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

Progress report

M.Tech Thesis Evaluation

May 2021

by

Pranjal Mankar

(2016IPG-046)

under the supervision of

Dr. Somesh Kumar

विश्वजीवनामृतं ज्ञानम्

ABV-INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT GWALIOR-474 010

Thesis Certificate

I hereby certify that the work, which is being presented in the report/thesis, entitled Automated Passive Income from Stock Market using Machine Learning and Big Data Analytics, in fulfillment of the requirement for the award of the degree of **Integrated Post Graduate** in **Information Technology** and submitted to the institution is an authentic record of my/our own work carried out during the period *July 2020* to *May 2021* under the supervision of Dr. Somesh Kumar. I also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

Signature:

Name: Pranjal Mankar
Roll No.: 2016IPG-046
Date: 17-05-2021

To the best of my knowledge, the statements made by Mr. Pranjal Mankar (Roll no. 2016IPG-046) in the above thesis certificate is correct.

Signature of Research Supervisor

ABSTRACT

With the evolution of Machine Learning, the Internet of Things (IoT), and Big Data technologies, digital data has increased exponentially. To handle and process such a large bunch of data, high-performance computers are used across various fields. 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 and sharp algorithms. Large financial institutions with big capital spend a lot of money on programmers and data analysts to develop best accuracy trading algorithms that can drive the overall market — the institutional computers fight with each other, which drastically increases markets' overall volatility. A novice trader or investor with less to no experience in financial markets feels it difficult to search for good trades or stocks to invest his/her hard-earned money on a short to long-term time horizon. They rely on expert advice for stock recommendations and sometimes end up making big losses. This paper 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 profitability. The historically traded big dataset of stock prices and investment reports of large financial institutions worldwide are gathered, and various machine learning and decision-making models are employed to perform technical and fundamental analysis across various securities. The output of the trend trading indicator is displayed on charting platforms, which can provide entry-exit levels at which even a novice investor can decide where to invest his/her money. Multi timeframe analysis is deployed to predict short-term, medium-term, and long-term overall trends, thus increasing the output accuracy. The indicator is useful for all kinds of retail traders and investors worldwide who struggle to earn profits from financial markets. Our proposed system was able to achieve profitability of 86.28% annual returns. The entire system, along with the trading orders, is automated so that anyone can earn extra passive income every month from the stock market.

Keywords: big data analytics, stock market, trend trading, multi time-frame analysis, technical analysis, trading strategies, return on investment, automation, profit and loss, passive incomes, future prediction.

Contents

List of Figures

List of Tables

1 Introduction

Many irregular fundamental factors such as government policies, GDP, interest rates, solvency, supply and de-mand, development, exchange rates, politics, and current events affect the daily price fluctuations of various stocks of the stock markets. Hence, it is not easy for a novice trader or investor with less financial knowledge to guess the upcoming trend of the financial markets under any given circumstance. Due to high-frequency buying and selling of stocks by large institutional investors, a bunch of big non-linear investment data [1] is created which is out of the reach of understanding of novice investors with the least financial intelligence, and they end up relying on expert advice or broker's stock picks which are highly risky.

The historical stock price data forms a non-linear pattern with much useful hidden information, which high-frequency mathematical and scientific algorithms can only operate. Large financial institutions use the latest technologies to make quick trading decisions. They can easily manipulate the market, thus creating extreme volatility where retail traders and investors lose a lot of their hard-earned capital. Due to the rise in big data, re-liable financial institutions' monthly investment reports are more complex to understand than ever. However, by retrieving useful information from these complex in-vestment reports, one can identify the overall structure of the upcoming market and will be able to judge at what particular sectors or specific stocks the fund managers are heavily bullish, bearish, or neutral. In 2014, S. Lauren and S. Harlili [2] used time-series analysis of a simple moving average to predict a stock's future trend. We, in our research, have used the simple exponential smoothing concept to enhance the trend predicting capability of our proposed model. Similarly, M. Makrehchi et al. [3], G. Attigeri et al. [4], and many other researchers [5] [6] [7] [8] of the same domain have made the use of sentiment analysis techniques to predict the short-term price movement as per the social media and news data. We retrieved useful data from investment reports to study how large bulk quantities of buying and selling of stocks by big financial institutions can affect stocks' short-term movement. We believed it is more necessary to be able to study the volatility of current prices for deciding whether the current market price of a given stock is suitable to take entry or exit, rather than focusing on predicting the future prices for next t days. Therefore, with the help of machine learning and decision making model, a universal trend trading indicator is designed to deal with the price movement of any given stock and identify its overall future trend. This indicator will analyze all the data collected from reliable stock exchange sources and large institutional funds buying and selling data.

The indicator can perform the following functions–

1. It can predict the entry-level at which even a novice investor with no prior investing knowledge or experience can enter and invest his/her money in any given stock.

2. The indicator also has a feature to predict the stock's overall future trend on a long-term basis over which a person can hold onto his/her in-vestments.

3. It can also decide the exit price at which the existing investor can sell or exit from his/her in-vestment of any given stock.

4. It can do multi-timeframe analysis [10] of any stocks or indices to predict its short, medium- and long-term trend range.

The trend trading indicator is a visual representation of the output of our decision-making model that we have proposed in our research. The three separate prediction models: price action model, technical analysis model, and simple exponential smoothing, are ensembled to form the decision-making model. Each one of these prediction models is briefly discussed in the later sections. The input datasets are individually fed to each of these prediction models, where each model performs its evaluations and provides its separate predictions. These distinct predictions are then combined and ensembled to form a single prediction as an output of the decision-making model in the visual form as a trend trading indicator.

