Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1) Satyam Jyoti Sankar
E-mail: satyamjyoti99@gmail.com

- Data analysis
- Approach towards plan.
- Data visualization
- Model Building
- Technical documentation.
- Project summery template.

2) Krushnagopal Brahma
E-mail: krbrahma54@gmail.com

- Data analysis.
- Data Visualization
- Feature Engineering
- Frame work of project.
- Model building.
- PPT presentation

Problem definition:

It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. This project tackles the problem of predicting the number of bikes which will be rented at any given hour in a given city. The main objective is to build various regression models and analyze their performance with respect to each other so as to get the best performing model, which could help in predicting the number of bikes required at each hour for the stable supply of rental bikes.

EDA on given Data set :

Digging into data we understand that

1. There is no null and missing value in the data set.
2. There are mainly 14 Features.
3. Dependent variable should be considered as Rented Bike count.
4. There are Factors affecting the dependent variable as seasons, Hours , Temperature, Weather , Working days and Holiday.
5. The heatmap gives good correlation between the Feature.

Models used :

- Linear regression
- Lasso regression
- Ridge regression
- Elastic net
- Decision Tree
- Random Forrest
- XGboost

Conclusion :

- As the temperature increases the more people use the rented bike vice versa the increase in snowfall the rented bike count decrease which clearly indicates the season and weather play role in the demand of rented bike.
- Bike count rent is highly correlated with 'Hour', which seems obvious. Demand for bike is mostly in morning (7 to 8) and in the evening (3 to 9) as people go to the work and returns the home.
- Lasso and Ridge Regression helps to reduce model complexity and prevent over-fitting which may result from simple linear regression, with Lasso, ridge and ElsasticNet regression We got r squared value of 0.75 each.
- When we compare the root mean squared error and mean absolute error of all the models, Random_forest and Gradient Boosting gives the highest R2 score ending with the accuracy of 89% . So, finally this two model are best for predicting the bike rental count on daily basis.

Please paste the GitHub Repo link.

Github Link:- https://github.com/brahma102/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)