Improving misspelled word solving for human trafficking detection in online advertising data

Chawit Wiriyakun, Werasak Kurutach
Faculty of Engineering and Technology, Mahanakorn University of Technology, Bangkok, Thailand

Article Info	ABSTRACT
	Social media is used by pimps to advertise their businesses for adult services due to easy accessibility. This requires the potentially computational model for law enforcement authorities to facilitate a detection of human trafficking activities. The machine learning (ML) models have been used to detect these activities mostly rely on text classification and often omit the correction of misspelled words, resulting in the risk of predictions error. Therefore, an improvement data processing approach is one of strategies to enhance an efficiency of human trafficking detection. This paper presents a novel approach to solving spelling mistakes. The approach is designed to select misspelled words, the replace them with the popular words having the same meaning based on an estimation of the probability of words and context used in human trafficking advertisements. The applicability of the proposed approach was demonstrated with the labeled human trafficking dataset using three classification models: k-nearest neighbor (KNN), naive Bayes (NB), and multilayer perceptron (MLP). The achievement of higher accuracy of the model predictions using the proposed method evidences an improved alert on human trafficking outperforming than the others. The proposed approach shows the potential applicability to other datasets and domains from the online advertisements.

Corresponding Author:

Chawit Wiriyakun
Faculty of Engineering and Technology, Mahanakorn University of Technology
140 Cheum-Sampan Road, Krathum Rai, Nongchok, Bangkok, 10530, Thailand
Email: chawitmild@gmail.com

1. INTRODUCTION

Human trafficking is a global problem that affects victims all around the world [1]. Women, men, and children can all fall victims to this crime, but women and children are particularly vulnerable due to their physical weakness compared to men. In the past, pimps used face-to-face communication to reach their customers, but this method made it easier for law enforcement authorities to track them down. Nowadays, pimps have shifted to use the internet to conduct their business because it is more convenient and secure [2]. Online advertisements are the primary means through which pimps and customers connect, making it a challenging task for law enforcement authorities to track them. Therefore, the use of a machine learning (ML) model to detect human trafficking activities in online advertisements is key to stop this vicious cycle. Such models have gained attention due to their abilities to efficiently handle large amounts of data and automate the screening process, ultimately reducing workload and time.

Machine learning [3] is one branch of artificial intelligence that focuses on teaching machines through data to build models to achieve goals, such as text classification, face detection, and face recognition. Datasets can be collected in various ways, including application programming interfaces (API) [4], organizations, and crowdsourcing [5]. However, the quality of the data strongly affects model performance,

such as misspelled words [6]. Several tools are required to solve this problem, including Pyspellchecker [7], autocorrect [8], and TextBlob [9]. Unfortunately, these tools are still lacking consideration for analysis of words and their contexts relevant to the domain of interest, which may cause prediction errors, indicating a need to improve data preparation processing.

This paper presents an approach to solving the misspelled words problem in human trafficking datasets. The approach detected and analyzed misspelled words and their contexts, then replaced them with suitable alternatives. For our experimentation, we used the Trafficking-10k dataset [10] provided by Marinus Analytics, which contains many misspelled words due to being collected from online advertisements. The experimental result shows that the proposed approach can improve an accuracy of models in detecting human trafficking activities in advertisements. The rest of the paper is organized as follows: related works are discussed in section 2, section 3 briefly describes the dataset used in the experiment, section 4 proposes the approach, section 5 describes the experiment and its results, section 6 discusses the experimental results, and finally, section 7 discusses the conclusion and future work.

2. RELATED WORKS

A model for detecting human trafficking activities has garnered attention for strengthening the detection of the online illegal activities. McAlister [11] proposed an approach to collect data related to human trafficking in Romania. This method involved scraping data to collect adult services advertisement in English, which was then translated into Romanian for use in searching for similar advertisements in Romania. Roshan *et al.* [5] developed a simple tool for data collection from a mobile application to detect human trafficking. The mobile application collected text and images while the identities of senders were concealed. Suspected information about human trafficking activities was then reported to the authorities for further investigation. However, this work was found to have repeated reports of alteration, which raised questions about the reliability of the data and the limitations of data cleansing, indicating an insufficient potential for practical usage. Mensikova and Mattmann [12] proposed an approach that applied a sentimental analysis to create a model for identifying advertisements potentially connected to human trafficking activities. However, the reliability of the dataset and the possibility of age-based biases led to a question in data prediction errors. Hernandez-Alvarez [13] demonstrated an approach for detecting human trafficking activities on Twitter. This method involved collecting data from Twitter using an API, then cleansing the data and selecting relevant features before creating a model. The model showed the potential for detecting human trafficking activities, but concerns were raised about the completeness of machine-based feature selection, indicating a need for the further improvement of the approach.

