computing@computingonline.net
www.computingonline.net

Print ISSN 1727-6209
On-line ISSN 2312-5381
International Journal of Computing

THE DEVELOPMENT OF TRUST MATRIX FOR RECOGNIZING RELIABLE CONTENT IN SOCIAL MEDIA

Zurina Saaya, Tham Weng Hong

Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka,
zurina@utem.edu.my, b031510130@student.utem.edu.my, http://ftmk.utem.edu.my

Paper history:
Received 6 November 2018
Received in revised form 27 December 2018
Accepted 10 January 2019
Available online 31 March 2019

Keywords:
Rule-based;
Trust matrix;
Social media.

Abstract: Social media has emerged as a popular platform for users to share information about real-world events, particularly during disaster emergencies. However, disaster managers often having problem to gather an accurate information of the current situation since not all observations are made by reliable users. In this project, we address this problem by developing a trust matrix framework to identify the trustworthiness of the information shared on social media. Specifically, this project is focusing on flood disaster management and information shared on Twitter platform. The first objective of this project is to create text corpus for statistical and keyword analysis of a flood event shared on social media. Second objective is to develop a trust matrix for the flood events from social media. Third objective is to evaluate trust matrix for flood event using crowd-sourced data. The evaluation of trust matrix is done using real flood event dataset which are gathered between June and July 2018. To establishing a ground-truth for trust matrix framework, each data is mapped with actual flood event from news portal.

1. INTRODUCTION

In its early appearance social media is used by most people simply as a means of broadcasting personal life and sharing information [1, 2]. Nowadays, social media has emerged as a popular platform for users to share information about real-world events, particularly during disaster [3, 22]. Early information from people on the scene of disaster often conveys timely and actionable information, which is greatly valuable for official authorities respond to emergencies occur during disaster event. However not all of the information is in good quality and related to the event, it may be fake, incorrect or noisy. Such false and incorrect information can lead to chaos and panic among people on the ground.

Furthermore, harnessing reliable information on disaster events from social media is very challenging, Twitter for example only allows its short textual messages called tweets of up to 140 characters. It thus encourages its users to use short form to reduce the length of their messages. In some condition, the large volume of emergencies information that is spread by many users also overwhelming. Therefore, disaster managers often having problem to gather an accurate information of the current situation since not all observations are made by validated users or reliable observers. Since early year of online application specifically for information sharing reliability is a major concern [4, 5]. There is a need for a high-level and abstract way of identifying and managing trustworthiness of information on online media, which can be easily integrated into applications and used in any domain.

This paper is focusing on identifying reliability of the information shared on the social media particularly related to flood event. Specifically, the objectives of this paper as follows. The first objective is to create text corpus for statistical and keyword analysis of a flood event shared on social media. Second objective is to develop a trust matrix for the flood events from social media. Third objective is to evaluate trust matrix for flood event using crowd-sourced data. The main contribution of this project is the text corpus and text analysis procedures for large unstructured dataset which

related to flood. These text corpus and text analysis procedures are used to develop the trust matrix that can identify trustworthiness of information shared on social media.

The paper is structured as follows. Section 2 presents in more details on research background. Section 3 describes methodology used in this research. Section 4 discuss the details of the trust matrix. Section 5 is regarding the results and discussion the results. Section 6 concludes the paper.

2. RESEARCH BACKGROUND

Although social media are used by many people as a way of sharing personal life information and day to day activities the insight provided by social media data goes far beyond. Researchers have been using social media to study relationship between happiness and mobility patterns [13], tourist origins and attractions [14] and disease outbreaks [15, 16].

In recent years, online social media has become an important role in real world events, especially during emergency events. Muralidharan at al. [18] for example, use data from Twitter and Facebook to study the usage of social media for disaster monitoring during the 2010 earthquake in Haiti. Guan and Chen [2], 10] develop a measurement tool to quantify the evolution of disasters in in term of temporal–spatial patterns based on Twitter activities during emergency events. In another research, Lu and Brelsford [17] investigate, the changes in human social behavior in response to extreme disaster like earthquakes, tsunamis as a mean to understand suitable emergency response and recovery.