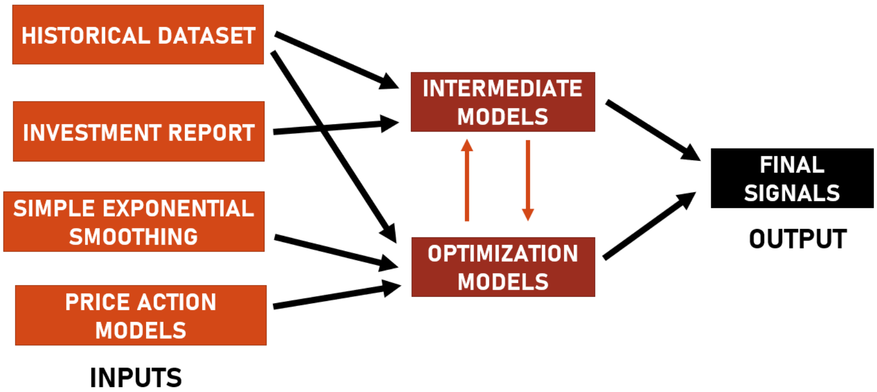

Figure 1: The modules of our system design.

As shown in Fig. 1, the historical stock price dataset and investment reports act as inputs to build intermediate models. The intermediate model consists of study tools used to perform technical analysis of the stock on candlestick charts. Technical analysis is required to get an idea of the stocks' volatility and demand and supply zones. We performed time series analysis into this historical stock price dataset using simple exponential smoothing and various price action models to obtain the optimized models of better accuracy. We will discuss more on our proposed model in the later sections. Our proposed system was able to achieve a profitability of 86.28% annual return on investment. We have taken the combined results of intermediate and optimized models as an output in the form of final signals. The ensembled output of the decision-making model as an indicator is deployed for automated buying and selling of stocks to generate passive income. Tradingview.com API and Streak API are used as a charting platform and automation platform, respectively [10] [11].

2 Review of key related research

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 [14].

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 [15].

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 [16].

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 [3].

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 [2].

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 [5].

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 [4].

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 [17].

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 [18].

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 [1].

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 [9].

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 [19].

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 [20].

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 [21].

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 [22].

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 [6].

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 [23].

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 [7].

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 [8].

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].

3 Objectives

1. For technical analysis, historical price data of stocks or indices are collected from reliable stock exchange sources. For fundamental analysis, large institutional fund buying and selling data are collected from their investment reports.

2. Collected big data is pre-processed, analyzed and classified by our unique decision making models while entry levels, future trend and exit levels of a given stock are

identified as an outcome.

3. Outcome results are backtested with the historical trend of the stock and the indicator's prediction accuracy is enhanced.

4. Indicator's prediction outcome is deployed with the Streak API platform for automated buying and selling of stocks and other securities [2].

4 Tools Required

Programming languages
Python, Pinescript

Tools and libraries
Pandas, Numpy, Keras, SciPy, Matplotlib, Scikit-learn, Pytorch, Streak.io API, Tradingview.com RestAPI

Datasets
NSE, NSDL

5 System architecture and Methodology

This section explains our system's overall architecture and the methodology to obtain the final proposed model. Fig. 2 shows the system architecture where each module is briefly discussed as follows:

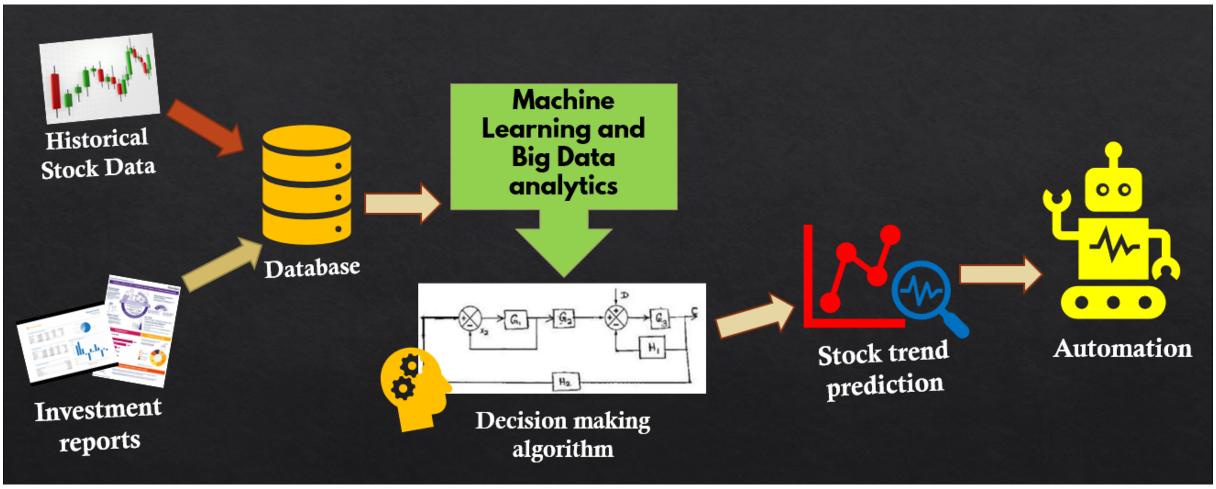

Figure 2: System architecture of our proposed model.

