Information Engineering Express
International Institute of Applied Informatics
2023, Vol. 9, No. 2, IEE774

# A Study of Sentiment Analysis based on Specific 6-emotion Category for Thai Language

Yasuto Nishiwaki [*], Masataka Oshima [*],
Kiyota Hashimoto [†], Kazuhiko Tsuda [‡]

## Abstract

Social media such as Twitter, Facebook, Web logs and Review sites are indispensable tools for our customers' communication sites. From a business perspective, it is important to improve customer satisfaction and customer insights by capturing and analyzing customer emotions in detail through social media, customer feedback from call centers, and questionnaire analysis. This paper presents an effective classification method for Thai. In order to solve the problem of linguistic difficulty in Thai, this method used sentiment analysis using 6-emotion as an aspect-based analysis method in addition to conventional sentiment analysis such as positive and negative. This paper describes the results of evaluating the usefulness of the 6-emotion analysis in helping to judge positive and negative in Thai sentences.

*Keywords:* Sentiment analysis, social media, Natural Language Processing

## 1 Introduction

In recent years, as people's values have diversified, customer needs have become more diverse and complex, and we are required to provide products and services that meet customer values in a timely manner. However, capturing the customer's sense of values and reaction depends on qualitative human intuition and the skills of the person in charge, so there is a problem that it is difficult to respond in a timely and uniform manner.

From a business perspective, it is important to capture detailed emotions of customers from social listening using social media, customer feedback from call centers, questionnaire analysis, etc., and analyze them to improve customer satisfaction. Setting aside positive opinions, negative opinions and complaints should be dealt with promptly to improve customer satisfaction, and these opinions will be useful information for improving new products and services.

For example, user comments on a new car include, "The seats are in a bad position and the

[*] Hitachi, Ltd., Tokyo, Japan
[†] Prince of Songkla University, Phuket, Thailand
[‡] University of Tsukuba, Tokyo, Japan

ride is uncomfortable," and regarding vacuum cleaners, "It's a hassle to throw away the garbage and it's hard to use." "I thought I'd try it myself, but the button placement was so bad that I had to try again two or three times." If such issues can be efficiently extracted, improvements related to seat positions can be achieved by developing functionality that allows users to easily adjust seat positions, and vacuum cleaners can be operated so that you can throw away the garbage with a single touch. It is possible to easily consider improvement measures such as improving the functionality and changing the UI of the web system to an intuitive layout. In order to deal with these problems, there is text mining as a method to solve them using artificial intelligence technology. Text mining technologies are described in Section 2.1.

In this paper, in order to solve the problem of linguistic difficulty in Thai, this method used not only conventional sentiment analysis such as positive and negative, but also sentiment analysis using 6-emotion as an aspect-based analysis method. Details of the 6-emotion are given in Section 2.2. This paper presents the results of an evaluation regarding the usefulness of the 6-emotion analysis in helping to judge positive and negative in Thai social media sentences.

# 2 Sentiment Analysis by Text Mining

## 2.1 Technology trends in text mining

Text mining is a technology that uses natural language processing technology to analyze documents written in natural language, understand the meaning and content, and statistically process the understood content to extract trends from the document.

Analyzing customer opinions such as inquiries and complaints using text mining is becoming common, but the problem is that development and operation costs are high. Furthermore, corresponding to the update of products and services requires a high cost and takes a lot of time, resulting in the case of not being able to update products and services in time. In addition, it is difficult for companies developing many products and services to prepare text mining that specialized for the features of each product. As a result, there are a lot of problems because of the existing large and heavy text mining tools. In order to solve these problems, a system that can be constructed and updated easily at a low cost is required.

## 2.2 6-emotion analysis

6-emotion analysis is an algorithm used in the initial stage of sentiment analysis in natural language processing technology. In Thai, there are some studies on classification such as positive and negative, but there is no studies and results that go into the details of emotions like the analysis of 6-emotion. Therefore, development of the Thai language processor is currently delayed due to the characteristics of the Thai language.

