**Capstone Project Submission: Bike Sharing Demand Prediction**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Contributor Name:** Ranajay Biswas  **Email ID:** - ranojoybiswas21@gmail.com  **Contributor Role:**   * Data collection * Descriptive Statistics * Data Cleaning & Pre-processing * Graphical Representation & Data Visualization * Feature Selection & Scaling * Model Building & Implementation * Hyper-parameters Tuning |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/RanojoyBiswas/Play-Store-App-Review-Analysis-  Google Drive Link:-  https://drive.google.com/drive/folders/1oeJz49GOuPqAy0HNN\_0LCQjlhexovFue?usp=sharing |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| 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.  To predict a continuous variable such as bike count, we needed to use regression models. But before doing that, we performed exploratory data analysis on the Seoul bike sharing dataset. Checked the distributions for different variables, their data-types etc.  We checked for null values and duplicates in the data. We transformed features into new and more useful sets of features making sure that we were not ignoring any important patterns of any kind.  We checked for multicollinearity in the data and removed any features which were showing multicollinearity and also removed features that were not useful for prediction.  There were few features that needed one hot encoding, some features were transformed by making dummy variables. Data scaling was also done on the input variables to make sure all the input values were on the same scale so that gradient descent can converge faster and distance based algorithms can work better. Scaling is done in general just to get better model performance.  Finally this scaled data was passed through several regression based machine learning models. Then we compared the models’ results.  Hyperparameters tuning was done using Grid Search Cross Validation. Then we picked the best performing model which in our case was XGBoost.  With that, the problem of bike demand prediction was solved. During this exercise, we also understood the relations between different features and the dependent variable and how they impact bike rentals. |