Bulletin of Electrical Engineering and Informatics
Vol. 7, No. 3, September 2018, pp. 479~486
ISSN: 2302-9285, DOI: 10.11591/eei.v7i3.1279

Evaluation of Support Vector Machine and Decision Tree for Emotion Recognition of Malay Folklores

Mastura Md Saad, Nursuriati Jamil, Raseeda Hamzah
Digital Image, Audio and Speech Technology, Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Article Info

Article history:

Received May 04, 2018
Revised Jul 20, 2018
Accepted Aug 17, 2018

Keywords:

Decision Tree
Malay language
Support Vector Machine
Textual emotion recognition

ABSTRACT

In this paper, the performance of Support Vector Machine (SVM) and Decision Tree (DT) in classifying emotions from Malay folklores is presented. This work is the continuation of our storytelling speech synthesis work to add emotions for a more natural storytelling. A total of 100 documents from children short stories are collected and used as the datasets of the text-based emotion recognition experiment. Term Frequency-Inverse Document Frequency (TF-IDF) is extracted from the text documents and classified using SVM and DT. Four types of common emotions, which are happy, angry, fearful and sad are classified using the two classifiers. Results showed that DT outperformed SVM by more than 22.2% accuracy rate. However, the overall emotion recognition is only at moderate rate suggesting an improvement is needed in future work. The accuracy of the emotion recognition should be improved in future studies by using semantic feature extractors or by incorporating deep learning for classification.

Corresponding Author:

Nursuriati Jamil,
Digital Image, Audio and Speech Technology, Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA,
40450 Shah Alam, Selangor, Malaysia.
Email: lizajamil@computer.org

1. INTRODUCTION

Emotion is the ability of a human to express their feelings when interacting with a living being or responding to an event. It plays an important role to express human feelings in daily communications and interactions [1]. However, a human being also has difficulty identifying with their own emotions. A human face and voice have the capability to express emotions [2]. However, emotions detection from a text is widely studied in psychology and has attracted attention in the human-computer interaction (HCI) field. In this field, images and texts are used to identify, recognize and interpret human emotions [1], [3]-[4].

Text emotion recognition is also important in information retrieval (IR) field and has been applied in many application domains [5], such as in data visualization in order to visualize emotion analysis results which can make the interpretation easier for the reader to read, self-organizing maps that can be used for automatically determine writer's emotion and attitudes, a novel application of multimodal emotion recognition algorithm in software engineering [6], and in educational gaming to produce more effectual learning experiences and get a better understanding of affective game design [7]. According to [5], emotion recognition from text has a high potential of useful applications to assist psychologist to understand their patients by analyzing their transcripts. They further stated that the use of emotion recognition from text can be widely used in social media such as on Twitter, Facebook, blog, storybook, poem and other emotional rich text-form documentation.

Text-based emotion recognition is not new and many approaches have been considered to perform the recognition. Our motivation for conducting this research is to recognize emotions from Malay children folklores to be used in an automated storytelling speech synthesis for the Malay language. In our earlier work of storytelling speech synthesis [8], [9], we have developed a storytelling speech synthesizer that is able to synthesize stories from text using a specific storytelling model. However, the synthesizer lacks emotions and also requires a specific storytelling model for a story. Therefore, an emotion recognition engine is essential prior to storytelling speech synthesis to automatically generate the emotions for the synthesizer. Once the emotion is classified, the synthesizer should then construct the emotion models to produce the emotions intonation.

Most works of text-based emotion recognition were done using the English language. In another work by [10], six languages that are English, Spanish, Czech, German, Czech2 and German2 are compared in emotion recognition using text. Since emotion recognition using text is language dependent, a different approach for text pre-processing and classification may be needed. As stated by [11]-[12], different languages have different language structures and not all text pre-processing used in English may be employed in other languages. Furthermore, even though some work on textual emotion recognition in the Malay language has been done, most of them [9], [13]-[14] focused on sentiment analysis using informal languages. Another known work by [15]-[16] used classical literature that is poems and proverbs. As children folklores are consist of simple formal language, we intend to investigate the use of common text feature extraction technique and classifiers to recognize the emotions. This paper is organized as such: Section 1 describes the motivation of this research supported by related literature in Section 2. In Section 3, the emotion recognition methodology is presented followed by the results and discussion of findings in Section 4. Finally, a conclusion and further work are deliberated in Section 5.

