MINING SOCIAL MEDIA FOR MENTAL HEALTH INSIGHTS: A TEXT ANALYTICS STUDY ON UUM STUDENTS

[1]Farzana Kabir Ahmad & [2]Nurul Farrah Bahazir

[1&2]School of Computing, Universiti Utara Malaysia, Malaysia

[1]*Corresponding author: farzana58@uum.edu.my*

ABSTRACT

This study explores the application of sentiment analysis using a BERT-based model on confession texts from a university social media platform to understand student mental health at Universiti Utara Malaysia (UUM). Mental health issues among university students are often shared anonymously through platforms like UUM Confession. This research aims to identify prevailing emotional trends and sentiments expressed by students. The methodology involves data crawling from Telegram, data preprocessing, and classification into positive, negative, and neutral sentiments using a fine-tuned BERT model. The study also uses keyword extraction and visualisation techniques to support straightforward interpretation of a massive dataset. Results reveal a high occurrence of negative sentiments, often associated with stress, anxiety, and relationship issues, highlighting the urgent need for mental health interventions. The findings contribute to the broader field of educational data mining and provide actionable insights for university counsellors and administrators.

Keywords: BERT, mental health, sentiment analysis, social media mining, UUM Confession

INTRODUCTION

The rising incidence of mental health issues among university students is a global concern, as intense academic demands, financial stress, and the emotional challenges of adolescence often characterise tertiary education settings. Various studies have confirmed data on rising anxiety, depression, and stress symptoms among university students that can largely influence their academic performance and life

quality (Babu & Kanaga, 2022; Kim et al., 2022). The same applies to Malaysia, where concerns over mental health provision and services are on the rise, particularly among students.

According to the World Health Organisation (WHO), one in every eight people in the world lives with a mental disorder. As mental health disorders increase globally, people mostly choose social media to share their struggles and find support. Social media also allows its users to express their mental health experiences without fear of judgment. In Malaysia, the majority of the younger population age group is interacting on social media, specifically those aged 25 to 34, comprising 32.9% of the population. Close behind is the 18-to-24-year age group with 24.5% (Howe, 2024). This high engagement among young adults emphasises how crucial social media is as a forum for conversations on mental health. It is because it gives them a way to express their emotions, connect with people who share the same struggles and find their support.

To cope with their dilemma, many students turn to online media, particularly anonymous social media forums, where they can freely express personal experiences and emotions without fear of judgment. One of these is the UUM Confession on Telegram, which has become an online forum for UUM students to express their emotional and psychological struggles, such as burnout, learning issues, and personal and relationship problems. These confessions, while anecdotal, are valuable data that can be explored to find patterns and trends in student mental health.

This study explores how sentiment analysis of these confession text messages, using Bidirectional Encoder Representations from Transformers (BERT)-based machine learning, can be employed to categorise them according to emotional tone. BERT offers a significant advantage because it understands language context in a bidirectional way, meaning it looks at words based on both the left and right context simultaneously. This allows it to capture subtle differences in meaning that older models like Word2Vec or GloVe often miss. Employing advanced Natural Language Processing (NLP) techniques, the study seeks to detect underlying mental health issues apparent in these posts. The ultimate goal is to enable learning institutions to identify prevailing trends in mental health, raise awareness, and develop timely intervention strategies that will enhance students' emotional health and academic performance.

The purpose of this project is to develop a system that applies advanced sentiment analysis using BERT to detect and evaluate the emotional state of students through their anonymous confessions. The project aims to support the university in identifying emotional trends, addressing emerging mental health concerns, and ultimately improving student support services through data-driven insights. In addition, by analysing the emotional tone, sentiment polarity, and contextual language patterns within these unstructured texts, the system aims to accurately detect students' emotional states, such as stress, anxiety, frustration, or well-being.

The overarching goal is to provide the university with a data-driven tool that can uncover underlying emotional trends within the student population over time. This insight will enable early identification of collective psychological distress or emerging mental health concerns. In turn, such information can inform targeted interventions, shape campus mental health policies, and improve support services. Ultimately, the project aims to enhance the university's ability to respond proactively to students' emotional needs, fostering a healthier and more supportive academic environment.

This study investigates how machine learning, particularly BERT-based sentiment analysis, can be utilised to process these unstructured confession texts and categorise them according to emotional tone. By leveraging advanced NLP techniques, the research seeks to uncover underlying mental health issues

reflected in these posts. The goal is to help academic institutions identify prevailing mental health trends, create awareness, and develop timely intervention strategies to improve students' emotional well-being and academic success. This paper is organised as follows: Section 2 elaborates on previous works, while a detailed description of the research methodology is given in Section 3. Section 4 presents the experimental results and an insightful discussion on this work. Finally, Section 5 presents the concluding remarks.

PREVIOUS WORKS

In the digital age, social media platforms have become significant sources of data that reflect public opinions, sentiments, and behavioural patterns. The vast amount of user-generated content on platforms like Twitter, Facebook, and Instagram has made social media data crawling, sentiment analysis, and text analytics key research areas in social computing, especially in academic environments such as universities.

Sentiment analysis, also known as opinion mining, refers to the computational study of opinions, emotions, and attitudes expressed in text. It categorises social media content into sentiments such as positive, negative, or neutral. Liu (2012) outlines the three main approaches to sentiment analysis: rule-based, machine learning-based, and hybrid models. More recent studies (Zhang et al., 2022) emphasise the effectiveness of deep learning methods such as LSTM (Long Short-Term Memory) and BERT for analysing sentiment in informal and short-form text like tweets or Instagram captions. Within a university context, sentiment analysis helps identify students' emotional responses to events, policies, or academic experiences (Mohammad et al., 2021).

Earlier research works in sentiment analysis have used different NLP methods to evaluate social media content. Conventional machine learning models like Naive Bayes and Support Vector Machines (SVM) were moderately successful but not contextually sensitive. Deep learning and attention-based transformer models like BERT (Singh & Singh, 2020) have been proven superior to earlier approaches because they can handle the bidirectional context of language. Text analysis includes a broader range of techniques, such as keyword extraction, topic modelling, and frequency analysis, to understand patterns in textual data. Tools like Natural Language Toolkit (NLTK), SpaCy, and Gensim are frequently used for preprocessing and analysis. In a study by Altrabsheh et al. (2015), text analysis of student feedback revealed key concerns and suggestions for improving educational services. Word clouds, term frequency-inverse document frequency (TF-IDF), and Latent Dirichlet Allocation (LDA) are commonly used for visualising and modelling themes in student-generated content.

Abbasi et al. (2018) also emphasised the importance of text analysis to enable sense-making on social media, noting that linguistic context is a key factor in understanding user-generated content. Recurrent and convolutional neural networks have also been combined in mental health prediction models, such as the CNN-BiLSTM model discussed by Abdurrahim and Fudholi (2024), which has been effective in classifying emotional tones from text. In recent years, researchers have increasingly applied text mining and AI-driven techniques to social media data to gain mental health insights among university students and adolescents. For instance, Jedeejit, Nuankaew, Nasa-ngium, and Nuankaew (2025) deployed text-mining analytics on social media posts from university students to build predictive models capturing emotional risk indicators. Similarly, a systematic review demonstrated that user-generated social media text is highly effective in predicting depression, with a substantial effect size (r = 0.63). It also highlights the influence of demographic, linguistic, and temporal features on model performance (Kaur et al.,

2025). Qualitative research has also enriched understanding, with evidence that university students perceive social media as having a dual role in moderating anxiety—certain features (e.g., social connection and escapism) help relieve anxiety, while others (e.g., social comparison and procrastination) exacerbate it (Balci & Sarac, 2023). Complementing these empirical studies, a study by Barrot et al. (2024) emphasised how text-mining, social media analysis, and machine learning are converging to unlock novel insights in mental health research, particularly around student burnout and pandemic-driven stress contexts. Although these studies examine mental health among students and adolescents, most of this research relies on a quantitative approach and does not offer any visualisation dashboards.

In the Malaysian context, Facebook and Twitter have been the focus of sentiment analyses (Howe, 2024), while others, such as Koh and Liew (2022), have examined loneliness during the COVID-19 pandemic through social media. Up until now, however, little work has been conducted using university-based platforms like UUM Confession. Social media provides an authentic insight into users' mental state, according to Herdiansyah et al. (2022). This presents a valuable opportunity to apply advanced NLP models to institution-specific, student-authored content, giving thought-provoking insights into their mental health journeys. Research by Chancellor and De Choudhury (2020) has highlighted the predictive capability of AI models in determining mental health states from online behaviour, suggesting that such methods could serve as early detection tools for psychological distress. This is echoed by Kim et al. (2022), who applied the PHQ-9 framework alongside NLP to detect depression indicators from social media posts, affirming the validity of text-based sentiment analysis in health screening. Moreover, Babu and Kanaga (2022) provided a systematic review that illustrates how sentiment analysis techniques have evolved into practical tools for mental health surveillance. Their work emphasised the increasing reliance on AI-driven platforms to detect depression symptoms, particularly among young adults who frequently use digital communication channels. Integrating these insights into educational contexts such as universities, where student mental health remains a pressing concern, reinforces the relevance and timeliness of the current study.

