STM JOURNALS

International Journal of
Data Structure Studies

ISSN: 3049-2440
Volume 4, Issue 1, 2026
January–June
DOI (Journal): 10.37591/IJDSS

https://journals.stmjournals.com/ijdss

Review IJDSS

Machine Learning Based Sentiment Analysis of Student Feedback in Higher Education

Ramani Jaydeep Ramniklal*

Abstract

Educational institutions routinely collect feedback from students to understand their perceptions of academic programs, infrastructure, and campus facilities, to improve the overall quality of the college environment. In current practice, feedback is often gathered using numerical or grade-based rating systems, which tend to oversimplify student opinions and may overlook important details related to their level of satisfaction. In contrast, open-ended textual feedback allows students to clearly express their views, concerns, and suggestions, offering valuable insights that can support meaningful institutional improvements. This study proposes an approach for analyzing student sentiment by applying sentiment analysis techniques to textual feedback obtained from students in Gujarat. Machine learning models, including multinomial Naïve Bayes (MNB), support vector machine (SVM), and random forest (RF), are employed to classify the sentiments expressed in the feedback data. A comparative performance analysis of these algorithms is conducted using evaluation metrics such as accuracy and F-score to determine their effectiveness. The experimental results demonstrate that, among the evaluated methods, the MNB classifier achieves superior performance in most cases. Its effectiveness is particularly notable in processing and classifying text-based data, making it a suitable choice for sentiment analysis tasks in educational feedback systems.

Keywords: Algorithms, Feedback, machine learning, multinomial Naïve Bayes, sentiment analysis, support vector machine (SVM)

INTRODUCTION

Feedback refers to the information provided to an individual regarding their past actions or performance, enabling them to modify their present and future behavior to achieve the desired outcomes [1]. In the context of education and learning, feedback plays a crucial role in helping learners acquire new knowledge, enhance their understanding, and avoid making mistakes. Within higher education, it is regarded as an essential mechanism for fostering students' growth as independent learners who can effectively monitor, evaluate, and regulate their own learning processes [2]. Effective feedback practices offer valuable insights that guide students to improve their academic performance.

***Author for Correspondence**
Ramani Jaydeep Ramniklal
E-mail: jaydeep.r.ramani@gmail.com

Associate Professor, Department of Computer Science and Management, B.H. Gardi College of Engineering and Technology, Rajkot, Gujarat, India

Received Date: November 06, 2025
Accepted Date: December 15, 2025
Published Date: March 20, 2026

Citation: Ramani Jaydeep Ramniklal. Machine Learning Based Sentiment Analysis of Student Feedback in Higher Education. International Journal of Data Structure Studies. 2026; 4(1): 1–10p.

Based on the type of response provided by the students, feedback can be categorized as either textual or grading (Likert-scale-based). In Likert-scale feedback, students respond using a five-point scale that focuses on predefined aspects but often fails to capture their precise emotions and opinions. To better understand students' true sentiments, textual feedback is employed, wherein learners express their thoughts in written form [3–5]. This approach benefits both instructors and academic administrators by providing deeper insights that help address institutional challenges and enhance the

overall learning environment of the course.

In this study, feedback from students across Gujarat was gathered through Google Forms and reflected a range of opinions. The main goal is to detect and classify opinion statements in the text as positive, negative, or neutral by applying machine learning methods [6].

Sentiment analysis involves determining the sentiment conveyed in textual data and has become increasingly important in the current context. Typically, there are three main approaches: machine learning based, lexicon-based, and hybrid approaches. Machine learning methods utilize either supervised or unsupervised algorithms, such as Naïve Bayes, support vector machines (SVM), and random forest (RF), to perform classification tasks [7–10]. In contrast, lexicon-based techniques identify sentiment polarity by referencing a predefined dictionary of words linked to specific sentiment scores. The hybrid approach combines these two methodologies to improve the overall accuracy, reliability, and analytical performance.

LITERATURE REVIEW

A significant amount of research has been conducted in the field of sentiment analysis. However, few studies have focused on text classification that specifically classifies a sentence into three classes: positive, negative, and neutral [11–15]. Prior work includes a hybrid approach using TF-IDF and a domain-specific sentiment lexicon, but its limitation was that it only computed the overall sentiment of the feedback. Another survey reviewed frameworks, feature extraction, and sentiment analysis methods, and discussed existing problems in those studies. Supervised learning techniques, such as SVM and Naïve Bayes (NB), are widely recognized as conventional approaches. SVM generally performs better with large datasets, whereas NB is preferred for smaller datasets [16–20].

