

Popularity of University based on Sentiment Analysis in Social Network Media

Chiu-Lin Koh, Su-Cheng Haw, Lay-Ki Soon

Abstract: *Since the emergence of the social media, many studies are conducted on social media to gain information on social media users. Among these studies are sentiment analysis which is an analysis of user sentiments and emotions towards an object, term, or event based on what they post. Sentiment analysis are often conducted on sites like Facebook and Twitter because of their huge number of users and popularity. This paper aims to create a GUI-based sentiment analysis application to find out popularity of universities based on Twitter user's sentiment. For this purpose, we firstly collected 600 tweets datasets, which is a mixture of 200 tweets each from Princeton University, Stanford University and University of Oxford for a period of 4 days (12/1/2018 to 15/1/2018). Second, the tweets were classified based on their sentiment into "positive", "neutral" and "negative" tweets. Finally, the results were being analyzed in terms of Precision, Recall and F1 score. These information will help universities to gather information of public sentiment towards their institution and allow them to recognize their strength and weakness. Universities can use that information to improve their public image if needed in the future.*

Index Terms: *sentiment analysis, crawler, university popularity, social network, opinion mining*

I. INTRODUCTION

In this day and age, communication over social network sites has become a norm where people share their emotions, opinion and sentiments online with their peer, friends, family and followers. Microblogging sites such as Twitter are well received by the public because of their ability to spread information through short sentences across large amounts of followers. The rise of such sites lead to emergence of technological research such as sentiment analysis. Sentiment analysis is useful in helping business and researchers by finding out how one's sentiment and opinion might affect others. This research can be applicable in helping a business improving its advertisement and public image, assist authorities in decision making during disaster management based on polarity of sentiments of social media users and political geography forecast.

On the other hand, universities are using social media sites to promote themselves among the public while at the same time students are using social media sites to ask others about

their experiences and opinion for universities. As such, universities can utilize sentiment analysis to have an overview of universities popularity in public and their public image. Universities can then work on improving their public image or use known public's positive opinion as an advantage in their promoting campaign.

Although there are many websites such as on global university ranking, Academic Ranking for World Universities (ARWU) or also known as Shanghai University Ranking, QS World University Rankings and Times Higher Education World University Ranking, these website does not analyze people opinion or perception towards the university. According to QS Ranking, "University and business school rankings represent a tried and tested outlet for prospective decision-making on their next study destination". These ranking list could help student to determine their choice of study, sort out which programs to register based on the most selective and prestigious program, and even to giving information on particular teaching and learning style, intellectual interests and talents that will pave the way toward finding the right place to study.

This paper focuses on performing sentiment analysis on university popularity through Twitter and creating an application that not only makes tweet crawling and sentiment analysis convenient to user but also allow user to download large number of tweets. Twitter is famous for their usage of tweets, short statements or microblog, to express opinions or spread information. We use tweet crawler to gather tweets from Twitter through the usage of Twitter API. After tweet crawling, Naïve Bayes classifier is used to perform sentiment analysis on gathered tweets after pre-processing the tweets and extracting the feature words into a vector. A pool of 600 tweets will be labelled by 3 students which a gold standard data set will be produced by selecting the sentiment of the majority count from the variation of the labelled data. The classifier will be trained using 300 of the 600 gold standard data set of tweets. The results are outputted into JSON and CSV files, convenient for the user to use for their own purposes. In addition, some statistical analysis will also be conducted. The evaluation will be conducted using precision, recall and F1 measure for each of the three sentiments, positive, negative and neutral against four testing sets. The four testing sets are selected from a pool of 1000 labelled tweets, 200 random tweets each for test set 1 to test set 3, while test set 4 include the entire pool of labelled tweets.

Revised Manuscript Received on August 18, 2019.

Chiu-Lin Koh, Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia.

Su-Cheng Haw, Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia.

Lay-Ki Soon, Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia.

II. LITERATURE REVIEW

Sentiment analysis (or

opinion mining, sentiment classification) is a field of study that involves computationally categorizing and identifying attitude, emotions, sentiments and opinions expressed from text, ranging from a word to a paragraph, about different entities such as products, brand name, services, individuals, events and organizations. It includes multiple fields of study such as Artificial Intelligence (AI), Natural Language Processing (NLP), Machine Learning (ML), computational linguistic, text analytics and Information Retrieval (IR). Sentiment analysis is researched mainly at three levels [1], i.e., the document level, sentence level, entity and aspect level.

Document Level: The objective at this level is to examine the overall sentiment of a document, for example, a product review or post event discussion. The analysis of this level assumes that the sentiments in this documents are directed to only one entity.

Sentence Level: The objective of analysis at this level is to examine the sentences and make a conclusion of the sentiment of the sentence, whether it is positive, neutral or negative. The sentiment 'neutral' is used when a sentence has been determined that there is no opinion involved, often use when spread of factual information is involved, thus, separating objective sentences from subjective sentences which expresses opinions, sentiments and emotions.

Entity and Aspect Level: The objective of analysis at this level is to examine the opinion from different aspect of itself instead of language constructs like documents and paragraphs. It is based on the idea that an opinion is consist of a sentiment, either positive or negative, and a target of the sentiment. This level of analysis provides a finer-grained analysis for each aspect found in the analysis.

Sentiment analysis is widely use over the years as the widespread use of the World Wide Web gave rise to social networks which become the huge focus for researchers of sentiments analysis. Twitter is one such social network or microblog site, currently ranked 12 globally in Alexa Traffic Ranking [2] and has a total number of monthly active Twitter users of 317 million as of 24 January 2017 [3], making it one of the most popular social networking sites around. As such, Twitter has been a target for many sentiment analysis researches on domains such as brand name, events [4], politics [5], election prediction [6], job recruitment [7] and so on.

One of the highlights of Twitter is the ability to retweet, reposting the same tweet of others on your wall. The ability to retweet allows one to spread information among one's followers with a single button click which in turn allow their followers to retweet among their followers. The process of retweeting a post that one have seen is fast and simple. A tweet can be retweeted to a network of people, if a tweet is retweeted, a tweet is expected to reach an average of 1000 users [8]. As such, the section below will be targeting sentiments of Twitter tweets.

High school students when searching for future tertiary educations institutes would visit the university sites, academic forums and social media for information regarding the university to aid them in making their choice. While

universities ranking sites by organizations like QS and Times gave a general overview and rating of universities' education standards, social media sites like Twitter is an important avenue for students to communicate with their peers or current university students for their opinions and experience in their university [9, 10]. They get know the course structures and environment better from senior students tweets. As such, popularity of a university among the students can be determined from tweets and interactions among students on Twitter through sentiment analysis.

This paper proposes to perform sentiment analysis on Twitter, and examined popularity of universities based on reactions and feedback from social media data gathered. Applications for data gathered from the analysis including aiding universities in their advertisements and how to improve their public image.

Using social media and microblogging sites such as Twitter as a source of information may pose some difficulties to the analysis. First of all, language usage in Twitter is informal, so there will be many statements with spelling errors, grammatical errors, slang usages and emoticons. Secondly, data gathered from Twitter may not be credible as statements from tweets are generally opinions and may be biased based on personal experience and views. Furthermore, Twitter has a limit of 140 characters per tweet, so users are forced to shorten their tweets with usages of phrase and abbreviations like FYI (for your info), IDK (I don't know), TY/ THX(Thank You/ Thanks) etc. Users may also end up spreading their statements across multiple tweet, thus taking out the context of their statement if their statements are examined as single tweets rather as continuous or connected tweets. Availability of data may also posed a problem if the data available is not sufficient to conduct an effective analysis with trusted results.

III. IMPLEMENTATION

A. The Architecture

The sentiment analysis program can be split into three major parts, Tweet Crawling phase, Sentiment Analysis phase, Display and Export phase. The program starts out by searching for tweets through tweet crawler and use the results obtained for sentiment analysis. After the analysis is done, the results will be shown and saved into a CSV file. Fig. 1 shows the overall flow of the program.

Fig. 2 shows the flowchart of the Tweet Crawling phase of the application, while Fig. 3(a) shows the sentiment analysis phase of the program, and Fig. 3(b) shows the Display and Export phase.

Subsequently, the pseudocode of some main functions such as start the tweet crawler, search and classify are depicted in Algorithm 1 to 3.

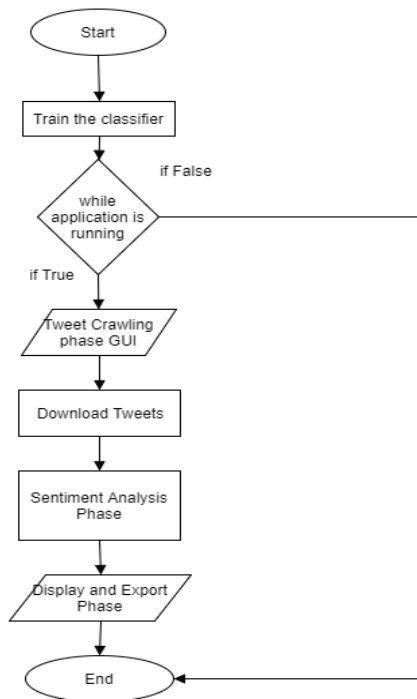


Fig. 1. Overall flow

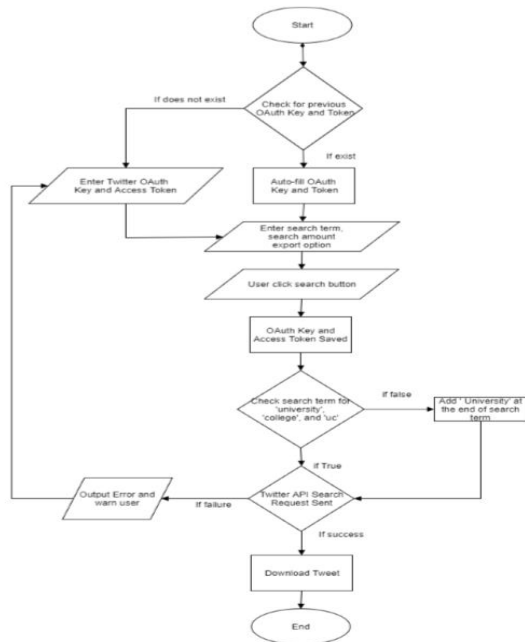


Fig. 2. Tweet crawling phase

B. The User Interface

The user will be greeted with the user interface, where there are six entry box for user to enter their Twitter OAuth consumer key and access token, search term and search amount. The default file format used by the application to export its result to is JSON and a checkbox below the entry box lets user choose whether the user wants to export the result into CSV file or not, in addition to the default JSON export. The Twitter OAuth consumer key and access token can be obtain by opening a Twitter Developer account. The search amount has no search limit as long as the tweets are

tweeted within the last seven days at the time of search.

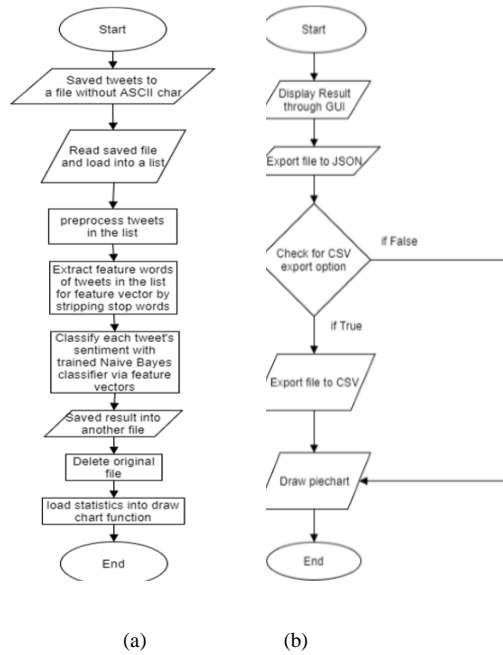


Fig. 3. (a) Sentiment Analysis Phase, (b) Display and Export Phase

Algorithm 1: Procedure Start the tweet crawler

```

{
    Tweetlist = []
    Add nltk corpus stopwords, URL and user tag into stopwords list
    Check for OAuth save file
    If file exist
        Fill in entry box with OAuth Consumer Key and Access Token
    If file not exist
        Entry box empty
    Open file trainingset.json file
    Loop for each line in file:
        Load each line in json into Python Object
        preprocess(tweet text) in Python Object
        TfeatureVector = extract feature words from processed tweet
        into vector and remove stopwords
        add TfeatureVector list into featureList
    append((TfeatureVector, tweetsentiment)) into TTweetlist
    featureList = list(set(featureList))
    put feature vector list into a trainingset list
    Set variable NBClassifier = trainNaiveBayesClassifier(training_set)
}
    
```

Algorithm 2: Procedure Search

```

{
    Save OAuth Consumer Key and Access token into a text file
    Read text file for Consumer Key and Access Token
    Filling OAuth parameters for API Authentication
    Get Search Term and Search Amount from Entry box
    Fill in search parameters
    Send API search request
    For each tweet in received
        Insert tweet data into object tweetslist{}
        Dump tweetslist into json file and remove any ASCII char
        Print downloaded tweet count
}
    
```

```

Algorithm 3: Procedure Classify
{
    Read each line from downloaded data from json file
    Load downloaded data into Python Object
    Pre-process tweets from Python Object
    Extract feature words into vector(pre-processed tweets)
    Sentiment = NBClassifier.classify(extracted feature vector)
    Dump tweet data and result into new result json file
    Display result on GUI
    Check GUI checkbox if user request export to CSV
        Export to CSV if true
    Save count of each sentiment into a tuple
    Delete old json file
    Draw piechart with statistics in the tuple
}
    
```

Fig. 4 shows the user interface which is empty if the user starts it for the first time. Fig. 5 shows an example where the user fills up the entry box in the user interface of the application and search. The TreeView box below will display the result, pink for negative tweets, light grey for neutral and green for positive tweet. Fig. 6 shows the pie chart with statistics of the sentiment analysis results mapped on it.

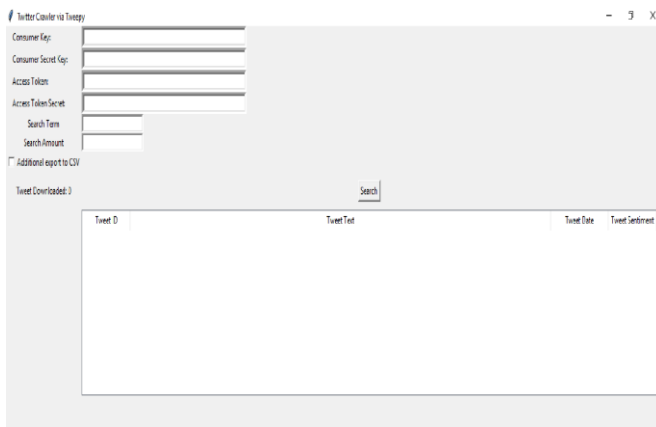


Fig. 4. Empty GUI when user first start the app



Fig. 5. GUI after user crawl for tweets

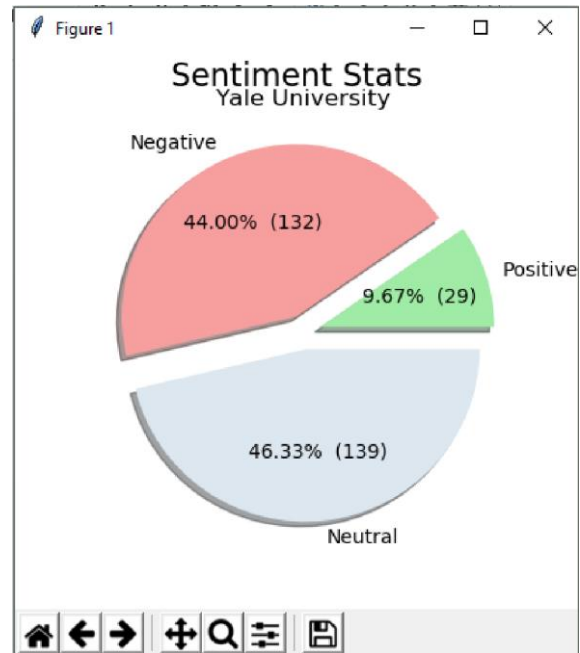


Fig. 6. Pie Chart with statistics from analysis

IV. EVALUATION

A. Test Set Evaluations

The original 600 tweets datasets are a mixture of 200 tweets each from Princeton University, Stanford University and University of Oxford through a period of 4 days from 12/1/2018 to 15/1/2018. The tweets are labelled by 3 students to ensure that the training set is unbiased and a gold standard is used by selecting the sentiment of the majority count from the variation of the labelled data.

300 of the golden standard data set are selected as training set with approximate balanced data of 100 positive tweets, 111 neutral tweets and 89 negative tweets are selected. After training the classifier with the training set, 4 testing sets are selected from another pool of 1000 labelled tweets, consisted from 200 tweets each from Harvard University, Princeton University, Cambridge University, Oxford university and Stanford University. Batches of 200 random tweets each, are selected for test set 1 to test set 3, while test set 4 include the entire pool of labelled tweets.

