Sentiment analysis of market demography For market research

by

Shadman Islam
16101304
Moshiur Rahman
19341028
Samina Tuz Zohura Ali
19141018
Tawhid Shahrior
19341026

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
December 2019

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Shadman Islam
16101304

Moshiur Rahman
19341028

Samina Tuz Zohura Ali
19141018

Tawhid Shahrior
19341026

Approval

The thesis/project titled "SENTIMENT ANALYSIS OF MARKET DEMOGRAPHY FOR MARKET RESEARCH" submitted by

1. Shadman Islam (16101304)
2. Moshiur Rahman (19341028)
3. Samina Tuz Zohura Ali (19141018)
4. Tawhid Shahrior (19341026)

Of Fall, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on December 10, 2019.

Examining Committee:

Supervisor:
(Member)

Name of Supervisor
Senior Lecturer
Department
Institution

Program Coordinator:
(Member)

Name of Program Coordinator
Designation
Department
Brac University

Head of Department:
(Chair)

Name of Head of Department
Designation
Department of Computer Science and Engineering
Brac University

Ethics Statement

We, hereby declare that this thesis is based on the results we obtained from our work. Due acknowledgement has been made in the text to all other material used. This thesis, neither in whole nor in part, has been previously submitted by anyone to any other university or institute for the award of any degree.

Abstract

Sentiment analysis,the science of understanding human emotions, has been touted as an important tool for business success.With the plan of helping various companies get key insights for specic market regions and their customers, and to gain a competitive advantage over others, we propose a model. In here, we estimated product perception of a demography based on emotions that were extracted from customers' facial cue and speech.Although many researches have been made in this eld, very few of them are multi-modal, integrated systems, where the different components rely on each other to produce an absolute result.We extracted the emotions of people by recording their facial cues and speech patterns as they interacted with a specic product of the market (e.g.: Mobile Phone). We took a video survey of people in real time as they judged a sample product(smartphones) based on their merits and features. Then we analyzed their facial expressions, extracted images from various time frames and analyzed those images using AWS Rekognition.A predictive model from those information were generated using XGBoost. We also merged that result with the sentimentanalysisthatwedidontheirspeechpattern.Weturned their speeches into texts and we analyzed those texts using an algorithm which has a mixture of Tensor ow, Keras, Sequential model and RNN. Having results from both sides, using Kfold cross validation with 3 folds and 5 iterations we have achieved an average accuracy of 81 percent approximately with .065 standard deviation.

Keywords: —Product market, Tensor ow, Keras, Sequential model and RNN, GRU , XGBoost, AWS Rekognition.

Dedication

A dedication is the expression of friendly connection or thanks by the author towards another person. It can occupy one or multiple lines depending on its importance. You can remove this page if you want.

Acknowledgement

Firstly and foremost, we would like to thank our Almighty for enabling us to conduct our research, give our best efforts and complete it. Secondly, we would like to thank our supervisors Golam Rabiul Sir And Mostafiz Akhand sir for Their feedback, support, guidance and contribution in conducting the research and preparation of the report. They encouraged us to conduct the research, give guidance to us and always were present to offer any help we could ask for. We are grateful to them for their excellent supervision to successfully conduct our research. We also like to extend our gratitude to our family and friends, who guided us with kindness and gave inspiration and with their suggestions. Last but not the least, we thank BRAC University for providing us the opportunity of conducting this research and for giving us the chance to complete our Bachelor degree.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thoughts behind working in the marketing landscape

In this ever changing world, the need of detecting underlying emotions is a necessity to detect motive behind a human decision which can range from purchasing a specific product, crime motivation, employee to employer relationship etc. Sentiment analysis is such a gateway where we can use the latest technological advancements to extract human emotions. Sentiment analysis-the automated process of understanding an opinion about a given subject from written or spoken language, or video capturing- is a hot topic in the current technological and socio-economical landscape.

Sentiment analysis (also known as opinion mining or emotion AI) refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

In a world where we generate 2.5 quintillion bytes of data every day, sentiment analysis has become a key tool for making sense of that data. With the help of sentiment analysis systems, this unstructured information could be automatically transformed into structured data of public opinions about products, services, brands, politics, or any topic that people can express opinions about. This data can be very useful for commercial applications like marketing analysis, product reviews, product feedback. Since The applications of sentiment analysis in business cannot be overlooked. Sentiment analysis in business can prove to be a major breakthrough for many companies as they look for the complete brand revitalization or even globalization, both of which we are focusing on. . Many companies have tried using social media review, survey based textual reviews or even video capturing to asses the sentiment of their target demography. However, in order to truly capture the actual perspective of a customer's view on a product, we evaluate how their facial structure along with their speech corresponds to a particular product. A robust mixture of sentiment extraction that have been rarely used to do a comprehensive market research .We are planning on bringing a radical new view where we will be extracting sentiment analysis from both facial expressions and speech from a customer as they look at specific products of the market in a shop, merge their results

to get an in depth analysis of their emotions towards any product of the market. These results will be sent towards the owners of the products, from which they will decide if their product has resonated with their target demography or are there any areas of improvement that they need to look at for the next iterations of their products. Where the Traditional metrics focus on quantity, such as number of views, clicks, comments, shares, etc. Sentiment analysis goes beyond plain demographics to the quality of the interactions between the public and the company brand.

1.2 Problem Statement

In a world, where to get the inside knowledge of the market's view of a particular product, the target demography's reaction and concern regarding the said product, it is imperative that a technology exist .A technology which would allow a company to not only understand their current position but also based on that information, position themselves to tackle the market from never before seen angle in the future that would help them gain a competitive advantage over their rival companies. However although many research has been conducted on this field of customer perception and sentiment towards a certain product, however there has not been a technology that would help the companies get a real time feedback about their products position within the market. There are very few technology that provides a multimodal system , which can be 'human like' in their approach of understanding the customer base's emotions. Moreover, without having a proper technique to extract this vital yet underlying emotions , the market is losing valuable insight which subconsciously stems from the customer base. An information through which they can penetrate the market in a more robust ways. Understanding a customer's sarcasm, ambiguous statements, doubts from facial cues, the slight giveaway from the tremble of his or her voice through a multimodal system can be revolutionary when it comes to market research.

1.3 Research Contribution

As customer opinion is a deciding factor in such cases , number of researchers have been doing various research on analyzing the sentiment of public reviews. However, this analysis has either been centralized to specific products such as phones or movies. It also has been only focused mostly on written reviews, or individual aspects of sentiments such as only face or only speech. Very few research has been done which combines various factors of emotions to give a proper result. In our research, we provide a multifaceted system, that provides an overall rating of a consumer base's view of a company's product which would be used for market research purposes.

1.4 Research objectives

Our research objective covered various ground which ranged from market research, to customer perception and lastly to feel the void of a tool which could let the marketers peek into their customers mind, a field that is yet to explode in the

current business sentiment analysis sector. Our objectives include but not limited to-

- Creating a robust sentiment analysis extraction system that not only takes the traditional speech to text method of sentiments but also extracting sentiments from facial cues as well.
- Through the combination of these two we will be hoping to have an in depth insight to a customer's mind
- We hope to answer some of the problem points of sentiment analysis such as lack of in depth human mind analysis rather than a classification of just happy and sad, A new of extracting emotions from customers besides the previously establish and aging social media metric of gathering information, uncovering sarcastic remarks by balancing between the results from speech and results from facial cues.
- A new method of market research that simultaneously gives information of a product's current performance, its selling points, its issues and ways of renovating the product in future iterations.
- Information that the companies can use in a multitude of ways, from product globalization, to rebranding and many uses in between.
- Understanding the recent market trends directly through the customer's point of view.

Chapter 2

Background Analysis

2.1 Related Works

There were many papers which tackled the issue of sentiment analysis through various methods. As we moved through various papers, we had found some papers that talk about our related field. First we needed to understand what the current

marketing landscape is looking for in terms of cutting edge technologies. In the paper,[1] , it talked about how due to the explosion of social media, there has been an influx of data from citizens. Policy-makers and citizen don't have an effective way to make sense of this massive surge of data. It detailed some of the long-term issues which includes sarcasm detection, Usable opinion mining tool for citizens. It recognized human analysis as one of the top market tools.

