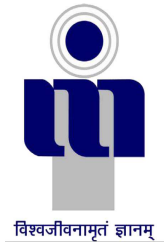
AUTOMATED PASSIVE INCOME FROM STOCK MARKET USING MACHINE LEARNING AND BIG DATA ANALYTICS

by

Pranjal Mankar

Roll. No.: 2016IPG-046

Submitted to: **Dr. Somesh Kumar**



ABV–INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT GWALIOR (M.P.), INDIA

Introduction

1. With the evolution of Machine Learning, the Internet of Things (IoT), and Big Data technologies, digital data has increased exponentially across various fields.
2. One of the fields has been the financial capital market of stocks, bonds, commodities, foreign exchange, and crypto-currencies where supercomputers largely trade securities with high computational ability.
3. A novice trader or investor with less to no experience in financial markets feels it really difficult to search for good trades or stocks to invest his/her hard-earned money on a short to long-term time horizon.
4. This project focuses on developing a universal trend trading indicator that can analyze and predict the overall future trend of any stock, bond, commodity, forex, or cryptocurrency with the highest possible accuracy.

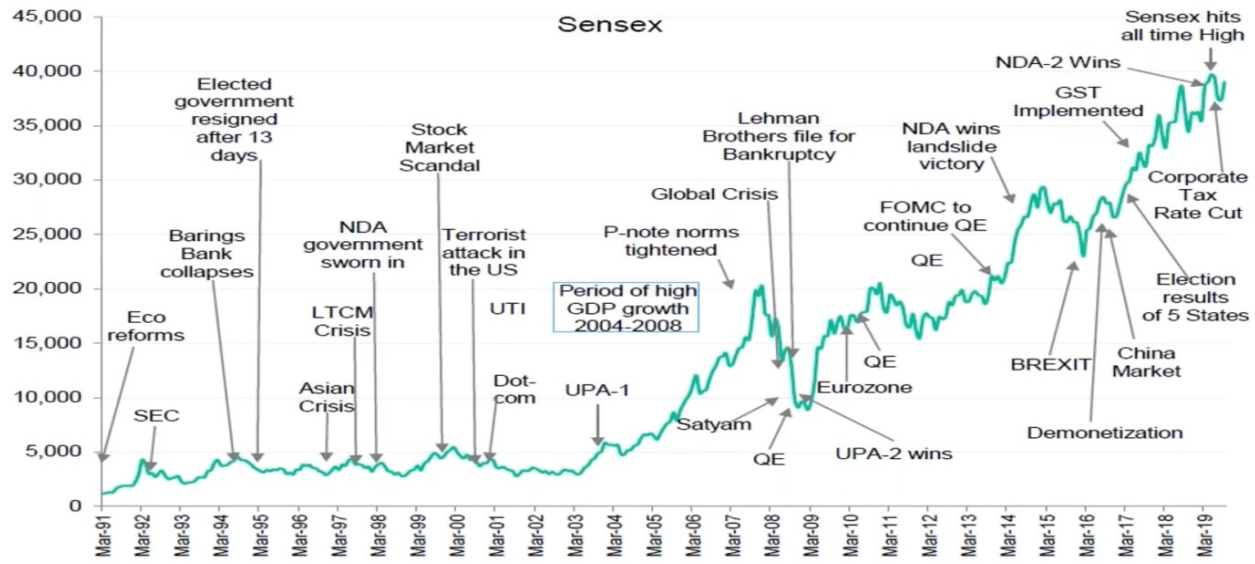
Motivation

[1]

[1] https://www.techjockey.com/blog/best-stock-trading-software.

Motivation

1. Many irregular fundamental factors such as government policies, GDP, interest rates, solvency, supply and demand, development, exchange rates, politics, and current events affect the daily price fluctuations of various stocks of the stock markets.



2

[2] **https://www.youtube.com/watch?v=ZdcOSuQtltI**.

Motivation

2. Novice investors with the least financial intelligence, end up relying on expert advice, or broker's stock picks which are highly risky.

[3]


[4]

[3] https://www.indiatoday.in/business/story/markets-in-red-after-sensex-sheds-over-400-points-nifty-tests-11-700-mark-1

[4] https://www.livemint.com/market/stock-market-news/coronavirus-impact-fear-index-soars-to-11-year-high-11584126

Literature review

- J. Patel et al. proposed experimentation based on a two-stage fusion method for predicting future values of two Indian stock market indices namely Nifty 50 and Sensex. The first stage is a single-stage fusion approach that involves the Support Vector Regression (SVR) model which takes the t^{th} day input values of ten popular technical indicators to predict the $(t+n)^{th}$ day future values of indices as the output. This output serves as the input to the latter two-stage fusion method where Artificial Neural Network (ANN), Random Forest (RF), and SVR form mapping hybrid models to predict accurate future closing price values for 1-10, 15, and 30 days in advance. The overall experiment is based on 10 years of historical price data of those two indices [5].
- A. Sheta uses the Takagi-Sugeno fuzzy model technique for two non-linear processes to predict the future values of the S&P 500 for the upcoming week. The TS fuzzy model has been developed with two basic steps which involve determining the consequent structure of membership functions and identification of rule-based antecedent quantities using the model input data for software effort estimation and stock market prediction [6].
- J. Mandziuk and M. Jaruszewicz presented a neuro-evolutionary method to predict short term future values of European and American indices. They gather historical data from the German Stock Exchange, Tokyo Stock Exchange, and New York Stock Exchange. Their approach involves the use of the genetic algorithm as the prediction engine to find a suboptimal set of input variables for short term price prediction [7].
- M. Makrehchi et al. used the sentiment analysis concept to propose an event-based novel approach to predict the future label and returns of indices and individual stocks by extracting contemporaneous text from social media sources like twitter posts. They assign each tweet a positive or negative label to train a model to create better trading strategies which can provide significantly higher returns over other baseline methods [8].
- S. Lauren and S. Harlili used a combination of time series data of simple moving average and financial news to predict the future trend of any given stock. They combined the previous one year of historical price data as well as the financial news to classify the trend as positive, negative, or neutral. An artificial neural network model is used to determine the result which also indicates that the prediction responsiveness can be improved with good quality financial news [9].

