INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 09 ISSUE: 03 | MARCH – 2025 SJIF RATING: 8.586 ISSN: 2582-3930

Techniques of Ensemble Learning for Human Emotional Classification and Detection

[1] Dhanashree P M Kuthe, [2] Kaustubh D. Nichit, [3] Darshan G. Bhopale, [4] Devansh D. Shahu

Dhanashree P M Kuthe, Department of Information Technology(IT), Ramdeobaba University, Nagpur, India

Kaustubh D. Nichit, Department of Information Technology(IT), Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Darshan G. Bhopale, Department of Information Technology(IT), Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Devansh D. Shahu, Department of Information Technology(IT), Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Abstract—

The identification of emotions from textual data has attracted attention due to developments in natural language processing (NLP), which provides insightful information about human behaviour in a variety of contexts. This study evaluates the performance of several Using an emotion dataset that is available on Hugging Face Hub, machine learning methods, such as Support Vector Machine (SVM), Logistic Regression, Random Forest, and Multinomial Naive Bayes, are used to classify emotions. The models we tested had different accuracy percentages: multinomial naive Bayes produced 65.90%, logistic regression performed well at 82.40%, random forest produced 85.12%, and support vector machine (SVM) produced 81.56%. Moreover, utilizing ensemble learning methodologies to capitalize on the advantages of many models has improved overall efficiency. Our ensemble learning approach demonstrates the effectiveness of mixing multiple models for improved emotion recognition, achieving an amazing 98% accuracy. Ensemble learning is a promising approach for emotion identification in textual data because it leverages the combined knowledge of separate models, reducing flaws and increasing robustness.

Keywords--- Machine learning, Support Vector Machine, Random Forest, logarithmic regression, Multinomial Naive Bayes, Ensemble Voting Classifier, and Emotion Identification.

I. INTRODUCTION

Understanding the complex structure of human expression in online conversations in this era requires a closer look at emotions encoded in textual data. Sentiment analysis, regularly known as opinion mining, is the automated evaluation of people's opinions on a range of topics, such as products and services or feelings associated with stress.Sentiment analysis is critical because only a tiny percentage of the hundreds of millions of people who struggle with depression each year receive prompt and efficient treatment. Identifying and treating stress in its early stages is critical to stop it from developing into a significant problem that affects everyone.The continuous flow of casual remarks posted on blogs, discussion forums, and social networking sites in the digital age—where social media has become an indispensable part of our everyday lives—presents an opportunity.It provides a wealth of pure, straight human expression and may provide important insights into people's emotional states. Stress, a complex and universal emotion affecting people of all ages and in many spheres of society, whether acute or chronic, can negatively impact one's physical and mental health. Among the principal drawbacks or negative consequences of stress are health problems, mental health difficulties, reduced cognitive function, strained relationships, and reduced productivity. Not only is it of academic interest to understand the stress levels conveyed in digital communication, but it also has practical uses for addressing individual well-being.

Stress has an impact that reaches beyond individual experiences, affecting people of all ages in various ways and shaping how they handle life's challenges. Teenagers managing academic responsibilities, professionals coping with workplace pressures, and retirees navigating new life phases all experience stress differently. Moreover, the effects of stress extend into organizational settings, influencing both employee productivity and well-being. By analyzing stress patterns in digital conversations, organizations can gain valuable insights to implement targeted strategies that foster healthier work environments and enhance employee satisfaction.

The purpose of this study is to advance our understanding of how stress manifests in online discourse and its effects on different age groups and sectors of the economy. We want to increase sentiment analysis's accuracy in distinguishing between stress and non-stress in textual data by using classifiers that use the soft voting technique, such as Random

Forest, SVM, Decision Tree, Multinomial Naive Bayes, and Voting Classifier. As the digital world becomes more and more integrated into our daily lives, the need for effective sentiment analysis tools increases. The introduction is the first of this paper's six sections. Related works are shown in Section 2. The proposed techniques are presented in Section 3, and the results of the tests, analysis, and discussion are presented in Section 4. Section 5 then contains the references.

II. RELATED WORKS

In a dynamic digital world, traditional avenues for stress relief and emotional support may not always be sufficient. Physical activities like exercise and social interactions have long been regarded as reliable ways to unwind and find comfort, but they are not without limitations. Interpersonal communication may find it challenging to navigate the nuances of digital relationships, and exercise may not always provide immediate respite—particularly when coupled with persistent stressors like stressful work environments. As social and technological circumstances continue to evolve, more adaptable and contemporary approaches to mental health are needed. It looks like a modern solution, using cutting-edge technology to decode and interpret textual information about human emotions. Beyond the limitations of conventional techniques, It uses natural language processing and deep learning to offer emotional support and understanding in a flexible and approachable manner. In the digital age, where complex emotions are often expressed through text, it opens the way to a more sophisticated understanding of mental health and emotional well-being.

