Preprints.org

Sentiments Extracted from News and Stock Market Reactions in Vietnam

Loan Thi Vu * , Dong Ngoc Pham , Hang Thu Kieu , Thuy Thanh Thi Pham

Article

Sentiments Extracted from News and Stock Market Reactions in Vietnam

Loan Thi Vu [1,*], **Dong Ngoc Pham** [2]**, Hang Thu Kieu** [1] **and Thuy Thi Thanh Pham** [1]

[1] VNU University of Economics and Business, Vietnam National University, Hanoi, Vietnam; hangkieu2001@gmail.com (H.T.K.), phamthanhthuyy1234@gmail.com (T.T.T.P.)

[2] VNU-University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam; dongpham120899@gmail.com

* Correspondence: loanvu.kttn@gmail.com

Abstract: News on the stock market contains positive or negative sentiments depending on whether the information provided is favorable or unfavorable to the stock market. This study aims to discover the news sentiment and classify news according to its sentiments with the application of PhoBERT, a Natural Language Processing model designed for the Vietnamese language. A collection of nearly 40,000 articles on financial and economic websites is used to train the model. After training, the model succeeds in assigning news to different classes of sentiments with an accuracy level of over 81%. The research also aims to investigate how investors concern about the daily news by testing the movements of the market before and after the news is released. The analysis results show that there is an insignificant difference in the stock price as a response to the news. However, investors tend to overreact to negative and positive news in the Vietnam stock market.

Keywords: behavioral finance; natural language processing; overreact; PhoBERT; textual analysis; underreact

JEL: G10; G14; G40; G41

1. Introduction

The pillar of traditional finance is the Efficient Market Hypothesis (EMH), which states that all relevant information is immediately reflected in the stock price. All investors are rational in making their trading decisions. However, because of its strict assumptions, traditional finance fails to explain the unusual, not "reasonable" phenomena that occur in the market such as the January Effect, the The-day-of-the-week effect, or the market bubbles leading to the stock market crash, ...Therefore, behavioral finance is developed to provide a different view from traditional finance by making the basic assumption that financial asset prices are not always driven by reasonable expectations of future returns. Behavioral finance supporters argue that human beings including market participants are entities of emotions, not just rational ones. Sentiment leads to overreaction, underreaction, or herding behaviors of investors. As the center of behavioral finance, sentiment including investor sentiment and news sentiment has been widely recognized and measured. According to Baker & Wurgler, 2007, finding an appropriate and realizable way to measure sentiment becomes a challenging and necessary task for researchers because of the increasing role of sentiment on stock pricing.

In Vietnam, investor sentiment has been analyzed to explain the investor's behaviors but research on this topic mainly focuses on constructing a comprehensive sentiment index (Phan et al. 2021). Despite its popularity, the sentiment index fails to reflect the emotions of investors in response to news and public information. There is a research gap in news sentiment analysis because of the lack of a well-trained textual analysis model in the Vietnamese language that focuses on financial news. Any news or articles posted daily on websites bring moods that depend on the content of the information. It can be good news if it is favorable, and it can be bad news if it is unfavorable to the stock market. Recently, with the development of internet access, websites specialized in stock market

analysis provides updated and adequate market information which can rapidly influence investors' expectation and behaviors. Therefore, sentiment from new information should be revealed and analyzed to understand its effects on the investors. Aiming at extracting sentiment from news, this study uses a textual analysis approach to extract sentiment by constructing Natural Language Processing model to read, analyze texts, and classify articles in the Vietnamese language according to their sentiments.

The findings of the study can make contributions to the investors as well as the firm's managers. For investors, revising trading decisions as a response to news helps them to know whether they are overreacting or underreacting to new information. For a company, the study's findings support them in assessing the role of relevant news on a company's stock price and examining the market expectation of the company. Moreover, this study is one of the first works which successes in training the Natural Language Processing model with a large dataset in the Vietnamese language. Therefore, the model after training can be applied easily and widely for securities companies, individual stock analyzers, and researchers in this field.

2. Literature Review

Behavioral finance has tried to examine sentiment in different aspects to explore the effects of sentiment on stock return patterns in the global market. The sentiment is defined as the belief about the expected returns and risks related to a stock without fact justification (De Long et al. 1990). Especially, this kind of belief may lead to pricing errors in a great number of traders. In general, the sentiment makes investors become bullish or bearish in their trading behaviors (Brown and Cliff 2004). According to Pandey, & Sehgal (2019), the existence of sentiment cannot be denied but the matter is about how to identify and measure this factor. The sentiment is recognized by a market-based method using comprehensive indexes or a survey-based method using questionnaires surveys. It is also identified by a text-based method of interpreting investors' messages on social media and financial news articles (Shen et al. 2023). In other words, sentiment can be measured by direct, indirect, or any approach that the sentiment-related data can be collected.

Sentiment Measures

In a direct method, the sentiments of investors can be collected by surveys taken regularly or by making an analysis of information-searching behaviors on the internet. According to Brown and Cliff (2005), the survey's results are an appropriate proxy for investor sentiment. The most popular surveys are the Confidence survey for Michigan consumers (Aggarwal 2018; Qiu and Welch 2004); Investors Intelligence - II (Brown and Cliff 2005; Verma and Verma 2008); and the American Association of Individual Investors – AAII (Fisher and Statman 2000; Brown and Cliff 2004; Verma and Soydemir 2009). The AAII sentiment index, for example, measures the percentage of respondents who are bullish, bearish, or neutral. It is conducted on a weekly basis to collect members' views on the stock market for the coming six months. The shortcomings of using direct surveys are mainly from the concerns of the difference between actual investors' behaviors and what they respond to in the survey. In addition, the value of the survey results depends significantly on the size of respondents as well as the response frequency of the survey (Aggarwal 2019).

