Modelling direction detection in selected stocks in

Indian BFSI sector

Anand Mohan  
*REVA Academy for Corporate Excellence(RACE)*  
*REVA University*Bangalore, India  
anandmohan.embedded@gmail.com

J.B Simha  
*REVA Academy for Corporate Excellence (RACE)*  
*REVA University*Bangalore, India  
jbsimha@gmail.com

ShinuAbhi  
*REVA Academy for Corporate Excellence (RACE)*  
*REVA University*Bangalore, India  
shinuabhi@reva.edu.in

*Abstract*— several research initiatives have been taken to predict stock market returns using historical data. During this capstone project, twenty-two years' price of the stock's daily close price is being utilized for direction detection. The objective of the project is to get the right stock and understand the data pattern using Exploratory Data Analysis and perform data preparation and then build the right models by using multiple Modelling techniques to predict whether the price will move up or move down. Closing prices are being utilized as 6 different feature variables for building the classification Model. The difference between the 7th and 8th-day Closing price is determined. The 0.7% difference, 1% difference, and 1.5% difference are different classes of direction for which the rule is being set to determine either positive change, negative change, or no change. A similar process is again repeated for a range of consecutive days to be utilized as the feature variable increased to 10 days and 14 days. Then momentum, trend, volatility, and volume indicators are utilized as feature variables and different classification models are built to determine upward direction detection. Random forest modelling has given the highest efficiency in direction detection. Logistic regression modelling done for percentage change in close price as 0.5% has given the highest efficiency for volume and momentum indicators whereas XG Boost Classifier provided the best prediction performance for trend and volatility indicators. The invaluable take away from the capstone is that various classification modelling techniques had been remarkably useful in direction detection for the stock under consideration.

Keywords— Direction detection, Stock Market, Technical Indicators, Classification Models, HDFC, KOTAK, SBI

# Introduction

Live validations are still becoming a grim prospect, because of several things like value variations, quiet news, and existing noise [15].

Several Machine-Learning associated techniques are developed which have created the potential to predict the market to an extent [16].

For the transaction of shares via a broker, there is mostly a fee paid to the broker for each buy and sale which will almost eat up the gains [6].

The requirement is to overcome the ambiguities of Fundamental and technical evaluation, and advanced development in the modelling strategies has pushed several researchers to check new strategies for stock value forecasting [13].

In the next section, some of the available literature will be scanned which would throw light on various related aspects of Machine-Learning methods and other methodologies, and also study and research other related issues which would help assist better in direction detection in Stock Market.

# LITERATURE REVIEW

## Fundamental and Technical Analysis

Rajkar et al in his paper comprehensively talk about the numerous parameters impacting value movements in varied sizes and layers in the stock market [12].

Therefore, Different analysis namely technical and fundamental analysis is being done to invest in stock markets.

Elbialy in his paper worked on Fundamental analysis and suggest that it helps to identify and implement short positions by selling the shares of companies showing downtrends and then covering these positions by buying back the shares of these companies when they start showing upward trends [4].

Fundamental analysis helps to identify stock quality and therefore, stock technical analysis done later performs better on the strong fundamental stock.

Thanekar & Shaikh in their paper concludes their study that Technical Analysis can demarcate and recognize commerce openings in the stock market by examining identifiable patterns similar to volume and price action movements [17].

Kimbonguila et al in their paper used many technical indicators like MACD, OBV, Moving average, etc on the past costs to identify better stocks for trading purposes [8].

## Algorithmic trading

Taking the discussion further, Hansen in his paper mentions Algorithmic trading which is a systematic method of trading without subjective assessment through a manual trader using computer programs. Hansen further observes that fast algorithms improve traders’ ability to scan the market and seize opportunities most appropriately [5].

Mukerji et al in the paper caution that though algorithmic trading gives better results than manual trading, regulators have restrained algorithmic trading following accusations of market manipulations [10].

## Supervised and Unsupervised learnings

This paper introduces a new stock market prediction model that includes three major phases namely feature engineering, Non-correlated feature selections, and finally direction detection.

