



# Comparative Analysis of Machine Learning and Deep Learning Models for Aspect-based Sentiment Analysis in Education

Sowndarya C.A.<sup>a</sup>, Shashi Dahiya<sup>b\*</sup>, Alka Arora<sup>b</sup>,  
Anshu Bhardwaj<sup>b</sup>, Mukesh Kumar<sup>b</sup>, Mrinmoy Ray<sup>b</sup>  
and Ramasubramanian, V<sup>c</sup>

<sup>a</sup> Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, India.

<sup>b</sup> ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India.

<sup>c</sup> ICAR-National Academy of Agricultural Research Management, Hyderabad, India.

## Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

## Article Information

DOI: <https://doi.org/10.9734/jsrr/2024/v30i122701>

## Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://www.sdiarticle5.com/review-history/127894>

Original Research Article

Received: 09/10/2024

Accepted: 12/12/2024

Published: 21/12/2024

## ABSTRACT

Aspect-Based Sentiment Analysis (ABSA) has emerged as a powerful technique for analyzing student feedback in educational settings, providing a deeper understanding of sentiments linked to specific aspects such as course content, instructor performance, assessment quality and technology support. Unlike traditional sentiment analysis, ABSA enables granular insights by

\*Corresponding author: E-mail: [dahiyashashi23@gmail.com](mailto:dahiyashashi23@gmail.com);

**Cite as:** C.A., Sowndarya, Shashi Dahiya, Alka Arora, Anshu Bhardwaj, Mukesh Kumar, Mrinmoy Ray, and Ramasubramanian, V. 2024. "Comparative Analysis of Machine Learning and Deep Learning Models for Aspect-Based Sentiment Analysis in Education". *Journal of Scientific Research and Reports* 30 (12):567-76. <https://doi.org/10.9734/jsrr/2024/v30i122701>.

extracting multiple aspects from a single review and assigning sentiments to each aspect independently. This study evaluates the performance of traditional Machine Learning (ML) models, including Logistic Regression (LR), Support Vector Machines (SVM), Naïve Bayes (NB), Random Forest (RF) and Gradient Boosting (GB), alongside advanced Deep Learning (DL) models such as Multi-Layer Perceptron (MLP), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT). The focus is on addressing the challenge of handling multiple aspects per review and performing aspect-specific sentiment classification. Experimental results demonstrate that BERT significantly outperforms other models in both tasks, offering superior precision, recall and F1-scores. Notably, BERT excels in handling complex, multi-aspect feedback, providing more accurate sentiment classification for each aspect. These findings highlight the importance of leveraging advanced models to analyze educational feedback effectively, enabling institutions to implement targeted improvements in key areas of learning and teaching.

**Keywords:** *Aspect extraction; BERT; student feedback; long short-term memory; random forest; sentiment classification.*

## 1. INTRODUCTION

The rapid expansion of online education has resulted in an overwhelming amount of student feedback, which serves as an essential source of information for improving educational services. As more institutions move toward digital learning environments, platforms such as Massive Open Online Courses (MOOCs) and other e-learning initiatives are receiving substantial feedback from students across the globe. Traditional Sentiment Analysis (SA) methods typically classify feedback as positive, negative or neutral. However, this method falls short of identifying and analyzing specific aspects of the learning experience, such as course content, instructor performance and technology support. To address this, Aspect-Based Sentiment Analysis (ABSA) has emerged as a more nuanced approach that links sentiment to specific aspects of the feedback (Hussain et al., 2024).

Unlike traditional SA, ABSA allows for a more granular analysis by categorizing feedback into different aspects, providing actionable insights for administrators and educators. For instance, feedback about course content can be classified separately from feedback on instructor performance or technology issues (Ngwira et al., 2023). This granular analysis is especially beneficial for institutions looking to improve specific elements of their courses or teaching strategies.