2. RELATED WORK

Since the main aim of this paper is to investigate the use of popular methods of feature extraction and classification for the task of textual emotion recognition in the Malay language, the literature on the existing methods are reviewed.

2.1. Text Feature Extraction

Feature extraction is the most important process before classifying the emotion from the text documents. Feature extraction techniques aim to represent the emotional value of the text that will help to classify the emotions into the correct category. There are many textual feature extraction methods such as sentiment analysis, Term Frequency-Inverse Document Frequency (TF-IDF), and unigrams.

Sentiment analysis is widely used to extract emotion recognition from a text document. It tries to understand the attitudes, opinion and emotion in the text by classifying it into either positive, negative or neutral. This technique is widely used to analyze attitudes, moods, and temperaments in social media, user profiling, news articles, and forum discussions. In [17], a survey was conducted showing the popularity of sentiment analysis used to extract emotions from text has been conducted regarding sentiment analysis. Sentiment analysis has also been used at sentence-, document-, aspect-, and user-levels to help extract opinions and emotions [18]. When sentiment analysis is used with natural language processing and machine learning, accurate sentiment results can be achieved. However, sentiment analysis techniques are mostly used to mine opinions from informal text documents comprising mainly spontaneous written speeches.

Term Frequency-Inverse Document Frequency (TF-IDF) is one of the most used text feature extraction technique as it provides a good insight into the important features of the text documents. TF-IDF is used by [19] to extract features from Malay poetry text documents. Other works that employed TF-IDF as feature extraction techniques are [20] and [21] where several classifiers are compared to classify emotions from Thai YouTube comments and Indonesian text documents. In [22], TF-IDF is also used to categorize relevant words in text documents to enhance query retrieval. This simple feature extractor is favoured by many due to its simplicity, robustness and is ideal for short text documents [20],[21]. Coupled with stop words, TF-IDF has shown to improve the classification of emotions from text documents. In this paper, TF-IDF is chosen as the feature extraction technique.

2.2. Text-based Emotion Classification

After the text features are extracted, the classification of these features is done to categorize the emotions into several categories such as happy, angry, sad or fearful. In this section, we review several classifiers that were used in textual emotion recognition in the literature. Li *et al*. [20] performed social emotion detection on short texts of news headlines and sentences (less than 4 words) using the hybrid neural network (HNN). Their method outperformed outperforms the baselines of SWAT used in SemEval-2007, Emotion Term method, Emotion Topic model, Multi-label supervised topic model, Sentiment Latent Topic model, and Affective topic model. Even though HNN shows promising results, the model is complicated and difficult to implement given our limited data.

In [21], four classifiers which are Naïve Bayes, K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and Machine-Sequential Minimal Optimization (SVM-SMO) are compared to recognize emotions in Indonesian folklore. One thousand documents ranging from 1-3 sentences are collected and the emotion of each sentence is labelled using the WordNet Affect List. The highest accuracy is achieved by SVM-SMO, followed by SVM, Naïve Bayes and KNN. Sarakit *et al*. [21] also compared three classification methods to categorize emotions from Thai language YouTube comments. A total of 2,771 comments from music videos and 3,077 comments from commercial advertisements are manually annotated and used as the experimental datasets. Naïve Bayes, SVM and decision tree classifiers are further used to recognize the emotions from the Thai comments. SVM outperformed Naïve Bayes and decision tree classifiers achieving an accuracy rate of 82.28%. Further work such as [1] has also shown that SVM produced better accuracy in the classification of emotions in 1000 news headlines from CNN and Google news. An SVM model is able to outdo three other systems that participated in the SemEval 2007 emotion annotation task. Literature suggested that SVM is a suitable classifier for emotion recognition from textual documents.

Another popular classifier is Decision Tree (DT) that is usually used in bioinformatics [6], data mining [23], and capturing knowledge in the expert system. DT offers flexibility and robustness due to its transparent nature by providing possible alternatives [24],[25]. The most important thing is that decision tree classification can reduce the ambiguity in decision making which leads it to a better classification. In [24], DT achieved an accuracy rate of 84.37%. In this paper, we compared SVM and DT performance to classify emotions into four categories that are happy, angry, fearful and sad.

3. METHODOLOGY

The main stages of textual emotion recognition are data collection, text pre-processing, feature extraction and emotion classification. Each stage is discussed further in this section.

