**Seoul Bike Sharing Demand Prediction**

**Contributor : Vibha Kumari**

**Define Your Problem Statement?**

**Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes.**

**Goals of this project are:**

**Understand the trends in the data and identify key factors affecting the hourly demand for rental bikes.**

**Build an appropriate regression model to forecast the number of rental bikes required per hour.**

**Dataset information:**

**Dataset used in this project is the Seoul Bike Share program data. This dataset contains information about the total count of rented bikes at each hour as well as the date of observation and meteorological information (Humidity, Snowfall, Rainfall, Temperature Season, and so on) for that hour. The observations in the dataset were recorded during a span of 365 days, from December 2017 to November 2018.**

**The Seoul Bike Dataset contains the following information**:

**Date - The date of each observation in the format 'year-month-day'**

**Hour - Hour of the day**

**Temperature - Temperature recorded in the city in Celsius (øC).**

**Humidity - Relative humidity in %**

**Wind speed - Speed of the wind in m/s**

**Visibility - measure of distance at which object or light can be clearly discerned in units of 10m**

**Dew point temperature - Temperature recorded in the beginning of the day in Celsius(øC).**

**Solar radiation - Intensity of sunlight in MJ/m^2**

**Rainfall - Amount of rainfall received in mm**

**Snowfall - Amount of snowfall received in cm**

**Seasons - Season of the year (Winter, Spring, Summer, Autumn)**

**Holiday - Whether the day is a Holiday or not (Holiday/No holiday)**

**Functional Day -Whether the rental service is available (Yes-Functional hours) or not (No-Non-functional hours)**

**Rented Bike count - Count of bikes rented at each hour (target variable)**

**Tools and Technologies used:**

**Pandas : For loading the dataset and performing data wrangling**

**Matplotlib: For data visualization.**

**Seaborn: For data visualization.**

**NumPy: For some math operations in predictions.**

**Stats models: For statistical computations**

**Sklearn: For the purpose of analysis, prediction and evaluation**.

**Steps involved:**

**Data Pre-processing : Checked for outliers, incorrect values, missing values, duplicates and performed data type correction.**

**Feature Extraction : Created new columns such as Day, Month, Year, Days\_of\_week and Weekend from Date column .**

**Exploratory Data Analysis : Performed Univariate, Bivariate, and Multivariate analysis with various graphs and plots to better understand the distribution of features and their relationships.**

**Feature Selection : Checked the VIF value (measure of multicollinearity) and dropped Dew point Temperature and Year which were highly correlated with other independent features.**

**Feature encoding : The categorical features present in the dataset Seasons, Holiday, Weekend, Functioning Day were dummified.**

**Feature Scaling : Brought features to a similar range using MinmaxScaler.**

**Implementation of Regression models**

**Hyperparameter tuning**

**Algorithms used:**

* **Ridge Regressor**
* **Elastic net Regressor**
* **Random forest**
* **Gradient Boosting.**

**conclusion**

* **Hour of the day holds most importance among all the features for prediction of**

**dataset**

* **It is observed that highest number bike rentals counts in Autumn/fall Summer**

**Seasons and the lowest in Spring season.**

* **We observed that the highest number of bike rentals on a clear day and the lowest on**

**a snowy or rainy day**

* **Peoples don’t use rented bikes in no functioning day**
* **for all the above experiments we can conclude that by gradient boosting we got the best result.**