Additional processing in data preparation is gaining the interest for completing feature selection and improving model performance. Diaz and Panangadan [14] introduced an approach for preparing the additional datasets for specific human trafficking cases. The dataset was firstly reviewed online by domain experts in sexual trafficking businesses. This certified dataset was then a combination of data from Rubmaps and Yelp, obtained using natural language processing (NLP). Rubmaps is a business review website that identifies businesses engaged in illegal activities. On the other hand, Yelp is an online review forum that provides data location. The approach combined datasets based on location data. The result of this approach was the ability to identify illegal sexual trafficking activities in specific locations. The use of a language model to identify features for model creation was proposed by Zhu *et al.* [15]. The model showed the capability to detect advertisements related to human trafficking activities.

Most of the reported approaches were developed based on text classification, omitting the feature of emojis [11]–[15]. This could reduce the reliability of model predictions when it was applied to real human trafficking activities online. As emojis are another factor expressing intention directly [16], Whitney *et al.* [17] presented an approach using knowledge management and NLP to explore emojis related to human trafficking activities. Unfortunately, the method was not applied to the further creation of the detected model. Additionally, the datasets used in most reports were not labeled by domain experts [11], [13], [15], reflecting that those results from the reported models could be questioned.

To improve the justification of data predictions obtained from the model, Wiriyakun and Kurutach [18] demonstrated the use of local interpretable model agnostic explanations (LIME) [19] to filter words related to human trafficking as an alternative approach to select features for model creation. The model showed the potential to detect advertisements relevant to human trafficking activities. However, the method only considered text classification, indicating a need for further modification of the model to include emoji feature selection to obtain the improved accuracy of model predictions.

To the best of our knowledge, there is no ML model for alerting human trafficking activities that considers an approach to improve the accuracy of data prediction based on the correction of the missing information such as misspelling words and emojis. This paper presents a novel approach for data preparation processing of the suspected datasets related to human trafficking activities. The dataset was prepared by

several steps that should be improved in terms of all aspects related to feature selection, such as text, emojis, and misspelled words, to achieve reliable model predictions for detecting human trafficking activities. The proposed approach was then applied with the labeled dataset, Trafficking-10k [10], to demonstrate the performance of model prediction on human trafficking activities, using the percentage of accuracy as an indicator.

3. DATASET

The four datasets used in this work were Trafficking-10k [10], emoji [17], dig-dictionaries [20], and an English word list [21]. Trafficking-10k is a dataset that contains the online adult service advertisements from backpage.com, which is provided by Marinus Analytics. Marinus Analytics is a company aiming to develop technology to disrupt human trafficking, child abuse, and cyber fraud. The dataset was labeled to identify the potential for human trafficking activities in the advertisements by domain experts. The level definitions are shown in Table 1.

The emojis used in this work were identified as those that pimps use instead of their messages to prevent the detection by law enforcement authorities. They use emojis to reference meanings such as price, young girls, and women's virginity. The emoji definitions are shown in Table 2.

A dictionary of the categorized words related to adult services by the University of Southern California [20] was used as a reference vocabulary database. Dig-dictionaries are dictionaries that collect the vocabulary from the internet, then classify those words into categories. In this work, we used the five categories which were cup size, escort service, place, nationality, and person name, of dig-dictionaries for classified the dataset. Examples are shown in Table 3.

