**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| RAHUL SINGH WALDIA  [Rahulwaldia2@gmail.com](mailto:Rahulwaldia2@gmail.com)   1. DATA MINING 2. GIS 3. PREDICTIVE ANALYTICS 4. INTELLIGENT TRANSPORT SYSTEM |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/RahulSinghWaldia/BIKE-SHARING-DEMAND.git |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| In a span of few decade, the sharing of bicycle system has seen enamours growth (Fishman, [2016](https://www.tandfonline.com/doi/full/10.1080/22797254.2020.1725789)). This system is a recently developed transportation system which provides people with bicycle for common use. Bicycle system provides user to rent a bike from one docking station, where user can ride and then return in another docking station. Today it is much easier for the public to rent bicycles. Global Positioning System enabled mobile application allows people to know the nearby bicycle station for renting the bicycle. Many countries have bike sharing system, such as Ddareungi is a bike sharing system in South Korea, which started in the year 2015, known as Seoul bike in English. It was started to overcome issues like greater oil prices, congestion in traffic and pollution in the environment and to develop a healthy environment for citizen of Seoul to live. In recent days, Pubic rental bike sharing is becoming popular because of is increased comfortableness and environmental sustainability.  For both the dataset, five statistical models were trained with optimized hyperparameters using a repeated cross validation approach and testing set is used for evaluation: (a) CUBIST (b) Regularized Random Forest (c) Classification and Regression Trees (d) K Nearest Neighbour (e) Conditional Inference Tree. Multiple evaluation indices such as R2, Root Mean Squared Error, Mean Absolute Error and Coefficient of Variation were used to measure the prediction performance of the regression models. The results show that the rule-based model CUBIST was able to explain about 95 and 89% of the Variance (R2) in the testing set of Seoul Bike data and Capital Bikeshare program data respectively. The results show that CUBIST algorithm improve the R2, RMSE, MAE and CV compared to RRF, CART, KNN and CIT models in both of the datasets. This shows that CUBIST algorithm can be used as an effective tool for bike sharing demand prediction. These findings provide a new option for researchers to predict hourly rental bike sharing demand and enrich the library of algorithms of empirical modelling-based hourly rental bike demand prediction. Future work will focus on district wise rental bike demand prediction by considering seasonal changes. |