

The Public Sentiment analysis within Big data Distributed system for Stock market prediction— A case study on Colombo Stock Exchange

M.V.D.H.P Malawana
Department of Computing &
Information Systems

Sabaragamuwa university of Sri Lanka
Belihuloya, Sri Lanka
Email:hashanmalawana@gmail.com

R. M. K. T Rathnayaka
Department of Physical Sciences and
Technology

Sabaragamuwa university of Sri Lanka
Belihuloya, Sri Lanka
Email: kapila.tr@gmail.com

Abstract— Stock price prediction plays an important role on the journey of investors on the stock market. The prices of the company stocks on the market are performed by different deliverables. Social media data sets, news sites, feedback and reviews are some kind of online tools that can affect the stock market. It is often worth using this context to predict the performance of market shares. We take the advantage of Sentiment analysis on Market related announcement and respective public opinions for stock market trend predictions for more accurate recommendations. Sentiment Analysis is a machine learning program for extracting opinions from a text section that is designed to support any product, company, individual or other entity (positive, negatively, neutral). In this research calculations and data processing were performed within machine learning approach with use of Spark model on Google cloud platform. Among most of the stock prediction researches, only few researchers have done their researches on sentiment analysis within big data distributed environment. Logistic Regression and Naïve Bayes perform well in sentiment classification. Main finding of this research is that public opinion significantly influences the fluctuations of market forces and economic factors such as monetarism, government reforms, unforeseen pandemics, interest rates, public trust, and faith in bond market trust. The detection of the feelings pattern will enhance the market prediction as it ensures the consistency of decision.

Keywords— Stock market, Sentiment analysis, Big data, Distributed environment, Machine learning

I.st INTRODUCTION

As it has been advanced into global economic environment Stock market[1][2] also advanced as one of the main forms of investment funds. There are many microeconomic and macroeconomic variables that have been reasonably impact on movement of price indices in the stock market around the world. Initially core concept and strategies behind the stock market investment depends on few deliverables. Examples are the prices of shareholders, when the company is running and overcoming its margins. Financial variables such as interest rates, country gross domestic product, economic outlook, inflation, deflation, share market announcements, changes in national economic and political policies, are directly caused by stocks increasing and decreasing in market prices with fluctuations of high volatility. Positive market conditions like positive mind set of investors, company financial margin levels, better government policies on industries and overall finance sector cause to market up movements while negative

market conditions like negative mind sets on stock market and industries, bad government policies on finance sectors cause market downward movements. Further these downward movements can occur for a wide variety of reasons, at any time even situations like Elections, War environments, bombing attacks, negative aspects of international relationships and unexpected situations like COVID-19 pandemic and after market conditions.

The factors influencing the decision of individual investors in share market investment is very important to assess. There are a number of factors that, according to a Khulna study[3], are industry attraction, corporate ranking, share price, historical data, expected dividends, stock index status, financial indicators, recent price volatility, not-stock attraction, portfolio diversification, capital gain, loss minimizations, financial statements and so on. These factors have a strict impact on decisions concerning stock markets.

Emotions of human beings have meaningful interactions with real situations. As shareholders of the stock market this is something very critical to look in emotional purpose on social networks like Facebook, Twitter, and Instagram. Social media or social networks are interactive computer-mediated technologies that facilitate the creation or sharing of information, ideas, career interests and other forms of expression via virtual communities and networks. By analyzing these social media sentiments regarding share market prediction, we can take a step further. It is called as sentiment analysis which is a new kind of text analysis which aims at determining the opinion and subjectivity of reviewers. We can get amazing insights into the level of discourse about the business and its history that is going on around the social networks. It seems to be its shareholders and investors attraction point or decision making point which effect to buy / sell their shares.

As a consequence, a range of applications to forecast business deliverables can be found. There are three main stock market prediction approaches: basic analyzes, technical (charting) analysis and technology (machine learning).