In order to get a better understanding, we specifically analyzed the situation in mobile market as it would be used as the main tool of our experiments. To better understand what are its' challenges, it was imperative in terms of the marketers point of view. The article [2] , gave us various insights. One of the first things it recognized as a precursor for a successful Mobile Marketing Strategies was having a sound understanding of a company's target customers. It strongly suggested that the retailers who better understand their demography and the customers' behavior can be more successful in their mobile marketing strategies than others. Continuous learning about the sentiment of the target demography is an imperative for a company to create value.

In the paper [3], the author proposed a model that talk about Automated identification of facial expression. Here the analysis has been performed based on Face

recognition using Viola-Jones algorithm. And for Facial expression recognition, Local Binary Pattern has been utilized. Support Vector Machines was used for Classification of the expressions and also for performing the Analysis. Using three stages which were, Face Detection, Feature Tracking and Face Recognition it extracted emotions from the images. To detect the outward appearance Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and the direct programming method were used. However, they did not use RNN here so there is no memory information which can detect pattern for sequential image. Thus it is unable to detect emotions from a video as it does not have sequential pattern recognizing characteristics. So it can not take images which come as sequences and will not be able to give a total estimation of sentiment analysis in a specific time span of sequential images each connected to each other.

From a text based sentiment analysis perspective,in another paper [4], they had proposed a model which had made significant progress towards understanding compositionality in tasks such as sentiment detection with richer supervised training and evaluation resources and more powerful models of composition. They introduced a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, the Recursive Neural Tensor Network was introduced. When trained on the new treebank, this model outperforms all previous methods on several metrics. It pushes the state of the art in single sentence positive/negative classification from 80 percent up to 85.4 percent. The accuracy of predicting fine-grained sentiment labels for all phrases reaches 80.7 percent, an improvement of 9.7 percent over bag of features baselines. Lastly, it is the only model that can accurately capture the effects of negation and its scope at various tree levels for both positive and negative phrases. Recursive Neural Tensor Network that can accurately predict the compositional semantic effects present in this new corpus. Recursive Neural Tensor Networks take as input phrases of any length. They represent a phrase through word vectors and a parse tree and then compute vectors for higher nodes in the tree using the same tensor-based composition function. RNTNs also learn that sentiment of phrases following the contrastive conjunction.
Thus it captured the compositional effects with higher accuracy.[4]
The corpus is based on the dataset introduced by Pang and Lee (2005) and consists of 11,855 single sentences extracted from movie reviews.
Problems include that it is only applicable in textual format. This memory centric design is not being applied in an image or video centric space where we can detect even finer grained emotions through facial expressions. This is limited in only extracting sentiments from texts. So, it is not robust and flexible.

In the paper [5], the author had proposed a model of sentiment analysis of different features of different company's mobile sets and rating them overall. They pre-

processed the gathered data to a supervised form and chose the most common features from train data. In the model, Naïve Bayes, Support Vector Machine, Logistic Regression, Stochastic Gradient Descent and Random Forest algorithms were used to compare performance. This model provides an average polarity of each features and an average polarity of the mobile phone which will give a rating of the device, thus assisting the customers to choose the best according to their desire. Moreover it provides an average polarity of each features and an average polarity of the mobile phone which will give a rating of the device, thus assisting the customers to choose the best according to their desire.

However, The aforementioned paper only deals with textual reviews gained from customers. It does not go in depth about the features of the phone sets. Thus, we do not get a comprehensive insight about the sentiment analysis. Moreover, this method does not use RNN or any other neural network that has memory storing capabilities. Thus it can't detect sequences between product reviews or predict about how a product can be due to the lack of pattern recognizing capabilities. So, it is very limited in its uses by only being textual review based. It also would not be able to detect sarcasm or hidden meaning behind opinions which is limited to actual human speech or facial characteristics.

The paper [6], dealt with removing "noisey web texts" which can cause significant problems both at the lexical and the syntactic levels. In this research , a random sample of 3516 tweets were used to evaluate the consumers' sentiment towards various mobile brands such as Nokia, T-mobile etc. The qualitative portion was conducted by using QDA Miner 4.) software package for coding textual data from twitter. From there, twitter, the plyr, stringr and the ggplot2 libraries in the R software were used to conduct the quantitative sentiment score. Using a graph, the bars are measured for particular brands for a certain score. This proved to be a great tool for brands to use to measure their standing in their target customer demography. However,Only using a limited dataset from twitter comments and by being only a text based sentiment analyzer, its functionality is pretty limited as this is not a multi-modal system. Also, this system would not be able to dissect the underlying emotions of a consumer who might not be able to express thoroughly, in a specific 140 word setting that twitter provides.

In terms of marketing perspective, The paper [7], focused on opinion mining which is domain-specific. They had built an architecture where in the first part , they collected information from various online shopping websites such as amazon, flipcart , snapdeal etc. The comments from the customers are used for feature extraction.

After feature extraction the comments are passed through the trainer classifier for finding the patterns of sentiments inside the comments , the patterns can be identified with techniques like N-Gram Extraction and part of speech Extraction are used by the trainer classifier. From here the collected comments are filtered by removing the irrelevant comments and clarity score is calculated. Through the use of trainer classifier and feature extraction from the comments the test classifier gives the feedback for a specified product which ranges from positive , negative and neutral. They hoped that by implementing this paper on the field of social networks, it can give an optimum solution to the problem of opinion acquisition. However , opinion mining from videos, which we are focusing on, tend to be much more accurate and goes far beyond the threshold of text based opinion mining.

Also, in a very similar manner to ours in the paper [8], The author constructed a system which would extract local feature from speech, and also from facial cue of patients using a handheld device with a camera or video cameras set around the room.
Using Support Vector machine or SVM it classifies these emotions and subsequently send the recognized state to a remote care center, healthcare professionals and providers for necessary services in order to provide seamless health monitoring. This paper demonstrated how effective the proposed approach could be with regards to face and speech processing. Our idea also matches the work pattern of this technique but we would implementing a modified and a more robust version of this from a business perspective where it could prove to be a very novel idea of market research.

We see that this multi-modal system could also be implemented in other fields where emotion could be used for media analysis such as news media as evidenced through the paper [9]. The paper presents a way of performing sentiment analysis of news videos, based on the linkage of audio, textual and visual clues extracted from their contents. Experimental results with a dataset containing 520 annotated excerpt of news videos from three Brazilian and one American TV newscasts show that the system achieved an accuracy of upto 84 percent in the sentiments classification task. This indicates that it has high potential to be used by media analysts in several applications, such as the journalistic domain.

For facial feature detection there have been many papers which gave multiple insights to track facial features to detect emotions. In the paper [10], it gives low computational cost while giving better detection performance for faces around the environment . Through a multi view face detection using color filtering, Adaboost

Learning Algorithm, a learning classification function and cascaded Detector it was able to properly detect faces even with various variables such as illumination, poses, various facial expressions, make-up, glasses etc.

We also tried to understand where the challenges of opinion mining that can be solved and integrated to our system. In the paper [11], they explored, analyzed various techniques, recent advancements and also future directions in the field of Sentiment Analysis. One of the challenges it presented was the ambiguity related to words in web user reviews. A word, in different context might mean different, which could be positive or negative. This conflict between the context and the words which are uttered can be resolved if the words are paired with facial expressions which then can provide a better understanding of what the customer is actually feeling about a certain product. Also since text based reviews can have many nuances such as contextual meanings, sarcasm gleaning over the actual emotions, capitalization to emphasis points, it is imperative to back the text based reviews up with other variables such as video analysis , audio analysis to detect emotions.

The paper [12], was one of the first papers to consider a multimodal sentiment analysis. It was a breakthrough in this field for that reason. They considered analyzing the audiovisual content along with text for sentiment analysis . From 47 videos that depicted a monologue , they chose 30-seconds excerpts which covered one topic and also transcribed the videos manually. Videos were graded along three catagories: positive, neutral and negative . The authors extracted sentiments from video feeds through the use of a commercial facial expression analysis software which took into account the duration of smiling and looking away by the interviewee. For language, word polarity was considered. Pitch and speech pause were extracted using openEAR. With these extracted features , Hidden Markov models were utilized for sentiment classification, which took this trimodal features as input. This trimodal sentiment detection exceeded the performance of unimodal ones.

