



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 Department of Banking and Finance, VNU University of Economics and Business, Vietnam National University, Hanoi 100000, Vietnam; hangkieu2001@gmail.com (H.T.K.); phamthanhthuy1234@gmail.com (T.T.T.P.)

² Department of Information Technology, VNU University of Engineering and Technology, Vietnam National University, Hanoi 100000, 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 news sentiments 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 are concerned with the daily news by testing the movements of the market before and after the news is released. The results of the analysis show that there is an insignificant difference in the stock price as a response to the news. However, negative news sentiments can alter the variance of market returns.

Keywords: behavioral finance; Natural Language Processing; overreact; PhoBERT; textual analysis; underreact

JEL Classification: G10; G14; G40; G41



Citation: Vu, Loan Thi, Dong Ngoc Pham, Hang Thu Kieu, and Thuy Thi Thanh Pham. 2023. Sentiments Extracted from News and Stock Market Reactions in Vietnam. *International Journal of Financial Studies* 11: 101. <https://doi.org/10.3390/ijfs11030101>

Academic Editors: Sahbi Farhani and Muhammad Ali Nasir

Received: 27 June 2023

Revised: 28 July 2023

Accepted: 2 August 2023

Published: 7 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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 day-of-the-week effect, or the market bubbles leading to the stock market crash. Behavioral finance was 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, rather than purely rational. Sentiment leads to the overreaction, underreaction, or herding behaviors of investors. As the center of behavioral finance, sentiments including investor sentiments and news sentiments have been widely recognized and measured. According to Baker and Wurgler (2006), finding an appropriate and realizable way to measure sentiments becomes a challenging and necessary task for researchers because of the increasing role of sentiments on stock pricing. Recently, with the development of public media and its attraction to investors, a news article about the financial market does not merely provide information but can also affect moods depending on the content of the messages inside. It can be positive news if it is favorable, and it can be negative news if it is unfavorable to the stock market. Therefore, sentiments from new information should be revealed and analyzed to understand its effects on the investors.

In Vietnam, sentiments have 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 take into account the emotional response of investors to public information related to the stock market. The use of textual analysis for news sentiments has been applied globally by many scholars such as Nguyen et al. (2015); Renault (2017); Huang et al. (2020); Petropoulos and Siakoulis (2021); and Liu et al. (2023). However, in Vietnam, there is a research gap in news sentiments because of the lack of a well-trained textual analysis model in the Vietnamese language. The complication of the Vietnamese language may prevent the application of the existing models which are working with texts in English. Therefore, this study expects to be one of the first works performing a Natural Language Processing model to read, analyze, and classify articles on the finance and security market in the Vietnamese language according to their sentiments.

Studies on news sentiments have reported the relationship between news sentiments and stock price movements, volatility, and trading volume. News sentiments even have the ability to predict the returns of specific stocks as well as the market index (Li et al. 2020). However, the level and the sign of the effects depend on the market characteristics and the time frame used for analysis. Therefore, another objective of this work is to assess the roles of news sentiments on the Vietnamese stock market—the market is still in the early stage of its development history.

Our research is expected to make contributions at some point. Firstly, it is one of the earliest studies training a language model with large news articles on financial and economic websites in Vietnamese to analyze news sentiments. Secondly, this study provides insights to explain the reactions of investors to news to understand investors' behaviors in the Vietnamese stock market. Therefore, a well-trained model is necessary for researchers, investors, stock analysts, and the firm's managers.

2. Literature Review

Efficiency exists when the stock prices fully reflect all available information (Fama 1970) and the stock prices follow a random route. From the general definition of market efficiency, researchers have conducted several tests for measuring the level of efficiency. For example, Chow et al. (2016) perform tests focusing on the variance ratio to assess the role of market liberalization in improving market efficiency in Latin American countries. The market efficiency theory assumes that the risk premium of a stock solely depends on its systematic risk because investors hold a well-diversified portfolio. Investors cannot earn higher returns without accepting more risk. However, in studies taken to recognize market efficiency, practitioners and scholars have discovered anomalies and market inefficiency or inadequacies. Schwert (2003) discusses the return anomalies which are the deviations from the returns expected from the traditional equity pricing models. Schwert (2003) mentions a few factors affecting the stock returns besides the stock's systematic risk such as the company's size, value, weekend, and dividend yield. Cho et al. (2007) also examine the the-day-of-the-week effect and discovered that bad news on Monday makes the returns worse than bad news on the other days of the week. Similarly, Chui et al. (2020) tried to analyze the calendar anomalies by interpreting the Halloween effect on Western markets.

Behavioral finance has tried to examine sentiments in different aspects to explore the effects of sentiments on stock return patterns in the global market. Sentiment is defined as the belief about the expected returns and risks related to a stock without fact justification (De Long et al. 1990). In particular, this kind of belief may lead to pricing errors in a great number of traders. In general, sentiments make investors become bullish or bearish in their trading behaviors (Brown and Cliff 2004). According to Pandey and Sehgal (2019), the existence of sentiments cannot be denied but the matter is about how to identify and measure this factor. Sentiment is recognized with a market-based method using comprehensive indexes or a survey-based method using questionnaires and surveys. It is also identified with a text-based method of interpreting investors' messages on social

media and financial news articles (Shen et al. 2022). In other words, sentiments can be measured using a direct, indirect, or any approach through which sentiment-related data can be collected.

