INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

# To Study and Analyze Sentiment Analysis of Customer Reviews Using Natural Language Processing Techniques

**Reshma Masurekar*, Deepashree Mehendale, Sonali Nemade, Ashwini Patil**

**Department of Computer Science, Dr. D. Y. Patil Arts, Commerce and Science College, Pimpri, Pune-18, Maharashtra, India**

**DOI: https://doi.org/10.51583/IJLTEMAS.2025.1413SP018**

**Received: 26 June 2025; Accepted: 30 June 2025; Published: 23 October 2025**

**Abstract:** Customer reviews are very important in today's digital world for influencing potential customers and for establishing brand perception. A Natural Language Processing (NLP) technique called sentiment analysis makes it possible to automatically read textual opinions and identify whether they are very negative, negative, neutral, positive and very positive. This study explores the use of machine learning algorithms and a variety of natural language processing techniques for sentiment analysis of customer evaluation. Text preprocessing, vectorization, model training, and performance assessment using metrics like accuracy, precision, recall, and F1-score are all included in the study. The findings show that when it comes to understanding contextual sentiment in customer evaluations, deep learning model, particularly LSTM perform better than conventional machine learning models.

**Keywords:** Sentiment Analysis, Customer Reviews, Natural Language Processing (NLP), Text Processing, Machine Learning, Deep Learning and LSTM.

## I. Introduction

In the modern digital age, user-generated content has increased at an unprecedented rate due to the quick growth of social media, online platforms, and e-commerce. Customer reviews are amongst the most valuable formats. The market dynamics, company reputation, and customer behavior are all significantly impacted by these reviews. Textual feedback from customers frequently conveys their thoughts, experiences, and degree of satisfaction and this information is very helpful to companies looking to enhance their goods and services.

The main goal of this study is to investigate several NLP-based methods for conducting sentiment analysis on customer evaluation. The entire process of data preparation, feature extraction, model training, evaluation and result comparison using several classification models is covered in the study. More complex deep learning models like Long Short-Term Memory (LSTM) networks are assessed with conventional machine learning techniques like Support Vector Machines, Naive Bayes and Logistic Regression.

This main objective of the study is to determine which approaches and models offer the most precise and contextually sensitive sentiment classification in customer reviews. By using these strategies, companies may improve customer service, track brand impression, automate the process of processing customer input and ultimately make better and informed decisions.

In order to understand the current situation of sentiment analysis approaches today, we start this study with a survey of relevant research. After that, we discuss methodology, which includes gathering data, preprocessing, feature extraction techniques and creating machine learning models. Our analysis findings are shown with a comparison of the models performance. We conclude up by outlining our findings, constraints and future research directions.

### Problem Statement:

In today's digital world, online reviews provide firms with a wealth of customer input. Although the human analysis is challenging due to the large volume and unstructured nature of textual data Conventional techniques provide a method for analyzing customer feedback. The emotional tone, meaning, and minute details that are inherent in human language are not adequately captured by traditional methods. In order to support decision-making, marketing strategies and customer relationship management, there is a rising need for automated systems that can precisely evaluate and categorize client sentiments.

The purpose of this study is to investigate and evaluate the application of Natural Language Processing (NLP) techniques to automatically recognize and categorize the attitudes conveyed in customer evaluations. The objective is to create and evaluate different deep learning and machine learning models that can categorize evaluations as neutral, negative or beneficial. Preprocessing unstructured text input, identifying significant features, training sentiment classification models and assessing the models' effectiveness with suitable metrics are all part of this process.

### Objective:

To clean and organize the unstructured textual data by performing text preparation on customer reviews using NLP techniques like tokenization, stop word removal, lemmatization, and normalization.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

- To apply and train a variety of deep learning and machine learning models including LSTM, Random Forest, SVM, Naive Bayes, and Logistic Regression for sentiment categorization.
- To determine which model is best for sentiment analysis by assessing the models' performance using criteria such as accuracy, precision, recall, F1-score, and confusion matrix.

**II. Methodology:**

A labeled dataset of customer feedback was used in this study on sentiment analysis of customer evaluations using Natural Language Processing (NLP) techniques. In order to train and assess machine learning models for sentiment categorization, the dataset is essential. It allows for supervised learning by combining text reviews with matching customer ratings.