The article [13], explored this issue by providing a multimodel system to extract sentiments from naturalistic video. Here , the system tries to extract sentiments from by combing results from audio analysis and video analysis. The dataset for the videos were created with videos where the speakers had different opinions , tonality, accent regarding various subjects. Using Keyword Spotting and Maximum entropy , speech from the videos were processed. The audio of the videos were processed with PocketSphinx. By finding keywords and tagging them and also calculating their frequency . Most frequent keywords are passed to maximum entropy algorithm. For the video part , a server within the JavaScript is used which handles requests for

video analysis which is run on the browser, all of which are controlled through a web-application. Through this the system predicts the emotions of a real-time video along with its audio.

In the paper [14], they have proposed a user independent fully automatic system for real time recognition of facial actions from the Facial Action Coding System (FACS). Their system automatically detects frontal faces in the video stream and codes each frame with respect to 20 Action units. They selected a subset of Gabor filters using AdaBoost and then trained Support Vector Machines on the outputs of the filters selected by AdaBoost. Their system was trained on FACS-coded images from 2 datasets. The first dataset was Cohn and Kanade's DFAT504 dataset. This dataset consists of 100 university students ranging in age from 18 to 30 years. 65 percent were female, 15 percent were African-American, and 3 percent were Asian or Latino. The second dataset consisted of directed facial actions from 24 subjects collected by Hager and Ekman. Here, they presented preliminary results for the performance of the system on spontaneous expressions. The system was able to detect facial actions despite the presence of speech, out-of-plane head movements that occur during discourse, and the fact that many of the action units occurred in combination. As a result of this, the system proved to be efficient enough to get a very accurate result.

In [15], it dealt with linking sentiments such as – Happy, neutral, sad with various emotions such as disgust, happy , anger etc through being treated as an application of GRU based inter-modal attention framework. This was evaluated through the benchmark dataset on multi-modal sentiment and emotion analysis (MOSEI) . The result of this experiment suggested that sentiment and emotion assist each other when learned in a multitask framework. This idea was used to evaluate how we could perceive the various emotions which would be extracted through our system.

In the paper [16], the authors have proposed the Multimodal EmotionLines Dataset (MELD) that is an extension and enhancement of EmotionLines. MELD contains about 13,000 utterances from 1433 dialogues from the TV-series "Friends". Here, each utterance is annotated with emotion and sentiment labels and encompasses audio, visual and textual modalities. They have used seven emotions for the annotation that include anger, disgust, fear, joy, neutral, sadness and surprise. They have again converted these seven emotions into more coarse-grained emotion classes such as positive, negative and neutral sentiments. In their work, they have included not only textual dialogues, but also their corresponding visual and audio counterparts. Also, they used popular toolkit openSMILE which extracts 6373 dimensional fea-

tures. They have used an LSTM model for unimodal audio for each audio utterance feature vector. They have also used DialogueRNN to model context by tracking individual speaker throughout the conversation emotion classification. For results, the best performance they could achieve 67.56 percent F-score by using DialogueRNN.

2.2 Supervised Learning

Supervised learning is the machine learning process of learning a function that uses an input to map an output based on example input-output pairs. It is simply a generalization of the idea of learning from example where a learner which is a computer program, is provided with two sets of data, a training set and a test set. The learner learn from a set of categorized examples in the training set so that it can identify unlabeled/uncategorized examples in the test set with the highest possible accuracy. The goal of the learner is to develop a guideline, a program, or a procedure that categorize new examples in the test set by analyzing examples it was previously given by the programmer that already have a class label. In supervised learning, the training set consists of n ordered pairs (x1, y1), (x2, y2),(x3, y3)..., (xn, yn), where each xn is some set of measurements of a single example data point, and yn is the label used for that data point. For example, an xn might be a group (which sometimes can also be called a vector) of four measurements for a patient in a hospital including height, weight, temperature,blood pressure. The corresponding yn might be a classification of the patient as either healthy or not healthy. The test data in supervised learning is another set of m measurement without labels: (xn+1, xn+2,xn+3 ..., xn+m). As described above, the goal is to make educated guesses, that follows a certain trend or path, about the labels for the test set by drawing inferences from the training data set with which it was trained with.

Generally, supervised learning operates with three main tasks:

- Binary classification: In the case of binary classification The algorithm classifies the data into two categories.
- Multiclass classification:In this classification the algorithm choose between more than two types of classification for a target variable.
- Regression:In regression model, it predict continuous values, but the classification models consider categorical ones.

Supervised Learning is imperative for for trend analysis because of the predictive nature it creates for itself based on past trends. It is designed to aid in building an cascading environment where the inputs are processed based on past records to produce new records . Below is a diagram of a supervise Learning System-

Supervised learning can be considered as the most business-oriented style of MachineLearning. It is less independent than unsupervised learning, where data is not classified as analysts might not know the target variables.Moreover It is also more practical than reinforcement learning, which only excels in closed game-like system architecture only.

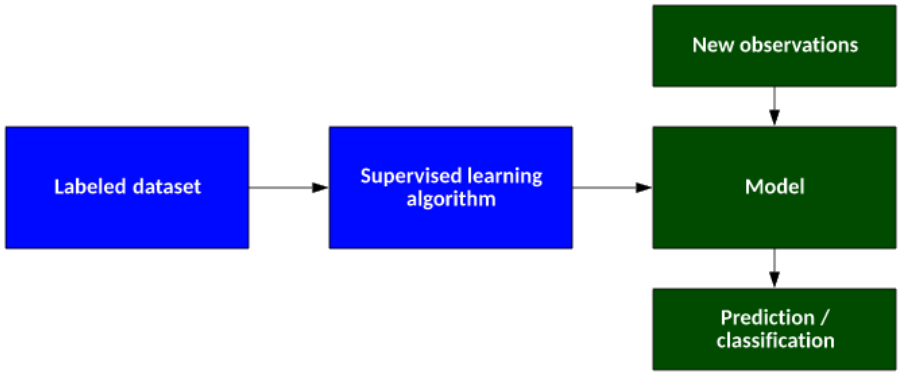

Figure 2.1: Supervised Learning Environment

2.3 Market Research

Companies which sell products and services can learn more about their current customers and target audiences and the reception of their products through market research. They can also use market research to learn even more about their business brands,reputation and other aspects of their organization.Market research can be defined as the process of gathering, analyzing, and interpreting information about a market, about a product or service to be offered for sale or already in circulation in that market, and about the past, present, and potential customers for the product or service.
Among the first steps in the business planning process is market research. Companies use it to help tackle such challenges as market segmentation,product requirements,design or product differentiation, the creation of an identity for a service or product to distinguish it from that of competitors. Market research companies gather, record, tabulate, and present data on marketing and public opinions of their current products and future products.

2.4 Amazon Rekognition

AWS Rekognition is a service that lets developers/programmers work with Amazon Web Services to add image,video analysis to their applications. With AWS Rekognition, application can detect, remember and recognize objects, scenes, and faces in images and videos. Moreover with Rekognition we can search and compare features for faces.
With the application programming interface provided by Rekognition, we can add a level of sophistication to our application environment with visual search and advanced image classification which is based on deep learning—the subset of machine learning which is aimed at modeling high-level abstractions using multiple nonlinear transformation neural networks and it's architecture.
The uses of this API include-

- Library of searchable images:Amazon Rekognition makes images and videos searchable, which enables developers and programmers to discover objects and scenes that appear in them.

- Facial verification of users: We can confirm the identities of users comparing their image in real time with their previously stored reference image.
- Opinion Analysis:We can Detect emotions such as happiness, sadness and surprise from facial images. A developer or programmer can analyze images in real time and store the emotional values of various emotional aspects.
- Facial recognition: We can store facial metadata and facilitate the search of your image of facial images with the IndexFaces function in the API

For emotion detection, it uses it api to give out a result which has a "String" type with numerical types attached with each of them. The valid values of emotional states are as the following-

- HAPPY
- SAD
- ANGRY
- CONFUSED
- DISGUSTED
- SURPRISED
- CALM
- UNKNOWN
- FEAR

For each of the item it provides a percentage which ranges from 0-1. With anything closer to 1 meaning that , that emotional state is the most likely state of a specific human being. Anything closer to 0 means, it is not the emotional state to consider as the likely emotional state of a person.

2.5 Algorithms

To implement our research on supervised machine learning for sentiment analysis, we had to use various different algorithms. The Algorithms that we had used were-XGBoost, RNN , SVM

2.5.1 XGBoost

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework.XGBoost is an ensemble learning method. Sometimes, it might be sufficient enough to rely upon the results of just a singular machine learning model. Ensemble learning provides a systematic solution to integrate the predictive power of multiple learners. The result is a single model which gives the aggregated output of several models.

