

MINI PROJECT
On
(Analysing and Predicting HDFC Bank Stock Prices)

In
Supervised Machine Learning
MASTER OF TECHNOLOGY
IN
Artificial Intelligence and Machine Learning

SUBMITTED BY

Akshay Marathe
(PRN - 23070149032)



Under the Supervision of

Dr. Kailash Shaw,
Associate Professor

SYMBIOSIS INSTITUTE OF TECHNOLOGY, PUNE – 412115
(A CONSTITUENT OF SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY))

2023-24

Table of Contents

Chapter 1 Problem Statement	4
1.1 Motivation	
1.2 Objective	
Chapter 2 Literature Review.....	5
Chapter 3 Introduction.....	7
3.1 Dataset Description	
3.2 Software Requirement	
Chapter 4 System Design.....	8
4.1 Flowchart	
Chapter 5 Methodology.....	9
5.1 Model Architecture	
Chapter 6 Result Analysis.....	11
Chapter 7 Conclusion.....	12
References.....	13

List of Figures

Figure 1 Google Colab	7
Figure 2 Dataset Header Rows	7
Figure 3 Flow Chart	8
Figure 4 Heat map	10
Figure 5 Scatter plot (linear regression)	12

List of Tables

Table 1 Literature Review	6
Table 2 Result Analysis	12
Table 3 Comparison Table	12-13

Chapter 1: Problem Statement

The aim of this project is to develop a machine learning model that tries to accurately forecast the closing price of HDFC Bank Ltd. stock by using historical stock data from 2019 to 2023 and relevant financial indicators.

- **Motivation:**

The financial market, especially the stock market is known for its high volatility and unpredictability. Investors, traders, and financial institutions require robust analytic and predictive tools to make well-informed decisions. Predicting stock market indicators can help in portfolio optimization, risk management, and can contribute to market efficiency.

- **Objective:**

The project's objective is to develop a machine learning model for forecasting the closing price of the stock of HDFC Bank Ltd... We proceed with gathering and preprocessing historical data followed by feature engineering and then selecting a more accurate and robust predictive model.

We need to first evaluate the model's performance, assess its accuracy, and, if successful, consider practical implementation. To incorporate we have added a wide range of technical indicators, such as moving averages, Relative Strength Index (RSI), and Bollinger Bands, as features in the predictive models to capture market trends and patterns.

The primary goal has been to provide investors and financial professionals with a reliable tool for making more informed investment decisions for HDFC Bank Ltd. Stock.

Chapter 2: Literature Review

TITLE	YEAR	PROBLEM STATEMENT	DATASET	METHODOLOGY	PERFORMANCE
A Cluster based Non-Linear Regression Framework for Periodic Multi-Stock Trend Prediction on Real Time Stock Market Data.	2020	To make a hybrid clustered model to find the relational multi-stocks. Here a cluster-based nonlinear regression multi-stock framework in order to predict the time-based trend prediction.	NSE	1. SVM 2. NN 3. RF 4. Naïve-Bayes 5. MultiStockFS 6. CNN	SVM = 89% NN = 91% RF = 89% NB = 92% MSF = 98% CNN = 90%
Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets	2022	To predict the future stock movement of shares using the historical prices aided with availability of sentiment data.	Yahoo Finance	Random Forest, LSTM	95% confidence level could be considered an approximate fit to explain the working of the model.
Stock Market Prediction Using Ensemble Learning and Sentimental Analysis	2021	To implement ensemble learning and show its use compared to single model.	Yahoo Finance	Ensemble Model: ARIMA, SVR, LSTM and XGBoost	XGBoost useful when less data is available and to avoid overfitting, for long patterns in large dataset LSTM is useful.
Technical Analysis Based Prediction	2022	To provide intraday trading strategies	NSE	Decision Tree, SVR, Naïve-Bayes, CNN, and GAN.	Decision Tree = 69.72% SVR = 67.13%

of Stock Market Trading Strategies Using Deep Learning and Machine Learning Algorithms		using machine learning and deep learning based technical analysis.			Naïve-Bayes = 62.15% CNN = 82.66 GAN = 79.08
Review of Stock Prediction using Machine Learning Technique	2021	Develop efficient and accurate stock market prediction models that guide investors, newcomers, and stakeholders in maximizing profitability while minimizing risk in the dynamic, non-stationary, and noisy stock market environment.	Yahoo Fianace	1.Naive Bayesian 2. Support Vector Machine, 3.K-Nearest Neighbour, 4. Artificial Neural Network, 5. Random Forest, 6. Neural Network (NN)	NB-90.38% SVM-76.65% KNN-89.90% ANN-not mentioned RF- not mentioned NN-not mentioned

