Education Journal
2026, Vol. 15, No. 6, pp. 256–270
https://doi.org/10.11648/j.edu.20261506.11

Research Article

Sentiment Analysis of Learner Feedback in Distance Education in Vietnam Using Machine Learning Models

Nguyen Thi Thanh Thuong, Ngo Phuong Linh, Pham Ngoc Tu, Khuat Minh Nguyet, Nguyen Dinh Ngoc Hoai, Do Thi Thanh Tam*, Nguyen Hung Long

Faculty of Applied Digital Technology, Thuongmai University, Hanoi, Vietnam

Abstract

In the context of the rapidly accelerating digital transformation of education, distance learning has become an indispensable part of higher education in Vietnam. To improve teaching quality, educational institutions typically rely on periodic surveys to collect student opinions; however, this approach reflects only the views of a small group at a particular point in time and lags behind reality, whereas discussions and reviews on social media platforms such as Facebook, Threads and TikTok are updated continuously and express learners' opinions objectively. To leverage this resource, the research team collected and constructed the V-LearnerSentiment dataset comprising 6,366 valid observations, filtered and manually annotated from these platforms, to support sentiment analysis and aspect-based sentiment classification. The study compares the performance of traditional machine learning models (Support Vector Machine - SVM, Decision Tree - DT, Random Forest - RF), a deep learning model (Transformer) and large language models (LLMs) based on Accuracy, Precision, Recall and F1-score, using macro and weighted averages to reflect sentiment classification effectiveness under imbalanced data conditions. Experimental results show that Gemini 2.5 achieves the highest Accuracy (86.22%), demonstrating its ability to effectively handle diverse and context-dependent feedback, while the specialized fine-tuned Transformer PhoBERT achieves an Accuracy of 82.6% and better identifies negative feedback belonging to the minority class. In addition to overall sentiment analysis, the team implemented aspect-based sentiment analysis (ABSA) through a two-step procedure: the LDA model was used as a data exploration tool to extract and propose an Aspect Schema consisting of 5 main groups (Tuition & Costs, Degree & Value, Program & Quality, Schedule & Learning Mode, Counseling & Procedures), and a large language model was then employed to extract the context and classify the corresponding sentiment polarity for each aspect, transforming unstructured text data into measurable quantitative indicators. This approach not only enhances the interpretability of the results but also clearly identifies learners' main concerns, thereby providing a practical basis for improving the quality of distance education in Vietnam.

Keywords

Distance Learning, Machine Learning, Sentiment Analysis, Latent Dirichlet Allocation

*Correspondence: Do Thi Thanh Tam (tam.dtt@tmu.edu.vn)

Received: 9 October 2026; **Accepted:** 3 November 2026; **Published:** 10 November 2026

Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

In the context of the digital transformation of education, distance learning is increasingly developing in Vietnam, with about 40 out of a total of 264 higher education institutions currently offering this form of training [1, 2]. Although it offers flexibility in terms of time, location and learning costs, distance learning still faces many problems related to program quality, learning materials, teaching methods, testing and assessment, and learner support services; dropout rates at some institutions have been reported to be high [3-5]. Therefore, directly exploiting learner feedback is necessary to obtain an additional objective source of information for evaluating training quality. In this context, Sentiment Analysis (SA), a task in Natural Language Processing (NLP), enables the automatic identification of users' attitudes from posts and has been increasingly widely applied in online education.

Research on SA in education can be summarized according to its development from traditional machine learning to deep learning and pre-trained language models. The traditional machine learning group typically combines linguistic features such as TF-IDF and bag of words with SVM, RF or Logistic Regression to classify positive, negative and neutral feedback. For instance, Demir and Alkan [6] used SVM, RF and Logistic Regression to analyze sentiment in 5,632 tweets about distance education, in which SVM achieved an accuracy of about 84%. Similarly, Osmanoğlu et al. [7] applied Logistic Regression to 6,059 feedback items on distance learning materials on an institution's LMS and achieved an accuracy of about 77.5%. The advantages of traditional machine learning are its simple structure, ease of implementation and suitability for medium-sized data; however, its ability to represent context and handle complex expressions remains limited.

The development of deep learning has expanded the ability to exploit contextual features and sequential relationships in text. Deep learning models such as Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM),... have been used to analyze sentiment in online course feedback, Moodle data and other educational sources; for example, the study by Onan [8] achieved an accuracy of 95.8% with LSTM combined with GloVe. However, these models still depend on the quality of word representations and face limitations when handling long contexts or highly volatile data. Transformers and pre-trained language models have brought about an important shift thanks to their better contextual representation capabilities. The study by Huang and Raga Jr. [9] achieved 96.08% with a BERT - BiLSTM - Attention architecture on MOOC data; in Vietnamese research, PhoBERT and hybrid architectures have shown high effectiveness on student feedback data, in which Duy Nguyen et al. [10] achieved 93.24% with PhoBERT-CNN-LSTM, higher than SVM at 88.91%. Overall, this group of models has an advantage in capturing context and linguistic features but requires higher computational costs; moreover, their generalizability still needs to be considered, especially for Vietnamese social media data.

Beyond overall sentiment analysis (Sentiment Analysis - SA) - which only identifies the general sentiment of the entire text - many studies have extended to Aspect-Based Sentiment Analysis (ABSA). Unlike traditional SA, ABSA makes it possible to identify each specific aspect/entity being evaluated (such as tuition fees, teaching quality or the curriculum) together with the learner's corresponding sentiment toward each aspect. Machine learning models such as SVM, RF or DT, as well as deep learning models, can all be applied to solve this task. Ton Nu Thi Sau et al. [11] studied aspect-based sentiment analysis on student feedback using a dataset of 5,010 sentences annotated with 11 aspects and three sentiment polarities. The combined BiLSTM-CNN model achieved an F1-score of 78.93% for aspect detection and 73.78% for the joint task of identifying aspects and sentiment polarity. This result shows that aspect-based sentiment analysis can provide detailed information about the specific issues learners are concerned with. However, the study's data were mainly collected within a single higher education institution, which limits the generalizability of the results to the broader learner community.