Omta et al in his paper used Machine learning and AI for the analysis of image-based cellular screens. It is suggested in the paper that exploratory data analysis should be performed as an initial step to gain a better data understanding before executing Machine learning algorithms. Machine learning can again be further categorized into supervised and unsupervised learning [11].

Alhomadi findings in his paper were that some literature has used both supervised and unsupervised machine learning techniques for securities market predictive modelling and located that both kinds of models will create predictions with satisfactory accuracy [2].

Dar researched further on unsupervised machine learning techniques by deeply studying principal component analysis and suggested that the central plan of PCA is to spot correlations and patterns in a dataset with high dimensionality and scale back it to a considerably lower dimension without losing any important info [3].

further, various supervised classification machine learning techniques have been used in this paper namely Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbours, and Extreme Gradient Boosting.

## Classification machine learning techniques

Al-Bairmani & Ismael worked further on exploring Logistic Regression and infers that Logistic Regression is used instead of linear regression in situations where the target variable is not numeric, but a nominal or an ordinal variable [1].

Jena & Dehuri studied Decision Tree and were of the view that the simple Linear modelling algorithms become more complex as the size of the datasets increases which is being handled using more advanced algorithms in Decision Tree for classification and regression problems [7].

Schonlau & Zou presented their studies on Random Forest modelling and were of the view that it is quite flexible to non-linearity in the dataset and is the most appropriate ensemble learning algorithm for medium-sized to very large-sized datasets [14].

Wang studied K-Nearest Neighbours and informs that it is the most popular statistical technique utilized in pattern identification over the last four decades [18].

Zhang et al researched on XGBoost which according to him was extensively recognized as an extremely useful ensemble learning algorithm. However, its performance needs more improvements ideally in scenarios where the dataset is imbalanced [19].

## confusion matrix for Classification Models

Various classification algorithms as discussed have to be built for the data. Subsequently, all these algorithms have to be tested. Confusion matrix for classification models is a step in that direction.

Markoulidakis et al. in his paper evaluate numerous performance metrics which include accuracy, precision, and recall [8].

# METHODOLOGY

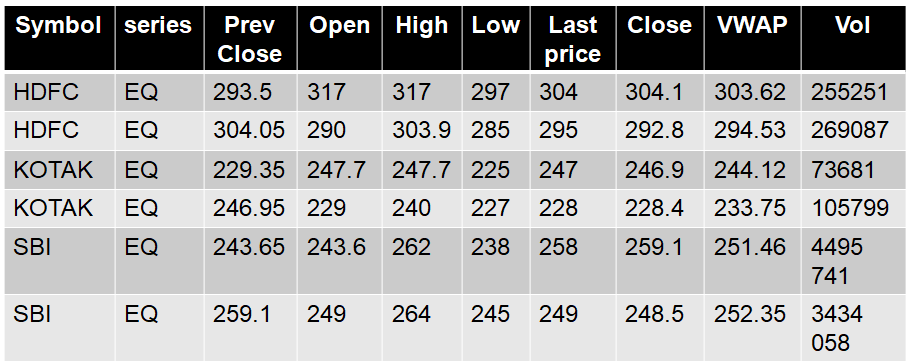
Initially Fundamental and Technical analysis of stocks under consideration is performed to demonstrate why a particular stock dataset has been used for this project The CRISP-DM framework has been used in this paper. In Data understanding the different feature variables used for the project are being studied and their Univariate analysis is performed. Based on the Data understanding phase, various steps are being taken in the Data preparation phase namely Handling Missing values, Features Addition, and Data Scaling using MinMax Scaler. Once the data has been prepared, different modelling algorithms are implemented on them namely Logistic Regression Classifier, Decision Tree Classifier, Random Forest Classifier, K Nearest Neighbour Classifier, and XG Boost Classifier. The data evaluation phase further examines the results of different Modelling techniques which were used in the Data Modelling phase. Deployment speaks about developing a front-end API for the deployment Dashboard.