3.1. Data Collection

The dataset used in this paper consists of Malay children short stories. The stories are collected from "Ollie Si Gajah" and "200 Kisah Teladan Haiwan". Only stories in dialogue form are selected because emotions are easily expressed in dialogue compared to narrations. A total of more than 200 short stories are collected, each story ranging from 20-50 words. Examples of two short stories are given in Table 1. The short stories are further broken down into sentences or phrases for emotion annotation. At this point onwards, each sentence or phrase is referred to as a document.

Table 1. Examples of Short Stories in our Datasets

No.	Short Story (in Malay)	Short Story (in English)	Story Title
1	*Tolong! Tolong!* *Tidak ada sesiapa yang mahu menolong saya.* *Saya terpaksa tinggal disini sehingga beberapa hari sehingga badan saya kurus.*	Help! Help! Nobody wants to help me. I have to stay here for a few days until I got thinner.	*Musang yang Tamak* The Greedy Fox
2	*Baik aku bersembunyi di dalam kandang lembu itu.* *Apakah yang kamu buat di sini?* *Tolonglah saya* *Saya diburu oleh seekor anjing pemburu* *Saya ingin bersembunyi di dalam kandang kamu.*	It's better for me to hide in the cow barn. What are you doing here? Help me I was hunted by a dog hunter I want to hide inside your barn	*Rusa yang Malang* The Unlucky Deer

3.2. Pre-processing

Pre-processing stage involves stop-word removal and stemming. Stopword is a common pre-processing process that filters out the meaningless or unnecessary words from each document [26]. Example of stopwords in English is such as 'is', 'for', and 'to'. Meanwhile, examples of Malay language stop words are 'ada', 'boleh', 'tidak', 'kamu', and 'yang'. For our work, we added 'si', 'sang', 'yang','adalah', 'kau' and 'aku' into the collections of stop words done by [27]. Next, the documents are stemmed using a Malay language stemmer to remove inflected words such as 'an', 'kan', 'men','meng', 'ter', 'pe','per' and 'ke', subsequently producing root words. For example in English, 'banks' is stemmed as 'bank' while for the Malay word 'termakan' is stemmed as 'makan'. Figure 1 displays some examples of stopword removals and stemming done on three documents. Once the stop words are removed from the documents and the words in the documents are stemmed, emotion annotation and text feature extraction are done.

Title	Dialog	Removing Stop Word	Remove Stemmer	Emotion
Musang yang Tamak	Tolong! Tolong!	Tolong Tolong	tolong tolong	Fear
1. 24	Tidak ada sesiapa yang mahu menolong saya	Sesiapa yang menolong	siapa tolong	Sad
	Saya terpaksa tinggal di sini hingga beberapa hari	Tinggal	tinggal	Sad
	Sehingga badan saya kembali kurus	Badan kembali kurus	badan kembali kurus	Sad
Rusa yang Malang	Baik aku bersembunyi di dalam kandang lembu itu	Bersembunyi kandang lembu	sembunyi kandang lembu	Fear
2. 61	Apakah yang kamu buat di sini	-	-	Angry
	Tolonglah saya	Tolonglah	tolong	Fear
	Saya diburu oleh seekor anjing pemburu	diburu seekor anjing pemburu	buru ekor anjing pemburu	Fear
	Saya ingin bersembunyi di dalam kandang kamu	Ingin bersembunyi kandang	ingin sembunyi kandang	Fear
	Tapi, kamu akan dibunuh sekiranya pekerja di sini ternampak kamu	dibunuh sekiranya pekerja ternampak	Bunuh kira kerja nampak	Fear
	Saya akan bersembunyi di belakang kamu	Bersembunyi belakang	Sembunyi belakang	Fear
	Nasib kamu agak baik	Nasib agak	Nasib agak	Happy
	Mereka tidak nampak kamu di sini	Nampak	Nampak	Happy
	Saya rasa lebih selamat berada di sini	Rasa selamat berada	Rasa selamat berada	Happy
Singa dengan Gajah	Kamu sepatutnya berbangga kerana mempunyai kekuatan yang lebih baik	Berbangga kekuatan haiwan	Bangga kuat haiwan	Angry
3. 48	Mengapakah kamu tidak pernah berpuas hati?	Berpuas hati	Puas hati	Angry
	Lagipun kamu tidak boleh menyalahkan saya	lagipun menyalahkan	Lagi salah	Angry

