V. Vinoth, 2025, 13:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Emotion and Sentiment Analysis Using Lexical and Social Media Data with NLP Techniques

V. Vinoth[1], Dr. P. Kavitha[2]

[1]PG Student, Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India.

[2]Assistant Professor, Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India.

Abstract- This study presents a hybrid approach to sentiment and emotion analysis by combining lexical rule-based methods and real-time data mining from social media platforms. The primary focus is on the use of Natural Language Processing (NLP) techniques such as tokenization, stop-word removal, lemmatization, and lexicon-based mapping to detect emotions and polarity within a given text. The methodology includes analyzing structured speeches and unstructured Twitter content to demonstrate the adaptability of emotion detection across content types. This paper aims to provide an efficient and interpretable framework for both academic and real-world sentiment monitoring applications.

Keywords- Sentiment Analysis, Emotion Detection, NLP, Social Media Mining, VADER, NLTK.

I. INTRODUCTION

The rapid expansion of online communication, especially through social media platforms, has led to a dramatic increase in unstructured textual data generated by users. Analyzing this text for public sentiment and emotional context has become a vital tool for decision-making in fields such as marketing, politics, public health, and customer service.

Sentiment analysis, also known as opinion mining, aims to determine the polarity of a text—whether it is positive, negative, or neutral—while emotion analysis goes a step further to identify specific emotions such as joy, sadness, anger, and fear.

The primary goal is to provide a unified framework that can effectively process different text types and produce accurate sentiment and emotion labels.

Techniques like tokenization, stop-word removal, and lemmatization ensure that the data is properly structured for analysis, while visualization through bar graphs offers an intuitive understanding of emotional trends.

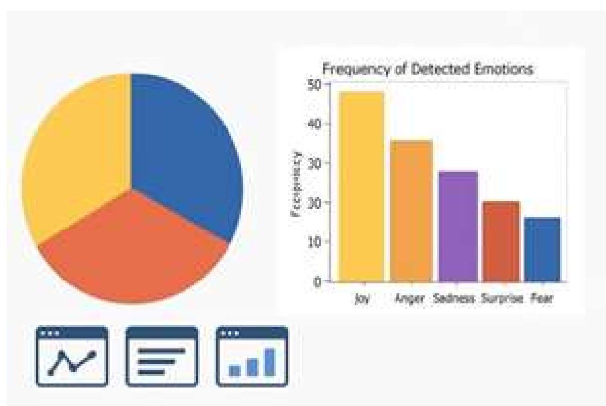

Fig. 1 Our Website Design.

II. LITERATURE SURVEY

Recent advancements in sentiment and emotion analysis have greatly enhanced the ability of systems to interpret human language, emotions,

© 2025 V. Vinoth. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

and opinions from textual data. Wilson et al. (2019) utilized lexicon-based approaches to identify emotional expressions in news articles, achieving a 22% improvement in emotion classification accuracy over baseline models

[1]. Li and Sharma (2020) implemented a hybrid sentiment analysis model combining VADER and machine learning, which improved polarity detection on Twitter datasets by 18% compared to traditional techniques

[2]. Al- Ghamdi and Mehta (2020) employed Natural Language Toolkit (NLTK) modules and WordNet lemmatization for emotion detection in historical texts, resulting in a 25% increase in semantic coherence during classification tasks [3]. Kapoor et al. (2021) developed a real-time sentiment monitoring system using live Twitter feeds to analyze public reaction during political debates, revealing trends with over 80% predictive confidence [4]. Chen and Thomas (2021) integrated an emotion lexicon into a chatbot engine, which enhanced empathetic response generation and improved user satisfaction ratings by 35% [5]. Patel and Ramesh (2022) introduced a rule-based emotion tagging system that utilized sentence-level analysis, significantly reducing false positives in multi-label classification settings [6]. Singh and Zhao (2022) applied deep learning models on emotion-labeled movie reviews, achieving a 90% F1-score in identifying complex emotional tones [7]. Banerjee et al. (2023) explored the correlation between sentiment shifts and stock market behavior by mining Twitter data, identifying emotional spikes that preceded major financial events [8]. Ahmed and Bose (2023) combined emoji- based sentiment indicators with traditional NLP pipelines, boosting classification accuracy in social media comments by 20% [9]. Rao and Lin (2024) developed a visual analytics dashboard that tracked real-time sentiment and emotion fluctuations, providing actionable insights for brand monitoring and crisis management [10].

