Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2025-23.1.007

RESEARCH ARTICLE

Improving Thai Sentiment Analysis Accuracy with Emoji Classification by Deep Learning and Stacking Models: A Case Study of Hotel Reviews

Walaithip Bunyatisai[1], Suwika Plubin[2], Knavoot Jiamwattanapong[3], Bandhita Plubin[4*]

[1,2,4] Department of statistics, Faculty of science, Chiang Mai University, Chiang Mai, Thailand, 50200
[3] Faculty of Liberal Arts, Rajamangala University of Technology Rattanakosin, Nakhon Pathom, Thailand, 73170

ARTICLE INFO

Received: May 21, 2024

Accepted: Aug 25, 2024

Keywords

Sentiment Analysis

Emoji Classification

Emoji

Emotion

Convolutional Neural Networks

Recurrent Neural Networks

***Corresponding Author:**

Bandhita.p@cmu.ac.th

ABSTRACT

Sentiment analysis presents unique challenges in hotel reviews, particularly in languages like Thai, renowned for nuanced expressions. Understanding customer opinions is pivotal, especially in domains such as hotel reviews, where subjective expressions prevail. This study delves into methodologies to refine sentiment analysis accuracy in Thai hotel reviews by integrating emoji classification with Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and stacking models. Through comparing models with and without emoji inclusion, we meticulously evaluate their performance metrics encompassing accuracy, precision, recall, and F1-score. The pinnacle achievement of 92.4% accuracy underscores the efficacy of advanced stacking techniques complemented by emoji integration. Our findings underscore the superiority of models encompassing emojis, affirming the value of amalgamating textual and emoji data. Leveraging sophisticated deep learning techniques and stacking models, our approach adeptly captures the subtle nuances of sentiment expressed in Thai text, resulting in heightened accuracy in sentiment analysis. This research underscores the paramount importance of embracing diverse data sources and sophisticated modeling strategies to elevate sentiment analysis accuracy in Thai hotel reviews.

INTRODUCTION

Sentiment analysis, the computational study of opinions, sentiments, and emotions expressed in text (Manalu et al. 2020; Razali et al. 2021), has become a critical tool in various applications, ranging from market analysis to customer feedback evaluation (Alharbi et al. 2021; Mukherjee and Bhattacharyya 2012; Salur and Aydin 2020). In the hotel domain, customer reviews on platforms such as Agoda, TripAdvisor, and Booking.com are essential sources of tourism information for travelers and play a prominent role in hotel selection. The UK Mintel Report (2013) showed that 86% of online travelers find consumer review websites helpful when making decisions about which hotel to book.

However, reviews are often in an unstructured form, containing numbers, symbols, abbreviations, and spelling errors (Khamphakdee and Seresangtakul 2021). In the context of the Thai language, sentiment analysis presents unique challenges due to its complex script and contextual nuances (Khamphakdee and Seresangtakul 2021; Poncelas 2020; Singkul et al. 2019). One emerging approach to enhance sentiment analysis accuracy is the integration of emoji classification, leveraging the expressive power of emojis as sentiment indicators. Emojis, widely used in digital communication, offer a rich source of sentiment cues that can complement textual data. They bridge gaps in understanding that may arise from textual ambiguity or brevity, especially in user-generated content such as social media posts and online reviews. By incorporating emoji classification, it is possible to capture a more comprehensive sentiment profile of a given text (Dey and Dey 2023; Felbo et al. 2017; Neel et al. 2023; Novak et al. 2015; Shiha and Ayvaz 2017). Traditional sentiment analysis models often struggle to accurately interpret sentiments expressed through emojis, leading to incomplete sentiment detection (Singla et al., 2022).

Recent advancements in deep learning have shown promise in handling the intricacies of natural language processing (NLP) (Mohammed et al., 2023; Chen et al., 2019). Techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants have demonstrated superior performance in various NLP tasks. Additionally, the strategy of model stacking—combining multiple learning algorithms to achieve better predictive performance—has been effective in improving accuracy and robustness (Zeng et al., 2018; Demirci et al., 2019; Kim and Jeong, 2019; Zhou et al., 2019; Mohbey, 2021; Yang et al., 2021; Xiang, 2021).

This paper addresses this gap by proposing a deep learning and stacking model approach that incorporates emoji classification to enhance sentiment analysis accuracy in Thai hotel reviews. The main contributions of this paper are as follows:

1. Develop a CNN-based model for emoji classification to enhance sentiment understanding.
2. Integrate emoji classification with RNN-based text sentiment analysis.
3. Employ a stacking model framework to combine multiple classifiers and improve overall sentiment classification performance.
4. Evaluate the proposed model's effectiveness using a dataset of Thai hotel reviews.