The indirect approach to measuring investor sentiment which is applied widely is the construction of a sentiment index from several proxies. Baker & Wurgler (2006) and Wurgler & Baker (2007) create a composite sentiment index from six proxies: the closed-end fund discount, the market turnover, the number of IPO and returns of the first-day trading of IPO stocks, the new issuance share, and the dividend premium. Baker et al. (2012) remove 3 variables including the closed-end fund discount, the new issuance share, and the dividend premium from the set of proxy components provided by Baker & Wurgler (2006), and add volatility premium as a new proxy for sentiment index. In the work of other researchers, investor sentiment is also recognized by interpreting some trading activities including margin borrowing, short interests change, and short sales of specialists (Brown & Cliff 2004). The investor sentiment index succeeds in presenting the market sentiment over a period, but it fails to measure how rapidly investors react to new information in the market.

Thanks to the development of machine learning and data mining approaches, textual analysis has been applied in reading, interpreting, and extracting sentiment from several online platforms. Sentiment can be discovered by analyzing the emotions behind the investor's comments on media such as message boards, Facebook, or Twitter. Moreover, sentiment can be taken from stock market news or economics news because news articles can transfer tone, and emotions through the texts inside. Therefore, news articles reflect the expectations of investors on the market in general and specific stocks according to the information they provide. This measure of sentiment is called media-based investor sentiment (Sun et al. 2016; Nguyen et al. 2015).

Antweiler and Frank (2004) measure sentiment by collecting and interpreting a great number of messages on finance websites such as Yahoo! Finance.   Others extract sentiment from investor's posts on media including Blog, Facebook, or Twitter (Barber and Odean, 2008; Bar-Haim et al. 2011; Bollen et al. 2011; Dougal et al. 2012). According to Li et al. (2020), there is increasing interest in textual sentiment analysis among researchers in financial behavior because this method can reduce bias that may be found in the survey-based sentiment approach (Schumaker and Chen 2009; Renault 2017; Shapiro et al. 2022). Liu et al. (2023) utilize news sentiment related to each firm to be a proxy of investor sentiment on the stock of that firm. The "Overall Sentiment Score" ranging from -1 to 1 reflects the level of optimism and pessimistic. The higher the score is put, the greater the investor's optimism about the stock.

News Sentiment and Investor Response

As an emotional entity, an investor may not be rational in interpreting new information to make essential responses to it. Behavioral finance recognizes bias in belief according to how the investors behave when news appears. *Barberis et al. (1998)* believe that investors overreact to information in some cases and underreact in other cases.

According to Montier (2002), the stock market tends to underweight the fundamental information on dividend payments or a firm's earnings. Therefore, the investors are recognized to have conservatism bias when they anchor their investments solely on their forecasts about the company. Conservatism bias makes investors react very slowly to the news. For example, when there is unfavorable information about the price of stocks that investors are holding, investors may hesitate to sell. Investors may hold stocks for too long before being forced to sell after suffering unnecessary losses. Meanwhile, De Bondt, and Thaler (1985) state that stocks having extremely low returns in the past usually obtain significantly higher returns in the future than the ones having high returns in the past. De Bondt and Thaler (1985) also explain this phenomenon by conservatism bias as the investors are too pessimistic about the bad stock's performance in the past. When they realize that they have overreacted to the bad news, their reversals can increase the stock returns.

Many researchers divide new information into different groups according to the moods or sentiments of the information (Vu et al. 2012; Nguyen et al.2015; Renault 2017; Costola et al. 2020). They all find a close relationship between sentiment from news and the movements of the stock market. Sentiments extracted from general new information (Nguyen et al.,2015; Renault, 2017) or news on COVID-19 pandemic (Costola et al. 2020; Baker et al. 2020) have a significant ability to predict future stock returns. Feng et al. (2022) explore the relationship between news sentiment and stock market volatility in Japan. The research findings show that news sentiment has realizable effects on the volatility of stock returns. Liu et al. (2023) suggest that firm-specific news sentiment can boost stock trading activity if it is optimistic. However, news with a pessimistic tone has stronger power in predicting stock returns than one with an optimistic tone.   Similarly, *Shen et al. (2022)* also present the role of the news tone on the volatility and stock returns in China.

Our research is expected to make contributions at some point. First, it is one of the first studies training a language model with large news articles on financial and economic websites in Vietnamese to analyze the news sentiment. Although there is a growing number of studies measuring sentiment by analyzing texts globally, limited research has been taken in Vietnam in this field. The complication of the Vietnamese language prevents the application of the existing models which are working with texts in English. Second, this study provides insights to explain the reactions of investors to news to

understand investors' behaviors in the Vietnam stock market. Therefore, a well-trained model can be necessary for researchers, investors, stock analysts, and the firm's managers.

3. Methodology

The news sentiment was extracted by training a Natural Language Processing model – NLP model to interpret texts in the daily news. In this study, news on the stock market, domestic economics, and international finance was collected from high-traffic financial and economic online platforms in Vietnam. News on specific firms was excluded because the study aimed to test the reactions of the market on general news instead of the news on any specific stock.

