Speech Recognition and Analysis

B. TECH SEM – VI Cloud Computing Lab Project
Dept. of Computer Science & Engineering

By

Pulak Jain	**20BCP029**
Sahil Anand	**20BCP010**
Jugal Soni	**20BCP018**

<u>Under the Supervision Of</u>

Dr. Manish Paliwal

SCHOOL OF TECHNOLOGY
PANDIT DEENDAYAL ENERGY UNIVERSITY
GANDHINAGAR, GUJARAT, INDIA

May 2023

ABSTRACT

Speech recognition and analysis is a complex task that requires the integration of multiple technologies to achieve accurate and efficient results. The project described here uses a range of technologies, including Streamlit, AssemblyAI's speech recognition API, TextBlob, and Spacy, to enable speech recognition and analysis. Also as input we give users an option to either select an existing audio file or record a fresh audio file.

Streamlit was used to build a graphical user interface for the project, which allows users to interact with the system in a user-friendly and intuitive way. The interface is designed to be easy to use and provides access to all the main features of the system.

AssemblyAI's speech recognition API is used to extract transcripts from audio recordings. The API is highly accurate and can extract speech from a wide range of audio sources, including recordings made in noisy environments. The extracted transcripts are then processed by TextBlob, a natural language processing library that provides easy-to-use APIs for tasks like sentiment analysis and part-of-speech tagging.

The project also utilizes Spacy, another popular natural language processing library, to perform extractive summarization on the transcripts. Spacy's features like named entity recognition and dependency parsing are useful for analyzing text data and generating summaries that capture the key points of the text.

Also, the project is hosted through Azure, a cloud computing platform that provides a wide range of services for building, deploying, and managing applications and services. By hosting the project on Azure, users can now be able access the system from anywhere with an internet connection, and the system can scale to meet demand as needed.

Overall, the project provides an effective tool for speech recognition and analysis, leveraging a range of technologies to achieve accurate and efficient results. The project is suitable for a wide range of applications, including speech-to-text transcription, sentiment analysis, and text summarization. And at last, this project can be accessible from anywhere thanks to cloud computing.

INDEX

1. Introduction

Speech recognition and analysis is an important area of natural language processing that involves the extraction of meaning from spoken language. In recent years, there has been a significant increase in the use of speech recognition and analysis technologies in a wide range of applications, including customer service, healthcare, education, and entertainment.

To enable speech recognition and analysis, a variety of technologies can be used. This includes natural language processing libraries like TextBlob and Spacy, which are capable of performing a range of text analysis tasks such as sentiment analysis, part-of-speech tagging, and extractive summarization. In addition, speech recognition APIs like AssemblyAI can be used to extract transcripts from audio recordings, which can then be analyzed using natural language processing techniques.

To make it easy for users to interact with these technologies, it is important to provide an intuitive and user-friendly interface. This is where Streamlit comes in, providing a powerful tool for building interactive web applications that allow users to access and use speech recognition and analysis technologies.

In this context, we have developed a project called "Speech Recognition and Analysis" that uses Streamlit, Assembly AI's speech recognition API, TextBlob, and Spacy to enable accurate and efficient speech recognition and analysis. This project provides a user-friendly graphical user interface that allows users to easily upload audio recordings and access a range of analysis features, including sentiment analysis and extractive summarization. Overall, this project demonstrates the potential of streamlit-based interactive web applications for enabling speech recognition and analysis, showcasing the potential of combining these technologies for a wide range of applications.

Also, at last, hosting the project correctly on cloud can help reach huge audience and would be scalable. Thus, Azure, a cloud computing platform that provides a wide range of services for building, deploying, and managing applications and services can be one of the good choice.

In the following sections, we will delve deeper to explain our methodology and showcase how we have deployed our project to cloud.

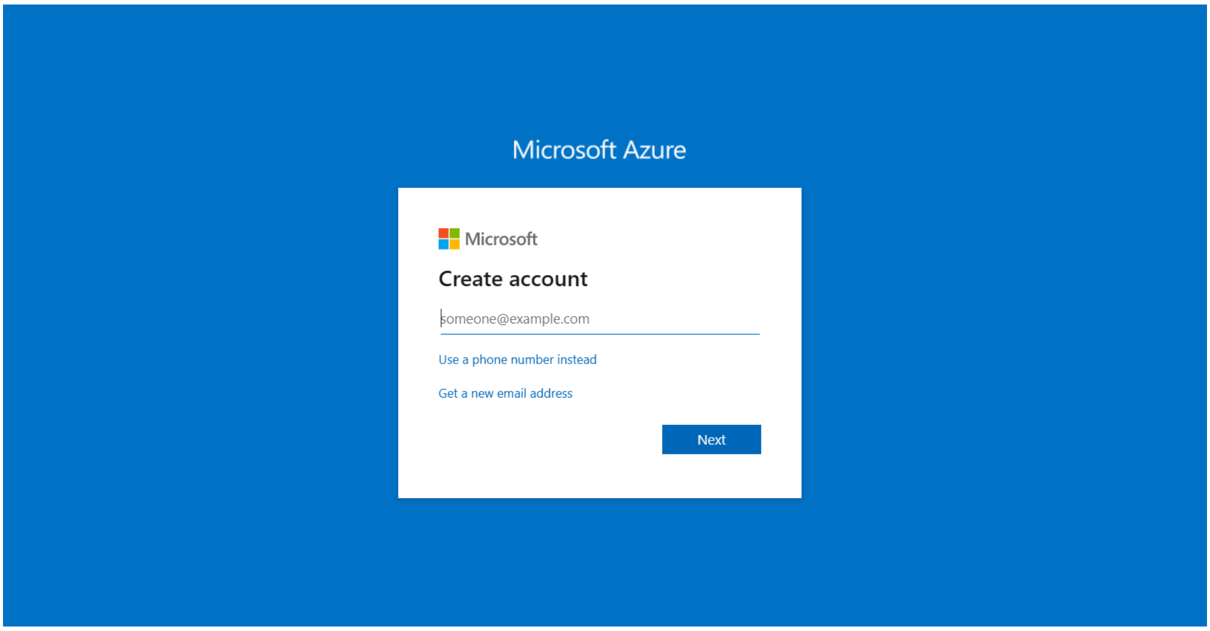
2. Methodology

1. Data Collection: The first step is to collect the audio data that will be used for speech recognition and analysis. This can be done by recording new audio or by using existing audio files.
2. Audio Transcription: Once the audio data is collected, the next step is to transcribe the audio into text using AssemblyAI's speech recognition API. This API can be used to automatically transcribe the audio data into a text format, which can be further analyzed using natural language processing techniques.
3. Sentiment Analysis: After the audio data is transcribed into text, the next step is to perform sentiment analysis using the TextBlob library. This library provides a simple API for analyzing text data and can be used to identify the sentiment of the spoken content in the audio recording.
4. Extractive Summarization: Once sentiment analysis is completed, the next step is to perform extractive summarization on the transcript. This can be done using Spacy, which is a popular natural language processing library that provides a range of features for analyzing text data.
5. GUI Development: To make it easy for users to interact with the system, a graphical user interface is developed using Streamlit. Streamlit provides a simple way to create web applications that allow users to interact with the system and view the results of the speech recognition and analysis process.
6. Integration: Finally, the various components of the system are integrated to create a functional speech recognition and analysis tool. This involves combining the audio transcription, sentiment analysis, and extractive summarization components with the GUI developed using Streamlit.

