

**RAJALAKSHMI INSTITUTE OF TECHNOLOGY**

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

**DEPARTMENT OF CSE (ARTIFICIAL INTELLIGENCE AND**

**MACHINE LEARNING) ACADEMIC YEAR 2025 – 2026**

**SEMESTER III**

**ARTIFICIAL INTELLIGENCE LABORATORY MINI PROJECT REPORT**

|  |  |
| --- | --- |
| **REGISTER NUMBER** | 2117240030135 |
| **NAME** | SARAH .S.S |
| **PROJECT TITLE** | SENTIMENT ANALYSIS ON REVIEWS |
| **FACULTY IN-CHARGE** | Mrs.M.BHAVANI |

**Signature of Faculty In-charge:**

**INTRODUCTION**

In today’s digital world, people express their opinions and feedback about products, services, or experiences through online platforms such as Amazon, Flipkart, and social media. Analyzing these large volumes of textual reviews manually is time-consuming and inefficient. **Sentiment Analysis**, a branch of **Natural Language Processing (NLP)** and **Artificial Intelligence (AI)**, helps to automatically determine whether a given review expresses a **positive**, **negative**, or **neutral** sentiment.

This project, **“Sentiment Analysis on Reviews,”** aims to develop an intelligent system capable of reading and classifying text reviews based on the emotions or opinions expressed. The system processes input data using text preprocessing techniques such as **tokenization**, **stop-word removal**, and **vectorization**, and then uses **machine learning algorithms** like **Logistic Regression**, **Naïve Bayes**, or **Support Vector Machine (SVM)** to predict the sentiment.

The main goal of this project is to help organizations understand customer opinions efficiently, make data-driven decisions, and improve product quality or customer service. By using AI-based text classification, the system ensures **fast**, **accurate**, and **automated** sentiment detection from large datasets of reviews.

**PROBLEM STATEMENT**

Analyzing large volumes of online reviews manually is **time-consuming** and **error-prone**. This project aims to develop an **AI-based Sentiment Analysis System** that automatically classifies customer reviews as **positive**, **negative**, or **neutral** using **Natural Language Processing (NLP)** techniques. The system uses **machine learning algorithms** to analyze text and predict sentiment accurately, helping organizations make better, data-driven decisions.

# GOAL

The goal of this project is to develop a **Sentiment Analysis System** that can automatically classify user reviews into **positive**, **negative**, or **neutral** categories using Artificial Intelligence and Natural Language Processing techniques. The system aims to:  
• Analyze large volumes of textual data efficiently  
• Provide accurate sentiment classification for better understanding of customer opinions  
• Assist businesses in making data-driven decisions based on customer feedback  
Ultimately, this project demonstrates how **AI and NLP** can simplify the process of opinion mining, making feedback analysis faster, smarter, and more reliable.

**THEORETICAL BACKGROUND**

**Automation in Data Analysis:**

Automation has become a key part of modern data-driven systems, helping organizations process large amounts of information quickly and accurately. In traditional review analysis, human evaluators manually read and interpret feedback, which is time-consuming, inconsistent, and prone to bias. Automation through Artificial Intelligence (AI) and Natural Language Processing (NLP) allows the system to automatically process thousands of reviews, extract useful information, and classify opinions efficiently. This improves decision-making by providing instant insights into customer satisfaction and brand perception.

**AI in Text Processing:**

Artificial Intelligence (AI) plays a major role in analyzing and understanding human language. In sentiment analysis, AI models learn from labeled datasets to recognize emotional tone in text. Techniques such as tokenization, stop-word removal, and vectorization help convert text into numerical data that can be processed by machine learning algorithms. The trained model can then predict whether a review expresses a positive, negative, or neutral sentiment. This automated classification helps businesses and organizations handle large-scale feedback without manual involvement.

**Natural Language Processing (NLP):**

NLP bridges the gap between human language and computers. It involves multiple preprocessing steps that prepare text for analysis, such as:  
• Tokenization – Splitting sentences into individual words.  
• Stop-word removal – Removing common words (like “is”, “the”, “and”) that do not affect meaning.  
• Stemming/Lemmatization – Reducing words to their root form (e.g., “running” → “run”).  
• Vectorization – Converting words into numerical form using methods like TF-IDF or Bag of Words.

