

# Project Report: Voice Transcription & Analysis App

## Objective

To build a Python application that

- Accepts an audio file as input
- Transcribes the content with timestamps
- Extracts key topics
- Analyzes sentiment
- Generates a summary (in Hindi or English, based on input)

## Approach

### 1. Environment Setup

- Created a virtual environment `sftwtrs`

```
C:\Users\kanch>sftwtrs\Scripts\activate
```

- Installed required packages like `ffmpeg`, `openai-whisper`, `transformers`, `keybert`, `torch`, etc.

### 2. Audio Transcription

- Used `whisper` (OpenAI's model) to transcribe audio into text with timestamps.

Romanized Hindi Transcription

He was not able to use the two legs. So he didn't know where he was. He was hit by someone or someone else. But he didn't even have blood loss. He didn't have the dog. He went to Amit Sar. Amit Sar took him to the gym. He took all the stuff he had. Amit Sar didn't even understand what he said. He said he is a bad man. He is only a pair of men. He didn't have the blood loss. He didn't have the blood loss. He took him to the gym. He gave him the pain t. He was resting. He didn't eat anything. He didn't eat anything. He just had to eat.

- For non-English audio (like Hindi), used `"small"` model with `language="hi"` to improve recognition.
- Parsed the segments with start and end times.

#### Segments with Timestamps

[0.0s - 3.0s] He was not able to use the two legs.  
[3.0s - 4.0s] So he didn't know where he was.  
[4.0s - 8.0s] He was hit by someone or someone else.  
[8.0s - 10.0s] But he didn't even have blood loss.  
[10.0s - 12.0s] He didn't have the dog.  
[12.0s - 14.0s] He went to Amit Sar.  
[14.0s - 18.0s] Amit Sar took him to the gym.  
[18.0s - 20.0s] He took all the stuff he had.  
[20.0s - 24.0s] Amit Sar didn't even understand what he said.  
[24.0s - 26.0s] He said he is a bad man.  
[26.0s - 28.0s] He is only a pair of men.  
[28.0s - 30.0s] He didn't have the blood loss.  
[30.0s - 32.0s] He didn't have the blood loss.  
[32.0s - 34.0s] He took him to the gym.  
[34.0s - 36.0s] He gave him the paint.  
[36.0s - 38.0s] He was resting.  
[38.0s - 40.0s] He didn't eat anything.  
[40.0s - 42.0s] He didn't eat anything.  
[42.0s - 44.0s] He just had to eat.

### 3. Key Topic Extraction

- Used [KeyBERT](#) with [distilbert-base-nli-mean-tokens](#) to extract key phrases and topics from the transcribed text.
- Limited to top 10 keywords to avoid redundancy.

#### Key Topics

- sar took (Score: 0.39)  
- sar didn (Score: 0.38)  
- blood loss (Score: 0.38)  
- legs didn (Score: 0.38)  
- didn eat (Score: 0.38)  
- didn blood (Score: 0.38)  
- dog went (Score: 0.36)  
- blood (Score: 0.36)  
- sar amit (Score: 0.34)  
- amit sar (Score: 0.34)

### 4. Sentiment Analysis


- Used Hugging Face Transformers pipeline with [mrms8488/bert-tiny-finetuned-sentiment](#) model.
- Analyzed sentiment of the full transcription.

Overall Sentiment: NEGATIVE (Score: 0.99)

### 5. Summarization

- For Hindi input, used multilingual model:  
`csebuetnlp/mT5_multilingual_XLSum`.
- Generated Hindi summary directly from the original transcription without converting to English.

Summary: He was not able to use the two legs so he didn't know where he was . He was hit by someone or someone else. He didn't even have blood loss. He went to Amit Sar. Amit Sar took him to the gym. He took all the stuff he had. He said he is a bad man. He is only a pair of men.

model.safetensors: 100%  1.22G/1.22G [09:30<00:00, 2.11MB/s]

## Challenges Faced

### 1. Whisper Language Detection

- Initially, Whisper misidentified the Hindi audio as Tamil.
- Resolved by forcing the language to "`hi`" during transcription.

### 2. FFmpeg Not Recognized

- Faced '`ffmpeg`' is not recognized error.
- Solved by manually adding FFmpeg's path to the system environment and verifying from the virtual environment.

### 3. KeyBERT Slow on Long Texts

- For long transcriptions, KeyBERT was slow and memory intensive.
- Used slicing and stop-word removal to improve performance.

### 4. Sentiment Analysis Model Conflict

- Faced `NameError: pipeline is not defined`.
- Solved by correctly importing the pipeline from `transformers`.

### 5. Summarization in Non-English

- Most summarization models work only in English.
- Used multilingual `mT5` model to handle Hindi summaries natively.

## Conclusion

The pipeline successfully:

- Transcribes speech
- Timestamps segments
- Extracts meaningful keywords
- Detects sentiment
- Summarizes in original language