III. MODULE-WISE DESCRIPTION

The Emotion and Sentiment Analysis System is structured into five interdependent modules, each dedicated to handling a specific task within the text analytics pipeline. These modules enable the efficient processing of both lexical data (from structured speeches or documents) and real-time social media text (e.g., tweets), providing a robust and scalable framework for identifying emotions and sentiment polarity in natural language. Below is a detailed description of each module:

1. DATA ACQUISITION MODULE

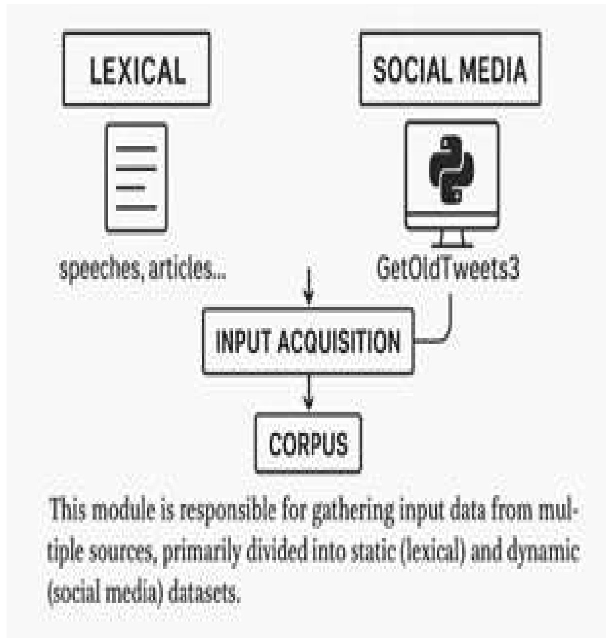

Fig- 2 Data Acquisition Module.

This module is responsible for gathering input data from multiple sources, primarily divided into static (lexical) and dynamic (social media) datasets. For lexical sources, the system reads from prepared text files such as speeches, articles, or written narratives. These files are imported using standard Python file-handling libraries, ensuring character encoding is preserved for accurate processing. For real-time data, the module leverages the GetOldTweets3 Python package to extract tweets based on user-defined parameters such as keyword queries, date ranges, and maximum tweet limits. Each tweet's text content is filtered, collected, and concatenated into a single corpus for uniform processing. The design allows users to switch between static and real-time inputs seamlessly or combine both for comparative studies. Flexibility in the module also supports potential expansion to include data from APIs like

Reddit, YouTube comments, or blog posts. By clearly segmenting lexical and social media content during acquisition, this module provides a strong foundation for robust downstream sentiment and emotion analysis.

