ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979444

BERT for Sentiment Analysis in Thai Hotel Reviews

Petcharat Phuttakij[1]; Bandhita Plubin[2*]; Walaithip Bunyatisai[3]; Thanasak Mouktonglang[4]; Suwika Plubin[5]

[1,2,3,5]Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand,
[4]Data Science Research Center, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand,

Corresponding Author: Bandhita Plubin[2*]

Publication Date: 2025/03/10

Abstract: The rapid growth of the tourism and hospitality industry has resulted in a significant increase in customer reviews shared online. These reviews help tourists discover new accommodations with a favorable atmosphere and reasonable prices; the volume of reviews makes it challenging for travelers to choose the right option. Negative reviews, in particular, can influence booking decisions and impact a hotel's image. Sentiment analysis that categorizes comments has become an important tool for analyzing customer feedback automatically. Moreover, the Thai language has unique characteristics, such as its diverse writing styles, punctuation, and multiple meanings of a single word, which pose language barriers for sentiment analysis. Our method, employing the Bidirectional Encoder Representations from Transformers (BERT) model, analyzes hotel reviews in Thai, classifying sentiments into three categories: positive, neutral, and negative. This study uses a dataset of 37,011 hotel reviews. Our experiment results show that the BERT model has an accuracy of 89.31% and an F1 score of 89.43%, outperforming prior research. The findings contribute insights to a deeper understanding of customer reviews for the hospitality industry, enabling hotel operators to respond to customer feedback more effectively and improve their services. Analyzing and distilling reviews from customer feedback may assist tourists and others in making quicker choices. Finally, the results of this study show that using BERT for sentiment analysis can help businesses grow and become more competitive in the quickly changing tourism market.

Keywords: *Sentiment Analysis, Thai Language, Hotel Reviews, BERT, Natural Language Processing.*

How to Cite: Petcharat Phuttakij; Bandhita Plubin; Walaithip Bunyatisai; Thanasak Mouktonglang; Suwika Plubin. (2025). BERT for Sentiment Analysis in Thai Hotel Reviews. *International Journal of Innovative Science and Research Technology*, 10(2), 1774-1782. https://doi.org/10.5281/zenodo.14979444.

I. INTRODUCTION

The tourism and hospitality industry is highly competitive, with hotels relying on past customer reviews to attract tourists. These reviews are a key source of information that helps travelers select the right hotel for their needs. However, it is challenging to manually analyze, analyze, and categorize all the customer reviews within a limited time (Sungsri&Ua-apisitwong, 2017). Sentiment analysis is a method to categorize reviews as positive, negative, or neutral. It has emerged as a crucial tool for tackling this issue.

Many languages widely use sentiment analysis, but Thai still has a low adoption rate. This is mainly due to the characteristics of the Thai language, including flexible sentence structures, various spellings, various writing styles (for example, abbreviations and repetitions), and that a single word can convey multiple meanings. (Porntrakoon et al., 2021). These linguistic complexities pose significant challenges for traditional models, hindering their ability to efficiently process and interpret Thai text.

At present, deep learning has seen significant advancements, particularly with transformer-based models. The Transformer model has been developed, relying on self-attention mechanisms that allow it to process words and their contexts more efficiently (Vaswani et al., 2017). Among the most popular is BERT, introduced by Devlin et al. in 2019. BERT can understand context more accurately than models that process text in only one direction. This two-way approach enables BERT to pick up on the subtleties and meanings of words based on their surroundings, improving language learning and the analysis of complex texts (Devlin et al., 2019; Luo et al., 2020).

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979444

This study seeks to leverage the capabilities of the BERT model to address the complexities of Thai text. The objective of the research is to apply a BERT-based model for sentiment analysis in the Thai hotel industry and evaluate its performance comprehensively. Additionally, this research aims to provide insights that can enhance hotel services and enable tourists to quickly summarize information and make informed decisions.

II. LITERATURE REVIEW

Sentiment analysis, or opinion mining, has become an important tool for analyzing customer sentiment in industries such as the hospitality industry. Previous research in sentiment analysis, such asPasupa and Seneewong Na Ayutthaya (2019), employed CNN, LSTM, and Bi-LSTM. These models still rely on features to increase accuracy in sentiment analysis. Still, it is insufficient to capture contextual nuances in complex Thai texts. For sentiment analysis in Thai, unique writing, such as cohesive sentences, consonant vowels, and space, has hindered effectiveness (Sangsavate et al., 2023). Pugsee and Ongsirimongkol (2019) used deep learning methods to study sentiment analysis over Thai text belonging to different categories, such as hotels, restaurants, tourist attractions, and airlines. They experimented with CNN as well as LSTM to check and compare the effectiveness.The work by Bowornlertsutee and Paireekreng (2022) uses Bag-of-Words and Term Frequency-Inverse Document Frequency in analytics. It employs machine learning to classify text. However, if more sophisticated text processing is required, a transformer-based model is necessary.