For example, there is a problem that methods such as dependency and syntactic analysis cannot be taken, which makes the language difficult for people to understand the meaning of sentences.

In addition, in terms of research themes, there are cases using classical classification methods such as SVM and Naive Bayes [1][2][3][4], and version research using deep learning methods [5][6], existing in large numbers.

Statistical machine translation, translation using a huge corpus, and deep learning are also used for translation engines that translate between languages. However, the Thai translation engine is still under development, and there are also process problems in terms of adding emotions after translation.

The analysis of 6-emotion defines a keyword corresponding to each piece of emotional information of joy, anger, sad, happy, love, and evil. This algorithm analyzes sentiment by counting the appearance frequency of these keywords in a sentence. The advantage of the 6-emotion analysis is to define the emotion method of 6-emotion and the keyword corresponding to it and consist of the logic of matching this keyword with the document. Due to such a simple algorithm, it is easy to update the 6-emotion methods and the corresponding keywords. Also, when updating the logic, it is at most a simple one that considers the correspondence to the negative form. Therefore, it is easy to handle with products and services.

### 2.3 Sentiment analysis in Thai

This time, in order to analyze word-of-mouth information in Thai, text mining using Thai language analysis technology was evaluated. As a result, it was able to acquire positive information relatively accurately, but it was difficult to acquire negative information.

In order to solve this problem, it is essential to improve the accuracy of Thai natural language processing technology, but this requires a lot of time. There are three reasons for this. First, there is no large-scale dictionary (corpus). Second, the number of researchers is small compared to other languages. The third point is that the language processing technology is not advanced, as there is no natural language processing engine with basic functions such as dependency analysis.

Depending on the results of this evaluation, it may be possible to utilize the advantages of the 6-emotion to analyze customer feedback, such as inquiries and complaints, even in Thai. This time, evaluation was performed using 6-emotion (happy, joy, angry, hate, sad, surprise) different from the usual 6-emotion. The reason for the adoption is that we aim to analyze the customer's emotions more accurately by conducting multimodal sentiment analysis that combines text (words) sentiment, facial expression emotions, and voice emotions. This 6-emotion is also based on Ekman's theory of facial expressions [7].

## 3 Methods of Thai 6-emotion Analysis

### 3.1 Method of analysis

Figure 1 shows the overall analysis method. Each rectangle shows the process of data analysis. Each drum contains of text data, word data and category types, etc. Bracket A indicates target test data. Bracket B shows 6-emotion dictionary, and Bracket C shows additional rules for handling the positive and negative judgement.