## Data Collection

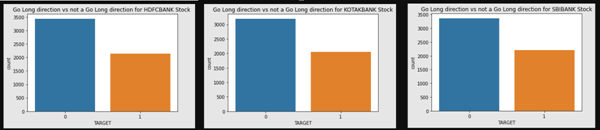
Daily Trading Data of HDFC, KOTAK, and SBI Bank from the year 2000 to 2022 are being used for this study. This study uses NSE Data.

The symbol column tells us the corporate symbol mentioned for the stock. The opening price is the first trade worth that was recorded throughout the day’s trading. The high and low is the highest and lowest value respectively at that a stock is listed during a period. The previous closing is going to be a consecutive session's opening price. The last price is the one at which the foremost recent transaction happens. The close is the last value recorded once the market is closed on the day. The volume-weighted average worth (VWAP) is a trading benchmark based on both volume and worth. Trading Volume shows the number of shares listed for the day, listed in lots of 100 quantities of shares.

Table1. Top rows of HDFC, KOTAK, and SBI stock dataset



## Data Exploration

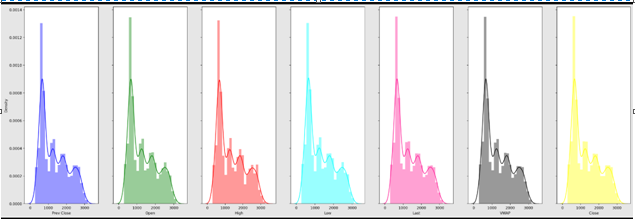


1. Class distribution For HDFC, KOTAK, and SBI stock

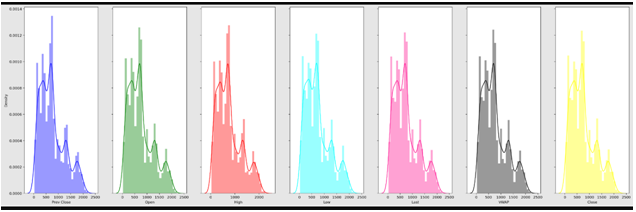
HDFC STOCK is moving 2140 times in an upward direction and is suitable for long trading whereas 3435 times, it is not moving in an upward direction. KOTAK STOCK is 2055 times suitable for long trading whereas 3199 times, it is not moving in an upward direction. SBI STOCK is 2211 times suitable for long trading whereas 3364 times, it is not moving in an upward direction.



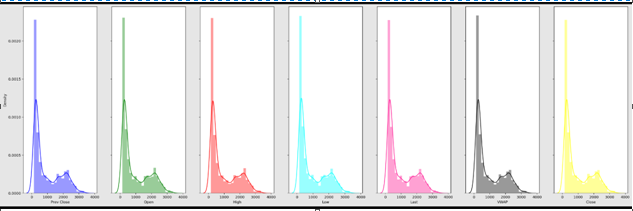
1. Close values of HDFC, KOTAK, and SBI stock from 2000 to 2022



1. Distribution Plot for the HDFC Stock

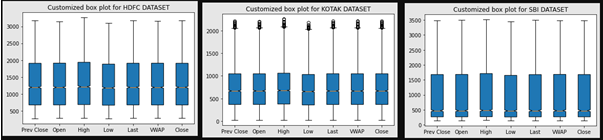


1. Distribution Plot for the KOTAK Stock



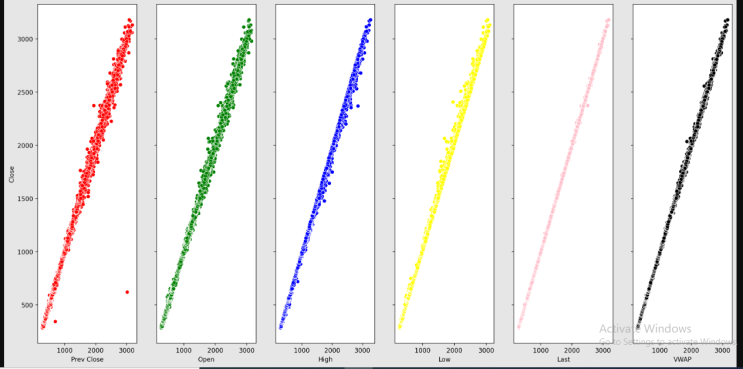
1. Distribution Plot for the SBI Stock

The Data has a positively skewed distribution which is observed in all 3 stocks namely HDFC, KOTAK, and SBI bank stock. SBIBANK stock is looking as the least volatile stock followed by HDFC and then KOTAK.