The release of BERT came out; it had a big effect on natural language processing (Devlin et al., 2018). People now widely use BERT for sentiment analysis, including tasks involving the Thai language. For example, Harnmetta and Samanchuen (2022) achieved the accuracy of WangchanBERTa and BERT in analyzing stocks' Thai content. According to the experimental results, the BERT model performed less accurately than the WangchanBERTa model, achieving an accuracy of 88.70% and an F1 score of 89.12% in the sentiment analysis of stocks. When compared to the baseline, this score is considered higher than the set threshold. Moreover, Bimaputra and Sutoyo (2023) used BERT to analyze customer reviews of hotels from TripAdvisor; this study focuses on aspect-based sentiment analysis, examining specific aspects such as cleanliness, service, and location, which confirms that the BERT model is effective in analyzing sentiments across different aspects of hotel services. Similarly, Wu et al. (2024) studied the BERT model to apply sentiment analysis approaches and also showed its accurate performance in prediction. After adjusting the model and parameters, the BERT model has better accuracy in detecting and capturing deep insights from the texts. Maity (2024) analyzes hate speech in Thai language messages. All of these demonstrate the capabilities of applying BERT.

Comparisons between BERT and traditional models further highlight its superiority. Areshey and Mathkour (2023) compared BERT with Naive Bayes, Support Vector Machines, and K-Nearest Neighbors in sentiment analysis tasks and found that BERT outperformed these models in accuracy and F1 scores.Another one,Innork et al. (2023), executed a study that contrasts multiple sentiment classification models for various classes with an application to English reviews. They achieved results such that BERT could attain 0.865 for both accuracy and the F1 score, better than other models.

Applied sentiment analysis in the hotel industry has been explored in various studies. For example, Matarat (2024) study on hotel reviews in terms of environment was conducted using sentiment analysis to classify Thai text with a machine learning model to understand the feelings of reviewers. Improving staff services, keeping the hotel environment clean, and offering friendly prices can significantly increase customer satisfaction and business efficiency. Moreover, Pramudya and Alamsyah (2023) studied the development of a hotel recommendation tool that is appropriate and satisfactory according to the needs of each customer by analyzing comments and text reviews from guests using RoBERTa and BERT models to improve the experience and satisfaction of using hotel services.

III. METHODOLOGY

A. Data Collection

Our study used a dataset of customer reviews from various hotels in Thailand. While the dataset features a total of 11 columns, this analysis specifically utilized only those that pertain to hotel reviews and their sentiment categories. We collected 37,011 data from Agoda and TripAdvisor. Table 1 categorizes the reviews into positive, neutral, and negative sentiments.

Table 1: Example of a Hotel Review Dataset

Reviews	**Label**
ห้องกว้าง เตียงใหญ่ สะอาด ราคาไม่แพง The room is spacious, the bed is large, it’s clean, and the price is affordable.	Positive
รูปห้องที่จองไม่เหมือนห้องที่พักจริง The room I booked does not look like the actual room I stayed in.	Negative
ที่พักใจกลางเมือง มีร้านอาหาร Accommodation in the city center with a restaurant.	Neutral

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979444

B. Data Imbalance

Because the amount of hotel data is unequal in Figure 1, we handle the imbalance of hotel review data using the oversampling and undersampling techniques (Lemaître et al., 2017). This oversampling technique increases the number of minority class data to balance with the majority class data. Duplicating existing data samples helps the machine learning model learn from more balanced data, which enhances accuracy.

