ML-FIESTA AI/ML HACKATHON

Indian Institute of Technology(IIT) Bengaluru

Synergy 2024

- Nitin S (M S Ramaiah college of arts, science and commerce)
- Grishma Yadav (Reva University)
- Ghamani (Reva University)

PROJECT OVERVIEW

The objective is to create a pipeline that processes and comprehends audio input associated with sandalwood cultivation by integrating AI and machine learning. Our main goal is to create:

* A Kannada language model for Automatic Speech Recognition (ASR).

* A question-answering system that uses speech to assist users in retrieving information from audio data.

PROBLEM OBJECTIVES

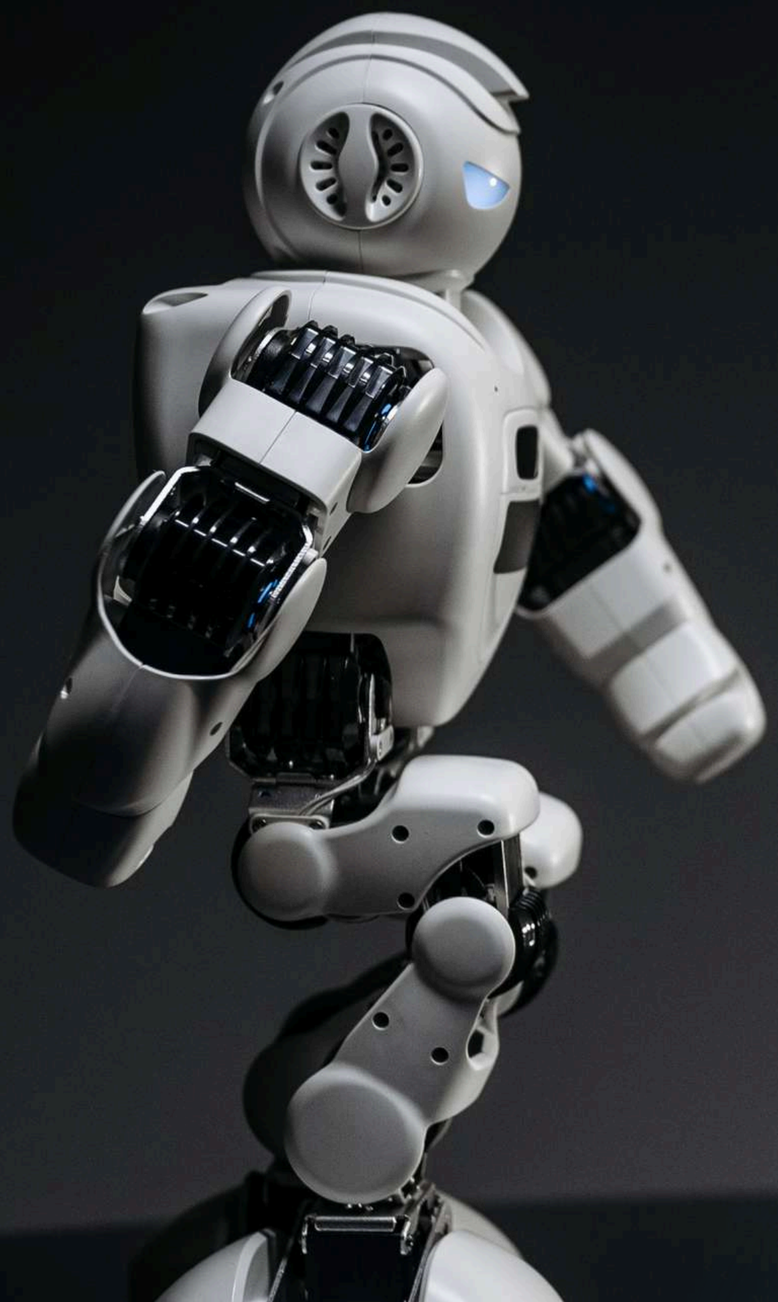
* Construct an ASR model that can recognise Kannada speech informally.

* Use the ASR output to create a searchable audio database.

* Put in place a question-answering system that lets users ask queries via speaking.

PROBLEM STATEMENT

An important source of sandalwood, which has great cultural, religious, and commercial significance in India, is Karnataka. However, a large portion of the traditional knowledge regarding the cultivation of sandalwood is passed down orally and recorded. It is necessary to digitise and conserve this indigenous knowledge because these resources are not readily available. Handling informal Kannada speech with background noise presents the most obstacle because it deviates from formal, standard language.



CHALLENGES:

* There is a dearth of digital data on sandalwood production.

* Noise and colloquial language are common in audio recordings.

* Conventional ASR models have trouble understanding colloquial language.

SCOPE

* Developing an ASR model for Kannada to recognise colloquial language.

* Transcribing audio files to create a searchable database.

