



1. Se tamanho da pétala  $\leq 2.35$  é Iris\_Setosa
2. Se tamanho da pétala  $> 2.35$ , largura da pétala  $\leq 1.75$  e tamanho da pétala  $\leq 4.95$  é Iris\_Versicolor
3. Se tamanho da pétala  $> 2.35$ , largura da pétala  $\leq 1.75$  e tamanho da pétala  $> 4.95$  é Iris\_Virginica
4. Se tamanho da pétala  $> 2.35$ , largura da pétala  $> 1.75$  e tamanho da pétala  $\leq 4.85$  é Iris\_Versicolor
5. Se tamanho da pétala  $> 2.35$ , largura da pétala  $> 1.75$  e tamanho da pétala  $> 4.85$  é Iris\_Virginica

**II. Das regras geradas, há apenas uma com cobertura por classe de 100%** - Verdade, apenas a classe da Iris\_Setosa possui cobertura de classe de 100%

**III. A menor cobertura por classe é de 6.8% e corresponde à classe Iris\_Virginica** - Falso, apesar da menor cobertura de classe da Iris\_Virginica realmente ser 6.8% no caminho Direita → Esquerda → Direita, esta não é a menor cobertura de classe. A menor cobertura de classe é de 2.7% da Iris\_Versicolor no caminho Direita → Direita → Esquerda

**Resposta:** Letra C - I e II, apenas

### Questão 3

|               |   |                       |    |    |    |
|---------------|---|-----------------------|----|----|----|
|               |   | Foi classificado como |    |    |    |
|               |   | A                     | B  | C  | D  |
| Era da classe | A | 10                    | 4  | 2  | 1  |
|               | B | 1                     | 15 | 2  | 0  |
|               | C | 2                     | 3  | 20 | 5  |
|               | D | 4                     | 1  | 2  | 50 |

Letra A - VP = 10 | FN = 4+2+1 = 7 | FP = 1+2+4 = 7 | VN = 15+2+0+3+20+5+1+2+50 = 98

Letra B - VP = 15 | FN = 1+2+0 = 3 | FP = 4+3+1 = 8 | VN = 10+2+1+2+20+5+4+2+50 = 96

Letra C - VP = 20 | FN = 2+3+5 = 10 | FP = 2+2+2 = 6 | VN = 10+4+1+1+15+0+4+1+50 = 86

Letra D - VP = 50 | FN = 4+1+2 = 7 | FP = 1+0+5 = 6 | VN = 10+4+2+1+15+2+2+3+20 = 59

|   | Precisão | Recall | F1Score | TVP   | TFN   | TFP   | TVN   |
|---|----------|--------|---------|-------|-------|-------|-------|
| A | 0,588    | 0,588  | 0,588   | 0,588 | 0,411 | 0,066 | 0,933 |
| B | 0,652    | 0,833  | 0,731   | 0,833 | 0,166 | 0,076 | 0,923 |
| C | 0,769    | 0,666  | 0,713   | 0,666 | 0,333 | 0,065 | 0,934 |
| D | 0,892    | 0,877  | 0,884   | 0,877 | 0,122 | 0,092 | 0,907 |

### Questão 4

A métrica Gini é utilizada para a construção de árvores de decisão. O número informado pelo atributo Gini é usado para medir a "pureza" de um nó em uma árvore de decisão. Com ele, a gente pode perceber o quão misturadas estão as classes em um conjunto de dados. Quanto mais puras forem as classes em um nó (ou seja, mais homogêneas), menor será o valor do índice Gini. No caso do algoritmo CART, o mesmo usa o atributo de gini como escolha principal para escolher a melhor divisão em cada nó. Com isso, as divisões são analisadas, e a melhor é aquela que minimiza a média ponderada dos índices de gini dos nós filhos, reduzindo assim a incerteza dos dados, devido a criação de nós mais puros.

### Questão 5

1. Balanceamento é uma técnica usada para lidar com conjuntos de dados desbalanceados, onde uma ou mais classes possuem significativamente mais instâncias do que outras. A falta de balanceamento pode causar problemas em modelos de machine learning, que tendem a favorecer as classes majoritárias.

2. Dados ausentes são dados faltantes ou em branco dentro de uma base de dados. A existência dos dados ausentes pode ocorrer por vários motivos, como: problemas nos equipamentos que realizam a coleta, a transmissão e o armazenamento dos dados ou problemas no preenchimento ou na entrada dos dados por seres humanos.
3. Dados inconsistentes são aqueles que possuem valores conflitantes em seus atributos, apresentando erros ou discrepâncias do resultado esperado. Dados redundantes podem se referir tanto a instâncias quanto a atributos. Uma instância é redundante quando ela é muito semelhante a uma outra instância do mesmo conjunto de dados. Um atributo é redundante quando seu valor para todas as instâncias pode ser deduzido a partir do valor de um ou mais atributos.
4. Conversão simbólica-numérica é um processo que converte dados simbólicos(nominais/categóricos) em valores numéricos. Neste processamento, existem 2 casos: 1 - quando o atributo é do tipo simbólico e assume apenas dois valores, se os valores denotam a presença ou ausência de uma característica podemos utilizar um dígito binário e 2 - quando o atributo é do tipo simbólico e assume mais de dois valores, a técnica utilizada na conversão depende de o atributo ser nominal ou ordinal.
5. Conversão numérico-simbólica é um processo que converte dados numéricos em valores simbólico. Isso ocorre devido ao fato de que algumas técnicas de AM foram desenvolvidas para trabalhar com valores qualitativos, sendo assim, é melhor discretizar o atributo para fazer uma análise mais correta.
6. Transformação de um atributo numérico é um processo que geralmente ocorre quando o valor numérico de um atributo precisa ser transformado em outro valor numérico, pois é possível que os limites inferior e superior de valores dos atributos são muito diferentes, o que leva a uma grande variação de valores, ou ainda quando vários atributos estão em escalas diferentes. Dessa forma, a transformação consegue evitar que um atributo predomine o outro, melhorando a performance dos modelos de machine learning.
7. Redução de dimensionalidade é utilizada para diminuir o número de variáveis de entrada em um modelo, combinando ou eliminando parte dos atributos irrelevantes. É possível realizar a redução de duas formas principais: 1 - Agregação, em que substituem os atributos originais por novos atributos formados pela combinação de grupos de atributos, porém levam à perda dos valores originais dos atributos, o que pode ser importante dependendo do contexto e 2 - Seleção de atributos, em que mantem uma parte dos atributos originais e descartam os demais atributos.