News Analytics for Business Sentiment Suggestion

Sirinda Palahan
School of Science and Technology
University of the Thai Chamber of Commerce
Bangkok, Thailand

***Abstract*—Business and economics news has become one of the factors businesses consider when making decisions. However, the exponential increase in the availability of business information sources on the internet makes it more difficult for entrepreneurs to keep up with and extract useful insights from many news articles. Although many preceding works focused on the sentiment extracted in the news, the results were intended for everyone. The sentiments based on a user's queries are needed to provide customized service. Hence, this paper proposed a system integrated into a chatbot to automatically understand users' queries and recommend sentiments based on news articles. The main objective is to provide entrepreneurs, especially those considering international trade and investment, with the sentiments embodied in the latest news articles to help them keep up with the business and economic trends relevant to them. The methodology is based on deep learning and transfer learning. A pre-trained deep learning model was fine-tuned for natural language processing tasks to perform sentiment analysis in news articles. A survey questionnaire was used to measure the effectiveness of the system. The survey result showed that most users agreed with the predicted sentiments from the system.**

Keywords*—*Sentiment analysis; deep learning; pre-trained model; natural language processing

I. Introduction

One of the important sources of business information for entrepreneurs is news articles. News, especially business and economics news, has become one of the factors that entrepreneurs consider when deciding to trade or expand the business to other countries. The news could be more than just a medium to report on what has happened or carry information about the topics or entities discussed in the articles. They could also carry sentiments or trends toward the main content, such as products or companies mentioned in the articles. In other words, news conveys how the authors and people feel about the entity. This public sentiment is crucial to business because it tells how the public feels about the product or service. It also significantly influences the future of the product or service industries. But with the advancement in internet technology, the amount of news articles is increasing rapidly. Therefore, it has become a challenge for entrepreneurs to keep up with the rapidly increasing number of news sources, filter out unreliable sources, find all relevant information they need, and extract sentiments manually.

Sentiment analysis is a natural language processing (NLP) technique used to determine and extract subjective information in a text. The information could be people's opinions, sentiments, or attitudes towards entities. Due to the rapid growth of data on social media, sentiment analysis has become one of the most active research topics in NLP. Moreover, due to its superior performance in many application domains, deep learning has gained popularity, thanks to advanced cloud-based technology and increasing computing power. Among the success of deep learning in various application domains, deep learning has been used in sentiment analysis [1-3]. Many deep learning models can be applied or adapted to NLP datasets to attain high accuracy [4]. Therefore, this paper focuses on applying deep learning to automatically analyze many news articles and recommend sentiments based on news articles. The suggested sentiment can be considered an indicator that reflects consumers' attitudes and economic outlook toward the specific product in the country. Therefore, it can help a user make proper business decisions.

There are two main contributions to this paper. First, a deep-learning-based model is proposed to automatically understand users' queries and analyze and recommend sentiments based on current news articles. Second, the proposed model was integrated into a chatbot to provide an end-to-end solution in a business and economic domain so that the model could be tested in practice. As a result of this system, users can stay up to date on business and economic trends relevant to their specific fields.

The rest of the paper is organized as follows. First, an overview of the related works is presented in Section II. Next, the methodology of the proposed system is discussed in Section III. Then Section IV shows the results and findings. Finally, Section V gives conclusions regarding this work.

II. Related Work

Sentiment analysis has begun to be used in economics and finance recently. Many studies are concentrating on using sentiment analysis for stock prediction. Several works [5-7] analyzed tweets' moods or opinions and used machine learning methods for prediction. [5] studied whether the collective mood states extracted from the Twitter feed can be used to predict the value of the Dow Jones Industrial Average (DJIA). They used a Self-Organizing Fuzzy Neural Network to predict the changes in DJIA closing values. The results showed that the prediction accuracy was significantly improved when including a calm sentiment. [6] calculated a sentiment score of each tweet using a dictionary-based method and created feature vectors of sentiment scores to train a support vector machine (SVM) to classify the stock trend. The best accuracy was 90.34%. [7] also used a dictionary-based method with eight sentiments to analyze Twitter data. However, the results showed that adding sentiment data did not significantly improve accuracy.