Table 1. Definition of the level indicating the potential for human trafficking activities labeled by Marinus Analytics

Level	Definition
0	Strongly likely not trafficking
1	Likely not trafficking
2	Weakly likely not trafficking
3	Unsure
4	Weakly likely trafficking
5	Likely trafficking
6	Strongly likely trafficking

Table 2. Emoji related to human trafficking activities

Emoji	Indicator	Reference
🌷 , ⊛	Price	Price
💕	Girl	Young girls
🍒 , 🌸	Virginity	Women's virginity, and a minor
✈️ , 🛫	Movement	Movement of the poster and a minor
👑	Pimp	Poster is a minor who is controlled by pimp

Table 3. Examples of word and category from dig-dictionary used for building the model for detection of human trafficking activities

Category	Word Example
Escort Service	Missionary
	Deep Throat
Cup Size	30B
	34A
Place	Guatemala
	Guam
Nationality	Austrian
	Bulgarian
Person Name	Jackelyn
	Codie

The English word list is a list containing words. This work used it to compare words in the Trafficking-10k dataset to find misspelled words. The misspelled word will be replaced with another English word that has the same meaning by this work.

4. METHOD

This work develops an approach for data preparation that can solve spelling mistakes in text classification and recover the missing information from the advertisements suspected to human trafficking activities. The dataset obtained from Trafficking-10k was processed using the proposed approach for preparation of the data before running by the conventional models later to demonstrate an improvement of the model's accuracy. The proposed approach consists of nine main steps as shown in Figure 1. Each step is explained in the following subsections as shown in subsections 4.1 to 4.9.
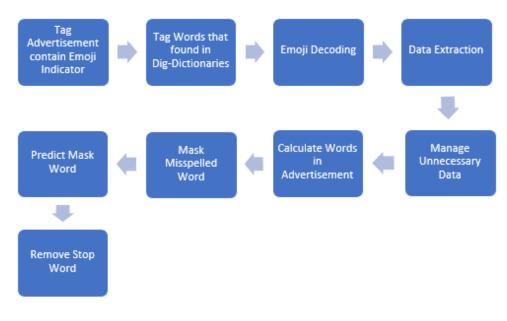

Figure 1. An overview of the proposed approach for data preparation

4.1. Tag advertisement contain emoji indicator

The dataset was first processed by this task involved tagging advertisements that comprised of emojis, which were used as indicators by pimps in their advertisements. The tags supported the detection of the model because they referenced the important data that pimps wanted to hide from detection, such as money, women, and girls. The indicators are shown in Table 2, while Table 4 shows examples of the results of this task. The advertisements are tagged based on the following rules:

- Tag "[[trafficking_emoji_price]]" If the advertisement contains price indicator.
- Tag "[[trafficking_emoji_girl]]" If the advertisement contains girl indicator.
- Tag "[[trafficking_emoji_ movement]]" If the advertisement contains movement indicators.
- Tag "[[trafficking_emoji_virgin_minor]]" If the advertisement contains women's virginity and a minor indicator.
- Tag "[[trafficking_emoji_pimp]]" If the advertisement contains pimp indicator.

Table 4. Tag indicator example

Before	After
💕 for you all time	💕 for you all time [[trafficking_emoji_girl]]
Our service 👑	Our service 👑 [[trafficking_emoji_pimp]]

4.2. Tag words that found in dig-dictionaries

This work used words in escort services, cup sizes, places, nationalities, and personal categories from dig-dictionaries because these categories are often associated with human trafficking activity. The task involved tagging advertisements containing words from these categories. This task is shown in algorithm 1.

4.3. Emoji decoding

Emojis are used in advertisements because they can convey mood, and pimps use emojis in their advertisements to promote their business. Emojis can often convey more meaning than words. However,

traditional text classification models often ignore emojis and focus only on words, allowing an escape route for spotting the human trafficking advertisements. To address this issue, this task involved decoding emojis into words based on the definitions provided by the Unicode Consortium [22]. Table 5 shows some examples of the results of this task.

