INTRODUCTION

1.1 What is NATURAL LANGUAGE PROCESSING:

NLP allows computers to comprehend natural language in the same way that people do. Natural language processing employs artificial intelligence to accept real-world data, interpret it, and make sense of it in a way that a computer can understand, whether the language is spoken or written.

Natural language processing is divided into two stages: data pre-treatment and algorithm development.

Data Pre-processing is the process of preparing and "cleaning" text data so that machines can examine it. Pre-processing transforms data into usable form and highlights textual elements that an algorithm can use.

1.2 CLOUD COMPUTING and Today's relevance in Industry:

- Cloud technology means that companies can scale and adapt at speed and scale, accelerate innovation, drive business agility, streamline operations, and reduce costs.
- Cloud computing is helping society to cope with future problems such as managing big-data, cyber-security, and quality control.
- Cloud computing enables us to reduce and increase demands as per our requirement.
- The truth is that data kept in the cloud is likely to be safer than data stored on your hard drive. However, this does not negate the need for businesses and individuals to be watchful. providers have effective methods for securing the cloud, and they place a strong emphasis on encryption and cloud security.

1.3 Amazon Web Services (AWS)

Amazon web service is a platform that offers flexible, reliable, scalable, easy-to-use and cost-effective cloud computing solutions.

AWS (Amazon Web Services) is a comprehensive, evolving cloud computing platform provided by Amazon that includes a mixture of infrastructure as a service (IaaS), platform as a service (PaaS) and packaged software as a service (SaaS) offerings. AWS services can offer organization tools such as compute power, database storage and content delivery services.

Fig 2. Amazon Web Services

1.4 Challenges:

- **Data Extraction:** extraction of the relevant and correct information from unstructured or semi-structured data using Information Extraction (IE) techniques.
- **Expectation:** users expect more accurate and specific results from Relational Databases (RDB) for their natural language queries like English.
- **Translation**: challenge with language translation is not in translating words, but in understanding the meaning of sentences to provide an accurate translation.
- **Processing:** Data ambiguities add more challenges to contextual understanding. Semantics are important to find the relationship among entities and objects.

Literature Survey

2.1 Overview of Amazon Web Services

[1]Amazon Web Services provides a wide range of worldwide cloud-based solutions, including compute, storage, databases, analytics, networking, mobile, developer tools, management tools, IoT, security, and enterprise applications, all of which are on-demand and pay-as-you-go. Over 175 AWS services are offered, ranging from data warehousing to deployment tools, directories to content delivery. Without incurring any upfront capital costs, new services can be offered immediately. This enables corporations, start-ups, small and medium-sized organisation, and public-sector customers to gain access to the building blocks they require to respond rapidly to changing business needs.

Amazon Web Services (AWS) began delivering IT infrastructure services to businesses as web services in 2006, a term that has now become synonymous with cloud computing. One of the most significant advantages of cloud computing is the ability to replace upfront capital infrastructure costs with low variable costs that grow with your company. Businesses no longer need to prepare for and purchase servers and other IT infrastructure weeks or months in advance thanks to the cloud. Instead, they can quickly spin up hundreds or thousands of servers and give speedier results.

Today, AWS provides a highly reliable, scalable, low-cost infrastructure platform in the cloud that powers hundreds of thousands of businesses in 190 countries around the world.

[1] "Overview of Amazon Web Services - AWS Documentation." 12 Apr. 2021, https://docs.aws.amazon.com/whitepapers/latest/aws-overview/introduction.html. Accessed 29 Jun. 2021.

2.2 Overview of Google Speech-to-Text: Automatic Speech Recognition

[2]The Speech-to-Text API enables developers to convert audio to text in over 120 languages and variants, by applying powerful neural network models in an easy to use API.

State-of-the-art accuracy: Google's most advanced deep learning neural network algorithms for automatic speech recognition (ASR).

Speech adaptation: Customize speech recognition to transcribe domain-specific terms and rare words by providing hints and boost your transcription accuracy of specific words or phrases. Automatically convert spoken numbers into addresses, years, currencies, and more using classes.

Domain-specific models: Choose from a selection of trained models for voice control and phone call and video transcription optimized for domain-specific quality requirements. For example, our enhanced phone call model is tuned for audio originated from telephony, such as phone calls recorded at an 8khz sampling rate.

Streaming speech recognition: Receive real-time speech recognition results as the API processes the audio input stream from your application's microphone or sent from a pre-recorded audio file (inline or through Cloud Storage).