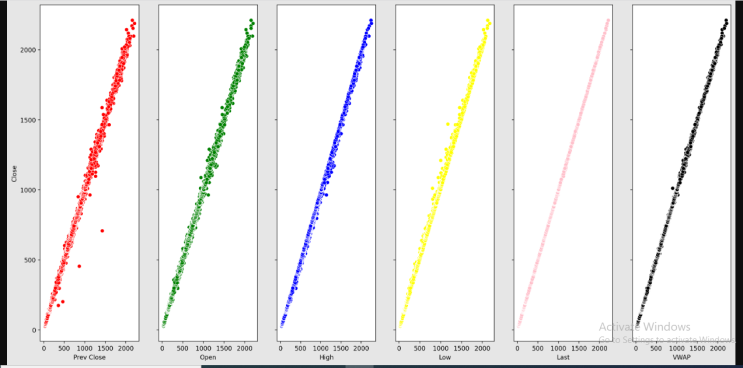


1. Box plot-HDFC, KOTAK, and SBI stock from 2000 to 2022

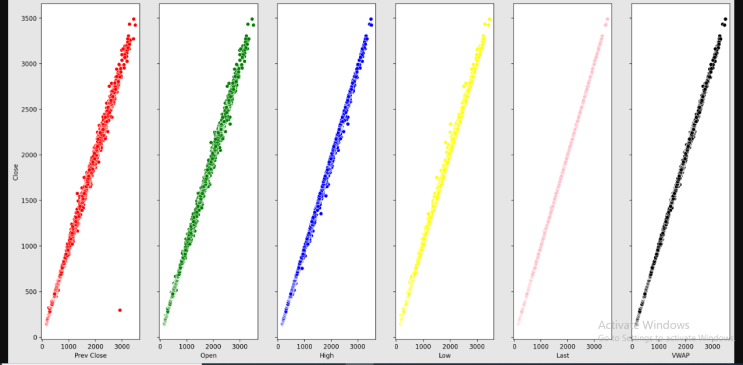
There is a large difference between the 75th %tile and max values of most of the feature variables for all 3 stocks. Therefore, it suggests that there are extreme values-Outliers in our data set.



1. Customized Scatter Plot against close price for the HDFC Stock from 2000 to 2022



1. Customized Scatter Plot against close price for the KOTAK Stock from 2000 to 2022



1. Customized Scatter Plot against close price for the SBI Stock from 2000 to 2022

A customized Scatter Plot is drawn for all feature variables against the close price of the HDFC, KOTAK, and SBI stock. It is observed that a linear relationship exists between Independent variables and the Target variable except for fewer outliers which is quite negligible.

## Data Pre-processing

The HDFC, KOTAK, and SBI data which are taken from NSE come with a lot of limitations that have to be processed.

Handling Missing values: Three of the features’ trades, ‘Deliverable Volume’, and’% Deliverable were dropped as they are having several missing values.

Features Addition: Computed variables added to the dataset are simple and exponential moving averages for rolling periods of 7, 13, 20,100, and 200 days. 1 day's previous lag values of volume are also added as features.6, 10, 14, and 30 days consecutive closing prices are tabulated week on week for the entire dataset and utilized as different feature variables. Momentum, trend, volatility, and volume indicators are also used as feature variables.

Data Scaling: Minmax Scaler is the data scaling approach that is being used. MinMax Scaler shrinks the data inside the given range, from zero to one.

## Data Modeling

Various Classification models namely Logistic Regression, Decision Tree, Random Forest, K Nearest Neighbour, and XG Boost Classifier are deployed and their prediction accuracy is compared.

When the majority of the 20 various models or all of them move in the same direction, a choice on whether to invest or not to invest in the stock under consideration must be made. Daily Trading Data of SBI and Kotak Bank from the year 2000 to 2022 are being used to repeat the entire process which had been implemented for the HDFC Bank dataset.

