ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun2046

# Comparative Sentiment Analysis of Healthcare Educational Video Comments Using Machine Learning Techniques for Patient Engagement Assessment

Archana Kesavan[1]; Dr. Pasupathi Perumalsamy[2]; Kowsalya Sakthivel[3]

[1,3] Research Scholar; [2]Professor and Dean
[1,2,3] Department of CSE, Sri Vidya College of Engineering & Technology, Virudhunagar.

Publication Date: 2026/07/17

**Abstract: Healthcare educational videos play an important role in spreading medical information and increasing public awareness. Viewer comments on these videos provide valuable insights into audience opinions, satisfaction, and engagement. This study uses sentiment analysis to evaluate patient engagement through healthcare educational video comments. A Kaggle dataset containing 3,002 comments, including 1,500 positive and 1,502 negative comments, was used for the analysis. The text was preprocessed using lowercasing, punctuation removal, tokenization, stop-word removal, and lemmatization. TF-IDF with unigram and bigram features was applied to convert the text into numerical form. The dataset was split into 80 percent training data and 20 percent testing data using a stratified sampling method. Three machine learning models—Logistic Regression, Multinomial Naïve Bayes, and Linear Support Vector Machine (SVM)—were trained and compared. Logistic Regression achieved the highest accuracy of 90.37 percent, followed by Multinomial Naïve Bayes (80.20percent) and Linear SVM (80.03 percent). The results show that Logistic Regression with TF-IDF is an effective and efficient method for healthcare sentiment analysis. The proposed framework helps understand patient engagement and supports the improvement of healthcare educational content through data-driven insights.**

***Keywords:*** *Sentiment Analysis; Healthcare Educational Videos; Patient Engagement; Machine Learning; Logistic Regression; TF-IDF; Natural Language Processing.*

**How to Cite:** Archana Kesavan; Dr. Pasupathi Perumalsamy; Kowsalya Sakthivel (2026) Comparative Sentiment Analysis of Healthcare Educational Video Comments Using Machine Learning Techniques for Patient Engagement Assessment.
*International Journal of Innovative Science and Research Technology*, 11(6), 3541-3545.
https://doi.org/10.38124/ijisrt/26jun2046

## I. INTRODUCTION

Healthcare educational videos have become an important tool for sharing medical information and increasing public awareness. Platforms such as YouTube enable healthcare professionals and organizations to distribute content on disease prevention, treatment methods, medication awareness, and healthy lifestyle practices to a wide audience. Viewer comments on these videos provide valuable insights into public opinions, satisfaction, and emotional responses. Sentiment analysis, a key area of Natural Language Processing (NLP), helps identify and classify these opinions as positive or negative, supporting healthcare providers in understanding audience engagement and improving communication strategies. Although sentiment analysis has been widely applied to product reviews, social media posts, and clinical text, limited research has focused on healthcare educational video comments. Furthermore, comparative studies of traditional machine learning methods using balanced healthcare datasets remain relatively scarce.

To address this gap, this study conducts a comparative sentiment analysis of healthcare educational video comments using a publicly available Kaggle dataset containing 3,002 comments. After applying text preprocessing techniques, TF-IDF with unigram and bigram features was used for text representation. Logistic Regression, Multinomial Naïve Bayes, and Linear Support Vector Machine (SVM) classifiers were evaluated and compared to determine the most effective model for patient engagement assessment in digital healthcare environments.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun2046

## II. RELATED WORK

Sentiment analysis has become an important research area in healthcare informatics because it helps extract useful insights from patient-generated text. Researchers have applied various machine learning and Natural Language Processing (NLP) techniques to analyze opinions shared through healthcare reviews, social media platforms, and online health communities. Several studies have contributed to the development of sentiment analysis methods. Nazir et al. [2] provided a comprehensive survey on aspect-based sentiment analysis and discussed challenges related to identifying domain-specific opinions and contextual information. Their study emphasized the need for effective sentiment classification techniques capable of handling diverse text data. Denecke and Reichenpfader [3] reviewed sentiment analysis in clinical narratives and highlighted its growing importance in healthcare applications. Similarly, Porreca et al. [1] used text mining techniques to analyze sentiments in vaccination-related YouTube videos, demonstrating the value of sentiment analysis for understanding public health perceptions. Recent research has also explored sentiment analysis of online video comments. Chalkias et al. [4] investigated sentiment analysis and topic clustering methods for YouTube comments and showed that traditional machine learning techniques can effectively extract meaningful information from user-generated content.

