Sentiment analysis of messages on Twitter related to COVID-19 using deep learning approach

Miss Chotika Imvimol

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2021
Copyright of Chulalongkorn University

การจำแนกอารมณ์จากข้อความบน twitter ที่เกี่ยวกับสถานการณ์การติดเชื้อโควิด-19 โดยวิธีการเรียนรู้เชิงลึก

น.ส.โชติกา อิ่มวิมล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2564

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title	Sentiment analysis of messages on Twitter related to COVID-19 using deep learning approach
By	Miss Chotika Imvimol
Field of Study	Computer Science
Thesis Advisor	Professor PRABHAS CHONGSTITVATANA, Ph.D.

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

.... Dean of the FACULTY OF ENGINEERING
(Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

THESIS COMMITTEE

.... Chairman
(Assistant Professor SUKREE SINTHUPINYO, Ph.D.)

.... Thesis Advisor
(Professor PRABHAS CHONGSTITVATANA, Ph.D.)

.... External Examiner
(Assistant Professor Tanasanee Phienthrakul, Ph.D.)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

โชติกา อิ่มวิมล : การจำแนกอารมณ์จากข้อความบน twitter ที่เกี่ยวกับสถานการณ์การติดเชื้อโควิด-19 โดยวิธีการเรียนรู้เชิงลึก. (Sentiment analysis of messages on Twitter related to COVID-19 using deep learning approach) อ.ที่ปรึกษาหลัก : ศ. ดร.ประภาส จงสถิตย์วัฒนา

สถานการณ์การระบาดของไวรัสโควิด-19 ที่ขยายวงกว้าง เป็นปัญหาสำคัญและจำเป็นต้องให้ความสนใจอย่างมาก เนื่องจากการระบาดนี้กระทบต่อเปลี่ยนแปลงวิถีชีวิตในหลายด้านเป็นเพื่อลดโอกาสการติดเชื้อจากไวรัสในระยะเวลาที่มีการระบาดนี้คนแสดงอารมณ์ในแบบต่างๆ นานา โดยในช่วงเวลาปัจจุบันคนจะแบ่งปันอารมณ์และความคิดต่างๆ บนโซเชียลมีเดียด้วย โดยเฉพาะบน Twitter ข้อความที่เกี่ยวข้องกับไวรัสโควิด-19 อาจจะสามารถบอกถึงอารมณ์ทางสังคมเกี่ยวกับเรื่องนี้ได้ การศึกษานี้ดำเนินการวิเคราะห์ข้อความเชิงอารมณ์ และข้อมูลย้อนหลังที่เกี่ยวกับไวรัสโควิด-19 และการสร้างแบบจำลองเพื่อจำแนกความรู้สึกในเชิงบวกและเชิงลบ 6 ประเภท ได้แก่ ความโกรธ ความขยะแขยง ความกลัว ความเศร้า ความยินดี และความประหลาดใจ ศึกษาข้อความที่ไม่ซ้ำกันทั้งหมด 120,642 ตัวอย่าง ระหว่างวันที่ 1 มกราคม 2020 ถึง 30 มิถุนายน 2021 โดยเปรียบเทียบประสิทธิภาพของโมเดลโครงข่ายประสาทเทียม 5 รุ่น ได้แก่ เพอร์เซปตรอนแบบหลายชั้น, RNN, LSTM, LSTM แบบสองทิศทาง และ GRU ผลการทดลองด้วยการวัดตัววัดประสิทธิภาพการทดลองหลายตัว ได้แก่ precision, recall, f1-score และ accuracy พบว่าแบบจำลอง LSTM พยากรณ์ผลได้ดีที่สุดบน precision เท่ากับ 77.7% และพบว่าแบบจำลอง LSTM แบบสองทิศทางได้รับคะแนนสูงสุดในวัดประสิทธิภาพการทดลองบน recall เท่ากับ 79%, f1-score เท่ากับ 78% และ accuracy เท่ากับ 79% ซึ่งแบบจำลองที่ได้จากการศึกษานี้สามารถเป็นประโยชน์ในการติดตามการเคลื่อนไหวของอารมณ์เชิงลบได้ นอกจากนี้งานศึกษายังให้ข้อมูลเชิงลึกจากการวิเคราะห์ข้อความเชิงอารมณ์และข้อมูลรายงานประวัติผู้ติดเชื้อ และข้อมูลการฉีดวัคซีน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่อนิสิต	....
ปีการศึกษา	2564	ลายมือชื่อ อ.ที่ปรึกษาหลัก	....

6272029521 : MAJOR COMPUTER SCIENCE
KEYWORD: Sentiment Classification Deep Learning COVID-19 Twitter

Chotika Imvimol : Sentiment analysis of messages on Twitter related to COVID-19 using deep learning approach. Advisor: Prof. PRABHAS CHONGSTITVATANA, Ph.D.

The widespread situation of the Coronavirus-19 (COVID-19) pandemic is a tangible and pressing concern. Many changes in terms of lifestyle are necessary to reduce the chance of infection. While citizens have gone through different emotions, they share their thoughts and interactions on social media, especially on Twitter. COVID-19 related messages can imply social emotion. This study performs sentiment analysis on tweets and annotated them into six classes of positive and negative feelings consisting of anger, disgust, fear, sadness, joy, and surprise. We analyzed both textual information and historical data. We collected 120,642 unique tweets datasets between 1 January 2020 and 30 June 2021. We compared the performance of five neural network models which are multi-layer perceptron, RNN, LSTM, Bidirectional LSTM, and GRU with several metrics consisting of accuracy, F1 score, precision, and recall. The results show that LSTM perform the best on precision with 77.7% while Bidirectional LSTM model achieved the highest score on metrics with 79% on recall, 78% on F1-score and 79% on accuracy. These models could be used to monitor the movement of negative emotions. In addition, we provide interesting insights from sentiment analysis with tweet data and historical reports of infected cases, and vaccination data.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Field of Study:	Computer Science	Student's Signature ....
Academic Year:	2021	Advisor's Signature ....

ACKNOWLEDGEMENTS

I need to thank my advisor, Professor Prabhas Chongstitvatana, for helping me to study for my thesis in the right way and utilizing deep learning knowledge to solve real-life problems. I really appreciate his hard work to help me, especially when we finish completing all the research. I appreciate Associate Professor Sukree Sinthupinyo as chairperson of thesis examination committee and Associate Professor Tanasanee Phienthrakul as a member of thesis examination committee who gave the great guidelines advice and recommendations to do my thesis. Moreover, I would like to thank my parents and friends who had been through thick and thin. They were super supportive and encouraged me until finishing. Lastly, I really appreciate myself for overcoming every problem and getting through this tough time.

Chotika Imvimol

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

TABLE OF CONTENTS

Page

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

LIST OF TABLES

Page

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

LIST OF FIGURES

Page

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter I
INTRODUCTION

1.1 Overview

Coronavirus Infection Disease 2019 (COVID-19) has been spreading around the world. This virus first appeared in December 2019 and the World Health Organization (WHO) reported it on 11 March 2020 in Wuhan, China and after that, the number of patient cases burst up in a short period in the local area. A month later, it spread abroad and the situation became worse with the rise of new cases. People in many countries around the world suffer from it, especially the USA, India, Brazil, and UK because they have a much higher number of patients.

In Thailand, the first found COVID-19 cases were found on 12 January 2020, after that the situation continually is getting worse because of the limit of the capacity of healthcare services to cover active cases. There have been significant increases in the number of Thai patients since the end of the first quarter of 2021 and became a crisis in August 2021 when the month of the highest rise of new cases was reported. After that, the situation in Thailand is getting better than in the middle of the year.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

This virus can easily infect people and it has effects from mild symptoms to severe. There are several symptoms from this virus, for instance, fever, cough, loss of taste or smell, sore throat, aches and pains, diarrhea, difficulty breathing, loss of mobility or speech even death. People need to be aware of these symptoms and take precautions to reduce the possibility of infection and need to be aware of it and live with it until the situation improves. The situation frequently affects people in bad ways, especially emotion. People use social media to express, and share thoughts with other people including the COVID-19 topic. This could imply that emotions are related to the situation.

Nowadays, there are many social media platforms for people, for example, Facebook, Twitter, Instagram, WhatsApp, Telegram and Line. The difference in the purposes of the platform makes people utilizing their functions to gain benefits as people wanted. Twitter is a popular platform among others and it is a social media platform where people can connect and communicate with short messages called tweets. The posting can be of various types which are 280 characters or less in text, link, picture, and a short video. Users can react to the posting by comment, like, retweet, and share with friends or another platform. Users can add hashtag character (#) into the context in the post to follow the topic easier and if many people post a lot about the topic in the same hashtag, it will show up on the top chart to inform that people are interested in this topic at that time.

Many Thai Twitter users share and express their opinions on many topics, including COVID-19 in Thailand. They share about infection situations, their emotion-related to COVID-19, and other stories. Those messages can imply their feeling with the situation. The feeling is described by word, phrase, sentence, collocations, context, tone, slang, and interjection. Natural emotions of humans can be divided into 6 categories which are anger, disgust, fear, sadness, joy, and surprise. There are many words that can express the six groups of feelings. For instance, Anger feeling is expressed by words which are furious, irate, incensed, and mad. Feeling of disgust is expressed by words which are queasy, weary, and nauseated. Fear is expressed by words which are scared, terror, dread, and fight. Sadness is expressed by words which are grief, heartache, heartbreak, mourning, and sorrow. Joy is expressed

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

by words which are delight, gladness, glee, charm, cheer, satisfaction, and happiness. Surprise is expressed by words which are shocked, amaze, stun, miracle, and astound.

Studied about classifying social media data which quickly change depending on thought, emotion and current situation. If people talk a lot about a specific topic, the topic can easily imply a real time situation. Analyzing messages can effectively forecast and predict behavior of users while facing pandemics in different areas, especially the spreading situation in Thailand which is totally different from foreign countries. Currently, there is a lot of information from tweet messages related to the pandemic. This studied aim to classify emotional messages during the spreading that people facing with the fluctuate in number of patients, died and relieve and vaccination people using deep learning to help to be aware of the hard time, decide to create the suitable policies to distract people from negative feeling that it often to hurt mentally and physically. These problems lead to the emotions of society turning to bad and it takes longer time to overcome the pandemic.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.2 Thesis Questions

1. Do emotional messages from tweets relate to infection situations?
2. Which neural network models perform the best on this dataset?

1.3 Thesis Objectives

Aim of this study was to use deep learning techniques with Thai language text messages from Twitter to monitor the emotions of people during COVID-19 infection situation in Thailand. Adapting machine learning techniques to get a suitable classification model with a specific domain. Proving the relation between emotion of people and the changing in real time situation provides useful information to relieve policy makers and others who need it.

1.4 Research Scope

Scope of study is utilizing deep learning techniques to classify only Thai language emotion messages related to the COVID-19 pandemic and finding trends and changes in emotion situations. This study classifies data into six categories which are anger, disgust, fear, sadness, joy, and surprise. Data is collected for analysis during 1 January 2020 until 30 June 2021.

1.5 Benefits

Result of studying classification emotion messages is getting the effective deep learning techniques model to classify others dataset in this specific domain. Proving about the official data that reported about COVID-19 situation is related to the emotions of people. Moreover, providing useful information and insight from data analysis and visualization.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

1.6 Research Methods

1. Studied method to annotate emotion dataset
2. Studied techniques and tools to classify Thai language contextual data
3. Used deep learning techniques to classify data and found the best model performance
4. Analyze data analysis and visualization
5. Found related data between emotion dataset with official reported dataset

6. Summarized finding of study

7. Composed academic article

8. Composed dissertation

1.7 Research Implication

The study proposed the greatest performance of a deep learning model to classify specific domain emotion messages and provide insights from data analysis and visualization. In addition, studies gave some examples to prove the relationship between emotion of people with official COVID-19 reported data, which is the number of patients and people who get vaccination.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter II
RELATED THEORIES

2.1 Natural Language Processing (NLP)

NLP is text processing by computer to understand meaning of human languages and utilize it to specific language-related tasks or applications. It is a subfield in artificial intelligence. This field has been developing machine translation language applications since the late 1940s. The application translates text from one language to others. Lately, there are many tasks in this filed for example, machine repose to spoken languages as voice assistance systems and automatic response chatbots; summarize the gist of large text as text summarizing; write new passages, article, or news as text generation; changing from text input to voice output as text to speech or speech recognition application; transforming from voice to text and understand meaning as speech to text software; detecting part of speech as part of speech tagging; fill words in the blank space to complete stories, and extracting subjective of text input as sentiment analysis. Moreover, all of tasks are benefits to uplift revenues and productivities for corporate and make work operation and processes easier and smoother to finish. In additional, nowadays there are several real word NLP cases in difference business industries. In general, an application consists of social media sentiment analysis, voice-enabled personal assistant, real times AI chatbots, speech recognition, translation, summarization. In healthcare industry, it is used in clinical diagnosis, clinical documentation, clinical decision support, patients trial matching, computer assisted coding. In finance area, it is applied to credit scoring, insurance claims management, financial reporting, auditing, fraud detection, contextual based stock price prediction. In retail and e-commerce, there are customer service chatbots, marketing intelligence, semantic based search. Lastly, in the human resource field, it is used in resume evaluation, recruiting bots, interview assessment and employee sentiment analysis.

Since text is unstructured data, they need specific knowledge and tools to attack these problems. There are many fields such as machine learning, deep learning, statistics and human linguistics to solve NLP tasks. To understand human languages and distinguish between type, purpose and level of language, machines need to know about 7 tiers of linguistics which consists of phonetic, morphological, lexical, syntactic, semantic, discourse, pragmatic level. Moreover, there are more things for machines to learn language such as synonyms, antonyms, homonyms, homophones, collocations, idioms, grammars, slangs, metaphors, and specific use in different areas of usages. Machines need to be taught the natural language rules by humans.

