Assessing the emotional feedback of teaching and learning service beneficiaries using machine learning on text comments

Wipawan Buathong and Pita Jarupunphol*

Department of Digital Technology,
Phuket Rajabhat University,
Phuket, Thailand
Email: w.buathong@pkru.ac.th
Email: p.jarupunphol@pkru.ac.th
*Corresponding author

Abstract: This study evaluates emotional responses to educational services using advanced machine-learning techniques to categorise sentiment in feedback. The dataset includes 1,033 comments from 402 individuals, collected via various platforms. Three algorithms were applied: random forest, Naïve Bayes, and long short-term memory (LSTM). The ten-folds cross-validation method ensured model robustness. Random forest achieved the highest F1-score of 0.833, LSTM at 0.827, and Naïve Bayes at 0.807. The analysis indicated that neutral sentiments were most accurately predicted, followed by positive and negative sentiments. Additionally, latent Dirichlet allocation (LDA) identified key themes within the feedback. Positive topics included teaching effectiveness, subject variety, and professional development. Negative topics highlighted issues with technology and resources. Word cloud dashboards focused on curriculum design, learning support mechanisms, and instructional quality. These insights are crucial for enhancing the effectiveness of teaching services, indicating areas of strength and potential improvement.

Keywords: sentiment analysis; machine learning; educational services; natural language processing; data visualisation.

Reference to this paper should be made as follows: Buathong, W. and Jarupunphol, P. (2025) ‘Assessing the emotional feedback of teaching and learning service beneficiaries using machine learning on text comments’, *Int. J. Innovation and Learning*, Vol. 38, No. 1, pp.22–49.

Biographical notes: Wipawan Buathong obtained her PhD in Information Technology from King Mongkut’s University of Technology North Bangkok, Thailand. Her research focuses on the application of data science, data mining, and machine learning across various fields. She is currently a Lecturer in the Department of Digital Technology at Phuket Rajabhat University, Thailand.

Pita Jarupunphol earned his PhD in Computer Science from the University of Auckland, New Zealand. His research areas encompass data mining, machine learning, cybersecurity, human-computer interactions, and the acceptance of technology by users. Currently, he serves as a Lecturer in the Department of Digital Technology at Phuket Rajabhat University, Thailand.

1 Introduction

Comments have evolved into a critical medium for communication and expression, teeming with emotions across various social media platforms. Technology is crucial in analysing customer and service recipient sentiment data in business, spanning the hospitality, dining, and healthcare sectors. This data is transformed into actionable insights that reflect the emotional and psychological impact on the respective businesses. Educational services are essential in education as they cater to individuals from childhood through adulthood, drawing keen public interest. Currently, a wealth of feedback, including criticisms and candid opinions, is shared online. However, some service-related issues remain unaddressed, potentially due to the slow recognition of problems. Traditional data collection methods, such as paper or digital questionnaires, can delay the accumulation of actionable data, hindering timely problem resolution and feedback presentation to decision makers.

Currently, sentiment analysis using machine learning is a well-explored field. Numerous studies have applied various algorithms to analyse feedback in different contexts, including education. Several notable works have focused on using natural language processing (NLP) and machine learning to extract and interpret sentiments from textual data. For example, sentiment analysis has been extensively used to understand customer feedback in the hospitality, healthcare, and retail sectors. Furthermore, previous research has explored sentiment analysis to evaluate teaching effectiveness, student engagement, and course content quality. Techniques like support vector machines, Naïve Bayes, and deep learning models have been employed. Despite the extensive application of sentiment analysis in various fields, there are specific gaps that this research aims to address. For example, while traditional models like Naïve Bayes and support vector machines have been used, applying more advanced models like long short-term memory (LSTM) networks, designed to handle sequential data effectively, in educational sentiment analysis is less explored. This capability is particularly useful for sentiment analysis, where context and sequence can significantly influence the interpretation of sentiments. In addition, a comprehensive approach integrating feedback from various platforms (e.g., social media, websites, questionnaires) is needed to provide a holistic view of student sentiments.

This research, of significant value, aims to assess the emotional responses received from users of teaching services. It also aims to develop sophisticated machine-learning models, a powerful tool capable of accurately classifying feedback sentiment. This addresses the problem of slow and inefficient recognition of service-related issues in educational settings due to traditional data collection and analysis methods. In addition, timely recognition and resolution of issues are crucial for maintaining and improving the quality of teaching and learning services. Quick and accurate sentiment analysis, a key benefit of machine learning, enables educational administrators to make informed decisions based on real-time feedback, leading to improved educational services. By identifying and categorising concerns swiftly, institutions can enhance the quality of educational services, ultimately leading to better student satisfaction and learning outcomes. Using machine learning models to automate sentiment analysis reduces the need for manual processing, saving time and resources. Moreover, machine learning techniques, a significant advantage, provide a more detailed and nuanced understanding of student sentiments, allowing for targeted interventions and improvements. This research utilises feedback data from students of the Digital Technology program at the Faculty of Science and Technology, Phuket Rajabhat University, including undergraduates, postgraduates, and alumni. Their feedback from social media, websites, and questionnaires will be used to develop a machine-learning model for sentiment classification and text analysis. In this case, the objectives of this research are below:

- to thoroughly analyse the emotional feedback received from users of teaching services and present the findings through word cloud dashboards
- to develop sophisticated machine learning models capable of accurately classifying the sentiment of the feedback
- to compare selected algorithms' performance (Naïve Bayes, LSTM, and random forest) against latent dirichlet allocation (LDA) to determine its effectiveness and suitability for educational sentiment analysis.

2 Literature review

Sentiment analysis, or opinion mining, is a field within natural language processing (NLP) that involves identifying, extracting, and categorising the sentiment expressed in textual data. The primary goal is to determine the attitude, opinion, or emotional tone behind a series of words to understand the sentiments conveyed by the text. Sentiment analysis is widely used across various domains, including business (customer feedback, product reviews), politics (public opinion analysis), healthcare (patient feedback), and education (student feedback). Sentiments are typically categorised into three main types: positive, negative, and neutral. Some advanced applications also consider finer-grained classifications (e.g., very positive, positive, neutral, negative, very negative). Techniques used in sentiment analysis range from simple rule-based approaches to advanced machine learning and deep learning models. Commonly used methods include lexicon-based approaches, traditional machine learning classifiers, and modern deep learning models.

2.1 Research on sentiment analysis

Sentiment analysis has long been critical in understanding consumer perceptions, particularly in business reviews, including extensive product evaluations across various sectors. The proliferation of social media platforms has significantly increased the volume of public opinion and emotional expressions, offering rich insights into consumer sentiments. Several studies have showcased the application of advanced analytical techniques in this field. For instance, Sajeevan and Lakshmi (2019) employed deep learning methods, specifically LSTM and CNN algorithms, to categorise sentiments in movie review articles. This approach highlights the evolving use of complex computational models to interpret nuanced expressions of opinion.

Similarly, Twitter data was utilised to analyse healthcare service discussions among patients in different locations and times within the USA (Sewalk et al., 2018). They developed a machine learning model using the support vector machine algorithm, noted for its superior classification performance. This study underscores the potential of machine learning in capturing and understanding patient experiences and sentiments. Furthermore, Jianqiang et al. (2018) explored Twitter data sentiment classification using a convolutional deep learning neural network. Their work contributes to the growing evidence of deep learning's effectiveness in processing and analysing vast and varied social media data. In addition, Hassan et al. (2020) demonstrated that regression machine learning models could enhance the R-square value by integrating new comments into a scientific domain dictionary. This approach suggests a method for continuously refining analytical models for better accuracy.

Haruechaiyasak et al. (2018) introduced S-Sense, an analysis framework for gauging the intensity of sentiments and intentions within social media messages. Their method bolsters marketing strategies and sales presentations through detailed sentiment classification. Moreover, Wicaksono and Mariyah (2019) delved into consumer sentiment analysis on Facebook comments, evaluating and contrasting the efficacy of dictionary-based approaches versus machine learning and sentiment analysis techniques. This comparison highlights the diverse methodologies available for sentiment analysis and their respective strengths in extracting meaningful insights from social media data. Primarily, Shah et al. (2018) contributed to the body of research by assessing sentiment analysis applications, further solidifying the importance of this analytical technique in understanding and leveraging consumer feedback across various digital platforms.

