



Academic Year: 2025-26
Class/Branch: T.E. DS

Semester: V
Subject: DWMLab

EXPERIMENT NO. 7

1. **Aim:** To study the WEKA environment and analyze the effectiveness of the Decision Tree and Naive Bayes algorithms for classification.
2. **Objectives:** From this experiment, the student will be able to
 - Learn about the Weka Data Mining tool.
 - To implement Decision Tree classifier-J48 using Weka.
 - To implement the Naïve Bayes classifier using Weka.

3. Theory:

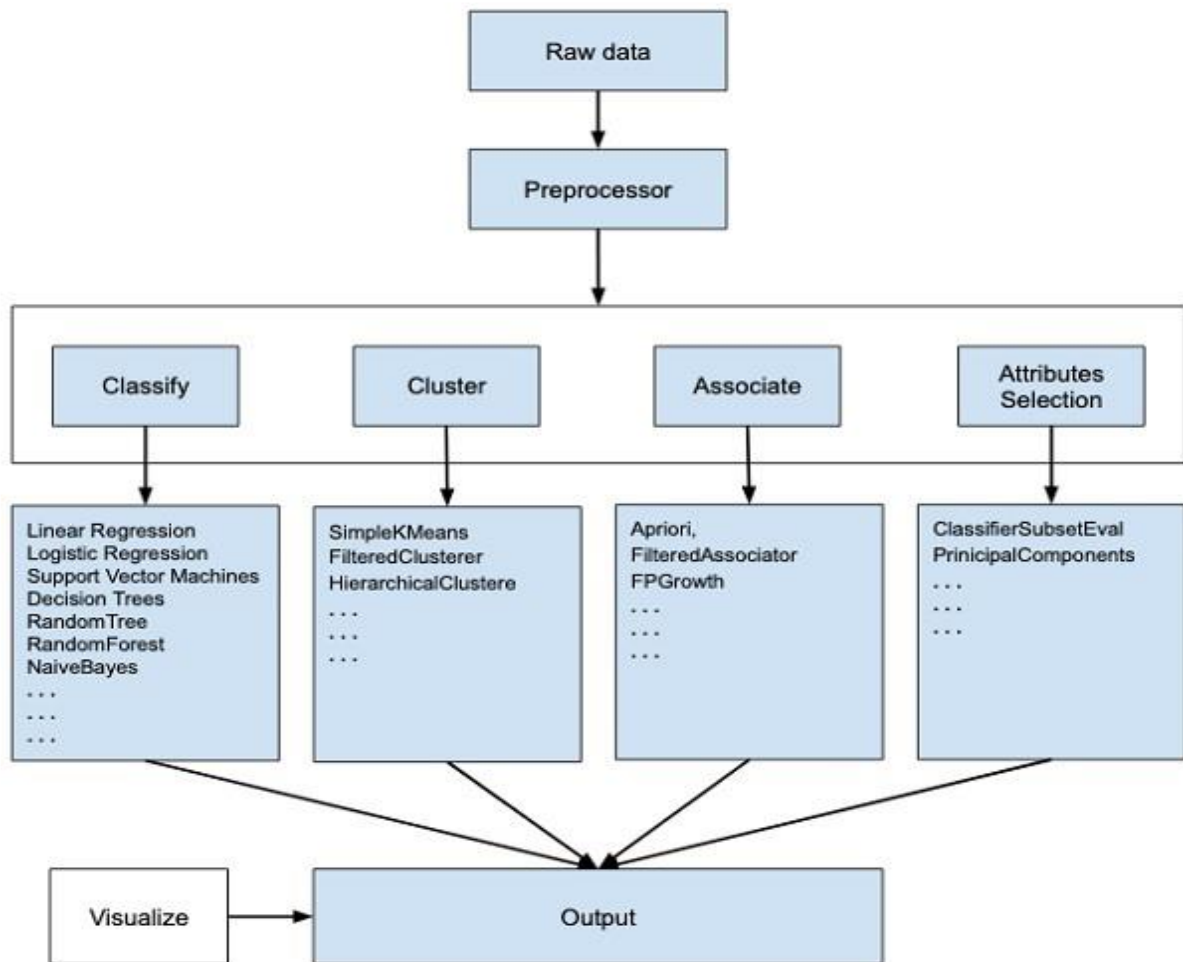
Waikato Environment for Knowledge Analysis - WEKA - an open source software provides tools for data pre-processing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real- world data mining problems.