There were two steps in the NLP model to discover news sentiment using textual analysis. In the first step, news articles from the websites were collected for model training. News was taken from 3 websites: Cafef.vn, Vneconomy.vn, and Stockbiz.vn from July 2021 to July 2022. The Pandas library was applied to crawl all the articles needed. After crawling and processing, the set of 40,000 articles was separated into a training sample (70%) and a testing sample (30%). The training sample was then divided into 3 parts and used for training, adjusting, and assessing. For constructing the training dataset, a set of 2,738 news was labeled into 2 groups having positive sentiment and having negative sentiment. The news was labeled manually by the researchers to avoid mistakes in sentiment recognition. News was classified into different sentiment groups by recognizing words showing emotions not only in the title but also in the whole content of the news. An example of news with positive sentiment was the one showing the positive expectation of the firm's earnings growth rate in 2021 from Dragon Capital, an Investment Fund in Vietnam. A collection of texts presenting optimistic sentiment were: "a significant increase", "attractive", "net buying", and "price ceiling increase". However, the news was considered pessimistic if it included "market washout", "net selling", "floor price", "decrease", and "shrinkage". News was labeled as 1 for having a negative sentiment and 2 for having positive sentiment.

In the second step, labeled news in the previous step was used for training the NLP model. The NLP model is an Artificial Intelligent application to analyze texts resembling a human brain. As all news was in the Vietnamese language, PhoBERT, an expansion of the BERT model was trained in this study. The BERT model is a shortcut to the Bidirectional Encoder Representation from the Transformers model developed by J. Devlin et al. (2019). For news interpreting, the Spacy Library was implemented to synthesize the opinion of each paragraph in a news and give a unified point of view of the whole news. The Spacy Library separated each sentence and combined them into a paragraph that met the standard of no more than 200 words. However, each paragraph sometimes did not represent the main content of the entire news. Meanwhile, the news headline was an important piece of information that summarised the main content that the whole text referred to. Thus, news headlines were incorporated in front of each paragraph to give more context and main content to the paragraph

The model's performance was assessed by its ability to identify the sentiment of news in the testing data and classify news according to its sentiment. The classification accuracies were measured by the estimations of Accuracy, Precision, Recall or Sensitivity, Specificity, and F1-score.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

TP: True positive, the number of news containing positive sentiment is classified as Positive
TN: True negative, the number of news containing negative sentiment is classified as Negative
FP: False positive, the number of news containing negative sentiment is classified as Positive
FN: False negative, the number of news containing positive sentiment is classified as Negative.

$$Presision = \frac{TP}{TP+FP} \qquad (2)$$

$$Recall = Sensitivity = \frac{TP}{TP+FN} \qquad (3)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (4)$$

$$F1 - score = \frac{2 Precision * Recall}{Precision + Recall} \qquad (5)$$

The accuracy is calculated in the whole dataset. It demonstrates the overall ability of the model to correctly classify news into a Negative sentiment group or a Positive sentiment group. The higher the accuracy the better the model in sentiment prediction. Precision measures the ratio of True Positive on the sum of True Positive and False Positive while Recall (Sensitivity) determines the fraction between the True positive and the whole actual Positive class. Similarly, the specification measures the relationship between True Negative and the whole actual Negative class. F1-score is the harmonic mean of Precision and Recall. As it is a representative of Precision and Recall, F1-score is an appropriate accuracy estimate of a classification model.

To understand the relationship between news and the market, the study also tried to reveal the changes in the stock market as a response to positive and negative news. The date that the news was released was the event date. The returns of the VN30 index were selected to be representative of market returns. VN30 index presented the comprehensive price index of 30 stocks leading in market capitalization in the Hochiminh stock exchange. Returns of the VN30 index were calculated within 10 days, including 5 days before and 5 days after each event date to examine the rapid reactions of the market to news. The market returns were also measured within 60 days: 30 days before and 30 days after the event date for checking the market movements in longer periods. The t-test was performed to explore the significant return changes from the effects of news sentiment. Specifically, there were two hypotheses used for testing.

H$_{01}$. *There is no difference in the market returns before and after the day releasing positive news.*

H$_{02}$. *There is no difference in the market returns before and after the day releasing negative news.*

4. Analysis Results and Discussion

The Vietnam stock market started its first operation in 2000. Despite being in the early stage of development, it is one of the fastest-growing markets in Asia. Currently, the number of investors in the market is about 2.77 million and the market capitalization accounts for 82,15% of GDP. The Vietnam stock market is also recognized as one of the 10 stock markets with the highest recovery rate from the COVID-19 pandemic. The new information about the market and the economy is updated daily and investors can access that information on several web platforms to form their investing decisions.

The daily news is collected from 3 popular websites in the stock market and Vietnam economy including Cafef.vn, Vneconomy.vn, and stockbiz.vn from July 2021 to July 2022. The number of articles on each website is shown in Table 1.

Table 1. Summary of the articles collected.

Websites	Number of Articles	Period
Cafef	30,471	21/10/2021-28/06/2022
Vneconomy	2,837	01/09/2021-28/06/2022
Stockbiz	9,374	12/07/2021-27/06//2022