Table 2– Modelling strategies and Model Evaluation Rule

|  |  |
| --- | --- |
| **Modelling Strategies** | **Model Evaluation Rule** |
| Direction Detection by 6, 10, and 14 days consecutive closing prices split week on the week. | percentage change on closing price>0.7% =>Positive Trend  percentage change on closing price<-0.7% =>Negative Trend  percentage change on closing price between 0.7 and 0.7% =>Neutral |
| Go Long Direction Prediction performed separately using Momentum, Trend, Volatility, and Volume Indicators. | percentage change on closing price>0.5% =>Positive Trend  percentage change on closing price<=0.5% =>Not Positive Trend |

# FINDINGS/DISCUSSION

The Data Evaluation phase is the result of the Data Modelling phase and discusses the Metrics utilized to determine the extent of the success achieved from the different Modelling Algorithms employed on the Target Variable.

## Model Evaluation using LR Classifier for Go Long Direction Prediction

Various Classification Models are utilized to predict the direction of the close value of HDFC, KOTAK, and SBI stock and estimate using different error metrics. All the results derived from the various models are examined below.

Table 3– Model Evaluation using LR Classifier

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelling Strategies** | **HDFC** | **KOTAK** | **SBI** |
| Direction Detection by 6,10,14 days consecutive closing prices split week on the week | Precision-0.35  Recall-0.60  Accuracy-0.35 | Precision-0.37  recall-0.74  Accuracy-0.36 | Precision-0.36  Recall-1.00  Accuracy-0.36 |
| Go Long Direction Prediction using Volume Indicators | **Precision-0.98**  **Recall-0.83**  **Accuracy-0.92** | **Precision-0.99**  **Recall-0.93**  **Accuracy-0.97** | **Precision-0.92**  **Recall-0.80**  **Accuracy-0.90** |
| Go Long Direction Prediction using Momentum Indicators | Precision-0.71  Recall-0.63  Accuracy-0.76 | Precision-0.73  Recall-0.61  Accuracy-0.75 | Precision-0.69  Recall-0.62  Accuracy-0.74 |
| Go Long Direction Prediction using Trend Indicators | Precision-0.83  Recall-0.59  Accuracy-0.80 | Precision-0.76  Recall-0.48  accuracy-0.72 | Precision-0.78  Recall-0.49  Accuracy-0.74 |
| Go Long Direction Prediction using Volatility Indicators | Precision-0.93  Recall-0.47  Accuracy-0.77 | Precision-0.90  Recall-0.40  Accuracy-0.74 | Precision-0.81  Recall-0.30  Accuracy-0.70 |

From Table 3, it can be observed that Go Long Direction Prediction using Volume Indicators has given considerable precision, recall, and accuracy in direction prediction.

## Model Evaluation using RF Classifier for Go Long Direction Prediction:

Table4.Model Evaluation using RF Classifier

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelling Strategies** | **HDFC** | **KOTAK** | **SBI** |
| Direction Detection by 6,10,14 days consecutive closing prices split week on the week | **precision-0.85**  **recall-0.89**  **accuracy-0.87** | **Precision-0.71**  **recall-0.79**  **accuracy-0.74** | **Precision-0.83**  **recall-0.88**  **accuracy-0.85** |
| Go Long Direction Prediction using Volume Indicators | **precision-0.93**  **recall-0.69**  **accuracy-0.85** | **precision-0.92**  **recall-0.79**  **accuracy-0.89** | **precision-0.90**  **recall-0.73**  **accuracy-0.86** |
| Go Long Direction Prediction using Momentum Indicators | precision-0.76  recall-0.51  Accuracy-0.75 | precision-0.78  recall-0.50  accuracy-0.75 | precision-0.72  recall-0.55  accuracy-0.74 |
| Go Long Direction Prediction using Trend Indicators | precision-0.87  recall-0.56  Accuracy-0.80 | precision-0.85  recall-0.44  accuracy-0.74 | precision-0.83  recall-0.57  accuracy-0.78 |
| Go Long Direction Prediction using Volatility Indicators | precision-0.92  recall-0.53  Accuracy-0.79 | precision-0.89  recall-0.50  accuracy-0.78 | precision-0.83  recall-0.61  accuracy-0.80 |

From Table 4, it can be observed that Direction Detection has given the highest precision, accuracy, and recall in prediction. Also, Go Long Direction Prediction using Volume Indicators has given considerable precision and accuracy in direction prediction but recall can still be improved.