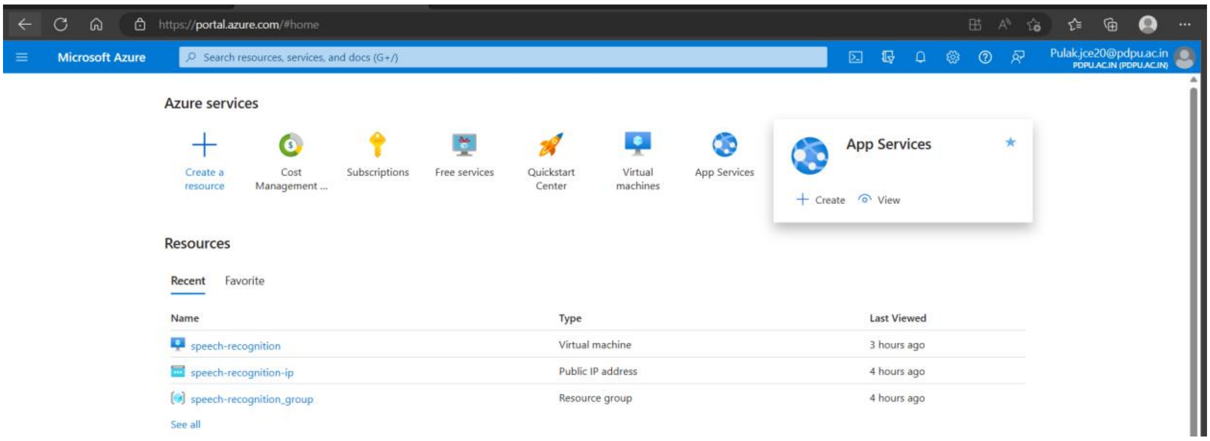
Overall, this methodology provides a structured approach for developing a speech recognition and analysis tool using a range of natural language processing technologies and graphical user interface development tools.

Steps for deployment:

1. Create and Login to Azure Students which provides free 100 credits to students.

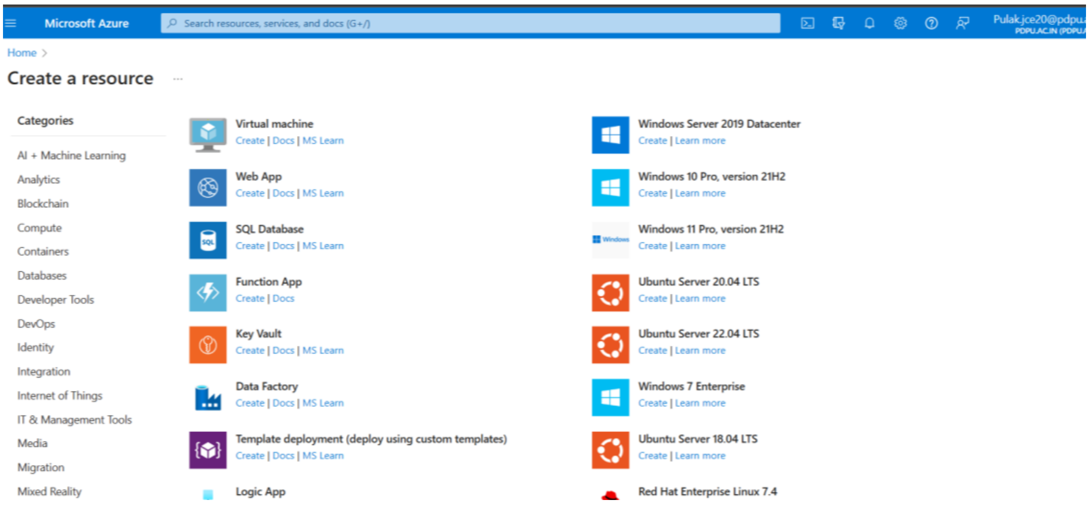


2. Now after successful login, there is create new resource – click on it

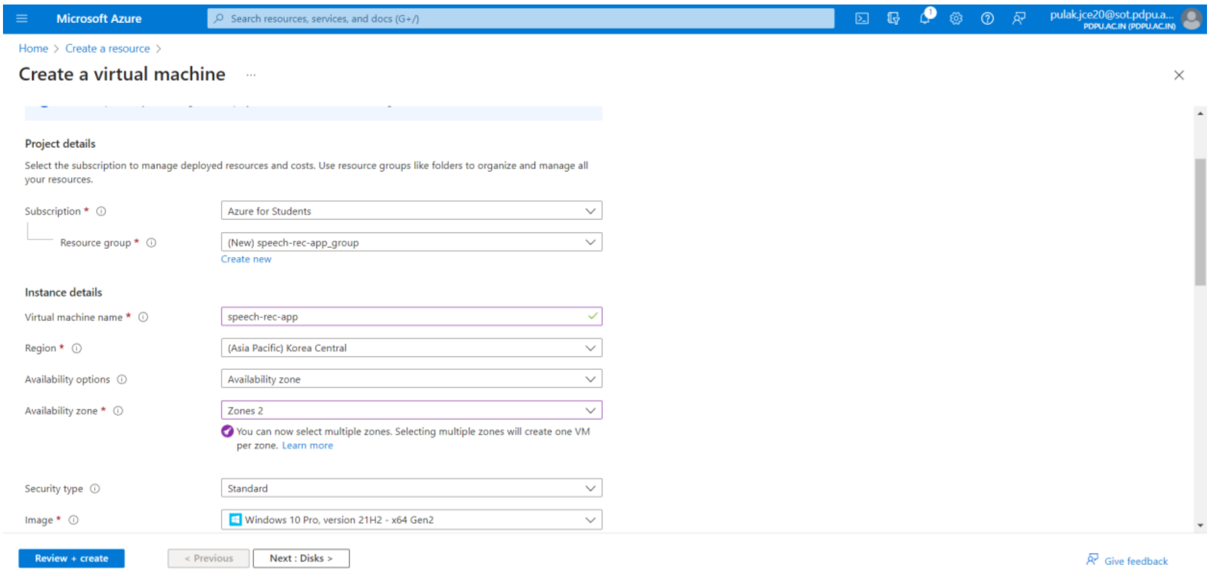


Name	Type	Last Viewed
speech-recognition	Virtual machine	3 hours ago
speech-recognition-ip	Public IP address	4 hours ago
speech-recognition_group	Resource group	4 hours ago

3. Here click on create button below virtual machine to create a new virtual machine



4. Now we filled the details to configure your virtual machine and then click on review + create button below



Microsoft Azure
Search resources, services, and docs (G+/)
pulak.jce20@sot.pdpu.a...
PDPU.AC.IN (PDPU.AC.IN)
Home > Create a resource >

