Shri Ramdeobaba College of Engineering and Management, Nagpur
Department of Computer Science and Engineering
Session: 2023-2024

Data Visualization and Analytics Lab   **VII Semester**

PRACTICAL NO. 1

Aim: (A) Introduction to Weka tool.
(B) Performing data understanding and preprocessing on the given data set in Weka.

Theory Questions:
1. What options are available on main panel?
2. What is the purpose of the the following in Weka:
     1. The Explorer
     2. The Knowledge Flow interface
     3. The Experimenter
     4. The command-line interface
3. Describe the arff file format.
4. What is the purpose of the following in the Explorer Panel?
     1. The Preprocess panel
          1. What are the main sections of the Preprocess panel?
          2. What are the primary sources of data in Weka?
     2. The Classify panel
     3. The Cluster panel
     4. The Associate panel
     5. The Select Attributes panel
     6. The Visualize panel.

Dataset: Weather.nominal.arff and
sick.arff (https://datahub.io/machine-learning/sick#resource-sick_arff)

Questions:
1. Press the Explorer button on the main panel and load the **weather dataset** and answer the following questions
     1. How many instances are there in the dataset?
     2. State the names of the attributes along with their types and values.
     3. What is the class attribute?
     4. How will you determine how many instances of each class are present in the data
     5. What happens with the Visualize All button is pressed?
     6. How will you view the instances in the dataset? How will you save the changes?
     7. Now, extend the dataset to include 50 instances in total.

2. Do as directed to apply Filter
     1. Use the unsupervised filter RemoveWithValues to remove all instances where the attribute 'humidity' has the value 'high'? Undo the effect of the filter.
     2. Remove the 'FALSE' instances of windy attribute and undo the effect.

3. Remove the attribute outlook and undo the effect.
4. Experiment with different filters and report their effects.

3. Application of Discretization Filters [use sick.arff dataset]
    1. Load the 'sick.arff' dataset.
    2. Apply the supervised discretization filter on different attributes.
    3. What is the effect of this filter on the attributes?
    4. How many distinct ranges have been created for each attribute?
    5. Undo the filter applied in the previous step.
    6. Apply the unsupervised discretization filter. [Use equal-width binning approach]
        1. In this step, set 'bins'=5
        2. In this step, set 'bins'=10
        3. What is the effect of the unsupervised filter on the dataset?
    7. Run the the Naive Bayes classifier after apply the following filters
        1. Unsupervised discretized with 'bins'=5
        2. Unsupervised discretized with 'bins'=10
        3. Unsupervised discretized with 'bins"=20.
    8. Compare the accuracy of the following cases
        1. Naive Bayes without discretization filters
        2. Naive Bayes with a supervised discretization filter
        3. Naive Bayes with an unsupervised discretization filter with different values for the 'bins attributes.
    9. Repeat steps 6 to 8 using equal-frequency binning approach and present your conclusion.