
Comparing Machine Learning Algorithms for Sentiment Analysis in Movie Reviews

Mamta Bisht*, Mukul Gupta, Vinod Kharakwal, Sparsh Jain
and Shubh Verma

*Department of Computer Science and Engineering (AIML), Inderprastha
Engineering College, Ghaziabad, India*

E-mail: mamtabisht@ipeccollege.org.in

**Corresponding Author*

Received 12 February 2025; Accepted 19 March 2025

Abstract

Sentiment analysis is a crucial natural language processing component that helps businesses to understand public opinion, customer feedback, and market trends. This study evaluates the performance of three machine learning models, namely Naive Bayes, Logistic Regression and SVM in classifying the sentiment of movie reviews. We analyzed 50,000 IMDB reviews from Kaggle using TF-IDF with three ML models using the following key metrics: accuracy, precision, recall, and F1-score. We further analyzed the confusion matrix for each model, which provided insights into the number of correctly and incorrectly classified reviews, helping identify potential areas for improvement. The findings of this study can guide the selection and optimization of algorithms for sentiment analysis, ultimately improving marketing, customer engagement, and social media strategies.

Keywords: Sentiment analysis, machine learning, Naive Bayes, logistic regression, support vector machines.

Journal of Graphic Era University, Vol. 13_1, 205–220.

doi: 10.13052/jgeu0975-1416.13110

© 2025 River Publishers

1 Introduction

Sentiment analysis, which is the process of identifying emotions in text, has gained significant importance in the digital age. With the increasing volume of consumer reviews and opinions, businesses and researchers have leveraged sentiment analysis to extract insights from user feedback and to enhance decision-making processes [1]. Early sentiment analysis methods were primarily lexicon-based but have evolved with advancements in machine learning (ML) and deep learning (DL) techniques [2]. The key aspect of sentiment analysis is polarity detection (positive/negative sentiments). One crucial application of sentiment analysis is in movie reviews, where understanding audience sentiment helps inform marketing strategies, content production, and audience engagement [3].

The existing literature on sentiment analysis using ML approaches highlights the importance of feature extraction and selection techniques in achieving optimal classification performance. Extensive research has been conducted on applying various machine-learning algorithms for sentiment analysis in diverse domains, such as product reviews, social media, and news articles [4–6]. Avinash and Sivasankar [7] examined the impact of term frequency–inverse document frequency (TF-IDF) and Document to Vector (Doc2vec) feature extraction on sentiment classification accuracy and found that these methods coupled with techniques such as stop word removal and tokenization can improve classifier performance. Kabir et al. [8] conducted a comparative study across multiple ML algorithms, including Logistic Regression, SVM, and Naive Bayes, for sentiment analysis on diverse datasets, including movie reviews.

Continuing this line of research, this study focuses on comparing the performance of Naive Bayes, Logistic Regression and Support Vector Machine (SVM) classifiers for sentiment analysis in the context of movie reviews. In other words, this study aimed to evaluate three ML models to determine the most effective for classifying sentiment in movie reviews. It is useful for automating sentiment analysis, aiding businesses in understanding audience responses, and improving decision making in entertainment and marketing. Our findings indicate that SVM outperforms the other models with an accuracy of 90%, followed closely by Logistic Regression (89%). Naive Bayes performed slightly lower (86%). These insights provide valuable guidance for selecting machine learning models for sentiment analysis.

The structure of this paper is organized as follows: Section 2 provides a comprehensive explanation of the methodology employed in this study, detailing the techniques, algorithms, and approaches used to conduct the research. Section 3 describes the evaluation metrics in detail, explaining their significance, computation methods, and how they are applied to assess the performance of the proposed approach. Section 4 presents the results obtained from the experiments and includes an in-depth discussion of the findings. This section analyses the implications of the results, compares them with existing methods, and highlights key observations. Section 5 concludes the paper by summarizing the main contributions and key takeaways. Additionally, it outlines potential future research directions to further enhance and extend the study.

