International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 9, pp 86-93, September-2024

https://doi.org/10.47001/IRJIET/2024.809011

SentiMix: A Unified Approach to Comprehensive Sentiment Analysis

[1]Kresha J. Shah, [2]Heta C. Shah, [3]Durva H. Patel, [4]Nilesh Marathe

[1,2,3,4]Department of Computer Science and Engineering (Data Science), SVKM's Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

***Abstract* - In the present era of digitalization, sentiment analysis plays a significant role for understanding public opinion, customer feedback, and social trends on different media channels. The sentiments are equally important for both businesses and individuals as these are now expressed through text as well as emoticons and images. With the vast growth of textual and visual data alongside emotions on the web, a need for an all-round sentiment analysis model has risen sharply. However most of the existing methodologies turn out to be myopic; they lack the ability to cohesively analyze sentiment from all three sources (text, emoticons, images). Our model seeks to address this limitation by adopting various machine learning techniques that enable seamless processing and interpretation of sentiment from diverse data repositories. We are proposing a comprehensive sentiment analysis tool that combines advanced techniques to perform sarcasm detection, rule-based and machine learning models for text sentiment analysis, emotion mapping for emoji-based sentiment analysis, and for image-based sentiment analysis. By integrating advanced machine learning techniques we look forward not just for providing but also packaging sophisticated details about public sentiment which will be visually delivered (like graphical reports) so that decision makers can better understand and act upon them.**

Keywords: Sentiment Analysis, Comprehensive Sentiment Analysis, SentiMix, Data Science, Customer feedback.

I. INTRODUCTION

As far as human communication goes, in today's digital era, it has changed a lot. Humans, now share their thoughts and opinions on the internet, i.e., online for example, social media, forums, and reviews. But understanding this opinion is difficult because we do not know the underlying meaning of the text. This is important because think about how reviews on any shopping websites affect what you decide to buy, how social media can shape you to see politicians during elections.

Communication nowadays does not just include words. It also includes emojis, images, etc. that can help us understand how people really feel. But we do not have a tool that can understand all these different forms of communication together. We need a tool that can understand emotions from words, emojis, and images together at once. This will help us understand people's sentiments in a better manner, especially when there are so many ways people express themselves on the internet.

The field of sentiment analysis is always shifting and does not take into account any challenge posed by the lack of capabilities for current methodologies to identify that. They struggle with being able to accurately detect sentiments across different data modalities because they are often unable to understand or interpret subtle cultural references, sarcasm or even emojis which are used as emotional indicators in text content. It is indeed challenging— a difficult task to say the least— trying to find corresponding emotional states behind countless collection of diverse emojis.

But we are not blind to these limitations; our paper sees them and thus proposes a unique sentiment analysis model that aims at closing this technological gap. We do not hold back; our approach is an integration of sophisticated methodologies. This includes sarcasm detection algorithms, rule-based machine learning models colliding for text sentiment analysis, and crossing with techniques for emoji emotion mapping. Still not satisfied with what we have on board, we go further by boosting our model: image-based sentiment analysis comes into play. We employ various algorithms meant to draw out emotions from visual contents as part of this additional enhancement in place.

The primary aim of our model is to integrate these different advanced methods in a way that would make the analysis of sentiment very smooth and easy to take from any source data. This will lead us to have a complete understanding of public sentiments. With this study, we hope to bring out more detailed findings on customer feedback and social trend as well as public opinions— which we plan to present using graphics that can easily be understood for decision-making. Section 1 provides a comprehensive Literature Review, followed by Section 2 outlining the Objectives and motivation, Section 3 defining the Problem statement, Section 4 detailing the Proposed Methodology,

ISSN (online): 2581-3048
Volume 8, Issue 9, pp 86-93, September-2024
https://doi.org/10.47001/IRJIET/2024.809011

Section 5 presenting the Result Analysis, and concluding with a References section.

Following this will be a comprehensive review that looks at the limitations, existing literature around methodologies employed during sentiment analysis.

