Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Ankit Sharma
Mail id ankit.sharma.in1088@gmail.com
Contribution:-

- Finding Column and shape of data
- Describing data
- Finding Duplicate value in data
- Breaking date column into weekdays and weekends
- Visualization by weekend
- Visualization of rented bike on weekend
- counting of rented bike on the basis of seasons
- Assigning and plotting numerical value into variable
- Finding relation between rented bike and temperature
- Plotting regression plot
- Finding correlation between variables using heatmap
- Model training
- Lasso Regression modelling
- Elastic net regression
- Random forest classification
- Hyperparameter tuning

PANKAJ KUMAR
Mail id erpankaj.kumar227@gmail.com
Contribution:-

- Finding Head and tail of data
- Finding Info of data
- Finding Unique values in data
- Finding Missing value and percentage of missing data
- Changing Column name
- Visualization of data by month
- counting of rented bike on the basis of hour
- counting of rented bike on the basis of Functioning day
- counting of rented bike on the basis of holiday
- Relation between count of rented bike and solar radiation and snowfall and rainfall and windspeed
- Data preprocessing for ML
- Correlation between variables
- One hot encoding
- linear regression modelling
- Ridge regression modelling
- Decision tree modelling
- Gradient Boosting

Please paste the GitHub Repo link:- https://github.com/ankit1088/Project-2-Bike-Sharing-Demand-Prediction

GitHub Link: - https://github.com/ankit1088

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

With the help of data that we got we prepared exploratory data analysis, visualizations, interactive plots, model training on Seoul bike sharing demand prediction.
In this dependent variable is rented bike count rest all are independent variable and target variable is rented bike counts..

We started with data preprocessing by checking null values, missing values and then we dd model training, checked for any duplicate values, performed EDA on target variable with respect to independent variable and concluded EDA part.
Secondly we performed train test split and did model training with the help of Linear regression, Lasso regression, Random forest, Decision trees, and finally evaluated the result of the model with the help of different evaluation metrics.

Conclusion:-

1.Hour of the day holds the most important feature.
2.Bike rental count is mostly correlated with the time of the day as it is peak at 10 am morning and 8 pm at evening.
3.We observed that bike rental count is high during working days than non working day.
4.We see that people generally prefer to bike at moderate to high temperatures, and when little windy
5.It is observed that highest number bike rentals counts in Autumn & Summer seasons & the lowest in winter season. We observed that the highest number of bike rentals on a clear day and the lowest on a snowy or rainy day. We observed that with increasing humidity, the number of bike rental counts decreases.
6. R squared Error of random forest regressor is- 0.99 which explains 99 percent of the variance of data so model is performing very good