Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Text-based Emotion Recognition using Machine Learning through Sentiment Analysis

Alokam Meghana[1], Bilakanti Vanshika[2], Karnam VedaSamhitha[3], Sreevidya B.[4] and Rajesh M.[5]

[1, 2, 3] Department of Computer Science and Engineering, Amrita School of Computing, Bengaluru, Amrita Vishwa Vidyapeetham, India
5,6 Department of Computer Science and Engineering, Amrita School of Computing, Bengaluru, Amrita Vishwa Vidyapeetham, India

Abstract. This application explores the smooth integration of emotion detection features into a web page. It follows a dual-pronged approach by leveraging natural language processing (NLP) for textual data machine learning models and deep learning models for analyzing the emotion in textual data. Powered by a flask backend, the platform utilizes a natural language toolkit (NLTK) for sentiment analysis, enabling users to put their text for customized emotion detection within an interactive web interface. This hybrid model provides a dynamic environment that not only analyses user emotions but also reacts to them in real time by seamlessly integrating machine learning and deep learning with web development. Its adaptability provides a sophisticated user experience. It extends to a multitude of applications, starting from sentiment-aware recommendation systems to interactive entertainment platforms. The effectiveness of the proposed model is verified by a comprehensive dataset consisting of text labeled with various emotions that are used to train and assess the suggested emotion detection algorithm. Quantitative results demonstrate that the Logistic Regression -based proposed model outperforms competing methods by accurately identifying and classifying emotions in textual data. The best suited model in all the ML algorithms and Deep Learning is Logistic Regression i.e 63%

Keywords: *Sentiment Analysis, NLTK, GANs, CNN, Emotions, Text data, Classification Models, Naïve Bayes, Logistic Regression, Machine Learning, FNN, RNN.*

1 Introduction

You will find In the current world Text-based emotion recognition is important due to the huge volume of digital communication, like in need for effective customer feedback, computer- human interaction, and in media. social media platform is the place where many people express their emotions publicly, analyzing these emotions in social media posts allows researchers, businesses and to respond to the issues effectively. Many organizations and governments can benefit from text-based emotion recognition, for analyzing public opinion on various issues, mainly for politics, social trends, and public policies. Emotion recognition helps to develop many personalized services and applications, like chatbots or virtual assistants they can understand us respond to users and create more natural and engaging interaction.

Sentiment analysis is also called opinion mining, it is a natural language processing (NLP) technique that determines and analyses the sentiment, or the emotion expressed in the text. the main objective of this sentiment analysis is to identify and categorize the sentiment given by the author whether the the given text is positive, negative, or neutral. Frequently used to gather information on public opinion and sentiments from user reviews, social media posts, customer feedback, and other textual communication formats. The main goal Is to not only identify if the sentiment is negative, neutral, or positive but also identify more specific features such as expressions of joy, sadness, fear, anger, and other emotions through the use of computer algorithms.

Pictures these days are like emotionally charged messages in the digital era. Images play a major role in conveying emotions, whether they are used in advertisements or social media posts. To implement this work, it is important to investigate text-Based Emotion Recognition, with a particular focus on applying sentiment analysis and machine learning techniques. The aim is to simplify the emotions concealed in pictures so that we may better comprehend the visual language of emotions. Consider all the images we view on the internet; they resemble a vast assemblage of feelings just begging to be deciphered. We aim to apply sentiment analysis to images in the same way that it is used to analyze text for emotions. This research explores machine learning for text-based emotion recognition. We want to delve further into the more nuanced emotions, such as happiness, sadness, rage, and more, rather than just identifying the fundamental ones like happy or sad. As we proceed, we will review the findings of other knowledgeable researchers, examine the approaches they have employed, and propose our approach to machine learning for image- based emotion recognition. The purpose of this post is to provide some insightful information about how machine learning, in particular sentiment analysis, might aid in our understanding of the emotions depicted in images. Let's investigate how machine learning for emotion recognition in photographs can be applied in practical scenarios. Applications on image-based sentiment analysis: Monitoring of Driver Emotion in the Automotive Industry, Entertainment, Healthcare, Determine Pain Level Advertising & Marketing.

