Sentimental analysis of audio based customer reviews without textual conversion

Sumana Maradithaya[1], Anantshesh Katti[2]
[1]Department of Information Science and Engineering, Ramaiah Institute of Technology, Bengaluru, India
[2]Continental AG, Bengaluru, India

Article Info

Article history:

Received May 23, 2023
Revised Jul 11, 2023
Accepted Jul 17, 2023

Keywords:

Deep neural network
Feature extraction
Mel frequency cepstral coefficients
Natural language processing
Sentiment analysis
Speech emotion

ABSTRACT

The current trends or procedures followed in the customer relation management system (CRM) are based on reviews, mails, and other textual data, gathered in the form of feedback from the customers. Sentiment analysis algorithms are deployed in order to gain polarity results, which can be used to improve customer services. But with evolving technologies, lately reviews or feedbacks are being dominated by audio data. As per literature, the audio contents are being translated to text and sentiments are analyzed using natural processing language techniques. However, these approaches can be time consuming. The proposed work focuses on analyzing the sentiments on the audio data itself without any textual conversion. The basic sentiment analysis polarities are mostly termed as positive, negative, and natural. But the focus is to make use of basic emotions as the base of deciding the polarity. The proposed model uses deep neural network and features such as Mel frequency cepstral coefficients (MFCC), Chroma and Mel Spectrogram on audio-based reviews.

Corresponding Author:

Sumana Maradithaya
Department of Information Science and Engineering, Ramaiah Institute of Technology
Bengaluru, India
Email: sumana.a.a@gmail.com

1. INTRODUCTION

Customer relation management system (CRM) is a system used by service providing companies to gather reviews or feedbacks from their customers based on the products they bought. These reviews are in reference to how the products or services were, How or what can be done in order to make the services or products quality improve. These reviews or feedbacks collected are in terms of e-mails, messages or textual form. The CRMs collect these reviews and feedbacks and they deploy sentiment analysis algorithms on them. The algorithm is trained with some pre-defined labeled data by which the model learns. The traditional sentiment analysis gives results in terms of three output namely, positive, negative and neutral known as polarities. Thus, the collected feedback is fed through the trained model and the results are produced. Using these results, the CRM informs the companies, the percentage of positive or negative or neutral polarities of the collected feedbacks and based on the results, the company can decide on which directions or steps to take in order to satisfy their customers best.

The sentiment analysis models are basically trained on millions of labeled text data where every sentence is broken down to individual words using different pre-processing techniques. The basic analogy of sentiment analysis can be given as, “I had a very pleasant day, today”. In this sentence, the word ‘pleasant’ is a positive word and when we consider the combination of it with its previous word which is ‘very’, it becomes a strong positive. Another example can be as, “The ocean is scary but without a doubt, beautiful”. In this example, ‘scary’ is a negative word but there is also the word ‘beautiful’ which is a positive. Such a combination of negative and positive words can lead to any of the three outcomes when training the model. In this statement, if only the first part is considered, then no doubt it is negative sentence. But if the entire sentence is considered, the other half dominates and thus it can be termed as either positive or neutral sentence.

In the traditional approach to calculating the sentiment where textual data is being used, the customers can express their concerns about the service or product only in terms of text. But with this, sometimes the customers can easily lie in text feedbacks. Sometimes they can be sarcastic where they mean one thing, whereas their words can mean a different thing. This can cause a problem for the companies and CRMs to provide with the best outcomes. In order to avoid this gap, audio feedbacks can be used as input data. But converting the audio data to text can lead to the same problem, so making use of audio data directly can result in better outcomes. In this paper, the use of audio data directly to train the model as well as to predict the results is done. Instead of just having positive, negative and neutral, this paper provides output in terms of neutral, calm, happy, sad, angry, surprised, disgust. A labeled Ryerson audio-visual database of emotional speech and song (RAVDESS) dataset is used to train the model.

