# International Journal of Multidisciplinary
## Research in Science, Engineering and Technology

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*

# Sentiment Analysis of Product Review using Natural Language Processing (NLP)

**Arpitha, Prakriti Bhandary**

PG Student, St Josheph Engineering College, Vamanjoor, Manglore, India

Assistance Professor, St Josheph Engineering College, Vamanjoor, Manglore, India

**ABSTRACT:** Sentiment analysis, also known as opinion mining, is an essential NLP technique used to analyze and extract subjective information from text data. This focuses on evaluating product reviews to determine customers' sentiments, categorizing them as positive, negative, or neutral. By leveraging NLP techniques and machine learning models, this aims to provide insights that can help businesses understand customer feedback better and improve their products and services. The implementation utilizes widely available libraries and tools, making it both accessible and straightforward.

## I. INTRODUCTION

In today's digital age, customers frequently share their experiences and opinions about products online. E-commerce platforms, social media, and review sites are flooded with vast amounts of user-generated content. Analyzing this data manually is impractical due to its sheer volume and complexity. Sentiment analysis automates this process, enabling businesses to systematically analyze customer feedback and derive actionable insights.

In recent years, sentiment analysis has emerged as a crucial tool in understanding customer opinions and behaviors from online reviews. The exponential growth of e-commerce platforms like Amazon has led to an abundance of customer reviews that contain valuable feedback. Sentiment analysis, also known as opinion mining, leverages Natural Language Processing (NLP) techniques to extract subjective information from these text data, enabling businesses to derive meaningful insights and enhance their products and services.

The exponential growth of e-commerce platforms such as Amazon, eBay, and Alibaba has significantly changed the retail landscape. Consumers now have access to a vast array of products and can make purchase decisions based on reviews and ratings from other customers. This user-generated content is a goldmine of information, offering insights into customer satisfaction, product performance, and areas for improvement.

Businesses can use sentiment analysis to monitor customer satisfaction and gather feedback on their products and services. By analyzing the sentiments expressed in reviews, companies can identify common issues, areas of improvement, and aspects that customers appreciate. This information is invaluable for making informed decisions and enhancing customer satisfaction.

## II. LITERATURE

The significance of sentiment analysis extends beyond merely categorizing reviews as positive, negative, or neutral. As highlighted in "Sentiment Analysis: An Overview and Comprehensive Comparison of Techniques and Applications," sentiment analysis helps businesses in decision-making processes, marketing strategies, and customer relationship management. By understanding the sentiments expressed in customer reviews, companies can identify strengths and weaknesses in their products, tailor their marketing efforts to target customer preferences, and proactively address issues that might lead to customer dissatisfaction.[1] Given the vast amount of customer reviews available on platforms like Amazon, sentiment analysis can provide valuable insights into customer opinions and experiences. "Sentiment Analysis and Opinion Mining: A Survey" highlights the potential of sentiment analysis in e-commerce, where businesses can analyze reviews to gauge customer satisfaction, identify trends, and improve product offerings. By categorizing reviews into positive, negative, and neutral sentiments, companies can better understand their customers' needs and preferences, ultimately leading to enhanced customer satisfaction and loyalty.[2]

TextBlob, a Python library built on top of NLTK and Pattern, provides a simple yet effective tool for sentiment analysis. As noted in "TextBlob: Simplified Text Processing," TextBlob offers an easy-to-use API for performing common NLP tasks, including sentiment analysis, part-of-speech tagging, and noun phrase extraction. TextBlob's sentiment analysis functionality is based on a predefined sentiment lexicon and can be used to quickly classify sentiments in text data with minimal effort.[3]

Preprocessing plays a critical role in sentiment analysis by transforming raw text data into a suitable format for analysis. "The Role of Preprocessing in Sentiment Analysis" emphasizes the importance of steps such as tokenization, stop word removal, stemming, and lemmatization in cleaning and standardizing the text data. Proper preprocessing not only enhances the performance of sentiment analysis models but also reduces computational complexity by eliminating noise and irrelevant information from the data.[4]

### III. METHODOLOGY

A. Research Design:
This methodology outlines the steps for performing sentiment analysis on Amazon product reviews using a combination of Python libraries, NLP techniques, and machine learning models. The process involves data collection, preprocessing, sentiment analysis, model training, evaluation, and visualization.