Algorithm 1. Tag words from dig-dictionaries

```
procedure TagWordsRelated(datas,digdict)
  N← length(datas)
  for i ← 1 to N do
    text← datas[i]["text"]
    escort_service_flag ← false
    cupsizes_flag ← false
    place_flag ← false
    nationality_flag ← false
    person_flag ← false
    tags← ""
    Tokens[]←TokenizeBySpace(text)     ▷ Tokenize content by Space
    Ntoken← length(Tokens)
    for j← 1 to Ntoken do
        if Tokens[j] in digdict["escort_service"] then
           escort_service_flag ← true
        else if  Tokens[j] in digdict["cupsizes"] then
           cupsizes_flag ← true
        else if  Tokens[j] in digdict["place"] then
           place_flag ← true
        else if  Tokens[j] in digdict["nationality"] then
           nationality_flag ← true
        else if  Tokens[j] in digdict["person"] then
           person_flag ← true
    if is_true(escort_service_flag) then
      tags ← "[[relate_escort_service]]" ▷Add escort service tag
    if is_true(cupsizes_flag) then
      tags ← tags + " "+ "[[relate_cupsizes]]" ▷Add cup size tag
    if is_true(place_flag) then
      tags ← tags + " "+ "[[relate_place]]" ▷Add place tag
    if is_true(nationality_flag) then
      tags ← tags + " "+ "[[relate_nationality]]" ▷Add nationality tag
    if is_true(person_flag) then
      tags ← tags + " "+ "[[ relate_person]]" ▷Add person tag
    datas[i]["text"] ← datas[i]["text"] + " "+ tags
  return datas
```

Table 5. Example of emoji decoding using the proposed approach

Before	After
🏩 girl sell fun time	: love hotel: girl sell fun time
👩 unlimited 🏩 all night	: woman: unlimited: love hotel: all night

4.4. Data extraction

The pimp's advertisements contained a lot of data, such as links and images, to attract their customers. Therefore, this data can be used as a feature for creating a model. To extract data from the advertisement's content, this task used regular expressions [23]. The extracted data included images, money, and locations:

− Image

This task involved extracting image data from HTML tags using regular expression patterns. Once the pattern was identified, it was replaced with the string "[[image]]".

− Money

This task involved replacing money data in HTML tags using a regular expression approach. Once the patterns were identified, they were replaced with the string "[[money]]". Examples of targets included $120 and €30.

− Location

This task involved extracting location data from HTML link tags that linked to Google Maps. Once the tags were identified, they were replaced with the string "[[location]]". An example of an HTML link is: *Map*.

4.5. Manage unnecessary data

The dataset generally contained a lot of unnecessary data, resulting in a waste of time during processing. We used well-known techniques to manage the data and solve this problem. These techniques are described: i) converted to lowercase, ii) added space between emoji and html tags, iii) removed HTML Tags, iv) changed abbreviation words to full form, and v) replaced all numbers from text content by "[[number]]".

4.6. Calculate words in advertisement

This step involved a calculation of an occurrence of English words [21] in Trafficking-10k and then created a dictionary. The keys were expressed as English words, and the values were listed to indicate the occurrence times in three different indexes; index 0 represented the total occurrence in class 0, index 1 represented the total occurrence in class 1, and index 2 represented the total occurrence in all classes. Table 6 shows some examples of keys and values. The task can be seen in algorithm 2.

Table 6. Example of key and value for words calculation

Key	Value	Mean
party	[0,1,1]	"party" occurrence one time in class 1, but not occurrence in class 0
night	[1,1,2]	"night" occurrence one time in class 0 and class 1

Algorithm 2. Generate occurrence English dictionary

```
procedure DictCreation(datas)
  DictWord{}
  N← length(datas)
  listWords = createList("English-words.txt")  ▷ Create Word List from dataset
  O ← length(listWords)
  for w ← 1 to O do
    DictWord[listWords[w]] ← [0,0,0]
  for i ← 1 to N do
    text← datas[i]["text"]
    Tokens[]←TokenizeBySpace(text)      ▷ Tokenize content by Space
    Ntoken← length(Tokens)
    for j ← 1 to Ntoken do
        word ← Tokens[j]
        if Tokens[j] in DictWord then
          totalInClass0 ← DictWord [word][0]
          totalInClass1 ← DictWord [word][1]
          totalAllClass ← DictWord [word][2]
          if datas[i]["label"] == 0 then
            DictWord[word] ← [totalInClass0+1, totalInClass1, totalAllClass+1]
          else
            DictWord [word] ← [totalInClass0, totalInClass1+1, totalAllClass+1]
  return DictWord
```

4.7. Mask misspelled word

For this task, the misspelled words that were not in the dictionary from the previous step were detected. Then, those misspelled words were replaced with "[MASK]". The task as shown in algorithm 3. These mask tags were later corrected in the next step.

