A

DISSERTATION REPORT

on

Speech and Text Data Analysis and Visualization Using Machine Learning Algorithm

Submitted By

Sangh Mitra Singh Rathore

2017PEB5477

Under the supervision of

Prof. Lava Bhargava

Professor, ECE Department,MNIT Jaipur

Submitted in partial fulfilment of requirement of degree of

MASTER OF TECHNOLOGY

Department of Electronics and Communication Engineering

Malaviya National Institute of Technology, Jaipur, India

May,*2019*

CERTIFICATE

Department of Electronics and Communication

Malaviya National Institute of Technology, Jaipur Rajasthan – 302017

This is to certify that this Dissertation report entitled **"Speech And Text Data analysis and visualization using Machine Learning Algorithm"** by **Sangh mitra Singh Rathore (2017PEB5477)**, is the work completed under my supervision and guidance, hence approved for submission in partial fulfilment for the award of degree of **Master Of Technology** in **Embedded Systems** to the Department of Electronics and Communication Engineering, Malaviya National Institute of Technology, Jaipur in the academic session 2018- 2019 for full time post-graduation program of 2017-2019. The contents of this dissertation work, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Date:

Prof. Lava Bhargava

Professor, E.C.E Deptt.

MNIT, Jaipur.

DECLARATION

I, hereby declare that the work which is being presented in this project entitled "Speech and Text Data Analysis and Visualization Using Machine Learning Algorithm" in partial fulfilment of degree of Master of Technology in Embedded Systems is an authentic record of my own work carried out under the supervision and guidance of Prof. Lava Bhargava in Department of Electronics and Communication Engineering, Malaviya National Institute of Technology, Jaipur. I am fully responsible for the matter embodied in this project in case of any discrepancy found in the project and the project has not been submitted for the award of any other degree. I also confirm that I have consulted the published work of others, the source is clearly attributed and I have acknowledged all main sources of help.

Sangh Mitra Singh Rathore
(2017PEB5477)

ACKNOWLEDGEMENT

I am grateful to my supervisor Prof. Lava Bhargava for his constant guidance and encouragement and support to carry out this work. His excellent cooperation and suggestions provided me with an impetus to work and made the completion of work possible. He has been great source of inspiration to me, all through. I am very grateful to him for guiding me how to conduct research and how to clearly & effectively present the work done. I would like to express my deepest sense of gratitude and humble regards to our Head of Department Prof. D. Boolchandani sir for providing necessary facilities in the Department. I am very thankful to all the faculty members of ECE Department, MNIT for their valuable assistance and advice. I would also like to thank my friends and seniors for their support in discussions which proved valuable for me. I am indebted to my parents and family for their constant support, love, encouragement, and sacrifices made by them so that I could grow up in a learning environment. A Prayer for a heartly Thanksgiving to GOD Finally, I express my sincere thanks to all those who helped me directly or indirectly to complete this work successfully.

Sangh Mitra Singh Rathore
(2017PEB5477)

Contents

List of Figures

List of Tables

Abstract

Audio sense analysis using automated speech recognition is an emerging research field where opinions or feelings are natural audio displayed by a speaker is detected. Still, text-based sentiment detection is relatively unexplored and very less used in the physical world despite its huge possible implementation area because talking and listening improves information if it is nice then it encourages to listen to you which helps in better understanding the things and to form a relationship. Talking have many emotions in it, this is not just a simple talk but also have anger, fear, criticism, or blame hence can be categorized in positive, neutral or negative communication. In this new architecture, we perform a predictive visual analysis of big data collected from various speaking sources and carry out graphical and non-graphical analysis and visualization. It is using machine learning algorithms. We accomplish weighted word cloud visualizations, which gives improved semantic insights. The results obtained by this work can help in finding good teacher, orator, speaker, and examine self-talking skill.

List of Abbreviations used

1.	NLP	Natural Language Processing
2.	RVIN	Random Valued Impulse Noise
3.	FVIN	Fixed Valued Impulsive Noise
4.	SMF	Simple Median Filter
5.	AMF	Adaptive Median Filter
6.	PSMF	Progressive Switching Median Filter
7.	FEMF	Fast and Efficient Median Filter
8.	FEDBA	Fast and Efficient Decision Based Algorithm
9.	MDBUTMF	Modified Decision Based Unsymmetrical Trimmed Median Filter
10.	PSNR	Peak Signal to Noise Ratio
11.	MSE	Mean Square Error
12.	SSIM	Structure Similarity Index
13.	ICM	Iterated Conditional Modes
14.	MRF	Markov Random Fields
15.	BN	Bayesian Network
16.	PGM	Probabilistic Graphical Models

Chapter 1

Introduction

1.1 Introduction

In today's scenario everyone become technologically powerful and educated, we become more connected and we can approach the entire world. Hence if we find someone to listen to his thoughts or a teacher, then the world is open to us we can find many options for a single requirement. Here it is important to find the right one for fulfilling our requirement. Here some kind of tool should be present which can help us by predicting the quality, level and sense of speaking of someone by analysing his/her previous speech or talking data set. This tool can find person and organization having a good economic, social and politic interest, opinion or ideology that is positive negative or neutral for the society or for a single person. According to the study, an average human speaks more than 5000 words per day it leads to generate massive data which is called big data. Once acquire the data, which is in high volume high variety and veracity information assets and by some innovative process this information process then that produce greater insights and better decision making. The text-based review is just one of the various ways by using it people can express their feelings/opinions about product, sports news and issues related to social concern, Audio / Videos/ text are emerging key methods to express opinion. There are millions of video products on YouTube film review, product United Nations boxing, political, social issues analysis and opinion on them. There are many audio platforms on the internet where people express their opinions. In addition, an audio format of info is more powerful than text and video is powerfull then audio in many situations. Because they provide a rich and natural information of the speaker about their opinion. This huge resource is unused and removing feeling/views about the Society, specific product or opinions, political or social conditions can be so useful for info data analysis. Detection of emotion in the audio mode is still unexplained. Audio based emotion finding is an budding and challenging area. In this study, 1 strong methods are presented. Remove expressions/opinions from various audio sources. A hybrid system has been developed which uses a strong automatic speech Recognition (ASR) system with NLP based emotion analysis techniques to detect the sense of the audio stream Contrary to text-based sources, audio sources have higher degrees Changes in terms of opinions as well as expressing the way of expression of opinion. There are many challenges for emotion figuring out in highly natural speech sources. Including: 1) Domains and vocabulary: The speaker can express opinions about any subject (e.g. Product, Movie, Politic, Social issue, game, etc.) Therefore, the Audio Speech Rrcognition system should be well suitable to handle a wide range of domains and terminology. The language model should be dealling with all elements 2) Speaker variation and speaker pronunciation: Audio Signal Recognition system speaker should be strong for variability in which one is included a broad range of English language accents from around the world. 3) Noise audio and channel: Incompatible recording device and distance of different modes / recordings, incompatible audiovisual and background environment conditions challenge the problem of finding the emotion. Other than this, Background music / talk, intentionally music mixing, issues of relation make the problem more difficult. 4) Natural and Intuitive: Distinguish Audio Spirit Natural and perfect speaker setting and all posible speaking or discussing scenario (i.e., 1-way talk, 2-way talk, or public speech etc.) is the challenge.

1.2 Motivation

Now a day's the Machine learning Algorithms are being used in every area the concept Machine learning was proposed in the middle nineties, hence the theoretical idea is not new but as processing devices increasing machine learning is implementing more and more. Hence people are excited to use this concept in every area, researchers are trying to take help of machine learning in such areas where problems were not that simple and needed so much effort to solve challenges. Machine learning basically deals with big data and categorizations of that data, which is also called big data analysis. This big data is very helpful to solve many real-time problems by predicting the most nearby result. This work is solving one of the most challenging problems like human emotions and sentiments detection. Which was probably an impossible thing before, for a machine like a computer and robots. Hence now by using machine learning, we can make certain decision which was dependent upon human sentiments.

