**List of Experiments**

**1.CENTRAL TENDENCY AND DATA DISPERSION MEASURES USING R-TOOL**

**Problem Statement:**

Download the dataset from UCI-repository or from the link given below and perform/implement the central tendency measures including mean, median, midrange and data dispersion technique including summary.

<https://vincentarelbundock.github.io/Rdatasets/csv/boot/cloth.csv>

**2.** **PLOTTING GRAPHS USING R**

**Problem Statement:**

Implement the R script to plot box plot, histogram, and scatter plot for the dataset downloaded from link below.use this link to download <https://vincentarelbundock.github.io/Rdatasets/csv/datasets/mtcars.csv>

**3.PERFORM CORRECTION ANALYSIS AND NORMALIZATION**

**Problem Statement:**

Perform the correlation analysis for numerical attribute using Pearson coefficient and for categorical attribute using chi-square also, perform the normalization technique using min max, Z score.

Implement the following operations using R .

a)Min-max Normalization

b)Z-score Normalization

c)Correlation Coefficient for numerical attributes

Download the dataset from this link

[https://vincentarelbundock.github.io/Rdatasets/csv/datasets/warpbreaks .csv](https://vincentarelbundock.github.io/Rdatasets/csv/datasets/warpbreaks%20.csv)

d) Find the correlation Coefficients for categorical attributes.

Use this link <https://vincentarelbundock.github.io/Rdatasets/csv/MASS/survey.csv>

**4.REGRESSION ANALYSIS USING R**

**Problem Statement:**

Perform the linear regression for the given dataset downloaded from link <https://archive.ics.uci.edu/ml/datasets/student+performance>

**5.DATA PREPROCESSING AND PREPARATION FOR KNOWLEDGE ANALYSIS USING WEKA**

**Problem Statement:**

For each attribute find the following information,

a. Attribute Type,

b. Percentage of missing values,

c. Find min, max, mean, standard deviation on the given dataset,

d. Are there any records that have a value but no other record has,

e. Write a note on class distribution for each of the attributes.

**6. K-MEANS CLUSTER ANALYSIS USING WEKA**

**Problem Statement:**

Apply K-means algorithm to your data experiment with the algorithm as follows by setting the number of clusters and seed of a random algorithm for generation initial cluster centers.

A.Choose a set of attributes for clustering give a motivation.

B.Experiment with attest 2 different numbers of clusters but with a same seed value.

(by default:Seed=10)

C.Try with different seed value .Explain what the seed value controls.

D.Do you think the clusters are good clusters? What does each cluster represent?

**7.DATA ANALYSIS BY EXPECTATION MAXIMISATION ALGORITHM THROUGH WEKA**

**Problem Statement:**

Analyse the dataset using expectation maximisation algorithm by setting the minimum standard deviation for normal density calculation and compare the results with simple k-means algorithm.

**8.DATA ANALYSIS BY COBWEB-HIERARCHAL CLUSTERING ALGORITHM THROUGH WEKA**

**Problem Statement:**

For the given data file find the following using Weka:

A. find if there is a strong correlation btw two items

B. other correlation given data in the files

**9.KNOWLEDGE MINIG USING ASSOCIATION RULE THROUGH WEKA**

**Problem Statement:**

A. Run the Apriori algorithm and explore the association rules by changing the following parameters :

1:Upper bound min support

2:Lower bound min support

3:Metric type

4:Output item sets

5. Implement the Apriori algorithm through R tool and compare the results obtained with results obtained through Weka.

**10. FP GROWTH ALGORITHM USING WEKA**

**Problem Statement:**

Run the FP Growth algorithm and explore the association rules by changing the following parameters:

A:

1:Upper bound min support

2:Lower bound min support

3:Metric type

4:Output item sets

B. Implement the FP Growth algorithm through R tool and compare the results obtained with results obtained through Weka.

**11.PREDICTION OF CATEGORICAL DATA USING DECISION TREE Algorithm Through Weka**

**Problem Statement:**

1.What attributes do you think might be crucial in making the credit assessment ? Come up with some simple rules in plain English using your selected attributes.

2. One type of model that you can create is a Decision Tree - train a Decision Tree using the complete dataset as the training data. Report the model obtained after training.

**12. PREDICTION OF CATEGORICAL DATA USING SMO ALGORITHM THROUGH WEKA**

**Problem Statement:**

Use the German credit dataset downloaded from UC repository and analyze the given tasks

1. Do we really need to input so many attributes to get good resource

2. Compare the results obtained by decision tree and SMO

3. Set the cost century evaluation and compare the obtained results

4. What is the significance of the following parameters:

a. Mean absolute error

b. Root mean square error

c. Relative absolute error

d. Total number of instances.

13. **EVALUATING THE ACCURACY OF THE CLASSIFIERS**

**Problem Statement:**

Compare the confusion matrix generated using Weka for the German credit dataset(downloaded from UC repository)

1.logistic regression

2.make base algorithm

3.j48

4.K nearest neighbor

5. Support vector algorithm

**14. PREDICTION ANALYSIS USING LINEAR REGRESSION THROUGH WEKA**

**Problem Statement:**

Using regression analysis, create a model to calculate the price of house .create the model based on other comparable houses in the neighborhood and how much they sold. for build dataset for Weka in arff file format and had the data into Weka and finally create regression model with Weka.

**15. ETL AND OLAP OPERATION USING KNIME**

**Problem Statement:**

For the movies database, perform ETL and OLAP Operations. Use row filter, rule based filters and visualize the obtained knowledge.