2 Methodology

This study utilizes the publicly available IMDB Dataset of 50 K Movie Reviews on Kaggle for sentiment analysis. The implementation was performed using Python and its relevant libraries, including scikit-learn, pandas, and matplotlib. The primary objective is to classify movie reviews as either positive or negative, based on their textual content.

The procedure begins with data pre-processing, which involves cleaning the text, tokenization, and transforming the data into numerical form using TF-IDF vectorization. Subsequently, three ML models—Naive Bayes, Logistic Regression and Support Vector Machine (SVM)—were trained using the scikit-learn library. The experiments were conducted using Google Colab with a Tesla T4 GPU and 12GB RAM, which provides cloud-based Jupyter notebooks with GPU support. The dataset was splitted into training (70%) and testing (30%) subgroups for the model evaluation.

Three widely used ML algorithms were employed for classification, as follows:

2.1 Naive Bayes

Naive Bayes assumes independence between features and often performs well in sentiment analysis tasks [9]. It is a probabilistic classifier based on Bayes' theorem and can be expressed as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where:

- $P(A/B)$ is the **posterior probability** (probability of class A, given evidence B).
- $P(B/A)$ is the **likelihood** (probability of evidence B given in Class A).
- $P(A)$ is the **prior probability** (the initial probability of class A before seeing evidence).
- $P(B)$ is the **marginal probability** (the probability of evidence occurring across all possible classes).

Naive Bayes assumes that each feature affects the class outcome independently, without depending on other features. This makes the calculations much simpler and faster. This assumption allows us to compute the probability of a class given multiple features as follows:

$$P(C|X_1, X_2, \dots, X_n) = \frac{P(C) \prod_{i=1}^n P(X_i|C)}{P(X_1, X_2, \dots, X_n)}$$

Where:

- $P(C|X_1, X_2, \dots, X_n)$ is the probability of class **C** given the features X_1, X_2, \dots, X_n .
- $P(C)$ is the prior probability of class **C**.
- $P(X_i|C)$ is the likelihood of feature X_i being given in class **C**.
- $P(X_1, X_2, \dots, X_n)$ is the overall probability that the features occur (acts as a normalizing constant).

Naive Bayes is easy to use, computationally efficient, and performs well even with small datasets. It is effective for high-dimensional data (e.g., text classification) and requires less training data than the other classifiers. With these advantages, this classifier has some disadvantages, such as the assumption that feature independence is often unrealistic and struggles with correlated features, thus reducing the accuracy [10]. Another disadvantage is that if a categorical feature value is not observed in the training data, it assigns zero probability (solved using Laplace smoothing).

2.2 Logistic Regression

In sentiment analysis of movie reviews, logistic regression is a statistical method used to classify reviews as positive or negative. It predicts the probability of a review belonging to a particular sentiment category using a

sigmoid function. Unlike linear regression, which predicts continuous values, logistic regression estimates sentiment probabilities and applies a threshold (e.g., 0.5) to classify a review as positive or negative based on the given text features [11]. The core of logistic regression is the sigmoid function, which transforms any real number into a value between 0 and 1, making it useful for probability estimation.

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Where:

- z is the linear combination of the input features and their weights:

$$z = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$$

- $\sigma(z)$ represents the probability of class
 - If $\sigma(z) \geq 0.5$, it was classified as a positive class (1).
 - If $\sigma(z) < 0.5$, it was classified as a negative class (0).
- e is Euler's number (≈ 2.718).

The output of the sigmoid function lies in the range of (0,1), making it interpretable as a probability.

Instead of using the Mean Squared Error (MSE) (which is unsuitable for classification), logistic regression uses the Log Loss (Binary Cross-Entropy) function:

$$J(w) = \frac{1}{m} \sum_{i=1}^m [y_i \log(\hat{y}_i)] + (1 - y_i) \log(1 - \hat{y}_i)$$

Where:

- m = Number of training samples
- y_i = Actual class label (0 or 1)
- \hat{y}_i = Predicted probability from the sigmoid function

The model is trained using Gradient Descent, which iteratively updates the weights to minimize the cost function.