5.1 Dataset Collection

5.1.1 Historical stock price dataset

The historically traded big datasets of all Nifty 50 stocks were collected from National Stock Exchange (NSE) web-site. These datasets mainly consisted of open, low, high, and closing prices of all 50 stocks from the year 2003 – 2019 as shown in Figure. These prices were used to de-termine the non-linear pattern of a particular stock for performing technical and multi-timeframe analysis [3].

Date	OPEN	HIGH	LOW	PREV. CLOSE	VOLUME
21-Sep-13	2,300.00	2,336.00	2,247.35	2,305.70	15519433
18-Sep-13	2,314.25	2,319.45	2,276.55	2,298.75	15264101
17-Sep-13	2,320.00	2,333.70	2,291.85	2,324.55	11919991
16-Sep-13	2,320.00	2,369.35	2,310.55	2,318.85	15669133
15-Sep-13	2,311.95	2,325.75	2,288.15	2,302.55	12543161
14-Sep-13	2,325.00	2,360.00	2,282.00	2,319.75	20335480

Figure 3: Historical stock price of Reliance Industries.

5.1.2 Investment reports

We collected the investment reports and trade-wise equity datasets of FPI/FII from the year 2003 - 2019 from the National Securities Depository Limited (NSDL) web-site. These reports and historical equity datasets, were used for determining where the big institutional inves-tors trade (buy/sell) large quantities of stocks at any giv-en point of time. We used this institutional trading data and the historically traded price patterns for predicting the future trend of stocks [4].

The entire historical price datasets and investment reports were gathered and stored in a local database to retrieve and derive all the useful information to further pass on to the decision-making model.

5.2 Decision-making model

5.2.1 Components of price action model

a. ***Higher-high and lower-low price action*** - The stock markets are open to trade 5 days a week for a limited amount of time in a day. Every day, prices of various securities in stock market moves up or down based on the demand and supply for that particular day. Someday, due to large demand than the previous day, the prices trade higher than the day before. Similarly, when there is a large supply of big institutions, the prices tend to trade lower than the previous day. So, when the current prices are higher than its previous day, it is said to be making a higher-high price action as shown in Fig. 4. b). In other words when the prices today are trading higher than the day before, this is a signal of extreme bullishness and a possible trend for further higher prices. Similarly, when there is a larger supply in the market, the prices today

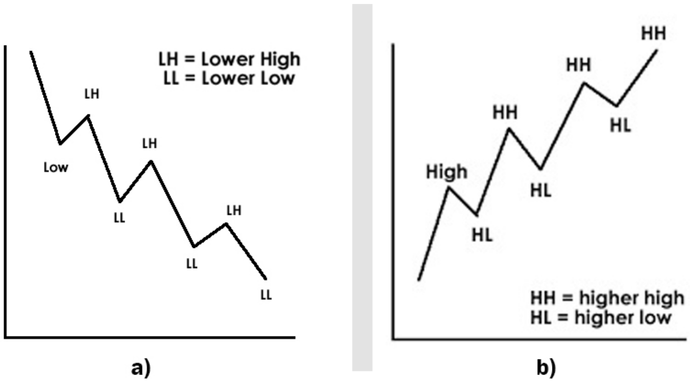

Figure 4: Higher-high and lower-low pattern diagram.

trade lower than the previous day, thus prices make lower-low price action as shown in Fig. 4. a), this suggests that the future trend is shifting to bearishness.

We used this concept in our research to predict the kind of price action the recent closing prices are making. We took the recent 14 days closing price of a particular stock to identify and analyze if it was making a higher-high or lower-low price action. If the closing prices were ascending with time, we set it as a possible uptrend signal. Similarly when the closing prices were descending in value with time, we set it as a possible downtrend signal. Thus, we implemented this price action concept to decide the current trend as an output in our decision-making model.

b. ***Pyramiding entry-exit*** - Pyramiding is a type of trading with multiple entries that are based on higher-high and lower-low price action concept as discussed previously. When the next day's prices trade higher than the previous day, we say that the trend is bullish and it is making higher-high price action. So, in such cases, we implement pyramiding into our trading style, by adding more buying quantities on the next day when the prices are trading higher than the previous day and making higher-high price action. Similarly, we can implement the same concept by adding more selling quantities for the lower-low price action.

   We used this concept in our proposed model to decide the entry-exit points in a particular trend. As discussed previously, there is a possibility of an uptrend in the higher-high price action of recent closing prices. So, in an uptrend, we only find buying opportunities or entering into the stock. Thus, the pyramiding concept helps us deciding at what particular level we should take our entry. We proposed that the third consecutive higher-low level should be our best buying point in higher-high price action. As shown in Fig. 5, Buy order 3 is our best buying point. And in lower-low price action of recent closing prices, the third consecutive lower-high should be our best selling or exiting point from the stock in a downtrend. Thus, we implemented the pyramiding concept to provide the entry-exit signals as an output in our decision-making model.

c. ***Averaging*** - The averaging concept works exactly opposite to the pyramiding con-

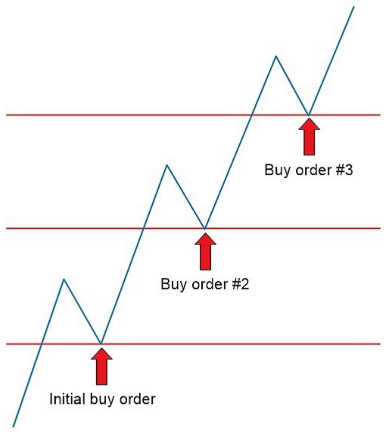

Figure 5: Pyramiding diagram for higher-high price action.

cept where we add more buying quantities if the current day prices are trading lower than the previous day. We do this to average out our entry price, with a conviction that the trend is still intact and the prices of the stocks can rise above our average price in the near future. Similar is the case for short selling or exiting the stock. We add more selling quantities if the current day price is trading higher than the previous day to average out our shorting/selling entry levels, expecting that the downtrend is still intact.