The growing complexity and volume of data in educational settings make ABSA a crucial tool for educational data analytics. While Natural Language Processing (NLP) and Artificial Intelligence (AI) have been instrumental in improving sentiment classification and aspect

extraction, the shift to online learning and the increasing variety of student feedback necessitate more advanced tools. This has led to the adoption of both Machine Learning (ML) and Deep Learning (DL) models to process large, unstructured feedback data. These models not only improve the accuracy of aspect extraction but also allow for deeper sentiment analysis, providing more detailed insights into the student learning experience (Ngoc et al., 2021).

While several studies have already employed ABSA using advanced NLP and DL techniques, these studies often fail to account for the multiple aspects present in each review. Instead, they typically assign a single aspect category to the entire feedback, along with a single sentiment classification for that aspect (Ngwira et al., 2023). However, student reviews frequently contain multiple aspects (e.g., course content, teaching quality, technology support) with distinct sentiments tied to each one. By not distinguishing between these aspects, previous models limit the depth of their analysis and the ability to address specific concerns raised by students.

This gap in existing research underscores the importance of the current study, which aims to extract multiple aspects per review and assign sentiment to each identified aspect separately. This approach allows for a much finer-grained analysis of educational feedback, which is essential for improving specific components of the educational experience, such as course materials, instructor effectiveness and the use of technology in course delivery. This study not only aims to compare the performance of traditional ML models and DL algorithms for ABSA but also to explore how deep learning models, particularly

BERT, can be used to handle multiple aspects per review and perform aspect-specific sentiment classification.

The primary objective of this study is to compare the performance of several ML and DL algorithms in two critical tasks of ABSA: aspect extraction and sentiment classification. This study evaluates a range of ML models, including Logistic Regression (LR), Support Vector Machines (SVM), Naïve Bayes (NB), Gradient Boosting (GB) and Random Forest (RF) and compares them to advanced DL models such as Multi-Layer Perceptron (MLP), Recurrent Neural Networks (RNN), Long Short-term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT). By evaluating these algorithms, this paper aims to determine which models are best suited to handle the complexity of educational feedback data.

The rest of the paper is structured as follows: Section 2 presents the literature review, summarizing previous research on ABSA, including the use of ML and DL models in education and highlighting gaps in handling multiple aspects per review. Section 3 details the methodology, explaining the models, algorithms and preprocessing techniques used in this study. Section 4 discusses the experiments and results, followed by Section 5, which provides a discussion of the findings and Section 6 concludes the paper with key takeaways and future research directions.

## 2. LITERATURE REVIEW

ABSA has become an increasingly valuable tool for analyzing student feedback in educational contexts, offering a more detailed understanding of specific aspects of the learning experience. Recent research in the field highlights how ABSA has been applied to educational datasets to extract meaningful insights related to course content, instructor performance and the use of technology in learning environments. For example, Kastrati et al. (2020) demonstrated the application of a weakly supervised ABSA framework to analyze student feedback from MOOCs. Their work showed how ABSA could be applied to large datasets with minimal reliance on manually annotated data, significantly reducing the need for expensive human-labeled datasets.

Moreover, Heryadi et al. (2022) applied ABSA to student feedback on online learning programs, revealing important insights that could help

educational institutions improve their course offerings. Their study highlighted the ability of ABSA to extract specific feedback related to various aspects of online courses, including content quality, teaching methods and course structure. Similarly, Alassaf and Qamar (2020) used ABSA for analyzing Arabic-language educational tweets, which underscored the importance of multilingual aspect extraction in diverse educational contexts.

ML models have been integral to ABSA, especially for tasks such as sentiment classification and aspect extraction. Traditional ML algorithms such as LR and SVM are effective in analyzing smaller datasets and handling simpler binary or multi-class sentiment classification tasks. However, these models face challenges when applied to large-scale, complex datasets where deeper linguistic structures need to be understood. SVM, for instance, has been shown to perform well on smaller datasets but often underperforms in more complex scenarios (Sindhu et al., 2019). However, ML models still face limitations in terms of their inability to handle the deep contextual dependencies that exist in educational feedback data. For instance, models like NB and LR fail to capture the subtle relationships between words in longer, more complex sentences, thus limiting their effectiveness in ABSA tasks (Kastrati et al., 2020).

