



Sentiment Analysis of Twitter and RSS News Feeds and Its Impact on Stock Market Prediction

Shri Bharathil* Angelina Geetha1 Revathi Sathiyarayanan1

¹*Department of Computer Science and Engineering, B.S.Abdur Rahman University, Vandalur, Chennai-600 048, Tamil Nadu, India*

* Corresponding author's Email: shribharathi01@gmail.com

Abstract In the recent, past accurate forecasting of stock market has become a challenging task. The proposed approach helps to achieve high precision in stock market prediction by combining the sensex points, with Really Simple Syndication (RSS) news feeds and Tweets. The algorithm focuses on the correlation between the stock market values, sentiments of tweets and RSS feeds for a particular period of time. In this algorithm, a trained model is used for stock market prediction rates. Experimental study focuses on the stock market sensex prices, RSS news feeds and tweets which are collected for the company ARBK from Amman Stock Exchange (ASE). This paper follows two types of hypotheses. Null Hypothesis H_0 : Stock level indicators predict the trend of stock rates at an allowable rate of minimum 80% above. Alternate Hypothesis H_a : Stock level indicators along with the sentiment analysis of RSS news feeds and tweets as stock enhances the accuracy of prediction. Our experimental study has proved the correlation between Stock level indicators and RSS news feeds and tweets and has shown significant improvement of 20% in prediction accuracy.

Keywords: Sentiment analysis, Stock market prediction, Twitter, RSS feeds, Sensex points, Stock market intelligence.

1. Introduction

Stock prices are considered as a basic measure in the prediction of stock market. Well defined strategies are used for earning high profit. The main aim of this work is to minimize the inaccurate forecasting of stock prices along with the help of sentiment analysis. Mathematically different indicators are there. However their level of accuracy in prediction is not very satisfactory. One way of improving this prediction process is by incorporating the sentiments and moods of public along with news events at that period of time. For capturing the sentiments of people, social media mining can be made use of. In our case we consider the tweets related to stock market as well as RSS news feeds. In order to capture the news events, we perform sentiment analysis on RSS news feeds.

In general stock market is volatile and needs

accurate forecast model. It is very difficult for an individual to consider all the current and past information for predicting future trend of a stock because it provides large amount of data every day. These high fluctuations affect the investor's belief. Financial stock market movements mainly depend on news, social networking sites like twitter, Facebook etc. current events and product releases and all these factors will have a significant impact on a company's stock value.

There are two standard measures used for predicting market trends. One is technical analysis and another one is fundamental analysis. Technical analysis considers daily ups and downs, trend indicators, highest and lowest values of a day, indices past price and volume to predict the future trend whereas Fundamental analysis involves reports and analysing its financial data to get some insights. The ability of both technical and

fundamental analysis is discovered by the efficient-market hypothesis which states that stock market prices are essentially unpredictable.

Nowadays, stock analysts can share their views through news document and social networking sites like twitter, Facebook etc. The behavior of investors is greatly affected by the sentiment of these mass media. Recently twitter has been used to detect and forecast trading and investing to break critical financial news faster. Twitter contains an enormous number of text posts and it grows every day. The collected corpus can be arbitrarily large. Thus if sentiments are correctly categorized and their polarity are correctly determined they can be helpful in enhancing a company's performance and making its investors happy. This research paper investigates the public sentiment, as expressed in large scale collections of daily twitter posts and RSS news feeds collected from stock related websites can indeed be used to predict the stock market.

A new novel approach is proposed to predict buy or sell signal to the investors in the stock market. This paper proposed a forecasting method by combining the stock related tweets and RSS news feeds with sensex points. Based on the combined result of opinion of sentences collected from tweets, feeds and sensex points of various stock related indicators the investors buy or sell their products.

This paper is organized as follows: Section 2 gives a survey on related work. Content retrieval procedures from twitter and RSS feeds are explained in Section 3. Section 4 describes Stock market prediction architecture. Section 5 presents the experimental results. Finally, Section 6 concludes our work with a scope for future work.

2. Related work

This paper explores a generic stock price prediction framework and considers textual documents as inputs and generates predicted price movements as outputs. For stock price prediction, a sentence level summarization model is applied to daily full-length news article [1]. For further improvement, higher level features such as sentiment analysis in the news pre-processing are needed to be considered.

This paper proposes a method which combines the technical indicator with the public opinion analysis. Performs the stock market prediction using the method of support vector machines [2]. In order to obtain more accuracy, this work needs to focus on public opinion data selection, public opinion influence factor analysis and improved prediction methods.

This work discovers subjective words and expressions automatically from unannotated texts. Also adds the learned nouns to an existing subjectivity lexicon, which use rule-based classifiers to bootstrap from the lexicon to harvest subjective and objective sentences from unannotated data [3]. For further increase in the performance of information extraction topic based text filtering and subjectivity filtering are considered.

