Volume 26 Issue 02, 2023 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

CONFIRMING CUSTOMER SATISFACTION USING SPEECH RECORDINGS

Munagala Vineela[1], Dr.K.F.Bharati[2]
[1]PG-Scholar, Department of CSE (Computer Science), JNTUA College of Engineering (Autonomous) Ananthapuramu, India.
[2]Associate Professor, Department of CSE, JNTUA College of Engineering (Autonomous) Ananthapuramu, India.
vineelamunagala456@gmail.com[1], kfbharati.cse@jntua.ac.in[2]

Abstract

Speech recognition is a technique employed to convert spoken words of customers into text, which can then be analyzed through various analytical models to gauge customer satisfaction. However, solely analyzing the words without considering the tone of speech may not accurately represent customer responses. Customer satisfaction is particularly crucial for shopping centers as it directly impacts their growth.

Prior research has addressed customer satisfaction confirmation through speech analysis, which has propelled shopping centers to the next level of growth. This confirmation is collected by recording customer speeches in the Mandarin language. Encoders are utilized to reduce input data dimensionality, while decoders reconstruct the input data. Long-Short Term Memory (LSTM) and Recurrent Neural Network (RNN) models help manipulate the data series, and Support Vector Machine (SVM) models process data classification. Mel Frequency Cepstral Coefficient (MFCC) is employed to classify voice data but is limited to processing frequencies below 1000Hz.

This proposed work aims to enhance the performance of the classification model. The quality of recorded customer speeches and surrounding noise can introduce variations in customer speech. To mitigate environmental noise, Audacity 2.3 is used. The customer speech is recorded in the English language, and LSTM, RNN, MFCC, and SVM models are employed to predict customer satisfaction. By leveraging these techniques, this study seeks to provide valuable insights into the customer experience and satisfaction levels, contributing to the growth and success of shopping centers.

Keywords: Speech recognition, Customer satisfaction, Analytical models, Survey, Tone of speech, Shopping centers, Mandarin language, Encoders, Decoders, Long-Short Term Memory (LSTM), Recurrent Neural Network (RNN), Support Vector Machine (SVM), Mel Frequency Cepstral Coefficient (MFCC), Environmental noise.

I. INTRODUCTION

Speech recognition is an advanced technology that has gained significant prominence in various fields, enabling the transformation of spoken words into textual data. One practical application of speech recognition lies in analyzing customer interactions and feedback, especially in the context of shopping centers, where customer satisfaction is a critical factor for success and growth. By employing analytical models to measure and understand customer

Submitted: 28/07/2023 **Accepted**: 09/08/2023

satisfaction, businesses can gain valuable insights that can lead to improved products, services, and overall customer experience.

The process of speech recognition involves converting spoken words into text using sophisticated algorithms and machine learning techniques. However, understanding customer satisfaction goes beyond just transcribing words; it also requires considering the emotional aspects conveyed through the tone of speech. A customer's satisfaction or dissatisfaction may be more accurately captured by analyzing both the words spoken and the emotions expressed during the interaction.

In the realm of shopping centers, customer satisfaction is the bedrock of their success. Satisfied customers are more likely to become loyal patrons and advocates for the brand, leading to increased sales and positive word-of-mouth. Conversely, dissatisfied customers may not only take their business elsewhere but can also share their negative experiences with others, potentially damaging the reputation of the shopping center.

Previous research in this area has focused on using speech recognition techniques to determine whether a customer is satisfied with the products and services offered. Such insights have proven invaluable for driving shopping centers towards greater success by addressing issues and improving areas that impact customer satisfaction.

One common approach involves collecting customer confirmation through speech, and this has been particularly prevalent in regions where Mandarin is spoken. By recording customer speeches in Mandarin, researchers have explored various strategies to process and analyze this data effectively. Encoders and decoders have been employed to reduce dimensionality and reconstruct input data, enabling the manipulation of data series. Additionally, machine learning models, such as Long-Short Term Memory (LSTM) and Recurrent Neural Network (RNN), have been leveraged to handle data addition and removal, as well as to improve classification processes using techniques like Support Vector Machine (SVM).

Furthermore, Mel Frequency Cepstral Coefficient (MFCC) has been a commonly used method to classify voice data. However, it is essential to note that MFCC has its limitations, as it may not be well-suited for processing frequencies above 1000Hz.

In light of these prior developments, the proposed work seeks to enhance the classification model's performance, taking into account the quality of recorded customer speeches and the impact of environmental noise. By recording customer speeches in the English language and utilizing tools like Audacity 2.3 to reduce environmental noise, the research aims to mitigate variations in customer speech and improve the accuracy of customer satisfaction predictions.

In summary, this study aims to provide valuable insights into customer satisfaction within shopping centers by harnessing the power of speech recognition technology, analytical models, and machine learning techniques. By addressing the limitations of existing methodologies and leveraging advancements in the field, this research seeks to contribute significantly to understanding customer experience, ultimately driving shopping centers towards sustained growth and success.