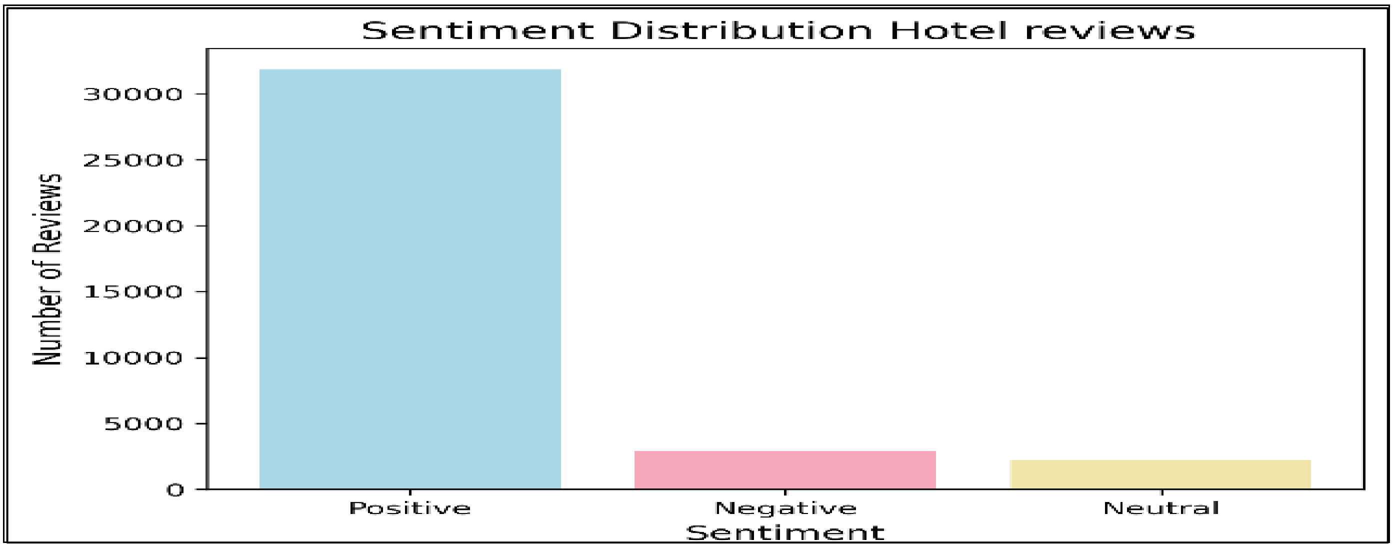


Fig 1: Shows the Sentiment Classes in the Hotel Review Dataset

C. Data Preprocessing

In the Thai language data preparation, we use the Auto Tokenization feature from the Hugging Face Transformers Library, which can automatically split the text into tokens. The library also provides other important data preparation functions, such as padding to make sentences longer and truncation to limit text length. This study split the dataset into 80% for training and 20% for testing, model development, and performance evaluation.

D. BERT Model

The BERT model, released by Google's research team, is an open-source language processing model that stands out for its profoundly bidirectional context analysis. It can concurrently understand word relationships in sentences from preceding and following contexts. At the same time, its operating principle is straightforward. BERT employs a bidirectional processing structure through the transformer encoder mechanism. Figure 2 illustrates this process. When receiving initial input vectors, they undergo analysis through multiple layers of encoders and decoders, generating output vectors that incorporate contextual relationships and essential features.

BERT has two important pre-processing components: Masked Language Model (MLM) and Next Sentence Prediction (NSP). BERT has a distinctive characteristic: employing a special token [MASK] as a tool for bidirectional context learning. This concept differs from traditional language processing models that analyze words in a single direction. BERT randomly masks words in sentences during its training process. Through this method, the model overcomes the limitations of unidirectional analysis, resulting in a higher understanding of word relationships within sentences.

$$L_{MLM} = -\sum_{t \in M} \log P(\hat{t} \mid t_{\setminus M}) \quad (1)$$

Where M is the set of tokens that have been altered or replaced with special tokens [MASK], $t_{\setminus M}$ is the tokens that have not been replaced with special tokens[MASK], and $\hat{t}$ represents the tokens selected for prediction in the task of guessing the masked tokens.

Next Sentence Prediction (NSP) is the process of seeing if there is a connection between the previous and the next sentence. For the NSP, it processes the input pairs of sentences. In 50% of the input, the last sentence will be connected to the previous sentence, marked as 'IsNext' and in 50% the other part consists of unrelated pairs of sentences. Since the last sentence is randomly selected, it is marked as 'IsNotNext'.

$$L_{NSP} = -[\, b \log P(\mathrm{IsNext} \mid C, D) - (1-b) \log P(\mathrm{IsNotNext} \mid C, D)\,] \quad (2)$$

Where C and D provide their input, let b be a binary label signifying whether sentence D follows sentence C.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979444