Speech emotion is a key when it comes to paralinguistic element for communication, which without a doubt has high subjectivity level. Sentimental varies among different languages and people [1] as the existing work is focusing on investigating emotional states. The research has created an “emotional speech ground truth database”, which consists of semantically and or emotionally “loaded” utterances speakers with five sentiments polarity. Kim and Hansen [2] proposes “an effective angry speech detection approach” by considering structure of the content of the input speech. A classifier which is focused on “emotional language” model for which the score is formed and combined with acoustic feature that includes Teager energy operator (TEO) based feature and Mel frequency cepstral co-efficients (MFCC). The proposed research has a 6.23% improvement in equal error rate (EER) which is gained by combining the TEO-based and MFCC features. A brief comparison between conversation sentiment analysis and single sentence sentiment analysis has been done in [3]. It introduces a model called bidirectional emotional recurrent unit (BiERU) for conservation sentiment analysis. This approach is focused on textual data and not speech or audio data. The BiERU uses a general neural tensor model with 2 channel classifiers to analysis the context of the conversation and perform the sentiment analysis. This model was compared with existing benchmark models like dialogue-recurrent neural networks (dialogue-RNN), dialogue-convolutional neural networks (dialogue-CNN) and attention gated hierarchical memory network (AGHMN) and the results in the proposed research outperform the existing models in terms of accuracy between various sentiments. The accuracy of BiERU was found out to be a claimed average accuracy of 66% and F1 score of 64% for textual conservational data. Garg and Sharma [4] discusses a process of sentiment analysis using different representations on twitter data. In [5], authors have worked on a popular word embedding approaches in identifying sentiments. Salehin *et al.* [6] discusses the use of support vector machine (SVM) and OpenCV in analysing student sentiments in an online class. Further, Moung *et al.* [7] have used an ensemble of methods in identifying sentiments in images. Novel approaches are required in finding sentiments from audio data. Mahima *et al.* [8] discusses approaches in identifying multiple emotions from textual data.

Bertero and Fung [9] proposes a real-time CNN approach for speech emotion detection. It is focused on training with audio using a dataset from “TED talks”, which are then elucidated manually into three emotions: “angry”, “happy” and “sad”. The research has achieved an accuracy averaging at 66.1%, which is 5% better than a feature based SVM baseline. Kantipud and Kumar [10] have proposed an interesting computationally efficient learning model to classify audio signal attributes. Each filter is activated at multiple frequencies, which is caused due to the amplitude-related feature learning. The proposed approach in Chen and Luo [11] uses audio-based inputs by introducing utterance based deep neural network. This approach uses a combination of convolutional neural networks (CNN) and audio sentiment vector (ASV). The process is based on using utterance from the audio inputs to analysis the content and finds the spectrum graph generated from the signals which are inputted to a long short-term memory (LSTM) model branch, making use of spectral centroid, MFCC which are the traditional acoustic features. The proposed model has a claimed accuracy of 57.74%. Maghilnan and Kumar [12] proposed a model has been implemented using the audio sentiment analysis as speaker discriminated speech data. The model has two starting points, one for speaker discrimination and the other for speech recognition. In [13] the discussion is based on the sentiment analysis for the customer relation management where the idea of have a customer loyal for a longer run after their first purchase is very important for a company and their products. The proposed concept in [14] is to use multi-aspect level sentiment analysis (MALSA) model in a CRM which not only helps find the sentiment but also has a recommendation approach to recommend the customers throughout the purchase cycle. The model was built on four discussions namely, Frequency based detection, Syntax based detection, supervised and unsupervised learning and finally hybrid models for analyzing the sentiments. This model is entirely based on textual data to find out the sentiments. Rotovei and Negru [15] is an extended study of [13] but with a slight addition to the existing model where a component of B2B with recommendation is proposed. They have used aspect term extraction that has four approaches like, finding frequent nouns and noun phrases, use of opinion and target relationships, supervised learning and topic modeling. And secondly uses aspect aggregation. In a natural language sentence, there are lists of aspect that are produced from aspect based sentiment analysis [16]. The proposed research is based on interactive multi-task learning (IMU) that is implemented for tokens as well as documents which are done simultaneously. A whole new algorithm has been proposed to train the IMU and the model was compared with different other models and the proposed model outperforms the existing models with a claimed accuracy of 83.89 and F1 score of 59.18.

The identification of sentiment polarity for specific targets in a context can be achieved by using aspect-level sentiment classification [17]. The proposed research introduces an approach of using targets as well as context where both are important and as well need special treatment which learns their representation using the proposed interactive attention network (IAN). It can represent its collective context as well as targets. The proposed model was compared with other existing models such as LSTM, temporal dependence-based long short-term memory (TD-LSTM), autoencoder based LSTM (AE-LSTM) where IAN outperforms these models by 4% to 5% with a claimed final accuracy of 72.1%. The dataset used in [18] is SemEval 2014 dataset. In natural language processing, the fundamental task is aspect-level sentiment analysis [19] and the main aim is to predict the polarity of sentiment of a given aspect term of a sentence. This research is similar to that of research in [17] but with an improved and additional feature which addresses the drawback. In [19] the research considers grammatical rules in an input sentence for sentiment analysis which was missing in previous researches. The dataset used is the SemEval dataset and the proposed approach has a 2% more better performing than the approach in [17] with a claimed accuracy of 74.89%. This proposed model was compared with the other existing models such as LSTM, AE-LSTM, attention-based LSTM with aspect embedding (ATAE-LSTM).

