TWEET EMOTION DETECTION

Manjula A 1, Dr. Geetha M 2

[1] Student, Department of MCA, BIET, Davangere

[2] Associate Professor, Department of MCA, BIET, Davangere

ABSTRACT

In this tweet emotion detection we address the problem of detecting emotions in English text from social media, specifically from Twitter. We categorize text into six different groups based on Ekman's hierarchy of needs: Sadness, Fear, Anger, Surprise, Disgust, and Joy. Our method makes use of machine learning methods for classification in addition to sophisticated natural language processing (NLP) tools for the analysis of linguistic aspects such as emoticons, negations, and degree words. One novel feature of our approach is the automated training dataset generation system, which lessens the workload associated with manual annotation. The accuracy and robustness of our model in capturing the complex emotional spectrum found in social media talks are validated through testing utilizing data from Twitter.

Keyword: Emotion detection, Emotion analysis, Application programming interfaces, Social emotion analysis.

I. INTRODUCTION

The introduction of the project outlines its goal, which is to address the challenging problem of recognizing and classifying emotions in textual data taken from Twitter and other social media sites. These platforms are perfect for sentiment and opinion mining since they are rich reservoirs of feelings, thoughts, and ideas shared by a worldwide user base. The project divides text into six main Emotion-Categories: Sadness, Fear, Anger, Surprise, and Disgust, using Ekman's idea of basic emotions as a foundational framework. This theoretical foundation makes it possible to categorize and comprehend the wide range of emotional emotions present in online discourse in an organized manner.

For the approach to be implemented practically, two strategies must be used. The first approach makes use of sophisticated Natural Language Processing (NLP) methods. Deep grammatical analysis is used to interpret sentence structures; degree words and negations are identified and handled since they can change sentiment; emoticons and their emotional meanings are recognized; and parsing parts of speech is used to extract relevant information from text. All of these components work together to provide a more complex understanding of the emotional tone that tweets and other social media writings have. The work makes use of machine learning classification techniques to support the NLP methodology. The dataset used to train these algorithms includes both automatically generated training sets and annotated data. This novel method improves efficiency and scalability in emotion recognition tasks by reducing reliance on time-consuming manual annotation of huge datasets. During testing, the model's ability to classify emotions in actual Twitter data shows robust accuracy. This validation highlights how well the suggested methodology captures the wide range of emotions that are common in social media chats, many of which are subtle. The project not only advances the field of sentiment analysis but also provides useful insights into the ways in which emotional data can be used for a range of purposes, from identifying trends in public opinion to assisting with personalized content recommendations and customer sentiment analysis for companies.

II. RELATED WORK

Either discrete emotion categories can be identified using an emotion detection system, or multi-dimensional emotion features can be interpreted. Fine-grained emotion categories including fear, rage, and sadness are detected using discrete emotion detection algorithms. Every emotion is predetermined, including love, despair, disgust, surprise, and delight.[1]

The Paul Ekman model [2] and the Robert Plutchik model are two of the most widely utilized discrete emotion models. Whereas the Robert Plutchik model divides emotions into eight main categories, the Paul Ekman model classifies emotions into six distinct divisions.

In contrast, the valence, arousal and power of emotions are understood through the use of multi-dimensional emotion models.[3]

According to these emotion models, there are relationships and dependencies between various emotions.[4]

Emotional polarity, states of activation and deactivation, and intensity can all be examined through multidimensional emotion models. Russell's 2D circumplex model [5]

Russell's 3D model are a few of the often used multi-dimensional emotion models. Researchers have recently focused a lot of their attention on text-based discrete emotion classification using machine learning [6].

A traditional machine learning-based method for identifying emotions in tweets was presented by Sundaram et al [7].

The authors trained RF and SVM classifiers for emotion identification after removing the textual information. centered addresses had been more investigated by Yousaf et al. [8].