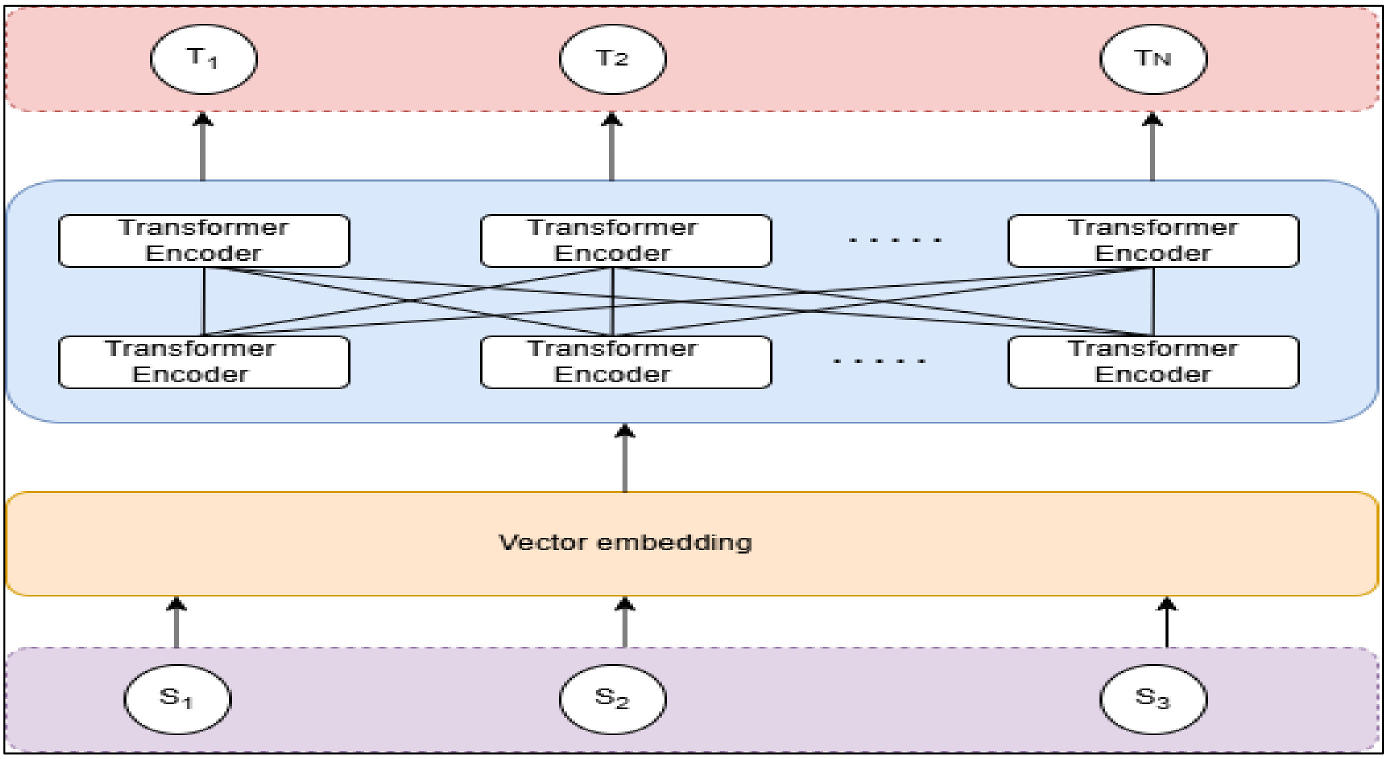


Fig 2: Structure of BERT

E. Hyperparameter Turning

Adjusting various configuration parameters can enhance the model's performance. Critical optimization elements include learning rate, batch dimensions, and training duration. Our experiment configures the parameters with 15 epochs, a batch size of 32, and a learning rate of 0.00001.

F. Performance Evaluation

In evaluating sentiment analysis model performance, we used several metrics to assess the performance of our models(Arrar et al., 2024).

- Accuracy: Indicates the proportion of correct predictions relative to the total number of samples, offering a general assessment of the model's overall correctness.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

- Precision: Indicates the accuracy of positive predictions, showing the proportion of correct predictions among all cases the model predicted as positive. This helps us understand how reliable the model is when it makes a positive prediction.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

- Recall: Measures the ability to identify all positive cases, showing the proportion of actual positive cases that the model successfully identified. This metric helps assess how well the model captures all relevant positive instances in the data.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

- F1 score: The F1 score is a metric that reflects the balance between precision and recall, utilizing their harmonic mean to evaluate overall effectiveness.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

IV. RESULT AND DISCUSSION

This section presents how we leverage the strengths of the BERT base model to analyze sentiment in Thai hotel reviews. In this experiment, The analysis revealsresults from its predictions. We evaluate the model's performance using four key metrics. Table 2 illustrates the sentiment classification performance model.