Motivated by the above limitations, this study proposes to address the problem by combining multiple methods from different groups, from traditional machine learning to advanced language models, within the same aspect-based sentiment analysis framework. Specifically, the methods are implemented on the same Vietnamese social media feedback dataset on distance education, with two closely linked tasks: (1) identifying the aspects mentioned by learners and (2) classifying the sentiment polarity toward each of those aspects. This approach makes it possible to exploit the respective strengths of each group of methods while fairly comparing their effectiveness on both tasks. The results not only reveal whether learners feel positive or negative in general, but also indicate which aspects they are satisfied or dissatisfied with, thereby providing useful information to help institutions take appropriate follow-up actions. In addition to technical metrics, the study also focuses on the interpretability of the models' results on real-world Vietnamese data in order to assess their reliability and applicability in educational practice.

The main contributions of this paper include:

1) Construction of a specialized standardized dataset: The study constructed and published a dataset of 6,366 real-world feedback items collected from popular social media platforms (Facebook, TikTok, Threads) on distance education in Vietnam. All data were preprocessed and manually annotated following a rigorous procedure, creating a high-quality ground-truth data source for further in-depth research in natural language processing for education.
2) Multi-model evaluation and comparison: The paper pro-

vides a comprehensive empirical evaluation of the performance of 05 classification methods from different approaches - from traditional machine learning (Decision Tree, Random Forest, SVM) to deep learning and large language models (PhoBERT, LLM) - on the same dataset and standardized testing procedure. The comparison results help identify the optimal solution for processing highly noisy, unstructured Vietnamese comment text.

3) Academic and practical contributions: By extracting detailed aspects and sentiment nuances, the study provides educational researchers with an empirical database for in-depth analysis of learners' psychology, attitudes and behaviors in the digital environment. At the same time, the experimental results offer a system of objective indicators that helps educational administrators accurately identify operational "bottlenecks", thereby developing learner care policies based on real data (data-driven decision making) and overcoming the time lag and subjectivity of traditional survey methods.

2. Materials and Methods

2.1. Overview of Sentiment Analysis

Sentiment Analysis (SA), or Opinion Mining, is a natural language processing task that aims to identify and classify the attitudes and emotions of a writer toward an entity mentioned in a text. According to the study by Liu in 2012 [12], an opinion can be formally represented as a quintuple (e, a, s, h, t), where *e* is the entity being evaluated, *a* is an aspect of the entity, *s* is the sentiment (positive, negative, neutral), *h* is the opinion holder and *t* is the time. In learner feedback, *e* may be a course or the institution, *a* may be tuition fees, lecturers, the training program, etc., and *s* is the learner's attitude toward *a*.

Regarding the level of analysis, depending on the granularity of the information to be extracted, SA is performed at three levels:

1) *Document level*: a single label is assigned to the entire text d, i.e., f(d) → s, where s ∈ {positive, negative, neutral}. This level assumes that each text discusses only one entity with a consistent attitude.
2) *Sentence level*: determines whether a sentence contains an opinion (subjectivity), then classifies its polarity, i.e., f(c) → s, where s ∈ {positive, negative, neutral} for each sentence c.
3) *Aspect level*: for each pair (text/sentence x, aspect a), the corresponding polarity is determined, i.e., f(x, a) → s. This is the most fine-grained level, because a sentence such as "The lecturers are enthusiastic but the tuition is too high" contains two opposing sentiments that the first two levels cannot distinguish.

According to Medhat et al. [13], SA methods are divided into three main groups: lexicon-based (lexicon-based), machine learning-based (machine learning-based) with models such as Naive Bayes, SVM, DT, RF, CNN, LSTM, BiLSTM, BERT and PhoBERT, and hybrid methods (hybrid) that combine sentiment lexicons with machine learning/deep learning models.

2.2. Data

2.2.1. Data Collection

Data were collected from public discussions on the three social media platforms currently most popular among Vietnamese students: Facebook, Threads and TikTok. These are the main interaction channels where learners frequently share opinions and reviews about service quality and learning experiences.

1) *Scope:* Data were crawled from official Fanpages, student communities and hashtags related to distance learning programs at major higher education institutions in Vietnam.
2) *Data collection period:* Feedback data were limited to the period from August 2023 to October 2025.
3) *Raw data obtained:* The collection process captured both posts and comments on the platforms. Since posts also contain content reflecting the author's opinions and evaluations, similar to a comment, the research team decided to treat each post as a comment unit and refer to all of them as "comments" throughout the dataset. Accordingly, the total raw data obtained comprised 49,773 comments, including 48,735 comments from Facebook, 600 comments (posts) from Threads and 438 comments from TikTok.
4) *Collection method:* The research team used the Facebook Graph API (version v18.0) to extract data from the targeted discussion groups. The system used token-based authentication to crawl all posts and comments, including text content, timestamps and engagement metrics, while applying random delays to ensure stability. The raw data were initially stored in JSON Lines format to preserve the hierarchical structure, then decoded and converted into a standard CSV table format for preprocessing.

2.2.2. Data Preprocessing

Data collected from social media are characterized by unstructured natural language containing much noise, such as advertisements, abbreviations, slang and spelling errors. To ensure the accuracy of the sentiment analysis models, the research team carried out a preprocessing procedure to standardize and improve the quality of the input data. The procedure consists of the following main steps:

a) *Filtering and cleaning raw data:* Data from the Excel file were preprocessed to remove components that carry no sentiment value.

1) *Automatic filtering:* Text was temporarily normalized (converted to lowercase, diacritics removed) to facilitate keyword and regular expression matching. Comments of a technical-interaction or promotional nature (e.g., “ib”, “inbox”, “contact”, “register”...) were removed. In addition, noise elements such as URLs, phone numbers, emojis, and comments that were too short or contained only special characters were also removed, along with stop words (e.g., “oh”, “oh yes”, “is that so”,...).
2) *Manual review:* Due to the limitations of automatic filters in handling the diversity of natural language, the rsearch team conducted a manual review to thoroughly remove remaining comments with no analytical value after the automatic filtering step. The review was carried out by five team members, each responsible for an equally divided portion of the data, reading each row of the spreadsheet to check and remove entries. Removal criteria included “seeding” comments such as “ib” (inbox), “interested”, standalone punctuation marks, hyperlinks, as well as comments that did not clearly express sentiment. To ensure the reliability of the review process, members cross-checked each other's results. For difficult cases, decisions were made by majority vote; if disagreement persisted, a third member was consulted to make the final decision.
3) *After cleaning, the remaining data comprised:* 557 comments from Threads, 5,685 comments from Facebook and 124 comments from TikTok, corresponding to a removal rate of about 87% of the initial raw data. This high removal rate stems from the characteristics of comment data in the distance education domain, where “seeding” (bait comments, fake engagement) is extremely common. Specifically, most posts in Facebook groups related to admissions or counseling “re framed as “requests for reviews”, and the comments below them have almost no analytical value - consisting of standalone punctuation marks (“.”, “…”), “interested”, “inbox me for support”, “inbox”, “check inbox” and similar comments. In addition, the automatic filtering step also removed advertising comments, admission solicitations, and comments that were too short or did not clearly express sentiment, resulting in a remaining volume of comments significantly smaller than the initial raw dataset.

