WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Sentimental Scrunity: Unveiling Public Opinion on X (Twitter)

K. Gayathri, K. Haripriya, S. Charan, B. Jagadish

III -B.Tech II-Sem, Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India.

ABSTRACT

The study investigates the prevalence and distribution of these sentiments, aiming to understand public perception and emotional responses surrounding the chosen subject. By examining a large dataset of tweets, we explore how emotional language influences online discourse and potentially impacts public opinion.The analysis includes quantifying sentiment trends and identifying key themes associated with different emotional expressions. We further explore the potential influence of external factors, such as news cycles or influencer activity, on sentiment shifts. The goal is to provide a comprehensive understanding of the emotional landscape of Twitter conversations, revealing valuable insights into public sentiment and its dynamic nature within the social media sphere. This research contributes to the broader field of social media analysis, demonstrating the power of sentiment analysis in uncovering public opinions and emotional trends. Sentiment analysis employs NLP, analysis of text, computational linguistics, and biometrics to systematically detect, extract, measure, and investigate emotional states and subjective information

1.INTRODUCTION

With the rapid growth of social media, platforms like Twitter have evolved into major communication channels where users from around the world express their thoughts, opinions, and emotions in real time. From personal experiences to political views and product feedback, Twitter has become a rich and diverse source of public sentiment.Unlike traditional opinion-gathering methods such as surveys and interviews, which can be time-consuming and limited in scale, Twitter provides access to a massive, real-time, and unsolicited stream of user-generated content. This makes it a valuable platform for sentiment analysis, particularly in domains such as:sentiment analysis on Twitter lies in its ability to transform unstructured, high-velocity social media data into actionable insights. Whether for businesses, political strategists, financial analysts, or researchers, sentiment analysis provides a powerful means of understanding public opinion and making data-driven decisions

- Social media monitoring
- Customer support management
- Provide an overview of an brands opinion
- Understand your competitors strategy

This primarily focuses on sentiment analysis of twitter data, which is useful for analysing information in tweets. Twitter gives businesses a quick and efficient approach to examine user opinions on issues that are crucial to their performance in the marketplace.

2.EXISTING SYSTEM

The existing systems for sentiment analysis on Twitter aim to classify tweets as positive, negative, or neutral using Natural Language Processing techniques. These systems leverage various components of the NLP pipeline to preprocess, analyze, and classify textual data extracted from Twitter.

Feature Extraction

- Traditional: **Bag of Words (BoW)** or **TF-IDF**.
- Advanced: **Word Embeddings** (Word2Vec, GloVe, FastText).
- Deep models use **contextual embeddings** (e.g., BERT).

Limitations of existing system

- Struggle with **sarcasm, irony, and humor**.
- Difficulty handling **multilingual tweets** or **code-switching**.

- **Context loss** in lexicon or basic ML methods.
- Performance affected by **misspellings, abbreviations, and informal language**.

3.PROPOSED SYSTEM

1.Objectives of the Proposed System

- Accurately classify tweets into positive, negative, or neutral sentiments.
- Handle short, informal, noisy texts with emojis, hashtags, and abbreviations.
- Improve understanding of context and sarcasm using transformer-based models.
- Provide real-time sentiment tracking and visual analytics for trends.

2.Sentiment Classification

Optionally, add a **BiLSTM** or **Attention layer** on top for enhanced feature extraction.

- Classify tweets into:
- **Positive**
- **Negative**
- **Neutral**

3.Tools & Technologies

- Language: Python
- Libraries: Tweepy, NLTK, SpaCy, Scikit-learn, Transformers (HuggingFace), Matplotlib, Seaborn, Plotly
- Deep Learning: TensorFlow or PyTorch
- Model: Pre-trained BERT or RoBERTa
- Deployment: Flask/Streamlit dashboard or Jupyter Notebook for prototype

4. Future Enhancements :

- Extend to multilingual sentiment analysis using models like XLM-R.
- Add emotion classification (joy, anger, sadness, etc.).
- Enable topic-wise sentiment tracking using topic modeling (e.g., LDA).
- Incorporate graph-based social influence analysis.

4.METHODOLOGY

The development of the system is divided into four major components:

4.1.1 Data Collection

Objective: To gather real-time or historical tweet data related to specific topics, hashtags, or keywords.