This research paper explores a large scale of twitter feeds which are correlated to different values over time [4]. This paper needs to examine the ability to learn from past experiences and can adjust their trading behavior accordingly.

This paper focuses on the prediction of stock market performance of Karachi Stock Exchange (KSE) on day closing using different machine learning techniques. Different attributes have adopted like Oil rates, Gold & Silver rates, Interest rate, Foreign Exchange (FEX) rate, NEWS and social media feed are used as input and predicts the opinion [5]. This paper explored only oil attribute and needs to focus on other attributes also.

This paper proposes a global dimensionality-reduction algorithm which is used to model the complex investor information environment and capture the links among various information sources in a tensor. Also this paper proposed a high-order tensor regression learning problem [6]. This work needs to focus on various applications like detecting moving objects in video data, context-aware mobile recommendations and health care monitoring.

In order to predict rises and falls on the Korea Composite Stock Price Index (KOSPI), compile a stock domain specific sentimental word dictionary from the news as big data and predict 'sentiment' for that Korean language news [7]. Needs further improvement in attributes such as market prospects, overseas news, and corporate performance

This research explores the opinion mining on stock market by combining the sensex points of moving average stock level indicator with RSS news feeds which obtained the high accuracy rather than individual sensex calculation [8]. This paper needs to focus on more than one stock level indicator with sentiment analysis.

For model optimization a novel sigmoid-based mixed discrete-continuous differential evolution algorithm is developed and incorporated into the proposed stock selection model for feature selection and weight optimization, which finely verifies the effectiveness of the variant in stock selection [9]. This work can be improved further into multiple objective models to provide different satisfactory portfolios according to different goals.

An effective and efficient method is focused for topic modeling based on Latent Dirichlet Allocation (LDA) that merges tweets in order to improve LDA performance [10]. Also needs to focus in the improvement of discarded tweets. Introduces a new technology which is used to collect the web news information from RSS website. Thus news gathering and filtering system based on RSS improves the efficiency [11]. This approach is similar to our approach of sentiment analysis.

This work discovers the accuracy of classification by including natural language processing techniques like semantics and Word Sense Disambiguation. Various methods of classification are applied to classify the data as Positive, Negative and Neutral [12]. For better performance randomized trees classification method needs to be considered. In order to predict the sentiments around the stock news a model is built and sentiment score values are analyzed for the stock news whether it is positive, negative or neutral [13]. For obtaining better accuracy negation and valence shifters also needs to be considered.

This research work explores the knowledge based historical prices of stock market and decision tree classifier is used to take the decision in the stock market for buying or selling of stocks [14]. There is a need for improvement in certain factors that affect the behavior of the stock markets, such as trading volume, news and financial reports are to be reconsidered which might impact stock price.

This paper analyses the relation between micro blog feeds and Chinese stock market [15]. To achieve higher prediction accuracy some of other moods like foreign mood state and political mood states are need to be considered.

This paper proposes an effective and accurate stock market prediction technique by combining the social media mining technology with the stock prices [16]. This analyse is similar to our approach of sentiment analysis.

3. Twitter and RSS news feeds

3.1 Twitter and its impact on stock market prediction

Twitter is a real-time micro blogging platform that allows people to communicate with short messages. Twitter post which is known as tweet comprises of 140 characters or less. In general users can instantly post anything to express their curiosity, how they feel, or just to give status updates to friends and family members. For the purpose of predicting stock market trends, excellent data source

which consists of many characteristics that are gathered from the twitter for sentiment analysis. Twitter is open for public consumption and any tweet can be retrieved without any privacy restrictions. Twitter has clean and well-documented API that enables developers to query for specific collection of tweets using certain keywords or based over a period of time.

In this proposed work, sentiments of the stock related tweets are analyzed and based on the polarity of the sentiments with a sentiment score ranging from +1 to -1. If the total sentiment score value is 0.0 to 1.0 then the sentence is positive sentence. If the total sentiment score value is -1.0 to 0.0 then the sentence is negative sentence. If the total sentiment score value is 0.0 then the sentence is neutral sentence. The sentiment score is correlated with the changes in the sensex price of stock for a certain period.

3.2 RSS news feeds and its effect on stock

Really Simple Syndication (RSS) is a format for delivering regularly changing Web content. Many news-related sites, Weblogs and other online publishers syndicate their content as an RSS Feed to whoever wants it. It is an XML document that facilitates content syndication. A RSS is reliable way to have the web content delivered to Internet since the data is small and fast-loading, it can be used with services like cell phones or PDA's, voice mails, and email. Unlike email an RSS feed is zero maintenance, the messages will never get blacklisted or filtered. With RSS, users can separate wanted information from unwanted information. RSS documents use a self-describing and simple syntax. Generally, RSS news feed contains author, title, and date information in addition to link and description.

If progressive and good news on current affairs is released then this has a positive impact on the stock market values. This is captured by analyses the RSS news feeds.