2.1. Sentiment Measures

In a direct method, the sentiments of investors can be collected via surveys taken regularly or through 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 sentiments. 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 Verma 2008). The AAI 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 how they respond to 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 2018).

The indirect approach to measuring investor sentiments which is applied widely is the construction of a sentiment index from several proxies. Baker and Wurgler (2006) created a composite sentiment index from six proxies: the closed-end fund discount, the market turnover, the number of Initial Public Offering (IPO) and returns of the first-day trading of IPO stocks, the new issuance share, and the dividend premium. Baker et al. (2012) removed 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 and Wurgler (2006), and added volatility premium as a new proxy for sentiment index. In the work of other researchers, investor sentiments are also recognized by interpreting some trading activities including margin borrowing, short interests change, and short sales of specialists (Brown and 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 sentiments from several online platforms. The textual analysis approach applies the Natural Language Processing model to texts in different categories depending on research purposes. Petropoulos and Siakoulis (2021) extracted the sentiments from the speeches of the central bank on the economic and financial outlook and relevant policies. Their work shows that the sentiment index from speeches can predict financial market turmoil. Huang et al. (2020) analyzed all public news on "a wide array of major news sources" about a corporation and found a relationship between news and institutional trading. Institutions mainly trade on the news tone immediately after the first news release and move the market returns in the weeks after that. Baker et al. (2019) also analyzed news on policies to track the volatility of the market based on news. Daudert (2021) even gave sentiment scores to the analysis of a company's financial performance.

Sentiments can be discovered by analyzing the emotions behind the investor's comments on social media such as message boards, Facebook, or Twitter. Moreover, sentiments can be taken from stock market news or economic news because news articles can transfer tone and emotions to the reader. Therefore, news articles reflect the expectations of investors on the market in general and for specific stocks according to the information they provide. This measure of sentiments is called media-based investor sentiments (Sun et al. 2016; Nguyen et al. 2015).

Antweiler and Frank (2004) measured sentiments by collecting and interpreting a great number of messages on finance websites such as Yahoo! Finance. Others extracted

sentiments from investor's posts on media including blogs, 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 the 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) utilized news sentiments related to each firm to be a proxy of investor sentiments on the stock of that firm. The "Overall Sentiment Score" ranging from -1 to 1 reflects the level of optimism or pessimism. The higher the score is, the greater the investor's optimism about the stock. Bali et al. (2016) stated that the increases in market volatility relate to unusual news. Unusual news induces investors' disagreement on valuing firms. "Given the high costs of short selling, pessimistic investors sit on the sidelines, while optimistic investors bid up stock prices to reflect their own valuations" (Bali et al. 2016).

2.2. News Sentiments 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.

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 the sentiments from news and the movements of the stock market. Sentiments that can be extracted from general news information (Nguyen et al. 2015; Renault 2017) or news on the COVID-19 pandemic (Costola et al. 2020; Baker et al. 2020) have a significant ability to predict future stock returns. Feng et al. (2022) explored the relationship between news sentiments and stock market volatility in Japan. The research findings show that news sentiments have realizable effects on the volatility of stock returns. Liu et al. (2023) suggest that firm-specific news sentiments 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 news tone on volatility and stock returns in China. Feng et al. (2022) explored the relationship between news sentiments and stock market volatility in Japan. The research findings show that news sentiments have realizable effects on the volatility of stock returns. Liu et al. (2023) suggest that firm-specific news sentiments can boost stock trading activity if they are optimistic. However, news with a pessimistic tone has stronger power in predicting stock returns than news with an optimistic tone. Similarly, Shen et al. (2022) also present the role of news tone on volatility and stock returns in China. Cho et al. (2007) analyzed the market reactions to negative news combined with the-day-of-the-week effect and concluded that a negative return on the previous Friday worsens the return on the next Monday. The underreaction of the market to news can be found on the other days of the week. Cho et al. (2007) also found different levels of reaction from different market indexes depending on the number of stocks that the indexes cover.

3. Methodology

The news sentiments were 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. Similar to the work of [Huang et al. \(2020\)](#), in this study, a large type of news that appeared in mass media is collected. However, [Huang et al. \(2020\)](#) constructed a “news cluster” that combined news on a particular firm.