II. LITERATURE SURVEY

To gain an understanding of the existing body of knowledge and research related to sentiment analysis, several papers were reviewed.

The paper written by V.Hymavathi2 [1] explores various sentiment analysis techniques, focusing on LSTM models for handling sequential data in finance contexts, although it lacks comparative analysis and detailed methodology.

The paper written by Uma Maheswari [2] investigates sentiment analysis on Twitter and Flipkart reviews, this study utilizes fuzzy logic for aspect-based sentiment analysis, yet it lacks comprehensive exploration of rule- based methods and model validation.

The paper written by Ms. Payal Yadav [3] proposes a hybrid sentiment analysis approach incorporating emoticons but lacks depth in discussing this method's effectiveness and the impact of multilingualism.

The paper written by Or Mohammad Aman Ullah [4] is a new set of sentences in English and Hebrew. explores sentiment analysis with ML and ensemble classifiers but faces limitations in emoticon data utilization, multilingual analysis, and word processing techniques..

The paper written by Ritu Agarwal [5] Focusing on Twitter data sentiment analysis, this paper introduces an Adaptive Neuro Fuzzy Inference System but lacks comprehensive emoticon coverage and comparative analysis.

The paper written by Monika Bhakuni [6] addresses sarcasm detection on Twitter, this study employs sentiment analysis techniques to offer insights into the evolving nature of sarcasm and proposes solutions for its detection challenges.

The paper written by Arif Ridho Lubis [7] tackles noisy data in social media platforms like Twitter and proposes a deep learning approach for enhanced sentiment analysis accuracy and sarcasm detection.

The paper written by , Shaina Gupta [8] focuses on proposing a hybrid model, this research aims to improve sentiment analysis accuracy, particularly in detecting sarcasm, by integrating multiple techniques.

The paper written by, M. Jeyakarthic [9] leverages bidirectional LSTM networks, this paper enhances sentiment analysis accuracy, specifically focusing on sarcasm detection in Twitter data.

The paper written by Yik Yang Tan1 [10] employs Deep Multi-Task Learning to improve sentiment analysis efficiency and tackle the complexities of sentiment expression, particularly sarcasm, on social media platforms.

The paper written by Joseph A. [11] investigates emotions using the International Affective Picture System, this research explores emotional categories evoked by images and minimal gender differences in categorization.

The paper written by Yun Liang [12] introduces DMN-HS, this method integrates image captioning into sentiment analysis, demonstrating superior performance in image sentiment classification through precise sentiment latent space construction.

The paper written by Namita Mittal [13] highlights the advantages of CNN in image sentiment analysis, emphasizing its efficiency, accuracy, feature extraction capabilities, and deep learning capabilities for hierarchical representation learning.

The paper written by Jana Machajdik [14] reviews the papers led to the conclusion that there is no integrated model for sentiment analysis that encompasses image, text, and emoticons.

These papers collectively contribute to the advancement of image sentiment analysis, covering emotional categories, advanced methods like DMN-HS, CNN dominance, and integration with psychology and art theory for affective image classification.

III. OBJECTIVES & MOTIVATION

From the identified gaps in the existing literature, we derived the motivation and objectives as mentioned below:

- Creation of a comprehensive and inclusive sentiment analysis tool with the inclusion of a text and emoticon-based analysis tool with image-based analysis to address existing shortfalls.
- Implementation of sarcasm detection along with sentiment analysis to improve accuracy and grasp of human language.
- Allowing public access to these sentiment analysis tools beyond technical experts and developers i.e. making it user friendly.
- Foster innovation in sentiment analysis by leveraging advanced algorithms and techniques to outperform

ISSN (online): 2581-3048
Volume 8, Issue 9, pp 86-93, September-2024
https://doi.org/10.47001/IRJIET/2024.809011

existing solutions and provide actionable insights across diverse communication modalities.

- Provide a valuable resource for businesses, researchers, and individuals seeking to understand and analyse sentiment effectively, driving informed decision-making and improving user experiences.