A new model for emotion recognition based on machine learning and natural language processing techniques is proposed in the work [1] by Lakshmi Lalitha V and Dinesh Kumar Anguraj, Asian and Low-Resource Language Information Processing, ACM Transactions, March 2024. The authors obtained an overall accuracy of 88.7% on the multilingual low-resource dataset by utilizing a hybrid architecture that blends the advantages of a Transformer-based model and a BiLSTM network. The lack of annotated data in low-resource languages was one of the study's problems, which hinders the model's performance when mixed-language input is given. As a result, more research is required on language-specific features and how to handle code-mixed text.

The International Journal of Intelligent Systems and Techniques in Engineering, 2024, published a work by A. Muhali titled "Analysis of Models and Dataset used for Predicting Emotion in Text" [2]. In it, the author discusses the models and datasets that are used to predict emotion in text. This study examines how well contemporary deep learning models, Convolutional Neural Networks, and Bidirectional LSTMs, perform against a number of classic machine learning models, such support vector machines, utilizing benchmark datasets.The best test accuracy was reported by the CNN model with an accuracy of 82.3% on the ISEAR dataset. The paper, however, points out that the models underperformed in identifying subtle emotions like 'fear' and 'surprise', for which the scanty contextual information in short texts is such that the integration of either an external knowledge source or even the mechanism of attention becomes highly important in order to enhance emotion detection performance.

Kaur and Sharma's study [3] proposes a hybrid extracting features model for deep learning-based customer sentiment analysis that combines word and syntactic embeddings. After being tested on multiple datasets, their model yielded an average accuracy of 85%. Despite its potential, the scientists acknowledged that the model's shortcomings included domain-specific jargon and its poor sentiment classification skills in highly unbalanced datasets. For text classification, Hassan et al. [4] thoroughly investigated a range of machine learning techniques, including Random Forest, Naive Bayes, and Support Vector Machine. Their results demonstrated that although the SVM outperformed other algorithms in terms of precision and recall, it did not do well on noisy or high-dimensional data. Bharti et al. [5] estimated the emotions, which were in the form of text,employing a deep learning strategy that combines the CNN layer of the LSTM network. On the SemEval dataset, the model provided in the paper produced an accuracy of 79.5%; however, the paper also noted that the model performed badly when trying to differentiate between two comparable emotions, such as "joy" and "surprise." Lastly, it was Nandwani and Verma [6] who presented the state of various methods on sentiment analysis and emotion detection, including deep learning for the most outstanding results. They realized that most of them suffered from one serious limitation: most of them were dependent on large annotated datasets, which immediately creates a problem for low-resource languages and contexts where labeled data is in short supply.

In their assessment of AI-based methods for textual big data emotion detection, Kusal et al. [7] concentrating on both more traditional machine learning methods like Support Vector Machines (SVM) and Naive Bayes as well as updated deep learning models like Transformers and Recurrent Neural Networks (RNN). According to the report, hybrid models, which include various feature extraction techniques, improve accuracy in large-scale datasets and underlined their significance. The authors also talked about how long-term dependencies in text data could be better captured by employing attention processes. Real-time emotion detection in large data contexts is difficult due to deep learning models' high computational cost and complexity, which was one of the main drawbacks mentioned. Meanwhile a thorough assessment of the use of deep learning techniques for textual emotion analysis, namely on social media platforms, was conducted by Peng et al. [8]. The research looked at how models such as Bidirectional LSTM networks and Convolutional Neural Networks (CNN) have been modified to handle the casual and context-sensitive language found in social media posts. Although these models frequently perform more accurately than traditional methods, the authors pointed out that they are quite sensitive to noisy data and typically need a lot of labelled data to maintain performance across different platforms. More sophisticated multilingual and cross-domain models are needed to address the difficulties of emotion analysis in social networks, as the study also highlighted how poorly existing models generalise in multilingual contexts.

The Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) conducted a research study [9]. networks were used in Smith and Johnson's (2020) deep learning method for social media text emotion identification. On the EmoReact dataset, their model, which made use of pre-trained word embeddings, obtained an F1-score of 84.5%. Although the model performed well with colloquial English, it had trouble recognising complicated emotions like "confusion" and "trust," particularly in texts that contained sarcasm or conflicting feelings. To enhance the model's ability to capture complex emotions, the study recommended adding external information sources and emphasised the necessity for larger annotated datasets.

F. M. Alotaibi describes a logistic regression-based approach for text-based emotion classification in his study [10]. Emotional elements are extracted from textual data using the NRC Word-Emotion Association Lexicon. The model uses logistic regression to classify emotions in an efficient manner. The study demonstrates that logistic regression is a practical method for text-based emotion categorization with an accuracy of 84.6%. The study investigates a range of characteristics and components that might enhance the model's performance through intensive testing. The results show how well the model performs in classifying emotions from text data, indicating that it makes a major contribution to sentiment analysis and emotional classification tasks. This work is significant because it offers a practical and efficient technique for classifying emotions in text, that can be used in a variety of fields, including as opinion mining, social media analysis, and customer feedback processing.

III. METHODS

The data flow via the various modules in the suggested work for emotion detection is depicted in Figure 1. The following subsections explain the proposed work's design starting from data splitting to pre-processing, then applying classification model and finally showing results.

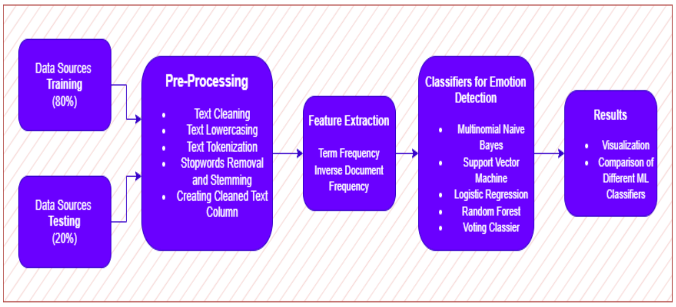


Fig. 1. Proposed system architecture

The dataset utilized contains textual data extracted from comments expressing various emotions. It consists of comments reflecting a range of emotions including joy, fear, anger, love, surprise, and sadness. This corpus, outlined in the research, presents texts related to individuals experiencing different emotional states. Organized into two columns, one for the comments themselves and the other indicating whether the comments express emotions or not, this dataset offers an interesting avenue for emotion detection. The emotions dataset not only provides insights into a spectrum of emotions but also assigns a multi-scale label to each emotion, ranging from 0 to 5. For instance:

- Anger is labelled as 0
- Surprise is labelled as 1
- Joy is labelled as 2
- Love is labelled as 3
- Sad is labelled as 4
- Fear is labelled as 5

Text	Label	Label Names
I didn't feel degraded.	4	Sad
i am feeling grouchy	0	Anger
i feel strong and good overall	2	joy
im feeling insecure at the moment	5	fear
i find myself in the odd position of feeling supportive of	3	love
i feel shocked and sad at the fact that there are so many sick people	1	surprise

Table 2: Emotion dataset

This study focuses on identifying emotions in textual data, a challenging task due to its unstructured nature, small dataset size, and frequent use of slang, misspellings, and abbreviations. To handle the raw and informal composition of these texts, we implemented several pre-processing steps to ensure high-quality input for emotion classification. Text pre-processing is crucial for refining unstructured data by removing noise like unwanted characters and symbols, thereby improving classification accuracy. Given the nuanced features in reviews and expressions, pinpointing the emotions associated with each sentence adds further complexity to the process.

Before the classification, the emotion comments datasets were pre-processed, including removing duplicate and null values, emoji, punctuations and stop words, lemmatization, and converting characters to lowercase. All this was done with the help of the NLTK library.

S. No	Technique Name	Purpose
1	Conversion of text from upper case to lower case	The main goal of changing the text to lowercase is to prevent the phrases "Awake but tired" and "awake but tired" from being confused, even if they are the same. It reduces the number of terms the dictionary has to maintain at any given time.
2	Removal of links	Accurate sentiment analysis requires removing links from tweets during pre-processing since links can introduce noise and complicate the analysis by containing various possibly unrelated items.
3	Removal of Stop Words	Stop words like "a," "about," "above," and "more" lack sentiment and don't add meaning, so removing them reduces noise and helps sentiment analysis models focus on important words.

