LP-III Machine Learning (2024-25) **Assignment 5:** Predict the price of the Uber ride from a given pickup point to the agreed drop-off location.	
Student Name:	**Roll No. :**
Batch:	**Division :**

Assignment 4:
Predict the price of the Uber ride from a given pickup point to the agreed drop-off location.
Perform following tasks:
1. Pre-process the dataset.
2. Identify outliers.
3. Check the correlation.
4. Implement linear regression and random forest regression models.
5. Evaluate the models and compare their respective scores like R2, RMSE, etc.

Title: Implement Linear Regression and Random Forest Regression algorithm to predict the price of the Uber ride from a given pickup point to the agreed drop-off location using dataset available at https://www.kaggle.com/datasets/yasserh/uber-fares-dataset.

Aim: Predict and Analyse Results of Linear Regression and Random Forest for Regression.

Prerequisites: Linear Regression, Random Forest, Decision Tree.

Theory:

Linear Regression Algorithm:

Regression Analysis is the process of estimating the relationship between a dependent variable and independent variables.

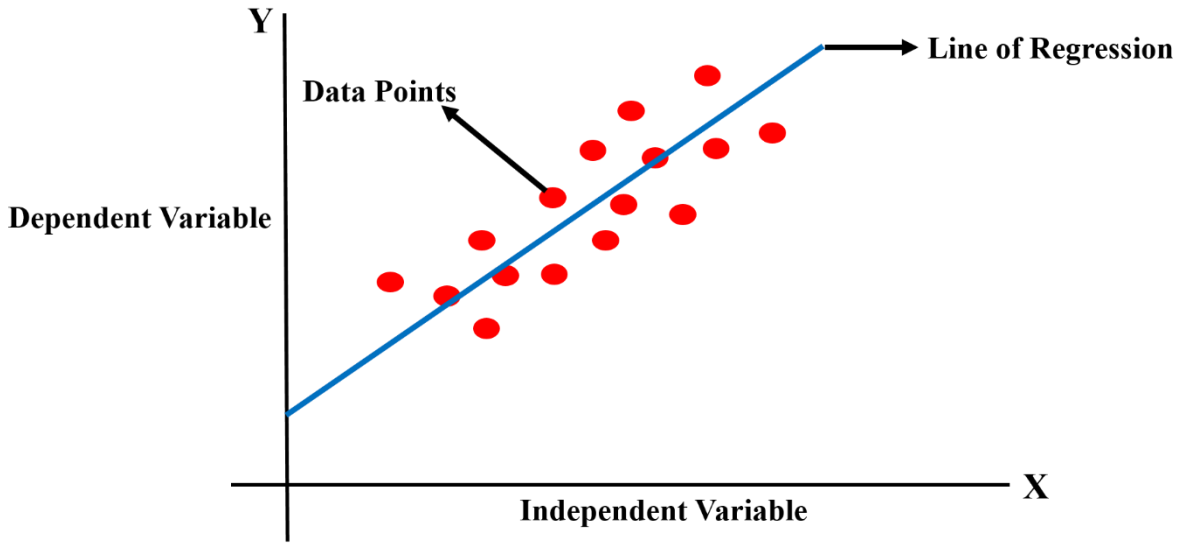

Figure 1: Linear Regression

Simple Linear Regression:

The number of **independent variables** is **one** and there is a **linear relationship** between the **independent(x)** and **dependent(y)** variable.

$$y = \propto_0 + \propto_1 (x) + \varepsilon$$

y = dependent variable

x = independent variable

$\propto_0$ and $\propto_1$ = Regression Coefficients

$\varepsilon =$ Residual Error

Multiple Linear Regression:

The number of independent variables is more then one and there is a linear relationship between the independent(x) and dependent(y) variable.

$$y = \propto_0 + \propto_1 x_1 + \propto_2 x_2 + \propto_3 x_3 + \ldots + \propto_n x_n + \varepsilon$$

y = dependent variable

$x_1, x_2, \ldots, x_n$= independent variable

$\propto_0, \propto_1, \propto_2, \ldots, \propto_n$ = Regression Coefficients

$\varepsilon =$ Residual Error

Random Forest:

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. Random Forest works in 2 phases:

1. First is to create the random forest by combining N number of decision trees
2. Second is to make predictions for each tree created in the first phase.

Algorithm:

Step-1: Select random K data points from the training set using Row Selection with Replacement and Feature Selection with Replacement.

Step-2: Build the decision trees associated with the selected data points (Bootstrap Samples).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Conclusion: Using concept of Linear Regression and Random Forest Regression algorithms, we have predicted price of the Uber ride from a given pickup point to the agreed drop-off location, compared and evaluated Linear Regression and Random Forest algorithm using evaluation metrics.

Questions:

1. What are Outliers?

2. Explain Intercept and Regression Coefficients?

3. What is Ensemble Technique?

4. Explain Bagging Technique?

5. Explain: 1. head() 2. shape 3. isnull() 4. drop()?