1.3 Scope of Work

Data is information and these kinds of information keep on gathering in either text

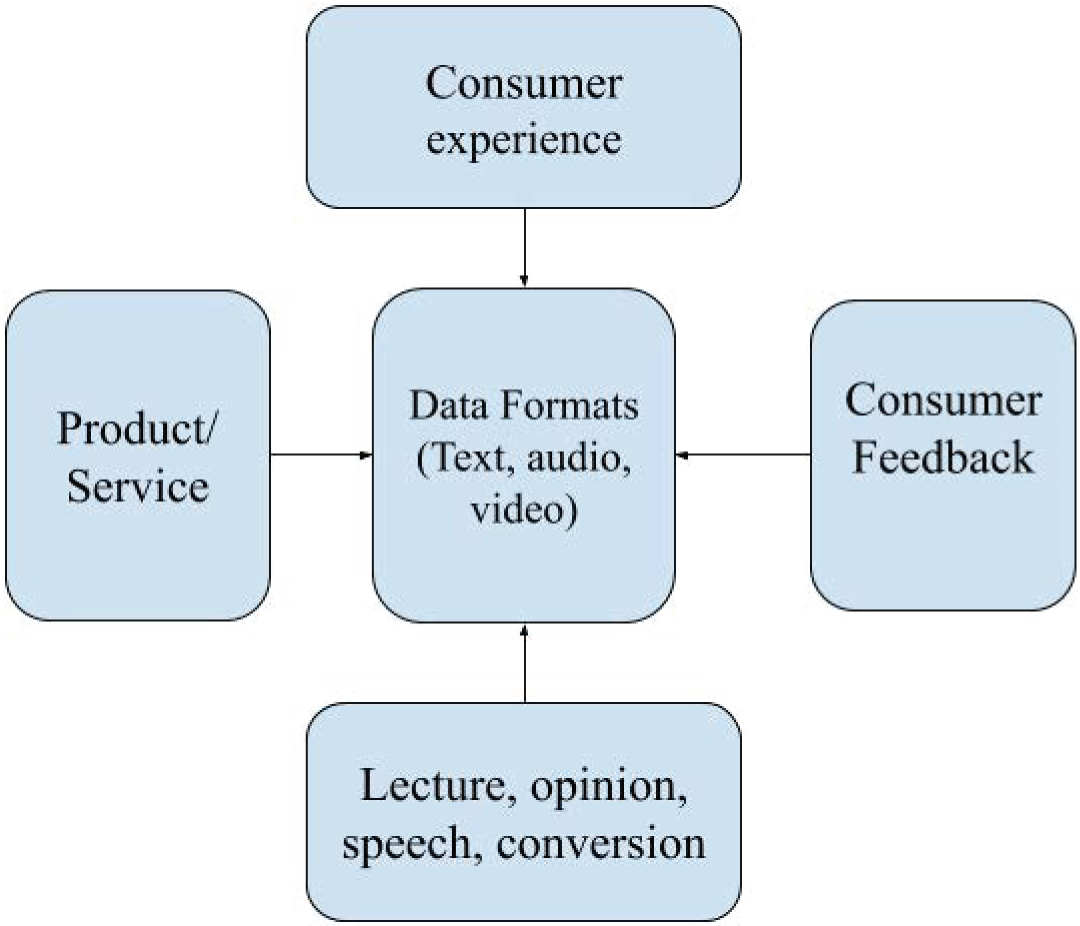

Figure 1.1: Data sources from various platform

form or audio form, all such information some time stores at big servers (if we considerweb data) or some kind of storage devices. The data producer and consumer both intended to find valuable data from that huge data. Sometimes a firm launch some product or service and want to know the consumer's reaction in very less time to provide their service more consumer friendly or bug-free then these firm have to identify some places from where they can get feedback. Moreover, if a consumer needs to distinguish among some products then again he/she has to go through a big amount of feedback data. When it comes to identifying good lecturer, teacher, leader or an opinion maker then we need to analyse audio data. For a single person, it will be hard to listen to all lectures and hours of the audio which was given in the past. Here needed a single and fast solution to identify someone in nutshell. The proposed method can be helpful in such problems.

1.4 Outline of Thesis

Thesis starts out with Chapter 1 Indroduction that brief about speech and text data

3 sentiment analysis techniques importance in the real word. It is explained how much important sentiment analysis in current scenerio. the application of this work can be in several diamension. Chapter 2 is the review of previous approaches. We have discussed about all literature related of sentiment analysis and, we also discuss about some important state of art methods in detail like text,audio analysis previous work. Chapter 3 is about the basic of Hidden Markov Chain Model and their properties with some examples. Here we will discuss the comparision amoung the all differnt voice detection methods which is available and why we choose perticular one method for voice detection. Chapter 4 is explaining the method of filteration,analysis and data mining techniques. Here we discuss about conversion from raw data to useful data. the algorithm Naive Bayes is explained here with some simple examples. This chapter completely explain about the basic principle of sentiment analysis. Chapter 5 contain all results, performance evaluation of proposed algorithm and its processing time with different data set and comaprison of time requirement in audio and text format. This part has a comparison of proposed method with state of art methods. Finally overall comparison is classified into two part 1) Time taken in different data format, 2) Time taken in differnt amount of data set. Chapter 6 has the content related to my contribution. here we will discuss the how we can enhance the earlier work and make this approach more eassier for all. what are the effect of error on the final result will also be discussed here. Chapter 7 conclude our thesis and also provide possible improvement in this thesis and will give the path for more research and future work.

Chapter 2

Related Work

2.1 Literature Review of Sentiment Analysis Techniques

This work has its kind of unique application as I know so far. But still assembled by pre-existing tools like voice recognition and text conversion technique. NLTK tool and sentiment detection library are being used which is already existing tools and available to use for open source to all. Some paper was presented before like "Automatic Sentiment Detection in Naturalistic Audio" [1], Hidden Markov chain method [13] for voice signal and voice to text conversion or vice-versa. There are some methods available for analysis data like speech data and text data.

2.1.1 Text data analysis method

2.1.1.1 Natural Language Processing

NLTK is a major platform for the creation of Python programs for working with human language data. It provides easy-to-use interface for wrappers for industrial-power NLP libraries as well as text processing libraries for more than 50 corporate and lexical resources such as Wordnet, classification, token, stemizing, tagging, parsing and semantic realization. And it do as an active discussion forum.

2.1.2 Audio Data Recognition Methods

GMM and HMM are the two well known method for pattern recognition in audio

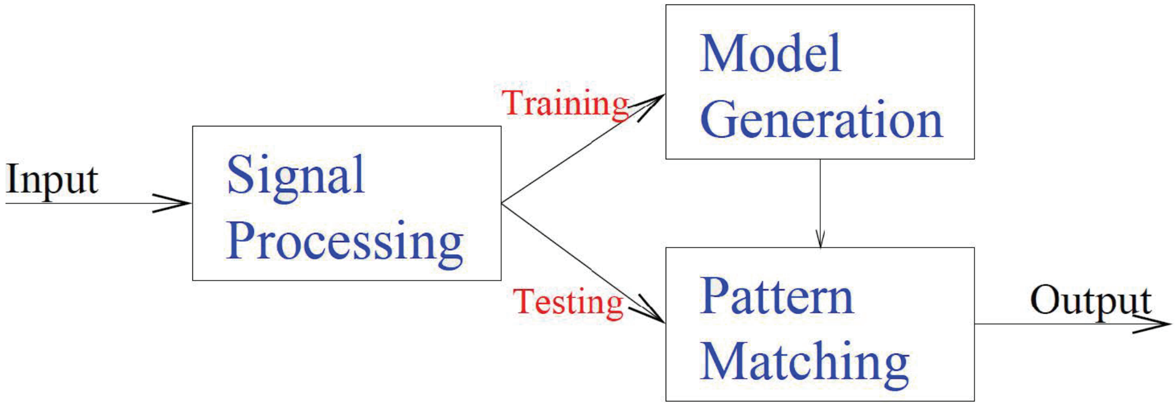

Figure 2.1: Pattern recognition technique block diagram

format.

2.1.2.1 Hidden Markov Model (HMM)

Hidden Markov Model is sequential model implicitly. Because in pattern recognition, next unit recognition depends on the previous pattern which is being recognized.

2.1.2.2 Gaussian Mixture Model(GMM)

It is same, like ANNs are universal approximators of functions, Gaussian Mixture Models are universal approximators of densities with condition that provided sufficient no. of mixtures are used. It is also true for diagonal GMMs as well.

2.1.2.3 Deep Feedforward Neural Network(DNN)

It is uses in many aspect of speech recognition like phoneme classification, audiovisual speech recognition and speaker adaption. Neural network needs comparatively big training data set. Though it is far better and well applied method in recent years but the variation in same language such as accent variation require huge training set.