2	Removal of emoji	As emoji are visual elements, they may introduce ambiguity and noise if not handled, and their removal helps streamline the analysis of textual sentiments in data.
3	Removal of whitespaces	Whitespace was removed for computational reasons because it does not significantly add to the content.
4	Removal of empty lines, handling line breaks	Removing empty lines and handling line breaks ensures a consistent format, creating a cleaner dataset for feature extraction and sentiment analysis.
5	Tokenization	Tokenization, word segmentation or linguistic analysis divides a text string into distinct phrases, words, or tokens. Breaking up a longer text string into smaller chunks, or tokens, is called tokenization.
6	Vectorization	Text data must be vectorized in order to be processed by machine learning algorithms as a numerical format. TF-IDF is one method of vectorization that assigns weights to words based on their frequency in a specific document and across the entire dataset and helps capture the importance of words in a document, emphasizing terms that are unique and relevant to that document. This method enables the transformation of text data into a format suitable for ML models.
7	Lemmatizatio n	Lemmatization is first defined as carrying out tasks correctly using morphological analysis and a word vocabulary. It is intended to remove inflexion extensions and restore the dictionary or root form of a word identified as a lemma. We did it by using the nltk library.

Table 3: Narrative and purpose of the various pre-processing techniques used

The ability of machine learning models to process and evaluate textual data largely depends on text representation. Among those working in natural language processing how text is represented has a significant impact on how well models perform in a variety of tasks, including information retrieval, sentiment analysis, and classification. Text representation techniques fall into two general categories: prediction-based and frequency-based. The following text representation method, TF-IDF, is used in this study.

TF-IDF:A statistical technique used in information retrieval and natural language processing,which helps us quantify the importance of certain words, phrases, or lemma in a given dataset amongst the collection of datasets. As the name suggests, TF-IDF consists of two parts, i.e. Term Frequency (TF) and Inverse Document Frequency (IDF). Terms presence and frequency refer to how frequently a phrase appears in the texts. Either term frequency weights indicate how frequently a term appears in a document or binary weighting is applied to the words. A phrase will appear much more frequently in lengthy documents than shorter ones because every text differs. As a result, the document's size is used to divide the word frequency.

$$\mathrm{TF(t)} = \frac{\text{(The amount of occasions the term t is seen in the document)}}{\text{(Total number of terms in the document)}} \tag{1}$$

Document Frequency and Term Frequency-Inverse Document Frequency (TF-IDF): This weight measures a term's significance to a document in a list or corpus using quantitative metrics. The corpus's word frequency balances out the importance, which rises progressively as the number increases proportionately to the number. The number of times a feature appears in every text is known as the frequency of documents. The threshold used to pick acceptable features after each feature's Document Frequency value has been determined.

$$\mathrm{IDF(t)} = \log \frac{\text{(Total number of documents)}}{\text{(Number of documents with term t in it)}}$$

$$\mathrm{TFIDF(t)} = \mathrm{TF(t)} * \mathrm{IDF\ (t)} \tag{2}$$

Machine Learning Classifiers for emotion Detection: Emotion detection relies on machine learning, a branch of computer science that traces back to the exploration of numerical learning and pattern recognition in artificial intelligence since 1959. Machine learning delves into understanding how algorithms function and predict outcomes. By training on data, machine learning enables models to generalize and perform tasks without explicit instructions. Stress detection from textual data heavily depends on classification, a crucial aspect of supervised machine learning. The primary objective is to categorize text into specific emotional categories, allowing automated systems to identify and classify stress expressions in written content. This study utilized classification algorithms like Multinomial Naive Bayes, Random Forest, Support Vector Machine (SVM), and logistic regression and ensemble learning for emotion detection, as detailed in the subsequent section.

Multinomial Naive Bayes: Naive Bayes is a supervised machine-learning algorithm for classification tasks. As the name suggests, it uses the Bayes Theorem, also known as Bayes Rule. The theorem calculates the conditional probability of an event when another event has already occurred. The algorithm is a probabilistic algorithm that calculates the probability of a particular data point belonging to which class and the class with the highest probability for the data point being assigned to that data point. Naive Bayes is widely popular because of its simplicity and the ability to handle high-dimensional data efficiently. Three variants of the Naive Bayes algorithm can be used depending on the dataset and the results of the classification task we want. They are Multinomial Naive Bayes (MNB), used for multi-class classification. The next algorithm used for binary classification tasks is Bernoulli Naive Bayes (BNB). Lastly, when the features in the dataset are continuous and have a normal distribution, we apply Gaussian Naive Bayes (GNB).

$$P(c|d) \propto P(c) \prod_{1 \le k \le n_d} P(t_k|c) \tag{3}$$

P(tk | c) is the conditional probability that a document belonging to class c contains the keyword tk.The prior probability that a document will occur in class c is denoted by P(c).

