THAMMASAT REVIEW
ONLINE ISSN: 2630-0303

Volume 28 No. 2 (July-December) 2025 [Page 1-16]

The Influence of Online News on Thai Investors' Stock Investment Decisions

Phum Tanggasemsun [a] and Pariyada Sukcharoensin [b, *]

[a] *Beryl 8 Plus Company Limited*
[b] *School of Development Economics,*
National Institute of Development Administration, Thailand

Received 13 July 2024; Received in revised form 11 May 2025
Accepted 17 Jun 2025; Available online 29 December 2025

Abstract

In this study, we explore the sentiment of publicly available financial news from online sources related to stocks in the Thai stock exchange markets, along with other relevant factors. The multiple datasets are gathered from economic and financial news across the internet during the easing period of COVID-19 restrictions. Sentiment is defined using Natural Language Processing (NLP) techniques, which analyze headlines with wordlists. We implement data mining and data analysis techniques in the financial field to enhance data gathering and processing. The advantages of this technological approach include repeatability, reliability over time, and the ability to handle large datasets. Conventional online behavior, such as trends, is included as a complementary variable alongside sentiment. Stock characteristics, including index, industrial category, and market, are included as sub-independent variables. The empirical results indicate that sentiment, trends, and other factors are related to stock movement, with magnitudes of each variable varying according to time differences.

Keywords

Conventional online behavior, Natural Language Processing, Sentiment, Time lag, Trend

* **Corresponding author:** spariyada@yahoo.com **DOI:** 10.14456/tureview.2025.17

Introduction

Nowadays, the internet influences our lifestyles in different ways. Since the internet globalizes the entire world, information is shared through online platforms such as websites and social media in various forms, including text, images, and videos. Since people can reach the same online information, online sources can also influence stock investment decisions of investors through rumors, opinions, facts, or analyzed data published on the internet.

The previous models explaining stock movement have been based on historical data. Some studies have used technology in forecasting stock movement. So, we realize the potential usage of technology in financial forecasting. According to the advancement of technology, it is possible to extract large amounts of information from online platforms in a short period of time. The idea of improving robustness by including textual data in forecasting stocks in the market has become popular as a multifactor model. Machine learning is also a popular research tool in predicting stock movement. The main advantage of machine learning is that it can process large amounts of data in less time compared with traditional methods handling the same amount of data.

However, feelings toward stocks, defined as sentiment, have also been considered in previous studies. These studies analyzed headlines or entire articles on multiple platforms, for example, Twitter, Reddit, Yahoo Finance, and Financial Times. Despite this, a gap remains between research and practice. Our goal is to simplify the implemented model while producing consistent and practical results. This study examines the behavior of Thai investors in making stock investment decisions using equivalent data from available online sources, along with other factors such as trending indicators and stock characteristics.

Literature Review

Market movement and prediction

Studies have shown that there are several options for forecasting with textual data, including focusing on headlines, whole articles, or both. Many agree to use headlines as a representative of sentiment instead of querying entire documents for analysis or opinion mining (Chiewhawan & Vateekul, 2020; Nguyen et al., 2015; Si et al., 2013; Tu et al., 2016). Further, Wansri et al. (2023) predicted stock trends in the tourism sector during the COVID-19 pandemic using sentiment analysis of news headlines combined with data mining techniques.

There are several concerns when using textual information, such as algorithms, window size, frequency, time lag, and wordlist selection. First, the algorithms most commonly used in prior research are deep learning and data analysis methods for sentiment. However, each methodology has its own strengths and weaknesses. Window size and frequency reflect differences in datasets retrieved from online platforms. Time lag has also been shown to significantly influence investor's decision-making (Luo et al., 2013). Wordlist selection can be classified into three categories: positive, neutral, and negative (Chiewhawan & Vateekul, 2020; Luo et al., 2013; Si et al., 2013; Sul et al., 2016; Vu et al., 2012). In addition, some studies are more specific in their data selection, such as analyzing messages containing hashtags, discussion threads, or posts categorized by specific keywords from authors' messages (Liu et al., 2015; Si et al., 2013; Sul et al., 2016; Tu et al., 2016). To measure forecasting performance, the implemented models are compared against historical stock data to determine accuracy rates.