2.1.3 Comparison Of Speech Recognition Technique

In speech recognition there is two component which work together for speech recognition, first is capturing acoustic representation and second is capturing temporal sequence representation. For capturing acoustic representation a generative model is used where as for capturing temporal sequence which is done by Hidden Markov Model(HMM). Hence it is need to work both model simultaneously in order to recognition like DNN-HMM and GMM-HMM. Now the question is which one combo is better for speech recognition?

Table 2.1: Relationship between three evaluation criteria and three field classifiers.

	KNN	SVM	Bayes
Precision	0.755	0.842	0.861
Recall	0.740	0.783	0.843
Value for all field	0.747	0.811	0.853

here we should know that GMM do not have the capacity to handle large amount of data where Deep Neural Network due to many hidden layers, will have large amount of parameters and capacity. Earlier people used to have ANN-HMM model for speech recognition but they did not know that is it better then GMM or not due to lack of data and computation, but after some research In 2009 GeoffHinton's lab, it was prove that DNN for acoustic modelling can be better then GMM-HMM and can give ~10% better result. The problematic concern with DNN is that it need high computational device like couple of GPUs. Now a days the DNN-HMM model is being opted for most of the service provider companies like(Google,Amazon,Microsoft etc.)

2.1.4 Sentiment Analysis using Machine Learning Previous work

sentiment analysis is basically a task of text categorisation. A big amount of corpus is used for training to get sentiment classification. Actually sentiment classifier is able to judge text sentiment tendency. B pange [5] were the earlier scholar which was involved in initial level sentiment analysis based on machine learning. Their team applied Naive Bayes and SVM in analysis some data, and their result performed better result as compare to other work approach. The sentiment feature is generated from a big sentiment covered dictionary. For short text classifier Wang et al[7] work was given effective result which was based on SVM. Guixian Xu[24] tried feasibility of Bayes theorem by comparing with SVM and KNN(K-Nearest Network)[16] classifications machine learning algorithms. The comparison results are Tabulated in the Table 2.1 and proves that Bayes theorem is more feasible comapre to others. Table 2.1 shows that for perticular data set all three algorithm being applied and the maximum precision[17], recall and values for fields is more comapre to KNN and SVM.

Chapter 3

Proposed Methods For Voice Detection

3.1 A Study Of Voice Detection Technique

A voice conversion system is a system capable of changing the speaker's speech (source speaker) so that his voice can be re-synthesized and confused with the voice of the second speaker (the target speaker). An important application of voice conversion is in speech synthesis, allowing the generation of different voices for text-to-speech systems, without the cost of building another speech database. Other applications that use voice conversion can be interpreted telephony, editing, and dubbing for speech signals, application of film industry, computer game applications or drug applications. Powerful speech analysis and synthesis model should be available to obtain high-quality voice conversion (or at least voice modification). A popular model is the harmonic plus noise model (HNM) [25] and its variations [21]. The parameters employed by these models are quite computational demand. One of these parameters is the maximum volume frequency of speech, which is shown in this paper FMV. Extraction of this parameter is a relatively unexplained subject in the literature, note that Stiarylu's proposed algorithm is widely accepted. In general, SPM provides reasonable estimates for FMV (Finite Volume Method)[22]

3.1.1 Types of speech

Speech recognition system can be divided into different classes by describing the type of utterances they can recognize

3.1.1.1 Isolated Words

Isolated words accepts indivitual words or single utterances entity at a perticular time and needs a break between two utterances hence it is having with it “Listen and Non Listen to state”. Isolated word recognizes the word where is usually each utterance entity to have end on either side of sample words windows.

3.1.1.2 Connected Words

Connected word are somewhat same as to isolated word it needed a fixed pause between two utterances, unlike an isolated word where a break is required.

3.1.1.3 Continuous words

Continuous voice/speech recognizers allows us to speak continuously like he used to be natural, this time recording, conversion and saving data goes on at the same time. This method requires parallel processing and more reliable hardware.

3.1.1.4 Spontaneous words

At a initial level, it can be thought of as voice/speech that is natural in listening/sounding and can not rehearsed. An ASR voice System with spontaneous speech/voice ability should be able to handle different/various words and variety of natural speech features like words being run together.

3.1.2 Speech Recognition Approach

Speech recognitions is an associative subdivision of computational lexical, which develops the mechanisms and technologies and techniques that enable the translation and recognition of a computer language spoken in text. It is also known as Automated Speech Recognition (ASR) , Computer’s Speech Recognition or Speech with Text [S.T.T.]. It involves knowledge and research in computer science, linguistics and electrical engineering field. For some speech recognition systems, some kind of "training" (can be called "nomination") is required, where a person reads text or separate terminology in the speaker system. The system analyses the person’s specific voice and uses it to correct the recognition of that person’s speech, which results in an increase in accuracy. The systems, which do not use their own training are called "speaker-independent" systems. Those systems, using the training are called "speaker dependent". Speech recognition applications has vrious possible work , include voice dealing (such as "set alarm"), call routing (such as

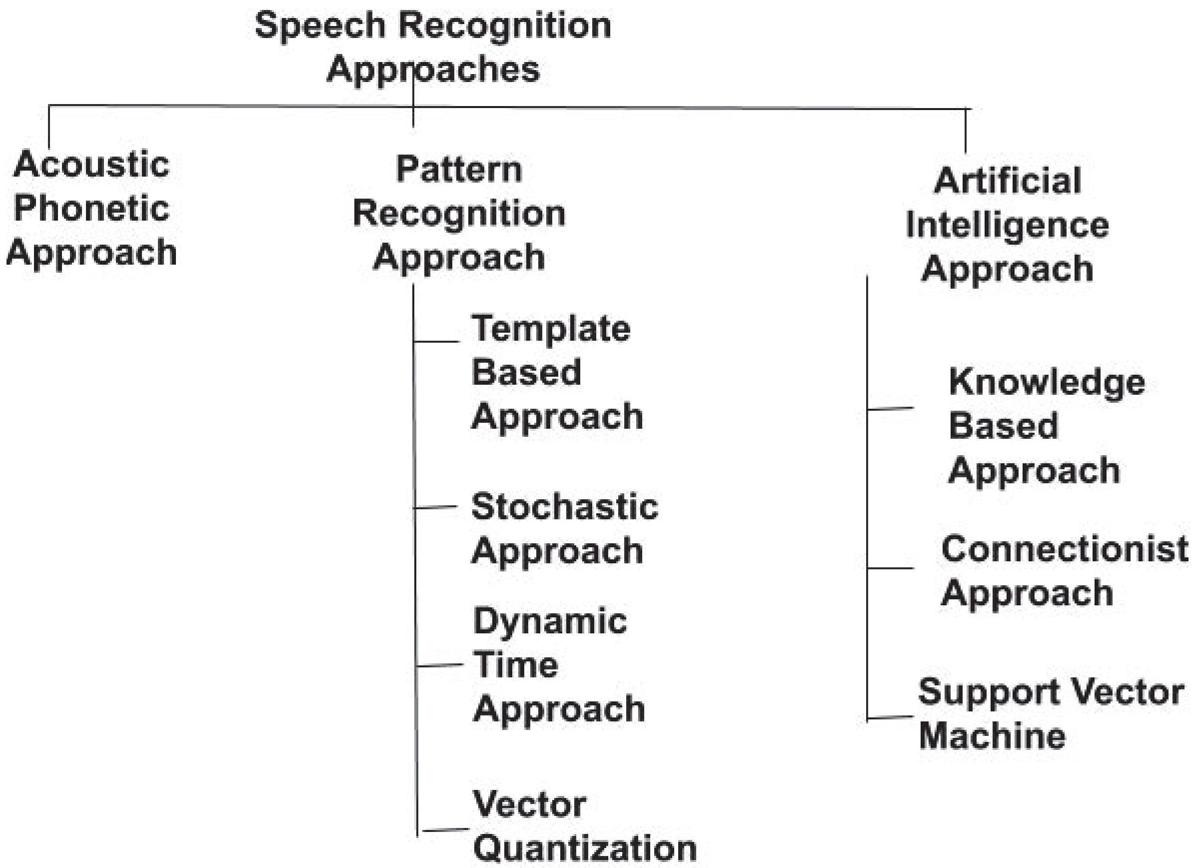

Figure 3.1: various Speech Recognition Approach

"i'd like to make a call"), domestic device control, searching (like search podcasts where special words were spoken) such as voice user interface) , data entry (e.g., storing entry of a credit card number), preparing structured documents (suchas radiology reports), speakers characteristics determining, [2] speech-to-text processing (e.g., word processor or email) and aircraft (commonly called direct voice input). Identification of the voice [3] [4] [5] or identification of speaker [6] [to] instead of identify the speaker refers to what they want to say. figuring out the speaker can solve the task of translating speech in those systems that have been trained on the voice of a particular person or used as a part of a security process to authenticate or verify the speaker's identity. From the point of view of technology, there is a long history of speech recognition with many waves of some big innovation. Recently, this area has been benefiting from deep education and advance research/innovaiton in large data. These advances are not only justify by the increase of academic papers published in the field, but it is important to adopt different types of intensive learning methods in this world to design and deploy speech recognition systems across the globe.