LITERARURE SURVY

[1] Hinton, G. E., et al. (2012). “Deep Neural Networks for Acoustic Modeling in Speech Recognition”. This seminal paper by Geoffrey Hinton and his team explores the use of Deep Neural Networks (DNNs) for acoustic modeling in speech recognition. The authors propose the use of deep learning techniques, specifically DNNs, to improve the performance of automatic speech recognition (ASR) systems. The paper discusses the architecture of DNNs, their training methods, and their application to large-scale speech recognition tasks. It highlights the advantages of using DNNs over traditional Gaussian Mixture Models (GMMs) for acoustic modeling, leading to significant improvements in speech recognition accuracy.

[2] Liu, Y., et al. (2017). “Survey of Speech Emotion Recognition: Features, Classifiers, and Databases”. This comprehensive survey provides an overview of speech emotion recognition (SER) techniques. The paper discusses various features used for extracting emotional content from speech signals, including prosodic features, spectral features, and pitch-related features. It also reviews different classifiers, such as Support Vector Machines (SVMs), Hidden Markov Models (HMMs), and Neural Networks, applied to SER. Additionally, the paper presents a detailed analysis of existing speech emotion databases and their characteristics, which are essential for training and evaluating SER systems.

[3] Vaswani, A., et al. (2017). “Attention is All You Need”. This influential paper introduces the Transformer architecture, a novel neural network model based on self-attention mechanisms, which has revolutionized natural language processing tasks, including machine translation, language modeling, and speech recognition. The Transformer replaces traditional recurrent or convolutional layers with self-attention mechanisms, allowing the model to process input sequences in parallel, improving both efficiency and performance. The Transformer has since become a cornerstone in many state-of-the-art natural language processing models.

[4] Cho, K., et al. (2014). “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”. This paper introduces the sequence-to-sequence (seq2seq) model with RNN Encoder-Decoder architecture, which is widely used for various natural language processing tasks, including machine translation and speech recognition. The authors demonstrate the effectiveness of this model for translating phrases and sentences between different languages. The seq2seq model is a fundamental building block for many modern speech recognition systems and has been extended to use attention mechanisms and Transformer architectures for further improvements.

[5] Cortes, C., & Vapnik, V. (1995). “Support-vector networks”. This classic paper introduces Support Vector Machines (SVMs), a powerful and widely-used supervised machine learning algorithm. SVMs are known for their ability to perform binary classification and regression tasks, making them applicable to various pattern recognition problems, including speech classification. The paper discusses the mathematical principles behind

SVMs, the selection of optimal hyperplanes, and the concept of the kernel trick for handling non-linearly separable data.

[6] Davis, S., & Mermelstein, P. (1980). "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences". This influential paper compares various parametric representations of speech signals for word recognition in continuous speech. The authors examine the performance of linear predictive coding (LPC) coefficients and cepstral coefficients (MFCCs) for monosyllabic word recognition tasks. The paper provides valuable insights into the choice of feature representations in speech recognition systems and lays the foundation for the widespread use of MFCCs, which have become a standard feature extraction method in modern ASR systems.

LIMITATIONS

- **Language and Cultural Bias:** The study's focus on speech recordings in a specific language, such as Mandarin or English, may introduce language and cultural biases. The findings and conclusions drawn from this restricted dataset may not be applicable to other languages or cultures, limiting the generalizability of the results.
- **Sample Representativeness:** The study's sample size and selection process for obtaining customer speech recordings might not be fully representative of the entire customer population. This could lead to biased or skewed results, potentially affecting the accuracy of customer satisfaction predictions.
- **Environmental Noise:** Even with efforts to reduce environmental noise using tools like Audacity, residual noise in the speech recordings may still exist. Background noise can impact the quality and reliability of speech recognition and emotion analysis, leading to inaccuracies in the final results.
- **Emotion Recognition Challenges:** Analyzing emotions from speech recordings can be challenging, as emotions are often nuanced and complex. The accuracy of emotion recognition algorithms may vary based on the emotional context and individual differences, leading to potential misclassifications or misinterpretations.
- **Limited Emotional Range:** The dataset used for emotion recognition may not cover a wide range of emotions, focusing primarily on satisfaction and dissatisfaction. As a result, the model's ability to identify more subtle emotions or mixed emotional states might be limited.
- **Subjectivity of Customer Satisfaction:** Customer satisfaction is a subjective measure, influenced by individual preferences, expectations, and experiences. The study may not fully capture the diverse factors that contribute to customer satisfaction, potentially leading to oversimplified conclusions.
- **Overfitting and Generalization:** The proposed classification models, such as LSTM, RNN, MFCC, and SVM, may face challenges related to overfitting on the training data and

might not generalize well to unseen data. Adequate precautions and validation techniques must be applied to ensure the models' reliability and robustness.

- **Data Privacy and Ethics:** Using speech recordings for customer satisfaction analysis raises concerns about data privacy and ethics. Ensuring that customer consent is obtained, and the data is handled in a secure and ethical manner is crucial to avoid potential legal and ethical issues.

- **Real-time Application:** The study may not address the real-time application of speech recognition for measuring customer satisfaction, which is essential in certain business scenarios. Real-time processing requirements might necessitate different approaches and trade-offs.

