OBSERVATION:

A. Attribute Type

S.no	Attribute	Type
1.	Date	Nominal
2.	Start time	Nominal
3.	Day of Week	Nominal
4.	Going To	Nominal
5.	Distance	Numeric
6.	Max Speed	Numeric
7.	Avg. Speed	Numeric
8.	Avg Moving Speed	Numeric
9.	Fuel Economy	Nominal
10.	Total time	Numeric
11.	Moving time	Numeric
12.	Comments	Nominal
13	Take 407 All	Nominal

B. Percentage of Missing Values

S.no	Attribute	% of missing values
1	Date	0%
2	Start time	0%
3	Day of week	0%
4	Going to	0%
5	Distance	0%
6	Max speed	0%
7	Avg. speed	0%
8	Avg moving speed	0%
9	Fuel Economy	8%
10	Total time	0%
11	Moving time	0%
12	Comments	88%
13	Takes 407 All	0%

Exp:21 DATA PREPROCESSING AND ANALYSIS FOR DATASET - WEKA

AIM: To create data preprocessing and analysis for dataset using Weka.

DESCRIPTION:

Consider a dataset of traveltimes.csv file where it contains the columns of attributes as Date, StartTime, DayofWeek, GoingTo, Distance, MaxSpeed, AvgSpeed, AvgMovingSpeed, FuelEconomy, TotalTime, MovingTime, Take407all comments.

PROCEDURE:

i) Download WEKA and install
ii) Start WEKA
iii) Open the data / iris.arff dataset
iv) Select and run an algorithm
v) Review the result

RESULT:

Thus, the dataprocessing and analysis for a dataset using weka tool has been successfully completed.

Exp:22

DATA SEGMENTATION BY K-MEANS CLUSTER - WEKA & R-tool

AIM: To create data segmentation by k-means cluster using weka and R-tool.

DESCRIPTION:

Consider a dataset of citycrimes.csv file of which it contains the attributes are City, Pop, WC, BP, Mur, Rap, Rob, Ass, Bus and car for the performance of the dataset by applying the K-Means algorithm in weka and as well using R-tool.

PROCEDURE:

i) Download WEKA and install
ii) Start WEKA
iii) Open the data / iris.arff dataset
iv) Select and run an algorithm
v) Review the results.

Experiment with atleast 2 different number of clusters:

- Compare 2 different clusters but with same seed value
- Change the number of clusters value and need not change the seed value.
- Apply k-means algorithm and start executing the algorithm

RESULT:

Thus, the k-means clustering analyzing using the weka tool has been successfully completed. In case of weka tool, the change in seed values lead to the decrease in the number of iterations.

Exp:23 DATA SEGMENTATION - EXPECTATION MAXIMISATION ALGO - WEKA

AIM: To create data segmentation by expectation maximisation algorithm through weka.

PROCEDURE:

Initially, load the dataset into the weka tool and check for all the attributes present in dataset.

Then move to cluster panel and apply to EM algorithm technique for dataset.

Finally, observe the results that are obtained

RESULT:

Thus, the data analysis the expectation maximisation algorithm using weka has been analyzed and observed properly.

Exp: 24

DATA SEGMENTATION - COBWEB HIERARCHIAL CLUSTERING

AIM: To create data segmentation by cobweb-hierarchial clustering algorithm using weka tool.

PROCEDURE:

1. Download weka and install
2. Start weka
3. Open the data / ins arff dataset
4. Select and run an algorithm
5. Review the results.

RESULT:

Thus, the data analysis of cobweb hierarchial clustering algorithm using weka tool has been analyzed and observed successfully.

Exp:23

FREQUENT PATTERN MINING - ASSOCIATION RULE

AIM: To create frequent pattern mining using association rule through Weka and R-tool.

