

## Document Information

Analyzed document	Investmentled.pdf (D148888115)
Submitted	11/7/2022 5:37:00 AM
Submitted by	Dr.V.Selladurai
Submitter email	selladurai.v@gmail.com
Similarity	0%
Analysis address	selladurai.v.annauniv@analysis.arkund.com

## Sources included in the report

**SA****University of Delhi, New Delhi / 18020590046 (Utkarsh).pdf**

Document 18020590046 (Utkarsh).pdf (D105660647)

Submitted by: ramanujancollege2010@gmail.com

Receiver: ramanujancollege2010.du@analysis.arkund.com

 1

## Entire Document

Analysis on Stock Market Tilak Vijayaraghavan S 1 , Sarathram M 2 , Radhamani V 3 1 Student (Department of Computing, Coimbatore Institute of Technology) 2, Student(Department of Computing, Coimbatore Institution of Technology) 3 Faculty (Department of Computing, Coimbatore Institute of Technology) \*Corresponding author's email: stvraghavan@gmail.com Article Type: Research Article Complete Detail of Each Author First Author's Full Name: Tilak Vijayaraghavan Highest Qualification: M. Sc. Decision and Computing Sciences Department: Decision and Computing Sciences Post/Rank (If a student, provide course name and course year): M. Sc. Decision and Computing Sciences 5 th year Affiliation (College/University/Institute) with postal address: Coimbatore Institute of Technology, Avinashi Road, Civil Aerodrome, Peelamedu, Coimbatore 641014 email id: stvraghavan@gmail.com ORCID: N/A Second Author's Full Name: Sarathram Highest Qualification: M. Sc. Decision and Computing Sciences Department: Decision and Computing Sciences Post/Rank (If a student, provide course name and course year): M. Sc. Decision and Computing Sciences 5 th year Affiliation (College/University/Institute) with postal address: Coimbatore Institute of Technology, Avinashi Road, Civil Aerodrome, Peelamedu, Coimbatore 641014 email id: sarathrammsr07@gmail.com ORCID: N/A Third Author's Full Name: Radhamani V Highest Qualification: M.Sc. Computer Science Department: Decision and Computing Sciences Post/Rank (If a student, provide course name and course year): Assistant Professor Affiliation (College/University/Institute) with postal address: Coimbatore Institute of Technology, Avinashi Road, Civil Aerodrome, Peelamedu, Coimbatore 641014 email id: vradhamani@cit.edu.in ORCID: N/A

**ABSTRACT** With never before seen dynamics in the world affecting the value of money across the globe, the option to save assets in the form of money is becoming more and more unstable. Investment becomes mandatory, investment in stocks have proven to be a viable solution to avoid inflation. Most of the current applications in the market who also provide the same variety or more varieties, Investmenttled is open-source and is free unlike most of the already existing applications. Most of the existing applications provide a freemium and premium version, with the premium version being mostly more feature rich and there is also the issue of service availability depending on regional compatibility. Investmenttled is designed to be a locally hosted application mostly dependent on certain data files to lookup information online using browsers. Investmenttled, which has been built so far, provides the end user the ability to track a set of stocks they select using a few tracking metrics, candlestick charts and daily returns charts. The proposed solution also provides information on the spread of the incoming data using a box plot to identify price fluctuations in the market seen across time. Investmenttled uses two main algorithms to provide its services. The first one being the portfolio optimiser to compute the discrete allocation of stocks. The second algorithm is the sentiment analyser which uses VADER to calculate the sentiment expressed by users in social media and other online sources. Investmenttled was built entirely with python. Investmenttled uses a number of libraries like pandas, numpy, pyportfoliopt, plotly, streamlit and much more to achieve its working efficiency. **Keywords:** investment, portfolio, optimization, stocks, visualization, python, sentiment analysis