Table 2: Performance Metric of BERT Model

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
BERT	0.8931	0.9023	0.8929	0.8943

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979444

➢ *A Summary of the BERT Model's Performance Metrics for Analyzing Sentiment in Thai Hotel Reviews is Presented Below.*

- Accuracy: The model BERT achieved an accuracy of 0.8931, indicating that approximately 89.31% of the predictions made by the model were correct. This highlights the model's overall reliability in classifying sentiments into three levels: positive, neutral, and negative.
- Precision: With a precision score of 0.9023, the model demonstrated an excellent ability to correctly identify positive predictions without including many false positives. This reflects the model's excellent predictive performance.
- Recall: The recall score of 0.8929 signifies that the model can identify the true positive sample across all sentiment classes; this ensures that all relevant sentiments are taken into account.
- F1 Score: An F1 score of 0.8943, which is a balance between precision and recall, demonstrates the overall performance of the model in handling sentiment classification tasks. This indicator confirms that the model is suitable for predicting Thai hotel reviews. Because it can maintain a well-balanced balance between precision and recall.

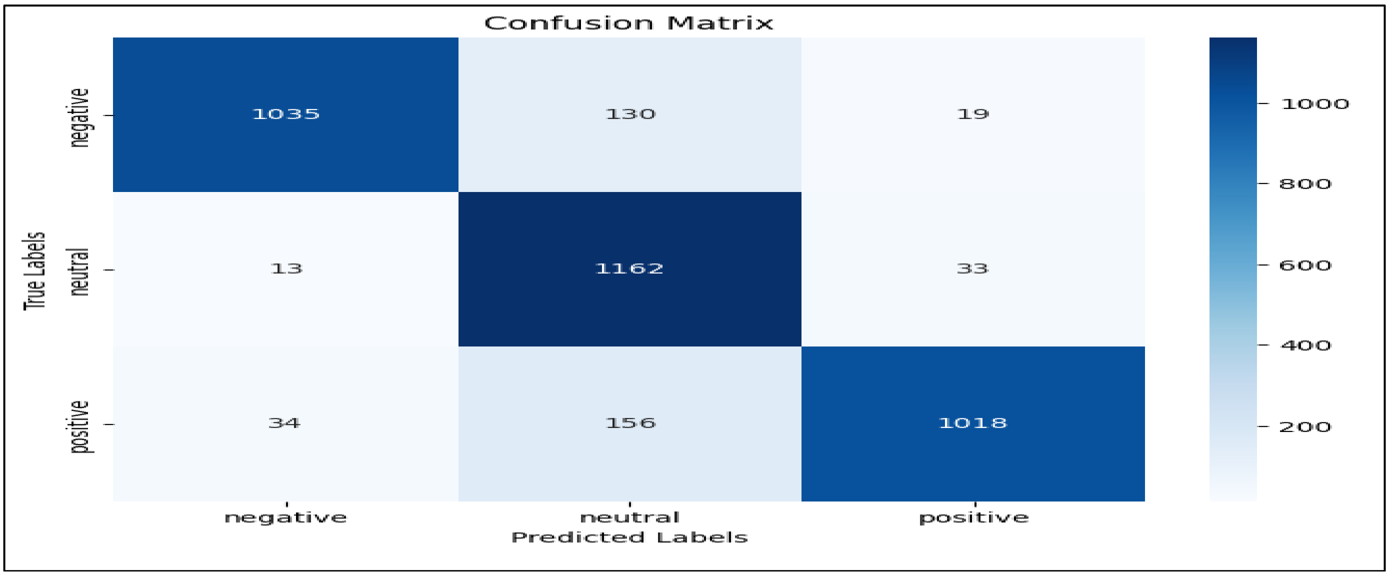


Fig 3: Confusion Matrix of BERT

In this experiment, we have shown the confusion metric in Figure 3 and the training loss function in Figure 4. The confusion matrix illustrates the performance of the BERT model in classifying sentiments into three categories: negative, neutral, and positive. The model correctly classified 1,035 negative, 1,162 neutral and 1,018 positive instances. Most misclassifications occur within closely related text categories. For instance, the model misclassified 130 negative messages as neutral and 156 positive messages as neutral. This shows that the model performs well. However, there are minor errors in distinguishing the sentiment of messages that are like the neutral class.