IV. PROBLEM DEFINITION

After identifying the gaps in the existing literature survey and coming up with the motivation and objectives, this paper aims to develop an integrated and comprehensive sentiment analysis tool that involves text, emoticons, and image-based analysis, in addition to implementing sarcasm detection to increase accuracy. The objective is to implement various sentiment analysis using various techniques such as rule-based for text sentiment analysis, emoji-based sentiment analysis, sarcasm detection & image sentiment analysis. We also integrate these techniques into a unified system. By reaching these goals, the project seeks to provide a holistic approach in understanding sentiments expressed via various modes of communication, enhancing accuracy and depth of sentiment analysis.

- To develop a user friendly and open-source sentiment analysis tool.
- To provide sentiment analysis capabilities for comprehensive methods: text, emoticons, image.
- To make the tool accessible and user-friendly for users with varying levels of technical expertise.
- To outperform existing state-of-the-art sentiment analysis solutions.
- To enable more accurate and actionable insights in areas such as social media monitoring and customer feedback analysis.

V. PROPOSED METHODOLOGY

Our proposed methodology as shown in Fig 1 encompasses a comprehensive approach to sentiment analysis, leveraging various techniques to analyze text, emoticons, and images. To begin with, we will employ Support Vector Machines (SVM) for sarcasm detection, training the model on labeled data using TF-IDF vectorization. This step ensures the accurate identification of nuanced language expressions before sentiment analysis. Moving forward, our text sentiment analysis will combine rule-based approaches and machine learning models to assess sentiment polarity and subjectivity, providing an overall sentiment score for text data. Additionally, we will analyze sentiments expressed through emojis by mapping them to emotions and applying sentiment analysis rules. For image-based sentiment analysis, Convolutional Neural Networks (CNNs) trained on augmented image data will be utilized to classify sentiment expressed in images accurately. Finally, we will integrate the results from text, emoji, and image analyses to offer a holistic understanding of sentiment across different communication modalities. Through this multifaceted approach, we aim to develop a robust sentiment analysis tool that empowers users to gain actionable insights from diverse forms of user-generated content.

A) Sarcasm Detection

Utilizing Support Vector Machines (SVM) for sarcasm detection involves advanced concepts and methodologies:

$$f(x) = sign(w \, . \, x + b) \quad \text{............} \quad (1)$$

SVM is a supervised learning algorithm used to classify data into two classes as mentioned in equation (1), such as sarcastic and non-sarcastic sentences. It optimally separates these classes with a hyperplane, maximizing the margin between them.

- Kernel Trick: SVM feature is that it employs kernel functions, including linear, polynomial, RBF, etc., to apply the non-linear distribution of data features more effectively by modifying the feature space through the ‘kernel trick’. It is crucial in our case given that the sarcastic and non-sarcastic sentences within the target news article dataset follow a non-linear situation.
- TF-IDF Vectorization: After defined sections, the text data is put through the TF-IDF vectorization technique, which converts the text data into numerical vectors. This numerical vector reflects the importance of the relative frequency of a word in the document to that of the entire corpus. This will allow the SVM model to meaningfully select key features.
- Training and Validation: the dataset is then split into training and validation, with hyperparameter tuning of kernel type, linear, polynomial, or RBF, plus regularization.
- Prediction and Evaluation: Trained SVM model predicts the sarcasm labels for new sentences. Evaluation metrics such as accuracy, Precision, Recall, and F1 score can evaluate this model. It will also help to inform how the model performs sarcasm detection. While machine learning models like Naive Bayes, SVM, and neural networks learn sentiment patterns from labeled data.
- Sentiment Polarity Calculation: Each word or phrase in the text is assigned a sentiment score, typically ranging from negative to positive. Aggregation methods derive an overall sentiment score for the text, categorizing it as positive, negative, or neutral.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 9, pp 86-93, September-2024