We discussed the system's broad structure and key components in Section 3 along with our planning and design process. After that, we covered the techniques and technologies we employed in Section 4 to go into more detail about how we implemented the system. Section 5 presents a detailed analysis of the obtained results and provides insights into the system's performance. A summary of our findings, their significance, and potential next steps are provided in Section 6, which also serves as a conclusion. In Section 2, we referred papers deeply into the work of other scholars in this field and situating our findings within the body of current knowledge.

1. Related Work

In recent research, a plethora of studies have delved into text-based emotion recognition using machine learning algorithms, with applications ranging from mental health tracking to social media analysis. Midhan et al.[1] address the concern of depression by evaluating ML algorithms for emotion classification in English text, emphasizing implications for mental health tracking systems. Chavan et al.[2] conduct a survey exploring the application of machine learning and deep learning in emotion detection from text, focusing on challenges in accurately depicting human emotions without audio or facial features. Their insights guide future research efforts, aiming to enhance emotion classification for text documents.Several studies, such as Jain et al.[3] and Tleubayeva et al.[4], leverage advanced models like BERT and Roberta for precise emotion detection in various contexts, including social media and surveys. Lathish et al.[6] compare the performance of SVM and KNN in text classification for emotion recognition, revealing SVM's superior accuracy. Julian et al.[7] specifically focus on emotion detection in tweets, finding Support Vector Machine to be particularly effective. Mahima et al.[8] introduce a hybrid approach for multiple emotion detection, outperforming traditional sentiment analysis. Furthermore, Andry et al.[9] achieve optimal results in text-based emotion recognition on social media using diverse machine learning techniques. These studies collectively contribute to the growing field of text-based emotion recognition, showcasing the versatility and efficacy of machine learning in understanding and categorizing emotions from textual data.

In a diverse array of studies exploring text-based emotion detection, Cecilia et al. [10] employ supervised machine learning with the SNoW architecture to predict emotions from children's fairy tales, aiming for expressive text-to- speech synthesis. Rajkumar et al.[11] delve into emotion detection from Twitter data, utilizing supervised learning techniques like K-means, Naive Bayes, and SVM, transforming emotions into an eight-category vector. E. C. C. Kao et al.[12] conduct a comprehensive survey categorizing text-based emotion detection methods into keyword-based, learning-based, and hybrid recommendation approaches, proposing solutions for practical enhancement in human-computer interactions. Zahid et al.[13] focus on identifying dominant emotions in email text using machine learning, achieving an average accuracy of 83% with three classifiers and feature selection methods. Additionally, Alaa et al.[14] review machine learning techniques for emotion detection and sentiment analysis, highlighting SVM and Naïve Bayes as popular algorithms. Lastly, Fatma et al.[15] present a novel system, PERS, integrating social media and machine learning for personality and emotion recognition, achieving high accuracy in personality detection and emotion recognition, with applications including identifying individuals at risk based on social media content. These studies collectively contribute to the evolving landscape of text-based emotion recognition, showcasing advancements in diverse domains and methodologies. These studies aim to

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

advance the subject of recognizing emotions from text by showcasing advancements in multiple domains and using a variety of approaches..

2 System Design

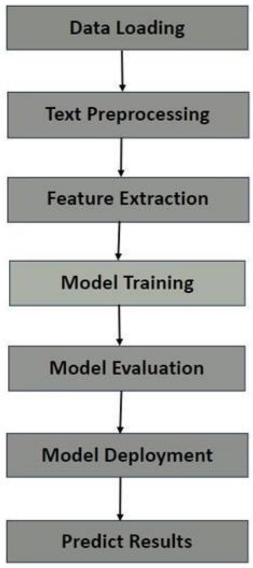

Fig. 1. Block Diagram

The Fig 1 shows the block diagram of the proposed system design.