There were two steps in the NLP model to discover news sentiments using textual analysis. In the first step, news articles from the websites were collected for model training. The news was taken from 3 websites: Cafef.vn, Vneconomy.vn, and Stockbiz.vn accessed on 12 July 2021. The Pandas (Beautifulsoup) library was applied to collate all the articles needed. After collecting and processing, the set of 40,000 articles was separated into a training sample (70%) and a testing sample (30%). According to [Petropoulos and Siakoulis \(2021\)](#), it is necessary to construct dictionaries for sentiments. In this study, the Convolutional Neural Network (CNN) layer, which will be discussed later in this section, was applied, so no sentiment dictionary was required. For model training, the news was labeled manually by the researchers to avoid mistakes. Specifically, a set of 2738 news articles was labeled as one of two groups: positive sentiment and negative sentiment. 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. Words and phrases presenting positive sentiment that could be found in the news were: “a significant increase”, “attractive”, “net buying”, and “price ceiling increase”. Words and phrases belonging to the negative or pessimistic group included “market washout”, “net selling”, “floor price”, “decrease”, and “shrinkage”. An example of news with positive sentiment was one showing the positive expectation of the firm’s earning growth rate in 2021 from Dragon Capital, an Investment Fund in Vietnam. News was labeled as 1 for having a negative sentiment and 2 for having a 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 Intelligence application resembling the human brain for analyzing texts. 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 [Devlin et al. \(2019\)](#). We developed a backbone-based model by starting with the PhoBERT model, a pre-trained language model, and fine-tuned it. To enhance its capabilities, we introduced several advanced techniques. In the PhoBERT base architecture, we integrated Convolutional Neural Network (CNN) layers with varying kernel sizes. These CNN layers were used to extract context representation vectors, thereby improving the model’s understanding of contextual information. This technique was proven to be very effective by [Pham et al. \(2021\)](#). However, because of the limitation of resources, we could not capture entire news articles in the training processing. The Spacy Library was implemented to synthesize the opinion of each paragraph in a news article and give a unified point of view of the whole piece of news. The Spacy Library separated each sentence and combined them into a paragraph that met the standard of no more than 200 words. However, sometimes each paragraph did not represent the main content of the entire news. Meanwhile, the news headline was an important piece of information that summarized 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 final sentiment of the news articles was summed up by all the points of view in each paragraph through majority voting.

For the training model, we employed the Adam Optimizer with a learning rate schedule that included linear warmup and linear decay. The peak learning rate was set to 1×10^{-5} . Before the classifier layer, a dropout with a rate of 0.1 was applied. During the training process, we used a batch size of 16 and conducted training for 20 epochs.

The model’s performance was assessed through its ability to identify the sentiment of news in the testing data and classify news according to its sentiment. The classification

accuracies were measured using the estimations of Accuracy, Precision, Recall or Sensitivity, Specificity, and F1-score.

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

TP: True positive, the number of news articles containing positive sentiments is classified as Positive.

TN: True negative, the number of news articles containing negative sentiments is classified as Negative.

FP: False positive, the number of news articles containing negative sentiments is classified as Positive.

FN: False negative, the number of news articles containing positive sentiments is classified as Negative.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

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

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

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

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 was at 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 classes. Similarly, the Specificity measures the relationship between True Negative and the whole actual Negative class. The F1-score estimate 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 attempted to reveal the changes in the stock market as a response to positive and negative news. To gain better knowledge of the market reactions to the news, the return movements of more than one index should be analyzed (Cho et al. 2007; Chow et al. 2016). We collected returns calculated from the 2 main indices including VN30-Index and HNX30-Index. VN30-Index and HNX30-Index presented the comprehensive price index of the top 30 market capitalization stocks in the Hochiminh stock exchange and the Hanoi stock exchange, respectively.

The date that the news was released was the event date. Returns of the VN30-Index and HNX30-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. To explore the effects of news on market returns, we performed several tests to compare the return variances and return means measured before and after the event dates. The variance ratio test was performed to assess the null hypotheses:

H₀₁. *There is no significant difference in the ratio of market return variances before and after positive news releases.*

H₀₂. *There is no significant difference in the ratio of market return variances before and after negative news releases.*

To highlight the importance of risks in investment analysis, [Cho et al. \(2007\)](#), [Chow et al. \(2016\)](#), and [Fang and Post \(2022\)](#) ordered the risks of returns collected in different periods using the stochastic dominance method. Therefore, in this work, the Cumulative Distribution Function (DCF) was estimated together with variance tests to rank the risks of each index. For mean comparison, the *t*-test was employed to explore the significant return changes from the effects of news sentiments. There were two hypotheses used for *t*-tests.

H₀₃. *There is no significant difference in the market returns before and after positive news releases.*

H₀₄. *There is no significant difference in the market returns before and after the negative news releases.*

All the above statistical tests were applied to 100 separate event dates. It includes 50 dates posting positive news and 50 dates announcing negative news.

4. Analysis Results and Discussion

4.1. NLP Results

The Vietnamese stock market started its first operation in 2000. Despite being in the early stages of its 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 Vietnamese 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. The daily news is collected from 3 popular websites in the stock market and Vietnamese 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	2837	01/09/2021–28/06/2022
Stockbiz	9374	12/07/2021–27/07//2022

Most of the articles were 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 updates on domestic and international economic news. Tables 2–4 provide the statistical descriptions of the number of news articles collected in one day from these web platforms.

Table 2. Descriptions of daily news numbers by category on the Cafef.vn website.

Name	Mean	Min	Max	Standard Deviation	Standard Error	Skewness
International finance	15.13	4	42	5.55	0.35	0.50
News	15.24	0	41	4.75	0.30	0.21
Business	15.06	0	40	8.24	0.52	−0.03
Life	15.19	0	46	7.39	0.47	−0.25
Macroeconomics	15.20	0	37	8.20	0.52	−0.67
Stocks	15.23	0	63	14.88	0.94	0.86
Properties	15.14	0	87	17.45	1.10	0.74
Market	15.20	0	89	20.50	1.29	0.84
All	121.40	5	280	59.48	3.75	0.30

Table 3. Descriptions of daily news numbers by category on the Stockbiz.vn website.