Data mining methodologies using text mining and a Naïve Bayes classifier have been used to rate faculties, though a drawback was not revealing the true sentiment of the students. The problem of processing a large number of end-of-semester feedback (a tedious and time-consuming job) has been addressed by analyzing feedback using machine learning techniques such as maximum entropy (ME), SVM, and Naïve Bayes [21–27]. The sentiment analysis conducted on Twitter data showed that the Decision Tree algorithm achieved better performance than the multinomial Naïve Bayes model, as indicated by higher values across evaluation metrics, including accuracy, precision, recall, and F1-score. The Naïve Bayesian approach has been used for text and document classification using n-gram features, with performance measured by recall, F-measure, precision, and accuracy [28–36].

METHODOLOGY

The student feedback collected from Gujarat students using Google Forms constituted the input data, which served as the training data for the system. After the test samples were received, the trained model processed and categorized each sentence into positive, negative, or neutral classes using machine learning techniques. The classification outcomes were subsequently displayed through graphical visualizations [37–42].

The suggested approach consists of six consecutive stages: preparing the training dataset, gathering feedback from students, extracting relevant features, training the model, testing and evaluating its performance, and visually presenting the results. This procedure is illustrated in Figure 1.

Preparing Training Data

Machine learning can be divided into two main types: supervised and unsupervised. In supervised learning, labeled data are utilized, where each example is associated with a known output. For every question, the training dataset was collected from students' answers, which were written in sentence form.

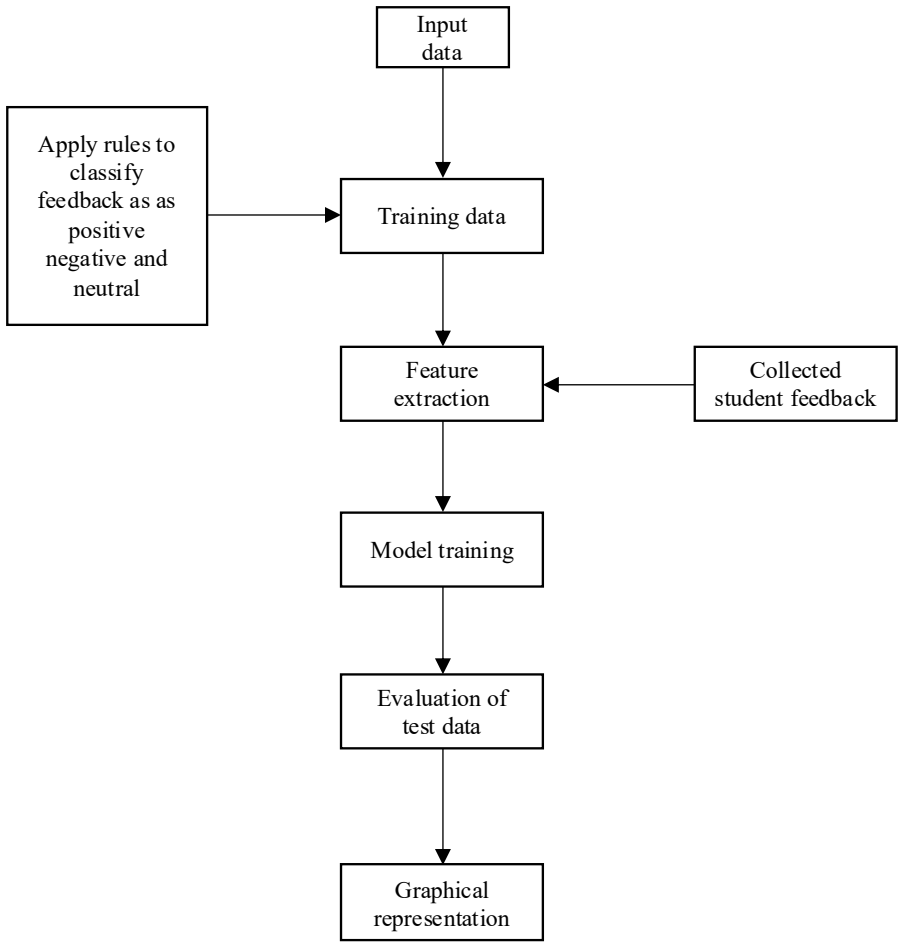

Figure 1. Methodology of sentiment analysis.