Create a virtual machine

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *: Azure for Students

Resource group *: (New) speech-rec-app_group

Create new

Instance details

Virtual machine name *: speech-rec-app

Region *: (Asia Pacific) Korea Central

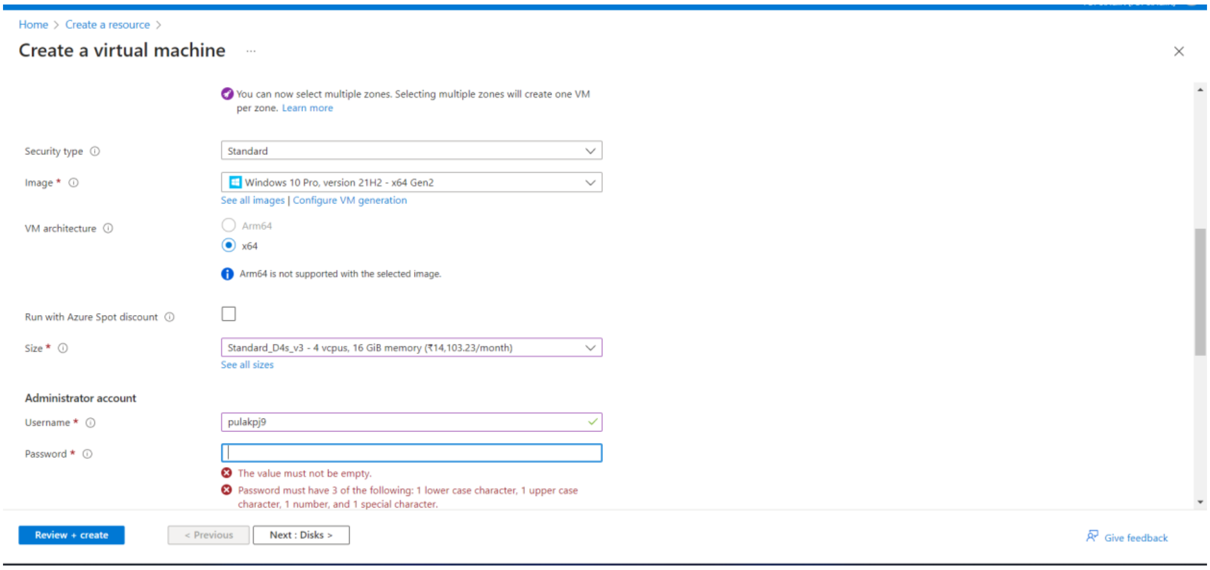
Availability options: Availability zone

Availability zone *: Zones 2

You can now select multiple zones. Selecting multiple zones will create one VM per zone. Learn more

Security type: Standard

Image *: Windows 10 Pro, version 21H2 - x64 Gen2



Create a virtual machine

You can now select multiple zones. Selecting multiple zones will create one VM per zone. Learn more

Security type: Standard

Image *: Windows 10 Pro, version 21H2 - x64 Gen2

See all images | Configure VM generation

VM architecture: ○ Arm64 ● x64

Arm64 is not supported with the selected image.

Run with Azure Spot discount: □

Size *: Standard_D4s_v3 - 4 vcpus, 16 GiB memory (₹14,103.23/month)

See all sizes

Administrator account

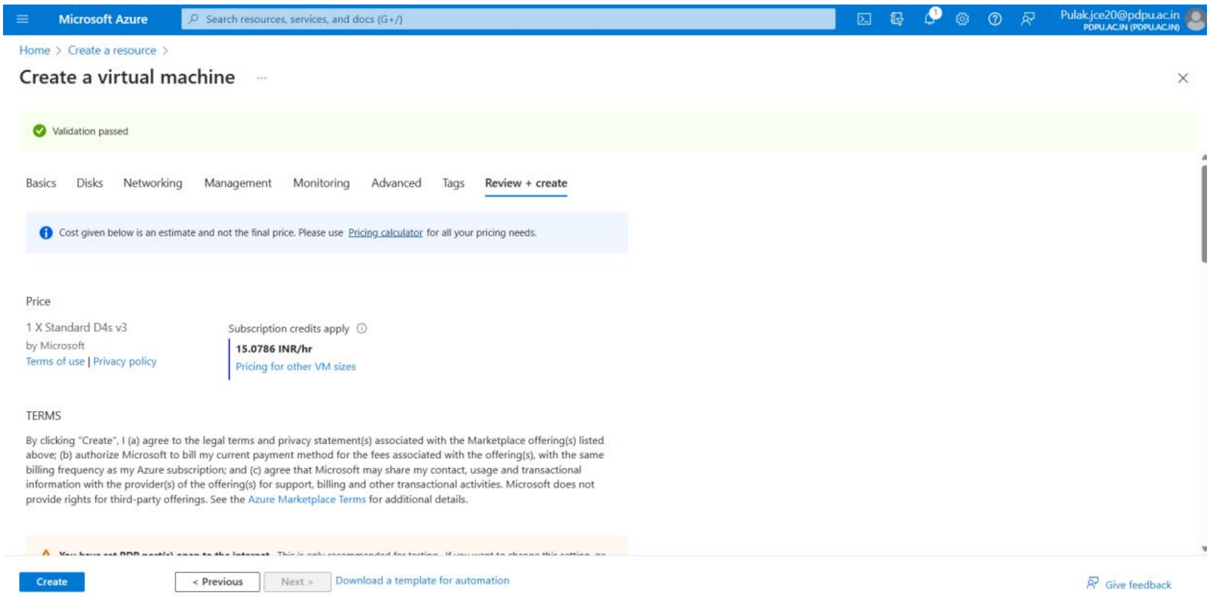
Username *: pulakpj9

Password *:

The value must not be empty.

Password must have 3 of the following: 1 lower case character, 1 upper case character, 1 number, and 1 special character.

5. Now as we saw validation passes message, we clicked on create button below and your virtual machine got deployed.



Microsoft Azure
Search resources, services, and docs (G+/)
Pulak.jce20@pdpu.ac.in
PDPU.AC.IN (PDPU.AC.IN)
Home > Create a resource >

Create a virtual machine

Validation passed

Cost given below is an estimate and not the final price. Please use Pricing calculator for all your pricing needs.

Price

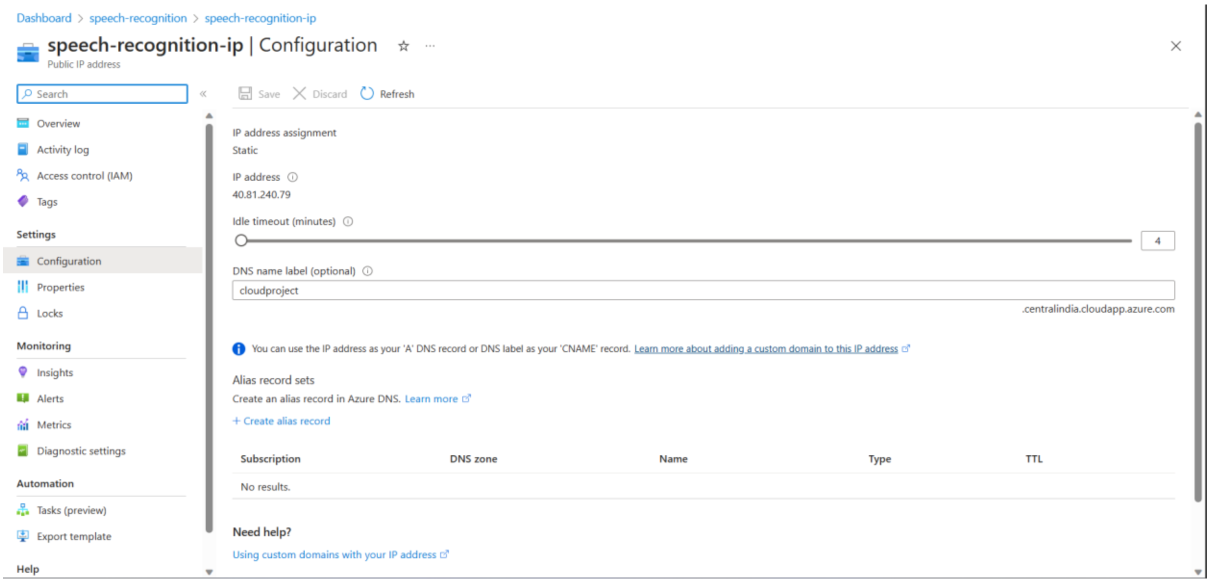
1 X Standard D4s v3
by Microsoft
Terms of use | Privacy policy

Subscription credits apply
15.0786 INR/hr
Pricing for other VM sizes

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the Azure Marketplace Terms for additional details.