[2] "Speech-to-Text: Automatic Speech Recognition | Google Cloud." https://cloud.google.com/speech-to-text. Accessed 29 Jun. 2021.

2.3 Natural Language Processing and Cloud Computing

[3]Natural language processing (NLP) is a fast evolving subject of artificial intelligence and data science that focuses on speech and text processing technology. The goal of this branch is to provide methods for automatically analysis and presenting human language.

Cloud services for NLP analysis are becoming increasingly popular among scientists and other consumers interested in the topic.

The task of recognising named entities is successfully accomplished by services, allowing them to be used both independently and as part of a specialised system.

Some morphological and syntactical analysis tasks have previously been solved and can now be implemented as software, such as "Software as a Service" (SaaS) (SaaS).

Developers who use the SaaS model create fully customised software that is hosted on their own servers, giving them complete control over the product. The fundamental benefit of SaaS for users is that they do not have to support software or dedicate processing power, storage, or equipment.

Monthly subscriptions and specific pricing structures allow users to use the programme just when they need it and only pay for what they use.

[3] "(PDF) Cloud Services for Natural Language Processing." 7 Dec. 2018, https://www.researchgate.net/publication/329484496_Cloud_Services_for_Natural_Language_Processing. Accessed 29 Jun. 2021.

2.4 AMAZON TRANSCRIBE : CONVERT SPEECH-TO-TEXT

[4]Amazon Transcribe makes it simple for developers to integrate speech-to-text functionality into their apps. Computers have a difficult time searching and analysing audio data.

As a result, before recorded speech can be used in apps, it must first be transformed to text.

To perform this task in the past, customers had to deal with transcription providers who needed them to sign expensive contracts and were difficult to incorporate into their technology stacks.

Many of these companies still rely on antiquated technology that doesn't adapt well to changing settings, such as low-fidelity phone audio, which is frequent in call centres and leads to poor accuracy.

Automatic speech recognition (ASR) is a deep learning approach used by Amazon Transcribe to convert speech to text rapidly and reliably.

To establish a completely searchable archive, Amazon Transcribe can be used to transcribe customer support calls, automate subtitling, and provide metadata for video assets.

Medical speech to text features can be added to clinical documentation apps using Amazon Transcribe Medical.

[4] "Amazon Transcribe – Speech to Text - AWS." https://aws.amazon.com/transcribe/. Accessed 29 Jun. 2021.

2.5 AMAZON COMPREHEND : DISCOVER INSIGHTS AND RELATIONSHIPS IN TEXT

[5]Comprehend by Amazon is a natural-language processing (NLP) service that employs machine learning to find information in unstructured data. Instead of sifting through documents, the process is streamlined, making previously overlooked data easy to comprehend.

Critical components in data, such as references to language, people, and locations, may be identified by the service, and text files can be categorised by relevant subjects. You can detect customer emotion in your content automatically and precisely in real time.

This improves customer experiences by allowing for more informed, real-time decision making. Comprehend not only finds but also redacts and conceals any text that contains personally identifying information.

Comprehend is completely managed, allowing you to get up and running immediately without having to start from zero with model training. By utilising the power of machine learning, you may begin processing millions of documents in minutes.

[5] "Amazon Comprehend - Natural Language.... " https://aws.amazon.com/comprehend/. Accessed 29 Jun. 2021.

2.6 Django - Creating a Project

[6]Every web app you want to make in Django is called a project, and a project is a collection of applications. The MVT pattern is used to create an application, which is a collection of code files. Let's pretend we want to create a website; the website is our project, and the forum, news, and contact engine are apps. Because each application is independent, this structure makes it easy to migrate an application between projects.

Whether you are on Windows or Linux, just get a terminal or a cmd prompt and navigate to the place you want your project to be created, then use this code −

```
$ django-admin startproject myproject
```

This will create a "myproject" folder with the following structure −

```
myproject/
   manage.py
   myproject/
      __init__.py
      settings.py
      urls.py
      wsgi.py
```

[6] "Django Tutorial - Tutorialspoint." https://www.tutorialspoint.com/django/index.htm. Accessed 29 Jun. 2021.

2.7 Reading and Writing CSV Files in Python

[7]A CSV file (Comma Separated Values file) is a type of plain text file that uses specific structuring to arrange tabular data. Because it's a plain text file, it can contain only actual text data—in other words, printable ASCII or Unicode characters.

