website: www.ijoetr.com

International Journal of Emerging Trends in Research

STOCK MARKET PREDICATION USING SENTIMENT ANALYSIS

Gaurav J.Sawale[1] *, Dr Manoj K.Rawat[2]

[1] Department of Computer Science & Engineering, Department of Computer Science & Engineering,
  SAGE University Indore, Madhya Pradesh, India,

[2] Department of Computer Science & Engineering, Department of Computer Science & Engineering,
  SAGE University Indore, Madhya Pradesh, India,

Abstract

 The goal of stock market forecasters is to develop a practical approach for predicting stock values. It is crucial to obtain the most significant outcomes and reduce erroneous stock price forecasts. This project aims to design and build a predictive system for stock market investing advice. Our method is unique in that it anticipates efficiently using Sensex data as well as Simple Syndication (SIGNAL VALUE) feeds. We argue that stock market prices are influenced by sentiment analysis of SIGNAL VALUE news streams. As a result, data from SIGNAL VALUE news feeds are combined with stock market investment data over time. Our sentiment analysis technology created the link between stock market prices and sentiments in SIGNAL VALUE news feeds. This is used to forecast stock market rates. The Amman Stock Exchange provided stock market values, and SIGNAL VALUE news feeds for the company RABK, which we used in our experiment (ASE).
Keywords: Stock market, dataset, Sensex facts, Sentiment mining
Keywords: factors analysis; facility layout; evaluation; selection;

1. Introduction

 Most investors and financial professionals have long been fascinated by stock price forecasting. All forecast of stock prices is a complex subject in finance, manufacturing, and mathematics. Due to its monetary success, it has garnered considerable interest from academic and corporate communities. Nonetheless, determining the optimal moment to purchase or sell has remained a challenging challenge for investors due to various variables that might affect stock prices.

Investors purchase and sell their items based on the joint outcome of phrase opinion and Sensex facts of moving average. A unique technique is given for predicting the buy or sell indication to stockholders in the prediction of stock market analysis. Stock market prediction is accomplished using a combination of stock-specific SIGNAL VALUE news feeds and Sensex points in the suggested technique.

Sentiment classification has grown in popularity as a problem in natural language processing, intending to forecast sentiment (view, emotion, etc.) from texts.

Nowadays, individuals post comments and ideas on social media platforms that others may share. Sentiment examination has /surpassed all other techniques for obtaining sentiment and evaluations from internet causes. Classification of sentiments may be accomplished at the word/phrase, sentence, or document level. Subjectivity examination classifies language components into two groups: objective and subjective, while sentiment examination classifies them into three groups: negative, positive, and neutral.

In the suggested system, two kinds of outcomes are obtained. The first is computed without considering sentiment, whereas the second is calculated with consideration of emotion. Both findings indicate that stock market prediction has improved.

2. Related Work

Numerous data mining methods for stock price prediction have been suggested. Based on the opinion sentences, the opinion summary is examined. This document provides solely factual information and is thus unsuitable for summary opinion [1]. To reduce the need for sophisticated spatial modeling of several sources, tensor-based information architecture for stock forecasting was created [2]. While this concept applies to various multidimensional learning issues, it does not address temporal interaction.

The combination of Self-Organizing Maps (SOM) and fuzzy – Support Vector Machines has been suggested as a model for predicting stock prices (f- SVM). This article presented a framework for extracting fuzzy rules from raw data using statistical machine learning models together [3].

Daily text material on Twitter was examined using mood monitoring methods, and a self-organizing fuzzy neural network was used to forecast changes in the DJIA's closing values [4]. This is comparable to how we monitor our moods. We employed a back proliferation neuronal network and methodological indications to anticipate fuzzy time series. The work's results indicated that ANNs were superior to time series models in predicting capability [5]. It improves signifying accuracy for known patterns but requires more significant effort for novel designs.

[6] The subjective assessments of social problems are the focus of this study, and subjectivity is very dependent on the phrases in a text. They planned a lexical-syntactical method for recognizing and classifying subjectivity at the judgment level and the part of numerous view terms, particularly verbs, in expressing views about social matters and the importance of focusing on the weakness and strength of objective sentences.

The backpropagation technique is used in this case, along with time and profit-based weight issues. They performed prediction using a feed-forward neuronal network and a primary, recurrent one [7].

