**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Abhishek Singh Rawat(abhisheksinghrawat12@gmail.com)**  * Analyzed Columns name and relation between the columns. * Understanding the variables. * Loading Data. * Cleaning Data. * Understanding the correlation between the graphs. * Feature Analysis * Hyperparameter Tuning * Feature Engineering * Linear Regression Modeling * Gradient Boosting  **2. Jatin(**[**guptajatin281@gmail.com**](mailto:guptajatin281@gmail.com)**)**  * **Analyzed all the columns.** * **Acquire and loading data.** * **Understanding the variables.** * **Exploring and visualizing.** * **Random Forest** * **Decision Tree** * **Multivariate Analysis** * **Data Understanding** * **Lasso Regression**   **Abhilasha.M(abhilasha.m03@gmail.com)**  * Performed EDA * Plotted different types of graphs * Data Understanding * Data Visualization * Elastic Net Regressor * Research Analytics * Ridge Regression |
| **Please paste the GitHub Repo link.** |
| Jatin’s GitHub Link - https://github.com/Jatingpt |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **About the Data Set:** The dataset contains weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.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 bike count required at each hour for the stable supply of rental bikes. This Data Set consists of 8760 Number of Rows and 14 Number of Columns.  Approaches we used:   1. **Loading the data:** In this section we just loaded the data in the Colab Notebook and read csv file. 2. **Data Cleaning and Processing**: In this section we have tried to remove null values and replaced all the null values by some integer values and with some appropriate values. 3. **Analysis and Visualisation**: In this section we have tried to explore all variables which can play an important role for the analysis. In the next part we have to explore the effect of one over the others. 4. **Future scope for further analysis**: The demand of the rented bikes are very low specially in the morning hour but when the evening start from 4pm to 8pm then the demand slightly increases. Generally **people use rented bikes during their working hour from 7am to 9am and 5pm to 7pm.**In the functioning day the rented bikes used by the people are very low or I can say that people don’t use rented bikes on functioning day.   **In summer season the use of rented bike is high and peak time is 7am-9am and 7pm-5pm.**  **In winter season the use of rented bike is very low because of snowfall.**    Conclusions: During the time of our analysis, we initially did EDA on all the features of our Data Set. We first analysed our dependent variable, 'Rented Bike Count' and also transformed it. Next we analysed categorical variable and dropped the variable who had majority of one class, we also analysed numerical variable, found out the correlation, distribution and their relationship with the dependent variable. We also removed some numerical features who had mostly 0 values and hot encoded the categorical variables.  Next we implemented 7 machine learning algorithms Linear Regression, Lasso, Ridge, Elastic Net, Decision tree, Random Forest and XGBoost. We did hyperparameter tuning to improve our model performance. |