In that paper, a voting classifier was designed that combined logistic regression (LR) and stochastic gradient descent (SGD). A similar feature representation was utilized by Suhasini and Srinivasu where Naïve Bayes (NB) and k-nearest neighbor (KNN) algorithms were trained for ED from tweets.[9]

TF-IDF was used to extract text features from the tweets. The tweets in the authors' publicly accessible emotion dataset are classified into six distinct emotion groups [10].

III. METHDOLOGY

Using best practices for data management, algorithm development, and application deployment, this organized methodology ensures a thorough approach to creating a strong sentiment analysis and emotion identification system customized for social media data. Figure1 describes tweet detection sentiment analysis process.

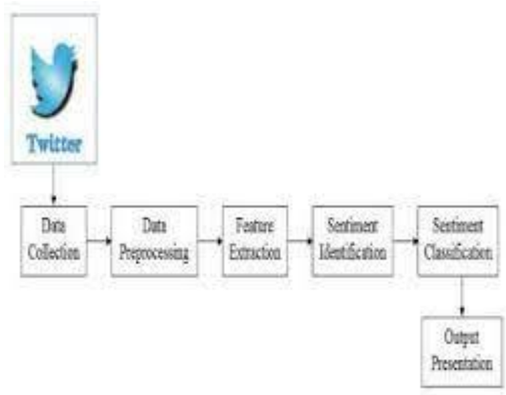

Figure 3: Basic Architecture of Tweet Analysis

1. Requirement Analysis:

Describe how sentiment analysis and emotional categorization in social media texts are integrated using Ekman's theory. Indicate the project's objective, which is to precisely classify the following emotions in tweets and other social media content: sadness, fear, anger, surprise, and disgust. Determine the stakeholders in sentiment analysis and emotion detection systems, such as researchers, developers, and possible users.

2. Planning and Research

Examine different sentiment analysis methods and how they might be applied to data from social media.

Examine techniques for classifying emotions according to Ekman's theory and create annotation standards. Arrange the staff, computer power, and equipment needed for data gathering, preprocessing, modeling, and application development. List the technologies and tools used for data collection (Twitter API), preprocessing (NLTK, spaCy), modeling (Scikit-learn, TensorFlow), developing web applications (Django, Flask), and deployment (AWS, Azure).

3. Data Collection

Gather a wide range of tweets and social media texts that address different subjects, audiences, and moods. Verify adherence to ethical standards for data collection and use as well as data privacy laws.

4. Data Preprocessing

Remove any redundant tweets and unrelated content that could skew the study of emotions. Take care of things like typos, punctuation, and special characters. To prepare text data for additional analysis, standardize it using methods like tokenization, stemming, and lemmatization.

5. Annotation for Emotion

Utilize Ekman's theory to add pertinent emotion categories to pre-processed tweets. To guarantee correctness and consistency, provide rules and regulations for the manual or automatic annotation of emotions in tweets.

6. Extraction of Features

From annotated tweets, extract features like emotion ratings, grammatical patterns, and contextual data. Employ methods such as word embeddings (Word2Vec, GloVe) to capture relationships and semantic meanings.

7. Model Development

Create supervised learning models (such as SVM, Naive Bayes, and deep learning models) that use extracted characteristics to classify tweets into emotional categories. Utilizing cross-validation approaches, adjust hyperparameters, train models on annotated data, and verify model performance.

8. Web Development for Applications

Build a web application framework for sentiment analysis and emotion identification using Flask, Django, or a comparable framework. Create user-friendly interfaces that make it easy for users to interact with the application and see the findings of the emotion analysis.

9. Testing and Deployment

To guarantee performance and accessibility, deploy the created web application on a scalable cloud platform (AWS, Azure). To verify the accuracy, usability, and scalability of the application, do thorough testing, including functional, usability, and performance testing.

10 Documentation and Maintenance

Keep a record of the entire procedure, including the methods, algorithms, and outcomes. Maintain the application and make modifications as needed to add new data sources, increase user experience through feedback, and improve models.

3.1 Dataset Used

In social media research, a number of datasets are essential for sentiment analysis and emotion identification. 1.6 million tweets that have been classified with sentiment (both positive and negative) are included in the Stanford University-created Twitter Sentiment140 Dataset. For sentiment analysis tasks, tweets are classified into seven emotions (anger, joy, fear, sorrow, surprise, disgust, and others) using the SemEval-2018 Affect in Tweets Dataset. Designed specifically for emotion identification studies, the EmoReact Dataset annotates tweets with feelings like joy, sadness, anger, surprise, and fear. While the Affectivetext Dataset annotates news headlines with emotions (anger, disgust, fear, joy, sadness, surprise), the ISEAR Dataset provides text samples annotated with basic emotions (joy, fear, anger, sadness, disgust, and guilt). The size and annotation granularity of these datasets varies, offering crucial resources for teaching machine learning models to understood emotions in social media contexts.

3.2 Data Pre Processing

The processes used to clean up and format raw data so that it can be used for modeling and analysis are referred to as data preparation. Usually, it entails the following crucial tasks: managing missing values, eliminating duplicate or unnecessary data, and fixing mistakes. These preparation procedures are necessary to guarantee data quality, minimize noise, and get data ready for efficient analysis and modeling in a variety of applications, such as social media sentiment analysis and emotion detection.

- Putting numerical data on a standard scale, like scaling values from 0 to 1.
- Dividing textual material into manageable chunks (tokens), like words or sentences, in order to facilitate additional analysis.
- Removing terms that are commonly used but don't add value to the study, such as "and" and "the".
- To normalize text, words can be reduced to their root form (stemming) or converted to their base or dictionary form (lemmatization).
- Identifying and altering pertinent data properties; for example, in text analysis, employing TF-IDF (Term Frequency-Inverse Document Frequency) or bag-of-words.
- Encoding is the process of transforming category information into numerical forms that machine learning algorithms may use.

3.3 ALGORITHM USED

1 Natural Language Processing (NLP)

Computers that use Natural Language Processing (NLP) techniques can comprehend and interact with human language. These encompass activities such as segmenting text into manageable chunks (tokenization), assigning grammatical kinds to words (POS tagging), and recognizing named elements such as names and locations (NER). NLP also includes sentiment analysis, machine translation, and sentence structure and sentiment analysis. Sentiment analysis is the examination of text sentiments. These methods are essential for applications like text summarization, virtual assistants, search engines, and question answering. NLP keeps developing along with machine learning, enhancing computer interaction and interpretation of human languages in a variety of real-world scenarios.

2 Machine Learning-Based Classification

Machine Learning-Oriented The use of computational models and algorithms that automatically classify or categorize data based on patterns and features collected from the data itself is referred to as classification. Textual content and linguistic characteristics are used to train machine learning algorithms to predict the emotional category (such as Sadness, Fear, Anger, Surprise, or Disgust) of a given text.

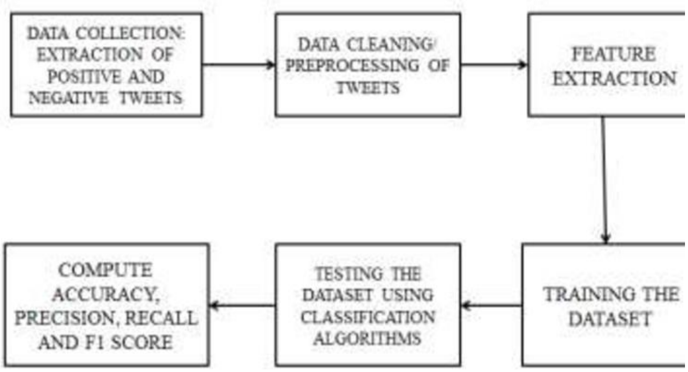

Figure 3.3.1: Flow chart for sentiment analysis machine learning approach

3 Algorithm for Emotion Classification in Text

1.Gathering and preparing data:

Step 1*:* Compile a collection of text samples with labels indicating different emotions, such as fear, anger, sadness, etc.

