Sentiment Analysis of Thai Financial News

Vatcharaporn Esichaikul
School of Engineering and Technology
Asian Institute of Technology
Pathumthani 12120, Thailand
Tel: 662-5245713
vatchara@ait.ac.th

Chawisa Phumdontree
Information Technology Group
Bank of Thailand
Bangkok 10200, Thailand
Tel: 662-2835353
chawisap@bot.or.th

ABSTRACT
Due to the big data and FinTech influencers, the novel SentiFine framework was developed to facilitate the financial analysts or specialists who need to understand the financial and economic circumstances from the daily news. The objective of this framework is to analyze the sentiment of Thai financial daily news by integrating the fine-grained sentiment analysis technique with the deep neural network. Based on the proposed SentiFine framework, the prototype of the SentiFine web-based system was developed. With the main feature "View Daily News", the users can view the daily Thai financial news detail, the date and time of the news, the source of news, and its sentiment which is categorized into three tones (positive, neutral, and negative).

CCS Concepts
• **Information systems→Sentiment analysis**

Keywords
Sentiment analysis; deep neural network; FinTech; thai natural language processing

1. INTRODUCTION
Due to the current technological advancement, big data has been generated and widely spreaded in every second via the online channels [1]. It is undeniable that the big data influences everyone in this digital age, including the financial domain as well. The available financial information such as the online news can manifest the current trend of the stock and bond prices in the efficient markets [2].

This can be implied that the online financial news, which is a part of the big financial data, is the essential information for investors, bankers, financial analysts, financial experts, regulators, governors, CFOs, and everyone who is interesting in the current situation of the financial markets because of their different purposes such as the firm's management, the financial investment, the financial risk assessment, and the government regulation [3].

Simultaneously, the term "FinTech" has been recently denominated for the combination of finance and technology domains. This jargon has been stated as the solution which big data might be involved with its technologies for figuring out several financial problems [4] such as the peer-to-peer lending, the economic forecast, the data analytics for business strategic management, and the sentiment analysis on financial trends [5]. Since the concealed truth in the big financial data is more valuable, it is the excellent opportunity to explore an efficient FinTech solution.

Normally, financial analysts or financial specialists have the responsibilities to identify the internal and external factors of their organization, to develop financial models under the legal limitation, to design and review policies, and to publish the sufficient news. To handle, analyze, and synthesize along with the essential duties of the financial experts, these groups of people have to update themselves by regularly digesting the information from the financial news. Every morning, they have to read the notable news which may affect the work that they are responsible for.

Nevertheless, according to the existing applications in Thailand, some organizations may use the paid service of online news clipping like iQNewsClip [6]. The main purpose of this service is to automatically send an email in every morning with the link to news. The content in the daily email contains headline with its hyperlink to the news clips, provider names, and the position of news in the physical paper. After clicking on each link, a user will be redirected to the news in the PDF format. However, it is difficult to determine the crucial or extrinsic news from the only headline news. Financial experts have to click most of the news link if they are unsure whether it is the essential one to read or not. Thus, even now, they still have to spend their morning time with this tedious routine because they have to click through all links to news, skim, and scan all information.

The main purpose of this study is to propose the SentiFine Framework which integrates the fine-grained sentiment analysis and the deep neural network for developing the FinTech application that can analyze the big data from the online financial news.

2. THEORETICAL BACKGROUND AND RELATED WORK
2.1 Sentiment Analysis
Sentiment Analysis is the method of recognizing whether a content in the form of unstructured data can imply its polarity. The tone or polarity or sentiment of content can be in positive, neutral or negative. Nowadays, many areas use the sentiment analysis as decision-making tools such as marketing, politics, or even in the finance domain. For instance, from the study of [7], the S-Sense sentiment analysis application is used for the social media sense to check the viewpoint of Thai people for the general topic whether it is positive or negative. In the case of the financial domain, the sentiment analysis is used to make the stock prediction [8]. However, to interpret the sentiment of content, from the study of [9], the Natural Language Processing (NLP) is an important

ICSEB '18, December 18–20, 2018, Zhuhai, China

ACM ISBN 978-1-4503-6127-9/18/12…$15.00
DOI: https://doi.org/10.1145/3301761.3301773

technique that will be involved in the text classification and text segmentation.