Data visualisation translates complex analytical results into understandable graphical formats. According to Ware (2012), compelling visualisations support decision-making by highlighting patterns and outliers. Tools such as Tableau, Power BI, and Python libraries like Matplotlib, Seaborn, and Plotly enable the creation of interactive dashboards. In an academic context, visualisations can display trends in student sentiment over time, map geographical data of posts, or show network relationships among social media users. Research by Dong et. al (2024) shows how visual analytics helps university management understand student discourse and engagement on social platforms. Studies targeting student populations, particularly in Southeast Asian contexts, have highlighted the growing reliance on social media to express academic and personal concerns. In Malaysia, Norazah et al. (2020) conducted a sentiment analysis of university students' tweets. They found significant patterns related to stress during exam periods, satisfaction with online learning, and reactions to campus events. However, there is limited literature specifically focused on Universiti Utara Malaysia (UUM), indicating a research gap. Exploring sentiment among UUM students through social media data would provide localised insights that can inform campus policies and student support services.

Table 1

Summary of previous works

Authors	Data	Preprocessing Method	Text Analytics	Evaluation	Dashboard
Kim et al (2022)	NA	stop word, spell checks	BERT, Word2Vec	Accuracy	NA
Babu, & Kanaga, (2022)	Twitter	Tokenisation and stemming remove irrelevant texts.	Text Blob SentiWordNet	Accuracy, precision and recall	NA
Owen et al. (2024).	Twitter, Reddit, and Facebook	tokenisation, stemming, and removing irrelevant texts. LIWC	LDA, Support Vector Machine (SVM), BERT.	F1-score, Receiver Operating Characteristics	NA
Herdiansyah et al. (2022)	Twitter	Data cleaning	Rapid Miner	Accuracy and Precision	NA

METHODOLOGY

This study adopts a quantitative and computational research methodology involving social media data extraction, NLP, sentiment analysis, and data visualisation. The methodology is divided into several stages: data collection, preprocessing, sentiment and text analysis, and visualisation, as depicted in Figure 1. Each step is tailored to focus on data generated by Universiti Utara Malaysia (UUM) students on the social media platform.

Figure 1

Research flowchart
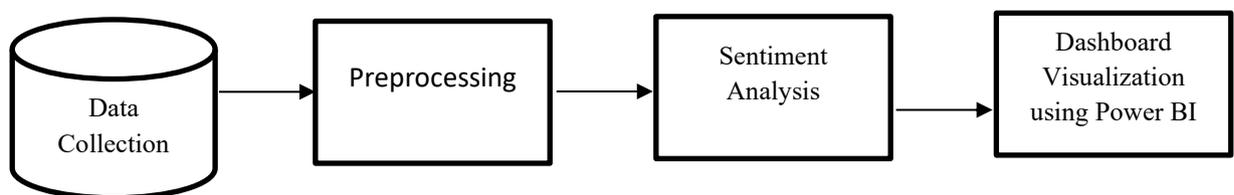
Data Collection

Confession data was collected from the UUM Confession Telegram channel using a custom-built Python script for data crawling. The crawler was developed to retrieve text-based posts while filtering out duplicate entries and non-textual data. In total, 3,172 confession entries were obtained between March and August 2024, covering a wide range of emotional and academic topics shared by students. In this study, the dataset was divided into a 70:30 ratio, with 70% allocated for training and 30% reserved for testing.

Preprocessing

The raw text data was run through several preprocessing steps to get it prepared for sentiment analysis. These included converting all text to lowercase, removing stop words, removing special characters and emojis, and normalising slang phrases and abbreviations commonly used in student confessions. As the dataset is bilingual (a mix of Malay and English), the language normalisation step was also included to maintain semantic meaning and prevent token mismatch. There was also manual cleaning of some entries to ensure a high-quality representation of the data.

Sentiment Analysis

To support thematic interpretation, sentiment analysis was conducted using the fine-tuned BERT model to classify student confessions into positive, negative, or neutral categories. In this study, the parameter setting of BERT followed the standard configuration as suggested in the original implementation (Devlin et al., 2019). This classification was visualised using frequency bar charts, sentiment distribution graphs, and sentiment-proportion pie charts to provide a more straightforward overview of emotional trends among students. These visualisations helped identify frequently discussed concerns such as academic stress, loneliness, financial pressure, and peer relationships, highlighting the emotional tone of discussions within the UUM Confession platform.

Dashboard Visualisation Using Power BI

In addition to static visualisations, an interactive dashboard was developed using Microsoft Power BI to enhance the exploration and presentation of sentiment analysis results. The dashboard displays dynamic sentiment trends over time, keyword frequencies, and student concern clusters categorised by topic. Filters and slicers allow stakeholders such as counsellors or researchers to drill down by sentiment category or specific keyword themes. This dashboard provides an intuitive and real-time interface for identifying emotional spikes and analysing shifts in student sentiment, supporting more informed mental health interventions on campus.

RESULT AND DISCUSSION

Data collected for this project is vital for subsequent analysis. Telegram is a platform widely used by UUM students to share personal thoughts, experiences, and emotions, often through anonymous confession channels. Telegram was chosen due to its popularity, accessibility, and the candid nature of the posts shared. A total of 5,243 posts were collected, of which 3,172 were considered relevant after filtering. Language analysis revealed that 58% of posts were in English and 42% in Bahasa Melayu, indicating a bilingual social media presence among UUM students. The dataset included contributions from 4,127 unique users, suggesting a broad and diverse representation of student voices. This variety enhances the reliability of the sentiment and text analysis by capturing a wide range of emotional expressions and perspectives.

Table 2

Data description

Metric	Value
Total Posts Crawled	5,243
Relevant Posts Analysed	3,172
Languages Detected	English (58%), Bahasa Melayu (42%)
Users	4127

For this study, the VADER sentiment analysis tool was used to analyse the emotional tone of 3,172 relevant Telegram posts submitted anonymously by UUM students. Given the informal, emotionally charged nature of the content, often related to personal stress, academic pressure, and mental health, VADER was selected for its effectiveness with social media and short text. In relative terms, sentiment classification thresholds are set as follows:

Sentiment Classification Thresholds (VADER)

Positive: Compound score ≥ 0.05

Neutral: Compound score > -0.05 and < 0.05

Negative: Compound score ≤ -0.05

The result of the VADER analysis shows that 44.9% of the comments are positive, 33.7% are negative, and 21.4% are neutral, as given in Table 3 and Figure 2. These findings have shown that almost half of the social media posts were positive, indicating generally optimistic attitudes among UUM students. In addition, negative posts peaked during exam weeks (June and July), often expressing stress, lack of sleep, or dissatisfaction with academic workload.
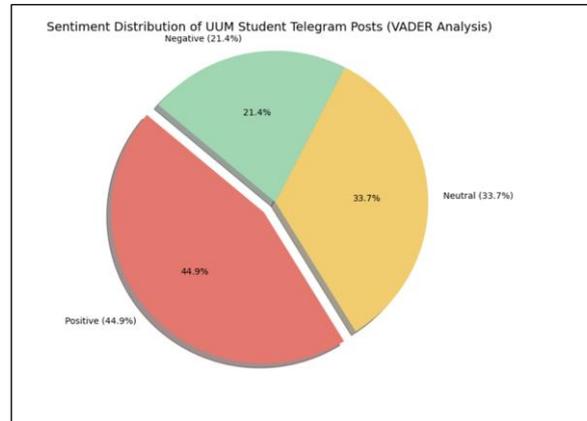
Table 3

Sentiment Distribution Results

Sentiment	Count	Percentage
Positive	1424	44.9%
Neutral	1069	33.7%
Negative	679	21.4%
Total	**3,172**	**100%**

Figure 2

Sentiment Distribution

To assess the performance of our sentiment classification model, we evaluated it using standard classification metrics: Precision, Recall, and F1 Score, computed for each sentiment class — Negative, Neutral, and Positive. Table 4 summarises the results along with the corresponding support (number of actual instances) for each class. The model demonstrates the strongest performance on the Negative class, with an F1 Score of 0.87, reflecting both high precision (0.85) and recall (0.89). This suggests that the model is highly effective in correctly identifying negative sentiments and making few false-positive predictions in this category. This strong performance may be partly attributed to the larger number of negative samples in the dataset (support = 1,600), which can help the model learn more robust patterns for this class.

In contrast, the Neutral and Positive classes exhibit comparatively lower performance. The Neutral class achieves a precision of 0.78 and a recall of 0.72, resulting in an F1 Score of 0.75. Similarly, the Positive class has the lowest scores across all three metrics, with an F1 Score of 0.71, which may be due to both its smaller representation in the dataset (support = 500) and the higher semantic ambiguity associated with positive expressions in text.

The macro-averaged scores — which treat all classes equally regardless of their support — are 0.79 (precision), 0.77 (recall), and 0.78 (F1 Score), indicating moderately strong and balanced overall performance. However, the disparities between class-wise metrics suggest that the model is biased toward the majority class (Negative), which is a common issue in imbalanced classification settings.

Table 4

Classification evaluation metrics

Class	Precision	Recall	F1 Score	Support
Negative	0.85	0.89	0.87	1,600
Neutral	0.78	0.72	0.75	700
Positive	0.74	0.69	0.71	500
Average/Macro	0.79	0.77	0.78	2,800

Dashboard Visualisation Interface

Interactive visualisations of these elements were done through the Power BI dashboard. Some of the key charts included the sentiment label distribution, which provided sentiment scores as a 5-star rating system, and the sentiment class breakdown, which represented how often each sentiment type was expressed. Visualisation of the emotion categories also revealed that among the most common emotions expressed were sadness, fear, and anger. A keyword frequency indicated that emotionally damaging words like "stress," "mental health," "depression," and "anxiety" were the most frequent, representing over 75% of the total keyword frequency. This affirmed the existence of emotionally negative material in the confessions.

The dashboard included interactive filtering by sentiment and keywords, providing real-time insights into students' emotional states and trends over time. Such analysis provides valuable resources for counsellors and university decision-makers to identify patterns and prioritise mental health support more effectively. Figures 3 to 8 show some of the dashboard visualisation results.

