**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email, and Contribution:** |
| **Team Member’s Role:-**   * **Debashish Das**   **E-Mail-** devashishdas40@gmail.com   * + Data Understanding   + Feature Analysis   + Feature Engineering   + Linear Regression modeling   + Random forest   + Gradient Boosting   + Hyperparameter tuning * **Lucky jain**   **E-Mail-jainluckycool@gmail.com**   * + Data Understanding   + Feature Analysis   + Data Visualization   + Multivariate Analysis   + Decision Tree   + Elastic Net Regressor   + Research Analytics * Technical documentation * **Vivek katolkar**   **E-Mail-** [**balarampanigrahy42@gmail.com**](mailto:balarampanigrahy42@gmail.com)   * + Data Understanding   + Data Visualization   + Multivariate Analysis   + Lasso Regression   + Ridge Regression   + Research Analytics * Technical documentation |
| **Please paste the GitHub Repo link.** |
| Github Link:- |
| **Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)** |
| The information was sourced from Seoul, which is a city. A bike-sharing system is a service that makes bikes available to people for short-term, shared use that can be paid for or provided for free. Many bike share programmers enable users to pick up a bike from a "dock," which is typically computer-controlled and where they enter their payment details to have the bike unlocked. Then, you can return this bike to a different system-affiliated dock. Date, hour, temperature, humidity, wind speed, visibility, dew point temperature, solar radiation, rainfall, snowfall, seasons, holiday, working day, and the number of rental bikes were among the factors in the data.  The goal was to create a machine learning model that, given other data, could forecast the number of leased bikes needed for an hour. Exploratory data analysis was the initial step in the activity, during which we attempted to glean insights from the available data. With the help of univariate and multivariate analysis, we were able to pinpoint certain patterns, connections, and correlations as well as identify the characteristics that may have had an influence on our dependent variable. The data needed to be cleaned up and changes made in the second stage. We eliminated pointless features and looked for missing values and outliers. The categorical variables were also encoded. On our split and standardized data, we ran many machine learning algorithms in the third phase. Several algorithms, including Linear Regression, Random Forest, and The most important features who had a major impact on the model predictions were; hour, temperature, wind-speed, solar-radiation, month and seasons. Demand for bikes got higher when the temperature and hour values were more. Demand was high for low values of wind-speed and solar radiation. Demand was high during springs and summer and very low during winters.  However, because the data is time-dependent, values of temperature, wind speed, sun radiation, etc., would not always be constant. The model functioned well in this instance. As a result, there will be situations in which the model may not work well. We must be ready for all eventualities and periodically evaluate our model because machine learning is a field that is rapidly changing. |
| Drive link: [https://drive.google.com/drive/folders/1ig8mXKKZ9mB4oBZiQarEUFS9uqXeK8FZ?usp=sharing](#_top) |