

Sentiment Analysis for Effective Online Stock Market Prediction using RSS News Feeds and Stock Level Indicators

SV. Shri Bharathi¹, Angelina Geetha²

¹Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai-602 105, Tamil Nadu, India

²Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai-603 103, Tamil Nadu, India

¹shribharathi01@gmail.com, ²angelinageetha@gmail.com

Article Info

Volume 83

Page Number: 3421-3426

Publication Issue:

May-June 2020

Abstract

In the field of stock market prediction, along with the historical prices of the stock market, RSS news feeds has great role in the forecast of financial exchange. Forecasting of the stock market is influenced by the RSS financial news feeds. This work is an endeavor to construct a system that identifies RSS news sentiments which may influence changes in stock information development patterns. The proposed approach analyses the sentiments in Really Simple Syndication (RSS) stock news feeds together with the historical prices of stock level indicators. Subsequently financial exchange RSS news feed information data is assembled for some time. Same way historical prices are also gathered for the same period of time. By combining both the results, the proposed method establishes an algorithm for sentiment analysis which can be used in the forecasting of stock market movement. In our exploratory investigation, the proposed technique finds that the RSS stock news feeds sentiment impact the forecast of stock value variances, regardless of whether up or down.

Article History

Article Received: 19 August 2019

Revised: 27 November 2019

Accepted: 29 January 2020

Publication: 12 May 2020

Index Terms: Stock Market Intelligence, Stock Market Forecast News, RSS Feed Reader, Sensex Points, Sentiment Analysis, Opinion Mining,.

1. Introduction

A great provocation problem in the field of fund is financial exchange determining. Now a day researchers have more attraction towards the forecast of stock exchange analysis. In the stock field, financial exchange and its movements are very unstable in nature. For most stock holders and money related investigators forecast of stock cost has consistently been a subject of interest. There are various elements that impact the financial exchange costs, which may influence the purchasing and selling of stock costs.

Sentiment Analysis (SA) is a latest sub-discipline of text categorization which expresses about the opinion of the text. It is otherwise called as opinion

mining, sentiment categorization or sentiment orientation analysis. In order to identify the sentiment, subjectivity, effectiveness, opinion mining is used.

Consistently, enormous amount of stock information are delivered and it is hard for a person to think about all the present and former data for foreseeing future pattern of a stock. Forecasting of stock market is divided into two methods. They are fundamental method and technical method. Fundamental method involves the business analysis and its financial data. Whereas technical analysis involves past historical data and volume.

By using news reports the proposed technique follows the basic investigation method to find future design of a financial data about an organization as prime data and attempts to group news as great (progressive) and terrible (undesirable). On the off

chance that the stock information assessment is certain, there are more possibilities that the stock cost will rise up and on the off chance that the news estimation is negative, at that point stock cost may go down.

The proposed work is well regulated as follows: Section 2 reviews the literature survey part. Area 3 talks about the Sentiment Investigation and RSS News data Feeds. Segment 4 depicts the News of the Stock Market. Section 5 depicts architecture diagram to identify the sentiments of Online Stock market information utilizing RSS Feeds and stock level pointers. Segment 6 presents the test results. At long last, Section 7 explains about conclusion.

2. Related Work

In modern years, critical endeavors have been placed into creating designs that can foresee the upcoming pattern of a particular financial data or by and large market. The majority of the current strategies utilize the specialized stock market techniques. The forecast of Chinese stock exchange with the assistance of open mind-sets is extricated from small scale blog takes care of [1].

The work explored the social issues related to the subjectivity [3]. The report of this firmly relies upon its sentences. A lexical-linguistic way was proposed to deal with perceive and characterize subjectivity at the sentence level. In this paper, retailer's tweets were put away and tweets assessments were dissected. Contrasted and standard model like SVM and MLP model, this work indicated great outcomes and improvement in the forecast exactness [4].

This work proposed a conventional stock value forecast structure to empower the utilization of various external signs to anticipate stock costs [5]. Assessments were accomplished at specific stock, section file, and stock index values. Alternative approach was suggested to identify Twitter information [6]. By removing vague literary tweet information using NLP systems to characterize open estimation, at that point utilize an information mining procedure to find designs between open assessment and genuine stock value developments. The proposed work has a decision tree classifier that classifies the information mining systems that assists with taking the choice in the stock exchange [7].