As shown in Fig. 6, we fed the historical stock price dataset as an input to each of the three components of the price action model. Each model performed its evaluation based on its parameters and logic, as discussed previously. Each of the three components provided its separate output signals, which were further merged into a signal assembler. The signal assembler takes inputs from each of the components as UPTREND or DOWNTREND. If all of the three components provided UPTREND signal at a time, the signal assembler also provides the Prediction 1 output as UPTREND. Similarly, if all of the three components provided a DOWNTREND signal at a time, the signal assembler also provides the Prediction 1 out-put as DOWNTREND.

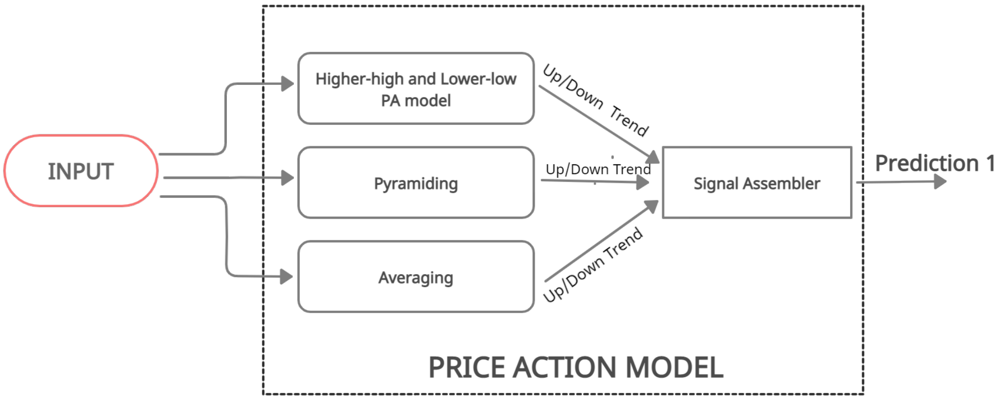

Figure 6: The architecture of the price action model.

5.2.2 Technical analysis

a. ***Average True Range (ATR)*** - It is a measure of the market volatility by taking the moving average of the series of 14 days of the true range which is taken as the maximum of: current candle high less the current candle low; the absolute value of the cur-rent candle high less the previous candle close; and the absolute value of the current candle low less the previous candle close. To obtain better accuracy of the volatility of a given stock, the default value of 14 days period could be adjusted to generate more trading signals.
The calculation of the Average True Range is-

$$True\,Range = max(high - low, abs(high - close[1]), abs(low - close[1]))$$

$$Average\,True\,Range = \frac{(TR_1 + TR_2 + \ldots\ldots. + TR_n)}{n}$$

where TR_i = A particular true range and n is time period

In our proposed model, we used this measure of stock volatility to predict if the stock is still in a trending or a consolidating zone. As per our study of previous re-search in the same domain, we have identified that stock volatility remains low in a trending phase. While in a consolidating zone, due to lack of directional movement, the volatility tends to increase with a decrease in trading liquidity. So, we used this approach to decide whether the stock's current trend is still intact when the volatility is low.

b. ***Commodity Channel Index (CCI)*** - CCI is an oscillator based on the momentum of price ac-tion used to determine the overbought and oversold re-gions of any given security. As it is a momentum-based indicator, it also gives us a clear-cut idea of the relative strength of the trend and direction of the security. CCI calculates the relative difference between the current price and the average price of historical data.

The range of the CCI oscillator can be from 0 to 100 for uptrend while 0 to -100 for the downtrend. If the readings of the oscillator cross above 100, it is a sign of extreme bullish strength, and the price of the security is well above the historical average. While readings of below -100 denote, the current price is way below than historical average price. The calculation of the Commodity Channel Index is-

$$CCI = \frac{(Typical\, price - moving\, average)}{(0.015 * \, mean\, deviation)}$$

where:

$$Typical\, price = \frac{(High + Low + Close)}{3}$$

We used this property of an oscillator in our proposed model to identify the current trend's strength of a given stock. If in an uptrend, as long as the value of current CCI remained above the average measure of the oscilla-tor, which is 50, we believed that the uptrend is still in-tact. Similarly, in a downtrend, if the value of current CCI remained below -50, we believed that the downtrend is still intact. When the CCI value ranges between -50 to +50, we set it as a consolidating zone for the underlying stock.

c. ***Exponential moving average*** - An exponential moving average measures the overall trend of any given security over time. It calculates the average of closing prices while giving more weightage to recent candles to form a current trend which can act as trending support or resistance to the equity price. By giving more weightage to recent closing price of candles, it is assumed that when the equity is above the longer period and shorter period EMAs, the trend is overall bullish. Similarly, when the equity is trading lower than the longer and shorter period timeframes, the trend is assumed to be bearish and the price might further fall.