I. Fundamental analysis

Fundamental[4] analysis requires the use of real data to determine the value of a stock. This method uses revenue, earnings, future growth, equity return, profit margins and other data to assess the intrinsic value and potential of a

business for future growth. According to fundamental analysis is mainly based on three essential aspects and they are as follows,

(i) Macroeconomic analysis such as Gross Domestic Product (GDP) and Consumer Price Index (CPI) which analyses the effect of the macroeconomic environment on the future profit of a company.

(ii) Industry analysis which estimates the value of the company based on industry status and prospect

(iii) Company analysis which analyses the current operation and financial status of a company to evaluate its internal value.

II. Technical analysis

Technical[5] analysis of stocks and trends is the study of historical market data, including price and volume. Using both behavioral economics and quantitative analysis, technical analysts aim to use past performance to predict future market behavior. Many new technologies and methods have been introduced over the years to predict stock prices via many avenues, thanks to the challenging and ever-changing landscape of stock markets. Furthermore, Hu et al[6] grouped the domains of technical analysis into sentiment, flow-of-funds, raw data, trend, momentum, volume, cycle, and volatility.

III. Technology analysis

Technologies like Machine learning applications are used in this analysis. By using data models and algorithms it is able to predict the stock prices. Advancements in stock analysis and prediction fall under four categories—statistical, pattern recognition, machine learning (ML)[7], and sentiment analysis.

IV. Sentimental analysis

The purpose of information extraction (IE) is to obtain a piece of data appropriate to the requirements of particular topics or users. Sentimental analysis[8][9][10] which also known as Opinion Mining is a technique that uses Natural Language Processing (NLP)[11] and text analysis that helps in building a system that identifies and extract information in source material. Opinions regarding news are differ from each other but the investor opinions move in random on the question of markets returns. So, analyzing the effect of a news text on the stock market can make a considerable difference on investment return. Since we can forecast public patterns on particular attractions in the industry by using social media sentiment analysis, it will help to predict investor market behaviors and trends that influence stock prices.

V. Social Media

One of the advances that have led to the emergence of big data is Web 2.0 focused on Hypertext Markup Language (HTML). It transforms from static websites to interactive websites created by users (UGCs). The popularization of Web 2.0 resulted in a vast number of blogs, puzzles, social networking and bookmarking services. In open or closed groups, users can build and exchange information leading to large data volumes. Twitter is a micro-blogging platform that incorporates blog features and content from social networks. Twitter was founded in 2006 and in the first years of operations witnessed rapid growth of usage. It currently has more than 500 million

registered users and more than 200 million active monthly users.

VI. Social network analysis

Social network analysis (SNA) is the most popular analytical technique used to investigate social network theory. SNA is the study of the interpersonal interactions between interacting network members – individuals, organizations, etc. – and the consequences of such relationships

VII. Big Data with Distributed environment

As of the several definitions "Big data[12] refers to databases whose size exceeds the capability of traditional database software systems to collect, store, handle, and analyze." Every day "2.5 quintillion bytes of data" is generated according to IBM reports. These numbers are increasing each year. This is due to the omnipresent access to the Internet and the increasing number of devices mentioned above. Data is generated and transmitted from various systems in real time. For example social media platforms aggregate constantly information about user activities and interactions e.g. one of most popular social sites Facebook has over 618 million daily active users. The system's output rate may also be important when almost real-time analyzes are required. Such an on - the-fly analysis is required in recommendation systems when the user's input affects website content; online retail platforms such as Amazon.com are good examples.

Big data analysis in conventional, standalone methods is not a simple task. The dispersed machine workers are there. Distributed computing and parallel processing techniques will simplify the processing of large data since they vary from the conventional method. The Cloud Infrastructure with distributed computer technology improves the speeds and capabilities of data processing. It reduces latency and data volume restrictions.

VIII. Spark

Apache Spark is an open source distributed cluster computing system, offering an interface with implicit parallel data and default tolerance for entire cluster programming. Spark is a major analytical engine with built-in Streaming, SQL, machine learning and specialized application Graphical processing modules for big data processing. The Spark API can be used in a number of programs (Scala, Java, Python, R). Pyspark (Python in a spark environment) in Dataproc is employed in this research.