## Model Evaluation using XG Boost Classifier for Go Long Direction Prediction

Table 5– Model Evaluation using XG Boost Classifier

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelling Strategies** | **HDFC** | **KOTAK** | **SBI** |
| Direction Detection by 6,10,14 days consecutive closing prices split week on the week | precision-0.35  recall-0.42  accuracy-0.40 | Precision-0.38  recall-0.41  accuracy-0.40 | Precision-0.38  recall-0.47  accuracy-0.37 |
| Go Long Direction Prediction using Volume Indicators | **precision-0.90**  **recall-0.73**  **accuracy-0.86** | **precision-0.92**  **recall-0.87**  **accuracy-0.92** | **precision-0.88**  **recall-0.82**  **accuracy-0.89** |
| Go Long Direction Prediction using Momentum Indicators | precision-0.70  recall-0.61  Accuracy-0.75 | precision-0.74  recall-0.59  accuracy-0.75 | precision-0.70  recall-0.59  accuracy-0.74 |
| Go Long Direction Prediction using Trend Indicators | precision-0.85  recall-0.65  Accuracy-0.82 | precision-0.82  recall-0.61  accuracy-0.79 | precision-0.83  recall-0.67  accuracy-0.81 |
| Go Long Direction Prediction using Volatility Indicators | precision-0.84  recall-0.69  Accuracy-0.82 | precision-0.81  recall-0.63  accuracy-0.79 | precision-0.80  recall-0.67  accuracy-0.81 |

From Table 5, it can be observed that Go Long Direction Prediction using Volume Indicators has given considerable precision, recall, and accuracy in direction prediction.

## Direction Detection and Go Long Direction Prediction using the best classifier model

Table6. Leader Board-comparison of Metrics for Direction Detection and Go Long Direction Prediction using the best classifier model

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Modelling Strategies** | **HDFC** | **KOTAK** | **SBI** |  |  |  |  |
| Direction Detection by 6,10,14 days consecutive closing prices split week on the week  (RF Classifier) | **precision-0.85**  **recall-0.89**  **accuracy-0.87** | **Precision-0.71**  **recall-0.79**  **accuracy-0.74** | **Precision-0.83**  **recall-0.88**  **accuracy-0.85** |  |  |  |  |
| Go Long Direction Prediction using  Volume Indicators  (LR Classifier) | **precision-0.98**  **recall-0.83**  **accuracy-0.92** | **precision-0.99**  **recall-0.93**  **accuracy-0.97** | **precision-0.92**  **recall-0.80**  **accuracy-0.90** |  |  |  |  |
| Go Long Direction Prediction using  Momentum Indicators  (LR Classifier) | precision-0.71  recall-0.63  Accuracy-0.76 | precision-0.73  recall-0.61  accuracy-0.75 | precision-0.69  recall-0.62  accuracy-0.74 |  |  |  |  |
| Go Long Direction Prediction using  Trend Indicators  (XG Boost Classifier) | precision-0.85  recall-0.65  Accuracy-0.82 | precision-0.82  recall-0.61  accuracy-0.79 | precision-0.83  recall-0.67  accuracy-0.81 |  |  |  |  |
| Go Long Direction Prediction using  Volatility Indicators  (XG Boost Classifier) | precision-0.84  recall-0.69  Accuracy-0.82 | precision-0.81  recall-0.63  accuracy-0.79 | precision-0.80  recall-0.67  accuracy-0.81 |  |  |  |  |

From Table 6, it can be observed that RFclassifier modelling has given the highest efficiency in Direction Detection among all Modelling techniques namely LR, DT, RF, KNN, and XG Boost Modelling. This has been tested and proven with 6, 10, and 14-day consecutive closing prices split week on week as 6, 10, and 14 feature variables. Also, LR classifier modelling has provided the best precision, recall, and accuracy for Go Long Direction prediction using Volume Indicators.