Sentiment analysis is not only limited to finding the sentiment straight forward from a textual sentence but can also be used to find the lost-won classification of complex deals [14]. The model is built upon using “Frequency, lexicon based and syntax-based detection”, “machine learning based approach” and “hybrid methods”. The concept of using deep learning for the prediction of sentiment analysis from a given input, be it a textual input data or a speech/audio input data. The same deep learning can be used for finding sentiments from a given input data has been mentioned and extensive research has been carried out in [1]. The research introduces all the possible approaches such as deep neural network, LSTM, auto encoders, word embedding, CNN, RNN, attention mechanism with recurrent neural network document level sentiment analysis and other few approaches. Zhang *et al.* [20] is an extensive research study of all the deep learning approaches that can be used for sentiment analysis with all the possible features, characteristics and related implementation approaches have been discussed. Sentiment analysis plays a crucial part in identifying customer gratification and opinion. Capuano *et al.* [21] proposes a “hierarchical attention networks” for analyzing the sentiment precedence of client’s feedbacks. The model can be seen for improvement using the reviews provided by CRM which is possible by an “integrated incremental learning mechanism”. Capuano *et al.* [21] also focuses on a prototype that has been developed and the dataset used for training with over large number of annotated items has been used. The accuracy in the proposed research averaged 0.85 for the F1-score.

“Interactive sentiment analysis” in [22] is an emerging and a sub branch of the natural language processing (NLP) problem. The implementation of new approaches is limited to the labeled datasets for interactive sentiment. In [22], a new conversational database was created-Scenario SA. The dataset was manually labeled. The increase in the use of smart mobile phones has enabled everyone to connect through social media where chats, comments on posts and products can be seen. The sentiment analysis product reviews entirely depend on lexicons. The generation of lexicons is a crucial thing that needs to be considered. In [23], “automatic approach for constructing a domain-specific sentiment lexicon” has been proposed in view of words consisting of sentiments and product feedbacks. The process chooses words consisting of sentiments from reviews and focuses on the relation that uses common data algorithm. High-resolution or HR information from social media Facebook or Quora presents great chance to CRM by analyzing arguments about commerce events [24]. Textual sentiment analysis has been greatly developed and researched by using mainly lexicon-based approaches. In [25] the research focuses on analyzing behavior of the customer for buying a product using the product feedback. The proposed work introduced in [26] is of “multi-mixed short text ridge analysis (MMSTR)” that is used to extract sentence and text feedbacks. The emotions are measured using the reviews. The sentiments that are expressed by the customers are quite important; and hence a swift analysis is a must. In [27], a “hierarchical approach” is put forward for sentiment analysis. Firstly, word embeddings of reviews are analyzed by using Word2Vec. The research also considers binate sentiment analysis, i.e., resolving of “positive” and “negative” as sentiments, an “extreme gradient boosting classifier” (xgboost) is used and on an average feedback vector. An overall accuracy with 71.16% is gained that uses categorization of 12 different classes that makes use of the Doc2Vec approach. Online shopping is growing day by day with popularity among customers, with these product feedbacks and reviews are received through customers [28]. Online analytical processing (OLAP) and Data Cubes techniques of data warehouse are used to analyze the sentences.

The main challenges natural language processing are sentiment analysis and opinion mining [29]. The work is based on methods used to identify text by which the opinions are considered i.e., “whether the overall sentiment of an individual is negative or positive or neutral”. The research also considers two advanced approaches along with experimental outcomes. By comparing three machines learning models for classification techniques such as logistic regression, hybrid bag-boost algorithm and SVM are considered, and results are analyzed. In [30], [31] discuss the various sentiments associated with reviews collected through twitter. The polarities of the sentiments are identified as positive, negative or neutral.

In the implemented methodology, pre-processing pipeline for audio data as in [32] was implemented. The pipeline consisted of loading, padding, feature extraction, normalization of audio data and finally saving the outputs in the form of pikle extension. Features extraction techniques such as MFCC, chroma, Mel spectrogram and features such as contrast and tonnetz are being extracted which are stored in the above said pikle file format. Once it is done, the next step is to train the model with RAVDASS dataset. The implemented model is based on a deep neural network for training from the dataset.

2. PROPOSED METHOD

Figure 1 shows the architecture of the model being implemented. The whole idea and mechanism start with data being collected by the company’s department responsible for collecting feedbacks and reviews in the form of audio recordings. Once all the data is being collected, it will be stored in a single storage location in the format of .wav file extension. This stored data is then prepared by passing through some pre-processing techniques such as loading all the audio files as batch, padding, extract features and finally save all the output of these steps in a desired location. The features that are extracted are using the MFCCs technique and chroma, Mel-spectrogram, contrast, tonnetz are the features that are directly extracted. All these features are saved as a .pkl or pikle file in a desired location. This pikle file will be used as input to the implemented deep neural network model for training. Once the training is completed, we can test out for a sample audio for prediction of sentiment analysis. For which we need to extract the same set of features as done before and pass it to the trained model for prediction. Based on this final prediction, the company can think of whether they need any improvement in their service/product or not.

