Practical 1: Visualization Features in WEKA

Solution:

1. Open WEKA → Explorer → Open File → Select dataset (e.g., iris.arff).

2. Go to the "Visualize" tab.

3. Scatter plot matrix appears showing relationships between attributes.

4. Identify patterns:

 - Iris-setosa is clearly separable in PetalLength vs PetalWidth.

 - Iris-versicolor and virginica overlap slightly.

Conclusion: Visualization helps detect separability and outliers.

Practical 2: Data Preprocessing & Association Rule Mining

Solution:

1. Load dataset → Preprocess tab.

2. Apply filters:

 - Remove → delete specific attributes.

 - Normalize → scale attributes.

3. To perform association mining:

 - Associate tab → Choose Apriori.

 - Set minimum support = 0.1, confidence = 0.9.

4. Click Start.

Sample Output Rule:

 petalwidth < 0.3 ⇒ class=setosa (confidence 1.0)

Conclusion: Apriori finds strong rules that describe attribute relationships.

Practical 3: Classification on Datasets

Solution:

1. Load dataset → Classify tab.

2. Choose classifier → J48 (Decision Tree).

3. Test Options:

 - Use Training Set

- Percentage split = 70%

4. Start.

Sample Result:

Accuracy: 94%

Confusion Matrix:

 a b c

a 50 0 0

b 2 44 4

c 0 3 47

Conclusion: J48 performs strongly on structured datasets like Iris.

4. Clustering in WEKA – Beginner Step-by-Step Guide

AIM:

To perform clustering on a dataset using WEKA tool.

SOFTWARE REQUIRED:

WEKA (Waikato Environment for Knowledge Analysis)

DATASET USED:

Iris dataset (iris.arff) – default dataset provided by WEKA.

THEORY (Very Simple):

Clustering is an unsupervised learning technique.

It groups similar data items together without using class labels.

STEP 1: Open WEKA

1. Launch the WEKA application.

2. Click on "Explorer".

STEP 2: Load Dataset

1. In the Preprocess tab, click "Open file".

2. Go to:

   C:\Program Files\Weka-3-8\data

3. Select iris.arff and click Open.

STEP 3: Understand Dataset

- Instances: 150 (rows)

- Attributes: 5 (columns)

STEP 4: Remove Class Attribute

1. Select the attribute named "class".

2. Click "Remove".

STEP 5: Go to Cluster Tab

1. Click on the "Cluster" tab.

STEP 6: Select Clustering Algorithm

1. Click "Choose".

2. Select clusterers → SimpleKMeans.

STEP 7: Set Parameters

1. Click on SimpleKMeans.

2. Set numClusters = 3.

3. Click OK.

STEP 8: Choose Cluster Mode

Select "Use training set".

STEP 9: Run Clustering

Click "Start".

RESULT:

The dataset is divided into 3 clusters based on similarity.

CONCLUSION:

Clustering groups similar data items without using predefined class labels.

(B) Hierarchical Clustering

Algorithm: HierarchicalClusterer

Why use it?

- Shows tree-like clustering

- Very easy to explain in viva

Steps (Short):

1. Cluster tab

2. Choose → HierarchicalClusterer

3. Click Start

👉 Output shows dendrogram

Write in practical:

Data is clustered hierarchically without specifying number of clusters.

_____

(C) EM (Expectation–Maximization) Clustering

Algorithm: EM

Why use it?

- Automatically finds number of clusters

- Probabilistic clustering

Steps:

1. Cluster tab

2. Choose → EM

3. Use training set

4. Start

Write in practical:

EM algorithm clusters data using probability distribution.

_____

2️⃣ Clustering with Different Number of Clusters (K Value)

Example:

- K = 2

- K = 3

- K = 4

👉 Show how results change

Write in practical:

Changing the number of clusters affects cluster size and centroid values.

5. **Sample Programs using German Credit Data (WEKA) – Beginner Guide**

AIM:

To study and perform sample machine learning tasks using the German Credit dataset in WEKA.

DATASET USED:

German Credit Data (german_credit.arff)

This dataset contains information about customers and is used to predict credit risk (Good or Bad).

ABOUT GERMAN CREDIT DATASET

• Each record represents one customer

• Attributes include age, credit history, loan amount, duration, etc.

• Class attribute shows credit risk: Good / Bad

• This is a real-world dataset used in banking systems

PROGRAM 1: LOADING GERMAN CREDIT DATASET

Steps:

1. Open WEKA.

2. Click on Explorer.

3. Go to Preprocess tab.

4. Click Open file.

5. Navigate to:

   WEKA → data → german_credit.arff

6. Click Open.

Result:

German Credit dataset is successfully loaded.

PROGRAM 2: DATA PREPROCESSING (REMOVE ATTRIBUTE)

Aim:

To remove an unnecessary attribute from the dataset.

Steps:

1. In Preprocess tab, select any attribute.

2. Click Remove.

3. Observe change in number of attributes.

Result:

Selected attribute is removed successfully.

PROGRAM 3: CLASSIFICATION USING NAIVE BAYES

Aim:

To classify customers into good or bad credit using Naive Bayes.

Steps:

1. Go to Classify tab.

2. Click Choose → bayes → NaiveBayes.

3. Select Use training set.

4. Click Start.

Result:

Naive Bayes model classifies credit risk successfully.

PROGRAM 4: CLASSIFICATION USING DECISION TREE (J48)

Aim:

To classify credit data using Decision Tree algorithm.

Steps:

1. Go to Classify tab.

2. Click Choose → trees → J48.

3. Select Use training set.

4. Click Start.

Result:

Decision tree is generated showing credit approval rules.

PROGRAM 5: CLUSTERING USING SIMPLE K-MEANS

Aim:

To group customers based on similar credit attributes.

Steps:

1. Go to Preprocess tab.

2. Remove class attribute.

3. Go to Cluster tab.

4. Choose → SimpleKMeans.

5. Set numClusters = 2.

6. Select Use training set.

7. Click Start.

Result:

Customers are grouped into 2 clusters based on similarity.

CONCLUSION:

German Credit dataset was successfully analyzed using preprocessing, classification, and clustering techniques in WEKA.