b) *Text normalization:* The team handled abbreviations and non-standard spellings commonly found in students' language. The procedure included frequency scanning to identify words likely to be abbreviations, followed by building a mapping table to convert them into their full forms. The replacement was performed twice using regular expressions to ensure accuracy and handle complex contexts.

c) *Data annotation:* The dataset of 6,366 clean comments was manually annotated by members of the research team. To ensure objectivity and avoid mutual influence, the data were divided into two equal parts; each part was annotated independently by two annotators in an anonymous (double-blind) manner. The labels consist of 03 classes: Positive, Negative and Neutral, assessed based on the overall context rather than individual keywords alone. After completion, the independent annotation results were compared: in cases of agreement, i.e., when the labels of the two annotators matched, that label was selected as the official label; for cases of disagreement after comparison, a linguistics expert discussed directly with the two annotators to reach a consensus. This complete dataset serves as the ground truth for model training and evaluation.

d) *Data splitting:* To objectively evaluate model performance, the preprocessed dataset (6,366 samples) was randomly split into two independent sets at a ratio of 80:20. Specifically:

1) *Training set:* Accounts for 80% of the data and is used by the models to learn linguistic features and optimize parameters.
2) *Test set:* Accounts for 20% of the data and serves as unseen data to verify generalization ability and compute the final performance metrics.

2.3. Sentiment Classification Models

The study compares five sentiment classification models, including three traditional machine learning models (DT, RF, SVM), a Transformer model (PhoBERT) and a large language model (LLM).

2.3.1. Decision Tree

Decision Tree algorithms are supervised machine learning methods suitable for text sentiment classification [14]. The model allows an intuitive representation of the decision-making process through IF-THEN rules, thereby clarifying the relationship between linguistic features (unigrams and bigrams extracted using TF-IDF) and sentiment labels. Decision Tree requires no assumptions about the data distribution and works effectively with high-dimensional data. Many previous studies have used Decision Tree as a foundational model in sentiment analysis. Therefore, in this study, Decision Tree was selected as the baseline model and combined with pruning techniques and hyperparameter tuning to reduce overfitting and improve generalization.

2.3.2. Random Forest

Random Forest is built by combining multiple independent decision trees, each trained on a randomly selected subset of the data and a randomly selected subset of features [15]. This mechanism increases diversity among the component models, thereby reducing variance and improving generalization compared with single decision trees. It is a highly stable ensemble machine learning algorithm that helps reduce overfitting and is effective when working with text feature representations

such as TF-IDF. The algorithm does not require strict assumptions about the data distribution and produces reliable results even with noisy and imbalanced data. Therefore, Random Forest is often used as a baseline machine learning model in sentiment analysis studies to evaluate the effectiveness of traditional methods and serve as a basis for comparison with more modern methods.

2.3.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm widely used in text classification and sentiment analysis [16]. The method works by constructing an optimal hyperplane that maximizes the margin between data classes in the feature space. When combined with the TF-IDF text representation technique, SVM is highly effective in processing high-dimensional and sparse natural language data- a common characteristic of online comment data. Owing to its stability, good generalization ability and lower computational cost compared with deep learning models, SVM remains a reliable traditional method in empirical studies on sentiment analysis.

2.3.4. Transformer (PhoBERT)

The Transformer is a deep learning model introduced in 2017, mainly used in the field of NLP. Its distinguishing feature is the self-attention mechanism, which allows the model to capture global semantic relationships within a text sequence without relying on a sequential structure as traditional models do [17].

In this study, the Transformer model was implemented using PhoBERT - a variant of BERT pre-trained specifically for Vietnamese on a large dataset, which enhances its ability to understand context and language-specific characteristics.

PhoBERT was fine-tuned for the multi-class sentiment classification task with three labels (Positive, Negative, Neutral) using a loss function (cross-entropy), allowing the model to simultaneously learn sentiment and context within the distance learning (online learning) training domain. This approach helps the model effectively distinguish between relevant and irrelevant comments, while improving the accuracy and reliability of sentiment labeling compared with traditional methods based on handcrafted features or classical machine learning models.

Fine-tuning configuration: PhoBERT was fine-tuned on the annotated dataset with the following hyperparameters: a learning rate of 2e-5, a training batch size of 16, 5 epochs and a weight decay of 0.01. The model was evaluated on the validation set and a checkpoint was saved after each epoch. The checkpoint with the best validation performance was selected for the final evaluation on the test set.

2.3.5. Large Language Model (LLM - Gemini 2.5 Flash)

Large Language Models (LLMs) are artificial intelligence models trained on large-scale text datasets to learn and represent natural language at a deep level. Based on the Transformer architecture and the self-attention mechanism, LLMs process text through token units and effectively model long-range contextual relationships between linguistic components [18]. Thanks to their ability to understand global context and perform semantic reasoning, LLMs can accurately identify complex sentiment nuances and indirect expressions in text. Therefore, LLMs are considered a suitable method for sentiment analysis in distance learning systems, outperforming traditional methods based on word frequency features.

This study uses Google Gemini (gemini-2.5-flash) with a zero-shot prompting strategy: the prompt contains no input–output examples, and the model is not fine-tuned on the annotated data. The entire inference behavior of the model is controlled through natural language instructions passed via the system prompt and the user prompt. Comments were processed in batches.

The prompt follows a multi-layer constraint design, including role assignment, data scope restriction, classification task specification and output format standardization. The main instructions are quoted below:

"You are an expert in analyzing comments on distance Higher Education (online university, distance learning). Only analyze comments related to distance Higher Education. If a comment is not related to distance higher education, skip it and do not return a result. Sentiment classification: Positive / Negative / Neutral (based on context). Return only a Json array. If the batch contains no valid comments, return an empty Json array.."

The model's output is mapped to the three sentiment labels directly from the structured JSON generated by the model, in which each element corresponds to one comment and includes the assigned sentiment label and aspect. This approach requires no complex post-processing and helps maintain label consistency when processing in batches.