In our proposed model, we used 50 days as shorter timeframe EMA to decide the shorter term trend and 200 day as longer timeframe EMA for deciding the long-term trend. If the current closing price of the underlying stock was trading above both 50 and 200 EMA, we said the stock to be in an uptrend. Similar would be the case for a possible downtrend if the current closing price were trading below both the EMAs. As shown in Fig. 7, we fed the historical stock price dataset as an input to each of the three components of the technical analysis model. Each model performed its evaluation based on its parameters and logic, as dis-cussed previously. Each of the three components: aver-age true range, commodity channel index, exponential moving average, provided its separate output signals, which were further merged into a signal assembler. The signal assembler takes inputs from each of the components as UPTREND or DOWNTREND. If all of the three components provided UPTREND signal at a time, the signal assembler also provides the Prediction 2 output as UPTREND. Similarly, if all of the three components provided a DOWNTREND signal at a time, the signal assembler also provides the Prediction 2 output as DOWNTREND.

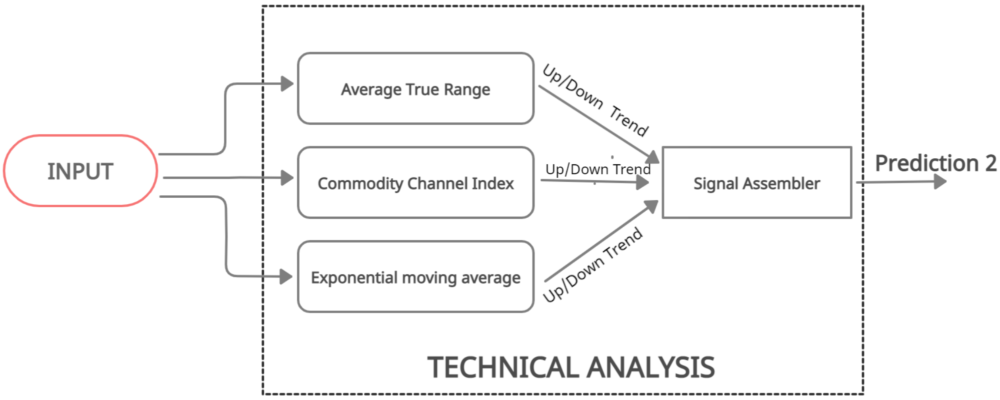

Figure 7: The architecture of the technical analysis model.

5.2.3 Simple exponential smoothing

It is a technique that is used to forecast the future values of a time series. Unlike weighted moving average, this technique considers all the previous historical values while still giving more weightage to the recent forecast-ed values in a linear invariant time series array. As per this technique, if the actual value at time t is higher than or equal to the forecasted value at time t, it is said that the actual value at time t+1 will be greater than the actual value at time t, and the uptrend will remain intact. Similarly, if the actual value at time t is now lower than the forecasted value at time t, it is assumed that the actual value will decline at time t+1, which results in a down-trend. 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.

5.3 Ensembling

As discussed in the previous subsections, each model is designed to make its own predictions based on its individual logic. However, a single model could not fulfill all the decision-making criteria to provide the final desirable predictions. So, to solve this problem, we ensembled the predictions of each model together to generate a more generalized single prediction as an output in the form of a trend trading indicator, as we have shown in the following subsection. While ensembling, equal weightage was given to each of the three previously dis-cussed individual models. The price action model consisted of higher-high and lower-low concepts, pyramiding entry, and averaging positions. The technical analysis model consisted of popular oscillators and moving averages: Average True Range, commodity channel index, and exponential moving averages. Simple exponential smoothing model is used to enhance the trend predicting capability of the decision-making model.

As shown in Fig. 8, the dataset input: historical stock price dataset and investment

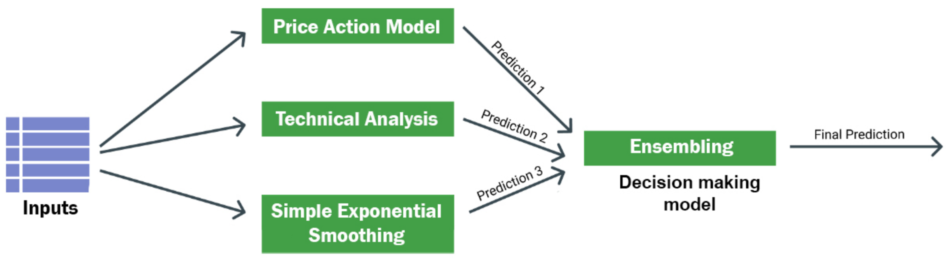

Figure 8: The ensembled architecture of decision-making model.

reports, is fed individually to each classifier model: price action model, technical analysis model, and simple exponential smoothing. Then each of the models performs their evaluations and provides their separate predictions: prediction 1, prediction 2, and prediction 3. These three predictions are then combined and ensembled to form a single prediction as an output of the decision-making model. As a result, when two or all of the three models provide a buy signal, the ensembled model tends to give a buy signal as well. Similar is the case for a sell signal. For trend prediction, when all of the three models provide an uptrend signal, the ensembled model tends to give an uptrend signal as well. Similar is the case for a downtrend signal.