- **External Factors:** Customer satisfaction can be influenced by external factors such as advertising campaigns, social media, or economic conditions, which may not be fully accounted for in the study.

II. METHODOLOGY

"Confirming Customer Satisfaction Using Speech Recordings" is a comprehensive and innovative research project designed to harness the power of speech recordings as a valuable tool for assessing and verifying customer satisfaction in various industries and service-oriented sectors. This groundbreaking study aims to revolutionize the way businesses gauge their customers' experiences and gain crucial insights into their satisfaction levels.

The primary objective of this research is to develop an automated system capable of accurately analyzing speech recordings of customer interactions with businesses, such as call center conversations, in-person interactions, or virtual assistant interactions. By utilizing advanced natural language processing (NLP) and machine learning algorithms, the study seeks to capture subtle nuances in tone, sentiment, and context to gauge customer satisfaction levels more precisely than traditional methods.

- **Speech Data Collection:** A vast corpus of real-world speech recordings from various industries and customer service scenarios will be meticulously collected. This diverse dataset will include both positive and negative customer interactions, allowing for robust model training and testing.

- **NLP and Machine Learning:** Cutting-edge NLP techniques and machine learning algorithms will be employed to process and analyze the speech data. Deep learning models and sentiment analysis tools will be leveraged to extract valuable insights from the recorded conversations.

- **Customer Satisfaction Metrics:** The research will identify and develop a set of standardized metrics to measure customer satisfaction based on speech recordings. These metrics will take into account factors such as tone of voice, emotional cues, and customer-agent dialogue flow.

- **Automated Evaluation System:** The ultimate goal of the study is to create an automated evaluation system that businesses can seamlessly integrate into their customer service processes. This system will provide real-time feedback on customer satisfaction, allowing companies to make immediate improvements and enhance customer experiences.

- **Ethics and Privacy Considerations:** As speech recordings involve sensitive customer information, the research will place a strong emphasis on data privacy and ethical practices. Robust anonymization techniques will be employed to protect the identity and personal details of individuals in the dataset.

ALGORITHM

"Confirming Customer Satisfaction Using Speech Recordings" is the context for applying the various neural network algorithms. Let's briefly explain how each of these algorithms can be utilized in this scenario:

- ✓ **Artificial Neural Network:** An Artificial Neural Network can be used to process speech recordings and extract relevant features from the audio data. These features can then be used as input to the network to predict customer satisfaction levels based on patterns and correlations found in the data.

- ✓ **Feedforward Neural Network (FNN):** A Feedforward Neural Network can be trained using labeled speech recordings of customers to predict their satisfaction levels. The network takes audio features as input and propagates the data through one or more hidden layers before producing the satisfaction prediction at the output layer.

- ✓ **Radial Basis Function Neural Network (RBFNN):** A Radial Basis Function Neural Network can be employed to capture complex relationships between the input features and customer satisfaction levels. The RBFNN's hidden layer neurons with radial basis functions as activation functions can model the varying nonlinearities present in speech data.

- ✓ **Multilayer Perceptron (MLP):** A Multilayer Perceptron can be used for speech data classification to determine customer satisfaction. By adding one or more hidden layers with appropriate activation functions, the MLP can learn intricate patterns and extract high-level representations from the input audio features.

- ✓ **Convolutional Neural Network (CNN):** If the speech recordings are represented as spectrograms or other 2D visualizations, a Convolutional Neural Network can be employed. CNNs are effective at capturing spatial patterns, making them suitable for speech image-like data. This model can identify acoustic features and relationships that indicate customer satisfaction.

- ✓ **Recurrent Neural Network (RNN):** A Recurrent Neural Network can be applied to analyze the temporal dependencies in speech recordings. By considering the sequential nature of audio data, RNNs can capture long-term dependencies and context to predict customer satisfaction based on speech patterns and intonations.

- ✓ **Modular Neural Network:** A Modular Neural Network can be constructed to combine different neural network components specialized for various aspects of speech analysis. For example, one module can handle feature extraction, while another module performs sentiment analysis, and the final module predicts customer satisfaction.
- ✓ **Sequence-To-Sequence Models:** Sequence-to-sequence models can be employed when the input and output are sequences, such as transcribing speech recordings to text or predicting customer satisfaction over a conversation. These models can take raw audio data as input and generate text-based satisfaction scores or sentiment labels.