Recent works [8-10] have developed a prediction model based on a deep learning approach. [8] used deep learning models to extract features from news headlines and predict stock prices. A Convolutional Neural Network (CNN) model was used to transform the sequence of words in the news title to the level of sentiment. The sentiment and other technical indicators were inputs to the Long Short-term Memory (LSTM) model to predict the price movement. The model achieved a 97.66% accuracy rate. [9] employed a similar approach where they extracted sentiments in the news content on Sina Weibo, China's largest online social network. They then input the sentiment features and technical indicators into an RNN-boost model to predict the stock volatility in the Chinese stock market. [10] enhanced LSTM by utilizing an attention mechanism to make predictions on the final output and informative outputs from hidden states. The results show that the attention-based LSTM improved prediction accuracy and reduced computational time.

Several other papers have focused on the economic sentiment embodied in the news and social media, especially Twitter. The author in [11] proposed a new technique to measure economic sentiment embodied in the news. Unlike survey-based economic sentiment measures, their index relies on extracting sentiment from news articles. The technique was applied to two applications. In the first application, they use their news sentiment index to predict consumer sentiment based on surveys. They found that the news sentiment is strongly predictive of Michigan Consumer Sentiment Index and the Conference Board's Consumer Confidence Index. Second, they investigated how the macro-economic response to sentiment shocks. They found that positive sentiment shocks increase consumption, output, and interest rates and temporarily reduce inflation. The author in [12] build a model to predict the sentiment index of a company based on news about that company. The authors evaluated their model by measuring the model's accuracy to predict the company's stock price movement. The result showed that the model has an average accuracy score of around 70.1%. The author in [13] proposed an automatic news chatbot that provides a variety of news articles organized into chatrooms based on news topics. When a user enters a chatroom, the chatbot provides the latest news articles on a given topic. A user can also ask specific questions regarding the topic, which a chatbot will answer with short sentences and provide a link to an article containing the answer. The drawback is a user cannot start a conversation by asking a specific question. Instead, a user must pick one of the topics to have conversations about that topic.

While most of the preceding work focused on the sentiment extracted in the news content where the results were intended for everyone, such as the consumer sentiment or sentiments for stock predictions, the sentiments based on a user's queries are needed to provide a customized service to the users. As a result, this work proposed a system that automatically understands users' queries and recommends sentiments based on current news articles. The system was integrated into a chatbot to provide easy access to the system for users. The system would help users to keep up with the latest business and economic trends in their fields of interest.

III. Methodology

This section describes the system architecture and how it processes users' queries to suggest product sentiments to users. The system consists of four components: a conversational interface, a keyword extraction module, a news search module, and a sentiment analysis module, as shown in Fig. 1. First, a user sends a query through a conversational interface. Next, the query is sent through the system. After receiving the query, the system extracts a keyword from the query and gathers news articles from trusted sources related to the keywords. It then analyzes the articles and suggests the sentiment. The news articles' list and their sentiment are then sent back to a user. The details of each component are explained in the following subsections.

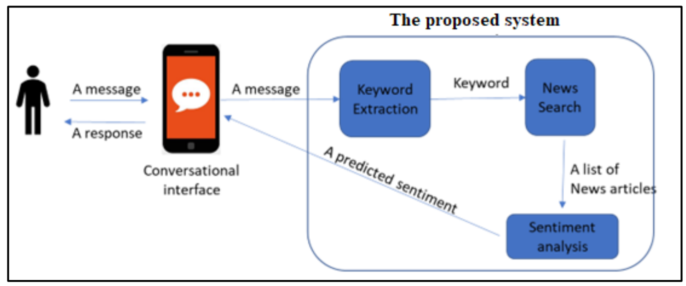
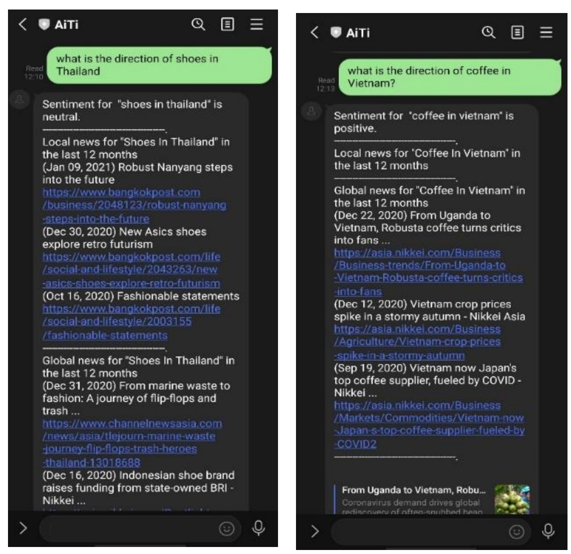


Fig. 1. The Architecture of the Proposed System.

