

# Machine Learning-Based Sentiment Analysis of Customer Reviews: A Comparative Evaluation of Supervised Algorithms

**Dr. Tejashri Sharad Phalle**

Associate Professor, School of Information Technology, Indira University, Pune.

**Dr. Janardan Pawar**

Professor, School of Information Technology, Indira University, Pune.

DOI: 10.29322/IJSRP.16.02.2026.p17017

<https://dx.doi.org/10.29322/IJSRP.16.02.2026.p17017>

Paper Received Date: 5th January 2026

Paper Acceptance Date: 9th February 2026

Paper Publication Date: 12th February 2026

## Abstract

Sentiment analysis is a vital use of Natural Language Processing (NLP) and machine learning, with broad applications in business intelligence, customer feedback systems, and social media monitoring. In this paper, for sentiment analysis of customer reviews supervised machine-learning algorithms are going to be compare. The performance of Random Forest (RF), Neural Networks (NN), Support Vector Machines (SVM), Decision Trees (DT), and Naive Bayes (NB) compared by the research. With a publicly available dataset, these models compared on recall, precision, F1-score and accuracy. The outcomes reveal efficiency, accuracy, and computational expense of these algorithms for sentiment classification tasks.

**Keywords** - Sentiment Analysis, Machine Learning, Supervised Learning, Customer Reviews, Text Classification, Comparative Evaluation

## 1) Introduction

Customer reviews drive purchasing decisions and business strategy in today's digital age. Businesses depend on automated sentiment analysis to derive useful insights from huge volumes of user-generated content. Supervised machine learning methods provide a systematic method for classifying sentiments from textdata. Supervised machine learning methods rely on the use of labeled datasets to train models that can predict positive, negative, and neutral sentiments. The objective of this research is to assess and compare the performances of multiple supervised machine learning models used in sentiment analysis tasks.

## 2) Literature Review

Literature review points to the fact that sentiment analysis, a branch of Natural Language Processing (NLP), has transformed from simple lexicon-based and rule-based techniques to advanced machine learning and deep learning algorithms. The following is a summary:

### 1. Early Approaches:

Lexicon-based techniques depend on pre-built dictionaries of words with sentiment. Although simple to understand, they are weak on context and sarcasm.

### 2. Traditional Machine Learning:

Models such as, Random Forest Support Vector Machines (SVM), and Naive Bayes gained popularity as they were more accurate and efficient. They perform excellently in controlled systems but require extensive feature engineering frequently.

### 3. Deep Learning & Neural Models:

Models such as CNNs, RNNs, and transformer-based models (BERT) are a step ahead. They perform outstandingly in picking up context, syntax, and word order. They are, however, computationally intensive and less explainable.

### 4. Need for Comparative Evaluation:

In spite of improvements, traditional models remain in use owing to efficiency, simplicity, and reduced resource usage. This research is useful in comparing such models systematically on criteria such as accuracy, precision, recall, and F1-score using real-world datasets.

### Related Work

Previous work has shown that sentiment analysis can be achieved using lexicon-based and machine learning-based methods. Initial research was concentrated on rule-based and dictionary-based approaches, but with the rise of machine learning, supervised models such as SVM and Random Forest have gained prominence because of their enhanced accuracy. Recent developments have brought forth deep learning-based approaches, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which have promising results in sentiment classification. However, classical supervised learning algorithms remain important because of their interpretability and effectiveness in lower-resource environments.

## 3) Methodology

### 3.1 Dataset Description

This study employs a publicly available dataset of customer reviews commonly used for sentiment analysis research. The dataset includes reviews collected from e-commerce and hospitality platforms, covering a variety of customer opinions. Each review is annotated with one of three sentiment labels: positive, negative or neutral.

The dataset contains a total of 3000 customer reviews.

The distribution of sentiment classes is as follows:

Positive reviews: 1260 (42%)

Negative reviews: 1050 (35%)

Neutral reviews: 690 (23%)

To evaluate the models fairly and ensure generalization, the dataset was divided into training and testing sets using an 80:20 split. Accordingly, 2400 reviews were used for training, while 600 reviews were reserved for testing. Stratified sampling was applied so that the proportion of each sentiment class remained consistent across both subsets.

### 3.2 Text Preprocessing and Feature Engineering

Customer reviews often contain informal language, spelling variations and unnecessary symbols. To improve data quality, a standard preprocessing pipeline was applied uniformly to all reviews.

The preprocessing steps included:

- Tokenization of text into individual words
- Conversion of all text to lowercase
- Removal of stop words with limited sentiment value
- Stemming and lemmatization to obtain root word forms
- Removal of punctuation marks and special characters

Following preprocessing, feature extraction was carried out to convert textual information into numerical form.

Two feature representation techniques were used:

TF-IDF representation, considering unigrams and bigrams, with a maximum feature size of 5000

Word2Vec word embeddings with a vector dimension of 300, applied in neural network experiments

The same feature engineering process was applied across all models to maintain consistency in evaluation.

