Final Project Report

Introduction to Machine Learning

ECGR 5105

HARSHATH MANCHIKA

Student ID: 801393748

GITHUB LINK:

https://github.com/harshath2000/Into_to_ML_FinalProject/tree/main

Age Prediction from the Voice data

Github Link:
https://github.com/harshath2000/Into_to_ML_FinalProject/tree/main

***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.)

Keywords—Logistic Regression, Convolutional Neural Networks, KNN, SVC .

I. INTRODUCTION

In the evolving landscape of data science and artificial intelligence, the analysis of voice data has emerged as a captivating frontier with broad-ranging implications. Our project is situated at the intersection of technological innovation and human interaction, focusing on the intriguing challenge of predicting age groups through a meticulous examination of audio features. Beyond the realm of age prediction, the potential applications of our research extend into practical domains, including security systems, crime-solving, and the optimization of AI-based assistants.

Voice, as a unique biometric identifier, harbours a wealth of information that extends beyond mere communication. The motivation behind our project is rooted in the belief that deciphering age-related patterns within voice data can unlock practical solutions with transformative impacts. Consider the implementation of a security system that not only authenticates based on voice but also gauges the likely age range of the speaker, adding an additional layer of verification.

In the context of crime-solving, our work is driven by the vision of forensic experts using voice recordings to estimate the age of an unknown speaker, providing valuable insights that could narrow down suspect lists. Moreover, envision AI-based virtual assistants that dynamically adapt their responses and interactions based on the inferred age of the user, creating a more personalized and intuitive user experience.

Our project is motivated by the prospect of translating theoretical advancements in voice data analysis into tangible, real-world applications. We seek to bridge the gap between cutting-edge research and practical implementations, recognizing that age prediction in voice data is not merely an academic pursuit but a versatile tool with the potential to address a spectrum of challenges.

The primary objective of our research is to extract salient features from voice data, enabling accurate age prediction. By employing a dual approach involving classical machine learning and neural networks, we aim to uncover subtle relationships within the audio features that contribute to age-related distinctions. Through this exploration, we aspire to not only advance the field of predictive modelling but also to pioneer innovative solutions that enhance security, contribute to law enforcement efforts, and optimize user interactions in the realm of artificial intelligence.

In essence, our work is driven by the conviction that the marriage of voice data analysis and age prediction can yield transformative outcomes, ushering in a new era of applications that integrate the intricacies of human expression with the precision of advanced machine learning techniques.

II. Approach

Our approach to age group prediction through voice data integrates both classical machine learning (ML) and neural network methodologies. For classical ML, we employ three distinct algorithms: Logistic Regression, Support Vector Classifier (SVC) with a radial basis function (RBF) kernel, and k-Nearest Neighbors (KNN).

A. Logistic Regression:

We leverage the logistic regression model to establish a baseline for age prediction. This algorithm is known for its simplicity and interpretability. We train the model on our extracted audio features and evaluate its performance using metrics such as accuracy, classification report, and confusion matrix.

B. Support Vector Classifier (SVC) with RBF Kernel:

SVC with an RBF kernel is employed to capture complex relationships within the data. The non-linear nature of the RBF kernel allows the model to discern intricate patterns that may be challenging for linear classifiers. Similar to logistic regression, we assess the model's accuracy, classification report, and confusion matrix.

C. K-Nearest Neighbors (KNN):

KNN is utilized for its simplicity and ability to capture local patterns in the data. By choosing the number of neighbors, we tailor the model's sensitivity to local variations. We evaluate the KNN model's performance using accuracy, classification report, and confusion matrix.

For the neural network approach, we implement both a Feedforward Neural Network (FNN) and a Convolutional Neural Network (CNN).

D .Feedforward Neural Network (FNN):

The FNN architecture comprises multiple fully connected layers with rectified linear unit (ReLU) activation functions. Dropout layers are incorporated to mitigate overfitting. The FNN is trained to learn complex relationships between input audio features and age groups, and its performance is evaluated using standard classification metrics.

E.Convolutional Neural Network (CNN):

The CNN architecture, designed specifically for one-dimensional audio data, utilizes residual blocks to capture hierarchical features. The network is trained to automatically extract relevant patterns from the input spectrogram. Similar to FNN, the CNN's performance is assessed through metrics such as accuracy and classification report.

In summary, our approach involves a comprehensive exploration of classical ML and neural network models to predict age groups based on voice data. Through a rigorous evaluation of these methodologies, we aim to discern which approach proves most effective for this challenging task.

III. Datasets and training setup

A. Data Source:

The dataset utilized in this project is sourced from Kaggle (www.kaggle.com), specifically the Common Voice dataset provided by Mozilla. The choice of this dataset is motivated by its ease of accessibility through various open-source platforms. Its size, approximately 2 GB, is deemed suitable for training on standard personal computers while providing sufficient data to train the model effectively for age prediction.

