**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1)Name :Vikrmaditya Sah  Email id : [vsah625@gmail.com](mailto:vsah625@gmail.com)  Contributions : 1)Data wrangling  2)EDA and Visualization parts  3)Contribution in PPT  2)Name : Sivaramaguhan S  Email id : [ssivaramaguhan@gmail.com](mailto:ssivaramaguhan@gmail.com)  Contributions : 1)Machine learning part  2)Technical documentation  3) Contribution in PPT |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/vikram0050/Bike-Sharing-Demand-Prediction |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| Seoul Bike is providing the city with a stable supply of rental bikes. It becomes a major concern to keep users satisfied. The crucial part is the prediction of bike count rents at each hour for a stable supply of rental bikes. It is important to make the rental bike available and accessible to the public, as it provides many alternatives to commuters in metropolises. Furthermore, it is the healthiest way to travel and it has many environmental benefits.  The studied dataset contains weather information which are the features (Temperature, Humidity, Wind speed, Visibility, Dew point, Solar radiation, Snowfall, Rainfall), the target is the number of bikes rented per hour and date information. The dataset presents the company's data between December the 1st of 2017 and finishes one year later.  This study shows that the rents of bikes are influenced by a lot of features. In this study, we understood that many people usually and mainly rent bikes during the weekdays, so we supposed that the main use is to go to school or work.  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 91%, 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.  We performed 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 CatBoost and LightGBM. |