Spark provides the following three different Data Frame Solutions through its APIs: RDD, Dataframe, Dataset.

IX. Google Cloud Platform

Google's cloud platform is a suite of cloud-based computing services that operate on Google's own network for end-user products like Google Search, Gmail or YouTube. Google's Cloud Platform services are accessible via the public internet or by means of a dedicated network connection to app developers, cloud administrators and other IT companies.

Dataproc is a Spark and Hadoop managed service, which allows us to benefit from open source data resources for batch processing, querying, streaming and machine learning. List of the open source (Hadoop, Spark, Hive, and Pig) and Google Cloud Platform connector versions are supported by Dataproc.

X. Colombo Stock Exchange

The principal stock exchange in Sri Lanka is the Colombo Stock Exchange (CSE)[13]. It is one of the South Asian exchanges, providing a platform for electronic trading. Since 1995 CSE headquarter is located in the World Trade Center (Colombo) Towers in Colombo.

There are currently two indices in the Colombo Stock Exchange;

1. The All Share Price Index (ASPI)
2. The S&P Sri Lanka 20 Index (S&P SL20)

Predicting future stock prices

This section is to explain the mechanism that forecasts future stock trends based on opinion detection of tweets in Twitter. System design consists of four components: retrieval of social media data, pre-processing and database saving, retrieval of stock data, model creation and prediction of future stock trends.

II." METHODOLOGY

Data retrieving from Twitter

Twitter API is performed to extract the recent tweets into CSV format. Python special package called GOT is performed to extract the historical twitter datasets as Twitter API has the bother limitation of time constraints where we can't get older tweets than a week. By using GOT package it is able to have the JSON formatted tweets according to our requested query. By this package we get the advantage of Twitter Search on browsers which can search the deepest oldest tweets.

In this work Tweets were collected for three use cases of each over 3 month's period in the years of 2009, 2015, 2019. Tweets were filtered by using keywords related to economic aspects. Tweets regarding Sri Lanka stock market, economy, interest rates, government, banking and finance sector, investment, politics, EPF/ETF fund, GDP, Colombo stock exchange etc. were performed to extract the public opinions on Twitter.

Further, it is possible to analyze the effects of market related announcements like Dividend announcements, corporate disclosures, Appointment of Directors, Debenture Issues, Dealing by Directors, Right Issues and New listings etc. by analyzing the public sentiment on Social Media contents. Each announcement makes different effects on market trends.

Table 1 Tweet data for each Year

Time period	Focused Situation	Number of Tweets
2009 Sep 10 - 2009 Dec 30	After War situation in Sri Lanka	59700
2015 Jan 02 - 2015 March 31	Presidential Election	15249
2020 May 02 - 2020 July 07	After COVID 19	9200

From here data pre-processing and sentiment classifications were performed within spark model which employee Google Dataproc.

Data Pre-processing

Tweets as the text contents includes many characters whether useful or not. Therefor several steps need to employee as we need the abstract content of the tweets. Only tweets in English are used here. Pre-processing in the sense include several steps to filter the exact text points as tweets contain lots of irrelevant characters ,textual and links as well.

A. Tokenization

Divided the text by spaces, thus forming a list of individual words per tweet were selected as set of text need to analyst for the final outcomes. We then used each word in the tweet as features to train our classifier. Having the exact impression in the text content is obviously a challenge that we have to take.

B. Removing Stop words / Symbols

This is another important activity for eliminating stop words from the words list, and a Stop Word dictionary is found in the Python Natural Language Practice library, a list of words that are neutral and unacceptable to analyze the feelings. The algorithm simply tested each word in the word list against the dictionary to eliminate the stop words from each text. When a word was in the list, the tweet was excluded. The stop word list includes posts, some preparations, and other terms that do not add a meaning to the feelings (or are, are, are etc.), and of course the "@" and "#" tweets include additional symbols and URLs as well. The phrase "@" is always a username on Twitter, we exclude it because it never makes sense about the tweet's feeling. Words that follow "#" (hashtags) are not filtered out as they can contain important tweet feeling details. They are also especially useful for categorization, as Twitter uses hashtags to build new databases that are collections of similar tweets. URLs will also be completely filtered out as they don't make any sense to the text and could be spam as well. Use the Words Bag, a process that divides the sentence into words by a combination of n-gram and group it together.