Chinese stock market forecasting using public sentiment data gathered from microblog feeds. This article introduces a unique stock collection model using discrete and ongoing variables. It is necessary to include investment risk and other capital markets to test its generalizations further.

Different textual representations of news stories were compared to linear regression with SVM. Other machine learning approaches, such as Relevance Vector Regression, must be studied to improve accuracy.

Provides excellent performance in nonlinear forecasting neural networks. Complex metaphors of the contributing factors were provided to forecast the stock market index price. Additionally, they compared the predictions produced by several Artificial Neural Network (ANN) models.

A strategy for automatically distinguishing previous from contextual polarity was presented, emphasizing comprehending the characteristics. Extensive vocabulary expansion enhances prediction by collecting the prior polarity of words and sentences.

This article examines a general framework for stock price forecast that accepts textual credentials as inputs and produces projected price changes. Daily full-length news item summaries are analyzed using a sentence-level summarization technique for stock price prediction. Additional enhancements to the present processing stages, such as applying sentiment examination in the news pre-processing step to generate higher-level structures, are required for future advancement.

A decision tree classifier, a kind of data mining tool, is used in the stock market to decide whether to purchase or sell stocks based on information derived from past stock prices. It is necessary to evaluate the elements that influence the stock markets' behavior to enhance the forecast, such as trading volume, news, and financial data.

It is abundantly clear from previous studies that sentiment analysis aids in prediction improvement [4, 6]. Our research considers the emotions expressed in newsfeeds to forecast stock market swings more accurately.

Additionally, we used Moving Average Stock Level Pointers in conjunction with sentiment mining to improve the accuracy.

3. Stock market forecast

The Sensex will either rise or decline. In most cases, the Sensex is determined.

Financial market information is accessible via various media outlets, including websites, Twitter, Facebook, and blogs. Generally, the price of a stock is determined by two variables, and there are two types of factors: fundamental and technical.

A company's statistics data mainly determine the critical component. It contains reports, the company's financial position, balance sheets, bonuses, and strategies of the corporations whose stock is to be monitored. The practical element consists of quantitative factors such as trend pointers, daily highs and lows, high and low values, stock volume, catalogs, and put/call ratios. In the technical issue, the ancient prices.

The market capitalization of the free float may be calculated using Eq (1).

$Free\ Float\ Market\ capitalization =$
$Share\ pric(shares\ out\ standing - locked\ in\ share)$ $\qquad$ (1)

The subsequent Eq. (2) is used to calculate the moving average for forecasting purposes. To begin, the website is queried for historical pricing for the specified firm. Various methods of

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

3.1 Moving average technique as stock level indicators

It is a technical examination tool that compares current index data to its average over a specified time. Moving averages are classified into three basic categories

The primary benefit of the moving average stock level pointer is that it provides a curved line and reduces noise on the price chart linked to other level pointers.

The Sensex is intended to capture the market mood in its entirety. It comprises 30 well-known equities spanning all major industries and is regularly traded. These are huge, well-established, and financially strong businesses operating in various sectors. The Sensex data is then subjected to a moving average to see if the next day's

F_t = prediction for the coming period,

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

N = number of periods to be averaged.

The suggested forecasting method used five-day, ten-day, and fifteen-day moving average computations.

The moving average calculation is compared in the following manner.

The outcome is impartial if the 5-day period is larger than the ten days and the ten days is less than the 15 days, or if the 5-day period is less than the ten days and the ten days is greater than the 15 days.

4. Proposed work

The system investigates the problem of independent sentiment analysis using SIGNAL VALUE news feeds and forecasts whether the stock market will rise or fall. The procedure for stock market prediction and analysis utilizing SIGNAL VALUE news feeds is shown in Fig. 1.
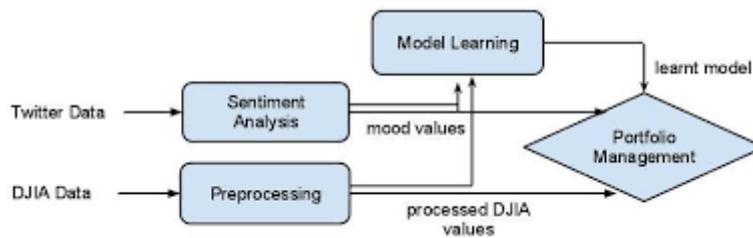
Figure 01: of Sentiment analysis for stock market prediction

4.1 SIGNAL VALUE stock news feed

SIGNAL VALUE feed readers extract needed news material from websites, including title, explanation, date, author, and link, in the form of an XML text. This pulls the most recent headlines from a financial news website. After configuring the site parameter, the final description information (sentence level) is fetched. Finally, it analyses the SIGNAL VALUE feed list's XML document. This SIGNAL VALUE feed contributes to creating a dataset of stock market news.

