JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Sentiment analysis neutral and happy using deep learning

Sentiment analysis using multimodal text and speech data aims to classify emotions like happy and neutral with machine learning for applications in mental health, smart assistants, and human-robot interaction.

[1]Neseebha,[2]Jithendra P.K Nayak
[1]M. Tech Student,[2]Associate Professor
[1] Computer Science and Engineering,
[1]Srinivas Institute of Technology, Valachil, Mangaluru,Karnataka, India

Abstract: Among the most crucial areas of research in human-computer interaction is sentiment analysis using multimodal data, including text and speech. The goal of this endeavour is to create an effective sentiment recognition system utilising methods from machine learning to categorise emotions from text and spoken audio input, especially happy and neutrality. The system can efficiently analyse and forecast emotional states by fusing real-time speech-to-text conversion with Natural Language Processing (NLP) methods including TF-IDF and Naive Bayes classifiers. Entails sing a publicly accessible emotion-labeled dataset for extraction of characteristics, model training, data preprocessing, and evaluation. Experimental results demonstrate satisfactory accuracy, highlighting the potential of integrating speech and text modalities for improved emotion recognition. This system can be used in various applications include mental health monitoring, smart assistants, and human-robot interaction.

***IndexTerms* - Sentiment analysis, Emotion recognition, Multimodal data, Speech emotion recognition, Text emotion recognition, Speech-to-text conversion, Natural Language Processing (NLP), Machine learning, Naive Bayes classifier, TF-IDF, Feature extraction, Model training, Data preprocessing, Happy and neutral emotion classification.**

I.INTRODUCTION

In today's digital era, sentiment analysis plays an important part of understanding human emotions through textual or spoken input. The purpose of this endeavour is to create a machine learning-based sentiment classification technique that is capable of discern between happy and neutral sentiments from voice and text inputs.

A labelled dataset of social media texts with various kinds of emotional expressions is being used to provide guidance the system. It leverages Natural Language Processing (NLP) techniques such as Count Vectorization, TF-IDF transformation, and a Naive Bayes classifier to learn and predict sentiment. The pipeline is built using Python’s Scikit-learn library, and it is capable of accepting real-time audio input through speech recognition.It leverages Natural Language Processing (NLP) techniques such as Count Vectorization, TF-IDF transformation, and a Naive Bayes classifier to learn and predict sentiment. The pipeline is built using Python’s Scikit-learn library, and it is capable of accepting real-time audio input through speech recognition. Sentiment Analysis, also known as opinion mining, is a branch of NLP, that deals with Finding and removing subjective information from the source material. It is widely used to analyze opinions, emotions, and attitudes expressed in text and speech. This project presents a practical implementation of sentiment analysis to detect and classify two specific emotional states: neutral,sad, angry and happiness.

The training data is taken from a pre-labeled dataset (tweet_emotions.csv), which contains tweets labeled with various emotions. The dataset is preprocessed and filtered to include only the neutral and happiness sentiments. This filtered dataset is split into training and testing sets to build and evaluate the model. The system processes both text input and live speech. Text is vectorized using CountVectorizer, transformed with TF-IDF, and classified using the Multinomial Naive Bayes algorithm—a common model for text classification due to its performance and simplicity. The speech input is captured using the speech_recognition library, transcribed into text using Google's speech-to-text API, and then analyzed using the trained model.

Speech is converted to text passed through the model for detecting sentiment. The classification of confusion matrix are used for evaluating the model’s performance. Only two sentiments are targeted to simplify training and enhance classification accuracy. The model is saved using Job lib for easy reuse in real-time applications. Visualization is done using Seaborn heatmaps for clear presentation of classification results. Applications include mental health monitoring, AI chatbots, and feedback analysis systems. The project highlights the importance of emotion detection in human-computer interaction. Detecting happiness can be useful in wellness apps and personalized assistants. Identifying neutrality helps determine emotional stability or disinterest. This focused classification approach allows better precision in binary emotion scenarios. The collection of data is preprocessed to enhancethe model and reduce noise. Text and voice integration demonstrates practical multi-modal sentiment analysis. This solution is lightweight, efficient, and deployable in real-world applications. The trained model is saved using Joblib, allowing it to be reused for real-time prediction without retraining. The efficiency of the approach is measured using a classification report and a confusion matrix visualized with Seaborn. The project is implemented using Python and libraries such as Scikit-learn, Pandas, Matplotlib, and Seaborn. It follows a modular pipeline design, making it easy to understand, maintain, and extend