2.2 Deep Learning

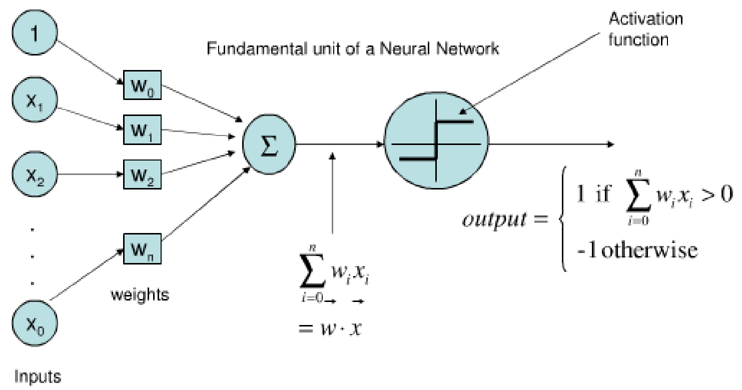
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Deep learning is a class of algorithms that have multiple layers of neuron structure like a human brain. The algorithms find the pattern in data by learning from examples. The hierarchy of algorithms learn input data and provide output. It can achieve many tasks, for example, prediction, classification, and time series analysis. It can handle both supervised and unsupervised learning tasks.

2.3 Neural Network

Neural networks are algorithms for recognizing relationships of data via processing them like cells in the human brain. There is a mathematical function in neurons in a neural network. The function receives and classifies data using the different architecture mechanism. It contains an interconnected node layer. Perceptron feeds information into neurons outputs values via activation function. The model learns a training dataset for predicting unseen data.

2.3.1 Perceptron



CHULALONGKORN UNIVERSITY

Figure 1 Neuron in neural networks

Perceptron is used for classify class data into category using output function and assigning variable as

x = input
w = weight
n = number of input feeding into perceptron

2.3.2 Learning rate

Learning a given dataset by adjusting weight in function every round when it is learned. The weights change following the function below.

$$Wi \leftarrow Wi + \Delta Wi$$

$$Wi = \alpha\ (\hat{y} - y)\ x\mathrm{i}$$

Assigning variable as

α = learning rate which is given number to let model learn and adjust data each step.

2.3.3 Activation Function

Activation function used in node to stimulate adjusting process in model. It makes models capture some features in data better and solve complex structures well. There are many activation functions to use in neural network mechanisms.

1) Sigmoid function

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

The Sigmoid function outputs in the range between 0 and 1. The sigmoid equation and graph are shown in the figure below.

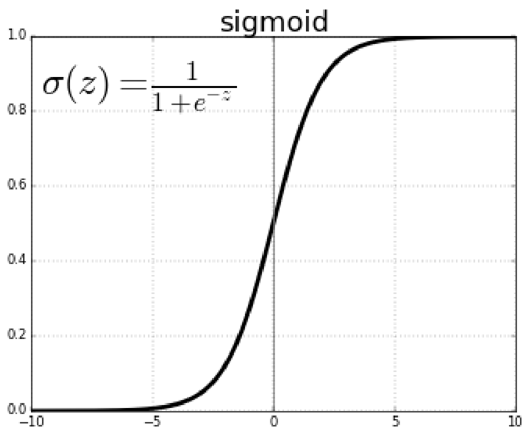


Figure 2 Sigmoid Function

2) Softmax function

Softmax functions squeeze output in the range between 0 and 1 as probabilities, the sum of output equals 1. The softmax equation is shown in the figure below.

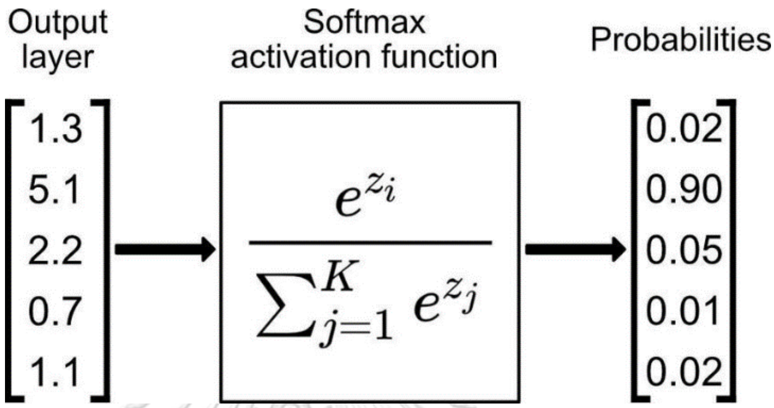


Figure 3 Softmax Function

3) Rectified Linear Unit Function (ReLU)

The ReLU function sends output values more than or equal to zero. The ReLU equation and graph are shown in the figure below.

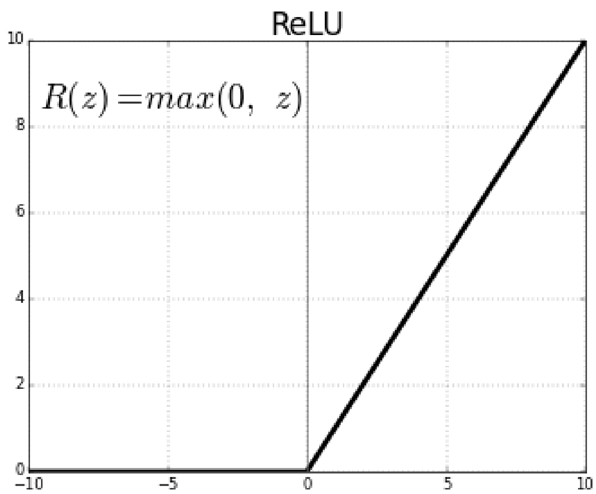


Figure 4 ReLU Function

2.3.4 Optimization

Optimization Algorithms are used when training machine learning models. To find the local minimum of a given function in iterative learning process depends on the convex function and model parameter. Many of them developed to help adjusting weight in models to make sure models learn samples better. It boosts the performance of the model to predict more correctly than without it. [1]

1) Gradient Descent or Batch Gradient Descent (Batch GD)

GD is an optimization algorithm to find a local minimum from a differentiable function in machine learning model. [2]

2) Stochastic Gradient Descent (SGD)

SGD is an iterative method for optimization algorithms which update their weight every training round. The more data in a model, the more variants and spending time increase. The flaws of this method are the unstable and complicated model. SGD optimization function for updating weight is shown below while w is weight and α is learning rate. [2]

$$w_{t+1} = w_t - \propto \frac{\partial L}{\partial w_t}$$

3) Mini-Batch Gradient Descent (Mini-Batch GD)

Mini-Batch GD is developed to get rid of cons from adjusting weights and pick the pros from Batch GD and SGD. It adjusts the weight every group of samples in one iterative process. [2]

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4) Adaptive gradient (AdaGrad)

AdaGrad was proposed by Duchi in 2011 [3] and this algorithm adapts the learning rate part by dividing the learning rate with the square root of S which accumulates current and past squared gradients following the time since the start. Given S initialized to 0.

$$w_{t+1} = w_t - \frac{\propto}{\sqrt{S_t + \epsilon}} \times \frac{\partial L}{\partial w_t}$$

where

$$S_t = S_{t-1} + \left[\frac{\partial L}{\partial w_t}\right]^2$$

5) Root mean square prop (RMSProp)

RMSprop is an adaptive learning rate that is an improvement of AdaGrad. The Algorithm takes the exponential moving average gradients instead of taking the cumulative sum of squared gradients as the AdaGrad algorithm. It was proposed by Hinton in 2012. [3] The algorithm tries to deal with the gradient problem which is some gradients may be tiny and others may be huge. Then it cannot adjust weight in the same amount for large and small gradients so algorithm would custom particular weight for adjusting them in each iterative.

$$w_{t+1} = w_t - \frac{\propto}{\sqrt{S_t + \epsilon}} \times \frac{\partial L}{\partial w_t}$$

where

$$S_t = \beta S_{t-1} + (1 - \beta)\left[\frac{\partial L}{\partial w_t}\right]^2$$

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

6) Adaptive moment estimation (Adam)

Adam was developed by Kingma & Ba in 2014. [4] The optimizer which can adjust the learning rate for each parameter in each training step and solve decay problems well. Adam is the most popular optimizer since it can solve the downsides of others such as making models to learn continually, saving a lot of training time and getting rid of fluctuations of parameters. The gradient part (V) and the exponential moving average of gradients and the learning rate part by dividing the learning rate (α) by the square root of S, the exponential moving average of squared gradients with V and S initialized to 0.

$$w_{t+1} = w_t - \frac{\propto}{\sqrt{\hat{S}_t + \epsilon}} \times \hat{V}_t$$

where

$$\hat{V}_t = \frac{V_t}{1 - \beta_1^t}$$

$$\hat{S}_t = \frac{S_t}{1 - \beta_2^t}$$

$$V_t = \beta_1 V_{t-1} + (1 - \beta_1)\frac{\partial L}{\partial w_t}$$

$$S_t = \beta_2 S_{t-1} + (1 - \beta_2)\left[\frac{\partial L}{\partial w_t}\right]^2$$

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.4 Feedforward neural network

The simplest structure of the neural network is the Feedforward Neural Network. The information is always fed in the forward direction. The network does not have any loops or cycles. The network consists of the input layer, many interconnected in the feed-forward hidden layer which is called multi-layer perceptron and output layer. Each layer has input, hidden and output nodes with an activation function.

Multi-layer perceptron networks mostly use a back-propagation method to learn data by applying the nonlinear transformation function to every neuron, creating output, comparing the output with the correct answer to calculate a predefined error-function. The error that the network gets is fed back to adjust weights to decrease error while learning. The algorithms continue to learn from datasets until they can get satisfying results. [5]

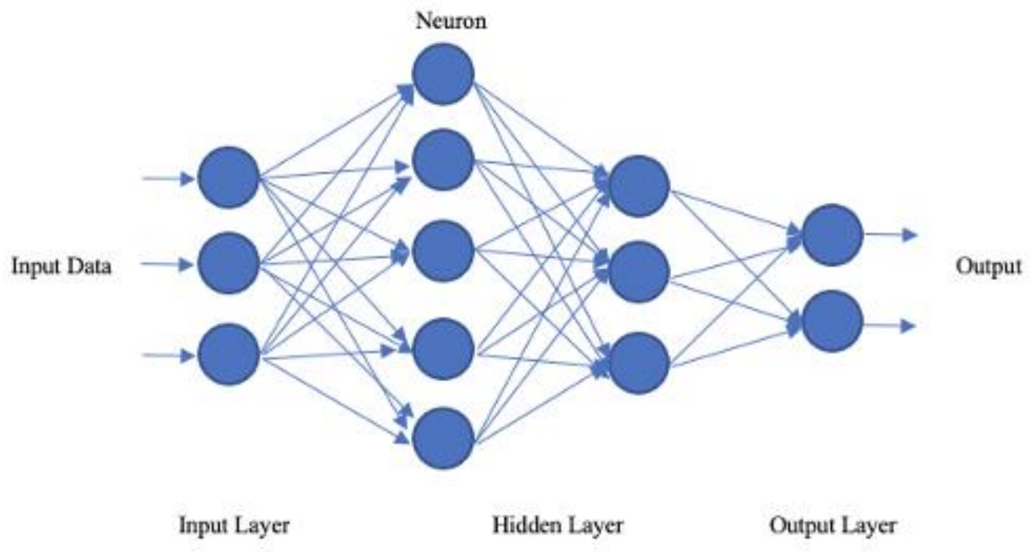


Figure 5 Multi-layer perceptron networks

The layers of multi-layer perceptron networks are input, hidden and output layers.

1) Input layer

The responsibility of the first layer is receiving information into the neural network and sending data to the next layer.

2) Hidden layer

The responsibility of this layer is getting information from the input layer to calculate information. Hidden layers can increase or decrease layers to fine tune model performance.

3) Output layer

The responsibility of this layer is to receive information from the hidden layer and calculate that information into output. This layer has the number of perceptron nodes equal to the class of predicting data.

2.5 Recurrent neural network (RNNs)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

RNNs are a subclass of neural networks which can handle sequential data or time-series data since they can remember prior inputs and use them to influence current input and output. This type of architecture performs well in speech recognition and natural language processing tasks. [6]

In Figure 6, Rolled RNN represents a one-time step of this network while unrolled RNN pictures show many times and use hidden layers to learn data as a sequence. Each layer of RNN uses the same weight for the entire of the network. Weights are adjusted through the learning process by backpropagation method.

However, the learning process of RNNs suffers from vanishing gradients when the size of the gradient is too small which means the model stops learning.

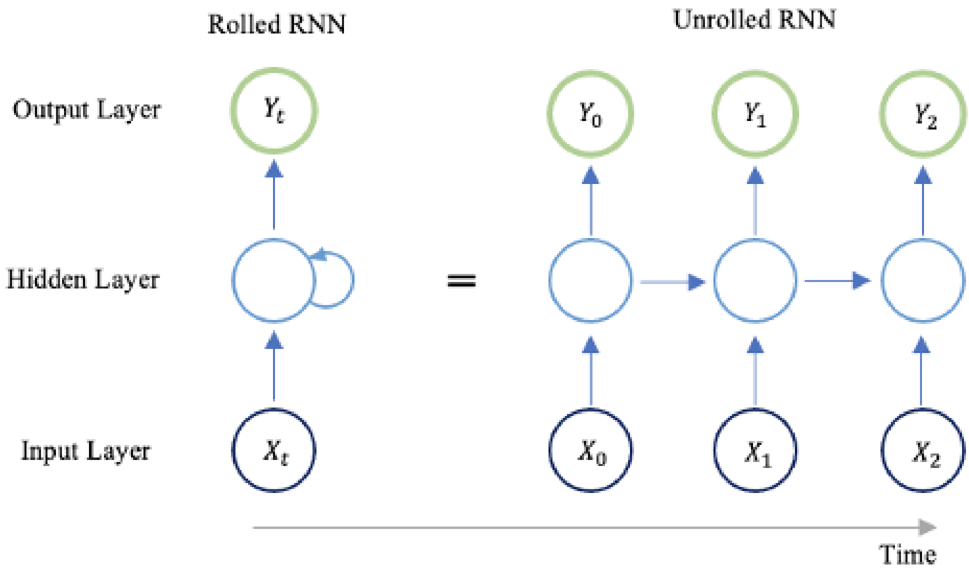


Figure 6 Recurrent neural network

Chulalongkorn University

To overcome this problem, the modified version of RNN, long short-term memory (LSTMs) was introduced by Hochreiter and Schmidhuber [7]. They create new cells in hidden layers, and there are three gates which are input, forget and output gate that use a suitable amount of previous data to predict output in the next step. In addition, combining with bidirectional data to feed data into networks, resulting in the bidirectional model. The prediction of this model works well for a sequence of text input, but they cannot be used for time series prediction because prediction problems cannot let the model know future data.

