



Original Article

Stock Market Prediction: The Application of Text-Mining in Vietnam

Le Hong Hanh^{1,*}, Nguyen Ngoc Nam², Nguyen Thuy Linh¹,
Nguyen Linh Diep¹, Nguyen Ngoc Hai¹

¹VNU University of Economics and Business,
No. 144, Xuan Thuy Road, Cau Giay District, Hanoi, Vietnam

²VNU University of Engineering and Technology,
No. 144, Xuan Thuy Road, Cau Giay District, Hanoi, Vietnam

Received: November 10, 2021

Revised: March 28, 2022; Accepted: April 25, 2022

Abstract: There are very few studies in Vietnam on the application of text mining in finance and Vietnamese language processing. The origin of this study comes from one of the leading studies on the use of machine learning to analyze text data from 4 well-known online newspapers in Vietnam to forecast the increase, decrease and neutrality of the VN-Index one day in advance. This study used nearly 70,000 articles from four reputable and reliable online newspapers in Vietnam as input data for machine learning models. These were: decision trees, random forests, KNNs and SVMs. After selecting the best model (SVM) and the best dataset (Vietstock), the techniques used to dig deep and refine the findings raised the accuracy to 60.1%. The end result is solid evidence that news about the financial and stock situation in the popular press affects the price movements of the VN-INDEX and the Vietnamese stock market.

Keywords: Machine learning, text-mining, stock market, decision tree, random forest, KNN, SVM, VN-Index.

1. Introduction

In these days and this age, the stock market in the world generally and Vietnam in particular, is attracting more and more investors. According

to the Vietnam Securities Depository [1], only in November 2020, the Vietnamese stock market had over 41,200 newly created accounts (the highest monthly increase ever). Currently with more than 2.7 million accounts (as of the end of

* Corresponding author

E-mail address: Hanhhongle1990@gmail.com

<https://doi.org/10.25073/2588-1108/vnueab.4715>

November 2020), Vietnam's stock market is attracting more attention and participating investors want to know more about the future of the market to invest more profitably. Therefore, effective market prediction will bring great benefits both at the macro and micro level, helping brokers to offer trading advice or that can be used as input for decision-making.

For decades, stock market analysis has had two approaches: fundamental analysis and technical analysis. The inputs for these methods are historical price data in coordination with financial indicators. However, these financial indicators and business financial performance are often summarized in a specific period (quarter, half year, year), but news related to a business would likely have an immediate impact on its share price. For example, when the board of directors of a company announce a merger and acquisition plan next year, this will immediately impact the share prices of the two companies, but in terms of financial data, it will reflect on the next financial reporting period. This demonstrates the fact that news has an evidently strong impact on the stock market, and for an emerging market like the Vietnamese stock market, this impact is even more pronounced.

There have been studies using algorithms to analyze market data through technical analysis by using ARCH or GARCH models, etc. In the past, the most popular method applied was regression analysis, but now with the development of science and technology, algorithms related to machine learning have been increasingly applied. Yu [2] indicated that technical analysis can only observe one piece of the overall picture of the market and using only this method of analysis will face many difficulties, as it is only applicable in specific circumstances. As for fundamental analysis, it is not only based on historical financial data but also relies on many other factors such as the political situation, business environment, news, etc. These factors can be in a structured number format (numeric data) or they can also be in a textual data format.

Vietnam's stock market has shown a strong development in size and structure and has played an important role in promoting the country's economic development and international integration. With more than 700 stock symbols representing more than 700 listed companies, every day, hundreds of news items related to these companies are published in newspapers and on social networks. According to the Efficient Market Hypothesis, asset prices reflect all available information and the price of a security will fully reflect its value. However, gathering all this news for each investor is not only a matter of sources but also a matter of time. Regarding sources, the Vietnamese stock market has only achieved a weak form of efficiency. This result is largely correlated with the lack of transparency of information on the Vietnamese securities market. A socialist economy with the state economy playing a leading role is being developed in Vietnam. However, the state-owned enterprises have been operating ineffectively and inefficiently for a long time. According to general statistics office of Vietnam in 2007, Vietnam has around 3,000 state-owned enterprises, out of which about 500 have 100%. In spite of equitization, many state-owned enterprises still hold a majority of stocks. This is a major barrier to information transparency [2]. According to a speed test sponsored by Staples, an adult has an average reading speed of 300 words per minute. On average, each A4 size sheet has between 400 and 500 words. Therefore, to read a 20 pages document, a person needs at least 30 minutes and even more time to understand, analyze, and summarize that amount of information. However, the increasing amount of information published on the Internet has increased the demand for tools to help readers search and summarize information [3].