Literature review

- M. Skuza and A. Romanowski also used the sentiment analysis model to predict future stock prices based on the analysis of social media services like Twitter microblogging and widely available historical stock market data. They developed the Map Reduce programming model to evaluate predictions for different time intervals to perform multi-time-frame analysis [10].
- G. Attigeri et al. proposed a big data approach to perform technical analysis and fundamental analysis over various stocks to predict the stock market. Historical price data is used as an input to perform the technical analysis while the fundamental analysis is done using social media data using sentiment analysis. The result produced by their model was used to show that the predicted values of their model are closely related to recent news and social data [11].
- D. Nelson et al. built a prediction model to study the usage of LSTM (Long-Short term memory) neural network along with suitable technical indicators to predict the future trend of various stocks. The empirical results obtained were promising, getting up to an accuracy of 55.9% when classifying the trend to be bullish, bearish or neutral [12].
- M. Vargas et al. proposed their deep learning methods to predict intraday directional movements of the S&P 500 index using a set of highly accurate technical indicators and financial news titles. Their proposed methods use Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) model which can detect and analyze complex charting patterns to forecast the directional move of any given stock [13].
- S. Xiao et al. with the help of IoT and big data technology proposed and back-tested a self-evolving trading strategy to trade commodity futures with a high yield to risk ratio. They gathered historical price data from the Shanghai Stock Exchange and Dalian Stock Exchange to back-test the trading strategy on other future contracts in an attempt to enhance its diversity. The root mean of the squared error (RMSE) and the mean absolute percentage error (MAPE) was used to evaluate the performance of the model [14].
- C. Lee and I. Paik collected real-time Twitter tweets, stock value, and latest news from reliable sources yet Apache Spark, Apache Storm, and Apache Flink framework were used for data storage, real-time text pre-processing, and live streaming purposes. News data was used to train the entire framework model which can classify the tweets with the best accuracy. They developed a visualization web interface to evaluate the performance and for a quick understanding of results. With the best 77% accuracy of Twitter data classification, they were able to predict 80% of the separation of stock market trends [15].

Literature review

- S. Maini and K. Govinda used the Random Forest model and Support Vector Machine model to propose an approach towards the prediction of the future trend of stock market indices. Their dataset consisted of historical news from Reddit World News Channel, the stock data of Dow Jones Industrial Average (DJIA) from 2008 to 2016, and data from the Guardian's restful API. Random Forest model was used for data regression and pre-processing which produced an accuracy of 84.3% - 86.2% while the Support Vector Machine model was as it is used for classification purposes with 82.2% - 84.6% of predicted accuracy [16].
- E. Sin and L. Wang attempted to explore the neurological relationship between the historic features of Bitcoin data and the next 50 days of a change in price using the Genetic Algorithm based Artificial Neural Network ensemble approach. They constructed a Multi-Layered Perceptron based model to predict the directional price of Bitcoin from a given set of 200 features of cryptocurrency. They also built and back-tested a trading strategy for over 50 days that generated a whopping 85% return on investment which out-performed their previously built trend-following models [17].
- L. Chen et al. used a combination of the deep learning prediction model, an autoencoder, and a restricted Boltzmann machine to compare three traditional artificial neural network models to find out which model better performs the prediction of the Chinese Stock Market. They took a 1 min high-frequency transactional data of the CSI 300 futures contract as an input to all three neural network models and found out that the deep learning method outperforms all three existing neural network models' performance. This suggests that deep learning captures the non-linear nature of transactional data and enhances predictability [18].
- X. Zhang et al. proposed accurate and robust predictions of directional stock movements by extending the Multiple Instance Learning model to integrate variable quantitative datasets from various reliable online platforms with the help of Restricted Boltzmann Machines and sentence2vec to produce desirable results [19].
- R. Camara et al. collected a high-volume dataset to present a computational intelligent tool which is based on fuzzy logic data analytics to study the effect of hurricanes on the stock market and specific individual sectors. They performed various classifications, sentimental analysis, and simulation through fuzzy logic based on probabilistic approaches on NYSE stock data. The resultant returns were in -66.0% to +59.5% range [20].