The rest of the paper is structured as follows. In the next section, we review related works on the challenges of sentiment analysis in Thai and the use of emojis. A concise outline of the proposed method is presented. The experimental setup, dataset description, and performance metrics are discussed in Sections 3 and 4. In Section 5, the investigational results with different metrics are presented. Finally, Section 6 concludes the paper.

LITERATURE REVIEW

Background and related work

Sentiment analysis

Sentiment analysis, a subfield of natural language processing (NLP), aims to determine the sentiment expressed in textual data, categorizing it as positive, negative, or neutral. Sentiment analysis, also known as opinion mining, has garnered significant attention in recent years due to its wide-ranging applications in understanding public opinion (Dehler-Holland et al. 2022), market trends (Hirata and Matsuda 2023), and consumer behavior (Salur and Aydin 2020). This literature review provides an overview of key studies, methodologies, and advancements in sentiment analysis.

Challenges in Thai sentiment analysis

The Thai language presents unique challenges for sentiment analysis due to its linguistic structure. Thai script lacks explicit word boundaries, meaning there are no spaces between words, which complicates the process of tokenization and syntactic parsing (Haruechaiyasak and Kongthon 2013). Additionally, Thai language relies heavily on context and tone, making sentiment detection more intricate compared to languages with clearer syntactic rules. Additionally, the extensive use of emojis in Thai text requires models that can effectively interpret these symbols.

Emoji lexicon

Emojis are Unicode graphic symbols used as shorthand to express concepts and ideas. They have become an integral part of digital communication (Alshenqeeti 2016), particularly in social media and mobile messaging. Unlike the limited number of emoticons, there are hundreds of emojis, each with its unique emotional content. Emojis provide a visual representation of emotions, adding a layer of sentiment expression that complements textual analysis. Emoji Initially, emoticons served as a means to convey emotions in environments limited to basic text; however, emoji have evolved as extensions of the character set, readily available on most digital devices (Novak et al. 2015). While emoticons and their corresponding emoji are often recognized or categorized equivalently — for instance, :-) and ▯ — emoji typically wield a more profound impact on individuals' mood (Ganster et al. 2012).

Table 1: Emoji meanings sample entry (face with tears of joy)

Emoji Icon:	Emoji Name	Emoji Meaning
▯	Grinning face	A classic yellow emoji with a big smile, shows happiness or friendliness.
▯	Grinning Face With Eyes	A classic yellow emoji with a big smile and prominent eyes expressing happiness, friendliness, and excitement.
▯	Smiling Face With Heart Eyes	A face blowing a kiss with heart-shaped eyes, conveying love and admiration.
▯	Crying Face	A teary-eyed face, representing sadness or sorrow.

Emoji's in sentiment analysis

Emojis have become an integral part of digital communication, serving as visual cues that convey emotions and sentiments beyond what is expressed through text alone (Novak et al. 2015). Incorporating emojis into sentiment analysis can enhance the accuracy of sentiment detection by capturing these non-verbal cues. Recent studies have demonstrated that emojis can significantly impact the perceived sentiment of a message and that their inclusion in sentiment analysis models can lead to better performance (Felbo et al. 2017; Shiha and Ayvaz 2017).

Hu et al. (2017) delved into the nuanced impact of emojis on sentiment in textual communication. Their research revealed intriguing findings regarding the influence of emojis on the perceived sentiment of verbal statements. Specifically, they noted that the inclusion of a positive emoji resulted in both neutral and negative verbal statements being perceived as more positive, although it had no discernible effect on already positive verbal statements. Additionally, Hu et al. (2017) observed that the presence of a neutral emoji predominantly affected positive verbal statements, leading to a more negative perception of such statements compared to when presented without an emoji. Similarly, negative emojis exerted a dampening effect on the positivity of positive and neutral verbal statements, while having no impact on negative verbal statements.

Building on this exploration, Boutet et al. (2021) uncovered that neutral texts were particularly susceptible to the influence of emojis; the presence of an emotionally-valent face emoji altered participants' perceptions of the message, subtly shifting the sentiment conveyed.

Novak et al. (2015) developed the first emoji sentiment lexicon, called the Emoji Sentiment Ranking, which provides a sentiment map of the 751 most frequently used emojis. This lexicon was created by analyzing the sentiment of tweets containing emojis. The study involved 83 human annotators who labeled over 1.6 million tweets in 13 European languages based on their sentiment polarity (negative, neutral, or positive). The results showed that most emojis are positive, especially the most popular ones. The sentiment analysis of emojis revealed that tweets with emojis have a higher inter-annotator agreement and different sentiment distribution compared to tweets without emojis. Consequently, the Emoji Sentiment Ranking was proposed as a language-independent resource for automated sentiment analysis.