**1. Introduction** A stock market is a place where buyers and sellers exchange stocks / shares which represent ownership of the company's stock being exchanged. In India, there are two main exchanges namely the Bombay Stock Exchange and the National Stock Exchange. Both of them are powered by state of the art technologies and provide screen trading. The exchanges in India offer trading via a demat account where everything is dematerialised to enable trading. Investment in the market usually is a bit of a complex process with a broker involved. The trades happening in the market are caused by a multitude of factors many of which can't be completely understood. By properly understanding the market movements, making profit from the market's movement becomes possible. By intelligently investing in stock the volume of expected return is usually higher than that of a regular bank deposit [8] . Investmenttled is a modular application which aims to provide its users with various stock information on demand. Investmenttled uses an open-source API to track stock information and process them to create meaningful information for the end-user. By considering various angles to interpret stocks Investmenttled aims to create new insights for users to understand the movement of stocks in the market. The tracker API tracks the movement of stocks using various techniques to generate possible buy / sell signal alerts. All the above paired with a sentiment analysis with data from twitter for the duration of past one or two days, enables the predictor to have a better understanding about the stock with regard to the stock's impact on social media. Built with open-source software, Investmenttled is stable and scalable due to modularity. All this packed together in a locally deployable application can be used on any system provided the necessary libraries are included.

2. Problem Definition There are various current applications which aim to provide similar services like stock tracking, portfolio optimisation, back testing, stock tracking indicators, brokerage deduction and many more. But, most of the features for a product currently available are limited either by having a paid version or by the nature of the application where the feature is absent. Most of the applications that provide this service are online web applications and they mostly require premium service. Also, the main drawback of most of the applications is that they are region specific and hence will not work outside their regions or will not work for foreign markets. This makes compatibility (compatibility in terms of availability of desirable service) a very big issue. The aim of the proposed solution is to create a decision tool (Investmentled) which when deployed will be able to assist an individual to know about the stock market and formulate an investment strategy based on a set of technical analysis and market trackers. Investmentled is a multi-module application which aims to provide a variety of services for people who wish to gain knowledge to invest in the stock market. This application aims to provide various details like stock trends, sentiment analysis, portfolio optimisation, stock movement visualization etc. Paired with a sentiment analysis with data from twitter for the duration of the past one or two days, enables the predictor to have a better understanding about the stock with regard to the stock's impact on social media and which would be a beneficial one too. Investmentled aims to provide its users with the following tools to gain insights, the tools provided include Knowledge about stocks, equities, and other technical terms, A tracker with various techniques, A portfolio analyser and optimiser. The proposed solution will provide newcomers with a wiki page to gain knowledge about the terminologies used in the market. There also exists aiding material along the various charts to interpret the said chart. Experienced users easily use the candlestick charts to their advantage and can benefit themselves from the sentiment analysis rather than reading through many articles.

3. Literature Survey In [1] Robert D. Edwards, John Magee, et.al proposed Usage of Technical analysis on the stock data have proven to yield substantial results to consider them for usage. Prevalent methods like candlestick charts, moving averages, convergence and divergence etc. prove useful. In [10] Matheus José Silva de Souza, et.al. A recent study by a group of researchers using a portfolio of stocks traded in the stock exchange of the BRICS countries done with the help of fundamental analysis and technical analysis proved to generate returns higher than the expected values. In [2] Michael J. Best proposed valuation of a portfolio using a Sharpe ratio on a frontier guarantees the most mathematically optimized solution. In [11] B. Kalayci, Okkes Ertenlice, et.al. anticipated a solution by evaluating a portfolio mathematically based on economic models and using a Ceteris Paribus way of evaluation, performance of an individual stock can also be recognised. A variance optimized portfolio often referred to as a VOP could either be a single variance or multi-variance optimization technique to achieve the best possible portfolio according to modern portfolio theory. In [3] Muhammad Iqbal designed candlestick proves to be an easier way to understand market trend rather than a normal candlestick. By capturing the trend in the data and smoothening the noise, the Heikin-Ashi candlestick chart gives a good representation of the trend present in the data.