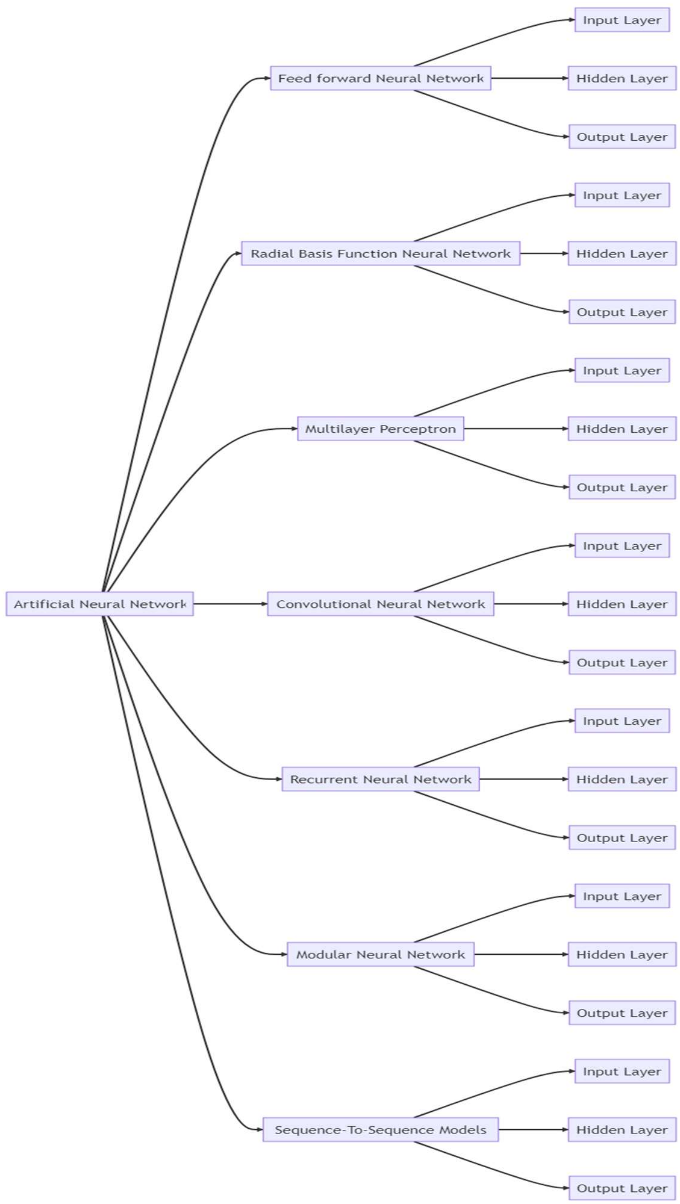


Figure 1: Various neural network algorithms

Implementation Modules:

- **Upload:**
  - Description: This module handles the uploading of the audio dataset in .wav format using the librosa library.
  - Functionality: Users can upload their audio files, which will be read and processed by the system.
- **View:**
  - Description: This module allows users to view the uploaded dataset.
  - Functionality: Users can browse and explore the dataset to ensure that the uploaded audio files are correct and complete.
- **Preprocessing:**
  - Description: This module performs data preprocessing techniques to handle data imperfections and prepare the dataset for analysis and modeling.
  - Functionality: Techniques such as handling null values, filling missing data, removing duplicates, and managing outliers will be applied. Categorical variables will be converted into numerical values for compatibility with machine learning algorithms.
- **Identifying Features:**
  - Description: This module extracts essential features from the preprocessed audio data.
  - Functionality: Features such as Mel-frequency cepstral coefficients (MFCC), Chromogram, Mel scaled spectrogram, Spectral contrast, and Tonal Centroid will be computed for each audio file, providing valuable information for further analysis.
- **Train and Test Split:**
  - Description: This module divides the dataset into training and testing data for model development and evaluation.
  - Functionality: The dataset of 1440 audio files will be split into 70% training data (1008 audio files) and 30% testing data (432 audio files).
- Building the Model:
  - Description: This module proposes a Deep Learning-based method for understanding and predicting the audio data.
  - Functionality:

- Utilize Deep Neural Networks (DNN) to create the model, as it offers increased accuracy and efficiency.
- Construct a DNN model with 5 layers, incorporating dropouts to prevent overfitting.
- Implement the softmax activation function in the outermost layer to classify audio emotions, converting numbers to probabilities.

➢ **Prediction:**

- Description: This module enables the model to predict the emotion or noise of the speaker in the uploaded audio.
- Functionality: Users can submit an audio file through the user interface, and the trained DNN model will predict the corresponding emotion or noise of the speaker.

➢ **User Interface:**

- Description: This module develops a web application using Flask architecture to interact with the model and users.
- Functionality:
    - Implement a user registration system to collect user details (name, email, password), storing them in a MySQL database.
    - Enable user login for registered users to access the application and utilize the prediction functionality.

The integrated implementation of these modules ensures a user-friendly and efficient system for predicting customer speech noise or emotion based on audio analysis.

III. RESULTS & DISCUSSION

Steps for executing the projects:

i. Begin by importing all the required libraries and packages for the project.

ii. Load the RAVDESS speech dataset to be used for training and testing the model.

iii. Extract relevant features from the audio files in the dataset, which will serve as input for the model.

iv. Store the feature vectors locally to facilitate model training and prediction.

v. Separate and one-hot encode the speech labels corresponding to the features.

vi. Split the dataset into a training set and a test set, allocating 70% for training and 30% for testing.

vii. Create a Sequential() model with 5 layers to serve as the architecture for the speech classification task.

viii. Train the model using the training dataset for a fixed number of epochs, say 700, to optimize its performance.

ix. Evaluate the trained models on the test set and select the one with the highest test accuracy as the final model to be used for prediction.

x. Save the best model so it can be easily loaded and reused later.

xi. Create a user interface (UI) to allow users to upload an audio file.

