Computer Society of India
Student Branch, VIT Pune

CODEFLIX 2025

CODEZILLA

[THE ULTIMATE HACKATHON]

VOCAL VAULT :
OUR SOLUTION TO VOICE SEPARATION

TRACK 3 SPONSOR 1 (PI-LABS.AI)

TEAM DETAILS

TEAM NAME :

VOCAL VARRIORS

TEAM MEMBERS:

1. HASAN RUPAWALLA (12414068 - FY CS E)
2. KABIR KHANUJA (12413584 - FY CS F)

Computer Society of India
Student Branch, VIT Pune

CODEFLIX 2025

PROBLEM STATEMENT

Speaker diarization involves separating individual speakers from a single audio file, even in challenging conditions like overlapping speech and noise.

Using the LibriCSS dataset, a system can be designed to optimize accuracy, speed, and memory efficiency, enabling robust applications in transcription, virtual assistants, and audio analytic.

TECH STACK

1. React.js and Tailwind

   A JavaScript library and CSS frame work for building user interfaces.

2. Firebase

   A backend-as-a-service platform for database and storage.

3. Python

   A programming language for machine learning models.

4. LibriCSS

   A large dataset used for training our speech models.

Computer Society of India
Student Branch, VIT Pune

CODEFLIX 2025

IDEAS AND APPROACH

LEVERAGING LIBRICSS: We utilize the powerful LibriCSS dataset to train our models, achieving high accuracy in speaker separation

OPTIMIZED FOR SPEED AND EFFICIENCY: Our solution is designed for rapid processing, minimizing wait times for users and delivering results quickly.

USER - FRIENDLY INTERFACE: We prioritize a simple and intuitive interface, making Vocal Vault accessible to users of all technical backgrounds.