2.1. Audio pre-processing

For the audio pre-processing, an entire pipeline was implemented that included loading the audio files, extracting features such as Mel spectrogram, MFCC technique, chroma, contrast, tonnetz. The results are then stored in a pikel file. The loader class is responsible to load all the audio files from the mentioned directory in batch. It is easier to work on batch files instead of loading each file individually and performing pre-processing on them which can be difficult and also time consuming. The padder class in the audio pre-processing pipeline is responsible to perform padding. The padder is used to add a zero-right padding to the audio files to match one single length.

2.2. Feature extraction

Feature extraction on the audio data is essential to precisely identify the essential properties that assist us to identify sentiments. The techniques used for features extractions are MFCC or Mel frequency cepstral coefficient, Mel spectrogram, chroma, tonnetz and contrast. In the proposed work, this is considered in the final implementation.

2.3. Mel frequency cepstral coefficient

MFCC is a major or most frequently used technique for extracting features from the audio signal such as windowing the signal which is used to detect the phones in the audio. As there are plenty of phones in a speech or an audio, the audios are broken down into smaller chunks of a said duration of 25 ms with 10 ms apart from each chunk. The next is taking applying log. A human is sensitive to only a set of frequency and cannot hear all types of frequency. In order to bridge this, log is applied so that the recorded audio can match the human hearable frequency. The next one is preemphasis where the magnitude is increased in the high frequency scale, performing discrete Fourier transform. The first order high pass filter is applied to pre-emphasis as seen in (1) where x is any taken signal, with time ‘t’ and ‘α’ which has a constant value of 0.95. The use of Mel-filter bank is done in MFCC as well, where the humans can distinguish between different frequencies, it is not true for all set of frequency band. But the machine can differentiate with any frequency band and this is done by using a formula as based on (2). It is used to convert Hz to Mel frequency scale. Here ‘f’ stands for frequency, 1,125 is the high frequency scale and 700 is low frequency scale.

$$y(t) = x(t) - a\,x(t-1) \tag{1}$$

$$Mel(f) = 1123\, log(1 + \frac{f}{700}) \quad (2)$$

MFCC technique comes pre-implemented with Librosa library and with the help of a dot operator; one can call and use this technique's algorithm to extract audio features and information. It also comes with analog to digital conversion and the audio can also be sampled with a specific frequency like 8 or 16 kHz. Figure 2 shows the visual representation of MFCC.