The models that form the ensemble, which is also known as base learners, could be either from the same domain learning algorithm or different learning algorithms. Bagging and boosting are two widely used ensemble learners.Its most predominant usage has been with decision trees.

In boosting, the decision trees are built in a sequential manner such that each subsequent tree reduce the errors of the previously generated tree. Each tree learns from its previous trees and updates the residual errors. So, the tree that grows next in the sequence will learn from an updated version of the residuals/error.
The base learners in boosting are weak/preliminary learners in which the bias is very high, and the predictive power is just a small amount of better than random guessing.However, Each of these weak learners contributes some important information for prediction, allowing the boosting technique to produce a strong learner by effectively integrating these weak learners. The final strong learner brings down both the bias and the variance.

In comparison to bagging techniques like Random Forest, in which trees are grown to their maximum depth, boosting makes use of these trees with fewer splits. Such small trees, which are not very deep, are highly interpretable. Parameters like the number of trees or iterations, the rate at which the gradient boosting learns, and the depth of the tree, could be optimally selected through validation techniques like k-fold cross validation. Having a large number of trees might lead to overfitting. So, it is necessary to carefully choose the termination criteria for boosting.
Boosting consists of three simple steps:

- An initial model F0 is defined to predict the target variable y. This model will be associated with a residual (y – F0)
- A new model h1 is fit to the residuals from the previous step
- Now, F0 and h1 are combined to give F1, the boosted version of F0. The mean squared error from F1 will be lower than that from F0:

$$F_1(x) < -F_0(x) + h_1(x)$$

To improve the performance of F1, we could model after the residuals of F1 and create a new model F2:

$$F_2(x) < -F_1(x) + h_2(x)$$

This can be done for 'm' iterations, until residuals have been minimized as much as possible:

$$F_m(x) < -F_(m-1)(x) + h_m(x)$$

Here, the additive learners do not disturb the functions created in the previous steps. Instead, they impart information of their own to bring down the errors.
A gradient boosting is a form of boosting which improves the gradient descent algorithm. Many gradient Boosting follow the below algorithm to minimize the objective function-
input:training set$(x_i, y_i)_{i=1}^{n}$, a differentiable loss function L(y,F(x)), number of iterations M.

Algorihtm:

1. initialize model with a constant value:

$$F_0(x) = argmin \sum_{i=1}^{n} L(y_i, \gamma).$$

(2.1)

2. m=1 to M:

   (a) Compute so-called pseudo-residuals:

   $$r_{im} = -[\delta L(y_i, F(x_i))/\delta F(x_i)]_{F(x)=F_{m-1}(x)} \tag{2.2}$$

   For i= 1,...,n

   (b) Fit a base Learner(e.g.tree) $h_m(x)$ to pseudo-residuals, i.e train it using the training set $(x_i, r_{im})_{i=1}^{n}$

   (c) Compute multiplier γ_m by solving the following one-dimensional optimization problem:

   $$\gamma_m = argmin \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)). \tag{2.3}$$

   (d) Update the model:

   $$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{2.4}$$

3. Output $F_m(x)$

The intuition is by fitting a base learner to the negative gradient at each iteration is essentially performing gradient descent on the loss function . The negative gradients are often called as pseudo residuals, as they indirectly help us to minimize the objective function.

XGBoost, in short for "Extreme Gradient Boosting" , takes this theory of Gradient boosting even further and refines it. As shown previously, GBM divides the optimization problem into two parts by first determining the direction of the step and then optimizing the step length. Different from GBM, XGBoost tries to determine the step directly by solving,

$$\frac{\delta L(y, f_{m-1}(x) + f_m(x))}{\delta f_m(x)} = 0 \tag{2.5}$$

For each x in the data set, by doing second-order Taylor expansion of the loss function around the current estimate $f(m-1)(x)$, we get-

$$L(y, f^{(m-1)}(x) + f_m(x))$$
$$\approx L(y, f^{(m-1)}(x)) + g_m(x) f_m(x) + \frac{1}{2} h_m(x) f_m(x)^2,$$

where $g_m(x)$ is the gradient, same as the one in Gradient Boosting, and $h_m(x)$ is the Hessian (second order derivative) at the current estimate:

$$h_m(x) = \frac{\partial^2 L(Y, f(x))}{\partial f(x)^2}_{f(x)=f^{(m-1)}(x).}$$

Then the loss function can be rewritten as :

$$L(f_m) \approx \sum_{i=1}^{n}[g_m(x_i)f_m(x_i) + \frac{1}{2}h_m(x_i)f_m(x_i)^2] + const.$$
$$\propto \sum_{j=1}^{T_m}\sum_{i\in R_{jm}}[g_m(x_i)w_{jm} + \frac{1}{2}h_m(x_i)w_{jm}^2].$$

Letting G_{jm} represents the sum of gradient in region j and H_{jm} equals to the sum of hessian in j region, the equation can be rewritten as=

$$L(f_m) \propto \sum_{j=1}^{T_m}[G_{jm}w_{jm} + \frac{1}{2}H_{jm}w_{jm}^2].$$

Now, with the fixed already trained structure, for each region, it is straightforward to determine the optimal weight :
$w_{jm} = -\frac{G_{jm}}{H_{jm}}, j = 1, ..., T_m$
Putting this back to the loss function, we get , $L(f_m) - \frac{1}{2}\sum_{j=1}^{T} m\frac{G_{jm}^2}{H_{jm}}$
This is the structure score for a tree. The smaller the score, the better and more optimized the structure. Thus for each split to make ,the proxy gain is defined as-

$$Gain = \frac{1}{2}[\frac{G_{jmL}^2}{H_{jmL}} + \frac{G_{jmR}^2}{H_{jmR}} - \frac{G_{jm}^2}{H_{jm}}]$$
$$= \frac{1}{2}[\frac{G_{jmL}^2}{H_{jmL}} + \frac{G_{jmR}^2}{H_{jmR}} - \frac{(G_{jmL} + G_{jmR})^2}{H_{jmL} + H_{jmR}}].$$

Taking regulariztion into consideration, we can rewrite the loss function as -

$$L(f_m) \propto \sum_{j=1}^{T_m}[G_{jm}w_{jm} + \frac{1}{2}H_{jm}w_{jm}^2] + \gamma T_m + \frac{1}{2}\lambda\sum_{j=1}^{T_m}w_{jm}^2 + \alpha\sum_{j=1}^{T_m}|w_{jm}|$$
$$= \sum_{j=1}^{T_m}[G_{jm}w_{jm} + \frac{1}{2}(H_{jm} + \lambda)w_{jm}^2 + \alpha|w_{jm}|] + \gamma T_m,$$

where is the penalization term on the number of terminal nodes, and are for L1 and L2 regularization respectively. The optimal weight for each region j is calculated as:

$$w_{jm} = \begin{cases} -\frac{G_{jm}+\alpha}{H_{jm}+\lambda} & G_{jm} < -\alpha, \\ -\frac{G_{jm}-\alpha}{H_{jm}+\lambda} & G_{jm} > \alpha, \\ 0 & else. \end{cases}$$

The gain of each split is defined correspondingly:

$$Gain = \frac{1}{2}[\frac{T_\alpha(G_{jmL})^2}{H_{jmL}+\lambda} + \frac{T_\alpha(G_{jmR})^2}{H_{jmR}+\lambda} - \frac{T_\alpha(G_{jm})^2}{H_{jm}+\lambda}] - \gamma$$

$$T_\alpha(G) = \begin{cases} G+\alpha & G < -\alpha, \\ G-\alpha & G > \alpha, \\ 0 & else. \end{cases}$$

This is how the XGBoost is defined, to produce a refined results when compared to other algorithm of the same kind.It deliver higher performance and accuracy when compared to other boosting algorithms.

2.5.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm which is used for both classification or regression. However, it is mainly used for test classification. In the algorithm each data will be defined as a plot as a point in n -dimensional space with the value of each feature in the feature set. n is the number of features in the domain. Then classification is done to find the hyper-plane to differentiate two classes very well. Normally this model draws lines to separate the groups according to patterns each group uniquely has. Below is a diagram that shows the classifications of two classes-

Figure 2.2: SVM Classifications of two classes

Here, we can see that there are two different class in the graph. Support Vector Machine will find a hyper-plane to separate two different classes as shown from the figure below-

The distance between the hyperplane and the nearest data point from any set is known as the margin. The goal is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, giving a greater chance of new data being classified accurately.To construct a hyper-plane line SVM employs an iterative training algorithm, which is used to minimize the error function.

