Sentiment Analysis of Social Media Presence

[1]Maheshwari Patil, [2]Divya Mandewal , [3]Nikita Patil, [4]Vedant Wadile, [5]Dr. Akash D. Waghmare

[1,2,3,4] UG Student, [5]Associate Profesoor Department of Computer Engineering, SSBT's College of Engineering and Technology Jalgaon, Maharashtra, India

Abstract: Social media has developed into a potent tool in the current digital era for people, groups, and brands to share experiences, voice opinions, and shape public opinion.By creating a sentiment analysis model that reliably categorizes user opinions posted on social media platforms as neutral, negative, or positive, this project seeks to address the challenge of comprehending public sentiment.Numerical feature vectors are created from raw textual data using the TF-IDF (Term Frequency-Inverse Document Frequency) method.Our approach includes data preprocessing (cleaning, tokenization, stop-word removal), feature extraction using TF-IDF.Our method consists of feature extraction using TF-IDF and data preprocessing (cleaning, tokenization, stop-word removal).Logistic Regression Model is used.
The outcomes demonstrate how well automated sentiment analysis can decipher trends in public opinion from social.

Index Terms **– Sentiment Analysis, Social Media, Natural Language Processing (NLP), Machine Learning, Public Opinion.**

I.INTRODUCTION

The quick development of social media sites like Facebook, Instagram, and Twitter in recent years has had a big impact on how people and organizations communicate. A vast amount of user-generated content reflecting public sentiments, emotions, and opinions is produced by these platforms. Understanding trends, consumer behavior, and societal responses to events, goods, or policies all depend on the analysis of this content. Opinion mining, another name for sentiment analysis, is a branch of Natural Language Processing (NLP) that focuses on identifying the emotional tone or sentiment of textual data. Usually, it entails grouping input text into categories like neutral, negative, and positive. Sentiment analysis offers a useful method for drawing insightful conclusions from vast amounts of data using machine learning algorithms and natural language processing techniques.

II. MOTIVATION

Social media's growing impact on public opinion has made it crucial for businesses, brands, and legislators to analyze user sentiment. Due to the enormous volume and dynamic nature of social media data, manual methods of determining public opinion are time-consuming and frequently inaccurate.
Using automated sentiment analysis to offer real-time insights into user opinions is what motivates this work. Businesses can use this to monitor brand reputation, enhance customer engagement, and improve marketing tactics. Furthermore, examining public responses to policies or events can help social organizations and governments make well-informed decisions. The difficulty of deciphering slang, emojis, sarcasm, and informal language on social media emphasizes how crucial it is to create an effective, intelligent sentiment analysis system.

III. LITERATURE SURVEY

This Literature Survey Table summarizes the existing research and systems in the field of sentiment analysis, highlighting their methodologies, techniques, and applications in analyzing social media content. Table 1.1shows Literature Survey.

Table 1.1 Literature Survey

Sr. No.	Year	Title	Authors	Implementation Details
1	2012	Sentiment Analysis and Opinion Mining.	B. Liu. Distinguished Professor of Computer Science at the University of Illinois at Chicago.	Established foundation of sentiment mining; emphasized challenges like sarcasm and domain-dependence.
2	2018	Sentiment Analysis to Predict Election Results.	Farha Nausheen & Sayyada Begum. In field of Computer Science, Data Science. Affiliated with institutions in India, presented at second international Conference Inventive System and Control.	Demonstrated political sentiment prediction using Twitter data with fair accuracy.
3	2016	Swachh Bharat Campaign Sentiment Analysis.	Tayal D. K. et al. Computer Engineering, Jaypee Institute of Information Technology, Noida, India.	Classified sentiments of tweets related to national campaigns.
4	2006	Adjectives and Adverbs in Sentiment Analysis.	Carmine C. et al. Computing and Multimodal Analysis. Department of Computing, Imperial College London, UK.	Concluded that adjectives and adverbs play a major role in identifying sentiment polarity.