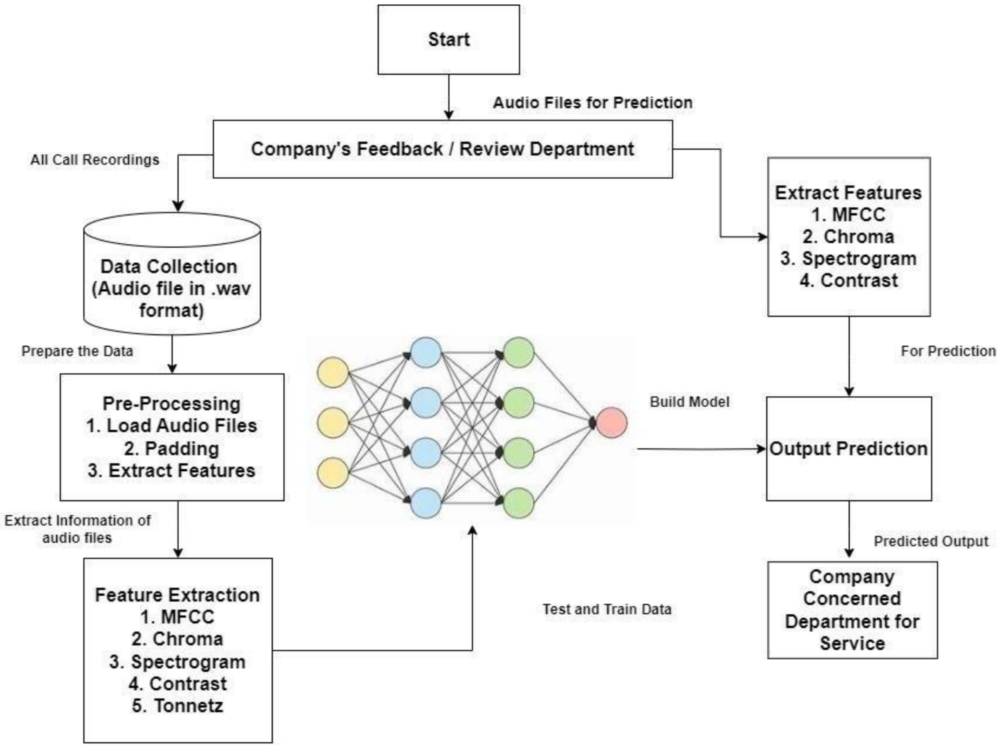

Figure 1. Architecture of audio sentiment analysis

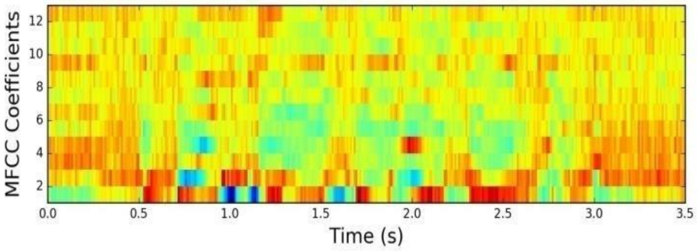

Figure 2. MFCC of a sample audio

MFCC is a "visual representation of the short-term power spectrum of the audio signal which is based on the linear cosine transform of a log power spectrum on the nonlinear Mel frequency scale". Figure 3 shows the Mel Spectrogram of a sample audio file. It is a representation of Mel-scale that is being converted from spectrogram. A spectrogram is basically a representation of frequency spectrum of an audio signal where the frequency spectrum is the frequency range of the audio signal which contains it. The bright color represents high intensity and the darker color represents low intensity in the audio signal. Figure 4 is the Chroma representation of a sample audio file. Chroma is basically a describes the representation of the tonal content in the audio signal which is in the condensed form. In Chroma, the intensity is decided from bright to dark where bright color represents low intensity and darker color represents high intensity.

Figure 5 shows the tonnetz representation of a sample audio file. Tonnetz is a concept representation of lattice diagram which in-term is used to represent the tonal space. The y-axis is made of some alpha-numeric values. This is basically best explained by taking the example of a musical instrument. In any musical instrument, there are some chords, which vary from instrument to instrument, like the guitar has chords like Emajor, and E minor. The y-axis alpha-numeric values in Figure 5 are the examples of those chords. All these features are shown in visual form which the machine or our model cannot understand. To make it possible, all these features are converted to numeric representation from their respective form so that our machine can understand which is done by using the Librosa library. Once these are converted to numeric form, they are then stored in a pikle file.