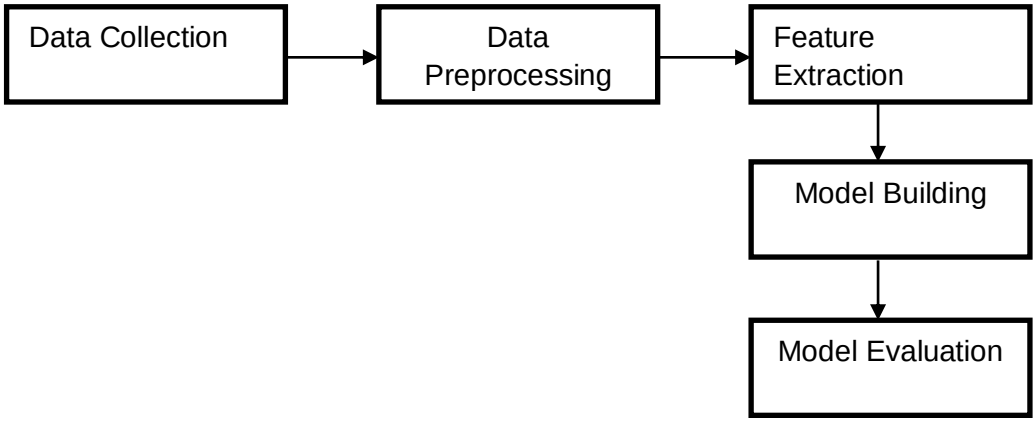
The dataset for study was taken from Kaggle Dataset which included customer review samples in this file. A unique identification (Id), the actual customer review (Review), and the customer's numerical rating (Rating) are all included in each entry. The goal variable for supervised learning is the rating values, which vary from 1 to 5.

- Ratings of **1** typically indicate very **negative sentiment**.
- Ratings of **2** typically indicate **negative sentiment**.
- A rating of **3** indicates **neutral sentiment**.
- Ratings of **4** represent **positive sentiment**.
- Ratings of **5** represent very **positive sentiment**.

These labeled examples are used to train and validate the sentiment classification models by providing ground truth for learning algorithms. To maintain consistency with the training data's format, every review in the test set contains an Id and a Review column. A sample format for submitting the test set's projected results is provided in the sample_submission.csv file. Id and Rating are its two columns. The Rating is the expected rating output from the sentiment classification algorithm, and the Id is the review's unique identification in the test set.

This study's methodology aims to investigate and assess different machine learning and natural language processing (NLP) approaches for doing sentiment analysis on customer reviews. Data gathering, data preparation, feature extraction, model creation, and evaluation are some of the methodical stages involved. For the sentiment classification process to be accurate and dependable, each step is necessary.

Figure1: Sentiment Analysis Methodology




Customer reviews frequently provide noisy, unpredictable, and unstructured raw textual data. A number of preprocessing procedures are used to clean and normalize the data in order to guarantee the efficacy and precision of the sentiment classification models. Text cleaning is the initial phase, which includes eliminating HTML tags, punctuation, special characters, and numeric values that don't significantly add to sentiment. In order to preserve consistency and prevent handling the same term differently because of case sensitivity, all text is then transformed to lowercase. Tokenization, which divides each sentence into discrete words or tokens for improved manipulation and analysis, comes next. To cut down on noise and enhance the model's focus on key phrases, stop words commonly used words like "is," "the," and "in" that don't have much semantic weight are then eliminated. The next step is lemmatization, also known as stemming, which groups many grammatical forms of a word by reducing it to its base or root form. Lastly, short, meaningless tokens and excess whitespace are removed by noise removal, producing a clean, organized text dataset that is prepared for feature extraction and model training.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Numerical input is necessary for machine learning algorithms to function properly; hence feature extraction techniques are needed to convert raw textual data into an appropriate numerical representation. Several popular techniques are employed in this study to turn text into feature vectors. Semantically related words are grouped together in dense, continuous vector spaces created by these mapping techniques.

