**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| Name - V Bhavya Reddy  Email - [bhavya.reddy0711@gmail.com](mailto:bhavya.reddy0711@gmail.com)  Contribution - whole project. |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/coolphotography/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)** |
| **problem statement:**  The goal of this project is to provide a model to predict stable supply of bike rentals to predict demand at any hour. This dataset contains hourly bike rental counts along with weather information and dates spanning an entire year.  **Approach:**  We begin by reviewing the data set overview, in which we briefly analyze the observations and note several characteristics of numeric and categorical. In addition, we find the number of NULL values - none were found.  The next step is Exploratory Data Analysis (EDA) on the dataset, where we analyze numeric and categorical features through bar charts, box plots, countplots, etc. We also use heatmaps to analyze correlation, and VIF calculations are carried out to check for multicollinearity.  In feature Engineering, we eliminate the outliers and drop unnecessary columns, and modify the dataset. Data encoding followed with train-test split are performed on the data set.  Using train data we train our models using Linear Regression, Lasso Regression, Ridge Regression, Decision Tree, and Random Forest Regression. So, a total of five regression models were used to predict the rented bikes per hour: Linear Regression, Lasso Regression, Ridge Regression, Decision Tree, and Random Forest Regression. We found that the Random Forest Model had the best/lowest RMSE of all the models. Ridge Regression and Lasso Regression models did not provide any improvement over Linear Regression. In order to gain a better understanding of the data, several plots are drawn (MSE, R2 score, Actual vs predicted, important features, etc.).Among all the 5 models, Random Forest Model has the best metric analysis. |