A. The Conversational Interface

The first component is a conversational interface acted as a chatbot to interact with users. The LINE messenger app was used as the conversational interface for this work because it is Thailand's most used messaging app [14]. Thus, the system can easily be accessible to Thai users.



(a) (b)

Fig. 2. The Conversational Interface. (a): Regular Response (b): Response with Empty List of Local News.

The conversational interfaces are shown in Fig. 2(a) and 2(b). Fig. 2(a) shows a response to the LINE messenger app when a user sends a message, "What is the direction of shoes in Thailand?". The first line in the response message is the sentiment of the keywords "shoes" and "Thailand." After the first line, it shows a list of local and global news articles related to the keywords. The local news is the news from local news providers such as bangkokpost.com or vietnamnews.vn. The global news is the news from global news providers such as channelnewsasia.com or asia.nikkei.com. If there is no related news, the list will be empty, as shown in Fig. 2(b) where the list of local news is empty.

The system suggests three types of sentiment: positive, neutral, and negative. The interpretation of the sentiments of keywords is as follows. The positive and negative sentiments mean the keywords have a positive or negative economic outlook. While the neutral sentiment means the keywords have neither a positive nor negative economic outlook.

B. *The Keyword Extraction Module*

The main objective of this module is to extract keywords in a user's query and pass the keywords to the news search module for searching relevant news articles. The keywords can be a single word or a combination of words in a query. Fig. 3 displays an overview of the input and output of the keyword extraction module. First, a user sends a message, "What is the trend of tourism in Myanmar?" via a conversational interface. Next, the message is sent to the keyword extraction module. The module then extracts keywords; in this example, the extracted keywords are "tourism" and "Myanmar." The News search module then uses this keyword to search for relevant news articles. If the module cannot identify any keywords, the system will tell a user that no keywords can be identified and ask a user to try a different query.

The module first tokenizes the sentences and identifies parts of speech (POS) tags to extract keywords. The keyword in the message is a combination of words tagged as nouns and proper nouns, both singular and plural. The result of tokenizing and POS tagging the sentence “What is the trend of tourism in Myanmar?' looks like this:

('What', 'WP')

('is', 'VBZ')

('the', 'DT')

('trend', 'NN')

('of', 'IN')

('tourism', 'NN')

('in', 'IN')

('Myanmar', 'NNP')

('?', '.')

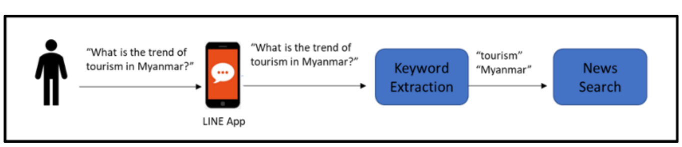


Fig. 3. An Overview of the Input and Output of the Keyword Extraction Module.

The tagging result has three nouns: 'trend,' 'tourism,' and 'Myanmar' in the sentence. After filtering out the word 'trend,' the keyword is "tourism Myanmar."

This work used the NLTK toolkit [15] for tokenizations and POS tagging. NLTK is a leading tool for working with NLP in Python. It provides various libraries for text processing, including tokenizations and POS tagging. Tokenization is how a sentence is broken into words and punctuations. For example, "What is the trend of tourism in Myanmar?" is tokenized as ['What', 'is', 'the', 'trend', 'of', 'tourism', 'in', 'Myanmar', '?']. After that, the module classifies words into parts of speech and labels them accordingly using a POS tagger. The POS tagger uses a UPenn Tagset, as shown in Table I.