Computer Society of India
Student Branch, VIT Pune

CODEFLIX 2025

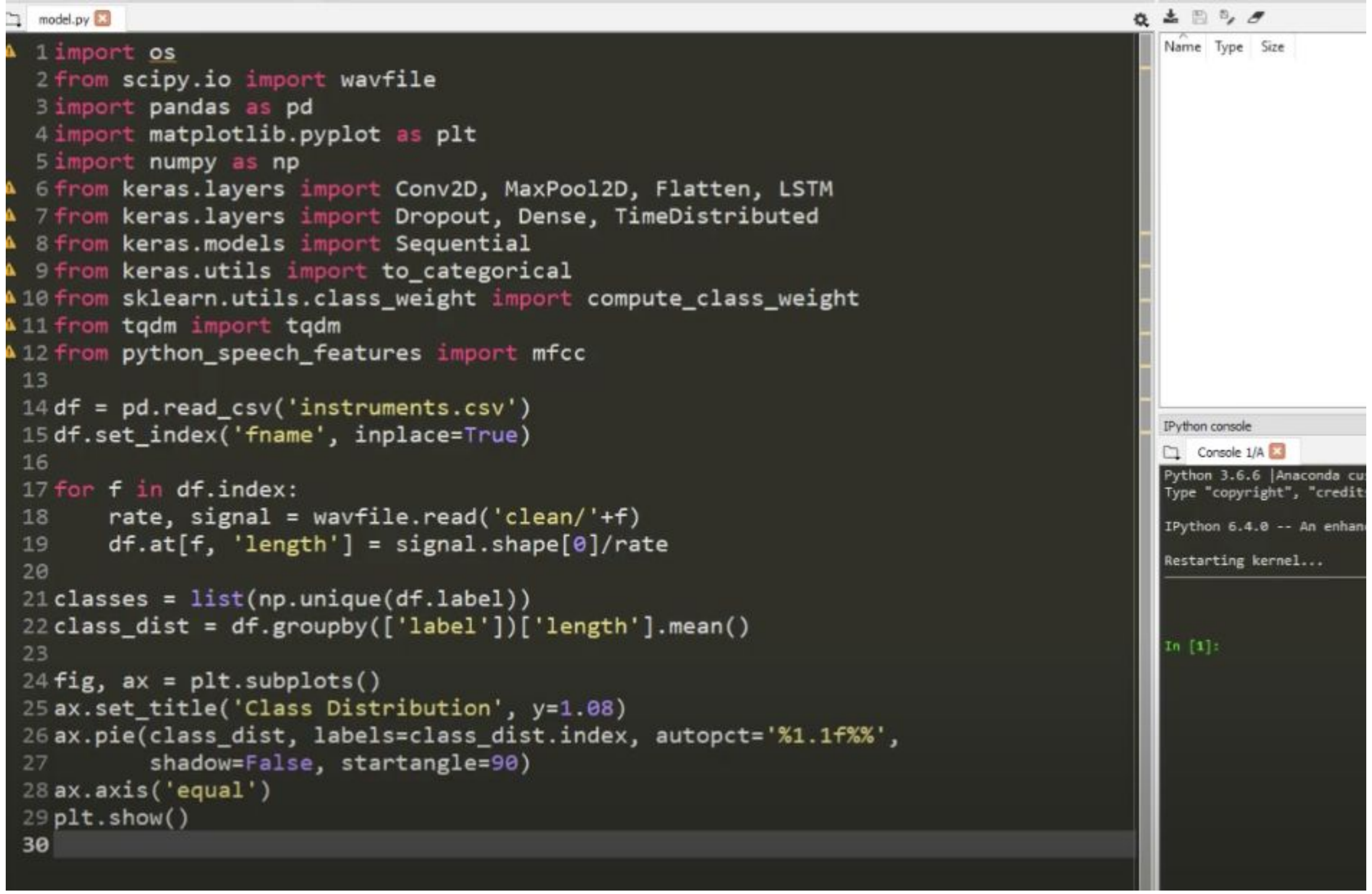
API MODEL USED

API Model :
https://music.ai/docs/api/reference/?utm_source=google.com&click_section=side_menu_docs_reference

Using Music ai’s api for vocal clarity as a first step towards vocal separation from unnecessary background noise.

Ml Model implementation:

Using libraries like keras as it’s built on top of TensorFlow and PyTorch- which would help analyze the data efficiently



Computer Society of India
Student Branch, VIT Pune

FEATURES

CODEFLIX 2025

Real-Time Diarization

Identify and separate speakers in real-time, providing immediate feedback.

Upload and Separate

Paste links of audios from YouTube or Spotify or upload local audio.

Audio Storage

Get to store audios of your friends and family members in personalized folders.

Downloadable Tracks

Output separated audio tracks with minimal loss in quality.

Computer Society of India
Student Branch, VIT Pune

METHODOLOGY

CODEFLIX 2025

FRONTEND

Developing intuitive screens for audio file uploads, YouTube link processing, and displaying separated tracks.

BACKEND

Using Firebase for backend services, including secure user authentication, file storage for uploaded audio files and separated tracks, and database management to store user profiles and processing history.

databaseStoring user profiles, uploaded audio files, separated tracks, and processing history.

ML MODEL

Training a speaker diarization model using the LibriCSS dataset, optimized for accuracy, speed, and memory efficiency.

AUDIO PROCESSING

Utilizing libraries like PyTorch/TensorFlow for voice separation and post-processing to enhance audio quality.

