Capstone Project Submission

Team Member's Name, Email and Contribution:

1) **Balaji B. Jadhav**

   E-mail: bjadhav015@gmail.com

   - Data sorting.
   - Approach towards plan.
   - Graphical representation.
   - Bar plot and Heat map.
   - Model selection and implementation.
   - Cross validation on model.
   - Implementation of grid parameters.

2) **Anant M. Patil**

   E-mail: mr.anantmpatil@gmail.com

   - Data visualization.
   - Sorting of values.
   - Pi-plot and Heat map.
   - Various model implement.
   - Suggestion of grid parameters.
   - Project summery template.

3) **Nishigandha Ingale**

   E-mail: Nishiingale28@gmail.com

   - Data analysis.
   - Approach towards multiline graph.
   - Frame work of project.
   - Histogram plot.
   - Model presentation.
   - Analyzing results of model.

Problem definition:

The main goal of the project is to Finding factors and cause those influence shortages of bike and time delay of availing bike on rent. Using the data provided, this paper aims to analyze the data to determine what variables are correlated with bike demand prediction. Hourly count of bike for rent will also be predicted.

EDA on given Data set:

Digging into data we understand that

- There is no null value in the data set.
- Dataset contains 13 independent features such as Date, Hour, Temperature, Humidity, Wind speed, Visibility, Dew point temperature, Solar Radiation, Rainfall, Snowfall, Seasons, Holiday, Functioning Day.
- Total 14 columns with values such as float, integer and object.
- Dependent variable should be considered as Rented bike count.
- Graphical representation according to various columns and with manipulation of columns.

Model selection and implementation:

By looking at distribution of data we can say that the data of dependent variable is skewed at the time of split we are going to apply it transformation. Various types of linear model are implemented on data such as Linear regression, Ridge regression, Lasso regression, Elastic net regression and Polynomial feature regression. As we can see from table the result accomplished are good but not satisfying. So we have to move towards some complex models such as Decision tree regressor, Random forest regressor, XG boost regressor and Cat boost regressor from these models we have good results and it can be said that after implementation of all models best results are obtained from XG boost with application of Grid search CV on that.

Model	MSE (Train)	RMSE(Train)	R2(Train)	MSE (Test)	RMSE(Test)	R2(Test)
Linear Regression	115071.02	339.22	0.72	112112.97	339.22	0.72
Ridge	115073.23	339.22	0.73	112092.51	334.80	0.72
Lasso	115071.14	339.22	0.72	112102.68	334.81	0.72
Elastic Net	115080.73	339.23	0.72	112074.51	334.77	0.72
Polynomial Feature	29566.78	171.94	0.92	40725.48	201.80	0.90
Decision tree regressor	0.0	0.0	1.0	63029.50	251.05	0.84
Random forest regressor	4811.62	69.36	0.98	33822.95	183.91	0.91
XG boost	10073.27	100.36	0.97	36204.99	190.27	0.97
Cat boost	101028.40	317.84	0.75	104605.09	323.42	0.75

Conclusion

Bicycle sharing systems can be the new boom in India, with use of various prediction models the ease of operations will be increased. The Nine algorithms are applied on the bike share dataset for predicting the count of bicycles that will be rented per hour. We got some good results and accuracy with random forest and XG boost. The accuracy and performance have been compared between the models using Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), R2 and Adjusted R2. If these systems include the use of analytics the probability of building a successful system will increase

Please paste the GitHub Repo link.

Github Link:- https://github.com/BalajiJadhav015/Bike-Sharing-Demand-Prediction.git

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