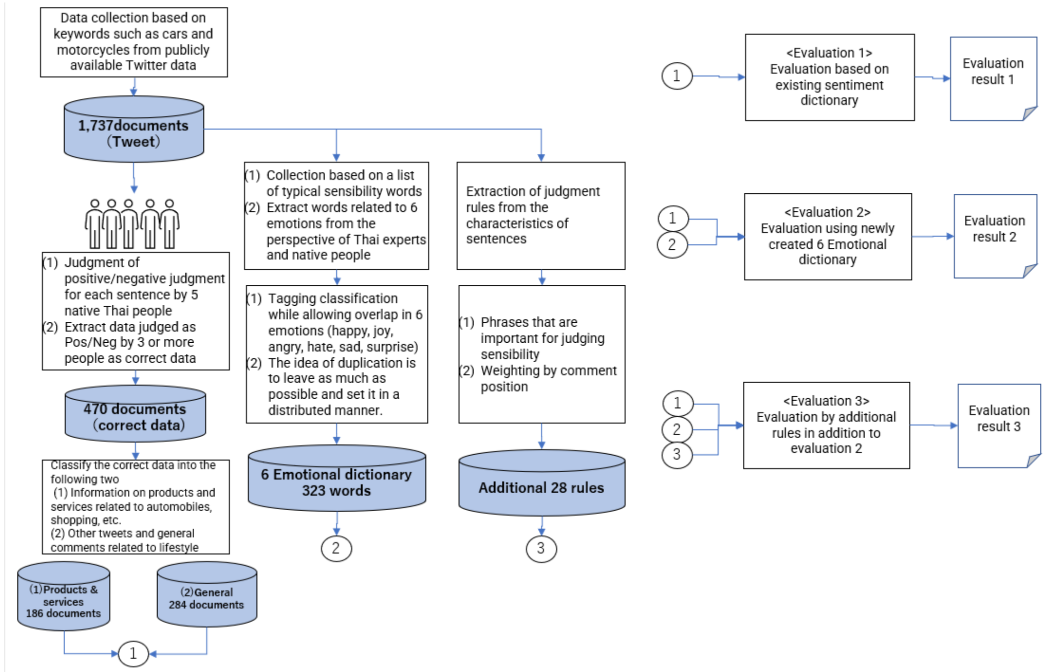


Figure 1 Overall analysis method

Each step of the analysis method is as below.

**(1)** Collected 1,737 documents of information data containing keywords such as car company name, car/motorcycle, job news, etc. from public Twitter data. And 470 records were targeted as correct data that were judged to be positive or negative by three or more people.

**(2)** Confirm that the correct data aggregated in (1) is unbiased from the perspective of the company that sells products and services, the perspective of the public and services, and general comments related to individual thinking and culture.

Case 1; Items related to product services (car-related, restaurants/food, beauty): 186 documents

Case 2; Others (politics, news, media, idols, daily life, lifestyle): 284 documents

**(3)** From the collected 1,737 documents of Twitter data, the parts related to typical emotional words were extracted by Thai experts and Thai speakers.

**(4)** 323 words were extracted from the sentiment/polarity dictionary, which have a relatively high number of expressions directly related to emotion in Thai. Focusing on the 6-emotion, happy, joy, angry, hate, sad, and surprise, and while allowing duplication, the words that could be used in both meanings were set in a distributed manner to the extent possible. As a result, it was created consisting of 519 words, allowing for duplication. There are 289 positive (joy, happy) words and 286 negative (angry, sad, hate) words, and the ratio is roughly half. Table 1 shows an example and Figure 2 shows the number of words by sentiment.

Table 1 6-emotion dictionary example

| | | | | joy | happy | surprise | angry | sad |
|---|---|---|---|---|---|---|---|---|
| # | Thai | Expressions directly related to sentiment | Transtate Thai into English | 喜 | 楽 | 驚 | 怒 | 哀 |
| 1 | มันเขี้ยว | （かじりつきたくなるほど）かわいい | Cute (so much that you want to bite) | 1 | | | | |
| 2 | เย็นชา | （心や言動が）冷たい | Cold (heart and behavior) | | | | | |
| 3 | ตัดใจ | あきらめる | Give up | | | | | 1 |
| 4 | ใฝ่ฝัน | あこがれる | Longing for | | 1 | 1 | | |
| 5 | สด | 鮮やかな | Brilliant | | 1 | | | |
| 6 | รสชาติแปลกๆ | 味が変な | The taste is strange | | | | | 1 |
| 7 | หวาน | 甘い | sweet | 1 | | | | |
| 8 | มักง่าย | いい加減な | Sloppy | | | | 1 | |
| 9 | น้อยใจ | いじける | Tweak | | | | | |
| 10 | แกล้ง | いじめる、困らせる | Tease, annoy | | | | 1 | |
| 11 | เต็ม | いっぱい | full | 1 | 1 | | | |
| 12 | หงุดหงิด | イライラする | get annoyed | | | | 1 | |
| 13 | ราคาญ | ウザい（若者言葉） | annoying (Youth language) | | | | 1 | |
| 14 | ลำไย | ウザい（若者言葉） | annoying (Youth language) | | | | 1 | |
| 15 | อิจฉา | うらやむ | envy | | | 1 | | 1 |
| 16 | ระเบ็ง | うるさい | noisy | | | | 1 | |
| 17 | เสียงดัง | うるさい | noisy | | | | 1 | |
| 18 | จอแจ | うるさい | noisy | | | | 1 | |
| 19 | หนวกหู | うるさい | noisy | | | | 1 | |