Literature review

- G. Liu and X. Wang gathered news corpus and historical data from China Security Index and Standard & Poor's 500 (S&P500) as an input to numerical based attention (NBA) method for dual information stock market prediction. The designed numerical model is firstly compared with existing basic model and efforts are made towards reducing the noise in predicting future values [21].
- D. Shah and H. Isah were able to achieve the accuracy of 70.59% in short-term stock movement prediction which worked on sentiment analysis of news data. They designed a dictionary-based sentiment analysis model for data pre-processing, retrieval, and classification models for gauging how news sentiment can affect pharmaceutical sector stocks. Hybrid models like Auto-Regressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) neural networks can be further applied for more accurate predictions across other sectors [22].
- M. Wen et al. introduced a new conventional method to reduce noise in financial time series data with the help of sequence reconstruction of frequently occurring patterns while utilizing a neural network model. Their experimentations outperformed the basic sentiment analysis approaches and time series trading patterns with at most 7% accuracy with hybrid models like motif extraction, sequence reconstruction and convolutional reconstructed series [23].
- Z. Peng with the help of a robust Cloudera-Hadoop framework handled a large dataset of selected US stocks to identify their day-to-day data gain or loss. Gathered large dataset is injected, stored, pre-processed in the Spark machine learning framework, and R Squared, Mean Average Error values of necessary inputs is calculated to enhance the accuracy of the outcome [24].

Objectives

- For technical analysis, historical price data of stocks or indices are to be collected from reliable stock exchange sources. For fundamental analysis, large institutional fund buying and selling data are to be collected from their investment reports.
- Collected big data is to be pre-processed, analyzed and classified by our unique stock trend prediction algorithms while entry levels, future trend and exit levels of a given stock are to be identified as an outcome.
- Outcome results are to be backtested with the historical trend of the stock and the indicator's prediction accuracy is to be enhanced.
- Indicator's prediction outcome is to be deployed with the Streak API platform for automated buying and selling of stocks and other securities [2].

System architecture and Methodology

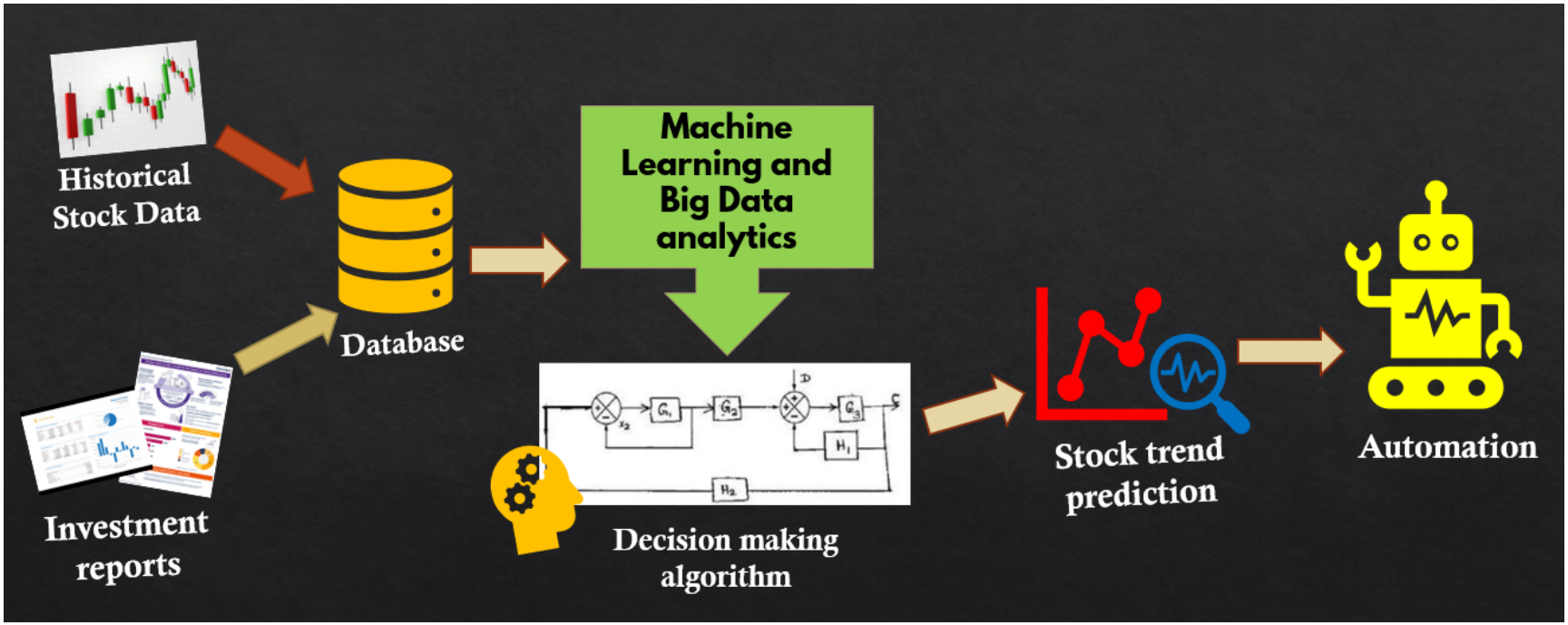


Figure 1: System Design.

System architecture and Methodology Steps

Step 1: The historically traded big datasets of all Nifty 50 stocks are collected from National Stock Exchange (NSE) website.

Step 2: The investment reports and trade-wise equity datasets of FPI/FII from the year 2003 - 2019 will be collected from the National Securities Depository Limited (NSDL) website.

