



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 7, July 2026

Emotion Analysis from Social Media using Deep Learning: A Systematic Review Focused on Twitter Data

Swadhin Nabin Sahu¹, Dr. Hetal Rahul Modi²

Assistant Professor, Shri Shambhubhai V. Patel college of Computer Science and Business Management, VNSGU,
Surat, India¹

Assistant Professor, Bhagwan Mahavir College of Computer Application, Bhagwan Mahavir University, Surat, India²

ABSTRACT: In the digital era, rapid technological advancements have led to increase large volume of textual data, most of which are unstructured, creating a strong need for effective methods to extract meaningful information. Emotion analysis from social media has become an important research area because of the abundance of user-generated content that reflects public opinion, mental health cues, and behavioral patterns. Twitter, a widely used microblogging platform, offers rich, short, and real-time text data that are highly suitable for emotion detection studies. This review systematically examines recent developments in textual emotion detection, focusing on machine learning techniques, emotional models, and commonly used datasets, and analyzes their limitations, challenges, and existing research gaps. It provides a detailed comparative evaluation of traditional deep learning methods and advanced transformer-based architectures, including CNN, LSTM, Bi-LSTM, GRU, BERT, RoBERTa, XLNet, ALBERT, and DeBERTa, highlighting their performance, strengths and weaknesses. This study also discusses preprocessing strategies, training techniques, and emerging trends, ultimately identifying unresolved challenges, suggesting future research directions to build more accurate, and robust emotion detection systems for social media data.

KEYWORDS: Emotion Detection, Machine Learning, Natural Language Processing, Deep Learning, transformer models, BERT, RoBERTa, LSTM, Fine Tuning

I. INTRODUCTION

Social media platforms, especially Twitter, have become a major medium for people to express personal emotions, political views, product opinions, and real-time responses to events. Understanding these emotions is beneficial for applications such as marketing analytics, disaster response, public health monitoring, and political forecasting and customer feedback analysis. Emotion analysis, aims to classify human emotions, such as joy, sadness, anger, fear, disgust, surprise, and mixed emotions. Unlike general sentiment analysis (positive/negative/neutral), emotion detection deals with more nuanced emotional states. With the increasing complexity of language, sarcasm, emojis, and short-text characteristics of Twitter, traditional machine learning models have struggled to capture contextual meaning. Deep learning models, especially those using attention mechanisms and transformer architectures, have improved accuracy by automatically learning contextual and semantic representations.

Over the past decade, deep learning (DL) models have become the dominant approach for emotion detection in tweets. By integrating machine learning algorithms (e.g., Naïve Bayes, Support Vector Machine, and Random Forest) and deep learning models (e.g., LSTM and BERT), sentiment classification can be significantly improved, enabling more accurate sentiment prediction. By identifying emotions such as frustration, anger, and disappointment, this multidimensional approach allows companies and consumers to derive meaningful insights and make informed decisions [1]. Platforms such as Twitter present unique challenges because of the informal and unstructured nature of user-generated text. Additionally, the use of sarcasm, irony, and cultural references further complicates emotion classification; therefore, advanced deep learning models like BERT (Bidirectional Encoder Representations from Transformers (BERT), A Robustly Optimized BERT Pretraining Approach (RoBERTa), and Decoding-enhanced BERT with Disentangled Attention (DeBERTa) are needed.

Kaushal et al. (2021) Emotion detection from text generally follows four major approaches: keyword-based, rule-based, classical machine learning, and deep learning methods. The keyword-based approach relies on directly matching emotion-related words present in the text; for example, if a sentence contains words like “happy” or “sad,” it is classified accordingly, making this method simple but limited in handling complex expressions. The rule-based approach uses predefined linguistic rules, such as detecting negation patterns like identifying “not happy” as a negative emotion making it more flexible than keyword matching but still dependent on handcrafted rules. In contrast, the classical machine learning approach trains algorithms like Naïve Bayes, SVM, or Logistic Regression on labeled datasets using features such as Bag-of-Words or TF-IDF, enabling the system to learn common emotional patterns for instance, classifying sentences like “I hate this” as anger. Finally, deep learning-based approaches such as LSTM, GRU, CNN, and transformer models like BERT capture deeper context and meaning in text, allowing them to detect subtle emotions and even sarcasm; for example, understanding that the sentence “Oh great, another rainy day” actually expresses frustration rather than positivity. These approaches collectively form the foundation of modern emotion detection systems, ranging from simple rule-based models to advanced context-aware neural networks[15].

