

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Team Member's Role:-

- **Ajinkya Dakhale**
Email- Ajinkya.dakhale2408@gmail.com
 - Bivariate Analysis
 - VIF
 - linear Regression
 - Ridge Regression
- **Harshjyot Singh Chawla**
Email- hs9158695878@gmail.com
 - Multivariate Analysis
 - Data wrangling
 - Feature Engineering
 - Lasso Regression
- **Suvir Kapse**
Email- suvirkapse@gmail.com
 - Data understanding
 - Univariate analysis
 - Decision Tress
 - Random forest
 - Hyperparameter tuning

Please paste the GitHub Repo link.

Github Link:-

https://github.com/Harshjyot-Singh-Chawla/Seoul-Bike-Sharing-Demand-Prediction_

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

PROBLEM STATEMENT:

Bike rental, public bike share scheme or bike sharing business helps an individual to rent out bikes for a period of time (for few hours). The customer can get a bike from a station for few hours to travel to a destination and then can drop to any station nearby. This business is helping in the enhancement of comfort mobility in the crowded urban cities.

This business requires making rental bike available at right time and at right place so that it decreases the waiting time of customer. Bike sharing decreases traffic, reduces energy consumption, decreases harmful gas emission and improves public health. Therefore for customer satisfaction there should be a stable supply of rental bikes at each station.

In this supervised machine learning regression project, the problem statement is to predict the bike count required at each hour for stable supply of rental bikes.

APPROACH:

- The first step includes loading of dataset and then inspecting the data through which we get to know the summary or description of data, shape of data, null value count in the data and about the data types of column.
- Second step includes extracting day month year from date column Then examine all categorical features and obtain their relationships by doing univariate and multivariate analysis. Then, plot histogram for numerical features. After plotting we can see that these features are either Right or Left Skewed. And to normalize all these features we tried log10, square, square root among which square root gave best results.
- Then, various regression models are applied and rented bike count is dependent variable and others are independent variable. We used Linear Regression, Regularized Linear Regression, Decision Tree Regression, and Random Forest Regression to get values for evaluation metrics.

CONCLUSION:

At last , conclusion is achieved on the basis of results shown through machine learning regression models evaluation metrics that:

- High demand of bikes on holidays or non working day.
- Maximum number of bikes are rented between 7 to 9 am and 5 to 7 pm in the evening which indicates office going customer rents the maximum number of bikes.
- Bike demand was least in monsoon and maximum during summer, autumn and spring because of beautiful weather.
- Temperature, Hour and humidity are the most important labels which are positive related to renting of bikes.
- We used different type of regression algorithms to train our model like, Linear Regression, Regularized linear regression (Ridge and Lasso), Random Forest Regressor, Decision tree. And Also we tuned the parameters of Decision tree, Out of them Random forest regression gave the best result on evaluation metrics.