

# Capstone Project Submission

## Instructions:

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

### **Team Member's Name, Email and Contribution:**

1. Zeeshan Ahmad ([zeeshanahmad00789@gmail.com](mailto:zeeshanahmad00789@gmail.com))

#### **Contribution:**

- Outlining project plan.
- Data wrangling and maintaining data integrity.
- EDA.
- Feature Engineering.
- Training and Testing Model.
- Hyperparameter Tuning.

### **Please paste the GitHub Repo link.**

Github Link:- <https://github.com/Zeeshan00789/Yes-bank-closing-price-prediction>

Drive Link:- <https://drive.google.com/drive/folders/1oMZ-dTTAzGikbuaITtoc7c7jAmiWEXRB>

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

The term stock price refers to the current price that a share of stock is trading for on the market. The price of a stock will go up and down in relation to a number of different factors, including changes within the economy as a whole, changes within industries, political events, war, and environmental changes. In this project, we are predicting yes bank stock closing price with the help of the regression models. The given dataset consists of features such as the open, close, low, and high price of the stock of yes bank on a monthly basis.

We followed step-by-step processes for the project like data collection, data cleaning, EDA, Visualization, Model Training and Testing, Hyperparameter Tuning, and Evaluation. In EDA we divided the analysis into several part to get a better idea of data. We checked the data distribution with the help of distplot, and normalize the data with the help of log transformation. The dependent variable was the closing price, we plotted scatter plot of 'open', 'low', and 'high' against 'close'. we found that the relation between them was linear.

We split the dataset into two parts a training dataset (80%) and a testing dataset (20%). The model gets trained from the training data, after training we use regression models such as linear regression, lasso regression, ridge regression, cross-validation and elastic net one by one and evaluated the results.

The test results of all the regression models are evaluated and compared. We checked performance metrics such as  $R^2$ , adj  $R^2$ , MSE, and RMSE etc. Unfortunately, results were not up to the mark with linear regression, ridge and lasso regression.

Further we tried other models such as random forest and Xgboost. With the help of this model we got better  $R^2$  scores and metrics.