Sentiment analytical model

Natural Language Processing (NLP) is introduced to close the gap between human and computer language. Sentiment analysis is based on NLP, which involves word classification to convert raw data into more reliable sentiment (Chiewhawan & Vateekul, 2020; Vu et al., 2012). For sentiment analysis, NLP is important for identifying sentiment words within an entire sentence. Good NLP implementation will provide a seamless process and produce accurate results. However, the Thai language is relatively hard to process for several reasons. Thai has unsegmented words, slang, and no officially published dictionary for NLP. There are several ways to deal with this issue, such as machine learning and the Naïve Bayes algorithm. Nonetheless, sentiment analysis can vary depending on models, inputs, user justification, and data sets. Thus, word selection for classifying sentiment categories related to market movements needs to be carefully made for the best results.

According to previous research, sentiment has been classified into three categories: positive sentiment, which is likely to cause an upward trend in a stock; neutral sentiment, which does not affect stock movement; and negative sentiment, which is likely to cause a stock to decline (Chiewhawan & Vateekul, 2020; Luo et al., 2013; Si et al., 2013; Sul et al., 2016; Vu et al., 2012). Chatchawan (2021, 2022) found that both overall and positive sentiment significantly explain equity returns on the Stock Exchange of Thailand, while negative sentiment does not have significant explanatory power. Further, positive news sentiment significantly reduces volatility in the SET50 and SET100 indices. Other studies have classified sentiment into four types: strong buy, strong sell, buy, and sell (Nguyen et al., 2015).

Conventional online behavior and time lagging

Previous studies have included conventional online behavior along with sentiment scores in stock movement models. Conventional online behavior includes specific factors such as search intensity and web traffic related to the subject. High search intensity implies that stocks are being looked up as news is published. By employing this method, stock selection becomes less biased.

Based on the assumption that investors may not react right after news is published but instead take some time to respond, Vector Autoregression was implemented to explain the relationship between time and stock movement. Previous empirical results show that the intensity value within six days after news is published online is significant (Luo et al., 2013).

Methodology

The main steps of this research consist of data mining, data processing, and regression analysis, which will be discussed later in this section. Most of the processes are implemented using Python programming. Related methods and ideas from previous research were considered in developing and improving our data gathering process. Our research framework is shown in Figure 1.

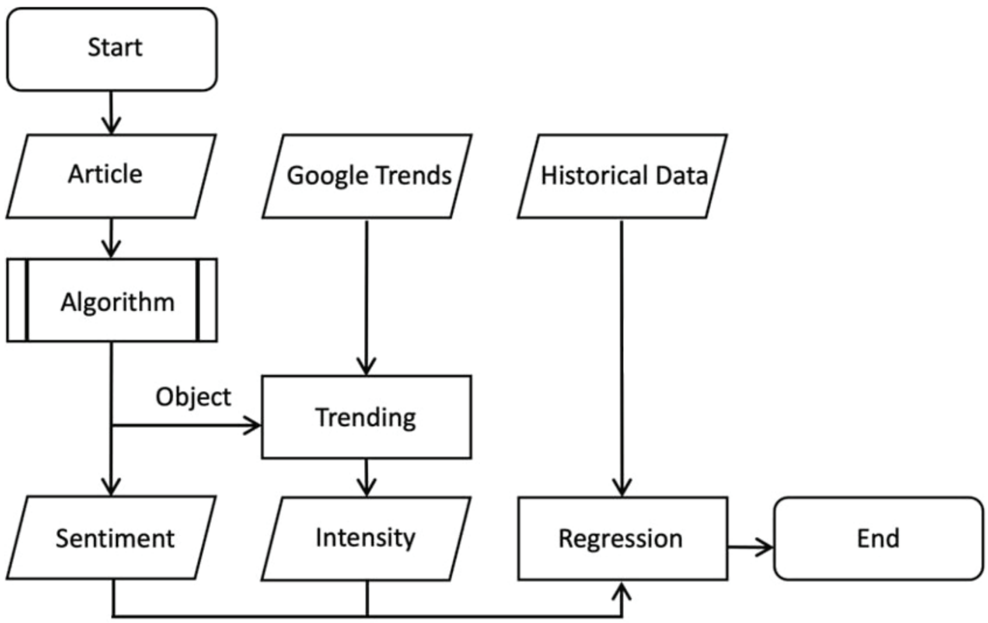

Figure 1 Research framework

Data Mining

Our research consists of four main data components: Thai financial and economic news, Google Trends, historical data, and stock characteristics. We focus on the impact of online news after the most challenging period of the COVID-19 pandemic, when restrictions began to ease. The first main dataset comprised news articles from online sources between August 2021 and December 2021, totaling approximately 6,549 articles. From available raw data, we split the dataset into training and testing sets for sentiment classification programs. After carefully selecting news related to stocks listed on the Stock Exchange of Thailand (SET) and the Market for Alternative Investment (MAI) and finalizing the data with our model algorithm, we obtained a processed dataset of 4,936 observations.

