Project Summary
Bike Sharing Demand Prediction

Currently rental bikes are introduced in many urban cities for the enhancement of mobility comfort. The purpose of this movement is to modernize cities and encourage people to head to a green world. Let's take the examples of Paris in 2007, where "velibs" were introduced and Amsterdam, where there are more bikes than cars. The goal is to facilitate the commute in Seoul and reduce the number of cars and the pollution. Indeed, the development of the way to commute reduced the use of cars to go to work and visit the city.

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.

Started by getting some basic information about the data types, null values as well as some descriptive statistics of the features. This allowed us to decide which columns and data to keep while removing the null values. The statistical analysis also allowed us analyze and keep the required data.

The given problem is solved by the regression model but from the pair plot, we can see that the relationship between the independent variable and the dependent variable is not linear. So, it shows that a linear regression fitting might not be the best model for it. A linear model might not be able to efficiently explain the data in terms of variability, prediction accuracy etc.

Then tried Lasso and Ridge Regression but there was no difference in the model evaluation metrics when compared to that of Linear Regression Then we have to make our model more complex for better discretion or move to tree and ensembling algorithm for better results.

Then carried out Decision Trees and Random Forest Classifier and tuned the Hyperparameter with Grid Search CV and the Adjusted R2 score was good for both Decision Tree and Random Forest. Random forest gives predictions better than a decision tree model. Predictions made by Random Forest is better than all the models used. The value of the Adjusted R-squared for the Random Forest is 0.796, which is good.

This study shows that the rents of bikes are influenced by a lot of features. And, also understood that many Koreans usually and mainly rent bikes during the week days, so we supposed that the main use is to go to school or work. There are also many conditions which contribute to the variation of number of rents like the day of the week, the moment of the day and weather conditions. Weather conditions are also very important because there are more rents during spring and summer. And as we expected more people are set to rent bikes when the weather is favorable.

Contributor Roles:

1. **Sharath Diwakar** *(mailmesharathd@gmail.com)*

    1) Data Cleaning
    2) Data Wrangling
    3) EDA (Exploratory Data Analysis):
        - Heat map
        - Bar plot, Pie plot, Line graph
        - Pair plots
    4) Feature Engineering:
        - Data pre-processing
        - Feature Selection
        - Scaling
    5) Classification Analysis:
        - Linear Regression
        - Lasso Regression
        - Ridge Regression
        - Decision Tree Classifier
        - Random Forest Classifier
        - Hyper Parameter tuning & Grid Search CV

GitHub Link: *https://github.com/Sharath2021/Bike-Sharing-Demand-Prediction*