Felbo et al. (2017) introduced DeepMoji, a model that uses transfer learning to predict sentiment based on emojis. Their work demonstrated that emoji-based models could significantly improve the performance of sentiment analysis tasks. As Erle et al. (2022) substantiated the idea that emojis equate to facial expressions, it follows that the perceived emotional tone of affectively neutral messages paired with emojis should align directly with the emotional valence conveyed by the emoji.

Table 2: Emoji meanings sample entry with sentiment analysis (face with tears of joy)

Emoji Icon:	Emoji Name	Emoji Meaning	Sentiment
?	Grinning face	A classic yellow emoji with a big smile, shows happiness or friendliness.	Positive
?	Grinning Face with Eyes	A classic yellow emoji with a big smile and prominent eyes expressing happiness, friendliness, and excitement.	Positive
?	Smiling Face with Heart Eyes	A face blowing a kiss with heart-shaped eyes, conveying love and admiration.	Positive
?	Crying Face	A teary-eyed face, representing sadness or sorrow.	Negative

Deep learning approaches for sentiment analysis

Deep learning models, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown superior performance in various natural language processing tasks, including sentiment analysis. CNNs are effective for tasks that involve feature extraction from images, making them suitable for emoji classification. Kim (2014) demonstrated the effectiveness of CNNs for sentence classification, which laid the foundation for their application in sentiment analysis. RNNs, especially Long Short-Term Memory (LSTM) networks, excel in handling sequential data and capturing long-term dependencies, which is crucial for understanding the context in text. Hochreiter and Schmidhuber (1997) introduced LSTMs, which have since become a standard in NLP tasks due to their ability to retain information over long sequences.

Ensemble learning techniques combine multiple models to improve prediction accuracy and robustness. Stacking, a type of ensemble learning, involves training multiple base models and a meta-model that combines their predictions, leveraging the strengths of each individual model (Wolpert 1992). This approach has been shown to enhance performance in various machine learning tasks by reducing the likelihood of overfitting and improving generalization (Ting and Witten 1999). In sentiment analysis, stacking can integrate different types of models, such as CNNs for emoji classification and RNNs for text analysis, to achieve better overall performance.

Recent advancements in deep learning have significantly improved the performance of sentiment analysis models. The integration of emojis into sentiment analysis has gained attention as emojis provide additional emotional context to text. Novak et al. (2015) developed the first emoji sentiment lexicon, the Emoji Sentiment Ranking, which maps the sentiment of the 751 most frequently used emojis. This lexicon was created by analyzing the sentiment of tweets containing emojis, demonstrating that the inclusion of emojis can enhance the accuracy of sentiment analysis.

Sentiment analysis of Thai text has been less explored compared to English and other widely spoken languages. Haruechaiyasak and Kongthon)2013(developed foundational tools for Thai text processing, which are crucial for subsequent sentiment analysis research. Despite the challenges, recent studies have shown that advanced NLP techniques, including deep learning and ensemble methods, can be effectively applied to Thai text for sentiment analysis.

METHODOLOGY

Data Collection

To study how emojis are used in positive, neutral, and negative expressions, we gathered Thai hotel reviews from popular travel platforms like Agoda, TripAdvisor, and Booking.com. Our dataset consists of text reviews and their associated emojis, collected from user comments between January

1, 2023, and December 31, 2023. We specifically selected sentences containing both text and emojis that expressed negative, neutral, or positive sentiments, resulting in a total of 12,450 sentences for analysis. These sentences were then classified into three categories: positive, neutral, and negative. Our positive word dictionary includes words like *"good," "happy," "love," and "enjoy,"* which represent positive feelings and opinions. Conversely, the negative dictionary contains words like *"bad," "sad," "hate," and "terrorism,"* reflecting negative sentiments. The three datasets consist of 5,200 positive sentences, 3,150 neutral sentences, and 4,100 negative sentences.

