Capstone Project Submission

Team Member's Name, Email and Contribution:

1) **Satyam Jyoti Sankar**
E-mail: satyamjyoti99@gmail.com

- Data Preprocessing
- Data analysis
- Data Visualization
- Approach towards plan.
- Feature Engineering
- Frame work of project
- Model Building
- PPT presentation
- Technical documentation.
- Project summery template.

Problem definition:

Rossmann operates over 3,000 drug stores in 7 European countries. Currently, Rossmann store managers are tasked with predicting their daily sales for up to six weeks in advance. Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. With thousands of individual managers predicting sales based on their unique circumstances, the accuracy of results can be quite varied.I have provided with historical sales data for 1,115 Rossmann stores.

My task is to forecast the "Sales" column for the test set. So that we predicted sales in current future.

EDA on given all Data set :

After EDA we get that

1. After EDA part we merge our all 2 data set and create a final data set.

2. In my final dataset we have 26 feature without any null values.

3. Dependent variable should be considered as sales.

4. The most important factor affected sales is avg customer.

5. The heatmap show good correlation between all the Feature.

Models used :

- Linear Regression (Baseline Model)
- Lasso (Hyperparameter)
- Decision tree by using decision tree regressor (Hyperparameter)
- Random forest regressor
- Gradient Boosting Regression
- Xg boost regressor (Hyperparameter)

Conclusion :

- Random forest regressor gives us high accuracy of 93% for our test data set in case of train data it show accuracy of 99% .
- Xg boost gives 90% in both train and test case both r2 score and adj r2.
- In case of random forest a little bit overfit occur but in xg boost it work properly.
- Our base model liner regression created a base line of accuracy of 74%.
- The lasso model although we use hyper parameter it gives us similar performance like linear regression.
- In case of Decision tree by using decision tree regressor (Hyperparameter) it gives us 78% accuracy in train and 79 accuracy for test case.

Please paste the GitHub Repo link.

Github Link:- https://github.com/satyam-jyoti-sankar/Rossmann-Sales-Prediction-

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)