https://doi.org/10.47001/IRJIET/2024.809011

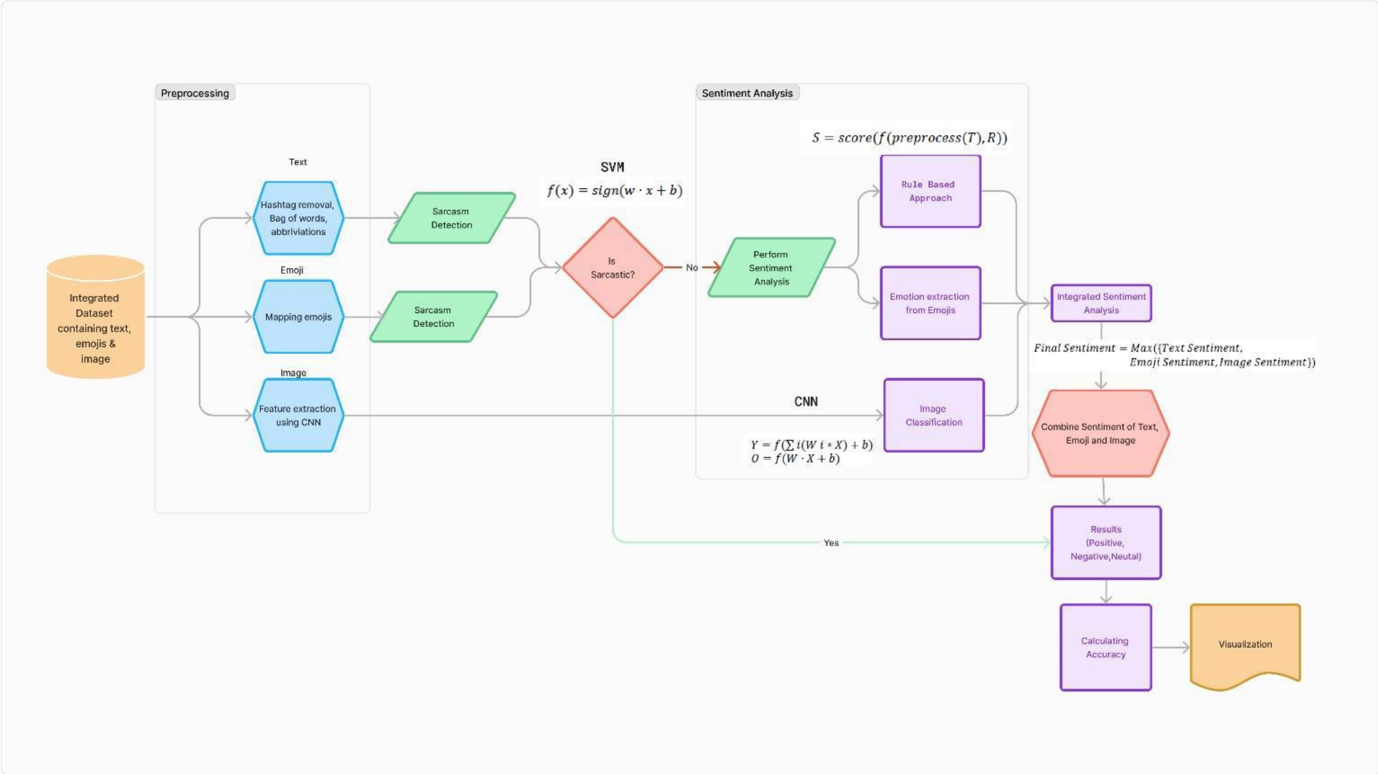


Figure 1: Architecture Diagram for Unified Sentiment Analysis

B) Text Sentiment Analysis

Text-based sentiment analysis includes various key steps and methodologies:

- Text Preprocessing: Preprocessing techniques like tokenization, normalization (e.g., lowercase conversion, punctuation removal), and stopwords removal clean the text data for sentiment analysis.
- Sentiment Analysis Methods: Lexicon-based approaches utilize sentiment lexicons or dictionaries to assign sentiment scores to words, Rule based approach:

$$S = score(f(preprocess(T), R)) \quad \ldots\ldots\ldots\ldots (2)$$

- Tools and Libraries: NLP libraries (e.g., NLTK, TextBlob), machine learning frameworks (e.g., scikit-learn, TensorFlow), and sentiment lexicons facilitate sentiment analysis tasks.

