Speaker Identification of Customer and Agent using AWS

G. Satyanarayana[1], Dr. Bhuvana J[2]

[1]PG Scholar, [2]Associate Professor,

[1, 2]Department of MCA, School of CS and IT, Jain (Deemed to be University), Bengaluru, Karnataka, India

ABSTRACT

As everyone knows that Sentimental analysis plays an important role in these days because many start-ups have started with user-driven content [1]. Only finding the voice is not be the real time scenario so finding the Sentiment analysis of agent and customer separately is an important research area in natural language processing. Natural language processing has a wide range of applications like voice recognition, machine translation, product review, aspect-oriented product analysis, sentiment analysis and text classification etc [2]. This process will improve the business by analyze the emotions of the conversation with respect to the customer voice separately and also agent voice separately. In this project author going to perform speaker identification and analyze the sentiment of the customer and agent separately using Amazon Comprehend. Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to extract the content of the voice. By using the speaker identification author can extract the unstructured data like images, voice etc separately so it is easy to analyze the business performance. Thus, will identify the emotions of the conversation and give the output whether the customer conversation is Positive, Negative, Neutral, or Mixed. To perform this author going to use some services from Aws due to some advantages like scaling the resources is easy compare to the normal process like doing physically such as support vector machine (SVM). AWS services like s3 is a object data store, Transcribe which generate the audio to text in raw format, Aws Glue is a ETL Service which will extract transform and load the data from the S3, Aws Comprehend is a NLP service used for finding sentiment of audio, Lambda is a server less where author can write a code, Aws Athena is a analyzing tools which will make complex queries in less time and last there is quick sight is a business intelligent tool where author can visualize the data of customers and also agents.

KEYWORDS: Sentimental analysis, NLP, Speaker identification, S3, Transcribe, Aws Glue, Aws Comprehend, Lambda, Athena, Quick sight, ETL

I. INTRODUCTION

Speaker Identification is one of the important thing which will help the business performance it means identifying the calls of the customer and separating the voice of the customer and agent separately and finding the sentiment of the customer its easy to analyze the business performance and also can know the drawbacks that they are making and can make them as a advantages. They are many which author can use for speaker identification without using cloud by there are some disadvantages when it comes to normal procedures like GMM and SVM and the disadvantages are a)firstly better security can't provide by using some brute force attack can get the access b)the GMM can provide only six speakers in the audio can be identified and if it is more than it will stop working c) For all doing this need physical servers and need to configure manually so it will take lot of time. The main objective of the project is by using cloud author can solve these all problems by using AWS services like AWS simple storage service (s3 which is used for dataset it also useful to secure the data by using key management service (KMS) and also for high availability by providing the Cross region replication so if the data is deleted in bucket it will help as a backup and if you fell the data is becoming more you can sent to the glacier which is used for long term storage and also for compressed data. AWS Transcribe is used to convert audio file to text file from S3 bucket and it will generate the text file to the other s3 bucket author can trigger by writing the lambda code and it will trigger the lambda function. AWS Comprehend is a NLP service which is used for finding the sentimental analysis of the customer and agent separate this is said to be speaker identification. AWS Glue is Extract Transform Load (ETL) service where author can extract data for analytics. AWS Athena is a query service is used to make a query for the data present in S3 bucket based on the S3 bucket data. AWS Quick sight is a data visualization tool so can use this and can visualize the data of customer and agent separately. Finally author can use Identity Access Management (IAM) where can create the roles and allows for specific services to that role and can create two factor authentication (TFA) which will help the use if the hacker know the password he can't access the account without knowing OTP. Cloud watch is a monitoring tools if there is anything gone wrong by seeing the logs can identify the errors can able to solve easily.

II. LITERATURE SERVEY

A. GAUSSIAN MIXTURE MODEL: [4]

Gaussian Mixture Model (GMM) is a function which is used for speaker identification. GMM provide more accuracy compare to other classification and regression algorithm. GMM Is having few disadvantages compare to latest algorithms and it is better to use when author is going to use the supervised learning so it can be easily handled compared to other algorithms and it is difficult to handle when it comes to unsupervised learning. When it comes to large data sets the Gaussian Mixture Model will not be suitable. Coming to speaker identification if the speakers in the audio should not be more than six in GMM if it is more than six speakers the algorithm will stop working and this function should configure manually and should tell the number of speakers should identify so classification is difficult due to this disadvantages the authors moved with the support vector machine.