Most of the articles are collected from Cafef.vn website and Stockbiz.vn website because these two websites obtain high traffic with many daily visitors. Vneconomy.vn is also a popular website providing updated domestic and international economic news. Figures 1 and 2 provide more details about the changes in the number of daily news during the studied period.
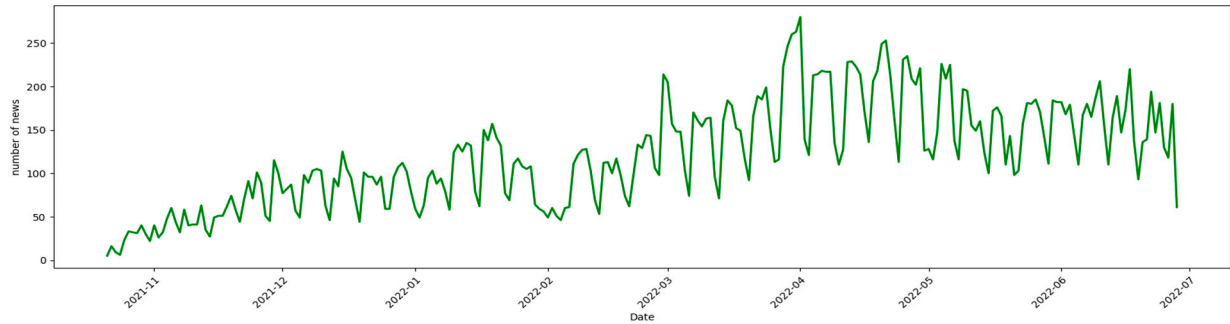
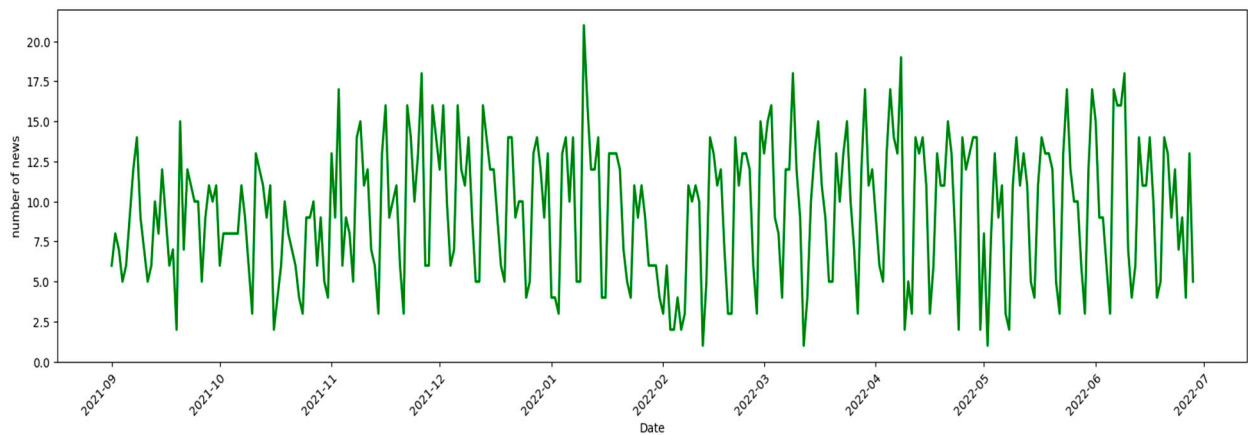**Figure 1.** Descriptions of daily news numbers on Cafef.vn website.

Figure 2. Descriptions of daily news numbers on Stockbiz.vn website.