There are two approaches for scoring the sentiment dataset which are coarse sentiment and fine-grained sentiment [10]. The coarse sentiment or binary sentiment is the simplistic classification of distinguished sentiments into positive or negative polar while the fine-grained sentiment analysis is more precise classification because its levels can be broken down into many sub-groups of emotion. For instance, in the study of [11], they categorized their dataset with fine-grained sentiment approach into five classes: very negative, negative, neutral, positive, and very positive while the study of [12], they categorized their dataset into three segments: positivity, negativity, and objectivity or neutrality.

The techniques for the sentiment analysis are composed of two approaches which are lexicon based and machine learning based techniques [13].

2.2 Related Work

There are several related works that apply the sentiment analysis with financial textual data in various aspects. Four most relevant studies are described here.

The first study [14] used data sources from Twitter to predict the positive sentiment for constructing a financial indicator. The study proposed a new indicator for predicting sentiment analysis in finance by using positive sentiment probability. The Granger causality which could represent the polarity of sentiment was used for determining a new indicator. Twitter feeds were used as a data source for forecasting the movements in stock closing prices, and the use of SVM (Support Vector Machine) to categorize tweets into positive, negative, and neutral. The outcome of this study found that the change in the values of positive sentiment probability could predict a similar movement of the stock prices.

The second study [15] used the sentiment analysis for providing the intraday stock recommendation system. The study researched an effectiveness intraday stock recommendation decisions system of augmenting market-based data with five different sets of textual news-based data. The concept of Bag of words was used as the text representation technique, and they also enhanced the sentiment analysis process with the calibrated sentiment scores developed by the linear regression technique. The outcome of this research was that the integration of market data and textual news could contribute to improving the modeling performance of an automated intraday stock recommendation system. Moreover, this study proved that more advanced textual data analysis could improve the accuracy of stock prediction.

The third study [16] introduced the fine-grained sentiment for analyzing financial text which is one of the important parts of our research. In this study, the objective was to propose a newly available labeled financial news dataset to the research community and a novel feature construction method using syntactic sentence patterns to more accurately predict financial news sentiment at fine-grained levels. The NLP technique was applied for text extraction. The sentiment analysis approach in the study was machine learning by applying with the SVM model and lexicon-based approach. As a result, the machine learning method using their novel features can more accurately predict financial news sentiment at fine-grained levels than the BOW lexicon methods. This paper suggested the new method for sentiment scoring which was more precise than the binary sentiment.

The last study [17] used the deep learning technique for the financial sentiment analysis on the financial news. The objective was to determine the influence of the investment from the various financial resources and to improve the accuracy of forecasting through deep learning. The input data source for this study came from four online news websites: NowNews, AppleDaily, LTN, and MoneyDJ. They also built their corpus called iMFinanceSD opinion words which were categorized by the type of sentiment (positive and negative). The feature used for the financial sentiment analysis consisted of 11 numeric features such as the total word number of news text, the number of new words, the total number of the positive words.

The sentiment analysis techniques that they used were lexicon-based and machine learning approaches. For the machine learning model, this paper used Deep Neural Network (DNN) for performing sentiment analysis and stock prediction. The outcome of the study provided a financial analytics method with deep learning in news sources for the stock price trend forecasts.

3. METHODOLOGY

The research methodology is comprised of two main steps: 1) SentiFine Framework Development and System Development.

3.1 SentiFine Framework Development

This study proposes a novel framework for analyzing the fine-grained sentiment from Thai Financial News. The proposed SentiFine Framework is shown in Figure 1.

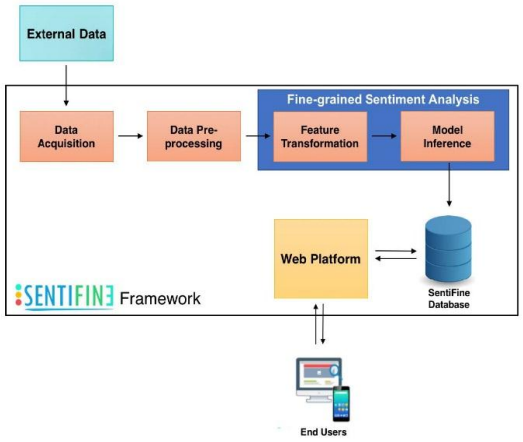

Figure 1: The proposed sentifine framework

The major components of the framework are consisted of as follows:

1. Data acquisition: for retrieving the external data source
2. Data pre-processing: for cleaning and preparing data
3. Feature transformation: for transforming text to word embedding which will be used for analyzing the fine-grained sentiment by using the deep neural network model
4. Model inference: for inferencing the headline news which is a part of fine-grained sentiment analysis by using the deep neural network model.

