ISSN: **3107-6513**

Sentiment Analysis of Social Media Post and Online Reviews Using Machine Learning

Buela X , Velumani T

Department of Computer Science, Rathinam College of Arts and Science (Autonomous), Coimbatore,

Tamil Nadu, India

buelaprincy10@gmail

ABSTRACT

Sentiment analysis of social media posts and online reviews using machine learning provides an efficient way to understand public opinions and user feedback. With the rapid growth of digital platforms, large volumes of textual data are generated daily, making manual analysis difficult and time-consuming. Machine learning techniques help in automatically classifying this data into positive, negative, or neutral sentiments. Various algorithms such as Naive Bayes, Support Vector Machine (SVM), and Random Forest improve prediction accuracy by learning patterns from textual data. The system involves data collection, preprocessing, feature extraction, and model evaluation to achieve reliable results. Visualization techniques are also used to represent sentiment distribution clearly. This approach helps businesses and organizations make better decisions, improve customer satisfaction, and analyze market trends effectively.

Keywords: *Machine Learning, Sentiment Analysis, Social Media, Data Preprocessing, TF-IDF, Naive Bayes, SVM, Model Evaluation*

I. Introduction

Sentiment analysis is an important application of Natural Language Processing (NLP) that focuses on identifying and extracting opinions from textual data. With the increasing use of social media platforms and online review systems, users frequently share their experiences, opinions, and feedback about products, services, and events. This data plays a crucial role in understanding customer behavior and improving decision-making processes.

Analyzing such a large amount of data manually is not practical. Therefore, machine learning techniques are used to automate the process of sentiment classification. These techniques can analyze patterns in text and categorize them into positive, negative, or neutral sentiments. This helps organizations monitor customer opinions, improve services, and enhance user experience.

Social media platforms like Twitter and review platforms such as Amazon generate valuable insights into public perception. By applying sentiment analysis, companies can identify trends, detect issues, and respond quickly to customer feedback. It also helps in brand monitoring, product improvement, and marketing strategies. Machine learning models such as Naive Bayes, Support Vector Machine, and Random Forest are widely used for sentiment classification. These models are trained using labeled datasets and evaluated using performance metrics. The increasing availability of data and computational power has further improved the efficiency of sentiment analysis systems.

ISSN: **3107-6513**

II. Existing and Proposed System

The proposed system aims to improve the accuracy and efficiency of sentiment analysis for social media posts or reviews by using advanced machine learning and Natural LanguageProcessing(NLP) techniques. Unlike traditional systems, this approach focuses on better text preprocessing, feature extraction, and model training to handle complex and unstructured data more effectively. In this system, the input text data is first cleaned and preprocessed using

techniques such as tokenization, stopword removal, and stemming. Then, feature extraction methods like TF-IDF are applied to convert the text into numerical form.

Machine learning algorithms such as Naïve Bayes, Random Forest, and Support Vector Machine (SVM) are used to classify the sentiments into positive, negative, and neutral categories. The proposed system is designed to handle large datasets and can be extended for real-time sentiment analysis. It also provides visualization outputs such as pie charts, bar graphs, and word clouds to better understand the sentiment distribution. Overall, the proposed system overcomes the limitations of existing methods by providing higher accuracy, better adaptability, and efficient processing of social media data

III. Advantages of the Proposed System

- Useadvancedmachinelearningalgorithms like Naïve Bayes, Random Forest, and SVM to provide more accurate sentiment predictions.
- Efficiently processes large volumes of social media posts and reviews.
- Handles complex language, including slang and informal text better than traditional systems
- Reducesmanual effort by automatically analyzing and classifying sentiments

III. System Specification

Hardware Requirements

- **Processor:** Intel Core i3 / i5 or higher
- **RAM:** Minimum 4 GB (Recommended 8 GB or above)
- **Storage:** At least 500 GB HDD or SSD
- **System Type:** 64-bit operating system
- **Internet Connection:** Required for dataset download and updates

Software Requirements

ISSN: **3107-6513**

The software requirements for developing and running the system include:

- **Operating System:** Windows 10/11, Linux, or macOS
- **Programming Language:** Python (Version 3.7 or above)

IV. Methodology

Data is collected from social media platforms such as Twitter and online review websites. The dataset contains textual data along with sentiment labels.

Preprocessing involves cleaning the text data by removing stopwords, punctuation, URLs, and special characters. Tokenization, stemming, and lowercasing are also performed.

Exploratory data analysis is performed to understand the distribution of sentiments and identify patterns in the dataset.

The dataset is divided into training and testing sets, typically using an 80:20 ratio to evaluate model performance.

V. Module Description

DATA GATHERING:

Data gathering is the first and most important step in the sentiment analysis project. It involves collecting relevant data from various sources such as social media platforms, online review websites, and public datasets. In this project, the dataset is mainly collected from sources like Kaggle, which provides a large number of labeled datasets for sentiment analysis.