2.4. Building the Aspect Schema with LDA and Aspect-Based Sentiment Classification

Latent Dirichlet Allocation (LDA) [19] is a probabilistic generative model commonly used to discover latent topics from text corpora. In this model, each document is assumed to be generated from a mixture of latent topics, where each topic is represented by a probability distribution over words. LDA follows a hierarchical Bayesian framework, in which documents are modeled as mixtures of topics and topics are modeled as distributions over the vocabulary. By identifying the words with the highest probability associated with each topic, the model enables the extraction of meaningful semantic patterns from text data. In this study, LDA was applied to learner comments to discover latent topics corresponding to potential aspects in feedback on distance learning. The extracted topics were then interpreted and grouped into higher-level aspect categories for subsequent sentiment analysis.

Let K be the number of topics and V the vocabulary size. The generative process of LDA can be described as follows:

For each topic k $\in$ {1,…,K}:

$$\emptyset_k \sim Dirichlet(\beta) \quad (1)$$

Each topic k is represented by a multinomial distribution over the vocabulary, where φk is the probability distribution of words associated with topic k. The parameter β controls the sparsity of the word distribution in each topic.

For each document d:

$$\theta_d \sim Dirichlet(\alpha) \quad (2)$$

Each document d is associated with a topic distribution θd, representing the proportions of topics appearing in the document. The hyperparameter α determines how concentrated or dispersed the topic mixture is.

For each word position n in document d:

$$z_{d,n} \sim Multinorminal(\theta_d) \quad (3)$$

A topic $Z_{d,n}$ is sampled from the topic distribution θd. This step determines which topic generates the current word.

Generating the observed word:

$$w_{d,n} = \sim Multinominal(\emptyset_{z_{d,n}}) \quad (4)$$

Finally, the observed word $W_{d,n}$ is generated from the word distribution of the selected topic $Z_{d,n}$.

Aspect identification configuration: In practice, a learner's comment may reflect one or more aspects of distance education, such as aspects related to the degree, value, time, training methods, etc. Therefore, the next task is to determine which aspects a learner's comment reflects. To address this problem, the team used the LDA algorithm to identify the set of words with the highest probability in an aspect. At the text vectorization stage, the parameters min_df = 6 and max_df = 0.6 were applied to filter out terms that were too rare or too frequent; here, min_df is the minimum number of comments in which an aspect term must appear, while max_df specifies the proportion used to remove common words from the dataset. The resulting document matrix was then used as input to the LDA model with the parameters num_topics = 10, alpha = 0.1 and eta = 0.1. For each topic, the team took the set of words with the highest probability and named each aspect. Being data-dependent, different studies identify different aspects.

After identifying the aspect groups using LDA, the next step is to determine the sentiment polarity associated with each aspect in each comment. According to Brauwers and Frasincar [20], polarity expresses the direction of sentiment or opinion toward an entity and can be represented by the labels Positive, Negative and Neutral. In this study, Positive means that the aspect is evaluated favorably by the learner or expresses satisfaction; Negative means that the aspect is evaluated unfavorably or expresses dissatisfaction; Neutral refers to cases where the comment does not clearly express a positive or negative evaluation of the aspect. On this basis, for each detected aspect, the related textual context was extracted and processed by a LLM to determine the corresponding polarity, thereby obtaining aspect – sentiment polarity pairs from learners' comments.

Prompt template

System Role: You are a natural language processing expert specializing in the analysis of educational feedback.

Task: Given a learner's comment about distance learning programs, (1) identify which aspects are mentioned; (2) determine the sentiment polarity expressed toward each aspect; (3) List the words/phrases expressing sentiment or evaluative opinions in the comment, serving as evidence for the polarity of each aspect.

The aspects must be selected from the following predefined groups: Tuition & Costs; Degree & Value; Program & Quality; Schedule & Learning Mode; Counseling & Procedures

Sentiment polarity is determined according to three labels:

1) Positive: the aspect is evaluated favorably by the learner, expressing satisfaction, agreement or a good evaluation.

2) Negative: the aspect is evaluated unfavorably by the learner, expressing dissatisfaction, objection or a poor evaluation.

3) Neutral: the aspect is mentioned, but the comment does not clearly express a positive or negative evaluation.

If a comment mentions multiple aspects, return all aspects together with their corresponding sentiments (Positive, Negative or Neutral).

2.5. Evaluation Tools

To objectively evaluate the performance of the sentiment classification models (DT, RF, SVM, Transformer, and LLMs), the study uses the Confusion Matrix as the basis for computing statistical metrics. This matrix allows comparison between the labels predicted by the model and the actual labels of the student feedback data.

The four standard performance metrics applied are:

a) Accuracy (Acc) This is the percentage of correct predictions over the total number of data samples. However, for educational feedback data that may be imbalanced (e.g., more positive comments than negative ones), this metric serves only as a general reference.

b) Precision (Pre) Measures the reliability of the model when making positive predictions. In the context of educational management, high Precision ensures that when the system flags an issue as "Negative", the probability that it is truly an issue needing to be addressed is very high, avoiding false alarms.

c) Recall (Rec) Measures the coverage of the model, i.e., how many actual feedback items the model misses. This metric is particularly important to ensure that institutions

do not miss important student input.

d) F1-Score (F1) The harmonic mean of Precision and Recall. This is the most important metric in this task because it evaluates the balance of the model and is particularly effective when handling real-world datasets with uneven label distributions (imbalanced data).

The F1-score is considered the key metric of the study because it reflects the balance between reliability and coverage and is particularly effective when handling real-world data with uneven label distributions.

Using this multidimensional set of metrics allows the research team not only to evaluate the rate of correct predictions (Accuracy) but also to understand the detailed behavior of each model (through Precision/Recall) in handling complex Vietnamese linguistic nuances.

3. Results

3.1. Sentiment Classification Results

After training and predicting sentiment using five algorithms, namely DT, RF, SVM, Transformer and LLM, the model analysis and evaluation results are summarized in Table 1 and Table 2.

Table 1. *Sentiment analysis results of comments on distance education using five algorithms.*

Algorithm	Positive (%)	Negative (%)	Neutral (%)
Decision Tree	13.12	5.21	81.67
Random Forest	17.29	7.45	75.26
SVM	15.67	6.66	77.67
Transformer	11.97	6.71	81.32
LLMs	15.20	15.44	69.35

Table 1 shows that all five algorithms predict the neutral class as the largest proportion, ranging from about 69% to over 81%. This reflects a common characteristic of educational comments, in which learners mainly provide informational feedback or objective evaluations rather than expressing clear emotions.

