**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| ***SARANYA . N***  ***email.id –*** [***snk4411@gmail.com***](mailto:snk4411@gmail.com)  ***Contribution:***   * ***Data understanding*** * ***Feature analysis*** * ***Feature engineering*** * ***Data visualization*** * ***Implementing algorithms*** * ***Presentation slide making*** * ***Technical documentation*** * ***Project report making*** |
| **Please paste the GitHub Repo link.** |
| Github Link:- [BIKE-SHARING-DEMAND-PREDICTION---CAPSTONE-PROJECT/Bike\_Sharing\_Demand\_Prediction\_ML\_Capstone\_Project.ipynb at main · SARANYA-4411/BIKE-SHARING-DEMAND-PREDICTION---CAPSTONE-PROJECT (github.com)](https://github.com/SARANYA-4411/BIKE-SHARING-DEMAND-PREDICTION---CAPSTONE-PROJECT/blob/main/Bike_Sharing_Demand_Prediction_ML_Capstone_Project.ipynb) |
| **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 contents of the data came from a city called Seoul. A bike-sharing system is a service in which bikes are made available for shared use to individuals on a short term basis for a price or free. Many bike share systems allow people to borrow a bike from a "dock" which is usually computer-controlled wherein the user enters the payment information, and the system unlocks it. This bike can then be returned to another dock belonging to the same system 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. Attribute Information: Date : year-month-day Rented Bike count - Count of bikes rented at each hour Hour - Hour of the day Temperature-Temperature in Celsius Humidity - % Wind Speed - m/s Visibility - 10m Dew point temperature - Celsius Solar radiation - MJ/m2 Rainfall - mm Snowfall - cm Seasons - Winter, Spring, Summer, Autumn Holiday - Holiday/No holiday Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)*  *The goal of the project is to predict the number of rental bikes required at each hour for stable supply of rental bikes. Various Regression machine learning algorithms have been applied on the dataset to get the best possible prediction. Some of the key steps in the exercise involved EDA . Performed Exploratory Data Analysis on the data to gain some insights. . It included univariate and multivariate analysis in which we identified certain trends, relationships, correlation and found out the features that had some impact on our dependent variable.Next we have to clean the data and perform modification . also check duplicates and missing values and outliners also removed irrelevant features . We also encoded the categorical variables. The third step was to try various machine learning algorithms on our split and standardized data. We tried different algorithms namely; Linear regression, Randomforest and XGBoost. We did hyperparameter tuning and evaluated the performance of each model using various metrics. The best performance was given by the Gradient boosting and Random forest model where the R2\_score for training and test set was 0.90 and 0.84 respectively.*  *In this exercise we can understand that hour , temperature , wind speed, solar radiation , month, and seasons have important roles while sharing bike . Demand of bike is higher in the morning and evening , also in summer seasons.*  *However ,these results are not the ultimate . as this data is time dependent , the values for variables like temperature, solar\_radiation, wind\_speed etc., Will not always be consistent. Therefore, there will be scenarios where the model might not perform well. As machine learning is an exponentially evolving field, we will have to be prepared for all contingencies and also keep checking our model from time to time. Therefore, having a quality knowledge and keeping pace with the ever evolving ML field would surely help one to stay a step ahead in future* |