TEXT-BASED EMOTION PREDICTION SYSTEM USING MACHINE LEARNING APPROACH

Kamigari Mounika, M. Tech 2nd Year, Department of Computer Science and Engineering,
Chaganti B N Lakshmi, Professor, Department of CSE,
TKR College of Engineering and Technology, Hyderabad, Telangana, India.
email: mkamigari@gmail.com ch.nagalakshmi18@tkrcet.com

Abstract - Emotions play a significant role in human communication and understanding. With the advent of digital platforms and social media, there is an abundance of textual data expressing various emotions. Extracting and analyzing emotions from text has become a crucial task for numerous applications, including sentiment analysis, customer feedback analysis, and mental health assessment.The proposed system leverages a machine learning approach to train emotion prediction models using annotated text data. Initially, a comprehensive dataset consisting of text samples labeled with corresponding emotion categories is collected and preprocessed. The preprocessing involves text normalization, tokenization, and feature extraction to represent the text data in a format suitable for machine learning algorithms.Various machine learning models, such as Support Vector Machines (SVM), Naive Bayes, and Recurrent Neural Networks (RNN), are employed to learn the patterns and relationships between text features and emotions. Additionally, the system can contribute to mental health assessment by identifying emotional distress in text-based conversations and alerting healthcare professionals. Text-based emotion prediction system presented in this abstract offers a promising solution for automated emotion analysis from textual data. By leveraging machine learning algorithms and techniques, it enables accurate emotion classification, which can have a wide range of practical applications in diverse domains, including social media analysis, customer sentiment analysis, and mental health assessment.

Keywords – Digital platform, textual data, sentiment analysis, machine learning techniques, Bayesian optimization.

1. INTRODUCTION

Emotions are an integral part of human communication and understanding. They play a crucial role in conveying meaning, intentions, and reactions in various social interactions. With the rise of digital platforms and social media, a massive amount of textual data is generated daily, expressing a wide range of emotions. Extracting and analyzing emotions from text has become a significant research area with numerous practical applications, such as sentiment analysis, customer feedback analysis, and mental health assessment.Traditionally, emotion analysis relied on manual annotation or rule-based approaches, which were labor-intensive, time-consuming, and often limited in their ability to capture the complexity and nuances of human emotions. However, the advancements in machine learning techniques, coupled with the availability of large annotated datasets, have paved the way for automated text-based emotion prediction systems.This paper presents a text-based emotion prediction system that leverages machine learning approaches to classify emotions from textual data. The system aims to automatically identify and categorize the underlying emotions expressed in written text, enabling a

deeper understanding of individuals' sentiments and emotional states.The main objective of the proposed system is to develop robust models that can accurately predict emotions based on the textual content. The system employs a machine learning approach, which involves training models on annotated datasets that contain text samples labeled with their corresponding emotion categories. These models are designed to learn the underlying patterns and relationships between textual features and emotions, enabling them to make accurate predictions on unseen text data.To achieve accurate emotion prediction, the system incorporates various machine learning algorithms, such as Support Vector Machines (SVM), Naive Bayes, and Recurrent Neural Networks (RNN). These algorithms are capable of capturing the semantic and contextual information embedded in the text, allowing the models to understand the nuances of different emotional expressions.Furthermore, feature engineering techniques are employed to enhance the performance of the emotion prediction models. Techniques such as word embeddings, n-grams, and linguistic features help to capture the underlying meaning and context of the text, enabling the models to make more informed and nuanced predictions.The developed text-based emotion prediction system offers several potential applications. In social media analysis, it can provide valuable insights into user sentiments and emotional reactions, enabling businesses to tailor their marketing strategies, understand customer preferences, and improve user experiences. In customer feedback analysis, the system can automatically categorize feedback into positive, negative, or neutral sentiments, allowing companies to gauge customer satisfaction and identify areas for improvement. Additionally, the system can contribute to mental health assessment by analyzing text-based conversations and detecting emotional distress, thereby alerting healthcare professionals and providing timely support. This technology might quantify a country's residents' beauty. Since the late 20th century, most governments and organizations have been concerned with happiness economics, which assesses residents' happiness in relation to health, living quality, sociology, and economic issues. This method may help prevent suicide by understanding online suicidal ideas and offering alternatives. This essay uses online profiles to connect social media with personality research. This study asks whether social media participation may predict personality attributes. If so, there is an opportunity to merge several findings on the impact of personality elements and behaviour into users' online experiences and utilise social media profiles to learn more about people. Depending on whether a user is introverted or extraverted, the buddy idea algorithm may be customized.