Among the traditional machine learning models, Decision Tree is strongly biased toward the neutral class (81.67%), with low positive and negative proportions. This result indicates that the model makes simple decisions and does not effectively exploit the context of complex texts. Compared with DT, RF is more balanced, with the highest positive proportion in the group (17.29%) and the neutral proportion reduced to 75.26%, owing to the combination of multiple decision trees that reduces overfitting and improves generalization. Meanwhile, SVM yields relatively stable results (neutral 77.67%), suitable for short texts when combined with TF-IDF features, although its capacity for deep semantic representation remains limited.

Despite using a deep learning architecture, the Transformer does not show a clear advantage, with a still high neutral proportion (81.32%), nearly equal to Decision Tree. This may be because the model was not thoroughly fine-tuned or the training data were not large enough, especially for Vietnamese.

In contrast, the LLM yields the most distinct sentiment distribution: the lowest neutral proportion (69.35%) and a negative proportion of 15.44%, markedly higher than the other models. This indicates that the LLM better recognizes sentiment nuances and context, reducing the tendency of traditional models to assign neutral labels. Results by institution also show that the LLM reflects sentiment differences between institutions more clearly, especially at those with many comments.

Table 2. *Performance of machine learning models in sentiment classification of Vietnamese comments.*

		Random Forest	Decision Tree	SVM	Transformer	LLMs
Accuracy		75.96%	74.63%	77.65%	82.60%	86.22%
Precision	Positive	0.91	0.69	0.66	0.70	0.86
	Neutral	0.75	0.80	0.78	0.95	0.83

		Random Forest	Decision Tree	SVM	Transformer	LLMs
	Negative	0.88	0.74	0.53	0.79	0.76
	Macro average	0.84	0.74	0.66	0.81	0.82
Recall	Positive	0.35	0.58	0.56	0.88	0.61
	Neutral	0.99	0.93	0.91	0.84	0.95
	Negative	0.19	0.31	0.21	0.96	0.54
	Macro average	0.51	0.61	0.56	0.89	0.70
F1 - score	Positive	0.50	0.63	0.61	0.78	0.72
	Neutral	0.85	0.86	0.84	0.89	0.88
	Negative	0.31	0.44	0.30	0.87	0.63
	Macro average	0.55	0.64	0.59	0.85	0.74

Table 2 presents a comprehensive comparison of the performance of the five models in the Vietnamese sentiment classification task. Overall, the deep learning-based models and the large language model consistently outperform the traditional machine learning methods on most evaluation metrics.

The LLM achieves the highest accuracy (86.22%), along with good Precision, Recall and F1-score in both macro and weighted averages. The small gap between these two types of averages indicates that the model classifies in a balanced manner across classes, including those with few samples, meaning it is less sensitive to class imbalance and captures complex contextual and semantic patterns of Vietnamese well. The Transformer ranks second (82.60%) with a competitive F1-score, confirming the effectiveness of the self-attention mechanism; however, its smaller pre-training scale and number of parameters may make it inferior to the LLM in capturing subtle nuances and long contexts.

In contrast, SVM, Random Forest and Decision Tree show lower performance, particularly for the negative class with low Recall and F1-score, indicating difficulty in identifying implicit, context-dependent sentiment expressions when the data are imbalanced. Nevertheless, Random Forest still has practical value owing to its interpretability and low computational cost.

Overall, the results highlight the advantages of modern language models in sentiment analysis in the higher education domain, where sentiment expressions are often nuanced and indirect. These findings support the view that sentiment modeling benefits considerably from methods capable of learning deep contextual representations, rather than relying solely on traditional feature-based methods. In summary, the experimental results show that applying modern language models is a suitable approach for Vietnamese sentiment analysis in the higher education domain, where linguistic expressions are often subtle and indirect. More broadly, these results support the view that sentiment should be modeled using methods capable of learning deep contextual representations, rather than relying solely on traditional machine learning models with handcrafted feature engineering.

3.2. Analysis of Sentiment Features

One of the challenges for machine learning models is the transparency and interpretability of classification results. To increase the interpretability of the labeling process, the research team performed manual labeling based on the context of each comment. At the same time, the LLM was used to extract words/phrases expressing sentiment based on a predefined set of keywords and sentiment criteria. These words/phrases were checked by the annotators against the context and used as the basis for determining the sentiment polarity (Positive, Negative or Neutral) toward each mentioned aspect.

Positive group: Feedback in this group often contains strong affirmations of the value and benefits of distance education. Specifically, when discussing the value of the degree, learners often use phrases such as “equivalent value” or “the mode of study is not stated” to express confidence in the legal validity of the degree and “necessary job opportunities”. Furthermore, satisfaction with the learning process is clearly expressed through keywords emphasizing autonomy and convenience, such as “flexible”, “self-directed”, “easygoing lecturers”, “caring about students” or “extremely convenient”.

Table 3. Positive sentiment.

Comment (translated)	Sentiment	Sentiment cues
The distance learning program of University X has a long-standing foundation in language training. In addition, since it is a key national university, the degree is more valuable. I find the lecturers there very competent and dedicated.	Positive	"Valuable"; "very competent"; "dedicated"
In S's distance learning program, the tutors are easygoing but not lax; they really care about students. Most classmates are aged 30–40, and some are even older than the tutors. Although the program is entirely online, both the academic advisors and the course lecturers are more caring than in regular full-time classes and even remember students' faces.	Positive	"Easygoing"; "really care"

Negative feedback: Feedback in this group often contains strong adjectives expressing disappointment, a sense of wasted resources, or unfavorable comparisons of training quality. Specifically, when evaluating quality and effectiveness, learners often use negative phrases such as "should not enroll", "extremely discouraging", "ineffective", "far inferior" or "nothing applicable"; "hard to get feedback from lecturers" to reflect the gap between expectations and reality. In addition, complaints about costs and benefits are often associated with keywords such as "buying a degree", "a waste of money", "high tuition" or "making money" (implying that the institution focuses only on profit). Regarding the mode of study, phrases such as "pre-recorded videos" or "padding one's résumé" are often used to criticize the lack of interaction and the low value of the degree in terms of actual competence.

Table 4. Negative sentiment.