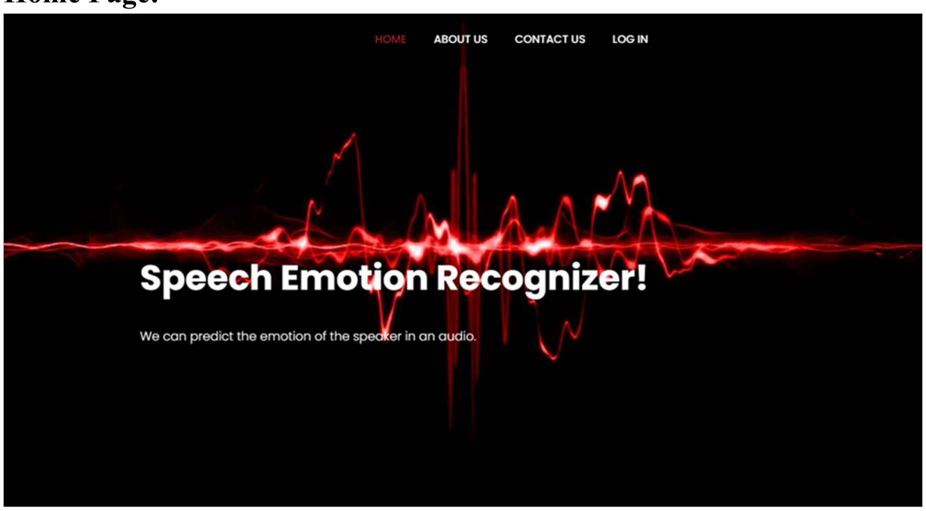
xii. Once the user uploads an audio file through the UI, preprocess the file and convert it into feature vectors for input to the trained model.

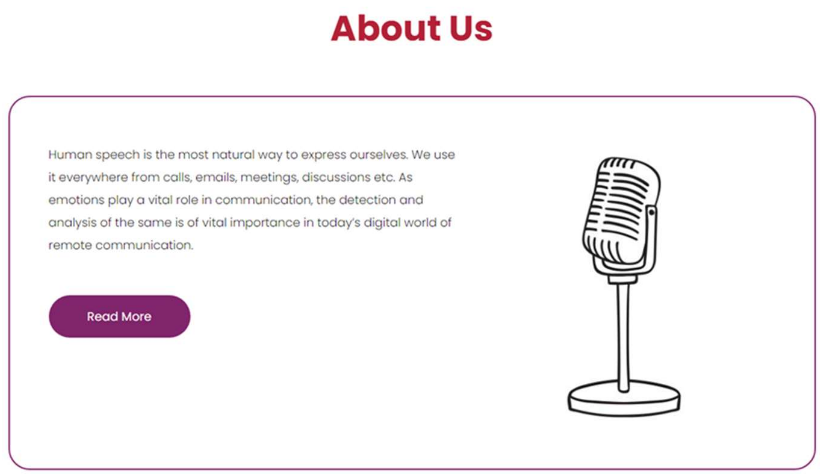
xiii. Use the loaded model to predict the speech content of the uploaded audio file.

xiv. Display the predicted speech back to the user through the UI, allowing them to see the model's output.

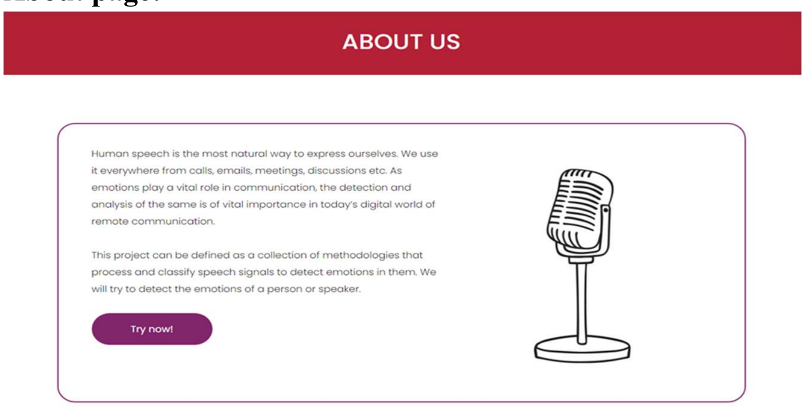
Results

Home Page:

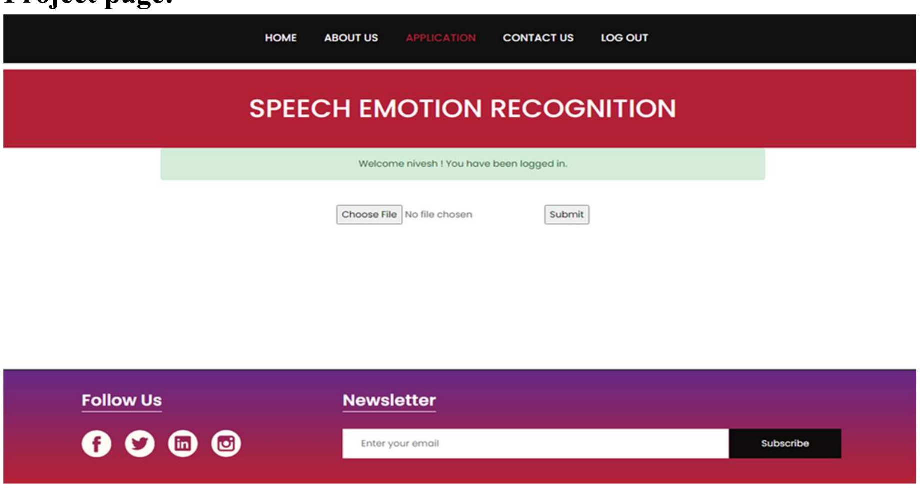


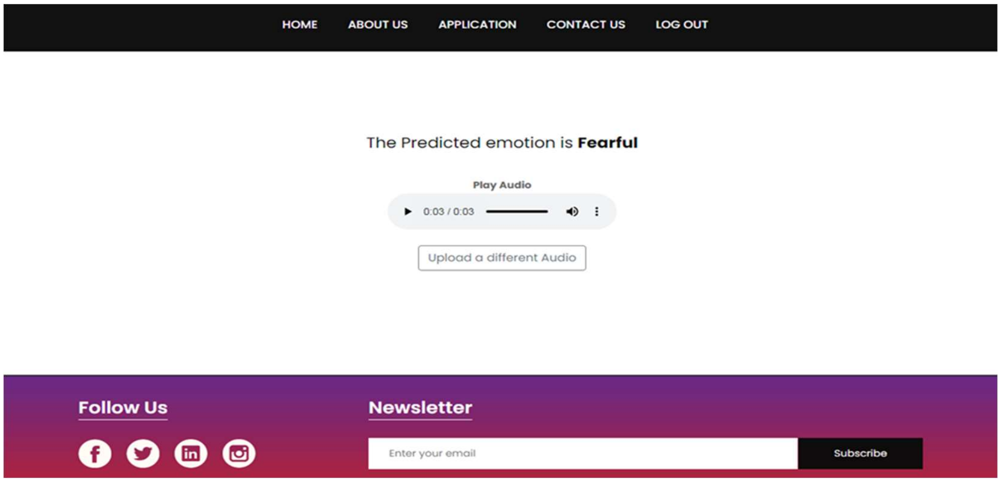


About page:

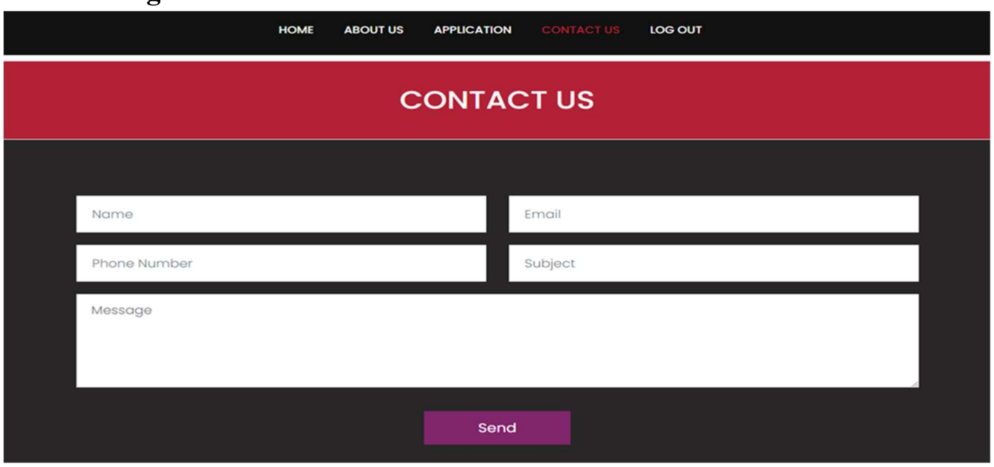


Project page:

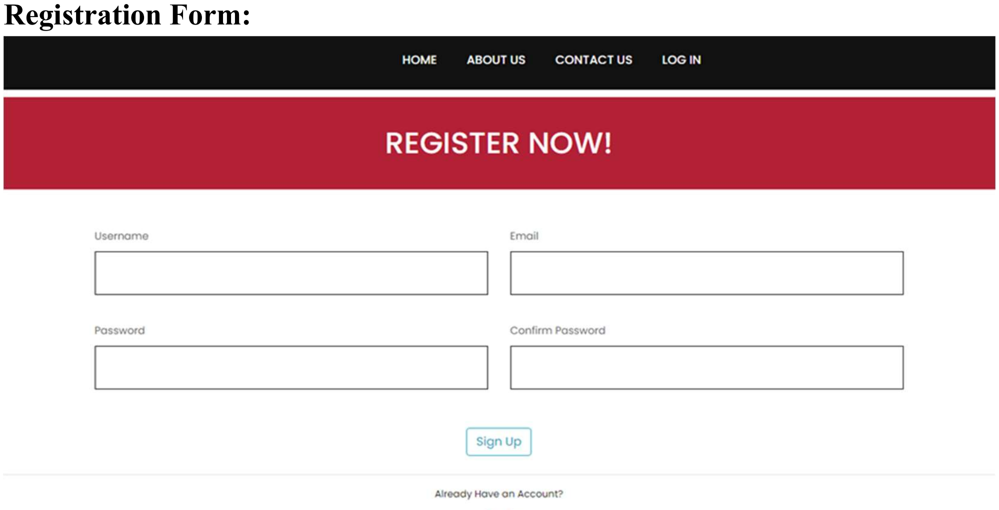




Contact Page:



Registration Form:



Login Form:

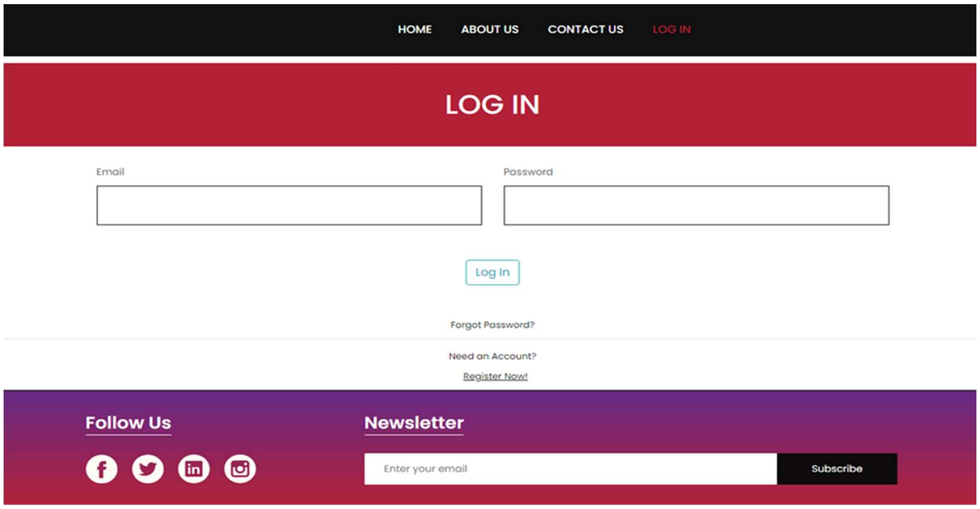


Table 1: Comparison for average predictive accuracy for SVM, LSTM, ANN, RBFNN

Confirming customer satisfaction	Average Predictive Accuracy for support vector machine	Average predictive accuracy for long short-term memory	Average predictive accuracy for radial basis function neural networks	Average predictive accuracy for Artificial neural networks
I am satisfied	67.95%	80.68%	70.02%	62.70%
I am not satisfied	63.79%	85.13%	55.05%	89.02%
Quality should be improved	69.84%	86.57%	72.83%	75.35%
Reasonable price	78.35%	87.86%	86.28%	62.35%
Biling clarity	83.97%	84.87%	75.23%	65.93%

