Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
1. Sanjeev Hegde: • Mail: sanjeevhegde454@gmail.com   • Contributions: Data cleaning, essential modifications to the data, Exploratory Data Analysis, Feature engineering, Fitting Linear Regression model, Regularized linear regression models, XGBoost Model for predicting Yes Bank closing price. PowerPoint Presentation, Project Summary and Project Technical documentation.
Please paste the GitHub Repo link.
1. Github Link for Sanjeev Repository: -   https://github.com/Sanjeevhegde009/Yes-Bank-closing-price-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)
 	
     Yes Bank is one of the prominent banks in the Indian Banking Sector. Since its listing on the Indian Markets after 2005 at Rs.13, Yes Bank shares witnessed phenomenal growth with an all time high of Rs.404 during 2018. However, during the same year, due to fraud cases registered against its Founder and CEO, shares tanked significantly wiping out all the gains by the stock. In this regard, with Yes bank monthly OHLC data(Open, High, Low, Close) given, supervised machine learning techniques have been deployed to successfully predict the monthly closing price of Yes Bank.
     The given data contained a total of 185 entries for OHLC data from Yes Bank listening month i.e., June 2005 to November 2020. Given data set contains no null values or duplicates. As part of EDA, appropriate data types have been assigned to necessary features. Further, it is found that the given data set is right skewed. All independent features(OHL) exhibit linear relation with the dependent variable(closing price). On plotting correlation heatmap, it was found that all features exhibited extremely high correlation.
     Feature engineering has been performed on the data set in order to reduce multicollinearity in the data set. Average of OHLC was calculated for each entry and lags were introduced in the data set using 14 previous values of average OHLC price thereby successfully reducing multicollinearity.
     Standard scaler was used to scale the data set for fitting regression. Sample size of 0.7 was considered for the same. Lags were considered as the independent features and closing price was predicted based on the same using regression models

and XGBoost. Two functions were created to induce modularity in the process. One of the functions plots actual vs predicted closing prices while the other gave evaluation metric scores of the machine learning models.

Linear Regression, Lasso and Ridge regression, elastic regression and XGBoost models were used to predict the closing price of the Yes Bank stock. It was found that the XGBoost model predicted the closing prices with the highest evaluation metric score for both training and testing data sets. R^2 score for testing data set was found to be 0.9087.