IV. PROBLEM STATEMENT

The vast amounts of unstructured text generated by social media make it challenging to manually decipher user sentiment. The goal of this project is to create an automated system that can categorize social media posts into three groups: neutral, negative, and positive. To ensure accurate sentiment analysis, it addresses issues like slang, emojis, and sarcasm while utilizing NLP and machine learning techniques.

v. ALGORITHM

The algorithm implemented in this project involves several key steps, starting from user input in the form of text data. The raw input is preprocessed to remove noise, converted into a machine-understandable format using natural language processing techniques, and then passed through a trained machine learning model — in this case, **Logistic Regression** — to analyze the sentiment.

1. Input Representation: Represent each input example as a feature vector $x \in \mathbb{R}^n$, where n is the number of features.

2. Initialize Parameters: Initialize the weights $W \in \mathbb{R}^{nxk}$ and biases $b \in \mathbb{R}^{1xk}$, where k is the number of classes.

3. Compute the Weighted Sum (Logits): $z = xW + b$

4. Apply the Softmax Function: Convert the logits into class probabilities:
$\text{Softmax}(z)_j = e^{(z_j)} / \Sigma_{i=1}^{k} e^{(z_i)}$, for j = 1, 2, ..., k

5. Prediction: Predict the class with the highest probability:
$\hat{y} = \arg\max_j \text{Softmax}(z)_j$

6. Loss Function (Cross-Entropy): For a true label $y \in \{1, ..., k\}$, compute the cross-entropy loss: $L = -\log(\text{Softmax}(z)_\gamma)$

7. Optimization: Use gradient descent or a variant (e.g., stochastic gradient descent) to update W and b by minimizing the loss.

8. Model Evaluation: Evaluate the model using accuracy, precision, recall, or F1-score on a validation/test set.

VII. RESULT AND DISCUSSION

As a result, the project offers users a useful and adaptable web application interface.Five important criteria related to sentiment analysis using the TF-IDF vectorization technique are used in this section to compare and assess four well-known machine learning models: K-Nearest Neighbors (KNN), Naive Bayes, Support Vector Machine(SVM),and Logistic Regression.
The bar graph's color-coding of each criterion is explained in more detail below:

1. Training Time:
(a) Naive Bayes (Score: 10): This model is the quickest to train because of its straightforward probabilistic methodology and feature independence assumption.
(b) Logistic Regression (Score: 8): For linear problems, this method offers decent optimization and reasonably quick training.
(c) SVM (Score: 5): Takes longer because it requires intricate high-dimensional space optimization.
(d) KNN (Score: 3): Low score due to the fact that it saves all training data and postpones computation until prediction time.

2. Time of Prediction:
(a) Logistic Regression Because of precalculated weights or probabilities, Naive Bayes (Score: 9) is incredibly effective at making predictions.
(b) SVM (Score: 8): Depending on the kernel, it performs well but a little more slowly.
(c) KNN (Score: 4): Slow prediction because it calculates the distances to each training point during inference.

3. TF-IDF performance:
Logistic Regression (a) SVM (Score: 9): Provides good classification performance and is very compatible with TF-IDF.
(b) Naive Bayes (Score: 7): Accuracy is somewhat constrained by feature independence assumptions, but it performs well.
(c) KNN (Score: 3): Underperforms because distance-based metrics are weakened by the high dimensionality of TF-IDF vectors.

4. Interpretability
(a) Logistic Regression (Score: 9): Highly interpretable due to linear model structure and feature weights.
(b) Naive Bayes (Score: 8): Provides probabilities and intuitive understanding of classification.
(c) SVM (Score: 7): Less interpretable, especially with non-linear kernels.
(d) KNN (Score: 4): Very difficult to interpret as decisions are based on raw distances.

5. Scalability

(a) Naive Bayes (Score: 10): Scales exceptionally well with large datasets.

(b) Logistic Regression (Score: 8): Efficient and scalable with proper regularization.

(c) SVM (Score: 6): Scalability becomes an issue with large datasets and complex kernels. Figure 1.2 shows Comparison Of Models for Sentiment Analysis using TF-IDF.