In fact, textual data provides just as important insights as numerical data. Textual information is easier to understand and helps the reader to grasp a general perspective. Especially for new investors with little experience in applying technical analysis for securities investment, most of them rely on textual

information provided by leading online news sites about securities or daily news articles published by securities companies. Hence, text analysis is of great importance and is complementary to the analysis of financial indices and valuation models.

Text mining, an artificial intelligence technique to solve problems, can generalize the main idea in a short time and calculate the correlation between words. Das [4] has exploited large-scale text, automatically processing plain text language in digital form to extract data converted into useful quantitative or qualitative information.

The use of text-mining methods for stock market forecasting is highly necessary and creative. However, there are only a limited number of studies on the topic in Vietnam. Therefore, the research team conducts this research to fill the research gap and contribute to the literature as well as the decision-making process of investors in the stock market. The objective of this study is to predict the volatility (three levels: rise, neutral, and decrease) of the VN-INDEX stock index one day ahead of time. The study data ranges from 2001 through 2021. The scope of this research is to employ the text-mining approach on a set of Vietnamese news in the financial and securities categories from 4 popular Vietnamese online dailies (such as Cafef, Thanh Nien, Vietstock, VNexpress).

The paper is organized as follows: Part 1 is the introduction; a literature review is presented in Part 2; Part 3 is about methodology; results are showed in Part 4; and Part 5 is the conclusion.

2. Literature review

The application domain of this research is the stock market also known as the equity market or share market. This is the market where shares of public listed companies are issued and traded. Empirically, studying stock markets movements reveal information about the driving factors of the markets. From a theoretical point of view, this can be viewed as assessment of existing asset pricing theories.

The skepticism about the predictability of stock prices can be traced to when the Capital Asset Pricing (CAPM) model and the Efficient Market Hypothesis (EMH) were introduced and proposed that the stock market is informationally efficient in the way that its prices always fully reflect all available information. Although the Efficient Market Hypothesis enjoyed great popularity in the past, two contradicting theories constitute a major school of thought in modern-day finance, namely, fundamental and technical analysis. In addition to the two traditional practices, the rising era of big data has led text mining to become another ubiquitous tool for stock market prediction. This is reflected in a large number of researchers investigating this particular topic. However, there are a variety of datasets, methods, and algorithms used among researchers, culminating in a variation in results as well.

With the current trends in big data and content creation on the Internet and the enormous amount of unstructured text data available through the mobile channels, and social network services, there are several research works that have attempted to predict stock movements using such text data. One such case is that of Dang [5] who applied the ARIMA model to predict the stock price based on 257 observations where the author forecast the closing price of the next 10 days. The results were quite accurate with high reliability, which also shows that the ARIMA model given for each security series in the thesis is quite suitable for short-term forecasting of stock prices. The author basically understands the process of using Eviews software to build ARIMA models for real-time data, to calculate forecast values for stock data series.

In Vu [6], the authors explored features from Twitter messages to capture the relevant public mood for the four Tech-companies to predict the daily up and down price fluctuations of their NASDAQ shares. The research proposed a new model that combined specific features such as positive and negative psychology, consumer confidence in the product in relation to the words

“Bullish” or “bearish” and the previous three days of stock market advocacy, as shown in Fig.1. Features used in the Decision Tree classification using multiple authentications delivered accuracy of 82.93%, 80.49%, 75.61%,

and 75.00% in anticipation of daily up-and-down changes by Apple (AAPL), Google (GOOG), Microsoft Stocks (MSFT), and Amazon (AMZN) respectively, in the 41-day market model.

