TWO-STEP MODEL FOR EMOTION DETECTION ON TWITTER USERS: A COVID-19 CASE STUDY IN MALAYSIA

Chempaka Seri Abdul Razak[1], Siti Hafizah Ab Hamid[2*], Hasni Meon[3], Hema A/P Subramaniam[4], Nor Badrul Anuar[5]

[1,2,4]Department of Software Engineering, Fac. of Computer Science & Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

[3]Department of Communication, Fac. of Communication, Visual Art & Computing, Universiti Selangor, Malaysia

[5]Department of Computer System & Technology, Fac. of Computer Science & Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

Email: chempakaseri96@gmail.com[1], sitihafizah@um.edu.my[2*](corresponding author), hasni@unisel.edu.my[3], hema@um.edu.my[4], badrul@um.edu.my[5]

DOI: https://doi.org/10.22452/mjcs.vol34no4.4

ABSTRACT

Human emotion is often a reflection of one's behaviour that leads to awareness of one's mental health. Emotion detection on Twitter users specifically gains attention as it generates information that is useful in the field of psychology and linguistics. However, most existing works related to the subject analysed the emotion of a tweet based on the number of keywords or phrases in it, thus resulting in false annotations and eventually false judgement of the user's emotion. This paper proposes a two-step model with high and reliable accuracy to detect the emotion of Twitter users based on the semantic meaning of the tweets they posted. The first step classified the emotion of each tweet using four different deep learning techniques. The second step detected the emotion of the user based on the proposed statistical post-level features and a boosting technique. Then, Kappa's agreement score method was implemented to validate both models. The first step achieved 0.9482 in accuracy while the second step achieved 0.9683 in accuracy. To further validate such highly accurate result, a case study was conducted, and a web-based system was developed to analyse the emotion of college students during the COVID-19 pandemic. Additionally, our analysis corresponded to most of the current reports on the pandemic which further proved the reliability of the developed model.

Keywords: Emotion Detection, Emotion Classification, Twitter, Statistical Post-Level Features, XGBoost, CNN, LS

1.0 INTRODUCTION

Twitter is one of the social media applications that constantly grows in popularity since it was first introduced in 2006 [1]. Twitter allows people to express their opinions or experiences as the expression records users' emotional state. Emotion is a major element of human nature, considering a significant number of studies on emotion by psychologists and behaviourists. Psychologists study emotion in Twitter to understand how individuals express their emotions and to interpret their mental state. Prior detection of users with mental health problems can impede serious consequences of mental health problems such as suicide. Emotion detection is a highly subjective matter since the interpretation of human emotion varies across individuals. Thus, researchers work with feature extraction algorithms to understand the subjectivity of emotions and to quantify emotions.

Emotion features cover the emotional perspectives of tweets such as the number of positive words, positive-affect words, dominance words, and negative-affect words in tweets. Existing studies of emotion detection on Twitter users showed that the emotion features were extracted directly from the users' tweets. In particular, studies by Choudhury et al. [2] and Li et al. [3] adopted emotion features as post-level features. However, both studies only focused on the number of positive words and negative words in the users' tweets based on existing libraries (i.e., LIWC and ANEW), thus affecting the reliability of the model accuracy proposed. On the other hand, the existing model classified the emotion of the user based on the overall features, superseding the importance of prior identification of the emotions of the tweets. In fact, studies showed that understanding the emotions of the tweets better interprets one's emotion and mental behaviour [4]. The emotions of the tweets need to be classified first before extracting the emotion features of the overall tweets to determine the emotion of the user and better understand each tweet semantically.

In this paper, we proposed a two-step model to detect the emotion of the Twitter users based on statistical post-level emotion features. The first model classified the tweets based on their emotions using deep learning techniques. Four different deep learning techniques were implemented to better classify the tweets. The second model exploited the boosting technique to detect the emotion of the Twitter users based on the statistical emotion features of the classified tweets. A two-step model is needed to classify the emotion of each tweet prior to extracting statistical emotion features and detecting the emotion of the Twitter users. The emotion detected described the intensity of the emotion class of the users to identify users with mental health conditions. The contributions of this paper are as follows:

a) Development of emotion detection model based on the semantics of bilingual tweets to ensure high accuracy in detecting the emotion of the Twitter users, using a two-step model, namely Emotion Classification Model (ECM) and Emotion Detection Model (EDM).
b) Analysis of the emotion of college students during the COVID-19 pandemic via Twitter to validate and verify the reliability of the model developed.

The remainder of the paper is structured as follows: Section 2 overviews related topics; Section 3 explains the creation of the datasets involved in the development of the proposed models; Section 4 outlines the proposed models to detect the emotion of Twitter users in relation to statistical-based emotion features of the tweets; Section 5 displays the results of the experiments conducted based on the proposed models; Section 6 illustrates the reliability of the proposed models by developing a system to analyse the emotion of Malaysian college students on Twitter during the COVID-19 pandemic; and Section 7 discloses the conclusion and future plans of this research.

2.0 RELATED WORKS

According to Li et al. [3], there are two levels of features involved in understanding the behaviour of the Twitter users, namely user-level features and post-level features as shown in Fig. 1. User-level features describe features based on the user profile such as the number of followings, the number of followers, and the number of groups that the user is interested in. Post-level features describe post-related features such as the number of tweets, the number of retweets, the number of positive words, and the number of negative words. Post-level features are categorised into three different groups, namely social interaction features, emotion features, and linguistic features. For this study, we specifically focused on the emotion features to detect the emotion of the Twitter users.

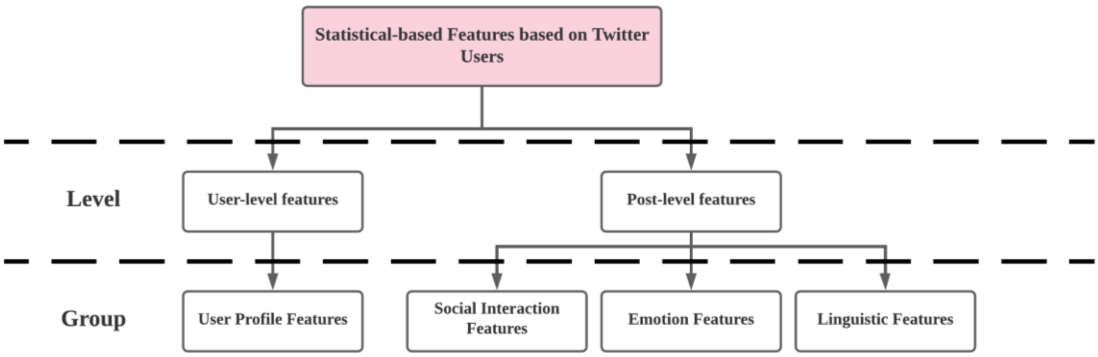

Fig. 1: Statistical-based features based on Twitter users

Choudhury et al. [2] predicted depression among Twitter users by measuring their depressive behaviour in consideration of the emotion features. They calculated the number of positive-affect words, negative-affect words, dominance words, and activation words in the users' tweets, on a daily basis. Support Vector Machine (SVM) algorithm was implemented in the prediction framework with 0.70 in classification accuracy. Meanwhile, Tsugawa et al. [5] estimated the degree of depression based on the users' social media activities where the ratio of positive-affect words to negative-affect words in their tweets was extracted. They adopted SVM and achieved approximately an accuracy of 0.69. On the other hand, Balakrishnan et al. [6] detected cyberbullying on Twitter based on the users' psychological features such as personalities, sentiment, and emotion where Indico API was used for emotion analysis in short texts. The cyberbullying detection models involved Random Forest and J48 that achieved 0.9188 in accuracy. Li et al. [3] implemented user-level and post-level features on Blued and Twitter to detect depression among men who have sex with men (MSM). They considered the average number of positive words and negative words, the total number of emoticons, and the average number of positive emoticons and negative emoticons in each post as the emotion features. They also implemented Extreme Gradient Boosting (XGBoost) algorithm to classify the users, thus achieved an accuracy of 0.9671. In practice, emotion features give a big impact in detecting Twitter users' emotion [2], [3]. However, their emotion needs to be identified based on their tweets prior to making any conclusions. In this study, we proposed several emotion features based on the classified tweets to detect the emotion of the Twitter users.