Logistic regression is easy and efficient for binary classification tasks. It performs good with small datasets and necessitates fewer computations. The disadvantage of this classifier is that it assumes linear relationships, which may not always hold true. In addition, this method is not suitable for highly complex data (e.g., images and speech).

2.3 Support Vector Machine (SVM)

In sentiment analysis of movie reviews, Support Vector Machine (SVM) is a powerful and reliable supervised learning algorithm used to classify reviews as positive or negative. It works by finding the optimal decision boundary (hyperplane) that best separates reviews based on their sentiment-related features. SVM selects this hyperplane to maximize the margin, ensuring the greatest possible separation between the closest reviews (support vectors) of each sentiment category, leading to accurate and robust classification.

SVM is highly effective for processing high-dimensional data, making it ideal for applications like image recognition, text classification, bioinformatics, and financial forecasting [12]. It can efficiently manage both linearly and nonlinearly separable datasets. When the sentiment features are not clearly separable, SVM applies kernel functions (such as polynomial, radial basis function (RBF), and sigmoid kernels) to transform the data into a higher-dimensional space, making it easier to distinguish between positive and negative reviews.

Additionally, SVM is known for its ability to handle outliers effectively, as it focuses only on the most relevant data points (support vectors), rather than considering the entire dataset. However, the computational complexity of SVM can be high, particularly for large datasets, requiring careful parameter tuning to achieve optimal performance.

Due to its strong theoretical foundation and versatility, SVM remains a popular choice for various machine learning problems, particularly when interpretability and accuracy are key concerns.

Figure 1 presents a flow chart describing the classification process used in Naive Bayes, Logistic Regression and SVM.

GridSearchCV was used for hyperparameter tuning in all the three ML models. It systematically searches for the best combination of hyperparameters by testing multiple values and selecting the one that delivers the best performance.

3 Evaluation Metrics

The performance of the models was evaluated in terms of precision, Recall, F1-Score and Accuracy. To understand these terms, refer to Figure 2, which illustrates the concept of a confusion matrix for binary classification. In this matrix, the terms TP, FP, TN, and FN correspond to True Positive, False Positive, True Negative, and False Negative, respectively.