Figure 3
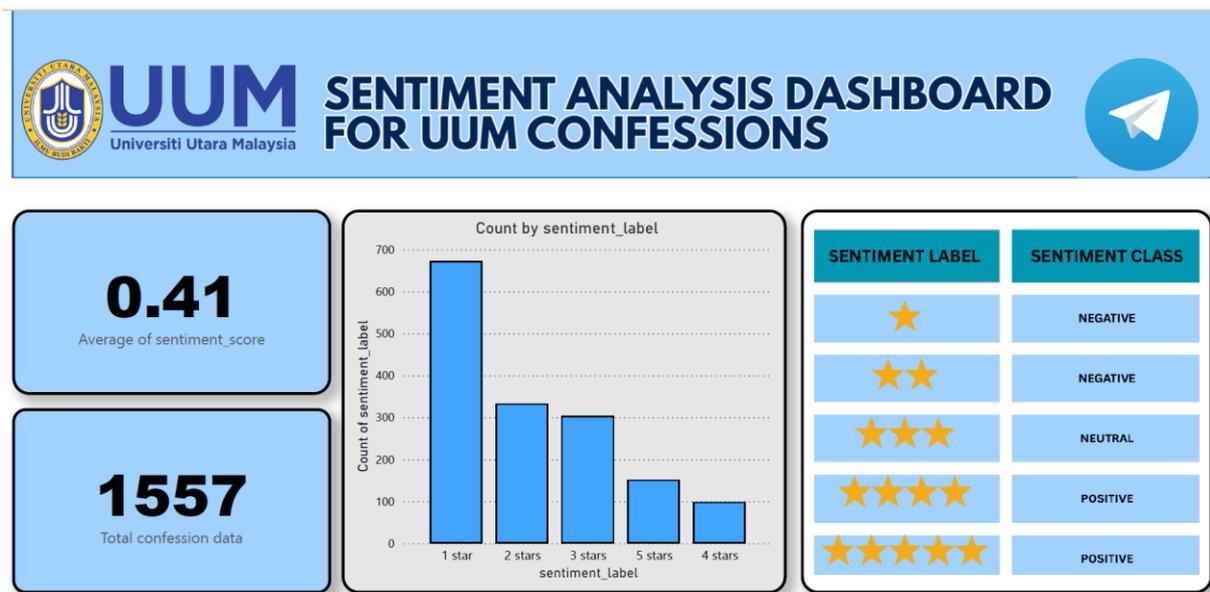
Main Page of Sentiment Analysis Dashboard for UUM Confessions

Figure 4

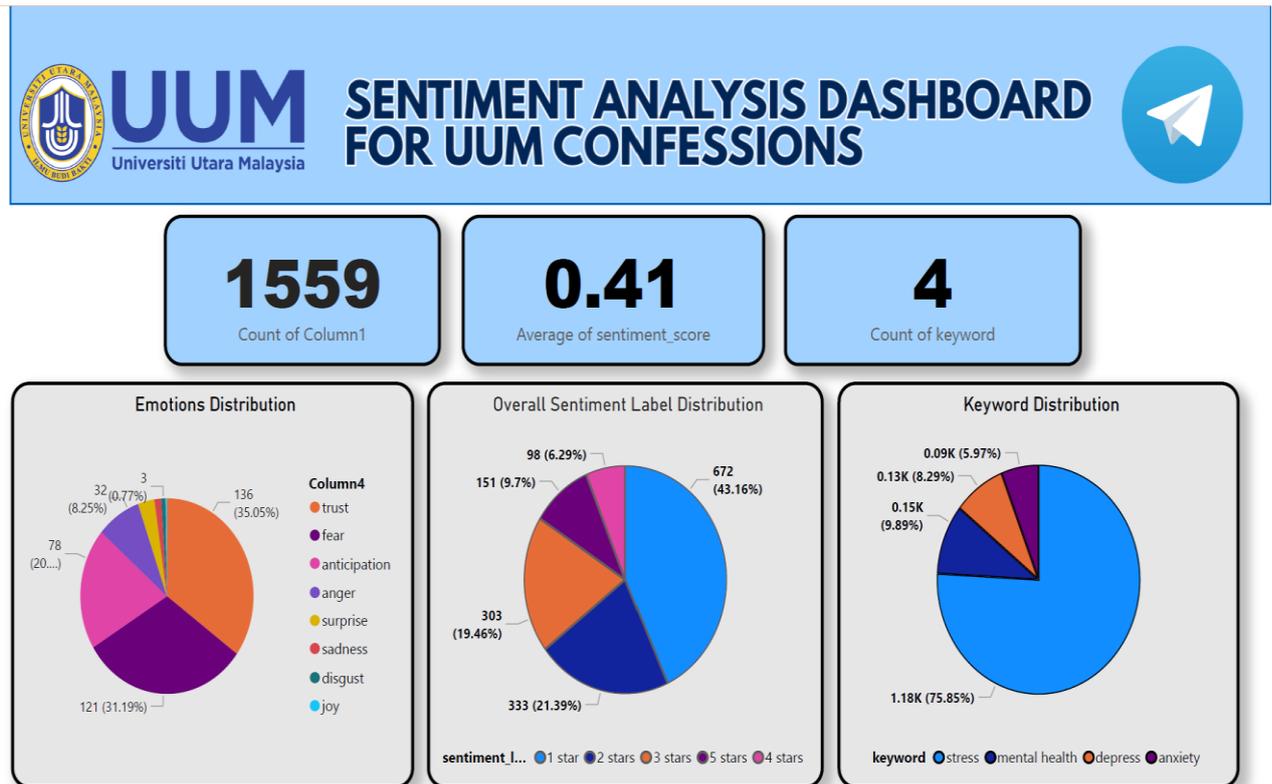*Keywords distribution dashboard for UUM Confessions*

