**Capstone Project (Bike Sahaing) Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Name:** Indugopal Maity **Email-Id:** [ig.maity@gmail.com](mailto:ig.maity@gmail.com)  **Contribution:**   1. Data Wrangling    * 1. Bike Count Dataset      2. Date, Hour, Temperature, Humidity, Wind Speed, Visibility, Dew Point Temperature, Solar Radiation, Rainfall, Snowfall, Season, Holiday and Functioning Day 2. Found Missing Values with Percentage 3. Calculated relationship between “Rented Bike” and “Different Month of Year”. 4. Calculated relationship between “Rented Bike” and “Weekdays and Weekend”. 5. Calculated relationship between “Rented Bike” and “Different Hours”. 6. Calculated relationship between “Rented Bike” and “Functioning Day”. 7. Calculated relationship between “Rented Bike” and “Season”. 8. Calculated relationship between “Rented Bike” and “Holiday 9. Analyzed different Numerical values (Temperature, Dew Point Temperature, Solar Radiation, Snowfall, Rainfall, Wind Speed). 10. Plotted Several Linear Regression Plots. 11. Plotted correlation between variables. 12. Plotted Heatmap. 13. Model Training 14. Wrote Colab Conclusion 15. Power Point Presentation   **Name:** Md. Tousif Ali **Email-Id:** [tousifali2018@gmail.com](mailto:tousifali2018@gmail.com)  **Contribution:**   1. Data Wrangling    * 1. Booking Dataset      2. Date, Hour, Temperature, Humidity, Wind Speed, Visibility, Dew Point Temperature, Solar Radiation, Rainfall, Snowfall, Season, Holiday and Functioning Day 2. Plotted Several Linear Regression Plots. 3. Plotted Several Lasso Linear Regression Plots. 4. Plotted Several Ridge Linear Regression Plots. 5. Plotted Several Elastic Net Linear Regression Plots. 6. Plotted Several Decision Tree. 7. Plotted Several Random Forest 8. Plotted Several Gradient Boosting. 9. Plotted Several Hyper Parameter Tuning. 10. Done Gradient Boosting Regressor with GridSearchCV 11. Importing Gradient Boosting Regressor |
| **Please paste the GitHub Repo link.**  **https://github.com/indugopal1991/Bike-Sharing-Demand-Prediction/blob/main/Indugopal\_Maity\_Bike\_Sharing\_Demand\_Prediction\_Capstone\_Project.ipynb** |
| Github Link:- https://github.com/indugopal1991 |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  As the first step, performed **“DATA Wrangling”** over the raw data, **“SeoulBikeData.csv**”. We can see that there is a combination of “**8760”** rows and “**14”** columns. The chart is consisting of several columns like **“Date”**, **“Rented Bike Count”**, **“Hour”**, **“Temperature”**, **“Humidity”**, **“Wind Speed”**, **“Visibility”**, **“Dew Point Temperature”**, **“Solar Radiation”**, **“Rainfall”**, **“Snowfall”, “Season”, “Holiday”** and **“Functioning Day”**.  Farther we have divided the complete project into seven different parts as per the **“Exploratory Data Analysis (EDA)”**, **“Feature v/s Bike Count”**, **“Regression Plot”**, **“Correlation Analysis”**, **“Modelling Overview”** and **“Feature Importance”.**  In the first part, we tried to find out **EDA** relation between **“Rented Bike Count”** to **“Different Month”, “Weekdays/ Weekend”, “Different Hours”, “Functioning Day”, “Seasons”** and **“Holiday”**.  In the second part, we tried to find out relation between **“Rented Bike Count”** to several **Numerical (Feature) Values** such as **“Temperature”, “Dew Point Temperature”, “Solar Radiation”, “Snowfall”, “Rainfall”** and **“Wind Speed”**.  In the third part, we tried to draw several linear Regression Plot in between **“Rented Bike Count”** to **“Temperature”, “Dew Point Temperature”, “Humidity”, “Visibility”, “Solar Radiation”, “Snowfall”, “Rainfall”** and **“Wind Speed”** to check whether they are positively or negatively correlated.  In the fourth part, we have checked **“Correlation Analysis – Heatmap”**, to check which are **positively correlated** and which are **negatively correlated**.  In the fifth part, we have used different **“Regression Model”** such that **“LASSO Regression”, “LIDGE Regression”, “ELASTIC NET Regression”, “Decision Tree”** and **“Random Forest”**. After using different types of regression model, we can see that **“Random Forest”** is the best with **R2** value **“0.98”**.  In the sixth part, we have modelled using **“Gradient Boosting GridSearchCV”** a type of **“Supervised Learning”** and evaluated using **“Root Mean Square Error (RMSE)”** and we have received R2 value as **“0.95”**.  In the seventh part, we have seen which **Feature** have the most importance for **Bike rent** and that is **“Temperature”**.  So, by the analysis we have figured out certain things, those are listed below:   1. No **“Overfitting”** is seen. 2. **Random forest Regressor** and **Gradient Boosting GridSearchCV** gives the highest **R2** score of **98%** and **95%** respectively for **Train Set** and **92%** for **Test set**. 3. **Feature Importance** value for **Random Forest** and **Gradient Boost** are different. 4. **We can deploy the Model.** |
|  |