Work with Google DataProc

These steps was followed to perform Spark job in DataProc.

I. Enabling APIs

GCP offers a broad variety of software, including Machine Engine, Cloud Storage, BigQuery, Cloud SQL, Cloud Dataproc. In order to use any of these resources in your project, you first have to activate them.

II. Installing Google Cloud SDK

Install the Google Cloud SDK so that we can interact with many services of GCP from the command-line.

III. Creating Bucket

Since we have installed Google Cloud SDK, we can either create a bucket from the command-line or from the web console.

IV. Creating Google Dataproc Cluster

Cloud Dataproc is a Google cloud service for running Apache Spark and Apache Hadoop clusters. As we expect it is ridiculously simple and easy-to-use and it only takes a couple of minutes to spin up a cluster with Google Dataproc.

V. Submitting Spark Job

Now, we are ready to run the training on Google Dataproc. The Python script (pyspark_sa.py) for the training is included in the Git repository we cloned earlier. Fig.1. shows the sequence of for submitting a spark job.

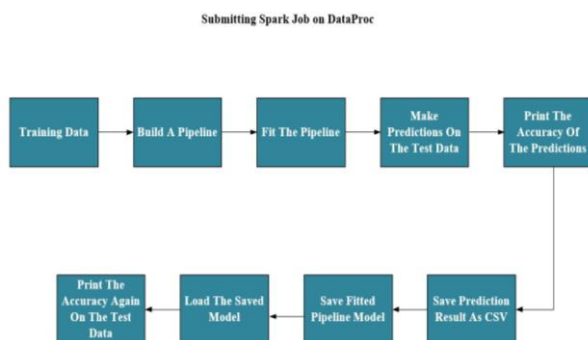


Figure II-1 Submitting Spark Job

VI. Checking the Results

Go to our bucket, then go into pyspark_nlp folder. We will see that the results of the above Spark job have been saved into “result” directory (for the prediction data frame), and “model” directory (fitted pipeline model).

Sentiment Analysis with PySpark

Pandas and Scikit-Learn are typically adequate to handle a pattern on the scale of the data we attempt to generate. However, that means we haven't had the ability to handle petabytes of data yet and we want to be prepared for a possible big-data.

Python's various modules and packages that lead to achieving data science objectives are excellent for data modeling. However, what if you can not fit into a single machine the knowledge we deal with? We can maybe carry out careful sampling on a single computer, but we can effectively execute the task in large data sets with distributed computing system, such as PySpark.

There are several programming languages available to Spark API (Scala, Java, Python and R). There are discussions as to how Spark output depends on the language we use, but since the main language we use is Python, we will concentrate on PySpark without going into the lengths that we should pick for Apache Spark.

Via its APIs Spark provides three different data structures: RDD, DataFrame (which is different to the Pandas data frame), Dataset. Resilient Distributed Datasets (RDD) is a fundamental data structure of Spark. It is an immutable distributed collection of objects. DataFrame is a distributed data set structured in designated columns in Spark. A broad variety of sources, including structured data files, Hive table, and external data bases or existing RDDs, can be created to build DataFrame. Dataset is a strongly typed data structure in SparkSQL that is a map of a relation scheme. Structured encoder queries are described. It's a data frame API extension.

“DataFrame is much faster than RDD because it has metadata (some information about data) associated with it, which allows Spark to optimize query plan.”

SparkMLlib is used with RDD, while SparkML supports DataFrame.

