Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Name : Nitesh Gajakosh **E-mail: niteshgajakosh1998@gmail.com** Contributor role : 1) Data Wrangling - i) Analyze Data set ii) Match data set according to requirement 2) Data cleaning- i) Delete unnecessary data. ii) Null value treatment 3) Data Visualization- i) Close price trend ii) Distribution of closing price iii) Distplot iv) Scatterplot v) Heatmap 4) VIF i) Important features. 5) Regression analysis- i) Linear Regression ii) Lasso iii) Ridge iv) Elasticnet 6) Cross validation i) Optimize hyper tuning parameter 7) Conclusion
Please paste the GitHub Repo link.
Github Link:- https://github.com/niteshgajakosh/yes-Bank-Stock-Closing_Price_Prediction G-Drive link -
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-4OO words)

Yes Bank Limited is an Indian private sector bank headquartered in Mumbai, India and was founded by Rana Kapoor and Ashok Kapur in 2004. It offers wide range of differentiated products for corporate and retail customers through retail banking and asset management services. On 5 March 2020, in an attempt to avoid the collapse of the bank, which had an excessive amount of bad loans, the Reserve Bank of India (RBI) took control of it.When fraud case news come out stock price of yes bank fall from 391 to 190 and now running price is around 13 Rs.

My First Step was to import the dataset through Pandas then data wranplinp and know features in dataset. I did not pet into the situation to remove NA values because there are 0 null values in the dataset.
Next step is EDA in that I briefly study related to close price of stock and find out dependent ,independent variable. After knowing dependent variable I plot distribution plot to check skewness of variable and I find data is rightly skewed so need to use lop transformation.
Now after transformation correlation been checked with help of heatmap, there is very high correlation among all variables so to check multicollinearity is VIF(variation inflation factor).we drop some features to prevent wrong prediction.

Prepare dependent and independent variable for train test split method. I apply Linear regression, lasso regression, Ridge regression,EIasticNet regression. As per model performance linear regression and ridge perform well. After cross validation and Hyperparameter tuning performance increase significantly.

Conclusion:

- Target variable(dependent variable) strongly dependent on independent variables.
- We get maximum accuracy of 82%.
- Linear regression and Ridge regression get almost same R squared value.
- Whereas Lasso model shows lowest R squared value and high MSE,RMSE,MAE,MAPE.