Capstone Project Submission

Supervised ML (regression) - Bike sharing demand prediction

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email, and Contribution:
Sandhya kumara Sah. E-mail:- sandhya.dis21@gmail.com • Data Exploration • Data Wrangling • Data Mining • Data Cleaning • Data Visualization • Observations • Summarization • Conclusions
Please paste the GitHub Repo link.
GitHub Link:- https://github.com/SandhyaSah22/Supervised-ML-regression---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) During the time of our analysis, we initially did EDA on all the features of our dataset. We first analyzed our dependent variable, 'Rented Bike Count' and also transformed it. Then we analyzed categorical variables and dropped the variable that had the majority of one class, we also analyzed numerical variables and found out the correlation, distribution, and their relationship with the dependent variable. We also removed some numerical features which had mostly 0 values and hot encoded the categorical variables. Next, we implemented 7 machine learning algorithms Linear Regression, lasso, ridge, elastic net, decision tree, Random Forest, and XG Boost. We did hyperparameter tuning to improve our model performance. The results of our evaluation are: • No overfitting is seen.

• Random forest Regressor and Gradient Boosting grid search cv give the highest R2 score of 99% and 95% respectively for Train Set and 92% for the Test set.

• Feature Importance values for Random Forest and Gradient Boost are different.

• We can deploy this model.

However, this is not the ultimate end. As this data is time-dependent, the values for variables like 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. Therefore, having quality knowledge and keeping pace with the ever-evolving ML field would surely help one to stay a step ahead in the future.