**Capstone Project Submission Bike Sharing Demand Prediction**

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Team Members: Email**  **1). Kartike -   kknirajkumar@gmail.com**  **2). Animesh Chakraborty - animmeshchakraborty.227@gmail.com**  **3). Anupam Mishra - mishraanu1991@gmail.com**  **Contributor Roles:**  **1). Kartike:**  **A). Explore the data**  **B). Preprocess the data**  **C). Check assumptions for Linear Regression**  **D). Data Splitting**  **E). Building different Models**  **F). Conclusion**  **2). Animesh Chakraborty:**  **A). Explore the data**  **B). Preprocess the data**  **C). Check assumptions for Linear Regression**  **D). Data Splitting**  **E). Building different Models**  **F). Conclusion**  **3). Anupam Mishra:**  **A). Explore the data**  **B). Preprocess the data**  **C). Feature Selection**  **D). Data Splitting**  **E). Building different Models**  **F). Conclusion** |
| **Please paste the GitHub Repo link.** |
| GitHub Link:- <https://github.com/Animesh227/Seoul-Bike-Sharing-Demand-Prediction> |

|  |
| --- |
| **Summary** |
| 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.  The dataset contains weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.  Steps for project:  1)Exploratory Data Analysis (EDA): In this part we have done some EDA on the features to see the trend.  2)Data Processing: In this part we went through each attribute and encoded the categorical features.  3) Model Creation: Finally in this part we created the various models. These various models are being analyzed and we tried to study various models so as to get the best performing model for our project.  In the project we have observed the Model Evaluation Matrices table, Linear Regression is not giving great results. Random forest & GBR have performed equally good in terms of adjusted r2.  We started with loading the data, then we did Exploratory Data Analysis (EDA), null values treatment, feature selection, encoding of categorical columns, and then model building. In all of these models, our accuracy ranges from 56% to 89%, which can be said to be good for such a large dataset. This performance could be due to various reasons like the proper pattern of data, large data, or because of the relevant features.  After performing variable importance analysis to find the most significant variables for all the models developed with the given data sets. We are getting the best results from Random Forest and Gradient Boosting |