Experiment - 21

Data Preprocessing and Analysis for Dataset using WEKA

Aim :- To create data preprocessing and analysis for dataset using weka.

Description :-

Consider a dataset of travel times .csv file where it contains the columns as Date, start time, Day of week, Going to, Distance, Speed etc.

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.

Preprocess :-

Observation :-

A ATTRIBUTE TYPE :-

S.NO	ATTRIBUTE	TYPE
1.	Date	Nominal
2.	Start time	Nominal

S.NO	Attribute	Type
3.	Day of week	Nominal
4.	Going to	Nominal
5.	Distance	Numeric
6.	maxspeed	Numeric
7.	Avg Speed	Numeric
8.	Avg moving Speed	Numeric
9.	Fuel Economy	Nominal
10.	Total time	Nominal
11.	Moving Time	Nominal
12.	Comments	Nominal
13.	Take 407 All	Nominal
14.	-	-
15.	-	-

B. Percentage of missing values:

S.No	Attribute	Percentage
1.	Date	0.%
2.	Start time	0.%
3.	Day ofweek	0.%
4.	Going to	0.%
5.	Distance	0.%

S.NO	Attribute	Percentage
6.	max speed	0.1,
7.	Avg speed	0.1.
8.	Avg moving speed	0.1.
9.	fuel economy	8.1,
10.	Total time	0.1,
11.	moving time	0.1,
12.	comments	88.1.
13.	Take 407 All	0.1.

B. min, max, mean, standard deviation.

Result:-

Thus the data processing and analysis for dataset, the using weka tool has been successfully completed.

Experiment - 22

Data Segmentation by k-means clustering :-

Aim :- To create Data Segmentation by k-means clustering using Weka and R-tool.

Description :-

Consider a dataset of city crimes.csv of which it contains the attributes are City, Pop, WC, BP, Mur, Rap, Rob, Ass, Bus & Car for the performance of data set.

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.

Using weka Tool :-

Step Involved :

- Choose a two different clusters but with the same seed values.
- Change the no.of clusters value and need not to change the seed value.
- Apply the k-means algorithm.

A. Experiment with atleast 2 different clusters but with same seed values :-

Steps Involved :-

- Compare the two different clusters but with same seed values.
- Change the no. of clusters and need not change seed values.
- Apply the k-means clustering.

Result :-

Thus the k-means clustering analyzing using weka tool has been successfully completed, in case of weka tool the change in seed value lead to decrease in no. of iterations.

Experiment - 23

Data Segmentation By Expectation Maximisation Algorithm Through Weka.

Aim :- To create data segmentation by exception maximisation Algorithm Through Weka.

Description :-

Consider a dataset of city crimes. csv file of which it contains the attributes are City, Pop, WC, BP, MUR, RAP, ROB, ASS,

Bus and Car for the performance of dataset by applying the k-means algorithm in weka and as well using R-tool.

Procedure steps :-

- Initially load the dataset into weka and check for all the attributes in dataset.
- Then move to cluster panel and apply the EM algorithm technique for the datasheet
- Finally observe the results that are obtained.

K-means Algorithm :-

Result :-

Thus, the data Analysis by the expectation maximization Algorithm using weka has been analyzed and observed properly.

Experiment - 24

Data Segmentation By CobWeb - Hierarchial Clustering Algorithm using Weka Tool.

Aim :- To create data segmentation by cobweb Hierarchial clustering Algorithm using weka.

Description :-

Consider a dataset net scan.csv where it contains the attributes of Newsgroups, groups Posts, posters, PPRatio, Returness, Replies, Replier UnRmsgs, xposts, xPTgs, etc.

Procedure :-

(i) Download weka and Install

(ii) Start weka

(iii) Open the Data / Iris and Dataset

(iv) Select and Run an Algorithm

(v) Review the Results.

Hierarchial Clustering :-

Result :-

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

Experiment - 25

Frequent Pattern Mining Using Association Rule through weka and R-Tools

Aim :- To create Frequent pattern mining using association rule through weka & R-tools.

Description :-

Consider a dataset for 2015. CSV file of which it contains the attributes are Refere no. Grid ref:, Easting, Grid Ref, Northing, No. of vehicles, Accident date etc.

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.

Using Weka Tool :-

Steps Involved :-

- Choose a set of attributes for clustering and for giving a motivation.
- Choose the dataset and import the dataset into

- Discretise the attributes from numeric to nominal to perform algorithm.
- Cluster the dataset and choose simple Apriori Algorithm.
- Set the upper bound min-sup and lower bound min-sup values.

Result :-

Thus the Apriori algorithm analyzing using both the weka tool and R-tool has been successfully completed. In case of weka tool the change in upper bound and lower bound values.

Experiment - 26

Frequent Pattern Mining using FP Growth Through Weka Tool.

Aim :- To create frequent pattern mining using FP Growth Through weka Tool.

Description :-

Consider a dataset of 2015.csv file of which it contains the attributes are Reference Number. Grid ref : Easting, Grid Ref : Nothing, No. of vehicles, Accident date, Time, Road, surface etc.

Procedure :-

(i) Download weka and Install.

(ii) Start weka.

(iii) Open the Data / Iris . arff dataset.

iii) Select and Run an Algorithm.

iv) Review the Results.