3.0 PROPOSED EMOTION DETECTION MODEL

We proposed a two-step model to detect emotions in tweets by firstly, constructing the Emotion Classification Model (ECM) to detect the emotion of the Twitter users based on statistical post-level emotion features. Secondly, building the Emotion Detection Model (EDM) using Extreme Gradient Boosting (XGBoost) as depicted in Fig. 2. We used Python as the main language to develop the predictive model. Our machine for model development was a Macbook Pro with a 2.3 GHz Dual-Core Intel Core i5 and an 8 GB 2133 MHz LPDDR3 RAM. The operating system used was macOS Big Sur 11.2.2, which ran on an Intel Iris Plus Graphics 640 1536 MB graphics. For this study, Russell's Circumplex Model of Affect was adopted in the proposed Emotion Detection Model (EDM). Russell suggested that the emotions were placed in two dimensions and separated into four quadrants of pleasant-active, pleasant-inactive, unpleasant-active, and unpleasant-inactive. The model covered neutral emotions unlike other models that only covered strong emotions. Therefore, to ensure that we included most of the important emotions, we chose *anger*, *afraid*, *happy*, *excited*, *sadness*, *bored,* and *relax*.

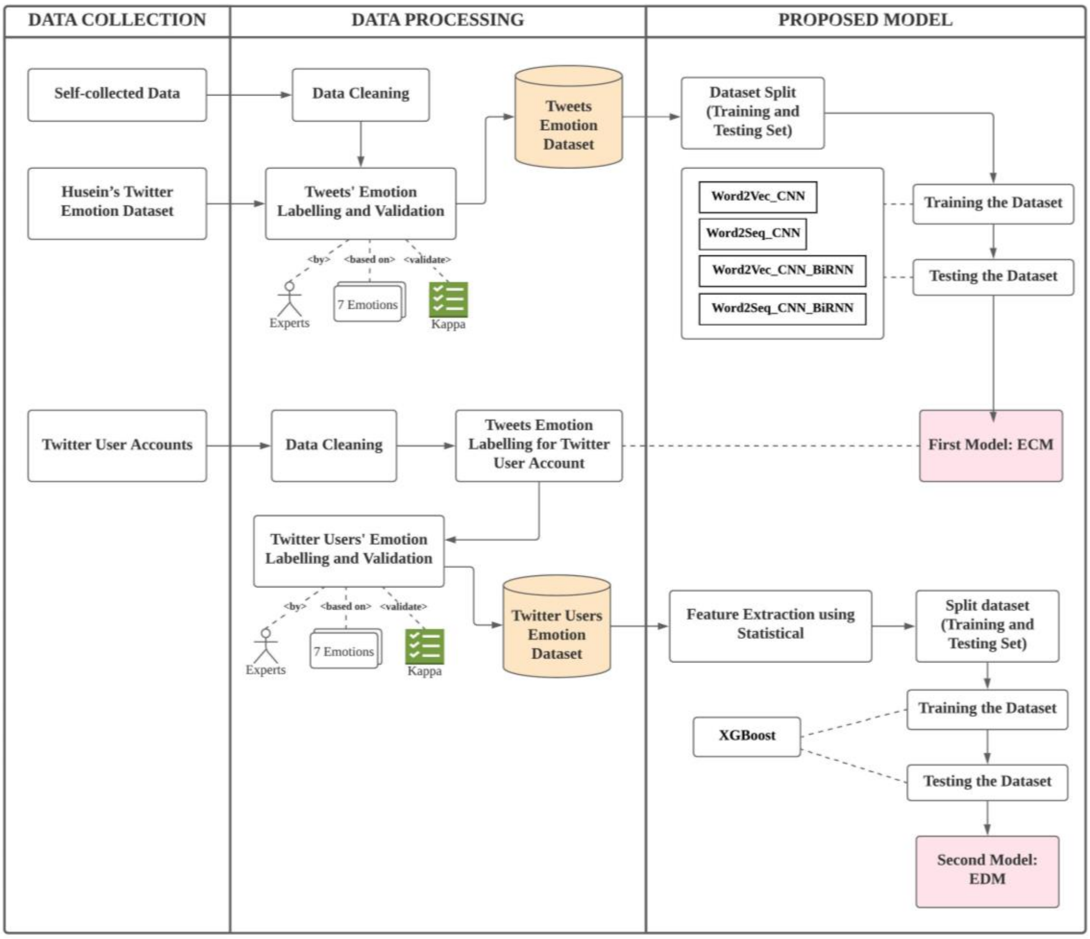

Fig. 2: The proposed framework

3.1 Step One: Emotion Classification Model (ECM)

The Emotion Classification model (ECM) was responsible to classify the tweets based on the seven emotions chosen. The ECM used the deep learning classifier to classify the tweets based on emotion. The process started with creation of Tweets Emotion Dataset. Tweets Emotion Dataset was for the ECM development where the tweets were classified based on seven emotions. Our tweets came from two resources, namely Husein/Malaya-Dataset/Twitter-Emotion-Data and self-collected tweets. For the latter, we wrote a customized Python HTTP API script via Twitter API; Tweepy. Tweepy returned a list of tweets from the Twitter database based on various parameters such as hashtags, links, and phrases by the chosen time. We used "*twitter_name*" to extract Malaysian Twitter users' tweets. Husein/Malaya-Dataset/Twitter-Emotion-Data [7] is a public dataset with six emotions (i.e., *anger*, *fear*, *happy*, *love*, *sadness,* and *surprise*), where raw tweets were extracted using Elasticsearch query. Once we gathered the tweets as data in the database, we then filtered them by removing URLs, username, redundant characters, non-alphanumeric characters, emojis, and duplicated tweets. For this study, we labelled the tweets or the data using sentence-level emotion annotation by employing a hundred annotators who were experts in social communication from Faculty of Communication, Visual Art and Computing, Universiti Selangor. We grouped them into five groups where each labelled 600 tweets. Then, to determine the consistency of the annotations, we validated the labelled data using Kappa agreement score based on multi-rater annotation. Finally, we only considered dataset with agreement score of 0.84. Table 1 shows the number of final data tweets for each emotion in the dataset.

