Capstone Project Submission

Bike Sharing Demand Prediction

Project Summary

Contribution - Individual Project

Member - Onkar Anil Pawar

E-mail – pawaronkar177@gmail.com

Problem Statement :

Currently, Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of the bike count required at each hour for the stable supply of rental bikes.

Linear, Lasso, Ridge, and Elastic Net:

Linear, Lasso, Ridge, and Elastic regression models have almost similar R2 scores(61%) on both training and test data. (Even after using GridserachCV I have got similar results as of base models).

Decision Tree Regressor:

On the Decision tree regressor model, without hyperparameter tuning, I got an r2 score of 100% on training data, and on test data, it was very less. Thus our model memorized the data. So it was an overfitted model.

After hyperparameter tuning, I got an r2 score of 88% on training data and 83% on test data which is quite good for us.

Gradient Boosting Regression(Gradient Boosting Machine):

On the Random Forest regressor model, without hyperparameter tuning, I got an r2 score of 86% on training data and 85% on test data. Our model performed well without hyperparameter tuning.

After hyperparameter tuning I got an r2 score of 96% on training data and 91% on test data, thus I improved the model performance by hyperparameter tuning.

Thus Gradient Boosting Regression(GridSearchCV) and Random forest(gridSearchCv) give good r2 scores. We can deploy this model.

Conclusions:

As I have calculated MAE, MSE, RMSE, and R2 scores for each model. Based on r2 score will decide our model performance.

Our assumption: If the difference in R2 score between the rain data and the Test is more than 5 % we will consider it as overfitting.

Linear, Lasso, Ridge, and Elastic Net:

Linear, Lasso, Ridge, and Elastic Net regression models have almost similar R2 scores(61%) on both training and test data. (Even after using Grid search CV we have got similar results as of base models).

GitHub Link:- https://github.com/Onkar-TAE/bike-sharing-demand-prediction-analysis.git