Collecting Student Feedback

The process of collecting student feedback in Gujarat begins with the meticulous design of a comprehensive set of questions carefully formulated to gather specific insights into students' educational experiences. These questions were then deployed using Google Forms, a highly accessible and user-friendly digital platform chosen for its efficiency in distributing surveys and capturing responses. Survey links were strategically disseminated to students across Gujarat through various official and informal digital channels, such as university email lists, student portals, and communication groups, to encourage broad participation. As students access the forms and submit their responses, the data are automatically and securely recorded by Google Forms. For subsequent analysis, the collected data were systematically exported, with the preferred format being a Microsoft Excel comma-separated values (CSV) file. This format was selected to ensure broad compatibility and seamless integration with a wide range of analytical tools. Within the CSV file, each row distinctly represents a complete feedback submission from an individual student, and commas are utilized to clearly separate each specific answer to the survey questions. This structured organization of raw feedback in a CSV file is crucial as it prepares the data for the ensuing stages of analysis, including feature extraction and model training, thereby enabling a robust evaluation of student sentiment and contributing to educational improvements.

Feature Extraction

This procedure transforms raw text data into a structured format that can be effectively used by machine learning models. During this process, both the training and testing datasets undergo feature extraction, which plays a vital role in preparing textual feedback for efficient model training and accurate evaluation. This intricate process involves converting raw, unstructured text from student

responses into a quantifiable, numerical representation that algorithms can readily interpret and process. Essentially, it is about identifying and isolating the most informative characteristics or "features" within the text that are relevant for sentiment analysis or classification tasks. At this stage, common methods involve breaking sentences into smaller units such as words or phrases through a process called tokenization, followed by applying techniques like TF-IDF (term frequency–inverse document frequency) or word embeddings to represent the textual data in a more meaningful and analyzable form (e.g., Word2Vec, GloVe) are applied to assign numerical weights or vectors to these tokens based on their importance and context. This transformation not only reduces the dimensionality of the data but also captures semantic relationships and patterns crucial for the learning model. It is imperative that both the training dataset, used to teach the model, and the test dataset, used to evaluate its performance, undergo identical feature extraction procedures to ensure consistency and prevent data leakage. By converting qualitative student feedback into a structured, numerical format, feature extraction effectively bridges the gap between human language and machine understanding, setting the foundation for accurate model training and insightful analysis.

Model Training

Building upon the foundation of feature extraction, the "training" phase is where the machine learning algorithms learn to classify the processed student feedback data. This study employs multiple robust text classification algorithms to leverage their distinct strengths. The multinomial Naïve Bayes classifier (MNBC) is utilized for its probabilistic approach, which fundamentally seeks to determine the likelihood of a text belonging to a particular class (e.g., positive, negative, neutral) by analyzing the joint probabilities of words and classes. It utilizes the frequency of words within a document as key features, making it highly suitable for text classification tasks in which word occurrences are crucial for identifying sentiment.

Another powerful algorithm is the SVM, a supervised learning model renowned for its efficacy in both linear and nonlinear classification challenges. The SVM operates by constructing an optimal hyperplane or a series of hyperplanes within a high-dimensional feature space. This hyperplane acts as a decision boundary that optimally separates the data points from different classes. When the data cannot be separated linearly in their original space, the method addresses this issue by transforming the data into higher-dimensional feature spaces, allowing for better separation. In the final stage, the RF algorithm was applied as an ensemble learning approach.

This technique constructs a collection, or "forest" of numerous decision trees, with each tree trained on a randomly chosen subset of both the data samples and the feature set. The strength of RF lies in its ability to combine the predictions from these multiple decision trees, where the final classification is determined by a majority vote or averaging. This ensemble approach significantly enhances the accuracy and robustness of the model by reducing overfitting and improving generalization, with accuracy typically increasing proportionally to the number of trees in the forest. The general steps for RF involve generating random vectors, building a multitude of diverse decision trees, and intelligently combining their outputs to derive a more stable and accurate prediction. The application of these diverse algorithms allows for a comprehensive and robust approach to classifying student feedback, potentially leading to a more reliable sentiment analysis model than previous studies.

