ITCS 5156 Project

Voice to Age Prediction

Dhanush Kumar Antharam
Student Id – 801282468

Primary Paper Reference

Title: Voice-based Gender and Age Recognition System.
Authors: V. S. Kone, A. Anagal, S. Anegundi, P. Jadhav, U. Kulkarni, and M. S. M.
Year: 2023
Conference/Journal Name: 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India.
Method of the Selected Paper: The paper proposes a system designed to recognize gender and age from voice recordings. It utilizes a combination of signal processing techniques and machine learning algorithms to analyze voice data. The voice signals are first preprocessed to remove noise and enhance quality, then features such as pitch, frequency, and intensity are extracted. These features are used to train models capable of distinguishing between different genders and age groups. The study evaluates the system's performance using a variety of metrics, demonstrating its effectiveness in gender and age recognition tasks.

Abstract

This research project delves into the realm of age group prediction through the analysis of voice data, utilizing a comprehensive set of audio features extracted from recordings. The objective is to discern age-related patterns within the extracted features, thereby enabling accurate age prediction. Two distinct methodologies, classical machine learning (ML) and neural networks, are explored to ascertain their effectiveness in this task.

For classical ML, a total of 23 audio features including "spectral centroid," "spectral bandwidth," "spectral_rolloff," "mfcc1" to "mfcc20," "Chroma Feature," "Spectral Contrast," "Tonnetz," and "RMS Energy" are extracted from the voice recordings. These features capture essential characteristics of the audio signals and form the basis for age prediction models. Logistic regression, Support Vector Machines (SVC), and k-Nearest Neighbors (KNN) are implemented and evaluated as part of the classical ML approach.

In contrast, the neural network approach involves the extraction of a more extensive set of 191 audio features. This expanded feature set aims to capture intricate nuances within the voice data for enhanced predictive capabilities. Feedforward Neural Networks (FNN) and Convolutional Neural Networks (CNN) are employed as neural network architectures to uncover complex patterns and relationships within the data.

The project evaluates and compares the performance of these methodologies, assessing their accuracy, precision, and recall in predicting age groups from voice data. Insights gained from this study contribute to our understanding of the efficacy of classical ML and neural networks in handling age prediction tasks based on audio features.

INTRODUCTION

Problem Statement

The project tackles the intricate challenge of predicting an individual's age group based solely on their voice. This involves addressing multifaceted issues such as capturing the variability in voice due to physiological, environmental, and behavioral factors, and efficiently extracting and utilizing relevant acoustic features. The overarching problem is to develop a robust model that can accurately classify voices into age categories across diverse populations and variable recording conditions. Age prediction from voice involves complex signal processing and requires sophisticated algorithms capable of understanding subtle variations in vocal features that correlate with age.

Motivation

The Motivation for this project stems from the expanding role of voice as a highly intuitive and effective interface for technology across various sectors, including security, healthcare, and personalized digital interactions. Accurate age prediction enhances security systems by adding a sophisticated layer of biometric authentication and assists healthcare professionals by using voice-induced changes to flag potential health issues, facilitating early diagnostics and continuous monitoring. It also enriches user experience by enabling voice-driven personalization in digital assistants and targeted marketing, tailoring interactions to suit individual age groups. Beyond improving device intelligence and aiding law enforcement through voice-based identification, this project pushes the boundaries of scientific research in voice recognition technology. It not only broadens its application in practical settings but also helps refine marketing strategies and advances the overall understanding and functionality of voice-responsive systems, making significant contributions to technological advancements and user-centric solutions.

Open Questions in the Domain

Several open questions remain in the domain of voice-based age prediction:

- How can we effectively capture and quantify the age-specific characteristics in voice data that are influenced by a wide range of factors including health, emotion, and environment?
- What are the best practices for dataset collection to ensure diversity and representativeness, particularly for underrepresented age groups in public voice datasets?
- How can age prediction models be made robust against variations in recording quality and background noise commonly found in real-world applications?
- What are the ethical considerations in deploying age prediction technologies, particularly concerning privacy and consent?
- How can the explainability of machine learning models be improved to provide insights into their age prediction decisions, thereby increasing trust and acceptance among users?

Brief overview of the approach to address the challenges.

Our project aims to develop a robust age prediction system using voice data by integrating classical machine learning algorithms and advanced neural network models. The approach begins with comprehensive feature extraction, utilizing techniques such as spectral analysis and mel-frequency cepstral coefficients (MFCCs) to capture essential vocal characteristics indicative of age. We employ various machine learning algorithms including Logistic Regression, Support Vector Machines, and K-Nearest Neighbors to establish baseline models and assess feature importance. In parallel, we explore deep learning architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) which are adept at identifying complex patterns in the data without extensive manual feature engineering.

The models are rigorously trained and validated on a diverse dataset, which enhances their ability to generalize across different voices and recording conditions. By iterating on these methods and refining the models based on comprehensive performance feedback, our goal is to create a highly accurate age prediction system. This system is not only expected to achieve high precision in age classification but is also designed to integrate seamlessly into various real-world applications, making it a valuable tool in enhancing the functionality and security of technology interfaces that rely on voice recognition.