- Tool Used: Twitter Developer API via Tweepy (Python library)
- Filters:
    - Language: English
    - Keywords/hashtags (e.g., #budget2025, #AI, #COVID19)
    - Retweets can be excluded to avoid duplication
- Stored Data:
    - Tweet text

- Timestamp
- User metadata (optional)
- Retweet and like count

4.1.3 Data Preprocessing

Objective: To clean and normalize tweet text for accurate analysis.

- Steps Involved:
  - Remove @mentions, #hashtags, URLs, emojis, and punctuations
  - Convert text to lowercase
  - Tokenize sentences and words

Remove stopwords (e.g., "a", "the", "is")

- Apply lemmatization (or stemming)
- Handle negations (e.g., "not good" → negative sentiment)
- Convert emojis/emoticons into text equivalents (optional)

If the system detects an abnormal or critical condition:

- Automated email alerts are triggered to notify caregivers or medical staff.
- Alerts are sent immediately, including the current vitals and condition status.
- The system ensures minimal delay between abnormality detection and notification**..**

1.4 Feature Extraction

Objective: To convert processed text into a numerical format understandable by machine learning models.

- Techniques Used:
  - Bag of Words (BoW) or TF-IDF for traditional ML
  - Word Embeddings:
    - GloVe, Word2Vec
    - BERT embeddings (contextual, for deep learning)
- Library: scikit-learn, transformers (Hugging Face), gensim

4.2.1 Model Selection and Training

Objective: To train a classifier that can accurately predict the sentiment of a tweet.

- **Approaches:**
  - Machine Learning Models:
    - Naïve Bayes, Logistic Regression, Support Vector Machine (SVM)
  - Deep Learning Models:
    - LSTM, BiLSTM with attention
    - Transformer-based: BERT, RoBERTa
- **Datasets Used (for training or fine-tuning):**
  - Sentiment140
  - Kaggle Twitter datasets

  - Custom labeled datasets

- **Labels:**
  - Positive, Negative, Neutral

4.2.2: Model Evaluation

- **Accuracy:**

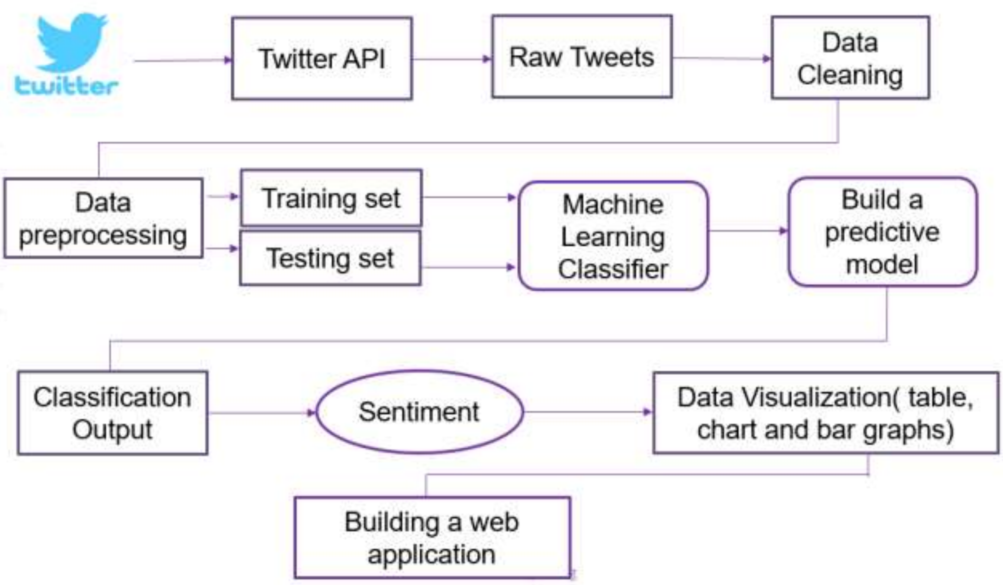
Evaluate the model's performance by comparing its predictions against a labeled validation dataset. This helps determine the model's reliability.