Several machine learning and deep learning models were created, trained, and assessed in this work in order to categorize sentiments from customer evaluations. Every model contributes unique strengths to the sentiment analysis problem. Because of its simplicity of use and interpretability, the popular linear model known as logistic regression works well for binary classification issues and provides an adequate baseline. Because it performs well on high-dimensional data, such as word vectors, Naive Bayes, which is based on Bayes' Theorem, is especially well-suited for text classification applications. Both linear and non-linear boundaries can be handled well by using the Support Vector Machine (SVM) model, which finds the appropriate hyperplane to divide the sentiment classes.

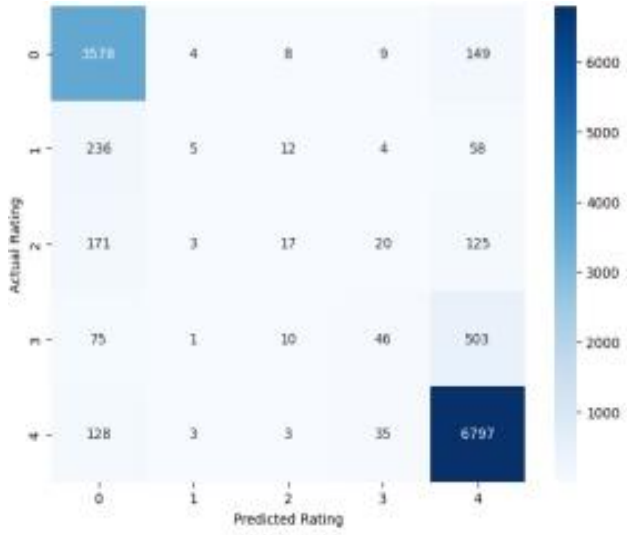
Several common evaluation measures are used to assess the sentiment categorization models' effectiveness and performance. One of the main metrics is accuracy, which shows what proportion of reviews the model properly anticipated in terms of sentiment. Accuracy gives an overall impression of performance, but when datasets are unbalanced, it can be deceptive. As a result, other parameters like recall and precision are also taken into account.

**III. Result Analysis and Performance:**

To evaluate the effectiveness of several machine learning and deep learning models in sentiment categorization, their performance was assessed on a dataset of customer reviews. Among the models used are Random Forest, Naive Bayes, Support Vector Machine (SVM), Logistic Regression. A preprocessed dataset was used to train each model, and common performance metrics like accuracy, precision, recall, F1-score, and confusion matrix were used to assess each model.

The models' great accuracy in forecasting extreme attitudes like "very positive" (rating 4 or 5) and "very negative" (rating 0 or 1) was demonstrated by the confusion matrix analysis. They had trouble with neutral or mid-range sentiments, though (rating 2 or 3), and frequently misclassified them as either positive or negative. This pattern demonstrates the difficulty of identifying complex or unclear sentiments in text data and suggests that the models are biased towards polar sentiment representations.

Figure 2: Confusion Matrix




The findings demonstrate how successfully the model classifies the two extreme sentiment groups, Very Positive (4) and Very Negative (0). The model's great ability to detect reviews with strongly divided attitudes is demonstrated by the large percentage of reviews in these categories that are correctly identified. For example, most of the real "Very Positive" evaluations were accurately predicted to be so, demonstrating the model's dependability in identifying distinct sentiment indicators.

But when it comes to mid-range thoughts, specifically Negative (1), Neutral (2), and Positive (3), the performance drastically declines. Particularly into class 0 (Very Negative) or class 4 (Very Positive), these classes are frequently misclassified. This implies that strong sentiment predictions are more likely to be predicted by the model than moderate or neutral expressions. Particularly problematic is the Neutral class (2), where many cases are mispredicted as either positive or negative, indicating a lack of contextual awareness in identifying complex attitude.

A more comprehensive evaluation is accomplished by utilizing precision, recall, and F1-score in addition to accuracy, which offers a broad perspective of model performance. Class imbalance and misclassification of some sentiment categories (e.g.,

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

neutral reviews) are frequent in multi-class classification issues, such as sentiment analysis, where these metrics are especially crucial.