Step 3: The entire datasets and investment reports is gathered and stored in a local database to retrieve and derive all the useful information and further pre-processing.

System architecture and Methodology Steps

Step 4: Dataset is taken as an input to our decision making model which will perform its own evaluations and provide entry-exit signals and predict the future trend of any given stock, in the form of trend trading indicator.

Step 5: Once the desirable accuracy of back-testing of outcome is achieved, the entire system will be deployed to the stock and securities across the globe to identify and filter out at what particular stocks or securities, the trend trading indicator will produce accurate predictions.

Step 6: All the stocks and securities with positive predictions and high returns on investments will be finalized for automated buying and selling using Tradingview.com RestAPI and Streak.io API [1-2].

Decision making model

Following are the components of our Decision making model:

1. Price action model
2. Technical analysis
3. Simple exponential smoothing

Components of price action model

1. **Higher-high and lower-low price action**: We implemented this price action concept to decide the current trend as an output in our decision-making model.
2. **Pyramiding entry-exit**: We implemented the pyramiding concept to provide the entry-exit signals as an output in our decision-making model.
3. **Averaging**: To average out our pyramiding entry-exit signals.

Components of price action model

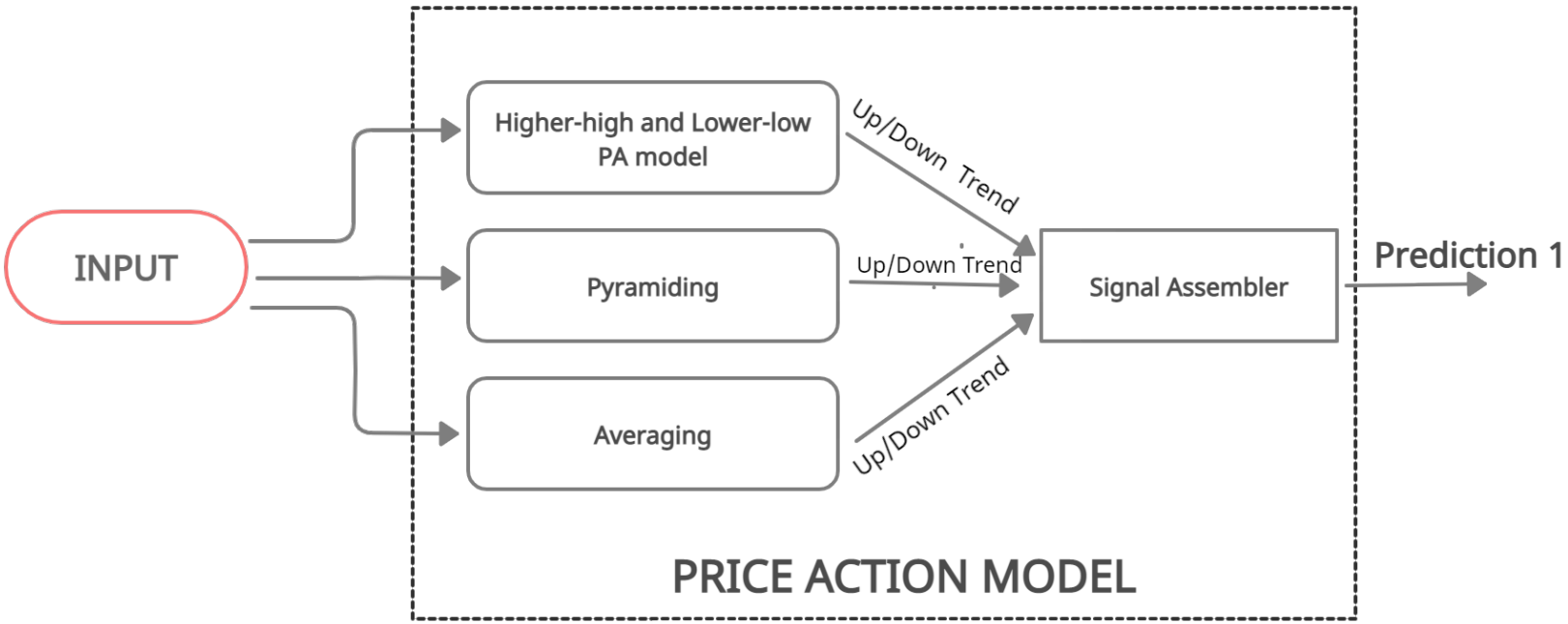


Figure 2: The architecture of the price action model.

Components of Technical analysis

1. **Average True Range (ATR)**: We used this measure of stock volatility to predict if the stock is still in a trending or a consolidating zone.
2. **Pyramiding entry-exit**: We used this property of an oscillator in our proposed model to identify the current trend's strength of a given stock.
3. **Exponential moving average**: Exponential moving average measures the overall trend of any given security over time.