6. Now in the next step we went to DNS and click on it to configure it, gave it the name ‘cloudproject’.



speech-recognition-ip | Configuration

Public IP address

IP address assignment
Static

IP address
40.81.240.79

Idle timeout (minutes)
4

DNS name label (optional)
cloudproject
.centralindia.cloudapp.azure.com

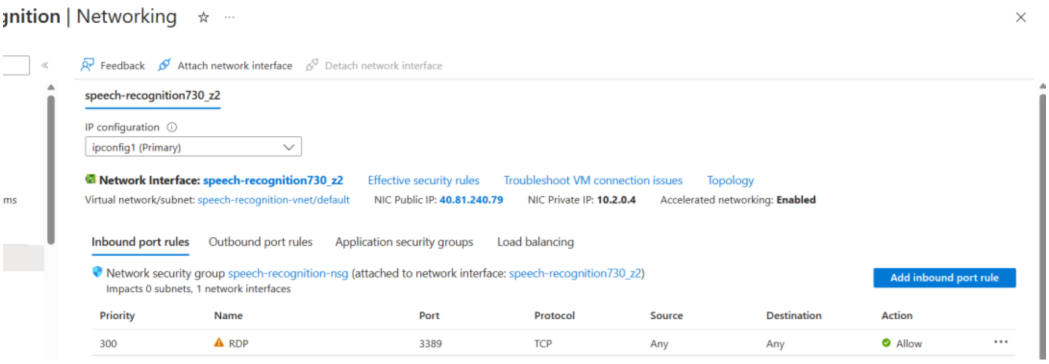
You can use the IP address as your 'A' DNS record or DNS label as your 'CNAME' record. Learn more about adding a custom domain to this IP address

Alias record sets
Create an alias record in Azure DNS. Learn more
+ Create alias record

Subscription	DNS zone	Name	Type	TTL
No results.				

Need help?
Using custom domains with your IP address

7. Then we went to the networking page and click on add inbound port rule



ɉnition | Networking

speech-recognition730_z2

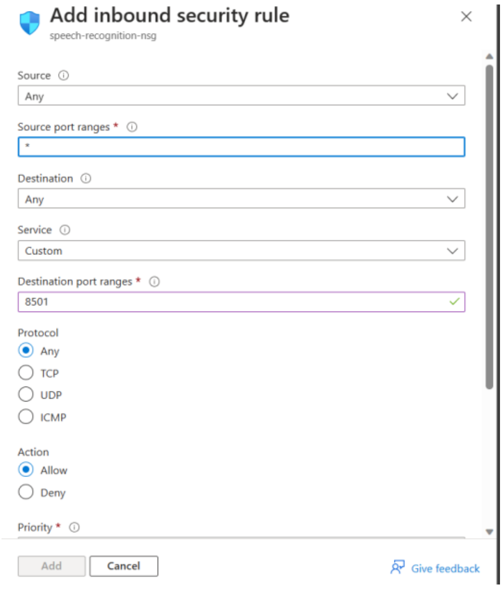
IP configuration
ipconfig1 (Primary)

Network Interface: speech-recognition730_z2 Effective security rules Troubleshoot VM connection issues Topology
Virtual network/subnet: speech-recognition-vnet/default NIC Public IP: 40.81.240.79 NIC Private IP: 10.2.0.4 Accelerated networking: Enabled

Network security group speech-recognition-nsg (attached to network interface: speech-recognition730_z2)
Impacts 0 subnets, 1 network interfaces

Priority	Name	Port	Protocol	Source	Destination	Action
300	⚠ RDP	3389	TCP	Any	Any	✓ Allow

8. Now we have added the port number 8501 in which our streamlit app was running and kept other as default and click on add.



Source: Any

Source port ranges *: *

Destination: Any

Service: Custom

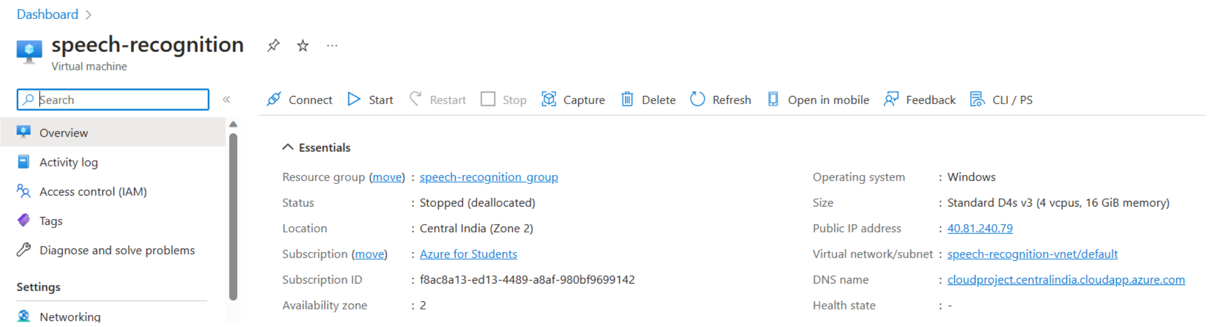
Destination port ranges *: 8501

Protocol: ● Any ○ TCP ○ UDP ○ ICMP

Action: ● Allow ○ Deny

Priority *

9. Now we have copied the public ip and run it on Remote Destop Connection in windows to check working of our app.



Resource group (move) : speech-recognition_group

Status : Stopped (deallocated)

Location : Central India (Zone 2)

Subscription (move) : Azure for Students

Subscription ID : f8ac8a13-ed13-4489-a8af-980bf9699142

Availability zone : 2

Operating system : Windows

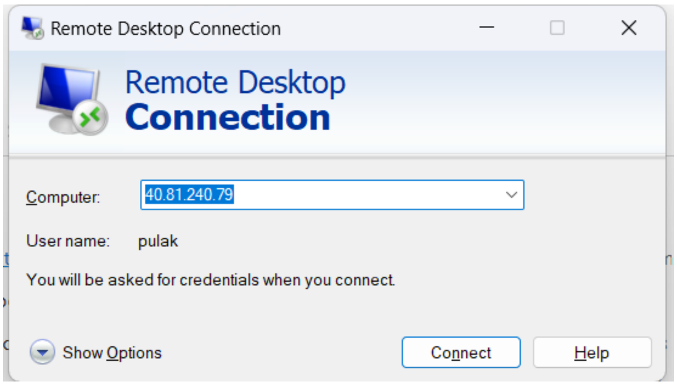
Size : Standard D4s v3 (4 vcpus, 16 GiB memory)