In [4][5] Bollinger Bands use a simple yet powerful technique to provide a means to find a significant point during a stock's movement. In [6] Peter Christofferson, Denis Pelletier, used Monte-Carlo simulation to deduce results prove better than other conventional retail models. By using a Monte-Carlo simulation which uses randomness, simulation of risk and uncertainty becomes possible and estimation of stock rates becomes a possibility. In [7] ThienHai, Nguyen, et.al. Planned Sentiment Analysis on stocks prove useful in providing insights to track their movement and confirms the effect of social media on markets. In [8,9] Chandra Prasanna, Jue Liu, Zhuocheng Lu, et.al. A recent analysis on stock news suggests the use of correlation as a factor to be considered while performing sentiment analysis. Correlation here refers to the connection that exists between the stock being monitored and a correspondingly correlated stock. This correlation can be traced using the Deep learning model by processing various relationships that exist between stocks. It is also proven that stock investors' psychology influences the market, and hence by understanding their mood swings, predicting whether the market will rise or fall becomes possible. From reviews from various articles and journals mentioned above and by considering various angles to interpret stocks this proposed work aims to create new insights for users to understand the movement of stocks in the market. Using a mathematical optimisation method like the efficient frontier is guaranteed to return the most optimal portfolio mathematically. Likewise, performing sentiment analysis on social media platforms yields a chunk of valuable information about the mindset of the individuals in the market as it is proven that man believes information found on the internet regardless of their origin and authenticity. Since the above techniques are implemented practically and provide an assurance for better efficiency, the proposed work approaches include help of candle stick, efficient frontier, sentiment analysis for better research and development.

4. Proposed Work Investmentled aims to overcome most of these drawbacks, mostly by being modular and open source. Since the application was designed to be modular, the datasource for the application can be any relevant form of data. Since the application computes the required indices ad hoc, one can design it to compute only necessary indices. Since the application is open source, incorporating newer modules for the proposed solution also shouldn't be a problem and anyone can use it for free of cost to manage their personal portfolios. This is a multi-module application which aims to provide a variety of services for people who wish to gain knowledge to invest in the stock market without having the necessity to spend money. The proposed solution will provide newcomers with a wiki page to gain knowledge about the terminologies used in the market. There also exists aiding material along the various charts to interpret the said chart. Experienced users easily use the candlestick charts to their advantage and can benefit themselves from the sentiment analysis rather than reading through many articles.

5. Dataset Since stock market movements are influenced by information shared through news, social media, articles and discussion, the prices of the stock dynamic in nature and change accordingly within a certain frequency of time interval mostly End of the Day (EOD) data. So here there is a need of making updates in the data thereby the proposed work uses live data updated from nseindia.com (National Stock Exchange). In case of server unavailability, data is sometimes obtained from Yfinance using another API. Figure 5.1 is a snapshot of the data collected till the 3rd of November The above snapshot was obtained from the nseindia.com website by using API calls done via the nsepy library. Figure 5.2 is a snapshot of the data collected till the 4th of November The above snapshot was obtained from the nseindia.com website by using API calls done via the nsepy library, it can be seen that the EoD prices update on all business days but not on national holidays and weekends. 6. Tools and Techniques Optimisation can be defined in many ways, one among them being, optimisation is the process of maximizing or minimizing a certain value while being subject to a set of mathematical constraints. The same can be rephrased for optimizing a portfolio, by considering the expected return and the percentage of risk willing to be taken, an area of possible solutions can be created and can be mathematically optimized to give the best possible solution. The boundary of the solution space is called the frontier and the solution often lies on the frontier as the frontier is the line that separates the set of possible and not possible solutions. 6.1. Efficient Frontier An efficient frontier is a solution line which spans at the edge of the solution space. This line when subject to further constraints will yield the desired result of optimisation [11][2]. Based on the additional constraints placed on the frontier, the portfolio represented by the frontier takes different names like, ● Minimum variance portfolio ● Return-efficient portfolio ● Parameter-efficient portfolio

Figure 6.1 An illustration of an efficient minimum variance portfolio Figure 6.2 An illustration of an inefficient portfolio The arbitrary constants  $\alpha_0$  and  $\beta_0$  are the expected return and variance of the minimum variance portfolio.

By moving along the frontier, the point of maximum efficiency can be found, at the point where the frontier meets the imposed constraints which is usually determined by the Sharpe Ratio which can be defined as the ratio of every unit gain against the volatility the stock faces (risk).

## 6.2. Candle Sticks

Visual representation is often preferred over raw numbers, candlestick charts prove to solve the problem by visualizing the market OHLC data (Open, High, Low, Close of a stock's price). There are two types of candlestick charts popularly used, 1. Normal Candlestick chart 2. Heikin-Ashi Candlestick chart [3] The normal candlestick charts provide the OHLC information about a stock and hence move along with the stock. This can cause the chart to capture all the stocks movement and noise. On the other hand, Heikin- Ashi charts which use a different formula tend to highlight stock movement trends along with their price movements. Both the visuals have their pros and cons. The normal candlestick chart highlights the actual movement along with the noise while Heikin-Ashi candlesticks smoothen the chart but does not give the actual values of the stocks in question (because they smoothen the data, a bit of data loss/transformation occurs).