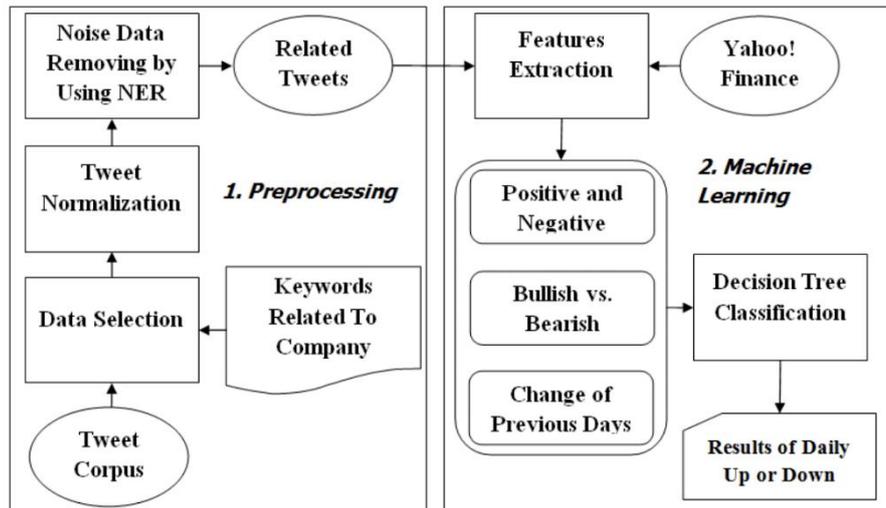


Figure 1: The stock price forecast model using Twitter data

Source: Vu [6].

Le [7] used a number of machine learning algorithm models to forecast the volatility trend of Vietnam's stock market index. The results show that the volatility of the Vietnamese stock market depends only on the volatility of the previous day, not on the days beyond, nor on the trading volume. Among models such as the Logistic regression model, Linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and KNN model, the KNN model has the best predictive accuracy with 55.6%.

The above researches mainly use pure numerical data to forecast stock prices without considering the impact of the news of the financial, economic and world situations on the volatility of stock prices. However, they are data that can cause stock prices to fluctuate by a large margin, creating new trends in the market that can lead to higher financial returns if accurate predictions are made.

Nassirtoussi [8], summed up the recent research on the application of text mining to

market forecasts, in which most of the systems proposed in the research works follow the architecture, as portrayed in Figure 2.

Another similar approach for finding the correlation between the content of news articles and stock prices for the purpose of predicting the stock markets was implemented by Kaya and Karsligil [9]. They collected news articles published in the last year and combined them with the stock prices for the same period. These articles were then labeled with a positive or negative sentiments categorization based on their effects on the stock price. Kaya and Karsligil's approach is a little bit different in the sense that for them it was important to use the price changes for the categorization of the news. While analyzing the textual data, the authors used an approach of word doubles of a noun and a verb as features and not only single words. The support vector machines (SVM) method was used in this case, which resulted in 61% accuracy.

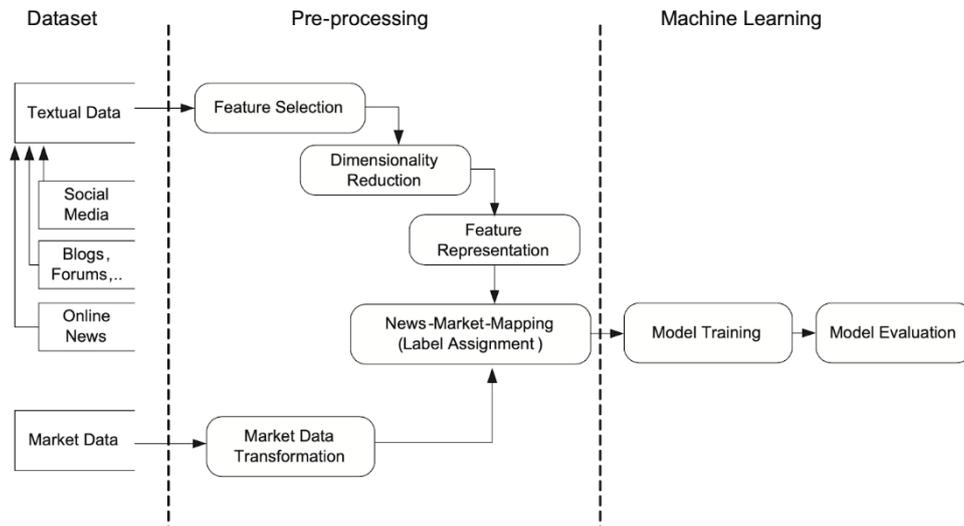


Figure 2: Generic common system components of the market forecasting system
 Source: Nassirtoussi [8].

