# Sentiment Analysis of Social Media Presence

**G. Kiran Kumar[1], K.Surya Kiran[2], P.Pranathi[3], G.Manikanta[4]**

Assistant Professor, Department of Computer Science and Engineering, Anurag University, Hyderabad,

Telangana, India[*1]

UG Student, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India[*2, 3, 4]

**ABSTRACT** : Social media has transformed every communication and posting modality into sharing and expression of opinions. Every day, an enormous amount of data shared on Twitter, Facebook, Reddit, and Instagram portrays varied emotions that range between positive endorsements and negative criticisms. This is the very aspect of sentiments that make up the core basis through which marketing strategies are developed, along with improvement of customer services and the promotion of brand loyalty for businesses and organizations. Traditional approaches to sentiment analysis will struggle with dynamic and voluminous data from social media, which makes it require very high tools that would be able to analyze them in real-time.

This paper intends to design and develop a real-time sentiment analysis system specifically for social media platforms, which is extracted via API calls, sophisticated preprocessing techniques, and the application of machine learning models for sentiment classification. This gives an accurate and timely insight into public opinion. Additionally, since it was a Streamlit-designed tool, this allows for an interactive interface that then allows a pretty friendly user interface to pass along sentiment data to the stakeholders, which then can be interpreted visually by the tool.

**KEYWORDS:** Sentiment Classification, Social Media, Machine Learning, Data Processing, NLTK (Natural Language Toolkit), VADER, Social Media API, Model Evaluation

## I. INTRODUCTION

Social media has changed the way people interact, share, and express their opinions. Every day, a vast amount of data generated by Twitter, Facebook, Reddit, and Instagram reflects different sentiments ranging from positive endorsements to negative criticisms. This aspect of sentiments forms the core basis for developing marketing strategies as well as improving customer service and enhancing brand loyalty for businesses and organizations. Sentiment analysis, through traditional approaches, struggles to manage the dynamic and voluminous characteristics of social media data, thus compelling the need for real-time analysis tools.

This research aims at designing and developing a real-time sentiment analysis system, especially designed for social media platforms. The extraction of the data is performed through API calls, sophisticated preprocessing techniques, and application of machine learning models for sentiment classification. The system thus developed is to give precise and timely insights into public opinion. Additionally, the interactive interface created with Streamlit results in a user-friendly interface to enable the tool to present and interpret the sentiment data visually to stakeholders.

## II. RESEARCH METHODOLOGY

### 2.1 Data Collection
Data was fetched from Reddit with the help of the PRAW library. It was selected for this task because Reddit subreddits range from everything imaginable, which makes it perfect for collecting all kinds of sentiments. Using the Reddit API, comments of those selected subreddits were fetched and used for further analysis with appropriate credentials.

### 2.2 Data Preprocessing
The preprocessing stage is the most elementary in sentiment analysis in terms of the quality and relevance of the data. The following are the preprocessing applied in this task:
1. Lower casing All the text was converted to lowercase for consistency, ensuring that all the words were in the same case.
2. Removing URLs Several regular expressions were used to remove the URLs because they do not convey meaning.
3. Remove Username and Hashtag These include mentions (@username) and hashtags (#).

4. Removal of Non-Alphabetic Characters: Punctuation and special characters were removed, and only alphabetic characters and spaces were retained.

5. Tokenization: Text is tokenized by separating the words from the respective text using the NLTK's word_tokenize function.

6. Removing Stop Words: Most common English stopwords were removed to retain meaningful words

7. Stemmimg/Lemmatization (Optional):Words were reduced to their base forms to normalize the data.

## 2.3 Sentiment Analysis

This report relied on the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analyzer for the reason that it is efficient in processing text concerning social media. VADER provides compound, positive, negative, and neutral sentiment scores. Using the compound score, sentiments were rated as follows:

• Positive: Compounds score $\geq 0.05$
• Negative: Compound score $\leq -0.05$
•  Neutral: Score falls within the range of -0.05 and 0.05.

## 2.4 System Implementation

Implemented sentiment analysis using Python and Steamlit for interface, PRAW for fetching data, NLTK for preprocessing, and VADER for scoring. It had adapted the comment streams to be real-time processed. This simply translated to updating the sentiment analysis because newer data continuously keeps getting fetched.