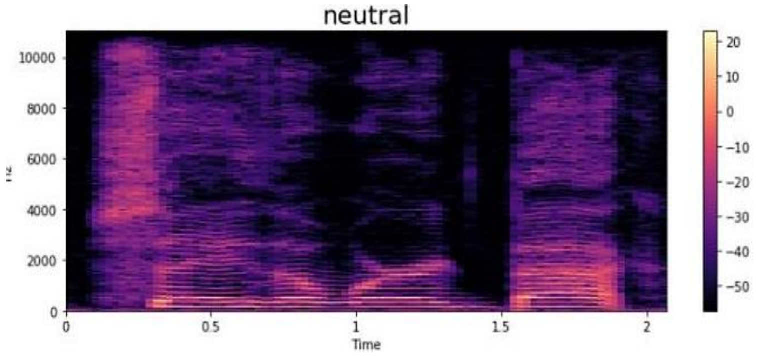

Figure 3. Mel spectrogram of a sample audio

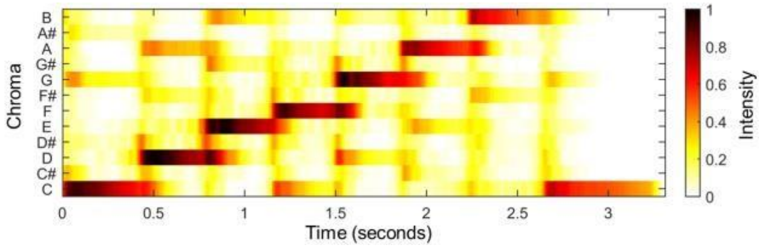

Figure 4. Chroma representation of sample audio

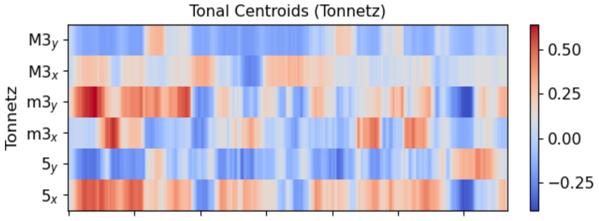

Figure 5. Tonnetz of sample audio file

3. RESULTS AND DISCUSSION

3.1. Training

The model was first trained with tanh as the activation function and rectified linear unit (ReLU) was the immediate second choice. Numbers of preceptron were also being changed. The optimizer being used at first was adadelta and later changing it to Adam. A few of these variations are being shown in Table 1. As the number of hidden layers increased after three, the results were almost the same and there was no improvement in accuracy. In order to maintain the efficiency of the model, it was best to use the least number of hidden layers and preceptron. Thus, the final parameters used are shown in Table 2. The model is trained by using the RAVDASS dataset which consists of about 1,400 audio files with each of them having 48 kHz and in wav format. The dataset is labeled with different emotions such as neutral, calm, fearful, sad, happy, angry, disgust, surprised. Using features such as MFCC, chroma, spectrogram, contrast, tonnetz and the labeled data, we pass these files through the neural network. The end results being obtained as 95% training accuracy and 75% testing accuracy.

Table 1. Model building iteration

Hidden layers	Activation function	Epoch	Preceptrons	Optimizer	Dropout	Testing accuracy	Training accuracy
1	Tanh	50	100	Adadelta	0.5	14.58	20.45
2	Tanh	50	100, 200	Adadelta	0.5	24.31	18.13
2	ReLU	50	200, 300	Adam	0.2	46.53	50.38
2	ReLU	100	200, 300	Adam	0.1	48.61	66..28
3	ReLU	200	300, 400, 300	Adam	0.1	67.89	89.97
4	ReLU	200	600, 800, 800, 600	Adam	0.1	68.75	96.14
5	ReLU	200	600, 800, 800, 800, 600	Adam	0.1	67.36	92.36

Table 2. Final parameters used for model

Hidden layers	Activation function	Epoch	Preceptrons	Optimizer	Dropout	Testing accuracy	Training accuracy
3	ReLU	200	400,600,400	Adam	0.1	75.09	95.49

3.2. Testing

A sample audio file with some spoken content can be inputted and the result can be predicted. But first, feature extraction should be performed over this audio file and with the help of them; we can obtain the result in terms of if it belongs to happy, sad, angry, neutral, surprised, calm categories. Table 2 shows the implemented model is providing with over 95% training accuracy and when the model was put to test with a test audio file, it performed with 75% accuracy. In [23] the results show about 89% accuracy, but this is because the model was trained and tested after the conversion of audio data to contextual data for analysis andprediction. The results are shown in Table 3.

$$accuracy\ =\ str\left(\left(\frac{count}{y2.shape[0]}\right)\right) * 100 \tag{3}$$

As shown in (3) represents the formula used to calculate the accuracy where y2 contains the prediction of model with respect to X_test and with a NumPy’s argmax value. Count is used to iterate over all the audio files featurespresent in the pickle file. And shape is used to give the dimensions of the array. In case the test audio file contains multiple speakers, then the need to identify the speakers and separate them in order to have only the customer’s utterance in the test audio file. In order to do so the concept can be referred to and used from [23]. Figure 6 represents the visual graph representation of the train and validation accuracies for the implemented model against epochs of 200.

Table 3. Comparison with different model’s accuracies

Model	Accuracy
XGBOOST	56%
Sentiment Analysis on Speaker Specific Speech Data [6] (Audio being converted to text)	89%
Proposed model	75%

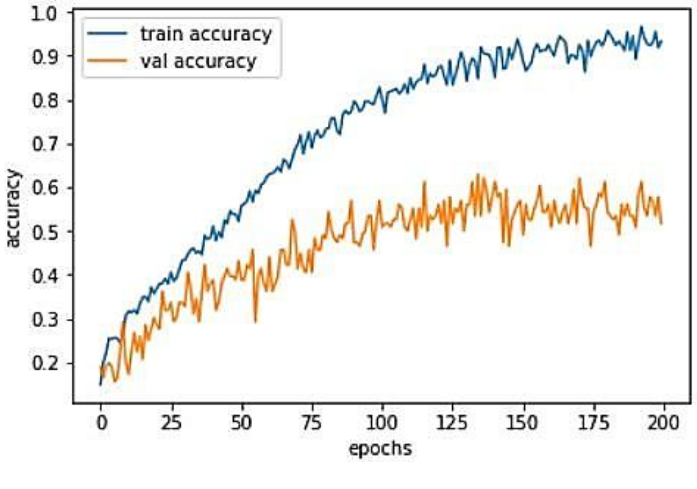

Figure 6. Graph of train/test accuracies

4. CONCLUSION

In the proposed model, the main focus is on the use of audio data for sentiment analysis, as most of the feedbacks are being captured via voice and also to get better understanding of the customer's reviews. In orderto avoid customers using sarcastic comments to confuse the system, audio reviews work best to prevent such things. The proposed model outperforms the existing systems or models by at least 10% in terms of accuracy. For preprocessing of data, to avoid any analysis or preprocessing of audio files manually or one at a time, a pipeline was proposed and implemented by which the entire preprocessing can be automated. This pipeline can accommodate n number of methods as per the implementer's requirements or needs. The applications of using audio as inputs for sentiment analysis does not only apply for CRMs but this application can be used medical domain for treatment of anxiety or depression.