Figure 1. Stopword removal, stemming and emotion annotation

3.3. Emotion Annotation

The next step is to create a ground truth dataset for the classification experiment. For this purpose, we hired a human annotator from a Language Academy to manually label the emotional states of all the documents and selected 100 documents. Each document is labelled using the words contained in the documents. For example, a document "*Tiada sesiapa yang mahu menolong saya"* is pre-processed producing the words "*siapa"* dan "*tolong*". These words are categorized as sad, thus the document is labelled as a sad emotion. If there are contradicting labelled emotions in the document, the highest frequencies of the emotional labelled words are used to determine the sentence's emotion. Out of the 100 documents, 25 documents are classified as Sad, 25 as Fear, 25 as Angry and 25 as Happy emotions. In the last column of Figure 2, the emotions of the documents are given. For classification purpose, eighty (80) % of the total documents are used for training and another 20% of the collected documents are used for testing. Figure 2 shows examples of TF-IDF of several words in the document. The emotion category of each word is also stated in the last column.

1	d	Words	Cumulative Frequency	Tf-idf	Happy	Angry	Sad	Fear
2	d	tolong	16	-1.751	1			
3	d	siapa	1	-0.8623				1
4	d	tinggal	2	-0.07656			1	
5	d	badan	2	-0.07656				1
6	d	kembali	3	-0.0662			1	

Figure 2. TF-IDF and its corresponding emotions

3.4. Text Feature Extraction

Term Frequency-Inverse Document Frequency (TF-IDF) is a text mining technique used to extract features from a text. TF-IDF measures how important the words are in the documents. Calculation of TF-IDF is shown as in equation 1. Term frequency will measure how frequent the words appear in a document. This is because every document has different length of words while inverse document frequency is to measure how frequent the word appears for all documents and all terms are considered important.

$$tf_{ij}idf_i = tf_{ij}log_2(\frac{N}{df_i}) \tag{1}$$

where; f_{ij}=frequency of term i in document

$tf_{ij}=\frac{f_{ij}}{max(f_{ij})}$

df_i=No. of docs containing term i

idf_i=idf of term i=$log_2\,(\frac{N}{df_i})$

N=Total no. of docs

3.5. Classification

Support Vector Machine (SVM) is a supervised machine algorithm and commonly used in classification and regression challenges. It plots each data items as a point in *n*-dimensional space which represents the numbers of features. Then, it will use hyper-plane to differentiate between features and class of emotion. The SVM model type that is used in this training is Fine Gaussian SVM, with kernel scale of 0.43 and box constraint level is 1. Figure 3 shows the Support Vector Machine basic flow diagram of the emotion classification, where 80% is used as training dataset and 20% as testing dataset.

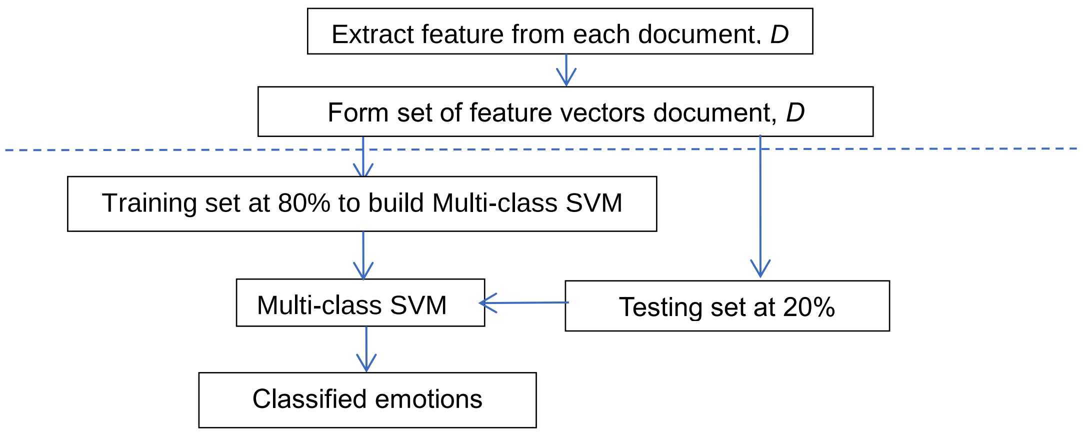

Figure 3. Support Vector Machine diagram

Decision Tree (DT) is a form of a tree structure used in classification and regression model. It works by breaking down the datasets into smaller subsets, incrementally developed them into nodes and leaves. The branches of the decision tree represent the category of the datasets. In this paper, we used DT of Complex Tree model type with the maximum number of splits set to 100. Split criterion is Gini's diversity index and the surrogate decision splits if off. The decision tree is split into 4 emotion classes: happy, sad, angry and fear, The goal of the decisions tree is to ensure it achieved maximum separation among classes at each level. Figure 4 shows the decision tree framework which is applied to four categories of emotion classes.