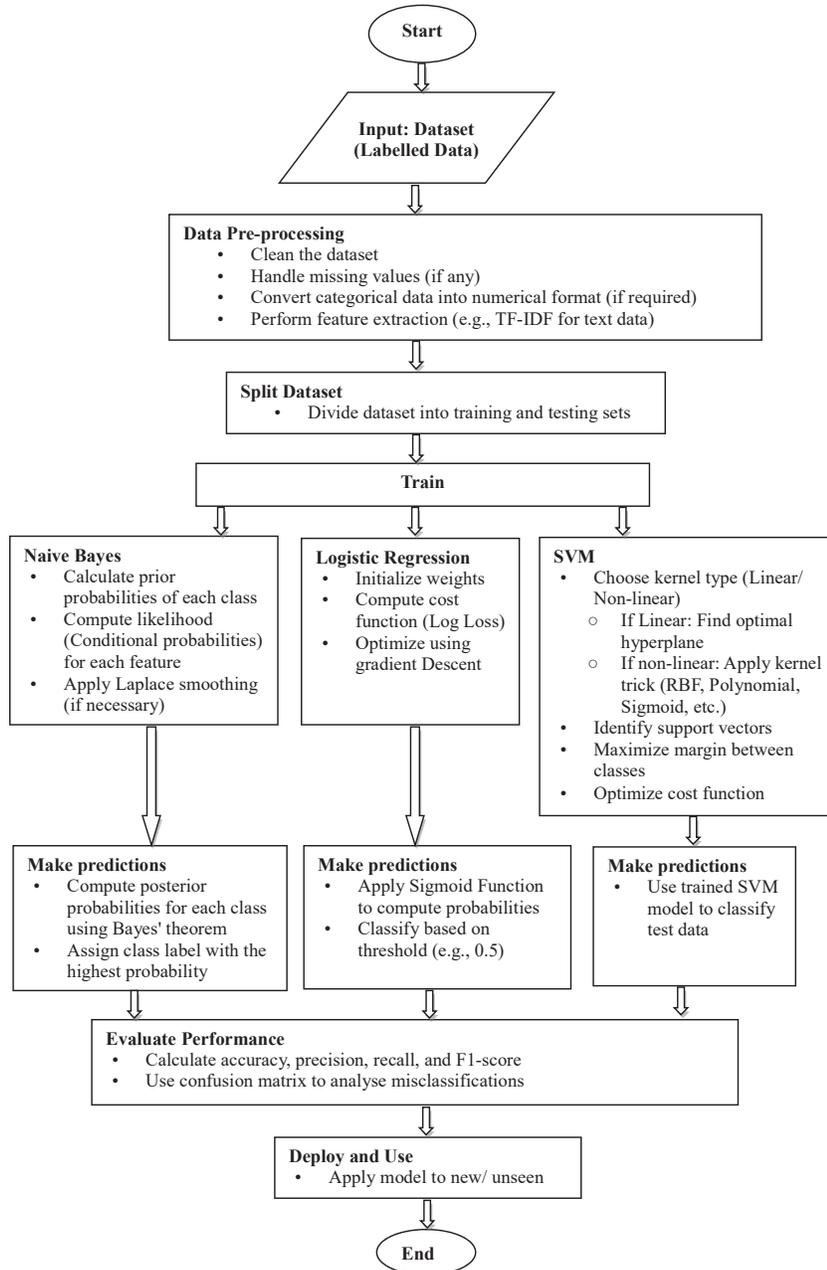


Figure 1 Flow chart describing the classification process used in Naive Bayes, Logistic Regression and SVM.

		Predicted label	
		Positive	Negative
True label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Figure 2 Confusion matrix for binary class.

The mathematical expressions for the evaluation matrices are as follows:

$$\text{Precision} = \frac{\text{True positives}}{\text{Predicted positives}} = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad (\text{The ratio of correctly predicted positive instances to the total predicted positive instances.})$$

$$\text{Recall} = \frac{\text{True positives}}{\text{All actual positives}} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad (\text{The ratio of correctly predicted positive instances to all actual positive instances.})$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{The harmonic mean of precision and recall, providing a balanced measure of classification performance.})$$

$$\text{Accuracy} = \frac{\text{Correctly classified samples}}{\text{Total samples}} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \quad (\text{The proportion of correctly classified samples.})$$

Since this study focuses on the binary classification of movie reviews as either positive or negative, these evaluation metrics are used to assess the performance of the sentiment analysis model.

4 Results and Discussion

The performance of the three machine learning models, namely Naive Bayes, Logistic Regression, and Support Vector Machine (SVM), was evaluated using key metrics: accuracy, precision, recall, and F1-score. The results summarized in Table 1 provide a comparative analysis of the effectiveness of the model in sentiment classification for movie reviews.

From these results, the SVM model achieved the highest accuracy (90%), followed closely by Logistic Regression (89%). Naive Bayes performed slightly lower (86%), which can be attributed to its assumption of feature independence, which may not hold true for natural language text data.

Table 1 Performance comparison of models

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.86	0.84	0.89	0.86
Logistic Regression	0.89	0.91	0.89	0.89
SVM	0.90	0.91	0.89	0.89

