**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| 1. **ANAS MUSTAFA:**   **Email:** [**Mustafaanas84464@gmail.com**](mailto:Mustafaanas84464@gmail.com)  **Contribution:**   * **Data Wrangling** * **Data Cleaning** * **EDA** * **Total rented bikes in 2017 and 2018** * **Effects of numerical features in rented bike count** * **Finding correlation between different variables** * **Linear Regression** * **Random Forest Regression** * **XGboost Regression** * **Important Features**  1. **SARTHAK RASTOGI:**   **Email:** [**sartakrastogi1@gmail.com**](mailto:sartakrastogi1@gmail.com)  **Contribution:**   * **Data Wrangling** * **Data Cleaning** * **Checking distribution of our dependent variable** * **Exploring numerical Variables** * **Average bike count per hour** * **Ridge Regression** * **Elastic Net Regression** * **Gradient Boosting Regression**  1. **CHETAN RAJPUT:**   **Email:** [**Chetan.rajput91@yahoo.com**](mailto:Chetan.rajput91@yahoo.com)  **Contribution:**   * **Data Wrangling** * **Data Cleaning** * **Total rented bikes on monthly basis** * **Rented bikes count in different seasons** * **Exploring categorical Variables** * **Finding Correlation of dependent variables with other variables** * **Lasso Regression** * **Decision Tree Regression** * **Hyperparameter Tuning** |
| **Please paste the GitHub Repo link.** |
| Github Link:- |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  We had completed our capstone project on seoul bike share prediction. As more number of rented bikes are being used in the cities nowadays, it becomes important for the company to predict the number of required rental bikes required across a day so that no demand supply gap would be generated for rental bikes. This project aims at providing necessary solution to predict the rental bikes demand using machine learning algorithms so that all the stakeholders of the business can be satisfied.  **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 bike count required at each hour for the stable supply of rental bikes.  The present scenario is about how good is the customer service is in any industry as the number of options at the customer’s disposal are unlimited. So, it becomes extremely important to make sure that the customers will not be made to wait for the rental bikes. It would also not be practical to keep a lot of bikes even when the demand is low. Hence , with the help of machine learning, this project aims at predicting the rental bike  demand so that no problems arise.  **Steps:**  The first step of our project is performing the EDA process on the dataset so that we can get the idea about the dataset i.e. the number of variables, the data type of the variables visualize the dataset for better understanding and decide the suitable methods and algorithms that might produce desired outcome.In EDA process we find the type of dataset and decide the approach, in this project the preprocessing steps would removing the punctuations, stopwords , generate count vectorizer and document term  matrix which would help in building up the model. After the data preprocessing is done then the data will be ready to be fit into machine learning models.For current problem statement topic modeling approach would be suitable .In topic modeling, a topic is defined by a cluster of words with each word in the cluster having a probability of occurrence for the given topic, and different topics have their respective clusters of words along with corresponding probabilities.  **Conclusion:**  The project comes to an end at this point. Beginning with loading the dataset, so far we have done EDA, pre-processing the data, Label encoding, Scaling the data, splitting the data into train and test data, applying various machine learning algorithms followed by hyper parameter tuning. We implemented 8 M.L. models. After comparing the mean square error and mean root square error of all the models, XGBoost has least mean square error and root mean square error. XGBoost has the highest accuracy of 91.9% among all algorithms. So, We can conclude that XGBoost is the best model to predict rented bike count. The number of business hours of the day and the demand for rented bikes were most correlated and It makes sense also. Highest number of bikes rented at the 18th hour of day. Total number of bike counts increased when there was favorable temperature. So, this can be an important factor in predicting underlying patterns of rented bike count. |