Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Contribution:

- Initial Libraries and Dataset Import setup.
- Data Exploration and Cleaning
  1. Pie Chart of Seasons , Holiday Column
  2. Check the Skewness
  3. Bar Plot of each Categorical Column
  4. Relations between categoricals and dependent Features
- Feature Engineering and Correlation
  1. Correlation
  2. Check multicollinearity
  3. Check Outliers from Box Plot
  4. Use InterQuantile range for removing Outliers
- Fit the model
  1. Linear Regression
  2. Lasso Regression
  3. Ridge Regression
  4. Elastic Net Regression
  5. Decision_Tree
  6. Random_Forest
  7. XGboost
  8. LightGBM
- Different Model Metrics

	Linear	Lasso	Ric
Mean_square_error	193577.938673	194075.891177	194107.6131
Root_Mean_square_error	439.974930	440.540454	440.5764
R2	0.545853	0.544685	0.5446
Adjusted_R2	0.540676	0.539494	0.5394

- Model Explainability
  1. SHAP
  2. ELI 5

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 & Approach:
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 bike count required at each hour for the stable supply of rental bikes.

Conclusion:
In this project, we explored several types of information that influence bike rental count. Below is a quick summary of exploratory data analysis conclusions from above pie chart:

Season: Most bikes have been rented in the summer season.least bike rent count is in winter season.autumn and spring seasons have almost equal amounts of bike rent count.

YearwiseRentedBikeCount: Most of the bikes have been rented in the year 2018.

MonthwiseRentedBikeCount in 2018: Very less bikes have been rented in December, which is the winter season.

Intensities of Rainfall: People tend to rent bikes when there is no or less rainfall.

Intensities of Snowfall: People tend to rent bikes when there is no or less snowfall.

Temperature: People generally prefer to bike at moderate to high temperatures. We see highest rental counts between 25 to 36 degree celsius.

Visibility Range: people tend to rent bikes when the visibility is between 300 to 1700.

Working or Non-working Day: We see 2 rental patterns across the day in bike rentals count - first for a Working Day where the rental count high at peak office hours (8am and 5pm) and the second for a Non-working day where rental count is more or less uniform across the day with a peak at around noon.

Hour of the day: Bike rental count is mostly correlated with the time of the day. As indicated above, the count reaches a high point during peak hours on a working day and is mostly uniform during the day on a non-working day.

Modeling Conclusion

We used 6 Regression Models to predict the bike rental count at any hour of the day - Linear Regression, Ridge, Lasso, Random Forest,XG Boost and LightGBM Model.

Below is a summary of the model performances:

Of all the models, we found a LightGBM Model providing the best/lowest RMSE score and highest R^2 score. Hour of the day is the most important feature in the respect of all independent features which provide the highest bike rented count. Thus, we have successfully built predictive models that can predict the demand for rental bikes based on different weather conditions and all other features. If the

model interpretability is important to the stakeholders, we can choose to deploy the Light GBM model.
https://github.com/Somya2119/Capston-Project-Bike-Sharing-Count-Prediction.git