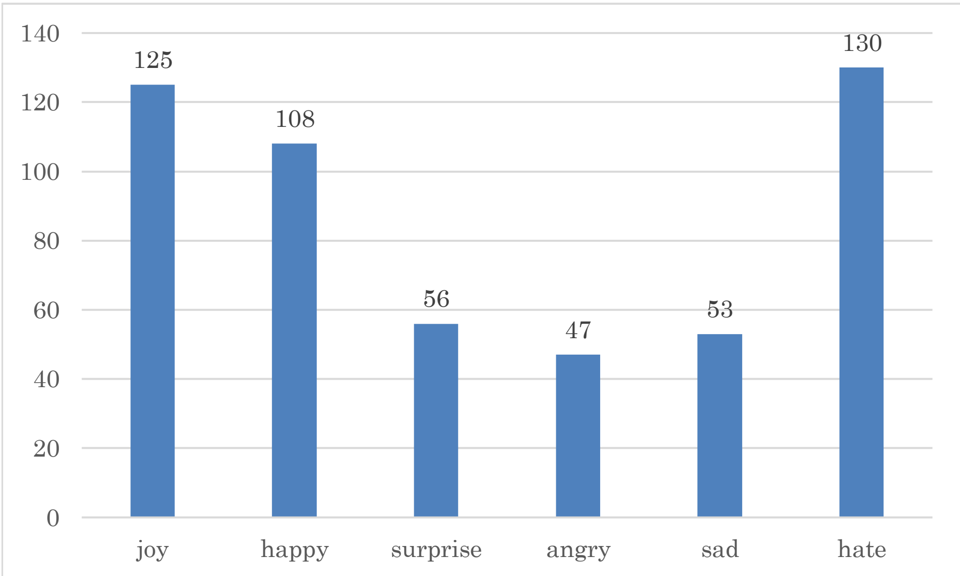


Figure 2 The number of words by sensitivity

**(5)** From the characteristics of sentences in 1,737 records of collected Twitter data, 28 patterns were extracted as two characteristic improvement rules necessary for judging emotions. The first is the addition of phrases related to the 6-emotion. Phrases to express necessary to judge sentiment as a result of data verification. Since it was found to be a frequent expression, it was judged as an additional logic judgment. The second is weighting by comment position. The content spoken at the end of the sentence was judged to be easily influenced as emotions and was used as an element for adding points of positive and negative. Table 2 shows the improvement rules.

**(6)** Evaluation 1 uses the existing sentiment analysis engine [8][9] of the polarity dictionary for judging positive and negative.

**(7)** Evaluation 2 uses a 6-emotion dictionary (323 words) created mainly for adjectives that have sentiment and evaluates positive and negative. A single word can have both positive and negative connotations. In such cases where the same word contained multiple sentiments, one word

was classified as positive or negative to avoid duplicate counts. For example, "surprise" present in both positive or negative is determined by adding the larger number of positive or negative words in the document. Finally, it was determined whether a sentence contained multiple sensitivities or whether the sentence was positive or negative depending on the number of emotions.
**(8)** For Evaluation 3, in addition to the 6-emotion dictionary used in Evaluation 2, evaluation is performed using the original improvement pattern created in (5).

Table 2 Example of additional rules (Total 28 patterns)

| # | Additional rules | Meaning |
|---|---|---|
| Rule1 | ไม่ชอบ | I don't like it |
| Rule2 | ไม่ชอบเลยนะ | I don't like it |
| Rule3 | จบใหม่ไม่เอา | not new |
| Rule4 | มันดูขี้อิจฉาเฉยๆอะ | It just looks jealous |
| Rule5 | ยังมีน้อยมากเล | There are still very few |
| Rule6 | มันก็ฟังไม่รู้เรื่องอยู่ดี | it doesn't sound good. |
| Rule7 | 555/ห้าห้าห้า | ha-ha-ha |
| Rule8 | เชื่อมั่น | confident |
| Rule9 | ก้าวไป | step on |
| Rule10 | ฝ่าฟันไปได้ | get through/ Overcome |
| Rule11 | มาไกลมาก | come a long way |
| Rule12 | ฮีล | heal |
| Rule13 | แรงใจในชีวิต | life motivation |
| Rule14 | เรื่องดีๆ | good story |
| Rule15 | รณรงค์ | campaign |
| Rule16 | ร่วมบริจาค | donate |