B. Dataset Characteristics:

The Common Voice dataset comprises diverse voice recordings from individuals spanning different age groups, making it a rich and varied source for both training and evaluating the machine learning model. The dataset includes features such as "spectral centroid," "spectral bandwidth," "spectral_rolloff," "mfcc1" to "mfcc20," "Chroma Feature," "Spectral Contrast," "Tonnetz," and "RMS Energy." These features are pivotal for capturing essential characteristics of the audio signals, enabling the model to discern age-related patterns effectively.

C. Feature Extraction:

To extract the features from the voice recordings, the Librosa Python audio analysis library is employed. The features are extracted using a dedicated function that encompasses spectral features like spectral centroid, spectral bandwidth, and spectral rolloff, as well as mel-frequency cepstral coefficients (MFCCs), chroma features, spectral contrast, tonnetz, and root mean square energy (RMSE).

D. Classical ML Setup:

For classical machine learning, the dataset is preprocessed by encoding the labels using the LabelEncoder from scikit-learn. This is applied separately for both the training and testing datasets. The labels are transformed into numerical values to facilitate the training process. Logistic Regression, Support Vector Classification (SVC), and k-Nearest Neighbors (KNN) are employed as classical ML algorithms. The labels are encoded using the LabelEncoder, ensuring compatibility with the classification algorithms.

E. Neural Network Setup:

In the case of neural networks, the labels are encoded using the LabelEncoder, and the data is split into training and validation sets using the train_test_split function. The number of classes and input shape are determined, and the data is transformed into PyTorch tensors. The training dataset is then organized into a DataLoader for efficient training.

F. Training Parameters:

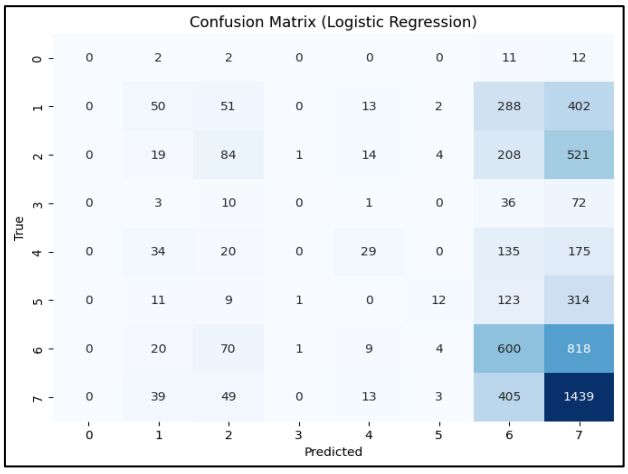
For both classical ML and neural network approaches, rigorous evaluation metrics, including accuracy, precision, recall, and F1 score, are employed to assess the model's performance. The datasets are split into training and validation sets to ensure the model's ability to generalize to unseen data and prevent overfitting. These datasets are then utilized to train and evaluate the models, ultimately providing insights into the efficacy of each approach for age group prediction through voice data analysis.

