**SENTIMENTAL ANALYSIS USING AUDIO (NLP)**

PAVITHRA M, PRIYAVARDHINI S R, SHRIPRIYA S

Department of Electronics and Communication

KCG College of Technology

Chennai

AIDA JONES

Assistant Professor

KCG College of Technology

Chennai

**ABSTRACT:** Sentimental analysis places a major role in marketing and sales. The main scope of this project is to understand human feelings and determining their emotions and opinions of a customer from voice call using natural language processing. This project is categorized as text and tone analysis. The Naïve Bayes algorithm is used for text analysis while Support Vector Machine (SVM) algorithm is used for tone analysis. Finally those results are combined in mapping to show the result as pie chart showing Positive, Negative and Neutral percentages in the audio. This research investigates a novel approach for detecting human emotions in their speech which is very much helpful in predicting their view towards a product or service. It is useful to improvise the sales and marketing department in any industry.

1. **INTRODUCTION**

Sentimental analysis places a major role in marketing and sales. The main scope of this project is to understand human feelings and determining their emotions and opinions of a customer from voice call using natural language processing. The emotions can be determined by getting the input audio from the customers and analyze their emotions regarding the products and services. For better accuracy the process is categorized into text and tone analysis. Initially noise will be removed from the audio. The audio data will be transcribed into text using speech recognition system. Under the text analysis, text pre-processing will be done followed by Tokenization, Stemming, Stop words removal, numbers removal and Lemmatization. The polarity of the text is also determined. Naive Bayes algorithm is used for the classification purpose and gives output of positive, negative and neutral.On the other side the tone analysis is being done.When amplitude of the audio crosses a particular frequency, the pitch of the tone will be analysed. After that the output of the text and tone analysis is mapped together and then the emotions like positive, negative and neutral can be determined as the output. The emotions will be displayed as an output in the pie chart representing their percentage in the speech.The voiced speech of a typical adult male will have a fundamental frequency from 85 to 180 Hz and that of a typical adult female from 165 to 255 Hz. Normally the range of human voice extends from 80 Hz to 14kHz but traditionally voice band or narrowband telephone calls limits audio frequencies to the range of 300 Hz to 3.4 kHz.

**2. PROSPOSED METHOD**

Sentimental analysis places a major role in marketing and sales.Sentimental analysis enables the computer to comprehend human emotions and make decisions based on those emotions. Sentimental analysis is a type of data mining that measures the inclination of people opinions through natural language processing. Using NLP statistics or machine learning methods to extract, identify or otherwise characterize the sentiment content of a voice. It is nothing but discovering the emotions

and opinions of a customer from voice call.The main scope of this project is to understand human feelings and determining the emotions and opinions of a customer from voice call using the natural language processing [1]. The emotions can be determined by getting the input voice from the customers and analyse their emotions.The input will be the recorded Voice from the telephone conversations [7].

**2.1 MACHINE LEARNING & NATURAL LANGUAGE PROCESSING**

Machine learning is a form of AI that enables a system to learn from data rather than through explicit programming. However, machine learning is not a simple process. As the algorithm follows training data, it is then possible to produce more precise models based on that data.

A machine learning model is the output generated when you train your machine learning algorithm trained with data. After training, when you provide a model with an input, the predicted output will be displayed. For example, a predictive algorithm will create a predictive model. Then, when you provide the predictive model with data, you will receive a prediction based on the data that trained the model.

NLTK has been called “a wonderful tool for teaching, and working in, computational linguistics using Python,” and “an amazing library to play with natural language.”

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over many datasets and lexicon resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum. NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac os x and Linux. NLP guides through the fundamentals of writing Python programs, working with corpora, categorizing text, analysing linguistic structure, and more.

**Fig 1. BLOCKDIAGRAM OF SENTIMENTAL ANALYSIS USING AUDIO (NLP)**

**3. AUDIO INPUT**

A telephone or phone is a telecommunications device that permits two or more users to perform a conversation when they are too far apart to be heard directly. A telephone converts sound, most efficiently the human voice into electronic signals that are transmitted via cables and other communication channels to another telephone which reproduces the sound of the speaker to the receiving user.