Components of Technical analysis

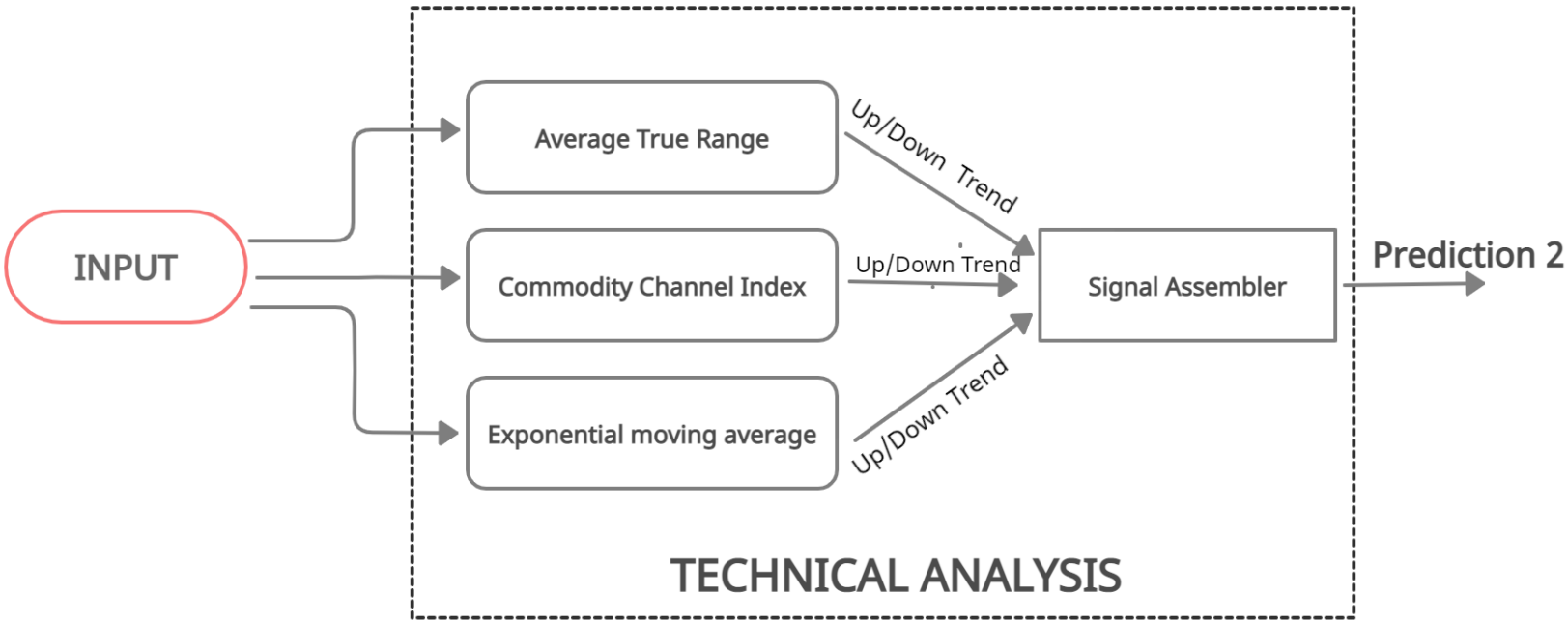


Figure 3: The architecture of the technical analysis model.

Ensembling

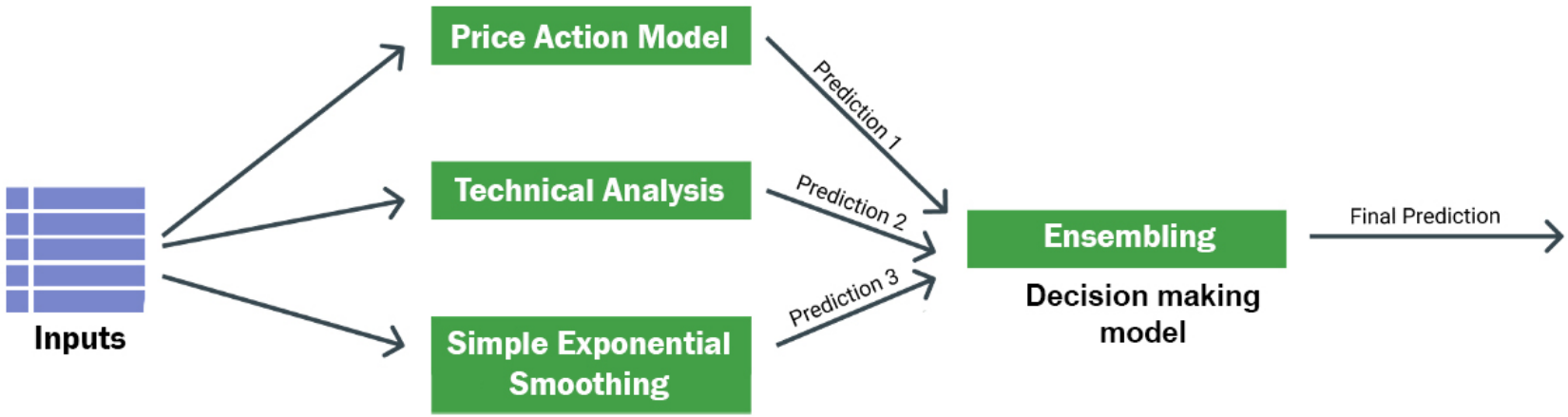


Figure 4: The ensembled architecture of decision-making model.

Visual representation of Trend Trading Indicator

Figure 5: Visual representation of trend trading indicator plotted on Tradingview.com charting platform.

Intermediate results

Table 1: Evaluation of profitability of various stocks for annualized returns.