Articles are selected in different categories such as General news, stocks, properties, business, international finance, macroeconomics, life, and market (Cafef.vn). Meanwhile, articles gathered from Vneconomy.vn are classified as general news, stock trading, property, business, international finance, life, and market. Figures 3 and 4 present the news categories on Cafef.vn and Vneconomy, respectively. The number of articles in each category varies according to the information provided. If the texts in an article do not contain moods or only involve a specific stock that may not impact the whole market, that article is not collected to have better undersanding of market reactions to new information. As can be seen in Figures 3 and 4, because of the unstable global economic condition from 2021, the articles mainly come from News, international finance, World economy categories.
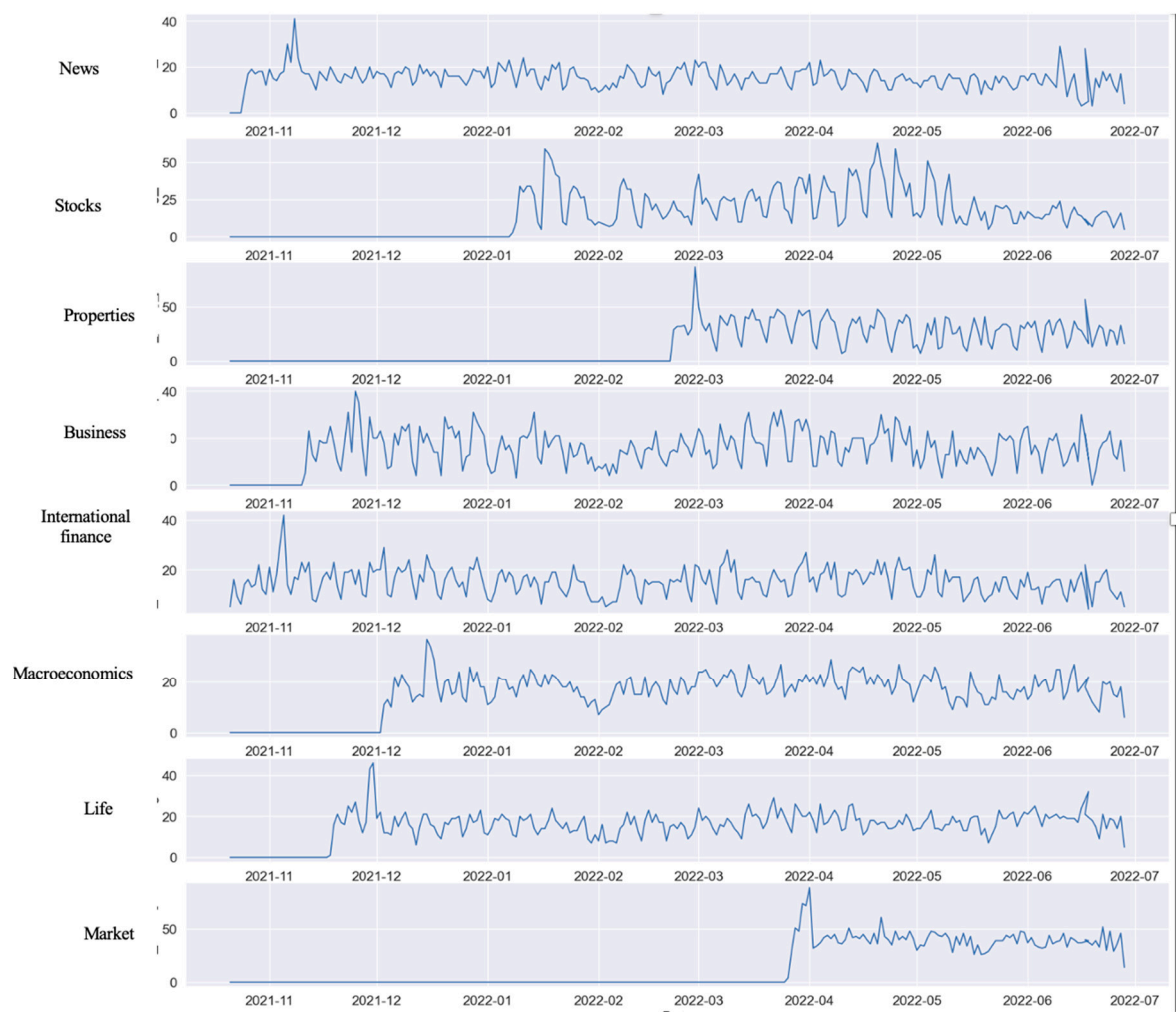
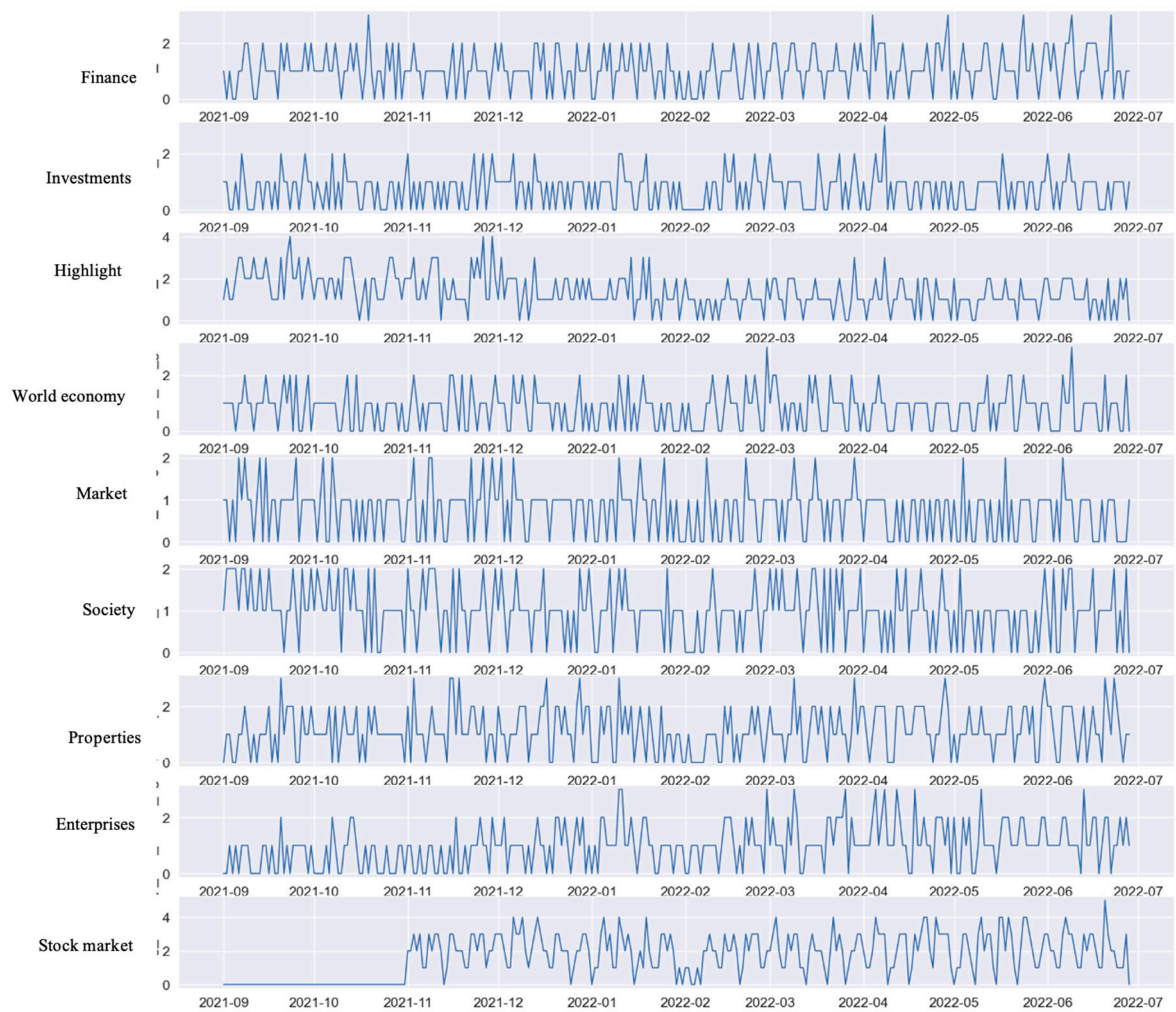**Figure 3.** Article categories on Cafef.vn website.

Figure 4. Article categories on Vneconomy.vn website.

For training the NLP language model, news on the training data is labeled according to the sentiments that the messages provide. Table 2 describes the news which is labeled for model training. News that is related to a specific company or industry is not chosen, for example, news with the title "DAG is estimated to achieve a revenue increase of 15% in the first quarter of 2022 compared to the same period last year". However, news "HPG hit the ceiling with record liquidity, VN-Index broke through nearly 20 points, surpassing the 1,500 points" is labeled positive because it presents an optimistic view of the market.