## Utility from the Business perspectives

For a stop loss of 2.0 reward-risk ratio for approximately 0.8 Precision would be 2\*.8/2\*.2=4:1 if a 0.5% difference in consecutive day close price for any stock is only 2.0.for higher percentage difference reward to risk ratio would be higher.

Here, Modelling Algorithms were provided for the close price of HDFCBANK, KOTAK BANK, and SBIBANK Stock over 20 years with the train test split of 70%:30%. If we invest Rs.10000 for 6 years and roughly calculate profit with 0.5% change on close price with the highest precision in detecting true positives then the following results are possible as per the formulae given below:



Using Trend Indicators with the highest precision of 0.85 for HDFCBANK stock, the confusion matrix provides information as below:



Figure 10. confusion matrix For HDFCBANK Stock using

Trend Indicators as Feature variables

Therefore, Net Returns will be:

0.5\*10000\*282\*0.85/100-0.5\*10000\*51\*0.85/100=Rs. 9817.5 profit which would be

9817.5 / (10000\*6)\*100=16.36% returns.

## Risk-Adjusted Returns

The real Data dump is imported for HDFC, KOTAK, and SBI stock between 2000 till 2022. Then the Return, Variance, and Volatility of these stocks are calculated following which the Annualized return to Risk ratio and finally, the Sharpe ratios are calculated. The Sharpe ratio for HDFC, KOTAK, and SBI Stock is calculated as 0.173818, 0.149589, and 0.005306 respectively.

Therefore, from the results obtained it becomes evident that HFDC shows a better Return vs. Risk performance over the specified period compared to KOTAK stock followed by the SBI stock which shows the least Return vs. Risk performance.

# CONCLUSION/IMPLICATIONS

The 6-day consecutive closing price for the stock under consideration is being taken. These 6 days' consecutive closing prices will be getting tabulated week on week for the entire dataset and will be utilized as 6 different feature variables for building the classification Model. The difference between the 7th and 8th-day Closing price is determined. The 0.7% difference,1% difference, and 1.5% difference are different classes of direction for which the rule is being set which is to be followed for computing the direction change as either positive change, negative change, or no change. Once it is determined say for example 0.7% difference has the best prediction accuracy among all different classes of direction then the similar process is again repeated for a range of consecutive days to be utilized as the feature variable increased to 10 days and 14 days using the Classifier Modelling algorithm which provided the best directional prediction. Similarly, All different types of technical indicators namely momentum indicators, trend indicators, volatility indicators, and volume indicators are utilized as feature variables based on the input dataset and Various Classification models namely Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbour, and XG Boost Classifier are deployed and their prediction accuracy is compared using Metrics namely precision, recall, f1-score, accuracy score, and ROC AUC Score. The construction of all 20 models was used to predict the direction of the close price for the stock under consideration. When the majority of the various models or all of them move in the same direction, a choice on whether to purchase or not to purchase the stock must be made.

This paper then solely focuses on predicting the direction of the close price of the HDFC stock using classification algorithms Techniques. A later similar process is applied for predicting the direction of the close price of other stocks in the banking sector namely SBI and KOTAK stocks. In the Future, there is a deployment Dashboard proposed. As per the proposal for future assignments, the dashboard takes API as an input Derived from the machine learning algorithms and can be utilized in predicting the direction of the close price for any stock in the Banking sector. Any stock on the stock market can utilize the same procedure to forecast buy or not to buy choices, which is helpful.

# RECOMMENDATIONS

This paper has not discussed how to address one major drawback of stock prediction, namely that over different periods the stock returns can change drastically.

In future projects, it can be shown how to define Bullish and Bearish regimes using modern machine learning techniques. The Sentiment Analysis Approach may also need to be explored using Text Analytics for predicting stock market returns. In the Future, there is a deployment Dashboard proposed. An intelligent automated system for Options Trading would be also the next step forward.