From Figure 2, the data acquisition process will automatically retrieve the daily Thai financial news from Thai news RSS links. The title, URL and other news information are stored in the SentiFine Database (Retrieved Financial News).

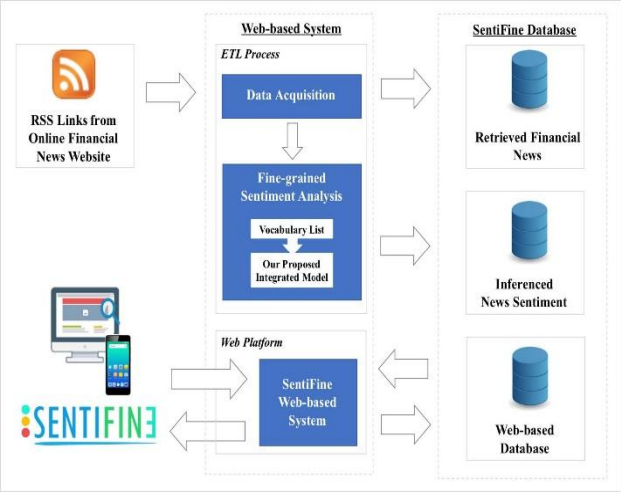

Figure 2. SentiFine environment.

3.2 System Development

System analysis is performed to gather all requirements from SentiFine's users. They can be divided into two groups: 1) senior analysts from the financial market division and 2) team executives from the financial payment policy division. Then functional requirements are determined. The main functions of the SentiFine web-based system are as follows:

a. View Daily News: This system allows users to view the daily Thai financial news detail, the date and time of the news, the source of news, and its sentiment which will be categorized into three tones (Positive, Neutral, and Negative). Moreover, this function should provide a feature that users can filter news from the news sources, news categories, time, and sentiments.

b. Export Daily News: This system allows users to export news from the system in the text format.

c. Sentiment Daily News Indicator: This system can summarize the daily polar of each news category by using the visualization indicator such as a gauge diagram.

For an interface, the user interface of the web page should make it easy to use in both platforms: a workstation and a mobile. The arrangement of interface's components in the display and report functions should be properly designed. In terms of performance, for all provided functions in SentiFine web-based system, the accuracy and the precision of news information and its retrieval time should be acceptable by users.

For system design, it covers physical design, logical, and architecture components of the SentiFine web-based system. The system architecture of SentiFine web-based system (see Figure 3) is composed of five parts: the data sources for our web-based system which is Thai News Websites, Back-end, Database, Front-end, and Client.

1. Thai News Websites: Starting from the input sources of our system. SentiFine system will retrieve Thai News Websites from ThaiPBS, PostToday, ThaiPR.NET, Manager, and BangkokBizNews by using the SentiFine application (ETL Process) at the back-end side.

2. Back-end: The back-end of SentiFine system is comprised of two parts: the Web Platform and ETL Process. Web Platform is composed of the main components for the website which are Model, View, and Template. ETL (Extract, Transform, Load) Process is another essential part of the system. Not only for retrieving news from the external website but it also involves other activities such as Data Cleansing, Feature Transformation, and Model Inference.

3. Database: After data has been retrieved, created, updated, or deleted by the back-end part, and it will be recorded in the unstructured database.

4. Client: The output of SentiFine web-based system will be produced in the responsive website form so that the client or computer/mobile with the web browser can use this application and CloudFlare will help by pointing to the SentiFine Server.

5. Front-end: In this front-end part, it includes HTML, CSS Bootstrap, D3.js, and JavaScript which is used for interacting with the user via the interfaces.

After the system analysis and design, the SentiFine system is implemented. The development covers two parts:

Data Model Formulation, and Web-based System Development. For data model formulation, after the dataset has been prepared, the next step is data modeling and model evaluation. In this part, the deep neural network models are used for our experiments. After the data modeling is completed, the best model is evaluated with accuracy and F1-score metrics [18].