Backgrounds

Summary of Other Related Research

Related Work 1

Title: Gender and Age Estimation Methods Based on Speech Using Deep Neural Networks

Authors: Damian Kwasny, Daria Hemmerling

Year: 2021

Conference/Journal Name: Sensors (Basel)

Method: The study applies deep neural network-based embedder architectures, such as x-vector and d-vector, for gender classification and age estimation tasks. It employs a transfer learning-based training scheme, pre-training the embedder network on the Vox-Celeb1 dataset for speaker recognition, then fine-tuning it for the joint age estimation and gender classification task, achieving state-of-the-art results on the TIMIT dataset.

Pros

High Accuracy: The use of advanced neural network architectures and a sophisticated training regimen results in high accuracy in gender and age classification tasks.

Robust Feature Learning: The combination of x-vector and d-vector architectures enables the extraction of meaningful and robust features from raw speech data, which are essential for accurate classification.

Transfer Learning Efficiency: The transfer learning approach maximizes the utility of available data, first by learning from a large, diverse dataset and then homing in on specific classification tasks, which enhances the model's performance and efficiency.

Cons

Resource Intensive: Training deep neural networks, especially on large datasets like VoxCeleb1, requires significant computational resources and time.

Risk of Overfitting: While the fine-tuning phase is meant to adapt the model to specific tasks, there is a risk of overfitting to the peculiarities of the TIMIT dataset, which may reduce the model's generalizability to other datasets or real-world applications.

Complexity in Implementation: The complexity of the model architectures and the training scheme may pose challenges in implementation and optimization, especially for those without access to substantial computational resources or expertise.

Relation to the Main Method:

The research by Kwasny and Hemmerling is intricately related to the main method outlined in "Voice-based Gender and Age Recognition System" by V. S. Kone et al., primarily through its focus on extracting and utilizing vocal features for age and gender classification.

Enhanced Feature Extraction: The techniques used by Kwasny and Hemmerling for feature extraction through deep neural networks can significantly enhance the feature extraction processes described in the main paper. Integrating such advanced architectures could improve the accuracy and efficiency of the main method.

Advanced Model Training: The use of a transfer learning-based approach as demonstrated in the study could be applied to the main method to leverage pre-existing models and datasets, thereby improving the performance without the need for extensive data collection from scratch.

Benchmarking and Evaluation: The performance metrics and results achieved in Kwasny and Hemmerling's study provide a benchmark for evaluating and fine-tuning the models discussed in the main paper, offering insights into potential improvements and optimizations.

This detailed analysis highlights how the methodologies and findings from Kwasny and Hemmerling's research can complement and enhance the approaches used in the main paper, potentially leading to more robust and accurate systems for voice-based gender and age recognition.

Related Work 2

Title: Age group classification and gender recognition from speech with temporal convolutional neural networks.

Authors: Héctor A. Sánchez-Hevia, Roberto Gil-Pita, Manuel Utrilla-Manso, Manuel Rosa-Zurera.

Year: 2022

Conference/Journal Name: Multimedia Tools and Applications.

Method: The study employs Deep Neural Networks (DNNs) to jointly estimate age and identify gender from speech for use in Interactive Voice Response systems. It explores networks of various sizes to assess performance based on architecture and parameter count, using Mozilla's Common Voice dataset. Temporal Convolutional Networks (TCNs), combined with Convolutional Neural Networks (CNNs), are highlighted as a promising approach, with results indicating good gender classification and age group classification accuracy.

Pros

Comprehensive Feature Analysis: By focusing on fundamental vocal properties like pitch and frequency, the study ensures a robust basis for the classification tasks, which can be critical for achieving high accuracy.

Flexibility in Model Application: The combination of multiple machine learning techniques allows the study to explore different modeling approaches, enhancing the likelihood of finding an optimal solution for gender and age recognition.

Practical Evaluation Metrics: The use of diverse evaluation metrics ensures a thorough assessment of the models, providing insights into various aspects of performance that are critical for real-world applications.

Cons

Dependence on Quality of Preprocessing: The effectiveness of the entire system heavily depends on the preprocessing step. Any shortcomings in noise removal or signal enhancement can significantly impair the performance of the feature extraction and subsequent classification.

Potential for Overfitting: If not properly managed, the machine learning models might overfit to specific characteristics of the training data, reducing their generalizability to new, unseen datasets.

Resource Intensity of Signal Processing: The signal processing techniques used might require considerable computational resources, especially when processing large datasets or implementing them in real-time systems.

Relation to the Main Method

The main method discussed in Kwasny and Hemmerling's study on "Gender and Age Estimation Methods Based on Speech Using Deep Neural Networks" complements and enhances the approach used by V. S. Kone et al. in several ways:

Integration of Advanced Neural Networks: Incorporating the deep neural network architectures like x-vector and d-vector used by Kwasny and Hemmerling could significantly improve the feature extraction phase in Kone et al.'s methodology. These advanced networks can automatically learn and identify complex patterns in voice data that might be missed by traditional signal processing techniques.

