

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

1) **Sourav Karmakar**

E-mail: karmakarsourav2024@gmail.com

- Framework establishment
- Line Plot
- Data Manipulation
- Data Preprocessing
- Exploratory Data Analysis
- Feature Engineering
- Error Handling
- One hot encoding
- Hyperparametric Tuning
- Heatmap & Correlation Analysis, VIF factor
- Tree Based Model Selection
- SHapley Additive exPlanations
- SHAP Summary
- Technical Documentation
- Linear Model Selection
- Evaluation Matrix

Please paste the GitHub Repo link.

GitHub Link:-

<https://github.com/souravkarma/Sourav-Karmakar-Bike-Sharing-Demand>

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:

Many urban cities introduce rental bikes for good accessibility purposes. It is important to make the rental bike available to the public at the right time as it makes our life easier and faster. 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.

Conclusion:

So we have come to the end of our project Bike Sharing Demand Prediction. What we did let's take a short recap. We have found our dataset info where we have found there are 8760 rows and 14 columns with 13 features as independent and one as dependent according to our problem statement that is Rented bike count on which we have done our prediction. We didn't find any null values or duplicates.

In feature engineering we have done one hot encoding, ordinal encoding on independent features. Then in correlation analysis we have dropped the temperature feature as it was showing very high correlation with Dew Point Temperature which would be very difficult for model prediction so as to interpret as well.

In model selection we have selected Gradient Boost Model with GridSearch to get the best estimator for our model prediction as it has given very less mean squared error and high r2 score of above 90% for both train and test data with a split of 50% each.

In model explainability we found that Winter Feature is giving high SHAP value while hour_8, Windspeed, Visibility is not contributing in Tree Based Models.

References

- GeekforGeeks
- Kaggle
- Analytics Vidya