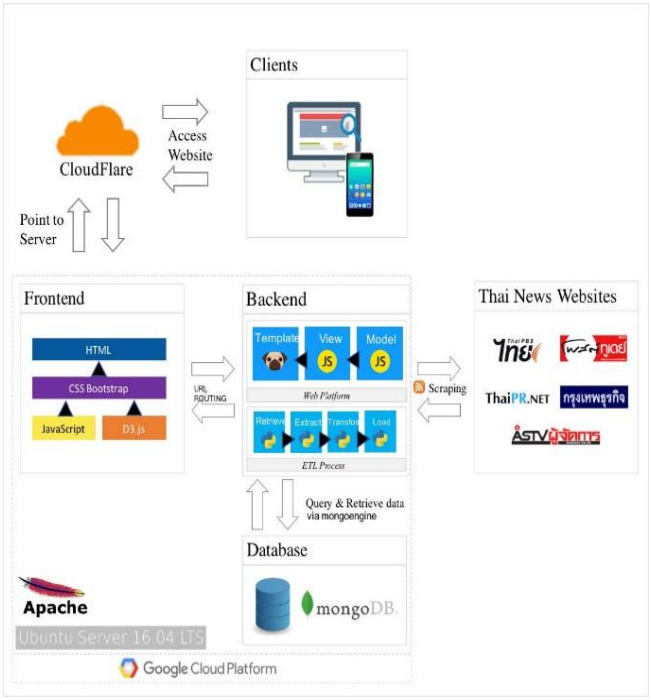

Figure 3. System architecture.

As a result from the experiments, Table 1 presents the F1 and accuracy results from top 5 models formulation. Model 2.3 CNN-Bidirectional GRU received the highest F1 score of 0.841065 and accuracy of 0.84233. Therefore, the model that we selected for our SentiFne Framework is 2.3 CNN-Bidirectional GRU which we renamed to United CNN Bidirectional GRU or UCBGRU.

Table 1. Results of the experiment

ExII.ID	Model	5-Fold Validation Accuracy	F1	Test Accuracy
2.1	CNN	0.839657937	0.834211	0.834848
2.2	Dynamic Seq2-CNN	0.830571887	0.826265	0.8279
2.3	CNN-Bidirectional GRU	0.838161411	**0.841065**	0.84233
2.4	Basic Dynamic CNN	0.830785676	0.832536	0.833244
2.5	CNN-Bidirectional LSTM	0.835275254	0.838125	0.839123

Then the web-based SentiFine system is implemented. For back-end implementation, the SentiFine back-end system was built on GCP by using Ubuntu 16.04 as the OS server, Apache v2.4.18 as the web server, and CloudFlare for automatically DNS configuration. For front-end implementation, after the back-end was built, then the last T(Template) was implemented by using pug which is the popular template engine for NodeJS. Some interfaces of SentiFine web-based system were shown in Figures 4-5.

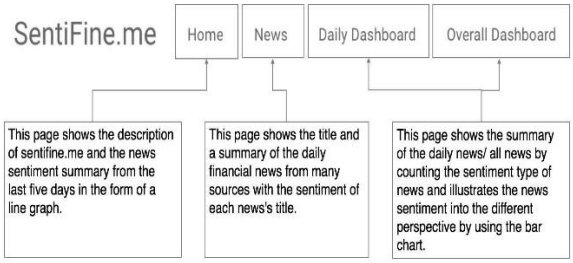

Figure 4. Navigation bar.

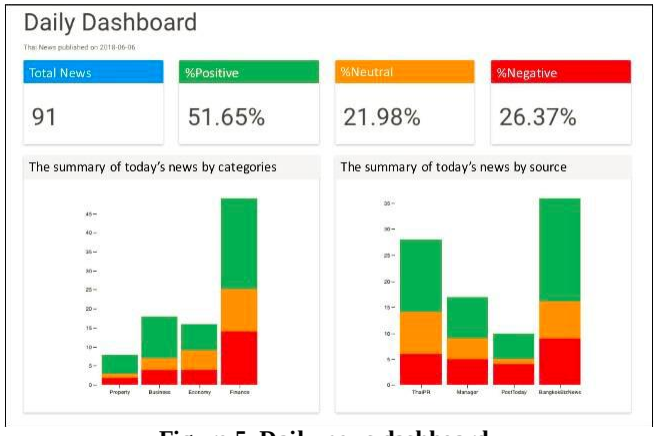

Figure 5. Daily news dashboard.

4. CONCLUSION

The objective of this study is to propose the SentiFine Framework for developing the FinTech application to analyze the sentiment of the financial news by using the fine-grained sentiment analysis and the deep neural network technique. To apply the framework, we also developed the prototype SentiFine web-based system along with the proposed framework.

Based on the literature review, there are several studies on the sentiment analysis and deep neural network, but the majority of them have rarely integrated with the news domain especially in Thai financial news. Thus, we proposed the framework with the novel processes and procedures by using Deepcut and pre-trained Thai2Vec resources with the deep neural network model.