Data Processing

The first step involves taking news articles from selected online sources. These sources are selected from public, well-known websites on the internet based on web traffic. High traffic density indicates the popularity of the sites as news references. According to related studies, we consider the headline, rather than the full article, as raw data, since the headline can reflect the sentiment of the news. Moreover, Alexa web traffic has shown that users spend little time on news websites. Thus, it is valid to consider only the headline of each article. Next, the article, as textual data, is processed using an algorithm, which will be discussed later in the algorithm subsection. After processing, we obtain outputs such as date, sentiment, and the object (stock name). The object and date serve as references in data mining and for defining dependent and independent variables. The trending process uses the Google Trend API to mine data, which provides a trending value ranging from 0 to 100. We also gather more general information about each stock, such as industrial type, index type, and market type. After obtaining the required data, we implement an econometric model, specifically multiple regression. Previous research has shown that news and trending information affect investors' decisions, often with a time lag. Thus, we collect multiple data points within a reasonable period from the news published date.

The algorithm primarily focuses on raw data collection, which is then normalized into sentiment using a classification system. This system consists of the following subparts: web scraping, word normalization and classification, and training. Raw data is collected from qualified online news sites, including articles and published data related to stocks in the Stock Exchange of Thailand (SET) and the Market for Alternative Investment (MAI). Although Facebook and Twitter are widely used in Thailand, Thai stocks and financial topics are not popular among Twitter users and are discussed mainly within private Facebook groups. Due to data limitations and privacy concerns, financial news, a publicly available data source, is used as the raw data for this research, along with Google search trends, which have been shown to be significant variables in previous studies. Sentiment is defined by applying Natural Language Processing (NLP) to analyze news headlines using wordlists. By normalizing articles into sentiment categories via NLP, we were able to classify and assign words into groups reflecting potential market directions and investors' feelings toward the articles. These groups are categorized as positive, neutral, or negative. However, the program requires manual training sessions before it can operate automatically, as the process is still under development. Due to processor limitations, only a four-month range is analyzed. From the collected news samples, we manually classified 350 articles into each category to train our NLP model. A train/test split was applied, by splitting 80 percent of samples into training set and 20 percent of samples into testing set. The model achieved an accuracy rate of 85 percent, which is considered satisfactory for a Thai language dataset with three sentiment categories. This trained model was then used for data normalization and categorization throughout the research. The processed data produced by the algorithm includes the stock name and corresponding textual data, which are then retrieved and processed for further analysis. All algorithmic processes are summarized in Figure 2.