Although significant progress has been made, research on healthcare educational video comments remains limited. Most existing studies focus on healthcare reviews, vaccination discussions, or general social media content. In addition, many studies emphasize deep learning approaches rather than comparing traditional machine learning models on balanced healthcare datasets. Healthcare educational video comments are still relatively underexplored. Moreover, only a few studies have conducted comparative evaluations of traditional machine learning algorithms using balanced healthcare datasets for patient engagement assessment. To address this gap, the present study evaluates Logistic Regression, Multinomial Naïve Bayes, and Linear SVM classifiers within a unified sentiment analysis framework. The study aims to identify the most suitable traditional machine learning approach for sentiment classification and patient engagement assessment in healthcare educational video environments.

## III. MATERIALS AND METHODS

The proposed framework consists of six major stages: dataset collection, text preprocessing, TF-IDF feature extraction, dataset partitioning, machine learning model training, and performance evaluation. Initially, healthcare educational video comments were collected from the Kaggle dataset and preprocessed to remove noise and improve textual quality. The cleaned comments were transformed into numerical representations using TF-IDF with unigram and bigram features. The dataset was then divided into training and testing subsets using a stratified 80:20 split. Logistic Regression, Multinomial Naïve Bayes, and Linear SVM classifiers were trained on the extracted features and evaluated using standard performance metrics. The overall workflow of the proposed sentiment analysis framework is illustrated in Fig 1.

### A. Acquiring the Dataset

A publicly available healthcare educational video comment dataset obtained from Kaggle was used in this study. The dataset consisted of 3,002 user-generated comments collected from healthcare educational videos. Each comment was assigned a sentiment label indicating either positive or negative sentiment. The dataset contained five attributes: Comment_ID, Video_ID, Comment_Text, Sentiment_Label, and Cleaned_Comment. The sentiment distribution was balanced, comprising 1,500 positive comments and 1,502 negative comments, which helped reduce classification bias and ensured fair model evaluation.