Evaluation of the Test Data

Following the rigorous training phase, the evaluation of the test data becomes paramount to ascertain the true efficacy and generalizability of the developed machine learning model. This critical stage involves using a completely unseen dataset, the test set, which is held separate from the training data, to objectively measure how well the model performs on new, real-world examples. The primary objective was to determine the working accuracy of the model by assessing its ability to correctly classify student feedback that it had not encountered before. This evaluation provides vital insights into whether the model has truly learned the underlying patterns rather than merely memorizing the training

data, thereby determining its practical applicability and reliability in real-world scenarios. This is a crucial step that validates the model's performance and ensures its robustness beyond the controlled training environment.

Graphical Representation

Graphical Representation serves as the final, crucial step in making the insights derived from the processed feedback data accessible and understandable. By transforming complex numerical outputs into intuitive visual formats, data visualization effectively communicates the model's performance and classification results. Python's powerful Matplotlib library is specifically utilized for this purpose, enabling the generation of a wide array of customizable plots, charts, and graphs. This visual output, exemplified by Figure 2, allows stakeholders to quickly grasp trends, identify patterns, and comprehend the sentiment distribution within the student feedback, facilitating informed decision-making based on clear, visual evidence.

Performance Analysis

In this section, a comparative evaluation is conducted using three machine learning algorithms, such as MNBC, RF, and SVM, implemented with unigram and bigram feature extraction methods. The assessment focuses on two primary performance measures: accuracy and F-score. Unigrams represent single-word tokens, whereas bigrams represent two consecutive tokens. The test dataset was used to evaluate the performance of the trained models based on the selected assessment metrics.

- *Accuracy*: This refers to the ratio of the model's accurate predictions to the total number of samples in the dataset.
- *F-score*: The weighted average of recall and precision.

Figures 3 and 4 illustrate the relationship between accuracy and the number of training samples for the unigram and bigram models, respectively, while the test dataset remains unchanged. The results indicate that as the volume of the training data increases, the accuracy of the models improves.

When comparing MNBC to RF, the accuracy of MNBC showed a steady linear improvement as the amount of training data increased, whereas the accuracy of RF fluctuated and did not follow a linear trend because of the inherent randomness involved in its training process.

Figures 5 and 6 show the F-score versus the number of training samples for unigram and bigram, respectively. The F-scores of both the MNBC and SVM show a steady linear improvement as the amount of training data increases.

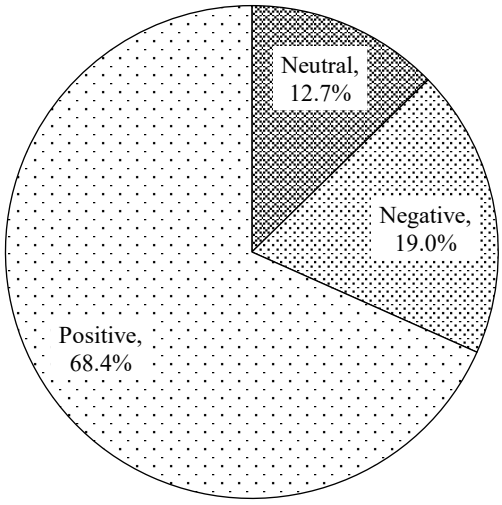

Figure 2. Graphical representation of the result.

Machine learning models that use unigrams tend to outperform those that rely on bigrams.

Figure 7 depicts the accuracy versus the number of test samples for unigrams, keeping the training data constant. The analysis shows that the accuracy of MNBC is 0.77%–1.27% better than that of RF, and SVM performs 0.78%–0.81% better than MNBC.