Figure 2.3: SVM Classifications using hyper-line

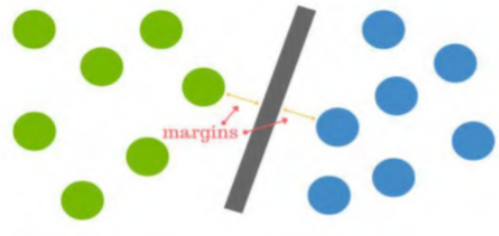

Figure 2.4: Best hyper-plane

From the above figure, we get a classic example of a linear classifier. A classifier that separates a set of objects into their respective groups (GREEN and BLUE in this case) along with a line. However,most classification tasks are not that easy, and often is more complex and complex structures are needed in order to make an optimal separation. In real world data is rarely as clean as the figure. Datasets will be mixed as the below figure-

Figure 2.5: SVM Classifications with jumbled data

To categorize the above jumbled dataset, it is necessary to move away from 2D view of the data towards a more 3D view. Normally for jumbled datasets it is imperative to represent all the data to a higher dimension. This is called "kernelling" .AS we consider our dataset in higher dimension our hyper-plane will not be a line this time, but it will be like the below-

In our research we used Scikit learn library to implement the support vector machine algorithm

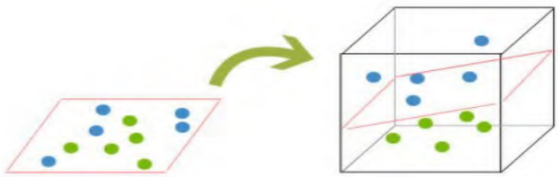

Figure 2.6: SVM Classifications in 3D dimension

2.5.3 Random Forest

Random forests, also known as random decision forests, are a popular ensemble method that can be used to establish predictive models for both classification and regression problems. It is a flexible, simple to use machine learning algorithm that generates, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because its simplicity. Also for the fact that it can be used for both classification and regression tasks.

Random Forest is a supervised learning algorithm. As with its name, it creates a forest and makes it random. The 'forest' it builds, is an accumulation of Decision Trees, most of the time trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result. Simply random forests build multitude of decision tress and merge them to get more accurate and refined and also a more stable predictions. There are three main choices to be made when constructing a random tree. These are

1. the method for splitting the leafs
2. the type of predictor to use in each leaf
3. the method for injecting randomness into the trees

. Each tree is grown as follows:

1. If the number of cases in the training set is N, sample N cases at random - but with replacement, from the original data. This sample will be the training set for growing the tree.
2. If there are K input variables, a number m¡¡K is specified such that at each node, k variables are selected at random out of the K and the best split on these k is used to split the node. The value of k is held constant during the forest growing
3. Each tree is grown to the largest extent possible. There is no pruning

Increasing the correlation between the trees increases the forest error rate. The strength of each individual tree in the forest make the forest better. A tree with a low error rate is a strong classifier.So increasing the strength of the individual trees decreases the forest error rate. Once the forest has been trained it can be used to make predictions for new uncategorized data points. To make a prediction for a query point x, each tree independently predicts as below equation -

$$f_n^j(x) = \frac{1}{N^e(A_n(x))} \sum_{Y_i \in A_n(x) I_i = e} Y_i \tag{2.6}$$

Here An(x) denotes the leaf containing x. Ne(An(x)) denotes the number of estimation points it contains.
The predictions made by each tree depend only on the estimation made by that tree. However, since points are assigned to the structure and estimation parts independently in each tree, structure points in one tree have the opportunity to contribute to the prediction as estimation points in another tree
In our research, we use random forest for Feature selection purpose. In python we use random forest(rf) library to implement the code.

2.5.4 NAÏVE BAYES

A Naive Bayes classifier is an algorithm that uses Bayes' theorem to classify data.A Naïve Bayes classifier is termed naïve because it assumes that all attributes of a data point are independent of each other. Naive Bayes classifiers assume strong, or naive, independence between attributes of various data points. The classifier uses probability theory to categorize objects. The main key theme of Bayes' theorem is that the probability of an event can be adjusted as new data are being introduced. A Naïve Bayes model is easy to build and it has no complex cascading parameters estimation, thus making it particularly useful for very large datasets [17]. As mentioned earlier Bayes theorem provides a way to calculate the posterior probability $P(c|x)$, from $P(c)$, $P(x)$ and $P(x|c)$. The mathematical representation of equation is given below -

$$p(c|x) = \frac{p(x|c)p(c)}{P(x)} \tag{2.7}$$

Here, $P(c|x)$ is the posterior probability of target class (c) given predictor (x). $P(c)$ is the prior probability of class. $P(x|c)$ is the probability of predictor given class. $P(x)$ is the prior probability of predictor class.
In Scikit learn there are three different types of Naïve Bayes model under the library.

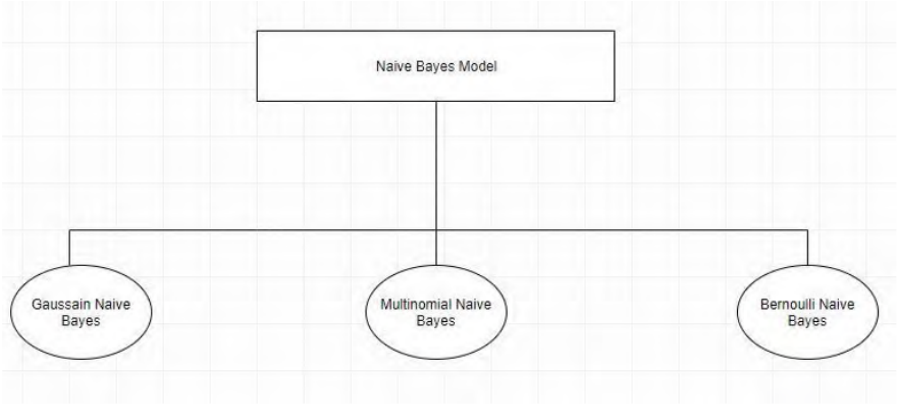

Figure 2.7: different types of Naïve Bayes model

In our research Paper We use Gaussian naive Bayes.

When dealing with continuous data, such as ours, a typical assumption is that the continuous values associated with each class are distributed according to a

Gaussian distribution.For example, suppose the training data contains a continuous attribute,x. we first segment the data by the class, and then compute the mean and variance of x in each class. Let μ_k be the mean of the values in x associated with class C_k, and let σ_k^2 be the Bessel corrected variance of the values in x associated with class C_k. Suppose we have collected some observation value v. Then, the probability distribution of v given a class C_k, $p(x = v \mid C_k)$, can be computed by plugging v into the equation for a normal distribution parameterized by μ_k and σ_k^2. That is,

$$\text{p(x=v}\mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\, e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}} \quad (2.8)$$

2.5.5 Logistic Regression

Logistic regression is the regression analysis to conduct when the dependent variable is binary. Like all regression analysis, the logistic regression is a predictive analysis. Logistic regression is used to give meaning to the data and explain the relationship between one dependent binary variable and one or more nominal, ordinal, ratio-level independent variables. This model tries to measure the relationship between categorical dependent variable and one or more independent variables by plotting the dependent variables probability scores on a graph. The mathematical equation of this is given below-

$$\log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m \quad (2.9)$$

Here,
β_i are the regression coefficients. x_i explanatory variables.

2.5.6 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is most commonly used as dimensionality reduction technique in the pre-processing step for pattern-classification and machine learning applications.
The goal is to project a dataset onto a lower-dimensional space with good class-separability in order avoid overfitting ("curse of dimensionality") and also reduce computational costs. LDA is like PCA , but it focuses on maximizing the seperatibility among known categories.In addition to finding the component axes that maximize the variance of our data (PCA), we are additionally interested in the axes that maximize the separation between multiple classes (LDA). The LDA Approach can explained in 5 steps.

