Sentiment Analysis on Thai Social Media Using Convolutional Neural Networks and Long Short-Term Memory

Chalisa Jitboonyapinit[1], Paralee Maneerat[2], and Nivet Chirawichitchai[3]

[1,2]School of Information Technology, Sripatum University, Bangkok, Thailand
[3]Faculty of Engineering and Technology, Panyapiwat Institute of Management, Nonthaburi, Thailand
E-mail: chalisajitboon@gmail.com, paralee.ma@spu.ac.th, nivetchi@pim.ac.th

Received: June 23, 2022/ Revised: July 5, 2022/ Accepted: July 21, 2022

***Abstract*—The objective of this research purposes a sentiment analysis of Thai social media using deep learning techniques consisting of a convolutional neural network, long short-term memory, and a gated recurrent unit. This research was used to test the algorithm with a wongnai product and service dataset and measured performance with accuracy. The experiment of this research found convolutional neural networks with long short-term memory outperform convolutional neural networks, long short-term memory, and gated recurrent units in classification accuracy, with an accuracy of 85.0%, followed by long short-term memory accuracy of 83.7%, convolutional neural network accuracy of 77.0% and finally gated recurrent unit an accuracy of 65.4% respectively. Therefore, the hybrid working model of a Convolutional Neural Network with long short-term memory is most suitable and effective for Thai sentiment analysis.**

***Index Terms*— Convolutional Neural Network, Gated Recurrent Unit, Long Short-Term Memory**

I. Introduction

Sentiment analysis is the process of analyzing customer opinion utilizing natural language processing, text analysis, and statistics. The finest companies are aware of their consumers' feelings—what they're saying, how they're telling it, and what they mean. Tweets, comments, reviews, and other sites where people mention your brand might reveal customer sentiment. Sentiment Analysis is the domain of using software to analyze these feelings, and it's a must-know for developers and business executives in today's workplace. Advances in deep learning, like many other domains, have pushed sentiment analysis to the front of cutting-edge algorithms. To extract and categorize the sentiment of words into positive, negative, or neutral categories, we now use natural language processing, statistics, and text analysis. For brand monitoring, sentiment analysis is used. One of the most well-known applications of sentiment analysis is to obtain a complete 360-degree perspective of how your brand, product, or company is perceived by customers and stakeholders. Product reviews and social media, for example, are widely available media that can give crucial insights into what your organization is doing properly or poorly. Sentiment analysis can also be used to assess the impact of a new product, ad campaign, or a consumer's reaction to recent company news on social media. Customer service agents frequently use a sentiment or intent analysis to automatically categorize incoming user emails into "urgent" or "not urgent" categories depending on the email's sentiment, proactively detecting unhappy users. The agent then prioritizes fixing the users with the most pressing issues first. Understanding the sentiment and intent of a specific case becomes increasingly critical as customer care becomes increasingly automated through machine learning. Sentiment analysis is used in market research and business intelligence to identify the subjective reasons why customers respond or do not respond to something (for example, why do consumers buy a product?). What are their thoughts on the user interface? Did the level of customer service fulfill their expectations?). Sentiment analysis can be used to examine trends, ideological bias, and opinions, assess reactions, and more in the fields of political science, sociology, and psychology. Sentiment analysis (also known as opinion mining) is a natural language processing (NLP) technique for determining the positive, negative, or neutral of data. Sentiment analysis is frequently used on textual data to assist organizations in tracking brand and product sentiment in consumer feedback and better understanding customer demands [1]-[3]. For the reasons mentioned above, the objective of this research was to create a research study to develop a sentiment analysis model on Thai social media using deep learning techniques to make an automatic model for classifying positive and negative customer opinions. We used a variety of preprocessing methods in conjunction with various deep-learning algorithms to construct a model for evaluation purposes. We compare the models' classification accuracy as well as the amount of time they need for training and building the model. The rest of the paper is organized as follows. Section II describes the literature review. Section III describes the research framework. Section IV describes the research methods. Section V describes the experiments and results. Finally, Section VI conclusions.