The structure of a CSV file is given away by its name. Normally, CSV files use a comma to separate each specific data value. Here's what that structure looks like:

```
column 1 name,column 2 name, column 3 name
first row data 1,first row data 2,first row data 3
second row data 1,second row data 2,second row data 3
```

If you understand the basics of reading CSV files, then you won't ever be caught flat footed when you need to deal with importing data. Most CSV reading, processing, and writing tasks can be easily handled by the basic CSV Python library. If you have a lot of data to read and process, the pandas library provides quick and easy CSV handling capabilities as well.

[7] "Reading and Writing CSV Files in Python – Real Python." https://realpython.com/python-csv/. Accessed 29 Jun. 2021.

2.8 Deep Neural Networks in Natural Language Processing (NLP)

[8]Over the years, success in deep learning for computer vision has followed closely on the heels of progress in natural language processing (or NLP, not to be confused with that NLP). We now have methods for transferring learning to new tasks with enormous pre-trained models like GPT-2, BERT, and ELMO, thanks to the introduction of pre-trained generalised language models. These and comparable models are doing real work in the world, both as a matter of routine (translation, transcription, and so on) and as a means of scientific discovery at the frontiers of knowledge (e.g. predicting advances in material science from publication text [pdf]).

Understanding language is a difficult subject, and despite extensive application in many fields, the difficulty of machine language understanding still poses many unsolved issues. Take a look at the ambiguous and unusual word or phrase pairs below. Despite the fact that the members of each pair ostensibly have the same meaning, they definitely express different subtlety. For many of us, the only nuance is a disregard for grammatical and linguistic correctness, but refusing to recognise everyday use meanings makes a language model look dumb.

[8] "The Unreasonable Progress of Deep Neural Networks in Natural...." 12 Jun. 2020, https://towardsdatascience.com/the-unreasonable-progress-of-deep-neural-networks-in-natural-language-processing-nlp-374443b21b00. Accessed 30 Jun. 2021.

2.9 Custom language models to supercharge Speech-To-Text performance

[9]You can utilise Custom Language Models (CLM) to train custom language models that target domain-specific use cases by submitting a corpus of text data. CLM is simple to use because it makes use of data that you already have (such as marketing assets, website content, and training manuals).
Train a custom language model for your speech-to-text use case using the data you already have**.**

Your text data is used by custom language models to increase transcription accuracy for your use case. Domain-specific text or audio transcripts can be included in your text data.

Website content, instruction manuals, and technical documentation are examples of domain-specific text data. Ground-truth audio transcripts make up audio transcript data. Ground-truth audio transcripts have been produced with extreme precision and are the most accurate depiction of the original audio.

[9] "Building custom language models to supercharge ... - Amazon AWS." 30 Sep. 2020, https://aws.amazon.com/blogs/machine-learning/building-custom-language-models-to-supercharge-speech-to-text-performance-for-amazon-transcribe/. Accessed 30 Jun. 2021.

2.10 Automatic Language Identification in Amazon Web Services

[10]Amazon Transcribe can now detect the main language in an audio recording automatically. This functionality eliminates manual tagging, allowing customers to create more effective transcription workflows. You may now use Amazon Transcribe to automatically recognise and transcribe voicemails, meetings, and any other type of recorded communication, in addition to the instances described above.

Amazon Transcribe can rapidly generate transcripts in the spoken language with a minimum of 30 seconds of audio without wasting time and resources on manual tagging. For all 31 languages, automatic identification of the prevailing language is possible in batch transcription mode. Language identification happens significantly faster than transcription, in a matter of seconds, thanks to sampling techniques. You can also limit the languages Amazon Transcribe attempts to recognise. If your company's call centre only receives calls in English, Spanish, and French, for example, limiting identifiable languages to this list will improve language recognition accuracy.

[10] "Amazon Transcribe Now Supports Automatic Language ...." 15 Sep. 2020, https://aws.amazon.com/blogs/aws/amazon-transcribe-now-supports-automatic-language-identification/. Accessed 30 Jun. 2021.

Methodology and Implementation

Problem Statement

1. To Obtain Meaningful Transcription for an Input Audio Stream.

2. Procure Unambiguous Information From Unpolished Data.

3. Format Information into Precise CSV Format.

4. Deploy the Obtained Solution in a User-friendly Interface.

Workflow

The initial solution was very case specific and offered very less dynamism.

Tools used: Speech Recognition, Google-trans.

- Then I delved into Google Cloud Services API which Showed much better results relative to using an SDK. Automatic Language Detection was in priority but not supported in Google cloud.

- Shifted focus from Google and considered other APIs such as sphinx and Microsoft Azure.

  While in amazon web services Indian regional Languages were supported and improved Transcription efficiency is noticed.