Sentiment and Stock Index Visualization

Using the best classifier is worth in sentiment analysis. Therefore our research work extended to find the best classifier among few Machine learning classifiers such as K Nearest Neighbor, Decision Tree Model, Support Vector Machine, Logistic Regression and Multinomial Naïve Bayes machine learning algorithms combined with sentiwordnet lexicon[14] approach and Using CountVectorizer to convert text into tokens/features. Observations of selected dataset show that Logistic Regression and Naïve Bayes perform well in all types of testing with the accuracy score of 72.352941 for LR and 71.764706 for NB. SVM at the next with 70.588235 of accuracy score. Then Decision tree classification with accuracy score of 64.117647. As of the result KNN scores lowest with 52.941176 of accuracy level. Logistic regression and Naïve Bayes models are performing respectively better in sentiment analysis and predictions.

We classified the tweets into one of three categories (positive or negative, neutral) with value of sentiment values by VADER tools. Then counted the numbers of positive and negative values for each day of the time series, and plotted them together with their difference and the daily stock indexes as ASPI (All share price index) or Sector's Indexes. For a day,

Total Sentiment = Positive sentiment - Negative sentiment

This fig.2. Shows the day wise sentiment values difference as above.

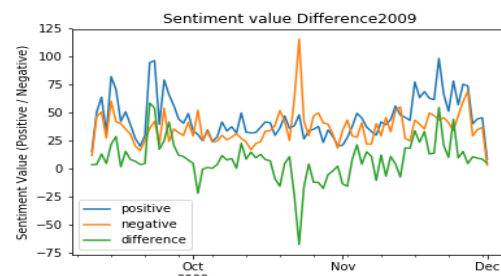


Figure 2 Sentiment value difference for 2009

III." RESULTS

Results of this paper can be presented in three scenarios. These plots show sensitive approach for the market trend prediction as investors need to be more wisely when the decisions to be taken.

Fig. 2. Showcase the respective stock market trend against the public sentiment in 2009 which affect after War period in Sri Lanka. Peaks show the days when people intensively tweeted about the stocks and the investment friendly environment in Sri Lanka. After War era in Sri Lanka most of the public look over the country in a positive manner. The figure itself shows the respective public opinions influence the overall market.

The proposed visual presentation of the sentiment time series for Presidential period in Sri Lanka can be seen in Fig. 2. Peaks show the days when people intensively tweeted about the stocks and the financial, political or economic perspectives in Sri Lankan. Two outstanding positive peaks can be observed in January and March 2015. One is a consequence of positivity of new government policies and the other is due to financial friendly state from the government. Decisions for



Figure 3 Sentiment Values and Stock market 2009

stock market prediction can be perform with this knowledge.

E.g.; How the market influence by the Political situations like elections?

How upcoming elections affects stock market? Will it drop or continue with usual pattern?

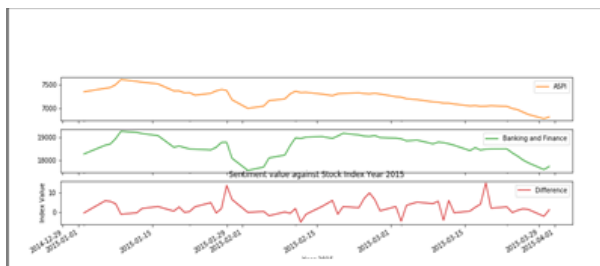


Figure 4 Sentiment Values and Stock market 2015

The proposed visual presentation of the sentiment time series for COVID 19 period in Sri Lanka can be seen in Fig. 3. Peaks show the days when people intensively tweeted about the market conditions and the good economic background in Sri Lankan. Two outstanding positive peaks can be observed in January and March 2020.

These plots shows sensitive approach for the market trend prediction as investors need to be more wisely when the decisions to be taken.

Further, sentiment analysis regarding financial aspects figure out critical factors related to the stock market which influenced the shareholders buying and selling. We are possible to grab these key factors buying sentiment analysis and predict the future market environment whether suitable for buying / selling or holding each sector's shares.