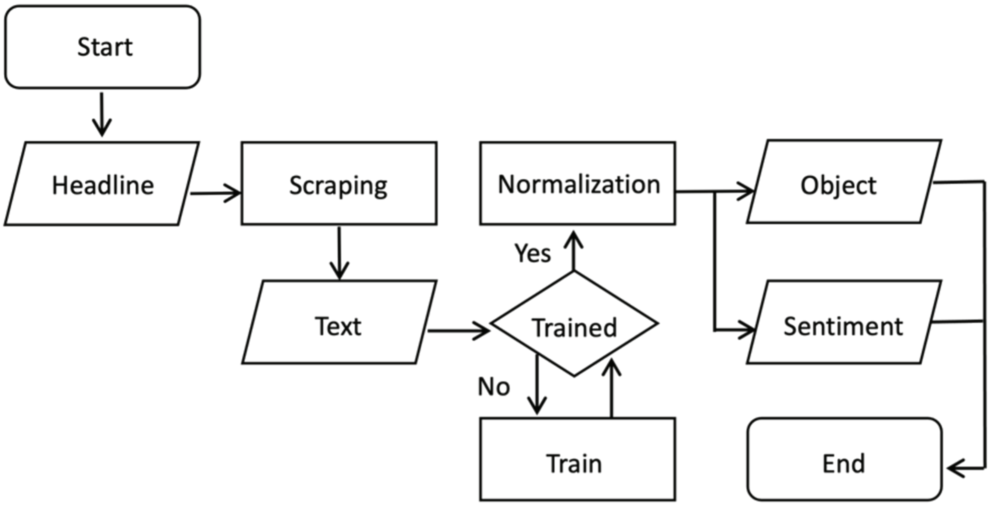

Figure 2 Algorithm

Regression Analysis

Since news variables are exogenous, multiple regression is employed in this research to analyze the impact of online news events on stock returns. We also consider the time lag in stock movements after the news is published. Thus, data at different time intervals are required as the dependent variable. Historical numerical data are used as the dependent variable, represented by the change in the adjusted closing price relative to the previous closing price. The time lag is measured for up to five days after the news is published. In the multiple regression model, we use sentiment, trend, and stock characteristics as independent variables. .

$$P_T = a + b_1Sen + b_2Avg + b_3Index + b_4Cat + b_5Market + e \quad (1)$$

Given that

P = Percentage difference between adjusted close and original price

T = Time after news published

Sen = Sentiment of news

Avg = Average trending value after news published

Index = Index reference

Cat = Industrial categories

Market = SET/MAI market reference

e = Error term

Hypotheses

Based on previous literature, we expect the following relationships between the dependent and independent variables:

Hypothesis 1: Stock price direction should move in the same direction as sentiment: an upward trend with positive sentiment and a downward trend with negative sentiment.

Hypothesis 2: Trends are factors that affect stock movement in a positive direction. When trends increase, stock prices should rise.

Hypothesis 3: Stocks included in an index have a greater effect on stock movement, whether upward or downward, compared with non-index stocks. We focus on the SET100 and SET50 indices.

Hypothesis 4: Some industrial categories may affect the stock movement more strongly than others and certain categories may have a positive effect.

Hypothesis 5: The SET may have a more immediate and stronger effect on stock movement than the MAI after the release of positive news.

Empirical Results

The finalized data from programming process was exported and cleaned into statistical variables. Outliers were removed and the final data returned.

Descriptive Statistics

The descriptive statistics, which provide measurements of central tendency and variability, are shown in Table 1 below.

Table 1 Descriptive Statistics

Variable	Mean	S.D.	Min	Max
P_T	-0.28	4.36	-42.67	50.00
P_{T+1}	0.11	3.35	-29.00	29.82
P_{T+2}	0.17	4.10	-29.00	40.20
P_{T+3}	0.22	5.16	-45.26	91.71
P_{T+4}	0.25	6.03	-61.68	91.71
P_{T+5}	0.20	6.67	-80.90	93.01
Avg	12.28	12.78	0.00	57.60

Factor variables are stored with their levels, which are sorted in alphabetical order. The first level of each factor variable is set as the reference variable. Stored values can be either numeric or string variables. In our dataset, factor variables are stored as string variables, as presented in Table 2.

Table 2 Factor Variables

Sen	Number of Observations	Variable
Negative (Reference)	540	
Neutral	1906	SenNeu
Positive	2490	SenPos
Index	**Number of Observations**	**Variable**
None (Reference)	2962	
SET100	799	IndexSET100
SET50	1175	IndexSET50
Cat	**Number of Observations**	**Variable**
AGRO (Reference)	540	
CONSUMP	226	CatCONSUMP
FINCIAL	489	CatFINCIAL
INDUS	482	CatINDUS
PROPCON	814	CatPROPCON
RESOURC	843	CatRESOURC
SERVICE	1039	CatSERVICE
TECH	503	CatTECH
Market	**Number of Observations**	**Variable**
MAI (Reference)	872	
SET	4064	MarketSET

Inferential Statistic

Multicollinearity refers to a high degree of intercorrelation between independent variables in multiple regression models, which can lead to misleading results. We apply the statistical technique known as the Variance Inflation Factor (VIF) to detect multicollinearity, as shown in Table 3. The results indicate that all variables fall below the maximum threshold (<10). This implies that there is no multicollinearity problem.