First, you will start with the raw data collected from the field. This data may contain several null values and irrelevant fields. You use the data pre-processing tools provided in WEKA to cleanse the data.

Next, depending on the kind of ML model that you are trying to develop you would select one of the options such as **Classify**, **Cluster**, or **Associate**. The **Attributes Selection** allows the automatic selection of features to create a reduced dataset.

Note that under each category, WEKA provides the implementation of several algorithms. You would select an algorithm of your choice, set the desired parameters and run it on the dataset.

What WEKA offers is summarized in the following diagram –



Pre-processing using WEKA:

Using the **Open file ...** option under the **Pre-process** tag select the **labor.arff** file.

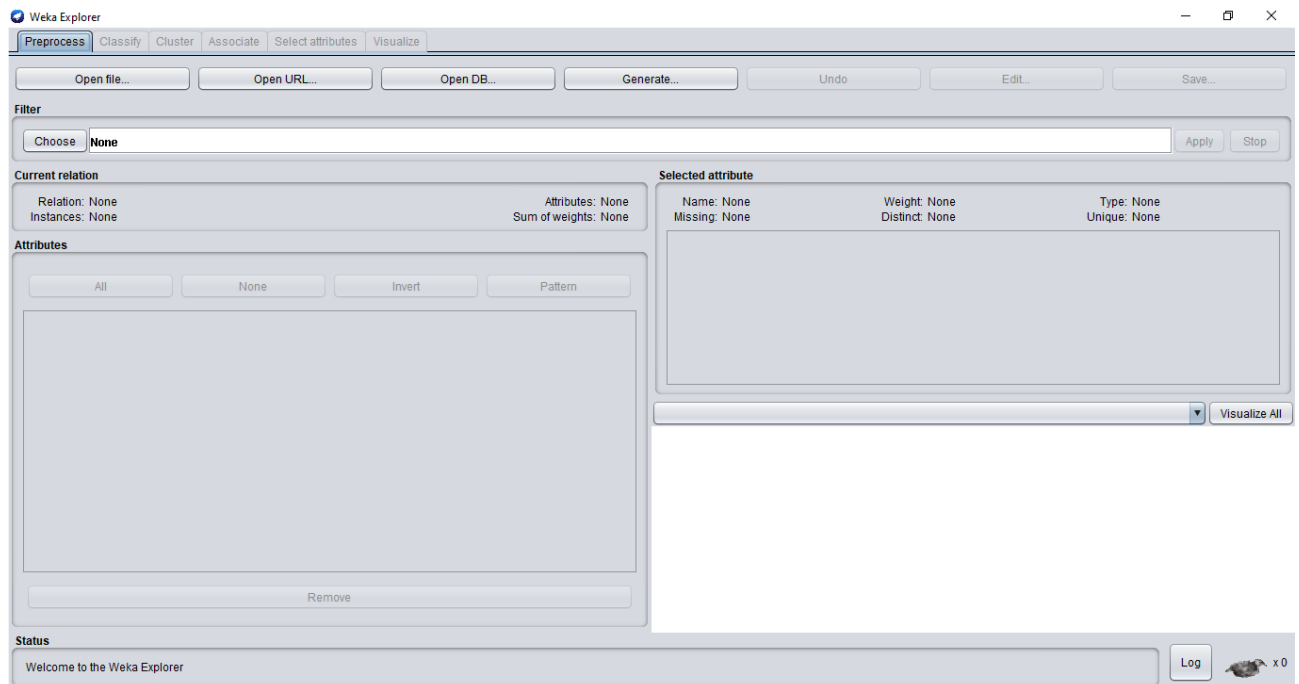
Using Filters:

weka → filters → unsupervised → attribute → ReplaceWithMissingValues

Classification using WEKA:

Decision tree:

A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision. Decision tree are commonly used for gaining information for the purpose of decision -making.



