Capstone Project Submission

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

Team Member's Name, Email and Contribution:
Name-Deepak Kumar Dubey Email- dkd31994@gmail.com
Please paste the GitHub Repo link.
Github Link:- https://github.com/dkd99/Seoul-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)
Problem statement- ● A **bicycle-sharing system**, **bike share program**, **public bicycle scheme** or **public bike share** (**PBS**) **scheme** is a shared transport service in which bicycles are made available for shared use to individuals on a short-term basis for a price or free. Many bike share systems allow people to borrow a bike from a "dock" and return it at another dock belonging to the same system. ● It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. ● The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes. ● We have around 8760 hours of observations for the whole year that cover all seasons and weather conditions to predict the demand of rental bikes at any particular hour using supervised machine learning algorithms. ● This is a regression problem that comes under supervised machine learning.

Approach-

This problem has been solved following these steps-

- Data understanding
- Data exploration
- Data preprocessing
- Feature engineering
- Model training
- Model validation
- Error analysis-Mean square error,mean absolute error,r2 score for test and train score,adjusted r2 score were calculated and compared for various models.

Conclusion-

1)We observed that the bike rental count is higher during no holidays than during holidays.

2)The rental bike count was at its peak at 8 AM in the morning and 6pm in the evening, an increasing trend can be observed from 5am to 8am, the graph touches the peak at 8am in the morning then dips a bit. Later we can see a gradual increase in the demand until 6pm, the demand is highest at 6 pm, and reduces until midnight.

3)People prefered to rent bikes when the temperature was between 20 degrees to 35 degrees celsius and even when it was a little windy.

4)Highest bike rental count was found in Autumn and summer seasons and the lowest is in the winter season.

5)Bike rentals were highest during clear days and lowest on snowy and rainy days.

6)When there was not a functioning day no bikes were booked.

7)Linear models such as linear regression ,lasso,ridge are performing good on both train and test data.R2 score is around 0.8 for both train and test set.

8)Polynomial regression is giving better results than linear models and giving a good r2 score on both test and train set.r2 score is around 0.85 for both train and test sets.

9)Ensemble models are giving quite good results on train data i.e R2 score more than 0.9 but are a little bit overfit i.e R2 score of 0.86-0.88 on test data.

10)Based on the feature importance plot, Temperature,whether day was functioning or non functioning,6:00pm evening time ,seasons were most important parameters in predicting the demand.