Project Work :

Final ISA(Review 4) / ESA 2020

Project Title	:	Voice based Age detection and recommending song based on age group .	
Project ID	:	PW20KKN01	
Project Guide	:	Prof. Kavitha KN	
Project Team	:	Chandhan Bhatta	01FB16ECS094
		D Ganesh Kumar	01FB16ECS102
		G Gautham	01FB16ECS118

Team Member:

Name: D Ganesh Kumar
SRN: 01FB16ECS102

Name: Chandhan Bhatta
SRN: 01FB16ECS094

Name: G Gautham
SRN: 01FB16ECS118

Guide Details:

Guide Name: Prof. Kavitha KN
Designation: **Assistant Professor**

Problem Statement

- Age detection from Voice : Since most of the technologies now are going ahead with voice based models it would be efficient and accurate to detect Age from the voice Sample.
- Song suggestion: suggest songs based on the age group the belong .

User Profile

- Age detection can be used in devices that use voice controlled tools like Apple's Siri, Amazon's Alexa etc.
- Music enthusiast
- Music scientists researchers
- Call centers

Literature Survey

Title

- Convolutional-recurrent Neural Network for Age and Gender Prediction from Speech
- IEEE – sept 2019
- Héctor A. Sánchez-Hevia , Roberto Gil-Pita, Manuel Utrilla-Manso

Datasets

- Mozilla's database "Common Voice" : 12 GB of labelled data

Accuracy

- Using CRNN – only one stage : 78.53%
- Using CRNN + GRU – two stage : 79.51%

Literature Survey

Accuracy

- Using CRNN –only one stage :

Gender classification error is 2.09%

- Using CRNN + GRU – two stage :

Gender classification error is 1.86%

Constraints

One of the main difficulties implementing this kind of systems is access to a dataset that could be used for training and testing the classifiers. Datasets of utterances sometimes include labels about gender, but rarely include information of speaker's age.

Source

- https://ieeexplore.ieee.org/document/8881961

Literature Survey

Title

- Voice-based Age and Gender Recognition Based on Learning Generative Sparse Models
- International Journal of Engineering – sept 2018
- S. Mavaddati

Datasets

- Generated dataset using 200 volunteers

Accuracy

- Using svm + GMM + SNFM: 73.89%

Literature Survey

Constraints

Performed under lab conditions

Source

- http://www.ije.ir/article_73289_c2e5f510617f3ef6a6a325f1582f0843.pdf

Literature Survey

Title

- End-to-End Deep Neural Network Age Estimation
- Interspeech -2018
- Pegah Ghahremani, Phani Sankar Nidadavolu, Nanxin Chen, Jesús Villalba, Daniel Povey, Sanjeev Khudanpur, Najim Dehak

Datasets

- NIST SRE08-10 dataset

Accuracy

- Using i-Vector : 84%
- Using x-Vector end-to-end : 88%

Literature Survey

Constraints

When training the network, we randomly select speech chunks from the full length training recordings. Those chunks can have a fixed duration or we can also sample chunk with random durations.