Table 2. Examples of news labeling.

	Title	Label
1	Trillions of billions of VND poured into Vietnamese stocks through ETFs	2
2	The Russia-Ukraine conflict added fuel to the fire, and the "ghost of inflation" began to haunt the Vietnamese stock exchange: Worried about leaving?	1
3	Dragon Capital: "Investors should not worry about short-term fluctuations from the Russia-Ukraine event but focus on the long-term prospects of the market"	2
4	HPG hit the ceiling with record liquidity, VN-Index broke through nearly 20 points, surpassing the 1,500 points	2
5	Domestic investors opened more than 210,000 new stock accounts in February	2
6	Government's financial strategy to 2030: Stock market capitalization reaches 120% of GDP	2
7	Unable to overcome the selling pressure, nearly 360 stocks were "on the floor", VN-Index dropped 60 points, lost the 1,270 points	1
8	Trading session 10/2/2022: Foreign investors suddenly net sold 740 billion dong on HoSE, selling hundreds of billions of VIC, HPG	1
9	Fertilizer, Petrol stocks all hit the floor, VN-Index lost more than 20 points in the first trading day of the week	1
10	Agriseco Research: Statistics since 2000, if inflation is below 10%, securities are still the most suitable investment channel	2

Note: 1 for negative news; 2 for positive news.

The performance of the NLP model is assessed by the power to classify news into 2 groups according to the emotion extracted from the content of the news. As shown in Table 3, there is not much accuracy difference in classifying any news into one of 2 groups: negative group and positive group. All estimates (Precision, Recall, and F1-score) described in Table 3 are bigger than 81% which demonstrates the significant power of the model to extract the sentiment from an article.

Table 3. Summary of the Model's Performance.

	Precision	Recall	F1-Score
Negative	0.817	0.825	0.821
Positive	0.817	0.809	0.813

For visualization, the receiver operating characteristic curve (ROC curve) is constructed to further examine the performance of the model. Each point in the ROC curve presents a combination of sensitivity and specificity corresponding to a threshold. The ROC curve should be closer to the upper left corner to obtain smaller error rates in classifying or predicting. The Area Under the ROC curve (AUC) is also a measure of the prediction accuracy of the model. The highest value of 100% shows that the model perfectly differentiates between positive and negative news. As presented in Figure 5, the AUC is nearly 90% showing that the model obtains a substantial level of accuracy in extracting the sentiment from any article.
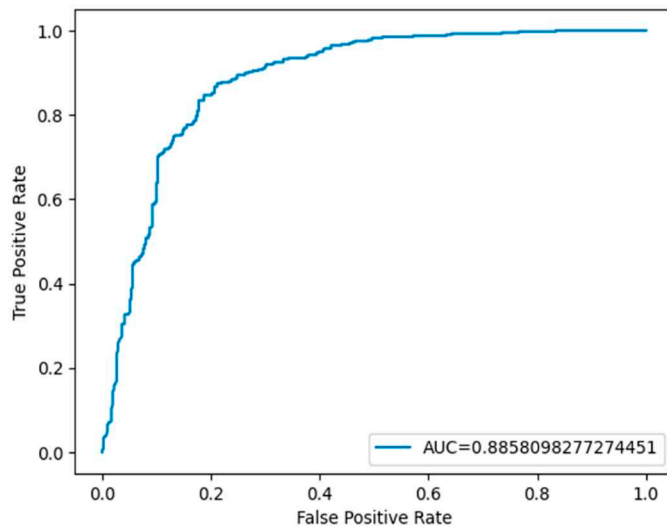
Figure 5. ROC curve.

The t-test, a statistical technique is applied to examine the effect of the new information on the movements of the VN30-Index, a representative of market changes. VN30-Index tracks the prices of 30 stocks having the largest market capitalization on the Vietnam stock market. A group of 100 event dates is selected. It includes 50 dates having positive news and 50 dates announcing negative news. The objective of the t-test is to find the statistical difference between market returns before and after the event dates. The first t-test measures the instant or rapid market reactions as it compares returns 5 days before the event dates with market return 5 days after the event dates. The second t-test works with the data sample consisting of a group of daily market returns within 30 days before the event dates and a group of market returns calculated 30 days after the event dates.

Tables 4 and 5 describe the two-sample t-test for positive news sentiment and negative news sentiment, respectively. Each table presents the mean and standard deviation of market stock returns calculated in different groups. The mean difference is calculated by taking the difference between the mean return after and before event dates. According to the tables, there is a difference between the market returns before and after the date of the news announcement whether it is positive or negative news, but the difference is not statistically significant.

Table 4 shows that the market tends to react strongly in the short-term with positive news sentiment as the return means market returns difference calculated within 5 days is higher than that of 30-day returns. However, Table 5 shows that the market decreases rapidly in response to negative news sentiment and recovers soon after that.

Table 4. Two-sample t-test with equal variances – positive news sentiment.

Group │	Mean	Std. err.	Std. dev.	[95% conf. Interval]	
0: 30 days after event date					
1: 30 days before event date					
0	.0001801	.0006097	.017062	-.0010169	.001377
1	-.0008694	.0004649	.0157799	-.0017816	.0000428
Combined	-.0004447	.0003709	.0163147	-.0011721	.0002826
Difference	.0010495	.0007555		-.0004321	.0025311
diff = mean(0) - mean(1)				t =	1.3892
H0: difference = 0				Pr(|T| > |t|) = 0.1649	
0: 5 days after event date					
1: 5 days before event date					
0	-.0004831	.0012393	.0172169	-.0029275	.0019613
1	-.0019054	.0009994	.017682	-.0038719	.0000611

Combined	-.0013629	.0007781	.0175026	-.0028916	.0001658
Difference	.0014222	.0016022		-.0017256	.0045701

<div align="center">

diff = mean(0) - mean(1)

H0: diff = 0

</div>

t =   0.8877

Table 5. Two-sample t-test with equal variances – negative news sentiment.