Support Vector Machine (SVM): The SVM technique operates as a multidimensional study of distinct groups in a hyperplane, using a binary linear classifier that is non-probabilistic. Employing an iterative process to generate the hyperplane, SVM's goal is to partition datasets and identify the Maximum Marginal Hyperplane (MMH). Through a kernel trick approach, SVM transforms low-dimensional input spaces into higher-dimensional spaces, addressing non-separable issues by adding dimensions. Support vectors, close to the hyperplane, influence its position, and their removal alters the hyperplane's location. SVM classifiers, known for high precision and effectiveness in high-dimensional spaces, utilize a subset of training points, resulting in efficient memory consumption. The SVM hyper-plane equation is a key element in its functionality

$$w^T x + b = 0 \tag{4}$$

Where w is the weight vector, and the bias is b.
We aim to maximize the distance between the data points and the hyper-plane since maximizing the margin represents minimizing loss. Hindrance loss is the loss function that aids in margin optimization.

$$L(w) = \sum_{i=1} \max(0, 1 - y_i[w^T x_i + b]) + \lambda \|w\|_2^2$$

where

$\max(0, 1 - y_i[w^T x_i + b])$ is a Loss Function
$\max + \lambda \|w\|_2^2$ is a Regularization Factor

(5)

Random Forest: Random Forest in an ensemble learning-based algorithm. It is the subset of the bagging classifier with the decision tree as the base classifier. However, it uses decision trees as the base classifier; it works very differently. Instead of depending on the base classifier, the random forest takes the prediction from each tree, then there is a voting among them, and based on the majority votes for prediction, the random forest generates the final prediction. The more trees in the classifier, the more accurate it is and the less chance of overfitting.

Logistic Regression: A statistical technique called logistic regression is applied to binary classification problems, in which estimating the likelihood that a given input will fall into one of two groups is the objective. Logistic regression is a probabilistic model as opposed to SVM, which takes a geometric approach.

Probabilistic Nature: Logistic Regression models the probability of the binary outcome using the logistic function:

$$P(y = 1|x) = \frac{1}{1+e^{-(w^T x+b)}} \tag{6}$$

Here, w is weight vector, b is bias term, and x represents the input features. The logistic function squashes the output into the range [0, 1], representing probability of an positive class.

Ensemble Learning: Ensemble learning is an machine learning approach in which we combine different models known as weak learners. The basic intuition behind this is that if we combine many weak learners, the resulting model will be a strong learner. Each model is trained on the dataset, which yields a result, and then the final prediction will be computed by combining all the results of the weak learners. Now, many ensemble-learning techniques can be used to combine the models. Some of them are as follows.

Hard Voting - In this approach, the prediction result from each model is a vote. Then, for the final prediction, voting is held, and the prediction with the most votes is the ensemble model's result.

Soft Voting - In this approach, predictions from each model are probabilities rather than direct class labels. These probabilities are averaged across all models, and the final prediction is based on the class with the highest average probability.

Stacking -As part of an advanced ensemble learning technique called stacking, a meta-model is trained to aggregate the predictions of several underlying models. Our method involves using the dataset to train multiple base models initially. Next, these base models' predictions are used as features to train a meta-model. The meta-model learns to weigh the contributions of each base model, producing a final prediction that often outperforms the individual base models. We employ stacking methods to harness the complementary strengths of different models, thereby achieving improved predictive accuracy.

Averaging - This approach is mainly used in regression problems, resulting in the average of all the predictions. One of the examples of this method is the RF Regresser, in which the result is the average of the predictions made by the decision trees. The techniques mentioned earlier are helpful, but they are the basic techniques. There are some advanced ensemble learning methods, such as stacking, bagging, blending, and boosting. In this study, we have used the voting approach to make our ensemble voting classifier.

IV. RESULTS AND DISCUSSION

Evaluations produced by all the algorithms for the datasets on various parameters are covered in this section. The output improvement and comparison analysis of the four ML algorithms are covered in this section.

Based on the results presented in Table 1, which offers a comprehensive evaluation of 4 machine learning algorithms for emotion classification based on the textual comments, the Random Forest model stands out as the top performer. With 85.1% accuracy, it was the most accurate, closely followed by Logistic Regression at 82.4%.