II. Literature Review

We discovered that most researchers constructed their analysis models based on data from the internet, particularly from an online platform, after doing a literature study in a specific area of sentiment analysis over comments submitted in the product review system. The majority of recent sentiment analysis model development efforts have focused on evaluating text data from online business platforms such as Amazon product reviews, Twitter, IMDB movie reviews, and Yelp trip suggestions. These data are continuously generated by users, and the quantity of the data is rapidly growing as a result of the enormous number of users that provide feedback and comments via the internet on a daily basis. These data are collected by platform owners and used to assess insight information in order to better support new product design or improve the product quality that mostly fits their customers in positive ways. Because the data is in a text format that the computer cannot understand, the analyst must first preprocess the data before proceeding to the major steps of the analysis process. The most common preprocessing technique used by analyzers is a bag of words. The analyzer then uses the preprocess data to create a machine learning-based text classification model [4]-[8].

Deep learning is a popular technique for generating models right now since it produces high-accuracy classification models without the requirement for feature selection. The use of word embedding to preprocess data and then developing a model based on a convolutional neural network (CNN), the application of transfer learning strategy based on the pre-train CNN to reduce additional model training time, and the use of word embedding to preprocess data and then applying sequence analysis algorithm including recurrent neural network (RNN) and long short-term memory (LSTM) for sentiment analysis [9],[10].

From the high accuracy of the deep learning method applied for sentiment analysis as reported in the literature, we are thus interested in empirically studying the performance of deep learning. We use several machine learning techniques as a benchmark to compare against deep learning performance. The framework of our comparative study is presented in the next section.

III. Research Framework

Our processes, as illustrated in Fig. 1, are a framework for a comparative investigation of preprocessing and learning strategies that produce the best results for sentiment analysis of product comments and reviews. To begin, we gather data for text cleaning and preprocessing, which includes converting texts to lowercase, removing stop words and punctuation, and removing prefix-suffix from terms, among other natural language processing stages. Then, applying text transformation and deep learning approaches, we analyze and construct models based on a bag of words and word embedding. Finally, we evaluate the results in terms of model accuracy, precision-recall, and F1 score.