##### References

Al-Bairmani, Z. A. A., & Ismael, A. A. (2021). Using Logistic Regression Model to Study the Most Important Factors Which Affects Diabetes for the Elderly in the City of Hilla / 2019. *Journal of Physics: Conference Series*, *1818*(1). https://doi.org/10.1088/1742-6596/1818/1/012016

Alhomadi, A. (2021). Forecasting stock market prices : A machine learning approach. *Digital Commons*, *11*(2), 16–36.

Dar, A. N. (2021). PRINCIPAL COMPONENT ANALYSIS (PCA) (Using Eigen Decomposition). *Gsj*, *9*(7), 240–252. www.globalscientificjournal.com

Elbialy, B. A. (2019). The Effect of Using Technical and Fundamental Analysis on the Effectiveness of Investment Decisions of Traders on the Egyptian Stock Exchange. *International Journal of Applied Engineering Research*, *14*(24), 4492–4501. http://www.ripublication.com

Hansen, K. B. (2020). The virtue of simplicity: On machine learning models in algorithmic trading. *Big Data and Society*, *7*(1). https://doi.org/10.1177/2053951720926558

Huang, Y., Capretz, L. F., & Ho, D. (2021). Machine Learning for Stock Prediction Based on Fundamental Analysis. *2021 IEEE Symposium Series on Computational Intelligence, SSCI 2021 - Proceedings*. https://doi.org/10.1109/SSCI50451.2021.9660134

Jena, M., & Dehuri, S. (2020). Decision tree for classification and regression: A state-of-the art review. *Informatica (Slovenia)*, *44*(4), 405–420. https://doi.org/10.31449/INF.V44I4.3023

Kimbonguila, A., Matos, L., Petit, J., Scher, J., & Nzikou, J.-M. (2019). Effect of Physical Treatment on the Physicochemical, Rheological and Functional Properties of Yam Meal of the Cultivar “Ngumvu” From Dioscorea Alata L. of Congo. *International Journal of Recent Scientific Research*, *10*, 30693–30695. https://doi.org/10.24327/IJRSR

Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem. *ACM International Conference Proceeding Series*, 412–419. https://doi.org/10.1145/3453892.3461323

Mukerji, P., Chung, C., Walsh, T., & Xiong, B. (2019). The Impact of Algorithmic Trading in a Simulated Asset Market. *Journal of Risk and Financial Management*, *12*(2), 68. https://doi.org/10.3390/jrfm12020068

Omta, W. A., van Heesbeen, R. G., Shen, I., de Nobel, J., Robers, D., van der Velden, L. M., Medema, R. H., Siebes, A. P. J. M., Feelders, A. J., Brinkkemper, S., Klumperman, J. S., Spruit, M. R., Brinkhuis, M. J. S., & Egan, D. A. (2020). Combining Supervised and Unsupervised Machine Learning Methods for Phenotypic Functional Genomics Screening. *SLAS Discovery*, *25*(6), 655–664. https://doi.org/10.1177/2472555220919345

Rajkar, A., Kumaria, A., Raut, A., & Kulkarni, N. (2021). Stock Market Price Prediction and Analysis. *International Journal of Engineering Research & Technology*, *10*(06), 115–119.

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. *Electronics (Switzerland)*, *10*(21). https://doi.org/10.3390/electronics10212717

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, *20*(1), 3–29. https://doi.org/10.1177/1536867X20909688

Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies*, *7*(2). https://doi.org/10.3390/ijfs7020026

Sonkiya, P., Bajpai, V., & Bansal, A. (2021). *Stock price prediction using BERT and GAN*. http://arxiv.org/abs/2107.09055

Thanekar, G. S., & Shaikh, Z. S. (2021). Analysis and Evaluation of Technical Indicators for Prediction of Stock Market. *International Journal of Engineering Research & Technology (IJERT)*, *10*(May), 341–344.

Wang, L. (2019). Research and Implementation of Machine Learning Classifier Based on KNN. *IOP Conference Series: Materials Science and Engineering*, *677*(5), 0–5. https://doi.org/10.1088/1757-899X/677/5/052038

Zhang, P., Jia, Y., & Shang, Y. (2022). Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*, *18*(6). https://doi.org/10.1177/15501329221106935