Pre-processing

For textual analysis, a pre-processing stage is imperative before commencing computational analysis. This is particularly crucial for the Thai language, which necessitates a distinct pre-processing approach. In this study, we will subject the acquired text to the following pre-processing steps:

- **Text cleaning:** Removing HTML tags, special characters, and irrelevant content.
- **Tokenization:** Segmenting text into words using Thai-specific tokenization tools.
- **Remove stop word:** is the removal of words that have no meaning (usually appear in large numbers).
- **Emoji extraction:** Identifying and isolating emojis for separate analysis.
- **TF-IDF** stands for Term Frequency-Inverse Document Frequency. It is a statistical measure that is used to evaluate how relevant a word is to a document in a collection of documents. TF-IDF is calculated by multiplying two metrics: term frequency (TF) and inverse document frequency (IDF). TF is the number of times a word appears in a document, divided by the total number of words in the document. IDF is the logarithm of the total number of documents divided by the number of documents that contain the word. (Saadah et al. 2013) The equation for TF-IDF is written in equations:

$$IDF(t) = log\frac{1+n}{1+df(t)}+1$$

$$TF-IDF(t,d) = TF(t,d)\cdot IDF(t)$$

Where n is the total number of documents and $df(t)$ is the number of terms appearing in all documents. Therefore, TF-IDF is normalized using the Euclidean norm in equation:

$$v_{norm} = log\frac{v}{\sqrt{v_1^2+v_2^2+\cdots+v_n^2}}$$

Where vi is TF-IDF of each term in the entire documents.

Train-test split

The dataset was divided into two parts: 80% for training and 20% for testing. This split ensures that the model is trained on a substantial portion of the data while preserving a separate subset for evaluating its performance.

Sentiment	Training Set	Testing Set
Positive	4,160	1,040
Neutral	2,520	6,30
Negative	3,280	8,20

Model selection by deep learning model for classification

Convolutional layer

A convolutional layer applies a set of filters (kernels) to the input image to extract various features. Each filter convolves with the input image to produce a feature map.

- **Convolution operation:** The convolution operation involves sliding a filter over the input image and performing element-wise multiplication followed by summation.

$$(f \times g)(t) = \sum_{a=-\infty}^{\infty} f(a) \times g(t-a)$$

For a 2D convolution, this can be expressed as:

$$Y(i,j) = (X \times K)(i,j) \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X(i+m, j+n) \times K(m,n)$$

Where:

  - Y is the output feature map
  - X is the input image
  - K is the kernel (filter)
  - M and N are the dimensions of the kernel

- **Output size:** The spatial dimensions of the output feature map can be calculated as:

$$output\ size = \frac{Input\ size - Filter\ size + 2 \times Padding}{Stride} + 1$$

- **Parameters:** The number of parameters in a convolutional layer is:

$$Parameters = (Filter\ width \times Filter\ height \times Input\ depth + 1) \times Number\ of\ filter$$

Pooling layer

Pooling layers are used to reduce the spatial dimensions of the feature maps, which helps in reducing the number of parameters and computation in the network.

- **Max pooling:** Max pooling selects the maximum value from each patch of the feature map.

$$Y(i,j) = Max_{m \cdot n}(X(i \times s + m, j \times s + n))$$

Where s is the stride, and m, n define the dimensions of the pooling window.

- **Output size:** The spatial dimensions of the output after pooling are:

$$output\ size = \frac{Input\ size - Pool\ size}{Stride} + 1$$

Fully connected layer

Fully connected layers (dense layers) are used towards the end of the CNN to perform high-level reasoning in the network. Each neuron in a fully connected layer is connected to every neuron in the previous layer.

- **Output:** The output of a fully connected layer is given by:

$$y = Activation(W \cdot x + b)$$

Where:

  - y is the output vector
  - W is the weight matrix
  - x is the input vector
  - b is the bias vector
  - Activation is the activation function (e.g., ReLU, Sigmoid)

Dropout regularization

Dropout is a regularization technique used to prevent overfitting by randomly setting a fraction of the input units to zero during training.

- **Dropout operation:** During training, each neuron is set to zero with a probability p.

$$y = Dropout(x, p) = x \odot Mask$$

Where $\odot$ represents element-wise multiplication, and Mask is a binary matrix with entries set to zero with probability p.

Recurrent neural networks (RNNS)

Recurrent Neural Networks (RNNs) are a class of neural networks designed for processing sequential data. Unlike traditional feedforward neural networks, RNNs have connections that form directed cycles, allowing them to maintain a memory of previous inputs and capture temporal dependencies. This makes them particularly suited for tasks such as language modeling, speech recognition, and time-series prediction

Basic structure of RNNs

An RNN processes input sequences one element at a time while maintaining a hidden state that captures information about previous elements. The hidden state is updated at each time step based on the current input and the previous hidden state.

