ISSN 2278-3091
Volume 8, No.4, July – August 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse10842019.pdf
https://doi.org/10.30534/ijatcse/2019/10842019

A New sentiment score based improved Bayesian networks for real-time intraday stock trend classification

Lakshmana Phaneendra Maguluri[1], R Ragupathy[2]
[1]Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamil Nadu 608002, India phanendra51@gmail.com
[2] Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamil Nadu 608002, India. cse_ragu@yahoo.com

ABSTRACT

The stock sentiment score is an important trend indicator for an enterprise and several factors may affect its value. Various events such as stock news, quarterly results and company announcements may differ in public emotion and sentiments, which may affect stock market trend. Due to variation in stock trend and news, stock market forecasting remains a challenging task in real-time market. However, most of the traditional sentiment analysis approaches are not directly predicting the sentiment score and its trend on the online stock market sites due to data sparsity and uncertainty problems. In order to overcome these problems, a hybrid probabilistic based stock sentiment prediction model is designed and implemented based on the real-time market data. In the proposed model, real-time stock data such as stocks technical, stocks news, stock momentum etc. are integrated to predict the buying or selling behavior of the investor. A novel stock sentiment score and data transformation methods are implemented to normalize the market trend for intraday data. Finally, an improved Bayesian networks technique is implemented on the filtered data to predict and test the trend of the stock on the filtered data. Experimental results show that the proposed framework is more efficient than the traditional sentiment classification models in terms of accuracy and error rate are concerned.

Key words: Stock market prediction, Bayesian networks, sentiment prediction, trend analysis.

1. INTRODUCTION

The stock market is considered as the most vital and active part of financial institutions and investors. The financial news articles and trend data are considered as the primary source for market trend prediction. Most of the organizations depend on the high computational systems in order to predict the market trend, based on the sentiment score and stock technical data. These predictions are used to filter positive and negative sentiment stocks and to take appropriate decisions by the investors. Hence, the modeling and analysis process of news articles are very much essential in order to make accurate predictions. The organizations can become market prominent, if they can attract the attention of more and more investors. Let us consider an example here, if an investor is having 100 stocks and it is not feasible to filter top positive and negative trend stocks based on the company announcements and technical data.

Also, the implementation of the sentiment analysis using the technical data and stock news is very expensive and time-consuming on real-time intraday market data. In this case, all of the public opinions and technical data are required to be monitored continuously from time to time. The social media platforms are widely used for idea exchange and expressing opinions. The process of stock price prediction is very vital during any business planning activities. It is very much difficult to construct an appropriate stock prediction model. Also, the stock market can be greatly influenced by the moods of society. The social mood for a particular organization is considered as the most vital variable that can influence the stock price. Due to the advancements in social networking, large volumes of sentiment data are available now. Hence, extracting essential information in the social media along with the market trend can enhance the predictive capability of models. In the last two decades, there exist various global stock crashes. The cause for these crashes can be given below:

1. The public market information and macroeconomic conditions are not capable enough to explain the unwanted slumps.
2. Modifications of stock market are unbalanced and asymmetric in nature. The market crash is possible within a short period of time. On the other hand, the increase of the actual stock price may require long time period.

The social media are capable to burst unstructured opinion content which is very important. Due to the variation in stock volume and velocity, it is very much complicated task to predict the stock market trend by the market analysts. The process of sentiment analysis is to predict the stock market trend along with individual stock trends by using supervised classifiers such as Support Vector Machines, Naive Bayes, ensembles, etc. The sentiment lexicons allow the generation of vital features in case of supervised sentiment analysis process. Basically, the sentiment lexicon is a list of words having a specific sentiment value. This value can be either positive or negative. Let us consider an example including a sentence “This dress is great”. This example is classified as positive category. Here, the term “great” signifies a positive value.

The sentiments present within financial reports or news articles can influence the stock return to a great extent. In order to determine the relationship among web media and stock market, supervised classification models are developed along with statistical analysis. Usually the correlation can be obtained through representing online financial news with the technical data. A bag-of-words technique is used to process the stock news articles and trend data by using the vector space model. Some of the bag-of-words techniques have better predictive capability when compared to other traditional static dictionary-based approaches. Most of the news articles are interpreted with the help of investors. After that, these are translated to market sentiments. The investors have the responsibility to take certain important decisions depending upon the sentiment interpretations.

According to a recent report [1], macroeconomic news is capable to explain one-third of the variance in stock returns. Traditional classification models use the financial news to restrict the short-term predictive capability on future stock prices. In these models, emotions can affect the investment decisions significantly. Hence, the combination of sentiment analysis and prediction framework is very complicated and time-consuming process. The support vector machine algorithm is implemented in order to resolve the classification issues. Kernel functions play a vital role during the prediction process and it also allow different users to operate on large feature sets. Different versions of kernel approaches are introduced to implement various kinds of data structure. Let us consider two examples of string kernel and graph kernel. In case of string SVM kernel, it permits support vector machine algorithm along with strings. In this case, these strings are not converted to fixed-length, real-valued feature vectors. On the other hand, graph kernel approaches permit support vector machine algorithm in order to work with graphs. According to different studies [2][3], financial news can influence the stock price significantly. The earnings announcement can give rise to high volatility in the stock price.

Intraday stocks are negotiated in stock market and most of the market investor has the responsibility to purchase and sell their products or services. The stocks which are directly associated with the indices are listed for intraday trading purpose. On the other hand, a negative direction indicates that, the closing value of the previous day is much higher as compared to the current value.

Now-a-days, the process of sentiment analysis is implemented in order to forecast the stock market variables. The micro blogging data are considered as the most important source to support stock market decisions. It also permits the real-time assessment process. The lexicons generated for financial and micro blogging domains are built easily. Some of the financial micro blogs use static trained lexicons for prediction process. Implementing a manual technique is not preferable practically. These lexicons are mostly not effective in real-time stock market prediction. The word “explosive” is mostly considered as the negative one, but in case of financial messages it is considered as positive.

The main contribution of the work is to find the sentimental based stock classification in order to improve the buying behavior of investors on the Futures market. Sentiment analysis reveals the effect of unstructured market data on investor emotions for decision making. The market feelings or the purchasing behavior of the merchants is based on the stock technical and the market trend.

2. RELATED WORKS

S. Akhtar et.al, focused on the stock salience and the asymmetric market influence of consumers’ sentiments [1]. They have included asymmetric announcement influences of various customers’ sentiments. This model predicts the sentiment of the market based on the stock review data.

