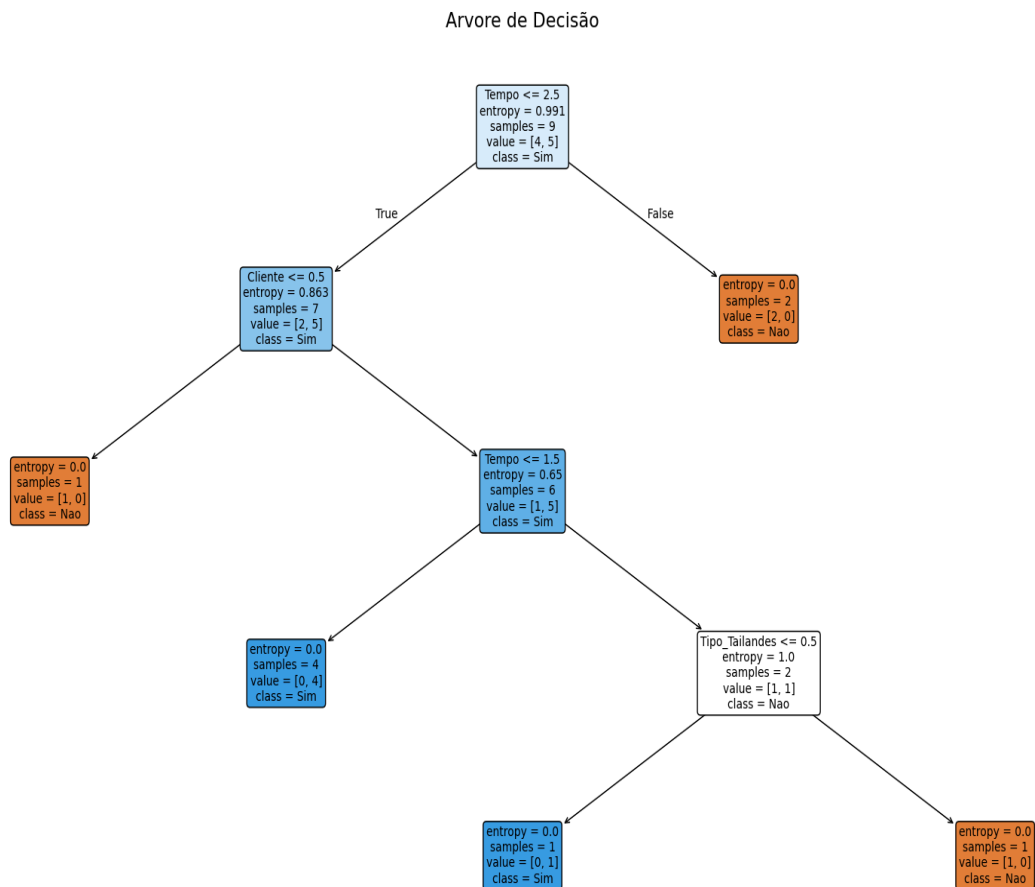
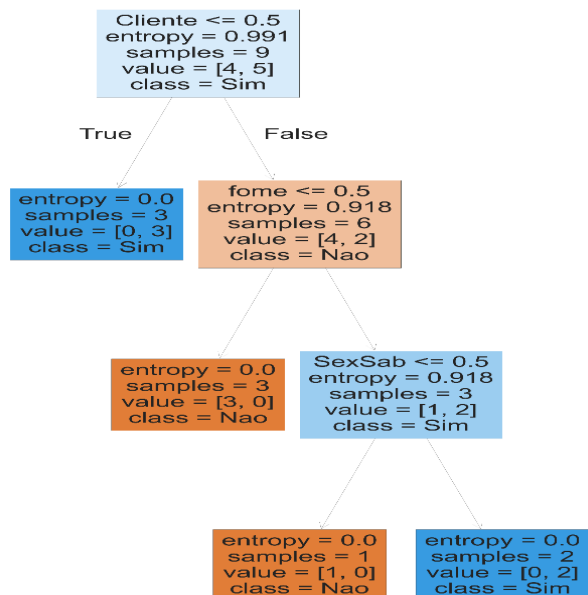


Questão 01

Árvore De Decisão Modificada com Atributo cliente 0,1 ou 2:



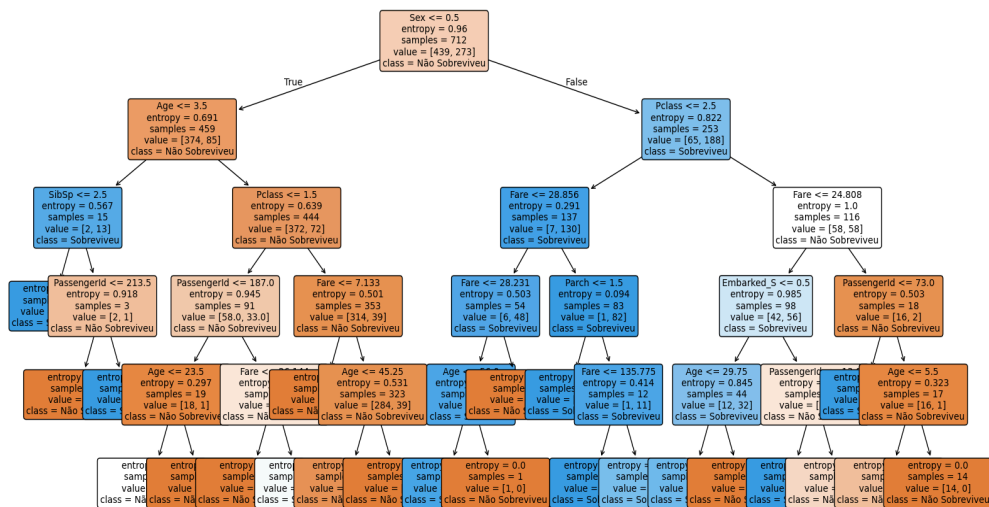
Árvore Original:



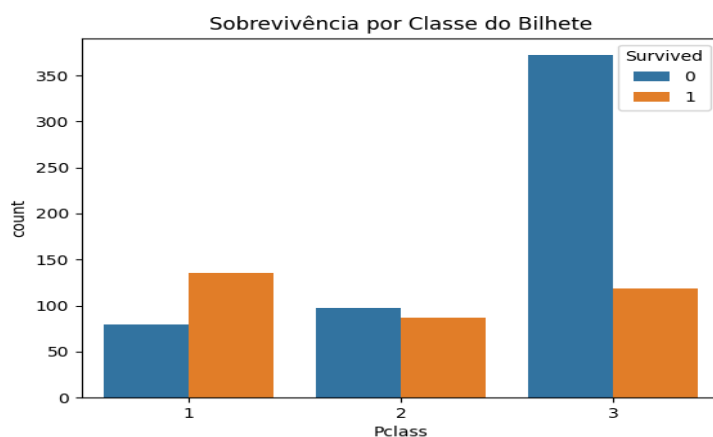
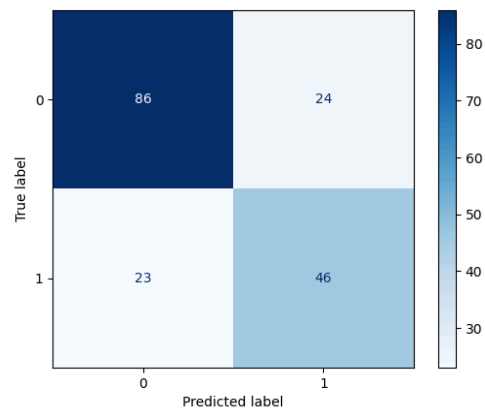
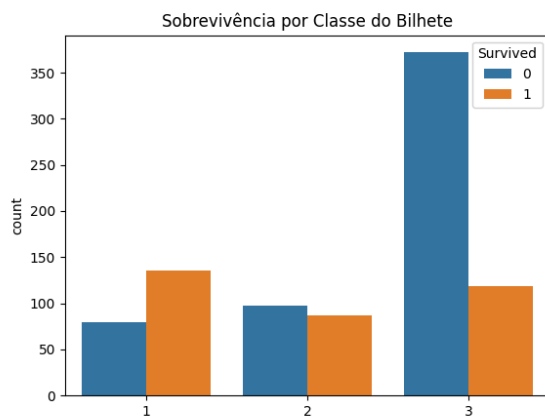
Questão 02

Base = "train.csv"

Árvore de Decisão do Titanic



Matriz de Confusão



REPOSITÓRIO GITHUB COM CÓDIGOS E IMAGENS GERADAS:

<https://github.com/andrielmark/PUCMINAS/blob/main/QUARTO%20PER%C3%8DODO/IA/LISTA3>

Questão 03

3.1) Questão 1: Diferenças entre ID3 e C4.5 O ID3 (Induction of Decision Tree) e o C4.5 são algoritmos de árvore de decisão desenvolvidos por J. Ross Quinlan, sendo o C4.5 uma extensão mais avançada do ID3. O ID3 utiliza o conceito de entropia e ganho de informação para selecionar o melhor atributo para construir a árvore. Uma de suas limitações é a sensibilidade a atributos com muitos valores, que podem ter um ganho de informação artificialmente alto. O C4.5 supera essa falha ao utilizar a razão de ganho, que normaliza o ganho de informação, evitando a preferência por atributos com múltiplos valores. Além disso, o ID3 é restrito a atributos nominais e não lida com valores ausentes, enquanto o C4.5 tem a capacidade de usar dados contínuos e lidar com valores desconhecidos. Outra diferença fundamental é que o C4.5 inclui uma técnica de poda (pruning) para reduzir o tamanho da árvore e melhorar a acurácia, algo que o ID3 não faz.

3.2) O algoritmo C4.5 gerencia atributos com valores contínuos de uma maneira que o ID3 não consegue. Para um atributo numérico, o C4.5 examina os valores no conjunto de dados de treinamento, ordenando-os em ordem ascendente. Em seguida, para cada valor, ele cria uma partição dos registros, dividindo-os em dois grupos: aqueles com valores menores ou iguais ao ponto de corte e aqueles com valores maiores. Para cada uma dessas partições, o ganho de informação ou a razão de ganho é calculado. Por fim, a partição que maximiza o ganho é selecionada, e esse ponto de corte é usado para criar uma regra de decisão binária no nó da árvore, permitindo que o modelo classifique dados numéricos de forma eficaz.

Questão 04

RESPOSTA: LETRA C: Iris_Versicolor, iris_Setosa, Iris_Versicolor, Iris_Virginica

Questão 05

RESPOSTA: LETRA C: I e II, apenas.

Questão 06

	Precisão	Recall	F1Score	TVP	TFN	TFP	TVN
A	10/17	10/17	0,5882	10/17	7/17	7/105	98/105
B	15/23	15/18	0,7317	15/18	3/18	8/106	98/106
C	20/26	20/30	0,7143	20/30	10/30	6/92	86/92
D	50/56	50/57	0,885	50/57	7/57	6/65	59/65