- **RNN cell:** The core component of an RNN is the RNN cell, which computes the new hidden state h_t at time step t.

$$h_t = \emptyset(W_{hx}x_t + W_{hh}h_{t-1} + b_h)$$

Where:

- h_t is the hidden state at t
- x_t is the input at time t
- $W_{hx}$is the weight matrix for the input
- W_{hh} is the weight matrix for the hidden state
- b_h is the bias term
- $\emptyset$ is the activation function, typically a non-linear function like tanh or ReLU

- **Output:** The output y_t at time step t is often computed using the hidden state.

$$y_t = \psi(W_{hy}h_t + b_y)$$

Where:

- y_t is the output at time t
- W_{hy} is the weight matrix for the output
- b_y is the bias term
- ψ is the activation function, often a softmax for classification tasks

Long short-term memory (LSTM)

LSTM is a type of RNN that addresses the problem of learning long-term dependencies. It introduces a memory cell and three gates (input gate, forget gate, and output gate) to control the flow of information.

- **LSTM equations:** The LSTM cell maintains a cell state C_t and a hidden state h.

$$f_t = \sigma(W_f \cdot [h_{t-1,x_t}] + b_f) \qquad \text{forget gate}$$

$$i_t = \sigma(W_f \cdot [h_{t-1,x_t}] + b_i) \qquad \text{input gate}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1,x_t}] + b_c) \qquad \text{cell candidate}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \qquad \text{cell state}$$

$$o_t = \sigma(W_o \cdot [h_{t-1,x_t}] + b_o) \qquad \text{output gate}$$

$h_t = o_t \odot \tanh(C_t)$ hidden gate

Where:

- f_t, i_t, and o_t are the forget, input, and output gates, respectively
- $\tilde{C}_t$ is the cell candidate
- $\odot$ denotes element-wise multiplication
- σ is the sigmoid function

Gated recurrent unit (GRU)

GRU is a simplified variant of LSTM with fewer gates, which makes it computationally more efficient while still addressing the vanishing gradient problem.

- **GRU equations:** The GRU combines the forget and input gates into an update gate and introduces a reset gate.

$Z_t = \sigma(W_z \cdot [h_{t-1,x_t}] + b_z)$ update gate

$r_t = \sigma(W_r \cdot [h_{t-1,x_t}] + b_r)$ reset gate

$\tilde{h}_t = tanh(W \cdot [r_t \odot h_{t-1,x_t}] + b)$ candidate hidden state

$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$ hidden state

Where:

- \- Z_t is the update gate
- \- r_t is the reset gate
- -$\tilde{h}_t$ is the candidate hidden state

Training RNNs

Training RNNs involves backpropagation through time (BPTT), which is an extension of the standard backpropagation algorithm. BPTT computes gradients by unfolding the RNN over time and applying backpropagation to the unfolded network.

The loss function for RNNs is typically the cross-entropy loss for classification tasks.

$$\mathcal{L} = -\sum_t \sum_k y_{t,k} \log(\hat{y}_{t,k})$$

Where:

- $y_{t,k}$ is the true label at time t for class k

- $\hat{y}_{t,k}$ is the predicted probability at time t for k

Stacking model

Stacking, also known as stacked generalization, is an ensemble learning technique that combines multiple machine learning models to improve overall predictive performance. The idea is to leverage the strengths of different models by training a "meta-model" to aggregate their predictions. This approach can often outperform individual models by reducing overfitting and bias.

Basic structure of stacking

Stacking involves two levels of models: base learners (level-0 models) and a meta-learner (level-1 model). The base learners are trained on the original dataset, and their predictions are used as inputs to the meta-learner, which is trained to make the final predictions.

1) Base learners:

- Multiple machine learning models are used as base learners.
- Each base learner is trained on the entire training dataset.

2) Meta-learner:

- The meta-learner is a model that takes the predictions of the base learners as input features.
- It is trained to learn the optimal way to combine these predictions to produce the final output.

Formulas

1) Training base learners:

Given a training set $\{(X_i, y_i)\}_{i=1}^{N}$, where X_i are the input features and y_i are the corresponding labels, the base learners $M_1, M_2, \cdots, M_k$ are trained to produce predictions $\hat{y}_i^{(j)}$ for each base model j

$$\hat{y}_i^{(j)} = M_j(X_i)$$

2) Creating meta-features:

The predictions of the base learners form the meta-features for the meta-learner. For each instance X_i, the meta-features are:

$$Z_i = \hat{y}_i^{(1)}, \hat{y}_i^{(2)}, \cdots, \hat{y}_i^{(k)}$$

3) Training the meta-learner:

The meta-learner M_{meta} is trained on the meta-features Z_i to predict the final output y_i:

$$\hat{y}_i = M_{meta}(Z_i)$$

Thus, the final prediction for a new instance X is:

$$\hat{y} = M_{meta}([M_1(X), M_2(X), \cdots, M_k(X)])$$

Cross-validation in stacking

To avoid overfitting and to ensure that the meta-learner is trained on unbiased predictions, a cross-validation approach is often used.