Table 1: Number of tweets in the dataset

Emotion	Labelled Dataset	Husein/Malaya-Dataset	Total
Afraid	9,543	20,443	29,986
Anger	8,984	21,026	30,010
Bored	29,001	0	29,001
Excited	29,004	0	29,004
Happy	9,670	20,462	30,132
Relax	29,010	0	29,010
Sad	9,623	20,799	30,422
Total	124,835	82,730	207,565

To ensure high accuracy, we developed the ECM model by implementing two deep learning algorithms called Convolutional Neural Network (CNN) and Bidirectional Recurrent Neural Network (Bi-RNN) with two word-embedding techniques, namely Word-to-Sequence (Word2Seq) and Word-to-Vector (Word2Vec). The algorithms were stacked as Word2Seq_CNN, Word2Seq_ CNN_Bi-RNN, Word2Vec_CNN, and Word2Vec_ CNN_Bi-RNN. Feizollah et al. [8] called this approach as stack algorithm where the approach produced a great performance for sentiment analysis. Word2Seq and Word2Vec are widely used, resulting in acceptable outcomes as explained below:

a) **Word2Seq:** A word sequencing approach to the text that ignored the weight of each word. Word2Seq mapped the word sequence into a matrix with the length (input size) and height (number of observations) that had been generalised.
b) **Word2Vec:** A feature extraction technique for text classification that used pre-trained models, namely Word2Vec [9] and GLoVe models [10]. The pre-trained model contained the weight of each word available inside the model. Thus, Word2Vec supplied the word sequence with the weight of each word, making the embedding or input layer, a vector representation of the texts.

The development of the ECM model used stacking of deep learning algorithms in the forms of layers. Different algorithms were stacked up and the results of one algorithm was passed on to another algorithm. This method proved that the weakness of one algorithm was compensated by the strength of another algorithm [11]. The development of the ECM used Keras library in Python to implement the neural network architectures to ensure easy implementation and modularity across different neural network architectures. The modularity reduced the complexity of building a powerful deep learning model. This allowed us to focus more on feature extraction and hyper parameter tuning rather than implementing the neural network architectures. Two different algorithms were implemented for the creation of ECM:

a) **CNN**: Convolutional Neural Network (CNN) utilised layers with convolving filters that were applied to local features [12]. It consisted of input layer, convolutional layer, fully connected layer, and output layer.
b) **Bi-RNN**: Bidirectional recurrent neural network was extended to a regular recurrent neural network [13]. The idea of Bi-RNN was to split the state neurons of a regular RNN in a part that was responsible for the first-time direction (forward states) and a part for the negative time direction (backward states).

To ensure that the inputs were of the same size, padding was added. The word embeddings were passed through the model. Table 3 shows the details for each algorithm used in the creation of ECM. Each word in the sentences was represented by its word embedding representation. The word embedding was passed through the layers. The final layer concatenated the vectors and passed through a fully connected layer with 250 neurons. Finally, the output of these 250 neurons was passed through another fully connected layer with 7 neurons (correspond to the emotion chosen) and the output of these neurons was converted to probability distribution of Softmax function.

Table 2: Details on each layer in the proposed model for each experiment

Model	Main Layer	Neural Network Layer	Output Size	Kernel Size	Max-Pool size	Activation function
Word2Seq_CNN	Input layer	Word2Seq Embedding Layer	300	None	None	None
	Hidden layer	CNN 1	300	3	50	ReLU
		CNN 2	300	2	10	ReLU
	Fully connected layer	Dense Layer 1	300	None	None	ReLU
		Dense Layer 2 (output layer)	7	None	None	Softmax
Word2Seq_CNN_ BiRNN	Input layer	Word2Seq Embedding layer	300	None	None	None
	Hidden layer	CNN Layer 1	300	3	50	ReLU
		CNN Layer 2	300	2	10	ReLU
		Bi-RNN Layer 1	300	None	None	ReLU
	Fully connected layer	Dense Layer 1	300	None	None	ReLu
		Dense Layer 2 (output layer)	7	None	None	Softmax
Word2Vec_CNN	Input layer	Word2Vec Embedding Layer	300	None	None	None
	Hidden layer	CNN 1	300	3	50	ReLU
		CNN 2	300	2	10	ReLU
	Fully connected layer	Dense Layer 1	300	None	None	ReLU
		Dense Layer 2 (output layer)	7	None	None	Softmax
Word2Vec_CNN_ BiRNN	Input layer	Word2Vec Embedding layer	300	None	None	None
	Hidden layer	CNN Layer 1	300	3	50	ReLU
		CNN Layer 2	300	2	10	ReLU
		Bi-RNN Layer 1	300	None	None	ReLU
	Fully connected layer	Dense Layer 1	300	None	None	ReLU
		Dense Layer 2 (output layer)	7	None	None	Softmax

3.2 Step Two: Emotion Detection Model (EDM)

The Emotion Detection Model (EDM) detected the emotion of the Twitter users based on the emotion features of their tweets. The development of EDM began by constructing the Twitter Users Emotion Dataset. First, we collected 735 user accounts using Twitter API with users' *twitter_name*, *tweets,* and *tweets_date*. Then, using a custom Python scripting, we classified the tweets in each user account based on the seven emotions generated by the ECM. The emotion of a particular user was concluded based on the emotion with the highest frequency, as derived from all tweets posted by the user. For example, if a user posted 2 *afraid* tweets, 45 *anger* tweets, 4 *bored* tweets, 2 *excited* tweets, 20 *happy* tweets, 39 *relax* tweets, and 17 *sad* tweets, then, the emotion of the user would be *anger*. For a user with the same number of two or more emotions, the average value of each emotion would be calculated and the emotion with the highest average value would be concluded as the emotion of the user. We invited two experts of social communication from Faculty of Communication, Visual Art, and Computing from Universiti Selangor (UniSEL) and two experts of English language and literature from Kuliyyah of Islamic Revealed Knowledge and Human Sciences, International Islamic University Malaysia (IIUM) to validate the emotion labelled by our model. Kappa agreement score was used to ensure consistency in validating the users' emotion and an agreement score of 0.83 was achieved. Table 2 shows the number of user accounts for each emotion in the datasets.

Table 3: Number of users for each emotion

Emotion	Number of User
Afraid	117
Anger	102
Excited	3
Happy	33
Relax	78
Sad	408
Total	735

The development of EDM included statistical feature extraction from the Twitter User Emotion Dataset. Table 4 shows the list of emotion features that were extracted for each Twitter users. We limited our features to 19 because the number of features was important in building a good and accurate detection model, but having too many features might lead to overfitting and increasing the complexity of the detection model [14]. The statistical features were described to be in the format of double-precision floating-point. The classes that represented the types of emotion were described as level 0 for *afraid*, level 1 for *anger*, level 2 for *bored*, level 3 for *excited*, level 4 for *happy*, level 5 for *relax*, and level 6 for *sad*.