Gated recurrent unit (GRUs) are the smaller version of LSTMs architecture because they have redesigned three gates to two gates which are reset and update gates to overcome a series of data problems. The gates are used to control information issues. The size of this model is smaller and lighter than LSTMs. (Figure 7) [8]

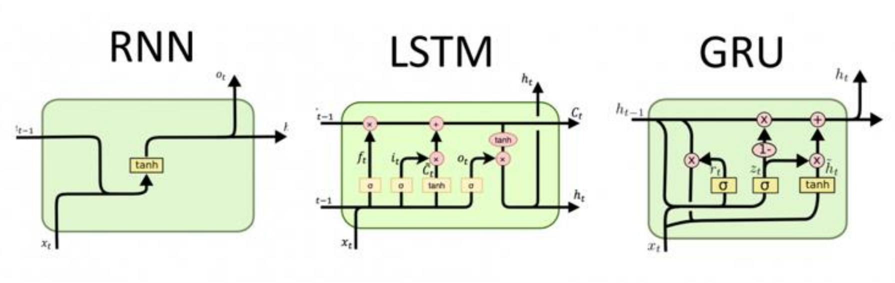


Figure 7 The cell in RNNs, LSTMs, and GRUs

2.6 Transformer

A transformer is an encoder-decoder-based neural network with an attention mechanism that determines the important relationship of input, for instance, collocation of words that both words are likely to be used to together. The transformer was proposed in 2017 by Vaswani, Shazeer [9]. This mechanism can solve sequences problems well, especially on text data, leading to the popularity for solving natural language processing problems. (Figure 8)

จุฬาลงกรณ์มหาวิทยาลัย

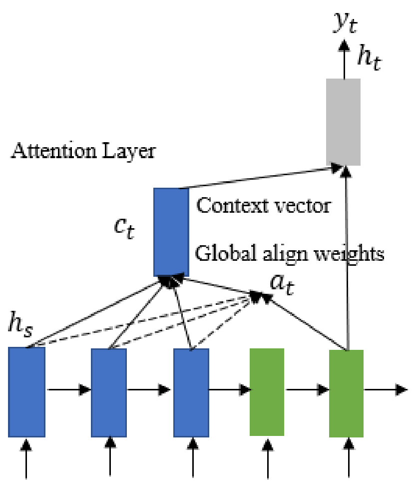


Figure 8 Attention layer

Bidirectional Encoder Representations from Transformers (BERT) is a language model which represents the relationship of the word in language in terms of statistical numbers. The BERT model consists of the encoder from the transformer-based technique with attention, feed-forward layers. It was proposed by Devlin, Chang [10]. BERTs are the state-of-the-art in natural language processing in these decades because they can achieve most tasks in greater performance than other models or techniques.

WangchanBERTa is pretraining transformer-based Thai Language Models. It was proposed by Lowphansirikul, Polpanumas [11]. They provided a model of Thai language which was trained on a large Thai language dataset that has a total size of 78 GB in several domains such as social media messages, news articles, and other public datasets. Their models outperform strong baselines in many downstream tasks such as sequence classification and token classification.

2.7 Developing and Evaluating estimator performance

2.7.1 K-fold Cross Validation

K- fold cross validation technique is popular method to evaluate models. This technique is used to get a more generalized and reliable model by splitting the training dataset into K-folds. The number of iterations of the training model equals K times and uses it as validation data to evaluate training performance with a different sample of dataset in every training step. The benefit is reducing the bias of cherry-pick samples to train. Training steps to find the best model and process to train this technique are shown below.

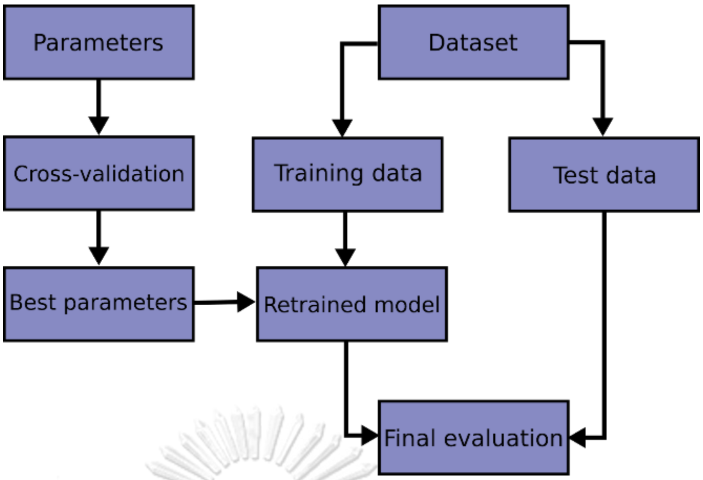


Figure 9 K-fold Cross Validation Step

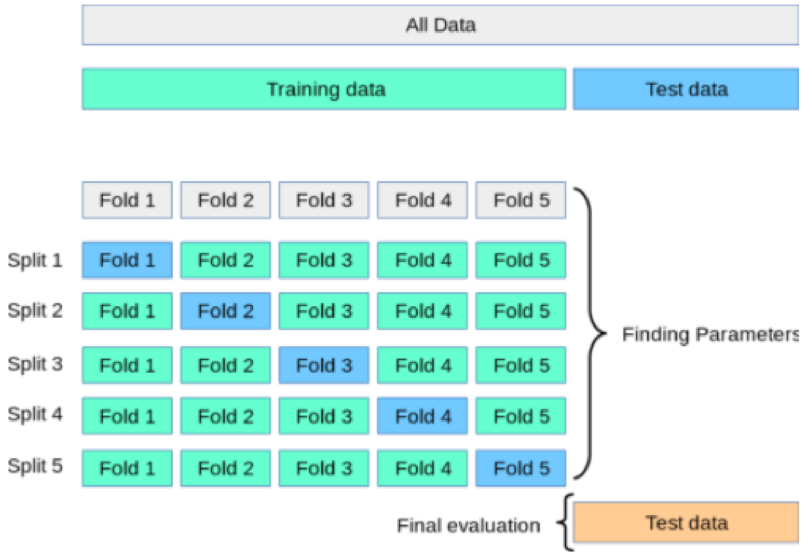


Figure 10 5-fold Cross Validation

2.7.2 Grid Search

Grid search technique is used to tune hyper-parameters in a model to get the best model performance and save time, effort and resources for training models. In deep learning, model experiments adjusting hyper-parameters such as batch size, training epochs, optimization algorithms, learning rate, weight initialization, activation functions, dropout regularization and K-fold cross validation.

2.7.3 Normalization

Max-Min Normalization is a rescaling technique to change the range of data between 0 and 1 by adjusting every sample with minimum, maximum and range of old data. The rescaling equation is shown below.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

2.7.4 Logarithm

The exponent or power to be a base for another number must be raised to yield a given number. It is inverse function to exponentiation. If logarithm of x to the base b, $\log_b x$. The common base of logarithm is base 10 and is called for decimal number. The base which is e of natural logarithm roughly equal 2.718. It commonly used in mathematics filed since it is simpler derivative or integral. Logarithmic scales are used for reducing large range of numbers to smaller range.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.7.5 Smoothing Adjustment Technique

Most of time series data have a wide range of short-term volatility in data, effects in harder to understand and analyze real trend of data. Smoothing techniques become to reduce and eliminate the volatility. One common technique is moving average, it is the window weighted average on focusing period. The benefit of this technique is to remain some volatile after adjustment in each round. The simple moving average formula is shown below.

$$\bar{y}_t = \frac{y_t + y_{t-1} + \cdots + y_{t-n-1}}{n}$$

where y is the number, t is the current time period, and n is the number of focusing time periods for average number in period. The n variable should select the suitable period for each specific domain. The larger n is given, the smoother the adjusted series are.

2.7.6 Correlation

Correlation helps to tell relationship between two entities how two variables connect. It can measure reliability and validity in dataset but it does not mean the relationship is causal. Correlation values range between -1 and 1. The two key components of a correlation value is sign of number, if sign is positive then the correlation relates in same direction but If sign is negative then the correlation relates in opposite direction. Another part is magnitude to tell how much correlate of two data. The larger of magnitude are, the stronger the correlation is.

Pearson correlation coefficient is the common of correlation between data. It is denoted by letter (r). The formula of Pearson correlation coefficient is shown below.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where, r is Pearson correlation coefficient, x is values in the first set of data, y is values in the second set of data, and n is total number of values.

2.7.7 Metrics

1) Confusion Matrix

Matrix for showing the summary of results from predicting classification problems. It shows the count of the number of correcting and non-correcting classify data in each class.

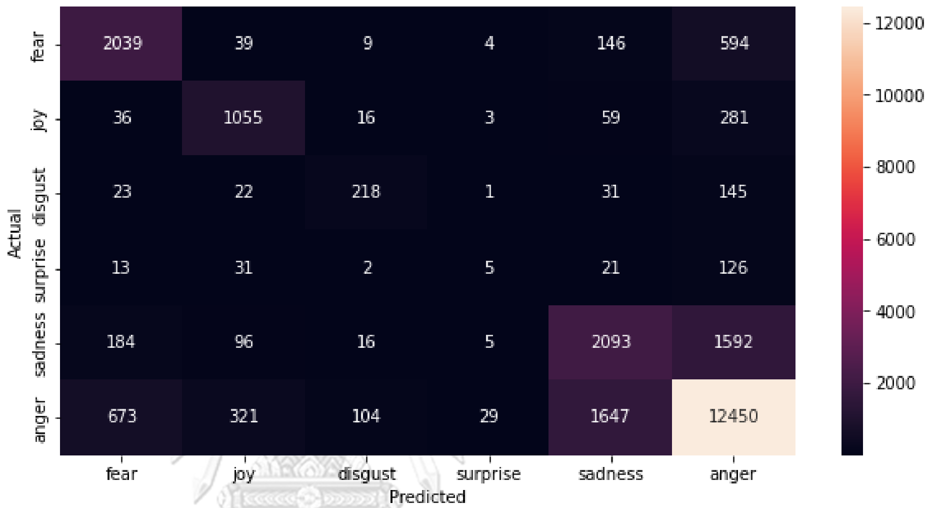


Figure 11 Confusion Matrix

2) Classification Metrics

From the confusion matrix,

Model predict class (P) and actual class (P) is called True positive.

Model predicts class (P) but actual class (N) is called False positive.

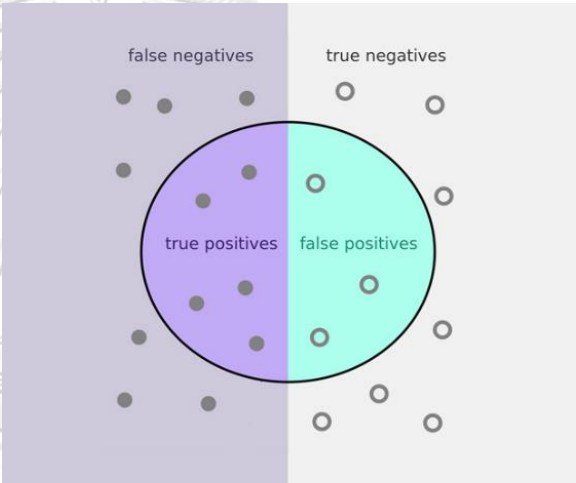
Model predicts class (N) and actual class (P) is called False negative.

Model predicts class (N) and actual class (N) is called True negative.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

		Predicted class	
		P	*N*
Actual Class	*P*	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Figure 12 Classification Metrics Table



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Figure 13 Classification Metrics Explain

- Precision is used for evaluation quality of being exact of model by calculate correcting of prediction in class P

$$Precision = \frac{TP}{TP + FP}$$

- Recall is used for evaluation how much correctly identify of model by calculate amount of correcting prediction in class P

$$Recall = \frac{TP}{TP + FN}$$

- F1 is used for evaluation of both precision and recall at the same time.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

- Accuracy is used for evaluation overall in model

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Chapter III
LITERATURE REVIEW

There are several researches on emotion classification on virus infection especially on COVID-19 related tweet messages in many languages. They utilize machine learning and deep learning knowledge to classify their data and find insights from specific emotion from specific users in local areas during pandemic situations. The related research will be outlined as follows.

Al-Laith and Alenezi [12] studied sentiment and symptom classification from January to August 2020 on Arabic language dataset. They labeled some datasets by manual labeling and trained a Fasttext model which is a neural network to label the rest. Their experiment on the LSTM model achieves 82.9% on F1-score. Moreover, they provide topic detection to find the important word in the dataset.

Garcia and Berton [13] classified sentiment and topic detection on both English and Portuguese language in USA and Brazil between April and August 2020 by using an auto-labeling library to create labels. They found that the best approach is logistic regression and Linear SVM which provide performance equal to 87% on F1-score.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Kausar, Soosaimanickam [14] reported emotional classification in many countries around the Persian Gulf area between 21 June and 20 July 2020 with 50000 tweets using an automated library to prepare a dataset and classify into 8 emotions (fear, joy, anticipation, anger, disgust, sadness, surprise, and trust).

Mathur, Kubde [15] studied on an English language dataset from 22 January to 15 April 2020 contains 30,000 tweets and classified them into nine classes which are anger, hope, disgust, fear, positive, negative, joy, sadness, and surprise.

Pasupa, Ayutthaya [16] studied sentiment analysis on Thai children tales dataset using different word embedding for semantic, POS-tag for grammar and emotion of word into emotional score. They compared results on deep learning models which are CNN, LSTM, BLSTM. The result showed CNN achieved the highest F1-score at 81.7%

Pasupa and Ayutthaya [17] trained many models which were deep learning consists of CNN and LSTM and hybrid models as BLSTM-CNN, CNN-BLSTM, BLSTM+CNN, and BLSTM×CNN on Thai-SenticNet5 corpus into positive, neutral and negative feelings and evaluated on three Thai language social media datasets which were ThaiTales, ThaiEconTwitter, and Wisesight. The highest performance was BLSTM-CNN on F1-scores at 74.36%, 77.07%, and 55.21%, respectively.