This experiment illustrates the use of j-48 classifier in weka. The sample data set used in this experiment is “weather.arff” data available arff format.

Selecting Classifier :j-48 classifier in weka

Click on the Choose button and select the following classifier –

**weka→classifiers>trees>J48**

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

- ☐ Use training set
- ☐ Supplied test set Set...
- ☒ Cross-validation Folds **10**
- ☐ Percentage split % **66**

More options...

(Nom) play

Start Stop

Result list (right-click for options)

- 03:54:46 - rules.ZeroR
- 03:58:49 - rules.ZeroR
- 04:01:34 - trees.J48
- 04:02:32 - trees.RandomForest
- 04:07:19 - trees.J48**

Classifier output

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	7	50	%
Incorrectly Classified Instances	7	50	%
Kappa statistic	-0.0426		
K&B Relative Info Score	7.4786	%	
K&B Information Score	1.0291 bits	0.0735 bits/instance	
Class complexity order 0	13.7612 bits	0.9829 bits/instance	
Class complexity scheme	3229.1699 bits	230.655 bits/instance	
Complexity improvement (Sf)	-3215.4087 bits	-229.6721 bits/instance	
Mean absolute error	0.4167		
Root mean squared error	0.5984		
Relative absolute error	87.5	%	
Root relative squared error	121.2987	%	
Total Number of Instances	14		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.556	0.600	0.625	0.556	0.588	-0.043	0.633	0.758	yes
	0.400	0.444	0.333	0.400	0.364	-0.043	0.633	0.457	no
Weighted Avg.	0.500	0.544	0.521	0.500	0.508	-0.043	0.633	0.650	

=== Confusion Matrix ===

a b <-- classified as

5 4 | a = yes

3 2 | b = no

Naive Bayes:

Naive Bayes uses a simple implementation of Bayes Theorem (hence naive) where the prior probability for each class is calculated from the training data and assumed to be independent of each other (technically called conditionally independent).

Selecting Classifier

Click on the Choose button and select the following classifier –

weka→classifiers>bayes>Naïve Bayes



PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

20:27:57 - bytes NaiveBayes

Classifier output

Naive Bayes Classifier

Attribute	Class	
	democrat (0.61)	republican (0.39)
handicapped-infants		
n	88.0	124.0
y	140.0	29.0
[total]	228.0	153.0
water-project-cost-sharing		
n	112.0	65.0
y	105.0	71.0
[total]	217.0	136.0
adoption-of-the-budget-resolution		
n	28.0	134.0
y	212.0	18.0
[total]	240.0	152.0
physician-fee-freeze		
n	224.0	3.0
y	13.0	147.0
[total]	237.0	150.0
el-salvador-aid		

Status

OK Log x0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

20:27:57 - bytes NaiveBayes

Classifier output

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	389	89.4253 %
Incorrectly Classified Instances	46	10.5747 %
Kappa statistic	0.7813	
Mean absolute error	0.0982	
Root mean squared error	0.2927	
Relative absolute error	20.6966 %	
Root relative squared error	60.1082 %	
Total Number of Instances	435	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.880	0.083	0.944	0.880	0.911	0.784	0.972	0.983	democrat
	0.917	0.120	0.828	0.917	0.870	0.784	0.972	0.955	republican
Weighted Avg.	0.894	0.097	0.899	0.894	0.895	0.784	0.972	0.972	

=== Confusion Matrix ===

a	b	<-- classified as
235	32	a = democrat
14	154	b = republican

Status

OK Log x0

4. Conclusion: