EFFECTIVENESS OF SENTIMENT ANALYSIS FROM OPINIONS USING GENERATIVE AI: A CASE STUDY OF THAI UNDERGRADUATE STUDENTS' COMMENTS IN ON-DEMAND LEARNING SYSTEMS

Nutthapat Kaewrattanapat[1], Jarumon Nookhong[2], Martusorn Khaengkhan[3], Napasri Suwanajote[4], Sittichai Pintuma[5]

Suan Sunandha Rajabhat University, Thailand
Email: nutthapat.ke@ssru.ac.th[1]; jarumon.no@ssru.ac.th[2]; martusorn.kh@ssru.ac.th[3]; napasri.su@ssru.ac.th[4]; sittichai.pi@ssru.ac.th[5]

ABSTRACT

This research investigates the effectiveness of Generative AI in conducting sentiment analysis on comments from Thai undergraduate students in an on-demand learning system. By designing and applying a novel methodology, the study examines the alignment between human expert sentiment classification and Generative AI predictions. A dataset of 200 comments was analyzed, with sentiments categorized into positive, negative, and neutral classes by both language experts and an AI model, specifically utilizing the capabilities of ChatGPT from OpenAI (gpt-3.5-turbo). The accuracy and efficiency of the AI's sentiment classification were evaluated using a Confusion Matrix, which revealed an overall accuracy of 73.63%. The results indicated a high level of precision in the positive and negative categories but highlighted discrepancies in the neutral category, underscoring the nuances and challenges inherent in automated sentiment analysis. These findings contribute to the field of AI-driven sentiment analysis by demonstrating both the promise and complexities of utilizing Generative AI in educational settings.

Keywords: Sentiment Analysis, Generative AI, Opinions Analysis

INTRODUCTION

On-demand learning systems offer several advantages over traditional classroom settings, including increased flexibility, personalized learning paths, and improved accessibility for geographically dispersed students (Kim & Hong, 2014; Xu & Wang, 2017). However, effectively assessing student engagement and sentiment within these online environments remains a significant challenge. Traditional methods, such as surveys and focus groups, are often time-consuming, resource-intensive, and lack immediacy, limiting their ability to provide real-time insights into student learning experiences (Concannon & Yang, 2018; Hwang & Chen, 2012).

The emergence of Generative AI has revolutionized natural language processing, enabling sophisticated analysis of vast textual datasets. In this context, data mining techniques play a pivotal role in analyzing and categorizing student behavior at the tertiary education level. These fundamental technologies are crucial for advancing into Artificial Intelligence (AI) (Kularbphettong & Tongsiri, 2012). Sentiment analysis, a critical subfield within AI, empowers us to decode the emotional undercurrents in written communication (Haddi et al., 2013). Within educational settings, deciphering student sentiment offers invaluable insights into their experiences and engagement, particularly in on-demand learning platforms where direct feedback is often scarce. Yet, despite their widespread adoption in universities, such systems often lack effective mechanisms for gauging student sentiment, hindering course improvement and student support. This challenge is further amplified in culturally diverse contexts, such as with Thai undergraduate students, where traditional analysis methods may overlook nuanced emotional expressions (Liu, 2012).

To address this critical gap, this research investigates the potential of Generative AI to analyze sentiment from Thai students' comments within an on-demand learning environment. By leveraging the advanced capabilities of AI, we aim to unlock a deeper understanding of student feedback, paving the way for enhanced educational experiences (Zhang et al., 2018).

Moreover, by assessing the effectiveness of Generative AI in this context, the study seeks to contribute to the burgeoning field of AI applications in education, illuminating both the promising potential and intricate limitations of current technology.

RESEARCH OBJECTIVES

This research study was aimed to

1) To design a process for sentiment analysis from opinions using Generative AI:
   A case study of Thai undergraduate students' comments in on-demand learning systems.
2) To evaluate the effectiveness of sentiment analysis from opinions using Generative AI:
   A case study of Thai undergraduate students' comments in on-demand learning systems.