Source

- https://www.danielpovey.com/files/2018_interspeech_age_estimation.pdf

Further Literature Survey

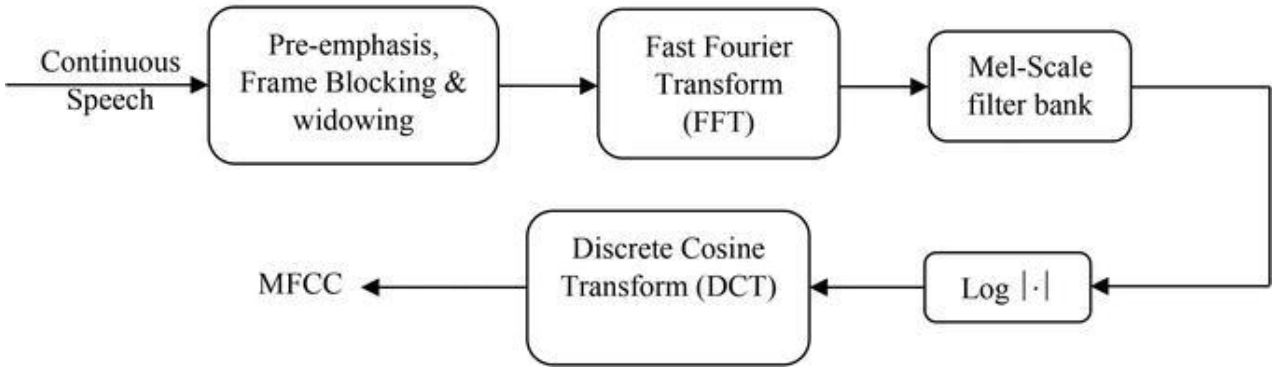
Title

- A Review on Feature Extraction and Noise Reduction Technique
- IJARCSSE -2014
- S. B. Magre, P. V. Janse, R. R. Deshmukh

This paper provides outline various feature extraction and noise reduction technique. This paper helps in selecting the technique beside on their relative advantages & disadvantages. This paper is concludes with the choice on feature direction for developing technique in human pc interface system.

Further Literature Survey

MFCC: Mel-Frequency Cepstral Coefficients



Disadvantages:

MFCC values are not very robust in the presence of additive noise, and so it is common to normalize their values in speech recognition systems to lessen the influence of noise.

Proposed Solution

- Web application where User has to login to use it, user then is prompted to record voice , this audio file is used for age prediction.
- Age is predicted into classes using multilayered perceptron model which consist of pyaudioanalysis tool which helps extract features from the audio file.
- Song is suggested based on the age group that is predicted by the model.
- The list of song suggested are displayed to the user in web app.

Why Your Solution is Better?

- We are providing a web UI which will visually help user.
- User can play the song which he is suggested.
- Age detection based on voice is accurate and effective.
- Fast and efficient method to detect the age group of user

Technologies / Methodologies

Technologies Used

- Flask , HTML, JS for Web UI
- Python , jupyter, pyaudioanalysis for preprocessing
- Colab , tensorflow , keras for model building

Dependencies and Risks

Dependencies:

- Dataset used for age detection was common voice from Mozilla dataset is very huge , had to make batches for preprocessing and feature extraction , takes a lot of time to preprocess.
- Dataset we used for age detection cannot be used for detection of precise age of a person, age can only be detected in groups like teens, twenties, thirties,etc.
- Used python tool called pyaudioanalysis which helps in preprocessing and feature extraction , performs operation on single file at a time .
- Model is built on google colab ,the reason being less computing power of our machine , colab provides Virtual environment to develop model.