B. SUPPORT VECTOR MACHINE: [3]

Support vector machine (SVM) is another supervised algorithm which is used for classification and also for regression analysis. SVM is difficult to train the data when it comes to long dataset it will take longer time but better when we compare with the GMM method. Since the final model is not so easy to see, we cannot do small calibrations to the model hence it's tough to incorporate our business logic. The SVM hyper parameters are Cost -C and gamma. It is not that easy to fine-tune these hyper-parameters. It is hard to visualize their impact.
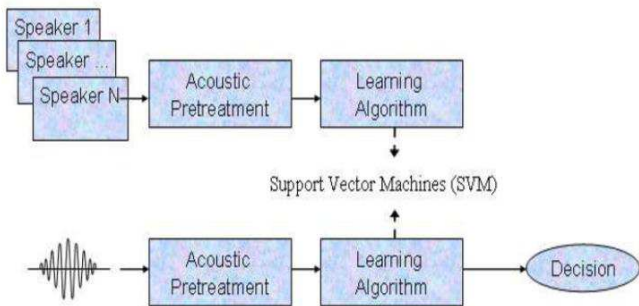

Fig.no:2.1 Super vector Machine Method

III. PROBLEM IDENTIFICATION

Previously to generate transcribe file they are using the dynamo DB which is a database in AWS and that cost you more and coming to architecture it is huge with many services. Training with large datasets will be difficult coming to the speaker identification it not possible if the speakers in the audio is more than six if you try more than the algorithm will going to stop working. Need to use complex formulae to find the speaker identification. Better security will not be possible when compare to GMM and SVM algorithms and that can crack by using some algorithms like brute force attack. Better scaling will not be possible as it cost more to buy physical resources. Better Automation is not possible by using the GMM and SVM.

IV. PROPOSED SYSTEM

In this proposed system there is four modules a) Generating text file by using AWS transcribe b)performing the speaker identification on the generated data c) Finding sentimental analysis after doing speaker identification d) Finally generating metadata from the given file and visualize the generated output. By using this proposed system author can reduce cost instead using database going to use simple

storage service (S3) data store. Can be scalable if the recourses needed no need to pay for the license all will be taken care by AWS. Author can perform the speaker identification by reducing the architecture which will reduce the complexity. Training a large data set will be easy when compared to GMM and SVM. No need to use Complex formulae. The proposed system will provide better security by using Identity Access Management (IAM) and also the key management service (KMS) which will provide security to the data which the data at rest and also by using s3 author can build high availability by creating the cross region replications which will help as a backup data.
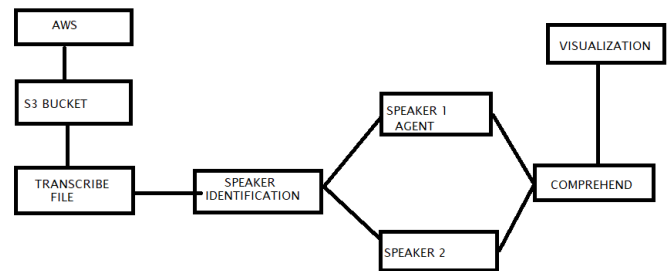
V. FLOW DIAGRAM

Fig no: 5.1 Flow Diagram

According to the flow chart all this process is done in AWS. Starting with S3 Bucket S3 is simple storage service where it can store the Audio files video files when the audio is uploaded to S3 bucket the lambda function will trigger the S3 bucket and generate the Text file using AWS Comprehend. The author will do speaker identification and separated with the customer and agent voice separately. Finally Author will find sentiment of the given file which is stored in another S3 bucket and visualize the data of customer and agent separately.

Visualization can be done in two ways can use the quick sight if the process is happening continuously the other method is using the excel sheet the data is copied to excel file and visualize the data by using the tableau software and this will take lot of time.

VI. METHODOLGY

In this proposed system author is going to upload a audio file into the s3 bucket and that bucket will be triggered from the lambda by writing the code author can generate a transcribe file from the audio which is generated in the transcribe bucket and that file consist of the raw data having of customer voice and agent voice and it is difficult to find out the result from the transcribe file so author is going to use speaker identification and separating the customer voice and agent voice separately and that two files will be saved in the other s3 bucket from that files able to find the sentiment of the customer and agent separately.
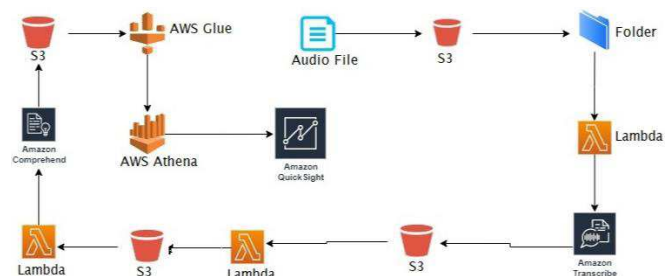

Fig no: 6.1 Architecture

After generating the sentiment file author can use to generate raw data by using AWS Glue It is a ETL service where it will extract, transform and load the data from s3 bucket by using Athena can make the SQL query which will saves the lot of time and Finally going to use the AWS quick sight which will help to visualize the generated file and for security purpose author is going to use identity Access Management (IAM) by creating to roles for transcribe, Comprehend full access and also creating the Cross region replication which will provide more high availability the things which u will upload that will automatically to the other s3 bucket this can use as a backup and also need to enable the key management service (KMS) which will provide a security to the data which is at the rest.