To query the audio corpus, a speech-based question-answering system is being developed.

* Using both supplied and openly accessible Kannada datasets, the ASR model is being fine-tuned.

OUT OF SCOPE

* Audio in languages other than Kannada is processed.

* Live stream transcribed in real time.

* Managing intricate conversations with multiple turns.

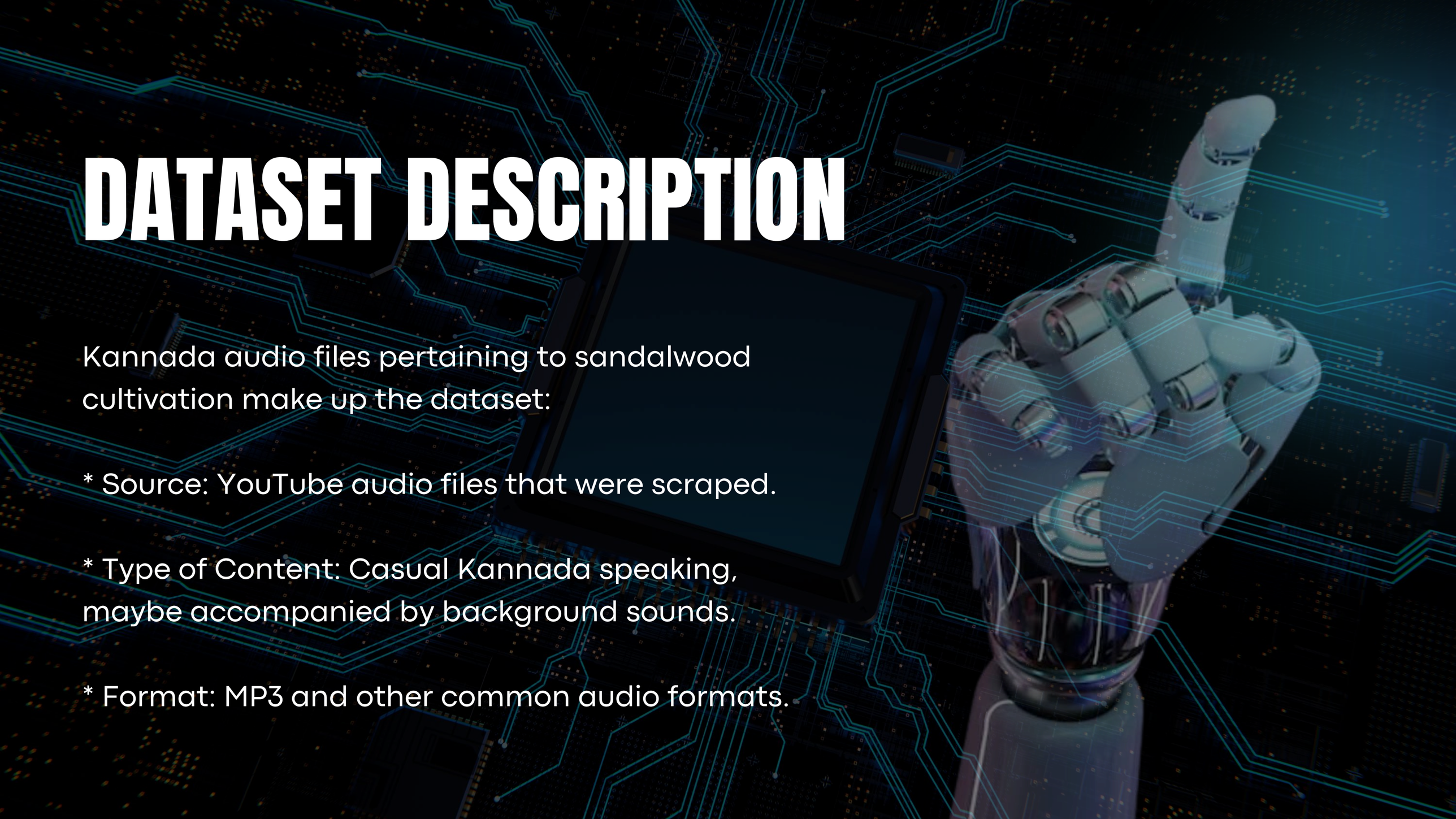
DATASET DESCRIPTION

Kannada audio files pertaining to sandalwood cultivation make up the dataset:

* Source: YouTube audio files that were scraped.

* Type of Content: Casual Kannada speaking, maybe accompanied by background sounds.

* Format: MP3 and other common audio formats.



DATASET CHALLENGES

* Loud audio recordings produced in public areas.

* The use of casual and colloquial language.

* Differences in dialects and pronunciation

FUNCTIONAL REQUIREMENTS

Task 1: Speech recognition

* Create an ASR model from audio files for Kannada speech.