Public IP address : 40.81.240.79

Virtual network/subnet : speech-recognition-vnet/default

DNS name : cloudproject.centralindia.cloudapp.azure.com

Health state : -

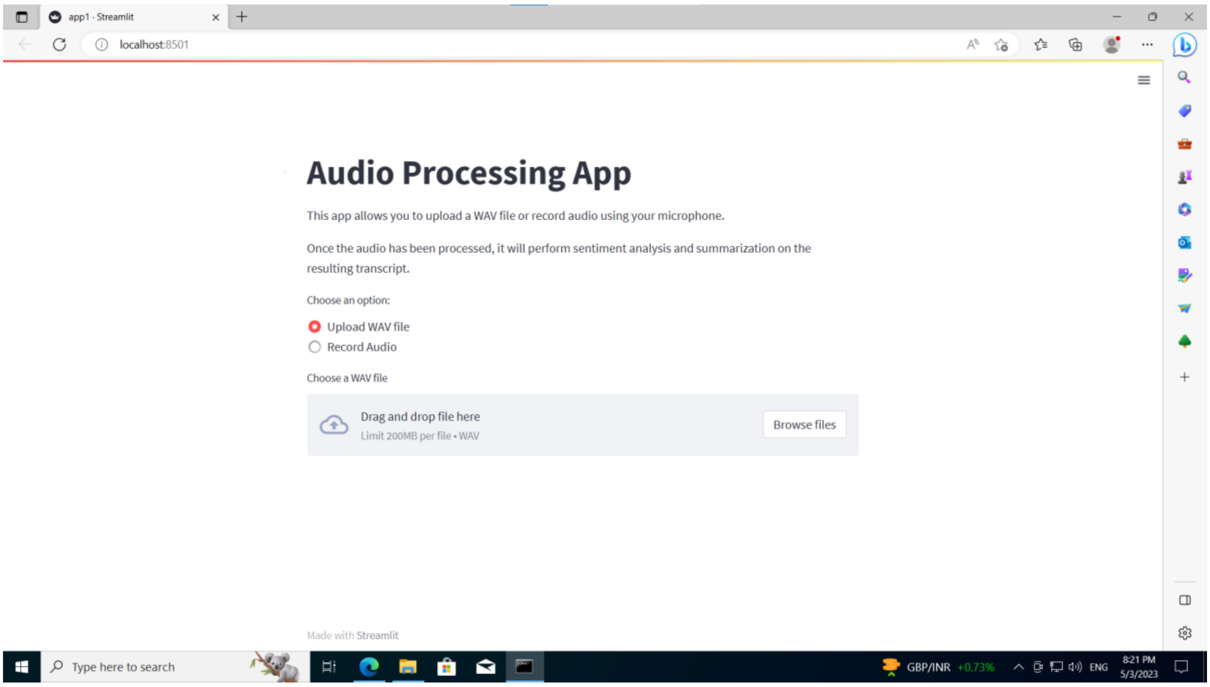


Computer: 40.81.240.79

User name: pulak

You will be asked for credentials when you connect.

10. Now the windows system opens up and then we uploaded our project in desktop and installed all the required libraries and also installed python and then at last ran it on the local host.



Audio Processing App

This app allows you to upload a WAV file or record audio using your microphone.

Once the audio has been processed, it will perform sentiment analysis and summarization on the resulting transcript.

Choose an option:

● Upload WAV file
○ Record Audio

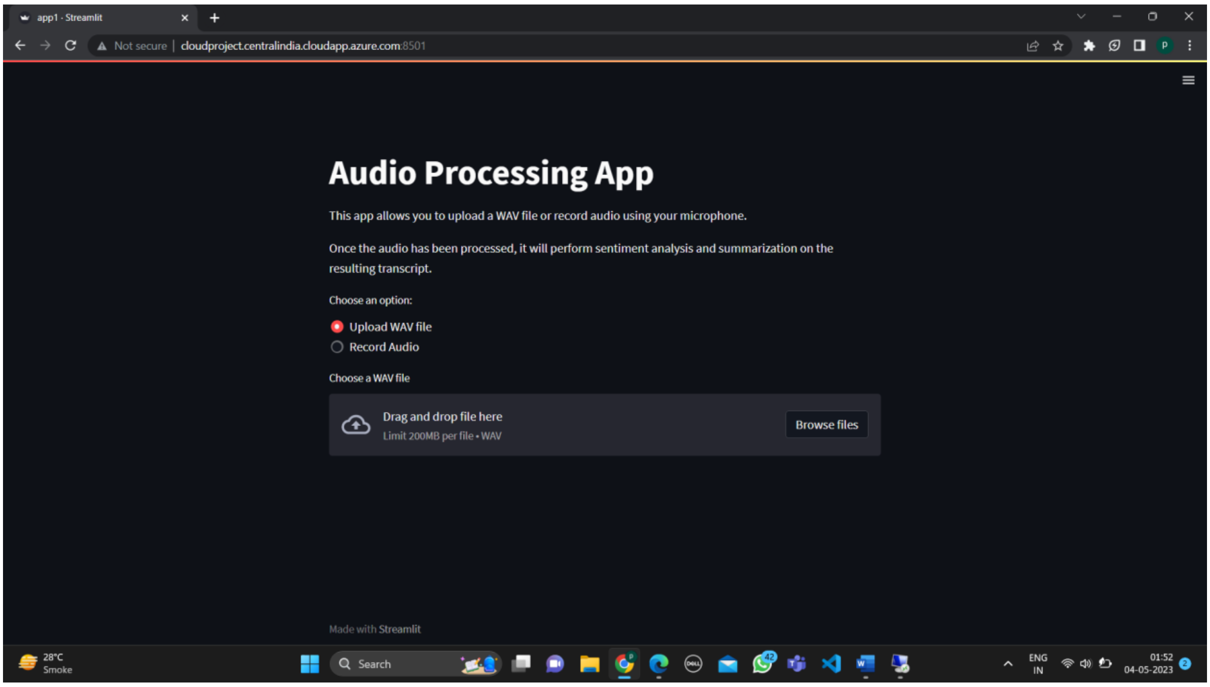
Choose a WAV file

Drag and drop file here
Limit 200MB per file • WAV

Browse files

Made with Streamlit

11. Now go to you desktop and open browser and type ‘cloudproject.centralindia.cloudapp.azure.com:8501’ and thus, we were able see our hosted app.


app1 · Streamlit
Not secure | cloudproject.centralindia.cloudapp.azure.com:8501
28°C
Smoke
Search
ENG
IN
01:52
04-05-2023

Audio Processing App

This app allows you to upload a WAV file or record audio using your microphone.

Once the audio has been processed, it will perform sentiment analysis and summarization on the resulting transcript.

Choose an option:

● Upload WAV file
○ Record Audio

Choose a WAV file

Drag and drop file here
Limit 200MB per file • WAV

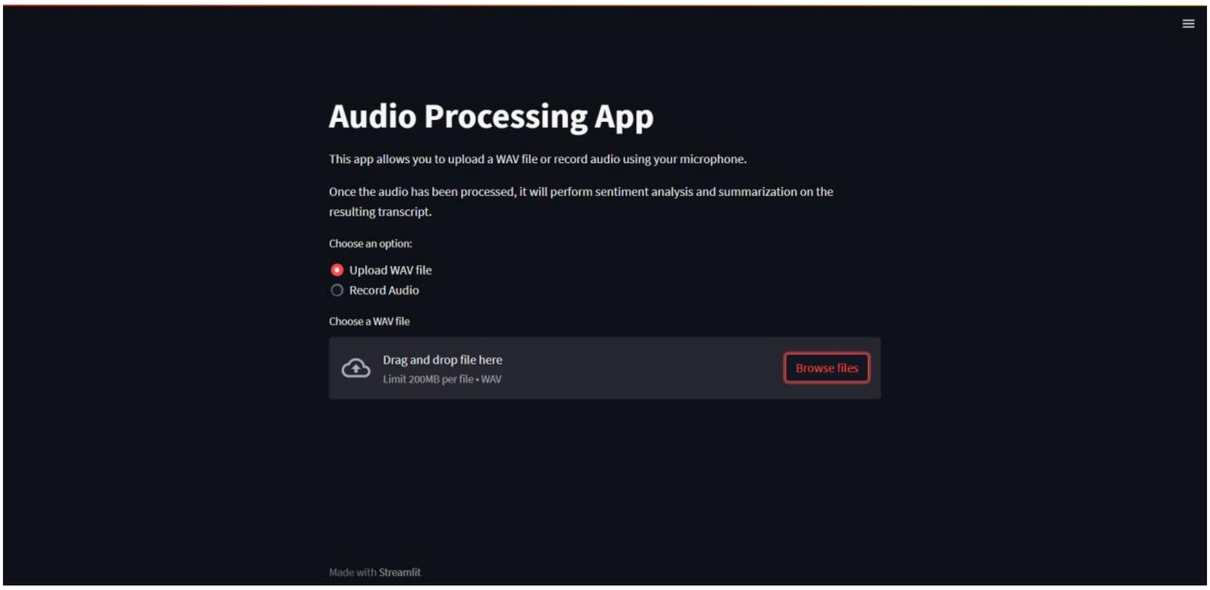
Browse files

Made with Streamlit

3. Results

The Output for our project are as follows:

Input page



Audio Processing App

This app allows you to upload a WAV file or record audio using your microphone.

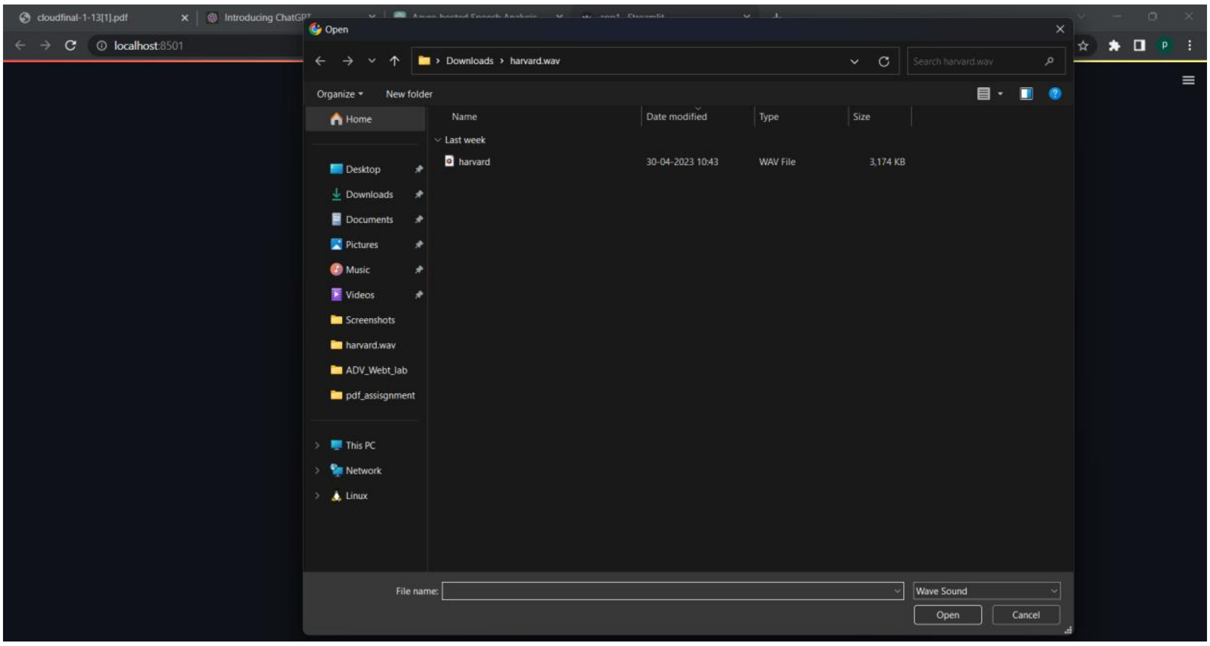
Once the audio has been processed, it will perform sentiment analysis and summarization on the resulting transcript.

Choose an option:

● Upload WAV file
○ Record Audio

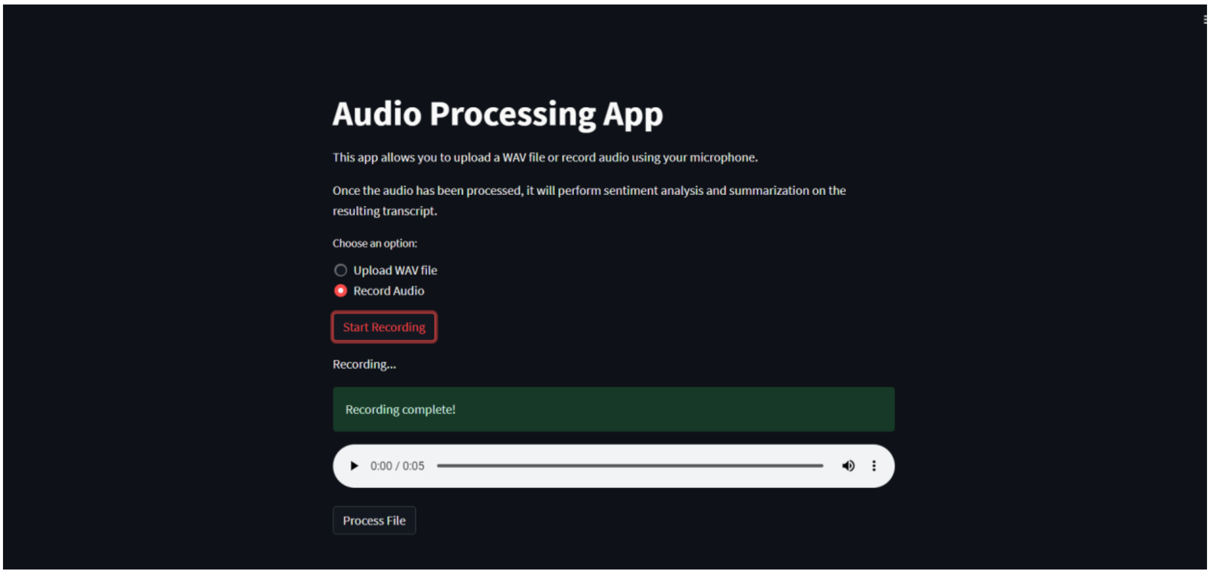
Choose a WAV file

Now below is fig is when user clicks on browse:



Name	Date modified	Type	Size
harvard	30-04-2023 10:43	WAV File	3,174 KB

Now below fig is for when user clicks on start recording:



Audio Processing App

This app allows you to upload a WAV file or record audio using your microphone.

Once the audio has been processed, it will perform sentiment analysis and summarization on the resulting transcript.