**Classification Report on Logistic regression:**

Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 1 | 0.85 | 0.95 | 0.90 | 3748 |
| 2 | 0.31 | 0.02 | 0.03 | 315 |
| 3 | 0.34 | 0.05 | 0.09 | 336 |
| 4 | 0.40 | 0.07 | 0.12 | 635 |
| 5 | 0.89 | 0.98 | 0.93 | 6966 |
| accuracy | | | 0.87 | 12000 |
| macro avg | 0.56 | 0.41 | 0.41 | 12000 |
| weighted avg | 0.82 | 0.87 | 0.83 | 12000 |

**Classification Report on Random Forest :**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 1 | 0.81 | 0.94 | 0.87 | 3748 |
| 2 | 1.00 | 0.00 | 0.01 | 315 |
| 3 | 0.00 | 0.00 | 0.00 | 336 |
| 4 | 0.40 | 0.00 | 0.01 | 635 |
| 5 | 0.87 | 0.96 | 0.92 | 6966 |
| accuracy | | | 0.85 | 12000 |
| macro avg | 0.62 | 0.38 | 0.36 | 12000 |
| weighted avg | 0.81 | 0.85 | 0.80 | 12000 |

**Classification Report on Support Vector Machine:**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 1 | 0.84 | 0.96 | 0.90 | 3748 |
| 2 | 0.33 | 0.00 | 0.01 | 315 |
| 3 | 0.35 | 0.07 | 0.11 | 336 |
| 4 | 0.45 | 0.02 | 0.03 | 635 |
| 5 | 0.89 | 0.98 | 0.93 | 6966 |
| accuracy | | | 0.87 | 12000 |
| macro avg | 0.57 | 0.40 | 0.39 | 12000 |
| weighted avg | 0.82 | 0.87 | 0.83 | 12000 |

**Classification Report on Naive Bayes:**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 1 | 0.83 | 0.95 | 0.88 | 3748 |
| 2 | 0.00 | 0.00 | 0.00 | 315 |
| 3 | 0.00 | 0.00 | 0.00 | 336 |
| 4 | 0.22 | 0.01 | 0.02 | 635 |
| 5 | 0.88 | 0.97 | 0.92 | 6966 |
| accuracy | | | 0.86 | 12000 |
| macro avg | 0.38 | 0.38 | 0.36 | 12000 |
| weighted avg | 0.78 | 0.86 | 0.81 | 12000 |

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

One of the simplest and most popular standards for assessing model performance is classification accuracy. It shows the proportion of accurate predictions the model made out of all the predictions.

**Table1: Accuracy of Models**

| Algorithm Used | Accuracy |
|---|---|
| Logistic regression | 87% |
| Random Forest | 85% |
| Support Vector Machine | 87% |
| Naive Bayes | 86% |

According to the comparative research, the best models for classifying the sentiment of customer reviews in the present investigation are Support Vector Machine and Logistic Regression, both of which achieved an accuracy of 87%. The computational efficiency and performance of these models are well-balanced. A competitive alternative that is quick and easy to use but has a little less accuracy is Naive Bayes. Because of its interpretability and resistance to over fitting, Random Forest is still a good option even though it is marginally less accurate.

**IV. Conclusion:**

The present research used a variety of Natural Language Processing (NLP) methods and machine learning models to investigate and evaluate sentiment analysis of customer evaluations. In order to help businesses better understand customer satisfaction and perspectives, the main goal was to categorize client sentiments based on textual comments, ranging from highly negative to very favorable.

The best-performing models among those tested were SVM and Logistic Regression, both of which achieved an accuracy of 87%. Naive Bayes and Random Forest came in second and third, respectively, at 86% and 85%. According to the confusion matrix study, the models performed best on reviews with high sentiment polarity (extremely positive or very negative), but they had trouble understanding more neutral or nuanced feelings. This highlights how crucial sophisticated methods are for capturing textual data's greater contextual value.

In conclusion, this research offers a strong basis for future advancements in intelligent customer feedback mechanisms and effectively demonstrates the possibilities and difficulties of utilizing NLP approaches in sentiment analysis.

**References**

1. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1–135.
2. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167.
3. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.
4. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543.
5. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.
6. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc.
7. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
8. Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text Classification Algorithms: A Survey. Information, 10(4), 150.