2.2 Advancements in sentiment analysis

Recent studies in sentiment analysis continue to explore the nuanced impact of customer-generated content, such as online reviews, on various industries (Lai et al., 2023). For example, a notable study investigated how online reviews could predict restaurant survival by categorising sentiments into five aspects: location, deliciousness, price, service, and atmosphere. This research demonstrated that aspect-based sentiment analysis, leveraging machine learning conditional forest models, significantly enhances the accuracy of survival predictions for restaurants (Li et al., 2023). Another study focused on customer satisfaction with Traveloka's services, analysing satisfaction levels among users based on data extracted from Twitter. Utilising algorithms such as the support vector machine, logistic regression, and Naïve Bayes for sentiment classification, the findings revealed the support vector machine as the most precise in distinguishing customer sentiments, particularly highlighting positive feedback related to promotions, campaigns, and discounts (Diekson et al., 2023). These studies underscore the critical role of sentiment analysis in interpreting the emotional content of messages, prompting further research employing various algorithms to refine the accuracy of sentiment evaluations in different contexts. For example, sentiment analysis entails evaluating the underlying messages in natural language, using a range of algorithms to enhance the accuracy of sentiment assessments, particularly within the context of education (Korkmaz, 2023; Fuller et al., 2023; Shaik et al., 2023). In terms of sentiment representation, word clouds have been widely utilised in various fields, including sentiment analysis, due to their effectiveness in visually representing the most frequently occurring words within a dataset. Word clouds provide an intuitive and immediate overview of key themes and sentiments expressed in textual data by displaying words in a size proportional to their frequency of occurrence. This visual representation allows researchers and practitioners to quickly identify prominent terms and patterns, making word clouds particularly useful for exploratory data analysis for highlighting significant insights and trends in large textual datasets (Heimerl et al., 2014; McNaught and Lam, 2010; Sinclair and Cardew-Hall, 2008). Table 1 summarises key studies related to using different algorithms for sentiment analysis, highlighting the algorithms used, the domain of application, and the key findings of each study.

Table 1 Summary of related studies on sentiment analysis algorithms

Study	*Algorithms used*	*Domain*	*Key findings*
Sajeevan and Lakshmi (2019)	LSTM, CNN	Movie reviews	Deep learning models effectively capture sentiment nuances.
Sewalk et al. (2018)	SVM	Healthcare	SVM performs well in classifying patient sentiments from Twitter data.
Jianqiang et al. (2018)	CNN	Twitter data	Convolutional neural networks excel in sentiment classification tasks.
Hassan et al. (2020)	Regression models	Scientific literature	Integrating new comments into domain-specific dictionaries improves sentiment prediction accuracy.
Haruechaiyasak et al. (2018)	Custom framework	Social media	S-Sense framework effectively gauges sentiment intensity for marketing insights.

This study will concentrate on text data from service users' opinions to analyse population sentiments, a crucial task in natural language processing (NLP) that supports business development (Shamshiri et al., 2024). The analysis will categorise sentiments into positive, negative, and neutral outcomes, employing machine learning models trained on accurate datasets for predictive accuracy. NLP is a pivotal technology in machine learning, enabling computers to interpret, manipulate, and understand human language, thus managing vast amounts of textual data. Today, organisations inundated with data from varied communication channels, including emails and social media, benefit from NLP's capacity to facilitate human-computer interaction in natural language. NLP applications range from reading and interpreting human speech to analysing underlying emotions and filtering key ideas for actionable use (Just, 2024). Despite the complexity and diversity of human language, NLP strives to organise unstructured information, overcoming challenges posed by language nuances, abbreviations, dialects, and accents to capture the essence of human communication and sentiment effectively.

2.3 Machine learning algorithms

In this research, we employed three distinct machine learning algorithms to analyse sentiments expressed in textual data related to educational services, including Naïve Bayes, long short-term memory, and random forest.

2.3.1 Naïve Bayes algorithm

The Naïve Bayes algorithm is a probabilistic classifier that utilises Bayes' theorem to classify data, under the simplifying assumption that all features are conditionally independent (Sano et al., 2023). Despite its simplicity, Naïve Bayes is remarkably effective for data classification, producing results on par with more complex algorithms like C4.5 (Dougherty et al., 1995). This efficiency stems from its basis in Bayes' theorem, simplified further by assuming data attributes operate independently. This approach is a straightforward probabilistic classifier efficient for large datasets and provides baseline performance for sentiment classification tasks. It has been widely used in sentiment analysis due to its robustness despite its feature independence assumption. The Naïve Bayes formula for calculating the probability of a class C given a feature vector $X = (x_1, x_2, \ldots, x_n)$:

$$P(C \mid X) = \frac{P(C) \prod_{i=1}^{n} P(x_i \mid C)}{P(X)} \tag{1}$$

where

- $P(C \mid X)$ is the posterior probability of class C given the feature vector X
- $P(C)$ is the prior probability of class C
- $P(x_i \mid C)$ is the likelihood of feature x_i given class C
- $P(X)$ is the evidence or the normalising constant.

2.3.2 Long short-term memory algorithm

LSTM networks, a subclass of recurrent neural networks (RNNs), are specially designed for processing sequential data (Choudhary and Chauhan, 2023). Unlike traditional neural networks, LSTMs have a unique architecture that allows them to remember and utilise previous computations for future processing. This capability makes them particularly suited for tasks involving sequences, such as processing a series of words in a sentence where each word's input $(x(t))$ and its corresponding output $(y(t))$, like its part of speech, are sequentially related. This sequence-to-sequence processing is a key strength of LSTM networks. Unlike traditional neural networks, LSTMs can remember previous inputs for long periods. This allows them to capture dependencies and relationships within the text that are crucial for accurate sentiment classification. Primarily, LSTM has demonstrated superior performance in various NLP tasks, including sentiment analysis, due to its ability to manage long-range dependencies and context in the data.

2.3.3 Random forest algorithm

Random forest is a versatile machine-learning algorithm that addresses classification and regression challenges. It is an ensemble learning method that combines multiple decision trees to improve classification accuracy and robustness (Sun et al., 2024). Each tree in the forest is trained on a distinct subset of the data, with considerations only for a randomly selected set of variables at each node. This diversity among trees enhances the model's decision-making capacity. For classification tasks, the final prediction is determined by a majority vote among the trees (Sullivan, 2017). Random forest is renowned for improving upon the decision tree method by assembling predictions from numerous trees, each informed by different data segments, thereby enhancing the prediction accuracy and reliability. It also provides insights into feature importance, which can be valuable in understanding which words or phrases most influence sentiment. Random forests aggregate the predictions of multiple decision trees. The prediction $\hat{y}$ for a regression problem can be written as:

$$\hat{y} = \frac{1}{M} \sum_{m=1}^{M} h_m(x) \tag{2}$$

Table 2 Comparison of machine learning algorithms

Algorithm	*Advantages*	*Disadvantages*	*Performance in study*
Naïve Bayes	Simple, efficient, good baseline	Assumes feature independence, less effective for complex patterns	Lowest F1 score (0.807)
LSTM	Handles sequential data, captures context, good for nuanced text	Computationally intensive, requires large datasets for training	Second highest F1 score (0.827)
Random forest	Robust, handles high dimensionality, provides feature importance	Can be slower to train, and may overfit with noisy data	Highest F1 score (0.833)

For a classification problem, the majority vote prediction $\hat{y}$ is:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_M(x)\} \tag{3}$$

where

- M is the number of trees in the forest
- $h_m(x)$ is the prediction of the m^{th} decision tree for input x.

These three methods comprehensively compare diverse machine learning algorithms. Naïve Bayes represents a traditional probabilistic approach, offering simplicity and efficiency in handling text data. LSTM exemplifies advanced deep learning techniques specifically designed for sequential data, capturing complex patterns and contextual dependencies in text. Random forest showcases ensemble methods known for their robustness and generalisability, leveraging multiple decision trees to enhance predictive performance. By comparing these methods, the study aims to elucidate the strengths and weaknesses of each approach within the context of sentiment analysis for educational feedback, thereby offering a well-rounded evaluation of different algorithmic strategies. Table 2 compares the advantages, disadvantages, and performance of different machine learning algorithms used in the study.

3 Methodology

The research process begins with data collection and sourcing information from a wide array of platforms to compile a dataset ripe for analysis. Following the principles of natural language processing, the collected data undergoes a selection phase to identify relevant phrases and sentiments. This dataset then serves as the foundation for developing a machine learning model, with comparative analysis across different algorithms to determine the most effective approach. The final step involves synthesising the data into dashboard reports designed to succinctly present the findings to decision-makers, facilitating ongoing improvements based on the insights garnered.

