**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email, and Contribution:** |
| 1. Name: - Shubham Narendra Kadu 2. Email - [shubhamkadu.ds@gmail.com](mailto:shubhamkadu.ds@gmail.com) 3. Contribution **–**  * Contributed to this project that first I have self-done. * Data processing * Data Cleaning * Data Analysis * Feature Engineering * EDA * Correlation Analysis * Multicollinearity Test * One hot encoding * Data Preprocessing * Machine Learning Model * Hyperparametric Tunning * Model selection * Conclusion * PPT |
| **Please paste the GitHub Repo link.** |
| GitHub Link:- https://github.com/shubhsk98/Cap-project---Yes-bank-stock-closing-price-prediction.git |
| **Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)** |
| **Problem Statement :**  Yes-Bank is a well-known bank in the Indian financial domain. Since 2018, it has been in the news because of the fraud case involving Rana Kapoor. Owing to this fact, it was interesting to see how that impacted the stock prices of the company and whether Time series models or any other predictive models can do justice to such situations. This dataset has monthly stock prices of the bank since its inception and includes closing, starting, highest, and lowest stock prices of every month. The main objective is to predict the stock’s closing price of the month.  **Objective**:  The dataset contains the monthly stock price details for Yes Bank [‘Date’, ‘Open’, ‘High’, ‘Low’, ‘Close’]. The main objective of this project is to predict the stock’s closing price of the month  **Summary**:  Before we start our ML Project process which is first Data Reading, Data Preparing, Data Cleansing, Data pre-processing, Data Analysis, Feature engineering, and Machine Learning Model.  Our First Step was to import the dataset and we check for duplicate and null values if there but luckily we see that there is no null values or duplicate values. But the Date feature has values in the object data type. So, we converted it into a proper date format.  Then performed, EDA in which trend of stock closing price, distribution of dependent variable have been examined. Plotted histogram of all variables with mean and median to check measures of central tendency is close to each other or far.  After that, the correlation has been checked among each other through a heatmap. And after selecting independent and dependent variables pass it to the next step which is to train models. then applied the machine learning model ( Linear Regression, Ridge regression, Lasso regression, and Elastic Net Regression) and applied cross-validation.  **Conclusion:**  In visualization we checked that from 2018 onwards there is a sudden fall in the stock closing price.  Target Variable is strongly dependent on Independent Variables.  we have performed VIF to reduce multicollinearity  Insights of all the models, A simple linear regression model was built and it was evaluated using accuracy, MSE, RMSE, r2\_score, and Adj\_R2, mean absolute percentage error.  Linear Regression, Lasso, and Ridge are performing better than Elastic Net models with training accuracy of 94.58%, 94.58%, and 94.58% respectively.  Apart from Linear Regression, Lasso, and Ridge, Elastic Net is also performing better but has less training accuracy.  Ridge and Elastic Net have performed far much better after Cross-validation which is R2 is about 95.25% and 95.09% respectively.  R2 and Adjusted R2 are around the range of 95% and 91% in each model. |