**Practical 7**

|  |
| --- |
| **Use WEKA to implement classification** |
|  |

**Title**

Introduction to **Use WEKA to implement classification part1**

**Aim**

To learn to **Use WEKA to implement classification**

**References**

* http://people.sabanciuniv.edu/berrin/cs512/lectures/WEKA/WEKA%20Explorer%20Tutorial-REFERENCE.pdf
* Witten, Ian and Eibe, Frank. *Data Mining: Practical Machine Learning Tools and Techniques.* Springer.
* weks dataset: <https://storm.cis.fordham.edu/~gweiss/data-mining/datasets.html>
* https://www.cs.ubc.ca/labs/beta/Projects/autoweka/datasets/

**Task:**

**(Attach screen shots of output)**

1. Load the iris 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. In the histogram on the bottom-right, which attributes are plotted on the X,Y-axes? How do you change the attributes plotted on the X,Y-axes?
   5. How will you determine how many instances of each class are present in the data
   6. What happens with the Visualize All button is pressed?
2. Load the weather dataset and perform the following tasks:
   1. Use the unsupervised filter Remove with Values to remove all instances where the attribute ‘humidity’ has the value ‘high’?
   2. Undo the effect of the filter.
   3. Answer the following questions:
      1. What is meant by filtering in Weka?
      2. Which panel is used for filtering a dataset?
      3. What are the two main types of filters in Weka?
      4. What is the difference between the two types of filters? What is the difference between and attribute filter and an instance filter?

4 Load the iris dataset and perform the following tasks:

1. Press the Visualize tab to view the Visualizer panel.
2. What is the purpose of the Visualizer?
3. Select one panel in the Visualizer and experiment with the buttons on the panel.

**Post lab:**

1. Give the steps to convert .xls file format into arff file format
2. Apply decision tree induction on following data. Give each step output with formulas.

compare ur result with weka result and comment on it.

@relation weather

@attribute outlook {sunny, overcast, rainy}

@attribute temperature real

@attribute humidity real

@attribute windy {TRUE, FALSE}

@attribute play {yes, no}

@data

sunny,85,85,FALSE,no

sunny,80,90,TRUE,no

overcast,83,86,FALSE,yes

rainy,70,96,FALSE,yes

rainy,68,80,FALSE,yes

rainy,65,70,TRUE,no

overcast,64,65,TRUE,yes

sunny,72,95,FALSE,no

sunny,69,70,FALSE,yes

rainy,75,80,FALSE,yes

sunny,75,70,TRUE,yes

overcast,72,90,TRUE,yes

overcast,81,75,FALSE,yes

rainy,71,91,TRUE,no