**Q1.**

1. **Implement of the R script to plot box plot, Bar plot, and Horizontal bar plot for the dataset "mtcars.csv".**
2. **Implement of the R script using a group of 12 sales price records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215. Partition them into three bins by each of the following methods.**

**(a) equal-frequency (equal depth) partitioning**

**(b) equal-width partitioning**

**(c) clustering**

**3. Perform the Association analysis for supermarket data set USING**

**Weka.**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q2.**

**1. Perform the Central Tendency and Data Dispersion Measures Using R-**

**Language a) Mean b) Median c) Mode d) Data frame**

**2. Use following group of data: 200, 300, 400, 600, 1000**

**(a) min-max normalization by setting min = 0 and max = 1**

**(b) z-score normalization**

**(c) z-score normalization using the mean absolute deviation instead of standard deviation**

**(d) normalization by decimal scaling**

**3. Implementation of the Data Segmentation by Expectation Maximization**

**(EM) Algorithm and K- means through WEKA using “cancer”**

**dataset. 1. Cluster 2. Visualize**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q3.**

**1. Perform Central Tendency and Data Dispersion Measures Using R-**

**Language a) Mean b) Median c) Mode d) Data frame**

**2. Implement using R language in which age group of people are affected**

**By blood pressure based on the diabetes dataset show it using scatterplot**

**and bar chart (that is Blood Pressure vs Age using dataset “diabetes.csv”)**

**3. Perform the data preprocessing and analysis for "travel-times" dataset**

**using WEKA. A) Attributes type b) Preprocess c) cluster**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q4.**

**1. Implement the R script to Histogram and Scatter plot for the dataset**

**"mtcars.csv".**

**2.Create a Boxplot graph for the relation between "mpg"(miles per**

**galloon) and "cyl"(number of Cylinders) for the dataset "mtcars"**

**available in R Environment.**

1. **Perform the Frequent Pattern Mining using WEKA using “2015”**

**dataset. 1. Cluster**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q5.**

**1. Perform the basic program using R-Programming.**

1. **Odd or Even b) Addition C) Subtraction**
2. **Obtain Multiple Lines in Line Chart using a single Plot Function in R. Use attributes “mpg” and “qsec” of the dataset “mtcars”.**
3. **Implementation of Data Preprocessing and Analysis for "diabetes"**

**Dataset Using WEKA. a) Attributes type b) Preprocess c) Cluster**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q6.**

**1. Perform the Central Tendency and Data Dispersion Measures Using R-**

**Language a) IQR b) Range c) Quintile d) Summary**

**2. Make a histogram for the “Air Passengers“ dataset, start at 100 on the x-**

**axis, and from values 200 to 700, make the bins 150 wide.**

**3. Implementation of Data Preprocessing and Analysis for "diabetes"**

**Dataset Using WEKA. a) Attributes type b) Preprocess c) Cluster**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q7.**

**1. Implement of the R script to Plotting graphs for histogram and scatter**

**plot.**

**2. Implement using WEKA for the given Suppose a database has five**

**transactions. Let min sup= 50% (2) and min con f = 80%. Transactions**

**Items,**

**T1 (M, O, N, K, E, Y)**

**T2 (D, O, N, K, E, Y)**

**T3 (M, A, K, E)**

**T4 (M, U, C, K, Y)**

**T5 (C, O, O, K, I, E)**

**Find all frequent item sets using Apriori algorithm.**

**3. Implementation of the Correlation Analysis and Z-score normalization**

**using Weka.**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q8.**

**1. Perform the Min Max normalization and decimal scaling normalization**

**for a) Min Max Normalization b) Mean c) Minimum d) Maximum**

**2. Implement using WEKA for the given Suppose a database has five**

**transactions. Let min sup= 50% (2) and min con f = 80%. Transactions**

**Items,**

**T1 (M, O, N, K, E, Y)**

**T2 (D, O, N, K, E, Y)**

**T3 (M, A, K, E)**

**T4 (M, U, C, K, Y)**

**T5 (C, O, O, K, I, E)**

**Find FP-Growth Tree algorithm.**

**3. Evaluating Accuracy of the classifier in WEKA using “germen credit”**

**dataset.**

* **Logistic regression**
* **Navie bayes algorithm**
* **J48 algorithm**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q9.**

**1. Perform the Min Max normalization and decimal scaling normalization**

**for a) Mean b) Minimum c) Maximum d) Min Max**

**2. Evaluating Accuracy of the classifier in WEKA using “germen credit”**

**dataset.**

* **J48 algorithm**
* **K-nearest neighbor**
* **SMO Algorithm**

**3. Analysis the dataset “diabetes. csv” how the diabetes trend is for different age people, using linear regression and multiple regression.**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q10.**

1. **Implement of the R script to plot box plot, Bar plot, and Horizontal bar plot for the dataset "mtcars.csv".**
2. **Evaluating of the Numerical Prediction Analysis using Linear Regression through WEKA using “house.arff” dataset.**

* **When cross validation folds = 5, 10, 15.**

1. **Prediction of Categorical Data using Decision Tree Algorithm through WEKA using any datasets. a) Tree b) Preprocess c) Logisti**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q11.**

1. **Perform the basic program using R-Programming.**
2. **Odd or Even b) Addition C) Subtraction**
3. **Implementation of the Correlation Analysis and Z-score normalization.**
4. **Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q12.**

1. **Implement of the R script to Histogram and Scatter plot for the dataset "mtcars.csv".**
2. **Prediction of Categorical Data using Decision Tree Algorithm through WEKA using any datasets. a) Tree b) Preprocess c) Logistic**
3. **Download the Dataset "water" From R dataset Link.Find out whether there is a linear relation between attributes"mortality" and"hardness" by plot function.Fit the Data into the Linear Regression model. Predict the mortality for the hardness=88.**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **Program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q13.**