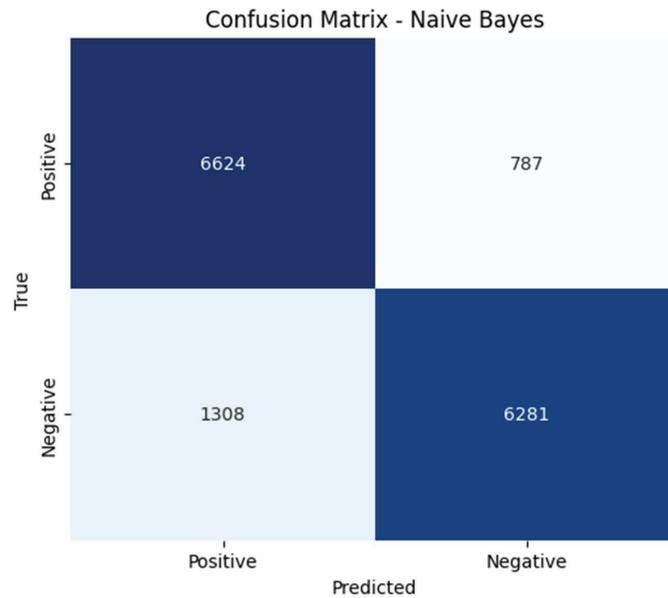


Figure 3 Confusion matrix for Naive Bayes.

Both logistic Regression and SVM achieved the highest precision (0.91), indicating that they were better at minimizing false positives (incorrectly classifying a negative review as positive). Naive Bayes had a slightly lower precision (0.84), suggesting that it misclassified a few negative reviews as positive reviews.

Naive Bayes achieved the highest recall (0.89), indicating that it effectively captured the most positive reviews. Logistic Regression and SVM also attained a recall of 0.89, meaning they performed comparably well in identifying actual positive instances.

The F1-score, which balances precision and recall, is highest for Logistic Regression and SVM (0.89), while Naive Bayes scored slightly lower at 0.86.

These findings highlight the trade-offs between precision and recall for different models. Although Naive Bayes is more recall-oriented, Logistic

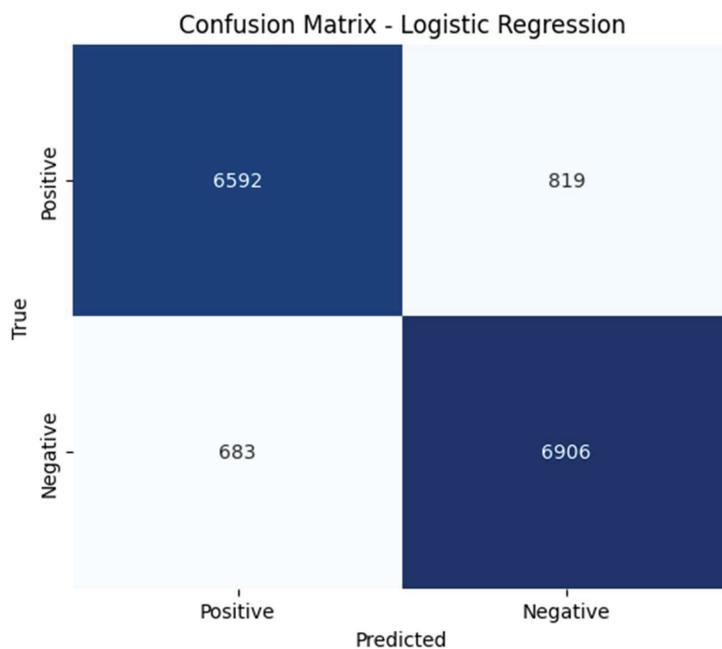


Figure 4 Confusion matrix for logistic regression.