Those related works are used for setting up and adjusting annotation labels of data and creating classification models.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Chapter IV
Research Method

4.1 Data overviewing

Data is collected from Twitter API called Twint library via many hashtags consist of nine hashtags which are โควิด (COVID), โควิด19 (COVID19), โควิดวันนี้ (COVIDtoday), โควิด19วันนี้ (COVID19today), ไวรัสโคโรนา2019 (VirusCorona2019), ไวรัสโคโรน่า (VirusCorona), ไวรัสโคโรน่า (CoronaViras), โคโรน่า (Corona). The data of this experiment are Tweets related to COVID-19 in Thai language between 1 January 2020 and 30 June 2021. The unique message in the dataset consists of 120,642 tweets. Moreover, the official reported data of new infection patients and vaccination people in Thailand from the Department of disease control, Ministry of Health in Thailand are also included. (see Table 1)

The attributes of this dataset are shown in Table2.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

date	tweet	language	hashtag	nlikes	nreplies	nretweets	label
2020-02-18	โอเค สบายใจขึ้นเยอะเลยยย	th	['โควิด']	2.0	0.0	0.0	joy
2020-02-20	เศร้ามาก เค้าตายไวมาก	th	['โควิด']	1.0	0.0	2.0	sadness
2020-02-30	กลัวติดอ่า	th	['covid19', ' โควิด'']	18.0	1.0	6.0	fear
2020-03-08	อีโควิท มึงจะทำกูเป็นบ้า ยาซึมเศร้าในมือนี่สั่นไปหมด 555	th	['โควิด']	1.0	0.0	0.0	anger
2020-04-16	หาซื้อมาส์กไม่ได้ กลัว	th	[‘โควิด-19', ' โควิด', ' โควิด19']	1.0	0.0	1.0	fear
2020-02-16	ปิดอะไรเยอะแยะ หาของกินไม่ได้	th	[' covid', ' โควิด', ' coronavirus']	8.0	0.0	11.0	anger
2020-02-16	สงสารแม่เค้าตายมาก	th	['covid', ' โควิด', ' coronavirus']	10.0	0.0	7.0	sadness
2020-10-20	เซอร์ไพร์สมาก คนหายป่วยเยอะขึ้นมาก	th	['โควิด']	1.0	0.0	2.0	surprise
2020-12-17	รังเกียจคนออกกฏเคอร์ฟิว	th	[' โคโรน่า', ' โควิด'']	1.0	0.0	0.0	disgust

Table 1 Example dataset

Number	Column name	Meaning
1	date	Date
2	tweet	Messages
3	language	Language of message
4	hashtag	hashtag
5	nlikes	Number of likes for message
6	nreplies	Number of replies for message
7	nretweets	Number of retweets for message
8	label	Emotion label of message

Table 2 Dataset Attribute

4.2 Data Filtering

This study filters only emotion messages to train the model. There are mixed up between emotion and non-emotion messages from the raw dataset. Overall, emotional messages are around 30% from COVID-19 messages. (see Table 3)

Tweets reaction	Average of Emotion messages/COVID related messages per day ratio
Tweets	36.4205%
Likes	33.9779%
Replies	31.5359%
Retweets	36.2164%

Table 3 Average of Emotion messages/COVID related messages per day ratio

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

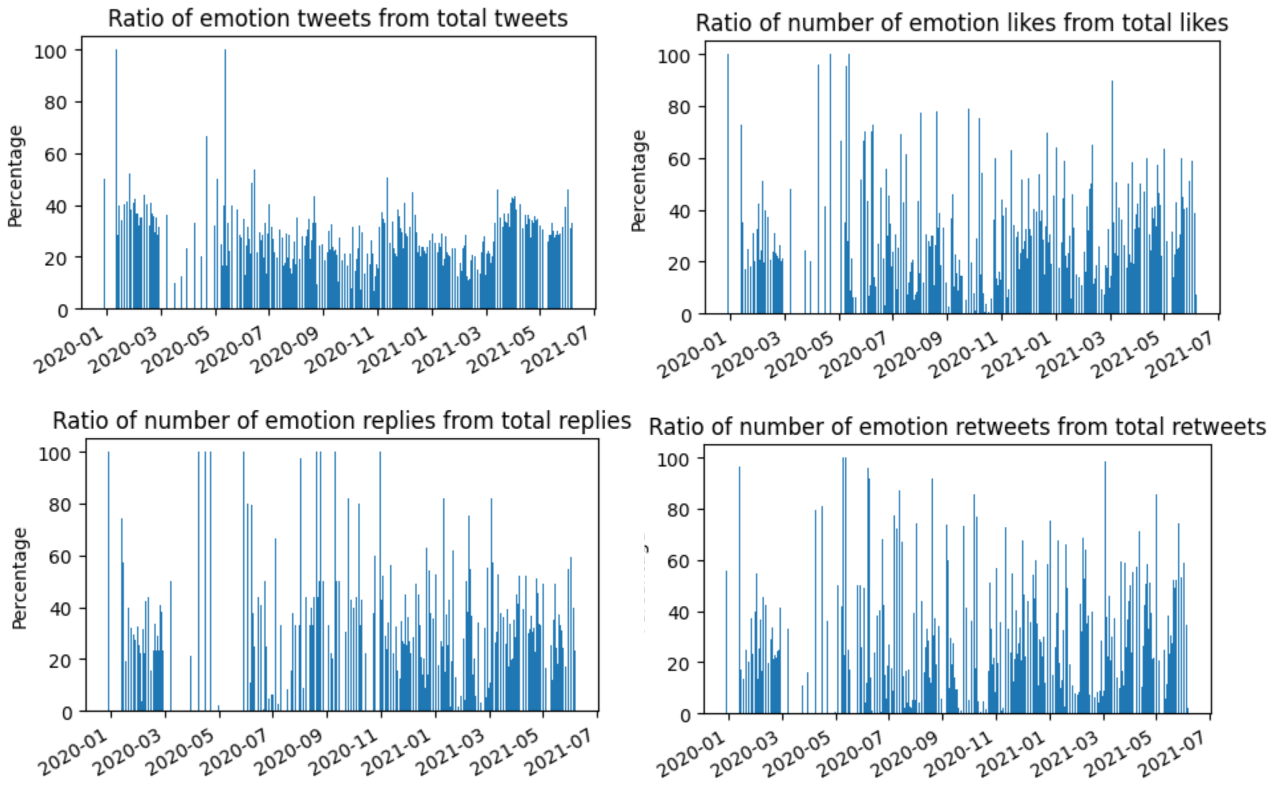


Figure 14 Volume of emotional data by specific categories of data

Figure 14 show daily the percentage of emotion messages over COVID related messages in each type of tweets. Average of emotion messages on different on each type of tweets are shown on Table 3. Moreover, the ratio on each class of emotion over all messages depending on each type of tweets are shown below on Table 4.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Label	Average Ratio of Tweets	Average Ratio of Likes	Average Ratio of Replies	Average Ratio of Retweets
anger	60.3331%	60.4661%	61.3827%	61.8089%
sadness	19.7507%	20.4538%	15.7935%	21.1076%
fear	14.7794%	15.3906%	15.0448%	15.0887%
joy	11.9640%	11.5916%	9.5695%	8.7742%
disgust	3.4824%	2.1820%	2.0150%	1.6995%
surprise	3.0401%	1.4805%	3.0404%	1.2969%
total				**100%**

Table 4 Average of Emotion messages/COVID related messages per day separate by emotion ratio

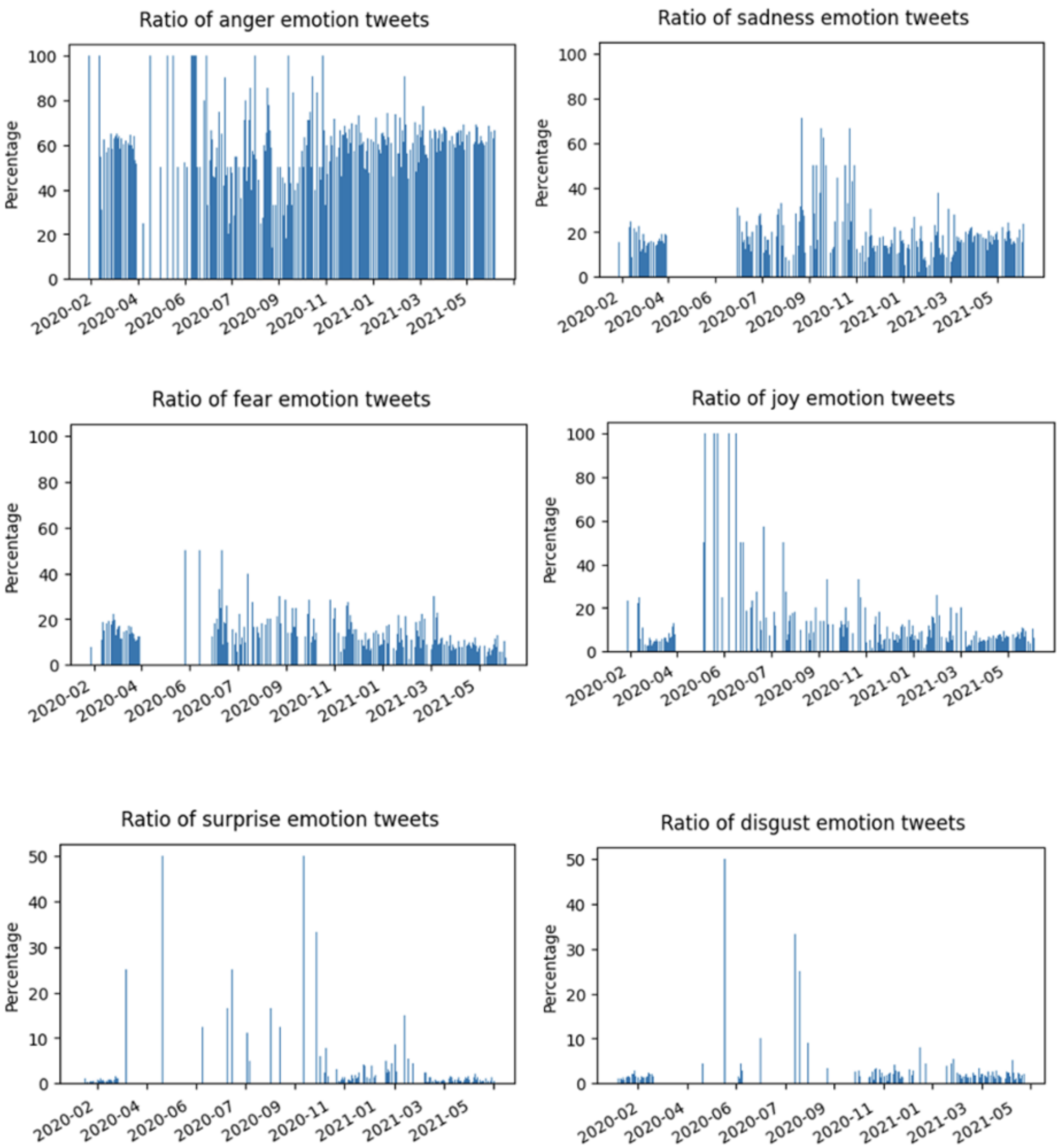


Figure 15 Ratio by emotion tweets

Figure 15 show daily the percentage of tweet separate by each class of emotion messages over COVID related messages. The daily of emotion tweets can be order in anger, sadness, fear, joy, disgust and surprise feeling. The details of average of tweet ratio on each class are shown on Table 4.

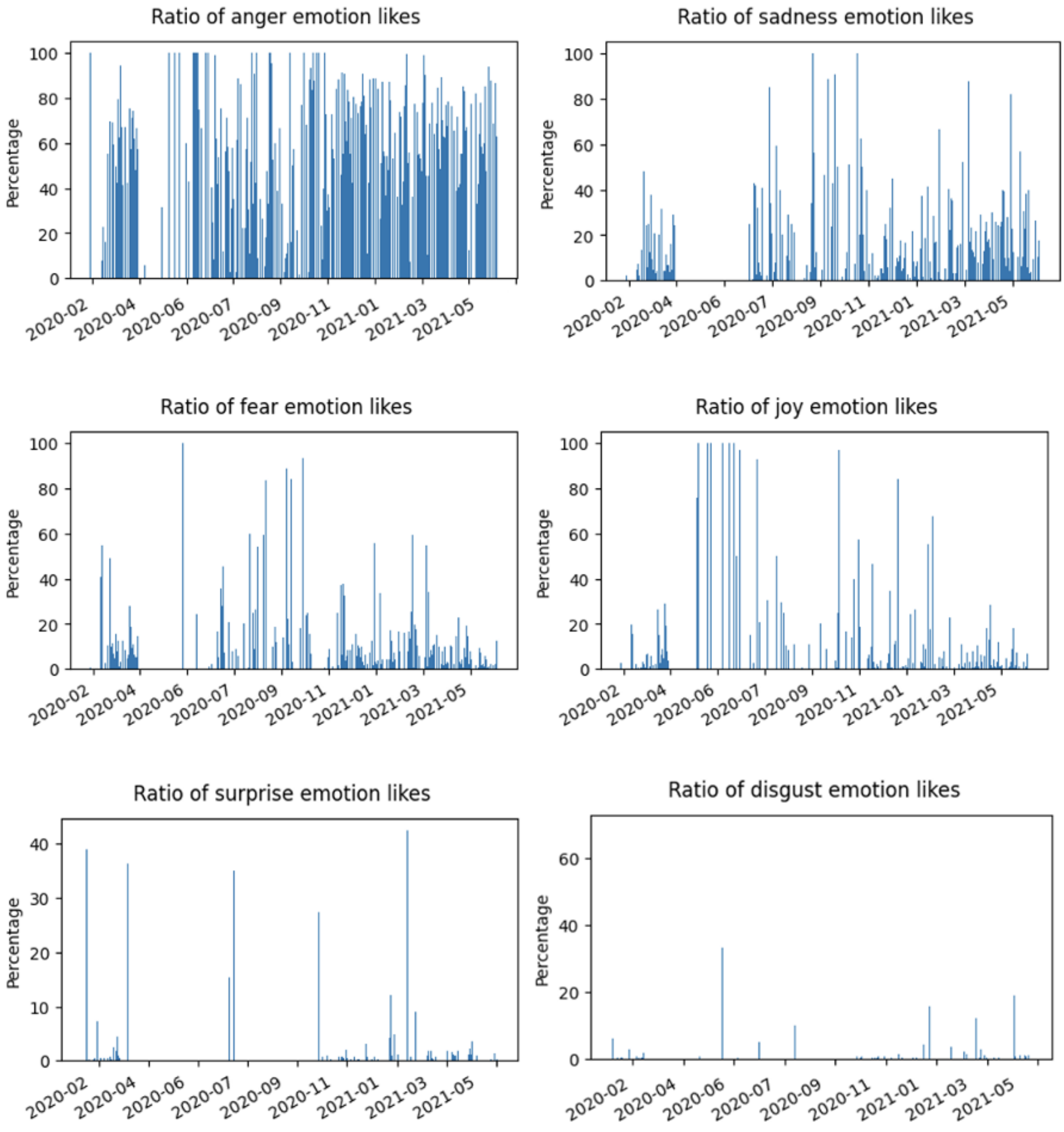


Figure 16 Ratio emotion likes

Figure 16 show daily the percentage of likes separate by each class of emotion messages over COVID related messages. The daily of emotion tweets can be order in anger, sadness, fear, joy, disgust and surprise feeling. The details of average of likes ratio on each class are shown on Table 4.