S. Basak et.al, tried to predict the flow of stock market prices with the help of tree-based classification algorithms [2]. Forecasting and diffusion modeling are considered as the major cause for numbers of different issues. Since, decreasing forecasting errors will directly influence the investment risk. In this work, they have considered a direction-predicting model. They have introduced an advanced framework in order to resolve the classification issue. Two different approaches known as gradient boosted decision trees and random forests are implemented in this work to classify the market data.

B. Li et.al, focused to detect public sentiment of the public companies using the market trend data [3]. Due to the popularity of various social media websites, both academicians and scientists have concentrated their research interests on the possibility of mining social media data during the analysis phase of public sentiments. According to the numbers of different research methodologies, public emotions described via Twitter will be helpful to correlate with the Dow Jones Industrial Average. There exists no clear mechanism that will explain how the public sentiment can be reflected in social media. It is very much essential to predict stock price movement of a specific organization. They implemented a new approach known as SMeDA-SA and this approach has the responsibility to mine Twitter data for the process of sentiment analysis. After that, the stock movement of certain organizations can be predicted. Initially, this algorithm extracts various ambiguous textual messages out of the tweets in order to build a list of words. After that, an efficient data mining technique is implemented in order to extend the word list through inclusion of emotional phrases.

S. Kim and D. Kim emphasized on investors’ sentiment on micro blog postings and the prediction of stock trend [4]. They have considered the investors’ sentiments from the year January 2005 to December 2010. They have considered two important mechanisms; those are inter-temporal and cross-sectional regression analyses. Here, the investors’ sentiment is incapable to predict future stock returns. On the other hand, the investors' sentiment is positively influenced through the prior stock price performance. Apart from this, the investors' sentiments from internet postings can predict volatility and trading volume. An efficient machine learning approach can be implemented for the process of classification. Further research efforts are required to extend and improvise the above model.

Q. Li et.al, focused on the influences of news and public sentiment on stock movements [5]. With the rapid advancements of technology, investors can get more useful information at the appropriate time. Here, the investor's decisions can be affected by peer to public emotions. In this research paper, they have introduced a quantitative media-aware trading scheme in order to analyze the influences of media on stock markets. The major objectives of this research work include:

1. Basic information related to organization-specific articles are used to enhance the investors' knowledge. It can also influence the trading patterns of investors.
2. The influence of media may vary according to the organizations' behavior and contents of the articles.

X. Li et.al, focused on the influence of news on stock price return with the help of sentiment analysis process [6]. The financial news articles are mostly influenced on the stock price return. Various textual news articles are measured and projected onto the sentiment space. Apart from this, instance labeling approach is discussed and evaluated. Following are the main problems of this approach:

1. Traditional sentiment classification models use sentiment polarity which is inefficient for predictions.
2. The models having two sentiment dictionaries for filtering and prediction.

K. Li performed a survey on Xiong'an new area scheme [7]. They have considered the Chinese stock market and the influence of news on stock market in their work. In this paper, they have analyzed the influence of various news on stock market categorized through different aspects. They have considered the on Xiong'an new area scheme in their work. They have gathered Xiong'an-related news and in the subsequent step, these are classified according to various aspects. The positive news is capable enough to enhance the performance of the stock market significantly. On the contrary, the negative news is the major cause of performance degradation. News related to government based sources can create much better influences as compared to the news from various other sources. Apart from the above, we can mention here that, news having viewpoints from academia may create greater influence as compared to news from government and other organizations.

W. Long et.al, introduced an advanced graphic kernel approach of stock price trend prediction which depends upon financial news semantic and structural similarity [8]. Numbers of different researchers have tried to develop an efficient and effective approach in order to predict the stock price movement with the help of a support vector machine algorithm. Most of these previously developed approaches emphasized on the news contents. On the other hand, few numbers of approaches include the concepts of information hiding. The relationships between news are also studied and analysed. In this work, they have presented an advanced kernel based model that depends upon the concepts of support vector machine. This approach is far better as compared to the linear kernel approach in terms of performance.

L. S. Malagrino et.al, introduced a new prediction strategy for stock market index intraday direction [9]. This approach includes the basic concepts of Bayesian network technique. They have considered two different network topologies for their experiment along with various stock indices. This approach is efficient enough to provide the next day closing direction.

T. H. Nguyen et.al, focused on the process of sentiment analysis on social media for stock market future prediction [10]. In this model, stock trend data and sentiment scores are automatically retrieved from the micro blog sites to predict the stock trend. In the subsequent time, additional research works can be performed in order to improve the overall performance.

Z. Ni, D. Wang et.al, concentrated on investors' sentiment and its nonlinear effect on stock returns [11]. Its influence is asymmetric and reversible in nature. In case of high returns in short term, it is positive as well as large. On the contrary, in case of long term, it is negative and small.

N. Oliveira et.al, emphasized on stock market sentiment lexicon acquisition with the help of micro-blogging data and statistical measures [12]. Lexicon acquisition is considered as a major problem during the process of sentiment analysis. In this work, they have introduced a new and fast technique in order to construct stock market lexicons. They have also implemented a lexicon in order to generate Twitter investor sentiment indicators. Due to the lower stock trend indicators and higher frequencies, the new technique is dependent on classical trend indicators.

J. L. Hui et.al, studied and analyzed influences of world class and text position in sentiment-based news classification process [13]. The sentiments gathered in the form of emotions to predict the feedback of reactions. These reactions are considered as the most important indicators for the social and political advancements. They have tried to automate the classification process of news texts according to the indicators. Most of the review texts include explicit words that can be interpreted directly during the process of sentiment classification. Generally, news texts include various facts and figures. In this work, a study is performed in order to analyze and identify the relevant key parts of news contents. It is very much essential during the sentiment-based classification process. There are two major factors those can affect the training of classifier; those are text part of speech and text position.

R. P. Schumaker et.al, emphasized during the evaluation of sentiment in financial news articles [14]. It is basically a financial news article forecasting system. This system is integrated with a sentiment analysis tool for better functionality. In future, further research works can be performed in order to modify and extend the above model.

S. Ik Seok et.al, designed a new investor-based sentiment framework to predict the stock market prediction on stock events or announcements [15]. The stock price reaction to positive ratings is more in case of organizations having very high sentiments. In other words, all the investors are more interested about the cash flows of good earning news in case of organizations having high sentiments. On the contrary, the stock price sensitivity is quite higher in case of organizations having low sentiments. The above scenario is actually insignificant due to the fact that, investors mostly never update bad earning information irrespective of the sentiment level. Additional research works can be performed in order to modify and extend the above presented approach.

