Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
1.Sonali Kaushal **Email - doll.kaushal181@gmail.com** • Feature engineering • Exploratory data analysis • Linear regression • Ridge regression • Lasso regression • Polynomial regression • XG Boost • Evaluation matrix **2.Somya Hingorani** **Email - somya.hingorani1899@gmail.com** • Feature engineering • Exploratory data analysis • Linear regression • Ridge regression • Lasso regression • Decision Tree • XG Boost • Evaluation matrix **3. Nitesh Verma** **Email - niteshavagabond@gmail.com** • Feature engineering • Exploratory data analysis • Linear regression • Ridge regression • Lasso regression • Random Forest • Gradient Boost • Evaluation matrix **4. Harsh Vardhan**

Email - harsh9vardhan9@gmail.com

- Feature engineering
- Exploratory data analysis
- Linear regression
- Ridge regression
- Lasso regression
- Elastic net regression
- Gradient Boost
- Evaluation matrix

5. Prateek Gupta

Email - prateekgupta0786@gmail.com

- Feature engineering
- Exploratory data analysis
- Linear regression
- Ridge regression
- Lasso regression
- Elastic net regression
- XG Boost
- Evaluation matrix

Please paste the GitHub Repo link.

GitHub Link: - https://github.com/niteshavagabond/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 become a major concern. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes.

Approach:

Exploring the dataset is the first step means Exploratory Data Analysis, finding null/missing values, and then looking for duplicate values if any, luckily dataset neither has null values nor duplicate values, after that we look for outliers on features, our target variable contained outliers and it was right-skewed, then we took the log of the target variable to convert into a normal distribution, and after that, we did some feature engineering on the dataset and convert columns which were in the wrong datatype, next we did encoding for categorical columns after that we did univariate, Bivariate and multivariate analysis, then we created functions for model creation and evaluation matrix and finally we choose the best model which is suitable for our problem.

Conclusion – EDA:

1. The highest demand for the rented bike was recorded during June but the least demand was recorded during January and February.
2. Summer is the busiest season and winter is the least busy season for the rented bike.
3. Around 6 pm is the busiest hour of the day.
4. On a working day, the rental bike demand is high in comparison to the holiday.
5. the highest demand for the rented bike is seen when there is the moderate temperature that is around 25°C to 35°C.

Conclusion – Model:

1. Based on R2 and Adjusted-R2, the Gradient Boosting model and eXtreme Gradient Boosting model are best as the accuracy of these models is above 90% and also their Adjusted-R2 values are less than R2 values.
2. Gradient Boosting model's R2 and Adjusted-R2 values are 0.904 and 0.903 respectively.
3. eXtreme Gradient Boosting model's R2 and Adjusted-R2 values are 0.922 and 0.921 respectively.
4. Linear model, Lasso model, Ridge model, and Elastic net model haveR2 and Adjusted-R2 values below 60%.