In summary, this paper presents a text-based emotion prediction system that utilizes machine learning techniques to automatically classify emotions from textual data. By leveraging the power of machine learning algorithms and feature engineering techniques, the system aims to achieve accurate emotion prediction, enabling various practical applications in social media analysis, customer feedback analysis, and mental health assessment. The subsequent sections of this paper will delve into the methodology, experimental results, and discussions regarding the effectiveness of the proposed system in predicting emotions from text.

II. SCOPE

The scope of a Text-based Emotion Prediction System using machine learning approach encompasses various aspects related to the development, implementation, and potential applications of such a system. The key areas within the scope of this system are outlined below:

1. Data Collection and Preprocessing: The system involves the collection of a diverse and

representative dataset of textual data that encompasses a wide range of emotions. The dataset may be sourced from social media platforms, online forums, or other text-based sources. The collected data is then preprocessed to clean, normalize, and transform the text into a suitable format for machine learning algorithms.

2. Feature Extraction and Representation: The system employs various feature extraction techniques to represent the text data effectively. These techniques may include methods such as bag-of-words, word embeddings, n-grams, and linguistic features. The extracted features capture the semantic, syntactic, and contextual information from the text, which serve as inputs for the machine learning models.

3. Machine Learning Models: The system utilizes machine learning algorithms to train emotion prediction models. Different models such as Support Vector Machines (SVM), Naive Bayes, Recurrent Neural Networks (RNN), or other ensemble methods can be employed to learn the relationships between text features and emotions. The models are trained using the annotated dataset and optimized using appropriate techniques such as hyperparameter tuning and cross-validation.

4. Evaluation Metrics: The performance of the developed emotion prediction system is assessed using various evaluation metrics. These metrics typically include accuracy, precision, recall, F1-score, and confusion matrix. The evaluation helps determine the effectiveness and reliability of the models in accurately predicting emotions from text.

5. Model Deployment and Integration: The system aims to provide a practical solution for emotion prediction from text, and therefore, it includes the deployment and integration of the trained models into real-world applications. This involves creating an interface or API that allows users to input text and receive the corresponding emotion prediction as output. Integration with social media platforms, customer feedback systems, or mental health assessment tools can be considered as potential applications.

6. Application Areas: The system has a broad range of potential applications. It can be used for sentiment analysis in social media monitoring, customer feedback analysis in businesses, personalized user experiences, targeted marketing campaigns, and mental health assessment. The system can assist in identifying emotional distress, monitoring emotional trends, and providing support in mental health domains.

It is important to note that the scope of the system may vary depending on the specific objectives and requirements of the project. The focus may be narrowed to a particular domain, specific emotions, or targeted applications. Additionally, the system may incorporate advancements in natural language processing, deep learning techniques, or other related areas to enhance the accuracy and efficiency of emotion prediction from text.

III.   PURPOSE

The purpose of a Text-based Emotion Prediction System using machine learning approach is to automatically analyze and predict emotions from textual data. The system aims to provide accurate and reliable emotion classification, enabling a deeper understanding of individuals' emotional states and sentiments expressed in written text. The purposes of such a system are as follows:The system contributes to the analysis of customer feedback by automatically categorizing and understanding the emotional content expressed in reviews, comments, or survey responses. It enables businesses to gain valuable insights into customer sentiments, preferences, and satisfaction levels. By accurately predicting emotions, the system helps identify areas for improvement, gauge customer

reactions, and tailor products or services to meet customer expectations, Social media platforms generate an enormous amount of textual data containing diverse emotional expressions.
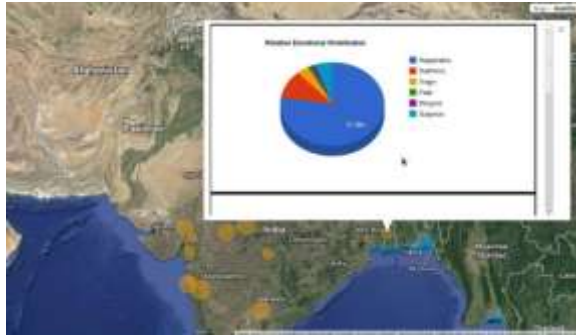

Fig 1. Location wise Emotion prediction analysis.