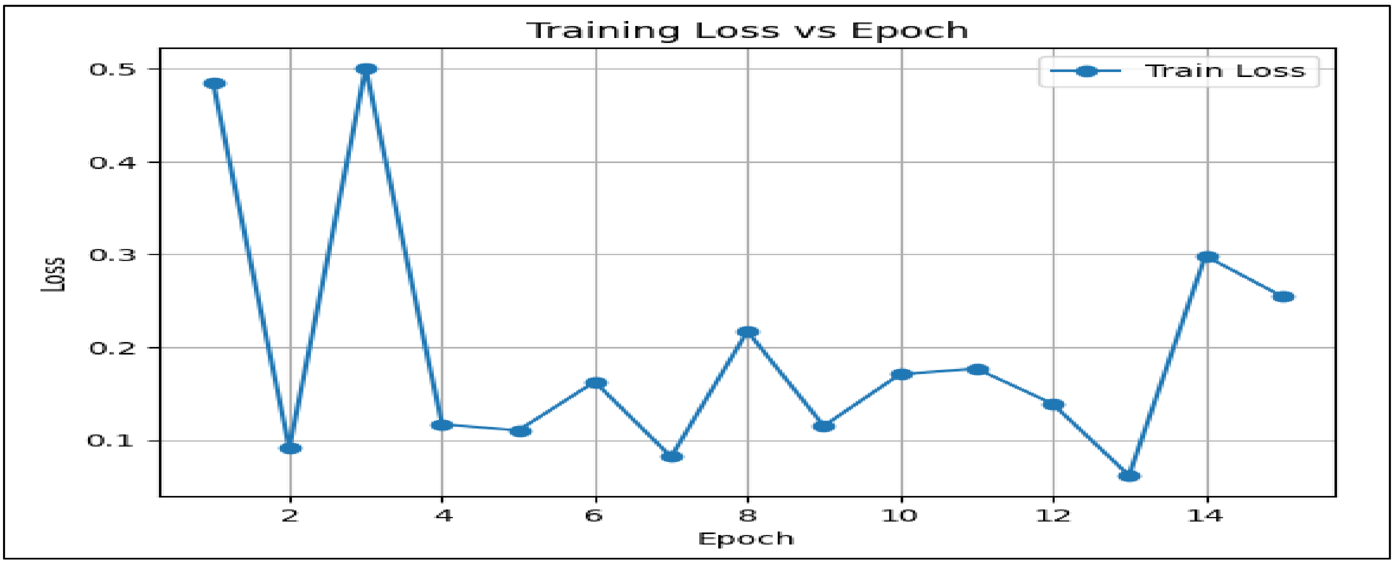


Fig 4: Training Loss per Epoch of BERT

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979444

The graph illustrates the training loss of the BERT model over 15 epochs. The loss goes down sharply in the first few epochs, which shows that the model is learning quickly. It then stays the same in later epochs, which shows that it has converged to a final loss value of 0.255. Minor fluctuations in the loss observed throughout the training process arise from the adjustments to parameters during optimization. The goal of these adjustments is to identify the model's optimal values, which will enhance its ability to identify patterns in the data and ultimately reduce errors.

The discussion reveals that the BERT model excels in sentiment analysis, specifically for hotel reviews in Thailand. The BERT model achieves an overall correctness percentage of 89.31%. The BERT model reliably classifies sentiment into three levels: positive, neutral, and negative, as evidenced by precision, recall, and F1 score metrics.This validates the model's strength and efficiency in analyzing sentiments. This overall score demonstrates the performance of the BERT model for sentiment analysis. Its bidirectionality enables it to capture the meaning of Thai text better than the original model.

The confusion matrix highlights the model's outstanding performance. However, some errors in classifying neutral text may be due to the use of ambiguous or unclear language in Thai reviews. The training loss graph supports the model's performance, showing a significant initial loss decrease, indicating effective learning of underlying patterns in the Thai dataset without significant overfitting.

Comparing this study's results with related research shows that the proposed BERT model outperforms previous sentiment analysis studies (Innork et al., 2023; Harnmetta&Samanchuen, 2022). The results demonstrate BERT's ability to efficiently process Thai text, confirming its applicability in sentiment analysis tasks. This capability is particularly valuable for industries that rely on understanding customer feedback.