Figure 6.3 Classical representation of a candlestick  
Figure 6.4 A candlestick chart for the Apple stock  
Figure 6.5 A Heikin-Ashi candlestick for Apple Stocks

By comparing Figure 3.4 and 3.5, a difference can easily be seen between a normal candlestick and a Heikin- Ashi candlestick. Figure 3.4 contains the actual stock prices and hence has more noise but the Heikin-Ashi candlestick has less noise but does not reflect the actual stock prices as the data used for the chart underwent smoothing.

## 6.3. Technical indicators

"Make hay while the Sun shines" is a classical proverb which says about the importance of action done at the right time, similarly buying and selling apt is important. The knowledge of when to buy or sell can be achieved by performing technical analysis using different techniques. Few of them being,

1. Relative Strength Index
2. Moving averages
- a. Simple moving averages
- b. Moving average convergence / divergence
3. Bollinger bands [4][5]

### 6.3.1. Relative Strength Index

Relative Strength Index (RSI) is an index which describes the amount of oscillation a stock faces. Relative strength of a stock is calculated as where Relative Strength (RS) is calculated by Where U and D are metrics calculated for change in closing price between two days. The function SMMA (), is a smoother or modified moving average taken for 'n' periods (days). If the market is rising, D is zero while U is the difference between the closing price of the stocks between two consecutive days, this logic is converse for a falling market where U is zero while D is the difference in closing price of consecutive days. RSI gives the current moment of the stock; this indicator is like a percent indicator. RSI values between 70 – 30 is considered as normal for a stock's momentum while outside of the said bounds could mean that the stock is overbought or oversold.

### 6.3.2. Moving Averages

Moving averages as the name suggests are averages of the stock prices taken at an interval of 'n' days where 'n' is the number of consecutive days considered for taking the average. Buy-sell signals can be generated using the help of moving averages, consider two sets of moving averages, one long-term and one short-term moving average. Whenever the short term moving average is greater than the long term moving average and vice-versa, a person could generate a signal to buy or sell respectively. This is a simple trading trick achieved by using moving averages. A more sophisticated trick is by using moving average convergence and divergence technique. This technique uses the idea of stock divergence to generate signals. Divergence is a term which is used to indicate that the "indicator" used and the stocks are moving in the opposite direction and a trend reversal is imminent.

### 6.3.3. Bollinger Bands

Bollinger bands are a simple and powerful technique to generate signals. By considering the stock price variation and setting a limit like 'x' standard deviations, one can create a boundary around the stock which when crossed signals a buy / sell. Figure 6.6 A Bollinger Band chart

## 6.4. Methodologies

### 6.4.1. Sentimental Analysis

The Internet is the current means through which information is traversing the world and social media has become the most influential network across the globe. By performing sentiment analysis on social media and news articles, users can capture an idea of the stock movement for a short term. The nltk library used VADER (Valence Aware Dictionary and Sentiment), a lexicon / glossary as well as a

70%

**MATCHING BLOCK 1/1**

SA

18020590046 (Utkarsh).pdf (D105660647)

rule-based sentiment analysis tool that is specially tuned to sentiments expressed on social media.

The API like snsrape and GoogleNews are used to gather information from the internet. Once all the raw data has been gathered, users can proceed to summarize the aggregate information to make meaningful analysis and an interpretation. [7] Figure 6.7 A chunk of raw data collected from the internet using Google News API. This info is useless by itself but, by generating a word cloud, it could make some useful interpretation. A word cloud is a representation which is used to represent words in the metadata on a website or showcases the occurrence of words in a given phase or paragraph.

Figure 6.8 a word cloud generated from the chunk of raw data collected This word cloud highlights certain phrases like 'price, petrol, target, stock' etc. Among other words. But, still the analysis can be made better by aggregating the emotion of the words / phrases by using nltk's VADER [12] using which one can gather the percentage of positive, negative and neutral emotions expressed in the internet. VADER is a short form for Valance Aware Dictionary and sEntiment Reasoner, which is a lexicon in nltk which uses rule based sentiment analysis to compute sentiment score.