II. LITERATURE REVIEW

Traditional sentiment analysis has largely focused on classifying text into categories such as neutral, negative, and positive. Early work by Pang et al. (2002) used algorithms like Naive Bayes sentiment classification of movie reviews. More recent work by Liu (2012) and others introduced lexicon-based and deep learning methods that consider semantic and syntactic structures for sentiment prediction. Speech-based emotion recognition has gained attention with the rise of virtual assistants and affective computing. Studies such as Schuller et al. (2003) and Eyben et al. (2013) explored acoustic features like pitch, energy, and spectral components to classify speech emotions. However, the integration text-based NLP models offers a simpler alternative by transcribing speech into text for further analysis.Recent research has emphasized the importance of real-time emotion detection using hybrid pipelines. Projects like DeepMoji (Felbo et al., 2017) and Google's Natural Conversations API show how integrating speech, text, and emotion can lead to more responsive AI systems. However, most production-level systems are still limited to text, primarily due to the noise and variability in live speech data.

III. PROPOSED METHODOLOGY

Implementing emotions with deep learning involves several steps, including signal processing, Feature extraction, learning environment and speech recognition. Below is a detailed explanation of how to implement the model: The below A Flow chart of the proposed emotion recognition is:

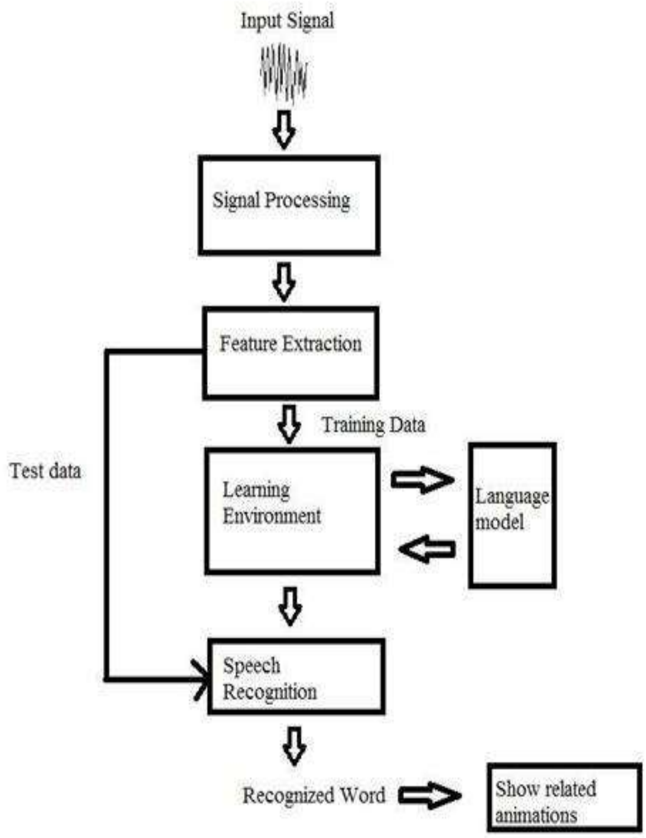

Figure 1: A flow chart of the proposed emotion recognition

The proposed user input text or speech, with the system detecting emotions in real-time. It provides instant feedback, including emotion classification and insights, in an intuitive and responsive design.