1. Compute the d-dimensional mean vectors for the different classes from the dataset.

2. Compute the scatter matrices(in-between-class and within-class scatter matrix)

3. Compute the eigenvectors(e_1,e_2,....e_d) and corresponding eigenvalues(λ_1,λ_2,..λ_d) for the scatter matrices.

4. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a $d \times k$ dimensional W (where every column represents an eigenvector).

5. Use this $d \times k$ eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the matix multiplication $Y = X \times W$ where X is a $n \times d$-dimensional matrix representing the n samples, and y are the transformed $n \times k$ -dimensional samples in the new subspace.

2.5.7 k-nearest neighbors(KNN)

The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.The KNN algorithm assumes that similar things exist in close proximity.In k-NN classification, the output is a class categorization. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. Below is the explanation of the algorithm-

1. Load the data

2. Initialize K to the chosen number of neighbors

3. For each example in the data,we need to calculate the distance between the query example and the current example from the data.

4. Add the distance and the index of the example to an ordered collection

5. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances

6. Pick the first K entries from the sorted collection

7. Get the labels of the selected K entries

8. If regression, return the mean of the K labels

9. If classification, return the mode of the K labels

From a mathemetical setting, we can say that Suppose we have pairs $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ taking values in $R^d \times \{1, 2\}$, where Y is the class label of X, so that $X|Y = r \sim P_r$ for $r = 1, 2$ (and probability distributions P_r). Given some norm $\|\cdot\|$ on R^d and a point $x \in R^d$, let $(X_{(1)}, Y_{(1)}), \ldots, (X_{(n)}, Y_{(n)})$ be a reordering of the training data such that $\|X_{(1)} - x\| \leq \cdots \leq \|X_{(n)} - x\|$.
We had used knn algorithm for feature selection.

Chapter 3

Proposed Model

3.1 Model Overview

Our model is split into two portions. The first portion is the video based sentiment analysis. The second portion deals with extracting the audio sentiment via texts that are generated from the video in real time. Our model in total consist of the following steps.

1. Data Collection through real time video capturing devices such as CC camera.
2. Data splitting into video and audio
3. Further splitting into train data and test data
4. Data Pre-processing to clean and prepare data.
5. Speech to text converting using speech-Recognition using python.
6. Choosing the most common features from the training dataset for text based Sentiment Analysis.
7. Training and testing various algorithms with the dataset.
8. Applying Sentiment analysis on the review data, with individual features to get polarity in text based sentiment analysis.
9. Frame to Frame sentiment extraction of the videos
10. Using opencv for Video to frame Splitting
11. Fusing the sentiment analysis of two different results
12. Getting a final feature based sentiment of the target customer

The figure of the work process of our system is given below.

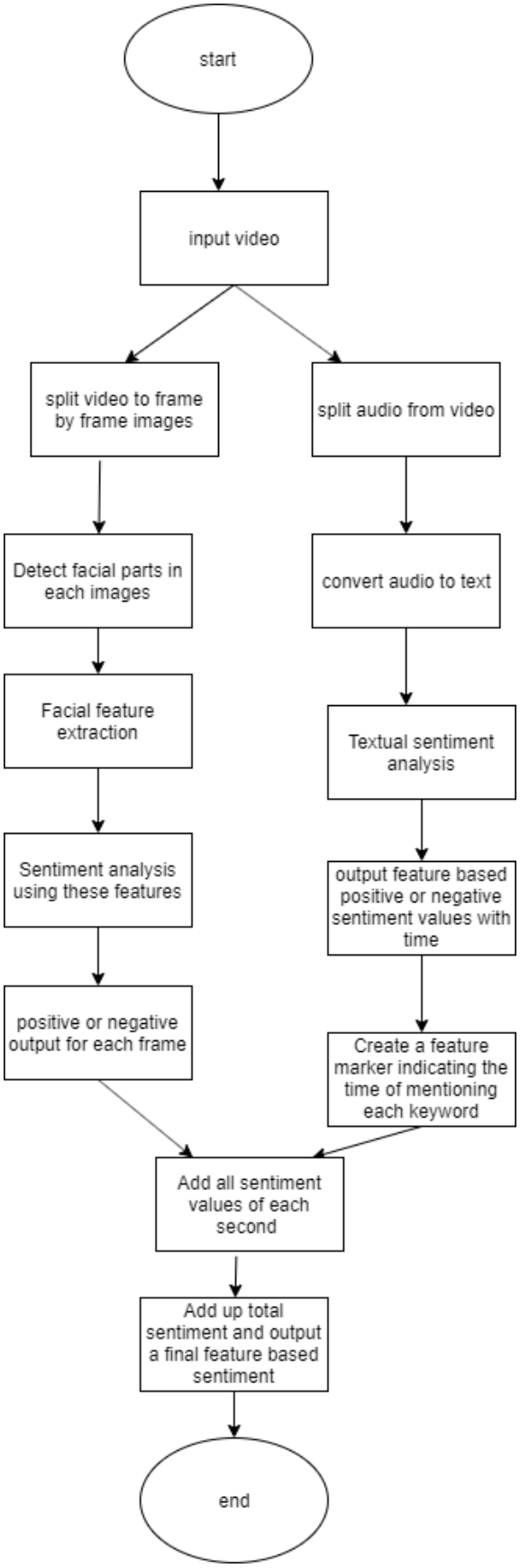

Figure 3.1: Work flow of the proposed system

3.2 Data Collection

In the research of sentiment analysis, the most technical and one of the most complex tasks is to find out the proper dataset applicable for the model.The data , which are videos of the customers who come to the shop to test out the products themselves. There would be multiple CC cameras around the shop. Any customer that enters the shop would be monitored via Camera. This way we get videos of the customers of a specific product. This would be a great way to observe a target demography of a certain product.

For our experiment purposes,we had taken a survey of smartphone users. We had used Smartphones as our chosen product for experimentation and data collection. We conducted a small video survey on various smartphone users.We asked 7 question. The questions were-

- What is your phone name?
- How is your camera?
- How is the storage of the phone?
- How fast is your phone?
- How is the battery life?
- How do you like the display?
- How is the performance compared to the price?

The reason we had chosen these question were because we would get specific responses for specific features of our product which in this case was smartphone. This would give us feature specific sentiments. However , it all boils down to if the customers are truly satisfied with their respective phones. A decisive answer would be used as a standard for our system . For that the customers are asked one final question which is-

- Are you satisfied with your phone?

If the answer to this question is yes, Then it would mean that overall sentiment for the product is positive. If the answer to this question is no, Then it would mean that overall sentiment for the product is negative.These answers will be used as the standards for our test. There would be an average sentiment output which we will get from the 7 questions. It could either be positive or negative. We will compare the result with the result we get from the final question.

For audio, we used the videos again. Here we had used the audio from the video. We used a Video to audio converter to extract the audio from the video. Then from the audio we used Speech-Recognition using python. We also had trained our model

for textual sentiment analysis using the Amazon's unlocked mobile phone review's dataset. There were about four hundred thousand data available in the dataset. The dataset contained in total six columns, which were 'product name', 'company name', 'review', 'rating', 'rating votes'. Three of the columns consists of textual data and rest are numerical data. The 'Product Name' column is the model name of the mobile phone whereas the 'Company Name' is the name of the company. The review column consists of reviews provided by users .Rating column is the rating of each products given by users at Amazon. We used this dataset to train our system for text based sentiment analysis. After training the system with this input data, we inserted the answers to the question to perform a sentiment analysis on what the customers say about the product.

3.3 Data Splitting

Data splitting is the process of splitting the dataset into training and testing data. This process is very useful and also needed for any machine learning process as the main idea of machine learning depends on training and testing data and finding the accuracy of the machine given output.
We actually perform two different data splitting for our system. As our system is a multimodal system at first we split the video and audio from the input video. We use a audio converter to split the audio. The videos are kept as it is.

We split the videos into frame and detect the faces using Opencv. We randomly split 64 percent of our dataset in training data, 16 percent in cross-validation set and 20 percent in testing set using train-test-split method of sklearn in python.

For the audio, we first trained our data using the Amazon mobile-phone unlocked dataset. First and foremost , we had to reduce our data to two thousand from the previously collected four hundred thousand as the processing power is not enough to handle this amount of data.Here the algorithms will be trained using the review and the rating column. We set the rating 5 and 4 as a positive sentiment and 1 and 2 as negative sentiment. The algorithms will be trained with those information.We will apply the trained algorithm to our test dataset and measure the accuracy of the system. Furthermore, we will use the extracted texts from the audio of the video survey as test data to gauge the accuracy of our system.