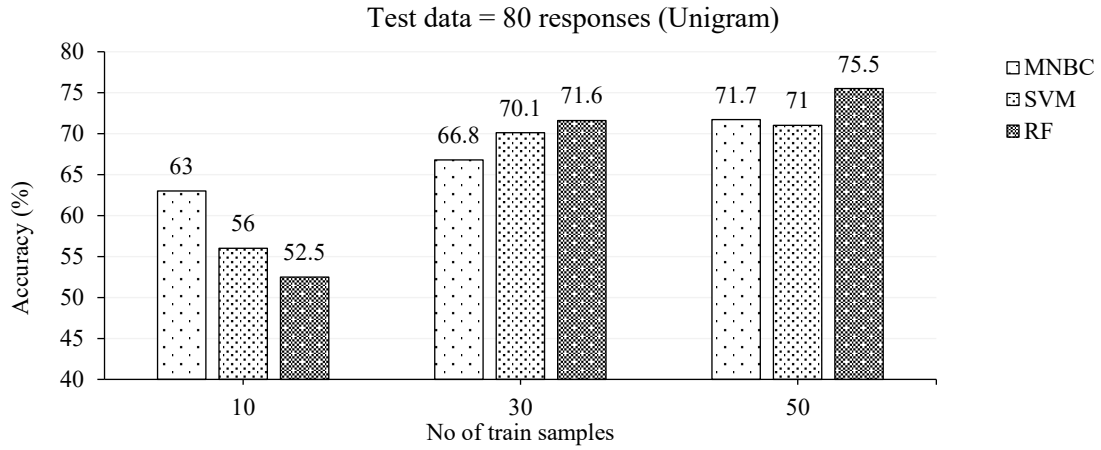

Figure 3. Plot of accuracy versus number of train samples for unigram.

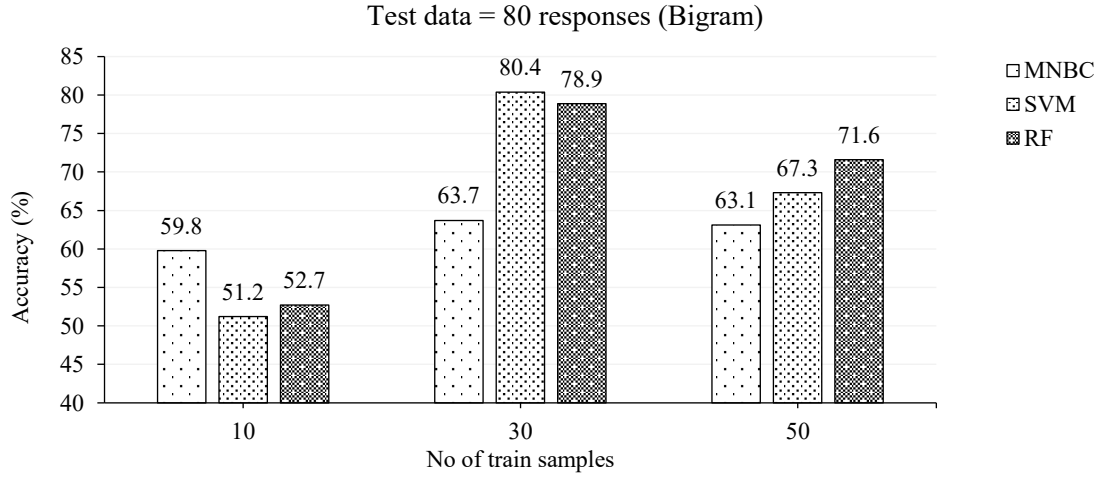

Figure 4. Plot of accuracy versus number of train samples for bigram.

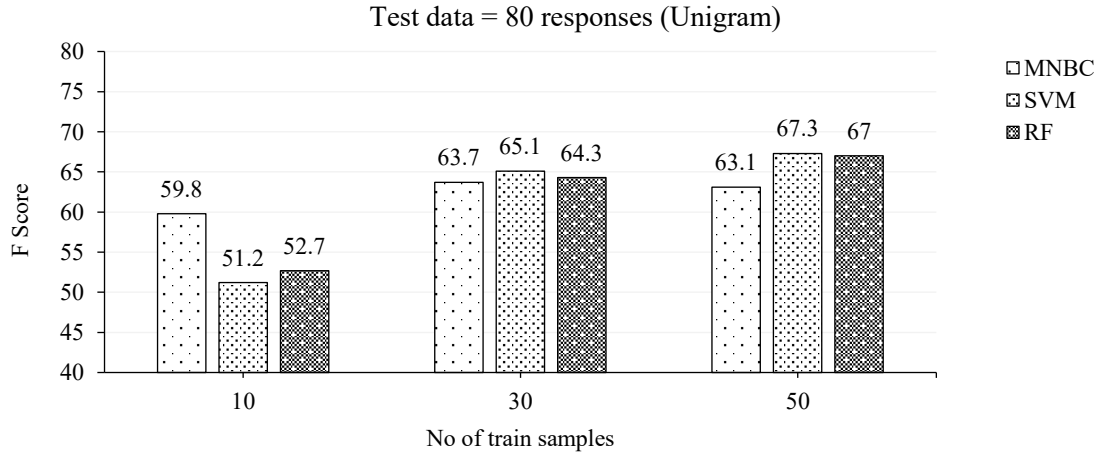

Figure 5. Plot of F-score versus number of train samples for unigram.