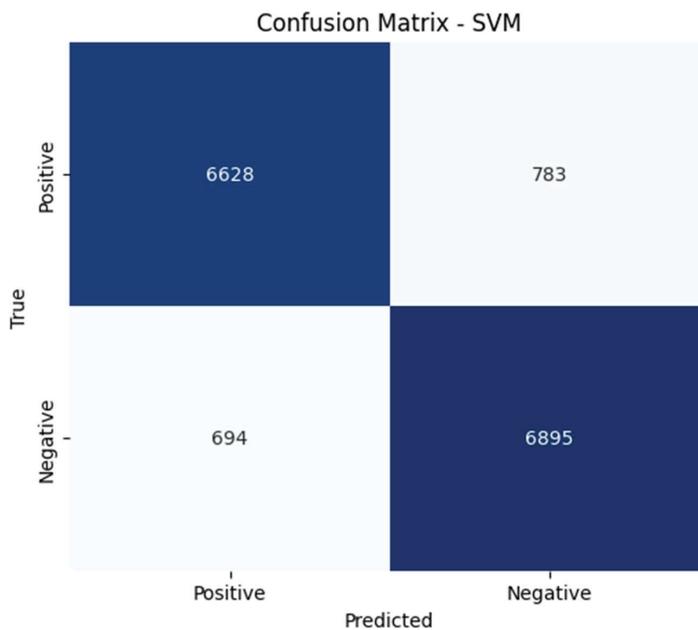


Figure 5 Confusion matrix for SVM.

Table 2 Performance comparison in term of overall accuracy (%) with other models

Authors	Dataset	Algorithms Gives Best Accuracy	Accuracy
Avinash & Sivasankar (2019) [7]	five different datasets of varying sizes	TF-IDF and Doc2Vec	85%
Kabir et al. (2021) [8]	Amazon review data, Yelp review data and IMDB review data	Maximum Entropy algorithm	94% ,97% and 91% respectively
Proposed	Kaggle	SVM	90%

Regression and SVM offer a better balance, making them more reliable choices for sentiment classification tasks.

To further analyze the model performance, the confusion matrices for each model are examined in Figures 3, 4, and 5. These matrices provide insights into the number of correctly and incorrectly classified reviews, helping identify potential areas for improvement.

The results indicate that the SVM and Logistic Regression outperform Naive Bayes in terms of precision and overall accuracy, making them more suitable for sentiment classification.

A performance comparison in term of overall accuracy (%) with other models is presented in Table 2.

5 Conclusion

This paper presents a comparative evaluation of three prominent machine learning algorithms – Naive Bayes, Logistic Regression, and Support Vector Machine in the context of sentiment analysis on a dataset of movie reviews. This study demonstrates that SVM achieves the best accuracy (90%), followed closely by Logistic Regression (89%), while Naive Bayes performed slightly lower (86%). The analysis of confusion matrices provides deeper insights into classification errors, revealing that Naive Bayes has a higher false positive rate, whereas SVM is the most precise model. While this work has shown promising results, there are still some limitations. The accuracy could be improved, especially in handling complex sentiment patterns. Additionally, the model's performance may vary depending on the dataset and feature selection. Future work can focus on hyperparameter tuning and feature engineering to further enhance the classification performance.

Future research can also focus on deep learning models, such as LSTMs and transformers, to improve sentiment classification.

References

- [1] Q. A. Xu, V. Chang, and C. Jayne, “A systematic review of social media-based sentiment analysis: Emerging trends and challenges,” *Decision Analytics Journal*, vol. 3, p. 100073, 2022.
- [2] C. Sahoo, M. Wankhade, and B. K. Singh, “Sentiment analysis using deep learning techniques: a comprehensive review,” *Int J Multimed Info Retr*, vol. 12, no. 2, p. 41, Dec. 2023.
- [3] Md. S. Islam et al., “Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach,” *Artif Intell Rev*, vol. 57, no. 3, p. 62, Mar. 2024.
- [4] R. B. Shamantha, S. M. Shetty, and P. Rai, “Sentiment analysis using machine learning classifiers: evaluation of performance,” in *2019 IEEE 4th international conference on computer and communication systems (ICCCS)*, IEEE, 2019, pp. 21–25.
- [5] M. D. Devika, C. Sunitha, and A. Ganesh, “Sentiment analysis: a comparative study on different approaches,” *Procedia Computer Science*, vol. 87, pp. 44–49, 2016.
- [6] M. Dorothy and S. Rajini, “The Various Approaches for Sentiment Analysis: A Survey,” *vol*, vol. 5, pp. 2014–2016, 2016.
- [7] M. Avinash and E. Sivasankar, “A Study of Feature Extraction Techniques for Sentiment Analysis,” in *Emerging Technologies in Data Mining and Information Security*, vol. 814, A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, Eds., in *Advances in Intelligent Systems and Computing*, vol. 814, Singapore: Springer Singapore, 2019, pp. 475–486.
- [8] M. Kabir, M. Md. J. Kabir, S. Xu, and B. Badhon, “An empirical research on sentiment analysis using machine learning approaches,” *International Journal of Computers and Applications*, vol. 43, no. 10, pp. 1011–1019, Nov. 2021.
- [9] P. P. Surya and B. Subbulakshmi, “Sentimental analysis using Naive Bayes classifier,” in *2019 International conference on vision towards emerging trends in communication and networking (ViTECoN)*, IEEE, 2019, pp. 1–5.

