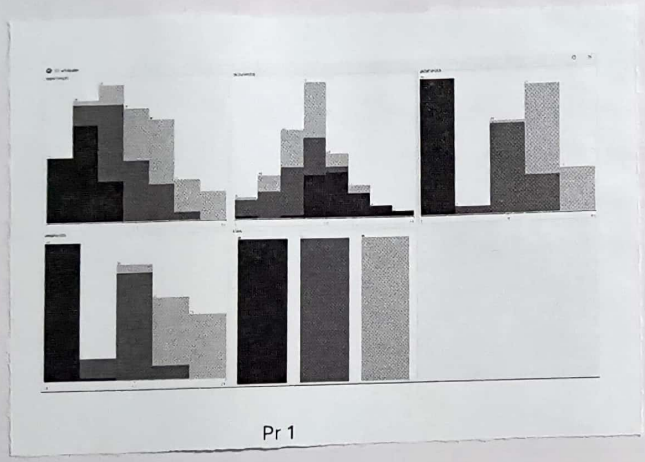


Pr 1

Parul® University | NAAC A++

Pg. No. : .....
Date : .....

Practical -1

Visualization Features in WEKA

Objective :- Explore the visualization features of WEKA for data analysis.

Steps

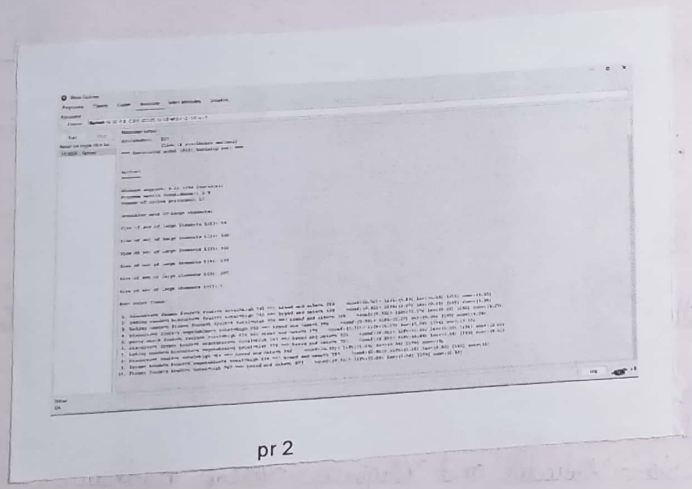
1). Open WEKA → Explorer → Open File → Local Disk (C:)
↓
Program Filed → Weka → data → Select dataset (eg - iris.arff).

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.



pr 2

Parul® University | NAAC A++

Pg. No. : .....
Date : .....

Practical - 2

Data Preprocessing & Association Rule Mining

Objective :- Perform data preprocessing and generate association rules.

Task
- clean and preprocess datasets
- Apply Apriori algorithms
- Generate association rules using support & confidence.

Steps

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 relation

Parul® University | NAAC A++

Pg. No. :
Date :

Practical - 3

Classification on Datasets

objective :- Apply classification algorithms using WEKA

Tasks
- use algorithms (J48, Naive Bayes, SVM).
- Train/test using percentage split.
- Compare accuracy metrics.

Steps

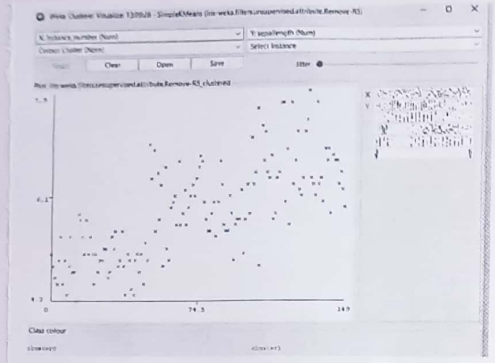
1). Load dataset → classify tab.

2). Choose classifier → J48 (Decision Tree).

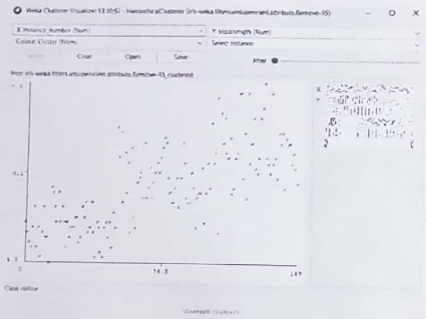
3). Test Options
- Use Training set
- Percentage split = 70%

4). Start
Sample Result
Accuracy : 94%
Confusion Matrix : a b c
a 5000
b 2444
c 0347

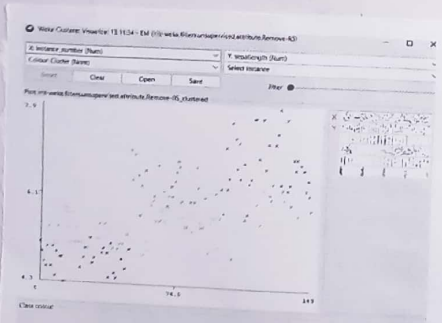
Conclusion :- J48 performs strongly on structured datasets like Iris.



pr 4(i)



Pr 4(ii)



pr(iii)

Parul® University | NAAC A++

Pg. No. : .....
Date : .....