On the other hand, ensemble models like RF and GB are highly effective in reducing overfitting and increasing the accuracy of sentiment classification. These models work by combining the predictions of multiple classifiers to improve overall performance. RF, in particular, has shown strong results in aspect extraction tasks, where nuanced relationships between words need to be captured from feedback data (Edalati et al., 2022).

Deep learning models, particularly RNN and LSTM networks, have shown significant improvements in ABSA by addressing the limitations of traditional machine learning methods. RNNs and LSTMs excel at capturing sequential dependencies in text data, which is crucial for sentiment classification and aspect extraction in feedback, where sentiments may be spread across multiple sentences or contexts. These models have been proven effective in handling the sequential nature of feedback, especially in reviews that contain rich, context-dependent sentiments (Sindhu et al., 2019).

However, BERT, a transformer-based model, has surpassed traditional ML and DL methods due to its ability to process bidirectional context. This ability allows BERT to understand the relationships between words in both directions, making it especially effective for ABSA in educational settings. Alshaikh et al. (2023) found that BERT significantly outperformed other models in both aspect extraction and sentiment classification tasks, particularly when analyzing feedback from MOOCs and traditional educational settings.

Additionally, BERT's capability to capture long-range dependencies and its ability to process complex feedback from students in diverse educational contexts have made it the model of choice for ABSA in education.

In summary, while BERT has proven effective in ABSA tasks, particularly for extracting sentiments from student feedback, most existing studies have focused on assigning a single aspect category and sentiment classification per review. This study aims to bridge this gap by extracting multiple aspects from each review and performing aspect-specific sentiment classification, offering a more detailed and accurate analysis of educational feedback.

### 3. METHODOLOGY

#### 3.1 Dataset

This study utilized a synthetic dataset consisting of 1,500 student feedback records, with 100 records corresponding to each of the 15 ( $2^4 - 1$ ) possible aspect categories. Each record contains six columns: *Review*, *Aspect\_Category*, *Course*, *Instructor*, *Technology* and *Assessment*. The *Review* column contains single or multiple sentences, each of which may address one or more aspects. For example, a single sentence might express sentiments about both the course content and the instructor's performance, while another sentence in the same review might address technology and assessment. The *Aspect\_Category* field represents combinations of four broad aspects: course, instructor, technology and assessment. There are 15 possible *Aspect\_Category* values, ranging from individual aspects (e.g., *course*) to combinations of multiple aspects (e.g., *course, instructor, technology, assessment*). The dataset includes columns for *Course*, *Instructor*, *Technology* and *Assessment*, each representing the corresponding sentiment polarity (positive, negative or neutral) for the aspects present in the

*Review*. This dataset provided a comprehensive foundation for both aspect extraction and sentiment classification tasks. Each record captured detailed feedback, enabling a nuanced analysis of the various elements of the learning experience.

#### 3.2 Data Preprocessing

Before training the ML and DL models, several preprocessing steps were applied to clean and normalize the text data. The following steps outline the preprocessing process:

- **Lowercasing:** All text was converted to lowercase to ensure that the models treat words like "Course" and "course" as identical, eliminating case sensitivity issues.
- **Removal of Numbers:** Numerical values, such as dates and scores, were removed using regular expressions. Numbers were not considered relevant for sentiment or aspect extraction and were thus excluded from the analysis.
- **Removal of Punctuation:** Punctuation marks were removed, except for full stops and commas, as these can be important for sentence boundaries and sentiment classification. Special characters and unnecessary symbols were also stripped away to avoid noise in the data.

After these text cleaning transformations were applied, the cleaned text was vectorized using the Term Frequency-Inverse Document Frequency (TF-IDF) technique. The TF-IDF vectorizer transformed the textual data into numerical feature vectors, which could then be used for training both machine learning and deep learning models. TF-IDF ensured that frequently occurring words in a review were given appropriate weight relative to their frequency across the entire dataset.

