Name: Diya Nikesh Jain

Class: TY9-A-21

DATA WAREHOUSE AND MINING

Experiment No. 8

Title: Demonstrate Classification, Clustering, Association using WEKA.

Aim: Perform data Pre-processing task and demonstrate Classification, Clustering, Association algorithm on data sets using data mining tool WEKA.

Introduction: Data mining is the process of extracting useful patterns from large datasets. WEKA is a powerful open-source tool that supports various data mining techniques through an easy-to-use interface. In this experiment, we use WEKA to demonstrate three key tasks:

- **Classification:** Predicting predefined class labels (e.g., spam detection).
- **Clustering:** Grouping similar data without prior labels.
- **Association:** Finding relationships between items (e.g., market basket analysis).

Before applying these algorithms, data preprocessing is done to clean and prepare the data for better accuracy.

Procedure: Steps to Perform Clustering in Weka Knowledge Flow:

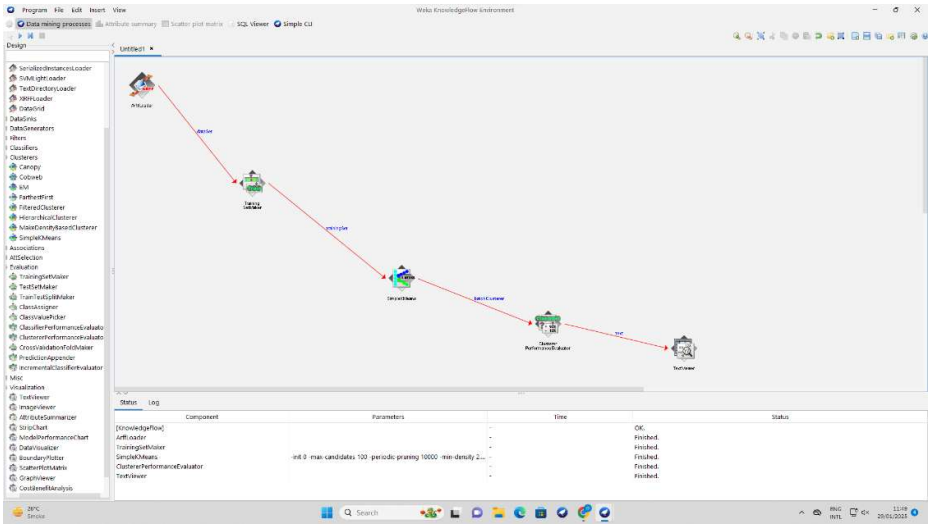
1. **Launch Weka Knowledge Flow**
    - Open *Weka 3.6* from the Program Files on your computer.
    - From the startup menu (Explorer, Experimenter, Knowledge Flow, etc.), select **Knowledge Flow**.
2. **Load Dataset with ArffLoader**
    - In the canvas, drag the **ArffLoader** component from the "Data Sources" section.

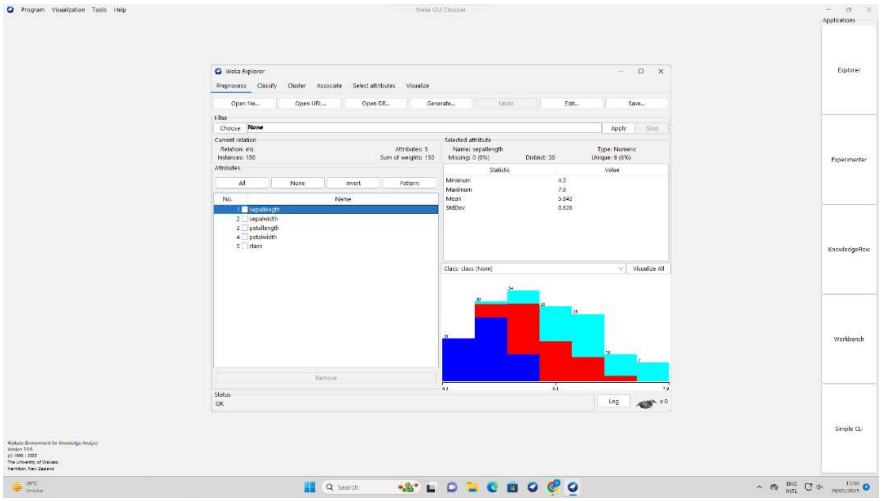
- Right-click on it → select **Configure**, then click **Browse** to locate and load a dataset (e.g., iris.arff from the Data folder).
- This will import your data into the flow.

