Journal of Information Science and Technology

JIST Journal of Information Science and Technology
Volume 13, NO 1 | JAN – JUN 2023 | 17-25
ISSN: 2651-1053 (Online)
ISSN: 1906-9553 (Print)

Mining Users' Intentions from Thai Tweets Using BERT Models

Nattapong Sanchan
School of Information Technology and Innovation, Bangkok University,
Pathum Thani, Thailand.

Received: January 11, 2023; Revised: June 14, 2023; Accepted: June 24, 2023; Published: June 30, 2023

ABSTRACT – In this paper, we explore the mining of users' intentions in text. We viewed that being able to identify the intentions of users expressed in textual data provides us to specifically know aims and what users want to do. In the experiment, we collected tweets, constructed a Thai intention corpus, and performed a binary classification task on the corpus. We investigated the intent classification results derived through the application of 3 different Bidirectional Encoder Representations from Transformers (BERT), Word Embedding, and Bag of Words models. The results revealed that BERT Based EN-TH Cased model outperforms other models in both classification and processing time aspects. It achieves the F1 Score of 0.81 and performs the classification task faster than other BERT models up to 15%.

KEYWORDS: Intent Mining, Intention Mining, Intent Classification, Intent Detection, Text Mining, Natural Language Processing

1. Introduction

Due to the exceptional growth of data, data mining has been playing significant role in extracting meaningful knowledge or patterns in a large amount of data. The extracted knowledge can be beneficially used in several areas. For example, in business areas, companies can predict customers' behaviors and their purchasing trends so that the companies are able to make business decisions, plan business strategies, enhance companies' day-to-day operations, and finally generate cost-effective solutions to increase profits. For financial institutions, banks can analyze their historical data to predict whether to give loans to customers [1]. This provides the financial institutions to save costs and time in the loan approval process. In medical areas, data mining can be also used to find better treatments or cures for different diseases. Moreover, governments can also use data mining techniques to help understand the public's opinions regarding their launched policies. For these reasons, data mining is an important component that facilitates us to discover hidden knowledge in order to be beneficially applied to several applications.

Nowadays, another trend in utilizing data mining is the analysis of user-generated content on social media. Sentiment analysis and opinion mining in textual data have been widely explored to discover opinions expressed in the data. However, the study of sentiment analysis and opinion mining seems to be insufficient. Opinions expressed may not fully specify the purposes or directions that users want to achieve. Being able to identify the intentions of users expressed in the data provides us to specifically know the aims and what users want to do. This leads to the exploration of intent mining.[1]

Whereas intent mining has been largely explored in several languages, there is no resource officially available to conduct such research in Thai. In this paper, we explore the mining of intentions in Thai tweets by collecting Thai tweets and constructing a Thai intent corpus for our experiment. We view that intentions can be discovered in two forms: *explicit intention* and *implicit intention*. When a speaker has an explicit intention, for instance; *I am going shopping*, the listener clearly knows the aim of the speaker that he or she is going to do something. For the implicit intention, the intention is semantically hidden in the statement. For instance, the statement *Despite the fact that we are all unique, we should all be treated fairly and with equal rights*, implies that

[1] The term intent mining and intention mining are used interchangeably in some research work.

Corresponding Author: nattapong.sa@bu.ac.th

the author wants everyone to be treated equally. We collected tweets that express intentions and later annotated those tweets.

In the classification task, we utilized 3 different Bidirectional Encoder Representations from Transformers (BERT) models, that were pre-trained using different Thai datasets, to classify whether the given tweets are *intent* or *not intent*. We observed which model would yield the best result in the classification of intentions. Moreover, we also compared the classification results against word embedding and Bag of Words models. The final results revealed that BERT Based EN-TH Cased Model outperforms the classification of intentions in Thai tweets and performs less time than other BERT models.