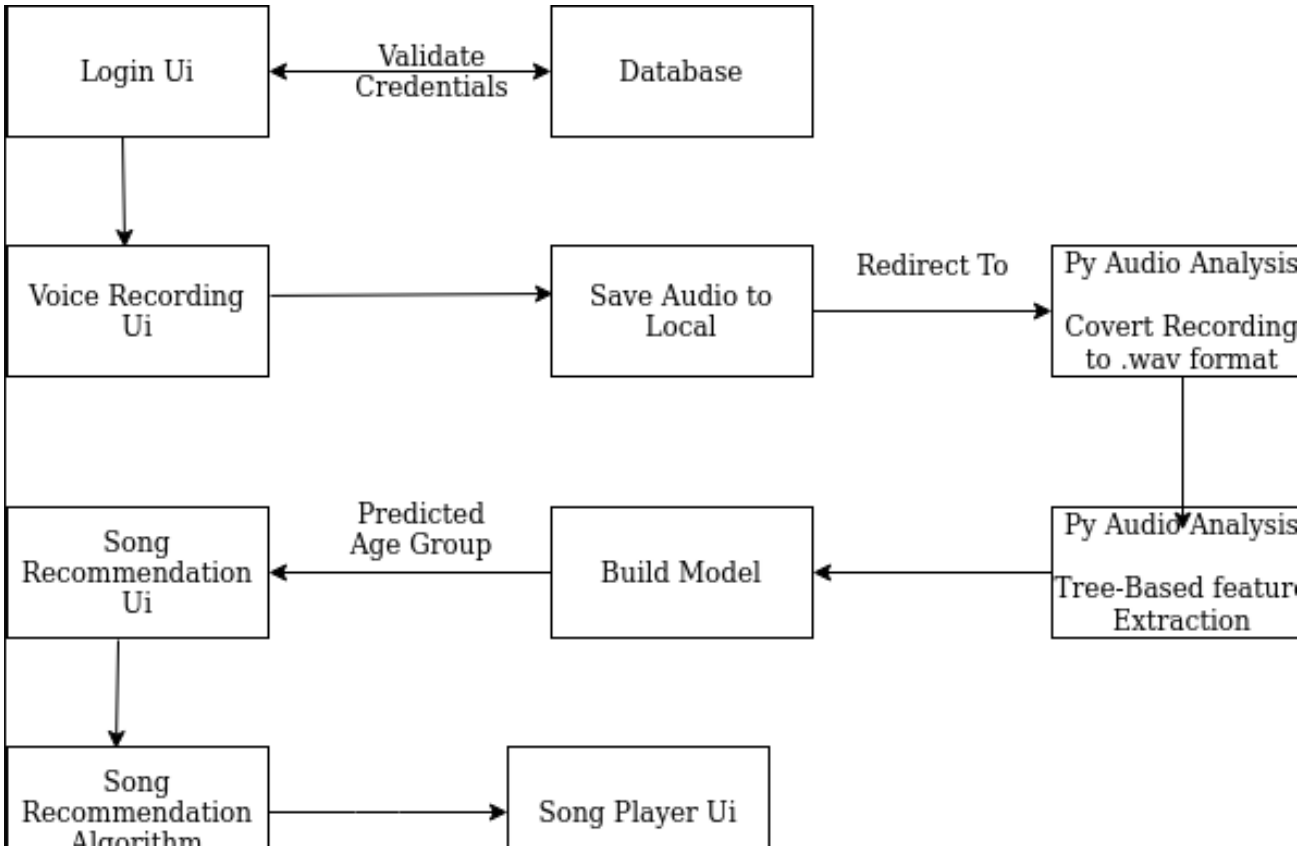
Risks:

- Might not provide accurate results if user has too much noise in recording mic .
- User might not like song suggested to him.