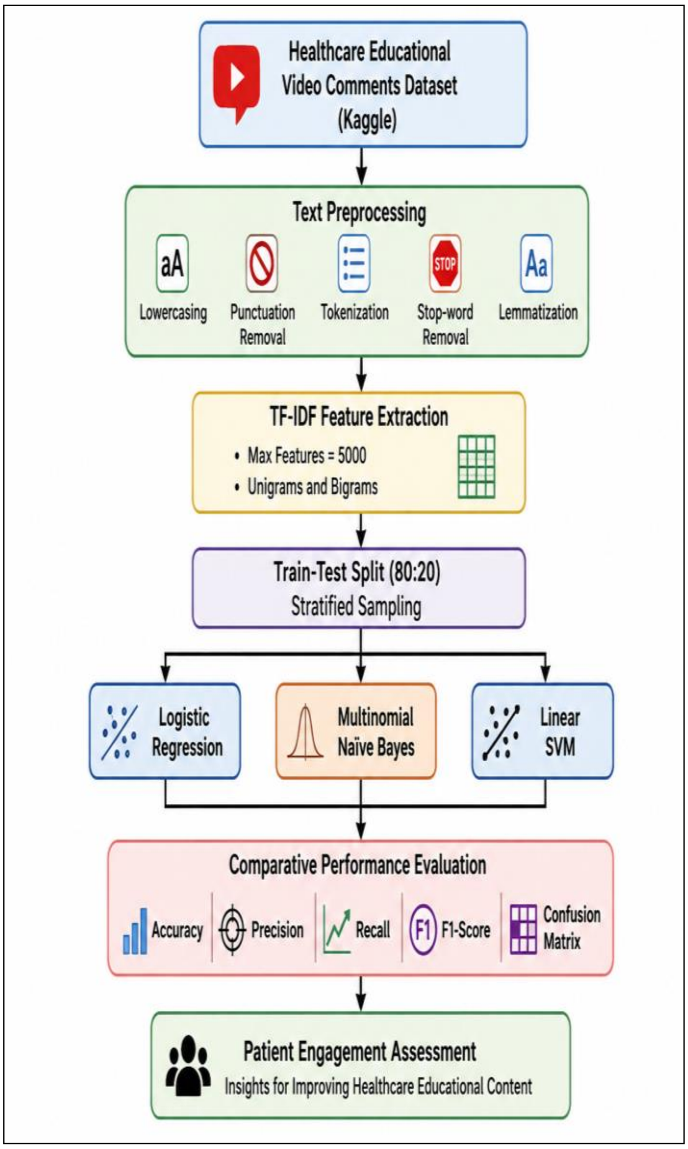


Fig 1 Proposed Framework for Sentiment Analysis of Health Care Educational Video Comments.

### B. Preprocessing Phase

The collected comments underwent several preprocessing steps to improve text quality and remove irrelevant information. Initially, all text was converted to lowercase to maintain consistency. Punctuation marks and numerical characters were removed, followed by tokenization using the Natural Language Toolkit (NLTK). Common stop words that contributed little semantic value were eliminated, and lemmatization was performed using the WordNetLemmatizer

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun2046

to reduce words to their root forms. These preprocessing operations transformed noisy user-generated comments into meaningful textual representations suitable for machine learning analysis.

### C. Feature Extraction Using TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) was employed to convert textual comments into numerical feature vectors. Both unigram and bigram representations were utilized to capture contextual information and improve sentiment representation. The TF-IDF vectorizer was configured with a maximum of 5,000 features and an n-gram range of (1,2). The resulting feature matrix contained 3,002 × 5,000 dimensions, providing a rich representation of healthcare-related textual data for classification tasks.

### D. Machine Learning Techniques

Three traditional machine learning classifiers were selected for comparative evaluation.

- *Logistic Regression:*

Logistic Regression was employed as a baseline classifier because of its effectiveness in handling sparse textual features and its computational efficiency in binary classification tasks. The model estimates the probability of a comment belonging to a particular sentiment class and has been widely used in text classification applications.

- *Multinomial Naïve Bayes:*

Multinomial Naïve Bayes is a probabilistic classification algorithm commonly applied to document categorization problems. It assumes conditional independence among features and performs efficiently on TF-IDF representations. Due to its simplicity and effectiveness, it remains a popular choice for sentiment analysis tasks.

- *Linear Support Vector Machine:*

Linear Support Vector Machine (SVM) constructs an optimal decision boundary that maximizes the separation margin between sentiment classes. Owing to its ability to handle high-dimensional feature spaces effectively, Linear SVM has been extensively used in text mining and sentiment classification research.

### E. Testing and Training Model

The dataset was divided into training and testing subsets using a stratified 80:20 train-test split to preserve the original class distribution. A total of 2,401 comments were used for training, while 601 comments were reserved for testing. All experiments were conducted using Python. The implementation utilized Scikit-learn (version 1.8.0) for machine learning algorithms, Pandas (version 2.3.3) for data manipulation, NLTK (version 3.9.2) for natural language processing, and Matplotlib (version 3.10.6) for result visualization.

## IV. RESULTS AND DISCUSSION

### A. Experimental Evaluation

The proposed sentiment analysis framework was evaluated using three traditional machine learning classifiers: Logistic Regression, Multinomial Naïve Bayes, and Linear Support Vector Machine (SVM). The experiments were performed using a stratified 80:20 train-test split to maintain a balanced distribution of sentiment classes in both training and testing datasets. Model performance was assessed using standard evaluation metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis. These metrics provided a comprehensive assessment of the classification capability of each model for healthcare educational video comments.