3.2 Artificial Intelligence Based Approach

3.1.1 Hidden Markov models

Modern voice recognition schemes for general use are normally based on the well known madel Hidden Markov model [13]. HMM is statistical model that generate symbol or quantity sequences. Hmm is used for speech recognition because a voice, audio, or speech signal can be viewed as a static signal piece by piece or as a stable signal for a short time. In a small period of time(e.g., 10 milliseconds),It is possible to estimate speech as a stable process. Speech can be considered for many stochastic reasons as a Markov model. Another reason is that HMMs are common as they can be trained automatically and these signals are simple to use and computationally more easy/feasible. In speech recognition, the HMM will produce 10 milliseconds in a sequence of n-dimensional real-value vectors (a small integer with n, like 10). Vectors include cepstral coefficients [23]. Those who adapt the spectrum to a Fourier transformation by converting a brief voice window and using cosine transforms then take the first (most significant) coefficient.There is a statistical distribution in each state in the Hidden Markov model, it is a mixture of Gaussian diagonal covariance, it will give every celebrated vector's possibility. Every word, or (somewhat more general voice identification schemes), each bit of sound will have a different output distribution. A HMM is developed for the sequence of words or sounds by combining distinct Hidden Markov models(HMM) for distinct words and sounds. The Hidden Markov Model (HMM) is an uncertified (i.e. concealed state) statistical Markov model in which the system is being simulated. It's possible to show the Hidden Markov model as the easiest dynamic Bayesian network. Baum and peers created mathematics behind HMM L. E.The state is viewed directly to the supervisor in the simple Markov model (like the Markov series), and thus state transition probability is the single parameter, whereas in the HMM ("Hidden markov model"), the state visiblity is not direct, but the output is lower in the expression(form) of data or "token"), depends on the state, appears. There is more distribution feasible in each state than the future manufacturing token. The sequence of state are provided some information which is generated by Hidden Markov model. HMM also known as pattern theory, grammar is an induction subject.

3.2.2 Hidden Markov model simple example

Consider two friends, Bob and Alice, who live at certain distance from each other and talk every day at the same time on the telephone on each day. Bob's interest are in three

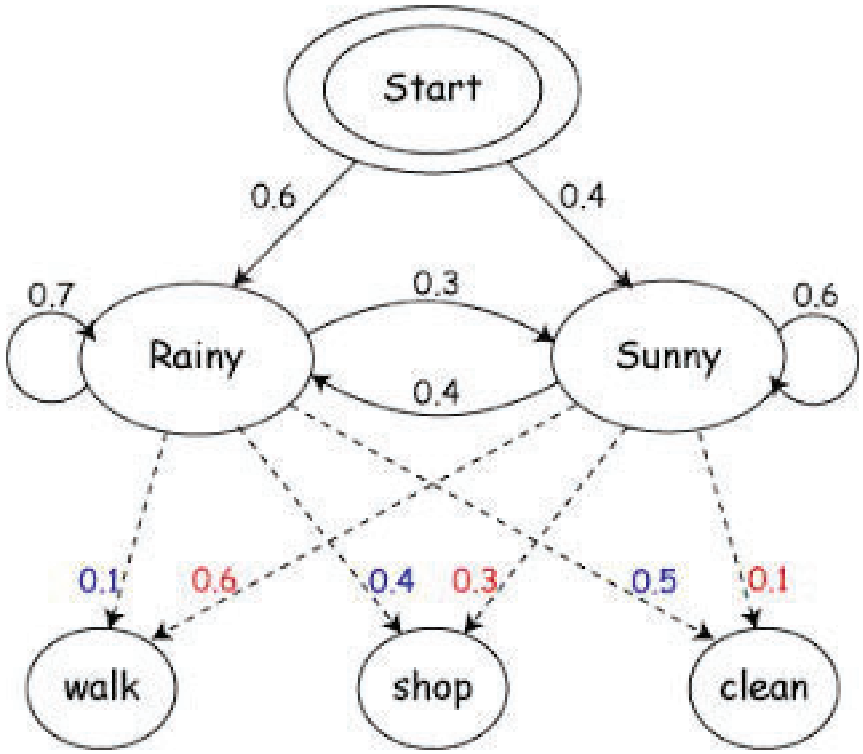

Figure 3.2: decision making by Hidden Markov chain method

operations: walking through the park, buying and cleaning his apartment. What are you going to do? Its option is set specifically on a given day by the weather. Alice doesn't have any definite weather data, but she understands the overall trends.Bob informed him what he is doing every day, on the grounds of which Alice is trying to imagine what the weather is like alice thinks the weather is acting as a discreet Markov chain."Rani" and "Sunny" are two states that are hidden from him, but they can't see them straight. There's a definite possibility every day that Bob will execute one of the following weather-based operations: "Walk," "Shop," or "Clean," as Bob informs Alice about his operations, he's an observation. The entire scheme consists of a concealed model of Markov (HMM).Alice understands the overall trends in the region, and on average Bob likes to do so. That is to say, HMM's parameters are known. They can be shown in Python as follows:

3.2.3 Understanding Using Markov Chain Model

The job of studying HMM's parameters is to figure out the output sequence or set of such similar to output sequences, the best set of prospect for state transit and emission. The job is generally to get the highest assessment of the probability of HMM parameters given the set of manufacturing sequences to fix this issue, no traceable algorithm is known, but using a BOm-Welch algorithm or buladi-chauvin algorithm,It is possible to achieve the highest local potential effectively. A unique case of expectation algorithm is the Baum-Welch algorithm. If HMM is used for time series prediction, More advanced biesian injection techniques such as the Monte Carlo Markov chain (MCMC) then demonstrate to be beneficial in discovering single highest probability models in terms of the precision and the stability. Since MCMC imposes a substantial computational burden in instances where computational scalability is also of interest, biased intentions may promote variable estimates, For example,[18] the predicted variable Conclusion Estimates — Comparative computational effectiveness is given to maximize, whereas precision is only for bases of the MCMC type Valued slightly lower than Information.

Chapter 4

Proposed Methods for Text Analysis

4.1 Maximum Entropy-Based Sentiment Analysis

Maximizing entropy is based on applying different level of filter to estimate the probability of certain linguistic context. To develop the MEBSA(Maximum Entropy based sentiment analysis) model, which extract relevant keywords which reflects polarities like positive ,negative, neutral etc. this method can be work for huge amount data that is further classified into two polarities like positive and negative. In this very initial stage of sentiment detection figure 1 model in which we extract useful word and word combinations. For example Table 4.1 has two tweets which was tweeted after the final match of IPL (Indian Premier League) 2019. These data having some specific keywords which is responsible for sentiment belongs to it. That sentiment presenting keywords is associated to their polarity. Polarity of the keywords can be positive negative and neutral. Here showing Comparison of two random tweets on the evening of IPL (Indian Premier League) each having its own kind of sentiment featured keywords and its polarity

4.2 Collecting Data and Creating Database

We can collect data through real time recording or already recorded audio and store on local or at cloud. Cloud computing is a mixture of a model of service[19] ("platform as a service, software as a service, unified communication as a service and infrastructure as a service"), a model of deployment (public, private, commodity or hybrid) and utility computing[20].Some audio data distribution after generation in different format recorder device will be needed to analyze real-time voice signal. In order to help external device and record the audio, appropriate operating system drivers need to be installed in real