TABLE I. A LIST OF PART-OF-SPEECH TAGS

Tag	**Description**
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential *there*
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
M.D.	Modal
N.N.	Noun, singular, or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
R.B.	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	*to*
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund, or present participle
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP$	Possessive wh-pronoun
WRB	Wh-adverb

C. News Search Module

This module searches for relevant articles across trusted news sources using the extracted keyword from the keyword extraction module. Two lists of trustworthy news sources were curated: a local and global list. The local list contains high-reputable local news sources from each country, while the global list contains high-reputable news sources covering news mainly from Asia and worldwide. The local news sources are as follows:

- https://en.vietnamplus.vn
- https://english.vov.vn
- https://aecnewstoday.com
- www.irrawaddy.com
- https://elevenmyanmar.com
- https://english.cambodiadaily.com
- www.phnompenhpost.com
- https://laotiantimes.com
- https://www.bangkokpost.com/

The global news sources are as follows:

- asiatimes.com
- www3.nhk.or.jp
- https://asia.nikkei.com
- https://www.asiaone.com
- https://www.businessnewsasia.com
- https://www.channelnewsasia.com
- http://annx.asianews.network
- www.businesstimes.com.sg

Google Programmable Search Engine (formerly known as Google Custom Search) was used to search across a collection of local and global news sources. A custom search JSON API was created with the Google Programmable Search Engine to retrieve search results automatically in JSON format via RESTful requests. In addition, this module was set to return news articles published within 12 months.

D. Sentiment Classification Module

This module classifies a sentiment related to the user's keyword. The sentiment of the keyword is based on sentiments of its related news articles. The keyword's sentiment depends on how the news providers and people feel about the content the keyword was discussed. This module consists of two models: a deep learning model and a classifier, as shown in Fig. 4.

Fig. 4 shows how the Sentiment Classification Module works. This example uses a user query "What is the trend of tourism in Myanmar?" where the news search module returns a list of news articles as follows:

- Article 1: Week in Review: Tourist arrivals fall 75% in 2020.
- Article 2: Myanmar sees 75% drop in tourist arrivals.
- Article 3: Mandalay eyes tourism revival post-pandemic.
- Article 4: Mandalay prepares to reopen hotels, revive tourism.

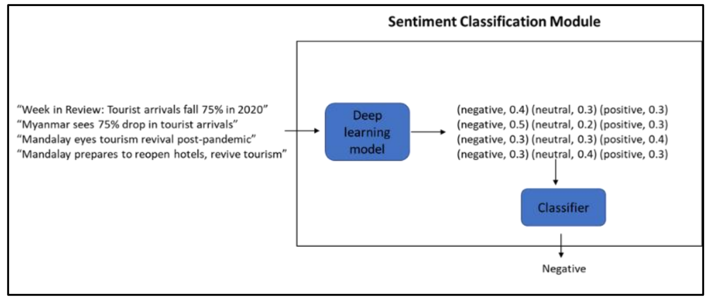


Fig. 4. An Example of How the Sentiment Classification Module Works.

After receiving a list of news articles, the sentiment classification module performs two steps. First, it calls a deep learning model to calculate sentiment scores for each article. There are three sentiments: negative, neutral, and positive. Hence each article is assigned three scores for the three sentiments. The result from the first step would look like this:

- Article 1: (negative, 0.4) (neutral, 0.3) (positive, 0.3).
- Article 2: (negative, 0.5) (neutral, 0.2) (positive, 0.3).
- Article 3: (negative, 0.3) (neutral, 0.3) (positive, 0.4).
- Article 4: (negative, 0.3) (neutral, 0.4) (positive, 0.3).

Second, it classifies a sentiment for the keywords by choosing the sentiment with the highest average scores from all news articles. From the example, the averages of the three sentiments from four articles are below:

- (negative, 0.375) (neutral, 0.3) (positive, 0.325)

The averages show that the negative sentiment has the highest score, so it classifies the sentiment's keywords as "Negative". The system then sends the result back to the user, saying, " The tourism trend in Myanmar is Negative.".

1) Sentiment classification module: Many works provide methods and tools for sentiment analysis. Recently, most meth-ods have been based on deep learning with transfer learning. Deep learning is a sub-field of machine learning methods based on neural networks. The networks typically have three or more layers, performing feature extraction and transformation via a cascade of multiple layers of linear and nonlinear processing nodes. The lower hidden layers near the input layer learn simple features, while the higher layers learn complex features derived from lower layers.

