Sr. No.	LP-1 Problem Statement
1.	Classification: Decision Tree Dataset: madfhantr.csv Dream Housing Finance company deals in all kinds of home loans. They have presence across all urban, semi urban and rural areas. Customer first applies for home loan and after that company validates the customer eligibility for loan. Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this process, they have provided a dataset to identify the customers segments that are eligible for loan amount so that they can specifically target these customers.
2.	Classification: Naïve Bayes Dataset: NaiveBayes.csv Use probabilistic approach to implement Classifier model. Evaluate the performance of the model.
3.	Clustering: K-Means Dataset: Cities_r2.csv Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Apply at K-Means clustering algorithms (based on total_graduates) to find the group of customers.
4.	Clustering: Hierarchical Dataset: Cities_r2.csv Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Apply at K-Means clustering algorithms (based on effective_literacy_rate_total column) to find the group of customers.
5.	Clustering: K-Means Dataset: Cities_r2.csv Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Apply at K-Means clustering algorithms (based on effective_literacy_rate_total column) to find the group of customers.

6.	Clustering: Hierarchical Dataset : hitters.csv Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Apply at K-Means clustering algorithms (based on CRuns column) to find the group of customers.
7.	Clustering: K-Means Dataset : Social_Network_Ads.csv Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Apply at K-Means clustering algorithms (based on EstimatedSalary column) to find the group of customers.
8.	Clustering: Hierarchical Dataset : 50_Startups.csv Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Apply at K-Means clustering algorithms (based on PROFIT column) to find the group of customers.
9.	Regression: Simple Linear Dataset : `diabetes.csv` Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Use any one feature of the dataset to train and test the regression model. Also calculate coefficients, residual sum of squares and the coefficient of determination
10.	Regression: Simple Linear Dataset: 1.01. Simple linear regression Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Explore the relationship between students SAT score and GPA using linear regression model. Also display the regression results and plot the regression line.
11.	Clustering: K-Means We have given a collection of 8 points. P1=[0.1,0.6] P2=[0.15,0.71] P3=[0.08,0.9] P4=[0.16, 0.85] P5=[0.2,0.3] P6=[0.25,0.5] P7=[0.24,0.1] P8=[0.3,0.2]. Perform the k-mean clustering with initial centroids as m1=P1 =Cluster#1=C1 and m2=P8=cluster#2=C2. Answer the following 1] Which cluster does P6 belongs to? 2] What is the population of a cluster around m2? 3] What is the updated value of m1 and m2?
12.	Regression: Simple Linear Dataset: advertising.csv Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Explore whether TV advertising spending can predict the number of sales for the product. Also display the regression results and plot the regression line.

13.	Regression: Simple Linear Dataset: advertising.csv Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Explore whether Radio advertising spending can predict the number of sales for the product. Also display the regression results and plot the regression line.
14.	Regression: Simple Linear Dataset: advertising.csv Apply Data pre-processing (Label Encoding , Data Transformation….) techniques if necessary. Explore whether Newspaper advertising spending can predict the number of sales for the product. Also display the regression results and plot the regression line.
15.	Market Basket Analysis: Apriori Algorithm Dataset: Order1.csv The dataset has 38765 rows of the purchase orders of people from the grocery stores. These orders can be analysed, and association rules can be generated using Market Basket Analysis by algorithms like Apriori Algorithm. Follow following Steps: a. Data Pre-processing b. Generate the list of transactions from the dataset c. Train Apriori on the dataset d. Visualize the list of datasets
16.	Market Basket Analysis: Apriori Algorithm Dataset: Order2.csv This dataset comprises the list of transactions of a retail company over the period of one week. It contains a total of 7501 transaction records where each record consists of the list of items sold in one transaction. Using this record of transactions and items in each transaction, find the association rules between items. There is no header in the dataset and the first row contains the first transaction, so mentioned header = None here while loading dataset. Follow following steps: a. Data Pre-processing b. Generate the list of transactions from the dataset c. Train Apriori algorithm on the dataset d. Visualize the list of rules
17.	Market Basket Analysis: Apriori Algorithm Dataset: Order3.csv The dataset has 20507 rows and 5 columns of the purchase orders of people from the bakery. These orders can be analysed, and association rules can be generated using Market Basket Analysis by algorithms like Apriori Algorithm. Follow following steps: a. Data Pre-processing b. Generate the list of transactions from the dataset c. Train Apriori algorithm on the dataset d. Visualize the list of rules

| 18. | Classification: Naïve Bayes

Dataset: pima-indians-diabetes.csv

Use probabilistic approach to implement Classifier model. Evaluate the performance of the model. |
| --- | --- |
| | |
| 19. | Classification: Naïve Bayes

Dataset: Social_Network_Ads.csv

Use probabilistic approach to implement Classifier model. Evaluate the performance of the model. |
| | |
| 20. | Classification: Decision Tree
Dataset: data.csv
Create & evaluate the decision tree.
Test the decision tree for the following example.
"Should I go see a show starring a 40-year-old American comedian, with 10 years of experience, and a comedy ranking of 7?" |
| | |
| 21. | Classification: Decision Tree
Dataset: pima-indians-diabetes.csv
Create & evaluate the decision tree.
Test the decision tree for any random sample. |