Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Name - Vinay Kumar Email – vkvinay6923@gmail.com Contribution – 1. Loading and Understanding Data 2. Exploratory Data Anaysis a. Multi-variate Analysis 3. Feature Engineering 4. Train-test split 5. Models a. Models Training b. Error Plots 6. Hyper parameter Optimization 7. Conclusion Name – Nikhil Lamje Email – Nikhillamje123@gmail.com Contribution – 1. Loading and Understanding Data 2. Exploratory Data Anaysis a. Univariate b. Bivariate c. Multi-variate Analysis 3. Feature Engineering 4. Train-test split 5. Models a. Models Training b. Error Plots 6. Hyper parameter Optimization
Please paste the GitHub Repo link.
Github Link:- https://github.com/vkvinay0627/Bike-Sharing-Demand-Prediction---Capstone-Project

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

In this dataset, we have rented bike data for a company in the state of Seoul. in which our task is to predict the number of bikes needed at a particular hour in the day. The data set we have consists of 8750 records and 14 columns.

First, we started with data wrangling, checking all the null values that we did have in our data, and then we went for the outliers. All the independent columns in our data set were in ranges, and for dependent variables we performed a squared root transformation to remove the skewness.

After that, we performed exploratory data analysis and did univariate, bivariate, and multivariate analyses and saw some interesting relationships between the variables. For example, we found that bike riders opt for a specific time of day to rent their bikes, and that varies on weekdays and weekends.

Using a heat map, feature selection was done on certain independent features that appear to be highly co-related to one another. Then, on categorical features such as seasons and days, one hot encoding was performed. This makes the data suitable for splitting into tests and training before importing them to machine learning algorithms.

We used four machine learning models, which are: linear regression, decision tree, random forest, and XGBoost, on test and training data and evaluated them on the basis of performance metrics like mean absolute error (MAE), root mean square error (RMSE), and adjusted R squared. We found that decision tree and random forest were overfitting on train data, but XGBoost gave the best results on both train and test data, with an MAE of 9.82 and an adjusted R-squared value of 0.87.

After that, we performed hyperparameter optimization on XGBoost to fine-tune the results for the lowest possible error, and we achieved an MAE value of 4.68 and an adjusted R-squared value of 0.92. The application of the proposed model will be a win-win solution for both the bike company and the customers.