INTERMEDIATE RESULTS-

Company	NET PROFIT	TOTAL TRADES	PERCENT PROFITABLE	PROFIT FACTOR	MAX DRAWDOWN	AVERAGE TRADE
ADANIPORT	299025.00 ↑	77	44.16%	2.136	45150.00 ↓	3883.44
ASIANPAINTS	530700.00 ↑	72	43.06%	1.609	195900.00 ↓	7370.83
BAJFINANCE	983550.00 ↑	79	37.97%	1.369	624825.00 ↓	12450.00
BPCL	30975.00 ↑	83	43.37%	1.174	31875.00 ↓	373.19
COALINDIA	16575.00 ↑	80	28.75%	1.225	19875.00 ↓	207.19
DRREDDY	53550.00 ↑	82	35.37%	1.029	424425.00 ↓	653.05
HDFC	293025.00 ↑	91	38.46%	1.203	447225.00 ↓	3220.05
ICICIBANK	210900.00 ↑	82	37.8%	1.906	52575.00 ↓	2571.95
INFOSYS	87000.00 ↑	86	30.23%	1.149	263700.00 ↓	1011.63
MARUTI	412200.00 ↑	85	36.47%	1.125	601200.00 ↓	4849.41
NESTLEIND	980775.00 ↓	79	35.44%	0.834	2197950.00 ↓	11737.97
RELIANCE	513600.00 ↑	76	47.37%	1.884	109950.00 ↓	6757.89
SBIN	187275.00 ↑	90	41.11%	2.106	37425.00 ↓	2080.83
WIPRO	187950.00 ↑	69	43.48%	2.227	27975.00 ↓	2723.91

quantities: 1500 (each)

Simple exponential smoothing

The forecasted value at time t+1 is based on the forecasted value at time t (and so indirectly on all the previous time values). In particular, for some α where $0 \leq \alpha \leq 1$, for all $t > 1$, we define

$$Y_{t+1} = \alpha X_{t+1} + (1 - \alpha) Y_t = Y_t + \alpha(X_{t+1} - Y_t)$$

Here,
Y_{t+1} = smoothed statistic, it is the simple weighted average of current observation X_t
Y_t = previous smoothed statistic
α = smoothing factor of data; $0 < \alpha < 1$
t = time period
Simple exponential smoothing technique was used in our proposed model to optimize the trend predicting capability of the indicator.

Optimized results

Table 2: Optimized evaluation of profitability of various stocks for annualized returns.

Company	NET PROFIT	TOTAL TRADES	PERCENT PROFITABLE	PROFIT FACTOR	MAX DRAWDOWN	AVERAGE TRADE
ADANIPORT	375800.00	65	81.23%	4.284	67955.00	4765.66
ASIANPAINTS	684500.00	72	77.97%	2.937	257980.00	12795.60
BAJFINANCE	1317580.00	47	62.30%	2.249	872495.00	34975.50
BPCL	33460.00	92	86.28%	3.246	47158.00	750.67
COALINDIA	27860.00	64	63.15%	2.237	41870.00	503.48
DRREDDY	61750.00	67	57.69%	2.276	784600.00	1537.55
HDFC	568920.00	75	42.24%	2.370	756890.00	7826.00
ICICIBANK	286450.00	64	73.20%	3.114	112795.00	4762.93
INFOSYS	96450.00	68	74.67%	2.775	467920.00	6729.77
MARUTI	759850.00	59	64.38%	2.048	824950.00	121750.37
NESTLEIND	438600.00	61	47.27%	1.657	4389500.00	43150.00
RELIANCE	716890.00	49	65.63%	3.943	167950.00	28640.86
SBIN	316800.00	62	69.54%	5.357	78160.00	4321.75
WIPRO	246720.00	55	57.23%	4.468	63790.00	6750.45

quantities: 1500 (each)