Beside text-based data from social media, Sun et al. [19] use the geotagged Flickr images as supporting features remote sensing in flood disaster map. Fohringer et al. [21] also utilize quantitative data that are gathered from photos that are shared by eyewitnesses in social media posts to support flood inundation mapping. Jongman et al. [20] study how the tweets from Twitter that are related to flood event and satellite observations of water coverage can be used to support early disaster response. Basically flood event were mentioned several days earlier on Twitter then reported to responsible organizations. However, they suggested that the pre-processing of social media data needs to be improved for operational use.

In another perspective, Sloman and Grandison [4] concern about the reliability of the information spread during disaster events. In some condition, it can be used for effective disaster alert or to spread rumours and fake news. For example, Gupta et. al, identified more than 10000 tweets which contained fake images that were spread over Twitter. They show that not all of the information shared by people are good quality with respect to the event, like it might be a fake or hoax [6].

Sakaki et al. used tweets as social sensors to detect earthquake events. In other words, the posted tweets are acted as sensor of the event [7]. For instance, user makes a tweet about a flood occurrence, then it can be considered that he or she act as a “flood sensor”, and classify it into true news. The tweets are classified using probabilistic spatio-temporal model.

There is a similar method is built but imperfection which is Alternative Trust Metric. It only focuses on the textual content of messages to determine if an event is spam or not spam. It is a method for reducing the fake news based on the number of users reporting the related event [8]. This paper aims to develop improve and perfect system to filter and reduce the fake news on online social media. The researchers consider each Twitter user as a sensor, and set a problem to detect an event based on sensory observations. Unlike Alternative Trust Metric, our version of trust matrix is built based on keywords analysis and context analysis to identify the trustworthy of the information.

Specifically, in this research, we develop a trust matrix diagram as an abstract overview to identify reliability of shared information on social media in particular related to flood event disaster. In this case, Trust is referred as reliable, relevance, and believable. The trust matrix is a tool which is able to filter the unreliable data based on based on two types of components, namely keyword and context.

3. METHODOLOGY

In this section, we discuss our research methodology in detail. First, we describe the methodology of collecting data from Twitter, followed by the various techniques that is used to developed the trust matrix. Fig. 1, shows the steps involved in the development of Trust Matrix.

Data collection is the first step of trust matrix development. In this step, data about flood event are extracted from Twitter. Twitter is a famous and widely used social network. It produces more than 200 million tweets per day. Even though it consists of millions of tweets the extraction process become easier because of Twitter Public Streaming API. In this research, we use Python Tweepy package of Twitter API for data collection. During this step, we also apply pre-filtering where we only keep tweets that contain keyword ‘flood’. After that, the tweets that are unrelated such as retweet is removed from the collected dataset.

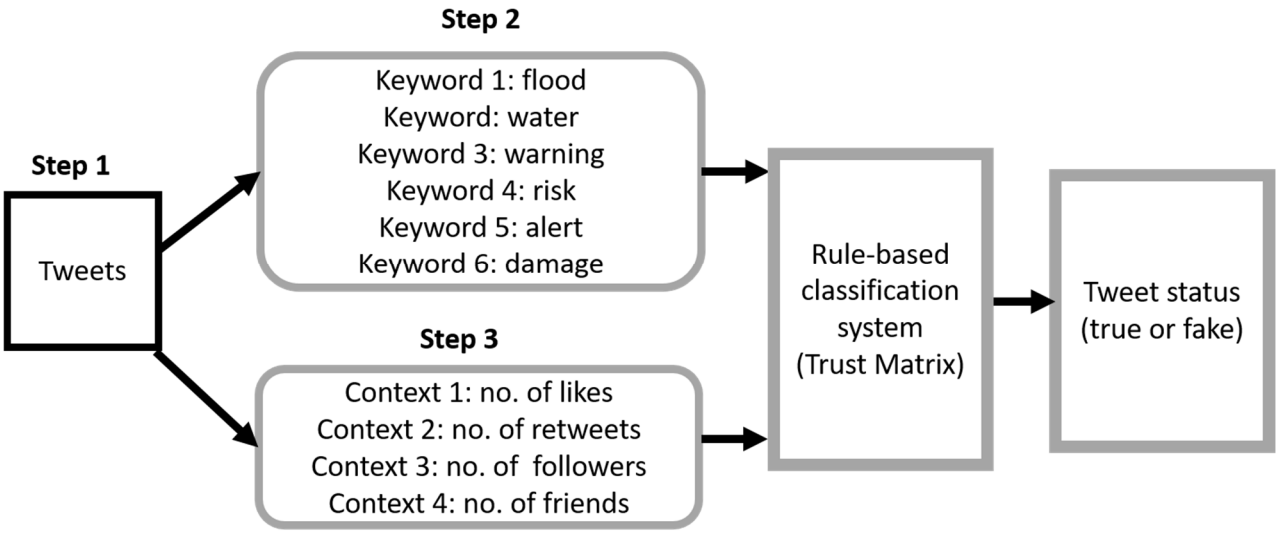

Figure 1 – Development of Trust Matrix

The next step is to identify significant keywords using keyword analysis. Keyword analysis is used to study the pattern of keywords consist in each tweet. The text analysis is carried out manually to increase the precision of the details of every single collected data. Although it is a time-consuming task, it can produce more accurate output for this research. Based on 40,000 sample data, we manage to identify five most frequent keyword or keywords combination as shown in Table 1. These keywords will be used as keyword component in the trust matrix.

Table 1. List of Keywords Combination

Elements	Count numbers
Flood + water	1497
flood + warning	28189
flood + risk	2097
flood + alert	4666
flood + damage	4666

Next, we identify context information that is relevant for flood event and trustworthiness of the tweets. The context elements are also identified based on previous work as discusses in the following paragraphs.

3.1 NUMBER OF FOLLOWERS

Every Twitter user can make a tweet. It can be a fake since Twitter does not require a proof of identity for tweets. The followers would not know the tweet about an event whether it is true. Therefore, the user will like to follow a verified person who have been manually verified by Twitter [9]. Verification is only initiated by Twitter. A verified user to be trusted because they must use their real-life identity and be well-known enough by the general public for Twitter such as Barack Obama and Ellen DeGeneres.

3.2 NUMBER OF FRIENDS

As a Twitter user, following the Twitter users in order to collect more information from them. The Twitter users have their own registered location. The news come from everywhere and the information will be obtained as following the others Twitter users. For example, user lived in Melaka, Malaysia but he or she heard a news about flood event in United State through Twitter since he or she had follow a trustable Twitter user from United State. The larger the number of friends, the more information getting from them.

3.3 NUMBER OF RETWEET

Retweet is a way to diffuse the information in Twitter. It is simple but powerful to disseminate the useful information [10]. At most of the situation, the greater the number of followers will be getting the greater the number of retweet. Retweet a tweet is one of the fast way to let others to know about the information. For example, a trustable Twitter user upload a tweet about the flood event will happen in 3 hours later. Retweet this news to let the friends and followers get through it. Through word of mouth, the news will reverberate very quickly and all the people can take their time to prevent the flood event. So, the greater the number of retweet, the more reliable the spread information. A threshold for number of retweet have to find out in order for estimating how many retweets is considered as trustable tweet.

3.4 NUMBER OF SIMILAR TWEETS

As mentioned before, multiple observations able to make information become more reliable. Hence, counting the similar tweets but from different users is carry out. Firstly, the threshold is set for number of similar tweets by ourselves after observing the count number for similar tweets. If the number of similar tweets is greater than its threshold, means that tweet is reliable. For example, three different users of Twitter upload the same real-time flood event with threshold of three and these tweets could be considered as reliable. This approach has also been applied by Eilander et. al. to develop a decision support tool for floods disaster [11].

Based on sample data 40000 sample data that already being tag as reliable tweet we identify the context element and the threshold for each element. Basically this sample data is used as training dataset to better understand the characteristics of the model to be develop. The last step is to design the abstract view for trust matrix based on main two components namely, keywords and contexts information. Context elements that are significant for flood trust matrix is shown in Table 2.