2. Related Work

The analysis of intention in natural language has been studied since early 1970. In [2], the term intention was defined as a concept of goals and plans, and later this concept was used in the development of a question-answering system in [3]. Additionally, intention was also defined as the information related to goals, plans, and beliefs that a speaker wants to deliver to a computer, in the form of natural language. The goals can be exemplified as *finding out the location of a shopping mall*, *calling a taxi*, or *finding out how to cook Spaghetti Carbonara* with respect to the plans: *to use the Google Map*, *phone to a local taxi service*, and *search for a recipe on the Internet*. This is known as the concept of *explicit intent* tweets in this paper. The following sections discuss related work that has applied different approaches for detecting intentions in text.

2.1 Term-Based Approach

Tweets have been regarded as a valuable resource for mining user intentions [4–7]. Related work has studied intent mining from commercial tweets as aiming to find new business opportunities between buyers and sellers. [4] developed a system to classify whether tweets contain commercial intentions. The tweets are considered intent when they contain at least one verb and express intentions to perform a commercial activity in a recognizable way. By employing a certain number of relevant intent keywords, together with different classification algorithms, the researchers reported the best recall score of 77.4% and the precision of 57.1% using a Complement Naïve Bayes classifier and a linear logistic regression classifier respectively. However, the researchers did not report the generalization of the classifiers to data in other domains. They only focused on commercial tweets related to buying and selling intentions.

[5] worked on a multi-classification method to classify intent tweets into different categories. They employed intent indicators expressing the intent of the users, which is formed by, for instance; a subject I followed by a verb want to in the sentence *I want to buy an iPhone*. In addition, they also defined the intent keywords which can be a noun, a verb (or a set of verbs), or compound nouns. For instance, the keywords *buy an iPhone* in the previous example. These two definitions were used to identify intent tweets and generate a semi-supervised graph for multiclassification.

2.2 Word Embedding

Word Embedding is a technique that represents words in continuous-valued vectors to capture the internal semantic and syntactic information in text [8]. Examples of well-known word embedding models are GloVe [9], Word2Vec [10], Fasttext [11], and Universal Sentence Encoder (USE) [12]. In intention mining, [11] developed a system to detect intentions from users' utterances through the pre-processing, vectorization, and classification processes. Fastext was used to generate word embedding vectors which were derived from a convolutional network classifier. The results indicate that there is a significant improvement in the classification using Fasttext word embedding. Moreover, [13] enriched their word embedding model using WordNet, The Paraphrase Database (PPDB), and the Macmillan Dictionary to help capture semantic representations in intent text. The researchers revealed that the intent detection results derived from the enriched model outperform those of the original GloVe vectors.

2.3 BERT

BERT is a language presentation model that has been pre-trained using a vast amount of unlabeled text, including Wikipedia text and book corpora. BERT is bidirectionally trained, meaning that, during the training process, the model considers text sequences from either left to right and right to left [14]. Figure 1 illustrates a general architecture of BERT where text representation can be extracted. In the figure, a sentence is tokenized before special tokens *[CLS]* and *[SEP]* are added to indicate the beginning and end of a sentence. *[PAD]* is added to sentences that are shorter than the maximum ones. Next, in an encoding step, each token is mapped to a unique integer representing the token in text. Finally, the vector representation of each term is derived.

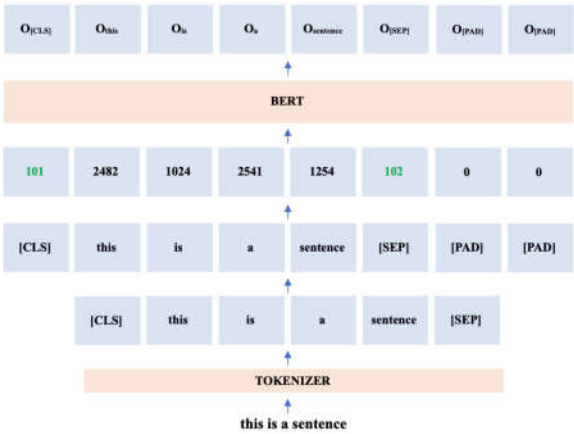

Figure 1. The representation of text extracted from BERT.