Using weka Tool:-

Steps Involved :-

(i) Choose a set of attributes to clustering and for giving a motivation.

(ii) Choose the data set and import dataset into weka tool.

(iii) Discretize attributes from all data types to nominal

(iv) Associate attributes with FP growth algorithm.

(v) Set the upper bound min-sup values & lower-bound min-sup values.

Observations :-

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 been successfully completed. In case of changing the upper bound and lower bound values there is change in no. of rules that are found.

Experiment -27

Prediction of Categorical Data Using Decision Tree Algorithm Through Weka

Aim :- To create prediction of Categorical Data Using Decision Tree Algorithm Through weka.

Procedure :-

(i) Download WEKA and Install

(ii) Start WEKA

(iii) Open the Data coff Dataset

(iv) Select and Run an Algorithm

(v) Review the Result.

Decision Tree :-

Visualize the decision Tree for the given dataset.

→ Cross Validation Analysis :-

when cross validation folds are 10 :

when cross validation folds are 05 :

Result :-

Thus, the observations and techniques done on the german credit dataset are analyzed. The decision tree has been successfully visualized.

Experiment - 28

Prediction of Categorical Data Using SMO Algorithm Through Weka

Aim :- To Create prediction of categorical data using SMO Algorithm through weka tool.

Description :-

Consider the german credit dataset which can be downloaded from UCI repository.

Procedure :-

(i) Download WEKA and Install

(ii) Start WEKA

(iii) Open and the Data.arff Dataset.

(iv) Select and Run an algorithm

(v) Review the Results.

Decision Tree :-

A tree has many analogies in real life and turns out that it has influenced a wide area of machine learning covering both classificatio and regression.

SMO Algorithm :-

The Iterative algorithm sequential minimal optimization is used for solving quadratic problems. One example where op problems are relevant is during training process of SVM.

1. Set the Cost sensitive evaluation and compare the obtained results:

Cost sensitive learning is a type of learning in data mining that takes the misclassification costs into consideration. The goal of this type of learning is to minimize the total cost. The key difference b/w cost sensitive learning.

Steps :-

(i) Classify the dataset with cost sensitive classifier techniques.

(ii) Change the cost matrix to 2*2 matrix.

Analysis :-

2. What is the significance of following parameters :-

a) Mean Absolute Error :-

Mean Absolute Error is similar to the mean squared Error. but it is absolute value instead of squaring. (the average error).

b) Total No. of Instances :-

The data present consists of various instances of class. In the case of german credit dataset. the total number of instances present.

Result :-

Thus, the observations and evaluations done on german credit dataset are analyzed. The comparision b/w decision tree and SMO has been successfully visualized.

Experiment -29

Evaluating Accuracy of classifiers

Aim :- To create evaluating accuracy of the classifiers using weka tool.

Description :- Consider the german credit dataset which can be downloaded from UCI repository.

Procedure :-

(i) Download WEKA and Install

(ii) Start WEKA

(iii) Open the Data.arff Dataset

(iv) Select and Run an algorithm

(v) Review the Results.

Analysis :-

a) Logistic Regression :-

Logistic Regression predicts the probability of an outcome, that can only have two values. The prediction is based on one or several predictors.

Steps :-

- Load the dataset into the weka tool and preprocess it.
- Apply the classification the logistic regression technique and execute for the result.

B) Navie Bayes Algorithm :-

The Navie Bayes Algorithm is based on Bayes theorem with the Independence assumptions b/w predictors. A Navie Bayesian model is easy to build.

Steps :-

- Load the dataset into the weka tool
- Apply the classification the Navie bayes technique and execute for result.

c) J48 Algorithm :-

Classification is the process of building a model of classes from a set of records that contain class labels. Decision Tree Algorithm is to find out the way the attributes vector behaves for a no. of instances.

Steps :-

- Load the dataset into the weka tool.
- Apply the classification the J48 technique.

d) K-Nearest Neighbour :-

K-Nearest Neighbour is one of the most basic yet essential classification algorithms in machine learning. It belongs to the supervised learning domain.

Steps :-

- load the dataset into the weka tool.
- Apply the classification KNN technique.

  and execute for the result.

Result :-

Thus the comparision of the matrix for all the methods and techniques. But of the comparing matrix with all the techniques there is a change in instances.

Experiment -30

Description Numerical Prediction Analysis Using Linear Regression through Weka

Aim :- To create description numerical prediction analysis using linear Regression through weka.

Description :- Consider a dataset of house.arff where it contains the attributes a house size, lot size, bed rooms granite, selling price.

Procedure :-

(i) Download weka and Install.

(ii) Start weka.

(iii) Open the Data arff Dataset

(iv) Select and Run an Algorithm.

(v) Review the Results.

Steps :-

- Load the dataset into the weka tool and check for attributes.
- Classify the data using linear regression analysis method.
- Check for validation folds where the value of the folds should be less than value of instance
- Observe the cross validation summary after applying the linear regression technique for price of house.

Observation :-

- when cross validation folds - 05.
- when cross validation folds = 10.

Result :-

Thus the house selling price has been observed while using Linear regression model. If the value of cross validation folds decreases time for creating model will be less than when folds value high.