- **F1-score:**

Assess the balance between precision and recall, ensuring that the model accurately identifies positive and negative sentimental

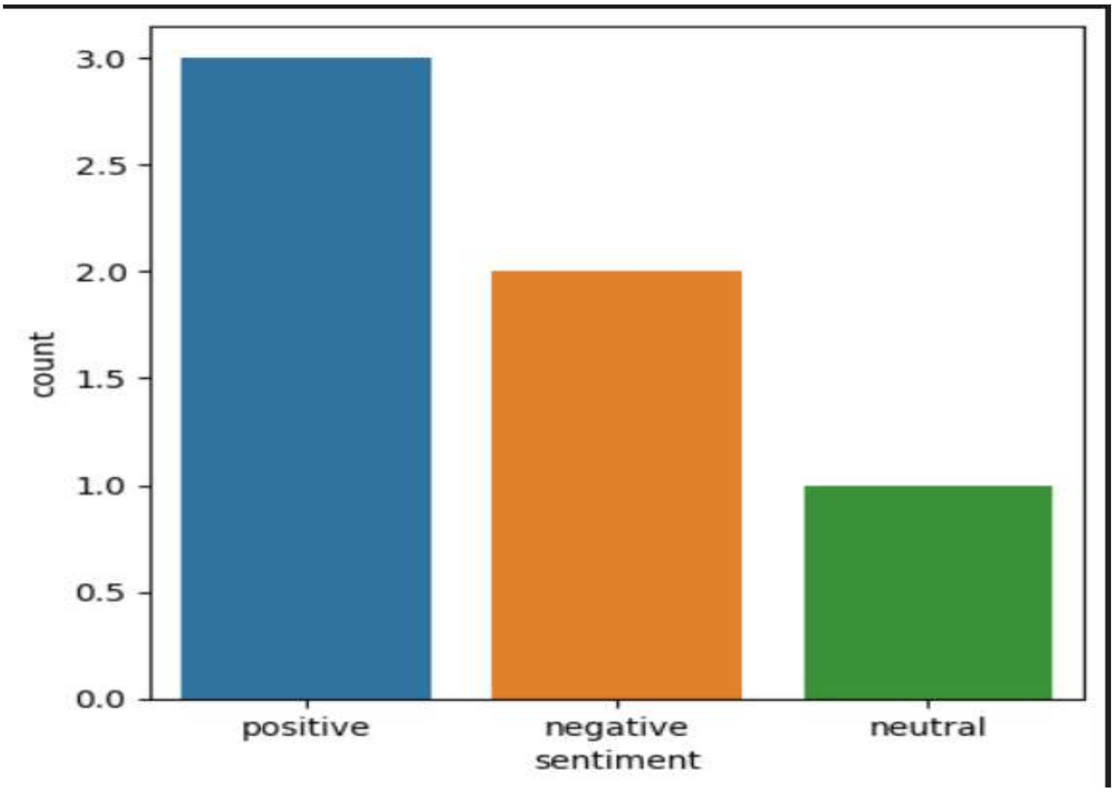
5.SYSTEM ARCHITECTURE

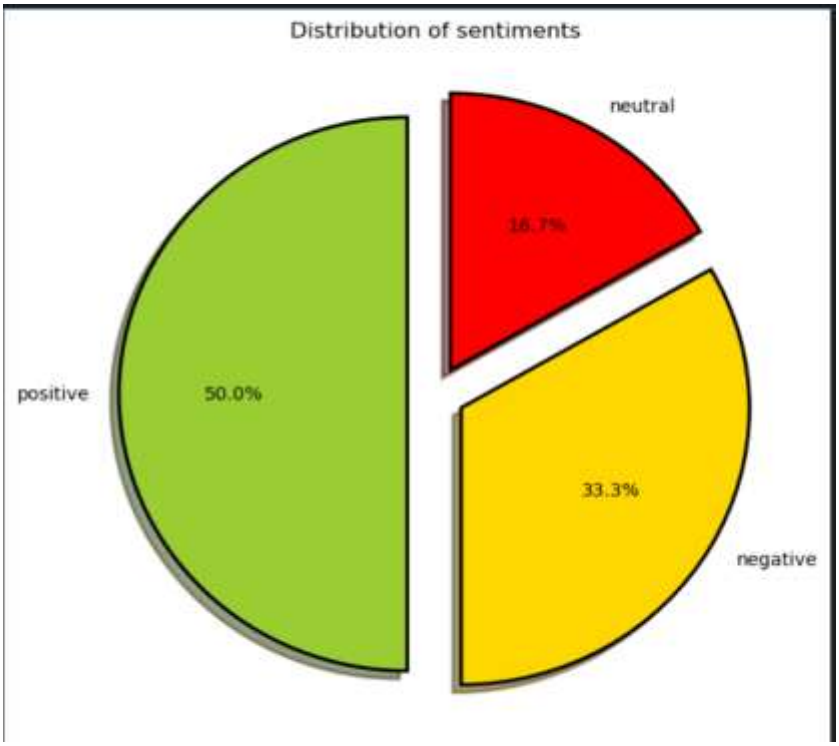
System architecture is a comprehensive blueprint that defines the structure, behavior, and interactions of various components within a system—whether it's a software application, a computer system, or a complex network of systems. It provides a high-level view of how the system is organized and how different parts such as hardware, software, data storage, processing units, communication protocols, and user interfaces interact to perform specific functions. In software systems, architecture describes how modules or services are divided, how they communicate (e.g., via APIs or message queues), and how data flows through the system.