3.1 Data collection and extraction

The study concentrates on students from the Digital Technology Department within the Faculty of Science and Technology at Phuket Rajabhat University. It encompasses 119 undergraduate students, 28 master's students, and 195 alumni, totalling 402 participants. These participants contributed feedback through various platforms, including websites, social media, LINE groups, and traditional and online questionnaires. The data collection phase spanned from March 2023 to May 2023, during which feedback from the specified demographics was gathered and analysed. A total of 1,033 comments were collected for analysis. The data collection methodology involves scraping, extracting, or harvesting information from websites. This process was facilitated using the Selenium library, which enables automated navigation and data collection from websites through a web driver. With Selenium, it is possible to automate the launch of websites and systematically gather the required data. The extracted data was organised and stored in a structured table format, such as Excel, CSV, SQL, or plain text files, making it ready for further analysis. This tabular data, often referred to as a data frame, is arranged in rows and columns, providing a clear and accessible way to review and process the information for in-depth analysis.

3.2 Data preparation

The initial stage involves preparing the structured data for analysis or further use in machine learning model development (Kalita et al., 2024). This preparation consists of several key operations. For example, comments were translated from Thai to English using Google Translate to standardise the data for broader accessibility and analysis. Translating the comments to English allows for using a wide range of well-established NLP tools and libraries that may not be as developed for the Thai language. Furthermore, many advanced machine learning models, including those for sentiment analysis, are primarily designed and trained on English text. Using English ensures that the research can leverage these sophisticated models. In addition, publishing the research in English allows for broader dissemination and engagement with the international research community. Translating the comments into English facilitates easier comparison with other studies and datasets, enhancing the generalisability of the findings. This translation aligns with the methodology employed by Shukla et al. (2023), who utilised Google Translate's API to translate the Bhagavad Gita from Sanskrit to English. Their approach involved pre-processing the text by systematically organising verses, removing verse numbers, converting verses into single-line formats, and performing sentiment and semantic analysis.

Furthermore, removing irrelevant text elements such as symbols, emojis, and special characters, and retaining only the textual messages for analysis was also required. All uppercase letters are also converted to lowercase to ensure uniformity across the dataset. This data-cleaning process also involves tokenisation, which divides text into constituent parts, dividing sentences into words or phrases. In English, this is typically done using spaces as delimiters. After that, words are converted to their base or root form, stripping away prefixes and suffixes to simplify analysis. Primarily, semantically insignificant words are eliminated from the text to focus on the more meaningful words for modelling purposes.

3.3 Machine learning model development

This part employs the prepared data to construct a machine-learning model for text sentiment classification. The research involves developing software for NLP to prepare the data and using data mining techniques for data classification (Just, 2024). It utilises algorithms such as Naïve Bayes (Libnao et al., 2023), LSTM (Choudhary and Chauhan, 2023), and random forest (Sun et al., 2024) to process and provide feedback. Several criteria are applied to evaluate these models' performance, including accuracy, precision, recall, and the F-measure (F1 score) (Witten et al., 2016). These metrics comprehensively assess the model's effectiveness in accurately classifying sentiments expressed in the text data.

3.4 Sentiment classification

The processed data was utilised to categorise the sentiment of each comment. This categorisation involves distinguishing between negative sentiments (indicated by –1), positive sentiments (indicated by +1), and neutral or no opinion sentiments (indicated by 0). For responses derived from questionnaires, the sentiment is determined based on a satisfaction score: scores ranging from 1 to 2 are considered negative, 3 signifies neutrality or no specific opinion, and 4 to 5 are interpreted as positive sentiments.

3.5 Feature engineering

The data underwent feature engineering, a crucial step in making it computationally understandable for machine learning models. This process involves converting the data into a format the models can interpret and analyse effectively. The term frequency-inverse document frequency (TF-IDF) was used to calculate the TF-IDF score for each word, identifying the significance of words within the document relative to a collection of documents (Sjarif et al., 2019). This technique helps highlight important words and distinguish them for sentiment analysis. The data also went through word embedding, which involves generating vectors for words based on their contextual similarity to other words, even when they have different meanings. Creating a feature vector for each word token in the dataset transforms words into numerical values suitable for computational operations. This process includes one-hot encoding of each word, ensuring the data is in a format quickly processed by computers.

However, the study focused on single-word features (unigrams) for the following reasons. For example, single words provide a straightforward and interpretable feature set for initial analysis. Moreover, using single words allows for a baseline comparison with other studies that may also use unigrams (Gulati et al., 2022). However, the study acknowledges the benefits of using more complex features like n-grams, capturing phrases or sequences of words (e.g., bigrams, trigrams) that can provide more context and improve the accuracy of sentiment analysis by considering word combinations. In addition, useful techniques like part-of-speech tagging, syntactic parsing, and named entity recognition could further enhance the feature set. These feature extraction techniques are critical in transforming textual data into a structured form that machine learning models can effectively interpret and analyse, enabling more accurate sentiment classification.

3.6 Sentiment analysis

Sentiment analysis is conducted through algorithmic modelling, which uses computer algorithms to identify, extract, and quantify subjective information from textual data. This technique is part of the broader field of NLP and data analytics, focusing on understanding the sentiments, opinions, emotions, and attitudes expressed in written language. This step enables the automated, large-scale analysis of sentiments, which would be impractical to perform manually, providing a powerful tool for extracting meaningful insights from vast amounts of textual data. In this research, however, the sentiment analysis involves preparing data through NLP and applying data mining techniques for classification using algorithms like Naïve Bayes, LSTM, and random forest. The NLP rules involved in the preprocessing steps included standard tokenisation, which used space and punctuation delimiters to separate words. A custom stop word list, tailored to the context of educational feedback, included both Thai and English stop words. Normalisation techniques, such as lemmatisation or stemming, were used to convert all words to their base forms, reducing variations.

In this section, keywords related to the study's main focus areas were anticipated, specifically targeting three primary domains. For teacher sentiment, expected keywords included positive terms such as teacher-friendly, understanding, teaching, good, and friendly, alongside possibly negative terms like workload and language. In the domain of curriculum sentiment, anticipated keywords comprised positive terms such as variety of subjects, programming, technology, and work, as well as negative terms like difficulty and relevance. Regarding supportive facilities sentiment, the expected mix of positive and negative keywords included modern internet, computer problem, slow computer, old device, internet, insufficient, and libraries. These anticipated keywords reflect the core themes and potential areas of concern addressed in the study.

3.7 Evaluation of machine learning model performance

This study's evaluation of machine learning models is based on the average outcomes from ten iterative experiments. To ensure the reliability of our model's performance, we utilised the 10-folds cross-validation technique (Malakouti et al., 2023). This method involves dividing the dataset into K equally sized subsets, using K-1 subsets to train the model and the remaining subset to test its accuracy. This process is repeated until each subset has been used for testing, allowing for aggregating and averaging the accuracy scores from each iteration. Performance metrics such as accuracy, precision, recall, and F-measure are calculated to comprehensively reflect the model's effectiveness.

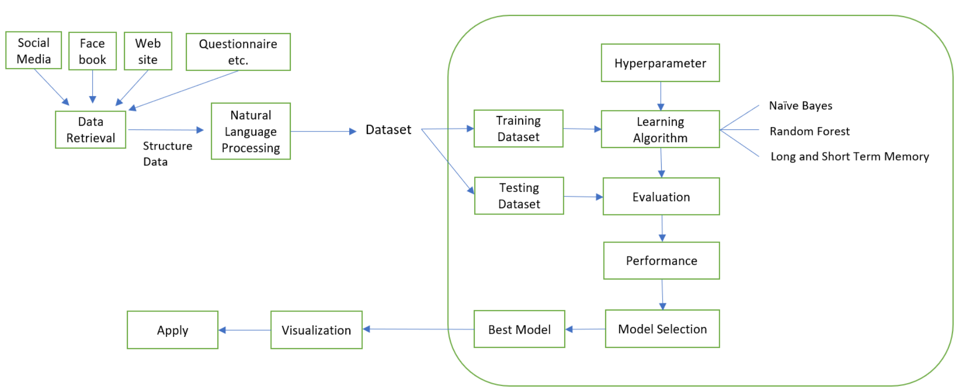
3.8 Data visualisation

The visual presentation of data plays a crucial role in exploratory data analysis (EDA), enabling data professionals to delve into, scrutinise, and familiarise themselves with datasets and the interrelations among various variables (Sharma et al., 2023). EDA leverages numerous tools and techniques, emphasising data visualisation to dissect and understand the data comprehensively. Such analysis aids in identifying the types of questions that can be addressed with the data at hand. For this research, Power BI, a widely adopted tool for data visualisation, was employed to present and analyse the data visually. This approach not only aids in understanding the data characteristics but also in conveying findings in an accessible and interpretable format.