5.4 Visual representation of trend trading indicator

Fig. 9 shows the visual representation of our trend trading indicator, which provides entry-exit points in the form of buy and sell signals. The buy signal at point 1 is represented with a blue triangular arrow mark against a particular candle, allowing an investor to enter/buy the stock. As long as the blue line follows the current candle price, the trend is said to be overall bullish, and the stock is assumed to rally up even further. At point 2, we get a red signal asking the investor to sell/exit the stock as the uptrend is said to have vanished for a while. This sell signal also indicates the confirmation of a downtrend or even an upcoming possible stock market crash. As long the red line follows the current candle, the trend is assumed to be bearish, and the investor should stay away from that stock until at point 3, the trend again shifts to-wards upwards, and the investor can re-enter at the current price of the candle.

Figure 9: Visual representation of trend trading indicator plotted on Tradingview.com charting platform.

6 Experimental Results

This section gives our results based on implementing our proposed model on popular stocks of the Nifty 50 index. As discussed in the previous section, we designed a trend trading indicator as an output of the decision-making model, which provides entry-exit levels, and can also decide the trend of any given stock based on its historical behavior. So in this section, we will be showing the payoff figures, profitability factors, and how we used the indicator signals to automate our trading activities.

6.1 Payoff figures

We strictly followed the entry-exit signals of our trend trading indicator to buy and sell stocks on a daily timeframe of candlestick charts. Whenever the indicator gave us a buy signal, we bought 1500 quantities of the stocks irrespective of the current market price. We sold/exited the stock whenever the indicator gave us the sell signal. We kept repeating this method on the historical price dataset from the year 2003 to 2019 and measured the payoff of the return on investment.
Calculation of the payoff:

1. Buy entry price = Rs. 1340

2. Sell exit price = Rs. 1463

3. Payoff = (sell price – buy price) * 1500 quantities = (1463 – 1340) * 1500 = Rs. 34,500/-

Figure shows the payoff graph of return on investment (ROI) of various popular stocks of the Nifty 50 index over the years. The blue part indicates the equity buildup or the increase in profit from buying and hold-ing the stocks over the years. In contrast, the upper red part is the drawdown in the existing profits from time to time, whenever the current exit level was lower than the recent entry-level, resulting in a loss in equity buildup.

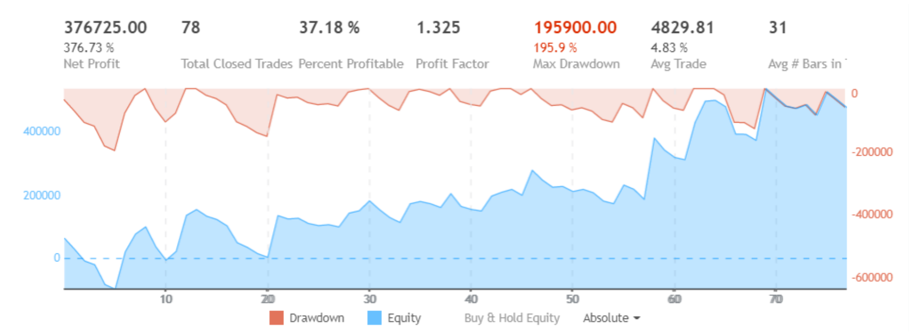

Figure 10: Payoff graph of Asianpaints stock buy-side return from the year 2003-2019.

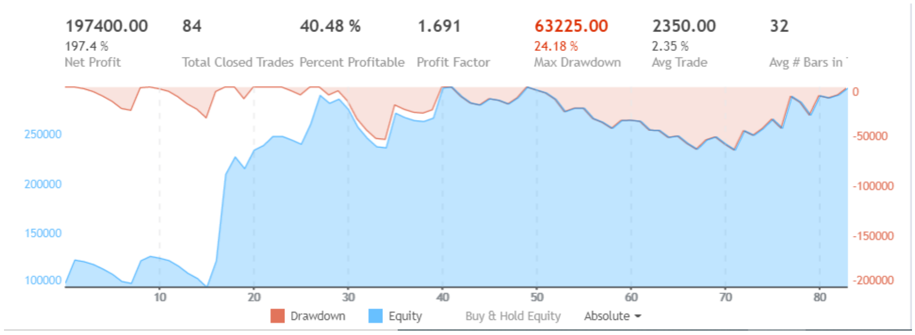

Figure 11: Payoff graph of Icicibank stock buy-side return from the year 2003-2019.

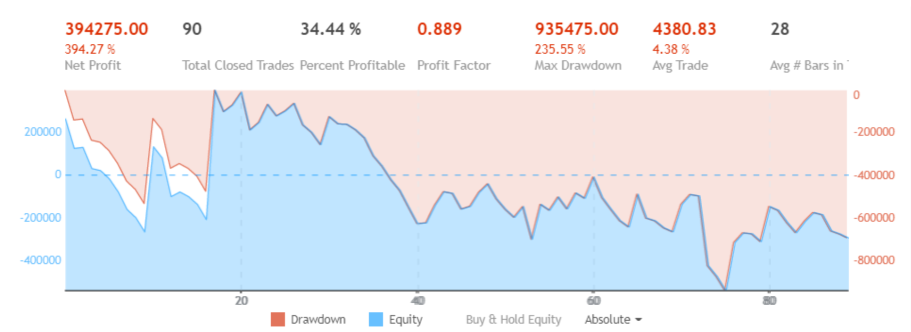

Figure 12: Payoff graph of Maruti stock buy-side return from the year 2003-2019.

We observed from the payoff graphs that some stocks that were quite trending and performing well historically generated a better equity buildup than those that were not much trending technically. As observed from figures 10, 11, 13, and 15 that the payoff of equity buildup appreciated over time as compared to figures 12 and 14 where the range-bound stocks had generated unstable payoff returns. The instability in the payoff returns was due to an improper one-sided trend in the stock, or it remained range-bound over a more extended period.

