**AISHWARYA S**

**18BCS029 CSE B**

**U18CSI6203T – DATA WAREEHOUSING AND DATA MINING**

**MODEL LAB: 10.04.2021**

**SET 4**

1. **Download a suitable dataset for classification from any Repository. List the attributes and its type in a word Doc.**

*For the ‘iris’ dataset which is already available in seaborn package, the following operations were performed to obtain the attributes and its types.*

| **sepal\_length numerical**  **sepal\_width numerical**  **petal\_length numerical**  **petal width numerical**  **species catagorical** |
| --- |

***df = sns.load\_dataset('iris')***

***df.head()***

| **sepal\_length** | **sepal\_width** | **petal\_length** | **petal\_width** | **species** |
| --- | --- | --- | --- | --- |
| **0** | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| **1** | 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| **2** | 4.7 | 3.2 | 1.3 | 0.2 | Setosa |
| **3** | 4.6 | 3.1 | 1.5 | 0.2 | Setosa |
| **4** | 5.0 | 3.6 | 1.4 | 0.2 | Setosa |

***df.info()***

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 150 entries, 0 to 149

Data columns (total 5 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 sepal\_length 150 non-null float64

1 sepal\_width 150 non-null float64

2 petal\_length 150 non-null float64

3 petal\_width 150 non-null float64

4 species 150 non-null object

dtypes: float64(4), object(1)

memory usage: 6.0+ KB

***df.isnull().any()***

sepal\_length False

sepal\_width False

petal\_length False

petal\_width False

species False

dtype: bool

***df.shape***

(150, 5)

**2. Load the dataset and set the target and feature variables. Split the dataset into training and test dataset. Build decision tree classifier with Entropy criteria. Perform Prediction for test dataset using Entropy and print the results in the form of confusion matrix, accuracy and classification report. visualize the decision tree**.

**3. Upload in your github account. Provide the link for access.**