This approach uses the consequences of a lexicon sentiment analysis investigation led on crowd-annotated news data to separate different sorts of common opinions from every day news stories. One more methodology was a approach to score news related articles with respect to public feelings, and to distinguish which news segments and sentiments because developments in a financial exchange record [8]. To manage the vulnerability factors, a fluffy logical approach dependent on grid partition is accepted in this work, for forecasting of the stock cost of any organization. Likewise the reason and resulting factors

of the knowledge rules are advanced in an optimized manner utilizing a hybrid neural learning system [9].

This work concentrated on the Joint Aspect/Sentiment (JAS) model, to together concentrate on aspect-dependent emotion lexicons from web client surveys [10]. This work consequently gathers the news substance from different web news media entryways and afterward solidifies them for the assurance of news biasedness. The idea of relationship between's the conclusion of RSS channels, twitter and at least one stock level pointers for the forecast of securities exchange are broke down [2, 11, and 14]. An epic stock determination model with discrete and persistent factors calculation is presented for model streamlining [12]. Examination of various textual indications of news contents were inspected to anticipate upcoming stock value, which was contrasted with linear regression through SVM [13].

3. Sentiment Analysis And RSS news Feeds

Sentiment analysis or opinion analysis states to actions that consequently distinguish an emotion, (for example, a progressive or destructive opinion) from a gathering of words, for example, a sentence or a document. In general, newspaper try to provide objective type of news, yet literary influence investigation in news records displays that numerous words convey progressive or destructive opinion.

Opinion categorization should be possible in word/axiom level, sentence level and text level. Sentiment mining has now a days becoming the leading methodology utilized for mining opinion and evaluations from online sources. Subjectivity investigation partitions the language units into different types: objective and emotional, while sentiment examination endeavors to separate the linguistic items into three classifications; destructive, progressive and impartial.

Three different levels of opinion mining are,

Document level: The sentiment of entire document is recognized as progressive, destructive or unbiased.

Sentence level: The sentiment of each sentence is known as progressive, destructive or impartial.

Entity and Aspect level: Extraction of the object value is identified as positive, negative or neutral

This proposed work automatically investigates the work of findings of proof of news sentiments with the assistance of RSS news Feeds. From the website pages, RSS news feed reader peruses the necessary news information's. This RSS feeds assists with gathering the financial exchange News as a Dataset.

Rich Site Summary (RSS) is an arrangement for conveying consistently, the regularly changing Web content. Numerous news-related websites, Weblogs and other online publishers syndicate their text as a RSS feed to whoever needs it. RSS for the most part, utilizes XML to convey refreshed substance on the Web. RSS feeds sources are broadly and effectively

accessible to the users from different news destinations without visiting the sites separately.

RSS feed identifier reads the necessary news values from the pages for example title, depiction, date, creator, interface and so on in the arrangement of XML. With RSS feed, users can lastly separate needed data from undesirable data. There will be a positive effect, if the RSS news feed is positive. Same way there will be a negative effect if the RSS feed is negative. In general negative affect financial exchanges also the stock exchange to go down.

4. Stock Exchange Forecast

Securities exchange forecast is one of the most problematic works in the economic related markets that are affected by numerous outside social-psychological and financial elements. At first the past prices of the certain company are taken from the site. Different strategies of stock level pointers are accessible to computing the stock money. Some of them are Moving Average, Stochastic RSI (Relative-Strength Index), Bollinger groups, Accumulation – Distribution, Typical Point (pivot point).

Stock Level Indicators

Moving Average Indicator

In this technique, the real index value is coordinated with its normal assumed control over some timeframe with the assistance of technical analysis tool. For stocks, normal timeframes for moving averages are 5 days, 10 days, 15 days, 21 days, 50 days, 100 days and 200 days.

The principle points of interest of moving average stock level indicator provides a flattened line and furthermore assists with cut down the measure of noise on charge plan contrasted with some other level of oscillators. In order to discover the senser movement the moving average value is used.

The moving average is determined by including the final price and afterward partitioning this summative by the number of timespans. Moving average value is shown in the following Equation 1,

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

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 suggested prognostic scheme thought about the calculation of the moving average for 5-days, 10-days and 15-days. The similar examination of the moving average calculation is as per the following.