Comment (translated)	Sentiment	Sentiment cues
As for the distance learning program of University T, I really advise those who haven't enrolled not to sign up, for example, for the Business Administration major. It is extremely discouraging; frankly, it is no different from buying a degree. If you have money to spare, it might be acceptable, but overall it is very wasteful and almost useless. Weekly quizzes can be retaken many times, and in-person exams can be cheated on if you pay the lecturer.	Negative	"Not recommended"; "extremely discouraging"; "buying a degree"; "wasteful"; "cheating"
I don't know about other universities, but I am studying at University K and find the program ineffective. Lecturers post pre-recorded lecture videos, students do assignments, and the final exam is held at the end of the semester. Tuition is high, and the university seems mainly focused on "making money".	Negative	"Ineffective"; "high tuition"

3.3. Aspect-based Analysis

To select the number of topics, LDA was trained with $K \in \{5,7,10,15\}$ on 80% of the comments and evaluated on the remaining 20%, repeated with 3 random splits (seeds 42, 43, 44). With K = 10, the model achieved a held-out perplexity of 296.3 ± 3.6 (document completion; a uniform distribution model has a perplexity equal to the vocabulary size V = 937) and the highest topic coherence (c_v = 0.636; NPMI = 0.103). Therefore, the team selected K = 10 and then merged the 10 topics into 5 aspect groups owing to their high prevalence and semantic similarity, as shown in Table 5. At the same time, for each aspect group, sentiment cues were identified through the evaluative words/phrases generated by the LLM during polarity analysis, serving as the basis for determining the sentiment corresponding to each aspect.

Table 5. *Aspect groups and top keywords in feedback on distance education.*

No.	Aspect	Scope description	Sentiment cues
1	Tuition & Costs	Discussions related to financial matters, tuition fees and support policies.	“tuition”, “cost”, “fee”, “money”, “expensive”, “cheap”, “payment”, “pay”, “support”, “discount”, “administrative fee”
2	Degree & Value	Evaluations of the legal status and practical value of the degree after graduation.	“degree”, “full-time”, “value”, “job application”, “graduation”, “credit transfer”, “recognition”
3	Program & Quality	Comments on course content, learning materials and lecturers' competence.	“program”, “learning materials”, "knowledge", "lecturer", "instructor", “quality”, “difficult/easy”, “good”, “boring”
4	Schedule & Learning Mode	Factors related to schedules, learning modes and assessment methods.	“time”, “class schedule”, “online”, “distance learning”, “exam”, “assessment”, “Zoom”, “weekend”
5	Counseling & Procedures	Administrative procedures, admissions and interactions with support services.	"admissions", "registration", "application", "enrollment", "counseling", "contact", "inbox"

The aspect–sentiment pairs obtained from learners' comments are illustrated in Table 6.

Table 6. *Examples of aspect-based sentiment annotation for feedback on distance education in Vietnam.*

No.	Aspect	Original user comment (translated)	Sentiment label	Aspect keywords
1	Degree & Value	If you study a distance learning program, the degree is a distance learning degree, but its format and value are equivalent to a regular (full-time) degree, ok	Positive	degree; value; full-time
2	Tuition & Costs	I wish I had asked the seniors before deciding. Continuing is both frustrating and expensive (500,000 VND per credit for online courses with pre-recorded videos), but quitting would waste both time and money.	Negative	“expensive”, “a waste of money”
3	Program & Quality	I am studying at University H, and honestly, it is not effective. Lecturers only post pre-recorded videos, students do assignments, and then take the final exam. Tuition is high and the university is clearly focused on "making money".	Negative	“ineffective”, “pre-recorded videos”, “making money”
4	Schedule & Learning Mode	In-person exams at designated exam centers.	Neutral	“in-person exams”, “exam centers”
5	Counseling & Procedures	You can look into the distance learning program of University Q. The university is admitting students to four majors in July: Business Administration, Accounting, Economic Law and Finance–Banking.	Neutral	“admissions”, “registration procedures”, “application submission”

No.	Aspect	Original user comment (translated)	Sentiment label	Aspect keywords
		Information on the curriculum, registration procedures, application submission and tuition is available on the website.		

Thus, the keyword dictionary acts as a semantic filter, transforming unstructured comments into clearly oriented information. This is an important input database that helps the research team quantify the level of interest and the sentiment nuances of students toward each aspect of distance learning programs.

4. Discussion

The experimental results show that the models differ in their sentiment classification capabilities, as reflected in both overall classification effectiveness and the ability to identify each sentiment class. Gemini 2.5 Flash achieves the highest Accuracy (86.22%), higher than PhoBERT (82.60%), but PhoBERT achieves a Macro Recall of 0.89 and a Macro F1 of 0.85, higher than Gemini's 0.70 and 0.74, respectively, especially for the negative class (F1 0.87 vs. 0.63). Since the neutral class accounts for a large proportion (69.35–81.67% depending on the model), Accuracy is considerably influenced by the dominant class and should therefore be combined with Macro Recall and Macro F1 for a more comprehensive evaluation. This difference may be because PhoBERT was fine-tuned on the research team's labels, whereas Gemini was run zero-shot and is therefore more cautious when assigning sentiment labels (Precision for the positive class 0.86 but Recall only 0.61), while the traditional models have low Recall for the negative class (0.19–0.31).

In this study, the comments after preprocessing were manually annotated into three classes, Positive, Negative and Neutral, and used as reference labels for training, evaluating and comparing the results of the models. Therefore, the Accuracy, Precision, Recall and F1 metrics reflect the degree of agreement between the automatic classification results and the manually defined label set. This evaluation approach allows the methods to be compared on the same data basis and clarifies the difference between overall classification ability and the ability to identify each sentiment class.

This trend is consistent with previous studies: contextualized models outperform traditional methods. The study by Duy Nguyen et al. [10] showed that PhoBERT-CNN-LSTM achieved 93.24% on UIT-VSFC, higher than SVM (88.91%). Similarly, in this study PhoBERT outperforms SVM by about 5 percentage points in Accuracy (82.60% vs. 77.65%) and achieves a Macro F1 of 0.85 compared with 0.64 for Decision Tree. However, the team's absolute results are lower than those of studies using more structured data: SVM achieved about 84% in the study by Demir and Alkan [6] (compared with 77.65% here), the MLP of Bakhsh et al. [16] achieved 89–91%, and the BERT-BiLSTM-Attention model of Huang and Raga Jr [9] achieved 96.08% on MOOC data. This gap may stem from the characteristics of natural feedback on social media (non-standard writing, frequent slang), as well as differences in data sources, the number of sentiment classes and the degree of imbalance; hence, direct comparisons should be made with caution. In addition, this study benchmarks traditional machine learning, Transformer and LLM on the same dataset, which has been limited in previous work. At the aspect level, positive feedback focuses on flexibility, convenience, degree value and lecturer support, while negative feedback relates to training quality, tuition fees, the level of interaction and the practical value of the program.