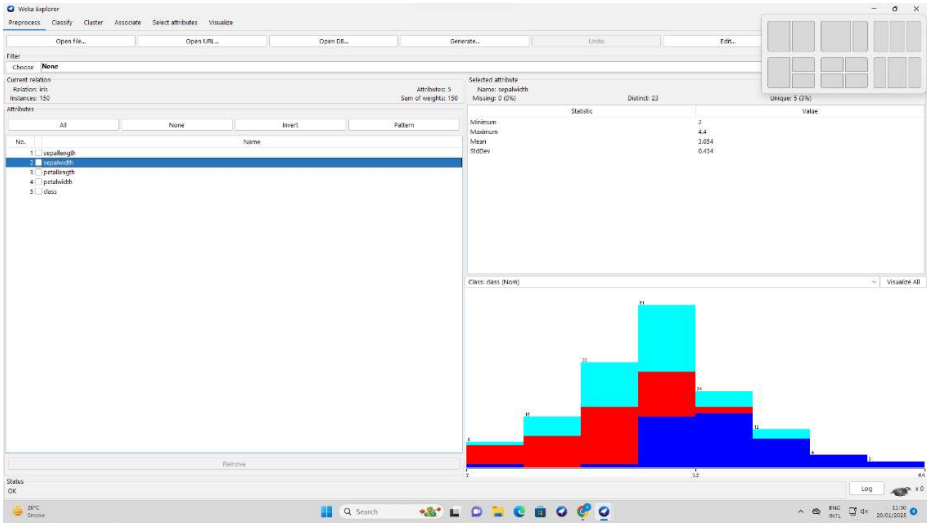
3. **Set Up Evaluation Component**
    - Add the **Evaluation** component to your workflow to evaluate the clustering output.
    - Set the **evaluation type** to **Static**, which uses the dataset as it is.
4. **Prepare Data for Training**
    - Drag the **TrainingSetMaker** component onto the canvas.
    - This prepares the data format for training purposes.
    - Connect it to the **output** of the ArffLoader.
5. **Add and Configure the Clusterer**
    - Insert the **Clusterer** component into your workspace.
    - Select **SimpleKMeans** as the clustering algorithm.
    - Configure the algorithm by setting parameters like the number of clusters and distance function.
6. **Evaluate Clustering Performance**
    - Add the **ClustererPerformanceEvaluator** component.
    - Link it to the output of the Clusterer to assess the model's performance.
7. **Set Up Output Displays**
    - Drag in a **TextViewer** component to display textual results like cluster assignments and summary statistics.
    - Add a **Visualization** component to see graphical representations of the cluster distributions.
8. **Connect Components and Execute Workflow**

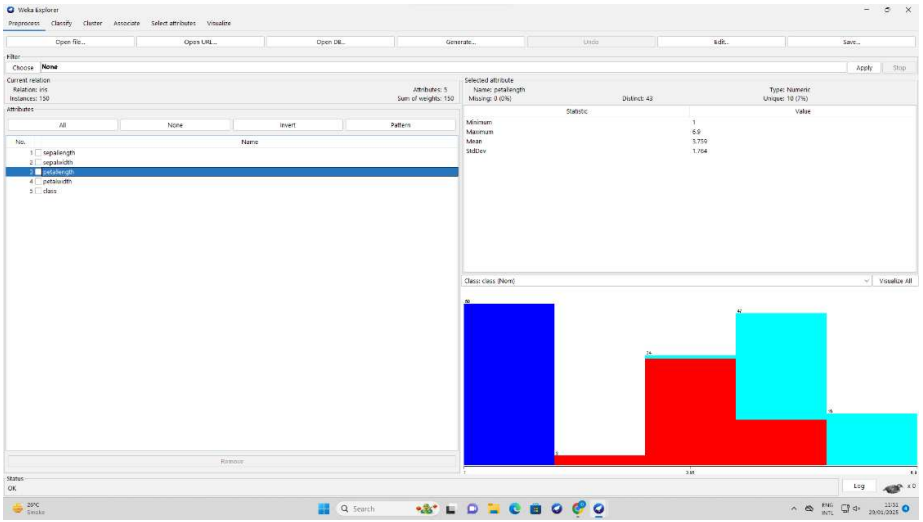
- Right-click each component and create the following connections: ArffLoader → TrainingSetMaker → Clusterer → ClustererPerformanceEvaluator → TextViewer / Visualization
- Once everything is connected, right-click on the final component and select **Start Execution** to run the entire process.