A transfer learning is a machine learning technique where a model is trained for one task and applied to a different but related task. The common approach to using transfer learning for deep learning is to use a pre-trained model. A pre-trained model is chosen from available models typically released by research institutes for this approach. The pre-trained model is then trained and fine-tuned on a new dataset for the task of interest.

Many pre-trained deep learning models can be employed for the task. Hence, four models were tested in the experiments, Bert, RoBERTa, XLM-R, and XLNet. The best model would be selected for the system. Google's Bert [16] is a bi-directional model pre-trained on an unlabeled text that can be used to create a wide range of tasks by fine-tuning with just one additional output layer. RoBERTa [17] extends the original BERT model where the researchers fine-tune the original BERT model with a huge dataset and improved input representation. XLM-R [18] is a cross-lingual language model trained with MLM (Masked Language Modeling) on one hundred languages and terabytes of texts. Finally, XLNet [19] is an auto-regressive language model which uses the context word to predict the next word. The model has also addressed some drawbacks of BERT and outperformed BERT in many tasks.

A linear layer was added on top of the pooled output to calculate sentiment scores, as shown in Fig. 5. Next, all models were trained and fine-tuned for a classification task where the outputs are probabilities of three classes: positive, negative, and neutral.

2) The dataset: The dataset used to train and fine-tune models was curated by [20]. The dataset contains about 5000 news titles from financial news resources, where each title was labeled as positive, negative, or neutral by annotators with good backgrounds in business and investment. The classes reflect an investor's perception of business and market conditions on the news. The positive class means the news title appears to positively influence the market and vice versa for the negative class. If the news title reflects neither positive nor negative influence, the news title is considered neutral. Each news title was trimmed to have the maximum sequence length of 256 for fast processing in a real-time environment. Examples of news titles and their sentiments are shown in Table II.

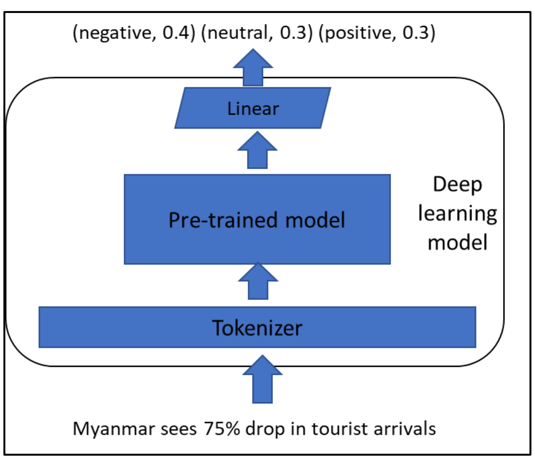


Fig. 5. The Architecture of the Proposed Deep Learning Model.

TABLE II. EXAMPLES OF NEWS TITLES AND THEIR SENTIMENTS

News Title	Sentiment
According to Gran, the company has no plans to move all production to Russia, although that is where the company is growing.	neutral
Technopolis plans to develop in stages an area of no less than 100,000 square meters to host companies working in computer technologies and telecommunications, the statement said.	neutral
With the new production plant, the company would increase its capacity to meet the expected increase in demand and would improve the use of raw materials, and therefore increase the production profitability.	positive
According to the company's updated strategy for the years 2009-2012, Basware targets a long-term net sales growth in the range of 20 % -40 % with an operating profit margin of 10 % - 20 % of net sales.	positive
The international electronics industry company Elcoteq has laid off tens of employees from its Tallinn facility; contrary to earlier layoffs the company contracted the ranks of its office workers, the daily Postimees reported	negative
A TinyURL link takes users to a scamming site promising that users can earn thousands of dollars by becoming a Google (NASDAQ: GOOG) Cash advertiser.	negative

IV. RESULTS AND DISCUSSION

This section shows the performance of different pre-trained deep learning models for model selection and the system's performance based on a user survey. Error analysis and some discussions are also provided.