# 4 Evaluation

## 4.1 Results of evaluation based on a conventional sentiment analysis dictionary

Results of evaluation based on a conventional sentiment analysis dictionary (evaluation 1) It used an engine [7][8] created based on Turney's evaluation method to carry out the evaluation. The result shows in Figures 3,4 and 5. Figure 3 shows the result of conventional analysis based on all data. Figure 4 shows the result of conventional analysis based on products and services data. Figure 5 shows the results of conventional analysis based on the general data.

| | | Actual condition | | | | |
|---|---|---|---|---|---|---|
| | | TRUE | FALSE | | Accuracy | 79.36% |
| Predicted conditon | TRUE | 307 | 9 | | Error_rate | 20.64% |
| | FALSE | 88 | 66 | | Recall | 77.72% |
| | | Accuracy rate of Pos | Accuracy rate of Neg | | Precision | 97.2% |
| | | 97.2% | 42.9% | | F1 | 86.4% |

Figure 3 Conventional analysis (Overall)

| | | Actual condition | | | | |
|---|---|---|---|---|---|---|
| | | TRUE | FALSE | | Accuracy | 78.49% |
| Predicted conditon | TRUE | 137 | 3 | | Error_rate | 21.51% |
| | FALSE | 37 | 9 | | Recall | 78.74% |
| | | Accuracy rate of Pos | Accuracy rate of Neg | | Precision | 97.9% |
| | | 97.9% | 19.6% | | F1 | 87.3% |

Figure 4 Conventional analysis (Products and services)

| | | Actual condition | | | | |
|---|---|---|---|---|---|---|
| | | TRUE | FALSE | | Accuracy | 79.93% |
| Predicted conditon | TRUE | 170 | 6 | | Error_rate | 20.07% |
| | FALSE | 51 | 57 | | Recall | 76.92% |
| | | Accuracy rate of Pos | Accuracy rate of Neg | | Precision | 96.6% |
| | | 96.6% | 52.8% | | F1 | 85.6% |

Figure 5 Conventional analysis (General)

Overall, both recall and accuracy were in the 70% range, which is not a state of high accuracy. In terms of evaluation patterns, the rate of negative correct answers was low at 19.6% to 52.8% for both (1) products/services and (2) general.

## 4.2 Results of evaluation based on 6-emotion dictionary

Figures 6,7 and 8 show the results of analysis using a dictionary of 6-emotion. Figure 6 shows the result of 6-emotion dictionary analysis based on all data. Figure 7 shows the result of 6-emotion dictionary analysis based on products and services data. Figure 8 shows the results of 6-emotion dictionary analysis based on the general data.

| | | Actual condition | | | | |
|---|---|---|---|---|---|---|
| | | TRUE | FALSE | | Accuracy | 85.53% |
| Predicted conditon | TRUE | 364 | 37 | | Error_rate | 14.47% |
| | FALSE | 31 | 38 | | Recall | 92.15% |
| | | Accuracy rate of Pos | Accuracy rate of Neg | | Precision | 90.8% |
| | | 90.8% | 55.1% | | F1 | 91.5% |

Figure 6 6-emotion dictionary analysis (Overall)

| | | Actual condition | | | | |
|---|---|---|---|---|---|---|
| | | TRUE | FALSE | | Accuracy | 89.25% |
| Predicted conditon | TRUE | 161 | 7 | | Error_rate | 10.75% |
| | FALSE | 13 | 5 | | Recall | 92.53% |
| | | Accuracy rate of Pos | Accuracy rate of Neg | | Precision | 95.8% |
| | | 95.8% | 27.8% | | F1 | 94.2% |

