

Capstone Project Submission

Project Summary

Vikas Chaudhary

Data Science Trainee,
Almabetter (Bengaluru)

Email Id:

Vikas.chaudhary53@gmail.com

GitHub Repo Link:

[vikaschaudhary53/bike-sharing-demand-prediction: Analyzing the Seoul Bike Sharing data for demand prediction using ML Regression \(github.com\)](https://github.com/vikaschaudhary53/bike-sharing-demand-prediction)

Problem Statements: Develop Supervised Machine Learning (ML) model on the 'Seoul Bike Sharing Dataset' to predict the demand of rented bikes.

Approach: The given database contains 8760 rows and 14 columns describing date, time, month, season, functioning/non-functioning days, holiday/non-holiday, climate and weather, number of rented bikes.

Our task is to develop ML models to predict the overall demand of bike at a certain condition or situation. While doing this a workflow that includes Data Preparation, EDA (Exploratory Data Analysis), Feature Engineering, Model Implementation.

The Following Supervised ML models are used:

- Linear Regression
- Lasso Regression
- Ridge Regression
- Elastic Net Regression
- Decision Tree
- Gradient Boosting Machine (GBM) Algorithm
- Random Forest

Conclusion:

- Features related to hours, work days, functioning days, seasons have a high impact on the overall bike rental operations.
- Non Functioning days have no bike rentals so there is no need to include it while training it.
- Domain knowledge can make a large difference during feature selections or dealing with outliers because the mathematical interpretation of outliers is not accurate.
- Since the project is about mobility so any situation or event that affects the overall transportation or movement of vehicles also affects these operations esp. the climate and weather conditions.
- Having a proper and efficient functioning model that can predict the number of bike requirements at a given point in time can lead to flawless operations.
- Since it is an emerging business or service it also has other ramifications in terms of business opportunities, efficient urban transport system, pollution, fuel consumption, employment etc.

Note: This project does not cover any unprecedented events or scenarios (pandemic, war etc) that can alter the operations; it is entirely based on the features that are covered in the database.