Name	Mean	Min	Max	Standard Deviation	Standard Error	Skewness
Real_estate	1.21	0	8	1.12	0.03	1.73
Finance	1.56	0	12	2.02	0.06	1.32
Economy	1.56	0	25	3.27	0.09	2.90
Market	1.55	0	16	2.88	0.08	1.96
World	1.56	0	18	3.38	0.10	2.28
All	7.43	1	60	10.02	0.28	2.16

Table 4. Descriptions of daily news numbers by category on the Vneconomy.vn website.

Name	Mean	Min	Max	Standard Deviation	Standard Error	Skewness
Finance	1.07	0	3	0.72	0.04	0.22
Investments	0.72	0	3	0.62	0.04	0.37
Highlights	1.36	0	4	0.85	0.05	0.49
World economy	0.79	0	3	0.68	0.04	0.42
Market	0.74	0	2	0.60	0.03	0.18
Society	0.99	0	2	0.66	0.04	0.01
Society	1.14	0	3	0.82	0.05	0.25
Corporate finance	1.00	0	3	0.78	0.05	0.46
Stock market	1.65	0	5	1.29	0.07	0.14
All	9.46	1	21	4.22	0.24	−0.04

Articles are selected in different categories such as general news, stocks, properties, business, international finance, macroeconomics, life, and market (Cafef.vn). The number of articles in each category varies according to the relevance of the information to the research purpose. 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 in order to have a better understanding of market reactions to new information. 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” is not included in the study because it is related to a specific stock. However, the news “HPG hit the ceiling with record liquidity, VN-Index broke through nearly 20 points, surpassing the 1500 points” is labeled positive because it presents an optimistic view of the whole market. For training the NLP language model, the news is labeled according to the sentiments that the messages provide. Table 5 illustrates the results of the labeling process by presenting examples of news that are labeled for model training.

The performance of the NLP model is assessed via the power to classify news into two groups according to the emotion extracted from the content of the news. As shown in Table 6, there is not much accuracy difference in classifying any piece of news into one of the two groups: the negative group and the positive group. All estimates (Precision, Recall, and F1-score) described in Table 6 are greater than 81%, which demonstrates the significant power of the model to extract the sentiment from an article.

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 1, the AUC is nearly 90%, showing that the model obtains a substantial level of accuracy in extracting the sentiment from any article.

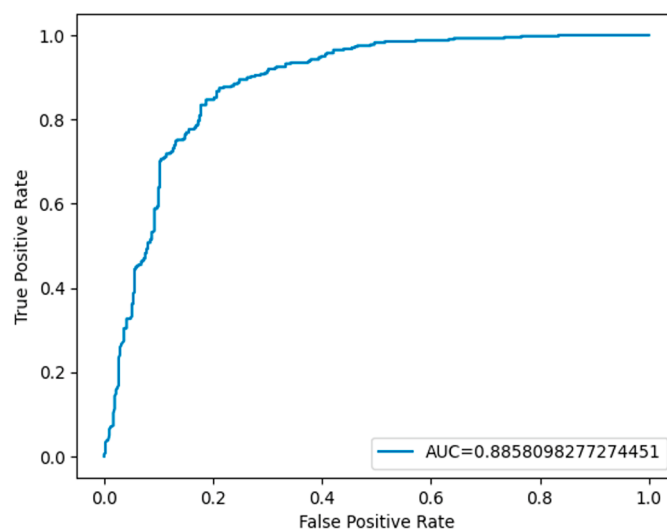
Table 5. Examples of news labeling.

	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 1500 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 1270 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.

Table 6. Summary of the model’s performance.

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

**Figure 1.** ROC curve.

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 highlight how the NLP model outperforms other models in its ability to extract sentiments from new information. For example, using BERT, the NLP models built by [Renault \(2017\)](#) can obtain an accuracy level from 75.24% to 90.75%.

4.2. Analysis Results on Market Reactions to News Sentiments

The return’s variance of stock is an important factor for decision making because variance estimates the risk belonging to a stock. Tables 7 and 8 present the variance ratio

test results to portray the changes in the variance of the market indices as reactions to news sentiments.

Table 7. Variance test results—positive news sentiments.

Group		Mean	Std. Err.		Std. Dev.
0: 30 days after event date 1: 30 days before event date					
VN30-Index					
	0	−0.00054	0.00041		0.01610
	1	−0.00026	0.00040		0.01574
Combined		−0.00041	0.00029		0.01592
H0: ratio = 1				f = 1.0464	2 × Pr(F > f) = 0.3719
HNX30-Index					
	0	−0.00143	0.00048		0.01919
	1	−0.00101	0.00048		0.01882
Combined		−0.00122	0.00029		0.01592
H0: ratio = 1				f = 1.0398	2 × Pr(F > f) = 0.4420
0: 5 days after event date 1: 5 days before event date					
VN30-Index					
	0	−0.00147	0.00096		0.01684
	1	−0.00195	0.00114		0.01813
Combined		−0.00122	0.00034		0.01901
H0: ratio = 1				f = 0.8911	2 × Pr(F < f) = 0.2178
HNX30-Index					
	0	−0.00213	0.00115		0.02006
	1	−0.00331	0.00136		0.02168
Combined		−0.00267	0.00088		0.02080
H0: ratio = 1				f = 0.8564	2 × Pr(F < f) = 0.1952

Table 8. Variance test results—negative news sentiments.