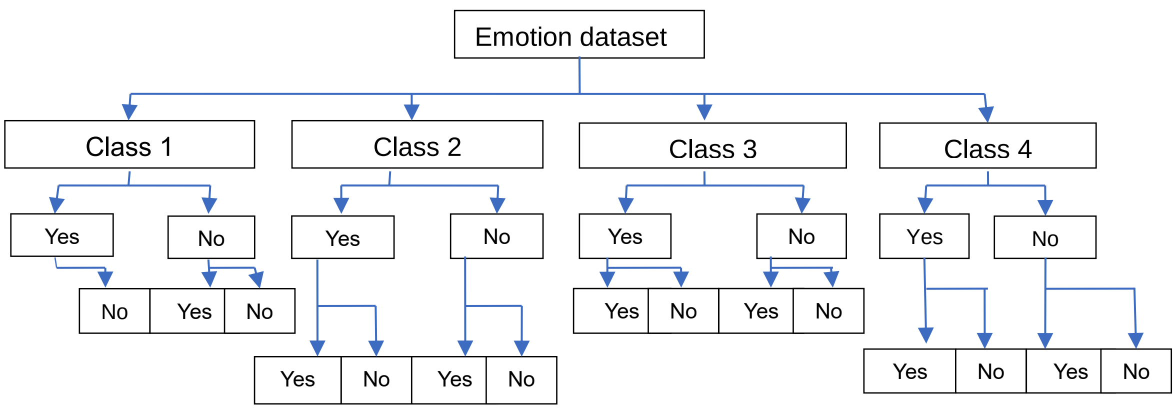

Figure 4. Decision Tree diagram

4. RESULTS AND ANALYSIS

Results of the SVM and DT classifications are presented based on training and testing datasets. Training dataset comprises 80 documents with a total of 320 words, while testing dataset is consists of 20 documents of 80 words. In Table 2, the results are presented and the findings are discussed. In this paper, we use recall, precision, F-measure and confusion matrix to measure the performance of the emotion classification. Precision also called positive predictive value is the number of documents correctly labelled as belonging to the positive class. On the other hand, recall or sensitivity is the number of documents which are not labelled as belonging to the positive class but should have been. Another measurement that combines recall and precision is F-measure. F-Measure indicates how precise the classifier is (how many instances are correctly classified) as well as its robustness (it does not miss a significant number of instances). The final measure is accuracy that refers to how well a given classifier works in classifying the document. Calculations of all the measurements are given in Equation 2 to 5.

$$Precision = \frac{True\ positive}{True\ positive\ +\ False\ positive} \tag{2}$$

$$Recall = \frac{True\ positive}{True\ positive\ +\ False\ negative} \tag{3}$$

$$Accuracy = \frac{True\ positive\ +\ True\ negative}{True\ positive\ +\ True\ negative\ +\ False\ positive\ +\ False\ negative} \tag{4}$$

$$F\text{ - }measure = 2\ x \frac{Precision\ x\ Recall}{Precision\ +\ Recall} \tag{4}$$

Table 2. Support Vector Machine and Decision Tree Classification Results

Classification Method	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
		Training Results		
Support Vector Machine	36.9	36.11	32.5	34.44
Decision Tree	53.1	28.75	28.75	28.75
		Testing Results		
Support Vector Machine	30.0	14.41	12.5	17.65
Decision Tree	62.5	23.72	25	23.32

Overall, DT classified a document better than SVM both using training and testing datasets by achieving 53.1% accuracy as compared to SVM at an accuracy of 36.9% during training and 62.5% and 30% in testing, respectively. This indicates that DT classifies the documents better than SVM. Recall, precision and F-measure are also calculated to further support the performance of the classification. As can be seen from Table 2, recall, precision and F-measure of DT outperformed SVM during testing. This indicates that DT has a higher sensitivity than SVM, precisely classified documents better and is more robust than SVM. However, results of the training dataset interestingly showed SVM achieved a higher percentage for recall, precision and F-measure. To understand the results better, we analyzed the emotions based on each emotion classification on the testing dataset.