The first telephones were directly connected to each other from one customer's office or residence to another customer's location. Being impractical beyond just a few customers, these systems were quickly replaced by manually operated centrally located switchboards. These exchanges were soon connected together, eventually forming an automated, worldwide public switched telephone network. For greater mobility, various radio systems were developed for transmission between mobile stations on ships and automobiles in the mid-20th century. Hand-held mobile phones were introduced for personal service starting in 1973. In later decades their analog cellular system evolved into digital networks with greater capability and lower cost. Convergence has given most modern cell phones capabilities far beyond simple voice conversation. They may be able to record spoken messages, send and receive text messages, take and display photographs or video, play music or games, surf the Internet, do road navigation or immerse the user in virtual reality [3].

The sentimental analysis plays a major role in improving the quality and also to strengthen the product standard. This helps to focus on a particular area rather than concentrating on whole process.

The main goal of this project is to improve the customer satisfaction in a company this can be used in much feedback analysis. The Initial input of the sentimental analysis is audio data. This project deals with recording the telephone

conversations in a company and finding the analysis of whether they are satisfied with our product or they have any query. This gives the output of percentage of Positive, Negative and Neutral. For better accuracy the process of sentimental analysis splits into two,

* Text Analysis and
* Tone Analysis

**4. NOISE REDUCTION**

Noise is unwanted sound judged to be unpleasant, loud or disruptive to hearing From a physics standpoint, noise is indistinguishable from sound, as both are vibrations through a medium, such as air or water. In communication systems, noise is an error or undesired random disturbance of a useful information signal. The noise is a summation of unwanted or disturbing energy from natural and sometimes man-made sources. Noise is, however, typically distinguished from interference, for example in the signal-to-noise ratio (SNR), signal-to-interference ratio (SIR) and signal-to-noise plus interference ratio (SNIR) measures.

In audio engineering, noise can refer to the unwanted residual electronic noise signal that gives rise to acoustic noise heard as a hiss. Sound is measured based on the amplitude and frequency of a sound wave. Amplitude measures the force of the wave. The energy in a sound wave is measured in decibels (dB), the measure of loudness, or intensity of a sound; this measurement describes the amplitude of a sound wave. Decibels (dB) are expressed in a logarithmic scale. On the other hand, pitch describes the frequency of a sound and is measured in hertz (Hz). All signal processing devices, both analog and digital, have traits that make them susceptible to noise. Noise can be random or white noise with an even frequency distribution, or frequency dependent noise introduced by a device's mechanism or signal processing algorithms.

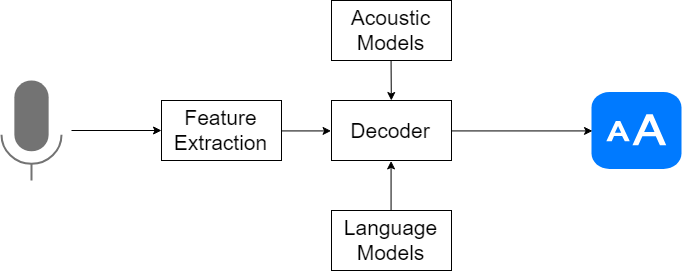
The noise level in an electronic system is typically measured as an electrical power in watts or dBm, a root mean square (RMS) voltage (identical to the noise standard deviation) in volts or a mean squared error (MSE) in volts squared. Noise may also be characterized by its probability distribution and noise spectral density in watts per hertz. Noise is a random process, characterized by stochastic properties such as its variance, distribution, and spectral density.

Hence for the proper analysis the noise reduction will be done with the appropriate filters like low pass filters .after this noise reduction ,the background unwanted noise will be remove and the audio will be ready for further analysis and processing.

**5. METHOD OF TEXT ANALYSIS**

**5.1 SPEECH - TEXT TRANSCRIBING**

Speech recognition is the inter-disciplinary sub-field of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers.

****

**Fig 2.PROCESS OF SPEECH TO TEXT CONVERSION**

It is also known as automatic speech recognition (ASR), computer speech recognition or speech to text (STT). It incorporates knowledge and research in the linguistics, computer science and electrical engineering fields.

Speech Recognition is a process in which a computer or device record the speech of humans and convert it into text format [4]. It is also known asAutomatic Speech Recognition (ASR), computer speech recognition or Speech To Text (STT). Linguistics, computer science, and electrical engineering are some fields that are associated with Speech Recognition.Implementing Speech Recognition in Python is very easy and simple. Two libraries are used

1. Speech recognition
2. Pyaudio

These Speech Recognition and Pyaudio are the library in the python to perform the operations like speech recognition, speech to text and also it helps to recognize the human voice and performs speech to text etc. The Pyaudio is the one of the library used to access the audio file. Speech Recognition is a library that helps in performing speech recognition in python. PyAudio provides Python bindings for Port Audio, the cross-platform audio I/O library.