Bayes' theorem describes the posterior or conditional probability of a hypothesis (S) based on prior knowledge of evidence (E) that might be related to the hypothesis. The posterior p(S|E) of H given e is defined as:

$$p(S,E) = \frac{p(S,E)}{p(E)}$$

$$p(S,E) = p(S|E).p(E) \text{ and } P(E,S) = p(E|S).p(S)$$

$$p(S|E) = \frac{p(E|S).p(S)}{p(E)}$$

$$p(S|E) = \frac{p(E|S).p(S)}{\sum_{H^*} p(E|S^*).p(S^*)} \quad (1)$$

This is the Bayes' rule and lies at the core of Bayesian inference whereas S^* in the denominator is a variable that takes on all possible hypotheses. Bayesian model averaging (BMA) is an application of Bayesian inferential analysis. It has been applied to model selection problems, where one combines estimation and prediction to produce a straightforward model choice criteria and less risky predictions. Suppose Ml is one of a set of models considered to fit the research question, Δ is the interested parameter, D is the dataset given, then the BMA-averaged Δ is the sum of specific model derived Δ_l weighted by the posterior model probability $(M_l|D)$

$$E(\Delta|D) = \sum_{l=1}^{k} \Delta_l p(M_l|D) \quad (2)$$

3. PROPOSED MODEL

In the proposed framework, a novel filter based Bayesian estimation model is designed and implemented to predict the market trend on the real-time market data. The proposed model is implemented in three phases as shown in Figure 1. In the first phase, real-time market data are taken from the stock market sites such as NSE, Upstox Wall mine etc. In the proposed model, stock data are captured from the wall mine and Upstox websites to filter the data and to find the stock sentiment score. Stock related technical factors such as symbol, price, ADX, ADR, RSI, MACD, news sentiment score, etc. are used as the training data. In the second phase, a novel stock sentiment score is computed on the captured stock related fields. Here, the sentiment score is computed on the intraday stock performance field and stock technical factors such as volatility field and the RSI. In this phase, a novel filtering method is implemented to normalize the data for the stock sentiment classification process. In the third phase, an improved Bayesian network technique is proposed on the normalized data to predict the accuracy of the stock trend on the intraday real-time data.

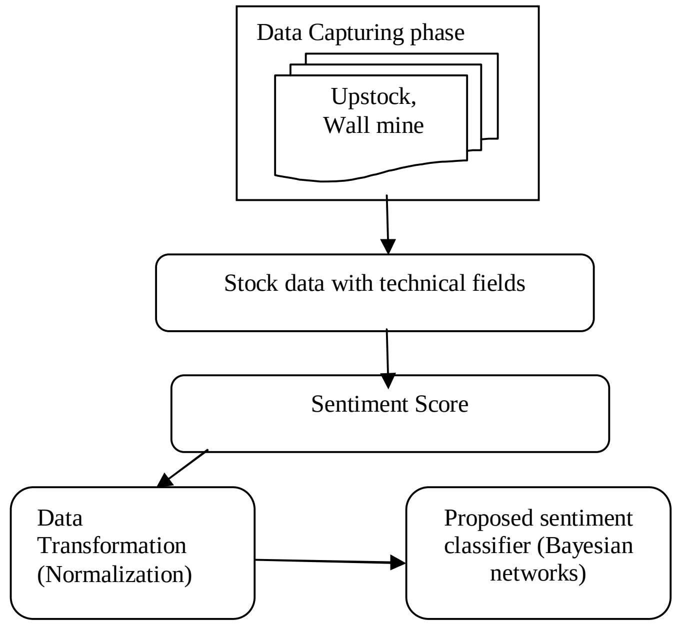

Figure 1: Proposed stock sentiment classification framework.

Real-time data is captured from the wall mine website with different stock technical factors. The sample stock market data along with technical factors are shown in Table 1.

Table 1: The sample stock market data along with technical factors.

Symbol	Price in Rs.	52 weeks high	Performance today
TCS	2,164.00	Rs.2,276 +5.17%	+0.42%
HINDUNILVR	1,738.65	Rs.1,870 +7.53%	-0.32%
ITC	300.55	Rs.323 +7.45%	-0.94%
HDFC	1,949.90	Rs.2,073 +6.31%	+0.79%
INFY	735.10	Rs.773 +5.16%	+0.95%
SBIN	305.40	Rs.340 +11.21%	+0.05%
KOTAKBANK	1,352.90	Rs.1,417 +4.74%	-0.18%
ICICIBANK	395.75	Rs.411 +3.85%	-0.19%
MARUTI	6,960.15	Rs.9,929 +42.65%	-1.26%
ONGC	166.95	Rs.192 +15.00%	+1.95%
AXISBANK	743.95	Rs.788 +5.87%	-1.23%
WIPRO	291.35	Rs.294 +1.01%	+0.09%
BAJFINANCE	3,059.35	Rs.3,130 +2.31%	+0.80%
COALINDIA	254.05	Rs.303 +19.44%	-0.57%
HCLTECH	1,125.30	Rs.1,133 +0.71%	+2.11%
ASIANPAINT	1,451.95	Rs.1,530 +5.37%	+1.46%
IOC	152.90	Rs.177 +15.89%	+2.34%