4.2 Pre-processing

This procedure eliminates inaccurate, insufficient, incorrectly formatted, or duplicated data. Contaminated data might contribute to data set misunderstanding. As a result, this unit cleanses the information by completing missing standards, smoothing noisy input, and finding and deleting outliers. The information is then sent to the following module after pre-processing.

4.3 Sentence splitting component

The sentence splitting component is in charge of parsing the cleaned news data.

The parsed news data is compiled into a text document to facilitate testing. The paper offers sentence-by-sentence summaries of SIGNAL VALUE news feeds.

4.4 Natural language processing (NLP) module

Sentiment examination, often known as view mining, is a technique for determining an individual's attitude, opinion, or sentiment toward a given subject. It identifies and extracts personal information from source materials through natural language processing (NLP) and text analytics.

Businesspeople often monitor reviews, ratings, suggestions, and other types of internet opinion to uncover new possibilities and protect their reputations.

This enables the identification of words indicative of emotion and the identification of their connections, which aids incorrect identification.

The emotion of words with a positive/negative/neutral connotation is determined using a scaling mechanism polarity of a sentence is determined by using a part-of-speech tagging tool, a dictionary-based technique, and the Sentence Sentiment Score (SSS) Procedure.

Positive phrases convey a positive emotion, such as joy, zeal, or benevolence. They are often regarded as having a favorable disposition. Likewise, negative statements are associated with negative emotions such as despair, hatred, aggression, and prejudice. They are often considered harmful, and they are classed as neutral when no feelings are suggested.

4.4.1 Dictionary-based method

Amongst the several unsupervised techniques obtainable, a dictionary-based method is utilized to identify the alignment of judgments. The vocabulary is used to ascertain the polarity of view terms. Dictionary-View publication stats

based approaches assess the feelings of words by examining substitutes, antonyms, and orders in WordNet (or other lexicons that provide sentiment data).

This technique categorizes each WordNet synset into three sentiment arithmetical scores: Obj(s), Pos(s), and Neg(s), indicating the degree to which the items in the synset are Objective, Positive, or Negative.

4.4.2 Sentence Score procedure for accessing data

The SSS Procedure is used to determine the overall consequence. The analytic technique is done to each phrase, and the findings are then summed to provide the document's overall result.

In general, the score for each synset is between 0.0 and 1.0, with a cumulative of 1.0. Initially, the POS tagger is helpful to each term, which describes the tag as a noun or adverb equal to words. Calculate the total of a sentence by assigning a score value to each word. When a sentence's score value is positive, it is favorable. If the value of the score is negative, the penalty is deemed harmful. If the value is 0, the statement is regarded as neutral.

4.4.3 Moving average stock pointer component

This component takes ancient values from the website and produces the moving average price of those values to assist stockholders in determining whether to purchase or sell their shares.

4.4.4 Stock market forecast

The findings of sentiment examination and Sensex are joint and analyzed to forecast the stock market. Table 1 illustrates the stock market's outcome prediction approach.

Sentiment Analysis Result	Sensex-Moving Average Result	Final-Result Prediction
Positive	Positive	Positive
Positive	Negative	Neutral
Negative	Positive	Neutral
Negative	Negative	Negative

Table 1. Sentiment and Sensex-Moving Average Final Result Prediction

The final forecast is also optimistic if the sentiment examination findings and the Sensex Moving Average are promising. If both are negative, the outcome will be wrong, and both of these combinations will result in neutrality

5. Experimental results

Stock market forecasts for the business RABK from the Amman Stock Exchange are gathered in this experimental research (ASE). The suggested algorithm's performance is compared to data mining approaches used to forecast the stock market [16]. The Amman Stock Exchange's (ASE) Oracle database provides historical pricing for each of the exchange's 230 listed businesses dating back to 2000. From 2005 to 2007, historical prices were compiled.

The moving average is calculated for April 2006. In the same way, April 2006 emotion is calculated. Finally, the observations are combined, and the numbers in Table 1 are used to make a stock market forecast. For sentiment analysis, the month and year are essential. The positive and negative terms in the News document are depicted in the accompanying Fig.2, displaying excellent news outcomes.