These techniques enable the machine learning model to interpret text data and identify underlying sentiment patterns.

**Machine Learning Algorithms:**

Several algorithms are used for sentiment classification, including:  
• Naïve Bayes Classifier – Works well for text data and assumes independence between features.  
• Logistic Regression – Efficient for binary and multi-class text classification.  
• Support Vector Machine (SVM) – Separates text data into distinct sentiment categories with high accuracy.  
These algorithms learn from labeled training data and generalize the knowledge to classify unseen reviews.

**Justification:**

Using AI and NLP for sentiment analysis is justified due to the need for speed, accuracy, and scalability in handling massive online data. Manual analysis is subjective and inconsistent, while automated sentiment detection ensures objectivity and efficiency. The system helps companies monitor customer opinions, evaluate product performance, and make timely improvements. Furthermore, AI-based systems can be retrained with new data, making them adaptive, reliable, and future-ready for continuous feedback analysis

**ALGORITHM EXPLANATION WITH EXAMPLE**

**The Sentiment Analysis on Reviews system follows a five-step process:**

**Step 1 — Data Collection:**Gather text data such as product or service reviews from datasets, websites, or user inputs. Each review includes the review text and its corresponding sentiment label (positive, negative, or neutral) for training.

**Step 2 — Text Preprocessing:**Clean and prepare the text by performing operations like tokenization, stop-word removal, stemming/lemmatization, and lowercasing. This ensures that the data is consistent and ready for analysis.

**Step 3 — Feature Extraction:**Convert the processed text into numerical form using techniques such as TF-IDF (Term Frequency–Inverse Document Frequency) or Bag of Words (BoW). These methods help represent textual data in a way that can be used by machine learning models.

**Step 4 — Model Training and Prediction:**Train a machine learning algorithm (e.g., Naïve Bayes, Logistic Regression, or Support Vector Machine) on the preprocessed and vectorized data. The trained model learns sentiment patterns and then predicts whether a new review is positive, negative, or neutral.

**Step 5 — Output Generation:**Display the classified sentiment for each input review, showing the review text, predicted sentiment, and accuracy score of the model. The results can be visualized or exported for further analysis**.**

**Complete Example Workflow:**

1. **Input Collection:**  
   The system takes input in the form of text reviews collected from users, datasets, or online platforms. Each review may include a text message and, in the case of training data, a known sentiment label (positive, negative, or neutral).
2. **Data Processing:**  
   The collected text is cleaned and prepared using Natural Language Processing (NLP) steps such as tokenization, stop-word removal, stemming/lemmatization, and lowercasing. This helps standardize the text and remove unnecessary words or symbols.
3. **Feature Extraction:**  
   After preprocessing, the text data is converted into numerical form using TF-IDF (Term Frequency–Inverse Document Frequency) or Bag of Words (BoW) methods. These numerical features represent the importance of words in the dataset and are used as input for machine learning models.
4. **Model Training and Prediction:**  
   A machine learning classifier such as Naïve Bayes, Logistic Regression, or SVM is trained using the processed data. The model learns sentiment patterns and predicts the sentiment of new or unseen reviews automatically.
5. **Display / Output:**  
   The system displays the original review text along with its predicted sentiment (Positive, Negative, or Neutral). The accuracy of the model is also shown, and results can be exported or visualized for further analysis.

**IMPLEMENTATION AND CODE**

The following Python code implements the AI-based medical diagnosis system:

from sklearn.model\_selection import train\_test\_split

from sklearn.feature\_extraction.text import CountVectorizer

from sklearn.naive\_bayes import MultinomialNB

from sklearn import metrics

# Sample data

reviews = ["Excellent food and service", "Worst experience ever", "Good value for money", "Not recommended", "Will buy again", "Very bad taste"]

labels = ["positive", "negative", "positive", "negative", "positive", "negative"]

# Text vectorization

vectorizer = CountVectorizer()

X = vectorizer.fit\_transform(reviews)

# Split data

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, labels, test\_size=0.3, random\_state=42)

# Naive Bayes model

model = MultinomialNB()

model.fit(X\_train, y\_train)

# Prediction and evaluation

pred = model.predict(X\_test)

print("Accuracy:", metrics.accuracy\_score(y\_test, pred))

**CODE EXPLANATION:**

**Step 1:**The dataset of customer reviews and their corresponding sentiment labels (positive, negative, or neutral) is created or imported. This serves as the input for training and testing the model.