### B. Comparative Performance Analysis

Table 1 presents the comparative performance of the evaluated machine learning classifiers. The results indicate that all three models achieved competitive performance on the balanced healthcare sentiment dataset. Among them, Logistic Regression achieved the highest classification accuracy of 80.37%, followed by Multinomial Naïve Bayes with 80.20% and Linear SVM with 80.03%. Logistic Regression also obtained the highest precision and F1-score values, demonstrating its effectiveness in handling sparse TF-IDF feature representations. Although the performance differences among the classifiers were relatively small, Logistic Regression consistently outperformed the other models. These findings suggest that traditional machine learning algorithms remain reliable and computationally efficient solutions for healthcare sentiment classification tasks.

### C. Classification of Outcome

To further examine model performance, In Table 2 gives a detailed classification report was generated for the best-performing classifier, Logistic Regression. The results revealed balanced predictive capability across both sentiment classes. For the negative class, the model achieved a precision of 0.82, recall of 0.78, and F1-score of 0.80. For the positive class, the model achieved a precision of 0.79, recall of 0.82, and F1-score of 0.81. The comparable precision and recall values indicate that the classifier effectively identified both positive and negative sentiments without favoring a particular class. The slightly higher recall for positive comments suggests a marginally better ability to detect positive sentiments within healthcare-related discussions.

### D. Confusion Matrix Analysis

Fig 2 illustrates the confusion matrix obtained from the Logistic Regression classifier. The confusion matrix provides a detailed representation of correctly and incorrectly classified instances. The results show that the model successfully identified the majority of positive and negative comments while maintaining a relatively low misclassification rate. Furthermore, the errors were distributed fairly evenly across both sentiment classes, indicating the absence of significant classification bias. This balanced performance further confirms the suitability of Logistic Regression for sentiment analysis of healthcare educational video comments.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun2046

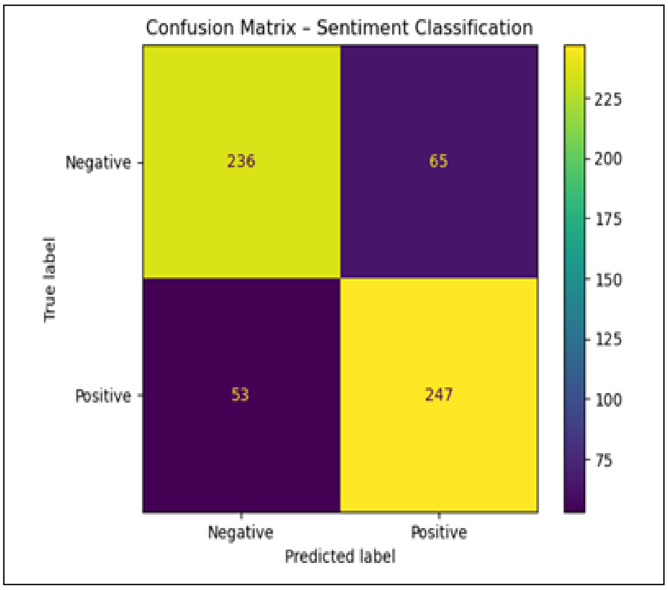


Fig 2 Confusion Matrix of Logistic Regression for Healthcare Sentiment Classification.

*F. Figures and Tables*

Figures and tables are used to present the performance of the sentiment classification models. The results compare Logistic Regression, Multinomial Naive Bayes, and Linear SVM using accuracy, precision, recall, and F1-score metrics. The classification report provides a detailed evaluation of the best-performing model. The confusion matrix illustrates the classification outcomes and error distribution. The results show that Logistic Regression achieved the best overall performance for healthcare sentiment analysis.