PROCEDURE:

- Choose a set of attributes for clustering and for giving a motivation.
- Choose the dataset and import the dataset into Weka.
- Discretize the attributes from numeric to nominal to perform the algorithm.
- Cluster the dataset and choose simple Apriori algm.
- Set the upper bound min-sup and lower bound min-sup values.

RESULT:

→ Thus the apriori algorithm analysing both the weka and R-tool has been successfully completed.

Weka tool → Change is upper bound and lower bound values lead to increase and decrease of number of itemsets and Rules.

↑ in absolute minimum support count value → R-tool

Exp.26

FP MINING USING FP GROWTH

AIM: To create frequent pattern mining using FP growth through weka tool.

PROCEDURE:

- Choose a set of attributes for clustering and for giving a motivation.
- Choose the dataset and import dataset
- Describe the attribute with FP growth algorithm.
- Associate the attributes with FP growth algorithm
- Set the upper bound min-sup and lower bound min. sup values.

OBSERVATION:

1) When the association rules are of values:
   a) Upper bound min.sup= 1.0
   b) Lower bound min. sup 0.1
   c) metric type = confidence

2) When the association rules are of values:
   a) Upper bound min.sup= 2.0
   b) Lower bound min. sup= 1.0
   c) metric type = confidence.

RESULT:

Thus the analysis of FP growth algorithm using weka tool has successfully completed.

Exp.27 PREDICTION OF CATEGORICAL DATA - DECISION TREE

AIM: To create prediction of categorical data using decision tree algorithm through weka tool

PROCEDURE:

* Decision tree

Visualize the decision tree for the given dataset.

* CROSS Validation Analysis

1. Download Weka
2. Open the data / iris.arff dataset
3. Select and run an algorithm
4. Review the result.

RESULT:

Thus, the observations and evations done on the german_credit dataset are analyzed. The decision tree has been successfully visualized.

Expt 8

PREDICTION OF CATEGORICAL DATA - SMO ALGM.

AIM: To create prediction of categorical data using SMO algorithm through weka tool.

PROCEDURE:

- Decision tree
- SMO algorithm
- Set the cost sensitive evaluation and compare the obtained results
- Classify the dataset with the cost sensitive classifier technique.
- Change the cost matrix to 2x2 matrix and execute.
- Analysis → total number of instances
  ↓
  mean absolute error.

RESULT:

Thus, the observation and evaluations done on the german_credit dataset are analyzed.
The comparision between decision tree and sequential minimal optimization (SMO) has been successfully visualized.

EX:25

EVALUATING ACCURACY OF CLASSIFIERS

AIM: To create evaluating accuracy of the classifiers during weka tool.

PROCEDURE:

a) Logistic Regression

- Load the dataset into the weka tool & preprocess.
- Apply the classification logistic regression technique & evaluate result.

b) Naive Bayes algorithm

- Load dataset into weka tool and preprocess it.
- Apply Naive Bayes technique and execute it.

c) J48 algorithm

- Load dataset
- Apply J48 technique and execute.

d) K-Nearest Neighbor

- Load dataset
- Apply K-Nearest Neighbor and execute

e) SMO algorithm

- Load dataset
- Apply SMO algorithm and execute

RESULT:

Thus the comparsion matrix for all the methods and technique.

Change in instance → on compalsion

Naive Bayes → most correct instance

↓

K-Means - Best.

Ex no 10 DESCRIPTION NUMERICAL PREDICTION ANALYSIS

AIM: To create description numerical prediction analysis using Linear regression.

PROCEDURE:

- Load the dataset
- Classify data using Linear regression analysis method
- Check the cross-validation folds where the value of the folds should be less than the value of the instances present in the dataset.
- Observe cross validation

OBSERVATION:

Cross validation folds - 05

Cross validation folds - 10.

RESULT:

Thus, the house selling price has been observed using linear regression model. If the cross validation folds ↓ time for creating model will be less than when folds value high, and the mean absolute error and root mean square value ↓ with ↑ in cross validation folds value.