4. Proposed work

Given in Fig. 1 the architecture of the proposed system. It basically accepts stock RSS news feeds and tweets automatically retrieved from twitter website. It also retrieves historical prices from respective stock market site. In our case it is <http://www.marketstoday.net/markets/jordan/Historical-Prices/10/en/#>. The task is divided into two parallel operations.

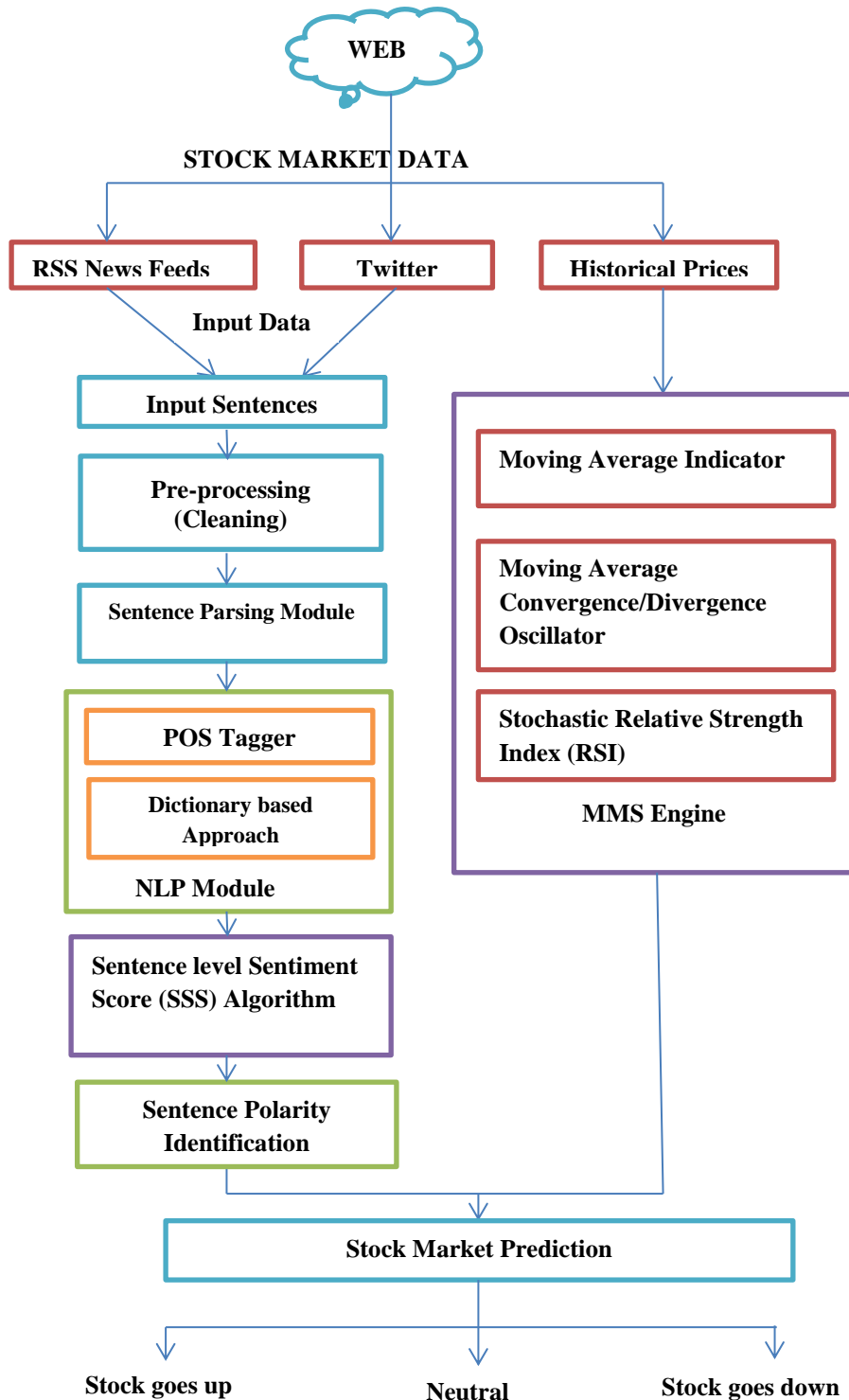


Figure.1 Stock market prediction architecture

One of the parallel task is performing sentiment mining on the contents of RSS news feeds and tweets retrieved from twitter site. RSS reader reads the required stock news content from the web pages such as title, description, date, author, link etc. in the format of XML document.

RSS feeds take the latest headlines of stock feeds from the related websites. All collected feeds are

small sentences and stored in the document. In the same way tweets from twitter contains an enormous number of text posts and it grows every day. Twitter aggregates users into communities and links users in a variety of ways, ranging from short dialogues to interest graphs. Short tweets are gathered and stored in a document. The collected data from RSS news feeds and tweets are noted as input sentences which

are not parsed as sentences. They are in the form of document. These sentences are passed to the pre-processing steps where only document cleaning is performed. Unwanted contents such as comma, semi colon, numerals, symbols, date and time etc. are termed as noises and are removed in this step. After performing the cleaning operation, Sentence splitting module will separates the document sentences into an individual sentence and stored it in a file.