REFERENCES

[1] N. Vryzas, L. Vrysis, R. Kotsakis, and C. Dimoulas, "Speech emotion recognition adapted to multimodal semantic repositories," in *2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, Sep. 2018, pp. 31–35, doi: 10.1109/SMAP.2018.8501881.

[2] W. Kim and J. H. L. Hansen, "Angry emotion detection from real-life conversational speech by leveraging content structure," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 5166–5169, doi: 10.1109/ICASSP.2010.5495021.

[3] W. Li, W. Shao, S. Ji, and E. Cambria, "BiERU: bidirectional emotional recurrent unit for conversational sentiment analysis," *Neurocomputing*, vol. 467, pp. 73–82, Jan. 2022, doi: 10.1016/j.neucom.2021.09.057.

[4] N. Garg and D. K. Sharma, "Sentiment analysis of events on social web," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 6, pp. 1232–1238, Apr. 2020, doi: 10.35940/ijitee.F3946.049620.

[5] A. Samih, A. Ghadi, and A. Fennan, "Deep graph embeddings in recommender systems: a survey," *Journal of Theoretical and Applied Information Technology*, vol. 99, no. 15, pp. 3812–3823, 2021.

[6] I. Salehin *et al.*, "Analysis of student sentiment during video class with multi-layer deep learning approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 3981–3993, Aug. 2022, doi: 10.11591/ijece.v12i4.pp3981-3993.

[7] E. G. Moung, C. C. Wooi, M. M. Sufian, C. K. On, and J. A. Dargham, "Ensemble-based face expression recognition approach for image sentiment analysis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, pp. 2588–2600, Jun. 2022, doi: 10.11591/ijece.v12i3.pp2588-2600.

[8] 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," in *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Sep. 2021, pp. 1–6, doi: 10.1109/ICSES52305.2021.9633843.

[9] D. Bertero and P. Fung, "A first look into a convolutional neural network for speech emotion detection," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 5115–5119, doi: 10.1109/ICASSP.2017.7953131.

[10] M. V. V. P. Kantipud and S. Kumar, "A computationally efficient learning model to classify audio signal attributes," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 5, pp. 4926–4934, Oct. 2022, doi: 10.11591/ijece.v12i5.pp4926-4934.

[11] F. Chen and Z. Luo, "Learning robust heterogeneous signal features from parallel neural network for audio sentiment analysis," *arXiv preprint arXiv:1811.08065*, Nov. 2018.

[12] S. Maghilnan and M. R. Kumar, "Sentiment analysis on speaker specific speech data," in *2017 International Conference on Intelligent Computing and Control (I2C2)*, Jun. 2017, pp. 1–5, doi: 10.1109/I2C2.2017.8321795.

[13] D. Rotovei, "Multi-agent aspect level sentiment analysis in CRM systems," in *2016 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Sep. 2016, pp. 400–407, doi: 10.1109/SYNASC.2016.068.

[14] D. Rotovei and V. Negru, "Improving lost/won classification in CRM systems using sentiment analysis," in *2017 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Sep. 2017, pp. 180–187, doi: 10.1109/SYNASC.2017.00038.

[15] D. Rotovei and V. Negru, "Multi-agent recommendation and aspect level sentiment analysis in B2B CRM systems," in *2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Sep. 2020, pp. 238–245, doi: 10.1109/SYNASC51798.2020.00046.

[16] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An interactive multi-task learning network for end-to-end aspect-based sentiment analysis," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 504–515, doi: 10.18653/v1/P19-1048.

[17] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Aug. 2017, pp. 4068–4074, doi: 10.24963/ijcai.2017/568.

[18] N. Vryzas, L. Vrysis, R. Kotsakis, and C. Dimoulas, "Speech emotion recognition adapted to multimodal semantic repositoriess," in *2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, Sep. 2018, pp. 31–35, doi: 10.1109/SMAP.2018.8501881.