- Amazon Transcribe is a viable path. Transcription results is not satisfactory but AWS offered custom language models.

- Amazon Transcribe outputs is recorded and Amazon's data retrieval service, Amazon Comprehend is used for Data Analytics. Amazon Transcribe was able to detect language automatically.
- For deployment, Django framework is used as it promises clean user interface and seamless processing.

- Exception Handling is implemented in the django framework and the interface is made user-friendly.

- Further, Custom language model is developed using training Dataset. We predict Custom language model to be precise when domain specific jargons are involved.

- Further, Microphone support is added. Output efficiency increases substantially when recording is done realtime.

- Working of the Algorithm:

  The language code is passed on to Google-Cloud module to extract transcriptions which poses as raw data for Amazon Comprehend to extract quality data.

  This data is formatted into a csv file. This solution is integrated in Django Framework and deployed.

Supported Audio format:

.mp3 | .wav | .ogg

Supported languages:

English | Tamil | Hindi | Telugu

The overall process of the algorithm working:

1. A user opens the interface and reads out the product details for the order using the microphone in his device.
2. This audio data is uploaded to our server from where it is first sent to AWS Transcribe server for automatic language identification.
3. After which the audio file with the language code is sent to Google Speech-To-Text server for transcription and translation.

   For ex: if the user reads out the details in Tamil, the audio with the language code for Tamil is sent to Google server,

   Where the audio is first transcribed into Tamil texts and then the text is translated into English text.

4. From here, the transcription is taken next to amazon transcribe again for further transcription & for custom language modeling transcription.
5. After that, the final transcription is sent to amazon comprehend for data Analytics, where the data is analyzed to structure it into readable form.
6. After data entities and key phrases are analyzed and the data is being divided to tabulate it, it is converted into CSV format using import CSV function for a much better readability and usability.
7. After it has been converted into CSV format, it is sent to the interface for the user to download and it is also sent to the owner server database.

   This is the whole process which takes place in the backend using the algorithm framed.

This is a very important project because of the bigger purpose it is a part of.

A lot of users and owners of unorganized sectors are uneducated and it is hard for them to read and write humongous amount of data, and there will be a lot of errors too.

This product solves the issue by getting information just from the user's speech, he/she just has to speak to order any amount of products.

This product can be used in a lot of food delivery apps too.

Manual typing of data is a thing of the past, speech and image processing technology has a rising need in almost every app and website.

This is a pre-existing problem in our day-to-day life a lot of people haven't even noticed, manual labor to carry out our everyday tasks should be and can be avoided using technology and that is what this product strives to do.

This is the real-world problem solved using this project.

Technologies Used:

1. Amazon Web Services
2. Google Cloud Platform
3. Django Framework
4. Microphone module

Screenshots & Sample Result

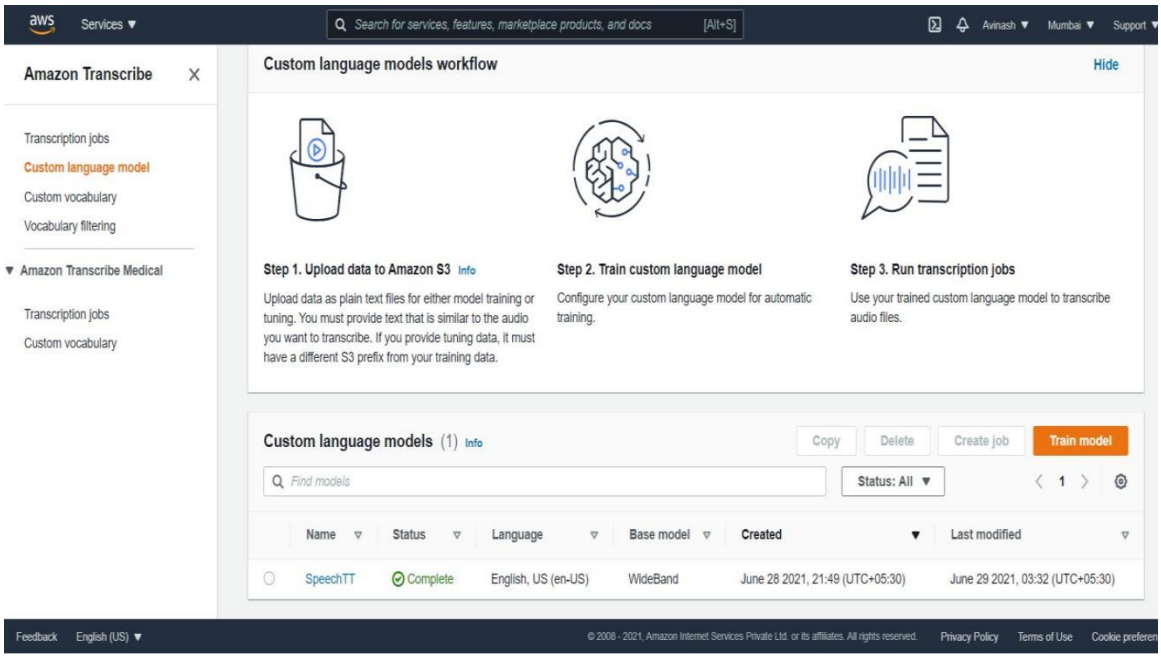

Fig.1 AWS custom model interface

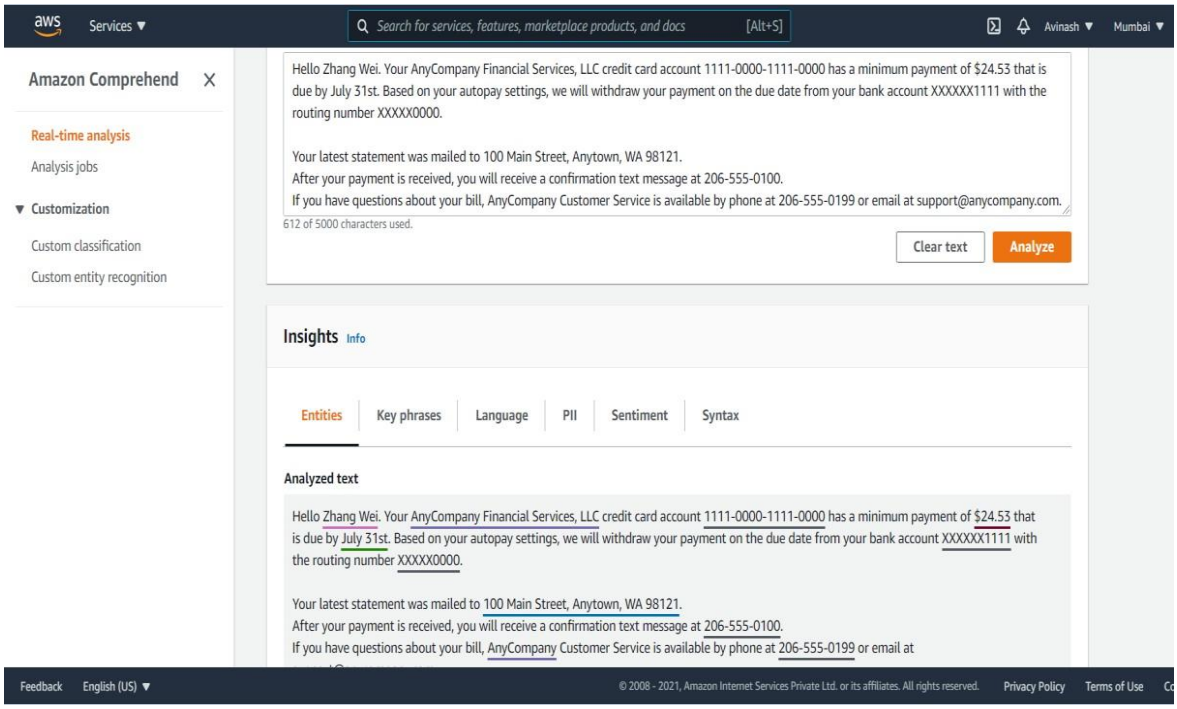

Fig.2 AWS comprehend interface