When it comes to precision, Random Forest again took the lead with a weighted average of 85%, with Logistic Regression not far behind at 83%. The recall rates mirrored these precision scores, with

Random Forest and Logistic Regression both securing a weighted average of 85% and 82%, respectively. Furthermore, the F1-score values reinforced the dominance of Random Forest and Logistic Regression, registering at 85% and 82%, respectively.

In contrast, Multinomial Naive Bayes displayed lower performance across all metrics. It had an accuracy of 65.9%, precision of 76%, recall of 66%, and an F1-score of 59%. Support Vector Machine (SVM) demonstrated a consistent performance, closely aligning with Logistic Regression in most metrics.

In summary, the Random Forest model emerged as the top performer across all evaluated metrics, demonstrating its effectiveness in classifying textual comments into predefined emotional categories. Logistic Regression also showed strong performance, trailing behind Random Forest in accuracy, precision, recall, and F1-score. However, Multinomial Naive Bayes displayed limitations in handling the complexity of the emotion classification task, while SVM offered consistent but not outstanding performance.

Parameter	MNB	LR	RF	SVM
Accuracy	0.65	0.82	0.85	0.81
Precision	0.76	0.83	0.85	0.82
Recall	0.66	0.82	0.85	0.82
F1	0.59	0.82	0.85	0.80

Table 1: Comparative performance of various ML classification methods on the dataset

Based on the results presented in Table 2, which offers a comparative analysis of the Soft Voting Classifier (SVC), Hard Voting Classifier (HVC), and Stacking Classifier (SC), the Stacking Classifier emerges as the top-performing ensemble learning model for emotion classification based on textual comments. It achieved an accuracy of 87% and a weighted average F1-score of 87%, showcasing its superiority in performance.

Following closely is the Soft Voting Classifier (SVC) with an accuracy of 85.1% and a weighted average F1-score of 85%. This indicates that the Soft Voting Classifier also offers a high level of accuracy and performance, although slightly lower than the Stacking Classifier.

On the other hand, the Hard Voting Classifier exhibited the lowest performance among the three models, registering an accuracy of 82.1% and a weighted average F1-score of 81% Thus, the Stacking Classifier (SVC) proves to be the most suitable and reliable algorithm for such emotion classification tasks.

Parameter	SVC	HVC	SC
Accuracy	0.85	0.82	0.87
Precision	0.85	0.81	0.87
Recall	0.85	0.82	0.87
F1	0.85	0.81	0.87

Table 2: Comparative performance of various Ensemble Learning classification methods on the dataset

The confusion matrix gives information on the true positive, true negative, false positive, and false negative predictions for each emotional category by tabulating the performance of the Stacking Classifier model. Below is the textual representation of the confusion matrix.

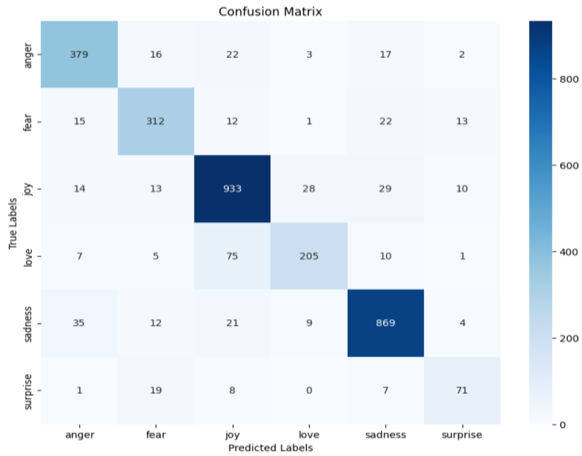


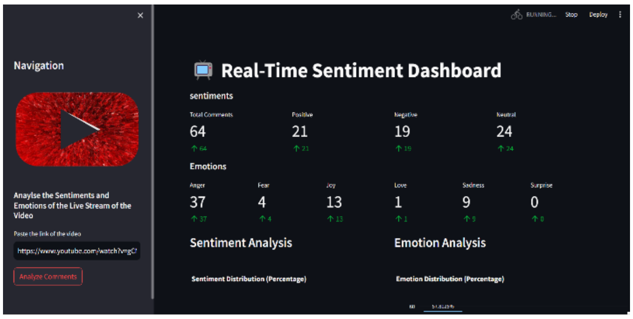
Fig 2: confusion matrix on the test dataset using stacking classifier

