

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Shubham Srivastava:

E-Mail- shubham.mach30@gmail.com

- Loading Dataset
- Data Inspection
- Exploratory Data Analysis
- Evaluating various models
- Conclusion

Please paste the GitHub Repo link.

Github Link:- https://github.com/Shubhamverse/Bike-Sharing-Demand-Prediction__Supervised-ML-Regression

Drive Link:-

https://drive.google.com/drive/folders/1UhhSvWf3eJDkxxg0fzeSKvqTPr_ELlvY?usp=sharing

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

In this Supervised Regression Project, we were provided with dataset i.e. Seoul Bike Sharing .

As the first step, perform Data Inspection after the loading of Dataset through which we get to know the summary of data, shape of data, null value count in data and detail about datatype of columns.

Next, extract day, month, year from date column. Then visualize all categorical features and infer their relationships. 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, we apply various regression models where rented bike count is dependent variable while others are independent variable. We applied Linear Regression, Lasso Regression ,Ridge Regression, Polynomial Regression, Decision Tree Regression, Random Forest Regression, Gradient Boosting Regression, XGBoost Regression, lightGBM Regression to get values for evaluation metrics.

Finally, we derive conclusion based on results shown through various regression models evaluation metrics that:

In holiday or non-working days there is high demands for bike.

People preferred more rented bikes in the morning than the evening.

When the rainfall was less, people have booked more bikes.

The Temperature, Hour & Humidity are the most important features that positively drive the total rented bikes count.

After performing the various models the lightGBM found to be the best model that can be used for the Bike Sharing Demand Prediction since the performance metrics (mse,rmse) shows lower and (r2,adjusted_r2) shows a higher value for the lightGBM! We can use lightGBM for the bike rental stations.