BERT has demonstrated its potential to provide state-of-the-art outcomes in a wide range of NLP tasks such as automatic summarization [15], sentiment analysis [16], machine translation [17], and intent classification [18]. However, focus on Thai intent classification. No research work officially reports Thai intent classification results. We explore the utilization of BERT, Word Embedding, and the Bag of Word models in the mining of intentions in Thai text. The following sections summarized key contributions made in this paper.

- Thai intention corpus. Currently, there is no Thai corpus available for conducting mining intention research, especially both explicit and implicit intentions. We, therefore, collected and annotated tweets related to gender non-discrimination. This will open research opportunities for NLP research communities to conduct research in Thai intent mining. We also provide an annotation guideline that could benefit the research communities for creating and expanding other genres of Thai intention datasets in future work.

- Investigation on Different BERT models. Currently, there are three BERT models that were pre-trained with the Thai dataset. One is Google's BERT-Based Multilingual Uncased model and the other two are those from the NLP research communities. At present, there is no observation on which model would provide the best intent mining result in Thai tweets. We, therefore, observe the classification results using the three models.

- Although BERT provides state-of-the-art results, one major drawback is the size of the pre-trained model which leads to the difficulty of the development time and the deployment of the system. [19] proposed a compact version of BERT which is claimed to reduce the processing speed and maintain the accuracy of results. Currently, there is no report to support this claim in Thai intent mining. We aim to observe this claim by comparing the processing time performed through each BERT model.

3. Methodology

In this paper, we view intention mining as a binary classification task. We compared the experimental results among 3 different pre-trained BERT models, word embedding, and Bag of Words approaches through 4 different classification algorithms. Figure 2 illustrates the experimental pipelines which will be discussed in the following sections.

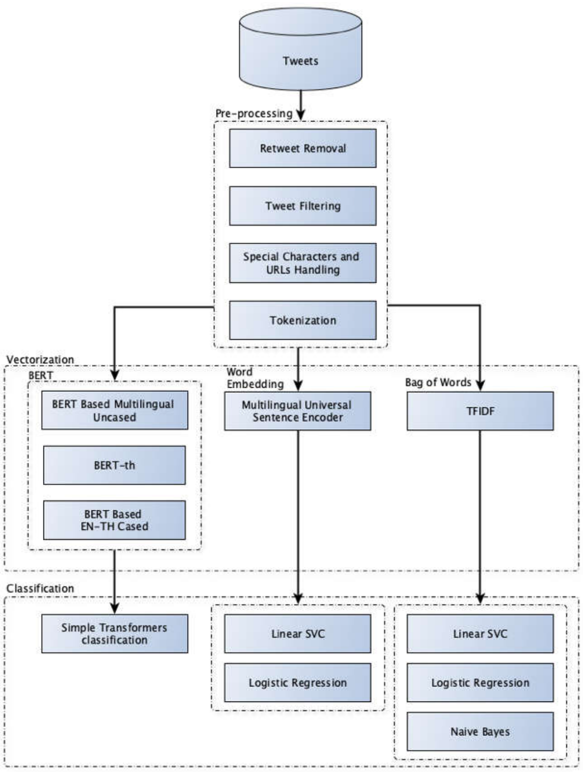

Figure 2. The experimental pipeline of the intent mining shapes 8 pipelines that generate vector representations through BERT, Word Embedding, and Bag of Words approaches.

3.1 Tweet Collection and Annotation

In order to conduct the experiment, we collected Thai tweets via the Twitter hashtag #MarriageEquality which discussed gender non-discrimination, same-sex marriage, and the legislation of same-sex marriage in Thailand, posted before February 2022. Originally, we collected 252,999 tweets. However, as the majority of tweets are redundant tweets and retweets, we used preprocessing techniques to clean the tweets and we obtained distinct 4,310 tweets in total, consisting of 382 intent tweets and 3928 non-intent tweets.