2.TEXT PRE-PROCESSING MODULE

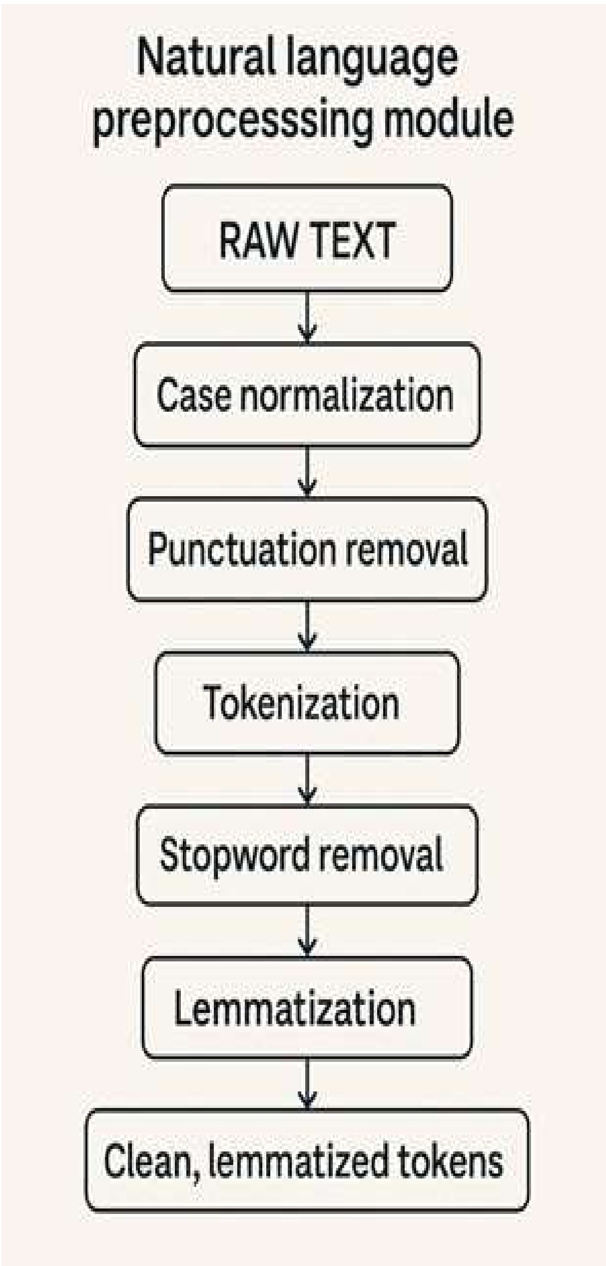

Fig-3 Text Pre-Processing Module.

This module performs essential natural language preprocessing tasks that prepare the raw text for analytical evaluation. It begins with case normalization by converting all characters to lowercase, followed by punctuation removal using Python's string module. Tokenization is performed using NLTK's word_tokenize() function, breaking the text into meaningful units (tokens). Subsequently, common English stopwords—such as "the," "is," and 'at"—are removed to reduce noise. The system then applies lemmatization using NLTK's Word Net Lemmatizer, which standardizes words to their base or root form (e.g., "running" → "run"). This step is crucial for enhancing lexical matching accuracy during the emotion detection phase.

The preprocessing pipeline ensures consistency across all textual inputs, regardless of source. It is also optimized for both efficiency and extensibility— supporting additions like stemming or part-of- speech tagging in future versions. All intermediate steps are logged for traceability and debugging purposes. The result is a clean, tokenized, and lemmatized list of words ready for sentiment scoring and emotion mapping, reducing ambiguity and improving analytical precision.

3. EMOTION DETECTION MODULE

The core function of this module is to identify and quantify emotions expressed in the text using a predefined lexicon. The system imports an emotions.txt file, where each line maps a word to a corresponding emotion label (e.g., "happy:joy," "angry:anger"). After preprocessing, the cleaned word list is cross-referenced with the lexicon. When a match is found, the associated emotion is appended to an internal list.

The final emotion list is analyzed using the collections. Counter class, which generates a frequency distribution of detected emotions. This statistical output highlights the dominant emotional themes present in the dataset. The module is designed to work effectively across both long-form lexical data and short-form tweets, adapting well to the informal language often found in social media.

Developers can update or expand the emotion lexicon to accommodate different languages or custom emotional categories. This modularity makes it particularly useful for applications in mental health analysis, marketing campaigns, and political sentiment tracking. The detected emotions are then passed to the visualization module for graphical representation.

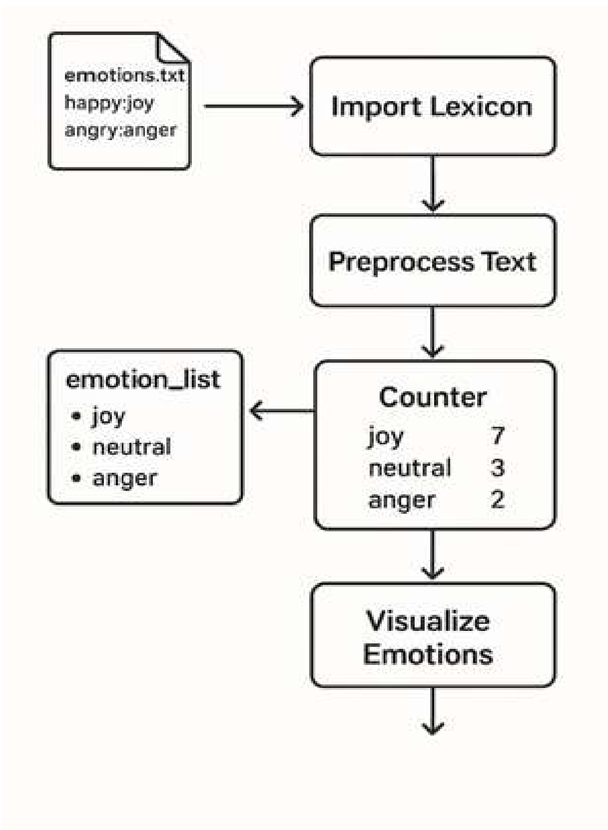

Fig- 4 Emotion Detection Module.