Figure 7 6-emotion dictionary analysis (Products and services)

| | | Actual condition | | | | |
|---|---|---|---|---|---|---|
| | | TRUE | FALSE | | Accuracy | 83.10% |
| Predicted conditon | TRUE | 203 | 30 | | Error_rate | 16.90% |
| | FALSE | 18 | 33 | | Recall | 91.86% |
| | | Accuracy rate of Pos | Accuracy rate of Neg | | Precision | 87.1% |
| | | 87.1% | 64.7% | | F1 | 89.4% |

Figure 8 6-emotion dictionary analysis (General)

By using the 6-emotion dictionaries, improvements in Accuracy and Recall were observed compared to cases based on conventional emotion dictionaries. However, the negative accuracy was still low (55.1%).

### 4.3 Results of evaluation based on 6-emotion dictionary with additional rules

Figures 9,10 and 11 show the results of analysis using a dictionary of 6-emotion with additional rules. Figure 9 shows the result of 6-emotion dictionary analysis with additional rules based on all data. Figure 10 shows the result of 6-emotion dictionary analysis with additional rules based on products and services data. Figure 11 shows the results of 6-emotion dictionary analysis with additional rules based on the general data.

| | | Actual condition | | | | |
|---|---|---|---|---|---|---|
| | | TRUE | FALSE | | Accuracy | 92.98% |
| Predicted conditon | TRUE | 390 | 28 | | Error_rate | 7.02% |
| | FALSE | 5 | 47 | | Recall | 98.73% |
| | | Accuracy rate of Pos | Accuracy rate of Neg | | Precision | 93.3% |
| | | 93.3% | 90.4% | | F1 | 95.9% |

Figure 9 6-emotion dictionary analysis with additional rules (Overall)

| | | Actual condition | | | | |
|---|---|---|---|---|---|---|
| | | TRUE | FALSE | | Accuracy | 96.24% |
| Predicted conditon | TRUE | 173 | 6 | | Error_rate | 3.76% |
| | FALSE | 1 | 6 | | Recall | 99.43% |
| | | Accuracy rate of Pos | Accuracy rate of Neg | | Precision | 96.6% |
| | | 96.6% | 85.7% | | F1 | 98.0% |

Figure 10 6-emotion dictionary analysis with additional rules (Products and services)

| | | Actual condition | | | | |
|---|---|---|---|---|---|---|
| | | TRUE | FALSE | | Accuracy | 90.85% |
| Predicted conditon | TRUE | 217 | 22 | | Error_rate | 9.15% |
| | FALSE | 4 | 41 | | Recall | 98.19% |
| | | Accuracy rate of Pos | Accuracy rate of Neg | | Precision | 90.8% |
| | | 90.8% | 91.1% | | F1 | 94.3% |

Figure 11 6-emotion dictionary analysis with additional rules (General)

The positive correct answer rate decreased from 97.2% to 93.3%, but the accuracy improved from 79.36% to 92.98%. Precision also improved significantly from 77.7% to 98.7%, and Negative accuracy rate improved from 42.9% to 90.4%. Comparing the harmonic mean F value, it increased by 9.6 points from 86.4% to 95.9%. When comparing (1) product/service categories and (2) general comments, there was no major deviations, and there was a similar tendency for the negative correct answer rate to improve significantly.

Looking at the usage of the 6-emotion words in Figure 12, 119 of the 6-emotion words (323 words) registered in advance for making emotional judgments were actually used in this data. In this data, there were 62 sentiment words used for positive judgment and 57 sentiment words used for negative judgment. A variety of sentiment words were used in both judgments, and the use of a few sentiment words did not determine most of the judgments.