B. Evaluation Metrics

The application will be evaluated based on precision, recall and F1 measure of each sentiment classifier.

- Precision is based on the amount of True Positive prediction among the prediction of a sentiment. The higher the precision, the lower the number False Positive.
- Recall is based on the amount of True Positive prediction among the actual number of a sentiment. The higher the value of recall, the lower the number of False Negative predictions.



- F1 measure is the combination of precision and recall. The score of F1 measure is helpful as it rates a system by both precision and recall using a single metric.

C. Results and Discussion

Test set 1 contains 72 positive tweets, 56 neutral tweets and 72 negative tweets. In test set 1, the F1 measure of the classifier for positive, neutral and negative sentiment is 30.66%, 33.63% and 50.66% respectively (see Fig 7). It can be observed that F1 measure of negative sentiment is significantly higher than the other 2 sentiments. The F1 measure of the remaining sentiments has a slight difference of 2.97%.

Test set 2 (see Fig. 8) contains 54 positive tweets, 71 neutral tweets and 75 negative tweets. In test set 2, the F1 measure of the classifier for positive, neutral and negative sentiment is 29.78%, 43.97% and 50.00% respectively. F1 measure of negative sentiment is significantly higher than the other 2 sentiments while F1 measure of positive sentiment is the lowest, only scoring 29.78%.

Balanced Training set 2	300 tweets	100 pos	111 neu	89 neg
Test Set 1	Labelled data:	Pos	Neu	Neg
Predicted data				
Pos		21	24	20
Neu		24	19	14
Neg		27	13	38
Precision Pos		32.31%		
Precision Neu		33.33%		
Precision Neg		48.71%		
Recall Pos		29.17%		
Recall Neu		33.93%		
Recall Neg		52.77%		
F1 Pos		30.66%		
F1 Neu		33.63%		
F1 Neg		50.66%		
Test Set 1 Labelled data	Pos:72	Neu:56	Neg:72	Total:200

Fig. 7. Evaluation Result of Test Set 1

Test Set2	Pos	Neu	Neg
Pos	16	24	17
Neu	18	31	21
Neg	20	16	37
Precision Pos	29.93%		
Precision Neu	44.29%		
Precision Neg	50.68%		
Recall Pos	29.63%		
Recall Neu	43.66%		
Recall Neg	49.33%		
F1 Pos	29.78%		
F1 Neu	43.97%		
F1 Neg	50.00%		
Test Set 2 Labelled data	Pos:54	Neu:71	Neg:75
			Total:200

Fig. 8. Evaluation Result of Test Set 2

Fig. 9 shows the test set 3. From the figure, it contains 44 positive tweets, 80 neutral tweets and 76 negative tweets. In test set 3, the F1 measure of the classifier for positive, neutral and negative sentiment is 28.57%, 49.69% and 49.68% respectively. F1 measure of positive sentiment is significantly lower than the other two sentiments while F1 measure of the remaining two sentiments has only a slight difference of 0.01%.

Fig. 10 depicts the test set 4, which contains of 221 positive tweets, 389 neutral tweets and 340 negative tweets. In this test set, F1 measure of neutral sentiment is the highest at 51.44%, followed by F1 measure of negative sentiment of 44.20% and F1 measure of positive sentiment 31.26%.

Test Set3	Pos	Neu	Neg
Pos	14	19	21
Neu	21	40	20
Neg	9	21	35
Precision Pos	25.93%		
Precision Neu	49.38%		
Precision Neg	54%		
Recall Pos	31.80%		
Recall Neu	50%		
Recall Neg	46%		
F1 Pos	28.57%		
F1 Neu	49.69%		
F1 Neg	49.68%		
Test Set 3 Labelled data	Pos:44	Neu:80	Neg:76
			Total:200

Fig. 9. Evaluation Result of Test Set 3

Test Set4	Pos	Neu	Neg
Pos	88	101	103
Neu	128	215	104
Neg	55	73	133
Precision Pos	30.14%		
Precision Neu	48.10%		
Precision Neg	51%		
Recall Pos	32.47%		
Recall Neu	55.27%		
Recall Neg	39%		
F1 Pos	31.26%		
F1 Neu	51.44%		
F1 Neg	44.20%		
Test Set 4 Labelled data	Pos:221	Neu:389	Neg:340
			Total:1000