Table 1

Chapter 3: Introduction

1. Software Used: -

Google Colab is popular for machine learning due to its free GPU/TPU access, cloud-based nature, no setup requirements, and integration with Google Drive. It simplifies the development process, promotes collaboration, and is widely used in education and research.



Figure 1

2. Data Set description

Date: This feature represents the specific date associated with the stock market data, indicating when the other metrics were recorded.

Open: The "Open" price refers to the stock's opening price at the beginning of the trading day, which is the initial price at which a stock is traded.

High: The "High" price represents the highest price at which the stock was traded during the trading day, indicating the peak price reached during the trading session.

Low: The "Low" price signifies the lowest price at which the stock traded during the trading day, reflecting the nadir of the stock's price during the session.

Close: The "Close" price denotes the stock's closing price at the end of the trading day, representing the final traded price for the stock.

Adj Close: The "Adjusted Close" price accounts for any corporate actions such as stock splits, dividends, or other adjustments, providing a more accurate representation of the stock's value over time.

Volume: "Volume" refers to the total number of shares of the stock traded during the trading day, indicating the level of market activity and liquidity for that particular stock.

	Open	High	Low	Close	Adj Close	Volume
Date						
2019-04-01	1162.6250000000000	1166.0000000000000	1152.3000488281250	1155.9000244140625	1115.3616943359375	4823538
2019-04-02	1160.0000000000000	1160.0000000000000	1142.7249755859375	1147.4250488281250	1107.1839599609375	5455248
2019-04-03	1146.0500488281250	1158.0000000000000	1144.1250000000000	1146.3750000000000	1106.1707763671875	3829726
2019-04-04	1146.3750000000000	1156.5000000000000	1140.0000000000000	1141.9749755859375	1101.9250488281250	4605776
2019-04-05	1143.9000244140625	1156.8750000000000	1141.0000000000000	1152.8000488281250	1112.3706054687500	4175590

Figure 2

Chapter 4: System Design:

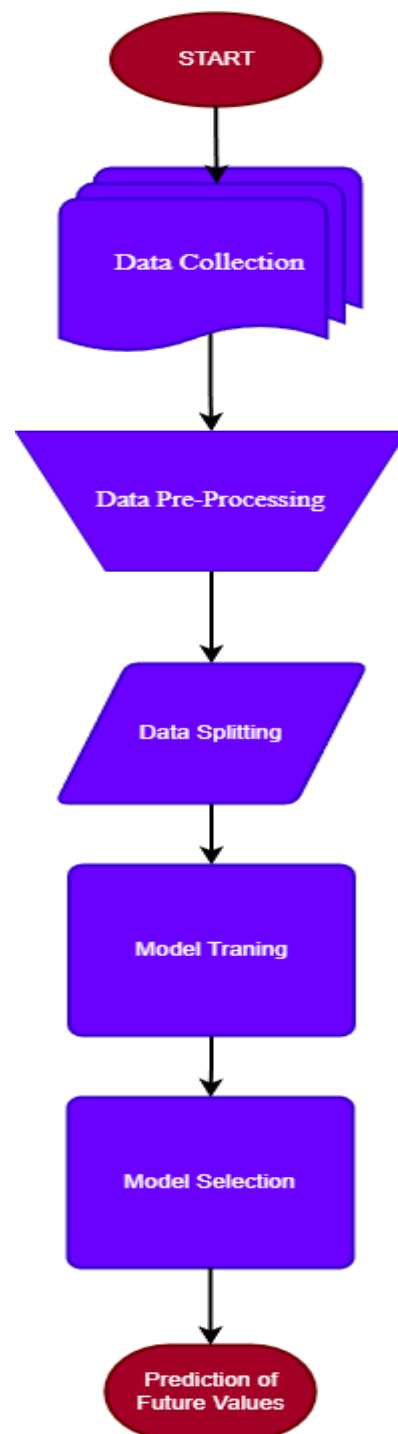


Figure 3.