A project by Joshi, Rao, and Bharathi [10] selected financial news articles about a given company, and they used those data to try to predict the future movements of the stock, again by applying sentiment analysis. The approach is similar to the other cases with the idea to identify how stocks have reacted if news has polarity. The approach in this paper is dictionary-based and contains positive and negative words that are built based on financially specific words. Further, the researchers have pre-processed the data, which resulted in having their own finance-specific stop words and dictionary. Using their own dictionary, they have implemented three models for classification and tested them. After comparing the results, they have concluded that the Random Forest algorithm resulted in better accuracy for the test cases, ranging from 88% to 92%. This algorithm was followed by Support Vector Machines with again very good accuracy of 86%. In their case, the Naïve Bayes algorithm performance was the lowest at 83%.

Many scholars have tried different approaches in research to prove that there is a potentially strong correlation among financial news articles and stock price fluctuations, as in Ayman et al. [11]. In their paper, they proposed

an approach to use sentiment analysis in financial news, along with features extracted from historical stock prices to apply predictions for the future behavior of stocks. According to their findings, the proposed model achieved high accuracy using sentiment analysis in categorizing news polarities by applying the Naïve Bayes algorithm. In their case, the accuracy of the model is up to 86.21%. By moving on with their experiment in prediction, during their next attempt in analyzing these news articles, they have included numerical attributes that increased the accuracy to 89.80%.

There is some promising research published that applies deep learning techniques and has resulted in higher accuracy ratings. Of interest is the published paper by Tabari et al. [12] that shows a comparison of diverse algorithms specifically applied in stock market tweets. This research shows quite promising results, with accuracy ranging up to 92.7% (using Convolutional Neural Networks).

Consequently, many researchers have explored the field of text mining to examine the effectiveness of text analysis for stock price movement prediction and produced a text mining prediction system to forecast the stock

price changes of companies (down, stay, or up) influenced by financial events documents. However, in Vietnam, research on stock market forecasts mainly approaches data mining based on purely numerical data. Thus, the use of Text-mining for Vietnam's stock market is still very crucial and might make a significant contribution to this field. In the following sections, we present a summary of text mining techniques, new models and test results.

3. Methodology

3.1. Data preparation

All systems take at least two sources of data as input, namely, the textual data from the online resources and the market data.

Input data:

Textual data: When deciding news sources, the basic importance and source field are both considered to guarantee the effectiveness and universality of the text data. Since main Vietnamese investors make their decisions relying on news in Vietnam newspapers, therefore, this research selects 69,947 sources from the 4 most popular and most visited websites. The given websites are collected based on their number of accesses, including both general news and finance news. Specifically, news related to the stock market tends to be divided into 8 parts: financial capital news; civil economic news; industrial-economic news; company news; international economic news; emerging market news; consumption news; financial news. The type of news is either general news or special financial news. The majority of the systems use financial news as it is deemed to have less noise compared with general news. What is being extracted here is the news text or the news headline. News headlines are occasionally used and are argued to be more straight-to-the-point and, hence, with less noise caused by verbose text.

Market data: The other source of input data for the systems comes from the numeric values in financial markets in the form of price-points

or indexes. This research approach is based on historical market prices of VN-INDEX from Investing.com and uses technical analysis “Beautiful soup” in the Python language program in order to remove redundant HTML tags, and filters to get necessary data such as the opening price, closing price, high and low price of the day and the trading volume. This data is used mostly for the purpose of training the machine learning algorithms and occasionally it is used to label each day. More specifically, each day is labeled with -1, 0 and 1, respectively, for the market to decrease, not to change and to increase. In other words, this labeling is done by subtracting the next day's closing price from the previous day, if this value is positive, it will be bullish, if it is negative, it will be bearish, and 0 will keep the price unchanged. As a result, the combination of both methods will be an ideal choice to increase the accuracy of the system.