Symbol	Price in Rs.	52 weeks high	Performance today
NTPC	133.35	Rs.147 +10.11%	-0.22%
BHARTIARTL	322.30	Rs.392 +21.76%	+1.83%
HINDZINC	273.90	Rs.334 +22.03%	-0.58%
ULTRCEMCO	4,131.95	Rs.4,494 +8.75%	-1.58%
SUNPHARMA	464.35	Rs.679 +46.29%	-0.90%
POWERGRID	190.60	Rs.216 +13.33%	-1.24%
TITAN	1,140.45	Rs.1,152 +1.01%	+0.96%
INDUSINDBK	1,686.95	Rs.2,038 +20.81%	+2.16%
BAJAJ-AUTO	3,040.10	Rs.3,195 +5.09%	+0.03%
GAIL	345.25	Rs.399 +15.68%	+1.71%
TATAMOTORS	219.45	Rs.352 +60.24%	-5.27%
BPCL	344.85	Rs.429 +24.45%	+2.53%
M&M	661.70	Rs.993 +50.07%	-0.44%
BRITANNIA	2,972.65	Rs.3,467 +16.64%	-0.61%
DABUR	399.35	Rs.491 +22.86%	-0.92%
TECHM	814.30	Rs.840 +3.16%	+1.16%
JSWSTEEL	289.10	Rs.428 +47.89%	+0.28%
GODREJCP	670.00	Rs.979 +46.07%	-0.28%
SHREECEM	19,007.90	Rs.19,820 +4.27%	-1.47%
PIDILITIND	1,220.00	Rs.1,313 +7.59%	+1.09%
TATASTEEL	518.70	Rs.648 +24.85%	-0.89%
GRASIM	859.10	Rs.1,115 +29.78%	-1.19%
YESBANK	232.55	Rs.404 +73.73%	-0.06%
BOSCHLTD	18,045.25	Rs.22,400 +24.13%	+0.55%
HAVELLS	772.75	Rs.783 +1.33%	+1.48%
MOTERSUMI	147.00	Rs.238 +61.88%	-2.07%
DRREDDY	2,832.80	Rs.2,878 +1.60%	+0.38%
MARICO	358.85	Rs.397 +10.52%	-1.17%
AUROPHRMA	779.80	Rs.830 +6.50%	+0.15%
CIPLA	556.30	Rs.678 +21.96%	-0.96%
AMBUJACEM	217.65	Rs.253 +16.13%	-4.03%
DIVISLAB	1,728.90	Rs.1,775 +2.66%	+2.49%
HINDALCO	198.30	Rs.260 +30.99%	-1.02%

Figure 2: Current trend of ICICIBANK in wall mine market data.

Figure 2 illustrates the current trend of the ICICIBANK in the wall mine market data. Here, different technical factors such as volume, exponential moving average and a simple moving average, high, open, low and close, and news trend are used to find the sentimental trend in ICICIBANK. Most of the traditional stock sentiment score is computed by using the stock price and the news positive and negative sentiment. The sample training data are shown in Table 2 for sentiment score and classification process.

3.1 Proposed Stock sentiment score:

Let α, β, γ represents three essential stock trend technical factors taking from the training data

α= Stock Performance (D)

β= Volatility (D)

γ=RSI (D)

$$Stocksentiment = \frac{(\alpha * \beta)}{\gamma} \tag{3}$$

Here, the computed stock sentiment score is used as the additional attribute in the training data as shown below.

```
@relation stockdata

@attribute Symbol
{SPYL,SUJANAUNI,LPDC,VHL,JETAIRWAYS,CUBEXTUB,Z
ENITHBIR,GAMMNINFRA,RELCAPITAL,DEEPIND,KAUSH
ALYA,GLOBALVECT,SURANAT&P,GKWLIMITED,PRECWI
RE,WSI,SUZLON,GRUH,GAL,ROHITFERRO,BSELINFRA,RA
MSARUP,PSL,MANINDS,SUPREMEINF,JBFIND,NEXTMEDI
A,PRADIP,MADHUCON,KKCL,JYOTISTRUC,MAHABANK,I
L&FSENGG,EDL,UNITECH,LUPIN,SUPREMEIND,SABTN,M
OTILALOFS,JPINFRATEC,TPNL,CURATECH,LGBFORGE,SY
MPHONY,EIHAHOTELS,CYBERTECH,ONGC,KGL,CUMMIN
SIND,SHILPAMED,GANESHHOUC,ABB,GTNIND,THOMASC
OOK,SYNCOM,JAYNECOIND,INDOSOLAR,ENOB,JINDCOT,
ASHIANA,MAHSCOOTER,AKSHOPTFBR,KOTHARIPRO,HI
NDOILEXP,BHUSANSTL,INDBANK,FCBP,RICOAUTO,LGBB
ROSLTD,KAMATHOTEL,BLBLIMITED,CALSOFT,MAGNUM
,HINDPETRO,EASUNREYRL
@attribute Price numeric
@attribute ADX numeric
@attribute ATR numeric
@attribute Volatility numeric
@attribute MFI numeric
@attribute 'Performance today' numeric
@attribute Sentiment {Buy,Sell}
@attribute senti numeric

@data
SPYL,0.25,28.95,0.05,25,30.04,11.11,0,-0.02,-7.92,2.25,24.99999
8,Buy,10.402795
SUJANAUNI,0.3,13.23,0.06,24,45.05,22.22,37.73,-0.01,-4.5,29.62
,20.000004,Buy,5.327415
LPDC,5.45,12.95,0.38,8.35,66.16,59.94,66.67,0.04,-0.25,86.76,19.
78021,Buy,1.248222
VHL,2281,13.22,92.49,4.47,42.56,49.58,40.17,-7.07,-2.09,51.42,1
0.140029,Buy,0.532494
KAUSHALYA,0.8,27.52,0.04,5.33,50.35,70.37,70.86,-0.03,-4.89,
83.47,6.666668,Buy,0.352863
GLOBALVECT,117.9,32.27,8.23,7.1,79.23,79.92,100,9.71,10.83,
95.23,6.40794,Buy,0.287116
SURANAT&P,4.3,9.49,0.2,4.82,45.4,53.48,48.49,-0.03,-0.08,64.0
9,6.172839,Buy,0.327677
```

Table 2: Sample training data taken from the NSE stock market website.