Figure 5

Sentiment analysis dashboard for UUM Confessions

Figure 6

Positive sentiment word cloud

Figure 7

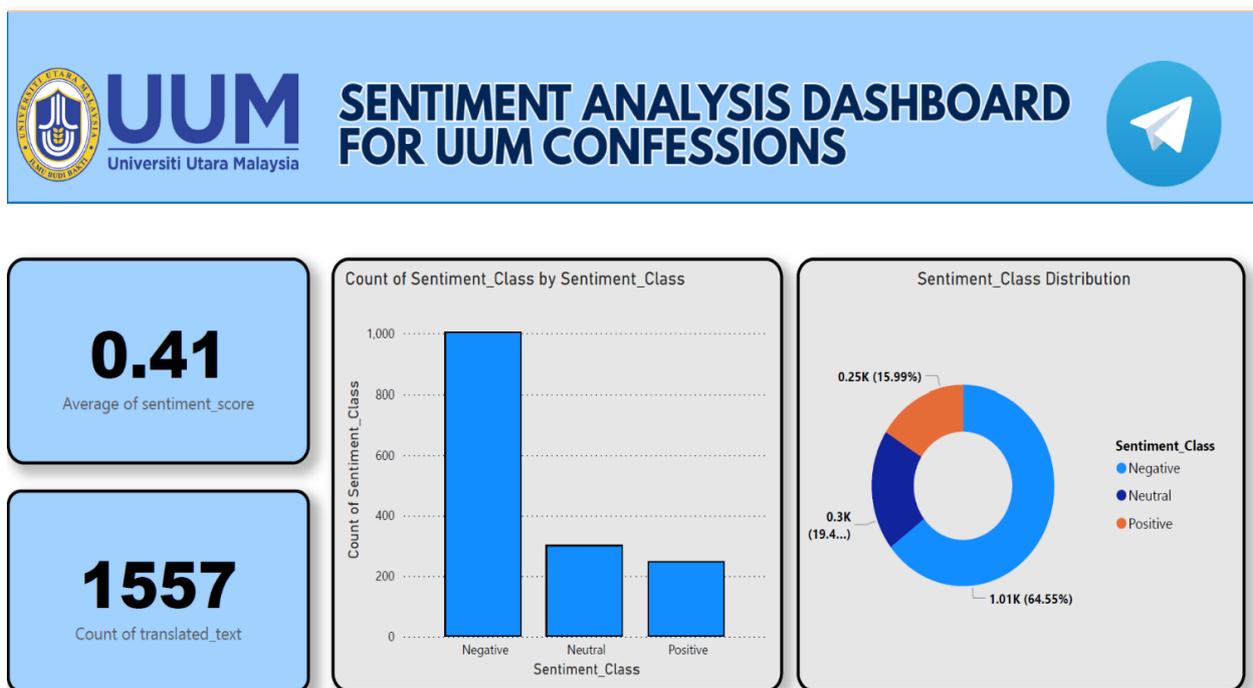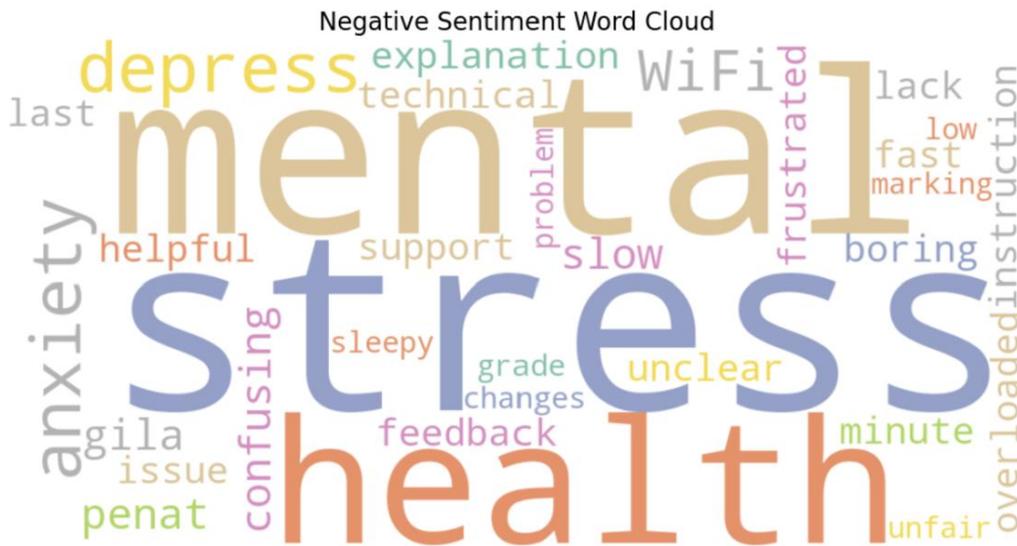*Neutral sentiment dashboard for UUM Confessions*