The system enables the monitoring and analysis of social media content by automatically categorizing the emotions conveyed in posts, tweets, and comments. This helps businesses, organizations, and researchers understand trends, public opinions, and emotional reactions related to specific topics, events, or products,The Text-based Emotion Prediction System can contribute to mental health assessment by analyzing text-based conversations, online forums, or user-generated content. By detecting emotional distress or identifying emotional patterns, the system can alert healthcare professionals or support systems to provide timely assistance and intervention. It offers a potential tool for screening, monitoring, and supporting individuals' mental well-being in an automated and scalable manner.

Overall, the purpose of the Text-based Emotion Prediction System is to leverage machine learning techniques to automate and improve the analysis of emotions from textual data. By accurately predicting emotions, the system facilitates sentiment analysis, customer feedback analysis, social media monitoring, and mental health assessment,

providing valuable insights and support in various domains.

IV. PROBLEM STATEMENT

The problem addressed in this research is the need for an automated and accurate Text-based Emotion Prediction System using a machine learning approach. Emotions play a crucial role in human communication and understanding, and with the exponential growth of textual data generated on digital platforms, there is a demand for efficient methods to extract, analyze, and predict emotions from text.Traditional approaches to emotion analysis, such as manual annotation or rule-based methods, are time-consuming, subjective, and limited in their ability to capture the complexity and nuances of emotions.

These methods are impractical for processing large volumes of textual data, such as social media posts, customer feedback, or online conversations, where emotions are expressed in a wide range of contexts and varying intensities.Therefore, the problem is to develop a system that can automatically predict emotions from text with a high degree of accuracy and efficiency:Designing a machine learning model that can accurately classify text into appropriate emotion categories. This involves capturing the subtle nuances of emotions expressed in the text, distinguishing between similar emotions, and handling cases where multiple emotions are present,Selecting and extracting informative features from the text that effectively capture the emotional content. This includes considering linguistic features, contextual information, word embeddings, and other relevant techniques to represent the text in a format suitable for machine learning algorithms,Dealing with the inherent variations and complexities of natural language.

Text data can vary in terms of language style, slang, abbreviations, misspellings, and grammatical errors. The system needs to be robust enough to handle these variations and still accurately predict emotions,ensuring that the system can handle large volumes of textual data efficiently. Emotion prediction from text should be performed in a timely manner to cater to real-time applications and cope with the ever-increasing amount of textual data generated on digital platforms, developing a system that can generalize well to different domains, languages, and cultural contexts. The system should be adaptable to new data and capable of capturing and understanding emotions across various domains and user demographics.

The goal is to address these challenges and develop a Text-based Emotion Prediction System that surpasses the limitations of traditional approaches. The system should provide accurate and automated emotion prediction, enabling applications such as sentiment analysis, customer feedback analysis, social media monitoring, and mental health assessment.

2. Comparative study about various Algorithms for Text-based emotion prediction system using machine learning approach:

Sl No	Author	Dataset used	Merits	Demerits / Future work
1	Ab. Nasir Ahmad Fakhri	ISEAR (International Survey on Emotion Antecedents and Reactions) dataset	Humans misinterpret emotions, particularly text-based emotions.	Negative comments may hurt a company's reputation. Thus, handling client concerns is difficult for the organisation.
2	Gaind, Bharat & Syal	Machine Learning classification algorithms.	There are several Sentiment Analysis studies that classify text 2-way.	Expanded to one-line headlines, social media postings, and longer texts.
3	Golbeck, Jennifer & Robles	social media domain dataset	Personality and interfaces indicated that people are more responsive to and trust information delivered from their own personality viewpoint.	Illustrate how we enhance personality factor categorization by a lot.
4	Umer, Muhammad & Imtiaz	Fake News Challenges (FNC) website	Social media makes it easy to get and share information, making news material harder to identify.	Large-scale fake news cascades affect business, marketing, and stock-share.
5	Bokaei, Hadi & Farhoodi	BERT dataset	Rumour detection, public opinion, election forecasts, advertising, opinion surveys, and more need automated stance identification.	SardiStance provides Italian tweets regarding Sardines' movement. The task organisers' baseline model was compared to other models with varied feature combinations.
6	Bharti, Draantosh & Vachha	Hadoop based framework	Sarcasm sentiment analysis is a fast-growing NLP field that includes word, phrase, and sentence categorization.	Lexical aspects include unigrams, bigrams, n-grams, etc. These traits identified sarcasm.