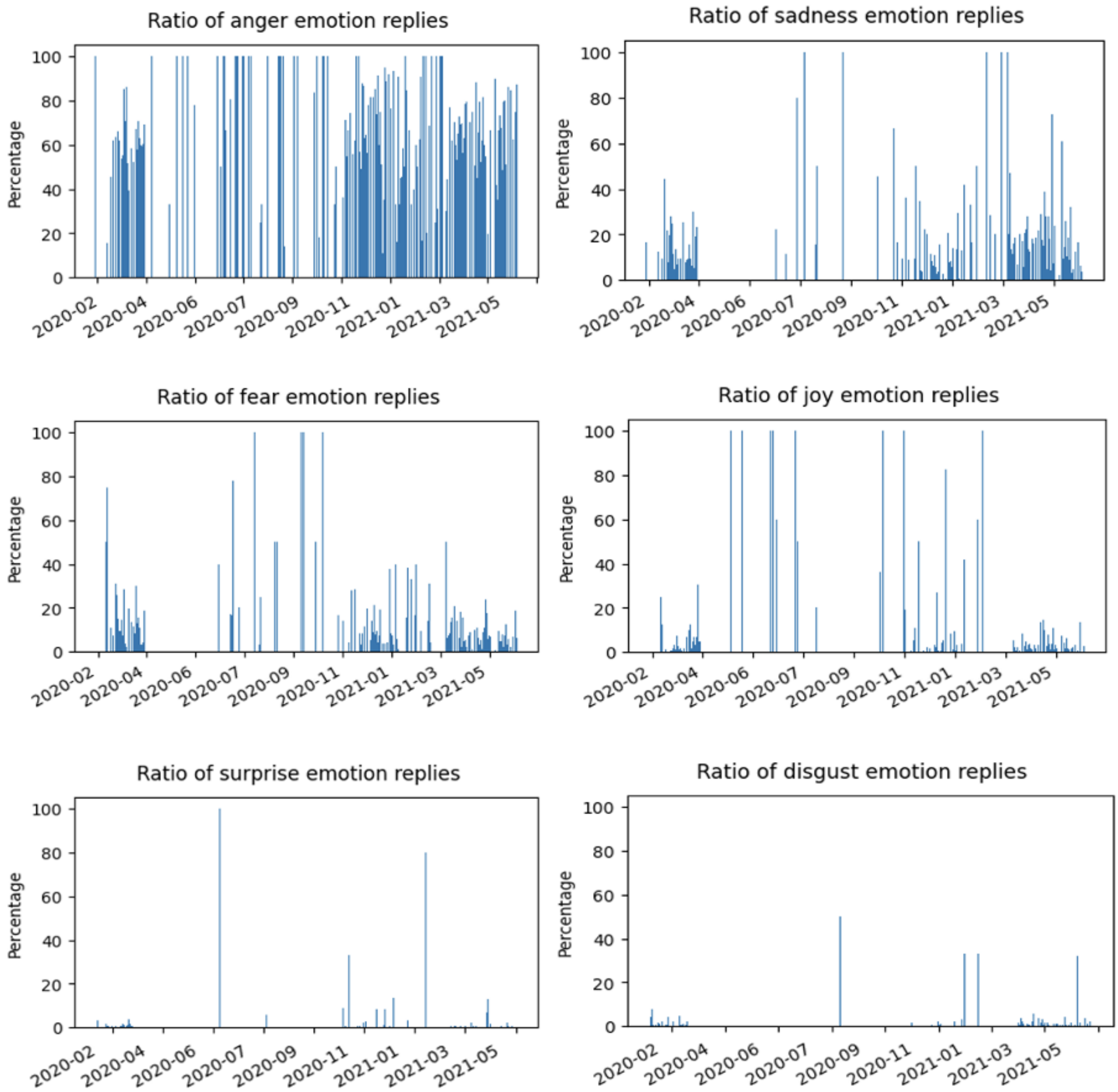


Figure 17 Ratio emotion replies

Figure 17 show daily the percentage of replies separate by each class of emotion messages over COVID related messages. The daily of emotion tweets can be order in anger, sadness, fear, joy, disgust and surprise feeling. The details of average of replies ratio on each class are shown on Table 4.

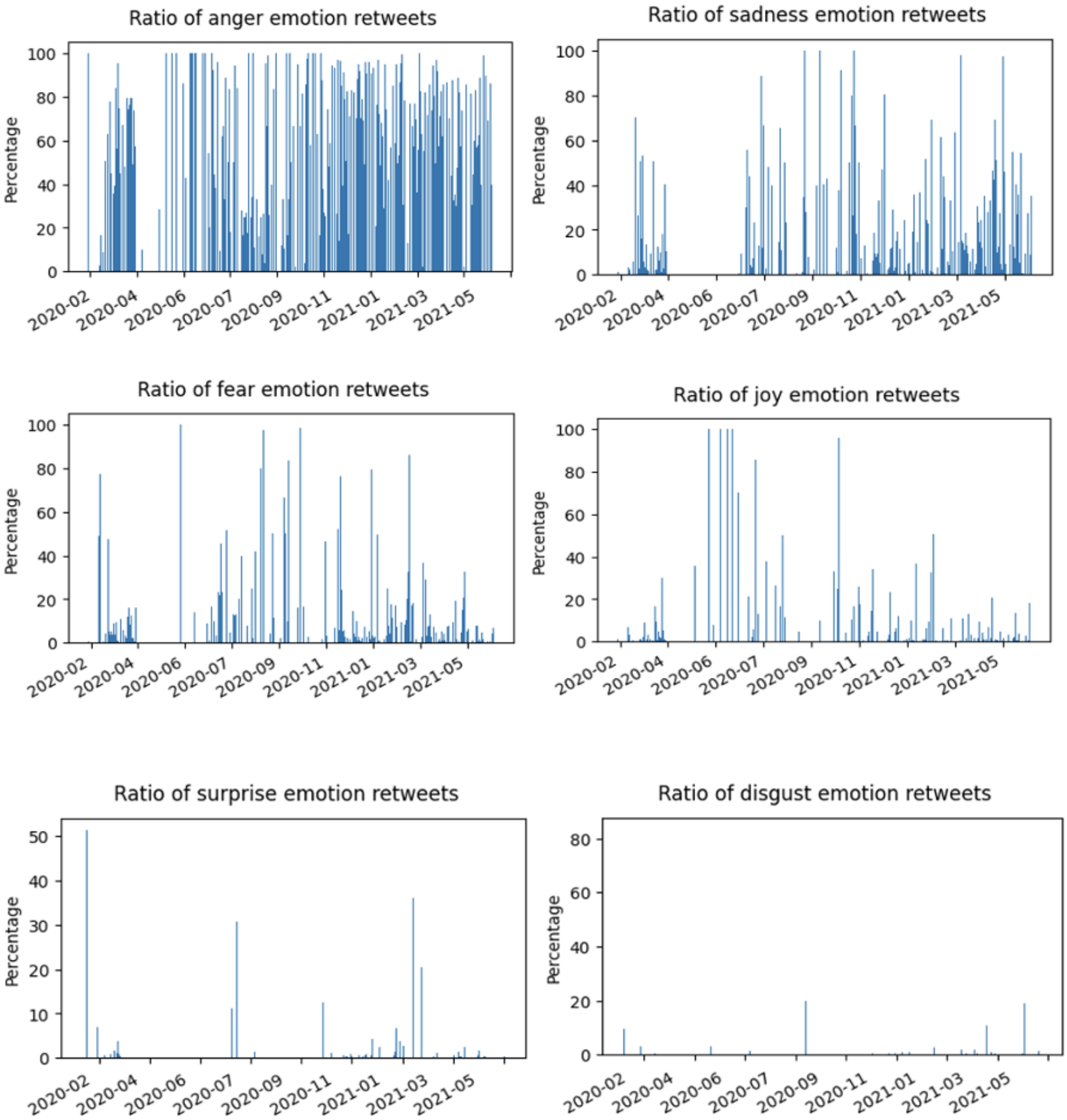


Figure 18 Ratio emotion retweets

Figure 18 show daily the percentage of retweets separate by each class of emotion messages over COVID related messages. The daily of emotion tweets can be order in anger, sadness, fear, joy, disgust and surprise feeling. The details of average of retweets ratio on each class are shown on Table 4.

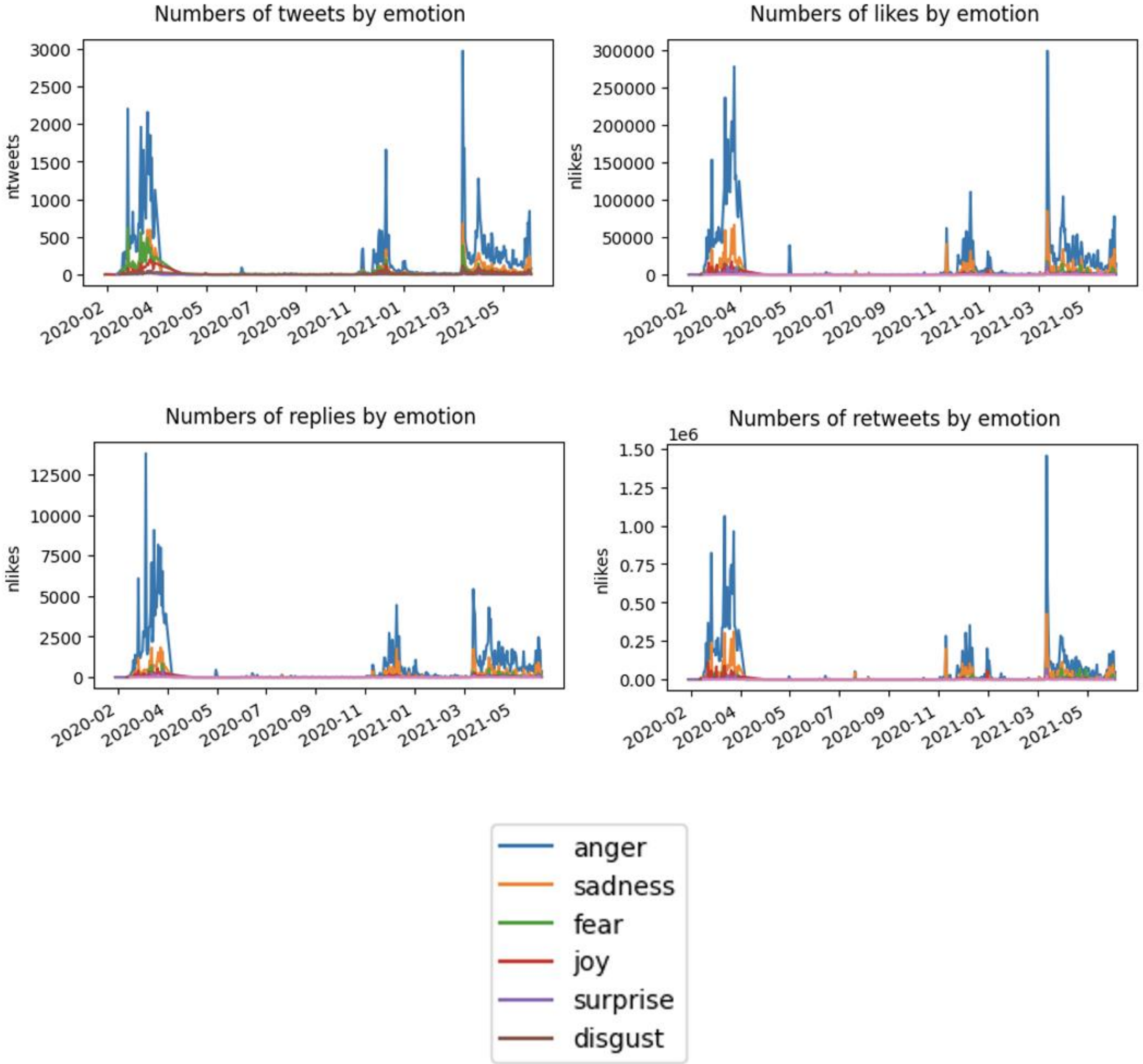


Figure 19 Emotion category trend

Figure 19 show movement plot of each category of emotion in different type of tweets which are tweets, likes, replies and retweets, respectively.

After filing out of data, the rest of data to train model and find data analysis and visualization and it was separated into ratio as table below.

Label	Number of examples	Ratio
anger	75,271	63.17%
sadness	19,630	16.47%
fear	13,986	11.74%
joy	7,114	5.97%
disgust	2,175	1.83%
surprise	989	0.83%
total	**116,379**	**100%**

Table 5 Data Statistics by label

4.3 Analysis tools

To classify and analyze emotion messages we use Python programming language (version 3.7.11) and collect classification performance, data insight and visualization. Annotation of unlabeled data using the pre-train dataset from WangchanBERTa with Transformers version 3.5.0, Thai2transformers version 0.1.2, and Pytorch version 1.4.0. We use an open-source system library called Keras for deep learning framework with TensorFlow version 2.6.0 as the backend and Python version 3.7.11.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.4 Data Pre-processing

The first step is to clean data by removing punctuation, tabs, blank space, number, hashtag, user mention, and special characters such as #,@ in the messages. The next process is to obtain unique messages for classification and data analysis purposes.

4.5 Data Labelling

Since the size of the dataset is too large to do hand labeling, we use WangchanBERTa pre-train model to predict 6 classes of labels for our dataset. There is around 10-15% non-accurate label compared with native judgment. We validate the transformer mechanism-based label by hand using a random pick sample to check around 3,000 tweets.