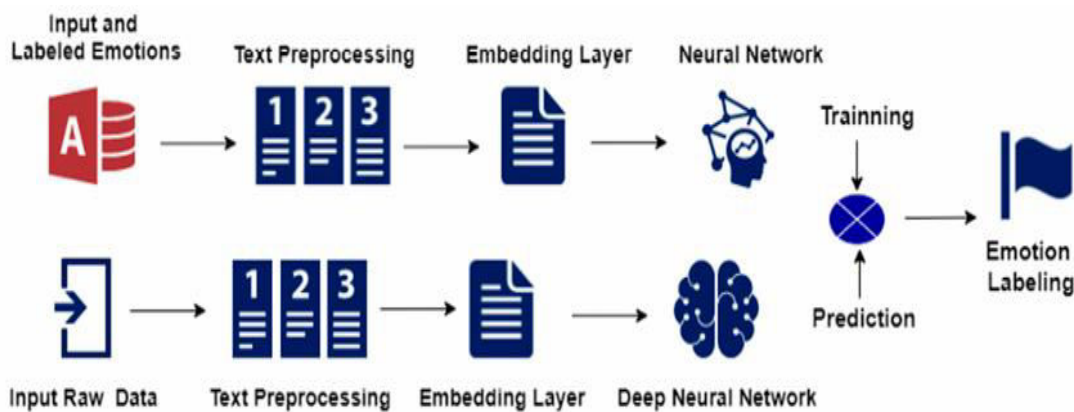


Figure.1.Deep Learning Based approach for emotion detection

II. LITERATURE REVIEW

Many studies have conducted extensive evaluations of transformer models compared to traditional and earlier deep learning approaches, consistently demonstrating their effectiveness in emotion classification. As the field continues to evolve, these advanced models play a crucial role in improving the accuracy, scalability, and contextual understanding required robust sentiment and emotion analysis systems to function.

Similarly, Krishna et al. (2024) aimed to classify emotions from real-time Twitter streams to understand public mood patterns and support domains such as crisis detection and social monitoring [3]. Their research emphasized automatic emotion recognition and demonstrated the importance of transformers and contextual embedding models in processing short, informal, and ambiguous social media texts. The incorporation of deep contextual models significantly improved the accuracy and reliability of emotion detection in rapidly changing online environments.

Sandeep et al. (2025) proposed a hybrid deep learning system that integrates emotion detection with sarcasm identification to address challenges present in product review datasets [1]. Their work leverages transformer-based models, such as BERT and RoBERTa, combined with traditional NLP and machine learning components to improve the interpretability and accuracy of sentiment classification. In another study, Golchine et al. (2025) focused on emotion classification in Twitter posts by categorizing user emotions into four distinct quadrants: Happy-Active, Happy-Inactive, Unhappy-Active, and Unhappy-Inactive [2]. Their goal was to analyze social media data to automatically interpret emotional states for practical applications, such as trend detection, public feedback assessment, and real-time sentiment monitoring. By leveraging deep learning and text modeling methods.

A comprehensive literature review table was prepared to provide a structured comparison of the studies discussed. This table summarizes the key aspects of recent research focused on emotion detection using deep learning and transformer-

based approaches. It includes essential information such as author names, publication year, model type, major findings, and research gaps. Organizing the studies in a tabular format makes it easier to identify patterns, differences, and advancements across various approaches. This structured overview helps in understanding the evolution of emotion detection techniques, enabling researchers to quickly compare methodologies, evaluate performance trends, and identify gaps for future research.

