Program 1 : 17/7/25 **Practice 1: Loading and Exploring Data in WEKA**

Step 1: Launch WEKA

- Open the **WEKA application**.

- From the **WEKA GUI Chooser**, select **"Explorer"**.

◆ **Step 2: Open a Dataset**

- You'll land in the **Preprocess** tab by default.

- Click the **"Open file..."** button.

- Select a dataset from your system:

    o .arff (Attribute-Relation File Format) – preferred by WEKA.

    o .csv – Comma-separated file (make sure it's properly formatted).

- Example datasets (bundled with WEKA):

    o weather.arff

    o iris.arff

    o contact-lenses.arff

📌 **Tip**: Sample datasets are usually found in:
C:\Program Files\Weka-3-8\data

◆ **Step 3: View Dataset Summary**

Once loaded:

- The **relation name** is shown at the top.

- The **attributes (features)** are listed on the left.

- For each selected attribute, you can see:

    o **Type**: Nominal or Numeric

    o **Distinct Values**

    o **Missing Values**

    o **Histogram** showing distribution.

◆ **Step 4: Explore Individual Attributes**

- Click on an attribute in the left panel.
- You'll see a **histogram** or **bar chart** for that attribute.
    - **Nominal attributes**: Show frequency of each category.
    - **Numeric attributes**: Show distribution in ranges.
- Observe class distribution using the color legend.

◆ Step 5: Basic Data Operations

Use these options for preprocessing:

- **Remove**: Delete unwanted attributes.
- **Rename**: Change attribute names.
- **Edit**: Launches a spreadsheet-style editor to change individual records.

◆ Step 6: Apply Filters (Optional)

- Use **filters** to preprocess your data.
    - Click on **"Choose"** under the Filter section.
    - Examples:
        - unsupervised.attribute.Remove: Remove specific attributes.
        - supervised.attribute.Discretize: Convert numeric to nominal.
        - unsupervised.instance.RemoveWithValues: Remove rows with specific values.

After choosing a filter, click **Apply** to process the data.

◆ Step 7: Save the Processed Data (Optional)

- Click **"Save"** to store the cleaned or edited data.
- You can save it in:
    - .arff (default)
    - .csv (optional)

Pg 2: 24/7/25 Practice 2:Data Preprocessing in WEKA

✅ **Step 1: Open WEKA Explorer**

- Launch **WEKA GUI Chooser**
- Click **"Explorer"**

✅ **Step 2: Load the Dataset**

- Go to the **Preprocess** tab
- Click **"Open file..."**
- Load an .arff or .csv file
    - Example: weather.arff or iris.arff (found in the data/ folder of WEKA)

✅ **Step 3: Explore the Data**

- See **Relation name** and **Number of instances**
- View the list of **attributes (features)**
- Click an attribute to see:
    - Type (Nominal/Numeric)
    - Distinct values
    - Mean, StdDev (for numeric)
    - Histogram (data distribution)

✅ **Step 4: Apply Filters (Preprocessing Tasks)**

◆ **A. Remove an Attribute**

1. Click **"Choose"** → unsupervised.attribute.Remove
2. Click the filter name to configure it
3. Enter index of attribute to remove (e.g., 2)
4. Click **"Apply"**

◆ **B. Replace Missing Values**

1. Click **"Choose"** → unsupervised.attribute.ReplaceMissingValues
2. Click **"Apply"**

◆ **C. Normalize Data (0–1 range)**

1. Click **"Choose"** → unsupervised.attribute.Normalize
2. Apply to scale all numeric attributes

◆ **D. Discretize Numeric Attributes**

(Useful when converting numeric data to nominal classes)

1. Choose unsupervised.attribute.Discretize
2. Click **"Apply"**

◆ **E. Convert String to Nominal**

1. Choose unsupervised.attribute.StringToNominal
2. Set attribute index (e.g., first or 1)
3. Click **"Apply"**

✅ **Step 5: Save the Preprocessed Data**

- Click **"Save"**
- Choose .arff or .csv format
- Example: weather_cleaned.arff
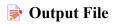

✅ **Step 6: Optional – Visualize Data**

- Click **"Visualize All"** button
- Check scatter plots of attribute relationships

Additional:

Objective: Clean and prepare weather.arff dataset.

Steps:

1. Open weather.arff
2. Remove the humidity attribute
3. Replace any missing values
4. Normalize numeric data
5. Save the final dataset as weather_preprocessed.arff

📝 **Output File**

- Contains cleaned, normalized data ready for modeling.

- Can now be used in the **Classify** or **Cluster** tabs.