Table 4: Extracted statistical-based emotion features based on the tweets

Label	Features	Category	Ref
F1	Total tweets for each user	Numerical	Proposed
F2	Percentage of afraid tweets for each user	Numerical	Proposed
F3	Percentage of anger tweets for each user	Numerical	Proposed
F4	Percentage of bored tweets for each user	Numerical	Proposed
F5	Percentage of excited tweets for each user	Numerical	Proposed
F6	Percentage of happy tweets for each user	Numerical	Proposed
F7	Percentage of relax tweets for each user	Numerical	Proposed
F8	Percentage of sad tweets for each user	Numerical	Proposed
F9	Average length of tweets for each user	Numerical	[15]
F10	Average Automated Readability Index (ARI) of tweets for each user	Numerical	[16]
F11	Average number of characters of words of each tweets for each user	Numerical	[17]
F12	Standard deviation of the emotion of the user	Numerical	Proposed
F13	Average probabilities value for afraid tweets for each user	Numerical	Proposed
F14	Average probabilities value for anger tweets for each user	Numerical	Proposed
F15	Average probabilities value for bored tweets for each user	Numerical	Proposed
F16	Average probabilities value for excited tweets for each user	Numerical	Proposed
F17	Average probabilities value for happy tweets for each user	Numerical	Proposed
F18	Average probabilities value for relax tweets for each user	Numerical	Proposed
F19	Average probabilities value for sad tweets for each user	Numerical	Proposed

The first emotion feature extracted was the number of total tweets for each user to show the frequency of the user's activity on Twitter. Each tweet was classified using the ECM and the model gave inputs for each emotion class probability. The emotion class that obtained the highest probability value was selected as the tweet's emotion. The next seven emotion features extracted for each user were *afraid* tweets (F2), *anger* tweets (F3), *bored* tweets (F4), *excited* tweets (F5), *happy* tweets (F6), *relax* tweets (F7), and *sad* tweets (F7). Equation 1 shows the formula to determine the percentage of each emotion class for each user as shown in Equation 1.

$$Percentage\ of\ afraid\ tweets = \frac{Total\ number\ of\ afraid\ tweets}{Total\ number\ of\ tweets\ for\ each\ user} \qquad \text{Eq. 1}$$

The average probability value of *afraid* tweets (F13), *anger* tweets (F14), *bored* tweets (F15), *excited* tweets (F16), *happy* tweets (F17), *relax* tweets (F18), and *sad* tweets(F19) for each user was extracted. These emotion features of tweets were extracted instead of the actual emotion of the user as the latter might have a high probability value but low percentage value of for each emotion class. Equation 2 shows the formula to calculate the average probability value of each emotion class.

$$Average\ Probability\ Value = \frac{\sum Probability\ value\ of\ each\ emotion\ class}{Total\ number\ of\ tweets\ for\ each\ emotion\ class} \qquad \text{Eq. 2}$$

Standard deviation of the user's emotion (F12) was calculated to measure the dispersion of the tweet's emotion to the mean of the user's emotion. The formula for the standard deviation is shown in Equation 3. If the data points are further from the mean, there is a higher deviation within the data set. Thus, the more spread out the data, the higher the standard deviation.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \qquad \text{Eq. 3}$$

Automated Readability Index (ARI) is an index score that identifies the level of readability of a text [19]. We extracted ARI to identify whether the text was understandable based on the grade group. Harris [17] leveraged the usefulness of the ARI to detect opinion spams in English texts. From that score, an assignment table was used to confirm that the text belonged to the respective grade group as shown in Table 5. Equation 4 shows the formula to calculate the ARI value. The average ARI of each user was calculated as shown in Equation 5. The implementation of the ARI in this study focused only on the calculation of the ARI scores without assigning texts to different groups.

$$4.71 \times \left(\frac{characters}{words}\right) + 0.5 \times \left(\frac{words}{sentences}\right) - 21.43 \qquad \text{Eq. 4}$$

$$Average\ ARI\ of\ tweets\ for\ each\ user = \frac{\sum ARI\ of\ each\ tweet}{Total\ number\ of\ tweets\ for\ each\ user} \qquad \text{Eq.5}$$

Table 5: The ARI score assignment table

Score	Age	Grade Level
1	5 – 6	Kindergarten
2	6 – 7	First grade
3	7 – 8	Second grade
4	8 – 9	Third grade
5	9 – 10	Fourth grade
6	10 – 11	Fifth grade
7	11 – 12	Sixth grade
8	12 – 13	Seventh grade
9	13 – 14	Eight grade
10	14 – 15	Ninth grade
11	15 – 16	Tenth grade
12	16 – 17	Eleventh grade
13	17 – 18	Twelfth grade
14	18 – 22	College

For the development of the EDM, we implemented Extreme Gradient Boosting (XGBoost) as the classifier for the model. Boosting was one of the many elements used in machine learning for the creation of a predictive model that utilised the data frame prepared in Python. Consequently, the data frame fed into different machine learning that utilised the boosting approach to build the predictive models. We carefully tuned by a grid with small but adaptive step to enumerate the classification accuracy rates in different XGBoost parameter settings to obtain the optimal value of parameter for the algorithm. The search ranges for learning rate and maximum tree depth are 0.3 and 3 respectively. With that in mind, the 70/30 percent splitting technique of data was applied on the training and testing set. We randomly sampled and evenly distributed the data based on its class to avoid any imbalanced data distribution. The randomised datasets were fed into the machine learning classifier to avoid any bias-related elements being considered when building the predictive model. The datasets were randomised multiple times for conformity of randomness.

4.0 RESULT AND ANALYSIS

Based on the prepared datasets for the ECM and the EDM, Table 6 presents the training results of the algorithms for the ECM based on its performance in emotion classification of the tweets. Meanwhile, the training result for the EDM was based on the performance of the model in emotion detection on the Twitter users. For the ECM, each tweet was separated into a single word to calculate the emotion score of each tweet and conclude the highest emotion score as the result. The training result showed that Word2Seq_CNN achieved the highest score of 0.9482 for both accuracy and recall metrics in which the latter determined the true positive rate of the accurately classified emotions in the tweets. The f-measure score showed the balance of the results in terms of precision and recall. Although the datasets contained an unbalanced tweet number for all emotion classes, the high f-score showed that the model could balance the classification of emotions. Word2Seq_CNN model acquired the highest f-measure score of 0.9482 and the highest precision score of 0.9493. The EDM achieved the highest accuracy and recall score of 0.9683 using XGBoost with proposed features. The recall score implied that XGBoost worked well in detecting the emotion of the Twitter users by achieving the highest positive rate. In terms of f-measure score, XGBoost with proposed features projected an evaluation score of 0.9672. It indicated the ability of the model to balance the positivity rate and false-positive rate. For precision score, XGBoost with selected existing and proposed features achieved the highest score of 0.9664.

Table 6: Training results for each algorithm

Model	Algorithm	Accuracy	Precision	Recall	F-Measure
ECM	Word2Seq_CNN	0.9482	0.9493	0.9482	0.9482
	Word2Vec_CNN	0.9326	0.9335	0.9326	0.9324
	Word2Seq_CNN_BiRNN	0.9406	0.9410	0.9406	0.9404
	Word2Vec_CNN_BiRNN	0.9274	0.9282	0.9274	0.9274
EDM	XGBoost (Without Proposed Features)	0.9412	0.9464	0.9412	0.9379
	XGBoost (With Proposed Features)	0.9683	0.9664	0.9683	0.9672

Table 7 shows the performance of each algorithm on each emotion class used by the model. For the Word2Seq_CNN model, the best performing emotion was *afraid* as the emotion obtained the highest f-measure score of 0.99 while *relax* emotion had the lowest f-measure score of 0.92. Word2Vec_CNN showed that *afraid* emotion achieved the highest f-measure score of 0.97 while *anger* and *sad* emotion achieved the lowest f-measure score of 0.91. *Afraid* emotion achieved the highest f-measure score of 0.98 and *sad* emotion obtained the lowest f-measure score of 0.91 for Word2Seq_CNN_BiRNN. Finally, for the Word2Vec_CNN_BiRNN model, the model achieved 0.97 for the *afraid* emotion that obtained the highest f-measure while *sad* emotion had the lowest f-measure score of 0.90. In conclusion, all models performed well on the *afraid* emotion but badly on the *sad* emotion.