Now the individual sentences are passed to the Natural language processing (NLP) module. NLP is used to identify and extract sentiment for the words having a positive, negative and neutral value. Here the part-of-speech tagger, dictionary based approach are used to find the polarity of the sentence, to conclude whether the sentence is positive, negative and neutral. Each separated sentences are passed to the Part-of-speech (POS) tagger in NLP Module. POS is a piece of software that reads text in some language and assigns parts of speech to each word, such as noun, verb, adjective, etc., After assigning the POS, each word carries either noun, verb or adjective etc. Now each word is passed to the dictionary based approach. The dictionary is used to find the opinion words and their polarities. Here orientation of each sentences are found and also used to determine the opinion of words and their polarities. The three types of classified opinions are positive, negative or neutral.

After NLP module, Sentence score sentiment (SSS) algorithm is used for finding the overall result of each sentence. After finding the results of all sentences, total score of summarization of opinions will explore the result of that document. In general, the score ranges from -1.0 to 0.0 and 0.0 to 1.0. The calculation of total score value of each sentence is performed by assigning the score value for each word of a sentence and their sum is calculated. Synset consists of a set of one or more synonyms. A sentence is said to be positive only if the score value occurs between 0.0 to 1.0. Whereas a sentence is said to be negative only if the score value ranges from -1.0 to 0.0. A neutral value is assigned if its value is 0.0.

The following Eq. (1) is used to calculate the Sequence of words.

$$\text{Sequence of words}(W) = W_1 + W_2 + \dots + W_n \quad (1)$$

n = Number of words.

The other parallel operation is computing the stock level indicators by the MMS engine. The

MMS engine consists of Moving average, Moving average convergence/ Divergence, Stochastic RSI stock level indicators.

The first indicator called the Moving Average stock level indicator is computed by considering the closing price of the security for a specified number of time periods and then dividing this total by the number of time periods. Common time periods for moving averages are 5 days, 10 days, 15 days, 21 days, 50 days, 100 days and 200 days.

Eq. (2) describes the simple moving average formula which is calculated by taking the average closing price of a stock over the last "n" periods.

$$F_t = \frac{A_{t-1} + A_{t-2} + A_{t-3} + \dots + A_{t-n}}{n} \quad (2)$$

F_t = Forecast for the coming period,

A_{t-1} = Actual occurrence in the past period for up to 'n' periods,

n = Number of periods to be averaged.

The second indicator is Moving Average Convergence/Divergence oscillator (MACD). It measures the difference between two Exponential Moving Average (EMA)s. If there is a positive indication in MACD then the 12-day EMA is above the 26-day EMA. This indicates that the rate-of-change of the faster moving average is higher than the rate-of-change for the slower moving average. Positive momentum is increasing and this would be considered rising.

If there is a negative indication then the 12-day EMA is below the 26-day EMA. If it is negative and declining further, then the negative gap between the faster moving average and the slower moving average is expanding. Since downward momentum is accelerating, this would be considered falling. MACD centerline crossovers occur when the faster moving average crosses the slower moving average. The formula for MACD, Signal Line and MACD Histogram are given as follows. The formula for MACD, Signal Line and MACD Histogram are given in Eqs. (3),(4) and (5) respectively.

$$\text{MACD} = (12 - \text{day EMA} - 26 - \text{day EMA}) \quad (3)$$

$$\text{Signal Line} = 9 - \text{day EMA of MACD} \quad (4)$$

$$\text{MACD Histogram} = (\text{MACD} - \text{Signal Line}) \quad (5)$$

Third indicator is Stochastic Relative Strength Index (RSI). It is an oscillator that calculates a value between 0 and 1 which is then plotted as a line.

This indicator is primarily used for identifying overbought and oversold conditions.

The formula for RSI is pretty straightforward: The Eq. (6) describes the RSI calculation of indicator whereas average gain and losses are explained in Eq. (7) and (8). The RS value is calculated using Eq. (9) the overall Stochastic RSI value calculation is shown in Eq. (10).

$$RSI = 100 - \frac{100}{1+RS} \tag{6}$$

$$Average\ Gain = \frac{Total\ Gains}{n} \tag{7}$$

$$Average\ Losses = \frac{Total\ Losses}{n} \tag{8}$$

n = number of RSI periods

$$RS = \frac{Average\ Gain}{Average\ Losses} \tag{9}$$

$$StochRSI = \frac{RSI(n) - RSI\ Lowset\ Low(n)}{RSI\ Highest\ High(n) - RSI\ Lowest\ Low(n)} \tag{10}$$

RSI = RSI indicator or current level.