#### 3.3 Data Splitting

The dataset was split into 80% for training and 20% for testing to ensure the models had a sufficient amount of data for learning, while leaving a portion of the data for unbiased evaluation. This split is commonly used in ML and DL studies to provide a robust estimate of model performance.

- **Training Set (80%):** Used for fitting the ML and DL models.

- **Testing Set (20%):** Used for evaluating the performance of the trained models.

The split was stratified to ensure that the distribution of aspect categories and sentiment classes remained consistent across both the training and testing datasets. This approach prevents any imbalance that could skew the model's performance on unseen data.

### 3.4 Aspect Extraction

The aspect extraction task was treated as a multi-label classification problem, where each review could be associated with multiple aspects. The goal was to predict one or more aspects relevant to each review. To handle overlapping aspects in a review (i.e., when multiple aspects are addressed in a single review), each aspect was treated as an independent label. A multi-output classifier approach was used, where each aspect category (course, instructor, technology, assessment) was predicted separately for each review. To achieve this, **Logistic Regression (LR)** was used as the base model, implemented within a MultiOutputClassifier to handle multiple aspect categories simultaneously. LR was chosen for its simplicity, interpretability and effectiveness in multi-class and multi-label classification tasks. As a probabilistic linear model, LR estimates the likelihood of a target outcome by modeling the relationship between input features and a binary or categorical response variable. Introduced by Cox (1958), LR has been a foundational algorithm in machine learning, particularly for tasks requiring probabilistic predictions. In this study, each aspect (e.g., course, instructor, technology, assessment) was treated as an independent label and the model was trained on the TF-IDF features of the reviews to predict the presence of one or more aspects. The model was trained using 80% of the data (training set) and its performance was evaluated on the remaining 20% (test set). Metrics such as precision, recall and F1-score were calculated for each aspect category to assess the model's effectiveness in predicting relevant aspects.

In addition to Logistic Regression, various ML and DL models were employed for aspect extraction, each bringing distinct strengths:

- **Support Vector Machines (SVM):** A model known for its ability to handle high-dimensional data efficiently. SVM is particularly effective in text classification tasks, leveraging hyperplanes to separate

classes in feature space (Cortes and Vapnik, 1995).

- **Naïve Bayes (NB):** A probabilistic classifier based on Bayes' theorem. While computationally efficient and easy to implement, NB assumes feature independence, limiting its performance in capturing complex dependencies in text data (Lewis, 1998).
- **Random Forest (RF) and Gradient Boosting (GB):** Ensemble learning methods that combine predictions from multiple classifiers to enhance robustness and accuracy. RF constructs multiple decision trees during training, while GB iteratively improves weak learners to minimize classification errors. These models are particularly suitable for multi-label classification tasks where multiple aspects may co-occur in a single review (Breiman, 2001; Friedman, 2001).
- **Multi-Layer Perceptron (MLP):** A feedforward neural network capable of capturing non-linear patterns in text. MLP is widely used in text classification tasks due to its adaptability and ability to model complex relationships between features (Rosenblatt, 1958).
- **Recurrent Neural Networks (RNN):** Designed to handle sequential data, RNNs are particularly useful in capturing dependencies within text. They process input sequentially, maintaining information from prior words to better understand context (Rumelhart et al., 1986).
- **Long Short-Term Memory (LSTM):** A variant of RNN capable of learning long-term dependencies. LSTMs are especially suited for understanding the relationships between multiple aspects in longer reviews, as they mitigate the vanishing gradient problem (Hochreiter and Schmidhuber, 1997).
- **Bidirectional Encoder Representations from Transformers (BERT):** A state-of-the-art transformer-based model that processes text bidirectionally, capturing nuanced relationships between words. Pre-trained on extensive corpora, BERT has demonstrated exceptional performance in extracting aspects from unstructured student feedback (Devlin et al., 2019).