Table 3 Variance Inflation Factor

Variables	GVIF
Sen	1.004
Avg	1.116
Index	1.106
Cat	1.022
Market	1.132

Heteroscedasticity is a problem when the variances or standard deviations of the predicted variables are non-constant. The problem can violate assumptions and affect statistical results. We applied the Breusch-Pagan test to detect heteroscedasticity. As shown in Table 4, all the models have p-values greater than 0.05, which indicates that the null hypothesis of the Breusch-Pegan test cannot be rejected. Thus, no heteroscedasticity violation is present.

Table 4 Breusch-Pagan test

Model	Test statistic	Df	p-value
Price T	285.67	13	0.000***
Price T+1	279.43	13	0.000***
Price T+2	277.42	13	0.000***
Price T+3	124.73	13	0.000***
Price T+4	119.98	13	0.000***
Price T+5	130.79	13	0.000***

*** denotes level of significance at 0.01.

Autocorrelation refers to incorrect specification of the model or related variables with time lags. The violation affects the standard error, leading to inefficient predictions. We used the Durbin-Watson statistics to test for autocorrelation. As shown in Table 5, the models satisfy the test, with values falling within the acceptable range of 1.5 to 2.5. Thus, there is no autocorrelation problem with the models.

Table 5 Durbin-Watson test

Model	Test statistic	p-value
Price T	1.820	0.000***
Price T+1	1.841	0.000***
Price T+2	1.795	0.000***
Price T+3	1.809	0.000***
Price T+4	1.774	0.000***
Price T+5	1.788	0.000***

*** denotes level of significance at 0.01.

Empirical Results

From the dataset of 4,936 samples collected between August 2021 and December 2021, we applied a multiple regression model. Six models were used to investigate the relationship between stock returns and the independent variables. The range of stock returns spans from the day the news is published to five days afterward.

Table 6 Multiple regression analysis of stock returns on the day of news publication

Variable	Coefficient	Std. Error	t-stat	p-value
C	0.089	0.308	0.289	0.773
SenNeu	-0.931	0.212	-4.385	0.000***
SenPos	-1.086	0.207	-5.250	0.000***
Avg	-0.005	0.005	-0.866	0.387
IndexSET100	-0.445	0.184	-2.416	0.016**
IndexSET50	-0.204	0.178	-1.149	0.251
CatCONSUMP	-0.185	0.357	-0.517	0.605
CatFINCIAL	0.508	0.276	1.843	0.065
CatINDUS	0.263	0.279	0.941	0.347
CatPROPCON	-0.026	0.243	-0.107	0.915
CatRESOURC	0.417	0.242	1.722	0.085*
CatSERVICE	0.808	0.232	3.475	0.001***
CatTECH	0.924	0.270	3.422	0.001***
MarketSET	0.391	0.183	2.136	0.033**

*, **, *** denote level of significance at 0.10, 0.05, and 0.01.

From Table 6, seven significant independent variables are found to influence stock returns on the day of news publication; sentiment, SET100, industrial categories, and market type. Positive and neutral sentiments show a negative relationship with stock returns. Moreover, stocks in the SET100 index also exhibit a negative relationship with returns, implying that news related to SET100 stocks negatively affects their performance. By contrast, resource stocks, service stocks, technology stocks, and stocks in the SET market show a positive relationship with returns. This implies that news concerning stocks, and stocks in the resource, service, or technology industries, or in the SET market, is positively associated with stock prices on the day of publication. However, stock closing prices and sentiment move in opposite directions on the day of news publication, indicating the need for further investigation. Therefore, we introduce the concept of time-lagged information in this research, which will be discussed in later sections. Nonetheless, our model shows that the sentiment variable is related to the returns at a confidence level greater than 99 percent.

Table 7 Multiple regression analysis of stock returns one day after news publication

Variable	Coefficient	Std. Error	t-stat	p-value
C	-0.821	0.236	-3.477	0.001***
SenNeu	0.493	0.163	3.021	0.003***
SenPos	0.528	0.159	3.321	0.001***
Avg	0.009	0.004	2.162	0.031**
IndexSET100	0.211	0.141	1.494	0.135
IndexSET50	0.228	0.136	1.668	0.095*
CatCONSUMP	0.213	0.275	0.775	0.439
CatFINCIAL	0.471	0.212	2.226	0.026**
CatINDUS	0.031	0.214	0.146	0.884
CatPROPCON	-0.190	0.187	-1.013	0.311
CatRESOURC	-0.115	0.186	-0.616	0.538
CatSERVICE	0.205	0.179	1.148	0.251
CatTECH	0.241	0.208	1.162	0.245
MarketSET	0.243	0.141	1.727	0.084*

*, **, *** denote level of significance at 0.10, 0.05, and 0.01.