Table 1: Summary of Literature Review

Author(s) / Year	Used Model	Key Findings	Research Gap
Sandeep et al. 2025 [1]	RNN and RoBERTa	RNN model used to detecting sarcasm from Twitter sarcasm corpus, and RoBERTa to produce a final sentiment classification, achieved an accuracy of 89%.	Models perform poorly on product feedback; While Transformer-based models are accurate, and computationally expensive and domain-specific.
Golchine et al. 2025 [2]	BiLSTM + CNN	The researchers trained a model that classifies emotions into four categories, hitting around 93% accuracy, which is a solid step up from past attempts.	The study addresses the challenge of accurately classifying emotions from social media data, which is often noisy and unstructured.
S. Krishna et al. 2024 [3]	Naive Bayes, SVM, Random Forest	SVM performed slightly better than other classifiers in most cases.	Contextual sarcasm and irony were difficult to detect. Hashtag-based labelling may introduce noise and misclassification.
Gupta et al. 2024 [4]	K-NN, LR, SVM, NB, Decision Tree, RF	SVM produced the best result for the model by scoring 96.42 % accuracy, 96.38% f1 score, 96.42% precision and 96.15 % recall.	only employs traditional machine learning classifiers, It does not explore deep learning or transformer-based models like BERT, LSTM, or CNNs.
Kaushik et al. 2024 [5]	Apriori Algorithm and FP-Growth Algorithm.	The Apriori model achieved 97.45% accuracy, especially useful for detecting cyberbullying and product dissatisfaction.	Models struggle to interpret sarcasm, irony, and contextual variations in emotional expression, also lose accuracy in different domains (e.g., product reviews vs. political texts).
Kumar et al. 2024 [6]	CNN and RNN models for emotion classification	Neural Network is a promising method for emotion predictions from text. Using neural network with an accuracy of 99.0%.	Gap in handling slang, and multilingual emotion detection, especially in informal online text (e.g., WhatsApp, social media).
Yohanes et al. 2023 [7]	LSTM, BiLSTM and GRU	GRU outperformed LSTM, BiLSTM, The accuracy of GRU 60.26%, accuracy of BiLSTM is 59.3%, and accuracy of LSTM is 57.65%.	The research focused solely on text-based data and did not incorporate multimodal data such as emojis, images, or context-rich metadata
Sunder et al.	RF,LR, Bidirectional LSTM	Both approach BiLSTM and BiGRU gave almost the same accuracy. BiLSTM achieved an accuracy of	Existing approaches for emotion detection from text are too simplistic and underperforming on nuanced emotional

2022 [8]	(BiLSTM)	(78.25%) and an f1 score is (77.78).	expressions.
Bharti at al. 2022 [9]	CNN + Bi-GRU + SVM	The Hybrid model achieved 80.11% accuracy, SVM achieved 78.97% accuracy and Bi-GRU (DL): 79.46% accuracy.	Hybrid models still struggle with sarcasm, and ambiguous context, so models like LSTM, BERT needed to explore more diverse and large-scale datasets.
Abas at al. 2022 [10]	propose a new model named BERT-CNN to detect emotions	The BERT+CNN model achieves an accuracy of 94.7% and an F1-score of 94% for semeval2019 dataset and an accuracy of 75.8% and an F1-score of 76% for ISEAR dataset.	The author primarily focuses on BERT for emotion detection; there transformer-based architectures like RoBERTa, XLNet, and ALBERT were not explored.
Akalya at al. 2022 [11]	propose CNN combined with LSTM model used	The training accuracy for CNN and CNN-LSTM model were found to be 92.284% and 93.707% respectively.	The study is limited to English text, leaving a gap for multilingual or cross-lingual emotion recognition.

III. DATASET OVERVIEW

In this paper emotion, detection research on social media relies predominantly on Twitter datasets, as tweets offer short, spontaneous, and context-rich user-generated texts that often contain explicit emotional cues. As highlighted by Alhuzali et al. (2022), Twitter remains one of the most widely used platforms for building emotion datasets because the structure of tweets is brief, expressive, and often accompanied by emojis or hashtags that support the efficient extraction of affective information [13].

Recent studies, including Kastrati (2024), have described various methods for collecting emotional data from Twitter, such as using the Twitter API, filtering tweets by hashtags or emotion-related keywords, leveraging emoji-based heuristics, and employing distant-supervision strategies for large-scale annotation [12]. Depending on the dataset size and task requirements, researchers commonly use manual annotation for smaller datasets, semi-automated methods for medium-sized corpora, and distant supervision for large-scale datasets. The growing emphasis on distant supervision techniques, such as using emojis or hashtags as weak labels, has allowed researchers to construct larger and more diverse emotion datasets while maintaining acceptable annotation quality.

Most emotion detection datasets follow classical psychological taxonomies, particularly Ekman's six basic emotions: joy, sadness, anger, fear, disgust, and surprise. Many Twitter-based emotion corpora developed in the past five years have adopted this six-class scheme, either directly or with extensions for finer emotional granularity. Standard public Twitter emotion datasets with these six classes are commonly used as benchmarking resources, whereas larger multi-platform datasets (e.g., derived from Reddit or comment forums) have influenced the design of expanded emotion taxonomies and transfer learning strategies. Together, these datasets enable the evaluation of various deep learning models, including transformer architectures (BERT, RoBERTa, and DeBERTa), recurrent networks (LSTM and GRU), and hybrid systems for both single- and multi-label emotion detection tasks.