B. Data Collection:
The first step in sentiment analysis is to gather a dataset of Amazon product reviews, which can be achieved either by selecting a relevant dataset from sources such as Kaggle, which should include both review text and corresponding sentiment labels if available, or by using web scraping tools to extract reviews directly from e- commerce websites like Amazon, with the data acquisition process involving downloading the dataset or utilizing APIs provided by data sources.

C. Data Processing:
Data preprocessing involves cleaning and transforming raw text data to make it suitable for sentiment analysis by performing several steps: first, text cleaning is carried out by removing punctuation marks to simplify the analysis, filtering out common stop words that do not contribute to sentiment analysis, and converting all text to lowercase to ensure uniformity; next, tokenization splits the text into individual words or tokens to facilitate further analysis; then, normalization is applied through stemming to reduce words to their root forms and lemmatization to convert words to their base or dictionary forms; finally, feature extraction techniques such as Bag of Words (BoW) and TF-IDF (Term Frequency-Inverse Document Frequency) are used to convert text into numerical formats that reflect token counts and the importance of words relative to the entire dataset.
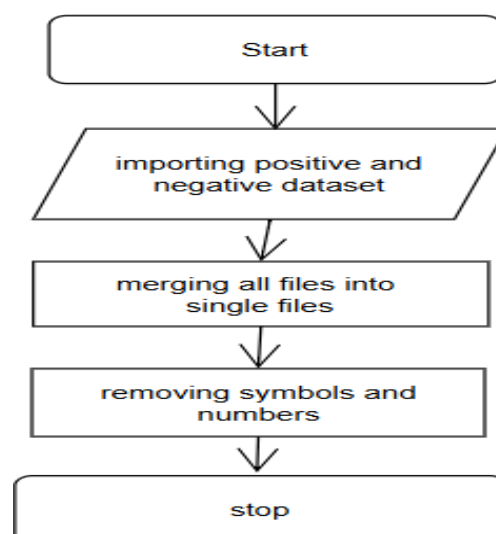


Fig1: Data Processing

**D. Sentiment Analysis:**

Sentiment analysis: Sentiment analysis involves classifying reviews into positive, negative, or neutral categories using various techniques; initially, TextBlob is employed to perform basic sentiment analysis by assigning a polarity score to each review, which reflects how positive or negative the sentiment is, with the score ranging from -1 to 1, where negative words such as "bad," "sad," and "terrible" fall towards -1, positive words like "good," "awesome," and "excellent" fall towards 1, and neutral sentiment is indicated by a score around 0.

**E. Model Training:**

Training involves using the preprocessed data to fit machine learning models by first splitting the data into a training set to train the model and a test set to evaluate its performance, then selecting appropriate models for sentiment classification, such as Logistic Regression or Naive Bayes, and finally fitting these models on the training data while adjusting hyperparameters as needed.
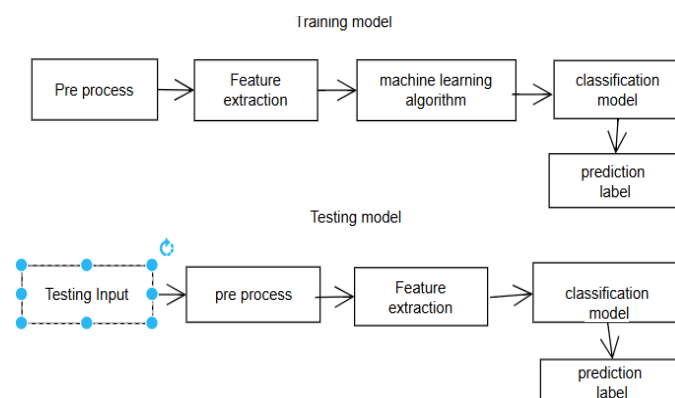
Fig2: Model test and training

**F. Model Evaluation:**

model evaluation assesses the accuracy and effectiveness of the sentiment classification through metrics like precision, recall, and F1 score.