.

VII. IMPLEMENTATION

This is JSON file content generated to S3 bucket where having the information that is spoken by both the customer and also agent content as a group.
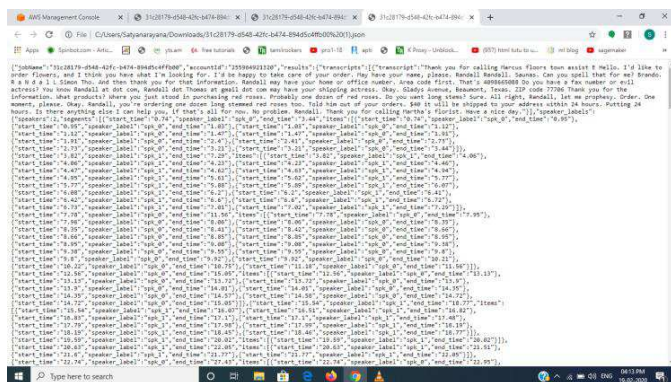

Fig no: 7.1 Generating transcribe file

Speaker identification is done by using the transcribe file separated the customer text and agent text separately in one file.
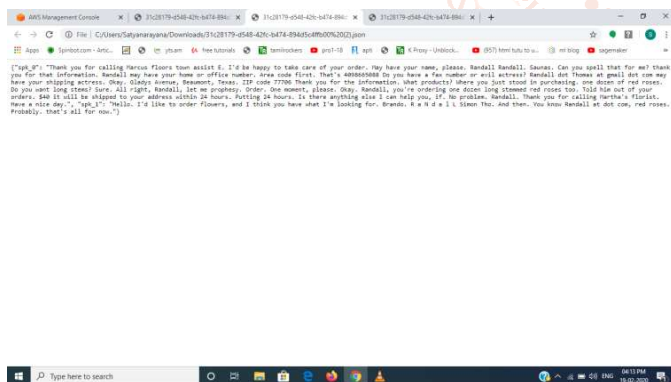

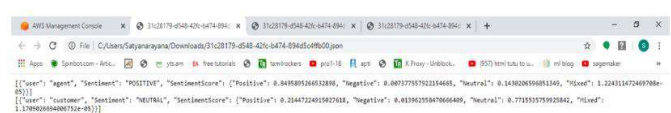**Fig no: 7.2 Performing Speaker identification**

Fig no: 7.3 Generating the Sentiment file

VIII. RESULT

Fig no: 8.1 Bar chart of Customer and Agent

IX. CONCLUSION&FUTURE WORK

Taking the audios and finding the sentiment of customer and agent separately is very important to improve the business. In this paper author is proposed to use aws cloud to reduce the cost by reducing the database and also some licence cost. It is automated and automation will reduce the work for the developers and also for administrators. In this proposed to identify the speaker identification author used some aws service as discussed above. In this proposed system the accuracy was 95% is the highest but the proposed system will solve the many security issues and also the cost.

X. REFERENCE

[1] B. Chethana Swetha, S. Divya, J. Kavipriya, R. Kavya, Dr. A. Abdul Rasheed, "*A NOVEL VOICE BASED SENTIMENTAL ANALYSIS TECHNIQUE TO MINE THE USER DRIVEN REVIEWS*", International Research Journal of Engineering and Technology.

[2] Indhraom Prabha M, G. Umarani Srikanth, "*Survey of Sentiment Analysis Using Deep Learning Techniques*".

[3] Nikunj V. Tahilramani, Ninad Bhatt, "*Performance Analysis of Speaker Identification System With And Without Spoofing Attack of Voice Conversion*" 2018 2nd International Conference on Micro-Electronics and Telecommunication Engineering.

[4] Rania Chakroun1, Leila Beltaifa Zouari1, Mondher Frikha1, and Ahmed Ben Hamida1 "*A novel approach based on Support Vector Machines for Automatic Speaker Identification*"

[5] P. Chaudhary and M. Vagadia, "*A Review Article on Speaker Recognition with Feature Extraction*", International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 2, February 2015,pp. 94-97.

[6] K. Kau and N. Jain, "*Feature Extraction and Classification for Automatic Speaker Recognition System – A Review*".

[7] *Hyung-suk kim "Linear Predictive Coding is All-Pole Resonance Modeling" Center for computer research in music and acoustics Stanford university.*

[8] *Mohamed Abou-Zleikha, Zheng-Hua Tan, Mads Græsbø Christensen, SørenHoldtJensen, "ADISCRIMINATIVEAPPROACHFORSPEAKERSELECTIONINSPEAKER DE-IDENTIFICATIONSYSTEMS" 2015 23rd European Signal Processing Conference (EUSIPCO).*