YESHWANTRAO CHAVAN COLLEGE OF ENGINEERING, NAGPUR.

(An autonomous Institution Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Speak2Summarize : Daily Recap using Machine Learning and Natural Language Processing

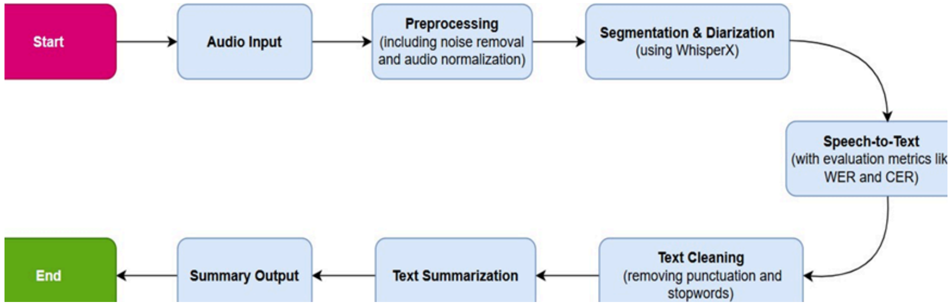
NAME OF STUDENT: Hemanshu Waghmare, Dhruv Dalvi, Rishabh Jain, Yuvraj Chavan, Sanket Asole

NAME OF GUIDE: Prof. Chanchla Tripathi

Abstract: In today’s information-rich world, quickly finding relevant content is often difficult. Automatic document summarization using natural language processing helps extract key details from lengthy texts. Whether extraction- or abstraction-based, text summarization techniques condense content into concise summaries. We propose a method that takes audio input and produces a written summary using these techniques. Our product, Speak2Summarize, tackles this challenge by providing clear, concise summaries of user interactions and daily meetings, complete with timestamps.

Introduction: With the rise of big data, analytics, and automation, audio has become a vital source of information in business, education, and media. Extracting insights from audio through Speech-to-Text (STT) and text summarization is essential for transforming raw data into meaningful, accessible content. STT systems convert spoken language into text, while summarization condenses lengthy transcripts into clear, concise overviews, saving time and enhancing comprehension. Early STT relied on statistical models like HMMs and GMMs, which had limitations with accents and noise. Deep learning introduced more accurate neural models like RNNs, LSTMs, and transformers such as Whisper, improving adaptability and performance. Similarly, summarization evolved from rule-based methods to advanced neural networks like BERT, GPT, and T5, enabling more coherent and context-aware summaries. Together, these technologies address the growing need for effective audio processing and information extraction.

Simulated Designs: For this project, with regard to the methodology the research consisted of two major steps, namely audio preprocessing and speech recognition with speaker diarization. These were achieved in order to create manageable phases for audio data, for purposes of transcription and also to create foundation for text summarization phase.


Start
Audio Input
Preprocessing
(including noise removal and audio normalization)
Segmentation & Diarization
(using WhisperX)
Speech-to-Text
(with evaluation metrics lik WER and CER)
Text Cleaning
(removing punctuation and stopwords)
Text Summarization
Summary Output
End

Simulation Result:

ML Model Results of Crop Recommendation System

1. A decision tree is a structured algorithm in which the data is divided into small subsets based on the input values, with the goal of predicting target variables.
2. Gaussian Naive Bayes is an algorithm based on Bayes' theorem that calculates probability by making a "naive" assumption about the independence of features.
3. Logistic regression is a popular algorithm for multi-class classification functions where it gives the probability of each class as a function of the input feature.

Conclusion and Future Scope:

Conclusion: The This project successfully completed two critical modules focusing on processing and understanding audio data:
Module 1: Achieved preprocessing of audio datasets by implementing speaker classification through speech diarization and segmentation techniques.
Module 2: Implemented a Speech-to-Text (STT) model that converts audio into accurate text. The model's performance was evaluated using metrics like Word Error Rate (WER) and Character Error Rate (CER).

Future Scope:

- Speaker diarization, EEND, multi-speaker handling
- Diverse datasets, languages, accents, real-world audio
- Cloud deployment, AWS/Azure, live transcription
- Scalability, enterprise, individual users, context expansion