To annotate the data, an annotation guideline was given to two annotators who are fluent in English and aged above 18 years old. We partially reported the inter-annotator agreement with Cohen's Kappa [20] which accounts for 0.53. According to Landis and Koch, this reached the agreement level of moderate agreement, indicating that the annotation is quality [21]. The following section elaborates on how we annotated the collected tweets into one of the two classes: *intent* and *not intent*. Note that, in this paper, we translated Thai tweets and described the exampled tweets in English.

1) Intent

We employed a previous work [4] to identify intent tweets based on the three conditions: 1) tweets have at least one verb that describes an action of the user; 2) elaborate the speaker's attention to perform an activity that relates or would relate to oneself (3) in a recognizable way. When the tweets are considered intent, they can be either *explicit intent* or *implicit intent*:

1.1) Explicit Intent: As its name implies, we refer to explicit intent as the tweets that explicitly express the intention of the authors. Usually, these tweets simply contain the intent keywords such as *want*, *confirm*, *expect*, *call for*, *agree*, *want to see*, *support*, etc. The following tweets exemplify the explicit intent.

- Example 1: "I *want* all marriages to be equal because genders are not divided."
- Example 2: "I still *confirmed* the original statement that what we asked for was the fundamental rights that everyone should have."

1.2) Implicit Intent: This category is opposite to explicit intent in which the tweets do not directly express the intent of the authors. In other words, intent can be hidden and semantically expressed such as in the forms of interrogative (Example 3) and imperative sentences (Example 4).

- Example 3: "How many sets of morals or committees must be used in order to be able to recognize that marriage equality is a fundamental right, not a privilege."
- Example 4: "Become a part of one million names to protect all of the loves and for gender diversity."

2) Not Intent

Tweets are considered as *not intent* when they do not indicate any intentions in the context. Usually, they only describe options related to people, situations, events, or activities. The following examples exemplify the *not intent* tweets:

- *Example 5: "The Equal Marriage Act hanged for another 60 days."* This is an example of a news headline that does not indicate an intention.
- *Example 6: "The Chiang Mai team is accepting volunteers to organize a great marriage equality campaign event on February 14th.*" In this example, no intention is indicated in this tweet. The main idea of this tweet is the description of what is happening.
- *Example 7: "If claiming marriage equality is against religious principles, the law should be repealed because sin should be judged by morality. Law and religion should be distinguished."* This tweet describes the opinions of the speaker, not the intention.

3.2 Preprocessing

Preprocessing techniques are the processes for cleaning textual data. We applied them for cleaning tweets as follows.

1) Retweet Removal

Retweeting is the action of reposting a user's tweet. It generally happens when a user likes and wants to share the tweet in his or her timeline. Retweets are usually indicated by having RT in the tweets. As retweets are redundant, we remove them from the dataset.

2) Tweet Filtering

We found the other kind of redundant tweets in the dataset that should be also removed. Such tweets mostly express the same content, are copied from other users' accounts, and are posted several times by the same users. To filter out these redundant tweets, a similarity metric, the Levenshtein distance, was used. We compared tweets from one to the others and computed the similarity scores among them. We observed that the tweets having similarity scores greater than or equal to 0.7 are usually redundant tweets.

3) Special Characters and URLs Handling

Generally, special characters, URLs, and emotions appearing in tweets do not usually contribute to the analysis of intention. We, therefore, removed URLs, emoticons, and characters @#/(){}? from the dataset

4) Tokenization

Tokenization is a process of breaking down a phrase, sentence, paragraph, or an entire textual document into smaller units i.e., phrases or words. These units are called tokens. In the detection of intentions, we applied a Stacked Ensemble Filter and Refine for Word Segmentation (SEFR CUT) [22] to tokenize Thai tweets into words. After the preprocessing was completed, the tweets were fetched and transformed into vector representations in the next step.

3.3 Vectorization

To investigate which models would provide the best classification results, we generated vector representations for Thai tweets using BERT, Word Embedding, and Bag of Words models.