[19] Q. Lu, Z. Zhu, G. Zhang, S. Kang, and P. Liu, "Aspect-gated graph convolutional networks for aspect-based sentiment analysis," *Applied Intelligence*, vol. 51, no. 7, pp. 4408–4419, Jul. 2021, doi: 10.1007/s10489-020-02095-3.

[20] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/widm.1253.

[21] N. Capuano, L. Greco, P. Ritrovato, and M. Vento, "Sentiment analysis for customer relationship management: an incremental learning approach," *Applied Intelligence*, vol. 51, no. 6, pp. 3339–3352, Jun. 2021, doi: 10.1007/s10489-020-01984-x.

[22] Y. Zhang, Z. Zhao, P. Wang, X. Li, L. Rong, and D. Song, "ScenarioSA: a dyadic conversational database for interactive sentiment analysis," *IEEE Access*, vol. 8, pp. 90652–90664, 2020, doi: 10.1109/ACCESS.2020.2994147.

[23] J. Feng, C. Gong, X. Li, and R. Y. K. Lau, "Automatic approach of sentiment lexicon generation for mobile shopping reviews," *Wireless Communications and Mobile Computing*, pp. 1–13, Aug. 2018, doi: 10.1155/2018/9839432.

[24] S. E. Griesser and N. Gupta, "Triangulated sentiment analysis of Tweets for social CRM," in *2019 6th Swiss Conference on Data Science (SDS)*, Jun. 2019, pp. 75–79, doi: 10.1109/SDS.2019.000-4.

[25] G. Chaubey, P. R. Gavhane, D. Bisen, and S. K. Arjaria, "Customer purchasing behavior prediction using machine learning classification techniques," *Journal of Ambient Intelligence and Humanized Computing*, Apr. 2022, doi: 10.1007/s12652-022-03837-6.

[26] A. Suriya, "Psychology factor based sentiment analysis for online product customer review using multi-mixed short text ridge analysis," *Journal of Critical Reviews*, vol. 6, no. 6, pp. 146–150, 2019, doi: 10.22159/jcr.06.06.20.

[27] M. Seyfioğlu and M. Demirezen, "A hierarchical approach for sentiment analysis and categorization of Turkish written customer relationship management data," in *Proceedings of the 2017 Federated Conference on Computer Science and Information Systems*, Sep. 2017, pp. 361–365, doi: 10.15439/2017F204.

[28] M. R. Yaakub, Y. Li, and J. Zhang, "Integration of sentiment analysis into customer relational model: the importance of feature ontology and synonym," *Procedia Technology*, vol. 11, pp. 495–501, 2013, doi: 10.1016/j.protcy.2013.12.220.

[29] S. K. Pathuri, N. Anbazhagan, and G. B. Prakash, "Feature based sentimental analysis for prediction of mobile reviews using hybrid bag-boost algorithm," in *2020 7th International Conference on Smart Structures and Systems (ICSSS)*, Jul. 2020, pp. 1–5, doi: 10.1109/ICSSS49621.2020.9201990.

[30] N. N. Alabid and Z. D. Katheeth, "Sentiment analysis of Twitter posts related to the COVID-19 vaccines," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 24, no. 3, pp. 1727–1734, Dec. 2021, doi: 10.11591/ijeecs.v24.i3.pp1727-1734.

[31] S. Yousefinaghani, R. Dara, S. Mubareka, A. Papadopoulos, and S. Sharif, "An analysis of COVID-19 vaccine sentiments and opinions on Twitter," *International Journal of Infectious Diseases*, vol. 108, pp. 256–262, Jul. 2021, doi: 10.1016/j.ijid.2021.05.059.

[32] A. Katti and M. Sumana, "Pipeline for pre-processing of audio data," in *Smart Innovation, Systems and Technologies*, vol. 312, Springer Nature Singapore, 2023, pp. 191–198.

BIOGRAPHIES OF AUTHORS

Sumana Maradithaya is an associate professor at Ramaiah Institute of Technology, India. She is a senior IEEE member and an IEEE-CIS member. She has paper publications in reputable conferences and journals. She was actively involved in several IEEE activities such as community outreach activities, industry interaction, webinars, and projects. She has mentored industry related projects. She is an active reviewer for several reputed journals and international IEEE (conecct and HPC) conferences. She has actively contributed towards several projects in the areas of machine learning, data science and deep learning. She can be contacted at email: sumana.a.a@gmail.com.

Anantshesh Katti has completed his M.Tech. in software engineering from Ramaiah Institute of Technology, India. He is currently working as a Software Developer at continental automotive working with C++, React and NodeJS Technologies. He has worked on projects related to machine learning and deep learning. He also has freelancer's experience in phone applications development using Java and flutter technologies. He can be contacted at email: anantsheshkatti@gmail.com.