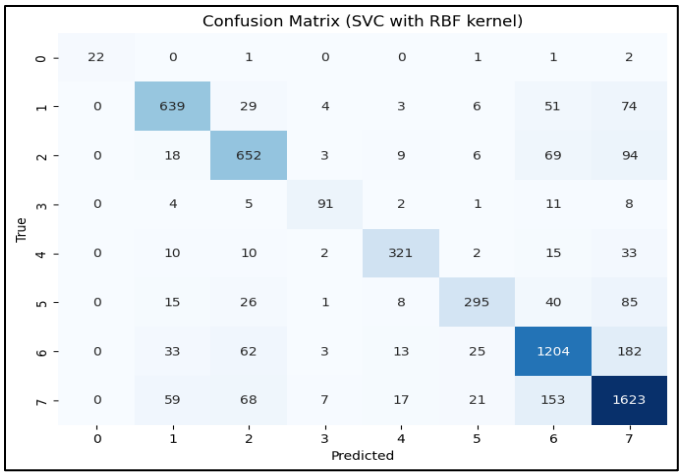
IV. Results and Analysis

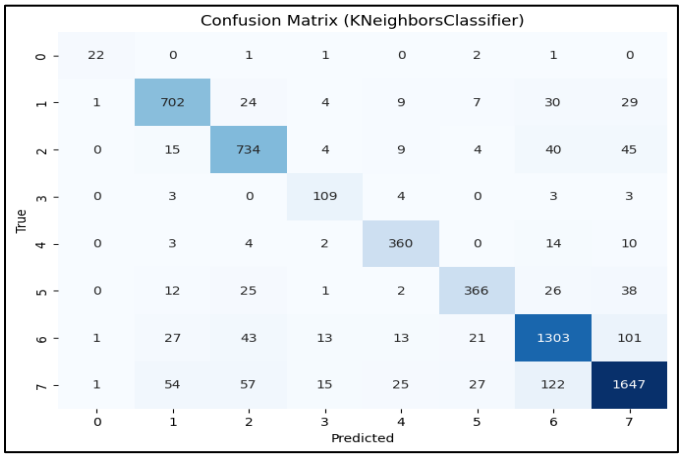
Models	Accuracy	F1 score	
		Macro avg	Weighted avg
Logistic Regression	36%	16%	29%
SVC with RBF Kernel	78%	80%	79%
K nearest neighbors	85%	85%	85%
Feed Forward Neural Network	72.70%	75%	73%
Convolutional Neural Network	73%	78%	73%

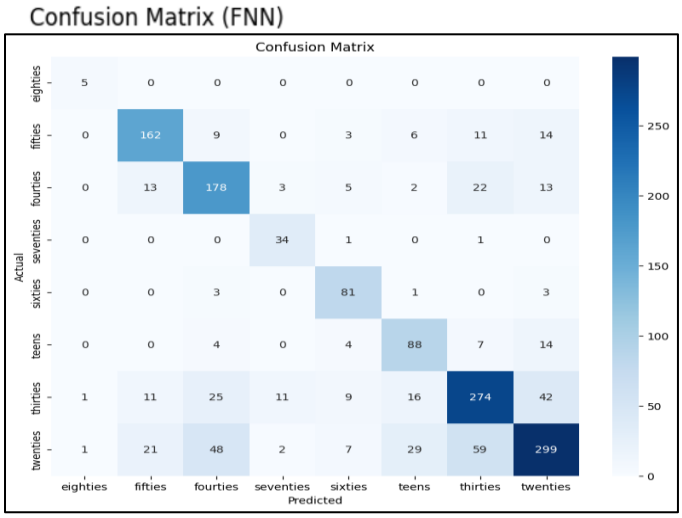
The table compares the performance of five different machine learning models, with the K nearest neighbors model outperforming others in accuracy and F1 score. Logistic Regression exhibits the lowest performance metrics across the board.

Confusion Matrices:

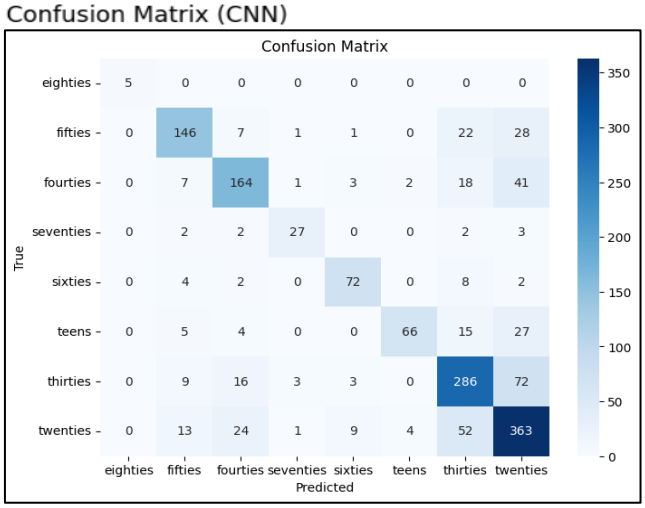








Confusion Matrix (CNN)



The confusion matrices indicate that the K Nearest Neighbors model outperforms others with the highest accuracy, especially for distinct classes, while Logistic Regression struggles the most with classification. Both neural network models, FNN and CNN, show competent classification with some inter-class confusion, performing better than Logistic Regression

V. Lessons Learned

The quality and diversity of the voice dataset are paramount for training robust age prediction models, as they must generalize across various speech patterns, accents, and intonations associated with different ages. Model selection is critical, while traditional models like Logistic Regression may falter on complex datasets, more sophisticated models like KNN and SVM with RBF kernel demonstrate superior performance. Neural networks (FNN and CNN) can capture non-linear patterns effectively, but their performance is heavily reliant on the architecture and hyperparameters used. Evaluation metrics and confusion matrices provide invaluable insights into specific areas where models excel or fail, guiding further improvements and iterations in the model development process.

CONCLUSION: This research project provides a comprehensive analysis of age group prediction using voice data through classical machine learning (ML) methods and neural network approaches. The study aimed to identify and compare the effectiveness of various methodologies in predicting age from audio features.

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

[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8309811

[3] https://link.springer.com/article/10.1007/s11042-021-11614-4

[4] https://medium.com/epfl-extension-school/age-prediction-of-a-speakers-voice-ae9173ceb322

[5] https://www.ije.ir/article_73289.html

[6] https://doi.org/10.1044/jshr.0902.