Choose an option:

○ Upload WAV file

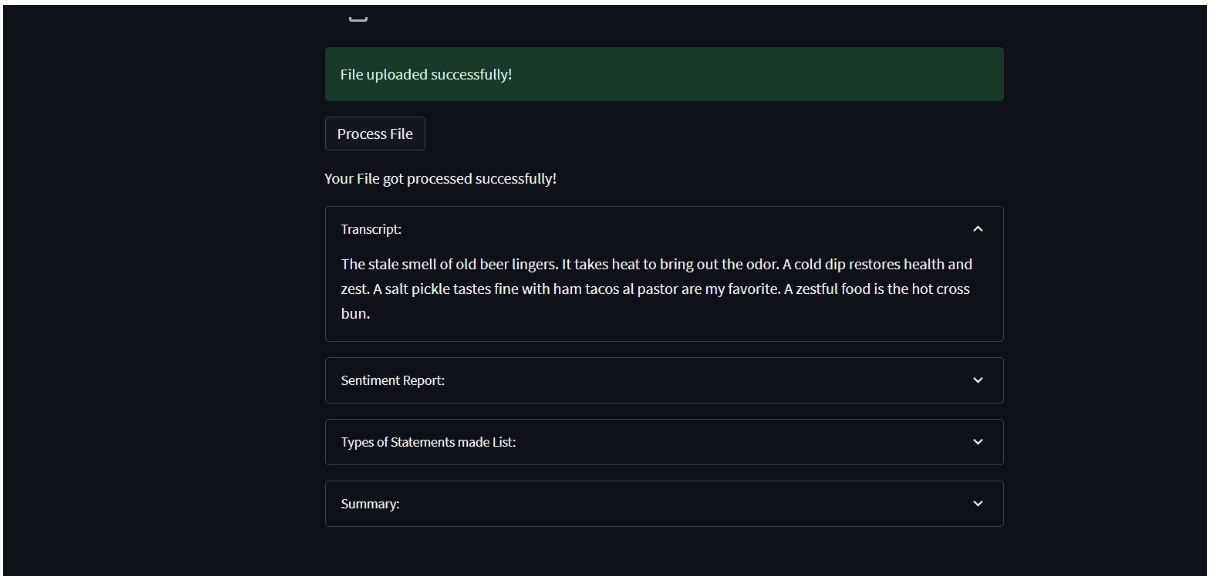
● Record Audio

Recording...

Recording complete!

Now, Below image is all reports generated:

Transcripts:



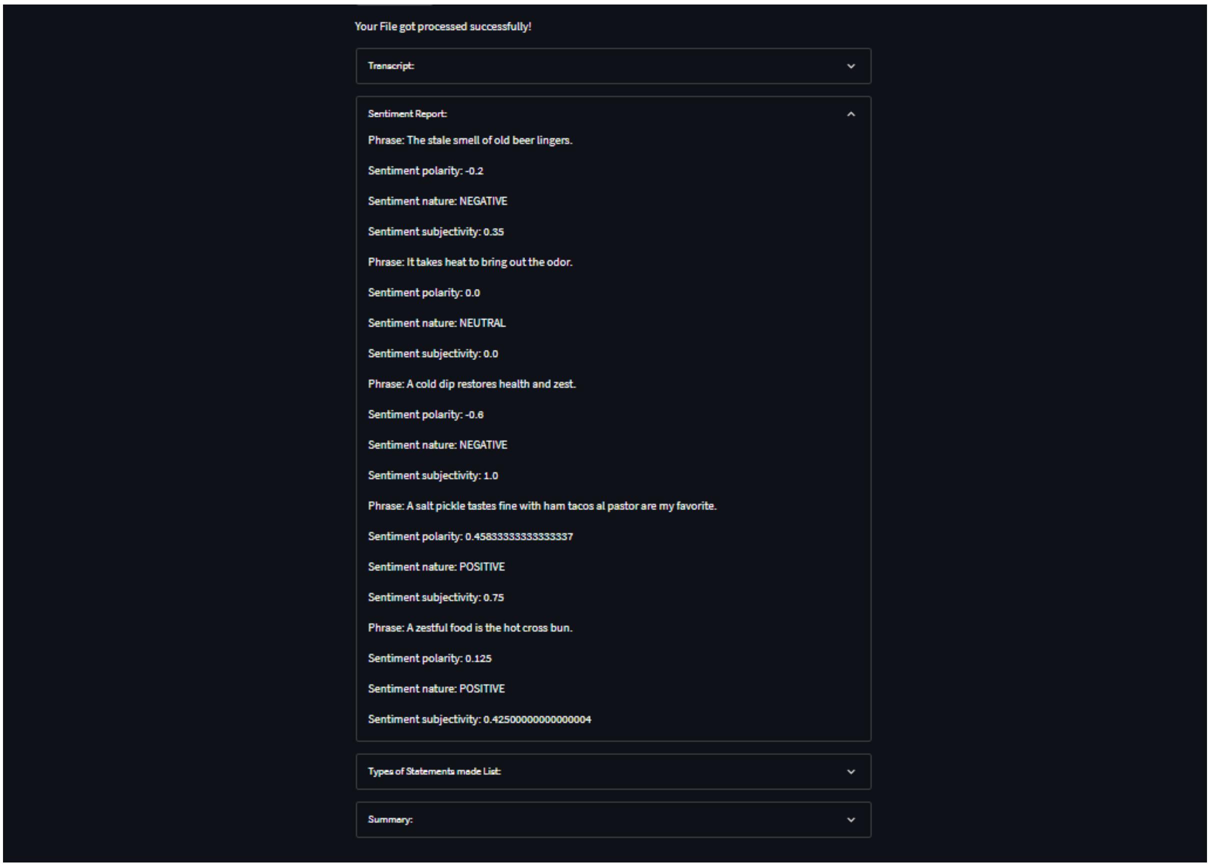
File uploaded successfully!

Your File got processed successfully!

Transcript:

The stale smell of old beer lingers. It takes heat to bring out the odor. A cold dip restores health and zest. A salt pickle tastes fine with ham tacos al pastor are my favorite. A zestful food is the hot cross bun.

Sentimental analysis report:



Your File got processed successfully!

Transcript:

Sentiment Report:

Phrase: The stale smell of old beer lingers.

Sentiment polarity: -0.2

Sentiment nature: NEGATIVE

Sentiment subjectivity: 0.35

Phrase: It takes heat to bring out the odor.

Sentiment polarity: 0.0

Sentiment nature: NEUTRAL

Sentiment subjectivity: 0.0

Phrase: A cold dip restores health and zest.

Sentiment polarity: -0.6

Sentiment nature: NEGATIVE

Sentiment subjectivity: 1.0

Phrase: A salt pickle tastes fine with ham tacos al pastor are my favorite.

Sentiment polarity: 0.45833333333333337

Sentiment nature: POSITIVE

Sentiment subjectivity: 0.75

Phrase: A zestful food is the hot cross bun.