Links to the news from the sources mentioned are collected using the “Requests” library that is provided in Python. After having all the links, we built a python script based on the Beautiful Soup framework that extracts the article data that is collected in the form of HTML definitions corresponding to each link. Finally, by using the “BeautifulSoup” library, it will analyze and collect necessary data such as main content, title, and time. This library provides idiomatic ways of navigating, searching, and modifying the parse tree and commonly saves programmers hours or days of work. Appropriate data cleansing has been applied for Stock Price Prediction to remove unnecessary HTML tags as well as to format the data from different sources to one standard.

The purpose of prediction varies from predicting the future stock trends to minimizing the volatility of the prices or the market trends. Market volatility is the indicator of fluctuation; therefore, higher volatility means the higher fluctuation of the corresponding stock prices.

Goal: The purpose of prediction varies from predicting the future stock trends to minimizing the volatility of the prices or the market trends. Market trends are the general direction of the

stock prices: upward or downward. Market volatility is the indicator of the fluctuation. A higher volatility means the higher fluctuation of the corresponding stock prices.

Time span: In this paper, we decided to stick with the long-term prediction from 2001 to 2021 generally, because when the important news about a company is published, investors will monitor this news, which impacts on their decision to buy, sell or keep shares that they hold, based on how they think it will affect the stock.

So far, there have been two different approaches in labeling documents. The first approach is to assign a class to the article manually, determined by an expert's opinions about the content of the article. Although the success rate is a bit higher by using this method, many articles will be relatively hard to classify by using only human effort. On the contrary, the second approach attaches a label to articles automatically according to their effect on stock prices. This method is less accurate than the first method because there are many different reasons

why the stock price change does not indicate the actual label of the article.

Therefore, in order to increase the effectiveness of the forecast, we combined the news of the above 4 news sites. The work of merging articles is done automatically by the computer program. Information synthesis is performed at both the data preparation phase for the training phase and the forecasting phase.

3.2. Documents labeling

Based on the approaches so far: 2 or 3 classes are defined for predicting market directions. The goal of processing news generally is to classify the news into two classes: either good news or bad news regarding the selected stock. Sometimes, this classification is extended and another category indicating neutral news is added. Therefore, in our paper, 3 directions of stock price change is collected: rise, drop and neutral, to reflect all the possible directions of the stock market, which are determined by the subtraction of the closing price of the VN-index on day $t+1$ and day t .

Table 1: Example of news classification by price history

Date of news (t)	Date of price (t+1)	The degree of price volatility (Change)	Label assign to news data
08/02/2021	09/02/2021	2.93%	0 (Increase)
07/02/2021	08/02/2021	-3.38%	1 (Decrease)
08/09/2019	09/09/2019	0.00%	2 (No increase or decrease)

Source: Author's compilation.

3.3. Data processing

All the news articles are collected in HTML extension, which contains many unnecessary tags of the HTML format, so we have to eliminate them first. Then we save exported content in plain text format for further processing. Now each document contains lots of sentences, and our next step is tokenizing each

sentence into words by using Regex101 in Python. This tool is proven to achieve more than 90 percent accuracy in tokenizing Vietnamese sentences. The collection of words will be the input for the next step. In the final step of this phase, the relevant words are extracted from the document. As usual, all words, as well as numbers, are considered to be features. After the set of features is extracted, stop words in the

Vietnamese language are removed; this will at least improve the efficiency of the target system and reduce the system resources. For this purpose, a list of Vietnamese stop words has been gathered and used as a stop list dictionary.

3.4. Features weighting

Each document is represented as a multidimensional vector based on all selected features as described in the document. Several methods of feature weighting have been proposed by researchers in the text-mining field. However, we used TF-IDF vectorization to increase the effect of words that appear many times and reduce the effect of words that appear too many times. After the vectorization step, the group obtains the values corresponding to each word. However, because a large number of words can negatively affect the prediction process of the model, the team selected 1024 features as a rule of thumb that are considered to have the most influence on the prediction results of the model. All this work is done by using the “scikit-learn” library with class “TF-IDF Vectorizer” and parameter “max_feature”. It is a free software machine-learning library for the Python programming language. It features various classification, regression and clustering algorithms and is designed to interoperate with the Python numerical and scientific libraries. Below is the formula for the algorithm:

$$w_{ij} = f_{ij} \log_2 \left(\frac{N}{df_i + 1} \right)$$

Where N is the number of documents, f_{ij} is the frequency of the term t_i in the text d_j , and df_i is the number of documents containing the term t_i .