Figure 8

Negative sentiment dashboard for UUM Confessions

Evaluation Strategy

The evaluation of the dashboard used a user-centred strategy with two major approaches: usability testing and a systematic questionnaire. Thirty UUM students participated, with the majority aged 18 to 24 and members of the UUM Confession Telegram group. This ensured feedback was collected from users who were already familiar with the platform, its tone, language usage, and popular themes present in confessions.

Usability Testing

Usability testing is a method used to evaluate how easily and effectively users can interact with a product, system, or interface. It is a core component of user-centred design. It is typically conducted during various stages of product development to ensure that the system meets the needs and expectations of its target users. In this project, the usability testing is used to examine the sentiment dashboard for UUM confession.

Participants were shown a walkthrough of the dashboard features, including sentiment score plots, emotion pie charts, and keyword visualisations. They were asked to perform tasks such as filtering by sentiment category, exploring keyword clouds, and viewing emotion trend timelines. Observers observed their usability, usability issues encountered, and interactions. Most participants reported finding the dashboard user-friendly and praised its logical structure and minimalistic graphics. Feedback also indicated that visual narration through graphs helped individuals better understand general emotional patterns.

Evaluation of the Questionnaire

Besides observational findings, a five-item Likert-scale survey was shared via Google Forms. The survey addressed five central topics as given in Table 5:

Table 5

Evaluation dimension of usability testing

No.	Evaluation Dimension	Question	Mean Score
1	Helpfulness	The dashboard helps me understand the overall emotional trends of UUM students.	8.67
2	User- Friendliness	The dashboard is easy to navigate and use without technical guidance.	8.33
3	Information Presentation	The visualisations (charts, graphs, and keyword cloud) are clear and informative.	7.21
4	Accuracy	The sentiment results shown on the dashboard match my interpretation of the confessions.	8.12
5	Overall Satisfaction	Overall, I am satisfied with the features and performance of the dashboard.	9.28

For accuracy, 73.3% of users said the sentiment analysis always matched the tone of the confessions, and 26.7% said it sometimes did, illustrating that the BERT-based model could capture fine-grained emotional tones. In addition, 80% of users found keyword categorisation to be useful, and 20% found it highly useful, illustrating that BERT was effective at extracting salient emotional themes, including stress, anxiety, and depression. Regarding BERT visualisation informativeness, 79% rated it as very informative and 21% as informative, confirming that organised keyword frequency plots helped users identify main subjects. Similarly, 86.7% of the users rated the word cloud as very useful, while 14.7% found it helpful for adequate comprehension of dominant vocabulary and emotional triggers. The visualisation of the dashboard itself was highly favoured, with 76.7% of the participants finding the visualisations very useful and 13.3% finding them helpful in rendering sentiment data more comprehensible. Finally, the dashboard interface was praised for being easy to use, with 90% finding it extremely easy and 10% relatively easy. This highlights the effectiveness of its navigation design, labelled charts, and visual cues in supporting both novice and experienced users. Together, the results point out the dashboard's potential in delivering accurate, accessible, and actionable data on UUM student mental health.

Respondents also left some qualitative comments, proposing that the dashboard be made available to student counsellors and mental health units in UUM for early intervention. Others even proposed future expansions, such as including timeline filtering, drill-down on specific confession examples, and incorporating emotion prediction in conjunction with sentiment classification. The assessment results

indicate that the dashboard was usable, acceptable, and potentially impactful for monitoring mental health trends within an academic setting. They also indicate that end-users can derive valuable insights from unprocessed machine learning outputs with minimal training, thereby enhancing the dashboard's utility in real-world decision-making contexts. These tests confirm the system's competency in delivering detailed sentiment analysis reports through a user-friendly and accessible dashboard.

CONCLUSION

This work succeeded in harnessing a BERT-based sentiment analysis model upon UUM Confession data to yield significant insights into student mental health. Interoperability with student-authored social media posts is a doorway to a stimulating research frontier of pre-emptive emotional monitoring of students. Emotional trends were realised using Power BI dashboards, making it easy for stakeholders to identify, interpret, and act upon seminal psychological challenges confronting students.