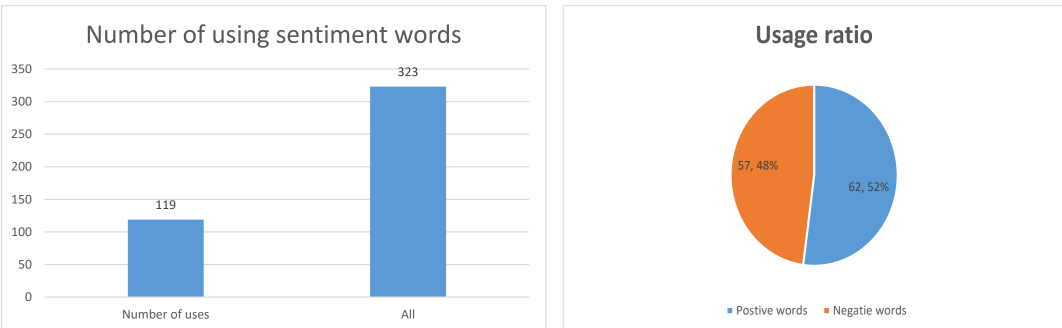


Figure 12 Number of using sentiment words and usage ratio

## 5 Conclusions

In addition to 6-emotion dictionary as a sentiment analysis method for immature natural language processor for Thai, weighting (the latter half of the sentence shows strong emotions) and rules for sentences that contribute to emotions were used in this study. As a result, compared to analysis using a conventional sentiment polarity dictionary, certain results were achieved in terms of increasing the precision of negative emotions, which is important in social media analysis such as word of mouth.

In this paper, by creating a dictionary of 6-emotion as an analysis based on emotions, it showed that the accuracy can be improved by developing the dictionary. In addition, sentences related to emotions are more likely to be reflected positively or negatively by being used in the latter half of the sentence, and by weighting emotions based on the business domain (in this case, the automotive domain). It was shown that the feeling is close to what the person expects.

For example, "no gas" is negative in places where gas is needed, but in the automobile field, as the shift to electric vehicles is progressing, no gas is a positive idea from an environmental perspective. In the future, this work will proceed with research on a methodology to further improve accuracy by weighting each word that expresses emotion, combining it with other words, and incorporating ironic expressions.

## References

[1] K. Takahashi, Doshisha University, Kyotanabe, Kyoto, Japan, ”Remarks on SVM-based emotion recognition from multi-modal bio-potential signals,” IEEE International Workshop on Robot and Human Communication (ROMAN), 22-22 Sept. 2004, DOI: 10.1109/ROMAN.2004.1374736

[2] Udit Jain, Karan Nathani, Nersisson Ruban, Alex Noel Joseph Raj, Zhemin Zhuang, and Vijayalakshmi G.V. Mahesh, ”Cubic SVM Classifier Based Feature Extraction and Emotion Detection from Speech Signals,” 2018 International Conference on Sensor Networks and Signal Processing (SNSP), 28-31 Oct. 2018, DOI: 10.1109/SNSP.2018.00081

[3] N. Sebe, M.S. Lew, I. Cohen; A. Garg, and T.S. Huang, ”Emotion recognition using a Cauchy Naive Bayes classifier,” 2002 International Conference on Pattern Recognition, 11-15 Aug. 2002, DOI: 10.1109/ICPR.2002.1044578

[4] Sagar K. Bhakre; Arti Bang, ”Emotion recognition on the basis of audio signal using Naive Bayes classifier,” 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 21-24 Sept. 2016,DOI: 10.1109/ICACCI.2016.7732408

[5] Peerapon Vateekul, ”A study of sentiment analysis using deep learning techniques on Thai Twitter data,” 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), https://ieeexplore.ieee.org/abstract/document/7748849

[6] Pasupa, K., Seneewong Na Ayutthaya, T., ”Hybrid Deep Learning Models for Thai Sentiment Analysis,” Cogn Comput 14, 167–193 (2022). https://doi.org/10.1007/s12559-020-09770-0;

[7] Ekman, P. (1972).,” Universals and Cultural Differences in Facial Expressions of Emotions,” In Cole, J. (Ed.), Nebraska Symposium on Motivation (pp. 207-282). Lincoln, NB: University of Nebraska Press.

[8] Kunpattanasopon, N., Tongtep, N., and Hashimoto, K, ”Noise Reduction Effect on Thai Social Texts Sentiment Analysis,” The Joint International Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2017): 104-115, 2017.

[9] Turney, P. D. ” Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews,” In: Proceedings of the 40th annual meeting on association for computational linguistics. 417-424, 2002