4. Results

Decision Tree, Random Forest, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) were used to evaluate the proposed model. In matrices, TU, TD, TN represent for the correct forecasts for trending, while FU, FD, FN indicate the number of false forecasts for trending. Accuracy is the proportion of the correct forecasts and the false forecasts in all the test data. The formula is as following:

$$\% \text{ Accuracy} = (TU + TD + TN) / (TU + TD + TN + FU + FD + FN) * 100$$

Regarding which method gained the highest result for our model, we performed different approaches and checked the performance of the model in each approach and we carried out a performance comparison of the four approaches for term weighting.

Table 2: Test results

Model	Accuracy
Decision Tree	51.23%
Random Forest	52.73%
K-Nearest Neighbor	51.38%
Support Vector Machine	52.8%

Source: Authors.

As we can see in Table 1, the Support Vector Machine (SVM) has the highest accuracy rate with 52.8%, which will be used to forecast the VN-Index in the following step. We have this result because the larger the number of articles, the more precise the accuracy is. In general, the fluctuation in accuracy is due to the noise in the news articles we gathered.

Table 3: Test result of input data experiment

	From date	To date	Sample	Result
Vietstock	03/01-2016	09/02/2021	1274	55.87%
Vnexpress	24/02/2001	11/02/2021	3408	53.37%
Thanhvien	28/01/2013	12/02/2021	3784	53.52%
Cafef	15/10/2008	04/02/2021	3065	53.48%

Source: Authors' compilation.

Next, to prove the possibility to predict a stock trend in a real scenario, we used the SVM model to forecast the VN-Index with input data from each of 4 financial newspapers. The results above indicated that Vietstock was ranked with the highest result (55.87%). Therefore, we chose the input data for the model to be financial and economic articles from the Vietstock website.

To improve the test program results, we use the technique of changing the C and Gamma parameters. The test group replaced the 2 main parameters for the model: C with values from 0.1 to 1000 and Gamma from 0.0001 to 1, and the kernel as "rbf".

With accuracy of 60.1%, the experimental result showed that the stocks trend prediction through the financial news had a high correlation with the actual price.

After improving the SVM performance, we got more feasible results than from the first testing step. The above results show that the financial and securities news on online newspapers that every Vietnamese reads every day has an impact on the VN-Index's stock price.

5. Conclusion

In the current era, the development of Information Technology has led to the development of many fields (including economics, finance, etc.). Nowadays, people do not have to struggle with data collection because computer systems and data transmission networks are deployed on a global scale. However, this change has led to a rapid increase in the amount of information. There are many newspapers and websites that post news daily and hourly. If investors in the stock market want to fully understand the market, they must search for useful information and classify information to capture the overall picture of the market. But with a large amount of information, fast processing and effective classification with the traditional method is impossible. Techniques of using Machine Learning and text-mining applications to automatically classify

information have been very effective for this purpose.

In this research paper, the authors have surveyed and summarized a number of recent works on stock market forecasts, from methods based on pure numerical data to methods using text mining. After that, the research team initialized a research model and tested the VN-Index forecasting program based on Vietnamese news which is downloaded from popular news websites in Vietnam, as well as researching the type of words that contribute to the best VN-Index forecast results.

After testing the most popular models and techniques such as: Decision Tree, Random Forrest, K-Nearest Neighbor (KNN) and Support Vector Machines (SVM), the authors decided to choose to use the Support Vector Machines (SVM) for the stock market forecast problem, with a number of reasons as follows:

(i) The number of features of the classified document does not heavily affect the system in using SVM; (ii) SVM is capable of classifying well in the test data set without the need for a large training data set; and (iii) SVM has been used a lot in the published scientific works on the application of mining forecasts for the stock market and has been proven to have good results.