C) Emoji-Based Sentiment Analysis

Analyzing sentiments conveyed through emojis involves specific procedures:

- Dataset Loading: The code imports a dataset containing text with emojis expressing emotions.
- Emotion Extraction: A function extracts emotions from emojis based on a predefined mapping of emojis to emotions.
- Sentiment Labeling: Another function assigns sentiment labels (positive, negative, neutral) to texts based on the extracted emotions.
- Application to Dataset: These functions are applied to specific dataset columns, associating sentiment labels with each entry.
- Displaying Results: The code prints text along with assigned sentiment labels, reflecting sentiments conveyed through emojis.

D) Image-Based Sentiment Analysis

The approach to image-based sentiment analysis involves several stages:

- Data Preparation: Directories for training and testing image data are defined, and the number of images per class is counted.
- Data Visualization: A few random images from the training set are visualized using Matplotlib and displayed to understand and review it.
- Image Data Augmentation: The training image is augmented by applying Rescale, shear, zoom, and horizontal flips to the training image, which helps improve the training ability and robustness.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 9, pp 86-93, September-2024

https://doi.org/10.47001/IRJIET/2024.809011

- Model Definition: In this stage, a CNN as mentioned in equation (3) model with multiple convolution and max-pooling layers and fully connected layers is defined with Keras.

$Y = f(\Sigma\, i(W\, i * X) + b)$ ………………. (3)

$O = f(W \cdot X + b)$

$softmax\ function: P(y = i \mid X) = eOi/\Sigma\, eO$

- Model Training and Evaluation: An augmented data CNN model is trained and tested on test data. The training of CNN model is assessed on various parameters, including accuracy and loss.
- Image Prediction: An example image is preprocessed, passed through the trained model, and its sentiment (class label) is predicted.
- Displaying Prediction: Using Matplotlib, the original image is shown, and the predicted sentiment for this image is displayed showcasing the models ability of extracting the sentiments.

E) Text + Emoji-Based Sentiment Analysis

Text-based sentiment analysis can be combined with emoji-based analysis by using both linguistic and emotive signals among them:

- Text Preprocessing: The text data is pre- processed using standard NLP preprocessing methods to prepare it for the sentiment analysis.
- Emoji Sentiment Mapping: The analysis totality is thus complemented by mapping emojis to sentiment categories, increasing the scores for the sentiment analysis.
- Integration: Text-based and emoji-based average sentiment scores are integrated in order to achieve the sentiment understanding, which may be done by ensemble or aggregation approaches.
- Contextual Analysis: The sentiment analysis accuracy and depth increases when the contextualization of the use of emojis with text is taking into account.

F) Text + Emoji + Image Based Sentiment Analysis

Integrating text, emoji, and image analyses into a combined sentiment analysis system comprises of the following steps:

- Importing Libraries: Essential libraries for NLP, machine learning, image processing, and visualization are imported.
- Load Pre-Trained Models: Pre-trained models for image classification and sarcasm detection are loaded for analysis.
- Data Preprocessing: Text data undergoes TF- IDF transformation, emojis are mapped to sentiment, and images are preprocessed for analysis.
- Sentiment Analysis: Functions for text sentiment, emoji sentiment, and image sentiment analysis are defined and applied to input data.
- Integrated Analysis: An integrated sentiment analysis function combines results from text, emoji, and image analyses, considering sarcasm detection for a comprehensive sentiment assessment.
- Displaying Results: The integrated sentiment analysis results are displayed, providing insights into sentiment across multiple modalities.