On the off chance that the 5-day is more prominent than 10-day, and 10-day is more noteworthy than 15-day, at that point it shows that the outcome is positive and the senser goes up for the following day. In the event that 5-day is under 10-day and 10-day is under 15-day, at that point it demonstrates that the

outcome is destructive and the senser drives down for the following day.

In the event that the 5-day is higher than 10-day, and 10-day is under 15-day, and furthermore If the 5-day is under 10-day, and 10-day is greater than 15-day, at that point it shows that the outcome is unbiased and the senser value keeps no variation for the following day.

Moving Average Convergence/Divergence oscillator (MACD).

In this method, distinction between two Exponential Moving Average (EMA)s are performed. In the event that there is a positive sign in MACD, at that point EMA of 12-day is over the EMA of 26-day. This demonstrates the faster moving average value changes its rate greater than slower moving average. In this case, the expansion of positive impact is rising.

In the event that there is a negative sign, at that point EMA of 12-day is underneath EMA of 26-day. On the off chance that it is destructive and decreasing further, at that point the destructive gap between the quicker moving average and the added slower moving average is increasing. Since downward momentum is quickening, this would be viewed as falling. If there is a cross over occurs in the faster moving average then it affects slower moving average which may cause MACD centreline crossovers. The MACD Histogram and Signal Line are shown as follows. The equation for MACD, Signal Line and MACD Histogram are given in Eqs. (2),(3) and (4) individually.

$$MACD = ((12 - day EMA) - (26 - day EMA)) \quad (2)$$

$$Signal Line = (9 - day EMA) of MACD \quad (3)$$

$$MACD Histogram = (MACD - Signal Line) \quad (4)$$

Stochastic Relative Strength Index (RSI)

This is an indicator that computes an incentive somewhere in the range of 0 and 1 which is then plotted as a line. Based on the overbought and oversold conditions indicator is principally used.

The RSI formula is truly direct:

RSI estimation of indicator is defined in equation (5). Also average increase and misfortunes are clarified in Equation. (6) and (7). Equation (8) defined the RS value and the general Stochastic RSI value estimation is appeared in Equation. (9).

$$RSI = 100 - \frac{100}{1 + RS} \quad (5)$$

$$Average Gain = \frac{Total Gains}{n} \quad (6)$$

$$\text{Average Losses} = \frac{\text{Total Losses}}{n} \quad (7)$$

$$RS = \frac{\text{Average Gain}}{\text{Average Losses}} \quad (8)$$

$$\text{StochRSI} = \frac{\text{RSI}(n) - \text{RSI Lowest Low}(n)}{\text{RSI Highest High}(n) - \text{RSI Lowest Low}(n)} \quad (9)$$

RSI = RSI indicator or current level.

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

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

n = calculation of number of periods.

At the point when the Stochastic RSI comes to greater than 0.80, then the outcome demonstrates extra bought. So also when it comes to beneath 0.20, and afterward it demonstrates oversold. In the event that it reaches 0.50 mark in an up direction it is recognized as overvalued circumstance as a purchase sign, and reaches 0.50 mark in a down direction it distinguished as negative circumstance as a trade signal.

So as to identify the stock exchange, the joined results of RSS news feeds and three stock level markers are investigated. Table 1 shows the conclusive outcome forecast procedure for Stock market.

Table 1: Final Result Prediction of Sentiment and Stock level Indicators

Sentiment Analysis Result	Stock level Indicators Result	Final-Result Prediction
Positive	Positive	Positive
Positive	Negative	Neutral
Negative	Positive	Neutral
Negative	Negative	Negative

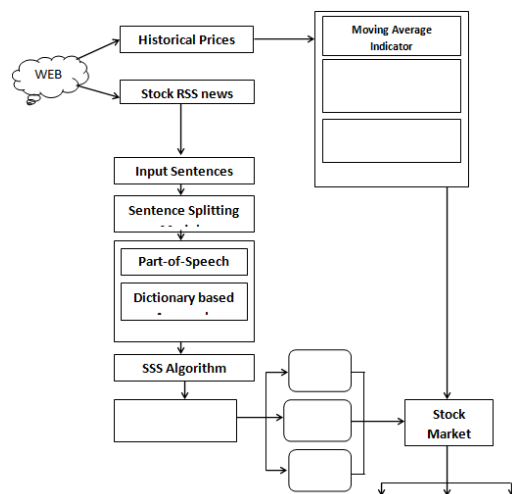


Figure 1. Architecture diagram for Sentiment Analysis for Online Stock market news using RSS Feeds and stock level indicators

In the event that Sentiment investigation results and Stock level Indicators outputs are progressive, at that point Final outcome is also progressive. On the off chance that both are negative, at that point result is additionally negative. Blends of both will results into impartial.