The initial part of this study was to gather data for the further training, validating, testing, and integrating with the newly implemented system. For the system integration part, news data were retrieved from the open source only which included PostToday, BangkokBizNews, Manager, ThaiPR, ThaiPBS. In addition, for data gathering, we also retrieved data from Twitter account @ThaiValueInvest, labeled all 9,352 news into the fine-grained sentiment (positive, neutral, and negative) and performed the data pre-processing such as text cleansing. After getting the input data, it was the time for the data model formulation. Currently, there are many deep neural network models. This study conducted the experiments by training 17 deep neural networks with the different optimizers and pre-train word embeddings. Finally, the study found the best one with the F1 score of 82.4%. The explored model has the combination between CNN and Bidirectional GRU. Therefore, we named it as united CNN Bidirectional GRU or UCBGRU. The whole SentiFine's processes and procedures have been fulfilled by using our explored model - UCBGRU.

In order to apply the proposed framework, the prototype of SentiFine web-based system was implemented by using the up-to-date technology based on Python and Javascript language. The main functions of SentiFine system include: view daily news, export daily news, and sentiment daily news indicator. For future work, we plan to evaluate system performance and usability of system with real users.

5. REFERENCES

[1] Desjardins, J. (2017). What happens in an internet minute in 2017? Retrieved September 7, 2017 from https://www.weforum.org/agenda/2017/08/what-happens-in-an-internet-minute-in-2017.

[2] Clarke, J., Jandik, T., & Mandelker, G. (2001). The efficient markets hypothesis. Expert Financial Planning: Advice from Industry Leaders, pp.126-141.

[3] Ross, S. A., Westerfield, R., & Jaffe, J. F.(2010). Corporate finance. McGraw-Hill/Irwin.

[4] Numnonda, A. P. D. T.(2016). Big Data in FnTech. Retrieved September 17, 2017 from http://www.sec.or.th/TH/Documents/FinTech/FinTech 220659 03.pdf.

[5] Techsauce (2017). iqnewsclip. Retrieved September 13, 2017 from https://techsauce.co/report/thailand-tech-startup-ecosystem%E2%80%8B-q1-2017/

[6] InfoQuest (2017). iqnewsclip. Retrieved September 16, 2017 from http://www.infoquest.co.th/en/news-services-2/iqnewsclip/

[7] Haruechaiyasak, C., Kongthon, A., Palingoon, P., & Trakultaweekoon, K. (2013). S-sense: a sentiment analysis framework for social media sensing. In Sixth International Joint Conference on Natural Language Processing (p. 6).

[8] Li, X., Xie, H., Chen, L., Wang, J., & Deng, X.(2014). News impact on stock price return via sentiment analysis. Knowledge-Based Systems, 69, pp.14–23.

[9] Sriphaew, K., Hiroya, T., & Manabu, O. (2009). Sentiment Analysis for Thai Natural Language Processing. In Proceedings of the 2nd Thailand-Japan International Academic Conference, TJIA 2009, pp. 123–124.

[10] Kohtes, R. (2014). From valence to emotions: How coarse versus fine-grained online sentiment can predict real-world outcomes. Anchor Academic Publishing.

[11] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., et al. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1631–1642.

[12] Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In LREC, Vol. 10, pp. 2200–2204.

[13] Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., et al. (2017). Sentiment analysis using deep learning techniques: A review. International Journal of Advanced Computer Science and Applications (IJACSA), 8(6), pp.424–433.

[14] Smailovi´c, J., Gr˘car, M., Lavra˘c, N., & ˘Znidar˘ si˘c, M. (2013). Predictive sentiment analysis of tweets: A stock market application. In Human-computer interaction and knowledge discovery in complex, unstructured, big data, pp.77–88. Springer.

[15] Geva, T., & Zahavi, J. (2014). Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news. Decision Support Systems, 57, pp.212–223.

[16] Meyer, B., Bikdash, M., & Dai, X. (2017). Fine-grained financial news sentiment analysis. In SoutheastCon, 2017, pp. 1–8.

[17] Day, M.-Y., & Lee, C.-C. (2016). Deep learning for financial sentiment analysis on finance news providers. In Advances in social networks analysis and mining (ASONA), 2016 IEEE/ACM International Conference, pp. 1127–1134.

[18] Theeramunkong, T. (2012). Introduction to concepts and techniques in data mining and application to text mining. TPA Book Centre.