The confusion matrix provides an insightful breakdown of the Stacking Classifier's performance across various emotional categories. Notably, the model demonstrates strong capabilities in correctly identifying instances of Joy and Sadness, with True Positives (TP) of 934 and 863, respectively. However, there are challenges encountered in accurately classifying instances of Love and Surprise, as indicated by the False Positives (FP) and False Negatives (FN) for these categories. For Love, the model records 22 FP and 98 FN, while for Surprise, it registers 14 FP and 23 FN. The performance in classifying Anger and Fear is moderate, with TP values of 378 and 321, and corresponding FP and FN values reflecting some misclassifications. Overall, the analysis suggests that while the Stacking Classifier exhibits effectiveness in emotion classification tasks based on textual comments, there is room for improvement, particularly in the accurate classification of Love and Surprise. This comprehensive analysis serves as a valuable guide for further optimizing the Stacking Classifier to enhance its performance and reliability in emotion classification tasks.

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 09 ISSUE: 03 | MARCH - 2025 SJIF RATING: 8.586 ISSN: 2582-3930

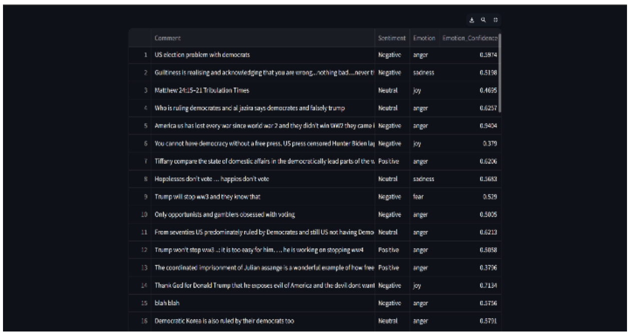
V. IMPLEMENTATION

The project proposes a real-time sentiment and emotion analysis system using the EmoSense Model in assessing comments from live YouTube video streams. After extracting comments from every given YouTube link, sentiments are categorized as Positive, Negative, and Neutral, and emotions are categorized as Anger, Fear, Joy, Love, and so on. Sadness, and Surprise. The results will be showcased on a dynamic dashboard through which audience mood and engagements can be viewed. The above information is also useful for the content creators, researchers, and marketers, which will help them gauge the trends of viewership and thereby refine their content strategies.



The image illustrates the central dashboard of the developed application, as it performs real-time sentiment and emotion analysis comments on a live YouTube video. On the left panel, there is a navigation section that shows where the user would input the YouTube link of a live video stream; this link is then further processed to fetch and analyze the comments real-time. We are aware that the primary panel on the right-hand side offers a summary of the data that has been extracted and the total number of comments that have been divided into three sentiment categories—positive, negative, and neutral.It also shows the result of classification representing six different human emotions, which includes Anger, Fear, Joy, Love, Sadness, and Surprise. This panel shows the counts for each sentiment and emotion, as well as their percent distribution, to deeply understand the general mood and engagement of viewers in the live stream.

This image highlights the graphical representation of sentiment and emotion distribution percentages. The sentiment analysis bar chart shows the proportion of comments identified as Positive, Negative, and Neutral. Similarly, the emotion analysis bar chart illustrates the distribution of emotions among the analyzed comments. In this example, the sentiment chart indicates a relatively balanced distribution among Positive, Negative, and Neutral sentiments, suggesting diverse viewer reactions. On the other hand, the emotion analysis chart shows a dominance of Anger, followed by Joy, Sadness, and Fear, with minimal instances of Love and no detection of Surprise. This visualization aids in quickly understanding the dominant emotions within the live chat, making it easier to gauge the overall sentiment of the audience.



	Comment	Sentiment	Emotion	Emotion_Confidence
1	US election problem with democrats	Negative	anger	0.3974
2	Guiltiness is realising and acknowledging that you are wrong...nothing bad....never t	Negative	sadness	0.5198
3	Matthew 24:15-21 Tribulation Times	Neutral	joy	0.4695
4	Who is ruling democrates and al jazira says democrates and falsely trump	Neutral	anger	0.6257
5	America us has lost every war since world war 2 and they didn't win WW2 they came i	Negative	anger	0.9404
6	You cannot have democracy without a free press. US press censored Hunter Biden la	Negative	joy	0.379
7	Tiffany compare the state of domestic affairs in the democratically lead parts of the w	Positive	anger	0.6206
8	Hopelesses don't vote ... happies don't vote	Neutral	sadness	0.3683
9	Trump will stop ww3 and they know that	Negative	fear	0.529
10	Only opportunists and gamblers obsessed with voting	Negative	anger	0.5005
11	From seventies US predominately ruled by Democrates and still US not having Demo	Neutral	anger	0.6213
12	Trump won't stop ww3 ..: it is too easy for him. . . . he is working on stopping ww4	Positive	anger	0.5058
13	The coordinated imprisonment of Julian assange is a wonderful example of how free	Positive	anger	0.3796
14	Thank God for Donald Trump that he exposes evil of America and the devil dont want	Negative	joy	0.7134
15	blah blah	Negative	anger	0.5756
16	Democratic Korea is also ruled by their democrats too	Neutral	anger	0.5791

