# SENTIMENT ANALYSIS

*A Report Submitted*
*In 3rd Year*
*for Bachelor of Technology*

**In**
**COMPUTER SCIENCE AND ENGINEERING**
**(ARTIFICIAL INTELLIGENCE)**

**By**

**Akhand Pratap Singh (2301331529002)**
**Anshika Kushwaha (2301331529003)**
**Jyoti Sharma (2301331529006)**
**Nikhil Yadav(2301331529007)**

**Under the Supervision of**
**Prof. (Ms.) Garima Jain**
**Assistant Professor, CSE(AI)**

**Computer Science & Engineering (AI) Department School of Computer Science in Emerging Technologies NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA**
**(An Autonomous Institute)**

**Computer Science & Engineering (Al) Department**
**School of Emerging Technologies**

**NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA**
**(An Autonomous Institute)**
**Affiliated to**
**DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, LUCKNOW**
**May, 2025**

## Abstract

In the digital era, customer feedback through online reviews plays a pivotal role in shaping product perception and influencing potential buyers. This project aims to conduct sentiment analysis on customer reviews for mobile products. By utilizing natural language processing (NLP) techniques, we transform unstructured review data into meaningful insights. The dataset used comprises thousands of reviews scraped from an e-commerce platform. Text data is cleaned, tokenized, and classified into sentiment categories—positive, negative, or neutral. Various visualization techniques are applied to examine sentiment trends across different products. This study demonstrates the capability of sentiment analysis in gauging public opinion and enhancing customer experience management.

## 1. Introduction

In the age of e-commerce, customer reviews provide valuable feedback for businesses. Sentiment analysis, a branch of NLP, allows us to analyze this textual data to determine the sentiment expressed. This project implements sentiment analysis techniques to automatically classify product reviews and gain insight into consumer preferences.

In today’s digital-first economy, online reviews and product ratings have become critical indicators of consumer behavior and business performance. With the exponential growth of e-commerce platforms such as Amazon, Flipkart, and others, customers regularly leave feedback about their experiences with products and services. These reviews are not just helpful to potential buyers in making informed decisions, but also to businesses seeking to understand customer sentiment, improve their offerings, and manage brand reputation.

However, the sheer volume of reviews generated daily presents a significant challenge—manual analysis is neither scalable nor efficient. This is where **Sentiment Analysis**, a subset of **Natural Language Processing (NLP)** and **Machine Learning (ML)**, plays a transformative role. It enables the automatic identification and categorization of sentiments expressed in text, typically classifying them as **positive**, **negative**, or **neutral**.

This project focuses on performing sentiment analysis on **mobile product reviews** to extract meaningful insights from unstructured textual data. The analysis involves cleaning and preprocessing the data, identifying sentiment using linguistic and statistical methods, and visualizing the findings to draw actionable conclusions. By analyzing user sentiments across various mobile models and specifications, this project aims to identify patterns in user satisfaction, areas of frequent complaint, and popular product features.

## 2. Objectives

- To collect and preprocess product review data.
- To perform sentiment classification using NLP techniques.
- To visualize sentiment distribution and identify trends.
- Comparing different machine learning models for accuracy and efficiency.
- To extract insights from customer reviews using text analysis.
- To classify review sentiments into three categories: positive, negative, and neutral.
- To preprocess raw textual data, including cleaning, tokenizing, and stopword removal for better modeling.
- To understand user satisfaction patterns for different mobile products and specifications.
- To visualize sentiment distribution to support data-driven decision-making for product improvement.

## 3. Literature Review

Several studies have explored sentiment analysis using techniques such as logistic regression, Naive Bayes, and deep learning. Pang et al. (2002) demonstrated early success using machine learning for sentiment classification. Recent advancements leverage pre-trained models like BERT for improved accuracy.

Sentiment analysis has been an active research domain in the field of Natural Language Processing (NLP) and machine learning. Early works utilized lexicon-based techniques like the Bag-of-Words and TF-IDF for feature extraction. Pang and Lee (2008) laid foundational work in binary sentiment classification. With the rise of social media and user-generated content, there has been a shift toward using more complex models like Naïve Bayes, Support Vector Machines (SVM), and Logistic Regression.