Table 2. List of Context Information

Elements	Threshold
Number of Likes	1
Number of retweet	1
Number of followers	30000
Number of friends	900
Number of similar tweet	3

There are many ways to confirm the reliability of certain information. Vosoughi, et. al, for example use fact-checking organizations to identify true or false online news [23]. As in our research, to establish a ground-truth for trust matrix framework, each data is mapped with actual flood event from Google News search. By using Google API we are able to extract flood event information using specific keywords. Specifically, each tweet is labelled whether it is true or false based on its exact situation gathered from the online news portal. For example, keywords "Alabama Flooding 6th June 2018" to find out whether Alabama suffer in flooding specified data. If the location of tweet is mapped with actual flood event location, the tweet is verified and considered as reliable and use as ground truth.

4. TRUST MATRIX

Trust Matrix is a matrix diagram used to recognize reliable tweets. Specifically, it consists of list of items and the existence of each item in each record. It is used to identify the presence of relationships between two or more items. Basically, it is a simple tool that allows us to analyze a relatively complex situation in a simple straightforward way to understand complex causal relationship. In other words, matrix diagram can be used in situation where we want to identify and assess the strength of relationships between a number of items. It is mainly useful for investigating the relationships between a set of vague and un-measurable items with a set of precise and measurable items. The component in Trust Matrix is can be used as a set of rules in a rule based system as a way to classify the trustworthiness of a tweet. As depicted in Table 3, the Trust Matrix consists of five rows vs. four columns. Rows are to represent keyword element while column are for context information.

Table 3. Matrix Diagram of Trust Matrix

		contextMatrix			
		No. of followers + no. of likes	No. of followers + no. of retweets	No. of followers + no. of friends	No. of followers + no. of similar tweets
keywordMatrix	Flood + water	/	/	/	/
	Flood + warning	/	/	/	/
	Flood + risk	/	/	/	/
	Flood + alert	/	/	/	/
	Flood + damage	/	/	/	/

Rule-based system is a simplest form of Artificial Intelligence. It represents the knowledge in terms of a set of rules for which to tell what to do or what to judge in non-identical situation [12]. A rule-based system can be created by a set of rules easily to classify the data.

Rule-based system consists of 3 important elements which are facts, rules, and terminator:

4.1 FACTS

A set of facts can be seen as a set of collection of data. The collected data should be anything relevant to the beginning state of the system.

4.2 RULES

A set of rules need to be created to determine and classify the facts. This contains all actions that should be taken within the specific scope of problem. Rules provide description of how to solve a

problem. A rule relates the facts in IF part (condition) to specific action or consequent in THEN part. Rules are relatively easy to create and understand. The system should contain only the relevant rules but not the irrelevant rules in order to ensure the well performance of the system (Irrelevant rules will affect the performance of the system). Simple structure of a rule as follows:

IF <condition TRUE>
THEN <consequent>.

4.3 TERMINATOR

This is a condition to determine that a solution has been found or not. It is necessary to terminate the rule-based system to avoid the infinite loops.

Rules are expressed as a set of IF-THEN statements. It can be applied for a large kind of problems but the area should not be large. For example, flood event is a bigger problem that always concerned by society could perform in a similar way when it meets the same facts (collection of data and conditions). Hence, the system need to be trained to make it become expert system which is capable to function with better zero-error. In this research, the rule-based is implemented to classify the trustworthy data by using a set of IF-THEN rules, but in specific area. Rule-based system may make the solution inefficient if the scope and range is wide.

5. RESULT AND DISCUSSION

We run experiments to evaluate how effective rule-based system as depicted in Fig. 1, can identify tweets that are reliable or not. Results are measured based of data accuracy. Specifically, accuracy is much easier to define. The accuracy of an experiment is how close the final result is to the correct or accepted value. The closer it is, the more accurate the experiment. The accuracy can be improved through the experimental method if each single measurement is made more accurate. Thus, for this experiment we use 30,000 pre-filtered tweets (tweet with keyword flood) collected every day in June 2018 (30 days) as test data set. These tweets then evaluated based on the set of rules or model as mentioned in the trust matrix. The system is able to identify the existence of the significant keywords and determine the context information and finally classify the tweet as true or fake. The accuracy of the final result is calculated against ground truth data which is collected from Google News API. For example, if tweet *id_01* is identified as reliable and the ground truth of tweet *id_1* also labelled as reliable then this tweet is counted as accuracy = 1.