Chapter 5: Methodology

Data Collection: Gather historical data for HDFC Bank Ltd. stock, including the features - "Open," "High," "Low," "Volume," and the target variable - "Close" or "Adjusted Close" prices. Ensure you have a time series dataset spanning the years 2019 to 2023.

Data Preprocessing: Clean the dataset by checking for missing values, outliers, or any data inconsistencies. You may also need to convert the "Date" feature into a date-time format for time series analysis.

Feature Engineering: Create additional features or technical indicators that may help capture stock market trends and patterns. Common technical indicators include Moving Averages, Relative Strength Index (RSI), Bollinger Bands, and more.

Data Split: Divide the dataset into training and testing sets. The training set is used to train the regression model, while the testing set is used to evaluate the model's performance.

Model Selection: Choose appropriate regression models for prediction. You can consider using Linear regression, random forest, support vector regression, or other regression algorithms. Train multiple models to compare their performance.

Model Training: Train the selected models on the training dataset using the features ("Open," "High," "Low," "Volume") to predict the target variable ("Close" or "Adjusted Close").

Model Evaluation: Evaluate the regression models using the testing dataset. Calculate key metrics such as:

R-squared (R²) Score: Measures the proportion of the variance in the target variable that is predictable from the features.

Mean Squared Error (MSE): Measures the average of the squared differences between predicted and actual values.

Mean Absolute Error (MAE): Measures the average absolute differences between predicted and actual values.

Root Mean Squared Error (RMSE): Measures the square root of the MSE, providing the error in the same units as the target variable.

Hyper-parameter Tuning: Fine-tune the hyper-parameters of the regression models to optimize their performance.

Model Selection: Choose the regression model that performs best based on the evaluation metrics (e.g., highest R2 score, lowest MSE, MAE, RMSE).

Predictions: Utilise the selected model to make predictions on future stock prices based on the provided features.

Reporting: Present the results, including R2 Score, MSE, MAE, and RMSE, in your project report. These metrics will provide insights into the accuracy and reliability of your regression model for predicting HDFC Bank Ltd. stock price