3.9 Applying the model

The application of sophisticated machine learning models enables the classification of sentiment from textual data and the analysis of word usage within sentences. This process facilitates the identification of key concern dimensions through automated data collection and swift data analysis. Such insights prompt recognition of emerging issues, allowing for the efficient communication of results to management via word cloud dashboards. This streamlined approach to data interpretation not only enhances problem awareness but also supports continuous service improvement initiatives. Figure 1 represents the methodological steps used in the research.

Figure 1 Research methodology (see online version for colours)



4 Results

The performance of various machine learning algorithms, random forest, Naïve Bayes, and LSTM, was evaluated using accuracy, precision, recall, and F-measure metrics. This evaluation helped identify the most effective algorithm for sentiment analysis in the context of educational services. The study found that the random forest algorithm outperformed LSTM and Naïve Bayes, achieving the highest F1 score. This result is consistent with the findings of Sun et al. (2024), who highlighted the robustness and accuracy of random forest in classification tasks. Moreover, the superior performance of LSTM compared to Naïve Bayes aligns with the results from Sajeevan and Lakshmi (2019), who demonstrated the effectiveness of deep learning models, such as LSTM, in capturing complex patterns in text data. In addition, this research provides a comprehensive understanding of the sentiments expressed by educational service recipients, particularly focusing on three key areas: teachers, curriculum, and supportive facilities. By employing machine learning algorithms, the study successfully classified and analysed the sentiments contained within textual feedback, offering valuable insights into the perceptions and experiences of students.

4.1 Model performance evaluation

In evaluating the performance of the models employing three distinct algorithms, including Naïve Bayes, LSTM, and random forest, it was determined that the accuracy of text classification or prediction varied among them. Based on data derived from the confusion matrix and statistical metrics utilised for assessing model performance, the algorithms ranked from highest to lowest in terms of classification accuracy (CA), F-measure score (F1), precision (Prec), and recall were random forest, LSTM, and Naïve Bayes, respectively. Table 3 presents the sentiment analysis results conducted using various algorithms, detailing their performance across different statistical metrics such as accuracy, precision, F-measure, and recall. These metrics comprehensively evaluate each algorithm's effectiveness in classifying sentiments in the text data. Table 3 presents the performance results of each algorithm used in this study, evaluated across various statistical metrics such as classification accuracy (CA), precision, F-measure (F1), and recall.

Table 3 Results of each algorithm with various statistical values

Model	*CA*	*F1*	*Prec*	*Recall*
Random forest	0.834	0.833	0.832	0.834
Naïve Bayes	0.819	0.807	0.830	0.819
LSTM	0.834	0.827	0.836	0.834

Upon analysing all statistical metrics to evaluate the overall efficacy of the models, it was observed that models employing the random forest algorithm exhibited performance comparable to those using the LSTM algorithm. However, the F1 score is chosen as the primary metric for this research to represent the comprehensive performance evaluation across models utilising different algorithms. The random forest algorithm emerged as the most efficient among the three algorithms compared, achieving the highest F1 score of 0.833. This indicates superior model performance when utilising the random forest approach. Close behind, the LSTM algorithm recorded an F1 score of 0.827, indicating nearly comparable efficiency. Meanwhile, the Naïve Bayes algorithm demonstrated the lowest efficiency, with an F1 score of 0.807.

Figure 2 The confusion matrix values (see online version for colours)

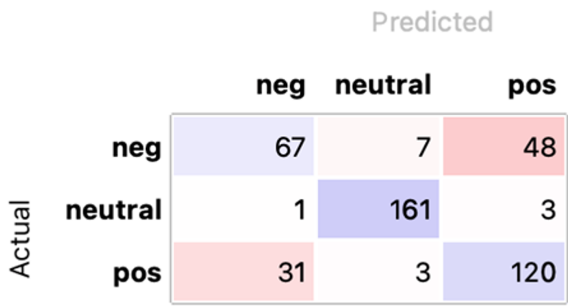


Figure 2 presents the confusion matrix for the testing dataset, which consists of messages that have undergone data preparation and cleansing of problematic inputs. This matrix specifically pertains to the random forest algorithm, identified as the most effective model in this research. Analysis of the matrix revealed that neutral sentiment data was predicted with the highest accuracy, followed by positive sentiment data. Conversely, negative sentiment data was the least accurately predicted category.

4.2 Data analysis results

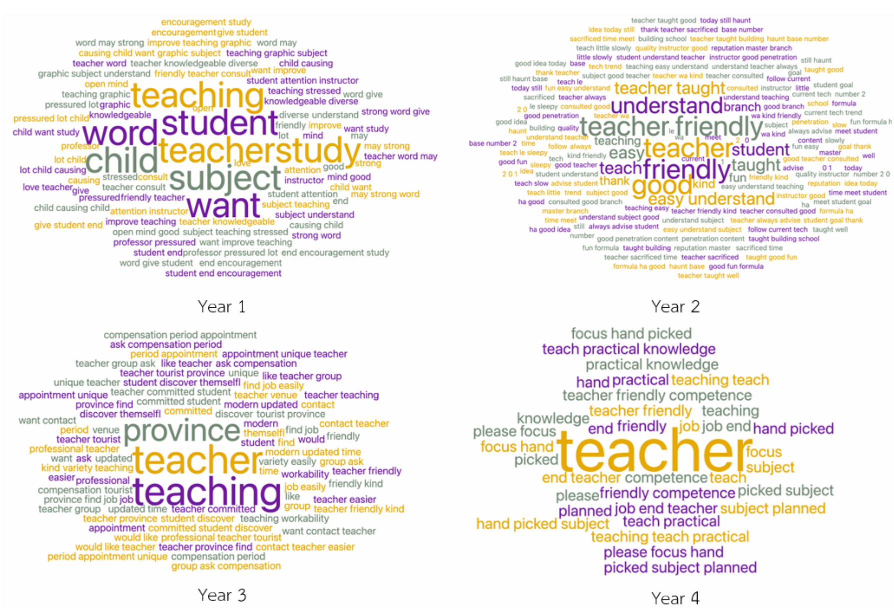
Implementing machine learning methodologies for predicting the sentiments of educational service users facilitates the creation of word cloud dashboards using Power BI. These dashboards effectively display the results of data analysis, offering deep insights into various aspects of educational services. This approach enhances problem recognition and enables clear communication of findings to management through user-friendly dashboards. The goal is to leverage these insights for continuous improvement in service provision. The analysis and reporting are segmented into three main categories: lecturer performance, curriculum quality, and learning support infrastructure.

4.2.1 Teacher sentiment

Sentiment analysis regarding instructors reveals a mixed but predominantly positive outlook. Teachers are praised for their diverse teaching methods, approachability, and effective communication skills, which simplify complex concepts for students. This positive feedback underscores a supportive teaching environment where educators are seen as friendly and engaging. On the flip side, there are calls for a reduction in the assignment workload and an appeal for language that resonates more closely with the student body, indicating areas for potential improvement in teaching practices. The sentiment analysis prominently features words like teacher-friendly, understanding, teaching, good, and friendly, which collectively paint a positive picture of the teaching quality. These terms suggest that teachers are well-regarded for their approachability, effective teaching methods, and ability to make concepts understandable. The frequent occurrence of such words indicates a strong appreciation for educators who are knowledgeable, genuinely supportive, and engaging in their interactions with students. Figure 3 provides a comprehensive visualisation of the sentiment analysis concerning teachers, encapsulating the range of emotions and opinions students hold towards their instructors.

Figure 3 Sentimental analysis on teacher (see online version for colours)

Figure 4 Sentimental analysis on teacher, separated by academic year (see online version for colours)



Our analysis of sentiments by academic year, from first to fourth year, reveals a fascinating evolution in student perceptions. Positive descriptors such as teacher-friendly, understand, teaching, good, and friendly consistently emerge, though their distribution varies across different years. Notably, the term 'teacher' becomes increasingly prominent by the fourth year, suggesting that as students progress through their academic journey, the significance of the teacher's role becomes more central in their educational experience. This trend indicates a growing appreciation for educators' impact, particularly in the later stages of students' academic careers, highlighting the lasting influence teachers have on them. Figure 4 presents a detailed breakdown of sentiment analysis regarding teachers, segmented by academic year, offering insights into how student perceptions of instructors evolve throughout their academic progression.