3.4 Pre-processing

Data preprocessing is the process of cleaning the data to an extent where the machine learning algorithm will understand the overall content of the data. Overall the data preprocessing deals with finding the inaccurate, irrelevant and incomplete data and removing or those data to clean up the dataset.

For the videos,Frame to Frame analysis is done using aws Rekognition API.Amazon Rekognition made it easier to do image and video analysis in our applications. we provided the frame to frame image of the video to the Rekognition API, and the service automatically identified the objects, people, text, scenes, and activities, as

well as detect any inappropriate content. Since Amazon Rekognition also provides highly accurate facial analysis and facial recognition on images and video that we provided, it was then used to create a statistic emotions of a person. The statistic provided a percentage of 9 emotions,which were-

- Happy
- Sad
- Angry
- Confused
- Disgusted
- Surprised
- Calm
- Unknown
- Fear

In the pre-processing portion we gather this percentage from the videos that were taken by us. This percentage would later be used as indicator of whether or not , the customer is happy or sad with their respective phones or a demo phone that they were presented with for judging.

For the audio, we first used an application from Google to extract the textual manuscript from the audio. After getting the textual manuscript, we first used tokenization from Keras pre-processing library. Tokenization is a step in which it splits longer strings of text into smaller pieces, or tokens. Larger chunks of text can be tokenized into sentences and sentences can be further tokenized into words.

After tokenizing , we decided to assign a number for each token/words. In order to that, we used text-to-sequences function from keras. Using this we had assigned an unique number for each tokens.

We then pad out the sequences of sentences to the same length of the longest sequence. Using Pad-sequence function, we do this. If a sequence is shorter than the longest sequence that it is padded out with a value at the end. Any sequences that are longer than the prescribed sequence length, are truncated so that they fit the desired length.

3.5 Feature Selection

Since our model concentrated on integrating various emotions of a person to give a comprehensive view of their emotional state(eg:positive, negative) towards a certain product, it was extremely important for us to focus on the correct features.

3.5.1 Feature importance

As previously discussed, using AWS Rekognition, we had extracted in total of 9 emotions from a person , and had a percentage attached to each corresponding emotion.We can get the feature importance of each feature of your dataset by using the feature importance property of the model. Feature importance gives you a score for each feature of your data, the higher the score more important or relevant is the feature towards your output variable. Decision Trees models which are based on ensembles (eg. Extra Trees and Random Forest) can be used to rank the importance of the different features. Knowing which features our model is giving most importance can be of vital importance to understand how our model is making it's predictions (therefore making it more explainable). At the same time, we can get rid of the features which do not bring any benefit to our model (or confuse it to make a wrong decision!).Below is a Feature importance plot for Random Forrest:

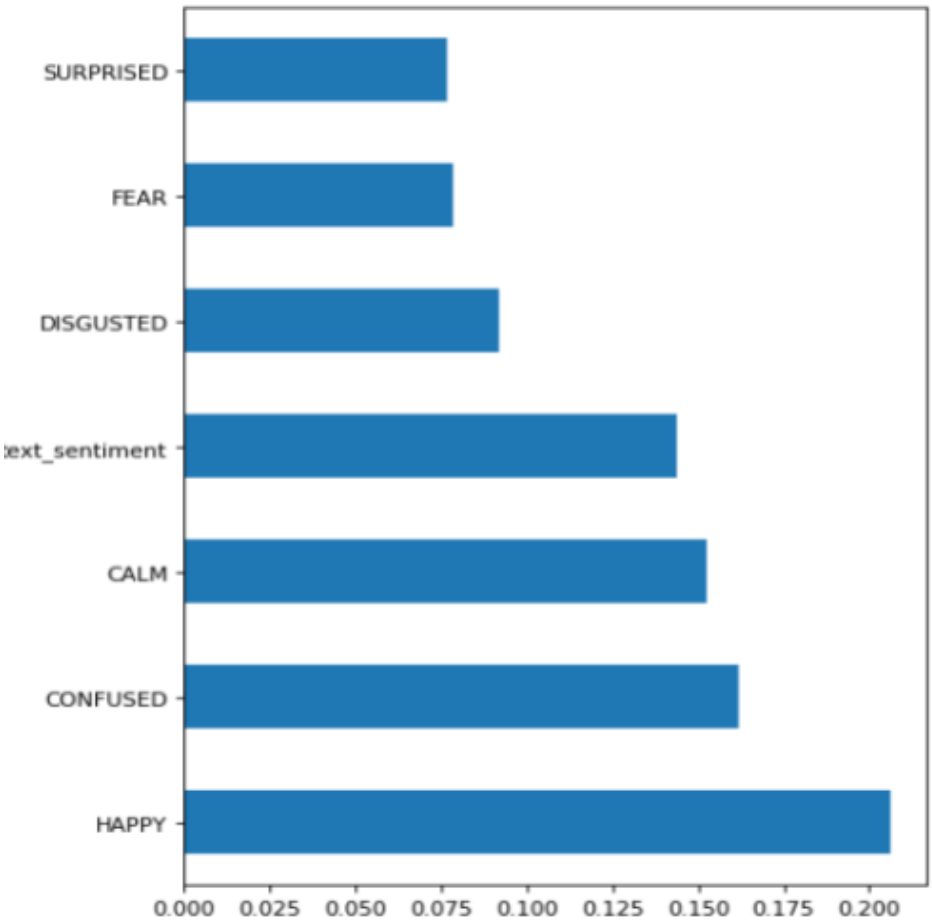

Figure 3.2: Feature importance plot for Random Forrest

We also had used XGBoost to get feature importance values for each features. The graph of that is given in figure 3.3-

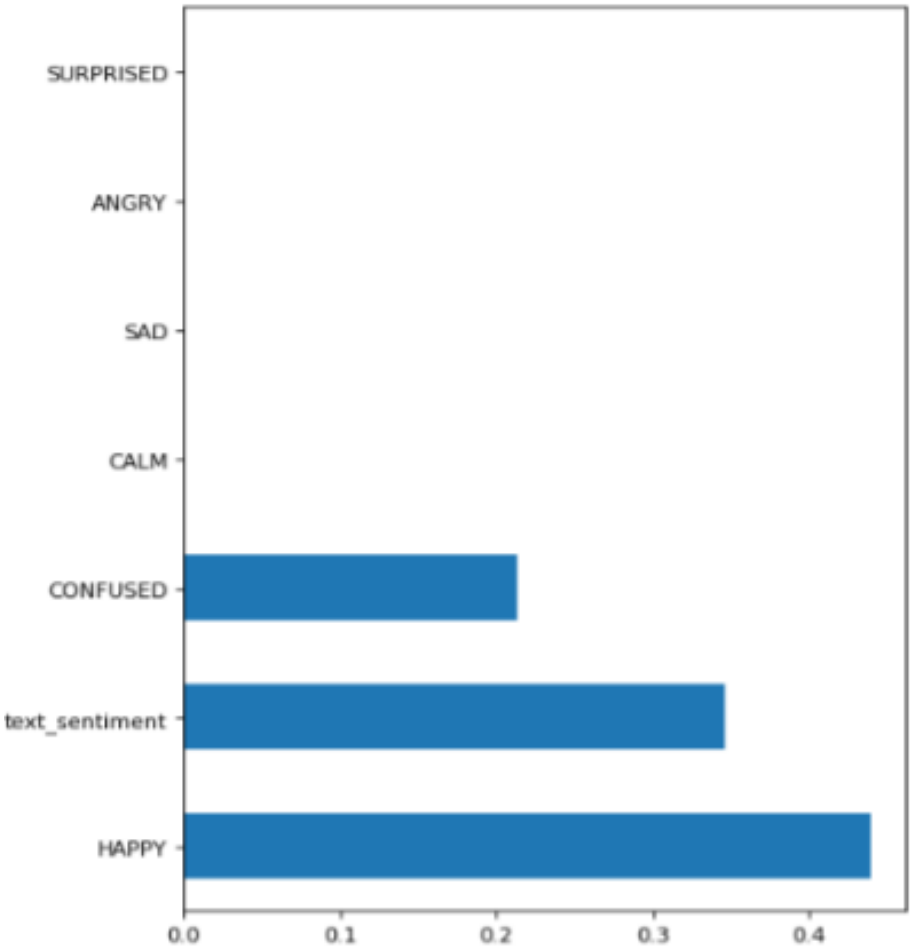

Figure 3.3: Feature importance plot from XGBoost

Feature importance plot for Extremely Randomized Tree is shown in figure 3.4.