Algorithm 3. Mask misspelled word

```
procedure MaskMisspelled(datas, dictWord)
  N← length(datas)
  for i ← 1 to N do
    text← datas[i]["text"]
    textNew← ""
    Tokens[]←TokenizeBySpace(text)      ▷ Tokenize content by Space
    Ntoken ← length(Tokens)
    for j ← 1 to Ntoken do
      if NotTag(Tokens[j]) then
        if Tokens[j] not in dictWord then
          Tokens[j] ← "[MASK]"
      textNew ← textNew + Tokens[j]
    datas[i]["text"] ← textNew
  return datas
```

4.8. Predict mask word

This step uses bidirectional encoder representations from transformers (BERT) [24] to predict words then replaces the masked words obtained from the previous task. BERT is a pre-training model that is widely used to train unlabeled data with no specific to any particular domain. Sometimes, the replaced words were not popular in the domain of interest. For example, BERT may replace the word '[MASK]' with 'beautiful' in the sentence "Our hotel has [MASK] woman". This resulted that the sentence in the advertisement would be read as "Our hotel has beautiful woman", which did not reflect to human trafficking purpose. Commonly, pimps prefer to use the word 'sexy' instead of 'beautiful' in their advertisements as it is more appealing to their customers. To address this issue, this task combined the probability of words in Trafficking-10k with BERT. This was calculated using (1).

$$P(w_i) \times P(w_{i_{class}}) \tag{1}$$

where $P(w_i)$ is probability of word index i from pre-trained, $P(w_{i_{class}})$ is probability of word index i in the advertisement class.

4.9. Remove stop word

The advertisements also contained words that did not affect the model's decision. Therefore, this task involved deleting the stop words from the advertisements to reduce training time and to improve the model's performance. Table 7 shows some example results of this task.

Table 7. Example of stop word removal

Before	After
our women are beautiful	women beautiful
woman is so hot	woman hot

The proposed approach was used to process the data suspected to human trafficking activities for recovering the missing information from misspelled words and emojis including unnecessary data removal. The processed data by this approach was then used to create the model to demonstrate its predictive performance of human trafficking alert.

5. EXPERIMENT

The experiment used the Trafficking-10k dataset to demonstrate the efficiency of the proposed approach for data preparation in a model creation. The results were then compared to those obtained from models used in other works [25]–[27]. The Trafficking-10k dataset, which was labeled as the seven levels as shown in Table 1, was relabeled into two classes: Class 0 representing levels 0-2 (not related trafficking) and Class 1 representing levels 4-6 (related trafficking). The advertisements marked as level 3 were excluded from the experiment because they were considered noise in classification. The experiment used a total of 7,698 advertisements, divided equally between the two classes. To create the models, the same learning parameters and dataset were used. Accuracy was employed to measure and compare the performances, calculated by (2). The results are shown in Table 8. The confusion matrix of KNN, NB, and MLP by using the proposed approach are shown in Figures 2 to 4, respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

where is true negative (TN): classifier predicts it as class 0 and the advertisements is class 0, false positive (FP): classifier predicts it as class 1 and the advertisements is class 0, true positive (TP): classifier predicts it as class 1 and the advertisements is class 1, false negative (FN): classifier predicts it as class 0 and the advertisements is class 1.

Table 8. Experiment result showing the percentage of accuracy obtained from each method

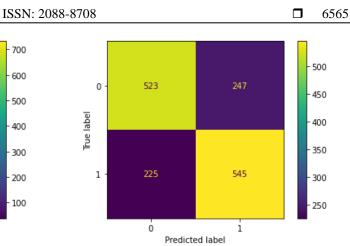| Model	Proposed approach	[25]	[26]	[27]
KNN	62.34	50.97	50.71	50.71
NB	69.35	52.40	52.60	52.73
MLP	75.32	51.88	52.47	52.53

Figure 2. Confusion matrix of KNN using the proposed approach

Figure 3. Confusion matrix of NB using the proposed approach
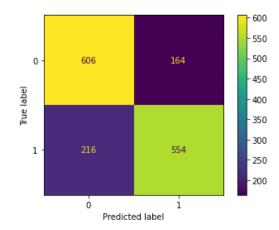

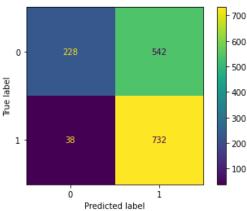Figure 4. Confusion matrix of MLP using the proposed approach