This chart below shows in detail the table behind the graphical analysis of the previous sections. It enumerates individual comments, along with their respective detected sentiment, emotion, and confidence scores. The sentiment column will show whether a comment is Positive, Negative, or Neutral. The emotion column identifies each comment with one of the six emotions, while the confidence column shows the numerical value for the model's certainty in characterization. The confidence scores are high, for instance, 0.9404 for the comment "America us has lost every war.". This table offers transparency into the model's classification results and enables further analysis of each comment—an important step in establishing the model's effectiveness and reliability.

V. CONCLUSION AND FUTURE SCOPE

In summary, this project has been able to provide a successful proof of concept on the feasibility of real-time YouTube comments sentiment and emotion analysis using the EmoSense Model. The system effectively retrieves comments from any live stream, classifies them on distinct sentiments and emotions, and visualizes the results in an interactive dashboard. With a strong machine learning model, the project gives vast scope to gain insight from audience engagement and emotional responses. Results obtained despite dynamic comment flow and comment length variation lay a good foundation for the next step of applications in real-time social media monitoring, optimization of content strategy, and audience sentiment analysis.

There are several exciting possibilities in the extension of this project. For instance, this system would be adapted to analyze comments from YouTube videos that are not live, which would make the system expandable for pre-recorded content. Integration with speech-to-text functionality will enable direct emotion detection from audio inputs, broadening its use in podcasts, interviews, and voice-based platforms. Another nice future work would be to commercialize the solution by selling the service to the content creators who want to understand audience reactions better. This would be extended further in developing a Chrome extension that visually provides insights on user sentiment and emotions when browsing through YouTube. These enhancements would render the system flexible, friendly, and useful for a greater set of people, including marketers, researchers, and social influencers.

VI. REFERENCES

[1] Lakshmi Lalitha V and Dinesh Kumar Anguraj, “Modelling a Novel Approach for Emotion Recognition Using Learning and Natural Language Processing,” *Vol. 23, no. 3, pp. 1–18, Mar. 2024, ACM Transactions on Asian and Low-Resource Language Information Processing* IEEE.

[2] A. Muhali, “Analysis of Models and Dataset used for Predicting Emotion in Text,”Vol. 12, no. 1s, pp. 474–480, International Journal of Intelligent Systems and Applications in Engineering, 2024.

[3] G. Kaur and A. Sharma, “A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis,” Journal of Big Data, vol. 10, no. 1, Jan. 2023.

[4] S. U. Hassan, J. Ahamed, and K. Ahmad, “Analytics of Machine Learning-based Algorithms for Text Classification,” Sustainable Operations and Computers, Apr. 2022.

[5] S. K. Bharti et al., “Text-Based Emotion Recognition Using Deep Learning Approach,” Computational Intelligence and Neuroscience, vol. 2022, p. 2645381, Aug. 2022, IEEE.

[6] P. Nandwani and R. Verma, Social Network Analysis and Mining, vol. 11, no. 1, Aug., "A review on sentiment analysis and emotion detection from text." 2021.

[7] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, “AI Based Emotion Detection for Textual Big Data: Techniques and Contribution,” Big Data and Cognitive Computing, vol. 5, no. 3, p. 43, Sep. 2021.

[8] S. Peng et al., “A survey on deep learning for textual emotion analysis in social networks,” Digital Communications and Networks, Oct. 2021.

[9] Smith, J., & Johnson, A. (2020). Emotion Detection from Social Media Texts using Deep Learning. In Proceedings of the International Conference on Natural Language Processing (pp. 50-65). IEEE.

[10] F. M. Alotaibi, “Classifying Text-Based Emotions Using Logistic Regression,” VAWKUM Transactions on Computer Sciences, pp. 31–37, Apr. 2019, IEEE.