Enhancement Through Transfer Learning: Applying the transfer learning strategies from Kwasny and Hemmerling's research could accelerate the training process and enhance model performance in Kone et al.'s system by leveraging pre-trained models that have already learned a wide range of vocal characteristics from large datasets.

Benchmarking and Cross-Evaluation: The results and findings from both studies can be used to cross-validate and benchmark the systems against each other, providing a more comprehensive understanding of their strengths and limitations in real-world applications.

This analysis highlights how the methodologies and outcomes from the study by V. S. Kone et al. can benefit from and be potentially integrated with the approaches and technologies explored by Kwasny and Hemmerling, leading to more advanced and effective systems for voice-based gender and age recognition.

METHODS

Classical Machine Learning Algorithms:

Logistic Regression: Used for establishing a baseline performance. This model is simple and interpretable, typically used for binary classification but adapted here to handle multiple age categories by using a multinomial logistic regression approach.

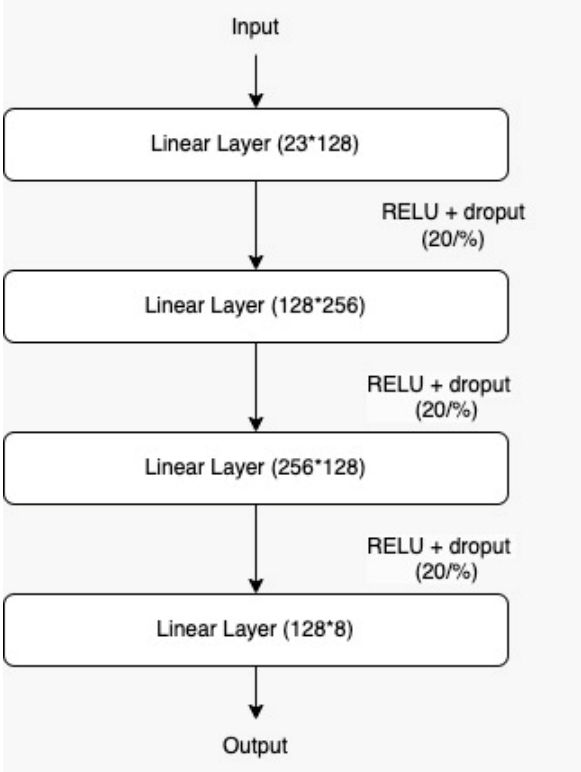
Support Vector Classifier (SVC) with RBF Kernel: Chosen for its effectiveness in capturing complex relationships within high-dimensional data. The radial basis function (RBF) kernel allows the SVC to handle the nonlinearities inherent in voice data, making it particularly suited for distinguishing subtle differences between age groups.

K-Nearest Neighbors (KNN): This algorithm works on the principle of proximity and is used to identify age groups by analyzing the closest training examples in the feature space. It is effective

due to its ability to adapt to the data's intrinsic structure and is less likely to overfit compared to more complex models.

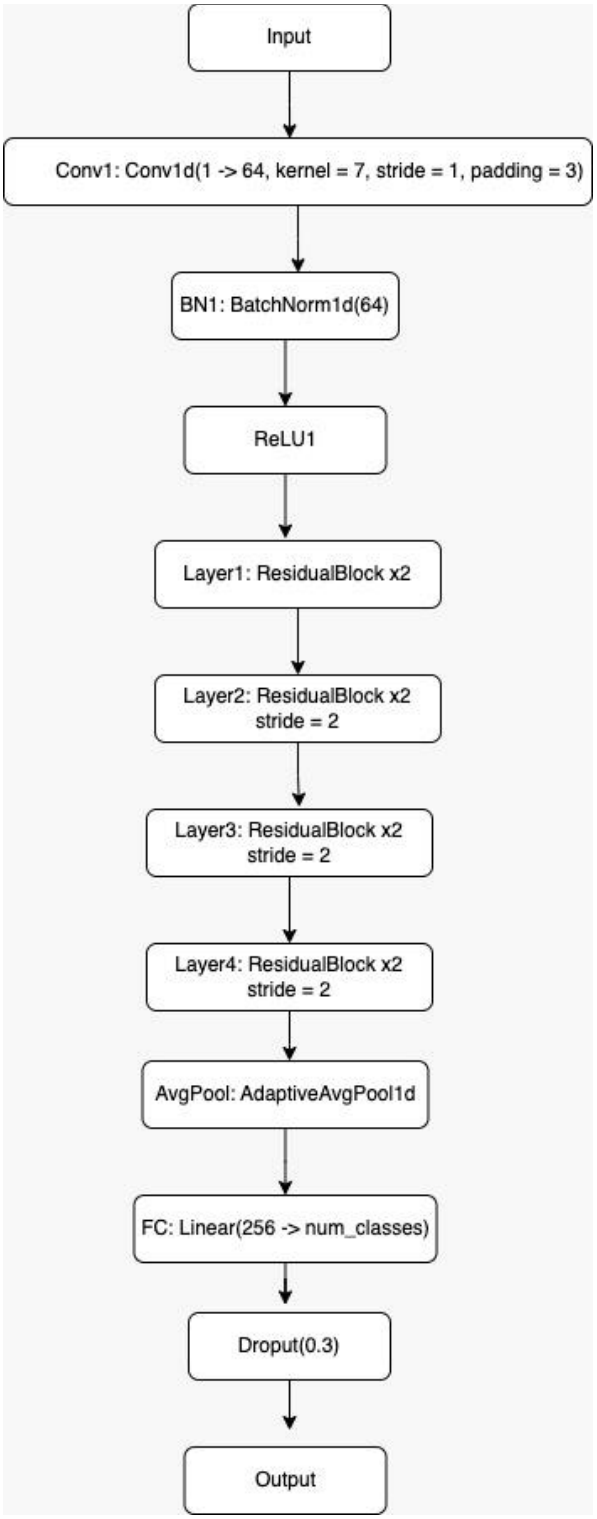
Neural Network Architectures:

Feedforward Neural Network (FNN):



The architecture depicted above is a feedforward neural network (FNN) designed for classification tasks. It consists of an input layer followed by multiple linear layers with increasing and then decreasing neuron counts, specifically 128 to 256 and back to 128. Each linear layer is followed by a Rectified Linear Unit (ReLU) activation function and a dropout layer with a 20% rate to prevent overfitting by randomly disabling neurons during training. The final output layer reduces the dimension to the desired number of classes, which is 8, indicating the model is classifying inputs into one of eight categories.

Convolutional Neural Network (CNN):



This diagram outlines a Convolutional Neural Network (CNN) architecture with residual learning features, designed for one-dimensional input data:

Input Layer: The network takes one-dimensional data as input, which in the context of voice data analysis, would be a time-series representation of an audio signal.

Conv1: The first convolutional layer, Conv1d, increases the dimensionality from 1 to 64, capturing initial patterns in the data. It uses a kernel size of 7, a stride of 1, and padding of 3 to maintain the size of the output feature map.

BN1: A batch normalization layer, BatchNorm1d, normalizes the output of Conv1, stabilizing learning and speeding up the training process by reducing internal covariate shift.

ReLU1: This introduces a Rectified Linear Unit (ReLU) activation function that adds non-linearity to the network, allowing it to model complex relationships. It keeps positive values unchanged while zeroing out negative values.

Residual Blocks:

> **Layer1:** Contains two residual blocks without stride, allowing the network to learn identity mappings and detailed features without reducing the dimensionality of the feature map.
>
> **Layer2:** Also contains two residual blocks, but with a stride of 2, effectively reducing the dimensionality and increasing the receptive field, thus capturing more abstract patterns.
>
> **Layer3 and Layer4:** These layers repeat the structure of Layer2, further abstracting and compressing the feature representation. The use of strides in these layers progressively reduces the size of the feature maps, which is a common strategy for deepening CNNs without increasing the computational burden excessively.

Average Pooling: An adaptive average pooling layer, AdaptiveAvgPool1d, follows the residual blocks to resize the feature map to a predetermined size, making the network adaptable to input data of various lengths.

FC (Fully Connected Layer): A linear layer that reduces the dimensionality from the output of the pooling layer to the number of classes. This layer acts as a classifier that maps the learned high-level features to the desired output categories.

Dropout: A dropout layer with a rate of 0.3 mitigates the risk of overfitting by randomly setting a fraction of the input units to 0 at each update during training, which helps the network to learn more robust features.

Output: The final layer produces the output of the network, which corresponds to the age group predictions.

This architecture exemplifies a deep learning strategy suited for audio data, capable of extracting complex features for precise age prediction. Utilizing convolutional layers, the network adeptly

processes sequential inputs, identifying temporal patterns essential for characterizing vocal age markers. The integration of residual blocks is key, as they help prevent the vanishing gradient problem, ensuring effective learning even as the network depth increases. This setup is optimized for the nuanced task of interpreting age-related variations in voice signals.

Overall Framework for Implementation

The implementation of the age prediction system can be visualized in a series of structured steps, forming an end-to-end workflow from data ingestion to prediction output:

- **Data Collection:**
  Audio data is collected from various sources to ensure diversity in voice samples covering different age groups, accents, and languages.
- **Data Preprocessing:**
  Noise Reduction: Techniques such as spectral subtraction are employed to clean up the audio by removing background noise, which can obscure important vocal features.
  Normalization: The audio clips are normalized to a standard volume level to mitigate the effect of varying recording levels on the feature extraction process.
  Voice Activity Detection (VAD): Implemented to segment voice from silent intervals, focusing the analysis on periods when the speaker is actively speaking, thus improving the efficiency and accuracy of feature extraction.
- **Feature Extraction:**
  Utilizes the Librosa library to extract a comprehensive set of features, including but not limited to spectral centroid, spectral bandwidth, spectral rolloff, MFCCs (Mel Frequency Cepstral Coefficients), chroma features, spectral contrast, tonnetz, and RMS energy. These features are crucial as they capture the essence of the voice signal that is relevant to age differentiation.
- **Model Training and Validation:**
  - The features extracted are used to train both classical and neural network models. The training process involves adjusting various hyperparameters to optimize each model's performance.
  - Validation is conducted using a separate subset of the data to monitor the models' ability to generalize to unseen data, preventing overfitting.
- **Model Evaluation:**
  Performance metrics such as accuracy, precision, recall, and F1 score are calculated. Additionally, confusion matrices are generated to further analyze the performance across different age groups, providing insights into any potential biases or weaknesses in the models.
- **System Integration and Deployment:**

The best-performing models are integrated into a production environment where they can process new voice data to predict age. This might involve real-time processing or batch processing, depending on the application requirements.