Speech recognition works using algorithms through acoustic and language modelling [6]. Acoustic modelling represents the relationship between linguistic units of speech and audio signals; language modelling matches sounds with word sequences to help distinguish between words that sound similar. Speech recognition performance is measured by accuracy and speed.**V / H z {\displaystyle \scriptstyle \mathrm {V} /{\sqrt {\mathrm {Hz} }}}**

**5.2 PREPROCESSING**

Data preprocessing is an important step in the data mining process. The phrase "garbage in garbage out” is particularly applicable to data mining and machine learning projects. If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult.The product of data preprocessing is the final training set.

This data preprocessing has several steps included for the proper data analysis they are

1. Tokenization

2. Lowercasing

3. Normalization

4. Stop words removal

5. Special characters removal

These are the steps in the preprocessing techniques. In machine learning the preprocessing is the initial process to clean and remove the irrelevant data. After this preprocessing the data will properly cleaned for the further text analysis.

**5.2.1 TOKENIZATION**

Tokenization is a step which splits longer strings of text into smaller pieces, or tokens. Larger chunks of text can be tokenized into sentences, sentences can be tokenized into words, etc.Tokenization describes splitting paragraphs into sentences, or sentences into individual words. For the former (SBD) can be applied to create a list of individual sentences.

Sometimes segmentationis used to refer to the breakdown of a large chunk of text into pieces larger than words (e.g. paragraphs or sentences), while tokenizationis reserved for the breakdown process which results exclusively in words .The text data is spitted into several pieces for the proper analysis.

**5.2.2 LOWER CASING**

The lowercasing is the one of the step in the pre-processing technique. The lowercasing is the process of changing the upper casing alphabets into lowercasing. This process is done for the proper alignment of text processing.

**5.2.3 NORMALISATION**

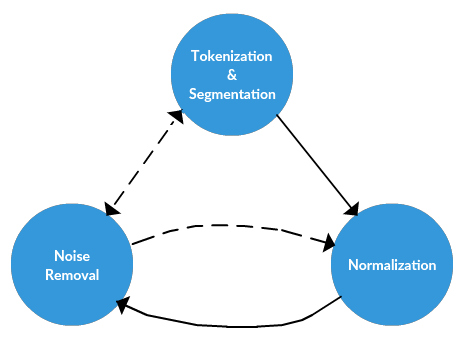
Normalization is an example of preprocessing data to remove or reduce the burden from machine learning (ML) to learn certain invariants, that is, things which make no difference in the meaning of the symbol, but only change the representation.

**5.2.4 STOP WORDS REMOVAL**

Text may contain stop words like ‘the’, ‘is’, ‘are’. Stop words can be filtered from the text to be processed. There is no universal list of stop words in NLP research; however the NLTK module contains a list of stop words. In natural language processing, useless words (data), are referred to as stop words. A stop word is a commonly used word (such as “the”, “a”, “an”, “in”) that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query.

**5.2.5 SPECIAL CHARACTERS REMOVAL**

This is the final step of preprocessing. This is nothing but the process of removing the special characters like {@#$%^&\*} from the text file. These special characters don’t have a part at analysis. This can be done with the python library.



**Fig 3. RELATIONS OF PRE-PROCESSING STEPS**

**5.3 NAIVE BAYES CALSSIFIER**

Naive Bayes is a family of algorithms based on applying Bayes theorem with a strong(naive) assumption, that every feature is independent of the others, in order to predict the category of a given sample [2]. They are probabilistic classifiers, therefore will calculate the probability of each category using Bayes theorem, and the category with the highest probability will be output. Naive Bayes classifiers have been successfully applied to many domains, particularly Natural LanguageProcessing (NLP).  Naive Bayes classifier is a straightforward and powerful algorithm for the classificationtask.

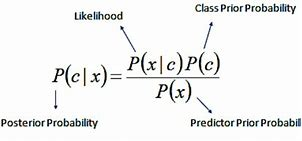
Naive Bayes is a kind of classifier which uses the Bayes Theorem. It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class [5].  The class with the highest probability is considered as the most likely class. This is also known as Maximum A Posteriori (MAP).