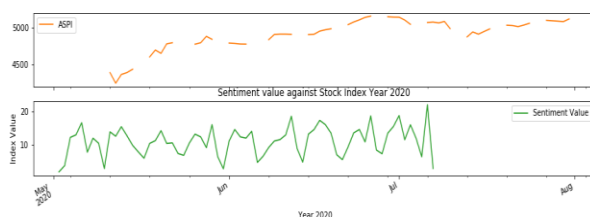


Figure III-1 Sentiment Values and Stock market 2020

E.g.; While interest rate getting low what will be the response in banking and finance sector shares prices? Is it worth to buy / sell or hold?

How the market sectors response in each situations like general elections, financial drops etc.

This fig.3. Shows the most frequent public opinion as a keywords where stock market sentiment is in a positive manner.



Figure III-2 Positive market condition

As the stock market patterns associates with the public sentiment we are possible to identify the negative market by analyzing negative opinions and factors as well.

This fig.3. Shows the most frequent public opinion as a keywords where stock market sentiment is in a positive manner.



Figure III-3 Negative market condition

Underlying Factors affecting the Stock Market

1. **Economic growth-** Better economic growth or better business expectations would allow businesses to be more competitive, as more demand for goods and services would increase. It will continue to raise earnings at the company and therefore share prices.
2. **Interest rates-** There are two reasons why lower interest rates[15] can make shares more attractive. Lower interest rates are helping to boost economic growth, making companies more profitable. Lower interest rates often make shares comparatively more appealing than a bank's saving capital or holding bonds.
3. **Stability-** Financial markets are not fond of surprises that could threaten economic stability and future development. And the reports of terrorist attacks or rises in the price of oil would continue to fall. This is what we possible to watch in year 2009 as Sri Lanka had War situation and end-up the mid of year 2009. They will also dislike political instability which can make strong economic policies difficult to pursue.

- 4." **Dealing by Directors** – How the company's directors move on with market conditions makes sense in the public behaviors on Stock selling and buying activities. These activities have the potentials to results selling and buying pressures on stocks.
- 5." **Dividends** – When companies announce their dividend dates and details shareholders response in different perspectives as their views. It influences the specific buying and selling pressures on stocks.
- 6." **Corporate Disclosures** – These factors highly motivate shareholders to hold, sell or buying specific shares of companies as these reports expose the corporate abilities and weaknesses as well. Shareholders keep on eye with company's performance as it directly decides their gain or loss of their investments.

An unexpected rapid rise in inflation would likely cause stock markets to fall. An increase in inflation would likely lead to a greater likelihood of interest hikes. This would reduce productivity and profitability. Higher inflation will also inspire investors to switch into more inflation-proof investments such as gold.

IV." SUMMARY AND CONCLUSIONS

Twitter feeds are a good source of data in the above cases to assess anticipated emotions, and frequent adjustments in positive feeling values are anticipated to predict similar price rises or decreases in advance. The mechanism could filter the posts by source and circumstance. I found that the prediction of stock prices is important by the use of the social media data sentiment analysis, because there is often no sense in technical and fundamental analysis of stock price fluctuations. The study shows that shifts in positive value can predict a similar shift in stock closing prices in circumstances where closing stock prices experience multiple fluctuations or drastic declines. The research is carried out in an environment with large data, where latency and data volume are not limited. For the training and testing of the classification, we experimented with various data sets to find the right classification for sentiment analysis. Policymakers should consider these macroeconomic impacts on stocks and take more proactive and accurate decisions. Analyzing the responses for market related announcement ensure the market decisions for effective stock market predictions of buying and selling activities.

The research work will continue to concentrate on Twitter, but must be connected to other social media such as Instagram, YouTube, and Facebook, etc. In the future. We foresee our approach to be used on data streams in order to forecast potential market price shifts.

It is also important to remember that stock forecast methods cannot predict sudden events called "black swans," but stock market feel analyzes are still important to ensure that market decisions are made.

Having a fundamentally and technically strong shares with better sentiment values makes investor's portfolio in a better status as it provides real expectations of investment in long term and short term manners.

