Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Sampanna Mishra - mishrasampanna1998@gmail.com

- Explaining Data Summary
- Data Preprocessing
- Exploratory Data Analysis
- Implementing ML Algorithms.
- Hyperparameter Tuning
- Conclusion

Please paste the GitHub Repo link.

Github Link:-
https://github.com/SampannaMishra/Bike-Sharing-Demand-Prediction

CONTINUED

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

Hyper urbanization, coupled with decentralization, has caused a host of transportation problems: gridlock, increasing travel demand,tailpipe emissions, and decreasing accessibility.While in recent years, dozens of cities have implemented bike share systems in efforts to mitigate some of these problems.Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated throughout a city.

In the problem the dataset provided is the Seoul Bike Sharing Demand.Using these Bike Sharing systems, people rent a bike from one location and return it to a different or same place on need basis.The main objective is to build a predictive model, which could help them in predicting bike count required at each hour for the stable supply of rental bikes.

Data Preprocessing:

- The dataset was checked for null and duplicate values.
- DataType of features were converted to appropriate form so that these could be used for further analysis.
- One hot encoding and label encoding was used on some categorical features.
- Outlier treatment and skewness was removed.

EDA:

- Line plots, Point plots, Bar plots, Density plot, Scatter plot, Regression plots were used to analyze relationships between dependent and independent variables.

Implementing ML Algorithms:

- Linear Regression model has been used.To enhance its results various regularization techniques have been used like Lasso,Ridge and ElasticNet.
- Datas have been trained and tested using Decision Tree also.To improve on its results ensemble techniques have been used like Random Forest and Gradient Boosting.

Hyperparameter Tuning:

- GridsearchCV hyperparameter tuning has been used over the gradient boosting algorithm.And the results were better than before.The accuracy of the test results had increased.

Evaluation Matrices:

- The accuracy and performance has been compared between the models using Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), R2 and Adjusted R2.

Conclusion:

Date column shares an important relationship with the rented bike count column.From visualization graphs it can be concluded that bike demand is greater on weekdays than on weekends. Feature importance graph it can be concluded that Temperature variable affects the most. Out of all the models Random forest gave the highest r2 score of 89 %.Also, there is much improvement in r2 score after hyperparameter tuning on gradient boosting.So the r2 score of our best model is 90% which can be said to be good for this large dataset.