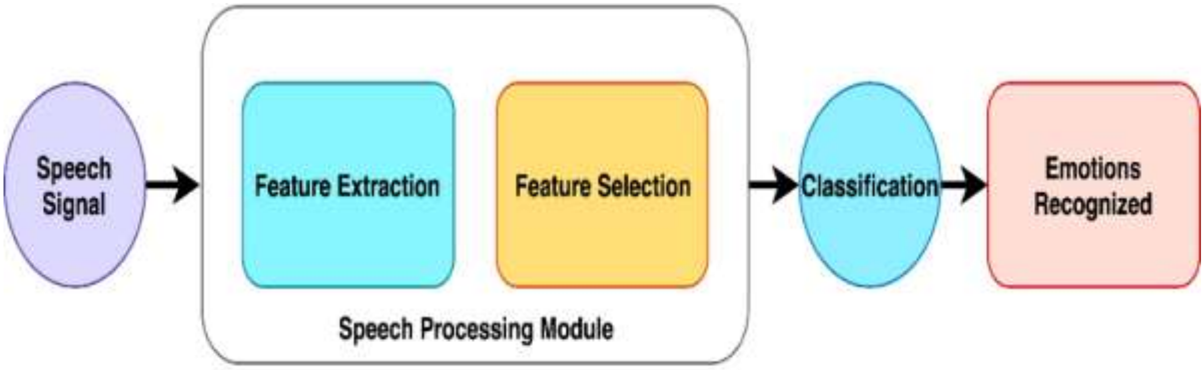

Figure 2: emotion detection system.

VI. RESULT AND DISCUSSION

The developed emotion recognition system effectively classifies emotions from speech inputs into categories such as neutral and happiness. The system integrates speech-to-text conversion, followed by emotion classification using a Naive Bayes classifier based on TF-IDF retrieved from the transcribed text. The model wastrained using a filtered dataset derived from the Tweet Emotion Dataset, containing only entries labeled as "neutral" and "happiness." After preprocessing and training, the model reached an accuracy.

We want to raise the quantity of emotion classes, use deep learning models to increased accuracy, and investigate multimodal fusion for more intricate emotion detection. The system is lightweight, quick, and appropriate for integration into real-time applications. It recognises emotions from speech and text data by combining text-based features and Naive Bayes classifier algorithms.

Figure 1: sentiment anyalysis

This command-line structure allows users to interact easily with your sentiment analysis tool, providing options for both text and audio input.

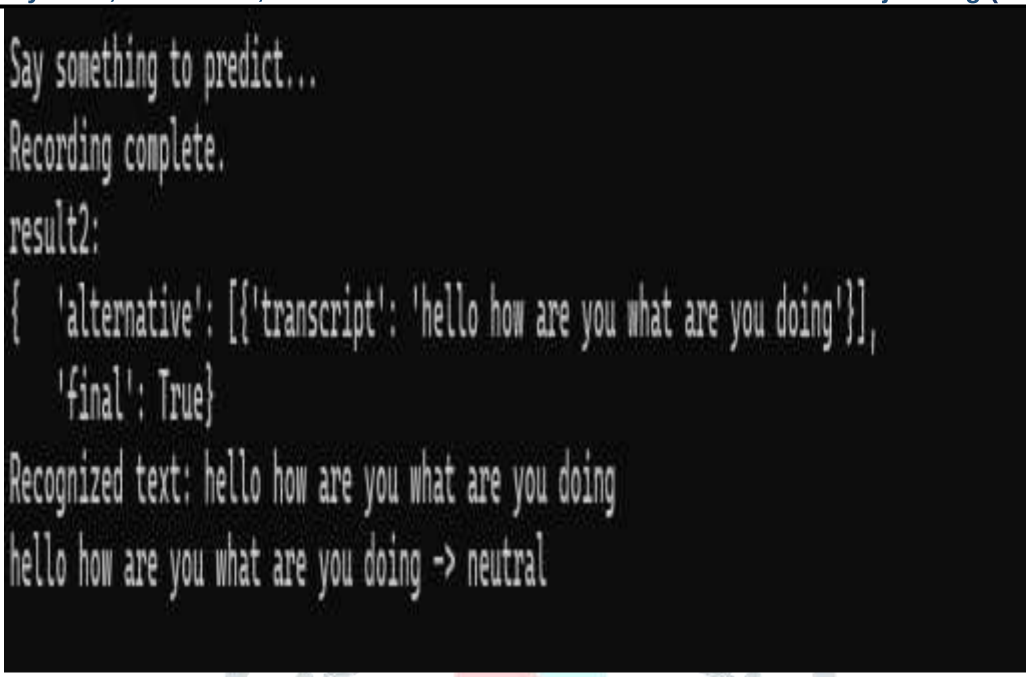

Figure 2: Emotion classified as neutral

The speech input is transformed into text, and the recognised statement and the anticipated emotion displayed.