6.5. Modules Figure 6.9 depicts the model flow diagram of the work. In this work there are three modules thus far implementing two key features of the planned work. Stock Tracker The first module in the work is the stock tracker module which uses various charts and indicators to track the movement of stocks in the market. This module implements a lot of technical indicators to give insights to the users. This module mainly shows outputs like EoD prices, average stock returns, cumulative stock returns, candlestick charts etc. Optimizer module The optimizer module takes the same set of stocks given as input in the first module and creates an optimal stock allocation list. This module also gives the maximum Sharpe ratio for the portfolio and an expected return of the portfolio combination. There is also a matrix representation of the relation between stocks. Sentiment Analyzer Module

The sentiment analyzer module is a separate module which looks up the internet at various sources like the news and twitter to understand the emotion / sentiment expressed by different people for a company and gives the summary in the form of a pie chart and a word cloud. This module helps us to interpret the variation in stock prices caused due to the emotion of the user who plays in the market. 7. Results and Discussions The user first selects the available stocks needed from the list obtained and has two different options to move further. First is the user can visualize the behavior of the selected stocks using Boxplot, candlestick chart, average returns. The other one is Stock Optimization which is used to optimize the portfolio by discrete allocation of stocks based on stock relations.

The Stock Tracker and Optimizer module of the work implements the task of collecting and processing information related to how a given stock performs in the market. By tracking the price movements, percentage of gains across time and calculation of returns generated by the stock per rupee invested. These indicators can be taken as key factors that reveal a stock's movement in a portfolio. By mathematically optimizing a basket of stocks, an optimal volume of investment can be visualized by this module. Fig 7.1 End of Day prices of TCS, BergePaint and ACC Figure 7.1 projects the closing price of stocks versus date over a certain time period of years. Consider the above image where the closing prices of TCS, BergePaint, ACC have been visualized and it shows that the price fluctuates in normalized manner for BergePiant where as compared to TCS and ACC the price falls at certain point of time and then gradually increases in a positive manner over some period of years. So it would be better to opt TCS and then ACC for investment purposes.

Fig 7.2 A sample portfolio allocation between Eichermotor, Mahindra and Mahindra and Coal India Portfolio Optimisation done using the pyportfoliopt module yields a result very much mathematically. The module prioritizes a stocks return over the nature of choice and it mainly uses the Sharpe ratio to calculate the portfolio allocation. Fig 7.2.1 Tabulation for possible Quantity of Stocks Figure 7.2 and 7.2.1 depicts a sample portfolio allocation and Tabulation for possible Quantity of Stocks it can hold between the companies selected(Eichermotor, Mahindra & Mahindra and Coal India) respectively. For a given amount of Rs. 10000 the portfolio can hold stocks on the count of one EicherMotor, three M&M, eight Coal India which corresponds to occupancy or allocation of 50%, 33.7% and 16.9% of mentioned stocks respectively.

The Sentiment Analyser module which pools in data from twitter and Google News to make sense of how a stock is being treated in the market. This is of necessity because social media has proven to influence many walks of life. By using the VADER lexicon dictionary, one can make out the sentiment of the words gathered from the internet which has been included in section 6.4.1 Fig 7.3 A Sentiment Pie Chart Figure 7.3 represents a Sentiment Pie Chart for Tata Consultancy Services which was obtained from the sentiment of the words gathered from the internet such as google news. The image assures that the talk about the stocks of TCS are flowing in a positive manner so there would be a positive change in price of stocks which helps to reduce the risk and to earn the profits and also ensures that the client/user can buy more quantity of stocks from TCS.