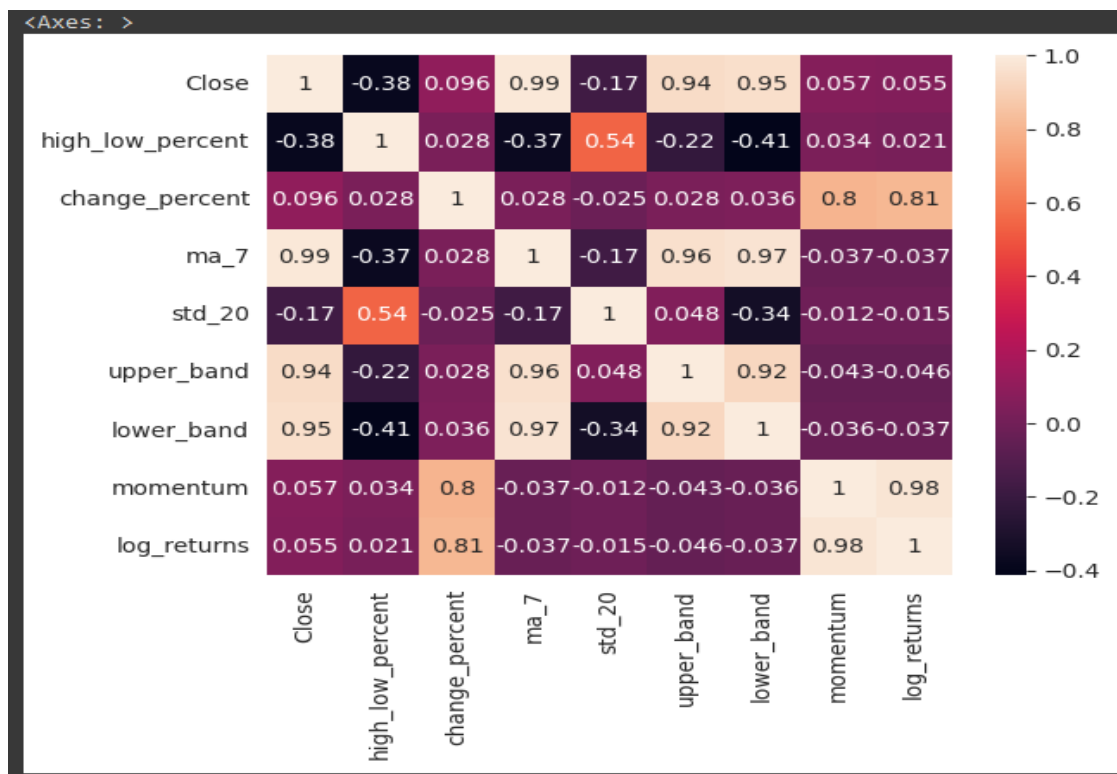


Figure 4

Chapter 6: Result Analysis

ALGORITHMS	ACCURACY
Linear regression	98.96%
Random Forest	98.58%
Support Vector Regression	97.93%

Table 2

The models are trained and tested and the evaluation metrics used are R2 score

- R2 score for linear regression: 0.9896238870151237
- R2 score for random forest: 0.9858356489722839
- R2 score for SVR: 0.9793275895270694

Based on the model accuracy of 0.9896238870151237, it can be concluded that the Linear Regression model is the most suitable model for this data.