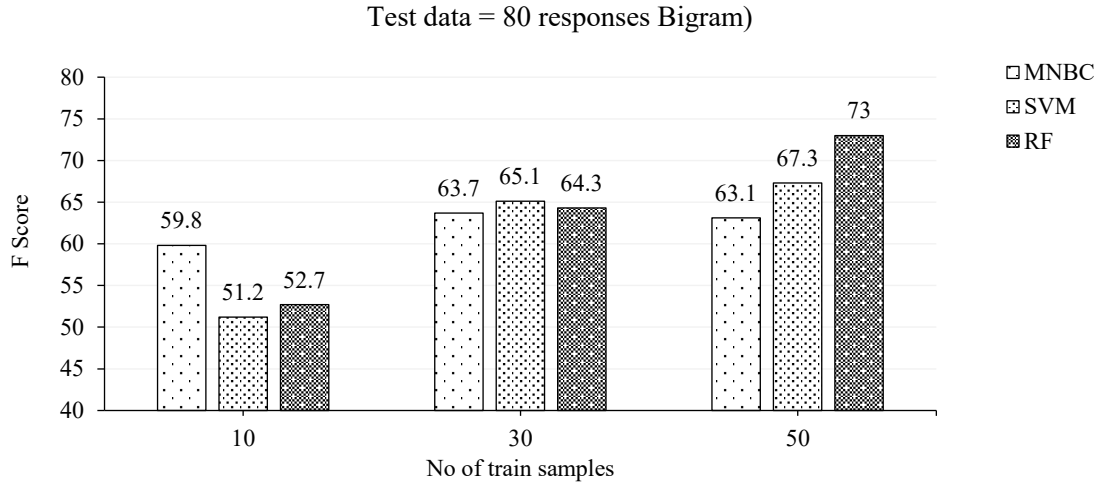

Figure 6. Plot of F-score versus number of train samples for bigram.

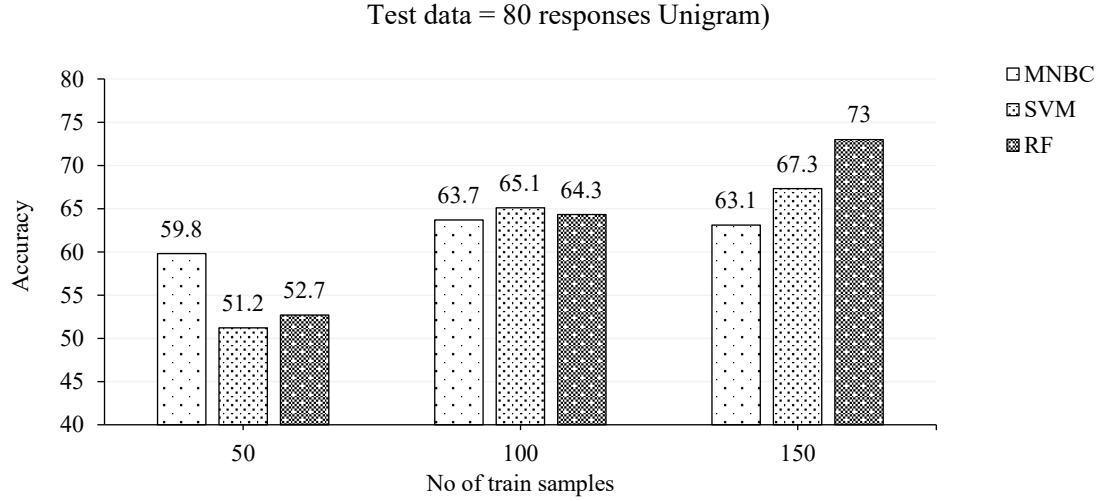

Figure 7. Plot of accuracy versus number of test samples for unigram.

CONCLUSION

The disadvantage of grading-based feedback is that it does not reveal the exact sentiment of the students; this is overcome by using textual feedback. Sentiment analysis helps institutions understand students' emotions and opinions from their feedback, enabling improvements in teaching quality and the overall campus environment. This project, utilizing feedback from Gujarat students, can also be applied to obtain students' opinions on events, workshops, and other relevant academic activities.