- [10] P. B. Pajila, B. G. Sheena, A. Gayathri, J. Aswini, and M. Nalini, “A comprehensive survey on naive bayes algorithm: Advantages, limitations and applications,” in *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, 2023, pp. 1228–1234. Accessed: Mar. 08, 2025.
- [11] R. Swathi, A. Sri, and P. Roshini, “Sentiment Classification of Movie Reviews with Logistic Regression,” in *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, IEEE, 2024, pp. 1534–1538.
- [12] D. N. Devi, C. K. Kumar, and S. Prasad, “A feature based approach for sentiment analysis by using support vector machine,” in *2016 IEEE 6th international conference on advanced computing (IACC)*, IEEE, 2016, pp. 3–8.

Biographies



Mamta Bisht is a passionate researcher and educator in Artificial Intelligence, Machine Learning, and Digital Communication. She earned her B.Tech. in Electronics and Communication Engineering from H.N.B. Garhwal University (A Central University, Srinagar, Uttarakhand) in 2011 and an M. Tech. in Digital Communication from B.T.K.I.T. Dwarahat (An Autonomous Institute of the Government of Uttarakhand) in 2015. She has pursued a Ph.D. in Electronics and Communication Engineering at Jaypee Institute of Information Technology, Noida, which she successfully completed in 2023. Her doctoral research, “Handwritten Text Recognition for Devanagari Script using Deep Learning Models,” explores innovative solutions in pattern recognition and AI-driven linguistic processing. Her research contributions are featured in reputed indexed journals and conferences. She actively explores cutting-edge developments in AI, Image Processing,

Pattern Recognition, Deep Learning, Signal Processing, and Communication. Currently, she serves as an Assistant Professor in the Department of Computer Science and Engineering (AI & ML) at Indraprastha Engineering College, Ghaziabad.



Mukul Gupta is a final-year B.Tech student in the Department of Computer Science and Engineering (AI & ML) at Indraprastha Engineering College, Ghaziabad. His academic interests include artificial intelligence, machine learning, and software development. He has worked on projects involving front-end development, API integration, and sentiment analysis. He is passionate about exploring new technologies and applying them to real-world problems.



Vinod Kharkwal is a final-year B.Tech student in the Department of Computer Science and Engineering (AI & ML) at Indraprastha Engineering College, Ghaziabad. He has strong interest in artificial intelligence, machine learning, and cloud computing. Over time, he has worked on various projects, including front-end development, Android applications, and sentiment analysis. He enjoys discovering emerging technologies and leveraging them to

solve practical challenges. His key focus areas include artificial intelligence, cybersecurity, and cloud computing.



Sparsh Jain is a final-year B.Tech student in the Department of Computer Science and Engineering (AI & ML) at Indraprastha Engineering College, Ghaziabad. His academic interests include artificial intelligence, cloud computing, and cybersecurity. He has worked on projects involving Web development, machine learning, or network security. He is passionate about leveraging technology to solve real-world challenges and is continuously exploring new advancements in his field.



Shubh Verma is a graduating B.Tech student specializing in Artificial Intelligence and Machine Learning at Indraprastha Engineering College, Ghaziabad. He is deeply interested in data analysis. He has hands-on experience in projects involving data analysis, front-end development, and sentiment analysis. Shubh is enthusiastic about adopting cutting-edge technologies to create impactful solutions and is always looking to refine his technical proficiency.