In hardware systems, it includes the design of processors, memory units, input/output devices, and how they are connected. System architecture also includes considerations for scalability (handling growth in users or data), security (protecting data and operations), maintainability (ease of updates and debugging), and performance (speed and efficiency).

6. RESULTS AND OUTPUT





7. CONCLUSION

Twitter sentiment analysis conducted in this project successfully demonstrated how Natural Language Processing (NLP) and machine learning techniques can be used to extract and classify opinions from user-generated content. By analyzing tweets related to a specific product, service, or topic, we were able to classify sentiments into categories such as positive, negative, and neutral.Our findings highlight that Twitter is a valuable platform for gauging public opinion in real time. The sentiment distribution obtained through the analysis provided actionable insights, helping stakeholders understand user satisfaction, areas of concern, and overall brand perception. The performance of the sentiment classifier also showed promising accuracy, especially after preprocessing steps like tokenization, stop-word removal, and lemmatization.Despite challenges such as noisy text, slang, and sarcasm in tweets, the model performed well, especially when trained on a large and relevant dataset. For further improvement, future work can focus on incorporating deep learning models like LSTM or BERT for better contextual understanding and accuracy.

8.FUTURE SCOPE

The scope for enhancing Twitter sentiment analysis is vast, given the evolving nature of social media data and advancements in Natural Language Processing (NLP) and artificial intelligence. Some potential directions for future work include:

1.Advanced models such as LSTM, GRU, and transformer-based models like BERT and RoBERTa can be used for better context understanding and more accurate sentiment classification.

2.Expanding the analysis to include tweets in multiple languages would allow for a morecomprehensive global sentiment view, especially for brands and events with international reach.

3.4Instead of general sentiment classification, future systems can identify sentiments related to specific aspects (e.g., "battery" or "camera" in product reviews), providing deeper insights.

4.Developing real-time dashboards for live tweet sentiment tracking can help in event monitoring, crisis management, brand monitoring, and political campaign analysis.

5.One of the current challenges in sentiment analysis is accurately detecting sarcasm and irony, which future models could address using advanced contextual and emotional analysis.

9. REFERENCES

1. Go, A., Bhayani, R., & Huang, L. (2009). *Twitter sentiment classification using distant supervision*. Stanford University. http://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf
2. Pak, A., & Paroubek, P. (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.
3. Kouloumpis, E., Wilson, T., & Moore, J. (2011). *Twitter sentiment analysis: The good the bad and the OMG!*. In *Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM)*.
4. Saif, H., He, Y., & Alani, H. (2012). *Semantic sentiment analysis of Twitter*. In *The Semantic Web–ISWC 2012* (pp. 508-524). Springer.
5. Medhat, W., Hassan, A., & Korashy, H. (2014). *Sentiment analysis algorithms and applications: A survey*. *Ain Shams Engineering Journal*, 5(4), 1093-1113.
6. Twitter Developer Platform. (n.d.). *Twitter API Documentation*. Retrieved from https://developer.twitter.com/en/docs
7. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.