**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Name: -** **Viral Bhatu Shewale**  **Email:-** [viralshewale1997@gmail.com](mailto:viralshewale1997@gmail.com)  **Capstone Project:- Seoul Bike Sharing Demand Prediction** |
| **Please paste the GitHub Repo link.** |
| GitHub Link: - <https://github.com/vir097/Bike-Sharing-Demand-Prediction---viral-shewale.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)** |
| **Seoul Bike Sharing Demand Prediction: -**  **Problem statement: -**   * 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 the bike count required at each hour for the stable supply of rental bikes. * The main objective is to build a predictive model, which could help them in predicting the demand for rental bikes. This would in turn help them in terms of business development.   **Approach: -**  **Exploratory Data Analysis: -**   * It is a good practice to understand the data first and try to gather as many insights from it. EDA is all about making sense of data in hand, before getting them dirty with it. * In Exploratory Data Analysis I performed initial investigations on data to discover patterns, spot anomalies, test hypotheses and check assumptions with the help of summary statistics and graphical representations.     **Data Visualization: -**   * Used different visualization methods (like Scatter plot, Bar plot, Pia chart, Distplot, Box plot etc.) to understand data more clearly.   **Encoding of categorical columns:**   * The categorical features, which are in string format cannot be understood by the machine and needs to be converted to the numerical format. Hence, I used one hot encoding to encode categorical features in binary integers 0 and 1.   **Fitting different models:**   * For modelling, I tried various regression algorithms like:   **1.** **Decision Tree**  **2. Random forest regression**  **3. XG Boost regression**  **Hyperparameter Tuning for better accuracy:**   * Tuning the hyperparameters of respective algorithms is necessary to get better accuracy, avoid overfitting and lead to a better-generalized model.   **Model evaluation:**   * Evaluating the performance of the built model using different metrics.   **Conclusion: -**   * The demand for rental bikes was **highest in the summer** season and **lowest in the winter** season. * **May-July are peak months** to rent a bike. Dec-Feb is the least preferred month for bike renting. * The rental bike demand was more on a weekday than on weekends. The majority of **clients belong to the working class.** * The **temperature of 20-30 Degrees**, evening time 4 pm- 8 pm and the **humidity between 40%-60%** are the most favourable parameters where the Bike demand is at its peak. * **Temperature, humidity, hour of day, solar radiation and functional day** are major driving factors for the bike rent demand. * The **XG Boost** model has the **lowest**    test **MAE**. A low MAE value indicates that the simulated and observed data are close to each other and show better accuracy. Thus **lower MAE is better for model performance**. (XG Boost model with an **accuracy of**88.52**%**). |