V. CONCLUSION

This study presents the benefits of the BERT model in analyzing the sentiment of hotel customer reviews by applying it to Thai text. This process utilizes BERT's bidirectional detection capabilities, enabling it to accurately predict sentiment as positive, neutral, and negative with an accuracy of 89.31% and an F1 score of 89.43%. The research findings on the use of transformation-based models like BERT for sentiment analysis may prove advantageous for hotels and the hospitality industry in understanding customer feedback, which will influence the enhancement of service quality and the development of strategies to increase future customer satisfaction. Future research could focus on improving the model in terms of detecting neutral categories and separating sentiment analysis for each service aspect or expanding the dataset to include a variety of review categories. This study demonstrates that the BERT model is an effective tool for developing sentiment analysis in Thai hotel reviews, which has important implications for both academic research and industry practice, enhancing competitiveness in the rapidly changing tourism market.

- **Grant/Funding:** Scholarship for Undergraduate Students to Pursue Graduate Studies from the Faculty of Science, Chiang Mai University.

REFERENCES

[1]. Areshey, A., &Mathkour, H. (2023). Transferlearning for sentiment classification using bidirectional encoder representations from transformers (BERT) model. *Sensors*, *23*(11), 5232.

[2]. Arrar, D., Kamel, N., &Refoufi, A. (2024). Implicit Sentiment Analysis of Hotel Comments for Disaster Management: A Comparative Study of Machine Learning, Deep Learning, and BERT Model. *In 2024 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)* (pp. 1-7). IEEE.

[3]. Bimaputra, D. S., &Sutoyo, E. (2023). Aspect-Based Sentiment Analysis of Hotels in Bali on Tripadvisor Using BERT Algorithm. *International Journal of Advances in Data and Information Systems*, *4*(1), 29-40.

[4]. Bowornlertsutee, P., &Paireekreng, W. (2022). The Model of Sentiment Analysis for Classifying the Online Shopping Reviews. *Journal of Engineering and Digital Technology (JEDT)*, *10*(1), 71-79.

[5]. Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[6]. Harnmetta, P., &Samanchuen, T. (2022). Sentiment analysis of thai stock reviews using transformer models. *In 2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 1-6). IEEE.

[7]. Innork, K., Polpinij, J., Namee, K., Kaenampornpan, M., Saisangchan, U., &Wiangsamut, S. (2023). A Comparative Study of Multi-Class Sentiment Classification Models for Hotel Customer Reviews. *In 2023 Research, Invention, and Innovation Congress: Innovative Electricals and Electronics (RI2C)* (pp. 88-92). IEEE.

[8]. Koroteev, M. V. (2021). BERT: a review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943*.

[9]. LemaÃŽtre, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of machine learning research*, *18*(17), 1-5.

[10]. Luo, H., Ji, L., Li, T., Duan, N., & Jiang, D. (2020). GRACE: Gradient harmonized and cascaded labeling for aspect-based sentiment analysis. *arXiv preprint arXiv:2009.10557*.

[11]. Maity, K., Poornash, A. S., Bhattacharya, S., Phosit, S., Kongsamlit, S., Saha, S., &Pasupa, K. (2024). HateThaiSent: Sentiment-Aided Hate Speech Detection in Thai Language. *IEEE Transactions on Computational Social Systems.*

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979444

[12]. Matarat, K. (2024). Enhancing Hotel Management: a Sentiment Analysis Approach to Assessing Customer Impressions on Environment-Based Reviews. *International Journal of Professional Business Review*, *9*(1), e04152-e04152.

[13]. Pasupa, K., & Ayutthaya, T. S. N. (2019). Thai sentiment analysis with deep learning techniques: A comparative study based on word embedding, POS-tag, and sentic features. *Sustainable Cities and Society*, *50*, 101615.

[14]. Porntrakoon, P., Moemeng, C., &Santiprabhob, P. (2021). Text summarization for Thai food reviews using simplified sentiment analysis. *In 2021 18th international joint conference on computer science and software engineering (JCSSE)* (pp. 1-5). IEEE.

[15]. Pramudya, Y. G., & Alamsyah, A. (2023). Hotel Reviews Classification and Review-based Recommendation Model Construction using BERT and RoBERTa. *In 2023 6th International Conference on Information and Communications Technology (ICOIACT)* (pp. 437-442). IEEE.

[16]. Pugsee, P., &Ongsirimongkol, N. (2019, November). A classification model for Thai statement sentiments by deep learning techniques. *In Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems* (pp. 22-27).

[17]. Sangsavate, S., Sinthupinyo, S., &Chandrachai, A. (2023). Experiments of Supervised Learning and Semi-Supervised Learning in Thai Financial News Sentiment: A Comparative Study. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(7), 1-36.