No	Symbol Nominal	Price Numeric	ADX Numeric	ATR Numeric	Volatility Numeric	RSI Numeric	STO Numeric	StochRSI Numeric	MACD Numeric	PPO Numeric	MFI Numeric	Performance today Numeric	Sentiment Nominal
1	ATLASCYCLE	63.1	26.67	4.23	6.39	26.88	11.4	18.55	-2.39	-8.51	19.1	-4.682777	Sell
2	CANDC	9.7	48.69	0.72	7.2	33.58	41.92	59.75	-1.15	-9.89	19.25	-3.000002	Sell
3	HERITGFOOD	485	26.05	17.8	3.62	36.95	12.56	0.0	-2.79	-1.71	19.41	-1.541421	Sell
4	BNDALAGRO	13.75	20.93	0.79	5.27	32.72	11.3	0.0	-0.38	-3.23	20.45	-0.36232	Sell
5	SSWL	821.2	23.88	20.88	2.51	32.3	31.74	0.0	-8.17	-1.23	20.92	-1.107901	Sell
6	NEXTMEDIA	19.65	57.49	0.91	4.67	34.34	28.63	86.2	-1.88	16.68	21.29	4.799998	Buy
7	VISASTEEL	7.0	15.53	0.53	7.57	43.32	17.04	0.0	-0.06	-2.31	21.41	0.0	Buy
8	VISASTEEL	7.0	15.53	0.53	7.57	43.32	17.04	0.0	-0.06	-2.31	21.41	0.0	Sell
9	KAWERITEL	4.6	8.42	0.32	6.67	41.51	37.5	55.89	-0.11	-1.84	21.53	-4.166672	Sell
10	PFOCUS	58.05	13.5	2.56	4.29	41.12	17.3	1.16	-0.54	-2.7	21.72	-2.763822	Sell
11	SAKHTISUG	11.05	14.08	0.6	5.45	46.71	59.23	33.33	-0.03	-2.54	22.0	0.454547	Buy
12	CENTURYPLY	180	20.96	7.45	4.08	39.37	7.04	33.33	-0.68	-1.32	22.18	-1.177442	Sell
13	CALSOFT	22.95	15.4	2.0	8.89	50.22	56.27	0.0	0.62	2.26	22.24	2.000003	Buy
14	IL&FSTRAINS	6.4	22.46	0.33	6.47	30.25	15.19	26.57	-0.4	-9.06	22.59	0.787405	Buy
15	ZODZARMKA	36.75	15.85	2.26	6.01	50.15	36.61	50.51	-0.18	-0.84	22.73	-2.260634	Sell
16	MASHEAMLS	474.3	10.99	9.56	2.02	44.72	20.72	0.0	-0.31	-0.5	23.18	0.327866	Buy
17	TIDEWATER	496	31.17	98.78	2.0	28.68	11.11	0.0	-79.95	-3.9	23.32	0.378406	Buy
18	GMDCLTD	73.5	16.16	2.04	2.74	30.9	9.58	32.03	-1.54	-3..53	23.33	-1.408455	Sell
19	JETAIRWAYS	169	20.07	21.94	14.1	19.9	32.48	0.0	-10.57	-1.42	23.42	9.728509	Buy

3.2 Stock Data Transformation

In this phase, a new stock trend transformation method is implemented on the training data to find the positive and negative trend normalization. The computational formula used to normalize the entire training data is given below.

Input: Sentiment score data in phase1 as training data T_D.
Output: Transformed normalized data for stock trend classification process N_D.
Procedure:
For each attribute A_T (i) instance T_D
Do
if (A_T (i)==Numerical)
Then
Normalize A_T (i) using eq 4. As

n= Stockperformace(T)

$$N\big(A_T(i)\big) = \frac{A_T(i)*n}{\sqrt{\sum_{i=1}^{n}(A_T(i)-\mu_T)^2*\sigma_T}} \quad (4)$$

End if
If(A_T (i)==Nominal)
Then
Continue;
End if
Done
Perform hybrid Bayesian classification algorithm.

3.3 Proposed Bayesian Network Classification model

In the proposed Bayesian network classifier, a directed acyclic graph is constructed on the normalized market data. Parameter estimation and statistical learning are the two methods used to improve the traditional weak Bayesian network. In the parameter estimation method, numerical or continuous attributes are estimated using the numerical parametric estimation and nominal or categorical attributes are estimated using the discrete parameter estimation measures. Finally, after parametric estimation Bayesian structure is learned using the optimized scoring function and joint probabilistic function as shown in Figure 3.

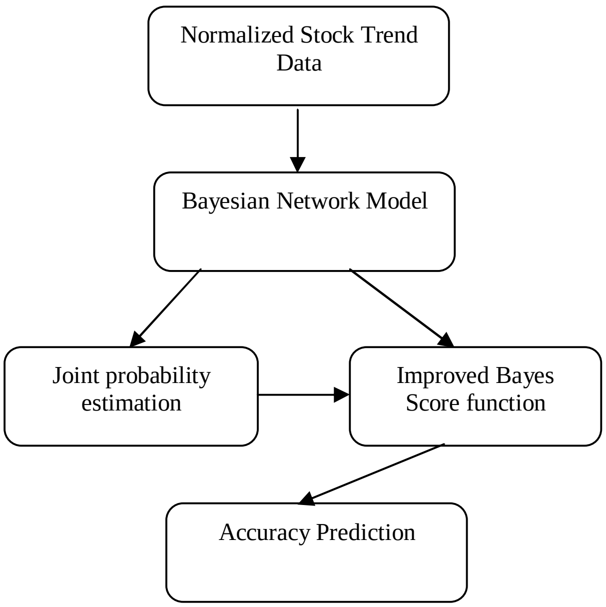

Figure 3: Proposed Framework.

3.4 Hybrid Bayesian Network for Stock Market Trend Prediction

The proposed Bayesian network has two steps to predict the stock market trend using the normalize data. In the first step, both numerical and nominal statistical parameters are predicted using the joint probability estimation. In the second step, statistical Bayesian DAG graph structure is learned using the stock trend Bayes scoring function.

Step 1: D: =Input normalized data.

Step 2: for each attribute A in D.

Step 3: Construct directed acyclic graph to the transformed data using Bayesian network.

Step 4: Computing conditional probabilities to each input variable for joint probability estimation.

Step 5: In Parameter estimation step discrete parameter estimation in the Bayesian network can be predicted using the following measure.

$$P\left(A_i = I_k / C_j\right) = \frac{N_{ijk}}{N_j} \qquad (5)$$

Where N_{ijk} is the number of instances of class c_j having the value I_k in attribute A_i.

Step 6: Continuous parametric estimation in the Bayesian network can be estimated using the following measure.

$$P\left(A_i = I_k / C_j\right) = G\left(I_k, \mu_{ij}, \sigma_{ij}\right)$$
$$G\left(I_k, \mu_{ij}, \sigma_{ij}\right) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{\frac{\left(I_k - \mu_{ij}\right)^2}{2\sigma_{ij}^2}} \qquad (6)$$

Here normal distribution is an approximation to a Gaussian distribution.

Step 7: Estimating Bayesian parameter using the traditional parameter estimation as

$$\theta_{ijk} = \frac{N_{ijk} + n_{ijk}}{N_{ij} + n_{ij}}$$
$$where\ N_{ij} = \sum N_{ijk} ; n_{ij} = \sum n_{ijk} \qquad (7)$$

Step 8: In Model score estimation statistical model selection is performed on the DAG graph to find the best fit to the input data. In the traditional Bayesian networks, constraint based and score based statistical models are used to find the best fit to the data. In the proposed approach, a novel Bayesian score is used to improve the prediction rate and to find the DAG conditional dependencies and independencies on each attribute. Basically, score based approach evaluates the dependencies and independencies in a structure that matches to the input data. This scoring measure optimizes the structure that maximizes the scoring value.