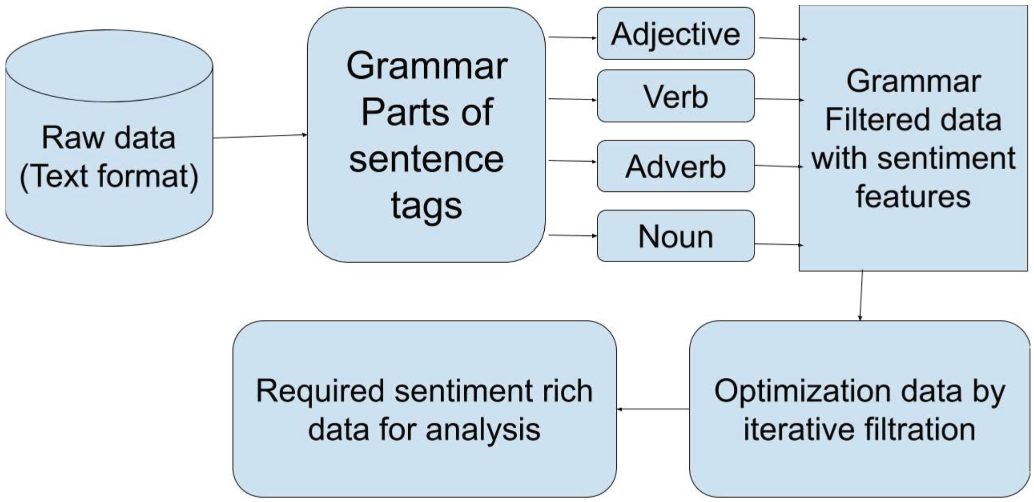

Figure 4.1: Step for generating sentiment rich optimized Maximum Entropy based sentiment model

time recording. Operating system itself provides access to third party software or script. When the script is executed, the recording unit begins recording while making an output file, if the external sound which is coming through recorder device is human language, not a noise sound which is being identified by language processing algorithms

4.3 Speech Recognition

When a speech is recognized its finest match with a specified information phrase (O) given word-sequence (W). O is a series of input vectors produced from the raw voice information. According to Bayes Theorem, this assignment can be formulated as this divides the job into two elements called acoustic model P(O) and p(W), which is called the language model. Hidden Markov Models (HMMs) can represent this model[13] and match the best likely fit stream.

4.4 Tokenization

Tokenization is the process which splitting text into token. NLTK provides many tokenizer in the tokenize module. The words of text file are blank space separated that can consider as one token. During the analysis of the stored text, there are distinct types of updated dictionaries. There is emoticon dictionary, lexical dictionary and stop words dictionary[7] Stemming method that significantly decreases the sparsity of the characteristics involved[11].

Table 4.1: Comparison of two random tweets on the evening of IPL (Indian Premier League) each having its own kind of sentiment featured keywords and its polarity

Tweet	Sentiment presenting keywords	Overall polarity
Thanks India for another wonderful experience. Thanks for inviting us into your homes once again. Your warmth and generosity of spirit continues to provide a rich experience. Until our paths cross again, be well @IPL , Twitter friends, and big-up @mipaltan & @ChennaiIPL	wonderful, Experience, warmth, generosityNegative, spirit, rich, big-up	Positive
#ipl2019final @IPL little bit disappointed by match result. #csk lost match at the edge. Dhoni got out by just bad umpiring decision otherwise this match would be snatched by @mipaltan 's hand	Little, disappointed, lost, match, out, snatched, decision, edge	Negative

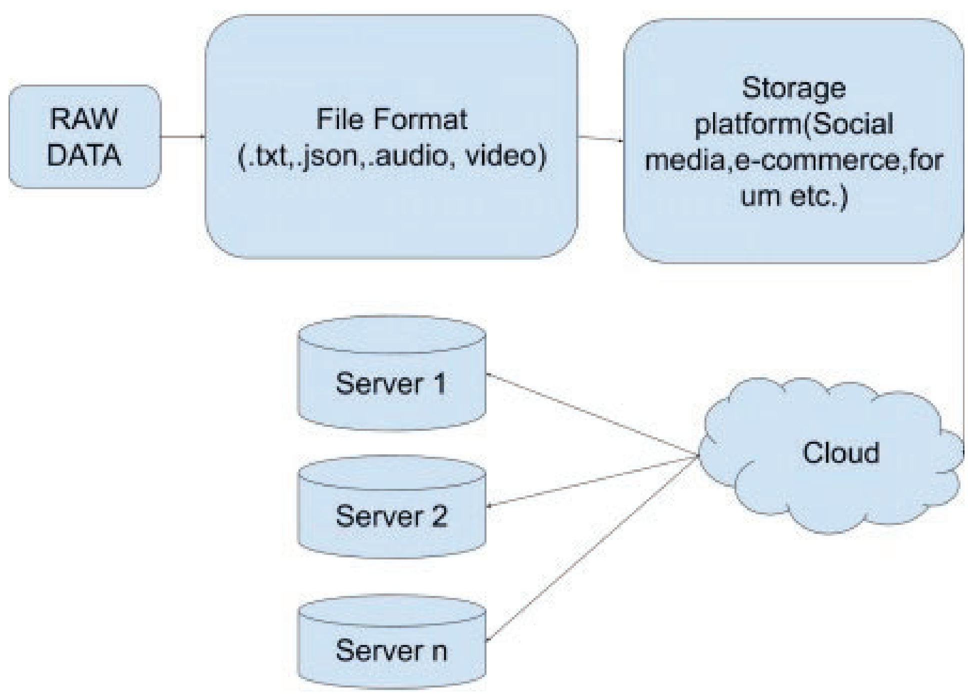

Figure 4.2: flow diagram of cloud-based data source

4.5 Machine Learning

The correct strategy is through machine learning to build a voice signal sentiment analyzer where we can classify views based on their nature. Machine learning has two kinds of classification, namely supervised machine learning and second machine learning without supervision[15].When training data is provided by developers, it is called supervised learning, whereas in unsupervised learning works based on clustering information that reacts based on commonalities and the lack of information commonalities that need to be grouped together. We used Machine Learning Algorithm [12] supervised by Naive Bayes (NB) to conduct voice information sentiment analysis. In a machine learning issue, Naive Bayes needs many variables as characteristics or predictors. This technique operates well with text categorization. As shown in formulae (1) and (2), C * is a class given to d, where ' f ' means ' function, ' the sum of characteristics was represented by ni, m indicates the real amount of characteristics current, and P(f) and P(c) are discovered using maximum probability estimates.

$$C* = argmac_c P_{NB}(c|d)$$

Another form,this classification technique is based on Bayes theorem with an assumption of independent amoung predictors. we can represent mutual independent as:

$$P(x_1, x_2, x_3, ..., x_n|c) = P(x_1|c) * P(x_2|c) * P(x_3|c)...... * P(x_n|c)$$

$$P_{NB}(d) = \frac{P(c)\sum P(f|c)^{n_i(d)}}{P(d)}$$

We used Natural Language Toolkit (NLTK) in this work[14]. NLTK works with information in the human language. NLTK offers significant and informative natural language characteristics such as tokenization, tagging, grammar identification, etc. This enables us to make automatic tokens in tree form and to identify the nature of the text as positive, negative or neutral by analyzing specific tree branches. Naive Bayes classifier's benefit is that the calculation/computation time required to train the dataset is short and accurate, hence its most frequently used algorithm.This algorithm will be so relevant for our job to find out the views of people on any subject and also for some research organizations. Naive Bayes algorithm has also help us to do better process, classify, predict, analyze feelings and recognize patterns in natural language.In turn, the findings acquired through the NB algorithm assist us visualize actual text material that investigates hidden semanthetic ideas. NB algorithm operating time to execute a specific assignment is much quicker than other current literature machine learning algorithms. These are some basic reasons which makes a preferable choice to use NB algorithm.

Table 4.2: Bag-of-words representation and words frequency of occurence

Word	Occurrence
Great	2
Love	2
Recommond	1
Laugh	1
Happy	1
..	..
..	..