Data Loading: The data was loaded and explored using pandas. We examined the shape, data types, and distribution of emotions in the dataset.

Text Preprocessing: Applied various text preprocessing techniques, such as stop word removal, punctuation removal, and extraction of the most common words for each emotion.

Feature Extraction: Used CountVectorizer to convert the cleaned text into numerical features that can be used for machine learning.

Model Training: Trained different machine learning models, including Naive Bayes, Logistic Regression, and Support Vector Machines on the training dataset.

Model Evaluation: Evaluated the models using accuracy, confusion matrix, and classification report. Interpretation of the model was performed using eli5.

Model Deployment: The final trained model was saved for future use.

Predict Results: The naive Bayes model achieved an accuracy of 0.56, whereas the Logistic Regression model achieved a better accuracy of 0.63.

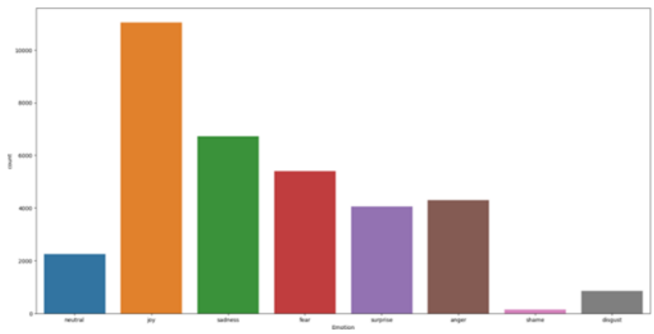

Fig. 2. Probability of success of different emotions

3 Implementation

Implementation starts by gathering this data from a MongoDB database, assuming the presence of information related to both the text itself and the associated emotions. The script then gets the data ready for analysis by converting it into a structured format using Pandas. Next, it cleans up the text by making it all lowercase and removing any characters that aren't letters or numbers. This ensures a uniform format for the text data. The dataset is then divided into two parts: one for training the model and the other for testing its accuracy. To understand the text better, the script converts it into numerical form using a technique called Count Vectorization. This essentially transforms words into numbers that a machine learning model can understand. Instead of using Naïve bayes to understand the text, it switches to another technique called Logistic Regression. This change might affect how accurately the model interprets emotions from the text. Finally, after training the model on the training data, it evaluates the model's performance using the testing data. It calculates how accurately the model predicts emotions in the text and provides a detailed report summarizing its effectiveness. Coming to deep learning models Each model underwent similar preprocessing steps to prepare the textual data and corresponding emotion labels. While the CNN focused on spatial patterns within the text, the FNN relied on traditional neural network layers to learn intricate relationships, and the RNN specifically targeted sequential data, retaining memory to capture contextual information. The Probability of success of different emotions is given in Fig 2.

4 Result

This study utilizes machine learning models and deep learning models to classify emotions in textual data using a dataset consisting of text categorized into eight different emotion labels: neutral, joy, sadness, fear, surprise, anger, shame, disguist The primary machine learning classifiers evaluated are the Multinomial Naive Bayes and Logistic Regression models. During testing, the Multinomial Naive Bayes classifier achieved 57% accuracy, whereas the Logistic Regression model achieved 63% accuracy. The primary deep learning models evaluated are CNN, RNN and FNN. During testing the CNN achieved 54.19% accuracy, FNN achieved 54.63% while RNN has achieved 56.03% accuracy. Those are depicted in following Fig 3 to Fig 7

The machine learning models have performed better than deep learning model and have attained more accuracy.The key aspect of this work revolves around the interpretability of the Logistic Regression model, it is highlighted through the presentation of feature weights. The visualization of the results provides valuable insights into the importance of every word in classifying the emotions from the texts.The dataset used in this work has a important role, containing annotated texts with one of the thirteen specified emotions. It serves as both the training and evaluation foundation for the emotion classification models.The results demonstrate the significance of the chosen classifiers and emphasize the interpretability of the Logistic Regression model. This offers a comprehensive overview of the performance in this emotion classification task.