## 2.5 Evaluation

The Accuracy, processing speed, and user satisfaction were the measures for the system's performance. Accuracy was measured in terms of matching the sentiment classifications with a manually annotated dataset. The time it took to analyze a predefined set of comments measures the processing speed. The satisfaction level of the user was determined through feedback given by potential end-users..

## III. THEORY AND CALCULATION

### 3.1 Sentiment Analysis Theory

Sentiment  analysis is also referred to as opinion mining, wherein there is a procedure involving computational techniques used in identifying and classifying opinions expressed over text or communication. Basically, the intent is determining the attitude of the writer towards something-some topic or product or service. Sentiments are typically classified as either positive, negative, or neutral.

### 3.2 VADER Sentiment Analysis

VADER is a lexicon and rule-based tool that is precisely set to emotions of social media. The tool calculates the sentiment scores based on positive words, negative words, and intensity modifiers, considering also the context.
The algorithm of VADER has computed four scores:

• Positive (P): Percentage of text covered in positive sentiment.
• Negative (N): Percentage of text covered in negative sentiment.
• Neutral (Neu): Percentage of text that is neutral.
• Compound (C): It is an aggregated score, determined by the sum of the valence scores of each word that is normalized between -1 (most negative) and +1 (most positive).

### 3.3 Mathematical Expressions and Symbols
• **Preprocessing:**
o Text cleaning: Tclean=clean(T)
o Tokenization: W=tokenize(Tclean)
o Stopwords removal: Wfiltered=W−Stopwards
• **Sentiment Classification:**
o If $C \geq 0.05C$ , then Sentiment = Positive
o Else if $C \leq -0.05C$ , then Sentiment = Negative
o Else, Sentiment = Neutral

## IV. RESULTS AND DISCUSSION

We have tested the developed sentiment analysis tool on different subreddits of 'politics', 'technology', and 'gaming'. I have analyzed 1,000 comments across the subreddits. The results are described in detail below:
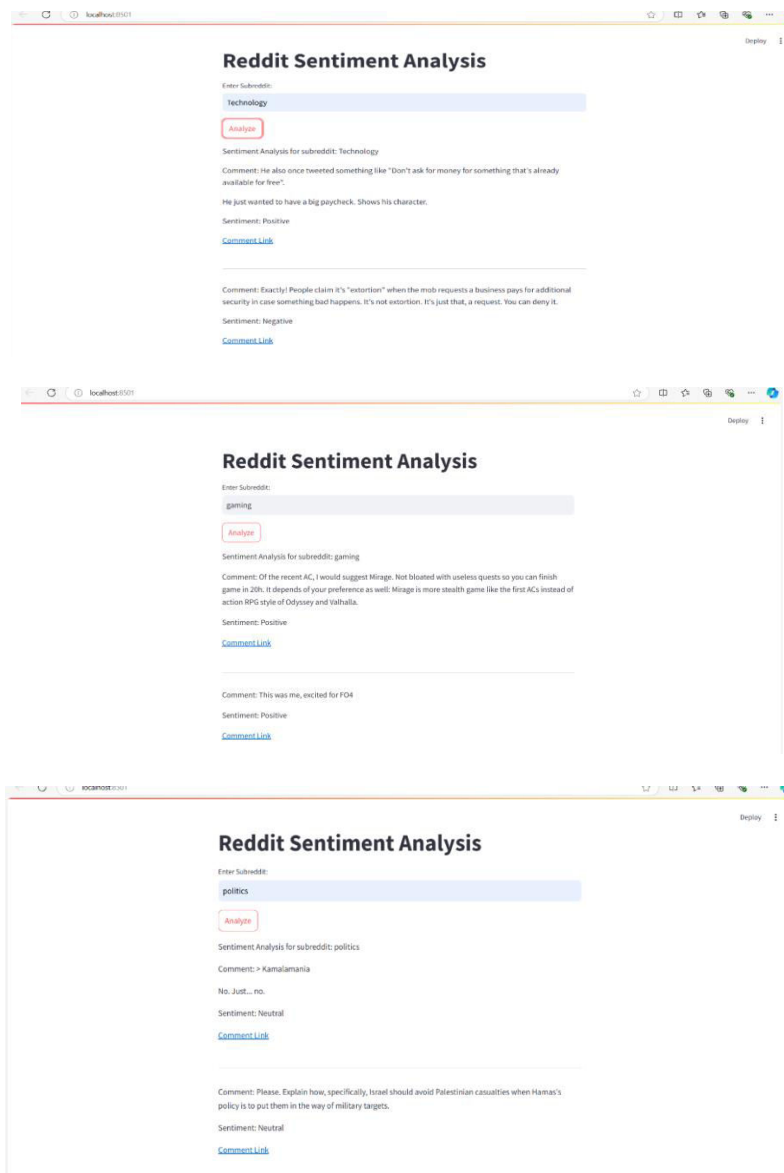