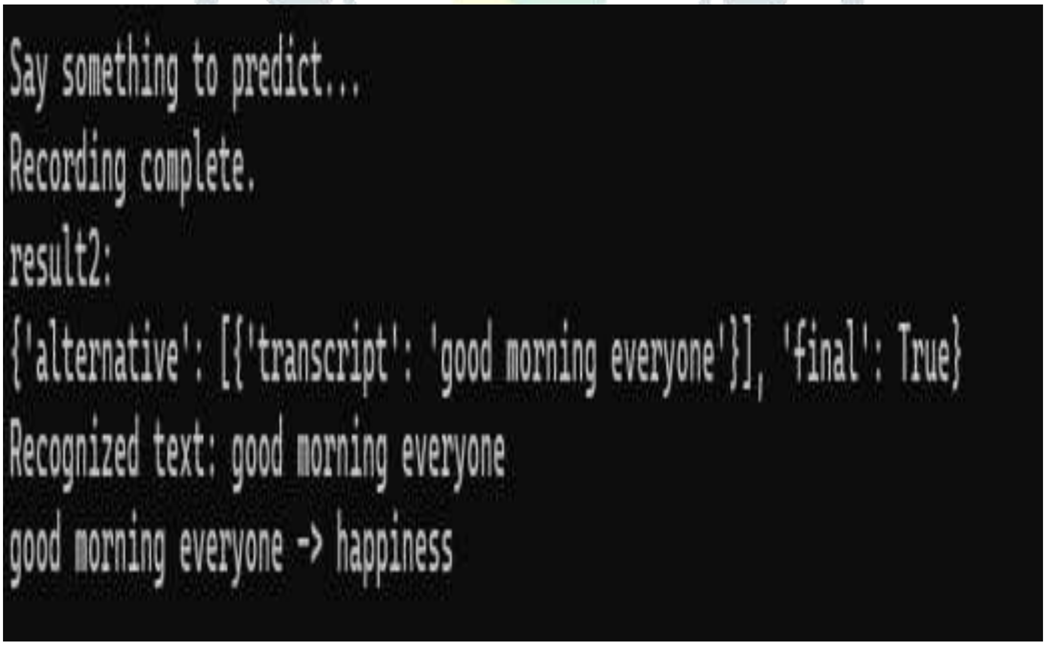

Figure 3: The emotion classified as happiness

The system successfully recognizes the spoken input "good morning everyone" and classifies the emotion as happiness, showcasing the effectiveness of real-time speech-based emotion detection through a text-based user interface