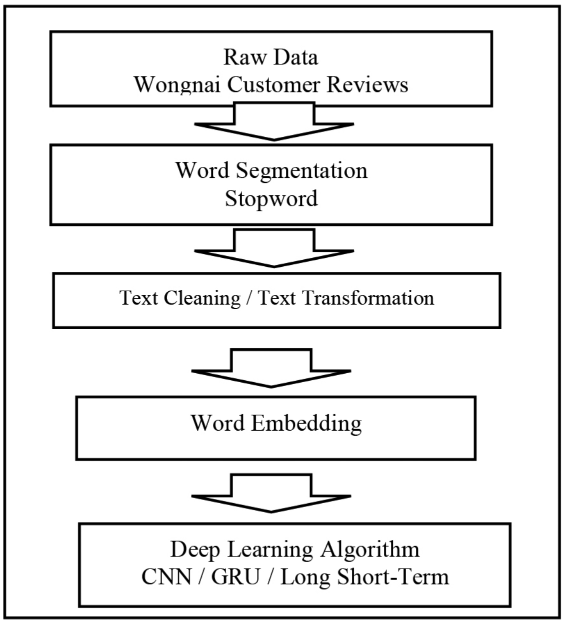

Fig. 1. Research framework

IV. Research Methods

A. Text Preprocessing

The first step in text classification is to transform documents, which typically are strings of characters, into a representation suitable for the learning algorithm and the classification task. For the Thai language, the main task of text processing is the segmentation of texts into word tokens. Thai texts are naturally unsegmented, i.e., words are written continuously without the use of word delimiters. Due to this distinct characteristic, preparing a feature set for Thai text categorization is more challenging than in Latin-based languages such as English, French, and Spanish. In Latin-based languages, a text string can easily be tokenized into terms by observing the word delimiting characters such as spaces, semicolons, commas, quotes, and periods. To prepare a feature set for the Thai social media corpus, we must first apply a word segmentation algorithm to tokenize text strings into a series of terms. Once a set of extracted words are obtained from the training news corpus, the removal of HTML tags, removal of stop-words, and then word stemming. The stop-words are frequent words that carry no information (i.e. pronouns, prepositions, conjunctions, etc.). By word stemming we mean the process of suffix removal to generate word stems. This is done to group words that have the same conceptual meaning, such as walk, walker, walked, and walking [1]-[3].

B. Word Embedding

The word embedding techniques are used to represent words mathematically. One Hot Encoding, GloVe, Word2Vec, and FastText, are frequently used word embedding methods. One of these techniques (in some cases several) is TM because of less parameter and more simple preferred and used according to the status, size, and purpose of processing the data. Word embedding is a pre-processing strategy for converting written content into a format that the deep learning system can understand. This approach turns words into vectors for the convenience of calculation when determining word similarity. Fig. 2 shows an example of vectors. Because it can evaluate sequence data, this word embedding method is ideal for usage with recurrent neural network (RNN) algorithms including Long Short-Term Memory and Gated Recurrent Units [9],[10].

Word2Vec is a state-of-the-art algorithm to generate a fixed-length distributed vector representation of all the words in the huge corpus. The effectiveness of Word2Vec is due to two reasons — One is the use of fixed-size vectors which means the vector size does not depend on the number of unique words in the corpus. Second, incorporates semantic information in the vector representations. Word2Vec vectors are highly efficient at grouping similar words together. The algorithm can make strong estimates based on the position of the word in the corpus. For example, “Kid” and “Child” are similar and hence their vector representation will be very similar [11].

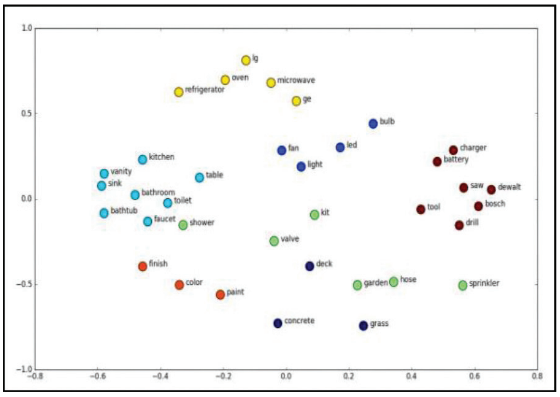

Fig. 2. Word2Vec

C. Gated Recurrent Unit

Gated Recurrent Units (GRUs) [12] are a gating mechanism in recurrent neural networks. GRU is a deep learning algorithm that reduces the complexity of LSTM networks by lowering the number of gates in each cell. As a result of the reduction, the number of parameters that must be computed in the network is reduced. Hence, the speed-up of computation time. GRU cell contains input, output, update, and reset gates. The update gate is for controlling change in the hidden state, while the reset gate is for resetting the value of the hidden state, as shown in Fig. 7. This cell structure makes GRU work faster than the LS structure.