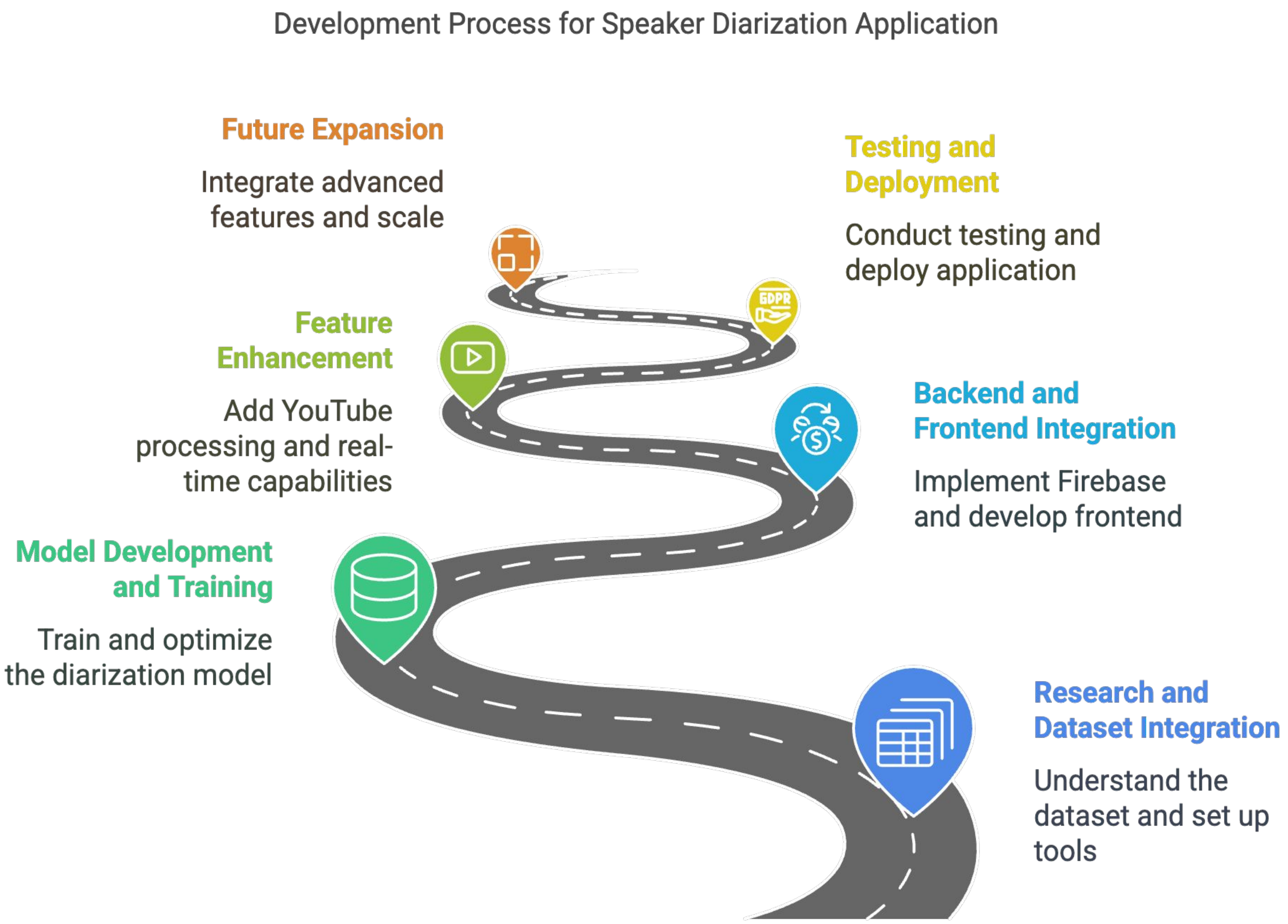
CLOUD INTEGRATION

Host the application and model on a scalable cloud platform to ensure efficient processing and storage for diverse user needs.