* The MAP for a hypothesis is:
* MAP(H)  
  = max( P(H|E) )  
  =  max( (P(E|H)\*P(H))/P(E))  
  = max(P(E|H)\*P(H))

**5.4 VADER SENTIMENT ANALYZER**

VADER (Valence Aware Dictionary for sentiment Reasoning) is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion. Introduced in 2014, VADER text sentiment analysis uses a human-centric approach, combining qualitative analysis and empirical validation by using human raters and the wisdom of the crowd.

VADER sentiment analysis combines a dictionary of lexical features to sentiment scores with a set of five heuristics. The model works best when applied to social media text, but it has also proven itself to be a great tool when analyzing the sentiment of movie reviews and opinion articles. The great thing about VADER sentiment analysis is that an open-source implementation in Python is available here. Sentiment analysis becomes a joy using the code.



**Fig 4. FORMULA OF LEXICON VADER SENTIMENT**

**6. METHOD OF TONE ANALYSIS**

Identifying a human based on the variations and unique characteristics in the voice is referred to speaker recognition. It has acquired a lot of attention from the research community for almost eight decades. Speech as signal contains several features which can extract linguistic, emotional, speaker specific information, speaker recognition harnesses the speaker specific features from the speech signal. Mel Frequency Cepstrum Coefficient (MFCC) is used for designing a speaker discriminant system. The MFCC’s for speech samples from various speakers are extracted and compared with each other to find the similarities between the speech samples.

Feature Extraction: The extraction of unique speaker discriminant feature is important to achieve a better accuracy rate. The accuracy of this phase is important to the next phase, because it acts as the input for the next phase.

MFCC— Humans perceive audio in a nonlinear scale, MFCC tries to replicate the human ear as a mathematical model. The actual acoustic frequencies are mapped to Mel frequencies which typically range between 300Hz to 5KHz. The Mel scale is linear below 1KHz and logarithmic above 1KHz. MFCC Constants signifies the energy associated with each Mel bin, which is unique to every speaker. This uniqueness enables us to identify speakers based on their voice. The steps in tone analysis are,

**Subjectivity dataset:** It is a dataset containing the list of positive and negative words. Subjectivity expresses some personal feelings, opinions, views, or beliefs. It is also used to reduce the distractions in the text that is to be analysed. Some in formations are considered to be true or trustworthy on the basis of subjectivity. Subjectivity is contrasted to the philosophy of objectivity, which is a reality or truth on an individual’s feelings or imaginings.

**Polarity classifiers:** The polarity analysis deals with the identification of amount of positive and negative terms that appearsin a sentence. The polarity can be broadly classified into Positive polarity item (PPI) and Negative polarity item (NPI). The Support Vector Machine (SVM) algorithm is used for this tone analysis.

**6.1 SVM CLASSIFIER**

In machine learning, Support vector machine (SVM) is a discriminative classifier designed by separating hyper plane used for classification, regression etc. this will build a model as points in space, so that the separated categories are divided by a clear gap that is as wide as possible.

The Support Vector Clustering is one of the most widely used algorithms in industrial applications. Clustering is a process of grouping a set of objects in a same group that are more similar to each other than others in separate groups.

Y

X

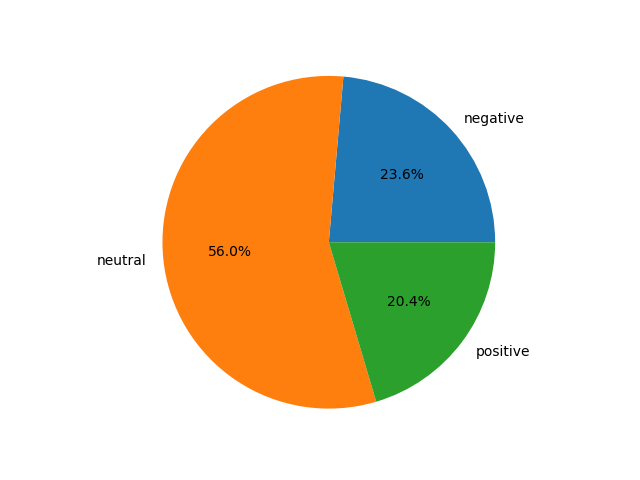
**Fig 5. SVM CLASSIFICATION MODEL**

**7. MAPPING**

Mapping is the process of combining or comparing the results obtained in text and tone analysis. Text analysis provide result as percentage of positive, negative and neutral while tone analysis gives result as percentage of happy, bored, sad, excited etc. Thus, these two results are mapped based on their values obtained on processes and the output is displayed in pie chart. The result pie chart shows the percentage of positive, negative and neutrality in the input audio.