$RSI\ Lowest\ Low\ (n)$ = RSI reached the lowest level over the last n periods.

$RSI\ Highest\ High\ (n)$ = RSI reached the highest level over the last n periods.

n = Number of periods used in calculation.

When the Stochastic RSI reaches above 0.80, then the result indicates overbought. Similarly when it reaches below 0.20, and then it indicates oversold. If it crosses 0.50 line in an upwards direction it is identified as oversold situation as a buy signal, and crosses 0.50 line in a downwards direction it is identified as overbought situation as a sell signal.

4.1 Prediction strategy

To predict the stock market, the results of both sentiment analysis of Twitter and RSS News feeds along with Sensex are combined and analyzed. Table 1 shows the final result prediction technique for Stock market. Table 1 shows the analysis of orientations for twitter data and RSS news feeds along with sensex points. After performing the optimizations the numerous number of combinations of sentiment results is reduced to 10 significant combinations.

5. Experimental results

In the experimental study the prediction of stock market was carried out for the company Arab Bank

Table 1. Sentiment analysis for Twitter, RSS news feeds and sensex-stock level indicators for result prediction method and previous methods

Sentiment Analysis for RSS News Feeds	Sentiment Analysis for Twitter	Stock level indicators Result	Final-Result Prediction
Positive	Positive	Positive	Positive
Negative	Negative	Negative	Negative
Neutral	Neutral	Neutral	Neutral
Positive	Positive	Negative	Positive
Positive	Positive	Neutral	Positive
Positive	Negative	Negative	Negative
Positive	Neutral	Neutral	Neutral
Positive	Negative	Neutral	Neutral
Negative	Negative	Neutral	Negative
Negative	Neutral	Neutral	Neutral

(ARBK) from Amman Stock Exchange (ASE) by collecting various Sentiment data and sensex prices. The Oracle database of Amman Stock Exchange (ASE) contains the historical prices of the 230 companies listed in the exchange from the year 2000. The historical prices are collected from the year 2014 to 2016.

In the previous work 20 items small –cap stocks were randomly selected from Chinese stock market and the social media data fetched from Sina Weibo, Tong Hua Network and Dong Fang Cai Fu networkShun and aims to predict the stock price accurately [16].

In our experimental study, for ARBK the RSS news feeds are collected from <http://investing.einnews.com/news/ase-stock> and for the same company the tweets are collected from <http://twitter.com/ArabBankGroup?lang=en>. Now the sentiment value is obtained by applying SSS algorithm for specific month. The sensex price for ARBK Company is also collected from <http://www.marketstoday.net/markets/jordan/Historical-Prices/10/en/#>.

Now three stock level indicators are applied and stock value is calculated for specific month. Finally the combined results of Twitter, RSS news feed and sensex prices for stock level indicators helps to predict the stock forecasters to decide whether buy or sell. Fig. 2 shows the polarity calculation of Sentiment for Tweets and RSS Feeds and Fig.3 shows the overall positive and negative sentences of Tweets and the implementation shows that the news result is positive.

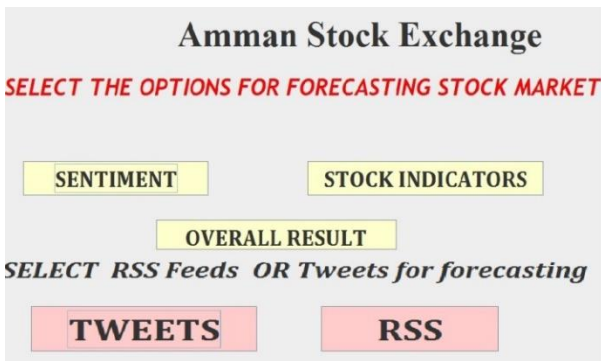


Figure. 2 Sentiment analysis for tweets and RSS feeds



Figure. 3 Polarity score value of Tweets

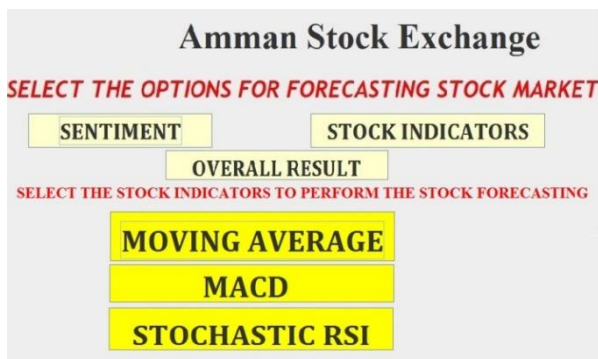


Figure. 4 Calculations of stock values for various stock level indicators



Figure. 5 Overall result of stock market forecasting

Calculation of sensex prices for various stock level indicators is performed in Fig.4 Combined result of all stock level indicators shows the overall

score value of stock market for sensex data. The overall traders choice will be shown in Fig. 5 by combining the result of RSS, Tweets and Stock level indicators value.