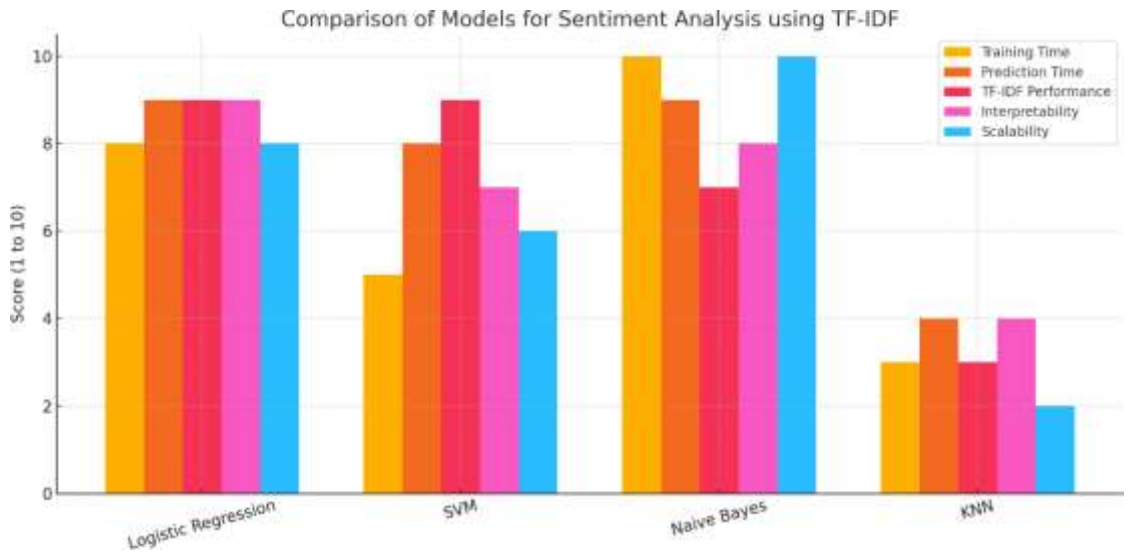

Figure 1.2 Comparison Of Models for Sentiment Analysis using TF-IDF.

6. The results from the graph indicate that Logistic Regression offers the best trade-off among performance, speed, and interpretability for TF-IDF-based sentiment analysis. SVM is a strong contender where accuracy is critical and computational resources are sufficient. Naive Bayes is ideal for rapid prototyping and baseline models. KNN,however, is not recommended for high-dimensional text classification due to scalability and performance limitations.

VIII. CONCLUSION

The effective application of sentiment analysis to social media data shows how machine learning can be used practically to interpret vast amounts of unstructured text. The project has found successful models in particular, Logistic Regression that provide precise and effective sentiment classification by using TF-IDF for feature extraction.This project has demonstrated how companies can gain important insights into public opinion, brand perception,and customer satisfaction by analyzing user sentiment. This type of analysis facilitates better audience engagement, strategic modifications, and more informed decision-making.The project's conclusions not only support the value of sentiment analysis in the current digital environment, but they also pave the way for further developments. These could involve adding sentiment analysis, multilingual support, real-time analytics, and deep learning models.

IX. REFERENCES

1. Fabrizio S. Andrea E., "Determining the Semantic Orientation of Terms through," October 31–November 5 2005..

2. Carmine C., Diego RFarah B., "Sentiment Analysis: Adjectives and Adverbs are better,"ICWSM Boulder, CO USA, 2006.

3. Lucas C., "Sentiment Analysis a Multimodal Approach,"Department of Comput ing, Imperial College London, September 2011.

4. Tayal, D. K Yadav. Sentiment Analysis on Social Campaign “Swachh Bharat Abhiyan” using unigram method. AI Society, 2016.

5. Farha Nausheen, Sayyada Begum, ”Sentiment analysis to predict election results using python”, Proceedings of the Second International Conference on Inventive Sys tems and Control (ICISC 2018) IEEE Explore Compliant, 2018.

6. Liu, B., ”Sentiment analysis and opinion mining Synthesis lectures on human lan guage technologies”, 2012.