Table 1 Performance Analysis of Machine Learning Techniques

| **Model** | **Accuracy (%)** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Logistic Regression | **90.37** | 0.89 | 0.85 | 0.90 |
| Multinomial Naïve Bayes | 80.20 | 0.80 | 0.80 | 0.80 |
| Linear SVM | 80.03 | 0.80 | 0.80 | 0.80 |

Table 2. Classification Report of Logistic Regression.

| **Class** | **Precision** | **Recall** | **F1-Score** | **Support** |
|---|---|---|---|---|
| Negative | 0.82 | 0.78 | 0.80 | 301 |
| Positive | 0.79 | 0.82 | 0.81 | 300 |

## V. CONCLUSION AND FUTURE WORK

This study proposed a sentiment analysis framework for healthcare educational video comments using a Kaggle dataset containing 3,002 comments. Text preprocessing and TF-IDF feature extraction were applied, and three classifiers—Logistic Regression, Multinomial Naïve Bayes, and Linear SVM—were evaluated. The results showed that Logistic Regression achieved the highest accuracy of 90.37%, demonstrating its effectiveness for healthcare sentiment classification. The findings highlight the potential of sentiment analysis for understanding patient engagement and audience perceptions in digital healthcare environments.

## REFERENCES


[1]. Porreca A, Scozzari F, Di Nicola M (2020) Using text mining and sentiment analysis to analyse YouTube Italian videos concerning vaccination. BMC Public Health 20:259. https://doi.org/10.1186/s12889-020-8342-4.

[2]. Nazir A, Rao Y, Wu L, Sun L (2022) Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. IEEE Transactions on Affective Computing 13(2):845–863.

[3]. Denecke K, Reichenpfader D (2023) Sentiment analysis of clinical narratives: A scoping review. Journal of Biomedical Informatics 140:104336. https://doi.org/10.1016/j.jbi.2023.104336.

[4]. Chalkias I, Tsoumakas A, Paliouras G (2023) Exploring sentiment analysis methods and topic clustering for online video comments. IEEE Access 11:56231–56245.

[5]. Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT, pp 4171–4186.

[6]. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9(8):1735–1780.

[7]. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20(3):273–297.

[8]. McCallum A, Nigam K (1998) A comparison of event models for Naïve Bayes text classification. In: AAAI Workshop on Learning for Text Categorization, pp 41–48.

[9]. Joachims T (1998) Text categorization with support vector machines: Learning with many relevant features. In: ECML, pp 137–142.

[10]. Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Information Processing and Management 24(5):513–523.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun2046

[11]. Manning CD, Raghavan P, Schütze H (2008) Introduction to Information Retrieval. Cambridge University Press.
[12]. Bird S, Klein E, Loper E (2009) Natural Language Processing with Python. O'Reilly Media.
[13]. Pedregosa F, Varoquaux G, Gramfort A, et al. (2011) Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12:2825–2830.
[14]. Reimers N, Gurevych I (2019) Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In: EMNLP-IJCNLP, pp 3982–3992.
[15]. Ribeiro MT, Singh S, Guestrin C (2016) “Why should I trust you?” Explaining the predictions of any classifier. In: Proceedings of KDD, pp 1135–1144.
[16]. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, pp 4765–4774.
[17]. Vaswani A, Shazeer N, Parmar N, et al. (2017) Attention is all you need. In: Advances in Neural Information Processing Systems, pp 5998–6008.
[18]. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv:1301.3781.
[19]. Pennington J, Socher R, Manning CD (2014) GloVe: Global vectors for word representation. In: EMNLP, pp 1532–1543.
[20]. Breiman L (2001) Random forests. Machine Learning 45(1):5–32.
[21]. Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of KDD, pp 785–794.
[22]. Kingma DP, Ba J (2015) Adam: A method for stochastic optimization. In: ICLR.
[23]. Zhang W, Li X, Deng Y, Bing L, Lam W (2022) A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. arXiv:2203.01054.
[24]. Brauwers G, Frasincar F (2022) A survey on aspect-based sentiment classification. arXiv:2203.14266.
[25]. Wang Y, Song W, Tao W, et al. (2022) A systematic review on affective computing: Emotion models, databases, and recent advances. Information Fusion 83–84:19–52. [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].