Result from evaluation is depicted in Fig. 2. The accuracy of displayed on daily basis. The average accuracy for a month is about 75%. As we can see from the graph, accuracy for every day in June 2018. Our specific task is to only identify the patterns of keywords and context that imply the trustworthiness of tweets particularly related to flood event. However, in our experiment we can only provide historic instances of flood disaster to train the rule-based model. As is it built upon the seen instances, a model could easily bias towards specific keywords or context associated with events as seen in the experiments.

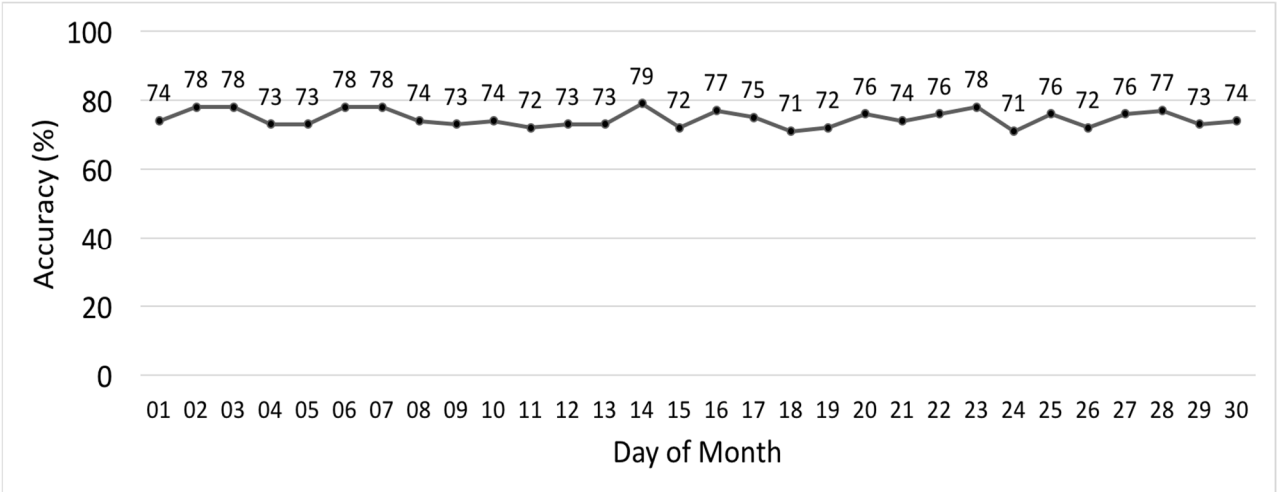

Figure 2 – Accuracy results

6. CONCLUSION

In this work, we focused on extracting relevant keywords and useful context information from Twitter during times of crisis especially in flood event to assist in better handling of critical disaster situations. The list of keywords can be used as text corpus for statistical and keyword analysis of a flood

event shared on social media. We explored the problem of filtering tweets, which are limited text messages on the popular social media platform, Twitter. We presented a rule-based framework in the form of Trust Matrix for filtering social media information into reliable and non-reliable messages. We highlighted an important component to identify reliability of certain content in social media is the context information such as location where the tweet come from, the followers for the person who send the tweets, the similar tweets that are also report the same event. Accuracy of the Trust Matrix is evaluated using ground truth data taken from online news portal.

Classification of text-based data is difficult without any additional features such as context information. In the future, we will consider providing knowledge from other resources to improve accuracy of our classification engine.

7. REFERENCES

[1] W. X. Zhao, J. Jiang, J. Weng, J. He, E. P. Lim, H. Yan, X. Li, "Comparing twitter and traditional media using topic models," *Proceedings of the European Conference on Information Retrieval*, Springer, Berlin, Heidelberg, April 2011, pp. 338-349.

[2] B. Krishnamurthy, P. Gill, M. Arlitt, "A few chirps about twitter," *Proceedings of the First ACM Workshop on Online Social Networks*, August 2008, pp. 19-24.

[3] X. Guan, C. Chen, "Using social media data to understand and assess disasters," *Natural Hazards*, vol. 74, no. 2, pp. 837-850, 2014.

[4] M. Sloman, T. Grandison, "A survey of trust in internet applications," *IEEE Communications Surveys & Tutorials*, vol. 3, issue 4, pp. 2-16, Fourth Quarter 2000.