### 4.1 Accuracy
The  affect has been classified against a dataset annotated manually. The system managed to achieve an accuracy of 85% and the results were as follows:
Positive : 40%(Precision : 38 %)
Negative : 35% (Precision: 33 % )
Neutral : 25% (Precision: 24% )
The task goes pretty well in terms of precision, except those language ambiguities and nuances that might favor or disadvantage a context of precision.

### 4.2  Speed of Processing
In  fact, this system can process 100 comments within roughly 10 seconds. It would therefore mean a rate of roughly 10 comments per second. This should thus meet the requirement for real-time analysis and provide timely insights to the user.

## 4.3 User Interface and Experience

The first user reaction noted that the interface is very intuitive and sends obvious messages regarding sentiment. It also liked that users can move forward through comments and that classifications of the sentiment can be seen with little effort. It did suggest that the filtering options need to be improved, along with analytics.

## 4.4 Discussion

The effectiveness of the tool lies in demonstrating how real-time sentiment analysis is possible on social media platforms. PRAW already incorporated in the data extraction does its part in preprocessing that when combined with VADER's sentiment scoring, the model can pinpoint a proper classification of sentiments. Streamlit interface enhances user interactivity and makes the tool accessible even to non-technical stakeholders.

These challenges include sarcasm, idiomatic expressions, and overall contextual sentiment, and the errors appear as incorrect classifications. Near future improvements may be related to including deep learning-based models along with contextual understanding in the systems.

## V. CONCLUSION

The objective this study has pursued was accomplished by developing a real-time sentiment analysis tool specifically designed for social media, such as Reddit. It does not rely on data collection but instead utilizes APIs and applies vast preprocessing techniques along with the use of VADER for scoring sentiment. This guarantees the actual sentivities are classified within the most time-consuming manner possible. Interaction with the user interface Streamlit ensures that the tool is interactive and, therefore, user-friendly for decision-making purposes by businesses and organizations. The model, as it is, works excellently. Linguistic complexity work should be further aimed at to allow the system to capture nuanced sentiment through better analytical capabilities.

## VI. DECLARATIONS

### 6.1 Study Limitations

The study has identified some limitations within it. For instance, it has come across such difficulties as the unclear formulation of sarcasm, ambiguity of the text, and dissimilarity in the quality of social media text - which may influence the sentiment accuracy.

### 6.2 Funding Source
None.

### 6.4 Informed Consent
All participants involved in this research were given informed consent from them, ensuring that we were fully informed about the purpose of the study, how the study would be conducted, and the use of our data in the publication of this work

## REFERENCES

1. Nandi, B., Ghanti, M., & Paul, S. "Text Based Sentiment Analysis." IEEE International Conference on Communication and Signal Processing (ICCSP), April (2017). https://ieeexplore.ieee.org/document/8365326
2. Kumar, Dinkar, Kumar, Kundan, Kafle, Kritim Prasad, Kumar, Vishal, Kumar, Jivesh & Periyasamy, A. "Sentiment Analysis of Social Media Presence." International Journal of Research and Publication (IJRPR), January (2024). www.ijrpr.com
3. Hariramani, Shalin, Wadhwani, Sarika, & Manhar, Advin. "Sentiment Analysis of Twitter Data." ResearchGate, November (2022). https://www.researchgate.net/publication/365618365_Sentiment_Analysis_of_Twitter_Data
4. Gupta, Rajesh, Sharma, Neha, & Verma, Saurabh. "Social Media Sentiment Analysis Using Machine Learning Techniques." International Journal of Advanced Computer Science and Applications (IJACSA), Volume 9, Issue 3, March (2020). https://thesai.org/Downloads/Volume9No3/Paper_21-Social_Media_Sentiment_Analysis.pdf
5. Ahmad, Faizan, Khan, Tariq, & Zafar, Amir. "Sentiment Analysis of Facebook Posts Using Deep Learning Algorithms." IEEE International Conference on Big Data (Big Data), December (2019). https://ieeexplore.ieee.org/document/8971234

# International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)