4.5.1 Feature Extraction Explaination

Feature extraction is done by the method maximum entropy based sentiment analysis as explained in Section 4.1. Here we have some continuous data set which is having many statement, each statement having so many words. We need to identify the sentiment feature rich words by using ME based, for example there is a paragraph which having some meaningful movie review: "i like this movie. it is nice, but ironic humor. The dialogue delivery are great and the action scene are excellent. it is combo of romantic and adventure together which make it perfect. I recommend this movie to everyone to watch it. after a long time i have seen such movie, i always be happy to watch it again" some feature rich word is being identified by maximum entropy based feature extraction, some of word can be appear in repeatative number then we arrange it in a table. It is known as bag-of-words. A bag of words (Table 4.2) shown here in a document. It contains specific words and also the unique words in a document and their number of repeatation. In Table 4.3 showing that document D(d1,d2,d3,d4...,dn) are n number of document available all document are mutually independent. These document containing corresponding words as bag-of-words, hence a ready data is available for Naive Bayes classifier. equation

$$P(x_1, x_2, x_3, ..., x_n|c) = P(x_1|c) * P(x_2|c) * P(x_3|c)...... * P(x_n|c)$$

represents above Table 4.3 in mathematical form.

4.6 Flow of the proces

It begins with gathering voice information by running voice recognition scripts that convert voice signal to text format in real time, maintaining an internal json / text file format at the same time and continuing to append statements. After that, by reading the internal file and tokenizing the information, the feature vector with appropriate

Table 4.3: Feature to pass Naïve Bayes classifier

Document	w1	w2	w3	w4	...	wn	Sentiment
d1	2	1	3	4	...	1	Positive
d2	2	1	1	3	...	1	Negative
d3	3	8	6	6	...	2	Positive
d4	2	4	1	0	...	1	Positive
d5	1	3	2	1	...	1	Negative
d6	2	3	5	3	...	4	Positive
...	...	...	...	...	...	...	...

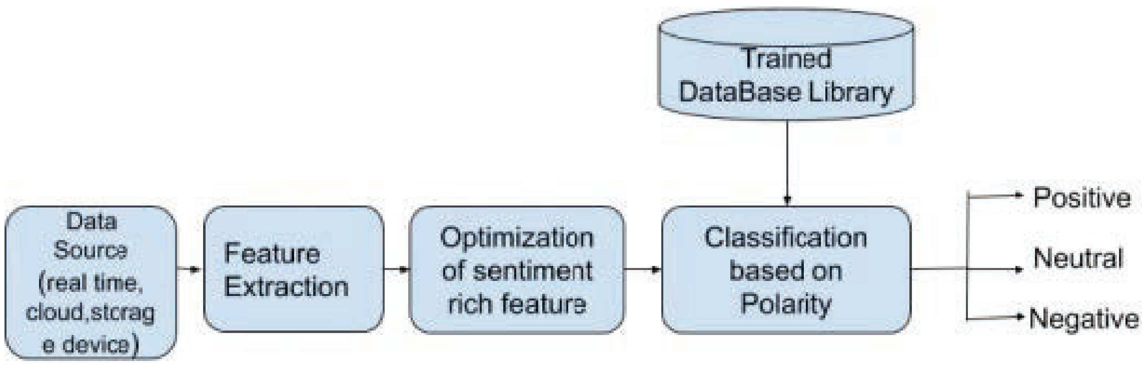

Figure 4.3: data flow blocks sentiment features identification filter stage and polarity decision block

and significant characteristics that will be helpful for assessment using the Naive Bayes algorithm must be constructed.

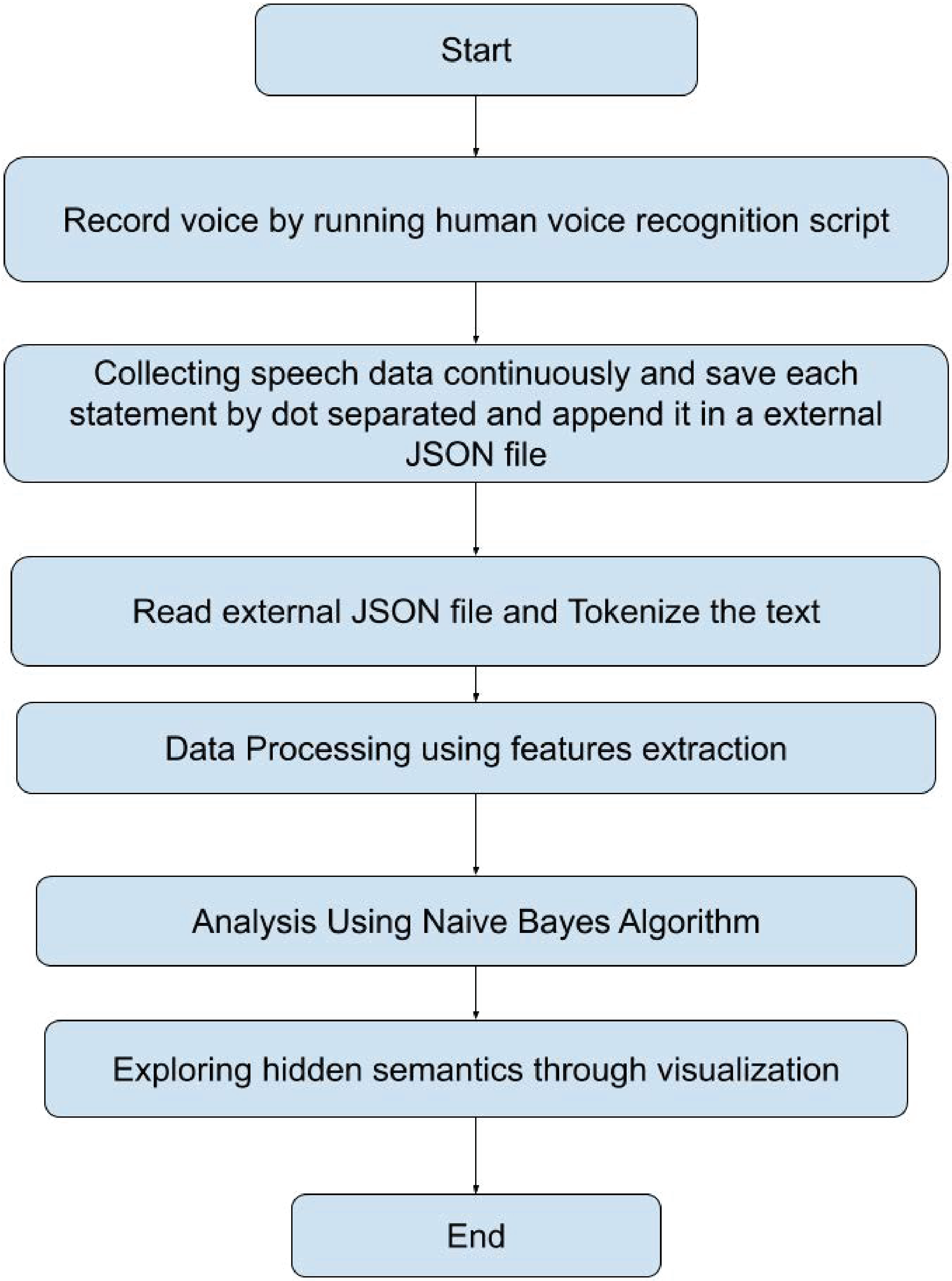

Figure 4.4: showing the flow block diagram of human voice data recognition and processing

Chapter 5

Results and Discussion

In this chapter we will see the result concluded by Sentiment analysis on the input data set. This chapter is going cover the following point

1. Understanding sentiment analysis algorithm from a practicle perspective
2. Formulating the problem statement into result
3. A different data set case study in python
4. How sentiment analysis can impact on bussiness acording to its result

5.1 Sentiment Analysis Work Methodology

In this work we categorized the analysis data in three-part first is speech dataset where we have collected human voice data about 500 statements. The collected data belongs to the same person who was presenting his views on the topic related to the education system. We collected each statement and processed that according to the given flow diagram in figure 2,3,4., the Python script was running real time and did 3 work simultaneously like recording audio, conversion from audio to text and analysis of text on the sentiment analysis point of view. Each statement separated by a dot (.) was taken as a single analysis entity. The whole speech was groups of such statement called a single assertion entity. Each assertion analyzed separately, the advantage of separate analysis is to identify the polarity of statement line by line to categorize result in more accurate and more precise data visualization. The second dataset is text data of about 15746 sentiment feature-rich data and the third data set is 26693 sentiment feature-rich data set. For this purpose, we have taken all speech of Mahatma Gandhi from 1916 to 1948 from an authentic website [www.inc.in]. All three is being analysed separately and compared

Table 5.1: Time and data quantity depending upon data formatting comparison table

S.No.	Data Format	Amount of sentiment featured words	Time taken in processing(millisecond)
1	Audio Format	2000	20678
2	Text Format	13737	1178
3	Text Format	26693	1956

on the basis of time taken in a different amount of data set analysis as shown in the table.