Table 7: Performance on each algorithm

Model	Algorithm	Emotion	Precision	Recall	F-Measure
ECM	Word2Seq_CNN	Afraid	0.99	0.99	0.99
		Anger	0.94	0.94	0.94
		Bored	0.96	0.95	0.96
		Excited	0.92	0.99	0.95
		Happy	0.96	0.94	0.95
		Relax	0.90	0.95	0.92
		Sad	0.98	0.89	0.93
	Word2Vec_CNN	Afraid	0.95	0.98	0.97
		Anger	0.95	0.87	0.91
		Bored	0.88	0.97	0.92
		Excited	0.96	0.95	0.95
		Happy	0.94	0.93	0.93
		Relax	0.93	0.92	0.93
		Sad	0.92	0.90	0.91
	Word2Seq_CNN_BiRNN	Afraid	0.98	0.99	0.98
		Anger	0.95	0.92	0.93
		Bored	0.93	0.96	0.95
		Excited	0.92	0.98	0.95
		Happy	0.97	0.92	0.94
		Relax	0.92	0.91	0.92
		Sad	0.91	0.91	0.91
	Word2Vec_CNN_BiRNN	Afraid	0.97	0.98	0.97
		Anger	0.91	0.90	0.90
		Bored	0.90	0.96	0.93
		Excited	0.95	0.96	0.96
		Happy	0.96	0.88	0.92
		Relax	0.92	0.90	0.91
		Sad	0.89	0.91	0.90
EDM	XGBoost (Without proposed features)	Afraid	1.00	0.91	0.96
		Anger	1.00	0.68	0.81
		Excited	1.00	1.00	1.00
		Happy	1.00	1.00	1.00
		Relax	0.95	1.00	0.98
		Sad	0.91	1.00	0.95

XGBoost (With proposed features)	Afraid	1.00	1.00	1.00
	Anger	1.00	0.77	0.87
	Excited	1.00	1.00	1.00
	Happy	1.00	1.00	1.00
	Relax	0.95	1.00	0.98
	Sad	0.91	1.00	0.95

5.0 CASE STUDY: MALAYSIAN COLLEGE STUDENTS' EMOTION DURING COVID-19 PANDEMIC

To ensure the reliability of the model accuracy, we developed a system, namely Tweep to analyse the Malaysian college students' emotion during the COVID-19 pandemic as a case study. Our main hypothesis was to ensure the soundness of the proposed model by applying it on a real case scenario. COVID-19 or Coronavirus Disease 2019 was first detected in Wuhan, China on December 31, 2019, and had grown rapidly and severely ever since. The coronavirus was declared as a pandemic on March 11, 2020, by World Health Organization (WHO), reporting 83,990,222 confirmed cases of COVID-19 worldwide as of December 31, 2020, with a death toll of 1,850,775. Malaysia was identified as the first Southeast Asia country that passed 500 COVID-19 confirmed cases and this had taken a toll on Malaysia educational system. All universities and colleges were forced to close down their campuses on March 18, 2020, and moved to online classes, thus affecting more than 500,000 students. All on-campus activities such as workshops, conferences, and sports were cancelled and relocated from in-person classrooms to online platforms. For the case study, we fetched the users' tweets from March 1, 2020, to December 31, 2020, following the rise in COVID-19 reported cases in Malaysia and the government announcement of total nationwide lockdown.

5.1 A Web-based System with Emotion Detection Model

We developed a system namely, Tweep that took a Twitter user's tweets with dates as the input of the system. Each user's tweets underwent tweet pre-processing where any unwanted information was removed such as URLs, links, username, images, and emojis. Next, each user's tweets were classified using the first step, ECM where four different algorithms were used to classify the emotions of the tweets. If two emotions had the same frequency, the average probability of the emotions would be calculated.

The emotion probabilities of each tweet were stored in the database before extracting the statistical post-level emotion features for each user. Later, the emotion features of each user were passed to the EDM to detect the emotion of the user. The system then produced two outputs as seen in Fig. 4 and Fig. 5. Fig. 4 illustrates the user's emotions through *word size* which visualised the number of tweets for each emotion analysed during the period, and *word saturation* which highlighted the low and high probability of the user's emotion. We employed the statistical features of the classified tweets to generate the saturation. Meanwhile, Fig. 5 displays the user's entire tweets and a time-varying distribution on the emotion of the tweets.

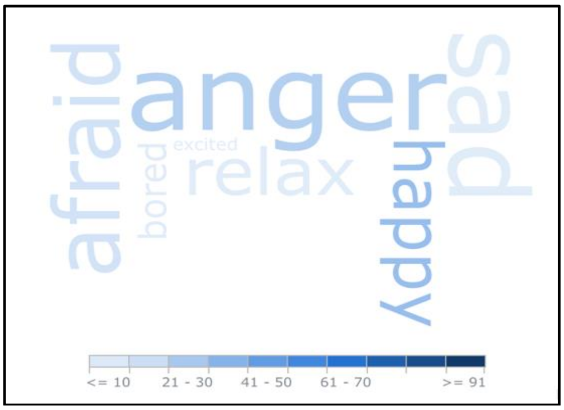

Fig. 4: Count vs Saturation Results

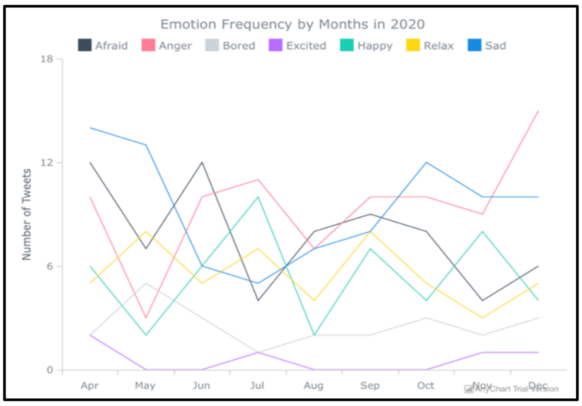

Fig. 5: Tweets Emotion Frequency

We limited the user population to students within the Klang Valley area. We also considered users that are Malay and Malay English (Manglish) speakers as Malaysians tend to post in Manglish on social media [17]. We retrieved a total of 153,336 tweets from 452 user samples posted within the timeframe as social distancing protocol and college closure were enforced nationwide. As we focused on the users' emotion during the pandemic, we ignored any topics tweeted by the users.