DATA PREPROCESSING:

Data processing involves cleaning and preparing raw text data for analysis. It includes removing unwanted characters, converting text to lowercase, and eliminating stopwords. The text is then converted into numerical form using techniques like TF-IDF. This helps improve the accuracy of the mod

DATA SPLITTING:

Data splitting is the process of dividing the dataset into training and testing sets. Usually, 80% of the data is used for training the model and 20% for testing. This helps evaluate the model's performance on unseen data and ensures better accuracy.

ALGORITHM SELECTION:

Algorithm selection involves choosing suitable machine learning algorithms for sentiment classification. In this project, algorithms like Naïve Bayes, Random Forest, and Support Vector Machine (SVM) are used. These algorithms are selected based on their efficiency, accuracy, and ability to handle text data.Algorithms:

VI. System Design

ISSN: **3107-6513**

File design refers to the organization and management of data files used in the sentiment analysis system. In this project, the dataset is typically stored in a structured format such as CSV (Comma Separated Values), which allows easy access, modification, and processing of data.

Input design focuses on how data is entered into the sentiment analysis system in a clear andefficient manner. In this project, the input mainly consists of social media posts or user reviews, which are provided in the form of text data. The input can be given through a dataset file (CSV) or directly entered by the user through the interface

VII. Data Flow and ER Design

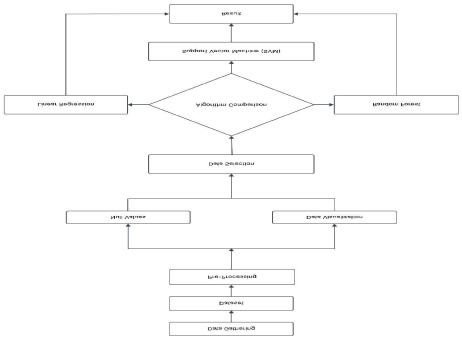

Figure 1. Data Flow Diagram of the proposed precision agriculture system

.

VIII. Testing and Implementation

System implementation is the stage where the sentiment analysis system is developed and executed.includes data preprocessing, model training, and testing. The system then generatesentiment predictions as output.

System development for sentiment analysis involves creating a system that can identify and classify emotions in text. It begins with data collection, gathering reviews, comments, or social media posts. The data is then preprocessed to remove noise, stopwords, and irrelevant information. Next, features are extracted, often using techniques like TF-IDF or word embeddings. A machine learning model, such as Random Forest or Support Vector Machine, is trained on the processed data.

ISSN: **3107-6513**

IX. Results and Discussion

The performance of different algorithms is compared based on evaluation metrics. Among the models, Support Vector Machine (SVM) provides better accuracy compared to other algorithms. Naive Bayes performs efficiently with faster computation, while Random Forest provides stable results.

Visualization techniques such as bar charts and word clouds are used to represent sentiment distribution and frequently used words in the dataset.

Table I. Model Performance Metrics

FIELD	DATA TYPE	DESCRIPTION
id	BIGINT (PK)	Unique record ID
username	VARCHAR(100)	Twitter handle of the user
User_id	BIGINT	Twitter user id
Tweet_text	TEXT	Full original tweet content
Cleaned_text	TEXT	Preprocessing text

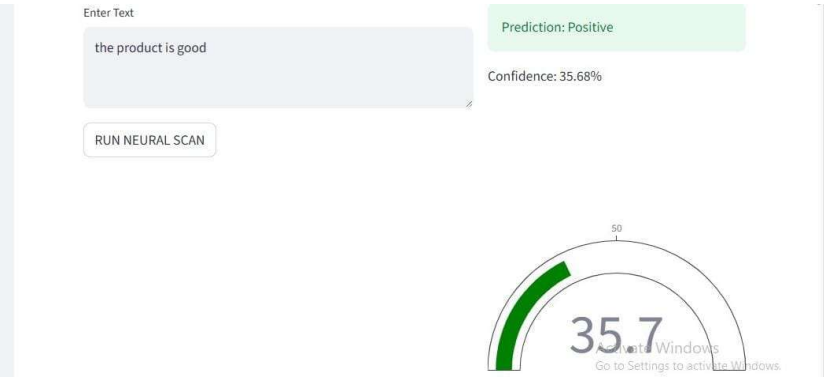

Figure 2. Abstract page from the uploaded final project report.

ISSN: **3107-6513**

X. Conclusion and Future Enhancement

Sentiment analysis using machine learning is an effective approach to analyze social media posts and online reviews. It helps in understanding customer opinions and improving decision-making processes. Among the algorithms used, Support Vector Machine (SVM) shows better performance in terms of accuracy.

The system can be further improved by using deep learning techniques and real-time data analysis. Future enhancements may include multilingual sentiment analysis and integration with business intelligence tools.

The sentiment analysis system can be improved to support multiple languages for global use. Deep learning models like LSTM or BERT can enhance accuracy and understand subtle emotions. Real-time monitoring can process social media and feedback instantly. Detecting sarcasm and contextual sentiments can make predictions more reliable

REFERENCES

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis.

Liu, B. (2012). Sentiment Analysis and Opinion Mining.

Jurafsky, D., & Martin, J. H. Speech and Language Processing.

Various IEEE papers on Sentiment Analysis.