Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Contributor's Role:

1. **Parth Sharma**
   Email: parthsharma01197@gmail.com.
   Contribution:
   - Understanding the Bike Sharing Demand Prediction Dataset.
   - Data Wrangling on the Bike Sharing Demand Prediction Dataset.
   - Discovering the underlying patterns.
   - Data Visualization on the Bike Sharing Demand Prediction Dataset.
   - Drawing conclusions from the visualizations.
   - Feature Engineering on the data set.
   - Testing different ML models on the data set.
   - Checking for the model giving the confirming results.
   - Conclusion drawn from the dataset.

Please paste the GitHub Repo link.

Github Link:- https://github.com/ParthSharma1197/EDA-Capston-Bike-Sharing-Demand-Prediction-.git

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

A Bike-sharing system is a shared transport service in which bicycles are made available for shared use to individuals on a short-term basis for a price. Many bike share systems allow people to borrow a bike from a "dock" and return it at another dock belonging to the same system. The user enters payment information, and the computer unlocks a bike. The user returns the bike by placing it in the dock, which locks it in place.

From the EDA done we can have some prominent observations that the demand that can be assumed through the count o bikes rented is maximum in the morning between 7 am to 9 am and then in the evening 5 pm to 7 pm which could be assumed to office timings so we can say that maximum bikes are rented by the office commuters.

Starting with data preprocessing, there were numerous operation that we required to shape the dataset in such a form that is acceptable by our ML models such as data cleaning, Data reduction, Feature Engineering was performed on the data in order to transform it in the required form. Then went to apply our ML models.

The linear regression was chosen for baseline Regressor. The best model is the model which has highest R2 score in this case as R2 is represents the goodness of fit of a model and minimum value of RMSE. Gradient Boosting Regressor with Grid Search CV was the model that provided us with the highest value of the R2 score. Compared to other prediction models Lasso performed worst.

Decision Tree model gave us fairly better results as compared with the Linear Regression, Lasso and Ridge regression. Most important feature according to the Decision Tress was Temperature i.e. the feature that affects the dependent feature the most is temperature only.

Ridge Regressor and base line Linear Regressor provided the same value of R2. Lasso regression model is worst, we can conclude that baseline algorithm was not beaten by it, but the other 3 algorithms successfully outperformed Linear regression. This indicates that the relationship between bike counts and the features are non-linear. Gradient Boosting Regressor has outperformed every traditional model which is not a big surprise.