6.   RESULTS AND DISCUSSION

Based on the results shown in Table 8, the models using the proposed approach for data preparation processing achieved accuracy rates of 62.34%, 69.35%, and 75.32% when using KNN, NB, and MLP, respectively. In contrast, the accuracy obtained from other models were in the range of 50.71% to 50.97% (for KNN), 52.40% to 52.73% (for NB), and 51.88% to 52.53% (for MLP). The results indicate that the higher accuracy rates achieved by the models using the proposed data preparation, evidencing the importance of this step for improving model performance. Notably, the misspelled words paralleled with text classifications should be considered for the correction to ensure full coverage of data filtration.

The proposed approach for data preparation has shown the potential for use in real applications for human trafficking detection. The applicability of the proposed approach will be demonstrated by applying it to different datasets for human trafficking detection. Furthermore, this approach could be applied to various domains of interest for data preparation purposes.

7.   CONCLUSION

This work presents the proposed method for solving the problem of spelling mistakes and recovering the missing information to improve the accuracy of model predictions for human trafficking detection. The proposed approach replaced the misspelled words with English words commonly used in human trafficking advertisements. Also, the missing information related to human trafficking activities was recovered by emojis translation. The unnecessary data was then removed to reduce the processing time for machine learning step for model prediction. The experimental results of the model prediction using the data prepared by the proposed method gave higher accuracy compared to those obtained from the other models, indicating the significantly improvement of human trafficking alerting. The proposed approach will be used

to further demonstrate for model creation using other datasets and to generalize the approach for creating models in other domains. Additionally, named entity recognition (NER) could possibly be applied with this approach to further improve the detection of human trafficking activities.

REFERENCES

[1] L. Li, O. Simek, A. Lai, M. Daggett, C. K. Dagli, and C. Jones, "Detection and characterization of human trafficking networks using unsupervised scalable text template matching," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 3111–3120, doi: 10.1109/BigData.2018.8622189.

[2] H. Alvari, P. Shakarian, and J. E. K. Snyder, "A non-parametric learning approach to identify online human trafficking," in *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, Sep. 2016, pp. 133–138, doi: 10.1109/ISI.2016.7745456.

[3] A. Geron, "Hands-on machine learning with scikit-learn and tensorflow: concepts, tools, and techniques to build intelligent systems," *O'Reilly Media*, 2017.

[4] A. Samreen, A. Ahmad, F. Zeshan, F. Ahmad, S. Ahmed, and Z. A. Khan, "A collaborative method for protecting teens against online predators over social networks: a behavioral analysis," *IEEE Access*, vol. 8, pp. 174375–174393, 2020, doi: 10.1109/ACCESS.2020.3007141.

[5] S. Roshan, S. V. Kumar, and M. Kumar, "Project spear: Reporting human trafficking using crowdsourcing," in *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*, 2017, pp. 295–299, doi: 10.1109/UPCON.2017.8251063.

[6] S. P. Ardakani, C. Zhou, X. Wu, Y. Ma, and J. Che, "A data-driven affective text classification analysis," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2021, pp. 199–204, doi: 10.1109/ICMLA52953.2021.00038.

[7] B. A. H. Murshed, J. Abawajy, S. Mallappa, M. A. N. Saif, S. M. Al-Ghuribi, and F. A. Ghanem, "Enhancing big social media data quality for use in short-text topic modeling," *IEEE Access*, vol. 10, pp. 105328–105351, 2022, doi: 10.1109/ACCESS.2022.3211396.

[8] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," *PLOS ONE*, vol. 15, no. 5, 2020, doi: 10.1371/journal.pone.0232525.

[9] U. Naseem, I. Razzak, S. K. Khan, and M. Prasad, "A comprehensive survey on word representation models: from classical to state-of-the-art word representation language models," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 5, pp. 1–35, Sep. 2021, doi: 10.1145/3434237.

[10] E. Tong, A. Zadeh, C. Jones, and L.-P. Morency, "Combating human trafficking with multimodal deep models," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017, doi: 10.18653/v1/p17-1142.

[11] R. McAlister, "Webscraping as an investigation tool to identify potential human trafficking operations in Romania," in *Proceedings of the ACM Web Science Conference*, Jun. 2015, pp. 1–2, doi: 10.1145/2786451.2786510.