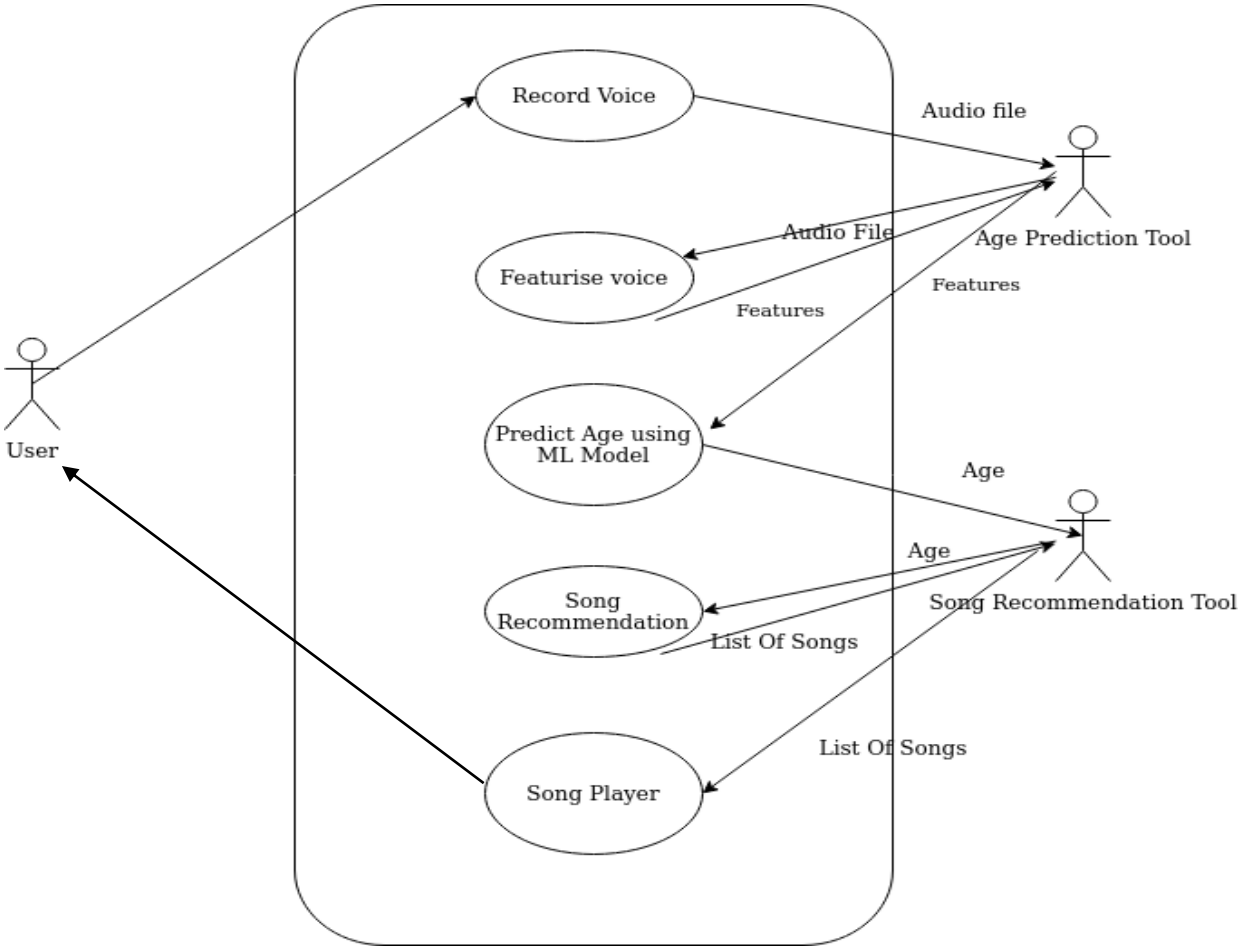
PES UNIVERSITY




System Architecture



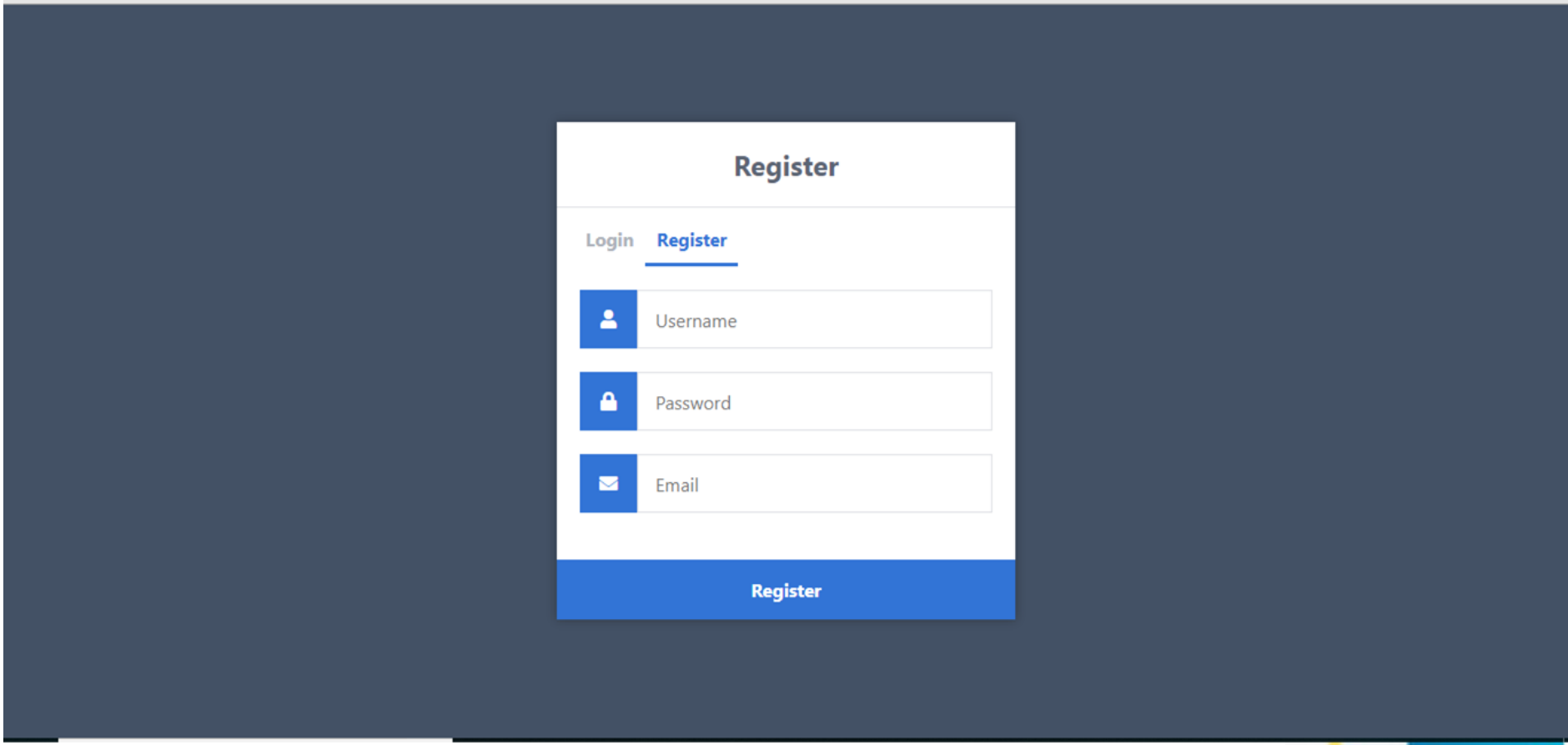
UI/ Use Case


Record Voice
Audio file
Age Prediction Tool
Audio File
Featurise voice
Features
Features
Predict Age using ML Model
User
Age
Age
Song Recommendation
Song Recommendation Tool
List Of Songs
List Of Songs
Song Player