1. **Implement of the R script to Histogram and Scatter plot for the dataset "mtcars.csv".**
2. **Prediction of Categorical Data using Decision Tree Algorithm through WEKA using any datasets. a) Tree b) Preprocess c) Logistic**
3. **Imagine that you have selected data from the All Electronics data warehouse for analysis. The data set will be huge! The following data are a list of All Electronics prices for commonly sold items (rounded to the nearest dollar). The numbers have been sorted: 1, 1, 5, 5, 5, 5, 5, 8, 8, 10, 10, 10, 10, 12, 14, 14, 14, 15, 15, 15, 15, 15, 15, 18, 18, 18, 18, 18, 18, 18, 18, 20, 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 25, 25, 25, 25, 25, 28, 28, 30, 30, 30.**
4. **Partition the dataset using an equal-frequency partitioning method with bin equal to 3**
5. **apply data smoothing using bin means and bin boundary.**

**Plot Histogram for the above frequency division**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q13.**

1. **Perform the Min Max normalization and decimal scaling normalization for a) Min Max Normalization b) Mean c) Minimum d) Maximum**

**. 2. Implementation of the Data Segmentation by Expectation**

**Maximization (EM) Algorithm and K-means through WEKA using**

**“cancer” dataset. 1. Cluster 2. Visualize**

**3.Assume the Tennis coach wants to determine if any of his team players are scoring outliers. To visualize the distribution of points scored by his players, then how can he decide to develop the box plot? Give suitable example using Boxplot visualization technique.**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q14.**

**1. Perform the Min Max normalization and decimal scaling normalization**

**for a) Mean b) Minimum c) Maximum d) Min Max**

1. **Implementation of the Correlation Analysis and Z-score normalization.**
2. **Create the following dataset using CSV file format. To perform cluster analysis using KMeans in WEKA. To change the cluster size and plot the graph and illustrate the visualization of cluster.**

**EmployeID Gender Age Salary Credit**

**111 Male 28 150000 39**

**222 Male 25 150000 27**

**333 Female 26 160000 42**

**444 Female 25 160000 40**

**555 Female 30 170000 64**

**666 Male 29 200000 72**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q15.**

1. **Implement of the R script to Plotting graphs for histogram and scatter plot.**

**. 2. Evaluating Accuracy of the classifier in WEKA using “germen**

**credit” dataset.**

* **J48 algorithm**
* **K-nearest neighbor**
* **SMO Algorithm**

**3. Prediction of categorical data using Naïve Bayes classification through WEKA using any datasets. Compare the Naïve Bayes algorithm with SVM using the summary of results given by the classifiers and plot the graph**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **Program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q16.**

1. **Perform the Central Tendency and Data Dispersion Measures Using R-Language a) IQR b) Range c) Quintile d) summary**

**. 2. Evaluating Accuracy of the classifier in WEKA using “germen**

**credit” dataset.**

* **Logistic regression**
* **Navie bayes algorithm**
* **J48 algorithm**

**3.The following table would be plotted as (x, y) points, with the first column being the x values as number of mobile phones sold and the second column being the y values as money. To use the scatter plot for how many mobile phones sold,**

**X - 4 1 5 7 10 2 50 25 90 36**

**Y - 12 5 13 19 31 7 153 72 275 110**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **Program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q17.**

**1. Perform the Min Max normalization and decimal scaling normalization**

**for a) Min Max Normalization b) Mean c) Minimum d) Maximum**

**2. Evaluating Accuracy of the classifier in WEKA using “germen credit”**

**dataset.**

* **J48 algorithm**
* **K-nearest neighbor**
* **SMO Algorithm**

**3.Assume the Tennis coach wants to determine if any of his team players are scoring outliers. To visualize the distribution of points scored by his players, then how can he decide to develop the box plot? Give suitable example using Boxplot visualization technique.**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q18.**

1. **Implement the R script to Histogram and Scatter plot for the**

**dataset "mtcars.csv".**

**. 2. Perform the Association analysis for supermarket data set.**

**3.Download the Dataset "water" From R dataset Link.Find out whether there is a linear relation between attributes"mortality" and"hardness" by plot function.Fit the Data into the Linear Regression model. Predict the mortality for the hardness=88.**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q19.**

**1. Perform the basic program using R-Programming.**

1. **Odd or Even b) Addition C) Subtraction**

**. 2. Implementation of the Correlation Analysis and Z-score**

**normalization.**

**3.Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**

**Q20.**

1. **Perform Central Tendency and Data Dispersion Measures Using**

**R-Language a) Mean b) Median c) Mode d) Data frame**

**. 2. Evaluating Accuracy of the classifier in WEKA using “germen**

**credit” dataset.**

* **J48 algorithm**
* **K-nearest neighbor**
* **SMO Algorithm**

**3.Implement of the R script using a group of 12 sales price records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215. Partition them into three bins by each of the following methods.**

**(a) equal-frequency (equal depth) partitioning**

**(b) equal-width partitioning**

**(c) clustering**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Name of the Assessment** | **Weightage** | **Marks** |
| **1** | **Procedure /Algorithm/pseudocode** | **20** |  |
| **2** | **program** | **20** |  |
| **3** | **Output and result** | **20** |  |
| **4** | **Technical questions by external** | **10** |  |
| **5** | **Github content** | **10** |  |
| **7** | **Debugging** | **20** |  |
|  | **Total** | **100** |  |

**Internal Examiner External Examiner**