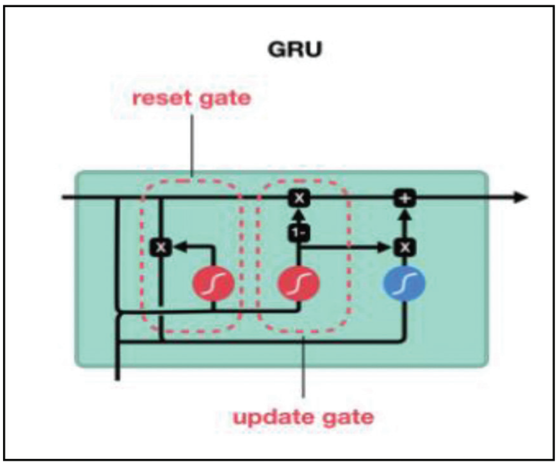

Fig. 3. Gated Recurrent Unit

D. Long Short-Term Memory

Long Short-Term Memory (LSTM) [13],[14],[17] is an artificial neural network used in the fields of artificial intelligence and deep learning. LSTM is the deep learning algorithm that has been developed from RNN for solving the gradient-descent variants problem and searching for long-term dependencies in the dataset. LSTM contains a cell as a subunit with a fundamental structure shown in Fig. 6. The cell contains a subsystem called an input gate to get input data, forget gate for weighing the significance of the memory cell state from the previously computed state, memory-cell state gate to compute a new memory cell state, and output gate for computing a new hidden state.

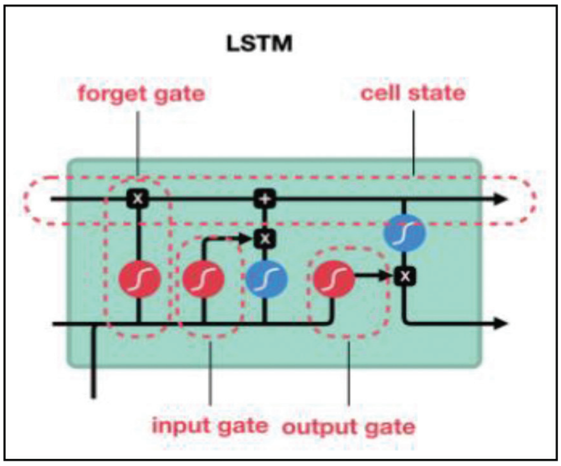

Fig. 4. Long Short-Term Memory

E. Convolutional Neural Network

A Convolutional Neural Network [10] (CNN) is a Deep Learning algorithm that can take in an input image, give importance (learnable weights and biases) to distinct aspects/objects in the image, and differentiate one from the other. CNN requires substantially less pre-processing than other classification techniques. While crude approach filters are hand-engineered, CNN can learn these filters with enough training.

Convolutional neural networks are a sort of feed-forward neural network that was first used in computer vision, recommender systems, and natural language processing. It is a deep neural network design made up of convolutional and pooling or subsampling layers that feed input to a fully-connected classification layer. Convolution layers extract features by filtering their inputs; the outputs of many filters can be merged. Pooling or subsampling layers reduce feature resolution, which can improve CNN robustness to noise and distortion. Fully connected layers perform classification tasks. An example of a CNN architecture can be seen in Fig. 5. The input data was preprocessed to reshape it for the embedding matrix. The figure shows an input embedding matrix processed by four convolution layers and two max pooling layers.

The first two convolution layers have 64 and 32 filters, which are used to train different features; these are followed by a max pooling layer, which is used to reduce the complexity of the output and prevent the overfitting of the data. The third and fourth convolution layers have 16 and 8 filters, respectively, which are also followed by a max pooling layer. The final layer is a fully connected layer that will reduce the vector of height 8 to an output vector of one, given that there are two classes to be predicted (Positive, Negative).

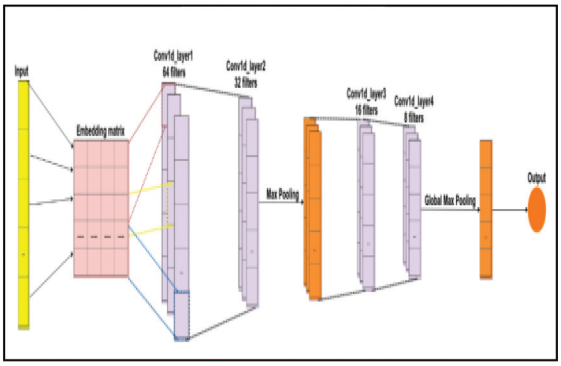

Fig. 5. Convolutional Neural Network

F. Dataset

In the research, we analyze and develop the models, we use product review data from the website wongnai.com. This dataset contains review texts in Thai as shown in some examples in Table I. This dataset has 5,000 records and 2 groups of reviews that contain 2,500 positive reviews that are labeled as positive and 2,500 negative reviews that are labeled as negative as shown on the rating column. We split data into two subsets as follows: 80% of them are the training set and the remaining 20% are the test set.