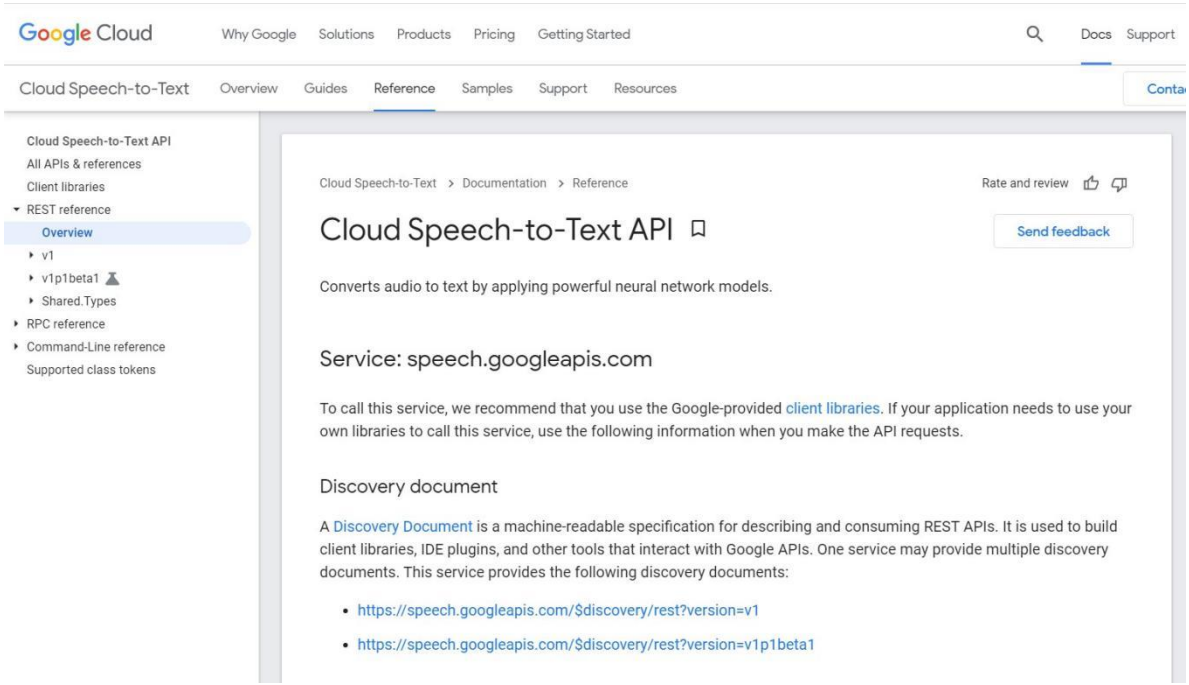

Fig.3 Google Speech-To-Text interface

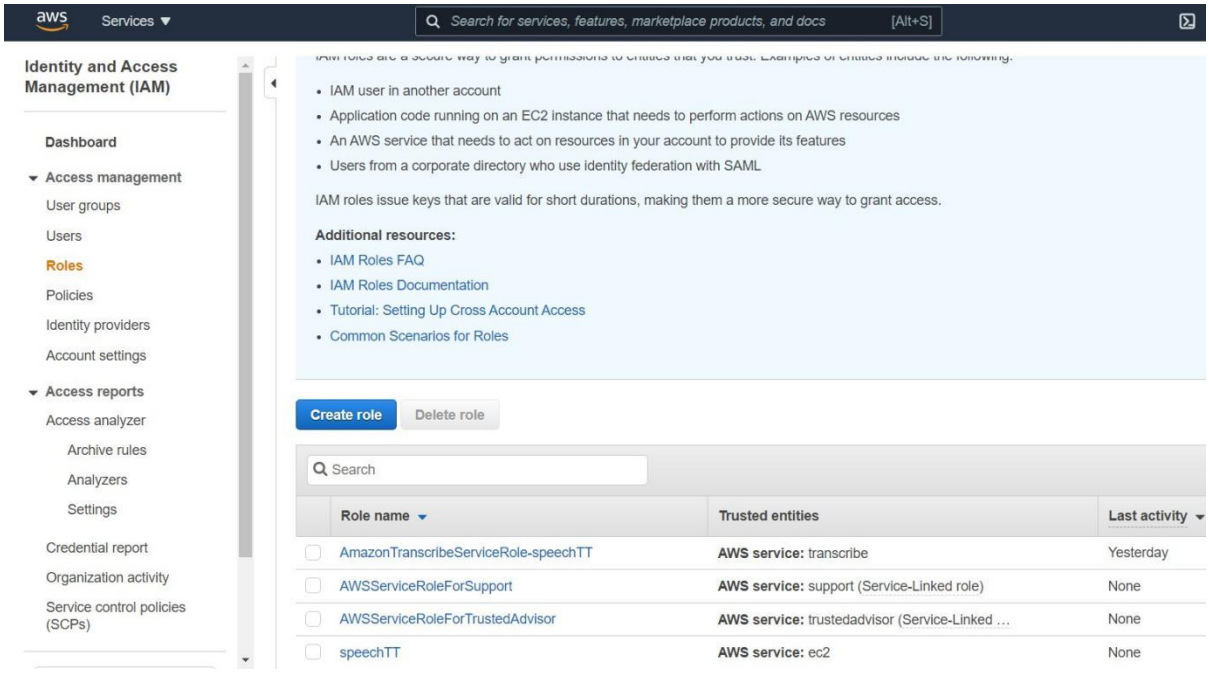

Fig.4 Setting IAM role for AWS Transcribe

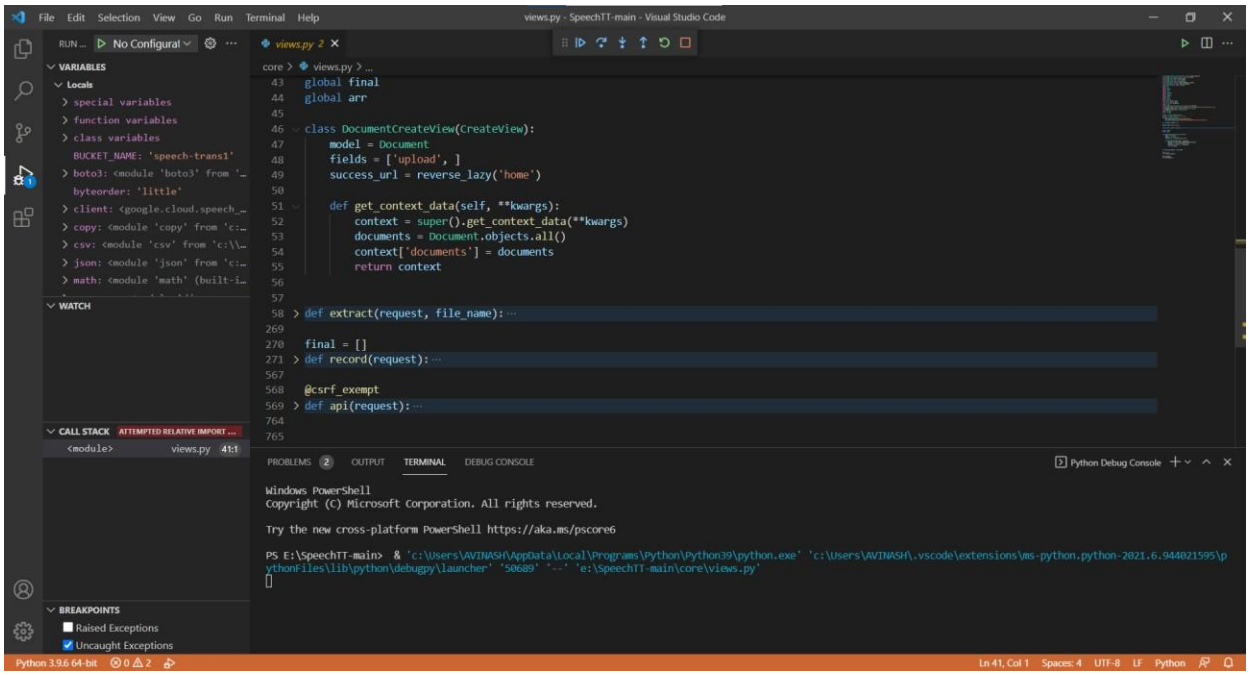

Fig.5 debugging interface

Fig.6 voice transcription interface

	A	B	C
1	Black Forest	1kg	500 rupees
2	German Black Forest	2kg	900 rupees
3	white forest	1kg	600 rupees
4	blackcurrant	2kg	800 rupees
5	chicken biryani	1 kg	300 rupees
6	choco almond	2kg	900 rupees
7	sugar	1kg	200 rupees
8	salt	2kg	300 rupees
9	Ginger	1kg	150 rupees
10	garlic	2kg	400 rupees
11	groundnut seeds	1kg	400 rupees