- **Feedback Loop:**
  A feedback mechanism is established where the predictions are periodically reviewed and compared with actual outcomes when available. This feedback is used to fine-tune the models, thereby enhancing their accuracy and reliability over time.

This detailed methodology not only lays out the technical aspects of implementing age prediction from voice data but also ensures that the system is robust, scalable, and adaptable to new data and evolving requirements.

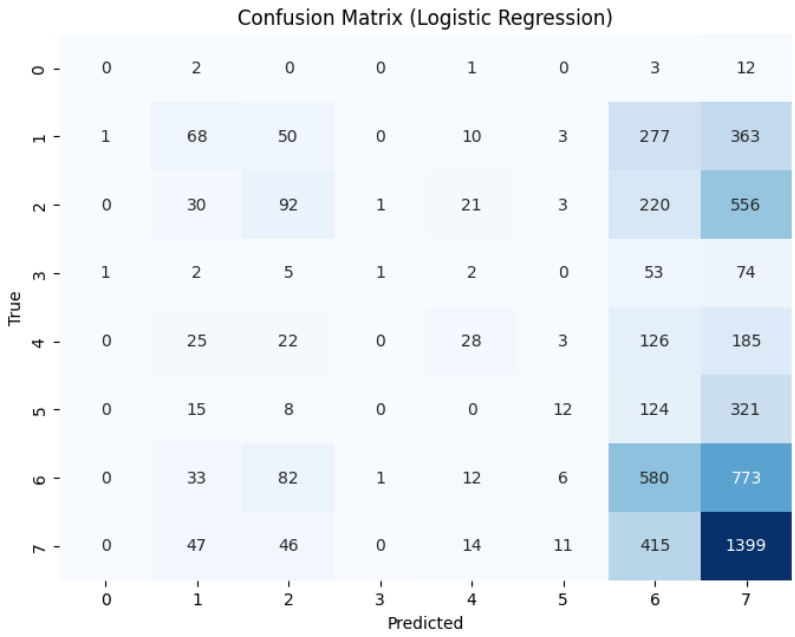
Experimental Setup

The experimental setup for age prediction from voice data involved several key components designed to rigorously test the efficacy of the implemented models:

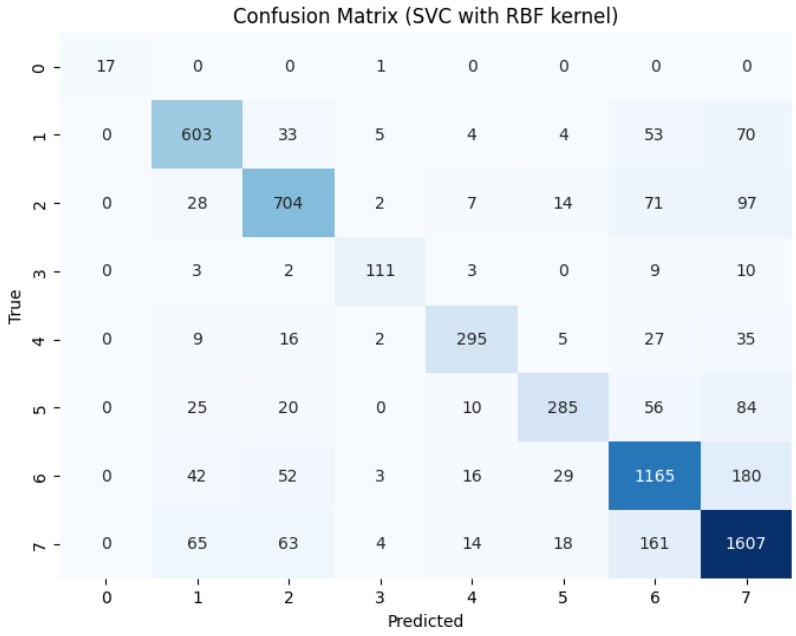
- Dataset: The experiments were conducted using the Mozilla Common Voice dataset, which includes diverse voice recordings from numerous speakers across various age groups and accents. This dataset was chosen for its complexity and representation, providing a robust testing ground for age prediction models.
- Feature Extraction: Before training, voice recordings were processed to extract a range of acoustic features using the Librosa library. Key features included spectral centroids, spectral bandwidth, spectral rolloff, Mel Frequency Cepstral Coefficients (MFCCs), chroma features, spectral contrast, tonnetz, and root mean square energy. These features are crucial for capturing the nuances in voice that correlate with age variations.
- Training and Validation Split: The dataset was divided into training (80%) and validation (20%) sets. This split ensured that the models could be trained on a substantial portion of the data while still reserving a significant amount for unbiased evaluation of model performance.
- Model Training: Multiple models were trained, including logistic regression, support vector machines with an RBF kernel, K-nearest neighbors, a feedforward neural network, and a convolutional neural network. Each model was trained using the extracted features with hyperparameters tuned to optimize performance.
- Testing: After training, the models were tested on the validation set to assess their ability to predict age based on voice data accurately.