A. Model Selection

Four deep learning models were tested for model selection. The models were Bert, RoBERTa, XLM-R, and XLNet. The Transformers and FastAI libraries [21, 22] with the default architecture were used to implement the deep learning models where weights were initiated randomly. First, the dataset was divided into a train and test set with a split ratio of 60:40. The training set was used to fine-tune the pre-trained models, and the test set was used to select the best model. Next, the models were trained with the Adam optimization algorithm with an accuracy rate as a metric. Adam optimization algorithm is an algorithm to update the weights in neural networks during training. The Adam optimization algorithm was chosen because it works well in many empirical results and is recommended as the default algorithm [23, 24]. The key metrics to evaluate the performance of the models are accuracy rate and computation time on the test set.

From Table III, the accuracy rates of Bert, RoBERTa, XLM-R, and XLNet are 0.8500, 0.8633, 0.7367, and 0.8667, respectively. From the results, RoBERTa is tied with XLNet regarding the accuracy rate, but it took the least computational time, so RoBERTa was chosen for the system.

TABLE III. EXPERIMENTAL RESULTS ON A MODEL SELECTION

Model	Accuracy Rate	Computational Time (Seconds)
Bert	0.8500	0.0428
RoBERTa	0.8633	0.0378
XLM-R	0.7367	0.3742
XLNet	0.8667	0.0399

B. Performance Evaluation

There is no public ground truth for evaluating the accuracy of the predicted sentiments of keywords because a sentiment of a keyword depends on the current economics, and it can change as time goes by. For this reason, a user survey was employed to measure the system's effectiveness from users' perspectives. First, the participants were asked to make a few queries using keywords of their choices and give feedback on the results. After that, users were asked two questions:

1) On a scale of 1 – 5, with 1 being highly irrelevant and 5 being highly relevant, how relevant each news article is to your keyword?

2) What sentiment would you assign to your keyword from your perspective?

The first question is to measure the relevancy of the news articles to the keyword. Participants were asked to rate an individual news article. A relevance rating for the keyword was calculated by averaging the ratings of all of its news articles. The second question measures the accuracy of the keyword sentiment from the users' perspective. It is used as a keyword accuracy rate, the fraction of predictions the system got right according to users' perspectives.

For example, a user searching for a 'What's the trend of shoes in Vietnam?' result is shown in Fig. 6. In the figure, the predicted sentiment for the keyword is "Positive," where the extracted keyword for the user's query is 'shoes Vietnam'. The system also returned related recent news articles to the user. Each title contains a link to its original news source.

Next, the participants were asked to rate each article in terms of relevancy to the keywords. The rating scale is 1 to 5. Table IV shows a news relevance rating for each news article and the news relevance score, which is an average of all ratings. For this example, the news relevance score is 3.88, indicating that the news articles are somewhat relevant to the keyword. The participants were also asked what sentiment they would assign to the keyword based on the news and their perspective. For example, for the keyword 'shoes in Vietnam', the user assigned a 'Positive' sentiment the same as the sentiment predicted by the system.

Search result for 'What's the trend of shoes in Vietnam?':

* The predicted sentiment is 'Positive'.

* Related news articles:

- Footwear and textile set for a strong bounce back - Vietnam News
- Leather and footwear on course for strong recovery: LEFASO – Vietnam News
- COVID-39 woes: Footwear exports likely to fall short of the target
- Coffee shoes help entrepreneurs tread new ground - Vietnam News
- Vietnam's bittersweet moment in Trump's spotlight .. Nikkei Asia
- Global manufacturers are flocking to Vietnam. Is it ready? ..Nikkei Asia
- Vietnam greenlights E.U. trade pact in a bid for China-exit deals ...
- Global footwear group's Vietnam operations were suspended for two days

Fig. 6. Search Result for a user's Keyword.

TABLE IV. RELEVANCE RATINGS

Keyword	News Title	News Relevance Rating
shoes in Vietnam	Footwear and textile set for a strong bounce back - Vietnam News	5
	Leather and footwear on course for strong recovery: LEFASO - Vietnam News	5
	COVID-39 woes: Footwear exports likely to fall short of the target	3
	Coffee shoes help entrepreneurs tread new ground - Vietnam News	5
	Vietnam's bittersweet moment in Trump's spotlight - Nikkei Asia	3
	Global manufacturers are flocking to Vietnam. Is it ready? - Nikkei Asia	3
	Vietnam greenlights E.U. trade pact in bid for China-exit deals ...	2
	Global footwear group's Vietnam operations were suspended for two days	5
	Relevance rating	**3.88**

Fifteen users participated in the survey. Seven of the users were entrepreneurs, and the remaining users were senior students with a major in international business management. Each user made 3 to 4 queries. The total number of queries was 40. The survey results show that the average relevance rating is 2.66 out of 5, and the accuracy rate of the keyword sentiment is 0.35. The average relevance rating and accuracy rate for each sentiment category were also computed, as shown in Table V. The results look promising, with a 62.50% accuracy rate for the positive articles and a 2.93 relevance rating for the negative articles.

