**Capstone Project Submission**

Now a days rental bikes are used in many urban cities around the world. It gives a flexibility to the people to make mobility very comfortable. 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, it will be providing the city with a stable supply of bikes becomes a major concern. Here the important part is to prediction of bike count required at each hour for the stable supply of rental bikes.

Bike sharing system is an innovative transportation strategy that provides individuals with bikes for their common use on a short-term basis for a price or for free. Over the last few decades, there has been a significant increase in the popularity of bike-sharing systems all over the world. This is because it is an environmentally sustainable, convenient and economical way of improving urban mobility. In addition to this, this system also helps to promote healthier habits among its users and reduce fuel consumption.

With the growing demand and user base for bike-sharing system, providing the city with a stable supply of rental bikes could eventually become a challenging task. The success of bike-sharing system relies in ensuring that the quality of facilities provided, meets the needs and expectations of the users. Therefore, it is important to ensure that rental bikes are available and accessible to the users at right time, as it reduces the waiting time. Forecasting the number of bikes required and identifying the key factors that influence the demand for rental bikes can greatly help in managing the bike-sharing system.

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.

In the first part we explore the dataset we have “SeoulBikeData.csv” how many rows and column are present and information regarding overall dataset like Null values, data type column name etc. Then we go for the clearing data process. Which is important step before we going for the EDA (Exploratory Data Analysis).

In the Second part we do the EDA with our modified dataset. There we have to analysis the independent variable relation with respect to dependent variable and independent variable. There are two type of relation scenario can happen like independent variable (Rented\_Bike\_Count) vs Categorical Variable

(Month, Weekdays and Weekend, Functioning Day and Non-Functioning Day, Season And Hour) and like independent variable (Rented\_Bike\_Count) vs Numerical Variable (Temperature, Windspeed, Dew\_Point\_Temperature, Solar\_Radiation, Snowfall, Rainfall).

In the third part we plot some regression plot to see the liner relation b/w numeric and target variable.

Here we will get two type of linear relation positively relation and negative relation. Next, we go for checking of correlation between dependent variable here in the project we find temperature and dew point temperature are positively correlated to each other. So we can eliminate dew point temperature.

Now in the fourth part we go for model training and testing. For that we must split the data in training and testing. Mostly we split in 80/20 or 70/30 ratio respectively.After split the data we go for various regression process( **LINEAR REGRESSION, LASSO REGRESSION, RIDGE REGRESSION, ELASTIC NET REGRESSION, DECISION TREE,RANDOM FOREST**). After this process we go for gradient boosting and hyper tuning.

At last, we go for final conclusion about your date that mostly random forest Regressor and Gradient Boosting gridsearchcv gives the highest R2 score of 99% and 90% respectively for Train Set and 92% for Test set. Feature Importance value for Random Forest and Gradient Boost are different.

**Contributor Role: -**

**N Narayan Santosh Ku. Choudhury**

* Data Wrangling
* Checking data frame
* Data frame Exploration
* Performing Exploratory Data Analysis
* Analysis of Independent variable wrt Dependent variable
* Categorical variables
* Numerical variables
* Model Training
* LINEAR REGRESSION
* LASSO REGRESSION
* RIDGE REGRESSION
* Understanding the importance of the features
* GRADIENT BOOSTING

**Kuresh Chandra tripathy: -**

* Data Wrangling.
* Mounting and read the data frame
* Clearing and Processing Data frame
* Regression plot, normalize dataset, checking correlation and modify as per requirement.
* Model Training
* ELASTIC NET REGRESSION
* DECISION TREE
* RANDOM FOREST
* Hyperparameter tuning

**CONCLUSION**

**GitHub link: -**

**Google drive link for dataset: -**

**Google drive link for project data: -**