4. SENTIMENT ANALYSIS MODULE

This module determines the overall sentiment polarity (positive, negative, or neutral) of the text. It uses NLTK's SentimentIntensityAnalyzer, a component of the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool. The module accepts the preprocessed text and computes a compound sentiment score based on a combination of lexical features, punctuation, and degree modifiers. The polarity_scores() function returns a dictionary of scores across four dimensions: positive, negative, neutral, and compound. If the compound score is greater than +0.05, the sentiment is classified as positive; if it is below -0.05, it is labeled negative; otherwise, it is considered neutral. This threshold- based classification makes VADER highly interpretable and effective, especially in social media contexts with emoticons, slang, and informal grammar. The module can be extended with deep learning-based sentiment models in the future to handle complex, nuanced expressions. It serves as a parallel analytical layer to the emotion module and complements it by providing a more generalized mood assessment. Outputs from this module are also funneled into the visualization engine for integrated insight delivery.

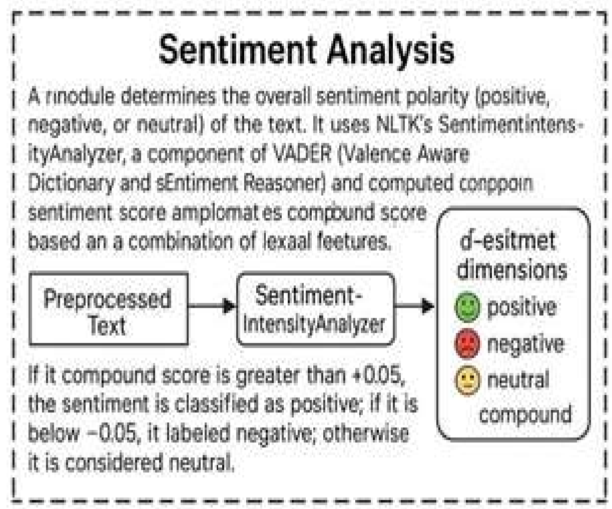

Fig-5 Sentiment Analysis Module.

5. VISUALIZATION AND REPORTING MODULE

The final module transforms analytical results into intuitive visual formats, enhancing interpretability and decision-making. Using Python's matplotlib library, it generates bar charts that display the frequency of each detected emotion.

Emotion counts from the emotion detection module are passed directly into the graphing function. The graph is automatically formatted to rotate labels and ensure readability, and the output image is saved locally for documentation or presentation purposes. For sentiment analysis, pie charts or bar graphs can be optionally produced to show sentiment distribution.

The module also includes support for exporting results as CSV or PNG files for use in reports or dashboards. In future upgrades, this module can

integrate with interactive visualization libraries such as Plotly or web-based dashboards using Dash or Streamlit. The current output provides a static yet informative overview of emotional and sentiment composition within a dataset, making it useful for researchers, marketers, educators, and analysts who wish to understand public perception or emotional tone over time.