The results of the comparative analysis revealed that the multinomial Naïve Bayes algorithm achieved higher accuracy than both the SVM and RF algorithms. The analysis showed that multinomial Naïve Bayes achieved an accuracy of 80%, making it the most effective among the three. Future studies should focus on improving the precision of the model by employing a more extensive and diverse training dataset.

REFERENCES

1. Nasim Z, Rajput Q, Haider S. Sentiment analysis of student feedback using machine learning and lexicon-based approaches. 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), Langkawi, Malaysia. 2017. p. 1–6. doi:10.1109/ICRIIS. 2017.8002475.

2. Nanli Z, Ping Z, Weiguo LI, Meng C. Sentiment analysis: a literature review. 2012 International Symposium on Management of Technology (ISMOT), Hangzhou, China. 2012. p. 572–576. doi:10.1109/ISMOT.2012.6679538.
3. Bhavitha BK, Rodrigues AP, Chiplunkar NN. Comparative study of machine learning techniques in sentiment analysis. 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India. 2017. p. 216–221. doi:10.1109/ICICCT.2017.7975191.
4. Krishnaveni KS, Pai RR, Iyer V. Faculty rating system based on student feedback using sentiment analysis. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India. 2017. p. 1648–1653. doi:10.1109/ICACCI.2017.8126079.
5. Ullah MA. Sentiment analysis of students feedback: a study towards optimal tools. 2016 International Workshop on Computational Intelligence (IWCI), Dhaka, Bangladesh. 2016. p. 175–180. doi:10.1109/IWCI.2016.7860361.
6. Devika MD, Sunitha C, Ganesh A. Sentiment analysis: a comparative study on different approaches. Procedia Comput Sci. 2016;87:44–49. doi:10.1016/j.procs.2016.05.124.
7. Jain AP, Dandannavar P. Application of machine learning techniques to sentiment analysis. 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Bangalore, India. 2016. p. 628–632. doi:10.1109/ICATCCT.2016.7912076.
8. Baygin M. Classification of text documents based on Naive Bayes using N-gram features. 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey. 2018. p. 1–5. doi:10.1109/IDAP.2018.8620853.
9. Fang X, Zhan J. Sentiment analysis using product review data. J Big Data. 2015;2(1):5. doi:10.1186/s40537-015-0015-2.
10. Isnan M, Elwirehardja GN, Pardamean B. Sentiment analysis for TikTok review using VADER sentiment and SVM model. Procedia Comput Sci. 2023;227:168–175. doi:10.1016/j.procs.2023.10.514.
11. Mitchell TM. Machine Learning. New Delhi: McGraw-Hill Education (India); 2013.
12. Huang Y, Li L. Naive Bayes classification algorithm based on small sample set. 2011 IEEE International Conference on Cloud Computing and Intelligence Systems, Beijing, China. 2011. p. 34–39. doi:10.1109/CCIS.2011.6045027.
13. Patil HP, Atique M. Sentiment analysis for social media: a survey. 2015 2nd International Conference on Information Science and Security (ICISS), Seoul, Korea (South). 2015. 2015. p. 1–4. doi:10.1109/ICISSEC.2015.7371033.
14. Qiang G. Research and improvement for feature selection on Naive Bayes text classifier. 2010 2nd International Conference on Future Computer and Communication, Wuhan, China. 2010. pp. V2-156-V2-159. doi:10.1109/ICFCC.2010.5497362.
15. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J. 2014;5(4):1093–1113. doi:10.1016/j.asej.2014.04.011.
16. Singh J, Singh G, Singh R. Optimization of sentiment analysis using machine learning classifiers. Hum Centric Comput Inf Sci. 2017;7(1):32. doi:10.1186/s13673-017-0116-3.
17. Xu S, Li Y, Wang Z. Bayesian multinomial naïve Bayes classifier to text classification. In: Park JJ, Chen SC, Choo KKR, editors. Advanced Multimedia and Ubiquitous Engineering. MUE/FutureTech 2017. Lecture Notes in Electrical Engineering. Singapore: Springer Singapore; 2017. p. 347–352.
18. Wang Y, Hodges J, Tang B. Classification of web documents using a Naive Bayes method. Proceedings. 15th IEEE International Conference on Tools with Artificial Intelligence, Sacramento, CA, USA. 2003. p. 560–564. doi:10.1109/TAI.2003.1250241.
19. Maier J, Ferens K. Classification of English phrases and SMS text messages using Bayes and support vector machine classifiers. 2009 Canadian Conference on Electrical and Computer Engineering, St. John's, NL, Canada. 2009. p. 415–418. doi:10.1109/CCECE.2009.5090166.