Group		Mean	Std. Err.		Std. Dev.
0: 30 days after event date 1: 30 days before event date					
VN30-Index					
	0	−0.00051	0.00039		0.01515
	1	−0.00082	0.00043		0.00043
Combined		−0.00066	0.00029		0.01586
H0: ratio = 1				f = 0.8346	2 × Pr(F < f) = 0.0005
HNX30-Index					
	0	0.00012	0.00046		0.01774
	1	−0.00095	0.00050		0.01918
Combined		−0.00040	0.00029		0.01586
H0: ratio = 1				f = 0.8558	2 × Pr(F < f) = 0.0027
0: 5 days after event date 1: 5 days before event date					
VN30-Index					
	0	−0.00070	0.00102		0.01743
	1	0.00001	0.00105		0.01640
Combined		−0.00040	0.00034		0.01846
H0: ratio = 1				f = 1.1306	2 × Pr(F > f) = 0.3198
HNX30-Index					
	0	−0.00028	0.00117		0.02005
	1	0.00025	0.00025		0.01971
Combined		−0.00040	0.00034		0.01846
H0: ratio = 1				f = 1.0349	2 × Pr(F > f) = 0.7826

The variance test results shown in Tables 7 and 8 disclose that the null hypothesis— H_{01} —cannot be rejected for both market indices before and after the release of positive news in media. However, there are significant differences in the variance of both the VN30-Index as well as the HNX-Index calculated 30 days before and 30 days after the release of negative news. Applying the same variance ratio test, [Chow et al. \(2016\)](#) also found a significant difference in variance as a response to the event of market liberalization in Latin American countries. Performing the Cumulative Distribution Function (CDF) mentioned by [Cho et al. \(2007\)](#), [Chow et al. \(2016\)](#), and [Fang and Post \(2022\)](#) for risk order, it can be seen that the two CDFs coincide when the news release is positive (Figures 2–4) except for CDFs built for negative sentiment (Figure 5).

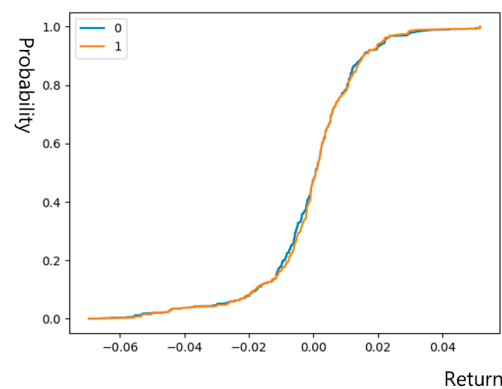


Figure 2. CDFs for VN30-Index (positive news sentiments).

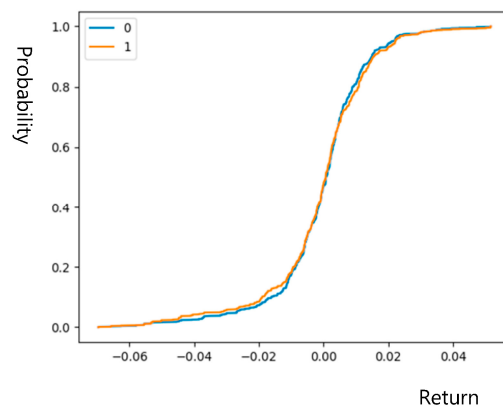


Figure 3. CDFs for VN30-Index (negative news sentiments).

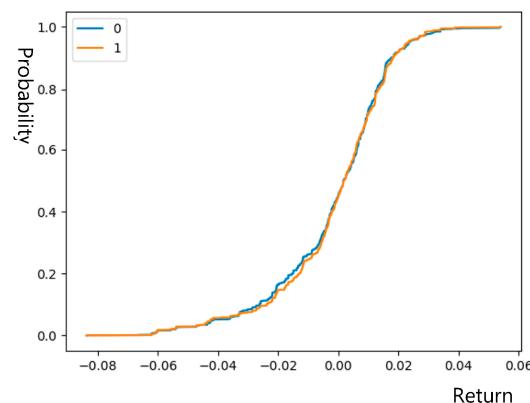


Figure 4. CDF for HNX30-Index (positive news sentiments).

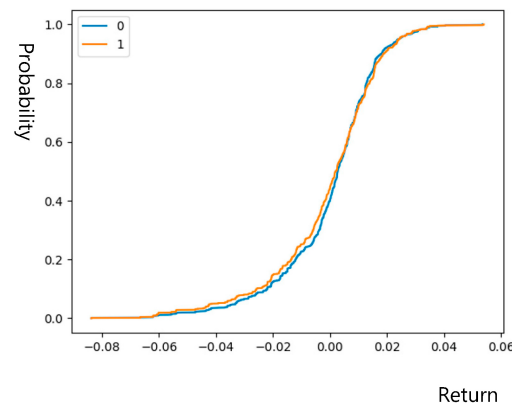


Figure 5. CDF for HNX30-Index (negative news sentiments).