Fig 7.4 A Gist of the News Articles collected from Google News Figure 7.4 shows the Gist of the News Articles collected from Google News. The work yielded expected results based on the mathematics involved. The tools used in the solution confirms that there is a definitive relationship between risk and return but not to the extent to say that all risky stocks pay-off. Fig 7.5 Stock with High Risk and Relatively High Return

Fig 7.6 is similar to Fig 7.5 Figure 7.5 reveals the Volatility of stocks with high risk and relatively high return by considering date versus daily simple returns. Volatility is the rate at which a stock price increases or decreases over a particular period of time . An investor can estimate the fluctuations that may happen in the future with higher stock price volatility. Figure 7.6 is a similar as figure 7.5 but this showcases more than 1 stock (depends on the number of stocks selected) Fig 7.7 Average Daily return (in %) Figure 7.7 showcases the average daily return(in %) on the selected stocks(in this case M&M and HDFC). Return on a stock is used to measure the day to day performance of stocks, it is the price of stocks at today's closure compared to the price of the same stock at yesterday's closure. The key advantage of ARR is that it is easy to understand. When comparing investments, the higher the ARR, the more attractive the investment. It means investors get yields that exceed what they expect from other investments with similar risks. Fig 7.8 Stock price change in observed time for a rupee's worth of investment Figure 7.8 depicts the average daily return of a stock based on Date factor which is the average price increase seen in the stock in the selected period of time. It can be considered as the gain in the price of stock for a rupee's investment. From the above image it is clear that M&M has a high growth of investment compared to HDFC at that particular period of time. Fig 7.9 Volatility of stocks M&M and HDFC in Box Plot

The Figure 7.9 showcases the maximum and minimum volatility seen in the selected stock visualized in the form of a box pot which reveals the distribution on price volatility among the selected stocks, here in this case regarding Mahindra & Mahindra and HDFC. There were also some relationships between the sentiment of the user and stock prices and the sentiment of the users involved in the market. A positive trend among the users shows signs of a bullish trend and vice versa. Figure 7.10 Word Cloud generated for Mahindra and Mahindra Figure 7.10 Depicts the word cloud generated for stock selected (Mahindra and Mahindra). Sentiment analysis aims to find the sentiment of users online, but in order to know more about the actuality of things, a word cloud visual comes in handy by displaying the words people mostly use to describe the things they search. But, still the analysis can be made better by aggregating the emotion of the words/phrases by using nltk's VADER [12] using which we can gather the percentage of positive, negative and neutral emotions expressed in the internet.

8. Conclusion Most of the features for a product currently available are limited either by having a paid version or by the nature of the application where the feature is absent. Most of the applications that provide this service are online web applications and they mostly require premium service. This makes compatibility (compatibility in terms of availability of desirable service)a very big issue. Investmentled also has aiding material along the various charts to interpret the said chart. Experienced users easily use the candlestick charts to their advantage and can benefit themselves from the sentiment analysis rather than reading through many articles. Investmentled aims to overcome most of the drawbacks, mostly by being modular and open source. Since the application was designed to be modular, the datasource for the application can be any relevant form of data. Since the application computes the required indices ad hoc, one can design it to compute only necessary indices. As the application is open source, incorporating newer modules for the proposed solution also shouldn't be a problem and anyone can use it for free of cost to manage their personal portfolios. The work also confirms the volatility of the market and its dependency on the mindset of investors, intrinsic and extrinsic factors like the nature of the company, economic conditions. 9. Limitations The dependency on the internet and python have been the most dragging aspect of the work done. The solution also provided / reconfirmed that stock market predictions using conventional and newer machine learning models still misses features that are necessary for accurate predictions. (An ARIMA model and an LSTM for predictions were tried but failed to yield substantial results as neither had an accuracy more than 80% accuracy). The portfolio optimisation module mainly prioritized return over choice which could be a draw-back should the user try to create their own portfolio with a notion in mind. The sheer volume of knowledge and information required to process the entire data proved monumental when compared to the time of the work. 10. Future Enhancements In Future this project will be enhanced by creating more/different visualizations of collected data. Creating more modularity to enable a less cluttered user interface, more independent modules to facilitate more features. Option to save a portfolio set. To have another module such as a Customized stock portfolio based on data interacted through social media platforms such as twitter and facebook and google news. Using the information collected from these media platforms one can build the strategy and allocate/provide a visualization of those browsed or researched stocks in an automated manner; which may also include a trend analysis on those stocks which would be beneficial for the customer for further movements and decision making purposes.

Hit and source - focused comparison, Side by Side

Submitted text	As student entered the text in the submitted document.			
Matching text	As the text appears in the source.			
1/1	SUBMITTED TEXT	15 WORDS	70% MATCHING TEXT	15 WORDS
rule-based sentiment analysis tool that is specially tuned to sentiments expressed on social media.				
SA 18020590046 (Utkarsh).pdf (D105660647)				