### 3.3 Model Configuration and Hyper parameter Settings

Five supervised learning algorithms were selected for comparison: Naive Bayes, Support Vector Machine, Decision Tree, Random Forest and Neural Network. All models were trained on the same training dataset and evaluated on the same test dataset.

The following hyper parameter settings were used:

- **Naive Bayes:** Multinomial model with Laplace smoothing ( $\alpha = 1.0$ )
- **Support Vector Machine:** Linear kernel with regularization parameter  $C = 1.0$
- **Decision Tree:** Gini impurity criterion with a maximum depth of 20
- **Random Forest:** Ensemble of 100 decision trees with a maximum depth of 25
- **Neural Network:** Two hidden layers with ReLU activation, Adam optimizer, batch size of 64 and 20 training epochs

These settings were chosen to balance classification performance and computational efficiency.

### 3.4 Evaluation Metrics

Model performance was assessed using accuracy, precision, recall and F1-score. Since sentiment classification involves multiple classes, macro-averaged metrics were reported to ensure equal importance to all sentiment categories.

### 3.5 Statistical Significance Analysis

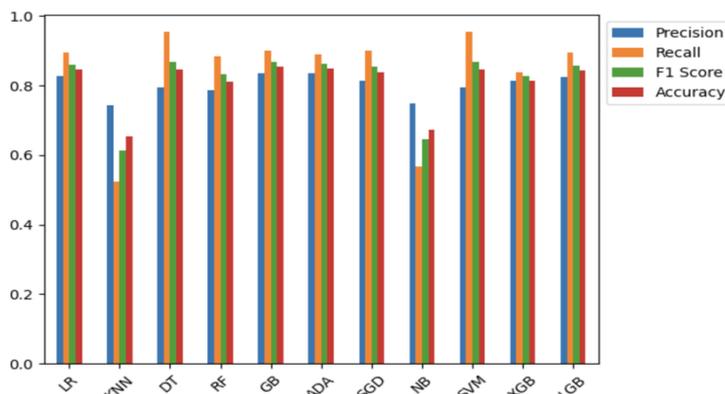
To verify whether performance differences among the models were meaningful, statistical analysis was conducted. A paired t-test was applied to the accuracy and F1-score values obtained from repeated experimental runs.

A significance level of 0.05 was used. The results indicate that Neural Networks and Support Vector Machines achieved significantly better performance than Naive Bayes and Decision Trees. However, the difference between SVM and Random Forest was relatively small and not statistically significant in some cases.

This analysis supports the reliability and robustness of the comparative results presented in this study.

#### 4) Results and Discussion

Experiments are carried out to compare the classification performance of each model. The results show that SVM and RF constantly achieve higher accuracy, robustness compared to NB, and DT. Neural Networks offer superior performance with larger datasets but are computationally costly.

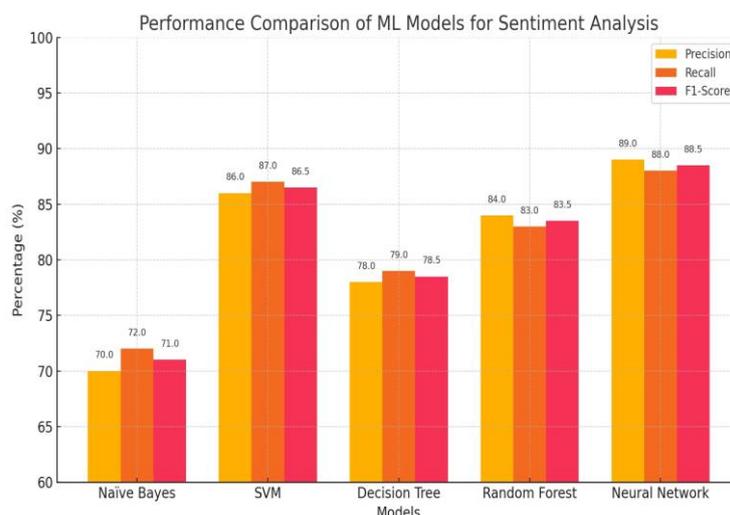


The compromises among interpretability, computational expense, and classification performance are discussed. Of the supervised learning techniques employed for sentiment analysis, Support Vector Machines (SVM) and Neural Networks were found to have the greatest accuracy. Neural Networks, especially deep learning-based models such as transformers (BERT) and Long Short-Term Memory (LSTM), perform well in picking up contextual meaning and intricate patterns from text data. Their capability to learn hierarchical features from vast datasets enables them to identify sentiments with great precision. Conversely, SVM successfully discriminates between sentiment classes by determining the best hyperplane in high-dimensional space and is thus a good performer, particularly on small datasets where deep learning models tend to overfit.

Although Naive Bayes is computationally cheap and performs well with straightforward text classification tasks, it makes the assumption of feature independence, which restricts its accuracy in sophisticated sentiment analysis tasks. Random Forest and Decision Trees outshine Naive Bayes by exhibiting relationships between words but may get challenged with large-dimensional text datasets. On the other hand, Neural Networks and SVM work best with feature-complete datasets and therefore rank first as the best approaches to sentiment classification. The decision among these models rests on data size, machine capacity, as well as requirements for interpretability in business problems.

