**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Sneha Raikar**  **Email:** [**sneharaikar650@gmail.com**](mailto:sneharaikar650@gmail.com)  **Individual project** |
| **Please paste the GitHub Repo link.** |
| GitHub Link : <https://github.com/sneraikar/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)** |
| **Introduction:** Today, bike-sharing systems are blooming across more than 1000 cities around the world, particularly in big or large cities like New York City, Paris, Washington DC, London, Beijing, and Barcelona. To complete a short trip renting a bike is a faster way when compared to walking. Moreover, it is eco-friendly and comfortable compared to driving.  Due to global warming, continuous pollution, and depletion of sources of energy. Many countries have been focused on using renewable energy which does not harm the environment and can be reused as well. South Korea is one of the countries which has adapted to it and their most used service is rented bikes in Seoul. But in order to avoid any difficulties such as waiting time it is necessary to have an estimate of future demand. Our goal here is to build a model that can predict bike sharing demand considering all the factors which have their effects. **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. **Approach**: Here first we imported a data set and performed EDA where we got valuable insights and further, we Encoded the Categorical Columns, Feature scaling and fitting into the models. At first, we tried with basic linear regression and with Lasso regularization technique but soon realized we will need a much more complex model and so we then used a Decision tree Regressor, XGB Model and Random Forest Regressor compared the results. |
| **Conclusion:**   * Most numbers of Bikes were rented in **Summer**, followed by **Autumn**, **Spring**, and **Winter**. **May-July** is the **peak** Bike renting **Season**, and **Dec-Feb** is the **least preferred** month for **bike renting**. * **Majority** of the **client** in the **bike rental sector** belongs to the **Working class**. This is evident from EDA analysis where bike demand is more on weekdays, working days in Seoul. * **Temperature** of **20-30 Degrees**, evening **time 4 pm- 8 pm**, **Humidity** between **40%-60%** are the most favorable parameters where the Bike **demand** is at its **peak**. * **Temperature**, **Hour of the day**, **Solar radiation**, and **Humidity** are major **driving factors** for the Bike rent **demand**. * Feature and Labels had a weak linear relationship, hence the prediction from the linear model was very low. **Best predictions** are obtained with a **Random forest** model with an accuracy of **0.875**. |