We considered age and gender entities to be our main descriptive attributes to gain general view of our user samples. According to Ministry of Education Malaysia, 552,702 students were enrolled in 2019 where 38.66% were male and 61.34% were female. For this study, we considered college students aged between 18 to 29 years old as the enrolled students [20]. Both attributes of age and gender were extracted using M3 Model, a deep learning system for inferencing the demographics of users based on four features on a Twitter profile that include user's name, screen name, biography, and profile image [21]. A gold-standard annotation was conducted to verify whether the collected Twitter users were college students or not. The human annotators were experts in English language and literature from Kuliyyah of Islamic Revealed Knowledge and Human Science, International Islamic University Malaysia (IIUM) where they validated the users' age and genders using Kappa agreement score, thus achieved the agreement score of 0.94.

5.2 Analysis of the Results

5.2.1 Demographics

We analysed a total of 452 Malaysian Twitter users to identify their emotion during COVID-19. The process took an average of 29 minutes for each user and collected a total of 153,336 analysed tweets. 38.5% of the collected Twitter users were male while the remaining were identified as female. Our data samples were dated from March 1, 2020, prior to the government's announcement of the Movement Control Order (MCO) or total nationwide lockdown on March 18, 2020. As the confirmed cases started to drop, the total lockdown shifted to a conditional one, Conditional Movement Control Order (CMCO) on June 10, 2020. During the CMCO, Eid celebration happened on May 24, 2020, as one of the major religious events in the country. As daily cases started to decline more prominently, the government announced the Recovery Movement Control Order (RMCO) on June 10, 2020. Unfortunately, the third waves of COVID-19 hit Malaysia on September 8, 2020, that coincided with two other major celebrations; Deepavali and Christmas on November 12, 2020, and December 25, 2020, respectively.

5.2.2 Analysis of Malaysian College Students' Emotion during COVID-19

Fig. 6 shows the emotion distribution of the Malaysian Twitter users during the defined period where the *anger* emotion took up 49.12%, involving 52.87% of the male users and 46.76% of the female users. Meanwhile, Fig. 8 depicts the emotion distribution of Malaysian users' tweets based on emotion and month from March 1, 2020, to December 31, 2020, where the *relax* emotion had the greatest number of tweets amounted to 24.62%. Our findings showed a significant trend of consistent negative emotion during the pandemic among college students, especially as the daily confirmed cases saw an alarming rise in March 2020 as shown in Fig. 7. Furthermore, the *sad* tweets had the highest percentage as the government enforced the Movement Control Order (MCO) on March 18, 2020, inferring the college students' sadness at the beginning of the MCO as they were stuck in campus away from their family [22]. However, as the number of daily reported cases began to decline, the *happy* tweets achieved the highest percentage as the college students felt more at ease, approaching the Eid celebration. Nevertheless, the percentage of *sad* tweets slowly re-emerged as the third wave hit the country and the trend remained throughout two major celebrations; Deepavali and Christmas that happened afterwards.

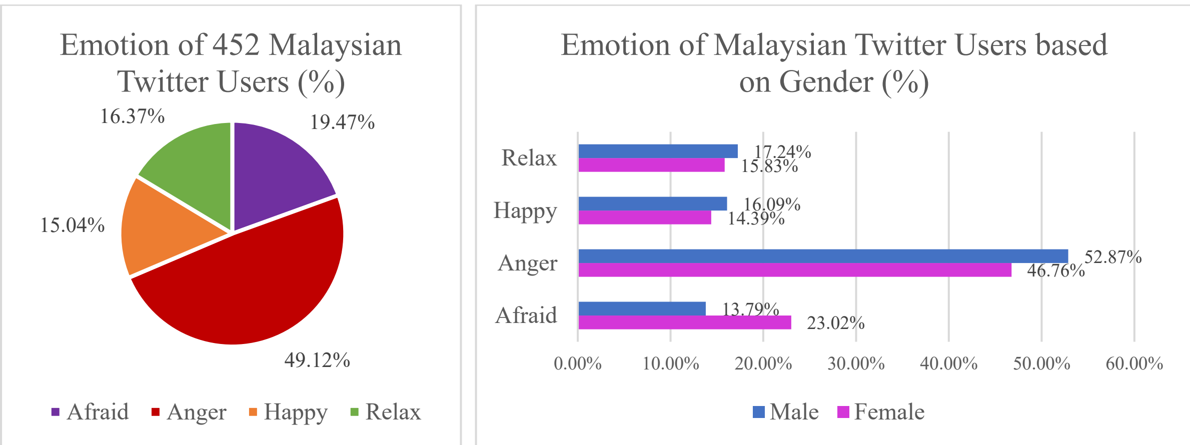

Fig. 6: Emotion distribution of Malaysian twitter users

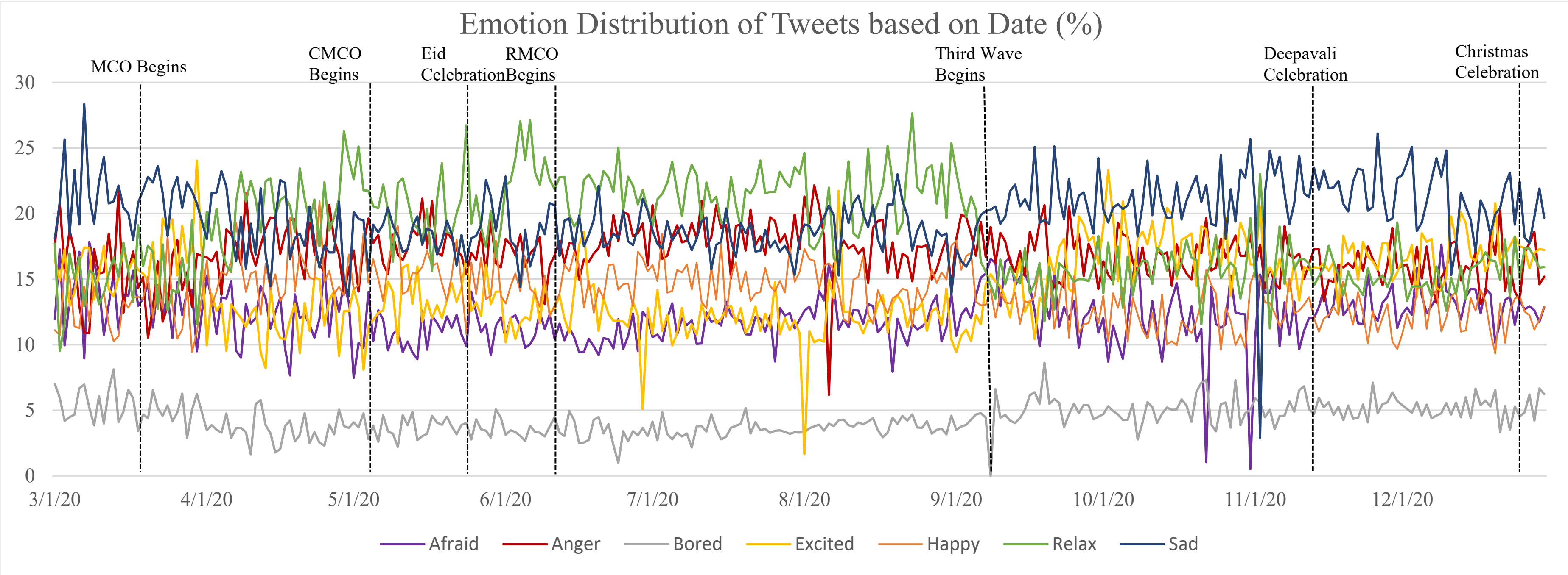

Fig. 7: Emotion distribution of Malaysian users' tweets from March 1, 2020, to December 31, 2020