Fig. 6. Wongnai dataset

V. Experiments and Results

We performed experiments using a collection of product reviews obtained from the wongnai website. There are two categories: positive and negative. A total of 5,000 comments were selected and trained. We used TensorFlow - Keras and a deep learning tool, to perform the experiments. For deep learning methods, we use Gensim as a pre-trained model for word embedding to use with LSTM and GRU. The vector in the word embedding step has 128 dimensions and the sequence contains 1,818 values. The architectures of LSTM are Bidirectional-LSTM = 256, activation = relu, fixed Dense = 128-64, optimizers = Adam EPOCHS = 50, Batchsize = 256, Learning Rate = 0.001 Dense output = 2, The architectures of GRU are GRU = 256, activation = relu, fixed Dense = 128-64, optimizers = Adam EPOCHS = 50, Batchsize = 256, Learning Rate = 0.001 Dense output = 2, The architectures of CNN are Conv1D filters = 32, kernel_size = 8, activation = relu, Dropout = 0.5, MaxPooling1D pool_size = 2, fixed Dense = 128, optimizers = Adam, Batchsize = 256, Learning Rate = 0.001, LSTM GRU and CNN are the same in that each having 1 layer. Each layer has 1,818 units with a dropout layer to prevent overfitting by defining a 0.2 dropout rate. Finally, we define a fully connected or dense layer as the last layer of deep learning architecture to perform the classification task by containing 1 node of the softmax function as an activation function. In the training process, this network has been configured for 50 epochs.

Classification effectiveness is usually measured using precision and recall. Precision is the proportion of true positive examples labeled positive by the system that was truly positive and recall is the proportion of true positive examples that were labeled positive by the system. The F-measure function which combines precision and recall is computed as [15]

$$F-measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

We tested all algorithms using the validation test set of 20%. The results in terms of precision, recall, and F-measure is the averaged values calculated across all cross-validation experiments. The experimental results of this word embedding scheme with respect to accuracy precision-recall and F-measure on the Thai product review corpus in combination with four deep learning are reported in Table I and Fig. 7 to Fig. 10.

TABLE I
ALGORITHM PERFORMANCE COMPARISONS

List	Accuracy	Precision	Recall	F-Measure
CNN	0.770	0.772	0.768	0.770
GRU	0.654	0.658	0.654	0.653
LSTM	0.837	0.839	0.837	0.837
CNN+LSTM	**0.850**	**0.852**	**0.848**	**0.850**

The experiment of this research found that convolutional neural networks with long short-term memory provided the most effective overall methods, with a classification accuracy of 85.0%, precision of 85.2%, recall 84.8%, and F-measure 85.0%, followed by Long Short-Term Memory an accuracy of 83.7%, precision 83.9%, recall 83.7%, and F-measure 83.7%, Convolutional Neural Network an accuracy of 77.0%, precision 77.2%, recall 76.8%, and F-measure 77.0%, Finally Gated Recurrent Unit an accuracy of 65.4%, precision 77.2%, recall 76.8%, and F-measure 77.0% respectively. The results obtained from the experiments in this research are consistent with the research of S. N. Murthy et al. [16], experiments with the IMDB movie review dataset and the Amazon product review dataset, found that CNN and LSTM were suitable and effective for sentiment analysis. This research is also consistent with the research of Kurniasari's [14] experiments with social media data in the Indonesian language found that the CNN and LSTM algorithm gives the same results of the model evaluation with good accuracy and good performance.