The moving average is used to analyze the historical data for Sensex. Fig 2 shows the 5-day, 10-day, and 15-day moving averages are computed for the Sensex point. The calculation of the moving average follows this.

In the graph below, five-day moving averages ten-day moving averages are exposed in blue, red, and green, respectively.
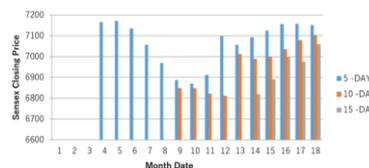

Figure 02: Result analysis of Sensex and stock market

If the 5-day moving average's Sensex points are more significant than the 10-day moving average's Sensex facts and the 10-day moving average's Sensex points are higher than the 15-day moving average's Sensex points, the outcome is positive. If the product is favorable, the stock will be favorable as well.

The table compares the proposed strategy to existing methods regarding categorization accuracy. The Sensex point moving average for 499 total cases is calculated using the suggested approach.

The sentimentality polarity value is derived similarly for the sentimentality news information from April 2006. Without sentimentality, the moving average achieves 64.32 percent accuracy, whereas, with the sentiment, the moving average achieves 78.75 percent accuracy, an increase of 14.43 percent. The suggested approach outperforms the earlier algorithms ID3 and C4.5 inaccuracy. This demonstrates how more efficient the proposed method is than the existing one.

Exactness and recall both contribute to the measure's accuracy. The ratio of true positives to all recovered examples is precision, whereas true positives to all positive samples are recalled.

Accuracy is determined by the number of accurate documents returned, and the precision analysis result picks prediction. The accuracy and recall values for instances are calculated using Equations (3).

Our experimental findings indicate that using Moving Average, ID3, and C4.5 techniques improves benchmarking accuracy and correctness metrics. Our future work will increase accuracy by combining one or more stock level pointers with stock news from SIGNAL VALUE feeds. Additionally, it is designed to include social media material like Twitter and Facebook to provide more accurate predictions.

$Recall = TP$

$(TP+FN)$ (3)

The proposed system gives a better outcome than the existing one, according to the above results.

6. Conclusion

Unlike traditional stock market prediction systems, our revolutionary approach blends ordinary people's sentiments with news feeds and Sensex data to forecast stock market behavior. The sentiment polarity of the news sentences is calculated for predicting stock news, whether positive, negative, or neutral, using the SIGNAL VALUE news feed of stock-related news for the RABK company. The stock market's Sensex points are calculated using the moving average stock level indicator. Finally, the stock market forecasters can use a combination of sentiment polarity news and Sensex points to determine when to buy or sell their stocks

References

[1] A. Balahur, MijailKabadjov, J. Steinberger, R. Steinberger, A. Montoyo, "Challenges and Solutions in the Opinion Summarization," Journal of Intelligent Information Systems, Springer, Vol.39, No.2, pp. 375-398, 2012.

[2] Q. Li, Y. Chen, L. L. Jiang and 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.

[3] D. Nguyen and M. Le, "A Two-Stage Architecture for Stock Price forecasting by combining SOM and fuzzy-SVM", International Journal of Computer Science and Information Security, Vol. 12, No. 8, pp. 1-6, 2014.

[4] B. Johan, H. Mao and X. Zeng, " Twitter Mood
Predicts The Stock Market", Journal of Computational Science, Vol.2, No.1, pp.1-8, 2011.

[5] H. Kunhuang & T.H.K Yu, "The Application of Neural Networks to Forecast Fuzzy Time Series", Physical A: Statistical Mechanics and Its Applications", Vol. 363, No.2, pp. 481-491, 2006.

[6] M. Karamibekr, A. A. Ghorbani, "Sentence Subjectivity Analysis in Social Domains" In: Proc. of IEEE/WIC/ACM International Conf. On Web Intelligence (WI) and Intelligent Agent Technology (IAT), Vol. 3, No. 1, pp.268- 275, 2013.

[7] N. L. D. Khoa, K. Sakakibara and I. Nishikawa, "Stock Price Forecasting using Back Propagation Neural Networks with Time and Profit Based Adjusted Weight Factors", In: Proc. of SICE-ICASE International Joint Conf. in Bexco, Busan, Korea, Vol.1, No.1, pp. 5484 - 5488, 2006.

[8] 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.