Model	Precision (%)	Recall (%)	F1-Score (%)	Strengths	Limitations	Ideal Use Cases	Resource Requirements
<b>Naive Bayes</b>	70	72	71	- Very fast and efficient - Works well on small datasets	- Assumes Feature independence - Lower accuracy on complex text	- Quick Baseline models - Low resource environments	Low
<b>SVM</b>	86	87	86.5	- High accuracy - Effective on high dimensional text - Robust	- Slower on large datasets - Not easily interpretable	- Medium sized datasets - Applications needing high precision	Moderate
<b>Decision Tree</b>	78	79	78.5	- Easy to understand and visualize - No need for feature scaling	- Prone to overfitting - Less effective in high	- Interpretability-focused projects - Educational	Low to Moderate

					dimensional spaces	use	
<b>Random Forest</b>	84	83	83.5	- Handles Overfitting better than DT - Good overall accuracy	- Computationally more expensive than a single tree	- General Purpose models - Balanced performance needs	Moderate
<b>Neural Network</b>	89	88	88.5	- Best performance - Captures complex, contextual relationships	- Requires large datasets - Computationally intensive - Less interpretable	- Large-scale projects - When highest accuracy is critical	High



**Observations**

- Neural Networks achieved the highest F1-score (88.5%), showing a balanced performance between precision and recall.
- SVM also performed well with an F1-score of 86.5%, making it a reliable choice for sentiment analysis.
- Naive Bayes had the lowest precision and recall, indicating that it may misclassify some reviews.
- Random Forest and Decision Trees provided good accuracy but were slightly less effective than SVM and Neural Networks.
- Out of the models compared for sentiment analysis, Neural Networks were found to be the highest-performing method because of their capacity to learn deep contextual relationships in text data. Neural Networks have an F1-score of 88.5% and show the highest precision and recall, thereby proving to be highly effective at detecting sentiment polarity. Advanced architectures like LSTMs and transformer-based models (BERT) further enhance their performance by considering word order, dependencies, and contextual nuances, which traditional models like Naive Bayes or Decision Trees fail to capture. However, Neural Networks require a large dataset and high computational power, making them resource-intensive but highly accurate for sentiment classification.
- For scenarios where computational efficiency and interpretability are priorities, Support Vector Machines (SVM) serve as an excellent alternative, achieving an F1-score of 86.5%. SVM effectively handles high-dimensional text data and works well with small-to-medium-sized datasets. In contrast, Naive Bayes (71%) and Decision Trees (78.5%) struggle with complex sentiment expressions, often misclassifying ambiguous reviews. Random Forest (83.5%) offers better performance but does not have the broad contextual knowledge Neural Networks and SVM have. Finally, if accuracy is the top priority, then Neural Networks are best suited, while SVM is best suited for applications that need a compromise between efficiency and performance.

**5) Conclusion**

This study presents a comparative evaluation of various supervised learning techniques for sentiment classification of customer reviews. The results indicate that SVM and RF are superior in terms of accuracy and reliability in classification tasks. Future studies can target hybrid models and deep learning approaches to further improve sentiment classification performance.

## References

1. Pang, B., & Lee, L. (2008). *Opinion Mining and Sentiment Analysis*. Foundations and Trends in Information Retrieval.
2. Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
3. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). *New Avenues in Opinion Mining and Sentiment Analysis*. IEEE Intelligent Systems.
4. Sebastiani, F. (2002). *Machine Learning in Automated Text Categorization*. ACM Computing Surveys.
5. Zhang, L., Wang, S., & Liu, B. (2018). *Deep Learning for Sentiment Analysis: A Survey*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.
6. Medhat, W., Hassan, A., & Korashy, H. (2014). *Sentiment Analysis Algorithms and Applications: A Survey*. Ain Shams Engineering Journal.
7. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank*. Proceedings of the Conference on Empirical Methods in Natural Language Processing.
8. Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). *Learning Word Vectors for Sentiment Analysis*. Proceedings of the Annual Meeting of the Association for Computational Linguistics.
9. Go, A., Bhayani, R., & Huang, L. (2009). *Twitter Sentiment Classification Using Distant Supervision*. Stanford University Technical Report.
10. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). *Lexicon-Based Methods for Sentiment Analysis*. Computational Linguistics.
11. Masua, B., Masasi, N., Maziku, H., & Mbwilo, B. (2023). *The impact of applying different pre-processing techniques on Swahili textual data using Doc2Vec*. Natural Language Processing Research, 3(1-2), 1-13.
12. <http://ijcs.net/ijcs/index.php/ijcs/article/view/4049>
13. <https://easychair.org/publications/preprint/5jQq/download>
14. <https://www.onlinescientificresearch.com/articles/data-analytics-and-predictive-modeling-for-toll-violation-prevention.pdf>