**RAJAGIRI SCHOOL OF ENGINEERING AND TECHNOLOGY**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**COMPUTER SCIENCE & BUSINESS SYSTEMS**

**SEMESTER V (2021 Admission)**

**101009/IT522S MACHINE LEARNING LAB**

**LAB CYCLE**

**Explore WEKA Machine Learning Toolkit**

Experiment No.1:

• Downloading and/or installation of WEKA data mining toolkit,

• Understand the features of WEKA toolkit such as Explorer, Knowledge Flow interface, Experimenter, command-line interface.

• Navigate the options available in the WEKA (ex. Select attributes panel, Preprocess panel, Classify panel, Cluster panel, Associate panel and Visualize panel)

Experiment No.2:

• Study the arff file format

• Explore the available data sets in WEKA

• Load a dataset (breast cancer)and observe the following:

a. List the attribute names and their types

b. Number of records in each dataset

c. Identify the class attribute (if any)

d. Perform necessary pre-processing (minimum 3)

d. Plot Histogram

**Explore Graphical Plots in R.**

Experiment No.3: Create a Scatter plot with the iris dataset using the ggplot package. Add legends, lines, and labels, and use the aes function in the plot.

Experiment No.4: Create a histogram with the titanic dataset using the ggplot package. Add legends, lines, and labels, and use the aes function in the plot.

**Classification**

Experiment No.5: Demonstrate pre-processing in soyabean dataset using R, with minimum of four preprocessing to be done, and prepare it for classification.

Experiment No.6: Demonstrate the following Classification algorithms using some public domain datasets in UCI ML repository and compute the accuracy of the classifier, considering few test datasets.

i. Naive Bayes classification

ii. Decision Tree

iii. Random Forests

iv. Ensembles of classifiers including bagging and boosting

Experiment 6: Demonstrate the following Classification algorithms using some public domain datasets in UCI ML repository and compute the accuracy of the classifier, considering few test datasets.

i. k-Nearest neighbor classification

ii. Support Vector Machines

**Micro Project:**

1. Implementation of clustering algorithms

2. Implementation of association rule mining algorithm (Apriori)

3. Implementation of anomaly detection algorithms

4. Implementation of EM algorithm for some specific problem

Theory (to be written in the record).

1. Details on the techniques used.

2. The domain selected.

3. Aim of the project.

4. Motivation behind the project.

5. Reason to choose the technique.

6. Pre-processing steps used.

7. Performance measures used, and accuracy obtained.

8. Inference from the implementation.

Output should include:

1. Code (Print)
2. Sample outputs of pre-processing, analysis and prediction.

3. Sample of performance evaluation.

4. Graphs, if any.

Faculty in-charges HoD

Ancy C A Dr Neeba E A

Tinku Soman Jacob

Viji Mohan

Jeshmol P J