**Department of Computer Science & Engineering**

**Final Year B. Tech. (CSE) – I: 2022-23**

**5CS462: PE5 - Data Mining Lab**

**Assignment No. 4**

**Name: Sanket Mote**

**PRN: 2019BTECS00113**

**Batch: B8**

**Title:** Design the data analysis tools (GUI) to perform the following pre-processing task.

1. Design the rule-based classifier: Extract the rules from decision tree build in assignment no. 3.

2. Tabulate the results and evaluate the performance of rules generated using following metrics: a. Coverage b. Accuracy c. Toughness (size)

3. Use the following categorical data sets from UCI machine learning repository: a. Balance Scale data set b. Car evaluation data set c. Breast-cancer data set

**Objective/Aim:**

1. To implement data analysis tool using python programming language.
2. To design the rule-based classifier.

**Introduction:**

Decision Trees are supervised [machine learning algorithms](https://www.analyticssteps.com/blogs/top-10-machine-learning-algorithms) that are best suited for classification and regression problems. These algorithms are constructed by implementing the particular splitting conditions at each node, breaking down the training data into subsets of output variables of the same class.

**Theory:**

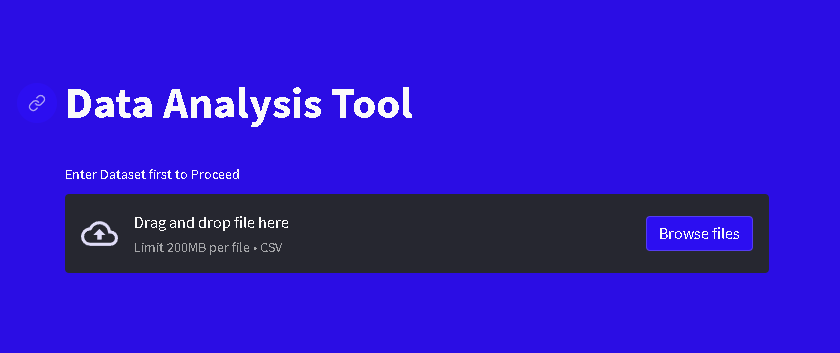
Decision Tree Classifier:

A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.

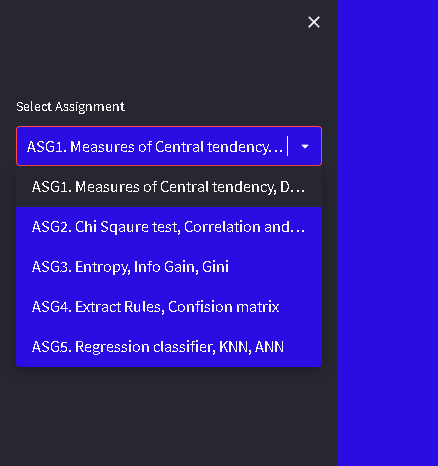
**Rule-based classifiers :**Rule-based classifiers are just another type of classifier which makes the class decision depending by using various “if..else” rules. These rules are easily interpretative and thus these classifiers are generally used to generate descriptive models. The condition used with “if” is called the **antecedent** and the predicted class of each rule is called the **consequent**.

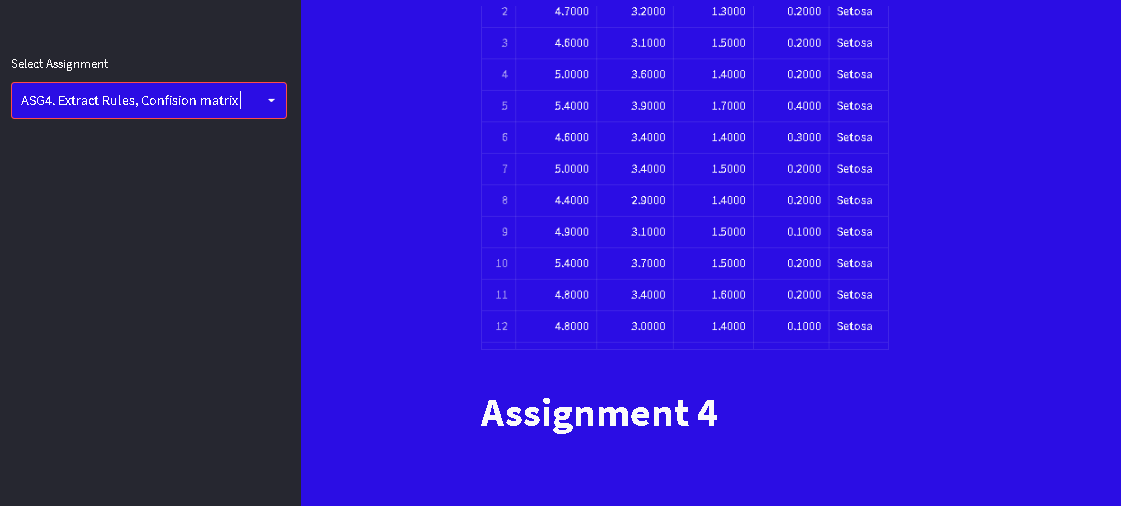
**Procedure:**

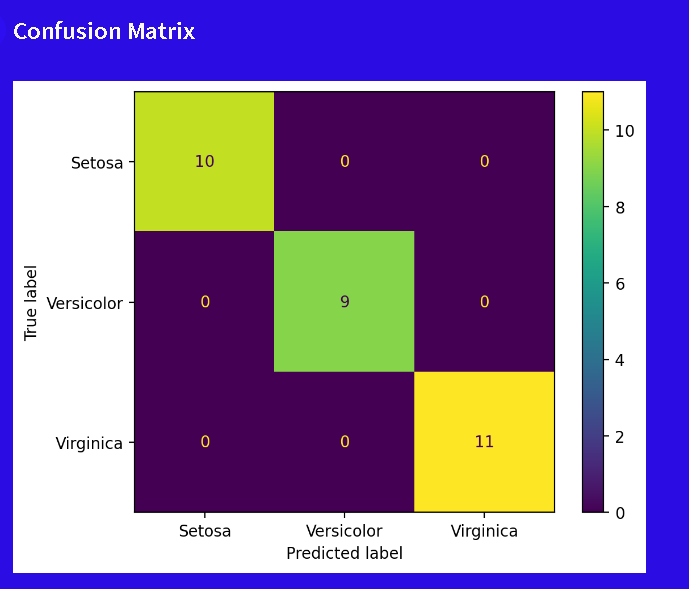
Given problem statement is solved using python programming language and specifically used streamlit module to implement GUI application and pandas module to load .csv file as dataset.

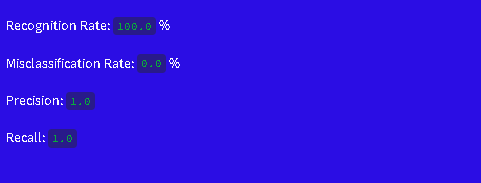












**Conclusion:**

Successfully implemented data analysis tool (GUI) for decision tree and rule based classifier for selected attribute of given dataset.