Fig.7 Sample output in CSV file format downloaded from interface

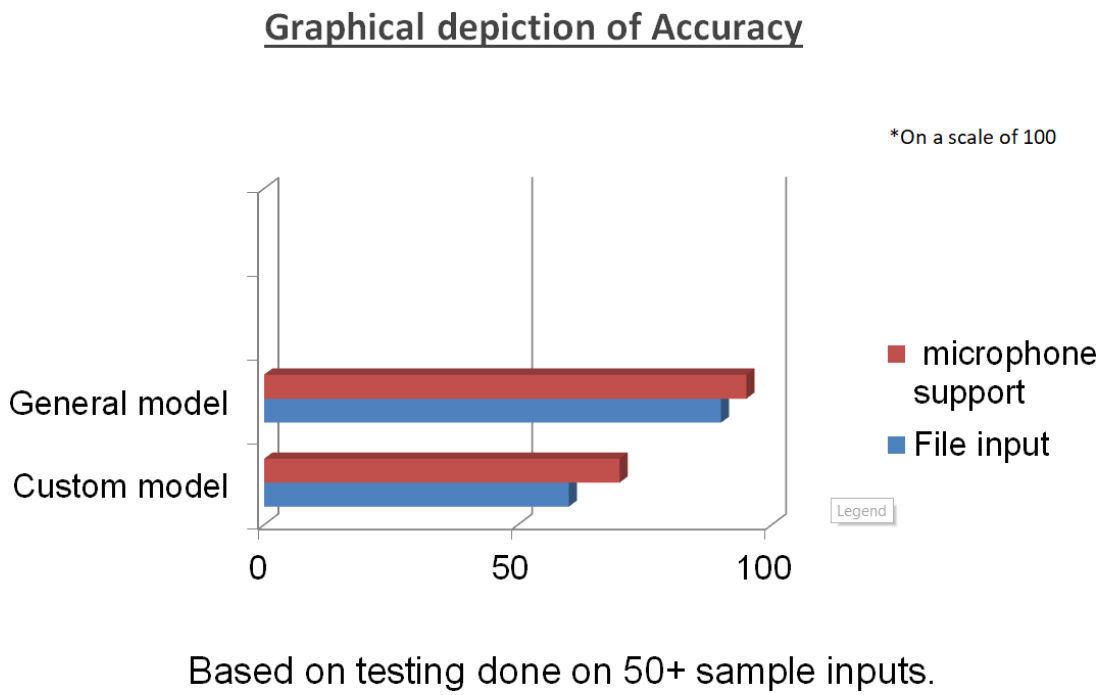

Fig.8 Accuracy depiction

CONCLUSION AND FUTURE ENHANCEMENT

The offered solution presents a one-of-a-kind platform to connect provisional stores from rural and semi-urban areas to the elsewhere population. Manual labor is terminated and every step of the process is automated with the help of cloud technology, thus making the entire process peerless.

The reason for choosing this project was my eagerness to solve a real-world problem faced by people in day-to-day basis. I hope this project does it by helping illiterate people and store owners to order products seamlessly and to increase profit margins and reduce the time spent.

This project can further be developed is by doing ground survey by giving people access to this product & note the experiences and use the data to create a better version of it & to carry out bug fixes too.

The more the product is used, the more data we get for training our model, which again will improve the usability of the product.

While doing this project, technically, I got well versed with the AWS Console and Cloud Technologies & also gained experience in Programming and Problem Solving. Overall, it presented a steep learning curve.

As for AWS and Cloud Technologies, The cloud computing business is likely to increase, which means there are numerous prospects for growth. According to Gartner, Inc., the global public cloud services market will rise 17.5 % in 2019 to $214.3 billion, up from $182.4 billion in 2018. According to recent Gartner surveys, cloud investments are among the top investment objectives for more than a third of enterprises. Cloud computing courses are also available at a number of Indian colleges.

REFERENCES

[1] Jabar H. Yousif, and Dinesh Kumar Saini ,"Part- Of-Speech Tagger Based Neural Networks" , Journal of Computing, Volume 3, Issue 2, pp59-65 ISSN 2151- 9617 ,NY, USA, February 2011.

[2] Ueffing, N. Macherey, K. & Ney, H. 2003. Confidence measures for statistical machine translation. New Orleans: MTSummit.

[3] International Journal of Computer Applications (0975 – 8887) Volume 77 – No.8, September 2013
Article: Natural Language Processing based Soft Computing Techniques.

[4] https://docs.aws.amazon.com/ ; Documentation for AWS.
Copyright Agent/Amazon Web Services
Amazon.com Legal Department

[5] Jabar H. Yousif, & Sembok, T., "Recurrent Neural Approach Based Part-Of-Speech Tagging", proceedings of International Conference on Computer and Communication Engineering (ICCCE'06), Vol 2, ISBN 983-43090-1-5© IEEE, KL-Malaysia, May 9-11, 2006.

[6] [Baldwin and Kim, 2010]
Timothy Baldwin and Su Nam Kim. Multiword expressions. In *Handbook of Natural Language Processing*. CRC Press, Taylor and Francis Group, 2 edition, 2010.

[7] Parsing by chunks. In Robert C. Berwick, Steven P. Abney, and Carol Tenny, editors, *Principle-Based Parsing: Computation and Psycholinguistics*, volume 44 of *Studies in Linguistics and Philosophy*. Kluwer, 1991.