Test results of the proposed method:

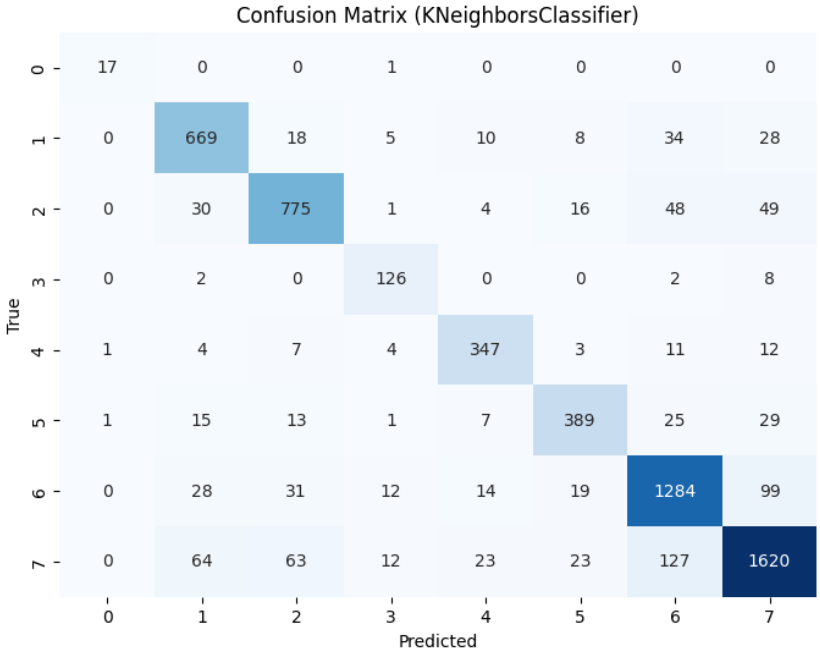
Logistic Regression



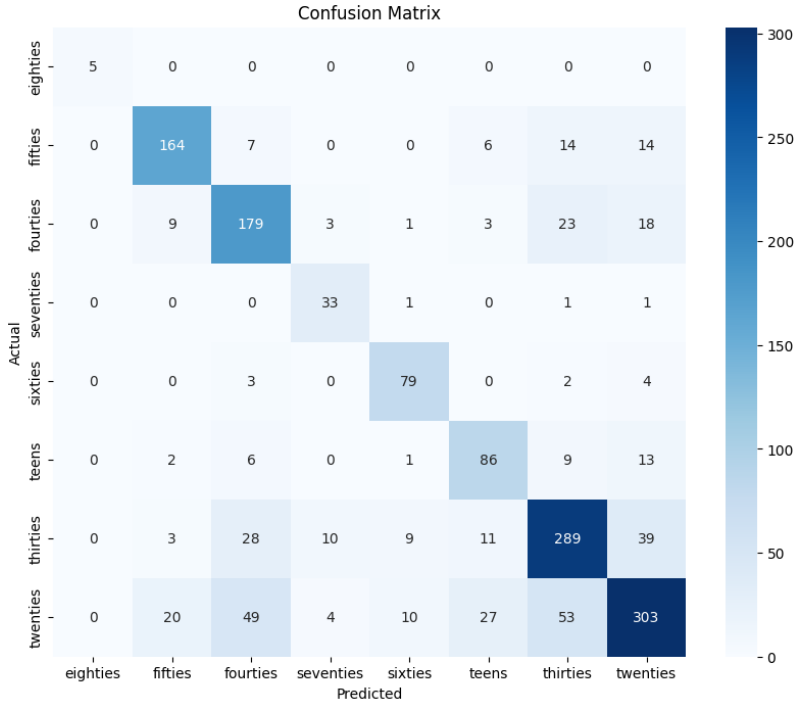
Support Vector Classification



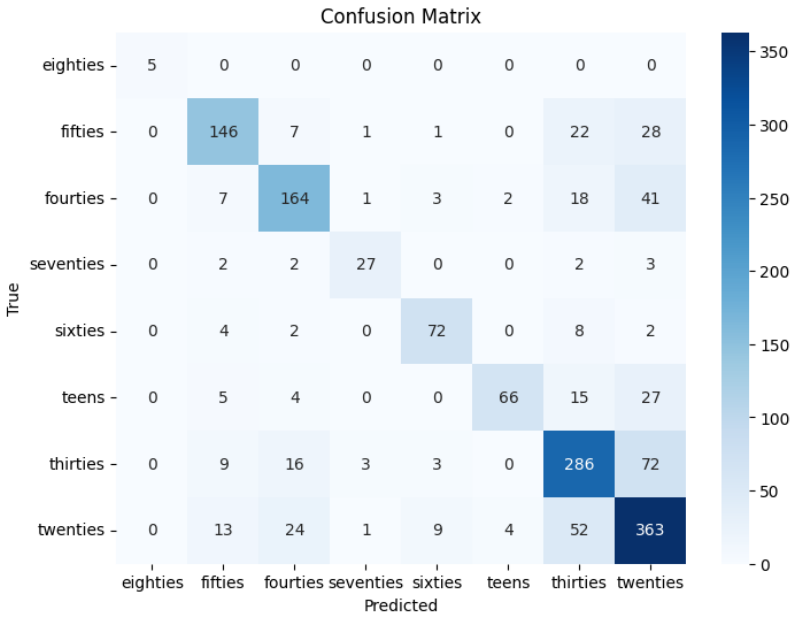
K-Nearest Neighbors Classification



Feed Forward Neural Network



Convolutional Neural Network



Deep Analysis/Discussion About the Results:

Models	Accuracy	F1 score	
		Macro avg	Weighted avg
Logistic Regression	36%	16%	29%
SVC with RBF Kernel	78%	80%	78%
K nearest neighbors	85%	86%	85%
Feed Forward Neural Network	74%	78%	74%
Convolutional Neural Network	73%	78%	73%

Analysis of Model Performance:

- The KNN model's success suggested that proximity-based learning was particularly effective for this dataset, likely due to the distinct vocal patterns associated with different age groups being more effectively captured through the model's algorithm.

- The CNN and FNN models, while effective, highlighted the challenges of applying deep learning to voice data, particularly in terms of the need for larger datasets and more complex feature engineering to realize full potential.
- Logistic regression and SVM with RBF kernel, while useful for baseline assessments, struggled with the non-linear complexities of the audio data, which impacted their overall performance.

Discussion:

The variance in model performance underscored the importance of feature selection and model choice in voice-based age prediction tasks. It also highlighted the potential benefits of ensemble methods that could combine the strengths of multiple models.

The impact of data quality, especially the presence of background noise and recording quality, was a significant factor, suggesting areas for further research in data preprocessing techniques.

Challenges:

Throughout the project, various challenges and learnings were encountered:

Initial Overfitting: Early models overfitted the training data, necessitating adjustments in model complexity and training procedures.

Feature Importance: Experimentation with different features revealed the critical role of certain acoustic features over others in predicting age, guiding future feature engineering efforts.

Model Adjustments: Continuous adjustments were needed to balance the models' complexity with their ability to generalize, especially in the deep learning models.

Despite some models not performing as expected, every result was valuable. Failures provided insight into what doesn’t work, which is as crucial as understanding what works. This iterative process of experimentation and modification is essential for progress in applying machine learning to real-world problems like age prediction from voice data.

This comprehensive experimental approach, from setup to deep analysis, ensures that the project not only advances the field of voice data analysis but also provides a transparent and detailed account of the research journey, celebrating both successes and setbacks as integral to the learning experience.