**Accuracy**: Measure the overall correctness of the model.

**Precision**: Assess the proportion of true positive results among all positive predictions.

**Recall**: Evaluate the proportion of true positive results among all actual positives.

**F1 Score**: Compute the harmonic mean of precision and recall to provide a balanced measure.

**G. Visualizations:**

Generate graphs to visualize performance. bar graphs that show the proportion of positive, negative, and neutral reviews. This helps in understanding the overall customer sentiment towards a product.
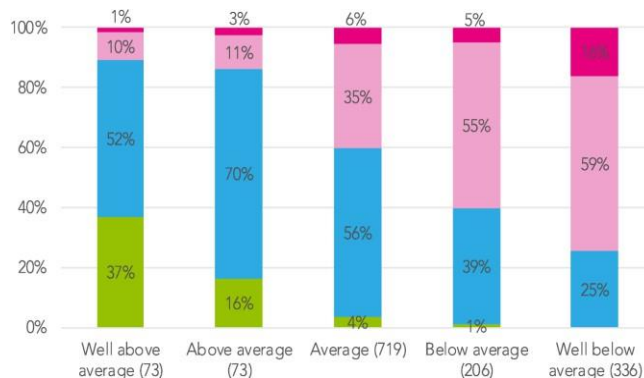
Fig3: visuals of reviews received

H. Interpretation and reporting:

Interpret the results of the sentiment analysis and generate a comprehensive report. Analyses the sentiment distribution and model performance to derive actionable insights. Sentiment scores range from -1 (very negative) to +1 (very positive).Positive sentiment (0.8) indicates a highly favorable review. Negative sentiment (-0.7) indicates a highly unfavourable review. Neutral sentiment (0.1) indicates a somewhat neutral or mixed review.

| Review ID | Processed Review | Sentiment Score |
|---|---|---|
| 1 | "product amazing love" | 0.8 |
| 2 | "terrible quality disappointed" | -0.7 |
| 3 | "okay best decent" | 0.1 |

Table1: Product reviews

## IV. IMPLEMENTATION

**A.** Algorithms Used

1. Text Preprocessing:

Text preprocessing involves cleaning and preparing raw text data for analysis. This includes removing noise such as special characters and punctuation, converting text to lowercase, tokenizing words, and eliminating stop words. In user reviews, preprocessing ensures that only relevant information is analyzed for sentiment classification.

2. Sentiment Analysis:

Sentiment analysis identifies and categorizes opinions expressed in user reviews as positive, negative, or neutral. It leverages NLP techniques to interpret subjective information, helping businesses understand customer feedback and preferences. This analysis enables companies to gauge overall sentiment towards their products and make data-driven decision.

3. Feature Extraction:

Feature extraction transforms preprocessed text into numerical representations for model training. Techniques like bag-of-words, TF-IDF, and word embeddings capture the key characteristics of user reviews. These features serve as inputs for machine learning algorithms, allowing them to detect patterns and classify sentiments effectively.

4. Model Training:

Model training involves using labeled data to teach a machine learning algorithm how to classify sentiments in user reviews. Algorithms such as Naive Bayes, SVM, or deep learning models are trained on extracted features to recognize sentiment patterns. The goal is to develop a model that accurately predicts sentiments on new, unseen reviews.

5. Model Evaluation:

Model evaluation assesses the performance of the trained sentiment analysis model using metrics like accuracy, precision, recall, and F1-score. This step ensures the model's reliability in classifying sentiments in user reviews. Effective evaluation helps in fine-tuning the model for better generalization and real-world applications.

B. Tools and Technologies Used:

1. Python:

Python is a versatile and widely-used programming language in sentiment analysis due to its simplicity and powerful libraries. For customer reviews, Python provides a robust environment to implement and combine various NLP and machine learning techniques, facilitating data processing, analysis, and model development.

2. NLTK (Natural Language Toolkit):
NLTK is a comprehensive library for NLP in Python. It offers tools for text preprocessing, tokenization, stemming, lemmatization, and more. In sentiment analysis of customer reviews, NLTK helps in cleaning and preparing text data, which is crucial for accurate sentiment classification.

3. TextBlob:
TextBlob is a Python library that simplifies text processing tasks. It provides easy-to-use interfaces for common NLP operations, including sentiment analysis. For customer reviews, TextBlob can quickly analyze the sentiment of the text, providing polarity and subjectivity scores that indicate the overall sentiment expressed.

4. scikit-learn:
scikit-learn is a powerful machine learning library in Python. It includes tools for feature extraction, model training, and evaluation. In sentiment analysis of customer reviews, scikit-learn can be used to build and train classification models, such as Naive Bayes or SVM, and evaluate their performance using various metrics.

5. Pandas:
Pandas is a data manipulation library in Python, ideal for handling structured data. For sentiment analysis of customer reviews, Pandas is used to load, clean, and transform datasets. It facilitates efficient data handling, making it easier to preprocess text data and manage feature extraction and model training workflows.