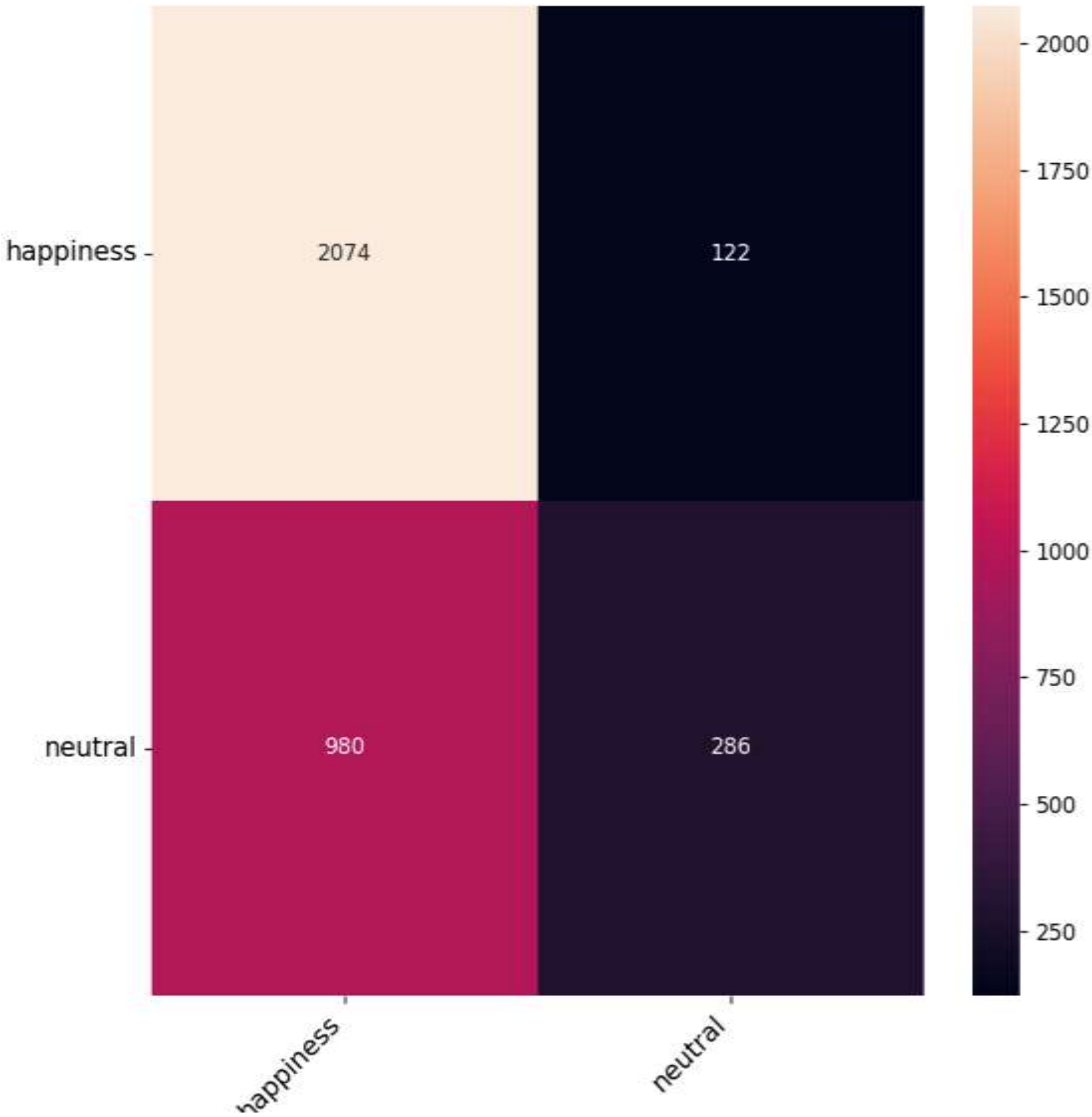

Figure 4: confusion matrix

The confusion matrix illustrates the model's performance in classifying 'neutral' and 'happiness' emotions, showing high accuracy with minimal misclassification between the two classes.

A classification problem's prediction outcomes are summarised in a confusion matrix. It enables you to compute several performance indicators by displaying the counts among false positives, false negatives, true positives, and true negatives.

VII. Conclusion

In this instance, we have developed and implemented a text and speech-based emotion recognition system capable of detecting emotions such as neutral and happiness from user input. By leveraging a Naive Bayes classifier trained on TF-IDF features extracted from the transcribed speech, the system efficiently classifies emotions from spoken sentences. The incorporation of speech recognition and processing of natural language enables the system to analyze and interpret user sentiments in real-time.

Our results demonstrate that the system performs reliably with high accuracy on the test data, with minimal misclassification between the selected emotion classes. The use of a filtered dataset helped in optimizing the classifier for more focused emotion categories. The system's text-based interface was tested with multiple speech inputs and produced expected and relevant emotion outputs, validating the level of accuracy and tenacity of the pipeline.

This project showcases the potential of combining machine learning, audio signal processing, and language modeling for emotion detection applications. It opens opportunities for integrating emotional awareness into a variety of systems, such as smart assistants, virtual customer support, and educational tools.