The Twitter Sarcasm Corpus, available through Kaggle, contains tweets labeled as "Sarcastic" or "Not Sarcastic" and is widely used for training sarcasm detection models [1]. The dataset included approximately 100,000 tweets for training and 33,000 tweets for testing, with tweets collected using emotion-related hashtags to ensure the presence of affective content [2]. Such datasets are essential when models integrate emotion recognition with sarcasm detection, as sarcasm frequently alters the emotional meaning of the text. In addition to Twitter, other well-established datasets have contributed substantially to emotion detection research. The International Survey on Emotion Antecedents and Reactions (ISEAR) dataset is frequently used as a benchmark for emotion classification tasks. Yohanes et al. (2023) described the dataset as a large-scale survey conducted across 37 countries with nearly 3,000 participants, resulting in

7,666 instances categorized into seven emotions: anger, sadness, disgust, shame, fear, joy, and guilt. Each category contained approximately 1,096 entries, making it a balanced and reliable resource for supervised emotion classification. Akalya et al. (2022) also employed the ISEAR dataset in their study, reaffirming the usefulness of its seven-emotion taxonomy for comparative model evaluation [11]. Similarly, Abas et al. (2022) utilized the complete ISEAR dataset, consisting of 7,666 labeled sentences covering the same seven fundamental emotions, demonstrating its continued relevance for modern deep learning approaches [10].

Overall, the diverse range of datasets, including Twitter-based emotional corpora, sarcasm-labeled collections, and established psychological datasets such as ISEAR, provides a rich foundation for evaluating emotion detection models. These datasets continue to support the development of advanced deep learning systems by enabling consistent benchmarking, cross-domain transfer learning, and robust comparative analyses of state-of-the-art transformer models.

Twitter-based datasets commonly found in reviewed studies:

Table 2: Twitter-based datasets

Dataset Name	Source	Emotions	Size	Notes
SemEval-2019 Task 3	Twitter	Anger, Fear, Joy, Sadness	10k	Most widely used
TweetEval	Twitter	7 emotions	70k	Standard benchmark
Kaggle Twitter Emotion Dataset	Twitter	13 emotions	40k	Balanced
GoEmotions (Google)	Web/Twitter extended	27 emotions	58k	Fine-grained
Custom Twitter datasets	API	Varies	10k	Manually or hashtag-labeled

IV. METHODOLOGY

To ensure robust and reliable experimentation, the dataset undergoes systematic pre-processing before model training. The text data is cleaned by removing special characters, tokenized into sub word units, and standardized through padding and truncation to maintain uniform input lengths. The Dataset split is applied to divide the dataset into training, validation, and testing subsets. The training set is used to help the model learn emotional patterns and linguistic features, while the validation set assists in hyper parameter tuning and detecting issues such as overfitting. The unseen test set is preserved for the final model evaluation, providing an unbiased assessment of performance on real-world data. Transformer-based models, including BERT, RoBERTa, and DeBERTa, are fine-tuned using the Hugging Face Transformers library to enable a fair comparison across architectures.

The complete methodology follows four major stages:

1. Data Preprocessing
2. Tokenization
3. Model Fine-tuning
4. Evaluation using metrics

1. Data Preprocessing:

The preprocessing phase focuses on preparing raw textual data for deep learning models. Tokenization is performed using transformer-specific tokenizers such as BERT Tokenizer, RoBERTa Tokenizer, and DeBERTa Tokenizer, which convert text into token IDs compatible with each architecture. Padding and truncation are applied to achieve a uniform sequence length across all samples. Emotion labels are numerically encoded so that the model can process them. This preprocessing pipeline ensures that the input data is clean, consistent, and structured appropriately for transformer-based learning.

2. Tokenization

Each sentence is transformed into a sequence of tokens based on the sub word vocabulary of the selected model. This process captures both semantic and syntactic components of language. Model-specific tokenizers preserve important linguistic features such as word boundaries, prefixes, and special characters. The tokenized sequences are aligned to the fixed maximum length through padding, enabling efficient batch processing during training.