CONCEPTUAL FRAMEWORK

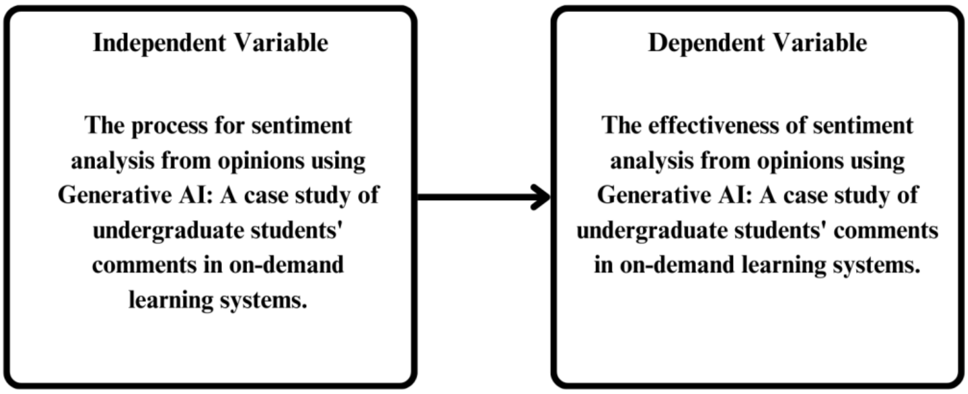

Figure 1. Conceptual Framework of the Study

The conceptual framework illustrated in Figure 1. depicts the relationship between the independent variables and the dependent variables in our study. The primary independent variable is the use of Generative AI, which refers to the designed methodologies and applications of artificial intelligence in analyzing student comments. These comments, characterized by their themes, tones, and complexity, constitute the secondary independent variable and are derived from undergraduate students participating in on-demand learning systems.

The dependent variables are centered around the outcomes of the Generative AI's sentiment analysis. The first dependent variable is the accuracy of sentiment analysis, indicating how effectively the Generative AI can identify and categorize emotions expressed in student comments. The second dependent variable is the overall efficiency of the process, encompassing considerations such as the time required for analysis, the system's capability to process large volumes of data, and resource utilization.

This framework is pivotal in exploring how Generative AI can be leveraged to discern and quantify emotional responses from textual feedback provided by students, thereby offering insights into the AI's operational efficacy in educational contexts.

METHODOLOGY

This study conducted a systematic collection and analysis of sentiment from comments provided by 200 undergraduate students engaged in a general education course delivered via an on-demand format. The data collection process involved the accumulation of 200 distinct comments, which corresponded to one comment per student participant.

To prepare the data for analysis, language experts were enlisted to categorize the comments into three sentiment groups: positive, negative, and neutral. This preliminary classification served as a standard for evaluating the subsequent automated sentiment analysis.

For the analysis phase, the student comments were meticulously recorded into a Spreadsheet (Google Sheets). This platform was chosen for its compatibility with the "GPT for Sheets" extension, which allows for the integration of the Generative AI's API. Specifically, the study utilized ChatGPT (gpt-3.5-turbo), powered by OpenAI, to perform the sentiment analysis. The formula =*GPT_CLASSIFY(Cell, "positive, neutral, negative")* was inserted into the spreadsheet to initiate the sentiment analysis process by the Generative AI.

The outcomes of the sentiment analysis by both the human experts and the Generative AI were then subjected to a comparative performance evaluation using a Confusion Matrix. This approach facilitated the computation of the accuracy metric, which quantified the efficacy of the Generative AI in sentiment classification against the expert-labeled dataset.

The Confusion Matrix is a tool used in evaluating the performance of classification systems and is often employed to measure the accuracy of sentiment analysis models. In a binary classification scenario, the Confusion Matrix comprises four main components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

The accuracy of a system is calculated using the formula:

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

Where:

TP (True Positives): The number of instances correctly classified as positive.

TN (True Negatives): The number of instances correctly classified as negative.

FP (False Positives): The number of instances incorrectly classified as positive (when they should be negative).

FN (False Negatives): The number of instances incorrectly classified as negative (when they should be positive).