The proposed Bayesian score is computed as:

$$fscore_k = \sum T(\alpha_i + N_i)$$
$$nsument = \sum \alpha_i * (\sqrt{x(\alpha_i, fscore_k}) + N_i$$
$$mfscore_k = fscore_k - \sqrt[3]{nsumentT\alpha_i}$$
$$BayesScore = (mfscore_k - |N|^* T\alpha_i) + NormDistri(|N|^* \alpha_i) \qquad (8)$$

Step 9: After computing the Bayesian score to each random variable, its class and attribute probability estimations are performed based on the following steps.

for each class in classlist
do
for each node in BN attribute list
do
if(node[i]==Class)
iCP=iCP*|C|+CIndex;
else
iCP=iCP*|P_A|+P_AIValue
Compute log class and exponential attribute distribution measure to each input attribute node and class node as
if (node[i] == Class)
Log class probability-LC Prob-Log(Node Prob(CIndex))
else

Exponential Attribute Probability=EAP=(Node Prob(P_A IValue))^2;
endfor
Total Class Probability=TCP[CIndex]-TLCProb
Total Attribute Probability=TAP[P_A IValue]=TEA
endfor

Step 10: Bayesian scores and computed class and attribute probabilities are used to choose the optimal Bayesian random dependency and independence variable that fits the structure to the data.

4. EXPERIMENTAL RESULTS

Experimental results are simulated using Java environment and real-time market data. The proposed model is compared to the traditional stock market classification models to verify the performance of the hybrid Bayesian model to the traditional models. Also, the proposed model is compared to the traditional techniques by using various statistical performance measures such as accuracy, true positive rate, recall, precision, false positive rate, ROC area etc. These performance metrics are analyzed and compared by using third party Java libraries.

Table 3 illustrates the experimental results of the hybrid Bayesian model on the training stock market data to predict the class label of the trend.

Table 3: Experimental results shows the class label trend.

KPIT,0.000011,0.004557,0.000165,0.000657,0.001104,0.00187
6,0.001616,0.000618,0.004805,0.000926,0.000412,Buy,0.0000
01 ====> Estimated stock trend value ::0.0
ITI,0.000008,0.001305,0.00008,0.000364,0.001702,0.001142,0,
0.000025,0.001643,0.001919,0.000349,Buy,0 ====> Estimated
stock trend value ::0.0
MARKSANS,0.000002,0.000649,0.000021,0.000361,0.001842,
0.001589,0,0.000004,0.000402,0.00136,0.000325,Buy,0 ====>
Estimated stock trend value ::0.0
BPL,0.000002,0.000697,0.000027,0.000471,0.001214,0.000222
,0,0.000031,0.002625,0.001328,0.000324,Buy,0 ====>
Estimated stock trend value ::0.0
JAGSNPHARM,0.000002,0.000534,0.000022,0.000366,0.0011
5,0.000161,0,0.000021,0.001777,0.001295,0.000299,Buy,0
====> Estimated stock trend value ::0.0
GANDHITUBE,0.000027,0.000951,0.000195,0.000246,0.0013
49,0.001014,0.000446,0.000112,0.000716,0.000363,0.000282,
Buy,0 ====> Estimated stock trend value ::0.0

207,Buy,0 ====> Estimated stock trend value ::0.0
KPRMILL,0.00004,0.001079,0.000237,0.000176,0.001644,0.00
1404,0.00141,0.000101,0.0002,0.001007,0.000191,Buy,0
====> Estimated stock trend value ::0.0
ZYDUSWELL,0.000084,0.000638,0.00039,0.000128,0.00142,0
.00055,0.000848,0.000055,0.00015,0.000491,0.000181,Buy,0
====> Estimated stock trend value ::0.0
THANGAMAYL,0.000021,0.001255,0.000236,0.000308,0.001
306,0.00044
THANGAMAYL,0.000021,0.001255,0.000236,0.000308,0.001
306,0.000441,0.000468,0.000061,0.000976,0.000841,0.000178,
Buy,0 ====> Estimated stock trend value ::0.0
KICL,0.000099,0.000935,0.000815,0.000197,0.001601,0.00131
7,0.000785,0.001113,0.002172,0.00098,0.000135,Buy,0 ====>
Estimated stock trend value ::0.0
NSIL,0.000055,0.000538,0.000488,0.000201,0.001369,0.00093,
0.001117,0.000233,0.001519,0.001232,0.00012,Buy,0 ====>
Estimated stock trend value ::0.0
INDLMETER,0.000002,0.000473,0.000036,0.000377,0.001059
,0.000285,0.000747,0.000015,0.001141,0.000542,0.000116,Buy
,0 ====> Estimated stock trend value ::0.0
TATAGLOBAL,0.000011,0.000947,0.000072,0.000147,0.0011
91,0.001317,0.000533,0.000085,0.001132,0.00106,0.000109,Bu
y,0 ====> Estimated stock trend value ::0.0
GODREJIND,0.000025,0.002384,0.000185,0.000153,0.000918,
0.00017,0.000531,0.000044,0.000162,0.000376,0.000108,Buy,0
====> Estimated stock trend value ::0.0
KIRLOSBROS,0.000008,0.001029,0.00007,0.000172,0.00118,0
.001063,0.000349,0.000065,0.000219,0.00111,0.0001,Buy,0
====> Estimated stock trend value ::0.0
HEIDELBERG,0.000008,0.000929,0.000066,0.000162,0.00107
7,0.000447,0,0.000059,0.000334,0.000881,0.0001,Buy,0 ====>
Estimated stock trend value ::0.0
NCLIND,0.000006,0.001006,0.000073,0.000227,0.000907,0.00
0247,0.000233,0.000022,0.000096,0.000713,0.000091,Buy,0
====> Estimated stock trend value ::0.0
CIPLA,0.000025,0.000896,0.000131,0.000099,0.001325,0.0011
47,0.000356,0.000136,0.000486,0.000863,0.000088,Buy,0
====> Estimated stock trend value ::0.0
MINDTREE,-0.000034,-0.000717,-0.000148,-0.000065,-0.0009
95,-0.000796,-0.000366,-0.000196,-0.00033,-0.000713,-0.00005
1,Sell,-0 ====> Estimated stock trend value ::1.0
MMFL,-0.000021,-0.000615,-0.000149,-0.000113,-0.000909,-0.
000735,-0.000418,-0.000073,-0.000196,-0.000855,-0.00006,Sell
.-0 ====> Estimated stock trend value ::1.0