PES UNIVERSITY

UI/ Use Case

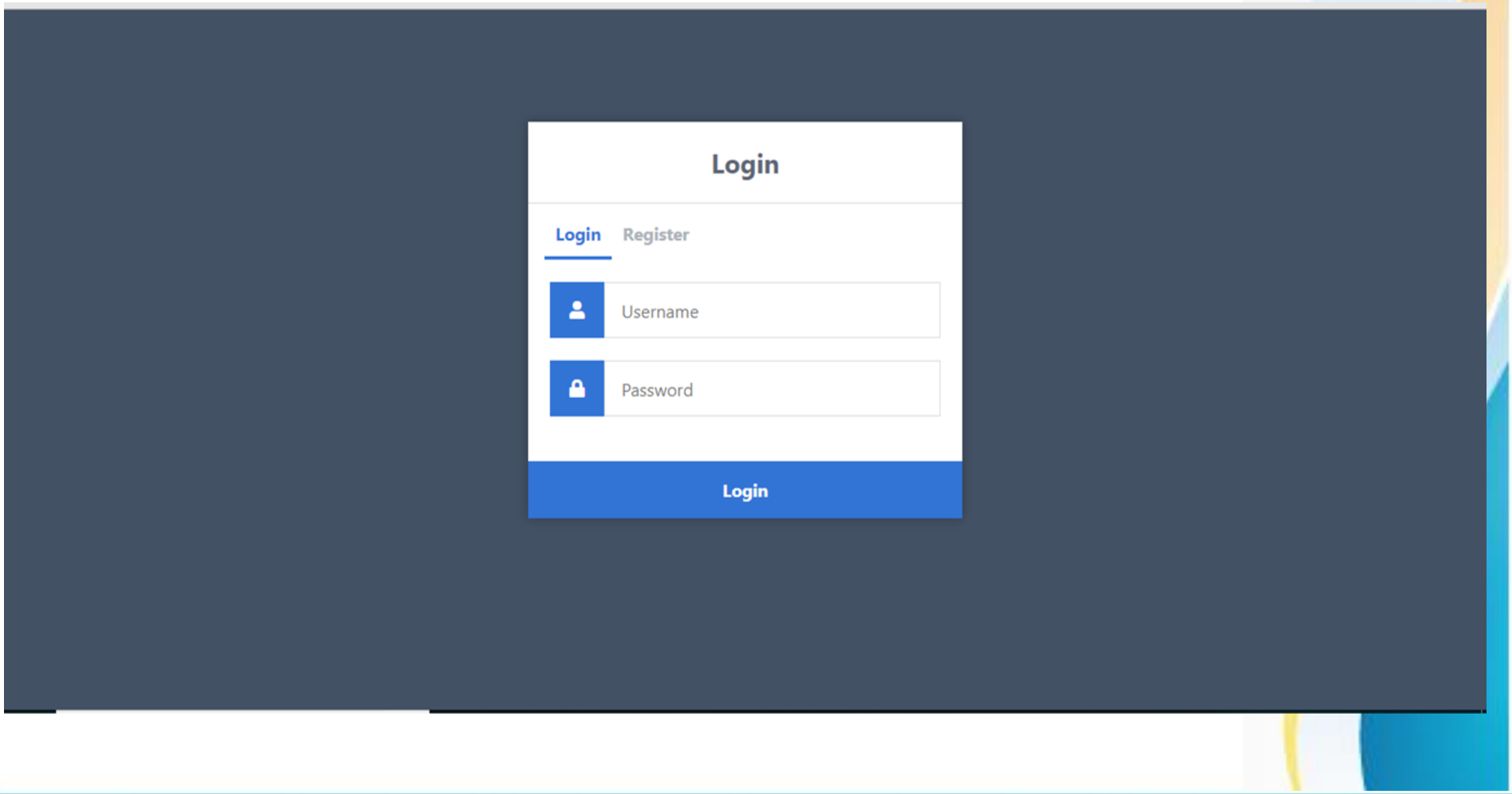
Register Page:



PES UNIVERSITY

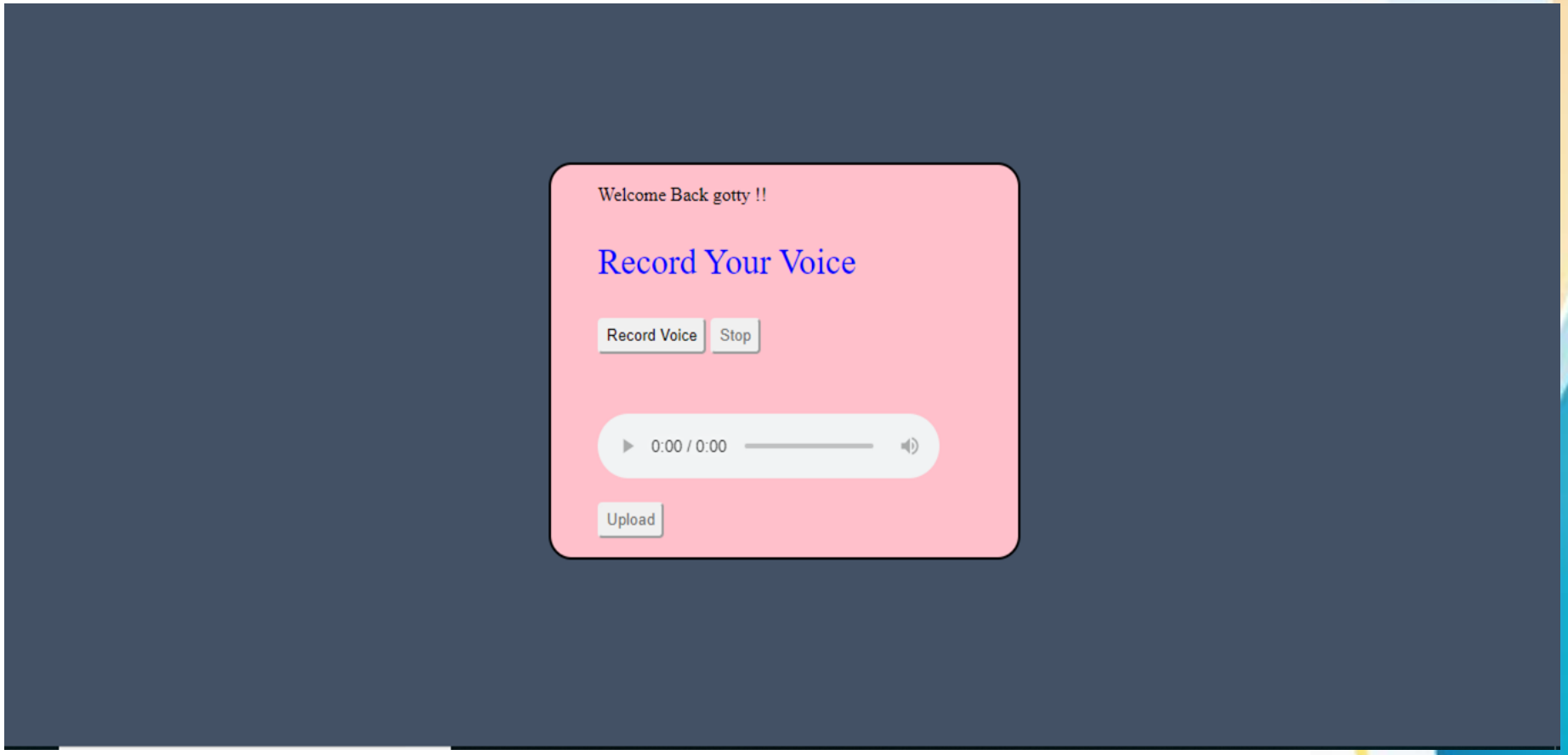
UI/ Use Case

Login Page:



UI/ Use Case

Voice Recording Page:



UI/ Use Case

Song Recommendation Page:

Recommended Songs for you !!!!

Undo

play Stop

Secrets

play Stop

Revelry

Fireflies

Modules

Modules used :

- Preprocessing : converting audio file to wav format using FFMPEG, extracting features using pyaudioanalysis tool
- Model building : multilayered perceptron model using keras tool.
- Web application: web page made using Html, javascript, flask for deploying the application on localhost.

Design Approach

- Preprocessing for Age detection includes converting audio file to wav , featurizing the audio file using python audio analysis tool , this tool gets common audio features like mel spectrogram feature coefficients(MFCC's), spectral centroid, spectral contrast, and spectral rolloff etc .
- we compute them in intervals and then take the mean,median, maximum, minimum, and standard deviation of all the extracted values, this includes 35 features , 5 statistical components of each 170 total features.
- Later we pass these features to multilayered perceptron model includes an input layer with nodes equal to the number of features (~80) and an output layer equal to the number of classes (9). There are four hiddenlayers.
- Using tensorflow , keras modules we built the multilayered perceptron model, using is MP we predict the which Age group class the user belongs to .
- For song suggestion we have built our own recommender model that takes song ID and user's age group and suggest song based on it .

Design Constraints, Assumptions & Dependencies

Constraints:

- Limitations in Voice dataset , song dataset with age group.
- Feature extraction is done through tool, it extracts 34 features * 5 statistical component resulting in 170 features in total.

Assumptions:

- We are assuming that the user has a perfectly working mic, as the noise increases it becomes hard to extract features.
- We are assuming that user speaks for more than 2 sec, less than that can impair the age prediction.

Test Strategy

Manual testing done.

Implementation Details

Lines of code :

Preprocessing – 110
Feature extraction – 52
Dataset visualization – 88
Model building – 380
Server for age Detection – 169
Song suggestion – 89
Web app – 670

Implementation Details

Preprocessing :

- cleaning the dataset includes converting audio file from mp3 to wav using FFMPEG tool.