In future research, the team plans to develop a hybrid classification system combining PhoBERT and LLM to improve the handling of complex comments, while extending the analysis along the time dimension to track learners' sentiment trends. In addition, the research aims to deploy the model as a monitoring and early warning system to support quality management of distance education.

5. Conclusions

The study constructed the V-LearnerSentiment dataset comprising 6,366 valid observations manually annotated from Facebook, Threads and TikTok, filling the gap in benchmark data for sentiment analysis of distance learners in Vietnamese. On this dataset, five models from three generations were compared: traditional machine learning (Decision Tree, Random Forest, SVM), Transformer (fine-tuned PhoBERT) and LLMs (Gemini 2.5, zero-shot).

Modern models outperform traditional ones, but no model wins outright. LLMs achieve the highest accuracy (86.22%), while the Transformer (accuracy 82.60%) leads in F1-score (macro 0.85) and identifies the negative class well (recall 0.96) owing to fine-tuning on imbalanced data. The traditional models (74.63%–77.65%) tend to concentrate predictions on the neutral class, missing most negative feedback, which constitutes an important quality warning signal.

In terms of sentiment, neutral dominates (69.35%–81.67%), while positive sentiment is generally higher than or equal to

negative sentiment. Positive sentiment is associated with degree value, flexibility and lecturer care; negative sentiment is associated with teaching quality, interaction and cost-effectiveness. Accordingly, institutions can shift from periodic surveys to near real-time feedback monitoring, prioritizing improvements in interaction and transparency of information on learning outcomes. Because the sample is unevenly distributed across universities and subject to self-selection bias, the results should be regarded as trend indicators. Future directions include fine-tuning LLMs for Vietnamese, building a hybrid PhoBERT–LLM system and integrating an early warning dashboard into the LMS.

Abbreviations

ABSA	Aspect-Based Sentiment Analysis
LLMs	Large Language Models
LMS	Learning Management System
SA	Sentiment Analysis
SVM	Support Vector Machine
LSTM	Long Short-Term Memory
BiLSTM	Bidirectional Long Short-Term Memory

Author Contributions

Nguyen Thi Thanh Thuong: Data curation, Resources

Ngo Phuong Linh: Investigation, Software

Pham Ngoc Tu: Formal Analysis, Software

Khuat Minh Nguyet: Visualization, Writing – original draft

Nguyen Dinh Ngoc Hoai: Investigation, Validation, Writing – original draft

Do Thi Thanh Tam: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing

Nguyen Hung Long: Formal analysis, Methodology, Validation, Writing – review & editing

Data Availability Statement

The data is available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Appendix

Table 7. Sentiment analysis results of comments on distance education using five algorithms.

Algorithm	Positive (%)	Negative (%)	Neutral (%)
Decision Tree	13.12	5.21	81.67
Random Forest	17.29	7.45	75.26
SVM	15.67	6.66	77.67
Transformer	11.97	6.71	81.32
LLMs	15.20	15.44	69.35

Table 8. Performance of machine learning models in sentiment classification of Vietnamese comments.

		Random Forest	Decision Tree	SVM	Transformer	LLMs
Accuracy		75.96%	74.63%	77.65%	82.60%	86.22%
Precision	Positive	0.91	0.69	0.66	0.70	0.86
	Neutral	0.75	0.80	0.78	0.95	0.83
	Negative	0.88	0.74	0.53	0.79	0.76
	Macro average	0.84	0.74	0.66	0.81	0.82
Recall	Positive	0.35	0.58	0.56	0.88	0.61

		Random Forest	Decision Tree	SVM	Trans-former	LLMs
	Neutral	0.99	0.93	0.91	0.84	0.95
	Negative	0.19	0.31	0.21	0.96	0.54
	Macro average	0.51	0.61	0.56	0.89	0.70
F1 - score	Positive	0.50	0.63	0.61	0.78	0.72
	Neutral	0.85	0.86	0.84	0.89	0.88
	Negative	0.31	0.44	0.30	0.87	0.63
	Macro average	0.55	0.64	0.59	0.85	0.74

Table 9. *Positive sentiment.*

Comment (translated)	Sentiment	Sentiment cues
The distance learning program of University X has a long-standing foundation in language training. In addition, since it is a key national university, the degree is more valuable. I find the lecturers there very competent and dedicated.	Positive	"Valuable"; "very competent"; "dedicated"
In S's distance learning program, the tutors are easygoing but not lax; they really care about students. Most classmates are aged 30–40, and some are even older than the tutors. Although the program is entirely online, both the academic advisors and the course lecturers are more caring than in regular full-time classes and even remember students' faces.	Positive	"Easygoing"; "really care"

Table 10. *Negative sentiment.*

Comment (translated)	Sentiment	Sentiment cues
As for the distance learning program of University T, I really advise those who haven't enrolled not to sign up, for example, for the Business Administration major. It is extremely discouraging; frankly, it is no different from buying a degree. If you have money to spare, it might be acceptable, but overall it is very wasteful and almost useless. Weekly quizzes can be retaken many times, and in-person exams can be cheated on if you pay the lecturer.	Negative	"Not recommended"; "extremely discouraging"; "buying a degree"; "wasteful"; "cheating"
I don't know about other universities, but I am studying at University K and find the program ineffective. Lecturers post pre-recorded lecture videos, students do assignments, and the final exam is held at the end of the semester. Tuition is high, and the university seems mainly focused on "making money".	Negative	"Ineffective"; "high tuition"

Table 11. *Aspect groups and top keywords in feedback on distance education.*

No.	Aspect	Scope description	Sentiment cues
1	Tuition & Costs	Discussions related to financial matters, tuition fees and support policies.	"tuition", "cost", "fee", "money", "expensive", "cheap", "payment", "pay", "support", "discount", "administrative fee"