1) K-fold cross-validation:

- Split the training set into K folds.
- For each fold, train the base learners on $K - 1$ folds and predict on the remaining fold.
- Aggregate the predictions for the meta-features.

2) Training with out-of-fold predictions:

- Use the out-of-fold predictions to train the meta-learner.

Benefits of stacking

- **Diverse model types**: Stacking allows combining models of different types (e.g., decision trees, SVMs, neural networks) to leverage their strengths.
- **Improved generalization**: By learning to combine predictions optimally, stacking can reduce overfitting and improve generalization.
- **Flexibility**: The meta-learner can be any model, adding flexibility to the stacking framework.

Evaluation metrics

Evaluation metrics can be described as measuring tools to measure the performance of classifiers (Mohammad and Sulaimam 2015). To calculate performance metrics, the formula as shown in Table 3.

Table 3: Evaluation metrics

Metric	**equation**
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{(TP+FP)}$
Recall	$\frac{TP}{(TP+FN)}$
F1-score	$\frac{2 \times precision \times recall}{(recall+precision)}$

Where TP, TN, FP and FN can be described as follows:

- TP (True Positive) is when the model correctly predicts the positive class.
- TN (True Negative) is when the model correctly predicts the negative class.
- FP (False Positive) is when the model incorrectly predicts the positive class.
- FN (False Negative) is when the model incorrectly predicts the negative class.

MODEL ARCHITECTURE

The two studies focus on improving Thai sentiment analysis for hotel reviews by exploring different methodologies. The key difference lies in the inclusion of emoji classification. The methodologies include traditional and advanced models, incorporating deep learning techniques and stacking models. Below is a textual description of what the flow chart would look like, followed by a more detailed explanation.

1) **Data preprocessing**
    - Common preprocessing steps for both models
        - Text Cleaning
        - Tokenization
        - Word Embedding
2) **Without emoji inclusion**
    - RNN for Thai Sentiment Analysis
        - Train RNN Model
        - Evaluate RNN Model
    - Stacking Models
        - Train RNN Base Models
        - Train Meta-Classifier
        - Evaluate Stacking Model
3) **With emoji inclusion**
    - CNN for Emoji Classification
        - Train CNN Model
        - Evaluate CNN Model
    - Hybrid Model (CNN for Emoji + RNN for Thai Text)
        - Train CNN and RNN Models
        - Combine Features
        - Train Hybrid Model
        - Evaluate Hybrid Model
    - Stacking Models
        - Train RNN Base Models
        - Train Meta-Classifier
        - Evaluate Stacking Model

RESULTS

The studies evaluated various classifiers for Thai sentiment analysis on hotel reviews, comparing models that include emoji classification using CNN with those that do not. The methods included Thai sentiment analysis using RNN and advanced stacking models. This comparison focuses on models with and without emoji inclusion.

Performance metrics

Table 4: Results obtained from all classifiers for sentiment classification model Without Emojis

Classifier	Accuracy	Precision	Recall	F1-Score
RNN for Thai Sentiment Analysis	87.8%	87.0%	87.5%	87.2%
Stacking Model 1 (RNN Base Models + Meta-Classifier)	90.0%	89.2%	89.5%	89.3%
Stacking Model 2 (RNN Only with Additional Layers)	91.0%	90.2%	90.5%	90.3%

- **RNN for Thai sentiment analysis**: Achieved an accuracy of 87.8%, with precision, recall, and F1-scores around 87%.
- **Stacking model 1**: Enhanced accuracy to 90.0%, demonstrating the benefit of combining base models with a meta-classifier.
- **Stacking model 2**: Further improvement to 91.0%, highlighting the advantages of additional layers and sophisticated stacking techniques.

Table 5: Results obtained from all classifiers for sentiment classification model With Emojis

Classifier	Accuracy	Precision	Recall	F1-Score
CNN for Emoji Classification	86.5%	85.7%	86.2%	85.9%
Hybrid Model (CNN + RNN)	89.6%	88.5%	89.0%	88.7%
Stacking Model 1 (RNN Base Models + Meta-Classifier)	91.2%	90.5%	91.0%	90.7%
Stacking Model 2 (Hybrid + Additional Layers)	**92.4%**	**91.8%**	**92.0%**	**91.9%**

- **CNN for emoji classification:** Achieved an accuracy of 86.5%, demonstrating the model's ability to capture sentiment from emojis.
- **Hybrid model (CNN + RNN):** Significant improvement to 89.6%, showing the benefit of combining text and emoji data.
    - **Stacking model 1:** Achieved 91.2%, indicating the effectiveness of combining base models with a meta-classifier and emoji classification.
    - **Stacking model 2:** The highest accuracy of 92.4%, showcasing the superior performance of advanced stacking techniques with emoji integration.