Accuracy provides the proportion of all correct predictions (both positive and negative) out of all predictions made. While it is a good standard measure for overall performance, accuracy may not be sufficient for analysis in certain scenarios, particularly in cases of class imbalance.

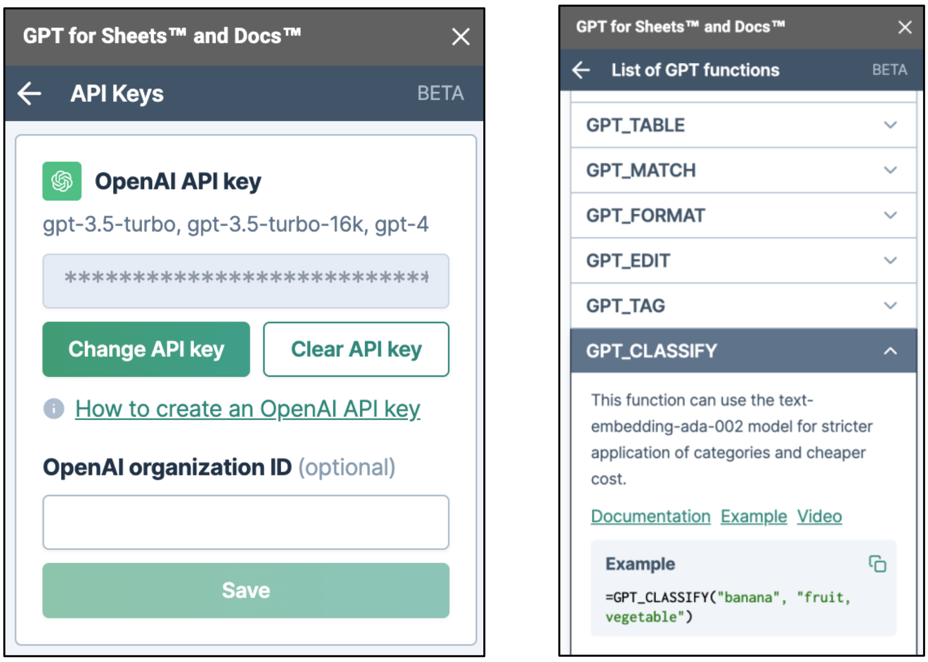

Figure 2. GPT for Sheets™ Extension Interface

The Figure 2. showcases the GPT for Sheets™ extension (Google, 2022) used for sentiment analysis in Google Sheets. On the left, the API Keys section displays where users can enter their OpenAI API key, essential for enabling the AI's capabilities within the spreadsheet. It also provides a space for the optional OpenAI organization ID and a "Save" button to secure the entered information. On the right, the List of GPT functions section outlines various functionalities available in the extension, such as GPT_TABLE, GPT_MATCH, and GPT_CLASSIFY, among others. The highlighted GPT_CLASSIFY function description explains its use of the text-embedding-ada-002 model for cost-effective and category-specific analysis, with an example provided for classifying a "banana" as "fruit" or "vegetable". This extension significantly streamlines the process of analyzing and categorizing sentiments from comments, as demonstrated in the research on Thai undergraduate students' feedback in an on-demand learning system.

RESULT

The Results section presents a detailed analysis of the sentiments expressed by Thai undergraduate students in their comments, as classified by human experts and predicted by Generative AI. The core objective was to measure the concordance between the expert and AI classifications to assess the AI's proficiency in sentiment analysis within an on-demand learning context. Utilizing a dataset formatted for clarity and precision, the study examined 200 comments, each labeled with an actual sentiment class and a corresponding AI-predicted sentiment class. The effectiveness of the Generative AI was quantified using a Confusion Matrix, offering a nuanced view of the model's performance across different sentiment categories.