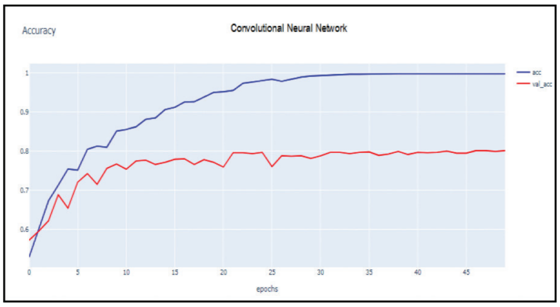

Fig. 7. Results of accuracy testing on Convolutional Neural Network algorithm

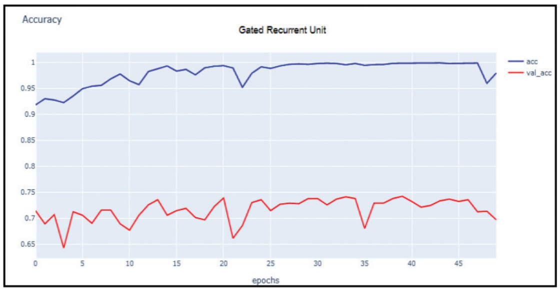

Fig. 8. Results of accuracy testing on Gated Recurrent Unit algorithm

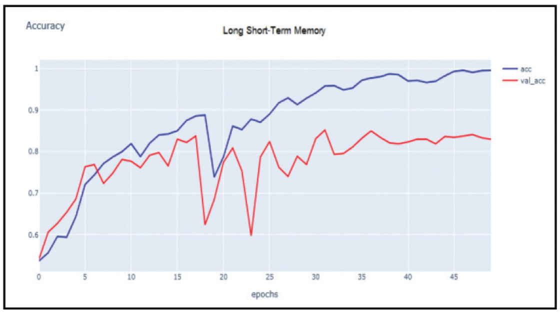

Fig. 9. Results of accuracy testing on the Long Short-Term Memory algorithm

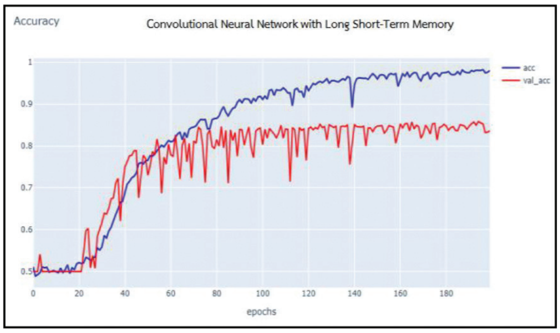

Fig. 10. Results of accuracy testing on Convolutional Neural Network with Long Short-Term Memory algorithm

VI. Conclusion

The experimental of this research purposes sentiment analysis on Thai social media using deep learning techniques consisting of the convolutional neural network, long short-term memory, and gated recurrent unit. This research was used to test the algorithm with the wongnai product and service dataset and measured performance with accuracy. This experiment of this research found convolutional neural networks with long short-term memory outperform convolutional neural networks, long short-term memory, and gated recurrent units in classification accuracy, with an accuracy of 85.0%, followed by long short-term memory accuracy of 83.7%, convolutional neural network accuracy of 77.0% and finally gated recurrent unit an accuracy of 65.4% respectively. Therefore, the hybrid model working with are convolutional neural network with long short-term memory is most suitable and effective for Thai sentiment analysis.

References

[1] D. Sharma, "Sentiment Analysis Techniques for Social Media Data: A Review," in *Proc. ICTSCI.*, 2019, pp.75-90.

[2] R. Ahmad and Y. Shaikh, "Opinion Mining and Sentiment Analysis for Classification of Opinions on Social Networking Sites Using Machine Learning Algorithms: Systematic Literature Review," *International Journal of Advanced Research in Computer and Communication Engineering*, vol.10, no. 5, May. 2011.

[3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," in *Proc. Conf. Empirical Methods in Natural Language Processing*, 2002.