Table 3 shows the performance evaluations of DT and SVM based on each emotion class. Using DT, happy emotion achieved the highest accuracy and performed moderately well for precision, recall and F-measure. This is followed by angry, sad and fear emotions. Similar to happy emotion, precision and recall of sad and angry emotions scored equally moderate. This implies that DT is able to correctly classify the happy, angry and sad moderately. However, fear emotion has a higher recall rate but a low precision rate. This indicates a high false positive rate for fear emotion. Upon further analysis of SVM, it shows that SVM performed miserably for fear and angry emotions. Happy emotion achieved the highest accuracy rate followed by a sad emotion. Their respective recall and precision rates are also equally moderate. In terms of classifying emotions into their respective classes, fear emotion seemed to be the most difficult emotion. However, for the other emotions, no conclusive findings can be drawn from the results. A confusion matrix is constructed for the testing dataset to further understand the emotion classifications. The matrix is shown in Table 4.

Table 3. Performance Evaluations based on Emotions

Emotion	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
		Decision Tree		
Fear	52.5	25.0	45.0	32.14
Sad	62.5	14.29	10.0	11.77
Angry	65.0	16.67	10.0	12.5
Happy	70.0	38.9	35.0	36.85
		Support Vector Machine		
Fear	0	0	0	0
Sad	50.0	18.75	15.0	33.75
Angry	0	0	0	0
Happy	70	38.9	35.0	36.85

Table 4. Confusion Matrix of the Testing Dataset

Decision Tree					Support Vector Machine				
Fear	9	0	7	4	Fear	0	4	16	0
Sad	13	2	1	4	Sad	1	6	13	0
Angry	10	5	2	3	Angry	7	11	0	2
Happy	4	7	2	7	Happy	7	11	1	1
	Fear	Sad	Angry	Happy		Fear	Sad	Angry	Happy

Table 4 shows that DT wrongly classifies fear emotion as mostly sad and angry. The same scenario can be seen for Support Vector Machine where 7 fear documents are classified as angry and another 7 documents as happy. DT also classifies 7 angry documents as fear, while 16 angry documents are classified as fear using SVM.

5. CONCLUSION

Out of the four emotions, happy achieved the highest accuracy rate for both Decision Tree and Support Vector Machine with a moderate rate of recall, precision and F-Measure. The overall emotion classification of Malay folklores performed averagely showing DT achieved better results than SVM. Upon analysis of each emotion, fear is the most complicated emotion to be classified. Even though SVM and DT are proven to be a robust classifier for other datasets in previous work, they seem to perform rather miserably producing inconsistent results making it difficult to reach a conclusive finding. We believed that the main problem is the emotion annotation process. When the manual annotation is done by the human annotator, the document is labelled based on the context of the document. For example, the word *'tolong'* can be categorized as sad or fear depending on the context of the document. This may reduce the precision of the classifier. For further improvement of the text-based emotion classifier, semantic text feature extraction is needed with a bigger dataset used for training.

REFERENCES

[1] Kirange D. K., "Emotion classification of news headlines using SVM," *Asian Journal of Computer Science & Information Technology*, vol/issue: 2(5), pp. 104-106, 2013.

[2] Hancock J. T., *et al.*, "Expressing emotion in text-based communication," *ACM SIGCHI Conference on Human factors in Computing Systems*. San Jose, California, pp. 929-932, 2007.

[3] Shivhare S. N. and Khethawat S., "Emotion detection from text," in Wyld D. C., *et al*., "Computer Science & Information Technology (CS & IT)," pp. 371-377, 2012.

[4] Strapparava C. and Mihalcea R., "Learning to identify emotions in text," *ACM Symposium on Applied Computing*, Ceara, Brazil, pp. 1556-1560, 2008.

[5] Herzig J., *et al.*, "Emotion detection from text via ensemble classification using word embeddings," *ACM SIGIR International Conference on Theory of Information Retrieval*, New York, pp. 269-272, 2017.

[6] Polaka I., *et al.*, "Decision tree classifiers in bioinformatics," *Scientific Journal of Riga Technical University Computer Sciences*, vol/issue: 42(1), pp. 118-23, 2010.