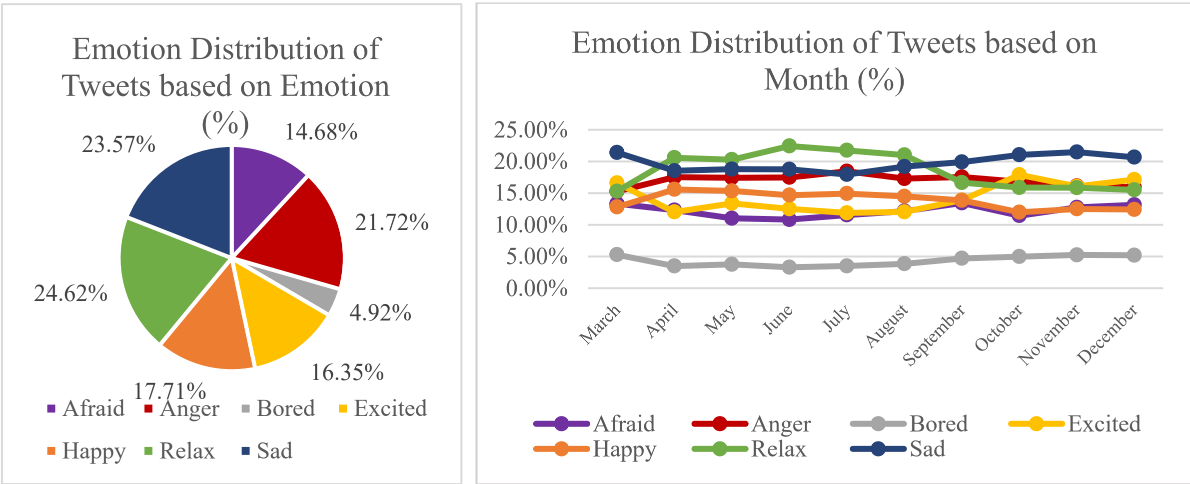

Fig. 8: Emotion distribution of Malaysian Twitter users' tweets based on emotion and month

5.2.3 Consistency Result with Other Reports

Due to the COVID-19 outburst in Malaysia, colleges and universities were closed, entailing the beginning of full online learning despite digital poverty in several rural places in Malaysia. These situations caused the college students to feel stressed as they needed to cope with online learning and limited social contacts, thus jeopardising their psychological health [23], [24]. Sundarasen et al. [25] studied the impact of Covid-19 on the anxiety level of university students in Malaysia. The study found that the main stressors included financial constraints, remote online teaching, and uncertainty about the future regarding academics and career as these were significantly associated with a higher level of anxiety. This problem raised concerns regarding mental health issues among the college students as they were stressed throughout the MCO period, fearing potential suicide [25].

The response to COVID-19 had a great impact on the public causing multiple challenges for the college students. Pragholapati [26] found that students often discussed topics that were familiar to their neighbourhood (i.e., school closing and local news) and the students were observed to tweet negatively during COVID-19. In addition, positive microscopic examinations and negative tweets revealed the student's problematic feelings during the spread of COVID-19, as well as adverse reactions to disturbances in their lives such as racial aggression. Duong et al. [27] focused on high-level attributes of tweets such as topic models and sentiments to understand the characteristics of users that discussed social issues associated with COVID-19 based on their age and gender. The study found that the college students' tweets were significantly more negative. The tweets revealed that they were overwhelmingly troubled amidst the spread of COVID-19, along with unfavourable reactions to the disruption of lives such as racism-aggression.

Our results were consistent with an article by The Straits Time that reported an increase in the number of suicides during the pandemic where 78 cases were documented from March 18 to June 9, 2020, compared to 64 cases during the same period in 2019 [28]. A study found that 45% of 1084 Malaysian respondents experienced varying level of anxiety and depression during the MCO. Similarly, our finding as well as studies by Li et al. [29], Li et al. [30], and Aucejo et al. [31] agreed to the trend of negative emotions during the lockdown and due to the pandemic. The similar results proved that our emotion detection model was not only accurate but also reliable.

6.0 CONCLUSION

In this paper, we proposed an Emotion Detection Model, a two-step model to detect the emotion of the Twitter users based on the statistical post-level emotion features of the classified tweets. The first model, ECM classified each user's tweets using four different models that implemented deep learning algorithm. The second model, EDM detected the emotion of the Twitter users based on statistical post-level emotion features of the classified tweets. The ECM achieved an accuracy of 0.9482 while the EDM achieved an accuracy of 0.9683. Our findings showed that the implementation of boosting model encouraged an increase in the performance of the second model or the accuracy of a detection model in general as it used the proposed statistical-based features. In addition, this research mainly focused on the statistical-based features of the tweets in detecting the emotion of the college students. Analysis on more fine-grained linguistics information in tweets during COVID-19 can be performed to gain further insights on the emotion of Twitter users during the pandemic. Based on the results, COVID-19 caused a negative impact on college students as they felt sad and depressed at the beginning of the MCO, being stuck in campus, far from their family. The college students needed to adapt to sudden changes as online learning was fully implemented and social distancing was enforced. Consequently, it took a toll on their mental health and worse, contributed to potential suicide. Regardless, the emotion distribution of the college students changed as soon as the number of confirmed cases declined and the total lockdown was lifted, shifting to a conditional one and eventually a recovery phase. The frequency of *happy* tweets increased rapidly as the students fully adapted to changes in their study and life environment. However, as the third wave of the pandemic began in September 2020, the percentage of *sad* tweets started to increase again and remained so despite major celebrations took place at the end of the period. In conclusion, the accuracy of the model was validated as it detected the emotion of the college students during the COVID-19 pandemic. The proposed approach can be implemented by healthcare professionals or counselling agencies to monitor and track a patient's emotional states, or to recognise anxiety or systemic stressors [32]. Furthermore, this technology can measure the public mood, which may help social scientists understand the quality of life of a population.

The advancement of technology and the introduction of boosting algorithms causes human interventions to decrease but the outcomes of this study suggested that human intervention is still needed to interpret the semantic meaning of tweets. Regardless, the application of boosting model is just as important in reaffirming the emotions of the tweets. It is therefore hoped that more studies can be conducted to increase the effectiveness of the emotion detection model in identifying the emotion of the Twitter users.

7.0 ACKNOWLEDGEMENT

We gratefully thank the Ministry of Education, Malaysia for the support provided under Fundamental Research Grant scheme, FRGS/1/2017/ICT04/UM/02/2.

REFERENCES

[1] M. Ahlgren, *50+ Twitter Statistics & Facts For 2021 You Should Know About. Website Hosting Rating. https://www.websitehostingrating.com/twitter-statistics/#:%7E:text=There%20is%20a%20total%20of%201.3%20billion%20Twitter%20accounts%2C%20but,monthly%20active%20users%20(MAUs)*, 2021.

[2] M. D. Choudhury, M. Gamon, S. Counts, & E. Horvitz, "Predicting Depression Via Social Media", in *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 7, No. 1, 2013, pp. 128-137.