6. Jupyter Notebook:
Jupyter Notebook is an interactive development environment that supports live code, visualizations, and narrative text. It is widely used for data analysis and machine learning projects. In sentiment analysis of customer reviews, Jupyter Notebooks allow for exploratory data analysis, iterative development, and visualization of results, making it easier to document and share findings.

## V. RESULTS

Sentiment analysis of product reviews using Natural Language Processing (NLP) is immensely beneficial for businesses, providing a range of insights and actionable data. By understanding customer sentiment, companies can gauge overall satisfaction and identify trends in customer opinions, enabling them to make informed decisions. This analysis helps in pinpointing specific product features that customers like or dislike, guiding targeted improvements and fostering innovation. It also enhances customer retention and loyalty by allowing businesses to address negative feedback proactively and reinforce positive experiences.

Market research and competitive analysis benefit from sentiment analysis by revealing customer perceptions of competitors' products, helping businesses identify market gaps and opportunities for differentiation. e-commerce platforms use sentiment analysis to monitor reviews and refine product recommendations, while retail and consumer goods companies leverage it for brand management and quality control. Service industries, such as telecommunications and airlines, apply sentiment analysis to improve customer support and service quality. Overall, sentiment analysis using NLP empowers businesses to enhance their products, services, and customer relationships comprehensively.
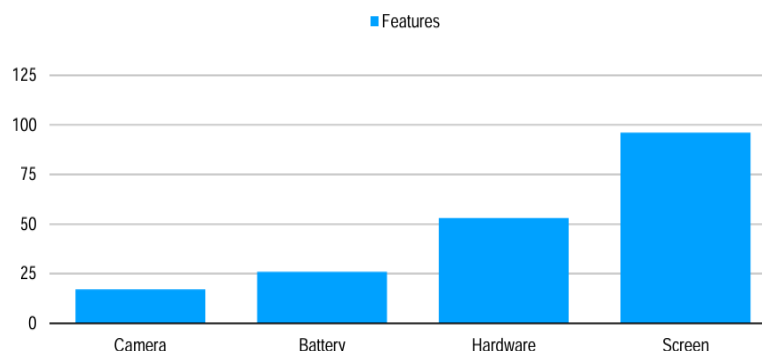


Fig4: result of a customer review

**International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)**

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

consideration of 4,762 reviews of the "Mi Max2" from the amazon website. The various features which had a positive and negative review are generated as follows. The matplotlib library is used to depict the result graphically. Features having a positive review.

## VI. CONCLUSION

This research on sentiment analysis of product reviews utilized a range of tools and methodologies to provide comprehensive insights into customer opinions. Python served as the primary programming language, facilitating the implementation of various NLP and machine learning techniques. NLTK (Natural Language Toolkit) was employed for essential text preprocessing tasks, including tokenization, stop word removal, and stemming, which are critical for preparing raw text data for analysis. TextBlob was used for simplified sentiment analysis, offering a user-friendly interface to determine the polarity and subjectivity of reviews.

Feature extraction was carried out using scikit-learn, which transformed preprocessed text into numerical representations suitable for machine learning models. This included techniques like TF-IDF to capture the importance of words in the context of reviews. Model training and evaluation were also performed with scikit-learn, where various algorithms, such as Naive Bayes and SVM, were used to classify sentiments and assess model performance using metrics like accuracy and F1-score. Pandas was instrumental in data manipulation, allowing efficient handling of structured data, while Jupyter Notebook provided an interactive environment for development, visualization, and documentation of the analysis process.

The final results revealed actionable insights, helping businesses understand customer sentiment, identify key areas for product improvement, and make data-driven decisions to enhance customer satisfaction and retention. By applying sentiment analysis, businesses can better manage their brand, refine product offerings, and tailor marketing strategies, ultimately leading to more informed and strategic decision-making.

## REFERENCES

[1] Harish Dutt Sharma and Parul Goyal (MDPI Engineering Proceedings) (MDPI)**:** " Sentiment Analysis: An Overview and Comprehensive Comparison of Techniques and Applications "
[2] Bo Pang and Lillian Lee:"Sentiment Analysis and Opinion Mining: A Survey"
[3] Steven Bird, Ewan Klein, and Edward Loper:"TextBlob: Simplified Text Processing"
[4] Pranav Bhardwaj and Divya Singh:"The Role of Preprocessing in Sentiment Analysis".