Table 1. Dataset Formatting and Sentiment Analysis Results

id	Comments in English	Actual Class	Predict Class
1	More knowledge about the university and Suan Sunandha Palace.	positive	positive
2	I have gained a lot of knowledge about the Suan Sunandha Palace.	positive	positive
3	Received diverse knowledge and understanding about the Suan Sunandha Garden.	positive	positive
4	Inspecting the project without specifying what is wrong, I thought that if the inspection is done, it should indicate what we did wrong. Just telling us where we were deducted points based on what we have provided, we wouldn't know.	negative	negative
5	Easy to understand content	positive	positive
6	How outdated is the content we are learning? It only teaches outdated knowledge and I don't know what I will use it for in the future.	negative	negative
7	The content of the lessons is interesting and the teachers who teach make it easy to understand.	positive	positive
8	The content is too deep.	negative	negative
9	It is a very good teaching, even though it is difficult to understand online, I understand it.	positive	positive
10	I made a mistake in my first year project and received 0 out of 30 points.	negative	negative

* Example from dataset of 200 comments

This table presents a snippet of the dataset used in the sentiment analysis, illustrating the format and the results of the predictive analysis performed by the Generative AI. Each row contains a unique identifier (id), the original comment in Thai, the translated comment in English, the Actual Class as labeled by language experts, and the Predict Class as determined by the Generative AI.

Column 1 (id): A sequential identifier assigned to each comment.

Column 2 (Comments in Thai): The original comments made by the students, written in Thai.

Column 3 (Comments in English): The English translations of the Thai comments.

Column 4 (Actual Class): The sentiment classification (positive, negative, neutral) assigned by human experts.

Column 5 (Predict Class): The sentiment classification as predicted by the Generative AI.

The table exemplifies the alignment between expert assessment and AI prediction, demonstrating the AI's capability in understanding and classifying sentiments in student comments. The small sample provided here reflects a consistent match in the positive category and an accurate recognition in the negative category, showcasing the potential of Generative AI in educational sentiment analysis.

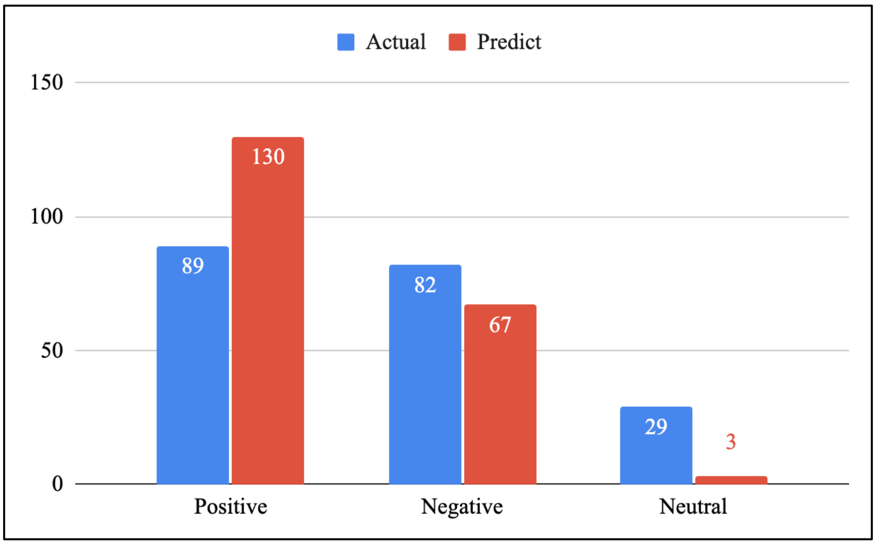

Figure 3. Comparison of Expert-Labeled and AI-Predicted Sentiment Classes

The bar graph in Figure 3. showcases a comparative analysis of sentiment classifications. The blue bars represent the actual number of comments labeled by language experts as Positive, Negative, and Neutral. The red bars reflect the classifications as predicted by the Generative AI using sentiment analysis on comments from Thai undergraduate students engaged in an on-demand learning system.

The Positive category shows a discrepancy, with experts labeling 130 comments as positive, whereas the AI predicted 89 comments as such.

For the Negative sentiment, experts labeled 82 comments, and the AI predicted 67 comments, indicating a closer agreement between the two.

The Neutral category exhibits a significant variance with experts labeling 29 comments, while the AI identified only 3.

This graph demonstrates the practical utility of Generative AI in sentiment analysis but also highlights the challenges in aligning AI predictions with expert opinions, particularly in nuanced categories such as Neutral.