ACKNOWLEDGMENT

This work was supported by Dr. Kapila Tharanga & Department head and lectures in Department of Computing and Information Systems. Also Lectures, Instructors and Demonstrators in Faculty of Applied Sciences in Sabaragamuwa university of Sri Lanka.

REFERENCES

- [1] A. M. Puertas, M. A. Sánchez-Granero, J. Clara-Rahola, J. E. Trinidad-Segovia, and F. J. De Las Nieves, "Stock markets: A view from soft matter," *Phys. Rev. E*, vol. 101, no. 3, 2020, doi: 10.1103/PhysRevE.101.032307.
- [2] A. Degutis and L. Novickytė, "The Efficient Market Hypothesis: a Critical Review of Literature and Methodology," *Ekonomika*, vol. 93, no. 2, pp. 7–23, 2014, doi: 10.15388/ekon.2014.2.3549.
- [3] F. Khan, "Factors Influencing Investors' Decisions in Stock Market Investment in Bangladesh [A Study on Khulna City]," *J. Financ. Account.*, vol. 3, no. 6, p. 198, 2015, doi: 10.11648/j.jfa.20150306.14.
- [4] D. Shah, H. Isah, and F. Zulkernine, "Stock market analysis: A review and taxonomy of prediction techniques," *Int. J. Financ. Stud.*, vol. 7, no. 2, 2019, doi: 10.3390/ijfs7020026.
- [5] Y. Hu, K. Liu, X. Zhang, L. Su, E. W. T. Ngai, and M. Liu, "Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review," *Appl. Soft Comput. J.*, vol. 36, pp. 534–551, 2015, doi: 10.1016/j.asoc.2015.07.008.
- [6] B. Dickinson and W. Hu, "Sentiment Analysis of Investor Opinions on Twitter," *Soc. Netw.*, vol. 04, no. 03, pp. 62–71, 2015, doi: 10.4236/sn.2015.43008.
- [7] M. Rushdi Saleh, M. T. Martín-Valdivia, A. Montejó-Ráez, and L. A. Ureña-López, "Experiments with SVM to classify opinions in different domains," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14799–14804, 2011, doi: 10.1016/j.eswa.2011.05.070.
- [8] D. Shah, H. Isah, and F. Zulkernine, "Predicting the Effects of News Sentiments on the Stock Market," *Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018*, pp. 4705–4708, 2019, doi: 10.1109/BigData.2018.8621884.
- [9] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011, doi: 10.1016/j.jocs.2010.12.007.
- [10] D. Shah, H. Isah, and F. Zulkernine, "Predicting the Effects of News Sentiments on the Stock Market," *Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018*, pp. 4705–4708, 2019, doi: 10.1109/BigData.2018.8621884.
- [11] F. Jiang *et al.*, "Artificial intelligence in healthcare: Past, present and future," *Stroke Vasc. Neurol.*, vol. 2, no. 4, pp. 230–243, 2017, doi: 10.1136/svn-2017-000101.
- [12] M. Skuza and A. Romanowski, "Sentiment analysis of Twitter data within big data distributed environment for stock prediction," *Proc. 2015 Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2015*, vol. 5, pp. 1349–1354, 2015, doi: 10.15439/2015F230.
- [13] P. Elangkumaran and J. . Jenitta, "Macroeconomic Variables and Stock Prices : A Study of Colombo Stock Exchange (CSE) in Sri Lanka," *Int. J. Res. Commer. IT Manag.*, vol. 4, no. January, 2014, doi: 2231-5756.
- [14] B. Zhao, Y. He, C. Yuan, and Y. Huang, "Stock market prediction exploiting microblog sentiment analysis," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2016-Octob, no. 1, pp. 4482–4488, 2016, doi: 10.1109/IJCNN.2016.7727786.
- [15] "Current and Future Developments in Artificial Intelligence," *Curr. Futur. Dev. Artif. Intell.*, no. March 2019, 2017, doi: 10.2174/97816810850291170101.