Comparative analysis result

- **Impact of emojis:** The inclusion of emojis through CNN classification significantly enhances the performance of sentiment analysis models. The hybrid model (CNN + RNN) outperforms the RNN-only model, demonstrating the value of integrating emoji data.
- **Stacking models:** Both with and without emojis, stacking models show improved performance over individual models. However, the stacking models that include emoji classification consistently achieve higher accuracy, precision, recall, and F1-scores.

CONCLUSION

This paper presents a deep learning approach to improving sentiment analysis accuracy in Thai hotel reviews by incorporating emoji classification and using a stacking model framework. By leveraging CNNs for emoji classification, RNNs for text analysis, and a stacking ensemble, the proposed model effectively captures the full spectrum of sentiments expressed by reviewers. The comparative results highlight the significant impact of incorporating emoji classification on the accuracy of Thai sentiment analysis. Models that leverage both text and emoji data, especially through advanced stacking techniques, achieve the highest performance metrics. This underscores the importance of considering multiple data sources and sophisticated modeling approaches in sentiment analysis tasks.

REFERENCES

Alharbi NM, Alghamdi NS, Alkhammash, EH, al Amri JF (2021) Evaluation of sentiment analysis via word embedding and RNN variants for Amazon online reviews. Math Probl Eng: 1-10. https://doi.org/10.1155/2021/5536560.

Alshenqeeti H (2016) Are emojis creating a new or old visual language for new generations? a socio-semiotic study. Advances in Language and Literary Studies 7(6):56–69.

Boutet I, LeBlanc M, Chamberland JA, Collin CA (2021) Emoji infuence emotional communication, social attributions, and information processing. Comput Human Behav. https://doi.org/10.1016/j.chb.2021.106722

Chen Y, Xiong J, Xu W, Zuo J (2019) A novel online incremental and decremental learning algorithm based on variable support vector machine. Cluster Comput 22(S3): 7435–7445.

Dehler-Holland J, Okoh M, Keles D (2022) Assessing technology legitimacy with topic models and sentiment analysis – The case of wind power in Germany. Technol Forecast Soc Change 175. 121354. 10.1016/j.techfore.2021.121354.

Demirci GM, Keskin SR, Dogan G (2019- December) Sentiment Analysis in Turkish with Deep Learning. In Proceedings of the IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2215–2221.

Dey P, Dey S (2023) Sentiment Analysis of Text and Emoji Data for Twitter Network. Al-Bahir Journal for Engineering and Pure Sciences. 3:1e10.

Erle TM., Schmid K, Goslar SH, Martin JD (2022) Emojis as social information in digital communication. Emotion 22(7): 1529–1543. https://doi.org/10.1037/emo0000992.

Felbo B, Mislove A, Søgaard A, Rahwan I, Lehmann S (2017) Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 1615-1625, Copenhagen, Denmark.

Ganster T, Eimler SC, Krämer NC (2012) Same same but different!? The differential influence of smilies and emoticons on person perception. Cyberpsychol Behav Soc Netw 15(4): 226–230. doi: 10.1089/cyber.2011.0179.

Haruechaiyasak C, Kongthon A (2013, October) LexToPlus: A Thai Lexeme Tokenization and Normalization Tool. In Proceedings of the Sixth International Conference on Information, Communications & Signal Processing, Nagoya, Japan.

Hirata E, Matsuda T (2023) Examining logistics developments in post-pandemic Japan through sentiment analysis of twitter data. Asian transpo stud 9. 10.1016/j.eastsj.2023.100110.

Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. Neural Comput 9(8): 1735-1780.

Hu T, Guo H, Sun H, Nguyen T, Luo J (2017) Spice up your chat: The intentions and sentiment efects of using emojis. In Proceedings of Eleventh International AAAI Conference on Web and Social Media 11(1): 102-111. https://doi.org/10.1609/icwsm.v11i1.14869.

Khamphakdee N, Seresangtakul P (2021) Sentiment Analysis for Thai Language in Hotel Domain Using Machine Learning Algorithms. Acta Inform Prag 10(2): 155–171. https://doi.org/10.18267/j.aip.155

Kim H, Jeong YS (2019) Sentiment Classification Using Convolutional Neural Networks. Appl Sci 9(11): 2347. https://doi.org/10.3390/app9112347

Kim Y (2014) Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, 1746-1751. arXiv: 1408.5882. https://doi.org/10.3115/v1/D14-1181.