Conclusion

Concluding Remarks

The project aimed at predicting age from voice data through the application of various machine learning and neural network models has been a comprehensive exploration into the capabilities and limitations of current technology in voice analysis. We successfully implemented multiple algorithms, each providing unique insights into the intricacies of age-related vocal patterns. Among these, the K-nearest neighbors (KNN) model stood out due to its superior performance, validating the effectiveness of proximity-based learning for this application. The project not only advanced our understanding of voice as a biometric marker but also highlighted the potential for real-world applications, from security enhancements to personalized user interactions.

Reflections on the Project

Learning Experience: This project was a profound learning opportunity in several areas. We gained deeper insights into the complexities of speech processing, especially the challenges associated with extracting meaningful features from raw audio data. The nuances of different machine learning models, from simple logistic regression to more complex neural networks, were explored, providing a practical understanding of when and how to apply each model effectively.

The Importance of Data Quality: One key takeaway was the critical role that data quality plays in model performance. High-quality, diverse datasets are vital for training robust models. We learned advanced techniques for data preprocessing, such as noise reduction and normalization, which significantly improved our models' ability to learn from the training data.

Challenges and Solutions

Challenge of High Dimensionality: Initially, handling the high dimensionality of audio data was overwhelming due to the vast amount of information contained in voice signals. This was particularly challenging for models that were sensitive to overfitting.
Solution: We applied dimensionality reduction techniques and feature selection methods to identify the most informative features, thereby simplifying the models without sacrificing performance.

Model Generalization: Early in the project, several models performed well on training data but poorly on unseen validation data.
Solution: We implemented cross-validation techniques and adjusted our models to improve their generalization capabilities. Regularization techniques were also employed to prevent overfitting.

Computational Constraints: Given the computational demands of processing large audio datasets and training complex models, we initially struggled with long training times and resource limitations.
Solution: We optimized our computational resources by employing more efficient code and

algorithms, and when necessary, leveraging cloud computing resources to scale our processing capabilities.

