Experiment – 1

Aim: Loading and Exploring Data in WEKA

Introduction:

WEKA (Waikato Environment for Knowledge Analysis) is a popular open-source tool used for data mining and machine learning. It provides a graphical interface through which users can preprocess data, apply classification or clustering algorithms, and visualize results without the need for programming. One of the first steps in any data mining process is to load and explore the dataset. The Preprocess tab in WEKA allows users to load data in ARFF or CSV format, view attributes, check data types (nominal or numeric), observe data distribution, and identify missing or abnormal values. Understanding the structure and distribution of the dataset is essential before applying any data mining technique.

Tools and Technologies Used:

- Software: WEKA Explorer (version 3.8 or higher)
- Supported File Formats: .arff (preferred), .csv
- Operating System: Windows/Linux

Procedure:

1. Open WEKA and select **Explorer** from the GUI Chooser.
2. In the **Preprocess** tab, click on **"Open file…"**.
3. Load a dataset such as weather.arff or iris.arff from the data folder.
4. View the dataset details: relation name, number of attributes, and instances.
5. Click on each attribute to check its type (nominal or numeric).
6. Observe attribute statistics like min, max, mean, standard deviation, or category counts.
7. Check the graphical visualization of each attribute shown below.

Observation:

The dataset `weather.arff` was successfully loaded in WEKA. It contains 5 attributes and 14 instances. "Outlook" is a nominal attribute with three categories: sunny, overcast, and rainy. "Humidity" is a numeric attribute with values ranging between 65 and 95. The graphical view displayed the frequency and distribution of each attribute clearly.

Output/Results:

The experiment was completed successfully. The dataset was loaded in WEKA using the Explorer's Preprocess tab, and all attributes were explored with their respective types, statistics, and visualizations.

Experiment – 2

Aim: Data Preprocessing in WEKA

Tools and Technologies Used:

- Software: WEKA (version 3.8 or above)
- Dataset Format: .arff (preferred) or .csv
- System Requirements: Windows/Linux with Java Runtime Environment
- Sample Datasets: weather.arff, iris.arff, or any other sample dataset

Procedure:

1. Open **WEKA** and click on **Explorer** from the GUI Chooser.
2. Go to the **Preprocess** tab and click **"Open file…"** to load a dataset.
3. Identify any **unwanted attributes** and select them from the list.
4. Click on the **"Remove"** button to delete the selected attribute.
5. To handle missing or unclean data, click on the **"Choose"** button in the Filter section.
6. Select filters like:
    - ReplaceMissingValues (for missing data)
    - Normalize (for scaling values)
    - Standardize (for mean-centering)
7. After selecting a filter, click **"Apply"** to preprocess the data.
8. Observe the changes in attribute values after applying filters.

Observation:

The dataset weather.arff was loaded in WEKA. Preprocessing was performed using filters like ReplaceMissingValues and Normalize. Missing values were replaced with suitable values, and numeric attributes were scaled to a standard range. The dataset was cleaned and standardized for further data mining operations.

Result:

The experiment was completed successfully. Using WEKA, basic data preprocessing techniques such as handling missing values, normalization, and attribute removal were applied. The dataset was made consistent and ready for the application of machine learning algorithms.

Experiment – 3

Aim: Data Visualization in WEKA

Tools Required:

- WEKA Software
- Sample dataset (iris.arff, weather.arff)

Procedure:

1. Open WEKA and select Explorer from the GUI Chooser.
2. In the Preprocess tab, load a dataset using the “Open file…” option.
3. Click on any attribute in the list to view its graphical distribution.
4. For nominal attributes, observe the bar chart showing category frequencies.
5. For numeric attributes, observe the histogram showing value ranges.
6. Now switch to the Visualize tab at the top.
7. In the Visualize tab, study the scatter plots comparing pairs of attributes.
8. Use the plots to understand attribute relationships and data spread.

Observation:

Nominal attributes like “Outlook” showed clear category frequencies through bar charts. Numeric attributes like “Humidity” showed a well-distributed histogram. The Visualize tab displayed scatter plots for comparing two attributes.

Result/Output:

Data visualization was successfully done using both the Preprocess and Visualize tabs in WEKA, helping to better understand data distribution.