Computer Society of India
Student Branch, VIT Pune

CODEFLIX 2025

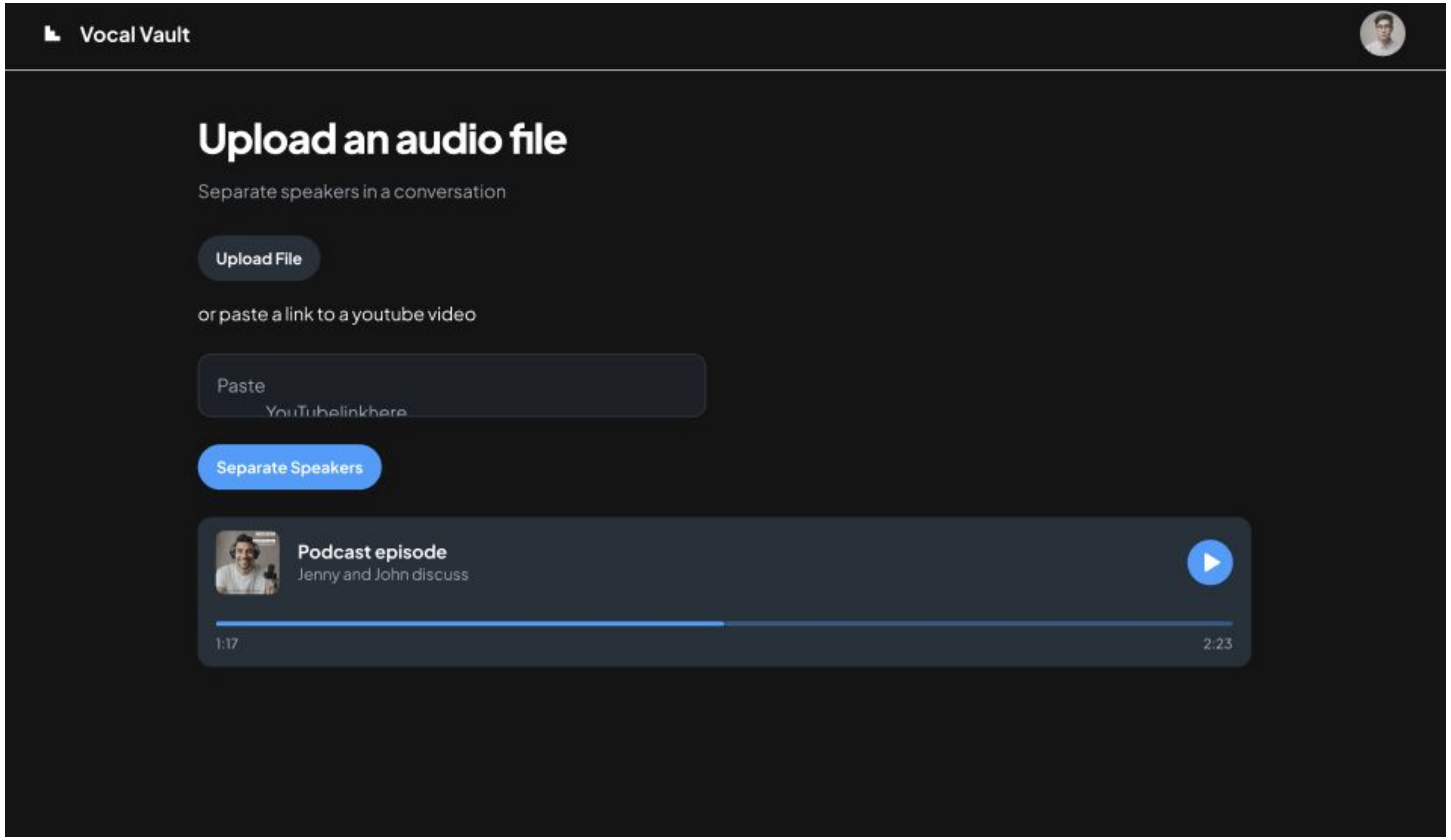
DIAGRAM

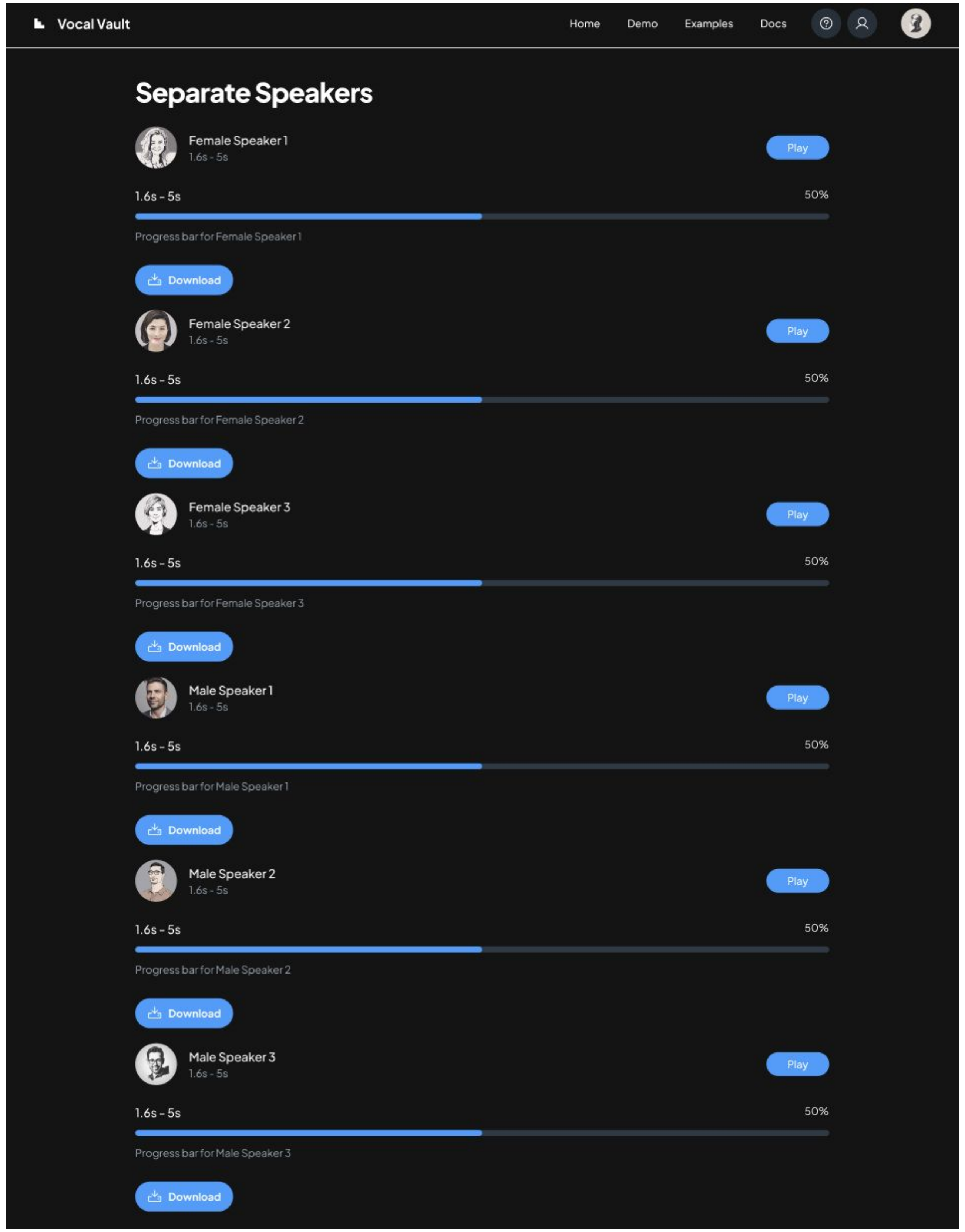


Computer Society of India
Student Branch, VIT Pune

USER INTERFACE

CODEFLIX 2025





CURRENT STATUS

We have successfully developed the frontend, ensuring a visually appealing and user-friendly experience.

Currently, we are delving into the machine learning aspect, which is critical for speaker diarization and audio separation. This phase involves implementing algorithms to accurately identify and segregate voices in the audio files.

Our goal is to seamlessly integrate the backend ML model with the frontend to deliver a cohesive and functional product.

As we advance, we are committed to refining both the UI and the audio-processing system to ensure optimal performance and usability, making the process faster and more efficient .

THANK YOU!