In an effort to gather evidence about the prediction accuracy a statistical hypothesis testing is conducted. There are two types of hypotheses such as null hypothesis and alternate or research hypothesis. Null hypothesis is generally termed as H_0 and alternative hypothesis is symbolized as H_a . In general, hypothesis is conducted to prove true or false with absolute knowledge. There are two possible outcomes. One is Reject H_0 and accept H_a because of sufficient evidence in the dataset in favour or H_a ; another one is do not reject H_0 because of insufficient evidence to support H_a .

In this proposed paper using one portion z-test statistics, hypothesis testing is conducted to H_0 .

H_0 : Stock level indicators predict the trend of stock rates at an allowable rate of minimum 80% above.

H_a : Stock level indicators along with the sentiment analysis of RSS news feeds and tweets as stock enhances the accuracy of prediction. The Eq. (11) and (12) shows the formula for one portion z-test statistics which is as follows,

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \tag{11}$$

$$\hat{p} = \frac{x}{n} \tag{12}$$

Where \hat{p} = the sample proportion.

p = hypothesized value of population proportion in the null hypothesis

x = number of samples occurred.

n = total sample size.

The two hypothesis of accuracy prediction of Sentiment and Stock Level Indicators are as follows and shown in the Table 2 and Table 3. Also Table 4 shows the Hypothesis testing decision about H_0 .

The conditions for making decisions about the hypothesis is as follows,

If the p-value is low, then reject H_0 where as if the p-value is high, then fail to reject H_0 . If Z test value is lower than Z then fail to reject H_0 where as If Z test value is greater than Z then reject H_0 . The Fig. 6, 7 and 8 shows the Z test curve for the significance level and Z.

Analysing the results of Z test and Z, α and p-value of stock level indicators we have enough evidence to reject null hypothesis (H_0).

Table 2. H_0 : A stock level indicator does not give accurate prediction

Company (ARBK)	Moving Average	MACD	Stochastic RSI
Total Instances	684	684	684
Correctly Classified	440	413	432
Precision %	64.32%	60.38%	63.15%

Table 3. H_a : Stock level indicators along with the sentiment analysis of RSS news feeds and tweets give the accuracy of prediction

Company	Sentiment Analysis for RSS Feeds + Moving Average Stock level Indicator	Sentiment Analysis for RSS Feeds + Stock level Indicators	Sentiment Analysis for RSS Feeds and Tweets + Stock level Indicators
Total Instances	684	684	684
Correctly Classified	542	548	575
Precision %	79.26%	80.11%	84.06%

Table 4. Hypothesis testing shows decision about H_0

S.No	Stock level Indicators	H_0	H_a	Z test	Z	α	p-value	Decision
1.	Moving Average	$p = 0.82$	$p > 0.82$	12.04	1.645	0.05	0.00001	H_0 : Rejected
2.	MACD	$p = 0.82$	$p > 0.82$	14.78	1.645	0.05	0.00001	H_0 : Rejected
3.	Stochastic RSI	$p = 0.82$	$p > 0.82$	12.84	1.645	0.05	0.00001	H_0 : Rejected

Table 5. Hypothesis testing shows decision about H_a

S.No	Stock level Indicators	H_0	H_a	Z test	Z	α	p-value	Decision
1.	SA RSS + Moving Average	$p = 0.82$	$p > 0.82$	1.362	1.645	0.05	0.0865	H_a : Accepted
2.	SA RSS + 3 Indicators	$p = 0.82$	$p > 0.82$	1.28	1.645	0.05	0.10027	H_a : Accepted
3.	SA RSS + Tweets+3 Indicators	$p = 0.82$	$p > 0.82$	1.403	1.645	0.05	0.80308	H_a : Accepted

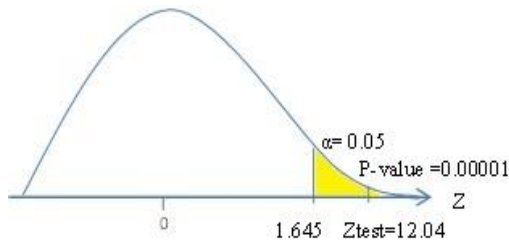


Figure. 6 Z test curve for moving average

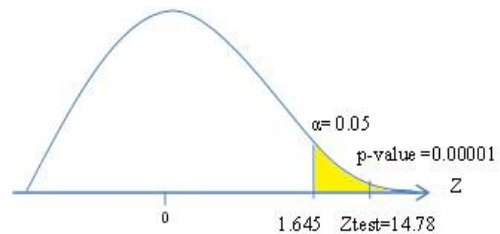


Figure. 7 Z test curve for MACD

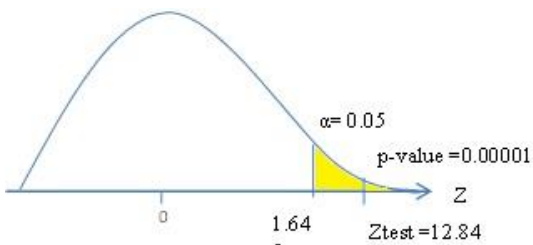


Figure.8 Z test curve for stochastic RSI

The P-value is the probability of observing a sample statistic as extreme as the test statistic. α is the significance level, generally 0.05.