Expected output:

What we hope for our solution is that it becomes a sentiment analysis tool that is all-inclusive and easy to understand. This model can analyze text, emoticons, and images. By knowing the sentiment behind your data through this mechanism, you will be able to gain more accurate and actionable outcomes. Our goal in sentiment analysis is to have an accurate sentiment behind the text so that decision making is done smoothly. For real-time open-domain named entity recognition as well as sentiment analysis improvement, we intend to include contextualized sparse representations too As such, market research could benefit from it greatly not forgetting social media monitoring, customer feedback analysis, tourism research among other areas. In summary words may be cheap but they can also change lives therefore ours seeks at being both efficient and effective in helping people do better things with their time.

VI. RESULT ANALYSIS

To gather insights for our model implementation, we conducted a survey with real-time responses from participants. We had collected in total 102 responses. The survey included the following questions:

1. "Describe your feelings about your latest movie or TV show binge-watching experience:"

- The graph analysis of this question gave maximum of positive responses for only text data.
- Accuracy: 0.82%.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 9, pp 86-93, September-2024

https://doi.org/10.47001/IRJIET/2024.809011

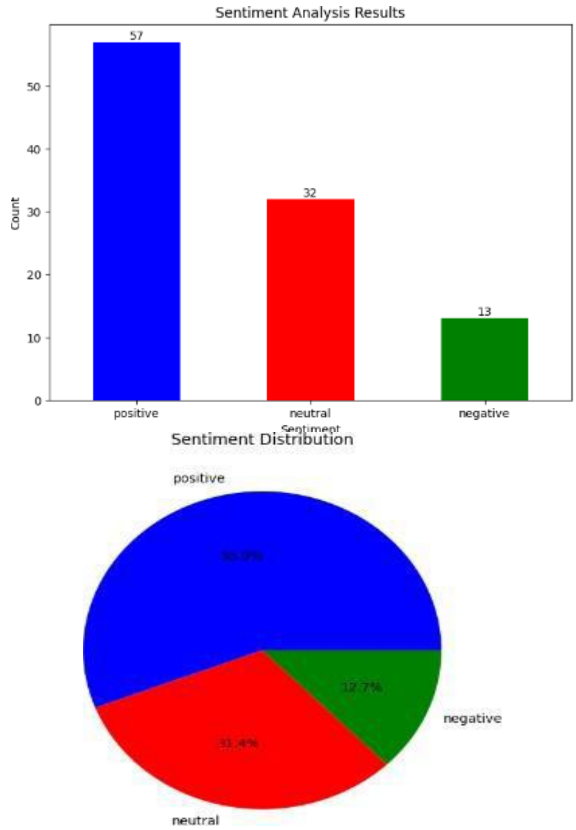


Figure 2: Text Sentiment Analysis

The figure above (Fig.2) illustrates a visual representation in the form of bar graphs and pie charts. It categorizes the data collected from text responses into positive, negative, and neutral sentiments.

2. "How would you express your mood today using emojis?"

- Participants expressed their mood using emojis, which will be analyzed for emoji-based sentiment analysis.
- Accuracy: 0.7666%.

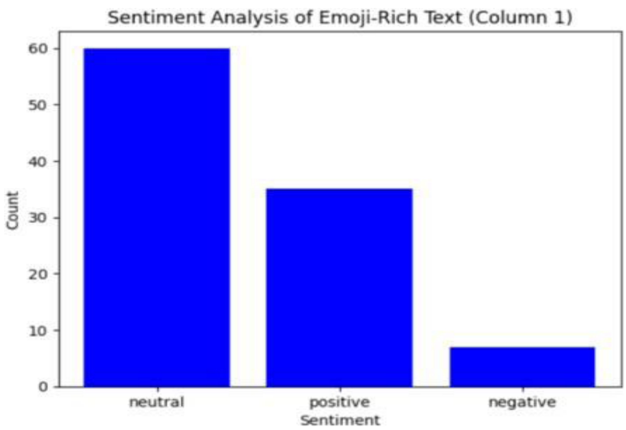


Figure 3: Emoji Sentiment Analysis

The figure above (Fig.3) illustrates a visual representation of emoji data collected from users without accompanying text, focusing solely on emojis and categorizing them into positive, negative, and neutral sentiments.