4.2.2 *Curriculum sentiment*

The sentiment analysis concerning the curriculum reveals a consistently positive perspective. Key terms such as variety of subject, programming, technology, and work frequently surface, indicating a high level of satisfaction with the curriculum's breadth, contemporary nature, and applicability. These positive sentiments reflect a well-received curriculum for its ability to equip students with the necessary skills and knowledge for their future careers. However, occasional negative feedback, focusing on the difficulty of some subjects and the relevance of certain general education courses, suggests areas for curriculum refinement to better align with student needs and industry trends. Figure 5 illustrates the sentiment analysis results related to the curriculum, providing a snapshot of students' attitudes and opinions towards their academic programs.

Figure 5 Sentimental analysis on curriculum (see online version for colours)

When analysing sentiments toward the curriculum across academic years, from the first to the fourth year, terms like programming, technology, graphics, and work consistently appear. However, their emphasis shifts through the years. Notably, phrases such as variety of subject, work, learning, and computer become particularly significant by the fourth year. This pattern suggests that as students advance through their academic journey, there's an increased appreciation for the curriculum's diversity, practicality, and focus on preparing them for the workforce. The emphasis on various subjects and the relevance of computer-related skills in the fourth year highlights the curriculum's success in meeting students' evolving educational and professional aspirations. Figure 6 displays the sentiment analysis of the curriculum, broken down by academic year, highlighting the shifting perspectives of students on their academic content over time.

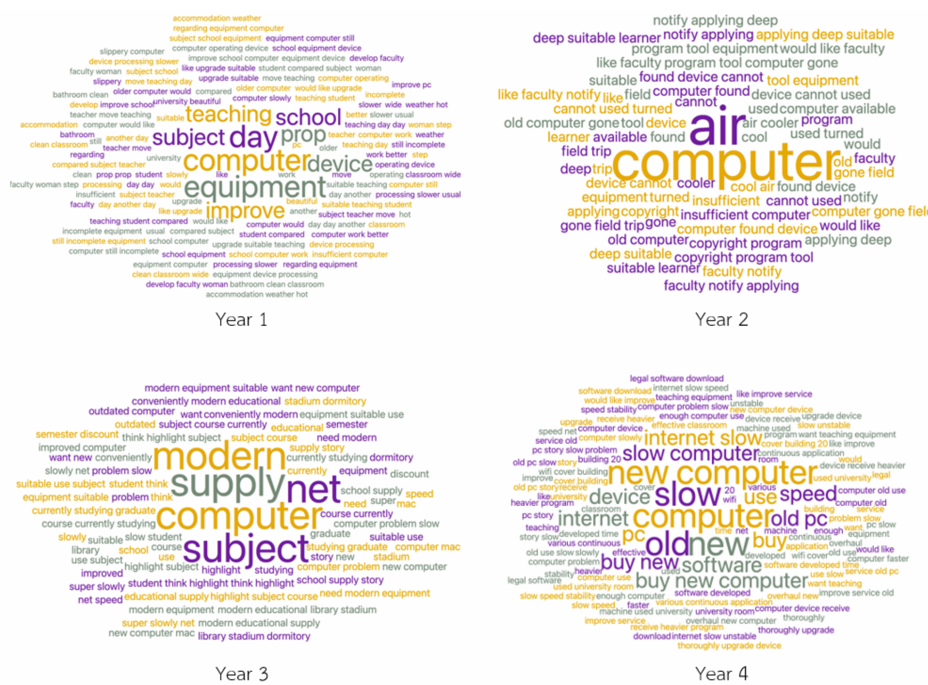
Figure 6 Sentimental analysis on curriculum, separated by academic year (see online version for colours)

4.2.3 Supportive facilities sentiment

The analysis of sentiments concerning supportive facilities reveals various perceptions, prominently featuring terms like slow computer, old device, modern internet, computer problem, internet, insufficient, and library. These keywords suggest a contrast in user experiences, where appreciation for modern internet capabilities coexists with frustrations over outdated equipment and insufficient resources. The recurrent mention of issues related to computer speed, outdated devices, and internet reliability indicates a critical need for upgrades and improvements in the technological and physical infrastructure supporting the learning environment. Additionally, the mention of the library alongside terms indicating insufficiency points to a desire for better-equipped and more resourceful study spaces. These mixed sentiments underscore the importance of updating and enhancing technological resources and infrastructure to meet students' growing needs and support an effective learning experience. Figure 7 depicts the sentiment analysis concerning supportive facilities, showcasing students' feedback on the available resources and infrastructure.

Figure 7 Sentimental analysis on supportive facilities (see online version for colours)

Figure 8 Sentimental analysis on supportive facilities, separated by academic year (see online version for colours)



When sentiments regarding supportive facilities are dissected by academic year, from the first through the fourth year, issues and praises such as slow computers, old devices, modern internet, computer problems, internet, insufficient, and libraries persistently emerge across the spectrum. This distribution indicates a consistent recognition of both the strengths and weaknesses in the supportive infrastructure throughout the academic journey. Notably, 'air' and 'computer' became significantly more prominent keywords in the second year. This suggests that students become particularly sensitive to the quality of air conditioning and computing resources by the second year, reflecting a heightened awareness or possibly an increased need for comfortable and technologically equipped learning environments as their coursework becomes more demanding. Figure 8 provides a segmented sentiment analysis of supportive facilities detailing student evaluations across different academic years.

4.3 Anticipated keywords and word cloud results

The word cloud results play a crucial role in this analysis, mainly corresponding to the anticipated keywords and confirming the initial hypotheses regarding student satisfaction and dissatisfaction. This alignment demonstrates the effectiveness of using word clouds in sentiment analysis to quickly identify key themes and issues based on the frequency of specific terms in the feedback data. For teacher sentiment, positive keywords included teacher-friendly, understand, teaching, good, and friendly, highlighting teachers' approachability, supportiveness, and effectiveness. Negative keywords such as workload and language pointed to concerns about assignment volume and language clarity. In the domain of curriculum sentiment, positive keywords like various subjects, programming, technology, and work reflected satisfaction with the curriculum's breadth, modernity, and practicality. Negative keywords such as difficulty and relevance indicate challenges and concerns with certain subjects and general education courses. Regarding supportive facilities sentiment, positive keywords like modern internet and new computer showed appreciation for updated facilities. In contrast, negative keywords like slow computer, old device, computer problem, insufficient, and library highlighted issues with outdated equipment and resource inadequacies.

5 Discussion

The discussion of the word cloud results provides valuable insights into the sentiments expressed by students regarding various aspects of their educational experience. By examining the prominent keywords associated with teacher sentiment, curriculum sentiment, and supportive facilities sentiment, we can better understand the areas where students feel satisfied and those that require improvement. These insights are crucial for informing targeted interventions and enhancing the overall quality of educational services. Below, we delve into the specific findings for each category, highlighting both positive and negative sentiments as revealed by the word cloud analysis.

5.1 Teachers

Sentiments regarding teachers celebrated the varied teaching methodologies employed, the approachability and friendliness of teachers, and their ability to simplify complex topics for better student comprehension. The word cloud results for teacher sentiment revealed several anticipated positive keywords such as 'teacher-friendly' and ‘understand’. These terms highlight teachers’ perceived approachability and supportive nature, and effectiveness in making complex concepts understandable to students. Conversely, negative keywords like ‘workload’ and ‘language’ were also prominent. This indicates concerns regarding the volume of assignments and tasks given to students and issues related to language use that might pose comprehension challenges for some students. The negative feedback indicated a desire for a reduced workload and a call for using student-friendly language in some courses, highlighting areas where teaching strategies could be refined for enhanced student learning experiences. This feedback from students is invaluable, as it provides educators with a direct insight into the areas where they can improve, making them feel valued and integral to the learning process.