5.2 Python Script and Emulator

For this work i have used python version "Python 3.7.0 (v3.7.0:1bf9cc5093, June 27 2018, 04:06:47)" for computer configuration [MSC v.1914 32 bit (Intel)] on win32. The script editor is IDLE (python 3.7 32-bit) which is available open source. Python script which is being used for sentiment analysis , audio to text conversion and analysis is as follow: Prerequisite for Twitter data mining: In order to mining data from some social media website we need some kind of permission from that website. Twitter allow as to let access its data. The Twitter data is only accessible to those who has made account on Twitter and have permission of access. Twitter gives some special and secret unique authentication key to the user. By using that consumer key and access token, a user can get access to twitter website data and can use it third party platform these keys are highly confidential user should not use it with anyone else because these key reflects your account and in any case of miss happening this user will be responsible for that, hence it is advised to keep it secret as it is possible.

5.3 Analysed Report and Result

The tabulated data shows how analysis time reduces when amount of data set relatively increases in text format and on the other hand the speech data takes more time as compared to text data because of speech to text conversion and text sentiment analysis algorithm both work is being executed. In the visualization of result we got an analysed presentation of that person's views on that topic. Which showing his positive, negative, strong positive, strong negative, weak positive and weak negative weightage of statement in percentile form.

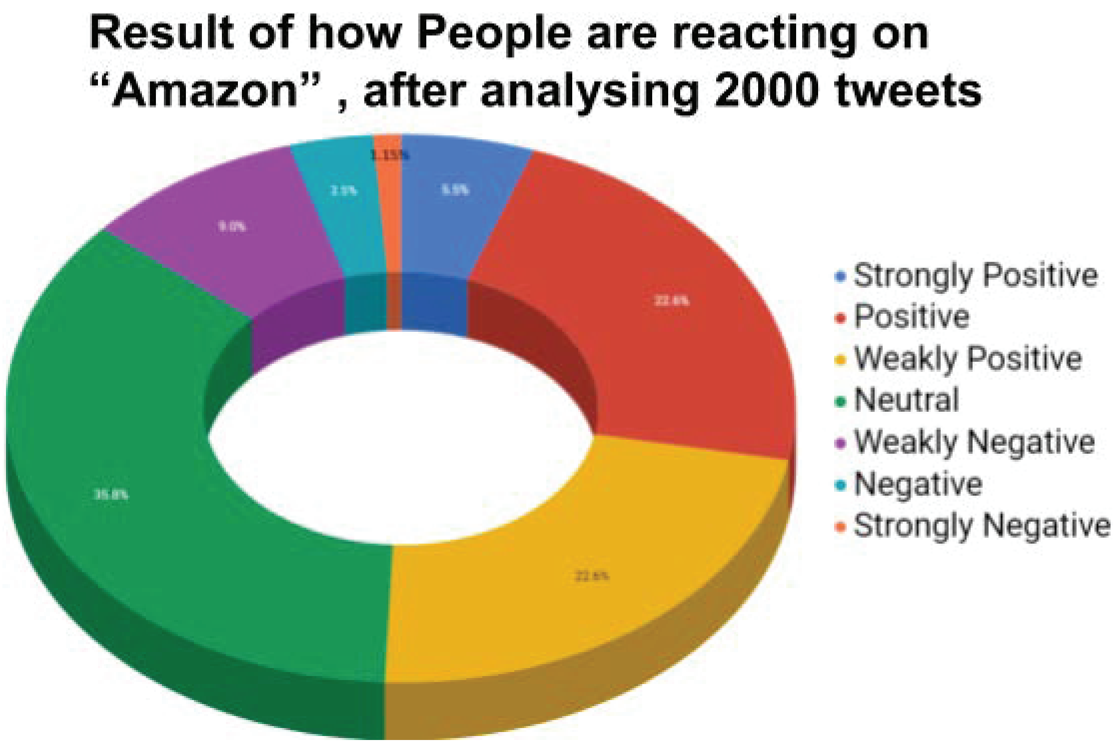

Figure 5.1: Visualization of speech signal sentiment analysis result

Chapter 6

Conclusion & Future Work

6.1 Review of Work

Audio sense analysis using automated speech recognition is an increasing research field where opinions or feelings are natural audio displayed by a speaker is detected. Still, text based sentiment detection is relatively unexplored and very less used in physical word despite its huge possible implementation area because Talking and listening improves information if it is nice then it encourages to listen to you which helps better understanding the things and to form a relationship. Talking have many emotions in it; this is not just a nice thing or good news, but also anger, fear, criticism, or blame hence can be categories in positive, neutral or negative communication. In this fresh architecture, we carry out predictive visual assessment of the collection of big data from different speaking sources and here we carry out graphical and non-graphical assessment and visualize it using machine learning algorithms (Naive Bayes).And we are performing numerical weighted word cloud visualizations that provide enhanced semantine insights. The findings acquired through this job can assist to find a successful teacher, speaker, speaker, and examine the ability to speak for themselves.

6.2 Error Effect on Sentiment Analysis

In speech recognition error should be considered because during recording an audio outer surrounding conditions does not remain ideal most off the cases moreover the language accent is important for particular language detection. Hence during conversion from audio to text form some misinterpretation of words may arrive for example Homophone word [9]. A homophone word is that words which sounds same but have different mean-

Table 6.1: comparison of data set having different amount of error words and its deviation from actual result.

S. No.	Number of Homophone word	Deviation from actual result (6.809159%)
1	10	0.007685%
2	20	0.029027%
3	30	0.032263%
4	40	0.047924%

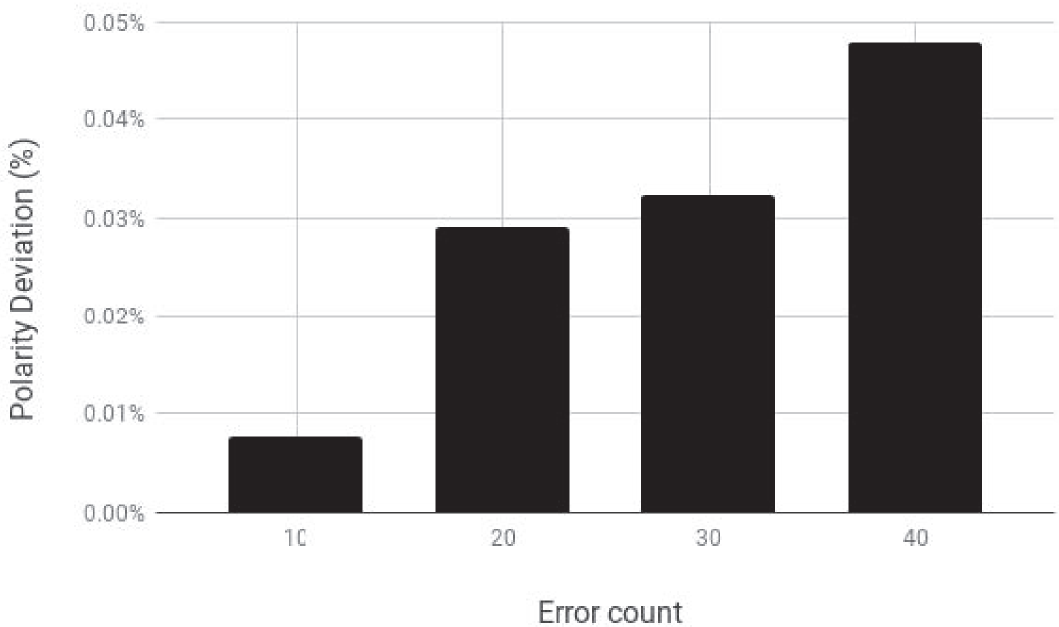

Figure 6.1: graph is showing the % variation in overall result due to increasing amount of error 4

ing. These homophone word can completely change the meaning of a sentence. Table4 is showing some homophone word, these words can be considered as error words in the recorded data. But the technique maximum entropy based sentiment analysis which is explained in theory framework and modelling paragraph, which identifying sentiment feature which is work as first iteration of error reduction. We took some data set with relatively increasing homophone error words and compared with actual result this comparison shown in figure 8 and table 4 as well. It is showing that due to some homophone words the final result effect very minutely and immune to change the final decision.