Table 2. Confusion Matrix of Sentiment Analysis

		Predicted			
		Positive	**Negative**	**Neutral**	
Actual	**Positive**	88	2	0	**97.78%**
	Negative	21	59	2	**71.95%**
	Neutral	22	6	1	**3.45%**
		67.18%	**88.06%**	**33.33%**	**Accuracy 73.63%**

The Confusion Matrix depicted in Table 2. demonstrates the performance of the Generative AI model in classifying sentiments from students' comments within an on-demand learning system. The table compares the predicted sentiment categories—Positive, Negative, and Neutral—against the actual sentiments labeled by researchers.

For the Positive sentiment, 88 comments were correctly predicted, resulting in a high classification accuracy of 97.78% for this category.

In the Negative category, 59 comments were correctly identified, with a classification accuracy of 71.95%.

The Neutral category had significantly lower classification accuracy, with only 1 comment correctly identified, yielding a 3.45% accuracy rate.

The rows of the matrix indicate the actual sentiments, while the columns show the predicted sentiments by the model. The classification precision of the model for each sentiment is given by the percentages on the diagonal of the matrix. The overall accuracy of the model is 73.63%, which is the sum of correctly predicted sentiments (88+59+1) divided by the total number of comments. The precision for each sentiment category is illustrated by the percentages in the respective rows, and the recall rates are depicted by the percentages in the columns.

CONCLUSION

This study set out to design and evaluate the effectiveness of sentiment analysis using Generative AI on a dataset of comments from Thai undergraduate students participating in an on-demand learning system. The investigation involved a dual approach to sentiment classification: one by human experts and the other by Generative AI, specifically ChatGPT powered by OpenAI.

The analysis revealed that Generative AI could match the expert labeling with high accuracy in positive and negative categories, as shown by the high classification accuracy rates of 97.78% and 71.95% respectively. However, the classification of neutral comments posed a challenge, with the AI model only achieving a 3.45% accuracy rate in this category. The overall accuracy of the AI in classifying sentiments was found to be 73.63%, which, while robust, indicates room for improvement, especially in the nuanced identification of neutral sentiments.

These findings suggest that while Generative AI exhibits substantial promise for sentiment analysis, its current application is more reliable for polarized sentiments rather than neutral or ambiguous ones. This research highlights the need for further refinement of AI models to better interpret the subtleties of human emotional expression in textual data. Additionally, it emphasizes the potential for AI to support and augment human expertise in educational contexts, providing a tool for understanding student feedback at scale.

Future work should focus on enhancing the AI's sensitivity to nuanced expressions and incorporating contextual understanding to improve accuracy further. Continued development in this field could significantly benefit educational institutions by providing deeper insights into student experiences and feedback.

ACKNOWLEDGMENTS

The authors would like to thank Suan Sunandha Rajabhat University, Bangkok, Thailand (http://www.ssru.ac.th/) to provide funding support to attend the dissemination of research on this and thank family, friends, colleagues, students in the field of Digital Innovation Management and Content, Digital Technology for Education and The Office of General Education and Innovative e-Learning for cooperation and provide the dataset in research, all of you.

REFFERENCES

Google. (2022). GPT for Sheets and Docs (Version 1.0). Retrieved from https://workspace.google.com/marketplace/app/gpt_for_sheets_and_docs/677318054654

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. Procedia Computer Science, 17, 26-32. https://doi.org/10.1016/j.procs.2013.05.005

Kim, S., & Hong, J. (2014). The impact of on-demand learning on student engagement: A systematic review of the literature. Computers & Education, 71, 181-190.

Kularbphettong, K. , Tongsiri, C. (2012). Mining Educational Data to Analyze the Student Motivation Behavior. World Academy of Science, Engineering and Technology, Open Science Index 68, International Journal of Information and Communication Engineering, 6(8), 1032 - 1036. https://doi.org/10.5281/zenodo.1079926

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Xu, J., & Wang, B. (2017). A review of research on on-demand learning. Educational TechnologyResearch and Development, 65(4), 823-848.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.https://doi.org/10.1002/widm.1253