[18]. Sungsri, T., &Ua-apisitwong, U. (2017). The analysis and summarizing system of thaihotel reviews using opinion mining technique. *In Proceedings of the 5th International Conference on Information and Education Technology* (pp. 167-170).

[19]. Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

[20]. Wu, Y., Jin, Z., Shi, C., Liang, P., & Zhan, T. (2024). Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis. *arXiv preprint arXiv:2403.08217*.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979444

AUTHORS' INFORMATION FORM

Paper Title	BERT for Sentiment Analysis in Thai Hotel Reviews
Corresponding Author (Author Name & Email)	Bandhita Plubin(Email: Bandhita.p@cmu.ac.th)

First Author – Information

First Name	**Petcharat**	**Last Name**	**Phuttakij**
Designation	-	Department	Statistics
University	Chiang Mai University	Mail ID	petcharat_ph@cmu.ac.th
Contact No.	0622640774	ORCID ID	
Residential Address	222/433 Dcondo CampusChiang Mai, Suthep,Mueang, Chiang Mai, Thailand, 50200.		

Second Author – Information

First Name	**Bandhita**	**Last Name**	**Plubin**
Designation	Asst. Prof.	Department	Statistics
University	Chiang Mai University	Mail ID	Bandhita.p@cmu.ac.th
Contact No.	+66951546659	ORCID ID	0009-0002-0115-2455
Residential Address	168/58 Palmview, Chang Phueak, Mueang, Chiang Mai, Thailand, 50300.		

Third Author– Information

First Name	**Walaithip**	**Last Name**	**Bunyatisai**
Designation	Asst. Prof.	Department	Statistics
University	Chiang MaiUniversity	Mail ID	Walaithip.bun@cmu.ac.th
Contact No.	0966822629	ORCID ID	https://orcid.org/0000-0002-6675-2948
Residential Address	323/180 Kulphan Ville Village 9, San Phak Wan Subdistrict, Hang Dong District, Chiang Mai Thailand, 50230		

Fourth Author– Information

First Name	**Thanasak**	**Last Name**	**Mouktonglang**
Designation	Asst. Prof.	Department	Mathematics
University	Chiang MaiUniversity	Mail ID	Thanasak.m@cmu.ac.th
Contact No.	081-716-0654	ORCID ID	0000-0001-7839-9937
Residential Address	106 M.10 T.Sanpooloy A. Doi Saket Chiang Mai, 50220		

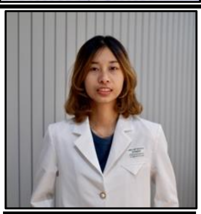
Fifth Author– Information

First Name	**Suwika**	**Last Name**	**Plubin**
Designation	-	Department	Statistics
University	Chiang Mai University	Mail ID	Suwika_j@cmu.ac.th
Contact No.	+66953629524	ORCID ID	0009-0000-8781-3861
Residential Address	168/58 Palmview, Chang Phueak, Mueang, Chiang Mai, Thailand, 50300.		

AUTHOR'S BIOGRAPHY

P. Phuttakij (PetcharatPhuttakij) graduated with a Bachelor's degree in Data Science from Chiang Mai University. She is currently a Master's student in Applied Statistics and Data Analytics at Chiang Mai University.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979444

	**B. Plubin)BandhitaPlubin(** is an Assistant Professor of Statistics at Chiang Mai University. Her research focuses on Text Analytics, Natural Language Processing)NLP(, and Multivariate Analysis. Statistical learning techniques will be applied to regression problems, particularly in the areas of text classification and language processing.
	**W. Bunyatisai (WalaithipBunyatisai)** graduated with Bachelor's degree in Statistics from Thammasat University. Then she obtained her Master's degree and PhD in Statistics from Chulalongkorn University and Kasetsart University, respectively. Currently, she is a lecturer at the Department of Statistics, Faculty of Science, Chiang Mai University, specializing in applied statistics.
	**T. Mouktonglang (Thanasak**Mouktonglang)is a faculty member at the Department of Mathematics, Faculty of Science, Chiang Mai University, Thailand. His academic and research expertise spans Data Science, Machine Learning, Deep Learning, Optimization, Numerical Methods, and Graph Theory, with a particular focus on their applications in climate change analysis, urban infrastructure planning, and carbon capture and storage. He has contributed to advancing knowledge through interdisciplinary approaches, integrating mathematical and computational models to address complex challenges.
	**S. Plubin (SuwikaPlubin**) is a Ph.D. student in Statistics at Chiang Mai University. Her research focuses on Text Analytics and Natural Language Processing (NLP). She works on projects related to Generative AI for market research applications.