### 3.5 Sentiment Classification

After extracting the relevant aspects from each review, the next step was to classify the sentiment (positive, negative or neutral) associated with each identified aspect. Sentiment classification was performed for each aspect independently. This allowed the handling of conflicting sentiments (e.g., praise for course content and criticism of the instructor) by assigning separate sentiment labels (positive, negative, or neutral) to each aspect. Separate classifiers were trained for each aspect category to predict the sentiment for reviews where that aspect was present. For instance, a LR classifier was trained for the "course" aspect on reviews where the "course" aspect was identified. The TF-IDF features were used as input to predict whether the sentiment was positive, negative or neutral for that aspect. Similar classifiers were trained for other aspects like "instructor," "technology," and "assessment." This allowed each classifier to focus only on the relevant reviews for each aspect.

In addition to LR, the following ML and DL models were used for sentiment classification:

- **SVM:** Leveraging its strength in high-dimensional feature spaces, SVM effectively classified sentiment for each aspect, particularly in sparse TF-IDF representations.
- **NB:** While efficient and computationally lightweight, NB was less effective for sentiment classification due to its inability to model complex dependencies between words.
- **RF and GB:** Both ensemble methods improved sentiment classification by combining the outputs of multiple classifiers, enhancing robustness and reducing overfitting.
- **MLP:** The neural network architecture of MLP captured complex patterns in text, improving sentiment classification performance for aspects with subtle contextual variations.
- **RNN and LSTM:** These models excelled in understanding sequential dependencies in text, enabling them to capture context-dependent sentiments effectively.

- **BERT:** With its pre-trained bidirectional transformer architecture, BERT achieved state-of-the-art results in sentiment classification. Its ability to process both word-level and sentence-level context allowed it to identify nuanced sentiments tied to specific aspects.

### 3.6 Evaluation Metrics

The models were evaluated on their performance in two key tasks: aspect extraction and sentiment classification. For aspect extraction, precision, recall and F1-score were calculated for each aspect category to measure how effectively the models identified the relevant aspects within student reviews. For sentiment classification, detailed classification reports were produced for each aspect, including precision, recall and F1-scores for the sentiment classes (positive, negative and neutral). The evaluation results provided a comprehensive overview of the strengths and weaknesses of each model in handling both aspect extraction and sentiment classification, enabling a clear comparison of their capabilities in aspect-based sentiment analysis.

## 4. RESULTS

### 4.1 Aspect Extraction

The results for aspect extraction underscore the strengths and limitations of both traditional ML and DL models in identifying multiple aspects within a single review. Among all models evaluated, BERT emerged as the top performer, achieving perfect precision, recall and F1-scores of 1.00. This outstanding performance demonstrates BERT's ability to effectively model complex relationships within unstructured text, particularly in reviews where multiple aspects coexist, such as feedback on course content, instructor performance and technology support. By leveraging its bidirectional attention mechanism, BERT was able to identify relevant aspects with unmatched accuracy, avoiding false positives and negatives.

DL models such as LSTM and MLP closely followed BERT, achieving F1-scores of 0.99. These models demonstrated their strength in sequential data processing, capturing dependencies between words that are essential for understanding the nuances of multi-aspect feedback. However, their reliance on sequential processing meant they could not fully match the efficiency and accuracy of BERT, particularly in

handling reviews with overlapping or subtle aspect mentions.

Traditional ML models such as RF and SVM also performed strongly, achieving F1-scores of 0.99. These models excelled in aspect extraction due to their robust feature-handling capabilities, particularly when paired with effective feature engineering techniques. GB and LR achieved slightly lower F1-scores of 0.98, indicating that while they are effective for structured data, they are less adaptable to the complexities of multi-aspect text. NB lagged behind, with an F1-score of 0.92, highlighting its limited ability to model interdependencies in feedback containing multiple aspects.

The results of aspect extraction are summarized in Table 1, which presents the precision, recall and F1- scores for each model. Fig. 1 provides a visual comparison of the models' performances, highlighting the superior capabilities of BERT and the competitive results of RF, SVM and LSTM.