[5] A. Jøsang, R. Ismail, C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision support systems*, vol. 43, issue 2, pp. 618-644, 2007.

[6] A. Gupta, H. Lamba, P. Kumaraguru, A. Joshi, "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy," *Proceedings of the 22nd ACM International Conference on World Wide Web*, May 2013, pp. 729-736.

[7] T. Sakaki, M. Okazaki, Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," *Proceedings of the 19th ACM International Conference on World Wide Web*, April 2010, pp. 851-860.

[8] T. Bodnar, C. Tucker, K. Hopkinson, S.G. Bilén, "Increasing the veracity of event detection on social media networks through user trust modeling," *Proceedings of the 2014 IEEE International Conference on Big Data*, October 2014, pp. 636-643.

[9] C. Lee, H. Kwak, H. Park, S. Moon, "Finding influentials based on the temporal order of information adoption in Twitter," *Proceedings of the 19th ACM International Conference on World Wide Web*, April 2010, pp. 1137-1138.

[10] B. Suh, L. Hong, P. Pirolli, E. H. Chi, "Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network," *Proceedings of the 2010 IEEE Second International Conference on Social Computing*, August 2010, pp. 177-184.

[11] D. Eilander, P. Trambauer, J. Wagemaker, A. Van Loenen, "Harvesting social media for generation of near real-time flood maps," *Procedia Engineering*, vol. 154, pp. 176-183, 2016.

[12] C. Grosan, A. Abraham, "Rule-based expert systems," *Intelligent Systems. Intelligent Systems Reference Library*, vol 17. Springer, Berlin, Heidelberg, 2011, vol. 17, pp. 149-185.

[13] M.R. Frank, L. Mitchell, P.S. Dodds, C.M. Danforth, "Happiness and the patterns of life: A study of geolocated tweets," *Scientific reports*, 3, 2625, 2013.

[14] S.A. Wood, A.D. Guerry, J.M. Silver, M. Lacayo, "Using social media to quantify nature-based tourism and recreation," *Scientific reports*, 3, 2976, 2013.

[15] C. Chew, G. Eysenbach, "Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak," *PloS one*, vol. 5, issue 11, e14118, 2010.

[16] R. Chunara, J.R. Andrews, J.S. Brownstein, "Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak," *The American Journal of Tropical Medicine and Hygiene*, vol. 86, issue 1, pp. 39-45, 2012.

[17] X. Lu, C. Brelsford, "Network structure and community evolution on twitter: human behavior change in response to the 2011 Japanese earthquake and tsunami," *Scientific reports*, 4, 6773, 2014.

[18] S. Muralidharan, L. Rasmussen, D. Patterson, J.H. Shin, "Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts," *Public Relations Review*, vol. 37, issue 2, pp. 175-177, 2011.

[19] D. Sun, S. Li, W. Zheng, A. Croitoru, A. Stefanidis, M. Goldberg, "Mapping floods due to Hurricane Sandy using NPP VIIRS and ATMS data and geotagged Flickr imagery," *International Journal of Digital Earth*, vol. 9, issue 5, pp. 427-441, 2016.

[20] B. Jongman, J. Wagemaker, B. R. Romero, E.C. de Perez, "Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals," *ISPRS International Journal of Geo-Information*, vol. 4, issue 4, pp. 2246-2266, 2015.

[21] J. Fohringer, D. Dransch, H. Kreibich, K. Schröter, "Social media as an information source for rapid flood inundation mapping," *Natural Hazards and Earth System Sciences*, vol. 15, issue 12, pp. 2725-2738, 2015.

[22] R. I. Ogie, R. J. Clarke, H. Forehead, P. Perez, "Crowdsourced social media data for disaster management: Lessons from the PetaJakarta. org project," *Computers, Environment and Urban Systems*, vol. 73, pp. 108-117, 2019.

[23] S. Vosoughi, D. Roy, S. Aral, "The spread of true and false news online," *Science*, vol. 359(6380), pp. 1146-1151, 2018.

Zurina Saaya *is a lecturer in Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM) where she involved with teaching computer networking topics. Her research focuses on technologies for information retrieval, data mining and recommender systems.*

Tham Weng Hong *is a student of Bachelor Degree in Computer Science majoring in computer networking. His research interest includes machine learning and computer networking.*