**Step 2:**The text data is preprocessed using Natural Language Processing (NLP) techniques such as tokenization, stop-word removal, and TF-IDF vectorization, converting text into a numerical format suitable for machine learning.

**Step 3:**A machine learning model (like Naïve Bayes or Logistic Regression) is trained on the preprocessed data. The trained model learns patterns in the text that correspond to each sentiment type and predicts the sentiment for new reviews.

**Step 4:**The system displays each review along with its predicted sentiment and shows the overall model accuracy and classification report in a neatly formatted output on the screen.

**OUTPUT:**

**A computer screen with a black background

AI-generated content may be incorrect.**

**OUTPUT EXPLANATION:**

When the program is executed, it displays the **review text** along with its **predicted sentiment** (Positive, Negative, or Neutral) in a clear and readable format. Each review is analyzed using **Natural Language Processing (NLP)** techniques and classified based on the features learned during model training.  
The output demonstrates how the system uses **AI and machine learning algorithms** to automatically identify the emotional tone of each review. After processing all reviews, the program also displays the **model’s accuracy** and a **classification report**, showing how effectively the system predicts sentiments.This confirms that the Sentiment Analysis model works correctly, providing an **accurate, automated, and efficient** way to understand user opinions.

**RESULTS AND FUTURE ENHANCEMENT**

### **Results:**

The developed **Sentiment Analysis on Reviews** system successfully demonstrates the application of **Artificial Intelligence (AI)** and **Natural Language Processing (NLP)** in understanding human opinions.It automatically analyzes and classifies reviews into **positive**, **negative**, or **neutral** categories based on emotional tone and word patterns.The system efficiently processes multiple reviews, identifies sentiment with high accuracy, and provides clear output showing the predicted label for each review.This project proves how **AI-based text classification** can help organizations quickly interpret large volumes of feedback, improving customer understanding and decision-making efficiency.

### **Comparison with Manual Methods:**

Traditional review analysis requires humans to manually read and interpret opinions, which is **slow, inconsistent, and prone to bias**.  
The automated sentiment analysis system, on the other hand, ensures **speed, objectivity, and scalability** by using trained machine learning models.While manual methods can capture emotional nuance, the proposed system delivers **faster and more reliable** results, making it ideal for analyzing thousands of reviews simultaneously.

### **Future Enhancements:**

• **Deep Learning Integration:** Use advanced models like **LSTM**, **BERT**, or **Transformers** for better contextual understanding and higher accuracy.  
• **Aspect-Based Sentiment Analysis:** Identify sentiment related to specific product features (e.g., camera, battery, delivery).  
• **Multi-Language Support:** Extend the system to handle reviews in different regional and international languages.  
• **Real-Time Data Analysis:** Integrate with APIs from platforms like **Twitter, Amazon, or Flipkart** to analyze live user feedback.  
• **Dashboard and Visualization:** Develop a **GUI or web dashboard** to display sentiment results graphically using charts and graphs.  
• **Database Connectivity:** Store and manage reviews dynamically for future model updates and performance tracking.

|  |  |
| --- | --- |
| **Git Hub Link of the project and report** | https://github.com/SarahStephen0912/MINI-PROJECT?tab=readme-ov-file#mini-project |

# 

# 

# REFERENCES

1. Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th Edition). Pearson Education.
2. Pang, B., & Lee, L. (2008). *Opinion Mining and Sentiment Analysis*. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
3. IBM Cloud Education. (2023). *What is Natural Language Processing (NLP)?* https://www.ibm.com/topics/natural-language-processing
4. GeeksforGeeks. (2023). *Sentiment Analysis using Python*. https://www.geeksforgeeks.org/sentiment-analysis-using-python/
5. TutorialsPoint. (2023). *Text Classification and Sentiment Analysis in Machine Learning*. https://www.tutorialspoint.com