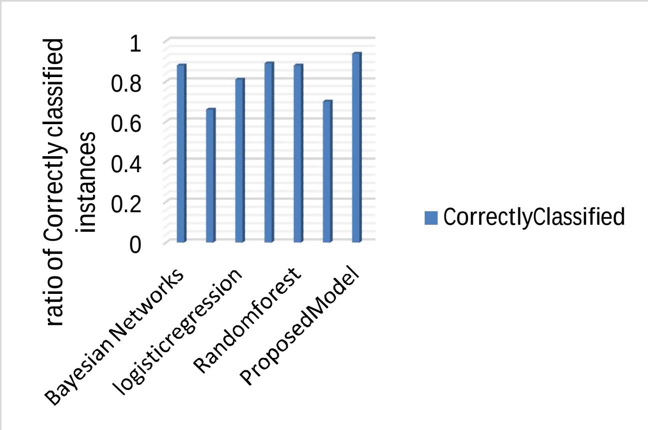

Figure 4: Performance analysis of proposed Bayesian technique to the traditional classification techniques on training stock market data by using correctly classified instances.

Figure 4 illustrates the comparison of the proposed Bayesian technique to the traditional methods such as naïve Bayes, logit boost, random forest, logistic, random tree, Bayesian networks on the training dataset. From the Figure 4, it is observed that the proposed Bayesian method has high correctly predicted stock trend than the traditional models.

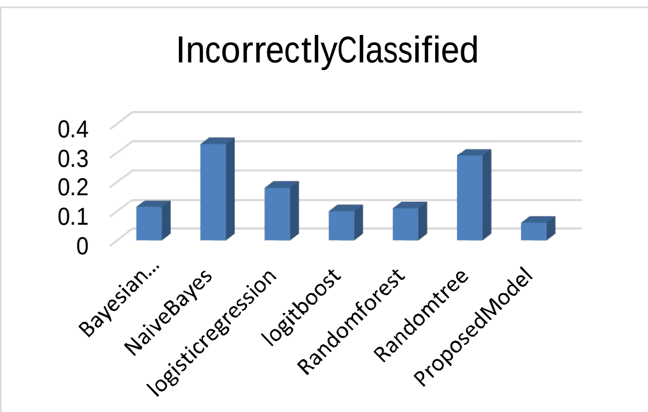

Figure 5: Performance analysis of proposed Bayesian technique to the traditional classification techniques on training stock market data by using incorrectly classified instances.

Figure 5 illustrates the comparison of the proposed Bayesian technique to the traditional methods such as naïve Bayes, logit boost, random forest, logistic, random tree, Bayesian networks on the training dataset. From the Figure 5, it is observed that the proposed Bayesian method has less incorrectly predicted stock trend than the traditional models.

Figure 6 illustrates the comparison of the proposed Bayesian technique to the traditional methods such as naïve Bayes, logit boost, random forest, logistic, random tree, Bayesian networks on the training dataset. From the Figure 6, it is observed that the proposed Bayesian method has high true positive rate for stock trend prediction than the traditional models.

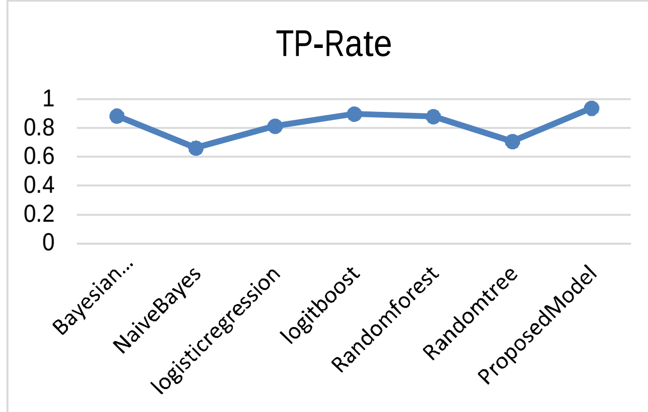

Figure 6: Performance analysis of proposed Bayesian technique to the traditional classification techniques on training stock market data by using true positive rate.

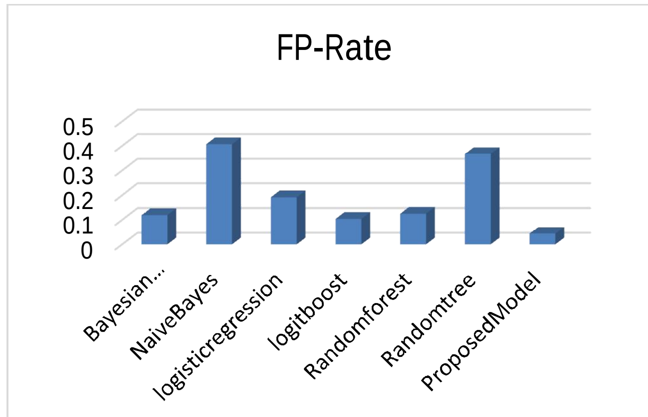

Figure 7: Performance analysis of proposed Bayesian technique to the traditional classification techniques on training stock market data by using false positive rate.

Figure 7 illustrates the comparison of the proposed Bayesian technique to the traditional methods such as naïve Bayes, logit boost, random forest, logistic, random tree, Bayesian networks on the training dataset. From the Figure 7, it is observed that the proposed Bayesian method has less false positive rate for stock trend prediction than the traditional models.

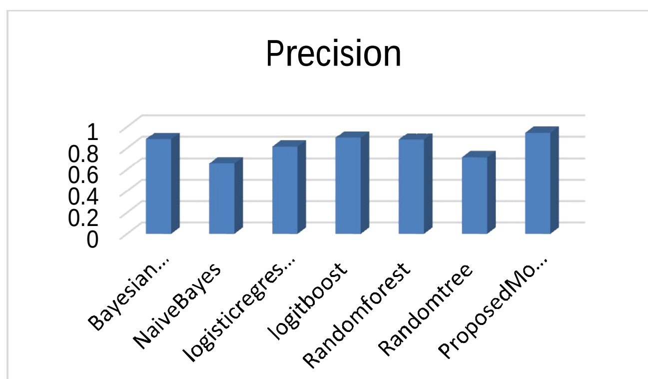

Figure 8: Performance analysis of proposed Bayesian technique to the traditional classification techniques on training stock market data by using precision rate.