1) BERT

In this paper, we investigated the classification results derived through different BERT models. Currently, there are three BERT models that support the Thai language, including BERT Based Multilingual Uncased, BERT-th, and BERT Based EN-TH Cased.

- BERT Based Multilingual Uncase: a BERT that was pre-trained with Wikipedia text over 102 languages using Masked Language Modeling (MLM). With MLM, 15% of the tokens are masked and were trained to predict the mask tokens. This allows the model to learn the sentence representation bidirectionally [14].

- BERT-th: a BERT model that was pre-trained using a 2.39 gigabyte-sized Thai Wikipedia dataset. To use the model, we employed the Hugging Face transformer to generate BERT-th vector representations from Thai tweets [23-24].

- BERT Based EN-TH Cased: a compact version of multilingual BERT. [19] described that due to the large size of the original BERT, it is difficult to meet the production requirements when the system is implemented and deployed. Therefore, they proposed a smaller version of the multilingual BERT, implemented by decreasing the vocabulary size which leads to the reduction of the total number of parameters. For this reason, the model can process faster while retaining state-of-the-art results. We additionally observed the classification efficiency and the processing time performed through each model. Later, we compared the processing time of the classification task performed through the BERT models.

2) Word Embedding

To investigate the performance of the classifier using the word embedding approach, we employed the Multilingual Universal Sentence Encoder from the pre-trained Tensor Flow saved model [12]. This model is an extension of Universal Sentence Encoder which was trained and optimized with a great number of sentences, phrases, and paragraphs. We employed the model to generate word embedding vectors. For instance, for each input tweet, the model encodes it and generates a 512-dimensional vector. We later used them as the input to the classification algorithms.

3) Bag of Word

Bag of Words is a representation of an unordered set of words that describes the occurrence of words within a document. In this paper, we represented tweets in the Bag of Word model and extracted Term Frequency-Inverse Document Frequency (TF-IDF). Equation 1 - 2 illustrates the equations of TF-IDF where tf_i indicates the number of times a word i occurs. IDF measures the commonality of a word i that occurs in the entire document set. N represents the number of documents and n_i indicates the number of documents containing the word w_i. The IDF value closer to 0 indicates the more common the word is. The high value of the multiplication of TF and IDF, the rarity of the word occurrence. The final output derived from the model is the TF-IDF vectors which will be fed into the classification algorithms.

$$idf(w_i) = log\frac{N}{n_i} \quad (1)$$

$$tf_i \cdot idf(w_i) \quad (2)$$

3.4 Classification

We employed four classification algorithms to classify intent tweets. The first algorithm is Simple Transformers Classification. Simple Transformers is a transformer-based library that can perform several machine-learning tasks, including text classification. After deriving vector representations through BERT models, we employed a text classification module in the library to classify intentions. Moreover, we also observed the results derived from Linear Support Vector Classification (Linear SVC) and Logistic Regression since recent work has reported that Linear SVC yields significant results in text classification tasks [25–29]. The final algorithm is a probabilistic algorithm based on the Bayes Theorem, Naïve Bayes. We investigated this algorithm as several systems use it as a baseline.

Table 1. The classification results of intent tweets are derived from each pipeline.

Models	Processing Time (Mins)	Precision	Recall	F1 Score
BERT				
BERT Based Multilingual Uncase	23.69	0.59	0.73	0.65
BERT-th	24.91	0.74	0.83	0.78
BERT Based EN-TH Cased	21.01	0.79	0.83	**0.81**
Word Embedding: Universal Sentence Encoder (USE)				
Linear SVC	0.66	0.69	0.83	0.75
Logistic Regression	0.56	0.88	0.69	0.77
Bag of Words				
Linear SVC	0.10	0.48	0.70	0.76
Logistic Regression	0.09	0.87	0.60	0.71
Naïve Bayes	0.07	0.46	0.50	0.48