MACHINE LEARNING

Accuracy: 0.5793935910331944

	precision	recall	f1-score	support
anger	0.65	0.50	0.57	836
disgust	0.83	0.02	0.05	202
fear	0.77	0.63	0.70	1104
joy	0.52	0.88	0.65	2214
neutral	0.70	0.03	0.06	481
sadness	0.57	0.56	0.56	1327
shame	0.00	0.00	0.00	23
surprise	0.69	0.28	0.39	772
accuracy			0.58	6959
macro avg	0.59	0.36	0.37	6959
weighted avg	0.62	0.58	0.54	6959

Fig. 3 Naïve Bayes

Accuracy: 0.6347176318436557

	precision	recall	f1-score	support
anger	0.62	0.57	0.59	836
disgust	0.62	0.18	0.28	202
fear	0.76	0.67	0.71	1104
joy	0.63	0.77	0.70	2214
neutral	0.61	0.72	0.66	481
sadness	0.60	0.58	0.59	1327
shame	0.80	0.70	0.74	23
surprise	0.55	0.42	0.47	772
accuracy			0.63	6959
macro avg	0.65	0.58	0.59	6959
weighted avg	0.63	0.63	0.63	6959

Fig. 4. Logistic Regression

DEEP LEARNING

```
Epoch 10/10
870/870 [==============================] - 16s 19ms/step
y: 0.5419
218/218 [==============================] - 2s 9ms/step -

Test Loss: 2.6000
Test Accuracy: 54.19%
```

Fig.5 CNN Model

```
y: 0.5450
Epoch 10/10
870/870 [==============================] -
y: 0.5463
218/218 [==============================] -

Test Loss: 3.0231
Test Accuracy: 54.63%
```

Fig.6 FNN Model

```
Epoch 10/10
870/870 [==============================] - 13s 15ms
y: 0.5603
218/218 [==============================] - 1s 5ms/s

Test Loss: 1.4181
Test Accuracy: 56.03%
```

Fig.7 RNN Model

5 Conclusion

The data set is classified into 8 categories so the accuracy will be a bit less than binary classification, when compared to all classifiers Logistic Regression model achieves an accuracy of 0.63%. In deep learning models the best accuracy is in RNN model which is very much lesser than the Machine learning model. As we explore the future of image analysis, the convergence of deep learning advancements, particularly in convolutional neural networks (CNNs), holds promise for more accurate and efficient visual data extraction. The integration of multiple modalities, such as text and image analysis, is a key trend, enhancing contextual understanding. Future achievements include the continual improvement of model interpretability, real-time analysis at the edge, and the responsible deployment of image analysis technologies. Expectations also include the evolution of Generative Adversarial Networks (GANs) for realistic image synthesis and increased automation in anomaly detection. Image analysis is poised to contribute significantly to environmental monitoring, sustainability efforts, and the seamless integration of augmented and virtual reality experiences. Considering image analysis in the future, a few predicted developments come to mind. First, improvements in multimodal analysis—which merges multiple sources, including text and images—are anticipated to enhance contextual knowledge. It is anticipated that interpretation and explanation will play a larger role in image analysis models going forward, addressing issues of trust and transparency. Real-time analysis at the edge, where models are installed closer to data sources to increase efficiency and decrease latency, is another expected development.

Generative Adversarial Networks (GANs) are poised to play a crucial role in realistic image synthesis, contributing to applications in various fields. The responsible and ethical use of image analysis technologies, including considerations for bias and privacy, is becoming increasingly important. Automation in anomaly detection, especially in domains like manufacturing and healthcare, is expected to advance, streamlining quality control processes. Lastly, image analysis is set to contribute significantly to environmental monitoring and sustainability efforts

6 References

[1] T. Madhu Midhan, P. Selvaraj, M. Harshavardan Kumar Raju., M. Bhanu Prakash Reddy. and T. Bhaskar., "Classification of Mental Health and Emotion of Human from Text using Machine Learning Approaches," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023.