4.6 Word Tokenize

This work used the attacut approach which is a fast and accurate neural network-based Thai word segmenter to cut sentence [18] into a single word.

4.7 Data Transformation

Normalize dataset for data analysis purpose by its minimum and maximum and scale all data between 0 to 1.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.8 Splitting data

Dividing the dataset into train, validation, and test dataset. There are 96,513 messages (80% of all dataset) for the training dataset and the rest is the testing dataset containing 24,129 messages (20% of all dataset). For validating model performance, K-fold cross validation with 20% of training data is used.

4.9 Training Model with Grid Search and K-fold Cross Validation

Experimented on deep learning-based models via Keras for deep learning framework, we define a structure for every model. Model needs to specify hyper-parameters before training. Hyper-parameters can change to fine tune the model for improving model performance, which are batch size, training epochs, optimization algorithms, learning rate, activation functions.

Studied the token of messages as input and trained the model on 6 emotion classes as output. The steps for the training model are shown as following.

4.9.1 Defining Network

Defining function of network structure starts from adding an index of tokenized words in the input layer with the assigned the highest length of sample in the training dataset. Next step is defining the embedding layer from the first layer and adding a deep learning model layer with nodes in the next layer. Then, add a dense layer with the number of nodes equal to the number of output classes. (Figure 23)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 100)]	0
embedding (Embedding)	(None, 100, 32)	2782720
gru (GRU)	(None, 32)	6336
dense (Dense)	(None, 6)	198

Figure 20 Defining Network

4.9.2 Compiling Network

This step transforms the predefined network and process training step. This compile step needs to assign hyper-parameters which are optimizer, loss function, and metric.

4.9.3 Grid Search and K-fold cross validation

Defining hyper-parameters for fine-tuning the model on this step. Assigning all of the model parameters and search for the set that achieves the best performance. This step is not only changing parameters but also validating the training model via K-fold cross validation.

4.9.4 Fitting Network

Fitting model with training data and label and defining hyper-parameters which is batch size and epochs.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.9.5 Evaluating Network

Testing models with testing data and labels to get the performance of the model.

4.10 Model Architecture

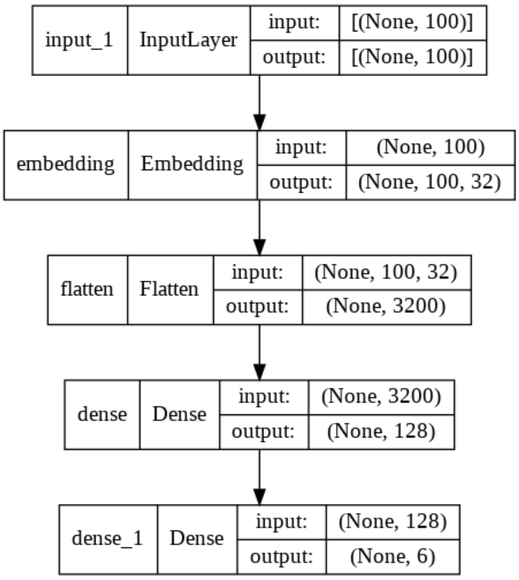


Figure 21 Multi-layer Perceptron Model Architecture

Figure 21 shows the best performance of Multi-layer Perceptron architecture with the greatest parameters which use input layer, embedding layer to prepare data, flatten layer to reshape matrix then feed into dense layer to calculate information in 128 neurons and send output out into 6 classes.

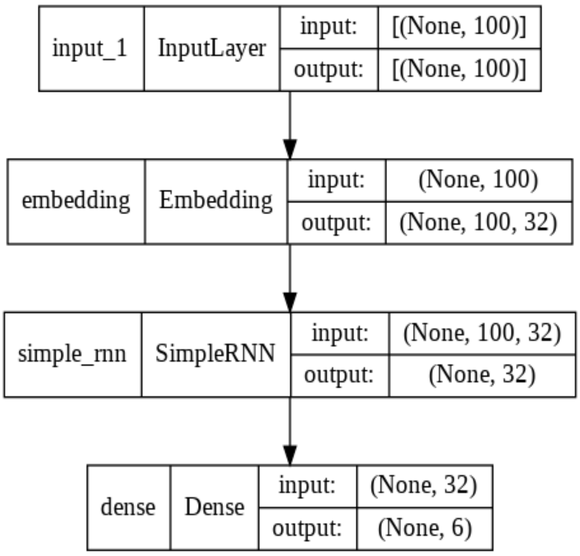


Figure 22 RNN Model Architecture

Figure 22 shows the best performance of RNN architecture with the greatest parameters which use input layer, embedding layer to prepare data, RNN layer to calculate information in 32 neurons and send output out into 6 classes.

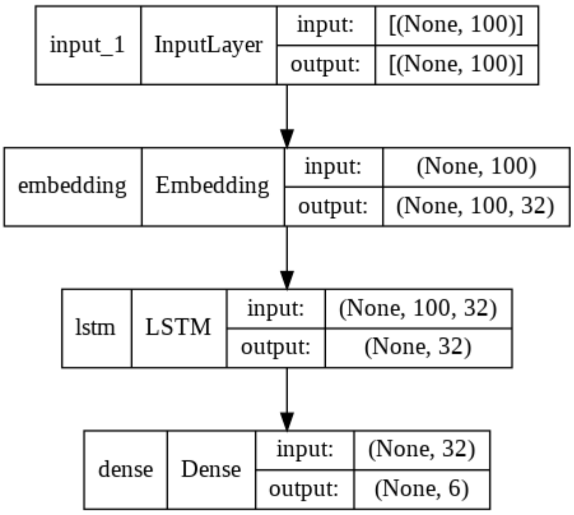


Figure 23 LSTM Model Architecture

Figure 23 shows the best performance of LSTM architecture with the greatest parameters which use input layer, embedding layer to prepare data, LSTM layer to calculate information in 32 neurons and send output out into 6 classes.

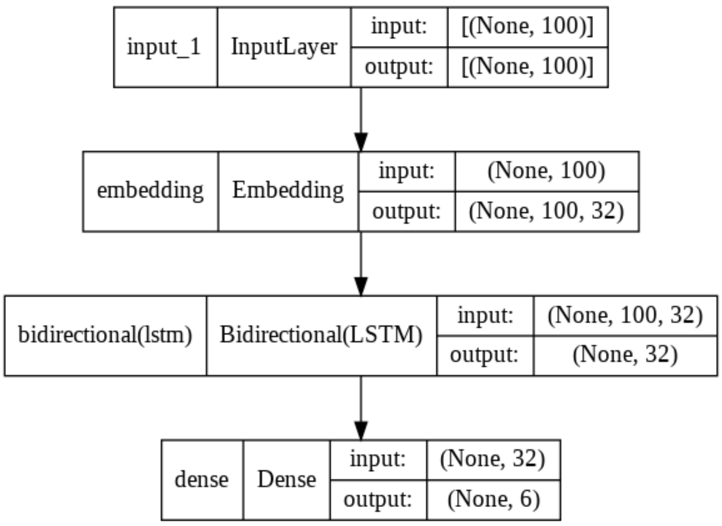


Figure 24 BiLSTM Model Architecture

Figure 24 shows the best performance of BiLSTM architecture with the greatest parameters which use input layer, embedding layer to prepare data, BiLSTM layer to calculate information in 32 neurons and send output out into 6 classes.

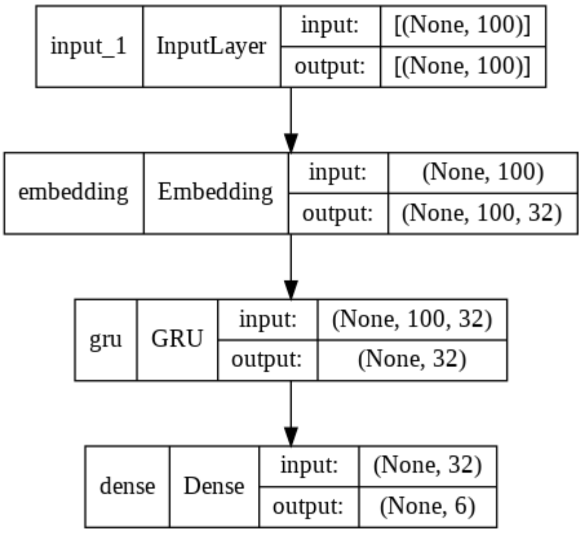


Figure 25 GRU Model Architecture

Figure 25 shows the best performance of GRU architecture with the greatest parameters which use input layer, embedding layer to prepare data, GRU layer to calculate information in 32 neurons and send output out into 6 classes.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4.11 Training Step Flowchart

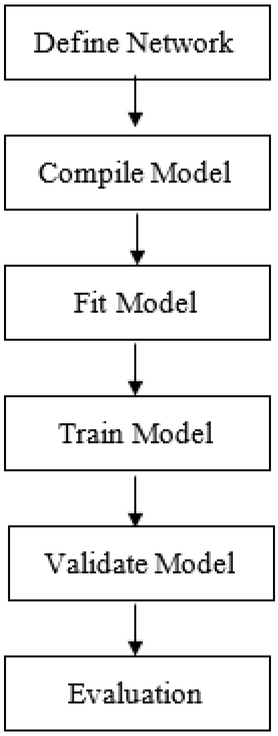


Figure 26 Train and Evaluation Model Flowchart

Figure 26 show training and evaluation step of experimental to get the best model on this dataset.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter V
RESULTS

5.1 Model Performance

Validation result on Multi-layer Perceptron model

Rank	Optimizer	Batch Size	Epochs	Average validation accuracy	Standard deviation
1	RMSprop	64	5	0.7517	0.0062
2	RMSprop	128	5	0.7485	0.0098
3	RMSprop	32	5	0.744	0.0064
4	Adam	128	5	0.7409	0.004
5	Adam	64	5	0.7375	0.0054
6	Adam	32	5	0.7327	0.0045
7	Adam	64	15	0.7188	0.0049
8	Adam	128	15	0.7168	0.0048
9	Adam	32	10	0.7167	0.0085
10	Adam	128	10	0.7163	0.005
11	Adam	32	15	0.7158	0.006
12	RMSprop	128	10	0.7133	0.0018
13	Adam	64	10	0.7122	0.0119
14	RMSprop	32	10	0.7052	0.0092
15	RMSprop	64	10	0.7003	0.0136
16	RMSprop	128	15	0.6967	0.0144
17	RMSprop	32	15	0.6928	0.0067
18	RMSprop	64	15	0.6897	0.009

Table 6 Grid search cross validation training result of Multi-layer Perceptron model

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Result of the best model of Multi-layer Perceptron from grid search and cross validation training on the highest average validation accuracy (Table 6) is the model with RMSprop as optimizer, 64 of Batch sizes, 5 of epochs and 5-Fold cross validation.

K-fold	Validation accuracy
1	0.7472
2	0.7437
3	0.7508
4	0.7616
5	0.7549
Average	0.7517

Table 7 5-fold cross validation result of the best Multi-layer Perceptron Model

Table 7 show validation accuracy result of each 5-fold cross validation and average accuracy result from the best hyper-parameters of Multi-layer Perceptron model.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Evaluation result on Multi-layer Perceptron model

	precision	recall	f1-score	support
fear	0.6870	0.7202	0.7032	2831
joy	0.6746	0.7276	0.7001	1450
disgust	0.5973	0.4955	0.5416	440
surprise	0.1064	0.0253	0.0408	198
sadness	0.5236	0.5251	0.5244	3986
anger	0.8197	0.8178	0.8188	15224
accuracy			0.7402	24129
macro avg	0.5681	0.5519	0.5548	24129
weighted avg	0.7366	0.7402	0.7380	24129

Figure 27 Multi-layer Perceptron Model Performance on Classification Report

Figure 27 is evaluation results both overall and each class of emotion on test data on 4 metrics. The overall results are 74.02% of accuracy, 73.66% of precision, 74.02% of recall, and 73.80% of F1-score. Model perform the best on anger emotion on 3 metrics which is precision, recall, and F1-score with 81.97%, 81.78%, and 81.88%, respectively.

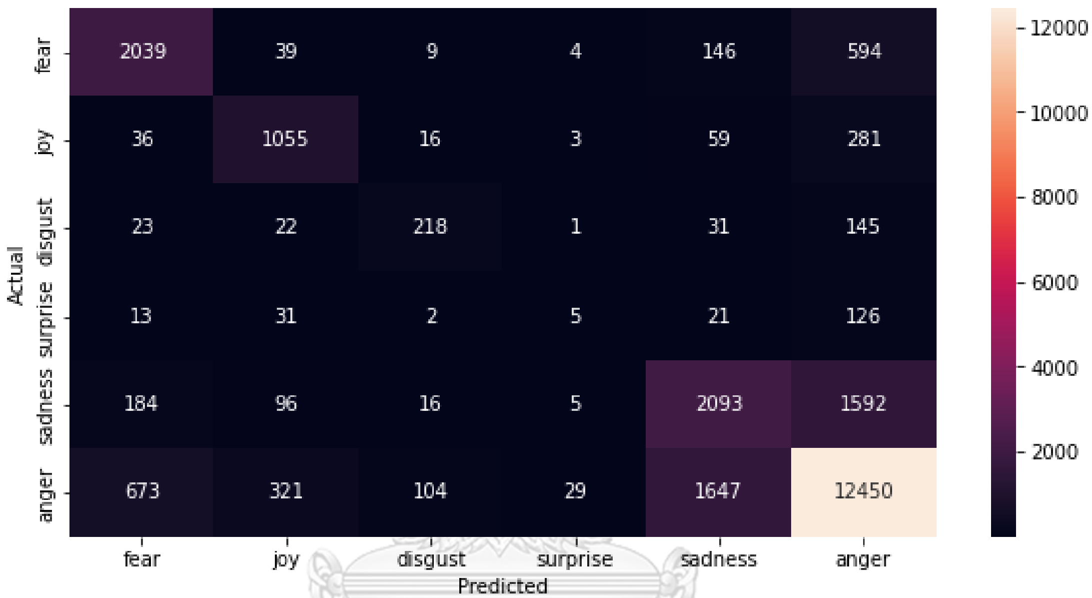


Figure 28 Multi-layer Perceptron Model Performance on Confusion Matrix

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Figure 28 is evaluation results from confusion matrix, model predict sample well on high sample classes consists of anger, sadness, fear and joy but model perform worse on low sample classes consists of disgust and surprise.

