Capstone Project Submission

Team Member's Name, Email, and Contribution:
Team Member's Role:- ▪ **Sanjaya Kumar Khadanga** 　eMail- skhadanga38@gmail.com 　o Data Understanding 　o Feature Analysis 　o Feature Engineering 　o Linear Regression modeling 　o Random forest 　o Gradient Boosting 　o Hyperparameter tuning ▪ **Bibhuti Bhusan Sahu** 　eMail- sahubibhuti45@gmail.com 　o Data Understanding 　o Feature Analysis 　o Data Visualization 　o Multivariate Analysis 　o Decision Tree 　o Elastic Net Regressor 　o Research Analytics 　　▪ Technical documentation ● **Balaram panigrahy** 　eMail- balarampanigrahy42@gmail.com 　o Data Understanding 　o Data Visualization 　o Multivariate Analysis 　o Lasso Regression 　o RidgeRegression 　o Research Analytics 　　▪ Technical documentation
Please paste the GitHub Repo link.
Github Link: https://github.com/Bibhuti-MLAI/Bike-Sharing-Demand-Prediction
Please write a summary of your Capstone project and its components. Describe

the problem statement, your approaches, and your conclusions. (200-400 words)

In the present scenario the demand for sharing bikes is increasing as to get the mobility and ease of travelling. We have the dataset to predict the demand of seoul bike sharing demand prediction. The data we have is having rented bike count as predictor or dependent variable and the independent features are like Hour, Wind Speed, Temperature, Humidity, days, etc.

First process we started with business understanding and the objectives and constraints. The objective is to predict the counts of bikes. We started with cleaning the data and then preprocessing the data as removal of outliers, renaming the columns, encoding the categorical features and type casting some features.

Then the important phase of the project starts, we started with Exploratory Data Analysis (EDA). In EDA we have analysed each feature by univariate analysis to know the trends and to analyse the outlier presence in data by box plot and also point plots to know the trend of the data. Then we moved to bivariate analysis where we analysed the correlation strength and and direction of the data points. Then in the final stage of EDA we have gone to multivariate analysis, in multivariate analysis we get the highest correlated data and analysed the demand of rented bike counts to other independent variables. The data is now ready for modelling.

I modelling step we started with multiple linear regression, Lasso regression, Ridge regression, Random Forest Regressor and decision Tree and to get the best fit line we also tuned the hyperparameter by GridSearchCV and we achieved the best fit line in Decision tree and gradient boosting.

For evaluating the models, we have calculated Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and R squared value and also adjusted R squared value. The model decision tree and gradient boosting shows best fit to solve the problem.