The *t*-test, a statistical technique, is applied to examine the effect of the news on the movements of the market. According to the variance test ratio results in the previous section, equal variance *t*-tests are run for two return groups having equal variance while unequal variance *t*-tests are performed for two return groups having unequal variance. For examining the instant market reactions, the pre-five-day period is compared with the post-five-day period while the pre-thirty-day period is compared with the post-thirty-day period for a longer investigation period. Tables 9 and 10 describe the two-sample *t*-test results for positive news sentiments and negative news sentiments, respectively. Each table also describes the mean and standard deviation of index returns calculated in different periods. 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, but the difference is not statistically significant.

Table 9. Two-sample *t*-test results—positive news sentiments.

Group	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0: 30 days after the event date 1: 30 days before the event date					
VN30-Index—Equal variance test					
0	0.0001801	0.0006097	0.017062	−0.0010169	0.001377
1	−0.0008694	0.0004649	0.0157799	−0.0017816	0.0000428
Difference	0.0010495	0.0007555		−0.0004321	0.0025311
H0: difference = 0			t = 1.3892	Pr(T > t) = 0.1649	
HNX30-Index—Equal variance test					
0	0.0001183	0.0004552	0.0177408	−0.0007745	0.0010112
1	−0.0009462	0.0005019	0.0191769	−0.0019307	0.0000383
Difference		0.0010645	0.0006776	−0.000264	0.0023931
			t = 1.5711	Pr(T > t) = 0.1163	
0: 5 days after the event date 1: 5 days before the event date					
VN30-Index—Equal variance test					
0	−0.0014716	0.0009625	0.0168371	−0.0033656	0.0004224
1	−0.0019548	0.0011351	0.0181257	−0.0041902	0.0002805
	−0.0016913	0.0007355	0.0174205	−0.0031359	−0.0002466
Difference	0.0004832	0.0014783		−0.0024205	0.0033869
H0: diff = 0			t = 0.3269	Pr(T > t) = 0.7439	
HNX30-Index—Equal variance test					
0	−0.002131	0.0011467	0.0200593	−0.0043874	0.0001255
1	−0.0033084	0.0013574	0.0216758	−0.0059816	−0.0006352
Difference	0.0011774	0.0017645		−0.0022883	0.0046432
H0: diff = 0			t = 0.6673	Pr(T > t) = 0.5049	

Table 10. Two-sample *t*-test results—negative news sentiments.

Group	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0: 30 days after the event date; 1: 30 days before the event date					
VN30-Index—Unequal variance test					
0	−0.0005062	0.0003886	0.015145	−0.0012684	0.000256
1	−0.0008217	0.0004339	0.0165781	−0.0016728	0.0000294
Difference	0003155	0.0005824		−0.0008265	0.0014576
H0: difference = 0			t = 0.5417	Pr(T > t) = 0.5881	
HNX30-Index—Unequal variance test					
0	0.0001183	0.0004552	0.0177408	−0.0007745	0.0010112
1	−0.0009462	0.0005019	0.0191769	−0.0019307	0.0000383
Difference	0.0010645	0.0006776		−0.000264	0.0023931
Ha: diff = 0			t = 1.5711	Pr(T > t) = 0.1163	
0: 5 days after the event date 1: 5 days before the event date					
VN30-Index—Equal variance test					
0	−0.0007025	0.0010168	0.017434	−0.0027036	0.0012986
1	0.0000139	0.0010475	0.016396	−0.0020494	0.0020772
Difference	−0.0007164	0.001468		−0.0036002	0.0021673
H0: diff = 0			t = −0.4880	Pr(T > t) = 0.6257	
HNX30-Index—Equal variance test					
0	−0.0002755	0.0011695	0.0200522	−0.0025771	0.0020262
1	0.0002452	0.0012593	0.019711	−0.0022353	0.0027257
Difference	−0.0005207	0.0017213		−0.0039019	0.0028605
H0: diff = 0			t = −0.3025	Pr(T > t) = 0.7624	

If the significant change in the market return's means is considered a signal of market reactions to news sentiments, the analysis results of this research show that news sentiments have no significant effects on the stock market in Vietnam. This finding is not aligned with the conclusions of related studies. For example, [Petropoulos and Siakoulis \(2021\)](#) claim that the sentiment of corporate news can forecast market turbulence. [Huang et al. \(2020\)](#) and [Li et al. \(2020\)](#) also confirm the role of news sentiments in stock return prediction. The disparity between the findings of this research and other works is down to two main factors. The first factor is the doubt about the leakage of both positive and negative news before it is public in the media so that the market may react before the official news release. Therefore, the choice of the time frame before and after the dates of news releases is crucial when it comes to interpretation and analysis. The second factor could be the limitation of the sample collected. If broader indices are investigated, the effects of news sentiments on stock returns would be more significant. Therefore, more event dates and stock indexes should be collected for future research to obtain an adequate estimate of market response to positive and negative news articles.