For Moving average stock level indicator the observed Z test value is 12.04. Since it is greater

than Z and Z falls within the rejection region, we reject H_0 . With a statistic of 12.04 and p-value of 0.00001 which is lesser than α value, at a 5 % level of significance, we have enough evidence to reject the null hypothesis and accept Alternate hypothesis.

For Moving Average Convergence/Divergence Oscillator (MACD), the observed Z test value is 14.78. Since it is greater than Z and Z falls within the rejection region, we reject H_0 . With a statistic of 14.78 and p-value of 0.00001 which is lesser than α value, at a 5 % level of significance, we have enough evidence to reject the null hypothesis and accept Alternate hypothesis. Table 5 shows the Hypothesis testing decision about H_a .

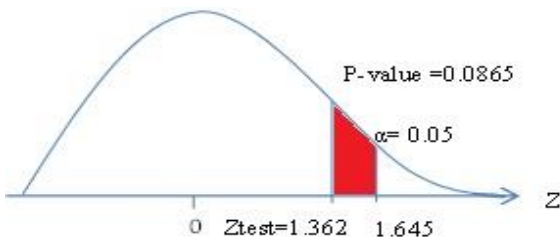


Figure.9 Z test curve for Sentiment analysis + Moving Average stock level indicator

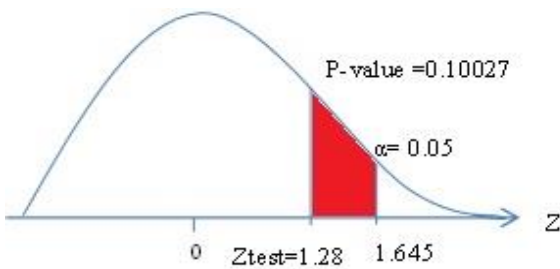


Figure. 10 Z test curve for Sentiment analysis (RSS) + 3 Stock level Indicators

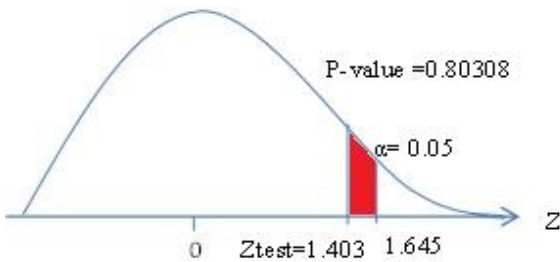


Figure.11 Z test curve for Sentiment analysis (RSS, tweets) + 3 Stock level Indicators

For Stochastic RSI stock level indicator the observed Z test value is 12.84. Since it is greater than Z and Z falls within the rejection region, we reject H_0 . With a statistic of 12.84 and p-value of 0.00001 which is lesser than α value, at a 5 % level of significance, we have enough evidence to reject the null hypothesis and accept Alternate hypothesis.

The Fig. 9, 10 and 11 shows the Z test curve for the significance level and Z for the H_a .

Analysing the results of Z test and Z, α and p-value of stock level indicators we have enough evidence to accept alternate hypothesis (H_a).

Z test curve for sentiment Analysis and moving average stock level indicator has the observed Z test value is 1.362. Since it is lesser than Z and Ztest does not falls within the rejection region, hence we accept H_a . With a statistic of 1.362 and p-value of 0.8065 which is higher than α value, at a 5 % level of significance, we have no evidence to reject the alternate hypothesis and accept Alternate hypothesis.

Z test curve for sentiment analysis (RSS) and 3 stock level Indicators has the observed Z test value is 1.28. Since it is lesser than Z and Ztest does not falls within the rejection region, hence we accept H_a .

With a statistic of 1.28 and p-value of 0.10027 which is higher than α value, at a 5 % level of significance, we have no evidence to reject the alternate hypothesis and accept Alternate hypothesis.

Z test curve for sentiment analysis of RSS, tweets and 3stock level Indicators has the observed Z test value is 1.403. Since it is lesser than Z and Ztest does not falls within the rejection region, hence we accept H_a . With a statistic of 1.403 and p-value of 0.80308 which is higher than α value, at a 5 % level of significance, we have no evidence to reject the alternate hypothesis and accept Alternate hypothesis.

6. Conclusions

In the conventional stock market analysis, stock level indicators such as Moving average, Moving average convergence/ Divergence, Stochastic RSI stock level indicators are used for stock market prediction.

