**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Name: Rushabh Tikale  Email: [rushabhtikale62@gmail.com](mailto:rushabhtikale62@gmail.com)  Contribution:    1.Import library.  2.Import csv data.  3.Checking null value.  4. Exploring Dependent and Independent variable  5.Checking and removing outlier.  6.EDA  7.Relation between dependent and independent variable.  8.Correlation in Variables.  9.Prepared data for train and test.  10.Build baseline model.  11.Use evaluation metrics.  12.Compare the best fited model.  13.Choose the bset predictor model.  14.Conclusion. |
| **Please paste the GitHub Repo link.** |
| Githubs Link:- <https://github.com/Rushabhtikale92/Bike-Sharing-Demand-Prediction.git>  Drive Link:- <https://drive.google.com/open?id=115dXbQ71tBrzSi3hbmnWHT8WsOSA0cg_&authuser=rushabhtikale62%40gmail.com&usp=drive_fs> |
| **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 Bike sharing is one of the major concern part for now a days life. Its basically use in major of city that every employee or any of the customer can take a benefit of this business so making this model is very useful for prediction of how many bikes will requored to fulfill demand of customer so any of the customer cant have to wailt for the bikes and it gives a lot bad impact for the successful run up business.  The dataset contain information about weather for the day when the bikes comes for the rent and the number of bikes rented per hour and date information. We extract some confusion data and try to make simple it, using some baseline model we try to predict the correct output for the model. Then we used some evaluation metrics for the model this helps us to know the residual error between actual and predicted dataset. Now the delicate part is use the hyperparameter tuning because a single wrong selection of the hyperparamater tuning will overfit the model. The **problem statement** is 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 **Approaches** is first of all we look over the data and try to understand the depth of dataset and sort data according to the requirement , actualy we clean the data and try to put it into the best fitted model. We used model such as Ridge, Lasso, ElasticNet, Random Forest and Decision Tree for analyse the data and make predictor model. Then we used some evaluation metrics to determine how the model is perfect for prediction. The R2 score and adjusted R2 score is used to predict the residual error between  actual and predicted dataset. We deal with the large dataset and try to make more comfortable to set for training and testing. Then after we calculate the MSE, MAE, RMSE.  MAE is the easiest to understand, because it's the average error. MSE is more popular than MAE, because MSE "punishes" larger errors, which tends to be useful in the real world. RMSE is even more popular than MSE, because RMSE is interpretable in the "y" units. All of these are loss functions, because we want to minimize them.  Here we come with the final **conclusion** from the business environment that the toughest part is to deal with the large data.The whole model depends on the data cleaning so the maximum part we spent on cleaning the data.We mostly get stuck in the preparation The data consists of large features so we have to deal with each and every subsets.Any wrong choosing for hypereparameter tuning can become the overfit model so its very essential to use best value for hyperparameter. So here we predicting that the Random Forest model gives us a better model so we can go ahead with this model. We train and test the data of regression by taking the sample of data. Then we used different baseline models such as lasso, ridge, ElasticNet which gives us the view for model good or bad.From this algorithm we did not get any good prediction for model by which we cant develop the business model. Then we used Random Forest and Decision Tree model. From this we get a good model by Random Forest.From this we can go with upnext business problem, and we can used as good predictor model. |