* Deal with casual and colloquial language.

* Encourage the fine-tuning of the model using more datasets.

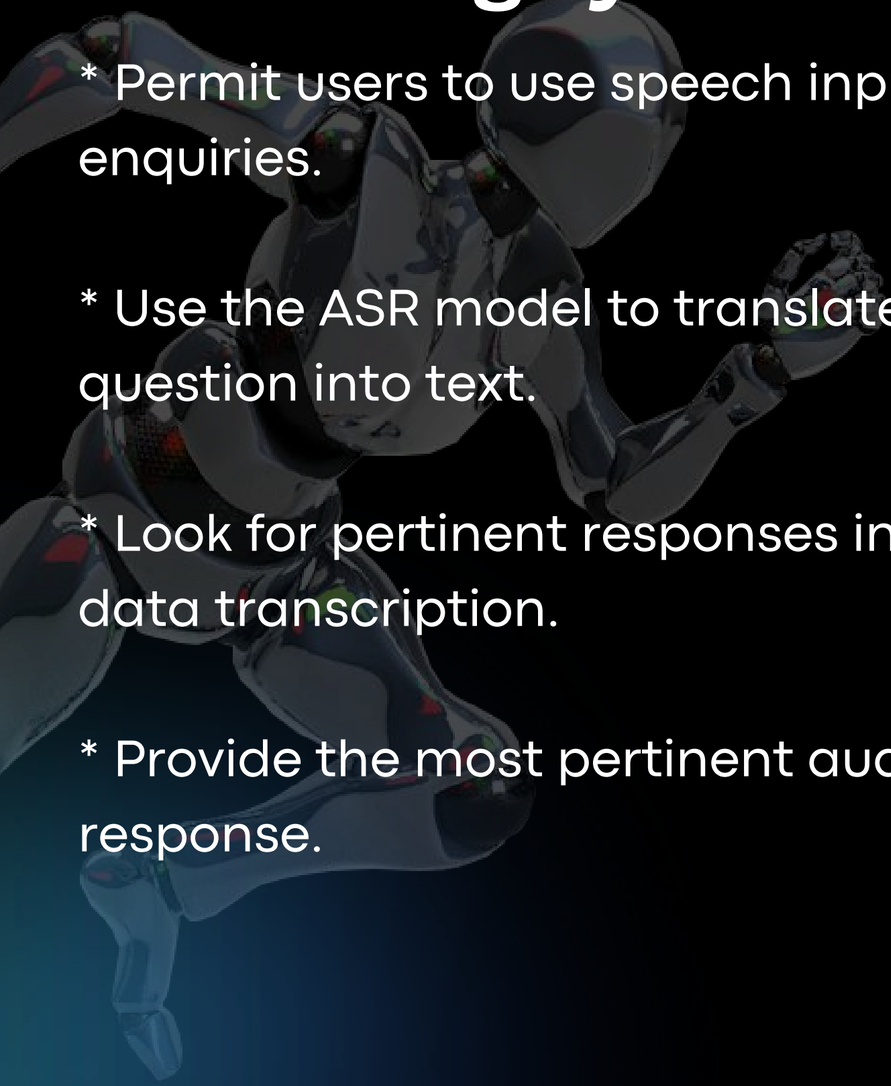
Key features

* Transcribing Kannada speech to text.

* Speech improvement and noise reduction.

* Take into account differences in dialect.

Task 2: Speech-based question answering system



* Permit users to use speech input to ask enquiries.

* Use the ASR model to translate the spoken question into text.

* Look for pertinent responses in the audio data transcription.

* Provide the most pertinent audio clip as the response.

Key features

* Retrieving accurate responses from speech queries.

* The transcribed corpus is efficiently searched and indexed.

* An intuitive interface for queries and answers.

TECHNICAL REQUIREMENTS

* Languages: Python.

* Libraries & Frameworks: PyTorch, Whisper (ASR), Hugging Face Transformers.

* Database: MongoDB.

* Deployment: Google Cloud Platform (GCP) or AWS for scalable processing.

* Hardware: GPU servers for training, cloud instances for deployment.

* Tools: Google Colab, Jupyter Notebooks, GitHub.

RISKS AND MITIGATION

RISKS

- Low-quality audio data
- Poor ASR accuracy on colloquial speech
- High query processing latency

MITIGATION

- Use noise reduction and data augmentation
- Fine-tune model with additional colloquial data
- Optimize search algorithms and indexing

REFERENCES

- GitHub Repository
- Youtube Videos
- Presentation Slides
- Project Documentation
- Whisper: A Speech Recognition Framework by OpenAI.
- Hugging Face Transformers for NLP models.
- Google Colab for collaborative coding.
- MongoDB for database management.
- PyTorch for deep learning models.
- GitHub for version control and collaboration.


THANK YOU!