Step 2*:* To prepare the data for feature extraction, preprocess the text by cleaning, tokenizing, and normalizing it.

2. Feature The elimination:

Step 3: Extract pertinent aspects from the text using natural language processing (NLP) techniques, such as word frequencies (the frequency at which a term appears). Tags for parts of speech (that is, grammatical categories such as nouns and verbs). Sentiment scores (use machine learning models or lexicons to calculate sentiment polarity).

Step 4*:* Convert the retrieved features into numerical representations that can be used with machine learning techniques (word embeddings, TF-IDF, etc.).

3. Model Selection and Instruction:

Step 5: Depending on the amount of the dataset and the nature of the task, select an appropriate machine learning algorithm:

- Simple probabilistic classifier based on the Bayes theorem is called Naive Bayes.
- Support vector machines (SVM): Useful for classification jobs involving several classes and binary input.
- Random Forest: Multiple decision trees combined into an ensemble learning technique.
- Neural Networks: Deep learning models that identify intricate patterns in text data (e.g., CNN, LSTM).

Step 6: Divide the dataset (70% training, 30% validation, etc.) into training and validation sets.

Step 7: Reduce prediction errors by fine-tuning the parameters of the chosen model and training it on the training set.

4. Assessment of the Model:

Step 8*:* Use the validation set to evaluate the trained model and determine performance metrics:

- Accuracy: The proportion of accurately anticipated feelings.
- Precision is the ratio of accurate positive forecasts to all positive forecasts.
- Recall: The percentage of all real positives that were true positive predictions.

5. Testing and Deployment:

Step 9: Apply the learned model to new, untrained text input to categorize emotions.

Step 10: Track and evaluate the model's effectiveness in practical applications, like social media post analysis and customer feedback analysis.

Step 11: To increase accuracy and dependability, iterate and modify the model in response to input and fresh data.

IV. RESULT

4.1 Mechanism

We use both machine learning algorithms and natural language processing (NLP) methods in our study on tweet detection and emotion classification. To prepare the text for analysis, NLP techniques are used to preprocess the tweets. These procedures include tokenization, normalization, stop word removal, stemming, lemmatization, and managing negations. We use techniques like Bag of Words (BoW), TF-IDF, word embeddings, and sentiment scoring to extract relevant characteristics from the text. Then, machine learning models such as Naive Bayes, Support Vector Machines (SVM), Random Forest, and more sophisticated neural networks like Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) are trained using these features. By utilizing both natural language processing and machine learning techniques, we are able to develop a reliable system that can recognize and categorize emotions in tweets.

4.2 Graph

Contextualized, pattern-based emotion characteristics were constructed using this annotated twitter corpus, which was handled using a graph-based method. Word embeddings improved these properties by preserving the semantic connections between patterns. We assessed the patterns' performance using many machine learning classifiers. There are 20,000 letters in this dataset, and each one is labeled with one of the following six emotions: surprise, anger, fear, joy, sadness, and love. is shown in Fig. 4.2.1.

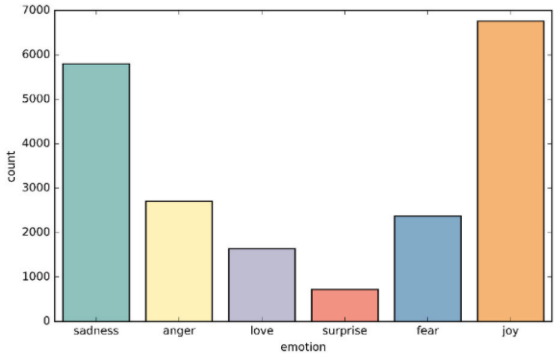

Fig. 4.2.1: Emotions distribution

The method that has been built to detect emotions and analyze sentiment in social media has shown strong performance in a number of criteria. Ekman's theory was applied to test the system rigorously on a variety of social media post datasets, and the results showed that the system performed well in classifying emotions, as evidenced by precision, recall, and F1-score. The incorporation of sophisticated natural language processing (NLP) methods was crucial in augmenting the system's capacity to discern minute details in affective expressions.