#### 4.2 Sentiment Classification

The sentiment classification task introduced additional challenges, as the models needed to associate sentiments (positive, negative or neutral) with each identified aspect. Once again, BERT demonstrated its superiority, achieving an F1-score of 0.97. Its ability to handle aspect-

specific sentiment classification was particularly notable, as it excelled at capturing contextual dependencies for each aspect, even in reviews with mixed sentiments across aspects. This performance further highlights BERT's ability to address the complexity of educational feedback, where sentiments are often distributed across multiple sentences or aspects.

Other DL models, such as GB, LSTM and RF, performed strongly, achieving F1-scores of 0.95, 0.94 and 0.94, respectively. These models showed robust capabilities in contextual sentiment analysis, particularly for well-structured feedback. SVM and MLP also demonstrated competitive performance, with F1-scores of 0.93 and 0.94. However, their reliance on fixed feature representations limited their effectiveness in capturing nuanced and context-dependent sentiments. LR (F1-score of 0.89) and NB (F1-score of 0.84) struggled significantly in this task, reflecting their inadequacy in handling the contextual intricacies required for aspect-specific sentiment classification.

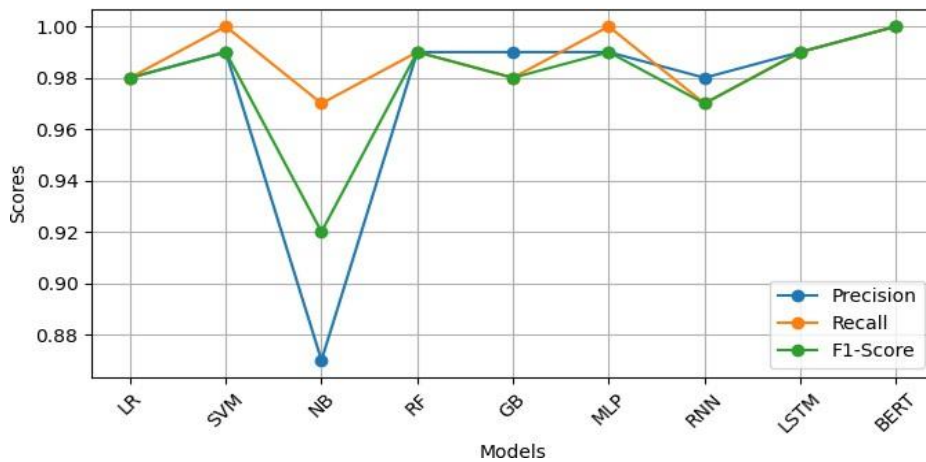
The detailed metrics for sentiment classification are presented in Table 2, which includes precision, recall and F1-scores for all models. Fig. 2 visually compares the performance of the models in sentiment classification, clearly showing the superior performance of BERT alongside the strong results of GB, LSTM and RF.

**Table 1. Results of Aspect Extraction**

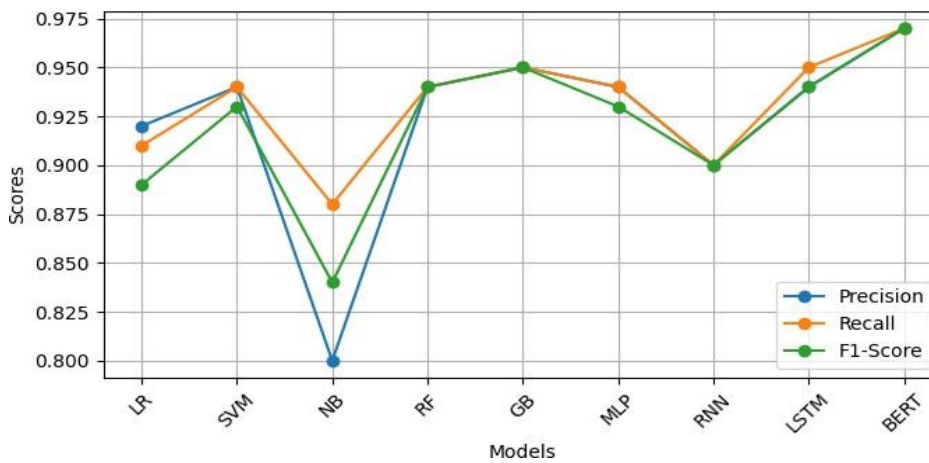
Model	Precision	Recall	F1-Score
LR	0.98	0.98	0.98
SVM	0.99	1.00	0.99
NB	0.87	0.97	0.92
RF	0.99	0.99	0.99
GB	0.99	0.98	0.98
MLP	0.99	1.00	0.99
RNN	0.98	0.97	0.97
LSTM	0.99	0.99	0.99
<b>BERT</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>