Fig. 10. Evaluation Result of Test Set 4



From the evaluation results of the four test sets, we can observe that the higher amount tweets of a sentiment in the test set improves the F1 measure of that sentiment. The results from Fig. 10 shows that the neutral sentiment, with the highest amount of tweets among the three sentiments among the test set, has the highest amount of True-Positive neutral sentiment and positive sentiment having the lowest F1 measure and True-Positive positive sentiment. The same pattern can be observed from test set 2 and test set 3 despite positive sentiment has the second highest amount of tweets in the training set. The negative sentiment, which has a higher amount of tweets than positive sentiment in each of the test sets, has a higher F1 measure score than positive sentiment despite having the lowest number of tweets in the sentiments.

V. CONCLUSION AND FUTURE WORKS

The results of the evaluation from the Tweet Crawler reveal that the balance of sentiments in a testing set does affect the effectiveness of the Tweet Crawler and F1 measure of the sentiment. The balance of tweets sentiment in training set while integral to effectiveness of the classifier, is not the main factor influencing the classifier effectiveness.

It should be noted that the training set and testing set in the evaluation are limited and in small numbers, future evaluations with larger amounts of tweets with good balance of sentiments may provide a more conclusive insight on how the balance of sentiments of testing set and training set affects the classifier's effectiveness. Future improvements that could be add to the Tweet Crawler include getting feedback from user on the sentiment of tweets that are Tweet Crawler fail to classify correctly and improve the classifier based on the feedback. The Tweet Crawler's Naïve Bayes classifier can be improve with usage of bigram and trigram in addition to the current unigram usage. Other approach such as Maximum Entropy can be added in conjunction with Naives Bayes to provide a better insight of the evaluation results with comparisons between two approaches.

REFERENCES

1. B Liu, "Sentiment analysis and opinion mining. Synthesis lectures on human language technologies", Morgan & Claypool Publishers, vol. 5 (1), 2012, pp. 1-167.
2. Alexa, "Traffic, Demographics, and Competitors", Available <http://www.alexametrics.com/siteinfo/twitter.com>, 2017.
3. S Aslam, "Twitter by the Numbers: Stats, Demographics & Fun Facts", Available <https://www.omnicoreagency.com/twitter-statistics/>, 2017.
4. M Thelwall, "The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength", *Cyberemotions*, 2013, pp. 119-134.
5. A Bakliwal, J Foster, J Puil, R O'Brien, L Tounsi, and M Hughes, "Sentiment analysis of political tweets: Towards an accurate classifier", *Association for Computational Linguistics*, 2013, pp. 49-58.
6. P. Salunkhe, and S. Deshmukh, "Twitter Based Election Prediction and Analysis", *International Research Journal of Engineering and Technology*, vol. 4(10), 2017, pp. 539-544.
7. S.A. Mardia, "Best practices for using social media as a recruitment strategy", *Strategic HR Review*, vol. 10(6), 2011, pp. 19-24.
8. H Kwak, C Lee, H Park, and S Moon, "What is Twitter, a social network or a news media? ", *ACM International conference on World Wide Web*, 2010, pp. 591-600.

9. H. Chade, G. Lewis, and L. Smith, "Student Portfolios and the College Admissions Problem", *The Review of Economic Studies*, vol. 81(3), 2014, pp. 971–1002.
10. S. Alhabash, and M. Ma, "A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students?", *Social Media + Society*, 2017, pp. 1-13.

AUTHORS PROFILE



Chiu-Lin Koh is a graduate from Faculty of Computing and Informatics, Multimedia University. The research on Popularity of University based on Sentiment Analysis in Social Network Media is conducted as part of his Final Year Project.



Su-Cheng Haw is Associate Professor at Faculty of Computing and Informatics, Multimedia University, where she leads several funded research on the XML databases. Her research interests include XML databases, query optimization, data modeling, semantic web, ontology, data management, and data warehousing. She has published more than 120 articles in reputable journals and conferences.



Lay-Ki Soon is a Senior Lecturer in School of Information Technology, Monash University Malaysia. She graduated with a PhD in Engineering (Web Engineering) from Soongsil University, Korea in 2009. Her main research interests are data mining, particularly text mining and natural language processing. She is also actively involved in research related to database technologies, where she leads projects on database integration for health data. To date, she has graduated six PhD students and two Master students. She is a Senior IEEE Member since 2018.