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 various moods of the news posted daily on three websites concerning finance and the economy in Vietnam. After being trained, the model 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. As one of the pioneers in building a textual analysis model in the Vietnamese language, 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 wider market or economic condition that can be favorable or unfavorable to stock investment. Therefore, investors can obtain an understanding of the current market without extensive reading. Researchers who are interested in textual

analysis should find the results of this study noteworthy because it provides a well-trained model in the Vietnamese language.

Regarding the research objective of examining the effects of news sentiments on the stock market, this study concludes that there is no significant change in the return means of the index before and after the news release. However, employing the variance ratio test and CDF presentation, this study reveals that there is a change in the index risk when a piece of negative news is available.

For future research, there should be a broader index that reflects the prices of a large number of stocks used to explore the reactions of the whole market to the news. Examples of current stock indexes in Vietnam that can be selected are Vn-Index, measuring the price changes of all stocks traded on the Hochiminh Stock Exchange, and HNX-Index, accounting for the prices of all stocks on the Hanoi Stock Exchange. To discover any news leakage, future research is intended to rearrange the time periods to detect any potential effects of news on the market before public announcements.

Author Contributions: Conceptualization, L.T.V., H.T.K.; methodology, L.T.V., D.N.P.; software, D.N.P.; validation, L.T.V., D.N.P. and T.T.T.P.; formal analysis, L.T.V.; investigation, D.N.P.; resources, T.T.T.P.; data curation, L.T.V., H.T.K.; writing—original draft preparation, L.T.V. and H.T.K.; writing—review and editing, L.T.V. and H.T.K.; visualization, T.T.T.P.; supervision, L.T.V.; project administration, L.T.V.; funding acquisition, L.T.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the VNU University of Economics and Business, Hanoi, grant number KT.21.05.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Acknowledgments: Many thanks to the VNU University of Economics and Business, Hanoi, for financing the Research Project number KT.21.05. This paper has been extracted from this research.

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

References

- Aggarwal, Charu C. 2018. Opinion Mining and Sentiment Analysis. In *Machine Learning for Text*. Cham: Springer, pp. 413–34. [\[CrossRef\]](#)
- Antweiler, Werner, and Murray Z. Frank. 2004. Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance* 59: 1259–94. [\[CrossRef\]](#)
- Baker, Malcolm, and Jeffrey Wurgler. 2006. Investor Sentiment and Cross-Section of Stock Returns. *Journal of Finance* 61: 1645–80. [\[CrossRef\]](#)
- Baker, Malcolm, Jeffrey Wurgler, and Yu Yuan. 2012. Global, Local, and Contagious Investor Sentiment. *Journal of Financial Economics* 104: 272–87. [\[CrossRef\]](#)
- Baker, Scott R., Nicholas Bloom, Steven J. Davis, and Kyle J. Kost. 2019. *Policy News and Stock Market Volatility*. NBER Working Papers 25720. Cambridge: National Bureau of Economic Research, Inc.
- Baker, Scott R., Nicholas Bloom, Steven J. Davis, Kyle Kost, Marco Sammon, and Tasaneeya Viratyosin. 2020. The unprecedented stock market reaction to COVID-19. *Review of Asset Pricing Studies* 10: 742–58. [\[CrossRef\]](#)
- Bali, Turan G., Andriy Bodnaruk, Anna Scherbina, and Yi Tang. 2016. Unusual News Flow and the Cross-Section of Stock Returns. *SSRN Electronic Journal* 64: 4137–55. [\[CrossRef\]](#)
- Barber, Brad M., and Terrance Odean. 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies* 21: 785–818. [\[CrossRef\]](#)
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. A model of investor sentiment. *Journal of Financial Economics* 49: 307–43. [\[CrossRef\]](#)
- Bar-Haim, Roy, Elad Dinur, Ronen Feldman, Moshe Fresko, and Guy Goldstein. 2011. Identifying and following expert investors in stock microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh: Association for Computational Linguistics, pp. 1310–19.
- Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2: 1–8. [\[CrossRef\]](#)

