**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email, and Contribution:** |
| * **Sk Samim Ali –** [**sksamima5@gmail.com**](mailto:sksamima5@gmail.com) **(Gmail)** * [**https://www.linkedin.com/in/samim-ali-62508b1b8/**](https://www.linkedin.com/in/samim-ali-62508b1b8/) **(LinkedIn)**   **Contribution :**   * **Data Understanding** * **Feature Analysis** * **Feature Engineering** * **Linear Regression modeling** * **Random forest** * **Gradient Boosting** * **Hyperparameter tuning** * **Technical Documentation** * **Mohd. Izhar -** [**mohdizhar678@gmail.com**](mailto:mohdizhar678@gmail.com) **(Gmail)** * [**https://www.linkedin.com/in/izhar-ahmad-38bb4a124**](https://www.linkedin.com/in/izhar-ahmad-38bb4a124) **(LinkedIn)**   **Contribution :**   * **Data Understanding** * **Data Visualization** * **Multivariate Analysis** * **Lasso Regression** * **Ridge Regression** * **Research Analytics** * **Technical Documentation** * **Sarath Haridas -** [**sarathhrds@gmail.com**](mailto:sarathhrds@gmail.com) **(Gmail)** * [**https://www.linkedin.com/in/sarath-haridas-596ba2196**](https://www.linkedin.com/in/sarath-haridas-596ba2196) **(LinkedIn)**   **Contribution :**   * **Data Understanding** * **Feature Analysis** * **Data Visualization** * **Multivariate Analysis** * **Decision Tree** * **Elastic Net Regressor** * **Research Analytics** * **Technical documentation** |
| **Please paste the GitHub Repo link.** |
| **GitHub Link: -**  [**https://github.com/Sarath-Haridas/Capstone\_Project\_2\_Bike\_sharing\_demand\_prediction**](https://github.com/Sarath-Haridas/Capstone_Project_2_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)**  **The information was sourced from Seoul, which is a city in South Korea. 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-sharing programs let users borrow bikes from "docks," which are typically computer-controlled and open when the user enters their payment information. 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, functioning days, and the number of rental bikes were among the factors in the data.**  **The goal is to create a machine-learning model that can forecast the number of leased bikes needed for an hour, given other factors. We tried to delve further into the data at hand in the first part of the activity, which comprised exploratory data analysis. We discovered certain patterns, connections, and correlations as well as the characteristics that had some influence on our dependent variable using univariate and multivariate analysis. The data needed to be cleaned up, then changes were made. We looked for missing values and outliers, and then we eliminated pointless features. The categorical variables were encoded as well. The final phase involved testing different machine-learning algorithms on our separated and standardized data. We experimented with several methods, including XGBoost, Random forest, and Linear Regression. Hyperparameter tuning was done, and several metrics were used to assess each model's performance. Gradient boosting and Random forest models delivered the greatest results, with R2 scores for the training and test sets of 0.93 and 0.91 for Gradient Boosting, and 0.84 and 0.81 for Random forest respectively.**  **The hour, temperature, wind speed, solar radiation, month, and seasons were the most significant variables that significantly affected the model's predictions. When the temperature and hour values were greater, demand for motorcycles increased. Low values of wind speed and solar radiation were in high demand. Demand peaked in the spring and summer and plummeted in the winter.**  **Although the model did well in this instance, the values of temperature, wind speed, solar radiation, etc. won't always remain constant as the data is time-dependent. As a result, there will be situations, where the model may not work well. We must be ready for every scenario and periodically evaluate our model because machine learning is a field that is rapidly changing.** |
|  |