Figure 8 illustrates the comparison of the proposed Bayesian technique to the traditional methods such as naïve Bayes, logit boost, random forest, logistic, random tree, Bayesian networks on the training dataset. From the Figure 8, it is observed that the proposed Bayesian method has high precision rate for stock trend prediction than the traditional models.

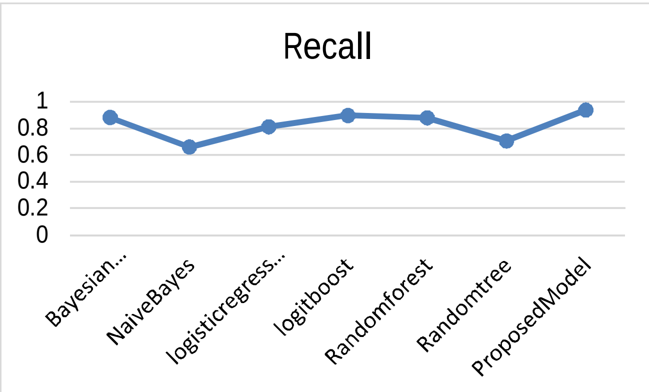

Figure 9: Performance analysis of proposed Bayesian technique to the traditional classification techniques on training stock market data by using recall rate.

Figure 9 illustrates the comparison of the proposed Bayesian technique to the traditional methods such as naïve Bayes, logit boost, random forest, logistic, random tree, Bayesian networks on the training dataset. From the Figure 9, it is observed that the proposed Bayesian method has high recall rate for stock trend prediction than the traditional models.

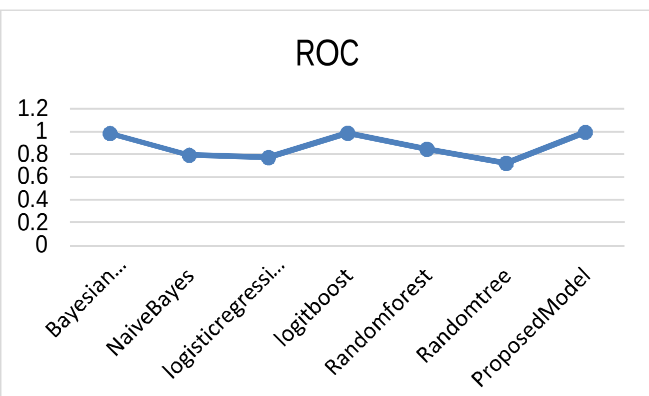

Figure 10: Performance analysis of proposed Bayesian technique to the traditional classification techniques on training stock market data by using ROC area.

Figure 10 illustrates the comparison of the proposed Bayesian technique to the traditional methods such as naïve Bayes, logit boost, random forest, logistic, random tree, Bayesian networks on the training dataset. From the Figure 10, it is observed that the proposed Bayesian method has high ROC rate for stock trend prediction than the traditional models.

5. CONCLUSION

In this paper, a novel filter based Bayesian estimation classifier is proposed to predict the stock trend on the intraday market data. In order to overcome data sparsity and uncertainty problems, a hybrid probabilistic based stock sentiment prediction model is designed and implemented on the real-time market data such as stocks technical, stocks news, stock momentum etc. and integrated to predict the buying or selling behavior of the investor. To normalize the market trend for intraday data, a novel stock sentiment score and data transformation methods are used. To predict and test the trend of the stock on the training data, an improved Bayesian networks technique is implemented on the filtered data. Experimental results show that the proposed framework is more efficient than the traditional sentiment classification models in terms of accuracy and error rate.

REFERENCES

1. J. Qiu, C. Liu , Y. Li and Z. Lin, **Leveraging Sentiment Analysis at the aspects level to Predict Ratings of Reviews**, information sciences.
2. S. Akhtar, R. Faff, B. Oliver and A. Subrahmanyam, **Reprint of: Stock salience and the asymmetric market effect of consumer sentiment news,** Journal of Banking & Finance 37 (2013) 4488–4500 https://doi.org/10.1016/j.jbankfin.2013.07.032
3. S. Basak, S. Kar, S. Saha, L. Khaidea and S. R. Dey, **Predicting the direction of stock market prices using tree-based Classifiers.**
4. B. Li, K. C.C. Chan, C. Ou b and S. Ruifeng, **Discovering public sentiment in social media for predicting stock movement of publicly listed companies**, Information Systems 69C (2017) 81–92. https://doi.org/10.1016/j.is.2016.10.001
5. S. Kim and D. Kim, **Investor sentiment from internet message postings and the predictability of stock returns,** Journal of Economic Behavior & Organization.
6. Q. Li, T. Wang, P. Li, L. Liu, Q. Gong and Y. Chen, **The effect of news and public mood on stock movements**, Information Sciences.
7. X. Li, H. Xie, L. Chen, J. Wang and X. Deng, **News impact on stock price return via sentiment analysis**, Knowledge based systems.
8. K. Li, Reaction to news in the Chinese stock market: **A study on Xiong'an New Area Strategy**, Journal of Behavioral and Experimental Finance.
9. W. Long, L. Song, Y. Tian, **A new graphic kernel method of stock price trend prediction based on financial news semantic and structural similarity**, expert systems with applications.
10. L. S. Malagrino, N. T. Roman and A. M. Monteiro, **Forecasting Stock Market Index Daily Direction: a Bayesian Network Approach**, expert systems with applications.
11. T. H. Nguyen, K. Shirai and J. Velcin, **Sentiment Analysis on Social Media for Stock Movement Prediction**.

12. Z. Ni, D. Wang and W. Xue, **Investor sentiment and its nonlinear effect on stock returns—new evidence from the Chinese stock market based on panel quantile regression model**, Economic Modelling 50 (2015) 266–274. https://doi.org/10.1016/j.econmod.2015.07.007
13. N. Oliveira, P. Cortez and N. Areal, **Stock market sentiment lexicon acquisition using microblogging data and statistical measures**, decision support systems.
14. J. L. Hui, G. K. Hoon and W. M. Nazmee, **Effects of Word Class and Text Position in Sentiment-based News Classification**, 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali,Indonesia.
15. S. Ik Seok, H. Cho and D. Ryu, **Firm-specific investor sentiment and the stock market response to earnings news**, North American Journal of Economics & Finance.