5. proposed work

The framework consequently recognizes the news sentiments with the assistance of RSS newscast feeds and forecasts the financial exchange development whether ascends or down.

The Figure 1 depicts framework for Opinion analysis on financial market newsdata forecasting that utilizing RSS News feeds and stock level indicators.

➤ Stock RSS News Feeds

From the applicable sites, RSS feeder peruses the necessary contents, for example, title, description are in the form of XML. All the news feeds are stored inside this module.

➤ Input Sentences

All the gathered RSS stock news feeds are put away inside the info sentence module all in all record.

➤ Sentence Splitting Module

The news reports and parts are cleaned first and then splitted into parsed sentences using the sentence splitting module. In order to test the data, all text data is collected as a parsed news information. The record contains the RSS news information as sentence by sentence

In order to recognize and remove individual data in main materials Natural language Processing (NLP) is utilized. So as to distinguish the new chances businessmen typically see the audits, appraisals, suggestions and different types of online supposition.

In order to decide the valuation of the words that is having positive/ negative/ impartial sentiment values this module is used. To discover the extremity of the sentence, part-of-speech tagger, is used. Word reference based procedure is used, to identify the sentence is progressive, negative and impartial.

Part-of- Speech Tagger: A POS (Part-Of-Speech) tagger is otherwise called syntactic labeling or word-class disambiguation. It is a bit of programming that peruses message in some language and relegates grammatical features to each word, for example adjective, verb, noun, and so on.,

Dictionary based Approach: In order to discover the sentiment words and their sentiments the dictionary is utilized. Word reference based systems use equivalent words, antonyms and chains of importance in WordNet (or different dictionaries with assumption data) to decide word conclusions. Mathematical scores are divided into three opinions are: Obj(s), Pos(s), and Neg(s) and portraying how Progressive, Objective, and Destructive synset are.

➤ Sentence Polarity

Every sentence polarization data is determined right now. When all is said in done, the score value varies from 0.0 to 1.0 and its whole value is 1.0 for every synset. On the off chance that the polarity identification of particular sentence is progressive, at that point that value of sentence is measured as a positive sentence. On the off chance that polarity data is negative, at that point it is considered as negative sentence. On the off chance that it is 0.0, at that point it is measured as an impartial sentence.

Equation (10) is utilized for the calculation of Sequence of words.

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

(10)

$S(W_{+ve})$ = Set of Positive Sentiment words.

$S(W_{-ve})$ = Set of Negative Sentiment words.

n = Number of words.

Sentence level Sentiment Score (SSS) Algorithm

After recognizing the sentimental value of every sentence, the polarization score data is named progressive assessment, destructive conclusion and nonpartisan. In the event that the assessment is certain, at that location value the stock goes high and if the feeling is negative the stock market goes downcast. In light of the outcomes the financial exchange pattern goes up, down or nonpartisan.

6. Experimental Results

The experimental case study shown that the financial exchange forecasting is gathered for the organization ARBK from Amman Stock Exchange (ASE). For Arab Bank (ARBK) Company, the opinion of RSS news feed is gathered from <http://investing.einnews.com/news/ase-stock> for the time of January 2017 to December 2019.

Opinion of RSS news data feed for Arab Bank (ARBK) Company is gathered from <http://investing.einnews.com/news/ase-stock> and afterward the senser rate for a similar organization is gathered from <http://www.marketstoday.net/markets/jordan/Historical-Prices/10/en/#>. The data set consists of 2006. The outputs of three stock level indicators and RSS news feeds are determined over some stretch of time and the last identification technique is analyzed by utilizing the Table 1. The classification accuracy was shown on the Table 2 which is used for utilizing the proposed strategy and past techniques.