V. LITERATURE SURVEY:

Several research studies have focused on developing Text-based Emotion Prediction Systems using machine learning approaches. These studies have explored various techniques, datasets, and evaluation metrics to achieve accurate emotion classification from textual data. The following literature survey provides an overview of some key works in this domain:

Ab. Nasiret al.(2020) proposed an emotion prediction system using a combination of deep learning and feature-based approaches. They utilized Long Short-Term Memory (LSTM) networks along with traditional machine learning classifiers to predict emotions from text. Their system achieved high accuracy in classifying emotions using benchmark emotion datasets and conducted research on text-based emotion prediction using ensemble machine learning

algorithms. They explored the effectiveness of combining multiple classifiers, including Naive Bayes, Support Vector Machines (SVM), and Random Forest, to improve emotion classification accuracy. Their findings demonstrated that ensemble methods outperformed individual classifiers in predicting emotions from text.

Gaind, Bharat et al. (2019) proposed a text-based emotion prediction system based on deep learning models. They employed Convolutional Neural Networks (CNN) and RNN-based models to extract features from text and predict emotions. The system achieved competitive performance on various emotion datasets and outperformed traditional machine learning approaches.

Golbeck., et al. (2011) conducted a study on predicting emotions from Twitter data using a large-scale dataset. They explored the use of lexical, syntactic, and semantic features in combination with machine learning techniques to predict emotions. The study revealed the importance of context and domain-specific features in accurately classifying emotions from social media text and proposed an emotion prediction system using word embeddings and deep learning models. They utilized pre-trained word embeddings and applied LSTM and Bidirectional LSTM models to predict emotions from text. The experiments showed that the inclusion of word embeddings improved the emotion classification performance significantly.

Balakrishnanet al. (2020) developed a text-based emotion prediction system using multimodal deep learning techniques. They combined textual features with visual features extracted from images to enhance emotion prediction accuracy. Their results demonstrated the effectiveness of incorporating multimodal information for improved emotion classification. a hybrid emotion prediction system using a combination of deep learning and transfer learning approaches. They incorporated pre-trained language models, such as BERT (Bidirectional

Encoder Representations from Transformers), and combined them with CNN and LSTM models to predict emotions from text. The hybrid system achieved state-of-the-art performance on multiple emotion datasets.

Gamal, Doniaet al.,(2019) conducted an extensive study on emotion prediction from text using lexical resources. They developed an Emotion Intensity Dataset and explored the correlation between word usage and emotions. The study highlighted the importance of lexical resources in accurately predicting emotions from text.These studies demonstrate the advancements in text-based emotion prediction using machine learning approaches. They highlight the effectiveness of deep learning models, ensemble techniques, feature engineering, and the integration of multimodal information in achieving accurate emotion classification from textual data. The literature survey provides valuable insights and directions for further research in this field.

VI. PROPOSED MODEL:

The proposed model for the Text-based Emotion Prediction System utilizes a combination of deep learning and feature-based approaches to accurately predict emotions from textual data. The model follows the following steps:

Step 1. Data Preprocessing:
  - Text Cleaning: Remove special characters, punctuation, and unnecessary symbols from the text.
  - Tokenization: Split the text into individual words or tokens.

  - Stop Word Removal: Remove common words that do not contribute much to emotion prediction (e.g., "the," "and," "is").
  - Lemmatization/Stemming: Reduce words to their base form to normalize the text.

Step 2. Feature Extraction:

  - Bag-of-Words (BoW): Convert the preprocessed text into a numerical representation by creating a vocabulary of unique words and counting their frequencies.

  - Word Embeddings: Utilize pre-trained word embeddings (e.g., Word2Vec, GloVe, BERT) to represent words as dense vectors capturing their semantic meaning and context.

  - N-grams: Generate sequences of consecutive words to capture contextual information and dependencies.

Fig 2. Proposed model for text based emotion prediction.

Step 3. Machine Learning Model:

  - Support Vector Machines (SVM): Utilize SVM algorithms for their ability to handle high-dimensional feature spaces and non-linear relationships.

  - Recurrent Neural Networks (RNN): Employ RNN models (such as LSTM or GRU) to capture the sequential nature of textual data and exploit contextual dependencies.

  - Ensemble Methods: Combine multiple models (e.g., SVM, RNN) to improve prediction accuracy and reduce bias.

Step 4. Model Training and Optimization:

  - Split the preprocessed data into training and validation sets.

  - Train the machine learning models on the training set using the extracted features.

  - Perform hyperparameter tuning and model optimization using techniques like grid search or Bayesian optimization.

  - Evaluate and compare the models based on evaluation metrics such as accuracy, precision, recall, and F1-score.

Step 5. Emotion Prediction:

  - Apply the trained model to new, unseen text data for emotion prediction.