5.2 Curriculum

Regarding curriculum sentiment, the word cloud analysis identified positive keywords such as ‘programming’ and ‘technology’. These reflect student satisfaction with the curriculum’s inclusion of modern technological advancements and the practical application of programming skills. The feedback highlighted the curriculum’s diversity, modernity, and emphasis on practical skills like programming and graphic design, suggesting strong job prospects post-graduation. This positive feedback should inspire educational administrators, as it validates the effectiveness of their decisions in designing a curriculum that meets the needs and interests of the students. Conversely, negative keywords like ‘difficulty’ and ‘relevance’ appeared frequently. These terms suggest that students found certain subjects particularly challenging and expressed concerns about the relevance of some general education courses concerning their main field of study. The negative remarks pointed to the difficulty of major subjects and questioned the relevance of some general education courses. This comprehensive analysis underscores the predominant neutrality in sentiment among educational service recipients, with a noteworthy inclination towards positive feedback on curriculum aspects. Negative comments, while present, were less frequent, offering specific areas for potential enhancement.

5.3 Supportive facilities

Regarding supportive facilities sentiment, the word cloud results showed positive keywords such as ‘modern internet’, ‘air condition’, and ‘new computer’, indicating student appreciation for updated internet facilities and new computer equipment. The positive feedback on supportive facilities highlighted the efficiency of air conditioning, the availability of new and modern computers, and well-equipped classrooms with diverse teaching aids. However, negative keywords such as ‘slow computer’ and ‘insufficient’ were also identified, reflecting complaints about computer speed and a lack of sufficient resources. These critiques emerged regarding outdated and slow computers, suggesting a need for upgrades. These mixed sentiments underscore the importance of continually updating and maintaining educational facilities to meet student needs and expectations.

5.4 Comparing machine learning models with LDA

The sentiment analysis results across the three main aspects have been discussed, revealing notable positive and negative keywords useful for informing further actions. To potentially enhance the insights derived from these results, incorporating latent dirichlet allocation (LDA) for topic modelling could provide a deeper understanding of the topics emerging from each sentiment cluster (Hou et al., 2024; Ma'ady et al., 2024; Zhang and Zhang, 2022). For instance, applying LDA to identify the primary issues within positive and negative sentiment clusters would allow for discovering abstract themes within the feedback data. LDA, as a generative probabilistic model, facilitates the identification of underlying issues in a collection of documents. Several techniques and steps were employed to perform sentiment analysis using LDA topic modelling on the provided datasets. Initially, data was loaded from CSV files and preprocessed using CountVectorizer from sklearn to remove common words (stopwords) and tokenise the remaining words, converting the text into a matrix of token counts. LDA, a generative statistical model that identifies abstract topics within a collection of documents by assuming documents are mixtures of topics and topics are mixtures of words, was then applied. To rank the topics, the distribution of each topic across the documents was calculated, and the mean probability of each topic was computed to measure their prevalence within the entire corpus.

Table 4 Positive topics identified by LDA

Topic	*Mean probability*	*Main words*	*Theme*
0	0.1647	Teacher, teaching, work, content, good, taught, interesting, penetration, building, deep	This topic revolves around the teaching process and teacher effectiveness, highlighting the quality of content and the depth of knowledge provided.
1	0.2616	Variety, programming, subjects, technology, graphics, learning, information, teachers, design, work	This topic focuses on various subjects, particularly programming and technology-related subjects, emphasising the learning of diverse skills and information.
2	0.1555	Knowledge, learning, computer, highlight, like, hard, used, work, lines, think	This topic emphasises acquiring knowledge and using computers, including the difficulty and appreciation of the learning process.
3	0.2161	Easy, branch, career, understand, subjects, teaching, good, computer, course, theory	This topic highlights the ease of understanding certain subjects and their relevance to career paths, emphasising good teaching practices and theoretical knowledge.
4	0.2022	Teacher, friendly, modern, professional, digital, pros, graduation, learn, job, course	This topic focuses on the professional and modern aspects of teaching, mentioning the friendliness of teachers and the preparation for professional careers post-graduation.

Table 5 Negative topics identified by LDA

Topic	*Mean probability*	*Main words*	*Theme*
0	0.1708	Net, equipment, computers, improve, speed, improved, content, computer, work, school	This topic deals with issues related to computer equipment, internet speed, and the need for improvement in these areas.
1	0.1794	Computers, insufficient, teaching, like, students, suitable, real, day, university, computer	This topic highlights the insufficiency of computers and teaching resources, addressing the inadequacy of equipment for student needs.
2	0.1891	Subjects, computer, slowly, study, little, slow, teach, new, subject, important	This topic focuses on the slow performance of computers and its impact on studying, mentioning the need for teaching new and important subjects.
3	0.2696	Slow, computer, internet, activities, field, study, knowledge, new, old, want	This topic addresses slow internet and computer issues, affecting various activities and fields of study.
4	0.1910	Teaching, want, computer, subject, teacher, course, lot, better, study, know	This topic covers general dissatisfaction with teaching and computer resources, expressing a desire for better teaching methods and materials.

5.4.1 Positive topics

LDA analysis of the positive feedback identified five distinct topics, each characterised by a set of top words and corresponding themes. Topic 1, with the highest mean probability of 0.2616, focuses on various subjects, particularly programming and technology-related subjects, emphasising learning diverse skills and information. Topic 3, with a mean probability of 0.2161, highlights the ease of understanding certain subjects and their relevance to career paths, emphasising good teaching practices and theoretical knowledge. Topic 4, with a mean probability of 0.2022, centres on the professional and modern aspects of teaching, mentioning the friendliness of teachers and the preparation for professional careers post-graduation. Topic 0, with a mean probability of 0.1647, revolves around the teaching process and teacher effectiveness, highlighting the quality of content and the depth of knowledge provided. Topic 2, with a mean probability of 0.1555, emphasises acquiring knowledge and using computers, including the difficulty and appreciation of the learning process. These topics collectively provide insights into the key strengths and focus areas of positive feedback. Table 4 summarises the top positive words for each topic identified, along with their themes and mean probabilities.

Table 6 Comparative analysis of LDA and machine learning models

Aspect	*LDA topic modelling*	*Random forest*	*Naïve Bayes*	*LSTM*
Focus	Identifying themes and topics	Classifying sentiment (neutral, positive, negative)	Classifying sentiment (neutral, positive, negative)	Classifying sentiment (neutral, positive, negative)
Output	Key topics with main words and themes	Highest F1 score: 0.833	Lowest F1 score: 0.807	Second highest F1 score: 0.827
Strengths	1 Identifies themes and trends 2 Provides detailed insights into specific areas of feedback	1 High accuracy 2 Provides feature importance 3 Handles high dimensionality well	1 Simple and efficient 2 Good baseline performance	1 Captures context and sequential data 2 Handles nuanced text well
Weaknesses	1 Does not directly classify sentiment 2 Requires interpretation of topics	1 Can be slower to train 2 May overfit with noisy data	1 Assumes feature independence 2 Less effective for complex patterns	1 Computationally intensive 2 Requires large datasets
Application	Understanding broader themes in feedback	Accurate sentiment classification	Baseline sentiment classification	Detailed sentiment classification
Interpretation	1 Provides insights into key issues and areas of strength or improvement 2 Highlights themes in both positive and negative feedback	1 Effectively predicts neutral sentiments with high accuracy 2 Useful for identifying feature importance	Effective for initial or simple sentiment analysis	1 Excellent for detailed and nuanced sentiment analysis 2 Captures the sequence and context of words

5.4.2 Negative topics

LDA analysis of the negative feedback revealed five distinct topics, each defined by a set of top words and corresponding themes. Topic 3, with the highest mean probability of 0.2696, addresses issues related to slow internet and computer performance, impacting various activities and fields of study. Topic 2, with a mean probability of 0.1891, highlights the slow performance of computers and its negative effect on studying, emphasising the need for teaching new and essential subjects. Topic 4, with a mean probability of 0.1910, covers general dissatisfaction with teaching and computer resources, expressing a desire for better teaching methods and study materials. Topic 1, with a mean probability of 0.1794, points to the insufficiency of computers and teaching resources, addressing the inadequacy of equipment for student needs. Topic 0, with a mean probability of 0.1708, focuses on computer equipment and internet speed issues, underscoring the need for improvements in these areas. Table 5 summarises the top words for each negative topic identified by LDA, including their themes and mean probabilities.

5.4.3 LDA and machine learning models

Our research findings have practical implications for understanding feedback. LDA topic modelling, by clustering similar words into topics, provides a deeper understanding of the underlying themes and issues in feedback. This approach helps identify specific areas of concern and strength. However, it does not directly classify sentiment. On the other hand, random forest, Naïve Bayes, and LSTM are focused on accurately classifying sentiment into predefined categories (neutral, positive, negative). Random forest showed the highest accuracy, while LSTM effectively handled sequential data and captured context. Naïve Bayes provided an effective baseline performance. By combining insights from LDA with the classification power of machine learning models, we can provide a comprehensive understanding of feedback, allowing for both detailed thematic analysis and accurate sentiment classification. This research can be applied in various fields where feedback analysis is crucial. Table 6 provides a structured and clear comparison of the results from LDA topic modelling and the three machine learning models used in the study.