TABLE V. THE RELEVANCE RATING AND THE ACCURACY RATE BY CATEGORY

Sentiment	Relevance Rating	Accuracy Rate
Negative	2.9342	0.3636
Neutral	2.5133	0.2380
Positive	2.7142	0.6250

C. Error Analysis

After the evaluation process, classification errors were examined to identify how to improve the system. First, a confusion matrix was constructed to see what categories the system misclassified and what categories had been predicted correctly. From the confusion matrix in Table VI, the system predicted Neural sentiments the most (21/40), whereas users assigned Positive sentiments the most (24/40). So the results of 14 keywords that the system predicted Neutral sentiments, but users assigned Positive sentiments were examined. The examination result shows that a few articles sound positive but were assigned Neutral sentiment by the system. The Neutral sentiment also received the lowest relevance rating and accuracy rate, as shown in Table V. These results may be due to the class imbalance in the dataset used to train the model, where 59% of articles were Neutral. In comparison, only 12% and 28% of articles were Negative and Positive, respectively. Based on this finding, the system's performance in the Neutral category could be improved by employing a resampling technique to deal with an unbalanced dataset.

TABLE VI. CONFUSION MATRIX

		User Sentiment			Total
		Negative	**Neutral**	**Positive**	
Predicted Sentiment	**Negative**	4	2	5	11
	Neutral	2	5	14	21
	Positive	2	1	5	8
Total		8	8	24	40

Second, the queries that received the correct sentiments were analyzed. The result shows that they also received a higher news relevance rating. The finding may indicate that higher relevant news articles will likely improve the sentiment accuracy. To test if higher relevant news articles would help boost the accuracy rate, queries whose average relevance rating is less than 3 were removed and re-calculated the accuracy rate. There are 22 queries whose average relevance rating is less than 2. After removing those queries, the accuracy rate went up to 0.5. From the analysis, the search engine should be improved to return more relevant news articles so that the algorithm has better information to calculate a sentiment.

Third, since the accuracy of keyword sentiments depends on the accuracy of predicted sentiments of news articles related to the keyword, the senior students who participated in the survey were asked to rate sentiments of the news articles in the search results from the survey. There were 268 news articles from 40 queries. Next, the accuracy rate and a macro-F1 score of the news sentiment were computed by comparing the predicted sentiment classifications from the model to the sentiments rated by the users. The accuracy rate and a macro-F1 score are 0.6679 and 0.4343, respectively. These results show that the system could accurately classify news sentiments, but those articles were irrelevant to their keywords. These results also emphasize the need to improve the system's search engine.

D. Discussion

The accuracy rate and a macro-F1 score of the news sentiment from the system are 0.6679 and 0.4343, respectively. These results are similar to the result in [11], where the authors proposed a new technique to measure economic sentiment embodied in the news. In [11], the macro-F1 score was calculated using the 100-article test set for which they compared predicted sentiments from various models to human-provided sentiments. The lowest macro-F1 score was 0.406 from using Lexicon only, and the highest macro-F1 score was 0.525 from their proposed method. According to these macro-F1 scores, sentiment analysis is still very challenging. The majority of the current sentiment analysis techniques are data-driven machine learning techniques. Hence they have limitations in terms of data size and inconsistency of ground truth [25]. Furthermore, labeling needs to be done by humans, and human ability to label large volumes of data is limited. Moreover, there is a great deal of subjective opinions when it comes to business news. A bad situation for one business might be good for another business. As a result, it is essential to provide clear and concise instructions to produce high-quality annotations. In this study, the participants were asked to make queries related to their businesses or expertise to evaluate results from those businesses' perspectives.