[2] D. Chavan, E. Anvekar, M. Dandapat, V. Bichave and J. Jagdale, "Machine Learning Applied in Emotion Classification: A Survey on Dataset, Techniques, and Trends for Text Based Documents," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023.

[3] G. Jain, S. Verma, H. Gupta, S. Jindal, M. Rawat and K. Kumar, "Machine Learning Algorithm Based Emotion Detection System," 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2022.

[4] A. Tleubayeva, A. Shaikhanova, B. Ospan, A. Sultan, M. Abu and N. Darmenkyzy, "Machine Learning Expert System for Recognizing Emotions in text “Umai Cloud Services”," 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 2023.

[5] Saravanakumar, S., & Thangaraj, P. (2019). A computer aided diagnosis system for identifying Alzheimer’s from MRI scan using improved Adaboost. Journal of medical systems, 43(3), 76.

[6] Kumaresan, T., Saravanakumar, S., & Balamurugan, R. (2019). Visual and textual features based email spam classification using S-Cuckoo search and hybrid kernel support vector machine. Cluster Computing, 22(Suppl 1), 33-46.

[7] Saravanakumar, S., & Saravanan, T. (2023). Secure personal authentication in fog devices via multimodal rank-level fusion. Concurrency and Computation: Practice and Experience, 35(10), e7673.

[8] Thangavel, S., & Selvaraj, S. (2023). Machine Learning Model and Cuckoo Search in a modular system to identify Alzheimer’s disease from MRI scan images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 11(5), 1753-1761.

[9] Saravanakumar, S. (2020). Certain analysis of authentic user behavioral and opinion pattern mining using classification techniques. Solid State Technology, 63(6), 9220-9234.

[10] R.R. Malagi, Y.R,S. P. T. K, A. Kodipalli, T. Rao and R. B. R, "Emotion Detection from Textual Data Using Supervised Machine Learning Models," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023.

[11] A. Lathish and T. Devi, "Comparison of Support Vector Machine and K-Nearest Neighbour Algorithm for Accurate text Classification," 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2022.

[12] C. R. Julian, G. R. Athevan, N. G. Hanjaya, H. Lucky and D. Suhartono, "Utilizing Machine Learning Algorithms to Detect Emotions from Tweets," 2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), Tokoname, Japan, 2022.

[13] M. A. Mahima, N. C. Patel, S. Ravichandran, N. Aishwarya and S. Maradithaya, "A Text-Based Hybrid Approach for Multiple Emotion Detection Using Contextual and Semantic Analysis," 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2021.

[14] Chowanda, A., Sutoyo, R. and Tanachutiwat, S., "Exploring text- based emotions recognition machine learning techniques on social media conversation", Procedia Computer Science, 179, pp.821-828, 2021.

[15] Alm, C.O., Roth, D. and Sproat, R., "Emotions from text: machine learning for text-based emotion prediction. In Proceedings of human language technology conference and conference on empirical methods in natural language processing pp. 579-586, 2005.

[16] Salam, S.A. and Gupta, R., "Emotion detection and recognition from text using machine learning", Int. J. Comput. Sci. Eng., Vol. 6, No. 6, pp.341-345, 2018.

[17] Kao, E.C.C., Liu, C.C., Yang, T.H., Hsieh, C.T. and Soo, V.W., "Towards text-based emotion detection a survey and possible improvements", 2009 International conference on information management and engineering, pp. 70-74, 2009.

[18] Halim, Z., Waqar, M. and Tahir, M., "A machine learning-based investigation utilizing the in-text features for the identification of dominant emotion in an email", Knowledge-based systems, Vol. 208, 2020.

[19] Alslaity, A. and Orji, R., "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions", Behaviour & Information Technology, pp.1-26, 2022.

[20] Talaat, F.M., El-Gendy, E.M., Saafan, M.M. and Gamel, S.A., "Utilizing social media and machine learning for personality and emotion recognition using PERS", Neural Computing and Applications, Vol. 35, No. 33, pp.23927-23941, 2023.