**Assignment 1**

Read the given datasets '[bank\_marketing.csv](https://drive.google.com/file/d/1zInGv4n8kIitrggNm2CxTB1KsJ6C_c_C/view?usp=sharing)'  
  
Kindly use this attachment [Bank Marketing Data Description](https://drive.google.com/file/d/18o7o7tHW4tJdNApu-bMdhXebvtStXlcd/view?usp=sharing)

Load the necessary libraries

 load the given dataset and prepare the dataset by following the steps given below:

• Ensure the datatypes of the columns are appropriate

• Drop the variable **“Unnamed: 0”**, missing values, duplicated values

• Convert the categorical variables into dummy variables

• Split the data into train (70%), test (30%) and set random state as 0

Build a KNN model using the train dataset (k=7) and. Using the model that has been built, answer the following question.

What is the value of accuracy of the model on the test dataset?

**Assignment 2**

Read the given datasets” bank\_train.csv” for training the model  
Read the given datasets” bank\_test.csv” for testing the model

Kindly use this attachment [Bank Marketing Data Description](https://drive.google.com/file/d/18o7o7tHW4tJdNApu-bMdhXebvtStXlcd/view?usp=sharing)

Drop the missing values for training and test data.

spilt the indpendent and dependent variables for test data and train data

Build a KNN model for classification using the train dataset (k=11)

What is the accuracy of the model?

How many samples are misclassified by the model?

**Assignment 3**

Read the data Uber dataset

The dataset contains the following fields:

* key - a unique identifier for each trip
* fare\_amount - the cost of each trip in usd
* pickup\_datetime - date and time when the meter was engaged
* passenger\_count - the number of passengers in the vehicle (driver entered value)
* pickup\_longitude - the longitude where the meter was engaged
* pickup\_latitude - the latitude where the meter was engaged
* dropoff\_longitude - the longitude where the meter was disengaged
* dropoff\_latitude - the latitude where the meter was disengaged

Perform following tasks:

1. Pre-process the dataset.

2. Identify outliers.

3. Implement linear regression model to predict the price of the Uber ride from a given pickup point to the agreed drop-off location

5. Find out RMSE score

**Assignment 4**

Read the data Uber dataset

The dataset contains the following fields:

* key - a unique identifier for each trip
* fare\_amount - the cost of each trip in usd
* pickup\_datetime - date and time when the meter was engaged
* passenger\_count - the number of passengers in the vehicle (driver entered value)
* pickup\_longitude - the longitude where the meter was engaged
* pickup\_latitude - the latitude where the meter was engaged
* dropoff\_longitude - the longitude where the meter was disengaged
* dropoff\_latitude - the latitude where the meter was disengaged

Perform following tasks:

1. Pre-process the dataset.

2. Identify outliers.

3. Implement random forest model to predict the price of the Uber ride from a given pickup point to the agreed drop-off location

5. Find out RMSE score

**Assignment 5**

Given a bank customer data (churn\_modelling.csv), build a neural network-based classifier that can determine whether they will leave or not in the next 6 months.

Dataset Description:

The case study is from an open-source dataset from Kaggle.

The dataset contains 10,000 sample points with 14 distinct features such as

CustomerId, CreditScore, Geography, Gender, Age, Tenure, Balance, etc.

Link to the Kaggle project:

Perform following steps:

1. Read the dataset.

2. Distinguish the feature and target set and divide the data set into training and test sets.

3. Normalize the train and test data.

4. Initialize and build the model.

5 Print the accuracy score and confusion matrix

**Assignment 6**

Given the medical data (diabetics.csv) having features like

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | Pedigree | Age |

Implement K-Nearest Neighbors algorithm on diabetes.csv dataset to predict whether person will be diabetic or not?

Compute confusion

matrix,

accuracy,

error rate,

precision

and recall on the given dataset.