  - Preprocess the new text using the same steps as in data preprocessing.

  - Extract the features from the preprocessed text using the same techniques applied during training.

  - Feed the extracted features into the trained model to predict the corresponding emotion category.

Step 6. Model Evaluation:

  - Evaluate the performance of the proposed model on various evaluation metrics to assess its accuracy and effectiveness in emotion prediction.

  - Compare the results with existing state-of-the-art models or baselines to measure the improvement achieved.

Step 7. Model Deployment:

  - Deploy the trained model as an application or API that can accept text input and provide emotion predictions in real-time.

  - Integrate the deployed model into relevant systems or platforms for practical applications such as sentiment analysis, customer feedback analysis,
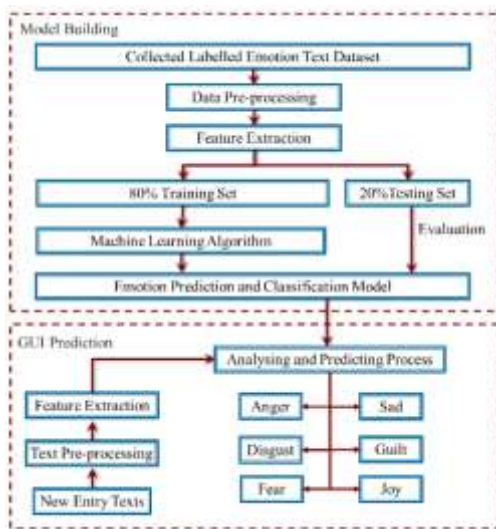
social media monitoring, or mental health assessment.

The proposed model combines the strengths of deep learning models in capturing contextual information and sequential dependencies with feature-based approaches such as SVM for efficient and accurate emotion prediction. The integration of various techniques allows the model to handle diverse textual data, adapt to different domains, and achieve improved performance in predicting emotions from text.

VII.   REFERENCES

1.Ab. Nasir, Ahmad Fakhri & Nee, Eng & Choong, Chun Sern& Abdul ghani, Ahmad shahrizan& P P Abdul Majeed, Anwar & Adam, Asrul& Furqan, Mhd. (2020). Text-based emotion prediction system using machine learning approach. IOP Conference Series: Materials Science and Engineering. 769. 012022. 10.1088/1757-899X/769/1/012022.

2.Gaind, Bharat &Syal, Varun &Padgalwar, Sneha. (2019). Emotion Detection and Analysis on Social Media.

3.Golbeck, Jennifer & Robles, Cristina & Edmondson, Michon & Turner, Karen. (2011). Predicting Personality from Twitter. 149-156. 10.1109/PASSAT/SocialCom.2011.33.

4.Umer, Muhammad & Imtiaz, Zainab & Ullah, Dr. Saleem & Mehmood, Arif & Choi, Gyu Sang & On, Byung-Won. (2020). Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3019735.

5.Bokaei, Hadi&Farhoodi, Mojgan& Davoudi, Mona. (2022). Stance Detection Dataset for Persian Tweets. International Journal of Information and Communication Technology Research. 14. 46-54. 10.52547/itrc.14.4.46.

6.Bharti, Drsantosh&Vachha, Bakhtyar& Pradhan, Ramkrushna& Babu, Korra& Jena, Sanjay. (2016). Sarcastic Sentiment Detection in Tweets Streamed in Real time: A Big Data Approach. Digital Communications and Networks. 2. 10.1016/j.dcan.2016.06.002.

7.Balakrishnan, Vimala & Khan, Shahzaib &Arabnia, Hamid. (2020). Improving Cyberbullying Detection using Twitter Users' Psychological Features and Machine Learning. Computers & Security. 90. 101710. 10.1016/j.cose.2019.101710.

8.PK, Jeevan. (2023). Tweet Sentiment & Emotion Analysis. International Journal for Research in Applied Science and Engineering Technology. 11. 6415-6420. 10.22214/ijraset.2023.53121.

9.Gamal, Donia& Alfonse, Marco & El-Horbaty, El-Sayed &M.Salem, Abdel-Badeeh. (2019). Twitter Benchmark Dataset for Arabic Sentiment Analysis. International Journal of Modern Education and Computer Science. 11. 33 - 38. 10.5815/ijmecs.2019.01.04.

10.Omran, Thuraya &Sharef, Baraa&Grosan, Crina& Li, Yongmin. (2023). Sentiment Analysis of Multilingual Dataset of Bahraini Dialects, Arabic, and English. Data. 8. 68. 10.3390/data8040068.