3. " Using emojis and a few words, express how you feel about the upcoming weekend"

- The graph analysis of this question gave maximum of positive responses for text and emoji combined.
- Accuracy: 0.63%.

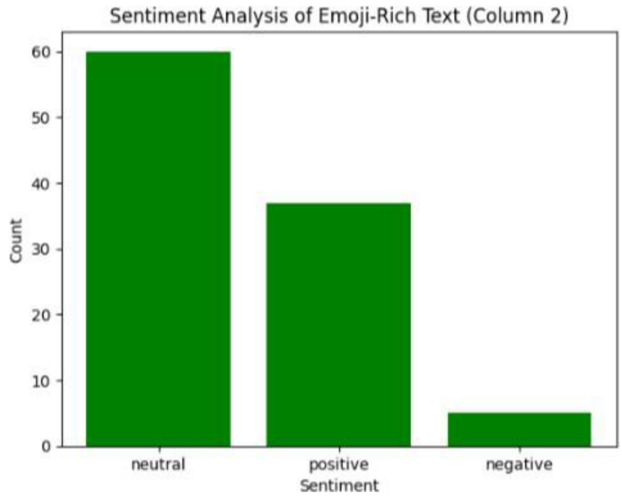
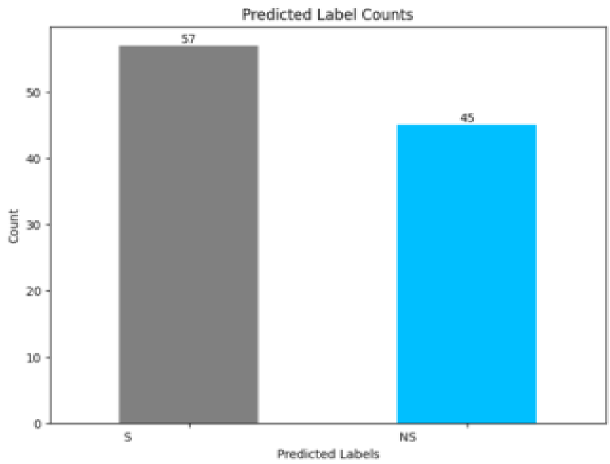


Figure 4: Text-Emoji Sentiment Analysis

The figure above (Fig. 4) presents a visual representation in the form of bar graphs, showcasing a combined analysis of text and emoji data collected from users. This integrated approach proves more efficient than individual text or emoji sentiment analysis, as it combines both types of data and categorizes them into positive, negative, and neutral sentiments.

4. "Share a lighthearted joke or sarcastic comment:"

- The graph shows the percentage of Sarcastic and non-sarcastic sentences, i.e., almost 50% of responses are sarcastic.
- Accuracy: 0.74%.



ISSN (online): 2581-3048
Volume 8, Issue 9, pp 86-93, September-2024
https://doi.org/10.47001/IRJIET/2024.809011

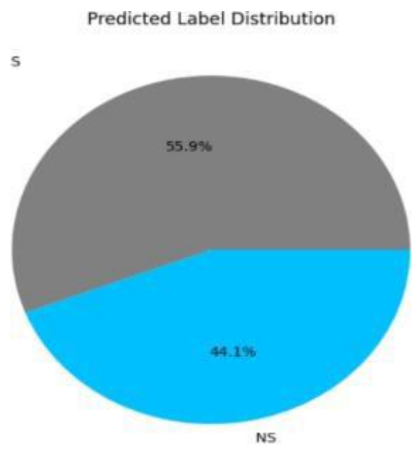


Figure 5: Sarcasm Detection

The figure above (Fig. 5) depicts a visual representation of data after categorizing it into sarcastic and non- sarcastic segments. This division is crucial in sentiment analysis as it helps overcome hurdles caused by sarcasm, ensuring a more accurate detection of the actual emotion conveyed in the sentence.