4. Experiment and Evaluation

In the evaluation of the classifiers, we utilized the 5-fold Cross Validation with a stratified option. In this setting, the examples in both classes are equally split in each fold. We followed the assignment of K = 5 from Kasthuriarachchy, Chetty, Karmakar, and Walls which explored the intent classification within a limited dataset [18]. We reported the classification results using three evaluation metrics, including Precision, Recall, and F1 Score as shown in Equation 3 to Equation 5. Note that TP, FP, and FN indicate True Positive, False Positive, and False Negative examples respectively. We calculated averaged precision, recall, and F1 Score from the aggregated confusion matrix derived in each fold. Precision is a ratio of *intent* examples that are correctly predicted by the classifier. Recall refers to from all *intent* examples how many *intent* examples are correctly captured. The final metric, the F1 score, is the harmonic mean of precision and recall.

$$\text{Precision(P)} = \frac{TP}{TP + FP} \quad (3)$$

$$Recall(R) = \frac{TP}{TP + FN} \quad (4)$$

$$F1\ Score = \frac{2PR}{P + R} \quad (5)$$

Table 1 presents the results of intent classification derived in each pipeline. In this experiment, we compared the classification results derived from three main vectorization approaches, including BERT, Word Embedding, and Bag of Words. By applying BERT, we explored different BERT models that were pre-trained using Thai datasets. Overall, the experimental results indicated that BERT Based EN-TH Cased model outperforms other models in detecting intentions in Thai tweets. The model achieved the averaged precision, recall, and F1 Score of 0.79, 0.83, and 0.81 respectively. Additionally, the intent classification using Word Embedding and Bag of Words yield nearly the same classification results, except those processed through Naïve Bayes.

Moreover, we also explored the processing time in the classification task performed by each model. It is a fact that the smaller size of the model commonly performs faster. However, emphasize large models like BERT. The processing time is longer due to the complexity of the model. Thus, Abdaoui et al. proposed the BERT Based EN-TH Cased which has a smaller number of language parameters so that it requires less memory and therefore loads faster [19]. In this paper, we compared the classification time among the three BERT models. We proved that the BERT Based EN-TH Cased model processes less time than other BERT models while still maintaining state-of-the-art results. It was able to classify the intent tweets faster than BERT-th and BERT Based Multilingual Uncased up to 11.32% and 15.66% respectively.

Aside from assessing the performance of classifiers with the averaged precision, recall, and F1 Score, we also conducted an error analysis to investigate possible factors that would lead to the miss-classification results. The examination of these factors could enhance the intent classification in future work. The tweets shown in Table 2 exemplify the miss-classification examples derived through the application of the BERT Based EN-TH Case model, which provides the best intent classification results.

Table 2. Miss-Classification Examples

Tweet Numbers	Tweets
1	"Regardless of gender, they should be able to love each other freely. We shouldn't block them just by gender."
2	"We are all human. We have equal rights. The era in that a male marries a female has ended. Male-male and female-female couples should have equal rights."
3	"We did not agree that you (the assemblymen) pass it (the Act) to the Cabinet, but the Pheu Thai Party confirmed in its resolution that they agreed with the Equal Marriage Act."

After investigating the classification results, we found that the classifier is substandard in understanding hidden semantics in the context. To illustrate, from the examples in Table 2, Tweet 1 illustrates an implicit intent tweet in which the intention is indirectly stated in the context. The sentence fragments "*should be able to love each other*" and "*shouldn't block them just by gender*" imply that the author intends to support gender non-discrimination. Likewise, the indication of should have equal rights in the second tweet also infers the author's intention. By understanding the semantics of the tweets, the classier would be able to interpret the overall meanings of the tweets. Moreover, the missing world knowledge identification is problematic in the detection of intention in our dataset. As shown in Tweet 3, the author expresses name entities i.e. the "*Cabinet*" and "*Phue Thai*" Party which are a part of the word knowledge in the gender equality domain. Additionally, the terms "*you*" and "*it"* refer to the assemblymen and the Equal Marriage Act led to the misclassification results.