6 Implications of the results

The study's findings have significant implications for various stakeholders within the educational ecosystem, which can be categorised into practical and theoretical implications.

6.1 Practical implications

The practical implications of this study are significant, offering actionable recommendations for improving educational services. For instance, the positive feedback on programming and technology courses suggests the need for educational institutions to expand and update their curriculum. This can be achieved through regular curriculum reviews and industry consultations. The study's findings on the positive perception of approachable and effective teachers underscore the importance of continuous professional development programs. This calls for the implementation of training workshops on teaching methodologies, communication skills, and workload management. Besides, addressing negative feedback on outdated and slow computers is also crucial. Universities should invest in upgrading computer labs, improving internet connectivity, and ensuring all technological resources are current and functional. By employing machine learning for sentiment analysis, academic institutions can quickly identify and address student concerns, leading to more responsive and adaptive management practices. In particular, understanding the specific areas of dissatisfaction, such as outdated facilities or excessive workloads, allows for targeted improvements and more efficient resource allocation.

This study's findings benefit several key stakeholders within the educational ecosystem. Academic institutions can use these insights to refine their services, customise curricular offerings, and enhance the overall student experience. By addressing the identified areas of improvement, institutions can foster a more engaging and supportive academic environment. Educational administrators can use the study's findings to formulate and implement policies and practices that address student concerns directly. This proactive approach ensures a more supportive and effective learning environment, improving student satisfaction and retention rates. For educators, the study provides valuable feedback on how their teaching methods are perceived by students. Understanding these perceptions allows teachers to adapt and refine their instructional strategies to better align with student needs and expectations, ultimately enhancing teaching effectiveness and student outcomes. Improved educational services, driven by student feedback, can lead to higher quality teaching, more relevant and engaging curricula, and better facilities, supporting a more positive and enriching educational experience.

6.2 Theoretical implications

The theoretical implications of this study contribute to the existing literature on sentiment analysis and educational research. This study demonstrates the application of advanced machine learning models, particularly LSTM and random forest, in analysing educational feedback, thereby expanding the scope of sentiment analysis in academic contexts. In this sense, the study's methodological framework, including integrating diverse data sources and real-time feedback processing, provides a comprehensive approach that can be adopted and refined in future research. By suggesting the application of LDA for topic modelling within sentiment clusters, this study opens new avenues for exploring detailed themes and issues in educational feedback. Combining sentiment analysis with topic modelling enhances the depth of insights derived from textual data, offering a richer understanding of student sentiments and their underlying concerns. Furthermore, the study's findings on the prevalent neutral and positive sentiments, with fewer negative sentiments, align with Haruechaiyasak et al. (2018), who found that educational feedback highlights positive aspects more frequently. In addition, the identified keywords related to supportive facilities and teaching methods correspond to the findings of previous research that emphasised the importance of infrastructure and instructor quality in educational satisfaction (Sewalk et al., 2018).

7 Conclusions

This study explored the sentiments held by recipients of teaching services, utilising machine learning to parse and understand the nuances of textual feedback. The objective was to extract and analyse the complex emotions associated with educational experiences, subsequently presenting the findings through word cloud dashboards and word clouds. To achieve this, robust machine learning models were developed to categorise sentiment within the collected data, employing algorithms such as random forest, Naïve Bayes, and LSTM. The evaluation of these models was based on the average outcome of 10 repetitive experiments, with the 10-folds cross-validation method ensuring the reliability of performance metrics. Key performance indicators, including accuracy, precision, recall, and the F1 score, were meticulously calculated, with the F1 score serving as the primary metric for assessing model efficacy.

The analysis also revealed that the random forest algorithm outperformed others with an F1 score of 0.833, indicating superior model efficiency. LSTM followed closely with an F1 score of 0.827, while Naïve Bayes lagged slightly with an F1 score of 0.807. Prediction accuracy was highest for neutral sentiments, followed by positive and negative sentiments. Leveraging data from current students and alumni of the Digital Technology major, the study provided insightful analyses through the word cloud, focusing on curriculum content, learning support, and instructional quality. Most feedback was neutral, with positive comments coming in second and negative feedback being the least common.

By analysing sentiments related to teachers, curriculum, and supportive facilities, the study provides valuable insights into the perceptions and experiences of students, contributing to a deeper understanding of educational sentiment dynamics. The research also uncovers the complexity of student sentiments, revealing contradictory views that may not be evident from primary keywords alone. The findings of this research have practical implications for educators and administrators, providing actionable insights for improving curriculum design, teaching practices, and supportive facilities based on student feedback. The LDA application results, identifying distinct themes within positive and negative feedback, highlight key areas of strength and concern. Future work could focus on conducting comparative studies across different cultural and educational contexts to explore how sentiments vary and what factors influence these differences. Furthermore, incorporating additional data sources, such as classroom observations, student interviews, and academic performance records, would enrich the sentiment analysis and provide a more holistic educational experience.

Declarations

Availability of data and material

The datasets utilised and/or analysed during this study are available from the corresponding author upon reasonable request. Additionally, the link to access the data has been uploaded to the submission system along with other files related to the manuscript.

Funding

The authors declare that no external funding was received for the research presented in this manuscript. This statement confirms that the study was conducted without financial support from funding agencies or organisations.

Acknowledgements

The authors thank the *International Journal of Innovation and Learning* for considering our manuscript for publication. We appreciate the opportunity to contribute to advancing research in educational technology and its application in higher education. We also thank Dr. Nasith Laosen from the Master's degree Program in Digital Technology for Business at Phuket Rajabhat University for his valuable suggestions and guidance on the LDA experiment.

References

Choudhary, P. and Chauhan, S. (2023) 'An intelligent chatbot design and implementation model using long short-term memory with recurrent neural networks and attention mechanism', *Decision Analytics Journal*, Vol. 9, p.100359 [online] https://www.sciencedirect.com/science/article/pii/S2772662223001996 (accessed 16 February 2024).

Diekson, Z.A., Prakoso, M.R.B., Putra, M.S.Q., Syaputra, M.S.F., Achmad, S. and Sutoyo, R. (2023) 'Sentiment analysis for customer review: case study of Traveloka', *Procedia Computer Science*, Vol. 216, pp.682–690 [online] https://www.sciencedirect.com/science/article/pii/S1877050922022621 (accessed 12 February 2024).

Dougherty, J., Kohavi, R. and Sahami, M. (1995) 'Supervised and unsupervised discretization of continuous features', in Prieditis, A. and Russell, S. (Eds.): *Machine Learning Proceedings 1995*, pp.194–202, Morgan Kaufmann, San Francisco (CA) [online] https://www.sciencedirect.com/science/article/pii/B9781558603776500323 (accessed 23 January 2024).

Fuller, K., Lupton-Smith, C., Hubal, R. and McLaughlin, J.E. (2023) 'Automated analysis of preceptor comments: a pilot study using sentiment analysis to identify potential student issues in experiential education', *American Journal of Pharmaceutical Education*, Vol. 87, No. 9, p.100005 [online] https://www.sciencedirect.com/science/article/pii/S0002945923000128 (accessed 16 January 2024).

Gulati, K., Kumar, S.S., Boddu, R.S.K., Sarvakar, K., Sharma, D.K. and Nomani, M.Z.M. (2022) 'Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to COVID-19 pandemic', *Materials Today: Proceedings*, Vol. 51, pp.38–41 [online] https://www.sciencedirect.com/science/article/pii/S2214785321032843 (accessed 27 January 2024).

Haruechaiyasak, C., Kongthon, A., Palingoon, P. and Trakultaweekoon, K. (2018) 'S-Sense: a sentiment analysis framework for social media monitoring applications', *Information Technology Journal*, Vol. 14, No. 1 [online] https://ph01.tcithaijo.org/index.php/IT_Journal/article/view/132091 (accessed 18 January 2024).

Hassan, S., Aljohani, N.R., Idrees, N., Sarwar, R., Nawaz, R., Martinez-Camara, E., Ventura, S. and Herrera, F. (2020) 'Predicting literature's early impact with sentiment analysis in Twitter', *Knowledge-Based Systems*, Vol. 192, p.105383 [online] https://www.sciencedirect.com/science/article/pii/S095070511930629X (accessed 22 February 2024).