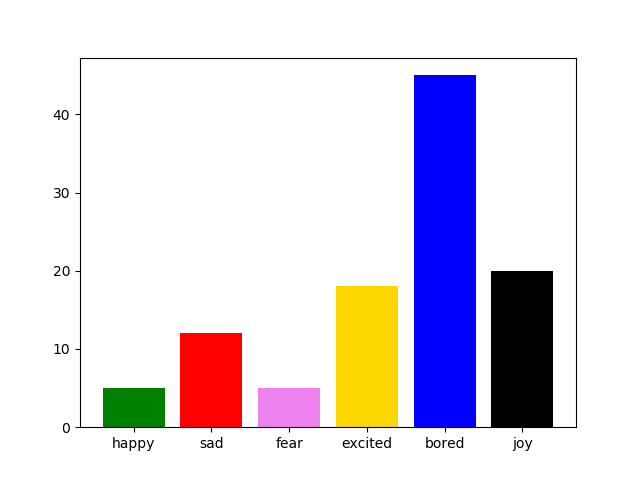
**8. RESULTS**

**8.1 TEXT ANALYSIS**

****

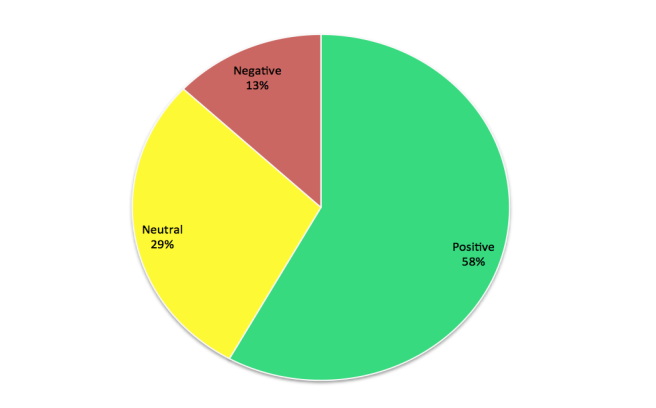
**Fig 6. TEXT ANALYSIS OUTPUT**

**8.2 TONE ANALYSIS**

****

**Fig 7. TONE ANALYSIS OUTPUT**

**8.3 MAPPING**

****

**Fig 8. MAPPING OUTPUT**

**9. CONCLUSION AND FUTURE WORK**

This research investigates a novel approach for detecting human emotions in their speech which is very much helpful in predicting their view towards a product or service. It is useful to improvise the sales and marketing department in any industry. This approach concentrates on customer satisfaction and in linear improvement of the company.

The benefit of detecting the sentiment of the customer or user is to identify the pros and cons of the products or services. Their valuable feedback is used to improve the standard and quality of the product manufactured or service provided.

The same project can be extended by analysing BigData. Using translator, different languages can be accessed and analysed to find the sentiment in it. Further research can be done to identify the gender of the speaker.

**REFERENCES**

[1]LakshmishKaushik, AbhijeetSangwan, John H. L. Hansen “Sentiment extraction from natural audio streams”, International Conference on Science and Technology 2014

[2]Sagar K. Bhakre, Prof.Arti Bang, “Emotion Recognition on The Basis of Audio Signal Using Naive Bayes Classifier”, International Conference on Advances in Computing, Communications and Informatics (ICACCI)2016

[3]Maghilnan,Rajesh Kumar,“Sentiment Analysis on Speaker Specific data”, 2017 International Conference on Intelligent Computing and Control (I2C2)

[4]Min Zhao, TaoZhenZhang, JianpingChai,“Based on SO-PMI Algorithm to Discriminate Sentimental Words’ Polarity in TV Programs’ Subjective Evaluation”,International Symposium on Computational Intelligence and Design 2015

[5]Christos Troussas, MariaVirvou, Espinosa, Kevin Llaguno, JaimeCaro“Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning”, International Conference on Advances in Communications and Informatics (ICACCI)2017

[6]Rohit Raj Sehgal, Gaurav, Raj ShubhamAgarwal“Interactive Voice Response using Sentiment Analysis in Automatic Speech Recognition Systems” ,International Conference on Advances in Computing and Communication Engineering (ICACCE-2018)

[7]SourayaEzzat, Neamat El Gayar, and Moustafa M. Ghanem “Sentiment Analysis of Call Centre Audio Conversations using Text Classification”

Figures [2], [3], [4] – Google (Wikipedia)