Practical 4

Clustering on Datasets

Objective:- Perform clustering techniques

Tasks
- Use K-Means, Hierarchical clustering
- Analyze clusters using visualization
- Interpret cluster grouping patterns.

steps

1). Load dataset → cluster tab

2). Choose simpleKMeans → set K = 3.

3) start

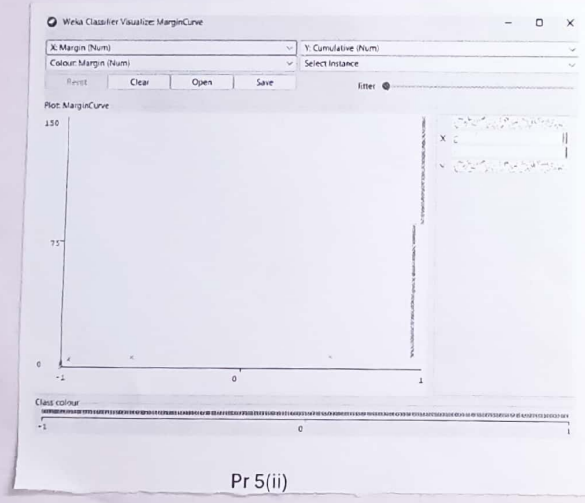
Output
cluster 0: mostly setosa

cluster 1: mixed Virginca / Versicolor

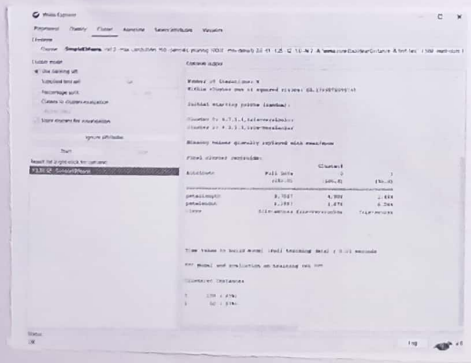
cluster 2: mixed Virginca / Versicolor

Conclusion:- K-means clusters based on attribute similarity but is unsupervised; labels are not used.

Pr 5(i)



Pr 5(ii)



Parul® University | NAAC A++

Pg. No. :
Date :

Practical - 5

Programs using German Credit data

Objective :- Apply ML techniques to German Credit dataset

Tasks :
- Load datasets into WEKA.
- Perform preprocessing.
- classify and evaluate results.

Steps

1). Load german_credit.arff → Preprocess.

2). Handle missing Values

3). Classify → choose NaiveBayes & J48.

4). Start

Result example
Accuracy : 72%

Conclusion : German Credit datasets is noisy and complex, giving la
accuracy than J48.

Program 2: Data Preprocessing (Remove Attributes)

Aim :- To remove an unnecessary attribute from the dataset.

Steps :

1). In Preprocess tab, select any attributes.

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 → Naive Bayes.

3). Select use training set.

4). Click start.

Result
Naive Bayes model classifies credit risk successfully.

Parul® University | NAAC A++

Pg. No.: .....
Date: .....

Practical 6.

Decision Tree Using Cross Validation

Objective :- Understand cross-validation in decision tree training.

Tasks
- Train decision tree using 10-fold cross-validation.
- Compare with training/testing split.
- Discuss accuracy differences.

Steps

1). Load dataset → classify tab.

2). choose J48 → Test option : 10-fold cross validation

3) Record accuracy

Compare:
- Training set accuracy : 98%
- Cross-validation accuracy : 94%

Reason: Training accuracy is higher due to overfitting.

Conclusion : Cross-Validation gives realistic performance measures.

Parul® University | NAAC A++ ACCREDITED UNIVERSITY

FACULTY OF IT & COMPUTER SCIENCE
PARUL UNIVERSITY

INDEX

Sr. No.	Date	Title of the Practical	Page No	Date Submission	Mark	Signature
1.	29/11/25	Visualization feature in weka		13/12/25		
2	13/12/25	Data Preprocessing & association rule mining		27/12/25		
3.	27/12/25	classification on DataSets		10/1/26		
4.	10/1/26	clustering on dataset		24/01/26		
5.	24/1/26	German credit data		31/01/26		
6.	31/1/26.	Decision tree with Cross Validation		31/01/26		