**Assignment to Introduction to Data Science**

Deadline: Tuesday, 2 May 2023

**Question 1: KNN Classification**

Vert Tige is a botanical laboratory in Cameroon. It manages many flower species and would like to acquire a software for intelligent classification of flower species from the following historical database:

|  |  |  |
| --- | --- | --- |
| **Sepal length** | **Sepal width** | **Species** |
| 5.3 | 3.7 | Setosa |
| 5.1 | 3.8 | Setosa |
| 7.2 | 3.0 | Virginica |
| 5.4 | 3.4 | Setosa |
| 5.1 | 3.3 | Setosa |
| 5.4 | 3.9 | Setosa |
| 7.4 | 2.8 | Virginica |
| 6.1 | 2.8 | Verscicolor |
| 7.3 | 2.9 | Virginica |
| 6.0 | 2.7 | Verscicolor |
| 5.8 | 2.8 | Virginica |
| 6.3 | 2.3 | Verscicolor |
| 5.1 | 2.5 | Verscicolor |
| 6.3 | 2.5 | Verscicolor |
| 5.5 | 2.4 | Verscicolor |
| 5.2 | 2.4 | ? |

Using the K Nearest Neighbour Algorithm (KNN), predict the species of the Heart of Mary flower that has 5.2 and 2.4 sepal length and width respectively if the 5 nearest neighbours are considered (K=5)

**Question 2: Decision Tree Algorithm / 10 Marks**

Build a Decision Tree of the following dataset. The set consists of 10 instances of weekends and 3 attributes: Weather, Parents, and Money. The decision is the output variable.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Weekend** | **Weather** | **Parents** | **Money** | **Decision** |
| W1 | Sunny | Yes | Rich | Cinema |
| W2 | Sunny | No | Rich | Tennis |
| W3 | Windy | Yes | Rich | Cinema |
| W4 | Rainy | Yes | Poor | Cinema |
| W5 | Rainy | No | Rich | Stay In |
| W6 | Rainy | Yes | Poor | Cinema |
| W7 | Windy | No | Poor | Cinema |
| W8 | Windy | No | Rich | Shopping |
| W8 | Windy | Yes | Rich | Cinema |
| W10 | Sunny | No | Rich | Tennis |