St. Francis Institute of Technology, Mumbai-400 103

**Department Of Information Technology**

A.Y. 2024-2025

Class: TE-ITA/B, Semester: VI

Subject: **Business Intelligence Lab**

**Experiment – 4: To implement a classifier- Decision tree using open source tool WEKA and ORANGE**

1. **Aim:** To Implement any one of the classifiers using WEKA (Decision Tree, Naïve Bayes, Random Forest)
2. **Objectives:** After study of this experiment, the students will be able to

Understand and knew about all the three classifiers.

1. **Outcomes:** After study of this experiment, the students will be able to

CO4:  Design and Implement the appropriate data mining methods like classification, clustering or Frequent Pattern mining on large data sets.

CO5:  Define and apply metrics to measure the performance of various data mining algorithms

1. **Prerequisite:** Introduction to all the three classifiers through algorithms & Problem solving approach.
2. **Requirements:** Personal Computer, Windows XP operating system/Windows 7, Internet

Connection, Microsoft Word, WEKA tool, Java/R/Python.

1. **Theory:**
   1. **What is Classification Data Mining?**

**ANS:**

1. Classification is a supervised learning technique in data mining that categorizes data intopredefined labels based on input features.
2. It involves training a model on labeled data, testing its performance, and using it to predict the class of new, unseen data.
3. Common applications include spam detection, medical diagnosis, fraud detection, credit risk assessment, and sentiment analysis.
4. Classification is a type of supervised learning, meaning it requires labeled datasets where the correct output is already known.
5. Popular classification algorithms include Decision Trees, Naïve Bayes, k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Neural Networks.
6. Performance is evaluated using metrics like accuracy, precision, recall, F1-score, and a confusion matrix to measure effectiveness.
7. Challenges include handling imbalanced datasets, avoiding overfitting, managing noisy data, and selecting the right algorithm for different problems.
8. Unlike clustering, which groups unlabeled data, classification works with labeled data and assigns new inputs to predefined categories.
   1. **Difference between supervised and unsupervised learning**

**ANS:**

| **SUPERVISED** | **UNSUPERVISED** |
| --- | --- |
| Uses labeled data to train models. | Works with unlabeled data to find patterns. |
| Predict known categories or values. | Discover hidden structures in data. |
| Techniques include Classification and Regression. | Techniques include Clustering and Association. |
| Produces specific class labels or values. | Groups data into clusters without predefined labels. |
| Examples: Spam detection, medical diagnosis. | Examples: Customer segmentation, anomaly detection. |
| Algorithms include Decision Trees, SVM, Neural Networks. | Algorithms include K-Means, DBSCAN, PCA. |

1. **Laboratory Exercise:** Implementation of Classification Algorithm in WEKA and Orange. Take printout of related snapshots.
2. **Post-Experiments Exercise**
   1. **Extended Theory:**
      * Explain about Decision Tree algorithm
      * Solve numerical on decision tree
   2. **Exercise:**
      * Simple CLI execution of classification algorithm in WEKA

**For training:** java weka.classifiers.trees.J48 -C 0.25 -M 2 -t directory-path\bank.arff -d directory-path \bank.model

**For Testing:** java weka.classifiers.trees.J48 -p 9 -l directory-path\bank.model -T directory-path \bank-new.arff

* 1. **Conclusion:**
     + Summary of Experiment
     + Importance of Experiment
     + Application of Experiment

| The training dataset : | Loading the training dataset in weka : |
| --- | --- |
| Applying J48 decision tree algorithm on training dataset : | Decision tree model generation in WEKA : |

| Viewing in separate window : | visualizing the tree : |
| --- | --- |
| Test Dataset : |  |

| After applying the test dataset : | visualise the classifier errors |
| --- | --- |
| Predicted pep values by the decision tree model : |  |

| **Exercise:**  Simple CLI execution of classification algorithm in WEKA  **For training:** java weka.classifiers.trees.J48 -C 0.25 -M 2 -t directory-path\bank.arff -d directory-path \bank.model | **For Testing:** java weka.classifiers.trees.J48 -p 9 -l directory-path\bank.model -T directory-path \bank-new.arff |
| --- | --- |
| Applying Decision tree algorithm in Orange : | Loading the iris dataset : |

| Iris dataset tuples view in the data table : | Applying tree algorithm on iris dataset : |
| --- | --- |
| Tree visible in the tree viewer : | tuples visible in the data table based on the selection in data tree : |

scatter plot generated based on the selection in data tree :