Comparison of profitability

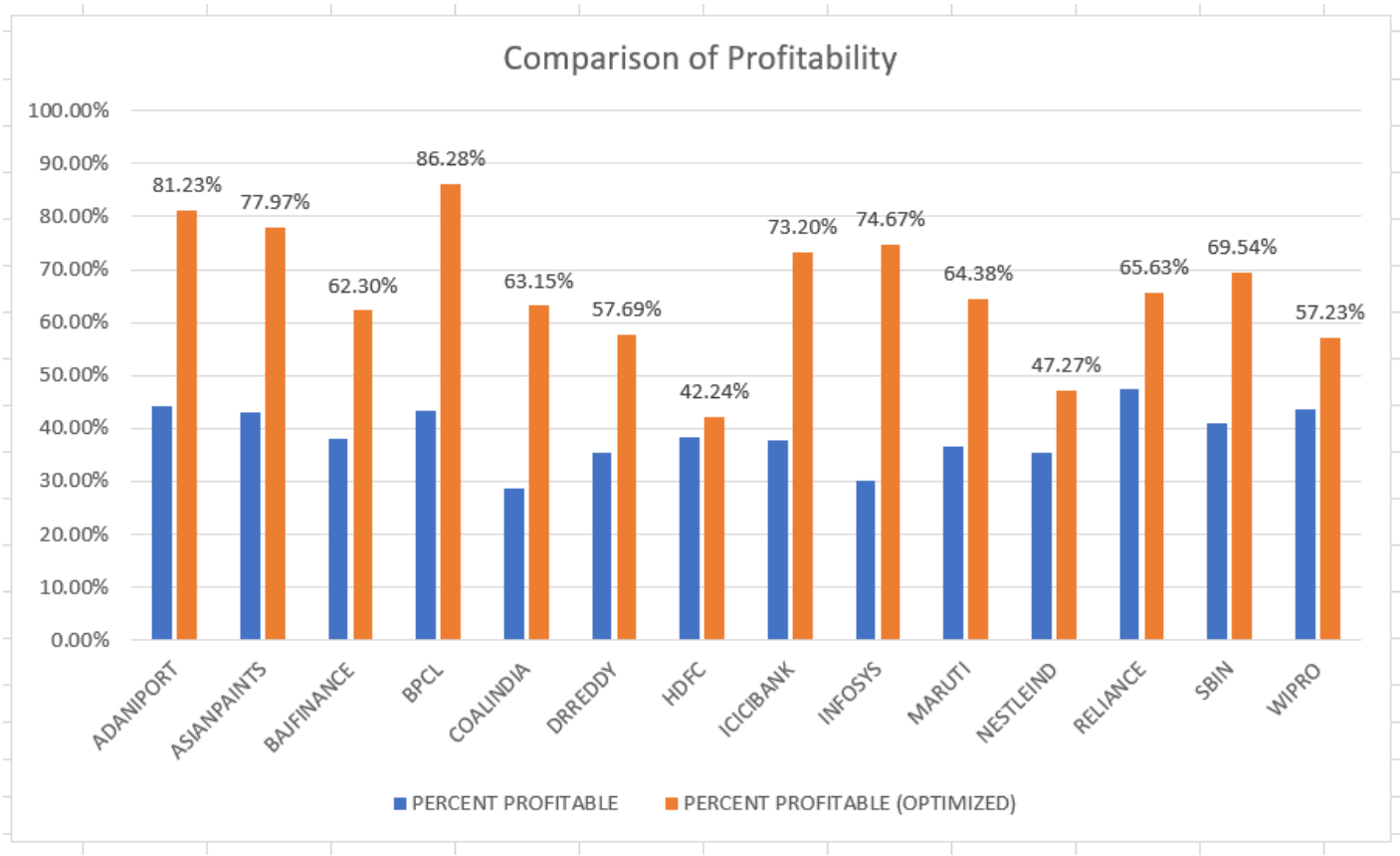


Table 3: Profitability comparison of the intermediate model with the optimized model.

Comparison with existing research

1. In my research, investment report dataset was used as a sentiment indicator for knowing the institutional interest in any given stock, which was not observed in any of the existing research works.
2. In the base paper of my research work, the researchers used set of simple moving averages and news classification and achieved the accuracy of 97.27% to predict the future trend. Whereas, in my project, I used set of exponential moving averages and price action models to predict the future trend and achieved a profitability factor of 86.28%

Technologies I used

1. **Charting platform**- Tradingview.com
2. **Programming languages**- Python (for machine learning models), Pinescript (for trend indicator tool designing)
3. **Tools and libraries**- Pandas, Numpy, Keras, SciPy, Matplotlib, Scikit-learn, Pytorch, Streak.io API, Tradingview.com RestAPI
4. **Datasets**- NSE (nseindia.com), NSDL (nsdl.co.in)
5. **Automation platform**- Streak (streak.zerodha.com)

Research outcome and Novelty of the proposal

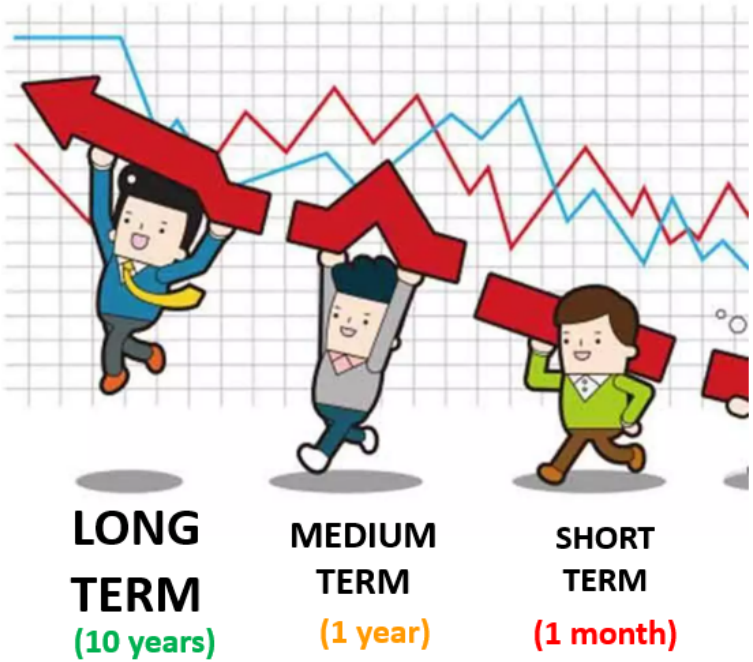
[1] Technical Indicator can provide entry level, future trend and exit levels of all stocks, indexes, commodities, foreign exchange securities.

Without my project

With my project

Expected research outcome and Novelty of the proposal

[2] Indicator can do multi time-frame analysis to decide the short term, medium term and long-term future trend of any stocks.