Sentiment polarity: 0.125

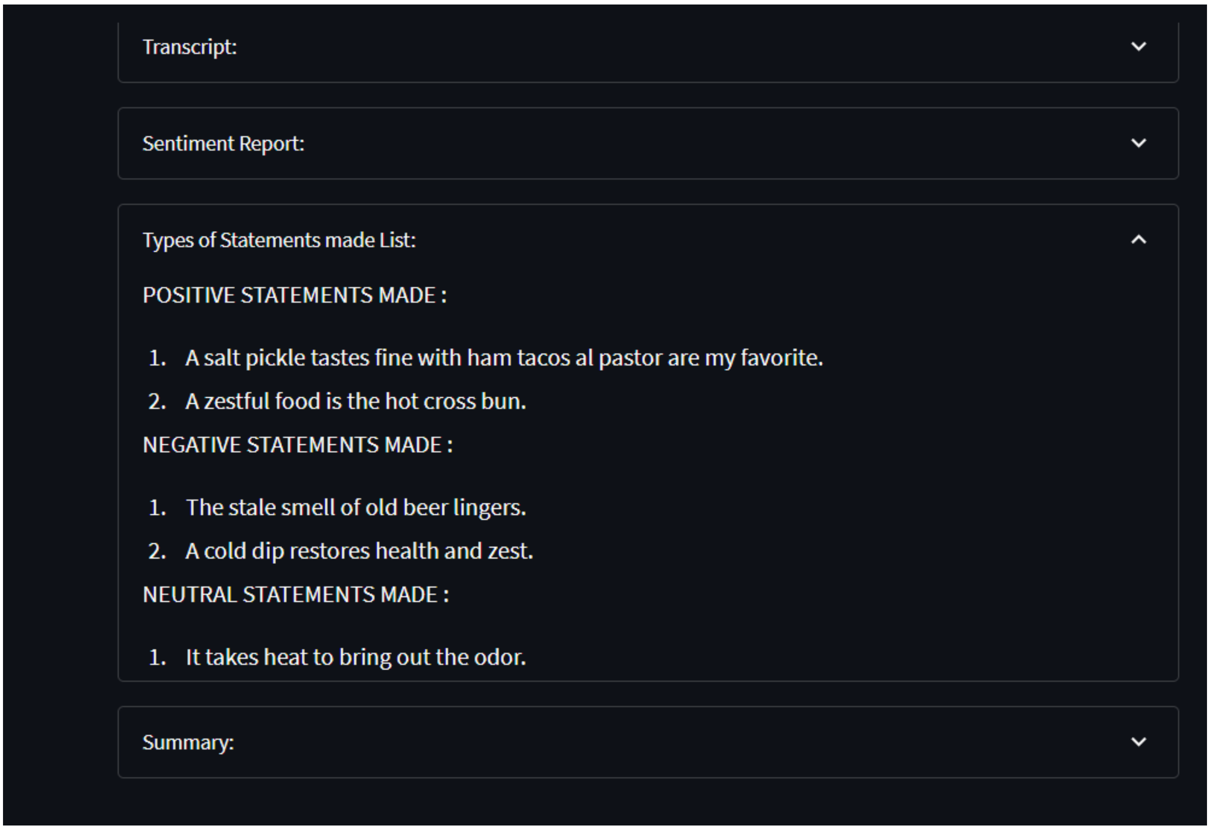
Sentiment nature: POSITIVE

Sentiment subjectivity: 0.42500000000000004

Types of Statements made List:

Summary:

Types of Statements:



Transcript:

Sentiment Report:

Types of Statements made List:

POSITIVE STATEMENTS MADE :

1. A salt pickle tastes fine with ham tacos al pastor are my favorite.
2. A zestful food is the hot cross bun.

NEGATIVE STATEMENTS MADE :

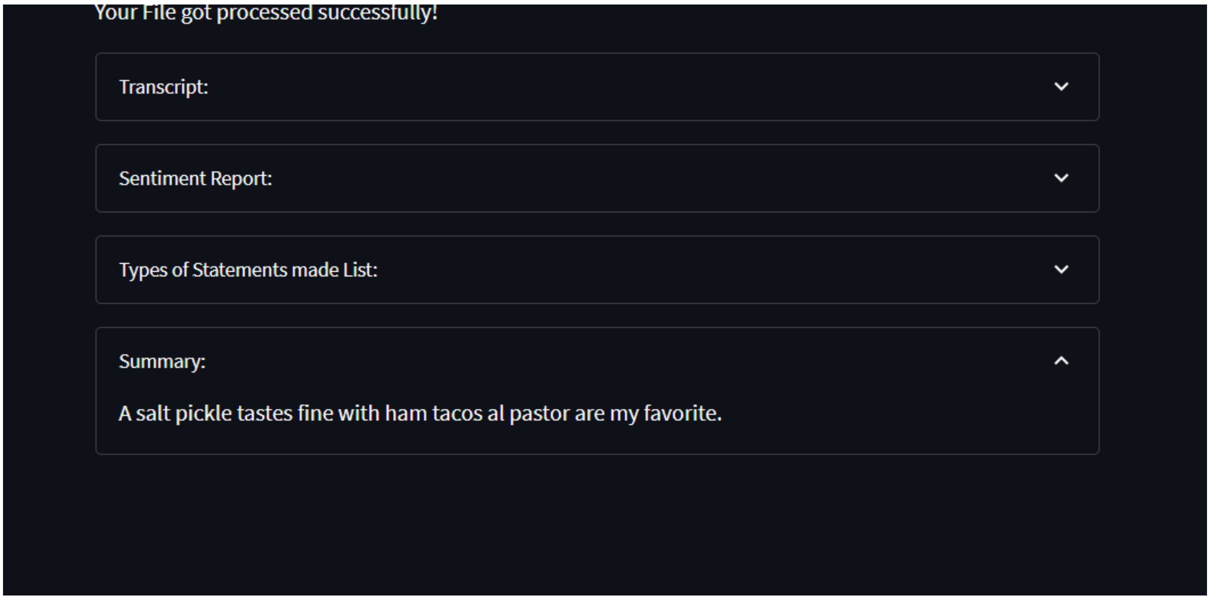
1. The stale smell of old beer lingers.
2. A cold dip restores health and zest.

NEUTRAL STATEMENTS MADE :

1. It takes heat to bring out the odor.

Summary:

Summarization:



Your File got processed successfully!

A salt pickle tastes fine with ham tacos al pastor are my favorite.

4. Future Scopes

1. Speaker identification: Currently, the system transcribes speech without identifying the speaker. Adding speaker identification capabilities would enable the system to identify and differentiate between different speakers in a conversation, which could be useful in applications like call center analytics or interview analysis.
2. Multilingual support: The system currently supports only one language. Expanding the system's capabilities to recognize and transcribe speech in multiple languages would increase its usefulness in a global context.
3. Real-time transcription: Currently, the system can only transcribe pre-recorded audio files. Adding real-time transcription capabilities would enable the system to transcribe speech as it's spoken, which could be useful in applications like live captioning for events or meetings.
4. Integration with other services: The system could be integrated with other services, such as machine translation or speech synthesis, to create a more comprehensive speech-to-speech translation tool.
5. Improved accuracy and speed: While the current system is highly accurate and efficient, there is always room for improvement. Future developments in speech recognition technology could be integrated into the system to increase accuracy and speed.

6. Customized analysis: The system could be extended to include more customized analysis based on specific use cases or industries. For example, sentiment analysis could be customized to identify emotions specific to a particular domain, such as healthcare or finance.

Overall, the potential future scope for this project is vast, and there are many directions in which it could be expanded and improved to meet the needs of various industries and applications.

5. Conclusion

In conclusion, this project has successfully demonstrated the capabilities of various technologies for speech recognition and analysis. With the integration of Streamlit, AssemblyAI's speech recognition API, TextBlob, and Spacy, the system is able to accurately and efficiently transcribe audio recordings, perform natural language processing tasks, and generate extractive summaries.

One of the key highlights of this project is the successful deployment on Azure, which provides a robust and scalable infrastructure for running the system. The use of Azure services, such as virtual machines, storage, and networking, ensures high availability, security, and performance of the system. Hosting the project on Azure also enables users to access the system from anywhere with an internet connection, and the system can be scaled up or down as needed to meet changes in demand.

Furthermore, this project has potential for future development, such as adding speaker identification, multilingual support, real-time transcription, and customized analysis. These developments could increase the versatility and usefulness of the system in various industries and applications.

Overall, this project showcases the power of integrating multiple technologies to create an effective tool for speech recognition and analysis. The successful deployment on Azure ensures that the system is accessible, secure, and scalable, making it a valuable tool for businesses and organizations in need of speech-to-text transcription, sentiment analysis, and text summarization.