**Table 2. Results of Sentiment Classification**

Model	Precision	Recall	F1-Score
LR	0.92	0.91	0.89
SVM	0.94	0.94	0.93
NB	0.80	0.88	0.84
RF	0.94	0.94	0.94
GB	0.95	0.95	0.95
MLP	0.94	0.94	0.93
RNN	0.90	0.90	0.90
LSTM	0.94	0.95	0.94
<b>BERT</b>	<b>0.97</b>	<b>0.97</b>	<b>0.97</b>



**Fig. 1. Models comparison for Aspect Extraction**



**Fig. 2. Models comparison for Sentiment Classification**

## 5. DISCUSSION

The results of this study reveal several important insights into the challenges and opportunities of ABSA in educational feedback. The ability to extract multiple aspects from a single review and assign sentiments to each aspect independently is a critical requirement for meaningful analysis and the performance of the models in this study highlights the relative strengths and weaknesses of different approaches.

BERT's outstanding performance across both tasks underscores its capacity to address the unique challenges posed by multi-aspect feedback. Its bidirectional context modeling enables it to capture subtle relationships between words, making it particularly effective in handling feedback with overlapping or conflicting sentiments. For instance, a single review might praise course content while criticizing instructor performance. BERT's

architecture allows it to identify and classify sentiments for each aspect separately, offering a level of granularity that is essential for actionable insights.

DL models such as LSTM and MLP also demonstrated their utility, particularly in aspect extraction, where their ability to model sequential dependencies resulted in strong performance. However, their reliance on sequential processing meant that they were less efficient than BERT in handling complex feedback where aspects and sentiments are distributed across sentences.

Traditional ML models, particularly RF and SVM, performed well in aspect extraction, largely due to their ability to handle structured data and capture relationships through feature engineering. However, their limitations became apparent in sentiment classification, where the ability to model contextual dependencies is crucial. NB, the weakest performer across both



tasks, highlights the inadequacy of simplistic models in addressing the nuanced nature of ABSA tasks.

The findings also highlight the importance of using task-specific classifiers for sentiment classification. By training separate classifiers for each aspect, the study ensured that models could focus on the unique characteristics and requirements of each aspect. This approach proved particularly effective in improving the accuracy of sentiment classification, as it allowed for a tailored analysis of aspects such as course content, teaching quality and technology support.

Overall, the results emphasize the need for advanced models like BERT to handle the complexities of multi- aspect educational feedback. By providing granular insights into student sentiments, these models enable educational institutions to identify specific areas for improvement and make data-driven decisions to enhance the learning experience.

## 6. CONCLUSION

This study provides a comprehensive analysis of ML and DL algorithms for ABSA, with a specific focus on extracting multiple aspects per review and performing aspect-specific sentiment classification. The findings demonstrate that BERT consistently outperforms other models in both tasks, achieving the highest precision, recall and F1-scores. Its ability to capture bidirectional context and model complex dependencies allows it to effectively handle the challenges of multi-aspect feedback, making it the most reliable and effective model for ABSA in educational settings.

