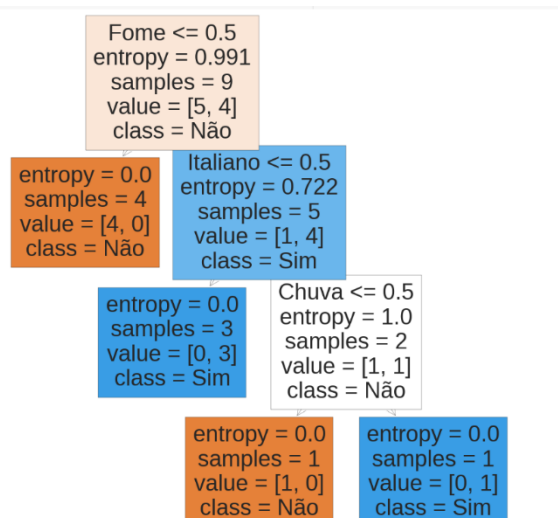


Questão 1 :

1)



- 2) Não houve alterações, pois a base de dados é muito pequena, assim impossibilita ver as mudanças. Entretanto, a mudança dos atributos de ordinais para nominais pode afetar os resultados dos experimentos com árvores de decisão. Ao perder a relação de ordem entre os valores, a árvore pode se tornar mais complexa, com mais nós e ramos. Isso pode levar a um risco maior de overfitting e menos eficiência na generalização para novos dados. Embora possa haver situações em que a conversão resulte em maior precisão e interpretabilidade, é importante pesar os prós e contras dessa mudança específica para o conjunto de dados em questão. Experimentar e comparar resultados será fundamental para determinar se a mudança trouxe ganhos.

Questão 2:

- 1) O algoritmo ID3 cria árvores de decisão a partir de dados rotulados, focando em atributos categóricos. Em contraste, o algoritmo C4.5 é uma versão aprimorada do ID3 que lida tanto com atributos categóricos quanto numéricos, ao dividir dados contínuos em intervalos discretos. Além disso, o C4.5 pode lidar com valores faltantes, atribuir diferentes pesos aos atributos, podar árvores após a criação para evitar superajuste e empregar um erro pessimista de previsão e sub-árvore elevada para melhorar a eficácia e a generalização da árvore resultante.
- 2) O C4.5 lida com atributos numéricos transformando-os em atributos categóricos. Ele faz isso ao definir intervalos para os valores numéricos e, em seguida, tratando esses intervalos como categorias discretas. Isso permite que o C4.5 trabalhe com dados numéricos em sua estrutura de árvore de decisão, mesmo sendo originalmente projetado para lidar com atributos categóricos.

Link do notebook:

https://colab.research.google.com/drive/1I2xUs4Al0XWnR14bWLMEUULc2IP7Adcq#scrollTo=fQb1CmKU_Wcz