Pg 3: 31/7/25 : Data Visualization in WEKA

Step 1: Open WEKA

- Launch **WEKA** from your system.

- Choose **"Explorer"** from the GUI Chooser window.

Step 2: Load a Dataset

- Click on the **"Open file…"** button under the **Preprocess** tab.

- Select a dataset (e.g., weather.arff, iris.arff, or any .csv/.arff file).

- Once loaded, you will see the attributes listed on the left and summary statistics on the right.

Step 3: Explore Attribute-Wise Visualization

- Click on any attribute name from the list on the left.

- WEKA displays a **histogram** for that attribute.

  o **Nominal attributes**: Bar chart showing class distribution.

  o **Numeric attributes**: Distribution graph.

- Class labels are usually color-coded.

Step 4: Use the "Visualize All" Option

- At the bottom of the Preprocess tab, click the **"Visualize All"** button.

- A new window opens showing **scatter plots** for each attribute pair.

  o Each point is an instance.

  o Colors represent different class values.

Step 5: Customize Visualization

- In the Visualize panel:

  o Click on **any scatter plot** to enlarge it.

  o You can set:

- **X-axis and Y-axis** attributes.

- **Coloring by class**.

  o Use zoom and drag for better inspection.

  o Right-click to save the plot as an image.

Step 6: Use Filters for Better Visualization (Optional)

- Back in the Preprocess tab, click **Choose** (in the Filter section).

- Try filters like:

  o **unsupervised.attribute.Remove** – remove unwanted attributes.

  o **supervised.attribute.Discretize** – convert numeric to nominal for better visual clarity.

- Apply the filter and recheck visualization.

Step 7: Save Modified Data (Optional)

- After filtering or preprocessing, click **"Save"** to store the modified dataset for future analysis.

**Additional**

Task: Load the iris.arff dataset and visualize sepal and petal dimensions.

Instructions:

1. Load iris.arff.

2. Go to **Visualize All**.

3. Select scatter plot with:

   o X-axis: petalwidth

   o Y-axis: sepallength

   o Coloring: by class

4. Observe cluster formation.

Association Rule Mining in WEKA – Step-by-Step Procedure

◆ **Step 1: Launch WEKA**

- Open WEKA.

- From the **GUI Chooser**, click **Explorer**.

◆ **Step 2: Load a Dataset**

- Go to the **Preprocess** tab.

- Click **"Open file..."**.

- Select an **ARFF** or **CSV** file suitable for association mining.

  o Use sample datasets like:

    ▪ weather.nominal.arff

    ▪ supermarket.arff

    ▪ Or create your own market basket .arff file with **nominal** values.

◆ **Step 3: Go to Associate Tab**

- Click the **"Associate"** tab.

- This tab is used for discovering **association rules** from data.

◆ **Step 4: Choose Algorithm**

- Click on the **"Choose"** button.

- Select **weka.associations.Apriori** algorithm.

◆ **Step 5: Configure Apriori Parameters (Optional)**

Click on **Apriori** to modify settings such as:

- LowerBoundMinSupport: Minimum support (default: 0.1)

- minMetric: Minimum confidence (default: 0.9)

- numRules: Number of rules to find (default: 10)

◆ **Example:** To find more rules, change numRules to 20.

Click **OK** after configuring.

◆ Step 6: Start Rule Mining

- Click the **"Start"** button.

- WEKA will generate association rules based on the dataset.

◆ Step 7: View Output

In the **Result List**, you will see output like:

markdown

CopyEdit

Best rules found:

1. outlook=sunny humidity=high ==> play=no   conf:(0.9)

2. outlook=overcast ==> play=yes   conf:(1.0)

...

Each rule shows:

- **Antecedent** (IF part)

- **Consequent** (THEN part)

- **Confidence** value

◆ Step 8: Interpret the Results

Understand key metrics:

- **Support**: Frequency of itemset in the dataset.

- **Confidence**: How often the rule is correct.

- **Lift**: Measures how much more often the rule occurs than expected if the items were independent.

✅ **Also Practise**

Objective: Perform association rule mining on a transactional dataset using the Apriori algorithm.

Steps:

1. Load supermarket.arff.

2. Go to **Associate** tab.

3. Choose **Apriori**.

4. Set numRules = 15.

5. Click **Start**.

6. Analyze the top 5 rules and note support/confidence.

Points to Note:

- Make sure your dataset contains **nominal attributes only**.

- Convert numeric to nominal using filters if needed.

- For custom datasets, save in **ARFF format** for compatibility.

- Use **"Save"** to store the rules or result output.