

# CapstoneProject Submission

## Regression Model

**DONE BY:**

**Praveen. S**

(Cohort Everest)

Email id: [spr87654@gmail.com](mailto:spr87654@gmail.com)

Github link:

<https://github.com/prav87654/Regression-YesBank-Stock-Prediction>

Google Drive link:

[https://drive.google.com/drive/folders/14QfDJadcQrFW\\_H2HJLimxglrjLckG1](https://drive.google.com/drive/folders/14QfDJadcQrFW_H2HJLimxglrjLckG1)

# SUMMARY

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 Machine Learning 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.

For the first step, we performed data wrangling over the raw data and concluded what are the features or columns should be used or dropped. Since the dataset doesn't contain any null values, we proceeded with the dropping of features. I have dropped the "Date" column as it was found irrespective of the target variable. I have dropped it.

I had done the EDA and Data Visualization with the remaining features and found the independent features are so correlated with the target variable.

For selecting the best models, I had imported pychart to compare different regression models and found linear, ridge and lasso regression models had higher accuracy by a micro margin and proceeded with the test train split and started to fit the models.

Tuning the hyperparameters of respective algorithms is necessary for getting better accuracy and to avoid overfitting and underfitting which was achieved by Grid search cv.

. In all of these models our accuracy revolves in the range of 80 to 83%. And there is no such improvement in accuracy score even after hyper parameter tuning. So the accuracy of our best model is 81% which can be said to be good for this very small dataset because of underfitting.