ACKNOWLEDGMENT

I sincerely thank Dr. Jithendra P.K. Nayak, AssociateProfessor, Department of Computer Science and Engineering, for his expert guidance, mentorship, and constant support during this research. I also extend my gratitude to the faculty of Srinivas Institute of Technology for their valuable feedback and motivation. I am deeply thankful to my family for their encouragement and patience throughout this journey.Sentiment analysis using multimodal text and speech data aims to classify emotions like happy and neutral with machine learning for applications in mental health, smart assistants, and human-robot interaction.. Finally, I appreciate the department for providing a supportive academic environment and access to essential resources for completing this stud

REFERENCES

[1] Zhao, Z., et al. (2023). Hierarchical Network with Decoupled Knowledge Distillation for Speech Emotion Recognition. *ICASSP 2023*. DOI: 10.1109/ICASSP49357.2023.10094889

[2] Xu, X., et al. (2023). Zero-Shot Speech Emotion Recognition Using Generative Learning with Reconstructed Prototypes. *ICASSP 2023*. DOI: 10.1109/ICASSP49357.2023.10094890

[3] Zou, H., et al. (2022). Speech Emotion Recognition with Co-Attention Based Multi-Level Acoustic Information. *ICASSP 2022*. DOI: 10.1109/ICASSP43922.2022.9746723

[4] Wu, X., et al. (2022). Neural Architecture Search for Speech Emotion Recognition. *ICASSP 2022*. DOI: 10.1109/ICASSP43922.2022.9746724

[5] Zhu, W., & Li, X. (2022). Speech Emotion Recognition with Global-Aware Fusion on Multi-Scale Feature Representation. *ICASSP 2022*. DOI: 10.1109/ICASSP43922.2022.9746725

[6] Sun, L., et al. (2021). Multimodal Cross- and Self-Attention Network for Speech Emotion Recognition. *ICASSP 2021*. DOI: 10.1109/ICASSP39728.2021.9413699

[7] Wang, X., et al. (2021). A Novel End-To-End Speech Emotion Recognition Network with Stacked Transformer Layers. *ICASSP 2021*. DOI: 10.1109/ICASSP39728.2021.9413700

[8] Xu, M., et al. (2021). Speech Emotion Recognition with Multiscale Area Attention and Data Augmentation. *ICASSP 2021*. DOI: 10.1109/ICASSP39728.2021.9413701

[9] Wang, X., et al. (2021). A Novel End-To-End Speech Emotion Recognition Network with Stacked Transformer Layers. *ICASSP 2021*. DOI: 10.1109/ICASSP39728.2021.9413700

[10] Xu, M., et al. (2021). Speech Emotion Recognition with Multiscale Area Attention and Data Augmentation. *ICASSP 2021*. DOI: 10.1109/ICASSP39728.2021.9413701

[11] Peng, Z., et al. (2021). Efficient Speech Emotion Recognition Using Multi-Scale CNN and Attention. *arXiv preprint arXiv:2106.04133*, 2021.

[12] Xu, M., et al. (2021). Speech Emotion Recognition with Multiscale Area Attention and Data Augmentation. *arXiv preprint arXiv:2102.01813*, 2021.

[13] Muppidi, A., & Radfar, M. (2021). Speech Emotion Recognition Using Quaternion Convolutional Neural Networks. *arXiv preprint arXiv:2111.00404*, 2021.

[14] Aftab, A., et al. (2021). Light-SERNet: A Lightweight Fully Convolutional Neural Network for Speech Emotion Recognition. *arXiv preprint arXiv:2110.03435*, 2021.

[15] Li, M., et al. (2021). Contrastive Unsupervised Learning for Speech Emotion Recognition. *arXiv preprint arXiv:2102.06357*, 2021.

[16] Xu, X., et al. (2021). Multi-Loss Auxiliary Learning for Speech Emotion Recognition. *ICASSP 2021*. DOI: 10.1109/ICASSP39728.2021.9413698

[17] Schuller, B., et al. (2021). The INTERSPEECH 2021 Computational Paralinguistics Challenge. *INTERSPEECH 2021*. DOI: 10.21437/Interspeech.2021-140

[18] Yang, S., et al. (2021). Progressive Multi-Granularity Training for Speech Emotion Recognition. *ICASSP 2021*. DOI: 10.1109/ICASSP39728.2021.9413703

[19] Li, M., et al. (2021). Contrastive Unsupervised Learning for Speech Emotion Recognition. *ICASSP 2021*. DOI: 10.1109/ICASSP39728.2021.9413702

[20] Sun, L., et al. (2021). Multimodal Cross- and Self-Attention Network for Speech Emotion Recognition. *ICASSP 2021*. DOI: 10.1109/ICASSP39728.2021.9413699