Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Team Member's Role:-

Nikhil Machave (machavenikhil@gmail.com)

Contribution:
o Data understanding
o Handling null or missing values
o Performing EDA
o Removing Outliers
o Linear Regression Model
o Multivariate analysis
o Lasso Regression Model

Aishwarya Methe (aishwaryamethe252@gmail.com)

Contribution:
o Data understanding
o Data visualization
o Multivariate analysis
o Gradient Boosting Model
o Ridge Regression Model

Aditya Tadas (adityatadas8075@gmail.com)

o Data understanding
o Data visualization
o Multivariate analysis
o Gradient Boosting Model
o Decision Tree Model
o Hyperparameter Tuning XG Boost Regressor

Please paste the GitHub Repo link.

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

The problem statement was to build a machine learning model that could predict the rented bikes count required for an hour, given other variables. The first step in the exercise involved exploratory data analysis where we tried to dig insights from the data in hand. It included univariate and multivariate analysis in which we identified certain trends, relationships, correlation and found out the features that had some impact on our dependent variable. The second step was to clean the data and perform modifications. We checked for missing values and outliers and removed irrelevant features. We also create dummy variables for categorical features. The third step was to try various machine learning algorithms on our split and standardized data. We tried different algorithms namely; Linear regression, Lasso and Ridge Model, Decision Tree and Gradient Boosting. We did hyperparameter tuning and evaluated the performance of each model using various metrics. The best performance was given Hyper parameter tuning Gradient Boosting where the R2_score for train and test data was 0.99 and 094 rep.

The most important features who had a major impact on the model predictions were; hour, temperature, Humidity, solar-radiation, and Winter. Demand for bikes got higher when the temperature and hour values were more. Demand was high for low values of Humidity and solar radiation. Demand was high during springs and summer and very low during winters.

The model performed well in this case but as the data is time dependent, values of temperature, wind-speed, solar radiation etc. will not always be consistent. Therefore, there will be scenarios where the model might not perform well. As Machine learning is an exponentially evolving field, we will have to be prepared for all contingencies and also keep checking our model from time to time