Feature extraction :

- python tool called pyaudioanalysis is used to extract features from the converted wav files , pyaudioanalysis calculates 35 basic features from the given audio file , each feature has 5 statistical values like mean , median, standard deviation , average max, average min, which results in total of 170 features .
- we store each of these as row elements in a csv file
- Pyaudioanalysis extracts features like mel-frequency cepstral coefficient spectral centroid, spectral contrast, and spectral rolloff.




Implementation Details

Data visualization:

- This plots graph showing the different data present in the training dataset, like class distribution, plotting mean energy, mean entropy, median energy, median entropy of the each class .
- This is done on cleaned data .

Model building :

- There is two part in building model first selecting the correct feature selector, then building final model

Baseline-model :

- includes creating a dummy model to choose correct feature selector , 2 layer neural network, with a hidden layer size of 20. There is a dropout layer with a 0.75 just after the hidden layer and before the output softmax layer. We use a standard scaler because our features are on different scales which will ensure we get more effecient/effective error minimization.
- We're going to see how this baseline performs with 3 different feature selection methods, so we don't have to try out these permutations later (it's difficult to integrate all these different feature selection methods with sklearn random search). So, we will simply choose the feature selection method that gives the best validation results on the baseline model.

Implementation Details

Baseline model:

- Baseline with Treebased feature selection, chi-squared feature selction, PCA dimentionality reduction.
- We've finished all feature selection methods for our baseline models. Luckily, all baseline models are decreasing training error after each epoch, which verifies that our neural network is fitting to the data.
- Tree-based feature selection seemed to yield the best results, so we will use this type of feature selection going forward.

Final model:

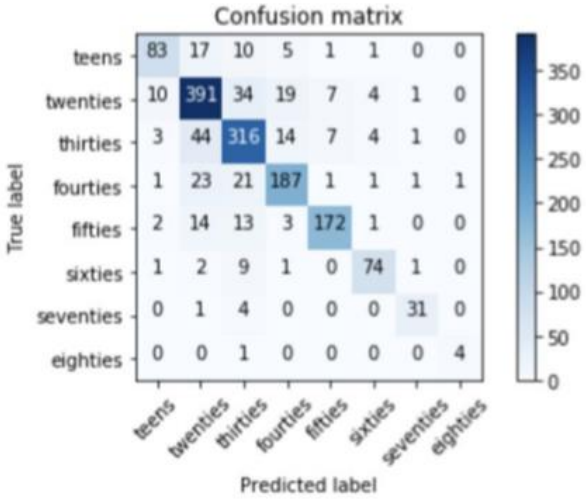
- we will train the final model and export it so it can be used in the web application.
- Our basic architecture for the model was an input layer with nodes equal to the number of features (~80) and an output layer equal to the number of classes (8). There are four hidden layers. The first one has 128 nodes, the second has 256, the third has 256, and the fourth has 128. Because there are is a relatively high number of layers, as well as nodes, the neural network can make more complex and accurate predictions.

Project Results

Accuracy On Training Data :- 89.23%

Accuracy On Test Data :- 84.72%

Highest Accuracy was found among Twenties as the data for the category was larger in number.



Planned Effort Vs Actual Effort

As per schedule proposed timeline:

- Preprocessing: we thought we could finish preprocessing data fast , but we faced a lot of issues regarding corrupted files , missing files so we had to clean data in batches since dataset was huge , (Planned :2 weeks , actual: 4 weeks)
- Developing model : initially we taught model could be completed in short time , but due to machine incapability we had to run on google colab , it was our first time using that software ,so we faced a lot of issues. (Planned : 2 weeks , actual :3 weeks)
- Finish Model, song suggestion, web UI :(Planned:2 weeks, actual: 2 weeks)

Lessons Learnt

- This Project was useful to us since we never integrated ML projects with Web application, it was a new experience for us.
- It was our first time working on audio based machine learning, learned new tools like FFMPEG, pyaudioanalysis, jinja.
- During building model we learnt new software like google colab , that can be used for future projects
- For future enhancement we can use Spotify API to play songs.

PES UNIVERSITY

Thank You