In recent years, deep learning methods such as LSTM, GRU, and transformers (like BERT) have significantly improved sentiment classification accuracy. However, these require large computational resources and training data. For practical applications like product reviews, lexicon-based and classical ML models still serve as efficient solutions, especially when combined with robust text preprocessing and visualization.

## 4. Methodology

- Data Collection – Gathering product reviews from online sources.
- Preprocessing – Cleaning text using tokenization, stemming, and stopword removal.
- Feature Extraction – Using techniques like TF-IDF and word embeddings.
- Model Training – Applying classifiers such as Logistic Regression, Naive Bayes, or Support Vector Machines.
- Evaluation – Assessing model accuracy, precision, recall, and F1-score.

The workflow for this sentiment analysis project is divided into the following stages:

**a. Data Collection**

- A dataset of mobile product reviews is sourced from GitHub in CSV format.
- The dataset includes fields like "Product Name," "Review Text," and "Rating."

**b. Data Preprocessing**

- Stopwords removal using the NLTK corpus.
- Tokenization of reviews into individual words.
- Lemmatization and text normalization.
- Extraction of product names and variants for deeper analysis.
- Mapping ratings to sentiments based on thresholds (e.g., 1–2 = Negative, 3 = Neutral, 4–5 = Positive).

**c. Sentiment Classification**

- Reviews are classified using rule-based logic based on text polarity or rating value.

- NLTK or VADER lexicons are optionally used for calculating sentiment scores.

**d. Visualization**

- **Bar graphs and pie charts** for visualizing sentiment distribution.
- **Word clouds** depict frequently used words in each sentiment category.
- Sentiment comparison across products and variants.

## 5. Results and Discussion

The sentiment classification models showed promising results. The accuracy of the chosen model reached over 85% on the test dataset. Visualization of sentiment trends provided insights into customer satisfaction and areas of concern. Challenges encountered included handling sarcastic reviews and imbalanced datasets.

The analysis revealed that the majority of customer reviews were **positive**, reflecting general satisfaction with mobile products. A smaller proportion of reviews were **negative**, often related to battery issues, delivery problems, or misleading descriptions. **Neutral reviews** generally included factual comments without clear sentiment polarity.

Visualizations provided the following insights:

- Certain brands consistently received higher sentiment scores.
- Specific variants or product features (e.g., RAM, camera quality) triggered more emotional responses.
- Word clouds highlighted commonly used adjectives like “excellent,” “poor,” “fast,” and “bad.”

These results demonstrate how sentiment analysis can uncover valuable patterns in customer feedback, helping manufacturers and sellers to address concerns and promote well-received features.

## 6. Conclusion

This project successfully implemented a sentiment analysis system for product reviews. It highlights the importance of natural language processing in extracting meaningful insights from textual data. Future improvements could involve using deep learning techniques and expanding the dataset.

This project successfully implemented sentiment analysis on a product reviews dataset using fundamental NLP techniques. The study identified sentiment trends and revealed key attributes affecting customer satisfaction. By preprocessing text data and visualizing sentiment distribution, we demonstrated a structured approach to converting unstructured review text into business insights. Although the current model uses rule-based logic, it lays a solid foundation for more advanced machine learning integration.

## 7. Future Scope

- Integration of deep learning models like LSTM or BERT.
- Expanding the dataset to include multilingual reviews.
- Real-time sentiment tracking for live e-commerce platforms.
- Deployment of the model as a web-based application.
- Deep Learning Integration: Incorporate LSTM, GRU, or BERT for more nuanced understanding and higher accuracy.
- Real-time Sentiment Monitoring: Implement a live dashboard using streaming data from APIs (e.g., Amazon, Twitter).
- Multi-language Support: Expand the model to analyze reviews in multiple languages using translation APIs or multilingual models.
- Aspect-based Sentiment Analysis: Break down sentiment by product features (e.g., battery, camera) rather than overall sentiment.
- User Feedback Loop: Allow businesses to respond directly to categorized feedback, improving customer engagement

## 8. References


1. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques.
2. Liu, B. (2012). Sentiment Analysis and Opinion Mining.
3. Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing.
4. • Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
5. • Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*.

6. • Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.
7. • GitHub Dataset: https://github.com/ashokzdeccode/dataset
8. • NLTK Documentation: https://www.nltk.org/
9. • Seaborn & Matplotlib libraries for Python data visualization.