Table 6.2: sub-classification of output of Naive Bayes classes in post Naïve Bayes Modification

Output Polar Range	Sub Classification
0.00 to 0.30	Weakly Positive
0.30 to 0.60	Positive
0.60 to 1.00	Strongly Positive
0.00	Neutral
0.00 to -0.30	Weakly Negative
-0.30 to -0.60	Negative
-0.60 to -1.00	Strongly Negative

6.3 Post Naïve Bayes Modification

The Naive Bayes classifier gives output as mathematical form. It gives the probability of the input case that is how nearby matching to a particular class. In this work we consider three classes of the output like Positive,Negative and Neutral. Positive class is identified by script if output is positive number and Negative class will be identified by its negative polarity, means a negative number, where as Neutral class will be represented by number zero as output. In probability function output range of a probability form is lies from 0 to 1. It means, if probability lies near to zero then we can conclude that particular input data set is less matching to a class but still having at least few properties of that class. Let consider if output lies near to 1 hence we can conclude that the input data set is strongly lies in a particular class and its maximum properties belonging to that class. If we consider a output number equal to 0.5 then it means that the input data set is half belongs to particular class but still having 50% properties matching to that class. Here it is necessary to further subcategories the output of Naive classifier in some more categories where we can clearly recognise that a input data set is belongs to particular class and how strongly, weakly or normaly belongs to that class. Table 5.2 showing some range and respective subcategories of that range. Hence if we divide range 0 to 1 into three categories then from 0 to 0.3 will be belongs to “weak” subclass, 0.3 to 0.6 belongs to “normal” subclass and 0.6 to 1 belongs to “strong” subclass as showing in Table 5.2.

Chapter 7

Conclusion And Future Work

In this work, the analysis shows us overall sentiment including audio and video is detected by some special words which is having sentiment feature. This study gives a path to research about focussing on particular terms that impact on decision making and identifying generalised public mass review. Overall system is immune to error because of losing few data either text or audio, due to some reason is very less effect on the final result. Mostly audio misinterpretation happened and actual word and printed word became different this problem often occurs in the audio to text conversion but resultant analysis remain immune to this problem. This study provide a new corpus for audio sentiment analysis evaluation area. Here this audio and text based sentiment detection, which is also provide a room for further improvement possibilities like traditional robustness problem of ASR[8](accent, noise, etc.) can have notable impact on performance.

Bibliography

[1.] A. Ratnaparkhi, “A simple introduction to maximum entropy models for natural language processing,” University of ennsylvania, Philadelphia, PA, USA, Tech. Rep. IRCS-97-08, 1997.
[2.] X. Bai, “Predicting consumer sentiments from online text,” Decision Support Syst.,vol. 50, pp. 732–742, 2011.
[3.] M. Ruiz and P. Srinivasan “Hierarchical neural networks for text categorization (poster abstract),” in Proc. 22nd Annu. Int. SIGIR Conf. Res. Develop. Inf. Retrieval, 1999, pp. 281–282.
[4.] N. H. Tou, G.Wei, and L.Kok, “Feature selection, perceptron learning, and a usability case study for text categorization,” in Proc. ACM SIGIRConf, 1997
[5.] R. Socher et al., “Recursive deep models for semantic compositionality over a sentiment treebank,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2013
[6.] S. L. Sonawane and P. V. Kulkarni, "Extracting sentiments from reviews: A lexicon-based approach," 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), Aurangabad, 2017, pp. 38-43. doi: 10.1109/ICISIM.2017.8122144
[7.] L. B. Shyamasundar and P. J. Rani. Twitter sentiment analysis with different feature extractors and dimensionality reduction using supervised learning algorithms. In 2016 IEEE Annual India Conference (INDICON), pages 1–6, Dec 2016
[8.] A. Ghule and P. Benakop, "Performance evaluation of frequency dependent speech signal fading under noisy environment for ASR using correlation technique," 2017 International Conference on Energy, Communication, Data Analytics and Soft omputing (ICECDS), Chennai, 2017
[9.] P. Ghosh, T. S. Chingtham and M. K. Ghose, "SLHAR: A supervised learning approach for homophone ambiguity reduction from speech recognition system," 2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, 2016
[10.] P. Mell and T. Grance. Final version of nist definition of cloud computing. Special publication 800-145, National institute of standards and technology, U.S. Department of 28 commerce, 2011
[11.] Yun Wan. An ensemble sentiment classification system of twitter data for airline services analysis. IEEE, pages 1318–1325, November 2015.
[12.] A. Goel, J. Gautam and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes," 2016 2nd International conference on Next Generation Computing Technologies (NGCT), Dehradun, 2016, pp. 257-261. doi: 1109/NGCT.2016.7877424
[13.] L. Cuiling, "English Speech Recognition Method Based on Hidden Markov Model," 2016 International Conference on Smart Grid and Electrical Automation (ICSGEA),Zhangjiajie, 2016, pp. 94-97. doi: 10.1109/ICSGEA.2016.63
[14.] N. K. Jha, "An Approach towards Text to Emoticon Conversion and Vice-Versa Using NLTK and WordNet," 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA), Changsha, 2018, pp. 161-166. doi: 10.1109/ICDSBA.2018.00036

[15.] H. Yan, P. Yu and D. Long, "Study on Deep Unsupervised Learning Optimization Algorithm Based on Cloud Computing," 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, China, 2019, pp. 679-681.doi: 10.1109/ICITBS.2019.00168

[16.] S. Taneja, C. Gupta, S. Aggarwal and V. Jindal, "MFZ-KNN — A modified fuzzy based K nearest neighbour algorithm," 2015 International Conference on Cognitive Computing and Information Processing (CCIP), Noida, 2015, pp. 1-5.doi: 10.1109/CCIP.2015.7100689

[17.] K. Hulliyah, N. S. A. A. Bakar and A. R. Ismail, "Emotion recognition and brain mapping for sentiment analysis: A review," 2017 Second International Conference on Informatics and Computing (ICIC), Jayapura, 2017, pp. 1-5. doi: 10.1109/IAC.2017.8280568

[18.] L. Jiang, K. Fang, H. Xue and S. Liu, "Emotion categorization based on probabilistic Latent Emotion Analysis model," 2016 IEEE 13th International Conference on Signal Processing (ICSP), Chengdu, 2016, pp. 618-623. doi: 10.1109/ICSP.2016.7877906

[19.] D. Das and S. Bandyopadhyay, "Sentence level emotion tagging," 2009 3rd International Conference on Affective computing and Intelligent Interaction and Workshops, Amsterdam, 2009, pp. 1-6. doi: 10.1109/ACII.2009.5349598

[20.] S. Rana and A. Singh, "Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques," 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), Dehradun, 2016, pp. 106-111. doi: 10.1109/NGCT.2016.7877399

[21.] Y. Long, Z. Yan, F. K. Soong, L. Dai and W. Guo, "Speaker characterization using spectral subband energy ratio based on Harmonic plus Noise Model," 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, 2011, pp. 4520-4523. doi: 10.1109/ICASSP.2011.5947359

[22.] S. Kahleras, A. Boubakeur and L. Boukezzi, "Numerical study using FVM of three cavities within XLPE insulation of HV cables," 2018 International Conference on 29 Electrical Sciences and Technologies in Maghreb (CISTEM), Algiers, 2018, pp. 1-5. doi:10.1109/CISTEM.2018.8613491

[23.] A. Kaur and A. Singh, "Power-Normalized Cepstral Coefficients (PNCC) for Punjabi Automatic Speech Recognition using phone based modelling in HTK," 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Bangalore, 2016, pp. 372-375. doi: 10.1109/ICATCCT.2016.7912026

[24.] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," in IEEE Access, vol. 7, pp. 43749-43762, 2019. doi: 10.1109/ACCESS.2019.2907772

[25.] P. D. Turney M. L. Littman "Measuring praise and criticism: Inference of semantic orientation from association" ACM Trans. Inf. Syst. vol. 21 no. 4 pp. 315-346 2003.