3. Model Fine-Tuning

The fine-tuning stage involves adapting pre-trained transformer models and the CNN–LSTM baseline to the ISEAR emotion dataset. Pretrained models from Hugging Face are initialized and then trained for three epochs with a batch size of 32. The AdamW optimizer is used along with a learning rate of $2e-5$ to ensure stable convergence. This phase adjusts model parameters to better capture emotion-specific linguistic patterns in the dataset. The uniform training configuration across models guarantees a fair and objective comparison.

4. Evaluation Using Metrics

To assess the effectiveness of each model, standard classification metrics are computed on the test set. These include:

- Accuracy – measures the overall rate of correct predictions.
- Precision – evaluates the proportion of correct positive predictions.
- Recall – measures the ability of the model to capture all relevant emotional instances.
- F1-score – provides a balanced measure combining precision and recall.

Together, these metrics offer a comprehensive understanding of each model’s strengths and limitations in emotion classification.

V. COMPARATIVE ANALYSIS

To clearly highlight the performance differences among the various deep learning and machine learning models used for emotion detection, a comparative analysis table was constructed. This table summarizes the key components of each study, including the model architecture; dataset used for experimentation, and achieved accuracy. Presenting the results in a tabular format enables a straightforward comparison of the performance of different models, such as transformer-based architectures, hybrid neural networks, and classical machine learning classifiers, across diverse datasets, such as Twitter emotion corpora and the ISEAR dataset. This structured comparison supports a deeper understanding of model behavior and guides researchers in selecting suitable architectures for future emotion-detection tasks.

Table 3: shows the performance of each model.

Model(s)	Dataset	Accuracy
RoBERTa[1]	Twitter Sarcasm Corpus	89%
BiLSTM + CNN[2]	Twitter Dataset	93%
KNN,LR,SVM, NB,RF and DT[4]	Emotion Dataset for NLP, The dataset contains 18000 samples.	96.42%
Apriori, FP-Growth [5]	Twitter Dataset, containing 10,000 posts curated using the Rapid Miner tool.	97.45%
BERT+CNN[10]	ISEAR dataset consists of 7666 sentences labelled with seven emotions	95.8%
CNN+LSTM[11]	The model is trained on the ISEAR dataset	93.7%

VI. SUMMARY & FUTURE DIRECTIONS

Although significant progress has been made in emotion detection, several research gaps persist owing to limitations in datasets, model generalization, and linguistic diversity. Most existing studies rely heavily on Twitter or other social media datasets; as a result, models trained on social media text often perform poorly when applied to product-review scenarios, mainly because of domain shift and contextual differences [1]. Additionally, Transformer-based models, while highly accurate, are computationally expensive and domain-specific, reducing their scalability and limiting their real-world deployment in low-resource environments [1]. Another challenge arises from the noisy and unstructured nature of social media data, which makes accurate emotion classification difficult and inconsistent across platforms [2]. Akalya et al. (2022) note that many existing systems struggle to capture contextual, implicit, or subtle emotions, particularly in short or ambiguous texts such as tweets or informal chats, and most studies remain restricted to English, leaving a large gap in multilingual or cross-lingual emotion recognition. Abas et al. (2022) further highlighted that much of the existing work focuses primarily on BERT without comparing alternative transformer architectures such as RoBERTa, DeBERTa, XLNet, or ALBERT, creating a gap in benchmark-level evaluation across diverse models [10].

Beyond these issues, several broader challenges exist, such as sarcasm and irony detection, code-mixed language handling, overlapping emotions within a single text, rapidly evolving Twitter slang, and imbalanced datasets.

VII. CONCLUSION

Emotion detection from text continues to be one of the most challenging tasks in natural language processing because of language ambiguity, overlapping emotional expressions, and the absence of explicit emotion words in many sentences. As noted by Kumari et al. (2023), even with multiple approaches, accurately identifying emotions, especially from short and noisy platforms such as Twitter, remains difficult, motivating the need for stronger deep learning solutions in future studies [14]. Over the past five years, significant progress has been driven by the shift from traditional CNN and LSTM architectures to transformer-based models such as BERT, RoBERTa, and DeBERTa, which consistently achieve state-of-the-art performance on emotion classification tasks. Despite these advancements, persistent challenges such as sarcasm interpretation, multilingual and code-mixed text understanding, emotion intensity prediction, and class imbalance continue to limit their real-world effectiveness. This review emphasizes the importance of developing more robust, multilingual, and real-time emotion detection systems that can adapt to the dynamic nature of social media language. Future research should explore the integration of large-scale multilingual datasets, personalization using large language models, and interpretable or explainable AI techniques to improve trust and usability in practical applications. Overall, this review provides insights into the recent advancements, existing limitations, and emerging opportunities that will shape the next generation of emotion analysis systems for social media.

