**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. **Mayur Prakash parande**   Email**:** [mayurparande25@gmail.com](mailto:mayurparande25@gmail.com)   * **Data description** * **Data cleaning** * **Data visualization** * **Data understanding** * **Linear regression modeling** * **ML learning modeling**  1. **Shubham Shrivastava**   Email: [shrivastavashubham360@gmail.com](mailto:shrivastavashubham360@gmail.com)   * **Exploratory data analysis** * **Decision tree regressor** * **Feature analysis** * **Elastic net regressor** * **Gradient boosting** * **ppt**  1. **Praful M. Gedam**   Email: [prafulgedam12@gmail.com](mailto:prafulgedam12@gmail.com)   * **Multivariate analysis** * **Model training** * **MSE** * **RMSE** * **MAE** * **Technical document** * **summery** |
| **Please paste the GitHub Repo link.** |
| GitHub Link: <https://github.com/PrafulGedam/bike-sharing-demand-prediction>  Drive link : <https://drive.google.com/drive/folders/135zJpyt3H2tZin0uBq6WR7eiIY3gRQGL?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)** |
| Problem statement:  entry 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  Approaches:  The first steps involved in the projects is Exploratory data analysis whare to tried to dig insight from the data in hand. The second steps was cleaning the data and perform modification. We checking the missing value and outlier and removed irrelevant features. We also encoded the categorical variable. The third step was try linear regression analysis algorithm, Random forest, Gradient boosting on our split the standardize the data. We did hyperparameter tuning and evaluated the performance the model using metrices. The best performance was given by the Gradient boosting and random forest model. The R2\_score and test set 0.789 and 0.784 respectively. Bas eon the analysis, we build random forest model to predict the number of bikes in rent. Also we use various type of model like linear regression, decision tree but this model we couldn’t get much score. So that finally we rent to random forest model.  Conclusion:  In the given dataset there was no strong linear relationship between ‘Rented bike count’ and dependent features. That’s why linear regression model and its other regularization variant model didn’t perform well. Out of all model apply decision tree and random forest model are most accurate. Reason for this is no specific relation between features a large data. Random forest perform best as is it is an ensemble model and result for multiple decision tree is average out to given to prediction |