Group |	Mean	Std. err.	Std. dev.	[95% conf. Interval]	
0: 30 days after event date					
1: 30 days before event date					
0	.001198	.0004864	.0153491	-.0021524	-.0002436
1	-.000231	.0004955	.0163078	-.0012034	.0007413
Combined	.0006943	.0003478	.0158593	-.0013764	.0000122
difference	.000967	.0006961		-.0023321	.0003982

<div align="center">

difference = mean(0) - mean(1)

t = -1.3891

H0: difference = 0

Pr(|T| > |t|) = 0.1649

0: 5 days after event date

1: 5 days before event date

</div>

0	-.0011582	.0010494	.0165927	-.0032251	.0009086
1	-.0007067	.0010371	.0179626	-.0027476	.0013342
Combined	-.0009119	.0007393	.0173392	-.0023642	.0005404
difference	-.0004515	.0014861		-.0033706	.0024676

<div align="center">

difference = mean(0) - mean(1)   t =   -0.3038

</div>

By reading and analyzing texts on the title and the content of the news in various fields that can be related to the stock market, the NLP model is trained and able to extract sentiments from any news article. The analysis shows that the model constructed in this research can classify new information in the Vietnamese language into the correct groups of sentiments with an accuracy level of over 81%. Other works conducted by Vu et al. (2012), Nguyen et al. (2015), Renault (2017), and Costola et al. (2020) also present the outperformance of the text analyzing model in extracting the sentiments in new information. For example, using BERT, the NLP models built by Renault (2017) can obtain an accuracy level from 75.24% to 90.75%.

The sentiment from the news is a factor to drive the change in the investor's expectations. As a result, the stock prices fluctuate. However, the reaction of investors is a complicated process that depends on the analysis of the information's effects on the stocks that they are holding. Investors may be irrational in analyzing new information and easy to be over-optimistic or over-pessimistic. As the t-test results describe, the market index increases quite impressively right after the announcement of good news, and the increase is smaller over time. Regarding the negative news, the market declines in response to the negative news within 5 days after the news is announced and recovers within 30 days. This finding can be considered evidence of overreaction to positive news and negative news which can be seen in the work of Montier (2002). However, more event dates and stock indexes should be collected for future research to obtain an adequate estimate of market response to the positive and negative information.

5. Conclusions and Future Research

New information can be favorable or unfavorable to investors and it requires careful assessment. This study constructs an NLP model to discover the moods of the news posted daily on 3 websites on finance and economy in Vietnam. The model after being trained obtains a high level of accuracy (over 81%) in assigning an article to a positive group or negative group by reading its title and

content. The study also explores the relationship between sentiment and market responses and finds that the market seems to overreact to positive news and underreact to negative news. The results of the study can contribute to the investors, researchers, and managers of companies whose stocks are trading in the Vietnam stock market and the State Securities Commission in Vietnam.

The NLP model in this paper can be applied by investors to extract the sentiment of any news article. It can be a tool for assessing the wide market or economic condition that may be favorable or unfavorable to stock investment. Therefore, investors can obtain an understanding of the current market without reading all news in a day. Researchers who are interested in textual analysis may find the results of this study important to them because it provides a well-trained model in the Vietnamese language.

From the evidence of investor overreaction to news sentiment, investors especially individual ones should review their investment strategies to avoid any potential loss. It also supports the company's managers to understand the behaviors of their stockholders so that the managers can find a suitable time to make financial decisions regarding dividend payouts, stock splits, and new issuance. For the State Securities Commission in Vietnam, which is a regulatory authority supervising the market, the research's results may be useful for making policies to prevent market manipulation from spreading fake news.

For future research, there should be more than a stock index used to explore the reactions of the broader market to the news. Examples of current stock indexes in Vietnam that can be selected are VnIndex, measuring the price changes of all stocks traded on HoChiMinh Stock Exchange, and HNX-Index, accounting for the prices of all stocks on Hanoi Stock Exchange. Regarding news sentiment, future research should classify news into 3 groups of sentiment: positive, negative, and neutral.

Conflicts of Interest: The authors declare no conflict of interest.

References

Aggarwal, C.C. 2018. Opinion Mining and Sentiment Analysis. In: Machine Learning for Text. *Springer, Cham*, 413-434. https://doi.org/10.1007/978-3-319-73531-3_13

Antweiler, W., Frank and M. Z. 2004. Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance* 59, 1259-1294. http://dx.doi.org/10.2139/ssrn.282320.

Baker, M., and Wurgler, J. 2004. A catering theory of dividends. *The Journal of Finance*, 59(3): 1125-1165.

Baker, M. & Wurgler, J. & Yuan, Y. 2012. Global, Local, and Contagious Investor Sentiment. *Journal of Financial Economics*, 104, 272-287. 10.2139/ssrn.1449929.

Baker, S.R., Bloom, N., Davis, S.J., Kost, K., Sammon, M., Viratyosin, T. 2020. The unprecedented stock market reaction to COVID-19. *Rev. Asset Pric. Stud*, 10 (4), 742–758. http://dx.doi.org/10.1093/rapstu/raaa008.

Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., and Goldstein, G. 2011. Identifying and following expert investors in stock microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP'11,1310–1319, Stroudsburg, PA, USA. Association for Computational Linguistics.

Barberis, N., Shleifer, A., Vishny, R. 1998. A model of investor sentiment. *Journal of Financial Economics*, 49, 307–343.