DL models such as LSTM and MLP showed strong performance, particularly in aspect extraction, but their reliance on sequential processing made them less efficient than BERT in handling the complexities of multi- aspect reviews. Traditional ML models like RF and SVM were competitive in aspect extraction but struggled with sentiment classification, highlighting their limitations in modeling contextual dependencies. NB consistently underperformed, reflecting its inability to address the intricacies of unstructured feedback.

The study underscores the transformative potential of ABSA in educational institutions. By

linking sentiments to specific aspects within feedback, ABSA enables targeted improvements in areas such as course content, teaching quality, assessment strategies and technology support. The ability to analyze multi-aspect feedback provides granular insights that traditional sentiment analysis methods cannot achieve.

Future research could extend these findings by exploring the application of BERT and other transformer- based models to multilingual and cross-domain datasets. Additionally, hybrid approaches that combine the strengths of traditional ML and DL techniques could further enhance the efficiency and scalability of ABSA systems. By continuing to refine and expand these methodologies, educational institutions can unlock the full potential of feedback analysis, driving impactful and data-driven improvements in the learning experience.

## DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of this manuscript.

## COMPETING INTERESTS

Authors have declared that no competing interests exist.

## REFERENCES

- Alassaf, M., & Qamar, A. M. (2020). Aspect-based sentiment analysis of Arabic tweets in the education sector using a hybrid feature selection method. In *2020 14th International Conference on Innovations in Information Technology (IIT)* (pp. 178-185). IEEE.
- Alshaikh, K. A., Almatrafi, O., & Abushark, Y. B. (2023). BERT-based model for aspect-based sentiment analysis for analyzing Arabic open-ended survey responses: A case study. *IEEE Access*.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- Cortes, C. (1995). Support-Vector Networks. *Machine Learning*.
- Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 20(2), 215-232.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep

- bidirectional transformers for language understanding. *NAACL*.
- Edalati, M., Imran, A. S., Kastrati, Z., & Daudpota, S. M. (2022). The potential of machine learning algorithms for sentiment classification of students' feedback on MOOC. In *Intelligent Systems and Applications: Proceedings of the 2021 Intelligent Systems Conference (IntelliSys) Volume 3* (pp. 11-22). Springer International Publishing.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189-1232.
- Heryadi, Y., Wijanarko, B. D., Murad, D. F., Tho, C., & Hashimoto, K. (2022). Aspect-based sentiment analysis for improving online learning program based on student feedback. In *2022 IEEE International Conference on Cybernetics and Computational Intelligence (Cybernetics Com)* (pp. 505-509). IEEE.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hussain, T., Yu, L., Asim, M., Ahmed, A., & Wani, M. A. (2024). Enhancing e-learning adaptability with automated learning style identification and sentiment analysis: A hybrid deep learning approach for smart education. *Information*, 15(5), 277.
- Kastrati, Z., Imran, A. S., & Kurti, A. (2020). Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs. *IEEE Access*, 8, 106799-106810.
- Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In *European Conference on Machine Learning* (pp. 4-15). Springer Berlin Heidelberg.
- Ngoc, T. V., Thi, M. N., & Thi, H. N. (2021). Sentiment analysis of students' reviews on online courses: A transfer learning method. In *Proceedings of the International Conference on Industrial Engineering and Operations Management* (pp. 306-314).
- Ngwira, B., Gobin-Rahimbux, B., & Sahib, N. G. (2023). A deep-learning framework for analysing students' review in higher education. *Computational Intelligence and Neuroscience*, 2023(1), 8462575.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386.
- Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986 Oct 9;323(6088):533-6.
- Sindhu, I., Daudpota, S. M., Badar, K., Bakhtyar, M., Baber, J., & Nurunnabi, M. (2019). Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. *IEEE Access*, 7, 108729-108741.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). This publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© Copyright (2024): Author(s). The licensee is the journal publisher. This is an Open Access article distributed under the terms of the Creative Commons Attribution 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.

Peer-review history:

The peer review history for this paper can be accessed here:  
<https://www.sdiarticle5.com/review-history/127894>