5. Discussion and Conclusion

While intent mining has been extensively studied in several languages, Thai has very limited resources for this genre of research. In this paper, we collected, annotated, and constructed a Thai intention dataset that is related to gender non-discrimination. For future research, we view that the dataset derived from this study could benefit the research communities for researching mining intentions in Thai text. Additionally, our annotation guideline could provide the researchers for creating and expanding other genres of intention datasets in future work.

Moreover, in the classification task, we explored 8 pipelines that generate vector representations through BERT, Word Embedding, and Bag of Words. The main algorithms used are Simple Transformers Classification, Linear SVC, logistic regression, and Naïve Bayes. The results revealed that BERT Based EN-TH Cased model outperforms other models in both classification and processing time aspects.

In our error analysis, we found that lacking hidden semantic and word knowledge identification approaches impedes the success of the intent classification task. In future work, further investigation could be performed on adding a semantic layer to BERT models in order to observe whether semantics would increase the classification performance. In addition, the investigation of world knowledge as new features could enhance the classification task. For instance, [30] used domain-specific and common-sense knowledge to build features. However, this may require time and effort as constructing domain-specific knowledge needs human experts. Furthermore, the work could be extended by performing an automatic summarisation to generate the Thai intent summaries regarding gender equality. The ability to generate summaries of the intentions would provide users to digest a large number of intentions expressed in tweets conveniently.

References

[1] Gupta, V. Pant, S. Kumar, & P. K. Bansal, "Bank Loan Prediction System using Machine Learning," 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART*)*, 4, 12, 2020, pp. 423 – 426.

[2] Schank RC, Abelson RP. Scripts, plans, goals, and understanding. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1997.

[3] Mckevitt P, "Analysing coherence of intention in natural language dialogue," Ph.D. thesis. Exeter: University of Exeter; 1991.

[4] Hollerit, B., Kröll, M., & Strohmaier, M, "Towards linking buyers and sellers: detecting commercial intent on Twitter," In Proceedings of the 22nd International Conference on World Wide Web. 2023, pp. 629-632.

[5] Wang J, Cong G, Zhao XW, Li X, "Mining User Intents in Twitter: A Semi-Supervised Approach to Inferring Intent Categories for Tweets, " The Twenty-Ninth AAAI Conference on Artificial Intelligence. '05, 2015, pp.339-345.

[6] Zhao XW, Guo Y, He Y, Jiang H, Wu Y., & Li, X, "We Know What You Want to Buy: A Demographic-based System for Product Recommendation on Microblogs," The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 08, 2014 pp.1935-1944.

[7] Purohit H, Dong G, Shalin V, Thirunarayan K, Sheth A, "Intent Classification of Short-Text on Social Media," IEEE International Conference on