Moreover, this research demonstrated the potential of interactive visual analytics to convert NLP technical outcomes into readable visualisations, helping stakeholders understand and engage better. The availability of performance metrics and user-focused testing ensured the system's usability and practicality, especially for institutions seeking low-barrier systems to enhance mental health awareness.

The results reinforce the importance of prospective mental health monitoring and demonstrate the potential of AI tools to enhance institutional acknowledgement and planning for intervention. This study mainly focused on exploring BERT on a text dataset to identify student emotion. In future works, other deep learning models will be used to test the effectiveness of the proposed method. Moreover, future research can explore more nuanced emotion classification beyond basic sentiment, expand multilingual capabilities, incorporate chatbot feedback loops, and collaborate with university counselling personnel to provide real-time notifications and feedback on emerging emotional patterns.

ACKNOWLEDGEMENT

REFERENCES

Abbasi, A., Zhou, Y., Deng, S., & Zhang, P. (2018). Text analytics to support sense-making in social media: A language-action perspective. *MIS Quarterly: Management Information* https://doi.org/10.25300/MISQ/2018/13239

Abdurrahim, A., & Fudholi, D. H. F. (2024). Mental Health Prediction Model on Social Media Data Using CNN-BiLSTM. *Kinetik: Game Technology, Information Systems, Computer Networks, Computing, Electronics, and Control*, *9*(1) https://doi.org/10.22219/kinetik.v9i1.1849

Altrabsheh, M. Cocea, and S. Fallahkhair. Predicting students' emotions using machine learning techniques. In the *17th International Conference on Artificial Intelligence in Education*, 2015.

Babu, N. V., & Kanaga, E. G. M. (2022). Sentiment Analysis in Social Media Data for Depression

Detection Using Artificial Intelligence: A Review. *SN Computer Science, 3* (1). https://doi.org/10.1007/s42979-021 00958-1

Balci, Ş., & Sarac, H. (2023). University students' perceptions of the effects of social media on anxiety: A qualitative study. *JMIR Formative Research, 7*(1), e43037. https://doi.org/10.2196/43037

Barrot, J. S., Llenares, I. I., & del Rosario, L. S. (2024). Editorial: Mining social media for mental health insights. *Frontiers in Research Metrics and Analytics, 9*, 1388691. https://doi.org/10.3389/frma.2024.1388691

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. *Digital Medicine, 3*(1). https://doi.org/10.1038/s41746-020-0233-7

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 4171–4186.

Dong, L., Du, H., & Dong, L. (2024). The research on Text Mining and Visualisation Analysis of ideological and political course texts in colleges based on natural language processing. *ICAIE '24: Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Education* (pp. 364–367). https://doi.org/10.1145/3722237.3722300

Herdiansyah, H., Roestam, R., Kuhon, R., & Santoso, A. S. (2022). Their post tells the truth: Detecting social media users' mental health issues with sentiment analysis. *Procedia Computer* https://doi.org/10.1016/j.procs.2022.12.185

Howe, S. (2024, May 6). Social Media statistics for Malaysia. Meltwater. https://www.meltwater.com/en/blog/social-media-statistics-malaysia

Jedeejit, P., Nuankaew, W. S., Nasa-ngium, P., & Nuankaew, P. (2025). Mining social media data for mental health: A text-mining approach among university students. *International Journal of Engineering Trends and Technology, 73*(4), 135–142.

Kaur, H., Gupta, R., & Sharma, N. (2025). Social media text as a predictor of depression: A systematic review and meta-analysis. *Journal of Medical Internet Research, 27*, e59002.

Kim, N. H., Kim, J. M., Park, D. M., Ji, S. R., & Kim, J. W. (2022). Analysis of depression in social media texts through the Patient Health Questionnaire-9 and natural language processing. https://doi.org/10.1177/20552076221114204.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Mohammad, S., Kiritchenko, S., & Zhu, X. (2021). Analysing Student Feedback Using NLP Techniques. *Computers & Education.*

Norazah, M. S., et al. (2020). Social Media and Sentiment of Malaysian University Students. *Journal of ICT in Education*.

Owen, D., Lynham, A. J., Smart, S. E., Pardinas, A. F. & Collades, C. J. (2024). Artificial intelligence for analysing mental health disorders in social media: a quarter-century narrative review of progress and challenges. *Journal of Medical Internet Research*, *26* (e59225). https://doi.org/10.2196/59225

Singh, R., & Singh, A. (2020). *Social Media Data Mining and Analytics*. Springer.

Ware, C. (2012). *Information Visualisation: Perception for Design*. Morgan Kaufmann.

Zhang, X., Wang, C., & Yu, Y. (2022). Deep Learning Models for Sentiment Analysis of Tweets. *IEEE Access*, 16(12), 432–448.