- Brown, Gregory W., and Michael T. Cliff. 2004. Investor sentiment and the near-term stock market. *Journal of Empirical Finance* 11: 1–27. [CrossRef]
- Brown, Gregory W., and Michael T. Cliff. 2005. Investor sentiment and asset valuation. *Journal of Business* 78: 405–40. [CrossRef]
- Cho, Young-Hyun, Oliver Linton, and Yoon-Jae Whang. 2007. Are there Monday effects in stock returns: A stochastic dominance approach. *Journal of Empirical Finance* 14: 736–55. [CrossRef]
- Chow, Sheung Chi, Yongchang Hui, João Paulo Vieito, and Zhenzhen Zhu. 2016. Market liberalizations and efficiency in Latin America. *Studies in Economics and Finance* 33: 553–75. [CrossRef]
- Chui, David, Wui Wing Cheng, Sheung Chi Chow, and Li Ya. 2020. Eastern Halloween effect: A stochastic dominance approach. *Journal of International Financial Markets, Institutions and Money* 68: 101241. [CrossRef]
- Costola, Michele, Oliver Hinz, Michael Nofer, and Loriana Pelizzon. 2020. Machine Learning Sentiment Analysis, COVID-19 News and Stock Market Reactions. *SSRN Electronic Journal* 64: 101881. [CrossRef]
- Daudert, Tobias. 2021. Exploiting textual and relationship information for fine-grained financial sentiment analysis. *Knowledge-Based Systems* 230: 107389. [CrossRef]
- De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. Noise Trader Risk in Financial Markets. *The Journal of Political Economy* 98: 703–38. [CrossRef]
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Edinburgh: Association for Computational Linguistics, pp. 4171–86.
- Dougal, Casey, Joseph Engelberg, Diego Garcia, and Christopher A. Parsons. 2012. Journalists and the Stock Market. *Review of Financial Studies* 25: 639–79. [CrossRef]
- Fama, Eugene F. 1970. Efficient Capital Market: A Review of Theory and Empirical Work. *Journal of Finance* 25: 382–417. [CrossRef]
- Fang, Yi, and Thierry Post. 2022. Optimal portfolio choice for higher-order risk averters. *Journal of Banking & Finance* 137: 106429.
- Feng, Lingbing, Tong Fu, and Yanlin Shi. 2022. How does news sentiment affect the states of Japanese stock return volatility? In *International Review of Financial Analysis*. Amsterdam: Elsevier, vol. 84.
- Fisher, Kenneth L., and Meir Statman. 2000. Investor Sentiment and Stock Returns. *Financial Analysts Journal* 56: 16–23. [CrossRef]
- Huang, Alan Guoming, Hongping Tan, and Russ Wermers. 2020. Institutional Trading around Corporate News: Evidence from Textual Analysis. *The Review of Financial Studies* 33: 4627–75. [CrossRef]
- Li, Xiaodong, Pangjing Wu, and Wenpeng Wang. 2020. Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management* 57: 102212. [CrossRef]
- Liu, Jun, Kai Wu, and Ming Zhou. 2023. News tone, investor sentiment, and liquidity premium. In *International Review of Economics & Finance*. Amsterdam: Elsevier, vol. 84, pp. 167–81.
- Montier, James. 2002. *Behavioural Finance: Insights into Irrational Minds and Markets*. Chichester: John Wiley and Sons Ltd.
- Nguyen, Thien Hai, Kiyooki Shirai, and Julien Velcin. 2015. Sentiment Analysis on Social Media for Stock Movement Prediction. *Expert Systems with Applications* 42: 9603–11. [CrossRef]
- Pandey, Piyush, and Sanjay Sehgal. 2019. Investor Sentiment and its Role in Asset Pricing: An Empirical Study for India. *IIMB Management Review* 31: 127–44. [CrossRef]
- Petropoulos, Anastasios, and Vasilis Siakoulis. 2021. Can Central Bank Speeches Predict Financial Market Turbulence? Evidence from an Adaptive NLP Sentiment Index Analysis Using XGBoost Machine Learning Technique. *Central Bank Review* 21: 141–53. [CrossRef]
- Pham, Ngoc Dong, Thi Hanh Le, Thanh Dat Do, Thanh Toan Vuong, Thi Hong Vuong, and Quang Thuy Ha. 2021. Vietnamese Fake News Detection Based on Hybrid Transfer Learning Model and TF-IDF. Paper presented at the 13th International Conference on Knowledge and Systems Engineering (KSE), Bangkok, Thailand, November 10–12; pp. 1–6. [CrossRef]
- Phan, Truc, Philippe Bertrand, Hong Hai Phan, and Xuan Vinh Vo. 2021. Investor sentiment and stock return: Evidence from Vietnam stock market. *The Quarterly Review of Economics and Finance* 87: 141–53. [CrossRef]
- Qiu, Lily, and Ivo Welch. 2004. *Investor Sentiment Measures*. NBER Working Paper. p. 10794. Available online: <https://ssrn.com/abstract=595193> (accessed on 12 July 2022).
- Renault, Thomas. 2017. Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance* 84: 25–40.
- Schumaker, Robert P., and Hsinchun Chen. 2009. Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems* 27: 1–19. [CrossRef]
- Schwert, G. William. 2003. Anomalies and market efficiency. *Handbook of the Economics of Finance* 1: 939–74.
- Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson. 2022. Measuring news sentiment. *Journal of Econometrics* 228: 221–43. [CrossRef]
- Shen, Shulin, Le Xia, Yulin Shuai, and Da Gao. 2022. Measuring news media sentiment using big data for Chinese stock markets. *Pacific-Basin Finance Journal* 74: 101810. [CrossRef]
- Sun, Licheng, Mohammad Najand, and Jiancheng Shen. 2016. Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance* 73: 147–64.

- Verma, Rahul, and Priti Verma. 2008. Are survey forecasts of individual and institutional investor sentiments rational? *International Review of Financial Analysis* 17: 1139–55. [[CrossRef](#)]
- Vu, Tien Thanh, Shu Chang, Quang Thuy Ha, and Nigel Collier. 2012. An Experiment in Integrating Sentiment Features for Tech Stock Prediction in Twitter. Paper presented at Workshop on Information Extraction and Entity Analytics on Social Media Data, Mumbai, India, December 9; pp. 23–38.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.