REFERENCES

- [1] Babu, M. S. S., Suryanarayana, S. V., Sruthi, M., Lakshmi, P. B., Sravanthi, T., & Spandana, M. (2025). Enhancing Sentiment Analysis with Emotion and Sarcasm Detection: A Transformer-Based Approach. *Metallurgical and Materials Engineering*, 794-803.
- [2] Golchin, B., & Riahi, N. (2025). Emotion detection in twitter messages using combination of long short-term memory and convolutional deep neural networks. *arXiv preprint arXiv:2503.20163*.
- [3] Krishna, P. S., Veera Babu, R., Othuri Soniya, Garikapati Bhanu Teja, Pasupuleti Rekha Madhuri, & Javvaji Meghana. (2024). EMOTION RECOGNITION USING TEXT-BASED ANALYSIS USING MACHINE LEARNING: a STUDY ON USER-GENERATED CONTENT [Research Article]. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 11(11), 256–257. <https://www.jetir.org>
- [4] Gupta, s. C., & goel, n. (2024). Emotion analysis of textual contents using natural language processing and text mining. *Journal of Theoretical and Applied Information Technology*, 102(22).
- [5] Kaushik, A., & Chawla, N. (2024). UNVEILING EMOTIONS THROUGH SENTIMENT ANALYSIS.
- [6] Kumar, S. A., & Geetha, A. (2024). Emotion Detection from Text using Natural Language Processing and Neural Networks. *International Journal of Intelligent Systems and Applications in Engineering*, 12(14s), 609-615.
- [7] Machova, Kristina & Szabóova, Martina & Paralic, Jan & Mičko, Ján. (2023). Detection of emotion by text analysis using machine learning. *Frontiers in Psychology*. 14. 10.3389/fpsyg.2023.1190326.
- [8] Sundar, B. S., Rohith, V., Suman, B., & Chary, K. N. (2022). Emotion Detection on text using Machine Learning and Deep Learning Techniques. *International Journal for Research in Applied Science and Engineering Technology*, 10(6), 2277-2286.
- [9] Bharti, S. K., Varadhaganapathy, S., Gupta, R. K., Shukla, P. K., Bouye, M., Hingaa, S. K., & Mahmoud, A. (2022). Text-Based Emotion Recognition Using Deep Learning Approach. *Computational Intelligence and Neuroscience*, 2022(1), 2645381.
- [10] Abas, A. R., Elhenawy, I., Zidan, M., & Othman, M. (2022). BERT-CNN: A Deep Learning Model for Detecting Emotions from Text. *Computers, Materials & Continua*, 71(2).
- [11] Akalya Devi C, Karthika Renuka D, & Sareena Antony. (2022). Emotion Detection using CNN- LSTM based Deep Learning Model on Tweet Dataset. *Journal of Advances in Computational Intelligence Theory*, 4(3), 1–16. <https://doi.org/10.5281/zenodo.7388631>
- [12] Kastrati, M., Kastrati, Z., Shariq Imran, A., & Biba, M. (2024). Leveraging distant supervision and deep learning for twitter sentiment and emotion classification. *Journal of Intelligent Information Systems*, 62(4), 1045-1070.
- [13] Alhuzali, H., Zhang, T., & Ananiadou, S. (2022). Emotions and topics expressed on Twitter during the COVID-19 pandemic in the United Kingdom: Comparative geolocation and text mining analysis. *Journal of Medical Internet Research*, 24(10), e40323.

- [14] Meena Kumari, P., Srinija Reddy, K., Nikhitha, J., Sai Charan, A., Kumar, A., UG Scholars, Department of CSE, AVN Institute of Engineering and Technology, Hyderabad, Telangana, India, & Associate Professor, Department of CSE, AVN Institute of Engineering and Technology, Hyderabad, Telangana, India. (2023). Emotion Detection using Twitter Datasets and Spacy Algorithm (Issue 07, p. 70) [Journal-article].
- [15] Kusal, S., Patil, S., Kotecha, K., Aluvalu, R., & Varadarajan, V. (2021). AI Based Emotion Detection for Textual Big Data: Techniques and Contribution. Big Data and Cognitive Computing, 5(3), 43. <https://doi.org/10.3390/bdcc5030043>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details