Table 2: Arrangement of accuracy utilizing the proposed technique and past strategies

ARBK Company	ID3	C4.5	Stock Level Indicators	SSS Algorithm
--------------	-----	------	------------------------	---------------

Total Instances	2006	2006	2006	2006
Correctly Classified	887	928	1267	1541
Precision %	44.21	46.26	63.16	76.81

In the proposed method, the calculation of stock level indicators using historical prices is considered over a period of time.

Similar way opinion sentiment value is determined for RSS news feeds over some undefined time frame. Without sentiment value stock level markers creates the precision of 63.16% and with sentiment the stock level pointers produce 76.81% with progress of 13.65%. In the past calculation ID3 and C4.5 the exactness is not exactly the proposed calculation. This work is proficient than the past calculation. Accuracy defines to the closeness of a calculated incentive to a standard or recognized worth. Precision defines to the closeness of at least two estimations to one another. The beneath condition (11) is utilized to ascertain the precision values.

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (11)$$

From the exploratory investigation the proposed framework delivers preferred outcome over the past one.

7. Conclusion

In the prediction of stock market system, the proposed approach builds a predictive model for the sentiments around the RSS stock news feeds. From Arab Bank (ARBK) Company, the real time RSS news feeds are collected and used to predict the sentiment scores which is positive, negative or neutral. This SSS algorithm is a useful tool for the stock investors to precede correct decision along with the stock level indicators. Finally combined results if sentiment and stock level indicators provide a productive outcome to the stock investors when to purchase or wholesale the stocks values. In future this work can be carried to evaluate the negation word, valence shifters what's more with supposition news which may improve the exactness.

References

- [1] 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.
- [2] Shri Bharathi and Angeline Geetha. "Determination of News Biasedness using Content Sentiment Analysis Algorithm," Indonesian Journal of Electrical Engineering and Computer Science, 16(2), 882 – 889, 2019.

- [3] M. Karamibekr, A. A. Ghorbani, "Sentence Subjectivity Analysis in Social Domains," In the Proceedings of IEEE/WIC/ACM International Conf. On Web Intelligence (WI) and Intelligent Agent Technology (IAT), pp.268- 275, 2013.
- [4] Xi Zhang, Jiawei Shi, Di Wang and Binxing Fang, "Exploiting Investors Social Network for Stock Prediction in China's Market", Journal of Computational Science, pp. 1-30, 2018.
- [5] X. Li, H. Xie, Y. Song, S. Zhu, and 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, Jan 2015.
- [6] B. Li, K. CC Chan and C. Ou, "Public sentiment analysis in Twitter data for prediction of a company's stock price movements," 11th International Conference On e-Business Engineering (ICEBE), pp. 232-239, 2014.
- [7] Q. A. Al-radaideh, A. A. Assaf, E.Alnagi, "Predicting Stock Prices using Data Mining Techniques," In the proceedings of International Arab Conference on Information Technology, pp. 1-8, 2013.
- [8] C. Wong; I. KO , "Predictive Power of Public Emotions as Extracted from Daily News Articles on the Movements of Stock Market Indices," IEEE/WIC/ACM International Conference On Web Intelligence (WI), Oct 2016.
- [9] A.Chakraborty, D.Mukherjee, A.Dutta, D.K.Kole, "An Effective Stock Price Prediction Technique Using Hybrid Adaptive Neuro Fuzzy Inference System Based on Grid Partitioning," Journal on Advances in Intelligent Systems and Computing, vol 385. Springer, 2016.
- [10] X. Xu, X.Cheng , S. Tan ,Y. Liu , H. Sh, "Aspect-level opinion mining of online customer reviews," China Communications. Vol. 10, no.3, pp. 25–41, 2013.
- [11] Shri Bharathi, Angelina Geetha, "Sentiment Analysis of Twitter and RSS News Feeds and Its Impact on Stock Market Prediction", International Journal of Intelligent Engineering and Systems, Vol.10, No.6, pp.68- 77, 2017.
- [12] 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.
- [13] P.Robert, Schumaker, and H.Chen, "Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZF in Text System", In the Proceedings of 12th American Conference on Information Systems, 2009.
- [14] Shri Bharathi and Angeline Geetha, "Sentiment Analysis for Effective Stock Market Prediction", International Journal of Intelligent Engineering and Systems, Vol.10, No.3, pp. 146- 154, 2017.