Heimerl, F., Lohmann, S., Lange, S. and Ertl, T. (2014) 'Word cloud explorer: text analytics based on word clouds', *2014 47th Hawaii International Conference on System Sciences*, IEEE, pp.1833–1842 [online] http://ieeexplore.ieee.org/document/6758829/ (accessed 28 January 2024).

Hou, J., Liu, W., Cao, Y., Wang, S. and Tang, O. (2024) 'Evaluating service quality of express logistics service based on online reviews using LDA-LSTM', *Journal of Management Science and Engineering* [online] https://www.sciencedirect.com/science/article/pii/S2096232024000143 (accessed 15 February 2024).

Jianqiang, Z., Xiaolin, G. and Xuejun, Z. (2018) 'Deep convolution neural networks for Twitter sentiment analysis', *IEEE Access*, Vol. 6, pp.23253–23260 [online] https://ieeexplore.ieee.org/document/8244338 (accessed 26 January 2024).

Just, J. (2024) 'Natural language processing for innovation search – reviewing an emerging non-human innovation intermediary', *Technovation*, Vol. 129, p.102883 [online] https://www.sciencedirect.com/science/article/pii/S0166497223001943 (accessed 22 February 2024).

Kalita, J.K., Bhattacharyya, D.K. and Roy, S. (2024) '3 – Data preparation', in Kalita, J.K., Bhattacharyya, D.K. and Roy, S. (Eds.): *Fundamentals of Data Science*, pp.31–46, Academic Press [online] https://www.sciencedirect.com/science/article/pii/B9780323917780000107 (accessed 22 February 2024).

Korkmaz, A.C. (2023) 'Public's perception on nursing education during the COVID-19 pandemic: sentiment analysis of Twitter data', *International Journal of Disaster Risk Reduction*, Vol. 99, p.104127 [online] https://www.sciencedirect.com/science/article/pii/S2212420923006076.

Lai, S., Hu, X., Xu, H., Ren, Z. and Liu, Z. (2023) 'Multimodal sentiment analysis: a survey', *Displays*, Vol. 80, p.102563 [online] https://www.sciencedirect.com/science/article/pii/S0141938223001968 (accessed 15 February 2024).

Li, H., Yu, X.B., Li, G. and Gao, H. (2023) 'Restaurant survival prediction using customer-generated content: an aspect-based sentiment analysis of online reviews', *Tourism Management*, Vol. 96, p.104707 [online] https://www.sciencedirect.com/science/article/pii/S0261517722002205 (accessed 24 January 2024).

Libnao, M., Misula, M., Andres, C., Marinas, J. and Fabregas, A. (2023) 'Traffic incident prediction and classification system using Naïve Bayes algorithm', *Procedia Computer Science*, Vol. 227, pp.316–325 [online] https://www.sciencedirect.com/science/article/pii/S1877050923016976 (accessed 15 February 2024).

Ma'ady, M.N.P., Rahim, A.F.A., Syahda, T.S.N., Rizqi, A.F. and Ratna, M.C.A. (2024) 'Malaysia citizen sentiment on government response towards COVID-19 disaster management: using LDA-based topic visualization on Twitter', Vol. 234, pp.561–569 [online] https://www.sciencedirect.com/science/article/pii/S1877050924003983 (accessed 15 February 2024).

Malakouti, S.M., Menhaj, M.B. and Suratgar, A.A. (2023) 'The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction', *Cleaner Engineering and Technology*, Vol. 15, p.100664 [online] https://www.sciencedirect.com/science/article/pii/S2666790823000691 (accessed 10 February 2024).

McNaught, C. and Lam, P. (2010) 'Using Wordle as a supplementary research tool', *The Qualitative Report*, Vol. 15, No. 3, pp.630–643 [online] https://nsuworks.nova.edu/tqr/vol15/iss3/8 (accessed 22 February 2024).

Sajeevan, A. and Lakshmi, K.S. (2019) 'An enhanced approach for movie review analysis using deep learning techniques', *2019 International Conference on Communication and Electronics Systems (ICCES)*, pp.1788–1794 [online] https://ieeexplore.ieee.org/document/9002043 (accessed 20 January 2024).

Sano, A.V.D., Stefanus, A.A., Madyatmadja, E.D., Nindito, H., Purnomo, A. and Sianipar, C.P.M. (2023) 'Proposing a visualized comparative review analysis model on tourism domain using Naïve Bayes classifier', *Procedia Computer Science*, Vol. 227, pp.482–489 [online] https://www.sciencedirect.com/science/article/pii/S1877050923017167 (accessed 16 February 2024).

Sewalk, K.C., Tuli, G., Hswen, Y., Brownstein, J.S. and Hawkins, J.B. (2018) 'Using Twitter to examine web-based patient experience sentiments in the United States: longitudinal study', *Journal of Medical Internet Research*, Vol. 20, No. 10, p.e10043 [online] https://www.jmir.org/2018/10/e10043 (accessed 28 January 2024).

Shah, D., Isah, H. and Zulkernine, F. (2018) 'Predicting the effects of news sentiments on the stock market', *2018 IEEE International Conference on Big Data (Big Data)*, pp.4705–4708 [online] https://ieeexplore.ieee.org/document/8621884 (accessed 23 February 2024).

Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y. and Galligan, L. (2023) 'Sentiment analysis and opinion mining on educational data: a survey', *Natural Language Processing Journal*, Vol. 2, p.100003 [online] https://www.sciencedirect.com/science/article/pii/S2949719122000036 (accessed 25 January 2024).

Shamshiri, A., Ryu, K.R. and Park, J.Y. (2024) 'Text mining and natural language processing in construction', *Automation in Construction*, Vol. 158, p.105200 [online] https://www.sciencedirect.com/science/article/pii/S0926580523004600 (accessed 15 February 2024).

Sharma, A., Sharma, M.K. and Dwivedi, R. (2023) 'Exploratory data analysis and deception detection in news articles on social media using machine learning classifiers', *Ain Shams Engineering Journal*, Vol. 14, No. 10, p.102166 [online] https://www.sciencedirect.com/science/article/pii/S2090447923000552 (accessed 28 February 2024).

Shukla, A., Bansal, C., Badhe, S., Ranjtan, M. and Chandra, R. (2023) 'An evaluation of Google Translate for Sanskrit to English translation via sentiment and semantic analysis', *Natural Language Processing Journal*, Vol. 4, p.100025 [online] https://www.sciencedirect.com/science/article/pii/S2949719123000225 (accessed 25 February 2024).

Sinclair, J. and Cardew-Hall, M. (2008) 'The folksonomy tag cloud: when is it useful?', *Journal of Information Science*, Vol. 34, No. 1, pp.15–29, SAGE Publications Ltd. [online] https://doi.org/10.1177/0165551506078083 (accessed 19 January 2024).

Sjarif, N.N.A., Azmi, N.F.M., Suriayati, C., Haslina, M.S., Yazriwati, Y. and Suriani, M.S. (2019) 'Sms spam message detection using term frequency-inverse document frequency and random forest algorithm', *Procedia Computer Science*, Vol. 161, pp.509–515 [online] https://www.sciencedirect.com/science/article/pii/S1877050919318617 (accessed 27 February 2024).

Sullivan, W. (2017) *Machine Learning Beginners Guide Algorithms: Supervised & Unsupervised Learning, Decision Tree & Random Forest Introduction*, CreateSpace Independent Publishing Platform.

Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M. and Liang, X. (2024) 'An improved random forest based on the classification accuracy and correlation measurement of decision trees', *Expert Systems with Applications*, Vol. 237, p.121549 [online] https://www.sciencedirect.com/science/article/pii/S0957417423020511 (accessed 18 February 2024).

Wicaksono, A.T. and Mariyah, S. (2019) 'Social network analysis of health development in Indonesia', *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, pp.1–6 [online] https://ieeexplore.ieee.org/abstract/document/8982482 (accessed 26 January 2024).

Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. (2016) *Data Mining: Practical Machine Learning Tools and Techniques* [online] https://doi.org/10.1016/c2009-0- 19715-5 (accessed 25 February 2024).

Zhang, Y. and Zhang, L. (2022) 'Movie recommendation algorithm based on sentiment analysis and LDA', *Procedia Computer Science*, Vol. 199, pp.871–878 [online] https://www.sciencedirect.com/science/article/pii/S1877050922001107 (accessed 20 January 2024).