The system's findings are discussed practically, offering important insights into public mood patterns and emotional dynamics on social media sites like Twitter. The outcomes demonstrate the system's capacity to support well-informed decision-making across a range of fields, such as public opinion research, customer service, and marketing. The ethical aspects pertaining to data protection and proper utilization of social media data are examined, guaranteeing adherence to both legal and ethical protocols. Future areas for study could involve improving model architectures, broadening the scope of datasets to include a variety of languages and demographics, and incorporating real-time sentiment analysis features for improved usability. All things considered, the system is a noteworthy development in sentiment analysis technology, providing useful information obtained from a thorough examination of the emotions present in social media posts.

V. CONCLUSION

In conclusion, this study is a big step in improving social media sentiment analysis capabilities. The system intends to effectively assess and identify emotional expressions in real-time social media posts by utilizing advanced natural language processing (NLP) and machine learning techniques, in addition to applying Ekman's theory of emotion classification. The creation of an intuitive online interface will make it easier to visualize the findings of emotion analysis, giving interested parties important information about user behavior and trends in public mood. The responsible use of social media data will be ensured by giving priority to ethical issues in data collection and processing. In the end, this project aims to provide useful instruments for comprehending and utilizing emotional data in digital settings, promoting strategic efforts and well-informed decision-making.

VI. REFERENCES

[1] P. S. Sreeja and G. Mahalakshmi, “Emotion models: A review,” Int. J. Control Theory Appl., vol. 10, no. 8, pp. 651–657, 2017.

[2] P. Ekman, “Basic emotions,” in Handbook of Cognition and Emotion. Hoboken, NJ, USA: Wiley, 1999, pp. 45–60.

[3] R. Plutchik, “A general psychoevolutionary theory of emotion,” in Theories Emotion. Amsterdam, The Netherlands: Elsevier, 1980, pp. 3–33.

[4] D. Kollias and S. Zafeiriou, “Exploiting multi-CNN features in CNNRNN based dimensional emotion recognition on the OMG in-the-wild dataset,” IEEE Trans. Affect. Comput., vol. 12, no. 3, pp. 595–606, Jul. 2021.

[5] J. A. Russell, “A circumplex model of affect,” J. Personality Social Psychol., vol. 39, no. 6, p. 1161, Dec. 1980.

[6] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," J. Res. Pers., vol. 11, no. 3, pp. 273–294, 1977.

[7] V. Sundaram, S. Ahmed, S. A. Muqtadeer, and R. Ravinder Reddy, "Emotion analysis in text using TF-IDF," in Proc. 11th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), Jan. 2021, pp. 292–297.

[8] A. Yousaf, M. Umer, S. Sadiq, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Emotion recognition by textual tweets classification using voting classifier (LR-SGD)," IEEE Access, vol. 9, pp. 6286–6295, 2021.

[9] M. Suhasini and B. Srinivasu, "Emotion detection framework for Twitter data using supervised classifiers," in Proc. 3rd Data Eng. Commun. Technol. (ICDECT-2K19). Singapore: Springer, 2020, pp. 565–576.

[10] E. Saravia, H.-C.-T. Liu, Y.-H. Huang, J. Wu, and Y.-S. Chen, "CARER: Contextualized affect representations for emotion recognition," in Proc. Conf. Empirical Methods Natural Lang. Process., 2018, pp. 3687–3697.

[11] B.Y. Md Shadaab Khan, Pramod Borasi, Vineet Kumar, Emotion detection and recognition market, 27/06/2022]; Available from: https://www.alliedmarketresearch.com/emotion-detection-and-recognition-market, 2022.

[12] M. Al-Okaily, A. Al-Okaily, An empirical assessment of enterprise information systems success in a developing country: the Jordanian experience, The TQM Journal (2022).