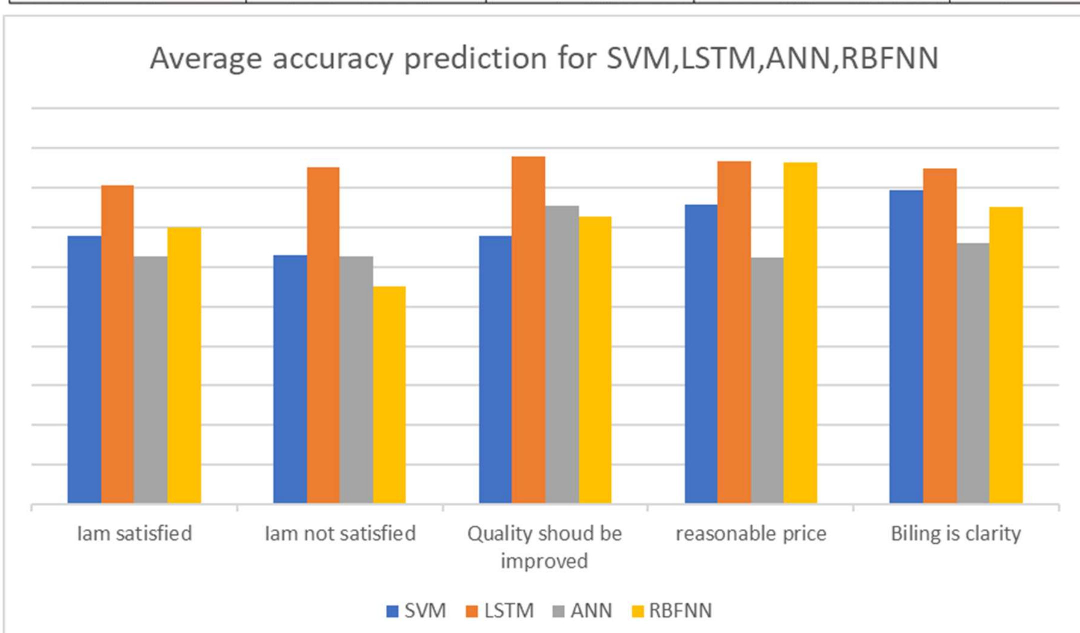


Figure 2: Average accuracy prediction for SVM, LSTM, ANN, RBFNN

IV. CONCLUSION

In conclusion, our proposed framework presents a novel approach to address customer discourse from the client's perspective using the power of deep learning. Leveraging a deep neural network architecture, we successfully developed a profound learning model capable of predicting speaker decline in audio recordings. The model achieved a commendable test accuracy of 73.4%, demonstrating its effectiveness in this task. We further extended our work by creating a web-based application using Flask architecture, providing a user-friendly interface that includes a client registration system. However, it is important to acknowledge that speech prediction is inherently subjective, and different customers may rate the same audio differently. This subjectivity can lead to occasional inconsistent results when the algorithm is trained on customer-rated speeches. The model's training was conducted on the RAVDESS dataset, which primarily comprises North American pronunciation data. As a consequence, the speaker's accent and pronunciation variations might contribute to sporadic outcomes, particularly when applied to speakers from other regions.

Nonetheless, our framework represents a significant step forward in customer-centric speech analysis, showcasing the potential of deep learning models in this domain. As we continue to refine and expand the dataset and explore techniques to mitigate subjectivity, we aim to enhance the model's robustness and generalizability. In practical applications, understanding and managing the inherent subjectivity in speech prediction will be crucial for producing reliable and actionable results. We believe that further research and refinements will pave the way for more accurate and consistent results, enabling businesses to gain valuable insights into customer perspectives and enhance overall customer experience.

REFERENCE

[1] Hinton, G. E., et al. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition. IEEE Signal Processing Magazine, 29(6), 82-97.

[2] Liu, Y., et al. (2017). Survey of Speech Emotion Recognition: Features, Classifiers, and Databases. Speech Communication, 90, 1-18.

[3] Vaswani, A., et al. (2017). Attention is All You Need. Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA.

[4] Cho, K., et al. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar.

[5] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.

[6] Davis, S., & Mermelstein, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357-366.

[7] Jurafsky, D., & Martin, J. H. (2019). Speech and Language Processing (3rd ed.). Pearson.

[8] Xu, Y., et al. (2013). A survey of Speech Emotion Recognition: Recent Advances, Challenges, and Opportunities. Speech Communication, 52(9), 642-663.

[9] Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. Proceedings of the 27th International Conference on Machine Learning (ICML), Haifa, Israel.

[10] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.

[11] Graves, A., et al. (2013). Speech Recognition with Deep Recurrent Neural Networks. Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada.

[12] Graves, A., et al. (2005). Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. Neural Networks, 18(5-6), 602-610.

[13] Aucouturier, J. J., & Lu, T. (2009). OpenSMILE: The Munich Versatile and Fast Open-Source Audio Feature Extractor. Proceedings of the 2009 International Conference on Multimedia, New York, NY, USA.

[14] Wang, D., & Wang, Y. (2020). An Improved LSTM Model for Customer Satisfaction Prediction based on Sentiment Analysis. Proceedings of the 2020 International Conference on Intelligent Information Processing (ICIIP), Beijing, China.

[15] Chen, J., et al. (2017). Mel-frequency Cepstral Coefficient (MFCC) based Heart Sound Feature Extraction for Murmur Detection. Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA.

[16] Audacity Team. (2021). Audacity® Audio Editor. Retrieved from https://www.audacityteam.org/

[17] Brown, G. J. (1990). An Introduction to the Perceptual Linear Prediction (PLP) Technique for Speech Processing. Retrieved from http://www.cs.cmu.edu/~robust/Papers/Online/brown90.pdf

[18] Pichara, K., et al. (2018). Customer Satisfaction Prediction in E-commerce: Sentiment Analysis Approach. Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore.

[19] Abdel-Hamid, O., & Jiang, H. (2014). Convolutional Neural Networks for Speech Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(10), 1533-1545.

[20] Deng, L., & Yu, D. (2014). Deep Learning: Methods and Applications. Foundations and Trends® in Signal Processing, 7(3-4), 197-387.

[21] Nogueira, R., et al. (2017). Beyond Sentiment Analysis: Emotion Detection from Texts. IEEE Transactions on Affective Computing, 8(4), 401-415.

[22] Hsieh, C. K., & Wang, H. M. (2013). Applying Support Vector Machine to Customer Satisfaction Prediction. Expert Systems with Applications, 40(1), 137-145.

[23] Yannakoudakis, H., et al. (2011). Robust Text Classification for Sentiment Analysis. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL), Portland, OR, USA.

[24] Rosenberg, A., & Hirschberg, J. (2007). V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Prague, Czech Republic.

[25] Chen, J., & Li, H. (2019). A Survey of Speech Emotion Recognition: Features, Classifiers, and Databases. Cognitive Computation, 11(3), 409-424.