Table 8 Multiple regression analysis of stock returns two days after news publication

Variable	Coefficient	Std. Error	t-stat	p-value
C	-0.836	0.290	-2.885	0.000***
SenNeu	0.444	0.200	2.220	0.027**
SenPos	0.522	0.195	2.677	0.008***
Avg	0.014	0.005	2.681	0.007***
IndexSET100	0.140	0.174	0.809	0.419
IndexSET50	0.035	0.167	0.211	0.833
CatCONSUMP	-0.082	0.337	-0.243	0.808
CatFINCIAL	0.577	0.260	2.220	0.026**
CatINDUS	0.210	0.263	0.798	0.425
CatPROPCON	0.093	0.230	0.406	0.685
CatRESOURC	-0.085	0.228	-0.374	0.708
CatSERVICE	0.296	0.219	1.353	0.176
CatTECH	0.449	0.255	1.764	0.078*
MarketSET	0.230	0.173	1.331	0.183

*, **, *** denote level of significance at 0.10, 0.05, and 0.01.

Tables 7 and 8 present stock returns following the news publication. We examine one- and two-day time lags using multiple regression. The independent variables that are significant in both models are sentiment, trend, and financial industry. The sentiment and trend variables are significant independent variables with a certain confidence level. The sentiment and trends are co-movement with the return, showing a positive relationship. The variables are significant with a time lag after the news has been published. We can conclude that sentiment is related to stock returns from the day of the news release to two days after. Moreover, trend variable is related to the returns from one to two days after the news was released. The stock in financial industry has a positive relationship with the return along with the sentiment and trends. However, some variables that are significant on the day of the news release do not maintain significance over the lagged periods.

Table 9 Multiple regression analysis of stock returns three days after news publication

Variable	Coefficient	Std. Error	t-stat	p-value
C	-1.005	0.365	-2.753	0.006***
SenNeu	0.121	0.252	0.482	0.630
SenPos	0.347	0.246	1.412	0.158
Avg	0.003	0.006	0.398	0.691
IndexSET100	0.308	0.218	1.411	0.158
IndexSET50	0.262	0.211	1.241	0.215
CatCONSUMP	0.787	0.423	1.856	0.064
CatFINCIAL	0.837	0.327	2.559	0.011**
CatINDUS	0.452	0.331	1.366	0.172
CatPROPCON	0.288	0.289	0.997	0.319
CatRESOURC	-0.094	0.287	-0.326	0.745
CatSERVICE	0.503	0.276	1.824	0.068
CatTECH	0.572	0.321	1.784	0.074
MarketSET	0.609	0.217	2.802	0.005***

*, **, *** denote level of significance at 0.10, 0.05, and 0.01.

Table 10 Multiple regression analysis of stock returns four days after news publication

Variable	Coefficient	Std. Error	t-stat	p-value
C	-0.754	0.427	-1.767	0.077*
SenNeu	-0.244	0.295	-0.830	0.407
SenPos	0.067	0.287	0.233	0.816
Avg	-0.004	0.008	-0.518	0.605
IndexSET100	0.365	0.255	1.429	0.153
IndexSET50	0.348	0.246	1.410	0.159
CatCONSUMP	1.015	0.496	2.048	0.041**
CatFINCIAL	0.889	0.382	2.325	0.020**
CatINDUS	0.267	0.387	0.689	0.491
CatPROPCON	0.196	0.338	0.581	0.561
CatRESOURC	-0.161	0.336	-0.479	0.632
CatSERVICE	0.577	0.323	1.790	0.074*
CatTECH	0.506	0.375	1.351	0.177
MarketSET	0.769	0.254	3.025	0.003***

*, **, *** denote level of significance at 0.10, 0.05, and 0.01.