Manalu BU, Tulus, Efendi S (2020- September) Deep Learning Performance in Sentiment Analysis. In Proceedings of the 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), Medan, Indonesia.

Mintel (2013) Holiday planning and booking process – UK. Mintel Report, November

Mohammed A, Himmi M, Bellafkih M, Talaghzi J, Bennane A (2023) Sentiment analysis in digital learning: Comparing Lexical, Traditional machine learning, and deep learning approaches. 1-6. 10.1109/SITA60746.2023.10373745

Mohbey KK. Sentiment analysis for product rating using a deep learning approach. (2021-March). In Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India. 121–126.

Mukherjee S, Bhattacharyya P (2012) Feature Specific Sentiment Analysis for Product Reviews. In: Gelbukh A. (eds) Computational Linguistics and Intelligent Text Processing. CICLing 2012. Lect Notes Comput Sc 7181. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-28604-9_39.

Neel LAG, McKechnie JG, Robus CM, Hand CJ (2023) Emoji Alter the Perception of Emotion in Affectively Neutral Text messages. J Nonverbal Behav 47:83–97.

Novak PK, Smailović J, Sluban B, Mozetič I (2015) Sentiment of emojis. PLoS ONE 10(12), e0144296. https://doi.org/10.1371/journal.pone.0144296

Poncelas A, Pidchamook W, Liu C.-H, Hadley J, Way A (2020- May) Multiple Segmentations of Thai Sentences for Neural Machine Translation. Proceedings of the 1st Joint SLTU and CCURL Workshop (SLTU-CCURL 2020), pages 240–244 Language Resources and Evaluation Conference (LREC 2020), Marseille. arXiv 2020, arXiv:2004.11472.

Razali NAM, Malizan NA, Hasbullah NA, Wook M, Zainuddin NM, Ishak KK, Ramli S, Sukardi S (2021) Opinion mining for national security: Techniques, domain applications, challenges and research opportunities. J Big Data 8(1): 150. doi: 10.1186/s40537-021-00536-5. Epub 2021 Dec 4. PMID: 34900516; PMCID: PMC8642766.

Saadah MN, Atmagi RW, Rahayu DS, Arifin AZ (2013) Information Retrieval of Text

Salur MU, Aydin I (2020) A novel hybrid deep learning model for sentiment classification. In: IEEE Access, 8: 58080–58093. https://doi.org/10.1109/ACCESS.2020.2982538.

Singkul S, Khampingyot B, Maharattamalai N, Taerungruang S, Tawunrat C (2019) Parsing Thai Social Data: A New Challenge for Thai NLP. 10.1109/iSAI-NLP48611.2019.9045639.

Shiha MO, Ayvaz S (2017) The Effects of Emoji in Sentiment Analysis. Int J of Comput Electr Eng 9(1): 360-369.

Singla C, Al-Wesabi N, Singh Pathania Y, Sulaiman Alfurhood B, Mustafa Hilal A, Rizwanullah, M, Ahmed Hamza M, Mahzari M (2022) An optimized deep learning model for emotion classification in tweets. Comput Mater Contin 70 (3): 6365–6380. https://doi.org/10.32604/cmc.2022.020480.

Ting KM, Witten IH (1999) Issues in stacked generalization. J Artif Intell Res 10: 271-289.

Wolpert DH (1992) Stacked Generalization. Neural Networks 5(2): 241-259.

Xiang S (2021-November) Deep Learning Framework Study for Twitter Sentiment Analysis. In Proceedings of the 2nd International Conference on Information Science and Education (ICISE-IE), Chongqing, China, pp. 517–520.

Yang L, Wang J, Tang Z, Xiong, NN (2021) Using conditional random fields to optimize a self-adaptive Bell-LaPadula model in control systems. IEEE Trans Syst Man Cybern Syst 51(7): 4505-4519. doi: 10.1109/TSMC.2019.2937551.

Zeng D, Dai Y, Li F, Sherratt R., Wang J. (2018). Adversarial learning for distant supervised relation extraction. Comput Mater Contin 55(1): 121–136. doi: https://doi.org/10.3970/cmc.2018.055.121

Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: A survey. Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery, 8(4): e1253.

Zhou J, Lu Y, Dai HN, Wang H, Xiao H (2019) Sentiment Analysis of Chinese Microblog Based on Stacked Bidirectional LSTM.162-167. 10.1109/I-SPAN.2018.00034.