5. Integrated Analysis Results:

- The integrated sentiment analysis tool combines results from text, emoji, and potentially image analyses to offer a holistic understanding of sentiment across different communication modalities.
- The integrated sentiment analysis tool combines results from text, emoji, and potentially image analyses to offer a holistic understanding of sentiment across different communication modalities.
- The analysis shows combined sentiment analysis of emoji+text+image.
- Accuracy: 0.69%.

Figure 6: Integrated Sentiment Analysis

The figure above (Fig. 6) illustrates an integrated approach that takes an image as input and produces emoji and text describing the image, detects sarcasm or non- sarcasm, and analyzes whether the sentiment is positive, negative, or neutral. This comprehensive analysis showcases the system's ability to process diverse inputs and provide multifaceted outputs.

VII. CONCLUSION

To conclude, this research aimed at examining and reviewing various different methods used by others for carrying out sentiment analysis but found a gap of not having comprehensive model and thus suggesting an all-inclusive solution for sentiment analysis of user generated data. The method proposed conducted an intensive study and assessment of different approaches such as text based, emoticon based and image based methods which had been previously used.

All these methods have their strengths as well as weaknesses; thus combining them would enable one to do more holistic and precise sentiment analysis of user generated data. This is achieved by coming up with a solution that accurately analyses such information while giving understandable visualization’s of sentiments.

REFERENCES

[1] Implementation of Text-Based Sentiment Analysis Using LSTM Model | IEEE Conference Publication | IEEE Xplore (mapmyaccess.com).

[2] S. Uma Maheswari and S. S. Dhenakaran, “Aspect based Fuzzy Logic Sentiment Analysis on social media Big Data.”, (July 28 - 30, 2020).

[3] SentiReview: Sentiment analysis based on text and emoticons | IEEE Conference Publication | IEEE Xplore (mapmyaccess.com).

[4] Sentiment Analysis using Ensemble Technique on Textual and Emoticon Data | IEEE Conference Publication | IEEE Xplore (mapmyaccess.com).

[5] Sentiment Analysis of Emoticon Based Neuro Fuzzy System | IEEE Conference Publication | IEEE Xplore (mapmyaccess.com).

[6] Monika Bhakuni,Karan Kumar, Sonia ,Celestine Iwendi, and Avtar Singh, “Evolution and Evaluation: Sarcasm Analysis for Twitter Data Using Sentiment Analysis”. (11 October 2022).

[7] Sentiment Analysis in social media: Handling Noisy Data and Detecting Sarcasm Using a Deep Learning Approach | IEEE Conference Publication | IEEE Xplore (mapmyaccess.com).

[8] A Hybrid Approach for Enhancing Accuracy and Detecting Sarcasm in Sentiment Analysis | IEEE

ISSN (online): 2581-3048

Volume 8, Issue 9, pp 86-93, September-2024

https://doi.org/10.47001/IRJIET/2024.809011

Conference Publication | IEEE Xplore (mapmyaccess.com).

[9] Optimal Bidirectional Long Short Term Memory based Sentiment Analysis with Sarcasm Detection and Classification on Twitter Data | IEEE Conference Publication | IEEE Xplore (mapmyaccess.com).

[10] Sentiment Analysis and Sarcasm Detection using Deep Multi-Task Learning | SpringerLink.

[11] https://link.springer.com/article/10.3758/BF03192732

[12] https://ieeexplore-ieee-org.svkm.mapmyaccess.com/document/9506701

[13] https://ieeexplore-ieee-org.svkm.mapmyaccess.com/document/8609672

[14] [PDF] Affective image classification using features inspired by psychology and art theory | Semantic Scholar.

Citation of this Article:

Kresha J. Shah, Heta C. Shah, Durva H. Patel, & Nilesh Marathe, (2024). SentiMix: A Unified Approach to Comprehensive Sentiment Analysis. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 8(9), 86-93. Article DOI https://doi.org/10.47001/IRJIET/2024.809011

****