Bollen, J., Mao, H., and Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2, 1–8.

Brown, G. W., and Cliff, M. T. 2004. Investor sentiment and the near-term stock market. *Journal of Empirical Finance,* 11(1): 1–27.

Brown, G. W., and Cliff, M. T. 2005. Investor sentiment and asset valuation, *Journal of Business,* 78(2):405–440

Costola, M.& Nofer, M. & Hinz, O. & Pelizzon, L.2020. Machine Learning Sentiment Analysis, Covid-19 News and Stock Market Reactions, *SSRN Electronic Journal*, 10.2139/ssrn.3690922.

Crawford, I. & Wang, Z.2010. Is the Market Underreacting or Overreacting to Open Market Share Repurchases? A UK Perspective. *Research in International Business and Finance*, 26. 10.1016/j.ribaf.2011.05.002.

Daniel, K.D., Hirshleifer, D., Subrahmanyam, A.1998. Investor psychology and security market under- and over-reactions, *Journal of Finance,* 53, 839–886.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J.1990. Noise Trader Risk in Financial Markets, *The Journal of Political Economy*, 98, 703-738.

Dougal, C., Engelberg, J., García, D., Parsons and Christopher, A.2012. Journalists and the Stock Market. *Review of Financial Studies*, 25 (3), 639–679.

Du, S., Liu, H., 2015. The overconfident trader does not always overreact to his information, *Economic Modelling*, 46, 384-390. https://doi.org/10.1016/j.econmod.2015.02.009

Dyl, E. & Yuksel, H. & Zaynutdinova, G. 2019. Price reversals and price continuations following large price movements, *Journal of Business Research*, 95. 1-12. 10.1016/j.jbusres.2018.08.036.

Wurgler, J. & Baker, M. 2007. Investor Sentiment in the Stock Market, *Journal of Economic Perspectives*, 21, 129-152, 10.2139/ssrn.962706.

Barberis, N., and Shleifer, A. and Vishny, R. W. 1997. A Model of Investor Sentiment. *NBER Working Paper*, 5926, Available at SSRN: https://ssrn.com/abstract=225707

Barber, M.B., Odean, T. 2008. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies*, 21(2): 785–818. https://doi.org/10.1093/rfs/hhm079.

Feng, L. & Fu, T. & Shi, Y. 2022. "How does news sentiment affect the states of Japanese stock return volatility?." *International Review of Financial Analysis*, Elsevier, vol. 84(C).

Fisher, K.L. & Statman, M. 2000. Investor Sentiment and Stock Returns. Financial Analysts Journal, 56, 16-23.

Li, X. & Wu, P. & Wang, W. 2020. "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong." *Information Processing & Management*, 57(5). 102212. 10.1016/j.ipm.2020.102212.

Liu, J. & Wu, K. & Zhou, M. 2023. "News tone, investor sentiment, and liquidity premium." *International Review of Economics & Finance,* Elsevier, vol. 84(C), 167-181.

Montier, J., 2002. *Behavioural finance: Insights into irrational minds and markets*, Chichester: John Wiley and Sons Ltd.

Nguyen, T.H., Shirai, K., Velcin,J. 2015. Sentiment Analysis on Social Media for Stock Movement Prediction. *Expert Systems With Applications*,doi: 10.1016/j.eswa.2015.07.052

Pandey, P. & Sehgal, S. 2019. Investor Sentiment and its Role in Asset Pricing: An Empirical Study for India, *IIMB Management Review,* 31. 10.1016/j.iimb.2019.03.009.

Phan, T. & Bertrand, P. & Phan, H.& Vo, X.V. 2021. Investor sentiment and stock return: evidence from Vietnam stock market. *The Quarterly Review of Economics and Finance*, 87, 10.1016/j.qref.2021.07.001.

Qiu, L.X. and Welch, I. 2004. Investor Sentiment Measures, *NBER Working Paper,* 10794. Available at SSRN: https://ssrn.com/abstract=595193

Renault, T. 2017. Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, 84, 25-40

Savor, P. 2011. Stock Returns After Major Price Shocks: The Impact of Information. *Journal of Financial Economics,* 106. 10.2139/ssrn.1306233.

Schumaker, R. P., and Chen, H. 2009. Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Trans. Inf. Sys,* 27(12),1–19.

Shapiro, A. & Sudhof, M. & Wilson, D. 2017. "Measuring News Sentiment." *Federal Reserve Bank of San Francisco*, Working Paper Series. 01-A2. 10.24148/wp2017-01.

Shen, S. & Xia, L.& Shuai, Y. & Gao, D. 2022. "Measuring news media sentiment using big data for Chinese stock markets." *Pacific-Basin Finance Journal*, Elsevier, vol. 74(C).

Tetlock, P. 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. Journal of Finance, 62, 1139-1168. 10.2139/ssrn.685145.

Sun, L. & Najand, M. & Shen, J. 2016. "Stock return predictability and investor sentiment: A high-frequency perspective." *Journal of Banking & Finance,* Elsevier, vol. 73(C), 147-164.

Vega, C. 2004. "Stock Price Reaction to Public and Private Information." *Journal of Financial Economics*, 82, 103-133. 10.1016/j.jfineco.2005.07.011.

Verma, R. & Verma, P., 2008. Are survey forecasts of individual and institutional investor sentiments rational?. *International Review of Financial Analysis*, 17(5), 1139-1155.

Verma, R. & Soydemir, G.2000. The impact of individual and institutional investor sentiment on the market price of risk, *The Quarterly Review of Economics and Finance*, 49, 1129-1145. 10.1016/j.qref.2008.11.001.

Vu, T. & Chang, S. & Ha, Q.& Collier, N. 201). "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*, 23-38.