Predictive Modeling

Project Summary-

### Problem Statement-

Finding the proper Algorithm has consistently shown to be the diffrence between the success and failure of the entire project while seeking for the best fit model for prediction.

### Variables Description

The dataset contains weather information (Hour,Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.

Attribute Information:

Date : year-month-day

Rented Bike count - Count of bikes rented at each hour

Hour - Hour of he day

Temperature-Temperature in Celsius

Humidity - %

Windspeed - m/s

Visibility - 10m

Dew point temperature - Celsius

Solar radiation - MJ/m2

Rainfall - mm

Snowfall - cm

Seasons - Winter, Spring, Summer, Autumn

Holiday - Holiday/No holiday

Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)

### **Flow-**

------------------------------------------------------------------------------------------------------- 11/09/2023 --

Topic Selection

Dataset Searching

------------------------------------------------------------------------------------------------------- 12/09/2023 --

Listing out Analysis techniques and Algorithms to be implemented

------------------------------------------------------------------------------------------------------- 13/09/2023 --

1. Importing Libraries

2. Dataset Study

a. Load Dataset

b. View Dataset (first and last rows)

c. Rows and Columns count (rows= 8760, columns= 14)

d. Duplicate Values (0)

e. Missing/Null values (None)

------------------------------------------------------------------------------------------------------- 15/09/2023 --

3. Understanding Variables

A]

a. dataset info

b. converting datatype of date

c. extracting day, month, day of week

B]

a. Dependent Variable

(bike count)

b. Independent Variables

1. Numerical Variables

(hour, temp, humidity, wind speed, visibility, dew point temp, solar radiation, rainfall, snowfall, day, month)

2. Categorical Variables

(Seasons, holiday, functioning day, day of week)

## Points to cover-

#### **1. Data Exploration:**

a. Data Wrangling

#### **2. Exploratory Data Analysis:**

#### **a. Univariate Analysis-**

i. Dependent Variable

ii. Numerical Variables

#### **b. Bivariate Analysis-**

i. Line plot numerical variables v/s rented bike count

ii. Spread of numerical variables across hours

iii. Categorical Variables v/s rented bike count

iv. Spread of Rented Bike Count across categorical variables

v. Numerical variable v/s rented bike count

vi. Spread of numerical variable across months

#### **3. Feature Selection -**

#### **4. Supervise Machine learning algorithms and implementation :**

a. Linear Regression

b. Lasso Regression

c. Ridge Regression

d. Elastic Net Regression

e. Gradient Descent Regression

f. Random Forest Regression

g. Gradient Boosting Regression

h. XGBoost Regression

5. Model Explainability -

Libraries Required

* import numpy as np
* import pandas as pd
* import matplotlib.pyplot as plt
* import seaborn as sns
* import datetime as dt
* from sklearn.model\_selection import train\_test\_split
* from sklearn.model\_selection import GridSearchCV
* from sklearn.metrics import mean\_squared\_error,mean\_absolute\_error, r2\_score
* %matplotlib inline
* sns.set\_style("whitegrid",{'grid.linestyle':'--'})
* import warnings
* warnings.filterwarnings("ignore")