No.	Aspect	Scope description	Sentiment cues
2	Degree & Value	Evaluations of the legal status and practical value of the degree after graduation.	"degree", "full-time", "value", "job application", "graduation", "credit transfer", "recognition"
3	Program & Quality	Comments on course content, learning materials and lecturers' competence.	"program", "learning materials", "knowledge", "lecturer", "instructor", "quality", "difficult/easy", "good", "boring"
4	Schedule & Learning Mode	Factors related to schedules, learning modes and assessment methods.	"time", "class schedule", "online", "distance learning", "exam", "assessment", "Zoom", "weekend"
5	Counseling & Procedures	Administrative procedures, admissions and interactions with support services.	"admissions", "registration", "application", "enrollment", "counseling", "contact", "inbox"

Table 12. *Examples of aspect-based sentiment annotation for feedback on distance education in Vietnam.*

No.	Aspect	Original user comment (translated)	Sentiment label	Aspect keywords
1	Degree & Value	If you study a distance learning program, the degree is a distance learning degree, but its format and value are equivalent to a regular (full-time) degree, ok	Positive	degree; value; full-time
2	Tuition & Costs	I wish I had asked the seniors before deciding. Continuing is both frustrating and expensive (500,000 VND per credit for online courses with pre-recorded videos), but quitting would waste both time and money.	Negative	"expensive", "a waste of money"
3	Program & Quality	I am studying at University H, and honestly, it is not effective. Lecturers only post pre-recorded videos, students do assignments, and then take the final exam. Tuition is high and the university is clearly focused on "making money".	Negative	"ineffective", "pre-recorded videos", "making money"
4	Schedule & Learning Mode	In-person exams at designated exam centers.	Neutral	"in-person exams", "exam centers"
5	Counseling & Procedures	You can look into the distance learning program of University Q. The university is admitting students to four majors in July: Business Administration, Accounting, Economic Law and Finance–Banking. Information on the curriculum, registration procedures, application submission and tuition is available on the website.	Neutral	"admissions, "registration procedures", "application submission"

References

[1] Giang, M. *The boom in non-regular university training [in Vietnamese]*. 2025; Available from: https://tuoitre.vn/bung-no-dao-tao-dai-hoc-khong-chinh-quy-20251005230256411.htm

[2] Anh, M. *National Economics University: Distance learning enrollment quota 3-5 times that of regular programs, admission scores from only 15 points [in Vietnamese]*. 2025; Available from: https://giaoduc.net.vn/dh-kinh-te-quoc-dan-chi-tieu-dao-tao-tu-xa-gap-3-5-lan-chinh-quy-diem-trung-tuyen-chi-tu-15-diem-post256608.gd

[3] Ky, X. *The "chaos" of non-regular university training [in Vietnamese]*. 2026; Available from: https://nhandan.vn/nhon-nhao-dao-tao-dai-hoc-khong-chinh-quy-post216222.html

[4] An, V. *Reducing dropout rates in online higher education: International experience and policy implications for Vietnamese universities [in Vietnamese]*. 2025; Available from: https://tapchigiaoduc.edu.vn/article/89797/222/giam-ty-le-bo-hoc-trong-giao-duc-dai-hoc-truc-tuyen-kinh-nghiem-quoc-te-va-ham-y-chinh-sach-cho-cac-truong-dai-hoc-viet-nam/

[5] An, N. T. H. and N. Q. Anh, *Reasons for dropout among online distance learning students at Hanoi Open University [in Vietnamese]*. Journal of Science, Hanoi Open University, 2023.

[6] Demir, H., Alkan, A. Sentiment Analysis of Students' Opinions on Distance Education During the COVID-19 Pandemic Using Twitter Data.

[7] Osmanoğlu, U. Ö., Atak, O. N., Çağlar, K., Kayhan, H., Can, T. Sentiment Analysis for Distance Education Course Materials: A Machine Learning Approach. Journal of Educational Technology and Online Learning. 2020, 3(1), 31–48. https://doi.org/10.31681/jetol.663733

[8] Onan, A. Sentiment Analysis on Massive Open Online Course Evaluations: A Text Mining and Deep Learning Approach. Computer Applications in Engineering Education. 2021, 29(3), 572–589. https://doi.org/10.1002/cae.22253

[9] Huang, S., Raga, R. C. Jr. A Deep Learning Model for Student Sentiment Analysis on Course Reviews. IEEE Access. 2024, 12, 136747–136758. https://doi.org/10.1109/ACCESS.2024.3463793

[10] Nguyen, D., Lau, T., Nguyen, H. A., Doan, T., Nguyen, H., Lu, S., Tran, N. A Hybrid PhoBERT-CNN-LSTM Model for Sentiment Analysis of Vietnamese Student Feedback. Asian Journal of Research in Computer Science. 2025, 18(5), 65–74. https://doi.org/10.9734/ajrcos/2025/v18i5639

[11] Ton Nu Thi Sau, Do Phuoc Sang, Pham Thi Thu Trang. Aspect-Based Sentiment Analysis on Student's Feedback in Vietnamese. TNU Journal of Science and Technology. 2021, 226(18), 48–55. https://doi.org/10.34238/tnu-jst.5101

[12] Liu, B. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, Vol. 5, No. 1. San Rafael, CA: Morgan & Claypool Publishers; 2012, pp. 1–167. https://doi.org/10.1561/1500000011

[13] Medhat, W., Hassan, A., Korashy, H. Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Engineering Journal. 2014, 5(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

[14] Kotsiantis, S. B. Decision Trees: A Recent Overview. Artificial Intelligence Review. 2011, 39(4), 261–283. https://doi.org/10.1007/s10462-011-9272-4

[15] Cutler, A., Cutler, D. R., Stevens, J. R. Random Forests. In Ensemble Machine Learning: Methods and Applications. Springer: New York, NY, USA; 2012, pp. 157–175. https://doi.org/10.1007/978-1-4419-9326-7_5

[16] Ahmad, R. I., Iqbal, M. J., Bakhsh, M. Sentiment Analysis of Social Media Contents Using Machine Learning Algorithms. Technical Journal. 2019, 24(4), 47–55.

[17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. Attention Is All You Need. In Advances in Neural Information Processing Systems 30, 2017.

[18] Zhao, W.X., Zhou, K., Li, J. et al. A Survey of Large Language Models. Front. Comput. Sci. 20, 2012627 (2026). https://doi.org/10.1007/s11704-026-60308-3

[19] Blei, D. M., A. Y. Ng, and M. I. Jordan, *Latent dirichlet allocation.* Journal of machine Learning research, 2003. 3(Jan): p. 993-1022.

[20] Brauwers & Frasincar (2022), A Survey on Aspect-Based Sentiment Classification, ACM Computing Surveys, 55(4). https://doi.org/10.1145/3503044