[3] Y. Li, A. Kazameini, Y. Mehta & E. Cambria, "Multitask Learning for Emotion and Personality Detection", in *IEEE Transaction on Affective Computing*, Vol. 1, No. 1, 2021.

[4] Y. Hswen, J. A. Naslund, J. S. Brownstein, & J. B. Hawkins, "Monitoring Online Discussions About Suicide Among Twitter Users With Schizophrenia: Exploratory Study", in *JMIR mental health*, Vol. *5*, No. *4*, 2018.

[5] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, & H. Ohsaki, "Recognizing Depression From Twitter Activity", in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 3187-3196.

[6] V. Balakrishnan, S. Khan, & H.R. Arabnia, "Improving Cyberbullying Detection Using Twitter Users", in Psychological Features and Machine Learning" in *Computers & Security*, Vol. 90, 2020.

[7] H. Zolkepli, "Malay-Dataset" in *GitHub. https://github.com/huseinzol05/Malay-Dataset#Twitter-Emotion,* 2018.

[8] A. Feizollah, S. Ainin, N. B. Anuar, N. A. B. Abdullah, & M. Hazim, "Halal Products on Twitter: Data Extraction and Sentiment Analysis Using Stack of Deep Learning Algorithms" in *IEEE Access*, Vol. *7*, 2019

[9] T. Mikolov, I. Sutskever, K. Chen, G, Corrado, & J. Dean, "Distributed Representations of Words and Phrases and their Compositionality"in *NIPS 2013*, 2013.

[10] J. Pennington, R. Socher, & C.D. Manning, "GloVe: Global Vectors for Word Representation*",* in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 1532-1543.

[11] Z. Ahmad, R. Jindal, A. Ekbal, & P. Bhattachharyya, "Borrow from Rich Cousin: Transfer Learning for Emotion Detection Using Cross Lingual Embedding*",* in *Expert Systems with Applications*, Vol. *139*, 2020.

[12] Y. Lecun, L. Bottou, Y. Bengio, & P. Haffner, "Gradient-based Learning Applied to Document Recognition", in *Proceedings of the IEEE*, Vol. *86*, No. *11*, 1998, pp. 2278–2324.

[13] M. Schuster, & K. K. Paliwal, "Bidirectional Recurrent Neural Networks", in *IEEE Transactions on Signal Processing*, Vol. 4*5*, No. *11*, pp. 2673–2681, 1997.

[14] A. Firdaus, N. B. Anuar, M. F. A. Razak, & A. K. Sangaiah, "Bio-inspired Computational Paradigm for Feature Investigation and Malware Detection: Interactive Analytics", in *Multimedia Tools and Applications*, Vol. *77*, No. 14, 2017.

[15] N. Jindal, & B. Liu, "Opinion Spam and Analysis", in *WSDM'08 - Proceedings of the 2008 International Conference on Web Search and Data Mining*, 2008, pp. 219-230.

[16] C. G. Harris, "*Detecting Deceptive Opinion Spam using Human Computation*" in *Human Computation AAAI Technical Report*, *WS-12-08,* 2012.

[17] M. Hazim, N. B. Anuar, M. F. A. Razak, & N. A. Abdullah, "*Detecting Opinion Spams Through Supervised Boosting Approach",* in *PLoS ONE*, Vol. *13*, No. *6*, pp. 1–23, 2018.

[18] E. A. Smith, & R. Senter, "*Automated Readability Index",* in *Wright-Patterson AFB, OH: Aerospace Medical Division. AMRL-TR, 66-22,* 1967.

[19] M. F. Rusli, M. A. Aziz, S. R. S. Aris, N. A. Jasri, & R. Maskat, "Understanding Malaysian English (Manglish) Jargon in Social Media", in *Journal of Fundamental and Applied Sciences*, Vol. 10 No. 2S, 2018, pp. 116-125.

[20] J. McCubbin, "*Student Demographics in Higher Education"* in Campus Blog, *https://blog.campus.app/student-demographics-in-higher-education#:%7E:text=Age&text=Age&text=In%202015%2C%20there%20were%2011.8,42.15%25)%20of%20all%20students*, 2018.

[21] Z. Wang, S. A. Hale, D. I. Adelani, P. Grabowicz, T. Hartman, F. Flöck, & D. Jurgens, "Demographic Inference and Representative Population Estimates from Multilingual Social Media Data*"*, in *Proceedings of the 2019 World Wide Web Conference*, 2019, pp. 2056-2067.

[22] P. Sahu, "Closure of Universities Due to Coronavirus Disease 2019 (COVID-19): Impact on Education and Mental Health of Students and Academic Staff*"*, in Cureus, Vol. 12, No. 4, 2020.

[23] A. Muhammad, & A. Kainat, "Online Learning Amid the COVID-19 Pandemic: Students' Perspectives*",* in *Journal of Pedagogical Sociology and Psychology,* Vol. 2, No. 1, 2020.

[24] J. Lee, "Mental Health Effects of School Closures During COVID-19*"*, in *The Lancet Child & Adolescent Health*, Vol. *4*, No. 6, 2020.

[25] S. Sundarasen, K. Chinna, K. Kamaludin, M. Nurunnabi, G. M. Baloch, H. B. Khoshaim, S. F. A. Hossain, & A. Sukayt, "Psychological Impact of COVID-19 and Lockdown among University Students in Malaysia: Implications and Policy Recommendations*",* in *International Journal of Environmental Research and Public Health*, Vol. 17, No. 17, 2020.

[26] A. Pragholapati, "COVID-19 Impact on Students", in *EdArXiv*, 2020.

[27] V. Duong, P. Pham, T. Yang, Y. Wang, & J. Luo, "The Ivory Tower Lost - How College Students Respond Differently Than the General Public to the COVID-19 Pandemic", in *2020 IEEE/ACM International*

Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2020, pp. 126-130.

[28] H. Hassan, "Coronavirus: Mental health issues rise during Malaysia's partial shutdown", in The Straits Times,https://www.straitstimes.com/asia/se-asia/coronavirus-mental-health-issues-rise-during-malaysias-partial-shutdown, 11 August 2020.

[29] S. Li, Y. Wang, J. Xue, N. Zhao, & T. Zhu, "The Impact of COVID-19 Epidemic Declaration on Psychological Consequences - A Study on Active Weibo Users", in *International Journal of Environmental Research and Public Health*, Vol. 17 No. 6, 2020.

[30] X. Li, M. Zhou, J. Wu, A. Yuan, F. Wu, & J. Li, "Analyzing COVID-19 on Online Social Media: Trends, Sentiments and Emotions", in arXiv preprint arXiv-2005.14464, 5 June 2020.

[31] E. M. Aucejo, J. French, M. P. U. Araya, & B. Zafar, "The Impact of COVID-19 on Student Experiences and Expectations: Evidence from A Survey", in *Journal of Public Economics*, Vol. 191, 2020.

[32] M. S. Rahman, N. C. Peeri, N. Shrestha, R. Zaki, U. Haque, S. H. A. Hamid, "Defending Against The Novel Coronavirus (COVID-19) Outbreak: How Can The Internet of Things (IoT) Help to Save The World?", in *Health Policy and Technology*, Vol. 9, 2020, pp. 136-138.