In our previous work, we have established the fact that the sentiment mining of RSS news fields helps in improving the stock market prediction along with stock level indicators. In this work, the impact of sentiments from tweets as well as RSS news feeds are analyzed. It is found that the sentiments from social media along with stock level indicators enhances the quality of prediction. Since our approach is an hybrid approach, the experimental analysis is carried out for hypothesis testing H_0 and H_a . Also H_a have shown significant improvement in the precision and correctness when compared with H_0 . H_a have enough evidence to reject the null hypothesis and accept Alternate hypothesis. This work can further be extended for prediction of buying patterns from customers by incorporating the sentiment mining from various social media such as RSS news feeds, Tweets, LinkedIn and Facebook contents.

References

- [1] X. Li, H. Xie, Y. Song, S. Zhu, Q. Li and F.L. Wang, "Does Summarization Help Stock Prediction? A News Impact Analysis", *IEEE Intelligent Systems*, Vol. 30, No. 3, pp. 26 – 34, 2015.
- [2] G. Wei, W. Zhang, and L. Zhou, "Stock Trends Prediction Combining The Public Opinion Analysis", In: *Proc. of the IEEE International Conf. On Logistics, Informatics and Service Sciences*, Spain, Vol.1, No.1, pp. 1-6, 2015.
- [3] J. Wiebe and E. Riloff, "Finding Mutual Benefit Between Subjectivity Analysis and Information

- Extraction”, *IEEE Transactions on Affective Computing*, Vol.2, No.4, pp. 175- 191, 2011.
- [4] J. Bollen, H. Mao, and X. Zeng, “Twitter Mood Predicts The Stock Market”, *Journal of Computational Science*, Vol. 2, No. 1, pp. 1-8, 2011.
- [5] M. Usmani, S. Hasan Adil, K. Raza, and S.S. Azhar Ali, “Stock Market Prediction Using Machine Learning Techniques,” In: *Proc. of the IEEE International Conf. on Computer and Information Sciences*, Vol.1, No.1, pp. 322-327, 2016.
- [6] Q. Li, Y. Chen, L. L. Jiang, P.Li, and H. Chen, “A Tensor Based Information Framework For Predicting The Stock Market”, *ACM Transactions on Information Systems*, Vol.34,No.2, pp.1-30, 2016.
- [7] Y. Kim, S.R. Jeong, and I. Ghani, “Text Opinion Mining To Analyze News For Stock Market Prediction”, *International Journal of Advances in Soft Computing and its Applications*, Vol. 6, No. 1, pp. 1-13, 2014.
- [8] S. Bharathi and A. Geetha, “Sentiment Analysis for Effective Stock Market Prediction”, *International Journal of Intelligent Engineering and Systems*, Vol.10, No.3, pp. 146- 153, 2017.
- [9] L. Yu, L. Hu and L.Tang, “Stock Selection With A Novel Sigmoid-Based Mixed Discrete Continuous Differential Evolution Algorithm”, *IEEE Transactions on Knowledge and Data Engineering*, Vol.28, No.7, pp.1891- 1904, 2016.
- [10] Z. Z. Alp and S. G.Oguduc, “Extracting Topical Information of Tweets Using Hash Tags”, In: *Proc. of the IEEE International Conf. on Machine Learning and Applications*, Vol.1, No.1, pp. 644-648, 2015.
- [11] M. Kanakaraj and R. M. R. Guddeti, “Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis Using NLP Techniques”, In: *Proc. of the IEEE International Conf. on Semantic Computing*, Vol.1, No.1, pp. 169-170, 2015.
- [12] Q.A. Al-radaideh, A.A. Assaf, and E. Alnagi “Predicting Stock Prices Using Data Mining Techniques”, In: *Proc. of the International Arab Conf. on Information Technology*, Vol.4, No.3, pp. 163-191, 2013.
- [13] S.V.S. Bharathi and A. Geetha, “Sentiment Analysis for Online Stock Market News using RSS Feeds”, In: *Proc. of the International Conf. on Computer Science and Information Technology*, Vol.1, No.1, pp. 01-07, 2017.
- [14] Z.Rui-Juan and Z.Y. Sen, “Design And Implementation of News Collecting And Filtering System Based on RSS”, In: *Proc. of the International conf. on Fuzzy Systems and Knowledge Discovery*, Vol.1, No.1, pp. 2295-2298, 2012.
- [15] D. Yan, G. Zhou, X. Zhao, Y.Tian, and F. Yang, “Predicting Stock Using Micro Blog Moods”, *Journal on China Communications*, Vol. 13, No. 10, pp.244- 257, 2016.
- [16] Y.Wang and Y.Wang, “Using Social Media Mining Technology to Assist in Price Prediction of Stock Market”, In: *Proc. of the IEEE International Conf. on Big Data Analysis*, Vol.1, No.1, pp. 48-51, 2016.
- [17] <http://www.bseindia.com/indices/indexarchivedata.aspx>