[7] Kołakowska A., *et al.*, "Emotion recognition and its application in software engineering," *6th IEEE International Conference on Human System Interaction (HSI)*, Sopot, Poland, pp. 532-539, 2013.

[8] Ramli I., *et al.*, "The first Malay language storytelling text-to-speech (TTS) corpus for humanoid robot storytellers," *Journal of Fundamental and Applied Sciences*, vol/issue: 9(4S), pp. 340-58, 2017.

[9] I. Ramli, *et al.*, "An improved pitch contour formulation for Malay language storytelling Text-to-Speech (TTS). *IEEE Industrial Electronics and Applications Conference (IEACon)*, Kota Kinabalu, Sabah, pp. 250-255, 2016.

[10] Agrawal A. and An A., "Unsupervised emotion detection from text using semantic and syntactic relations," *2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology*, Los Alamitos, vol. 1, pp. 346-353, 2012.

[11] Jamal N., *et al.*, "Poetry classification using support vector machines," *Journal of Computer Science,* vol/issue: 8(9), pp. 1441, 2012.

[12] Mat A. N., *et al.*, "Morphological differences between Bahasa Melayu and English: Constraints in students' understanding," *The Second Biennial International Conference on Teaching and Learning of English in Asia (TELiA2). Langkawi,* pp. 1-11, 2007.

[13] Samsudin N., *et al.*, "Bess or xbest: Mining the Malaysian online reviews," *IEEE 3rd Conference on Data Mining and Optimization (DMO),* Bangi, Selangor, pp. 38-43, 2011.

[14] Kaur J. and Saini J. R., "On classifying sentiments and mining opinions," *International Journal of Emerging Technologies in Computational and Applied Sciences,* pp. 210-214, 2014.

[15] Jamaluddin M. N., *et al.*, "Classification and quantification of user's emotion on Malay language in social network sites using Latent Semantic Analysis," *IEEE Conference on Open Systems (ICOS)*, Langkawi, Kedah, pp. 65-70, 2016.

[16] Noah S. A. and Ismail F., "Automatic classifications of Malay proverbs using Naïve Bayesian algorithm," *Information Technology Journal*, vol/issue: 7(7), pp. 1016-22, 2008.

[17] Povoda L, *et al.*, "Sentiment analysis based on Support Vector Machine and Big Data," *39th IEEE International Conference on Telecommunications and Signal Processing (TSP),* Vienna, Austria, pp. 543-545, 2016.

[18] Kaur H. and Mangat V., "A survey of sentiment analysis techniques," *2017 International Conference on InI-SMAC (IoT in Social, Mobile, Analytics and Cloud (I-SMAC),* India, pp. 921-925, 2017.

[19] Ahmed K., *et al.*, "Sentiment analysis over social Networks: An overview," *2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* Hong Kong, pp. 2174-2179, 2015.

[20] Li X., *et al.*, "Hybrid neural networks for social emotion detection over short text," *2016 International Joint Conference on Neural Networks (IJCNN),* Vancouver, Canada, pp. 537-544, 2016.

[21] Sarakit P., *et al.*, "Classifying emotion in Thai youtube comments," *2015 6th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES),* Hua-Hin, Thailand. Mac, pp. 1-5, 2015.

[22] Jamal N., *et al.*, "Poetry classification using support vector machines," *Journal of Computer Science,* vol/issue: 8(9), pp. 1441, 2012.

[23] Wilkinson P., "Affective educational games: Utilizing emotions in game-based learning," *2013 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES),* Bournemouth, United Kingdom, pp. 1-8, 2013.

[24] Winarsih N. A. and Supriyanto C., "Evaluation of classification methods for Indonesian text emotion detection," *IEEE International Seminar on Application for Technology of Information and Communication (ISemantic),* Semarang, Indonesia, pp. 130-133, 2016.

[25] Sriram S. and Yuan X., "An enhanced approach for classifying emotions using customized decision tree algorithm," *IEEE Southeastcon,* Orlando, Florida, pp. 1-6, 2012.

[26] Wilbur W. J. and Sirotkin K., "The automatic identification of stop words," *Journal of Information Science,* vol/issue: 18(1), pp. 45-55, 1992.

[27] Sadjirin R. and Rahman N. A., "Efficient retrieval of Malay language documents using latent semantic indexing," *2010 International Symposium in Information Technology (ITSim),* Kuala Lumpur, vol. 3, pp. 1410-1415, 2010.