Expected research outcome and Novelty of the proposal

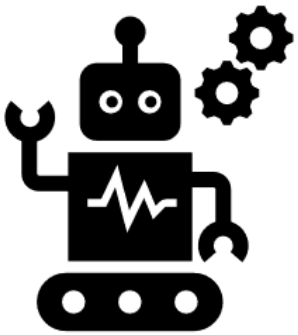
[3] Indicator can also predict and give prior alert and warnings about upcoming short-term corrections, medium term recessions and long-term stock market crashes to investors. So, investors can save their investments from losses.

Without my Indicator

With my Indicator

Expected research outcome and Novelty of the proposal

[4] A person can also deploy this indicator with their Demat account to do automated algorithmic short-term trading and long-term investing. So, anybody can create passive income from my indicator.



References I

[1] in.tradingview.com/rest-api-spec/

[2] streak.readme.io/

[3] www.nseindia.com/market-data/live-equity-market

[4] www.fpi.nsdl.co.in/web/Reports/ReportsListing.aspx–

[5] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using a fusion of machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2162-2172, 2015.

[6] A. Sheta, "Software Effort Estimation and Stock Market Prediction Using Takagi Sugeno Fuzzy Models," in Proc. *IEEE International Conference on Fuzzy Systems*, Canada, 2006, pp. 171-178.

[7] J. Mandziuk and M. Jaruszewicz, "Neuro-evolutionary approach to stock market prediction," in Proc. *International Joint Conference on Neural Networks*, USA, 2007.

[8] M. Makrehchi, S. Shah and W. Liao, "Stock Prediction Using Event-Based Sentiment Analysis," in Proc. *IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, USA, 2013, pp. 337-342.

[9] S. Lauren and S. Harlili, "Stock trend prediction using simple moving average supported by news classification," in Proc. *International Conference of Advanced Informatics: Concept, Theory, and Application (ICAICTA)*, Indonesia, 2014, pp. 135-139.

References II

[10] M. Skuza and A. Romanowski, "Sentiment analysis of Twitter data within big data distributed environment for stock prediction," in Proc. Federated Conference on Computer Science and Information Systems (FedCSIS), Poland, 2015, pp. 1349-1354.

[11] G. Attigeri, P. Manohara, P. Radhika and A. Nayak, "Stock Market Prediction: A Big Data Approach," in Proc. *TENCON 2015 - 2015 IEEE Region 10 Conference*, China, 2015.

[12] D. Nelson, A. Pereira, and R. Oliveira, "Stock market's price movement prediction with LSTM neural networks," in Proc. *International Joint Conference on Neural Networks (IJCNN)*, USA, 2017, pp. 1419-1426.

[13] M. Vargas, B. Lima, and A. Evsukoff, "Deep learning for stock market prediction from financial news articles," in Proc. *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, France, 2017.

[14] S. Xiao, H. Yu, Y. Wu, Z Peng, and Y. Zhang, "Self-Evolving Trading Strategy Integrating Internet of Things and Big Data," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 2518-2525, 2017.

[15] C. Lee and I. Paik, "Stock market analysis from Twitter and news based on streaming big data infrastructure," in Proc. *IEEE 8th International Conference on Awareness Science and Technology (iCAST)*, Taiwan, 2017, pp. 312-317.

References III

[16] S. Maini and K. Govinda, "Stock market prediction using data mining techniques," in Proc. *International Conference on Intelligent Sustainable Systems (ICISS)*, India, 2017, pp. 664-661.

[17] E. Sin and L. Wang, "Bitcoin price prediction using ensembles of neural networks," in Proc. *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, China, 2017, pp. no. 666-671.

[18] L. Chen, Z. Qiao, M. Wang, C. Wang, R. Du, and H. Stanley, "Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market?" *IEEE Access*, vol. 6, pp. 48625-48633, 2018.

[19] X. Zhang, S. Qu, J. Huang, B. Fang, and P. Yu, "Stock Market Prediction via Multi-Source Multiple Instance Learning," *IEEE Access*, pp. 50720-50728, 2018.

[20] R. Camara, A. Cuzzocrea, G. Grasso, C. Leung, S. Powell, J. Souza and B. Tang, "Fuzzy Logic-Based Data Analytics on Predicting the Effect of Hurricanes on the Stock Market," in Proc. *IEEE International Conference on Fuzzy Systems (FUZZIEEE)*, Brazil, 2018.

[21] G. Liu and X. Wang, "A Numerical-Based Attention Method for Stock Market Prediction with Dual Information," *IEEE Access*, vol. 7, pp. 7357-7367, 2018.

References IV

[22] D. Shah, H. Isah and F. Zulkernine, "Predicting the Effects of News Sentiments on the Stock Market," in Proc. *IEEE International Conference on Big Data (Big Data)*, USA, 2018, pp. 4705-4708.

[23] M. Wen, P. Li, L. Zhang, and Y. Chen, "Stock Market Trend Prediction Using High-Order Information of Time Series," *IEEE Access*, vol. 7, pp. 28299-28308, 2019.

[24] Z. Peng, "Stocks Analysis and Prediction Using Big Data Analytics," in Proc.*International Conference on Intelligent Transportation, Big Data Smart City (ICITBS)*, China, 2019, pp. 309-312.

The End

Thank You
For Your Attention