Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1). Chinmay Rojatkar

E-mail: chinmayrojatkar4@gmail.com

- Framework establishment
- Line Plot
- Data Manipulation
- Data Pre-processing
- Feature Engineering, Correlation Analysis

2). Bipasha Zade

- Tree Based Model Selection
- Model Deployment
- Feature Importance
- Shapley Additive explanations

3). Deepali Mahajan

E-mail: deepali2062@gmail.com

- Debugging Error
- Data Sorting
- Technical Documentation
- ppt Presentation
- Approach Towards Plan
- Seaborn, matplotlib
- Heatmap
- Linear Model Selection
- Evaluation Matrix

4). Kunal Gawande

E-mail: gkunal8019@gmail.com

- Data Sorting
- Matplotlib
- ppt Presentation
- Data Visualization
- Technical Documentation
- Approach toward Plan
- Line Plot, Bar plot , Histogram
- Heatmap
- Linear Model Selection
- Data Preparation

5). Nikhil Aggarwal

E-mail: nickagg30899@gmail.com

- Data Cleaning
- Data Analysis
- Error Handling
- One hot encoding

Please paste the GitHub Repo link.

Kunal Gawande Link:- https://github.com/gkunal8019

Chinmay Rojatkar Link:- https://github.com/ChinmayRojatkar

Bipasha Zade Link:- https://github.com/Bipashazade

Nikhil Aggarwal Link:- https://github.com/Nikhil8815

Deepali Mahajan Link::- https://github.com/deepali2062

Introduction:

Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network and locations throughout a city. The goal of this project is to combine the historical bike usage patterns with the weather data to forecast bike rental demand.

Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city. Using these systems, people are able rent a bike from a one location and return it to a different place on an as-needed basis. One of the most important problems from a business point of view is to predict the bike demand on any particular day. While having excess bikes results in wastage of resource (both with respect to bike maintenance and the land/bike stand required for parking and security), having fewer bikes leads to revenue loss (ranging from a short-term loss due to missing out on immediate customers to potential longer-term loss due to loss in future customer base), Thus, having an estimate on the demands would enable efficient functioning of these companies. The goal of this project is to combine the historical bike usage patterns with the weather data to forecast bike rental demand.

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

Please paste the GitHub Repo link.

GitHub Link:-

https://github.com/gkunal8019/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.