[12] A. Mensikova and C. A. Mattmann, "Ensemble sentiment analysis to identify human trafficking in web data," in *Workshop on Graph Techniques for Adversarial Activity Analytics (GTA 2018), Marina Del Rey, CA, USA*, 2018, pp. 5–9.

[13] M. Hernandez-Alvarez, "Detection of possible human trafficking in twitter," in *2019 International Conference on Information Systems and Software Technologies (ICI2ST)*, Nov. 2019, pp. 187–191, doi: 10.1109/ICI2ST.2019.00034.

[14] M. Diaz and A. Panangadan, "Natural language-based integration of online review datasets for identification of sex trafficking businesses," in *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, 2020, pp. 259–264, doi: 10.1109/IRI49571.2020.00044.

[15] J. Zhu, L. Li, and C. Jones, "Identification and detection of human trafficking using language models," in *2019 European Intelligence and Security Informatics Conference (EISIC)*, Nov. 2019, pp. 24–31, doi: 10.1109/EISIC49498.2019.9108860.

[16] Q. Bai, Q. Dan, Z. Mu, and M. Yang, "A systematic review of emoji: Current research and future perspectives," *Frontiers in Psychology*, vol. 10, 2019, doi: 10.3389/fpsyg.2019.02221.

[17] J. Whitney, M. Jennex, A. Elkins, and E. Frost, "Don't want to get caught? don't say it: The use of EMOJIS in online human sex trafficking ads," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018, doi: 10.24251/HICSS.2018.537.

[18] C. Wiriyakun and W. Kurutach, "Feature selection for human trafficking detection models," in *2021 IEEE/ACIS 20th International Fall Conference on Computer and Information Science (ICIS Fall)*, 2021, pp. 131–135, doi: 10.1109/ICISFall51598.2021.9627435.

[19] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?,'" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[20] P. Szekely, "Dig dictionaries," GitHub, Szeke. Accessed: 23 Apr, 2022. [Online]. Available: https://github.com/usc-isi-i2/dig-dictionaries

[21] Nelson, "English words," GitHub, Nelsonic. Accessed: Jun. 10, 2022. [Online]. Available: https://github.com/dwyl/english-words

[22] "Full emoji list," Unicode, https://unicode.org/emoji/charts/full-emoji-list.html (accessed Apr. 27, 2022).

[23] Y. Zhenjun and J. Xiangyu, "A simplified application of regular expressions: With the extraction of Chinese cultural terms as an example," in *2009 ISECS International Colloquium on Computing, Communication, Control, and Management*, 2009, pp. 439–442, doi: 10.1109/CCCM.2009.5268087.

[24] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Prepr. arXiv.1810.04805*, Oct. 2018.

[25] I. D. Buldin and N. S. Ivanov, "Text classification of illegal activities on onion sites," in *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, Jan. 2020, pp. 245–247, doi: 10.1109/EIConRus49466.2020.9039341.

[26] N. Save, O. Thopate, V. Waghe, M. Bansode, and N. Ghatte, "Artificial intelligence based fake news classification system," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2021, pp. 1–5. doi: 10.1109/ICCCNT51525.2021.9579759.

[27] L. S. Gopal, R. Prabha, D. Pullarkatt, and M. V. Ramesh, "Machine learning based classification of online news data for disaster management," in *2020 IEEE Global Humanitarian Technology Conference (GHTC)*, 2020, pp. 1–8, doi: 10.1109/GHTC46280.2020.9342921.

BIOGRAPHIES OF AUTHORS

Chawit Wiriyakun received M.S. (Information Technology) degree from Kasetsart University, Thailand in 2014. Currently, he is an engineer at the National Science and Technology Development Agency (NSTDA) in Thailand. He is also a Ph.D. student at the Faculty of Engineering and Technology, Mahanakorn University of Technology. He can be contacted at email: chawitmild@gmail.com.

Werasak Kurutach received Ph.D. in Computer Science and Engineering from The University of New South Wales, Australia, in 1996. Currently, he is an associate professor in Computer Engineering at Mahanakorn University of Technology, Bangkok, Thailand. He can be contacted at email: werasak@mut.ac.th.