20. Bouazizi M, Ohtsuki T. Sentiment analysis in Twitter: from classification to quantification of sentiments within tweets. 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA. 2016. p. 1–6. doi:10.1109/GLOCOM.2016.7842262.
21. Sharma N, Singh M. Modifying Naive Bayes classifier for multinomial text classification. 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India. 2016. p. 1–7. doi:10.1109/ICRAIE.2016.7939519.
22. Khatri SK, Srivastava A. Using sentiment analysis in prediction of stock market investment. 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India. 2016. p. 566–569. doi:10.1109/ICRITO.2016.7785019.
23. Liu B. Sentiment Analysis and Opinion Mining. Cham: Springer International Publishing; 2022. doi:10.1007/978-3-031-02145-9.
24. Pang B, Lee L. Briefly noted. Comput Linguist. 2009;35(2):311–312. doi:10.1162/coli.2009.35.2.311.
25. Aggarwal CC. Mining text data. In: Data Mining: The Textbook. Cham: Springer International Publishing; 2015. p. 429–455. doi:10.1007/978-3-319-14142-8_13.
26. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. Comput Intell. 2013;29(3):436-65. doi:10.1111/j.1467-8640.2012.00460.x.
27. Cambria E, White B. Jumping NLP curves: A review of natural language processing research. IEEE Comput Intell Mag. 2014;9(2):48-57. doi:10.1109/MCI.2014.2307227.
28. Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng AY, et al. Recursive deep models for semantic compositionality over a sentiment treebank. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA. 2013. p. 1631-42. doi:10.18653/v1/D13-1170.
29. Kazanova M. Sentiment140 dataset with 1.6 million tweets [Online]. Kaggle; 2017. Available from: https://www.kaggle.com/datasets/kazanova/sentiment140
30. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-2014), Ann Arbor, MI, USA. 2014. p. 216-25. doi:10.1609/icwsm.v8i1.14550.
31. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA. 2019. p. 4171-86.
32. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:5998-6008.
33. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space [Preprint]. 2013. arXiv:1301.3781. doi:10.48550/arXiv.1301.3781.
34. Pennington J, Socher R, Manning C. GloVe: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar. Stroudsburg, PA, USA: Association for Computational Linguistics; 2014. p. 1532-43. doi:10.3115/v1/D14-1162.
35. Maas AL, Daly RE, Pham PT, Huang D, Ng AY, Potts C. Learning word vectors for sentiment analysis. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA. Stroudsburg, PA, USA: Association for Computational Linguistics; 2011. p. 142–50.
36. Turney PD. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews [Preprint]. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, (2002), Philadelphia, Pennsylvania. 2002. p. 417–24. arXiv:cs/0212032 [cs.LG]. doi:10.48550/arXiv.cs/0212032.
37. Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques [Preprint]. 2002. arXiv:cs/0205070 [cs.CL]. doi:10.48550/arXiv.cs/0205070.
38. Baccianella S, Esuli A, Sebastiani F. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta. Paris, France: European Language Resources Association (ELRA); 2010. p. 2200–4.

39. Hu M, Liu B. Mining and summarizing customer reviews. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA. 2004. p. 168–77. doi:10.1145/1014052.1014073.
40. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau RJ. Sentiment analysis of Twitter data. In: Proceedings of the Workshop on Language in Social Media (LSM 2011), Portland, OR, USA. Stroudsburg, PA, USA: Association for Computational Linguistics; 2011. p. 30-8.
41. Omar N, Albared M, Al-Moslmi T, Al-Shabi A. A comparative study of feature selection and machine learning algorithms for Arabic sentiment classification. In: Jaafar A, et al., editors. Information Retrieval Technology. AIRS 2014. Lecture Notes in Computer Science. Vol. 8870. Cham: Springer; 2014. p. 429–43. doi:10.1007/978-3-319-12844-3_37.
42. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. WIREs Data Min Knowl Discov. 2018;8(4):e1253. doi:10.1002/widm.1253.