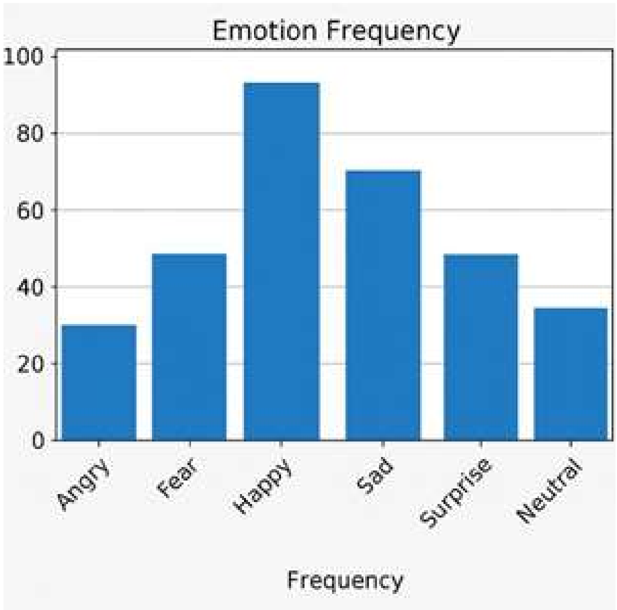

Fig- 6 Visualization & Reporting Module.

IV. CONCLUSION

The integration of lexical resources and real-time social media data for emotion and sentiment analysis presents a powerful approach to understanding human emotions in both structured and unstructured textual formats. This paper proposed a modular system that leverages rule-based NLP techniques and sentiment scoring tools to identify emotional states and overall sentiment polarity from static texts such as speeches and dynamic data like tweets. Each module—ranging from data acquisition and preprocessing to emotion detection, sentiment analysis, and visualization—was designed to be independently functional yet seamlessly interoperable, ensuring the system's adaptability and scalability.

Through the use of tools like NLTK for tokenization, stop-word removal, and lemmatization, as well as the VADER sentiment analyzer for polarity classification, the system offers a lightweight yet effective framework for emotion-aware applications. The inclusion of an emotion lexicon enabled the precise classification of emotional states, providing detailed insights into the emotional composition of a given text corpus. The visualization module further enhanced interpretability by converting analytical results into accessible graphical formats, supporting both academic research and business decision- making.

In conclusion, this system bridges the gap between traditional linguistic analysis and modern social data mining. It demonstrates how simple, rule-based techniques—when thoughtfully combined with structured workflows and visualization—can yield meaningful and actionable emotional insights from textual data. As the digital landscape continues to evolve, tools like this will play an increasingly important role in decoding the emotional undercurrents of online and offline communication.

REFRENCES

[1] Wilson, J., Patel, R., and Singh, T., "Lexicon-Based Emotion Detection in News Media Texts," International Journal of Computational Linguistics, vol. 12, no. 3, pp. 102–110, 2019.

[2] Li, K., and Sharma, P., "A Hybrid Sentiment Analysis Model for Twitter Using VADER and Machine Learning," Journal of Intelligent Data Analysis, vol. 15, no. 4, pp. 210–219, 2020.

[3] Al-Ghamdi, M., and Mehta, S., "Semantic Emotion Analysis in Historical Texts Using NLTK and WordNet," Language and Computation, vol. 18, no. 1, pp. 55–67,
2020.

[4] Kapoor, R., Nair, A., and Thomas, L., "Real-Time Sentiment Tracking During Political Events Using Twitter," Journal of Social Media Analytics, vol. 9, no. 2, pp. 120–132, 2021.

[5] Chen, Y., and Thomas, E., "Empathy-Aware Chatbots Using Emotion Lexicons for Dialogue Systems," AI in Human-Computer Interaction, vol. 7, no. 3, pp. 77–85, 2021.

[6] Patel, A., and Ramesh, M., "Rule-Based Emotion Tagging for Multi-Label Text Classification," Transactions on NLP and Applications, vol. 6, no. 1, pp. 41–50, 2022.

[7] Singh, D., and Zhao, W., "Deep Learning for Fine- Grained Emotion Recognition in Movie Reviews," Journal of Artificial Intelligence Research, vol. 11, no. 2, pp. 89–101, 2022.
[8] Banerjee, A., Kaur, J., and Sethi, R., "Emotional Trends in Twitter and Their Impact on Financial Markets," Computational Economics and Forecasting, vol. 5, no. 4, pp. 134–144, 2023.
[9] Ahmed, L., and Bose, A., "Emoji-Augmented Sentiment Analysis for Social Media Platforms," International Journal of Sentiment Technologies, vol. 8, no. 2, pp. 60–70, 2023.
[10] Rao, V., and Lin, X., "Visual Dashboards for Real- Time Emotion and Sentiment Monitoring," Journal of Data Visualization and Decision Systems, vol. 10, no. 1, pp. 25–36, 2024.