Table 11 Multiple regression analysis of stock returns five days after news publication

Variable	Coefficient	Std. Error	t-stat	p-value
C	-0.948	0.472	-2.007	0.045**
SenNeu	-0.139	0.326	-0.426	0.670
SenPos	0.266	0.318	0.838	0.402
Avg	-0.008	0.008	-0.902	0.367
IndexSET100	0.439	0.283	1.551	0.121
IndexSET50	0.330	0.273	1.208	0.227
CatCONSUMP	0.924	0.549	1.682	0.093*
CatFINCIAL	0.974	0.423	2.302	0.021**
CatINDUS	0.433	0.428	1.010	0.312
CatPROPCON	0.391	0.374	1.045	0.296
CatRESOURC	-0.056	0.372	-0.151	0.880
CatSERVICE	0.740	0.357	2.071	0.038**
CatTECH	0.692	0.415	1.667	0.096*
MarketSET	0.663	0.282	2.356	0.019**

*, **, *** denote level of significance at 0.10, 0.05, and 0.01.

From Tables 9, 10, and 11, neither sentiment nor trends are significantly related to stock returns. Thus, we can conclude that sentiment and trends do not affect stock prices three days after the news is published. Further investigation is required to explain other significant industry factors

Conclusion and Recommendation

The purpose of this research is to identify the relationship between sentiment, conventional online behavior, and stock returns. We used samples of financial and economic news from online platforms between August 2021 to December 2021, during the easing of COVID-19 restrictions in Thailand. Independent variables were generated by including two main variables, sentiment and trends, along with stock characteristics as alternative independent variables in the model. The dependent variables are stock returns with lags from the publication date to five days afterward. From the empirical results, we can confidently conclude that sentiment and trends are factors that affect stock returns. Sentiment influences returns two days after the news is published. On the other hand, trends begin to influence returns from one to two days after publication. Several other variables also need to be considered. Indices and markets are factors that influence stock prices within a day of publication, while certain. industries can be influential at different points in time, depending on their characteristics. The relationship between sentiment and returns on the day of publication requires further investigation.

This research aims to study how news sentiment affects stock movements. For further theoretical implications, our model can be implemented and applied as a foundation or reference for sentiment-based models across various fields. Additionally, the data gathering and processing methods in our model can be improved and applied to any field, without limitations on dataset size. As a practical implication, the magnitude and direction of the independent variables can be used to predict stock movements within a specific time frame.

Sentiment refers to subjective opinions that vary from person to person. Manual input of sentiment is recommended for small datasets. However, implementing a program is more suitable for large datasets or big data. As technology improves, machines will be able to handle larger datasets, leading to more accurate results. In Thai language processing, new techniques are continuously improving. Thus, while the general concepts and thought process remain similar, future implementations may vary from this research. Furthermore, economic factors could be incorporated into future research.

References

Chatchawan, S. (2021). Can sentiment from news headlines explain stock market returns?: Evidence from Thailand. *Thammasat Review*, *24*(1), 317-333.

Chatchawan, S. (2022). How does news-based sentiment affect stock market volatility in the Stock Exchange of Thailand? *KKBS Journal of Business Administration and Accountancy*, *5*(3), 161-182.

Chiewhawan, T., & Vateekul, P. (2020). Explainable deep learning for Thai stock market prediction using textual representation and technical indicators. In *Proceedings of the 8th International Conference on Computer and Communications Management* (pp. 19-23).

Liu, L., Wu, J., Li, P., & Li, Q. (2015). A social-media-based approach to predicting stock comovement. *Expert Systems with Applications*, *42*(8), 3893-3901.

Luo, X., Zhang, J., & Duan, W. (2013). Social media and firm equity value. *Information Systems Research*, *24*(1), 146-163.

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603-9611.

Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., & Deng, X. (2013). Exploiting topic based Twitter sentiment for stock prediction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp. 24-29).

Sul, H. K., Dennis, A. R., & Yuan, L. I. (2016). Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, *48*(3), 454-488.

Tu, W., Cheung, D. W., Mamoulis, N., Yang, M., & Lu, Z. (2016). Investment recommendation using investor opinions in social media. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 881-884).

Vu, T., Chang, S., Ha, Q. T., & Collier, N. (2012). An experiment in integrating sentiment features for tech stock prediction in Twitter. In *Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data* (pp. 23-38).

Wansri, T., Phaksuwan, A., Iamtae, K., & Chaturantabut, S. (2023). Prediction models for tourism stock market trends during COVID-19 pandemic using news sentiment analysis with data mining: A case study. *Science and Technology Asia*, *28*(3), 98-115.