We performed the same experiment to evaluate the pay-off of profitability percentage

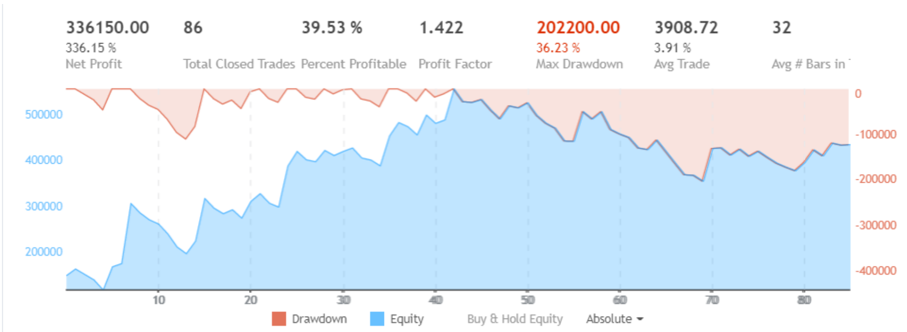

Figure 13: Payoff graph of Reliance stock buy-side return from the year 2003-2019.

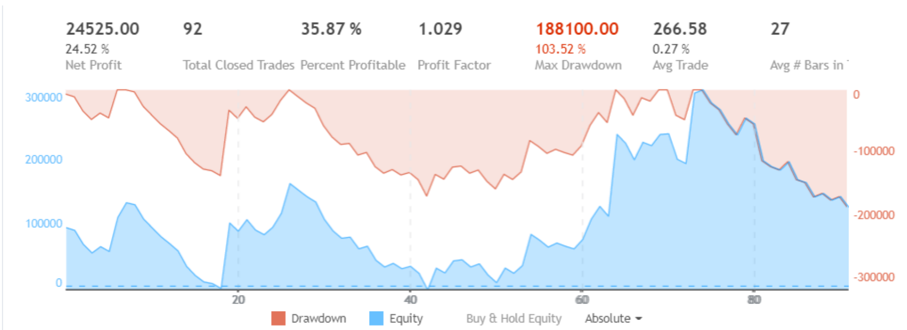

Figure 14: Payoff graph of Titan stock buy-side return from the year 2003-2019.

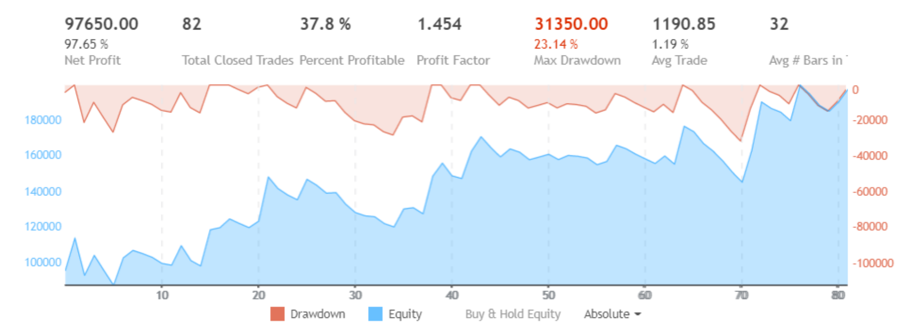

Figure 15: Payoff graph of Wipro stock buy-side return from the year 2003-2019.

and annualized return on more Nifty 50 index stocks. The following Table 1 shows the result of our evaluation.
We were able to make the following observations from Table 1:

1. The highly trending stocks generated the most profitable annualized return on investment of 75-80%. The indicator provided smooth entry-exit signals out of which most were profitable with a profitability factor of as high as 5.357.

2. The range-bound stocks generated an annualized return on investment in the range

of 40-70%. While the indicator struggled to provide precise entry-exit levels, one of the stocks Nestle even provided a negative payoff return and max drawdown due to its non-trending nature.

Table 1: 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)

6.2 Role of Simple exponential smoothing on enhancing profitability

As we discussed in the previous section, simple exponential smoothing technique was used in our proposed model to optimize the trend predicting capability of the indicator. The intermediate model consisted of a combination of price action components and technical analysis to provide the entry-exit signals and predict the current trend of the underlying stock. But by introducing the time series analysis in the form of simple exponential smoothing and its combination with the intermediate model enhanced the profitability of the annualized re-turns up to 1.8 times. Here is a performance summary of the intermediate model with the optimized model after introducing simple exponential smoothing.

We performed the same backtesting of signals provided by our trend trading indicator to obtain the annualized returns on the same underlying stocks. The blue bars in the Fig. 16 is the result of annualized returns of intermediate models. In comparison, the orange bars are the annualized returns of optimized model which are upto 1.8 times better.