R2_Score: 0.9896238870151237

MSE: 0.010739208383851956

MAE: 0.07991844506631797

RMSE: 0.10363015190499315

✓
0s



```
sns.scatterplot(x=y_test, y=y_test_predict)
```

<Axes: xlabel='Close'>

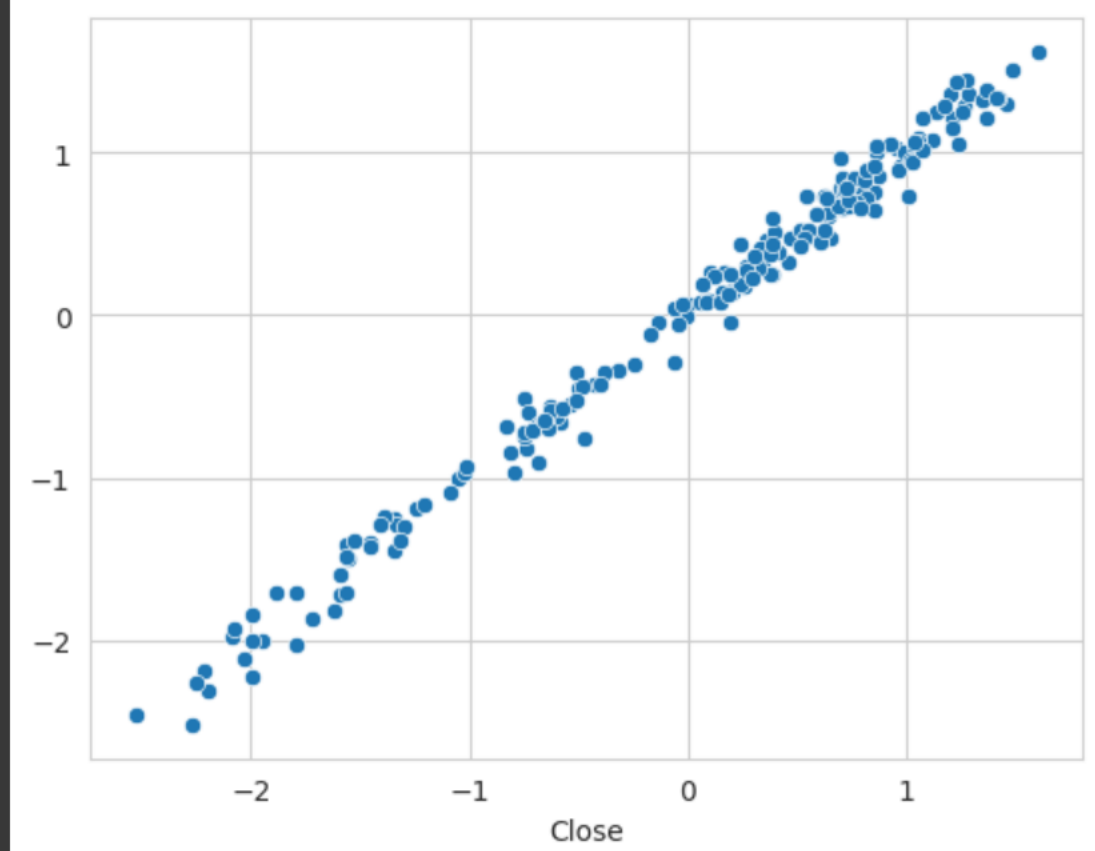


Figure 5

Reference No.	Models Used	Results
1	PCA and A multivariate linear regression model.	Accuracy: - 82% RMSE rate: - 4.73 MAE rate: - 2.13
2	Random Forest and LSTM	~ 95% Accuracy
3	1. Decision Tree 2. SVR 3. Naïve-Bayes 4. CNN 5. GAN	Decision Tree = 69.72% SVR = 67.72% Naïve-Bayes = 62.15% CNN = 82.66 GAN = 79.08
4	1. Naïve-Bayes 2. SVM 3. K-NN	NB-90.38% SVM-76.65% KNN-89.90%
Proposed Method	1. Linear Regression 2. SVR 3. Random Forest Regression	Linear Regression – 98.96% SVR – 98.58% Random Forest Regression – 97.93%

Table 3

Chapter 7: CONCLUSION:

Linear Regression model will be used to predict the closing price of HDFC Bank Ltd. stock. After that, we will evaluate the model by checking the R2 score, mean absolute error, and mean squared error. We will then use this model to make predictions on the future closing prices of Tata Motors stock. Then plot the Actual and Prediction Data using the Scatter Plot and Inference Chart is plotted to show all how the predicted price support the momentum for long term.

References:

1. Mallick, Aswani Kumar, Alok Kumar Mishra, and Iti Vyas. "Forecasting stock prices of five major commercial banks in India and stress testing: A multivariate approach." *Journal of Public Affairs* 21, no. 3 (2021): e2239.
2. Darapaneni, Narayana, Anwesh Reddy Paduri, Himank Sharma, Milind Manjrekar, Nutan Hindlekar, Pranali Bhagat, Usha Aiyer, and Yogesh Agarwal. "Stock price prediction using sentiment analysis and deep learning for Indian markets." *arXiv preprint arXiv:2204.05783* (2022).
3. Singh, Tinku, Siddhant Bhisikar, Satakshi, and Manish Kumar. "Stock Market Prediction Using Ensemble Learning and Sentimental Analysis." In *Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021*, pp. 429-441. Singapore: Springer Nature Singapore, 2023.
4. Patel, Ramkrishna, Vikas Choudhary, Deepika Saxena, and Ashutosh Kumar Singh. "Review of stock prediction using machine learning techniques." In *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 840-846. IEEE, 2021.
5. Sakhare, Nitin Nandkumar, and S. Sagar Imambi. "Technical Analysis Based Prediction of Stock Market Trading Strategies Using Deep Learning and Machine Learning Algorithms." *International Journal of Intelligent Systems and Applications in Engineering* 10, no. 3 (2022): 411-421.
6. Ling, W., and Mei, Fu. D-"Estimation of missing values using a weighted k-Nearest Neighbors algorithm". International Conference of Environmental Science and Information Application Technology (ESIAT), 2009.
7. Chatrath, A., Miao, H., Ramchander, S., Villupuram, S., 2014. Currency jumps, cojumps and the role of macro news. *Journal of Money and Finance*, 40, 42-62, 2014.
8. Aloysius Edward, JyothiManoj – "Forecast Model Using Arima For Stock Prices Of Automobile Sector", Volume 6, Issue 4 (April, 2016) (ISSN 2231- 5985), International Journal of Research in Finance and Marketing (IMPACT FACTOR– 5.861)