Capstone Project Submission

Team Member's Name, Email and Contribution:

1).**Saugata Deb**
   E-mail: saugatad56@gmail.com
- Framework establishment
- Line Plot
- Data Manipulation
- Data Preprocessing
- Feature Engineering
- Correlation Analysis
- Tree Based Model Selection
- Model Deployment
- Feature Importance
- SHapley Additive exPlanations
- Shap Summary

2).**Chukkapalli Naga Sai**
   E-mail: nagasaichowdary1111@gmail.com
- Debugging Error
- Data Sorting
- Technical Documentation
- ppt Presentation
- Approach Towards Plan
- Seaborn,matplotlib
- Heatmap
- Linear Model Selection
- Evaluation Matrix

3). **Shreyash Movale**
   E-mail: shreyash9m@gmail.com
- Data Sorting
- Matplotlib
- ppt Presentation
- Data Visualization
- Technical Documentation
- Approach toward Plan
- Line Plot,Barplot,Histogram
- Heatmap,VIF factor
- Linear Model Selection
- Evaluation Matrix
- Data Preparation

4). **Ankit Patil**
   E-mail: ankit.patil67@gmail.com
- Data Cleaning
- Data Analysis
- Error Handling
- One hot encoding

- Feature Engineering
- Normalization
- Correlation Analysis
- Tree Based Model Selection
- Model Deployment
- Hyperparameteric Tuning

Please paste the GitHub Repo link.

Github Link:-

https://github.com/SaugataDeb/SaugataDeb-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:
Currently, Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. 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 the 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 GradientBoost 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_21,hour_8,WindSpeed is not contributing in Tree Based Models.

References
- GeekforGeeks
- Kaggle
- Analytics Vidya