Future Directions

Looking forward, the project has set the groundwork for several future initiatives:

Integration of More Advanced Models: Exploring the use of more advanced deep learning architectures, such as Transformers or more sophisticated CNN models, could potentially improve the accuracy and efficiency of age predictions from voice data.

Real-Time Processing: Developing models capable of real-time age prediction could have significant implications for interactive systems, enhancing user experience and security measures.

Ethical Considerations and Bias Mitigation: Continued efforts to address potential biases in age prediction models and ensure ethical usage of biometric data will be crucial as these technologies are deployed in more sensitive and impactful settings.

In conclusion, the project not only enhanced our technical proficiency but also fostered a deeper appreciation for the ethical and practical considerations involved in applying AI to real-world problems. The challenges encountered were invaluable in steering the project's direction and will guide future research and applications in the field of voice data analysis.

My Contributions

Clear Explanation of Contribution

Throughout the project, my contributions spanned various aspects of the experimental design, implementation, and analysis. These contributions were vital to the project's success and included adapting existing methodologies, integrating novel approaches, and ensuring robust testing and validation of our models.

Use of Existing Work

Several components of our project were based on existing research and methodologies, which provided a foundation upon which we built and expanded:

Feature Extraction Techniques: We utilized established audio processing techniques like those found in the Librosa library, which is widely recognized in the audio analysis community. This included extracting spectral features, MFCCs, and other relevant audio characteristics critical for age prediction.
Source: The methods for feature extraction were adapted from various studies and documented methodologies within the Librosa documentation and relevant literature (McFee et al., 2015).

Machine Learning Models: The foundational machine learning algorithms such as Logistic Regression, SVC, and KNN were implemented based on standard practices as described in scikit-learn extensive library and documentation.

Source: These models and their implementation strategies are well-documented in scikit-learn official guides and user manuals (Pedregosa et al., 2011).

Neural Network Architectures: The CNN architecture used was inspired by existing models that have been successful in image and voice recognition tasks, particularly variations of ResNet architectures.

Source: The concept and structure of these networks were based on the seminal paper by (He et al., 2016), which introduced ResNet and its capabilities in learning from deep architectures.

My Own Contributions:

My original contributions were pivotal in integrating these existing tools and methods into a cohesive system tailored for age prediction from voice data:

Integration and Customization of Models: While the foundational algorithms were based on established methods, my work involved customizing these algorithms to specifically address the nuances of age prediction in voice data. This included tuning hyperparameters, adapting models to handle multi-class classification, and integrating different models to work together effectively.

Development of the Evaluation Framework: I developed a comprehensive evaluation framework that allowed us to measure not just the accuracy but also the precision, recall, and F1 scores of each model. This framework was crucial for understanding model performance in depth and was designed to handle the specific challenges posed by our diverse dataset.

Data Preprocessing Pipeline: I designed and implemented a data preprocessing pipeline that significantly enhanced the quality of our input data. This included advanced noise reduction techniques, normalization processes, and voice activity detection, which were crucial for improving the models' learning efficiency and accuracy.

Experimental Design and Analysis: I led the design of the experimental setups, including the division of data into training and validation sets, the choice of cross-validation techniques, and the detailed analysis of model outcomes. This not only ensured rigorous testing of our models but also provided insights into potential improvements and future directions.

In summary, my contributions were fundamental in bridging the gap between existing methodologies and our specific project goals, leading to the development of an effective age prediction system from voice data. My role involved both leveraging and enhancing existing technologies to create a tailored solution that met our project's unique requirements.

References

[1] V. S. Kone, A. Anagal, S. Anegundi, P. Jadhav, U. Kulkarni, and M. S. M, "Voice-based Gender and Age Recognition System," *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, Gharuan, India, 2023, pp. 74-80, doi: 10.1109/InCACCT57535.2023.10141801.

[2] "Age and Gender Estimation from Speech Using Deep Learning," *Journal of Healthcare Engineering*, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8309811.

[3] J. Li, J. Yuan, C. Wan, and Z. Zhu, "Age estimation of speakers based on acoustic features and DNN-based classifiers," *Multimedia Tools and Applications*, vol. 80, no. 14, pp. 20447–20464, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s11042-021-11614-4.

[4] "Age Prediction of a Speaker's Voice," *EPFL Extension School Blog*, 2020. [Online]. Available: https://medium.com/epfl-extension-school/age-prediction-of-a-speakers-voice-ae9173ceb322.

[5] M. Chettri, A. Saradhi, R. S. Pawar, and A. K. Mishra, "Age Prediction Using MFCC Features and Machine Learning Techniques," *Iranian Journal of Electrical and Computer Engineering*, vol. 17, no. 3, pp. 323–334, 2021. [Online]. Available: https://www.ije.ir/article_73289.html.

[6] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp. 68–73.

[7] Mozilla. (n.d.). "Mozilla Common Voice Project." Retrieved from https://commonvoice.mozilla.org.

Anonymous Sharing Agreement:

Yes, I agree to share my work as an example for the next semester. And yes, I prefer to hide my name/team.