The first test results showed that the highest level of accuracy SVM reached was 52.8%. In order to improve the model to achieve higher reliability, the research team tested it a second time with a separate dataset of newspaper pages and finally selected the first dataset with the best results from the online newspaper from Vietstock.

Changing the data set range, the research team has forecast the price trend of the VN-INDEX to reach an accuracy rate of 60.1%. This proves that the news about the financial and securities situation in the popular press affects the price trend of the VN-INDEX. The forecast results of the model in the study are an objective and valuable reference source for managers and investors in making their decisions on the stock market. However, the forecast results of the study are not very high, according to the research team, partly due to the unpredictable volatility of

the stock market, and partly due to the lag of news data.

In addition, the model in the study can be used in other systems such as exchange rate forecasting, opinion analysis in product reviews, gold price forecasting, news classification by topic, and analysis in email type and contextual customizable news pages.

In comparison to prior studies, the prediction accuracy is quite poor. There are two key causes for this limitation, according to us: (1) The textual analysis tool for Vietnamese is limited in comparison to other popular languages, causing the prediction algorithm to be less efficient; (2) The Vietnamese stock market is developing, and available news has recently become more popular on the internet, but in the previous period was recorded as less available.

Finally, despite the interesting findings, this research contains a few caveats; particularly, the data limitation in terms of diverse sources and frequency. The results relied on news samples obtained from only one provider and news is not published on an everyday basis. Hence, one of the further extensions of this research is to extend news articles to capture a longer time frame, greater number of sources and with a higher frequency, which can possibly lead to better forecasting performance. Consequently, the size of news samples may become larger; hence, more feature selections are required to drop insignificant word vectors. Moreover, despite optimizing some parameters, there are still more parameters and functions in the algorithms; hence, fine-tuning each algorithm would be another way to further the future work.

References

- [1] Yu, R., He, X. & Liu, Y. "Glad: Group Anomaly Detection in Social Media Analysis," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10 (2) (2015) 1-22.
- [2] Pham T. Nghia, Tatiana K. Blokhina, "Improving the Efficiency of the Vietnam Stock Market," *Advances in Business-Related Scientific Research Journal*, 11 (1) (2020) 75-86.
- [3] Aas, K., & Eikvil, L., "Text Categorization: A Survey," *Technical Report*, Norwegian Computing Center, 1999.
- [4] Das, R.S., "Text and Context: Language Analytics in Finance," *Foundations and Trends(R) in Finance*, 8 (3) (2014) 145-261
- [5] Dang Hong Phu, "Application of Microsoft Time Series to Build a Forecasting System for Vietnam's Stock Market," Master Thesis, University of Information Technology, Vietnam National University HCMC, 2008.
- [6] Vu, Tien Thanh, et al., "An Experiment in Integrating Sentiment Features for Tech Stock Prediction in Twitter," *Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data*, 2012.
- [7] Le Van Tuan, "Application of Some Machine Learning Models in Forecasting the Direction of Movements of the Vietnamese Stock Market," Master Thesis, Thuongmai University, 2021.
- [8] Nassirtoussi, A. K. et al. "Text Mining for Market Prediction: A Systematic Review," *Expert Syst. Appl.*, 41 (2014) 7653-7670.
- [9] Kaya, M., and M. E. Karsligil, "Stock Price Prediction Using Financial News Articles," *2010 2nd IEEE International Conference on Information and Financial Engineering*, 2010, 478-82.
- [10] Joshi, Kalyani, Bharathi H. N, and Jyothi Rao, "Stock Trend Prediction Using News Sentiment Analysis," *International Journal of Computer Science and Information Technology*, 8 (3) (2016) 67-76.
- [11] Ayman E. Khedr, S.E.Salama, and Nagwa Yaseen, "Predicting Stock Market Behavior Using Data Mining Technique and News Sentiment Analysis," *International Journal of Intelligent Systems and Applications*, 9 (7) (2017) 22-30.
- [12] Tabari, N., A. Seyeditabari, T. Peddi, M. Hadzikadic, and W. Zadrozny, "A Comparison of Neural Network Methods for Accurate Sentiment Analysis of Stock Market Tweets," in *ECML PKDD 2018 Workshops*, edited by C. Alzate et al., Cham: Springer, 2019.