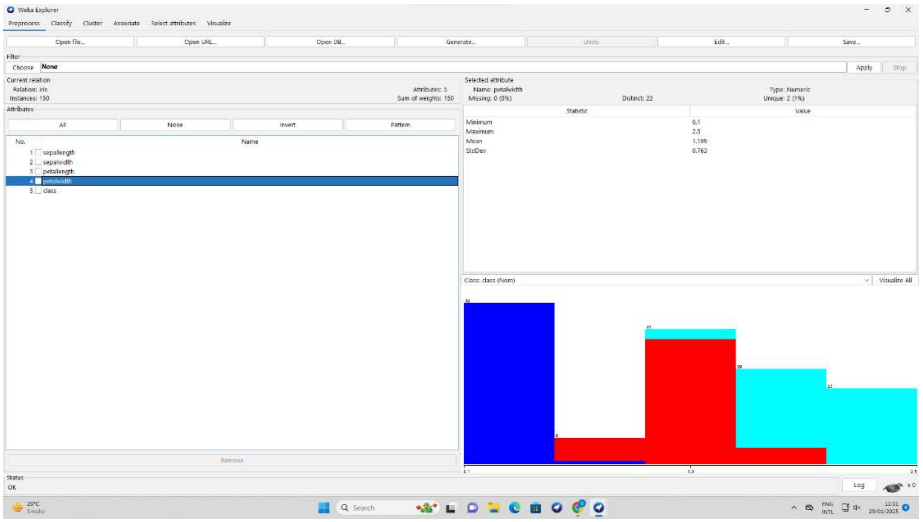
Implementation/Outputs:

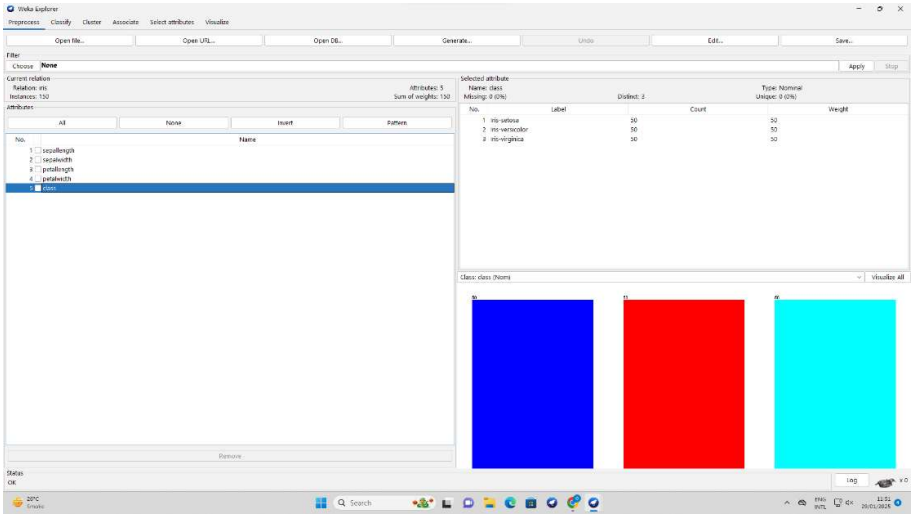


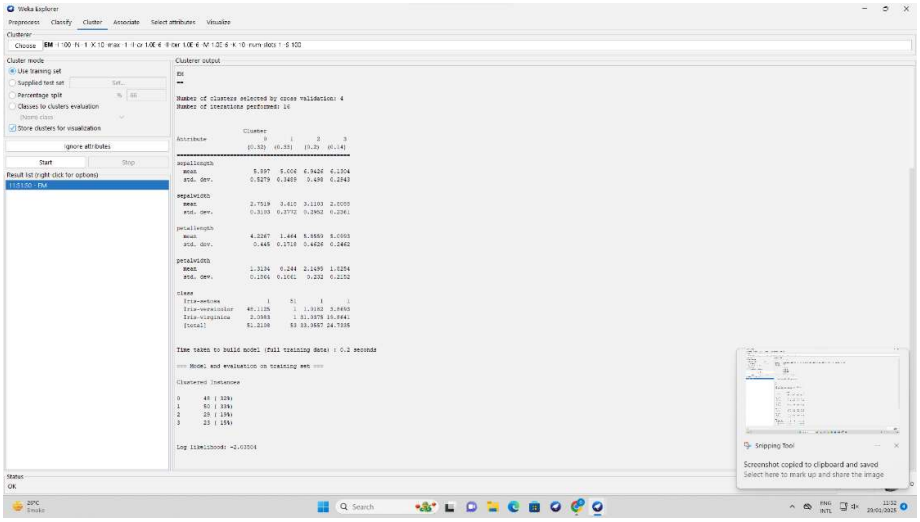












Conclusion: We effectively demonstrated data preprocessing and implemented essential data mining techniques—Classification, Clustering, and Association—using the WEKA tool. Thanks to WEKA's user-friendly interface and integrated algorithms, we were able to seamlessly load datasets, configure models, and visualize outcomes. This hands-on experience provided valuable insights into classifying data, grouping similar instances into clusters, and uncovering hidden patterns and associations—skills that are crucial for real-world data analysis and informed decision-making.

GitHub Link: https://github.com/diyaaaa294/DWM-experiments/tree/main