Smart City/Socialcom/ Sustaincom (Smartcity). 12, 2015, pp. 222-228.
[8] Li, Y., Yang, T, "Word Embedding for Understanding Natural Language: A Survey,"In: Srinivasan, S. (eds) Guide to Big Data Applications. Studies in Big Data, Vol 26. Springer, Cham.
[9] Pennington J, Socher R, Manning CD, "GloVe: Global Vectors for Word Representation," In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* 10, 2014, pp. 1532-1543.
[10] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J, " Distributed representations of words and phrases and their compositionality". Advances in neural information processing systems. The 26th International Conference on Neural Information Processing Systems, 05, 12, 2013, pp. 3111-3119.
[11] Bojanowski P, Grave E, Joulin A, Mikolov, "Enriching word vectors with subword information," Transactions of the association for computational linguistics. 2017, pp. 135-146.
[12] Balodis K, Deksne D, "Intent detection system based on word embeddings," International Conference on Artificial Intelligence: Methodology, Systems, and Applications. 12, 08, 2018, pp. 25-35.
[13] Chidambaram, M., Yang, Y., Cer, D., Yuan, S., Sung, Y. H., Strope, B., & Kurzweil, R, "Learning Cross-Lingual Sentence Representations via a Multi-task Dual-Encoder Model," In Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019), '02, 08, 2019, pp. 250 – 259.
[14] Kim JK, Tur G, Celikyilmaz A, Cao B, Wang Y Y, "Intent detection using semantically enriched word embeddings," The 2016 IEEE spoken language technology workshop (SLT). 13-16, 12, 2016, pp. 414-419.
[15] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K, " BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol 1, 2019, pp. 4171 – 4186.
[16] Grail Q, Perez J, Gaussier E, "Globalizing BERT-based transformer architectures for long document summarization." The 16ndConference of the European Chapter of the Association for Computational Linguistics. 19-23, 04, 2021, pp. 1792-1810.
[17] Hoang M, Bihorac OA, Rouces J, "Aspect-based sentiment analysis using BERT," The 22nd Nordic Conference on Computational Linguistics.
[18] Imamura K, Sumita E. "Recycling a pretrained BERT encoder for neural machine translation," The 3rd Workshop on Neural Generation and Translation. '04, 11, 2019, pp. 23-31.
[19] Kasthuriarachchy, B., Chetty, M., Karmakar, G. & Walls, D. Pre-trained Language Models with Limited Data for Intent Classification. In 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom, 19 – 24 July 2022, pp. 1 - 9.
[20] Abdaoui A., Pradel C., & Sigel, G, "Load What You Need: Smaller Versions of Multilingual BERT," In Proceedings of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing, 20, 11, 2020, pp. 119 – 123.
[21] Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1), pp. 37 – 46.
[22] Landis, J.R. & Koch, G.G. (1997). The Measurement of Observer Agreement for Categorical Data. Biometrics, 33, pp.159.
[23] Limkonchotiwat, P., Phatthiyaphaibun, W., Sarwar, R., Chuangsuwanich, E., & Nutanong, S, "Domain Adaptation of Thai Word Segmentation Models using Stacked Ensemble.," In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 16 – 18, 11, 2020, pp. 3841 – 3847.
[24] Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Rush, A, "Transformers: State-of-the-Art Natural Language Processing," In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 16 – 20, 11, 2020, pp. 38 – 45.
[25] T. Team, "Thaikeras: BERT Pre-training in Thai Language.".[Online]. Available: https://huggingface.co/Geotrend/bert-base-en-th-cased. [Accessed: May 1, 2020].
[26] Kaewnoo, P., & Senivongse, T, "Identification of Software Problem Report Types Using Multiclass Classification," Proceedings of the 2019 3rd International Conference on Software and e-Business. 9 – 11, 12, 2019, pp. 104 –109.
[27] Abdullah Y. Muaad, G. Hemantha Kumar, J. Hanumanthappa, J.V. Bibal Benifa, M. Naveen Mourya, Channabasava Chola, M. Pramodha, R. Bhairava, "An effective approach for Arabic document classification using machine learning," Global Transitions Proceedings, 3(1)(2022), pp. 267 – 271.
[28] Sikarwar, S.S., & Tiwari, D.N, "Analysis of the Sentiments of Amazon Reviews Dataset By Using Linear SVC and Voting Classifier,"

International Journal of Scientific & Technology Research, 2020, pp. 461 – 465.

[29] Gulati, K., Kumar, S. S., Boddu, R. S. K., Sarvakar, K., Sharma, D. K., & Nomani, M. Z. M, "Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to COVID-19 pandemic" Materials Today: Proceedings, 51, pp. 38 – 41.

[30] Sanchan N., Bontcheva K. & Aker, A. "An Adoption of a Contradiction Detection Task to Assist the Summarization of Online Debates," 5th International Conference on Information Technology (InCIT), 21 – 22, 10, 2020, pp. 185 – 190.

[31] Gabrilovich, E., & Markovitch, S, "Feature Generation for Text Categorization Using World Knowledge," IJCAI International Joint Conference on Artificial Intelligence, 30, 07, 2005, pp. 1048 – 1053.