[4] K. Floyd, R. Freling, S. Alhoqail et al., "How Online Product Reviews Affect Retail Sales: A Meta-Analysis," *Journal of Retailing*, vol. 90, no. 2, pp. 217-232, Jun. 2014.

[5] S. Poria, E. Cambria, D. Hazarika et al., "A Deeper Look into Sarcastic Tweets Using Deep Convolutional Neural Networks," in *Proc. Int. Conf. 26th on Computational Linguistics*, 2016, pp. 1601-1610.

[6] R. Taemung and N. Chirawichitchai. "Thai Sentiment Analysis of Product Review Online Using Support Vector Machine," *Journal of Engineering Siam University*, vol. 18, no 34, pp. 1-12, Jun. 2017.

[7] N. Chirawichitchai. "Developing Term Weighting Scheme Based on Term Occurrence Ratio for Sentiment Analysis," *Information Science and Applications*, vol. 339, pp. 737-744, Feb. 2015.

[8] N. Chirawichitchai, "Emotion Classification of Thai Text-Based Using Term Weighting and Machine Learning Techniques," in *Proc. Conf. 11th JCSSE.*, 2014, pp. 91-96.

[9] V. Q. Nguyen. "Real-Time Event Detection Using Recurrent Neural Network in Social Sensors," *International Journal of Distributed Sensor Networks*, vol. 15, no. 6, pp. 1-15, Jun. 2019.

[10] N. Kalchbrenner, E. Grefenstette, and P. Blunsom. "A Convolutional Neural Network for Modelling Sentences," in *Proc. ACL.*, 2014, pp. 655-665.

[11] S. Li and B. Gong. "Word Embedding and Text Lassification Based on Deep Learning Methods," in *Proc. CSCNS2020,* 2021, pp.1-5.

[12] J. Chung, C. Gulcehre, K. Cho et al., "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," in *Proc. Int. Conf. NeurIPS.,* 2014. pp. 1-9.

[13] P. Srivastava. (2019, Dec. 31). *Essentials of Deep Learning: Introduction to Long Short-Term Memory*. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm

[14] L. Kurniasari, "Sentiment Analysis Using Recurrent Neural Network-LSTM in Bahasa Indonesia," *Journal of Engineering Science and Technology*, vol. 15, no. 5, pp. 3242-3256, Oct. 2020.

[15] N. Chirawichitchai, "Sentiment Classification by a Hybrid Method of Greedy Search and Multinomial Naïve Bayes Algorithm," in *Proc. ICT&KE.,* 2013, pp. 1-4.

[16] S. N. Murthy, S. R. Allu, B. Andhavarapu et al. "Text-Based Sentiment Analysis Using LSTM," *International Journal of Engineering Research & Technology*, vol. 9, no. 5, pp. 299-303, May. 2020.

[17] X. Shi, Z. Chen, H. Wang et al., "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," in *Proc. Int. Conf. NeurIPS.,* 2015, pp. 802-810.

Chalisa Jitboonyapinit received her B.Sc. and M.Sc. in Information Technology, Dhurakij Pundit University, Thailand. Currently, she is a Ph.D. candidate in the Doctor of Philosophy Program in Information Technology at Sripatum University.

Paralee Maneerat received her B.Sc in Computer Science from Rangsit University, M.Sc. in Information Technology, Mongkut's Institute of Technology Ladkrabang, Thailand, and Ph.D. degree in Technopreneurship and Innovation Management from Chulalongkorn University. She currently has the rank of director, of the Master of Information Technology Program, Faculty of Information Technology, Sripatum University.

Nivet Chirawichitchai received his B.B.A. in Industrial Management and B.A. in Mass Communication from Ramkham haeng University, M.Sc. in Information Technology, and Ph.D. degree in Information Technology from King Mongkut's Institute of Technology North Bangkok, Thailand. He currently has the rank of director, Master of Engineering Program in Engineering and Technology, Faculty of Engineering and Technology, Panyapiwat Institute of Management, Thailand.