Model perform the best on anger class which has highest sample on dataset by predicting right samples equal 81.77% and the wrong equal 18.22% while the worse on surprise class which has lowest sample on dataset with 2.53% of right prediction and 97.47% of wrong prediction.

Validation result on RNN model

Rank	Optimizer	Batch Size	Epochs	average validation accuracy	Standard deviation
1	RMSprop	64	5	0.7592	0.0142
2	RMSprop	128	5	0.7569	0.0065
3	RMSprop	32	5	0.7543	0.0101
4	Adam	64	5	0.7463	0.0077
5	Adam	128	5	0.7458	0.002
6	RMSprop	32	10	0.7384	0.0162
7	RMSprop	64	10	0.735	0.017
8	RMSprop	128	10	0.731	0.0259
9	Adam	128	10	0.731	0.0076
10	Adam	64	10	0.73	0.0197
11	RMSprop	128	15	0.7289	0.0142
12	Adam	32	15	0.7278	0.0154
13	RMSprop	64	15	0.7256	0.0213
14	Adam	32	10	0.7243	0.009
15	RMSprop	32	15	0.7224	0.0154
16	Adam	64	15	0.7218	0.0248
17	Adam	128	15	0.7216	0.0211
18	Adam	32	5	0.7212	0.0435

Table 8 Grid search cross validation training result of RNN model

Result of the best model of RNN via grid search and cross validation training on the highest average validation accuracy (Table 8) is the model with RMSprop as optimizer, 64 of Batch sizes, 5 of epochs and 5-Fold cross validation.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

K-fold	Validation accuracy
1	0.7705
2	0.7689
3	0.7545
4	0.7689
5	0.7332
Average	0.7592

Table 9 5- fold cross validation result of the best RNN Model

Table 9 show validation accuracy result of each 5-fold cross validation and average accuracy result from the best hyper-parameters of RNN model.

<u>Evaluation result on RNN model</u>

```
              precision    recall  f1-score   support

        fear     0.7641    0.7358    0.7497      2831
         joy     0.6812    0.6779    0.6796      1450
     disgust     0.4834    0.5295    0.5054       440
    surprise     0.0000    0.0000    0.0000       198
     sadness     0.6776    0.3901    0.4951      3986
       anger     0.7989    0.9017    0.8472     15224

    accuracy                         0.7701     24129
   macro avg     0.5675    0.5392    0.5462     24129
weighted avg     0.7554    0.7701    0.7544     24129
```

Figure 29 RNN Model Performance on Classification Report

Figure 29 is evaluation results both overall and each class of emotion on test data on 4 metrics. The overall results are 77.01% of accuracy, 75.54% of precision, 77.01% of recall, and 75.44% of F1-score. Model perform the best on anger emotion on 3 metrics which is precision, recall, and F1-score with 79.89%, 90.17%, and 84.72%, respectively.

Figure 30 RNN model on Confusion Matrix

Figure 30 is evaluation results from confusion matrix, model predict sample well on high sample classes consists of anger, sadness, fear and joy but model perform worse on low sample classes consists of disgust and surprise.

Model perform the best on anger class which has highest sample on dataset by predicting right samples equal 90.17% and the wrong equal 9.83% while the worse on surprise class which has lowest sample on dataset with no right prediction on evaluation dataset.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Validation result on LSTM model

Rank	Optimizer	Batch Size	Epochs	average validation accuracy	Standard deviation
1	RMSprop	32	5	0.7837	0.0081
2	RMSprop	64	5	0.781	0.0067
3	RMSprop	128	5	0.7803	0.0038
4	Adam	64	5	0.7724	0.0039
5	RMSprop	64	10	0.7712	0.0126
6	Adam	32	5	0.7675	0.0052
7	RMSprop	32	10	0.763	0.0072
8	Adam	128	5	0.7612	0.0104
9	RMSprop	128	10	0.76	0.0205
10	RMSprop	128	15	0.7566	0.0035
11	Adam	128	10	0.7509	0.0093
12	RMSprop	64	15	0.7494	0.0072
13	Adam	64	10	0.7487	0.0055
14	RMSprop	32	15	0.748	0.013
15	Adam	128	15	0.7316	0.0108
16	Adam	64	15	0.7314	0.0062
17	Adam	32	10	0.7311	0.0047
18	Adam	32	15	0.7194	0.0029

Table 10 Grid search cross validation training result of LSTM model

Result of the best model of LSTM via grid search and cross validation training on the highest average validation accuracy (Table 10) is the model with RMSprop as optimizer, 32 of Batch sizes, 5 of epochs and 5-Fold cross validation.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

K-fold	Validation accuracy
1	0.7840
2	0.7947
3	0.7697
4	0.7869
5	0.7832
Average	0.7837

Table 11 5- fold cross validation result of the best LSTM Model

Table 11 show validation accuracy result of each 5-fold cross validation and average accuracy result from the best hyper-parameters of LSTM model.

<u>Evaluation result on LSTM model</u>

	precision	recall	f1-score	support
fear	0.7625	0.7542	0.7583	2831
joy	0.8654	0.6076	0.7139	1450
disgust	0.6981	0.5045	0.5858	440
surprise	0.0000	0.0000	0.0000	198
sadness	0.6882	0.4767	0.5632	3986
anger	0.8069	0.9134	0.8569	15224
accuracy			0.7892	24129
macro avg	0.6369	0.5427	0.5797	24129
weighted avg	0.7770	0.7892	0.7762	24129

Figure 31 LSTM Model Performance on Classification Report

Figure 31 is evaluation results both overall and each class of emotion on test data on 4 metrics. The overall results are 78.92% of accuracy, 77.70% of precision, 78.92% of recall, and 77.62% of F1-score. Model perform the best on anger emotion on 3 metrics which is precision, recall, and F1-score with 80.69%, 91.34%, and 85.69%, respectively.

Figure 32 LSTM Model Performance on Confusion Matrix

Figure 32 is evaluation results from confusion matrix, model predict sample well on high sample classes consists of anger, sadness, fear and joy but model perform worse on low sample classes consists of disgust and surprise.

Model perform the best on anger class which has highest sample on dataset by predicting right samples equal 91.34% and the wrong equal 8.66% while the worse on surprise class which has lowest sample on dataset with no right prediction on evaluation dataset.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Validation result on BiLSTM model

Rank	Optimizer	Batch Size	Epochs	average validation accuracy	Standard deviation
1	RMSprop	32	5	0.7888	0.0048
2	RMSprop	128	5	0.7841	0.0033
3	RMSprop	64	5	0.7822	0.0053
4	RMSprop	64	10	0.7743	0.0054
5	RMSprop	32	10	0.7711	0.0037
6	Adam	32	5	0.7692	0.0058
7	Adam	128	5	0.7682	0.0033
8	RMSprop	128	10	0.7674	0.0132
9	Adam	64	5	0.7653	0.0048
10	RMSprop	64	15	0.7594	0.0062
11	RMSprop	128	15	0.7587	0.0124
12	RMSprop	32	15	0.7584	0.0041
13	Adam	64	10	0.7479	0.0043
14	Adam	128	10	0.7442	0.0067
15	Adam	32	10	0.7379	0.0087
16	Adam	128	15	0.7373	0.0091
17	Adam	64	15	0.7363	0.0062
18	Adam	32	15	0.7228	0.0032

Table 12 Grid search cross validation training result of BiLSTM model

Result of the best model of BiLSTM via grid search and cross validation training on the highest average validation accuracy (Table 12) is the model with RMSprop as optimizer, 32 of Batch sizes, 5 of epochs and 5-Fold cross validation.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

K-fold	Validation accuracy
1	0.7827
2	0.7933
3	0.7849
4	0.7953
5	0.7881
Average	0.7888

Table 13 5- fold cross validation result of the best BiLSTM Model

Table 13 show validation accuracy result of each 5-fold cross validation and average accuracy result from the best hyper-parameters of BiLSTM model.

<u>Evaluation result on BiLSTM model</u>

	precision	recall	f1-score	support
fear	0.7821	0.7467	0.7640	2831
joy	0.7373	0.7124	0.7247	1450
disgust	0.7338	0.5136	0.6043	440
surprise	0.0000	0.0000	0.0000	198
sadness	0.6637	0.5095	0.5765	3986
anger	0.8199	0.8971	0.8567	15224
accuracy			0.7900	24129
macro avg	0.6228	0.5632	0.5877	24129
weighted avg	0.7764	0.7900	0.7800	24129

Figure 33 BiLSTM Model Performance on Classification Report

Figure 33 is evaluation results both overall and each class of emotion on test data on 4 metrics. The overall results are 79% of accuracy, 77.64% of precision, 79% of recall, and 78% of F1-score. Model perform the best on anger emotion on 3 metrics which is precision, recall, and F1-score with 81.99%, 89.71%, and 85.67%, respectively.

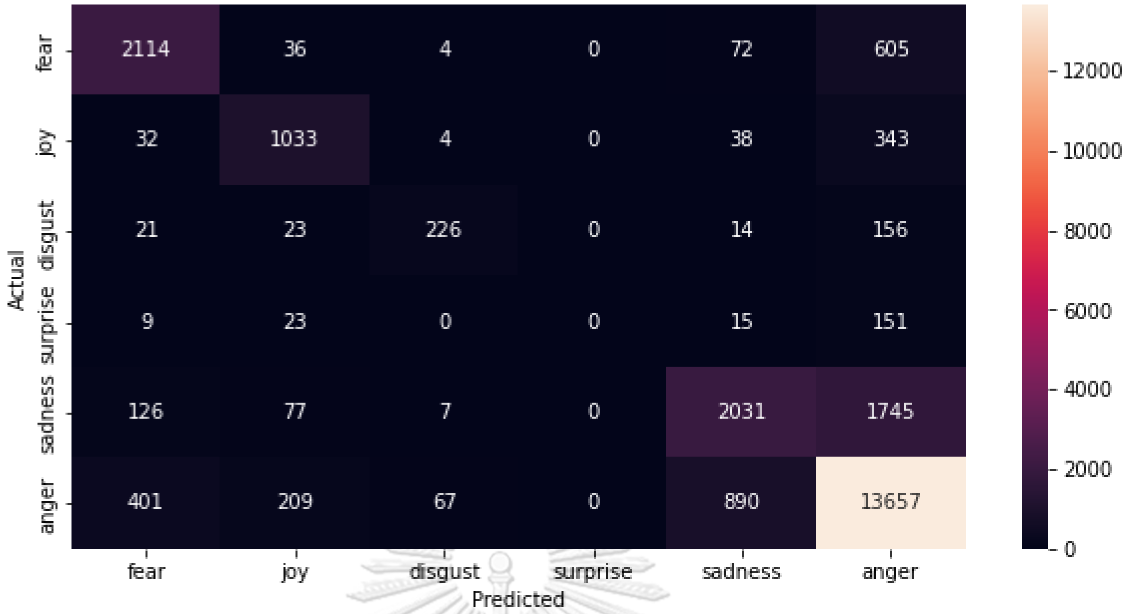


Figure 34 BiLSTM Model Performance on Confusion Matrix

Figure 34 is evaluation results from confusion matrix, model predict sample well on high sample classes consists of anger, sadness, fear and joy but model perform worse on low sample classes consists of disgust and surprise.

Model perform the best on anger class which has highest sample on dataset by predicting right samples equal 89.71% and the wrong equal 10.29% while the worse on surprise class which has lowest sample on dataset with no right prediction on evaluation dataset.

Validation result on GRU model

Rank	Optimizer	Batch Size	Epochs	average validation accuracy	Standard deviation
1	RMSprop	64	5	0.788	0.0038
2	RMSprop	32	5	0.7858	0.003
3	RMSprop	128	5	0.7803	0.004
4	Adam	32	5	0.7692	0.0052
5	Adam	64	5	0.7691	0.0041
6	Adam	128	5	0.7665	0.0058
7	RMSprop	64	10	0.7661	0.0045
8	RMSprop	128	10	0.7574	0.0095
9	RMSprop	32	10	0.7564	0.0076
10	RMSprop	32	15	0.753	0.0083
11	RMSprop	64	15	0.7458	0.0102
12	RMSprop	128	15	0.7426	0.015
13	Adam	128	10	0.7407	0.0038
14	Adam	64	10	0.7341	0.0056
15	Adam	128	15	0.7273	0.0047
16	Adam	32	10	0.7266	0.0085
17	Adam	64	15	0.7182	0.0081
18	Adam	32	15	0.7124	0.0048

Table 14 Grid search cross validation training result of GRU model

Result of the best model of GRU via grid search and cross validation training on the highest average validation accuracy (Table 14) is the model with RMSprop as optimizer, 64 of Batch sizes, 5 of epochs and 5-Fold cross validation.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

K-fold	Validation accuracy
1	0.7872
2	0.7938
3	0.7828
4	0.7906
5	0.7858
Average	0.7880

Table 15 5- fold cross validation result of the best GRU Model

Table 15 show validation accuracy result of each 5-fold cross validation and average accuracy result from the best hyper-parameters of GRU model.

Evaluation result on GRU model

```
              precision    recall  f1-score   support

        fear     0.7826    0.7298    0.7553      2831
         joy     0.8559    0.6062    0.7097      1450
     disgust     0.7675    0.3977    0.5240       440
    surprise     0.0000    0.0000    0.0000       198
     sadness     0.6502    0.5203    0.5780      3986
       anger     0.8101    0.9069    0.8558     15224

    accuracy                         0.7875     24129
   macro avg     0.6444    0.5268    0.5705     24129
weighted avg     0.7758    0.7875    0.7762     24129
```

Figure 35 GRU Model Performance on Classification Report

Figure 35 is evaluation results both overall and each class of emotion on test data on 4 metrics. The overall results are 78.75% of accuracy, 77.58% of precision, 78.75% of recall, and 77.62% of F1-score. Model perform the best on anger emotion on 3 metrics which is precision, recall, and F1-score with 81.01%, 90.69%, and 85.58%, respectively.

Figure 36 GRU Performance on Confusion Matrix

Figure 36 is evaluation results from confusion matrix, model predict sample well on high sample classes consists of anger, sadness, fear and joy but model perform worse on low sample classes consists of disgust and surprise.