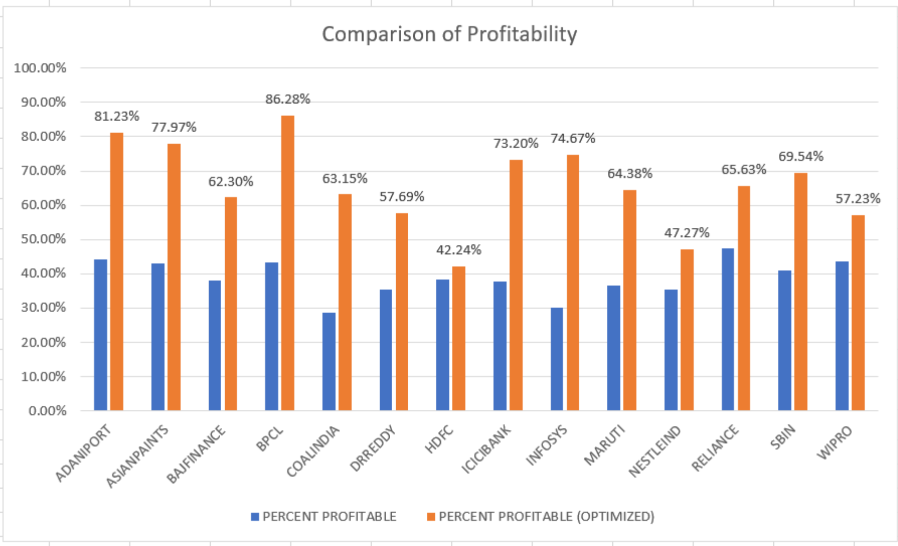

Figure 16: Profitability comparison of the intermediate model with the optimized model.

7 Novelty

1. Algorithms and the technical Indicator can provide entry-level, future trends, and exit levels of all stocks, indexes, commodities, foreign exchange securities. So, a novice investor who doesn't even have any idea of the stock market can use this indicator to decide where he/she can invest his/her money.

2. The Indicator can do multi-time-frame analysis to decide the short-term, medium-term, and long-term future trends of any stocks or any securities all over the world.

3. 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 this indicator.

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

5. The entire output of algorithms and indicators will be deployed to the Streak platform for automated buying and selling of stock or indices for generating passive income.

8 Timeline and Gantt chart

To achieve the objectives of this project, I have divided the project work into various tasks. The following shows the timeline and gantt chart of all the activities.

Table 2: Timeline and gantt chart of all the activities

	Start Date	End Date	Timeline	Status
Tasks	**Aug 22, 2020**	**Jun 25, 2021**		
Topic discussion with mentor	Sep 23, 2020	Sep 25, 2020		Complete
Subject exploration of the topic	Aug 22, 2020	Sep 10, 2020		Complete
Project planning	Sep 15, 2020	Sep 30, 2020		Complete
Gathering of historical big dataset	Oct 2, 2020	Oct 12, 2020		Complete
Programming of algorithms	Oct 9, 2020	Oct 25, 2020		Complete
Training the algorithms with training dataset	Oct 26, 2020	Nov 5, 2020		Complete
Testing of algorithm on testing dataset	Nov 10, 2020	Nov 15, 2020		Complete
Optimize the algorithm for best accuracy	Nov 25, 2020	Dec 15, 2020		Complete
Displaying of outcome and results on stock charts	Dec 15, 2020	Dec 20, 2020		Complete
Designing of mathematical Trend Trading Indicator	Dec 12, 2020	Dec 25, 2020		Complete
Intermediate results	Jan 1, 2021	Jan 5, 2021		Complete
Examining of outcome for multi time-frame analysis	Jan 7, 2021	Jan 28, 2021		Complete
Backtesting and validation of outputs	Feb 1, 2021	Feb 25, 2021		Complete
Implementation of Automated trading	Feb 28, 2021	Mar 17, 2021		Complete
Passive income returns evaluation	Mar 20, 2021	Apr 4, 2021		Complete
Conclusion	Apr 15, 2021	Apr 20, 2021		Complete
Final Report Submission to Supervisor	May 1, 2021	May 15, 2021		Complete
Project presentation	May 1, 2021	Jun 25, 2021		Active
		Burndown		

9 Current Work Progress

1. ***Tasks completed***:

   - Various subject and topic-related literature work from popular journals and conference papers were thoroughly studied and all useful information regarding the project was gathered as mentioned in the above Literature review section.
   - In project planning, Figure 2. system was designed as mentioned above in the methodology section.
   - The historically traded big datasets of all Nifty 50 stocks are collected from National Stock Exchange (NSE) website [3].
   - The investment reports and trade-wise equity datasets of FPI/FII from the year 2003 - 2019 are collected from the National Securities Depository Limited (NSDL) website [4].
   - Various machine learning and big data analytics algorithms for data pre-processing, classifying, decision making, and back-testing of desirable outputs are programmed using my own mathematical models such as higher high, pyramiding, and averaging concept. Unique trend trading logic is implemented to design the trend trading algorithm and indicator.
   - 30% of the total pre-processed and categorized datasets are used for the training of decision-making algorithms.
   - Trained decision-making algorithms is used to test remaining 70% of the total datasets.
   - For best possible accuracy of outcome, decision-making algorithm was further optimized using time series analysis, to obtain better results.
   - The optimized results were displayed on stocks charts using Tradingview.com charting platform in the form of trend trading indicator.
   - Based on the signals provided by indicator, the intermediate results and payoff of annualized returns were calculated.
   - To improve the efficiency of intermediate results, multi time-frame analysis was performed to develop trend predicting capability as per suitable period of time horizon.
   - The expected output was thoroughly back-tested and validated on historical prices of dataset and performance summary table was formed.
   - The entire system, was deployed to Streak platform for automated buying and selling of stocks with highest profitability.
   - Last two months data was taken as an input to evaluate passive income return of automated system.
   - Finally the project was concluded and discussed with the supervisor.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] streak.readme.io/

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

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

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[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.