V. CONCLUSION

This paper presents a methodology to automatically understand a user's query about a product and provide its sentiment embodied in news articles. The system is based on deep learning and transfer learning to build a model using a pre-trained deep learning model fine-tuned for sentiment analysis in news articles. The news articles are automatically searched and collected by the news search module from the lists of trustworthy news sources both locally and globally to ensure the quality of news articles. Finally, the model was integrated into a chatbot and tested in practice. The satisfaction survey shows participants agreed with the results, with a relevance rating of 2.66 and an accuracy rate of 35%. The evaluation by category shows that the positive articles received the highest accuracy rate of 62.50%, while the negative articles received the highest relevance rating of 2.93. In the future, the news search module could be improved to return relevant search results with higher precision by adding more capabilities such as semantic understanding to understand user intention better. Moreover, the sentiment classification module could be improved by re-training the deep learning model with different parameters.

REFERENCES

[1] Chen, Y., Convolutional neural network for sentence classification. 2015, University of Waterloo.

[2] Dos Santos, C. and M. Gatti. Deep convolutional neural networks for sentiment analysis of short texts. In Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers. 2014.

[3] Guggilla, C., T. Miller, and I. Gurevych. CNN-and LSTM-based claim classification in online user comments. In Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers. 2016.

[4] Zhang, L., S. Wang, and B. Liu, Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018. 8(4): p. e1253.

[5] Bollen, J., H. Mao, and X. Zeng, Twitter mood predicts the stock market. Journal of computational science, 2011. 2(1): p. 1-8.

[6] Meesad, P. and J. Li. Stock trend prediction relying on text mining and sentiment analysis with tweets. In 2014 4th World Congress on Information and Communication Technologies (WICT 2014). 2014. IEEE.

[7] Porshnev, A., I. Redkin, and A. Shevchenko. Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis. In 2013 IEEE 13th International Conference on Data Mining Workshops. 2013. IEEE.

[8] Liu, Y., et al. Stock price movement prediction from financial news with deep learning and knowledge graph embedding. In Pacific rim knowledge acquisition workshop. 2018. Springer.

[9] Chen, W., et al., Leveraging social media news to predict stock index movement using RNN-boost. Data & Knowledge Engineering, 2018. 118: p. 14-24.

[10] Jin, Z., Y. Yang, and Y. Liu, Stock closing price prediction based on sentiment analysis and LSTM. Neural Computing and Applications, 2020. 32(13): p. 9713-9729.

[11] Shapiro, A.H., M. Sudhof, and D.J. Wilson, Measuring news sentiment. Journal of econometrics, 2020.

[12] Chowdhury, S.G., S. Routh, and S. Chakrabarti, News analytics and sentiment analysis to predict stock price trends. International Journal of

Computer Science and Information Technologies, 2014. 5(3): p. 3595-3604.

[13] Laban, P., J. Canny, and M.A. Hearst, What's The Latest? A Question-driven News Chatbot. arXiv preprint arXiv:2105.05392, 2021.

[14] Corporation, N., 2020 NAVER Annual Report. 2020: Online.

[15] Loper, E. and S. Bird, Nltk: The natural language toolkit. arXiv preprint cs/0205028, 2002.

[16] 1Devlin, J., et al., Bert: Pretraining of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[17] Liu, Y., et al., Roberta: A robustly optimized Bert pretraining approach. arXiv preprint arXiv:1907.11692, 2019.

[18] Conneau, A., et al., Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116, 2019.

[19] Yang, Z., et al., generalized autoregressive pretraining for language understanding; 2019. Preprint at https://arxiv.org/abs/1906.08237 Accessed June, 2021. 21.

[20] Malo, P., et al., Good debt or bad debt: Detecting semantic orientations in economic texts. Journal of the Association for Information Science and Technology, 2014. 65(4): p. 782-796.

[21] Howard, J. and S. Gugger, Fastai: a layered API for deep learning. Information, 2020. 11(2): p. 108.

[22] Wolf, T., et al. Transformers: State-of-the-Art Natural Language Processing. 2020. Online: Association for Computational Linguistics.

[23] Kingma, D.P. and J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[24] Ruder, S., An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747, 2016.

[25] Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev (2022). https://doi.org/10.1007/s10462-022-10144-1.