Model perform the best on anger class which has highest sample on dataset by predicting right samples equal 90.69% and the wrong equal 9.31% while the worse on surprise class which has lowest sample on dataset with no right prediction on evaluation dataset.

Overall, every model predicted test sample are perform well on high classes sample well. The problem is on the low classes especially on surprise class that only Multi-layer Perceptron can predict on some sample but RNN-based model cannot predict the right samples even the model performances are better than it.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

5.2 Comparing Models

The performance on five neural network models are reported in Table I. The performance metrics are accuracy, F1 score, precision, and recall on test dataset.

Model	Precision	Recall	F1-Score	Accuracy
Multi-layer perceptron	0.7366	0.7402	0.7380	0.7402
RNN	0.7554	0.7701	0.7544	0.7701
LSTM	0.7770	0.7892	0.7762	0.7892
Bidirectional LSTM	0.7764	0.7900	0.7800	0.7900
GRU	0.7758	0.7875	0.7762	0.7875

Table 16 Model Performance

Table 16 show the best performance in our study is BiLSTM which outperforms other models. LSTM perform the best on precision with 77.7% while BiLSTM achieved the highest performance with 79% on Recall, 78% on F1-Score and 79% on accuracy. However, BiLSTM and LSTM performance are really close, leading to confirmation that LSTM is the most suitable model on this dataset. In addition, RNN based models perform better than the simple neural network.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter VI
DISCUSSION

6.1 Finding Summary

The best performance in our study is the LSTM and BiLSTM model which outperforms other models on these metrics. LSTM provides the highest precision at 77.7% while BiLSTM has the best on recall at 79%, F1-score at 78% and accuracy at 79%. RNN based models perform better than the simple neural network.

6.1.1 Emotion movement during COVID-19

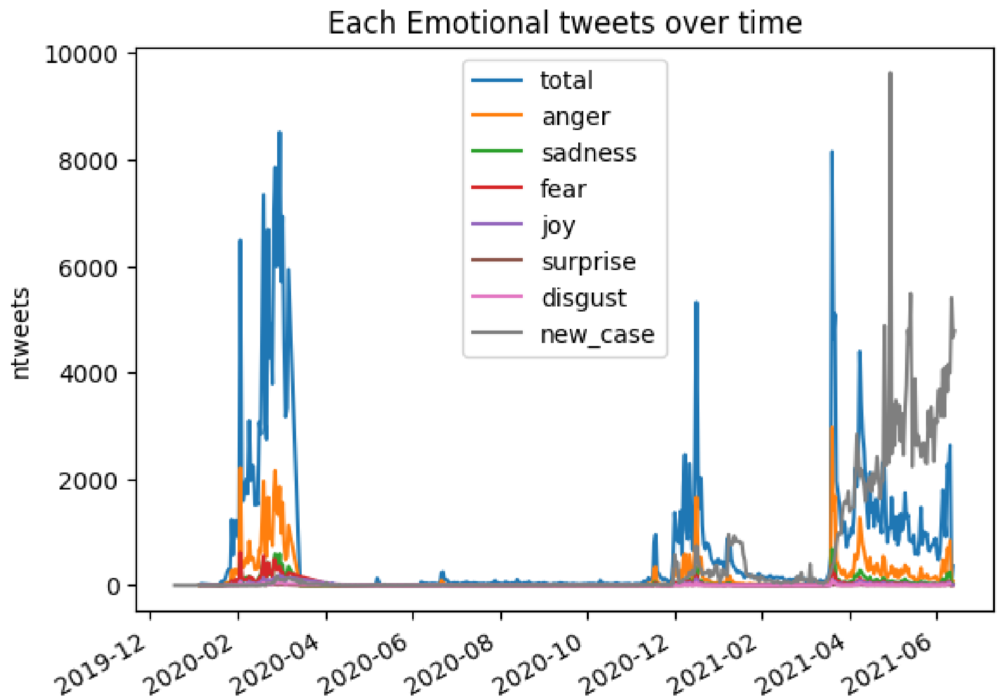


Figure 37 Emotion movement during COVID-19 pandemic

Emotion during the COVID-19 pandemic of Thai language Twitter users was at its peak 3 times in a period. The first peak in the number of emotional messages is around February to April 2020. It was during this first period of this virus spreading that there is news about running out of face masks in Thailand [19], [20]. Following between December 2020 and January 2021, users feel angry with the finding of huge clusters in Shrimp Samutsakhon Market and Lumpini Boxing Stadium [21],[22]. Lastly, the third peak of emotional messages from users around April to May 2021 was during the sharp rise in the number of new cases and deaths mainly in Bangkok and from prison [23],[24].

6.1.2 Anger emotion with official records of infection cases

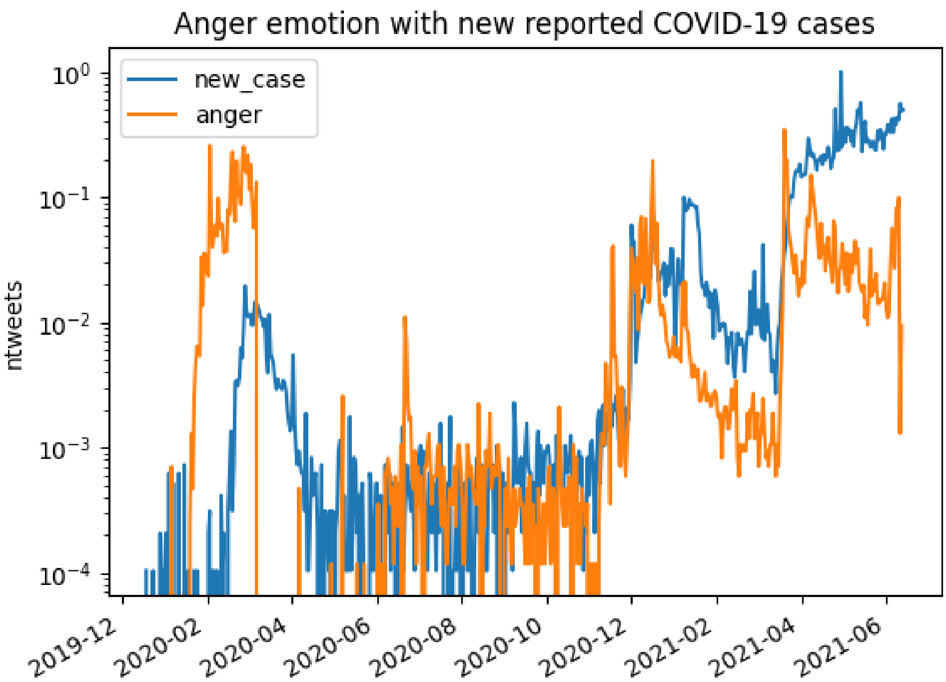


Figure 38 Anger emotion related with COVID-19 new patients

Plotting Anger emotion during COVID-19 infection trend compared with daily reported cases using normalizing data into the same scale and plot in log scale. The figure shows most of the anger feelings are fluctuating in the same period except the first and last quarter of the entire period. Relationship of amount of anger messages and new COVID-19 cases can describe by correlation. Their correlation of logarithm scale after normalization of both data equal 68%. Moreover, the correlation from adjusted series by smoothing both series with moving average in difference period are 1 day-window smoothing equal 67%, 3 days-window smoothing equal 73%, 5 days-window smoothing equal 73% and 7 days-window smoothing equal 71%. Correlation value can tell medium to strong relation in same direction of both, meaning amount of anger message is related with new COVID-19 cases.

6.1.3 Fear emotion with official records of vaccinated people

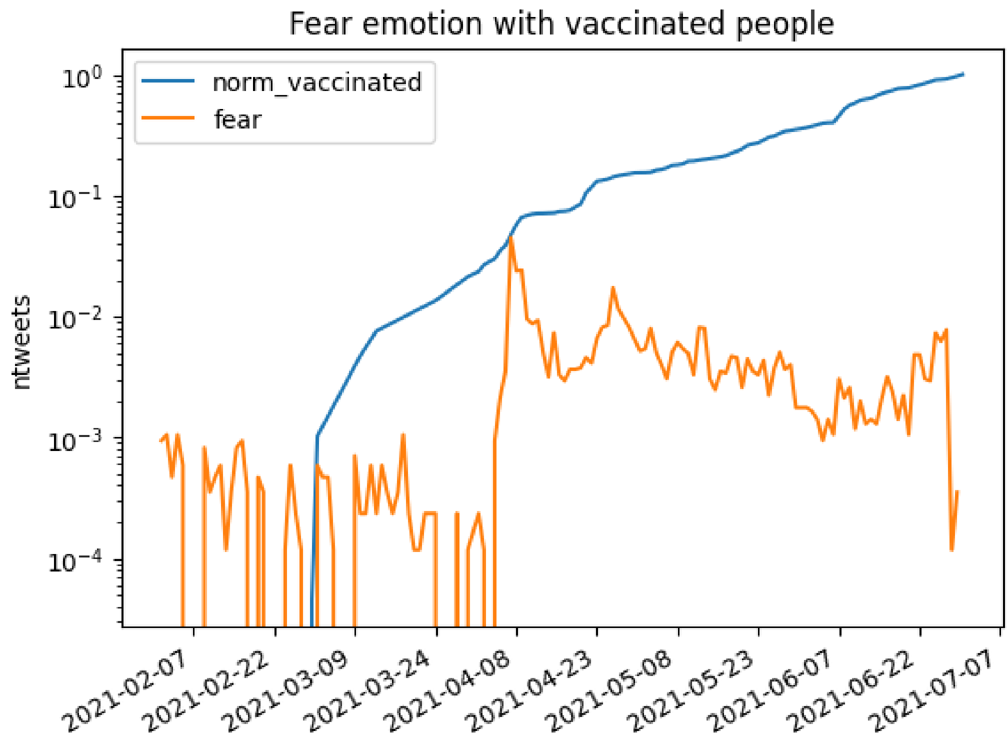


Figure 39 Fear emotion related with vaccinated people

Plotting fear emotion and vaccination trend by normalizing data into the same scale and plot in log scale. The figure shows most fear feelings are fluctuating up and down between February to April 2021. After that, the number jumped up in a short period and it was likely to be a downward trend in the late of April during the rise in the progress of vaccination people. Since the graph show a small relationship between fear messages and vaccination people, amounts of both data are too little in entire period. The relationship of them is hard to describe in correlation values.

6.2 Future Work

Collecting more tweets messages to get a broader scope of data related to the COVID-19 infection situation. Trying to utilize other knowledge bases to classify datasets to gain higher performance than currently studied. Moreover, upgrading the computing level of the training platform and using native judgement or better model performance to annotate dataset in the future.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

REFERENCES

1. Le, Q.V., et al. *On optimization methods for deep learning*. in *ICML*. 2011.
2. Sun, R.J.A., *Optimization for deep learning: theory and algorithms.* 2019. **abs/1912.08957**.
3. Mukkamala, M.C. and M. Hein. *Variants of rmsprop and adagrad with logarithmic regret bounds*. in *International Conference on Machine Learning*. 2017. PMLR.
4. Kingma, D.P. and J.J.a.p.a. Ba, *Adam: A method for stochastic optimization.* 2014.
5. Fine, T.L., *Feedforward neural network methodology*. 2006: Springer Science & Business Media.
6. Zaremba, W., I. Sutskever, and O.J.a.p.a. Vinyals, *Recurrent neural network regularization.* 2014.
7. Hochreiter, S. and J.J.N.c. Schmidhuber, *Long short-term memory.* 1997. **9**(8): p. 1735-1780.
8. Chung, J., et al., *Empirical evaluation of gated recurrent neural networks on sequence modeling.* 2014.
9. Vaswani, A., et al. *Attention is all you need*. in *Advances in neural information processing systems*. 2017.
10. Devlin, J., et al., *Bert: Pre-training of deep bidirectional transformers for language understanding.* 2018.
11. Lowphansirikul, L., et al., *WangchanBERTa: Pretraining transformer-based Thai Language Models.* 2021.
12. Al-Laith, A. and M.J.I. Alenezi, *Monitoring people's emotions and symptoms from Arabic tweets during the COVID-19 pandemic.* 2021. **12**(2): p. 86.
13. Garcia, K. and L.J.A.S.C. Berton, *Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA.* 2021. **101**: p. 107057.
14. Kausar, M.A., A. Soosaimanickam, and M. Nasar, *Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak.*
15. Mathur, A., P. Kubde, and S. Vaidya. *Emotional Analysis using Twitter Data during Pandemic Situation: COVID-19*. in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. 2020. IEEE.
16. Pasupa, K., T.S.N.J.S.C. Ayutthaya, and Society, *Thai sentiment analysis with deep learning techniques: A comparative study based on word embedding, POS-tag, and sentic features.* 2019. **50**: p. 101615.
17. Pasupa, K. and T.S.N.J.C.C. Ayutthaya, *Hybrid deep learning models for thai sentiment analysis.* 2021: p. 1-27.
18. Chormai, P., P. Prasertsom, and A.J.a.p.a. Rutherford, *AttaCut: A Fast and Accurate Neural Thai Word Segmenter.* 2019.
19. The-Standard, *January 12, 2020 – found the first COVID-19 patient in Thailand.* 2021.
20. BBC-Thai-News, *Coronavirus: Situation of medical masks in Thailand, Ran out of it-Free- Complimentary-Recycle.* 2021.
21. BBC-Thai-News, *Covid-19: Closed the shrimp market in Mahachai, Samut Sakhon province, after three days, found infected patients from the local area in*

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Thailand. 2020.
22. Thairath-News, *Throwback to the time found a large number of people who infected from the boxing area in Thailand.* 2020.
23. The-Bangkok-Insight, *Covid-19 situation update on 4 May 2021.* 2020.
24. BBC-Thai-News, *COVID-19: What is causing the coronavirus outbreak in prisons? After the number of infected people surpassed 10,000.* 2020.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

VITA

NAME	Miss Chotika Imvimol
DATE OF BIRTH	12 March 2538
PLACE OF BIRTH	Phichit
INSTITUTIONS ATTENDED	Chulalongkorn University
HOME ADDRESS	72/16 Soi Ngamwongwan 28 yak 1-2, Ngamwongwan Road, Thung Song Hong Sub-District, Lak Si District, Bangkok 10210
PUBLICATION	The 25th International Computer Science and Engineering Conference 2021 (ICSEC2021)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University