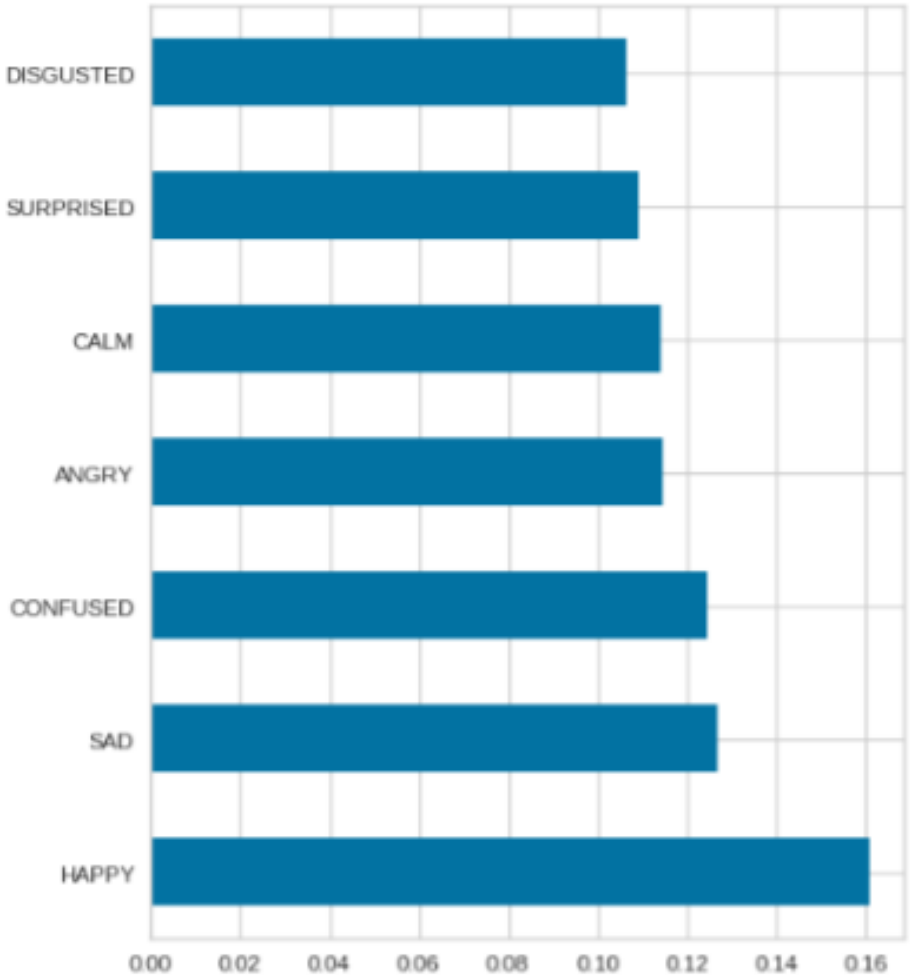

Figure 3.4: Feature importance plot from Extremely Randomized Tree

We can also understand how to perform Feature Selection by visualizing a trained Decision Tree structure.The features which will be at the top of the tree structure are the ones our model retained most important in order to perform its classification. Therefore by picking just the first few features at the top and discarding the others could possibly lead to creating a model which an appreciable accuracy score.

Using Decision tree we can visualize the data.

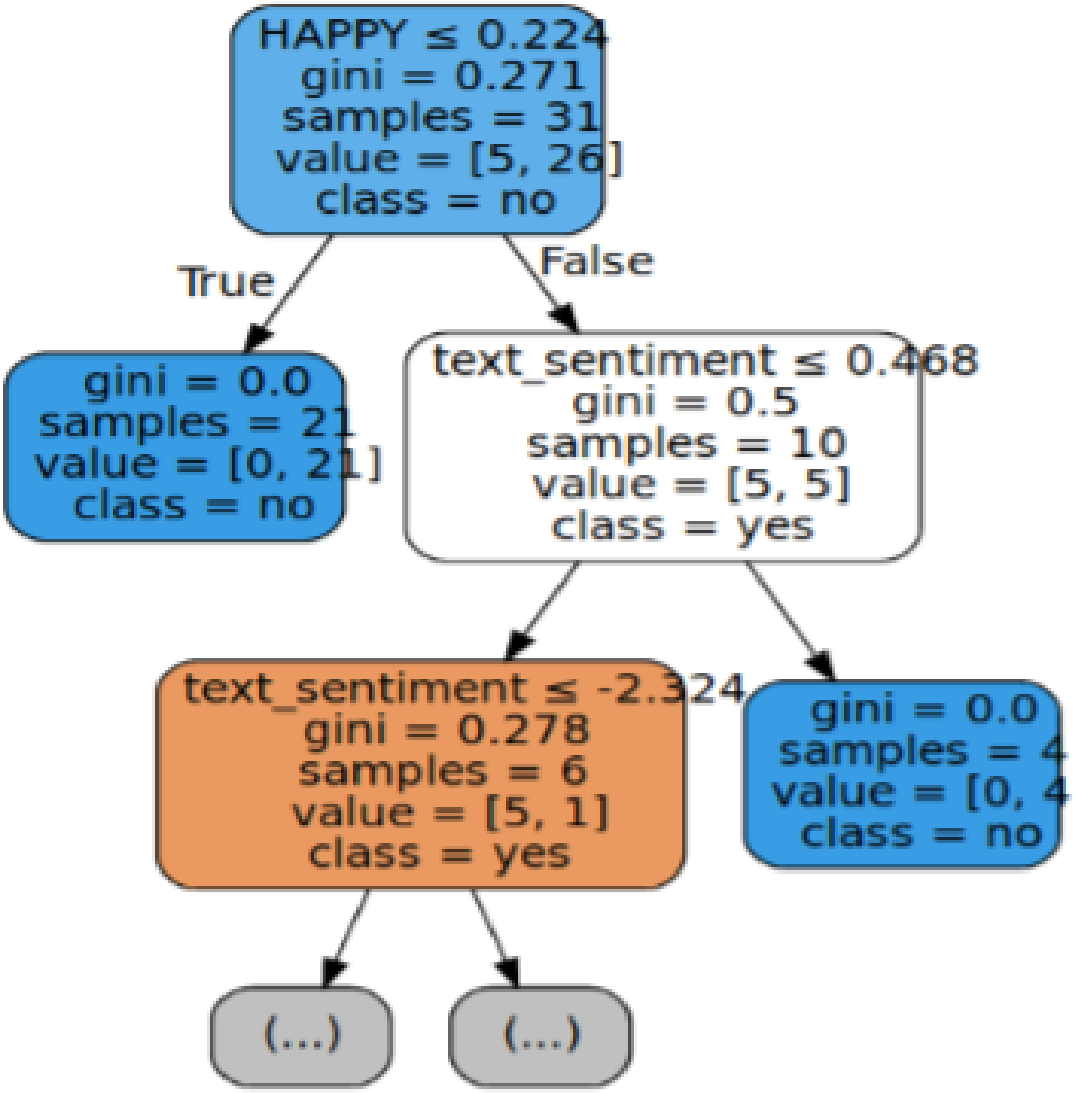

Figure 3.5: Decision Tree visualization

In figure 3.5 the top 2 features indicated by the decision tree are- Happy and Text-sentiment

3.5.2 Recursuve Feature Elimination(RFE)

Recursive feature elimination (RFE) is a feature selection method that fits a model and removes the weakest feature (or features) until the specified number of features is reached. Features are ranked by the model's `coef_` or `feature_importances` attributes, and by recursively eliminating a small number of features per loop, RFE attempts to eliminate dependencies and collinearity that may exist in the model. RFE requires a specified number of features to keep, however it is often not known in

advance how many features are valid. To find the optimal number of features cross-validation is used with RFE to score different feature subsets and select the best scoring collection of features. The RFECV visualizer plots the number of features in the model along with their cross-validated test score and variability and visualizes the selected number of features.

Using Stratified K-fold cross validation with 4 folds and XGBoost classifier as the model the RFE curve is as follows -

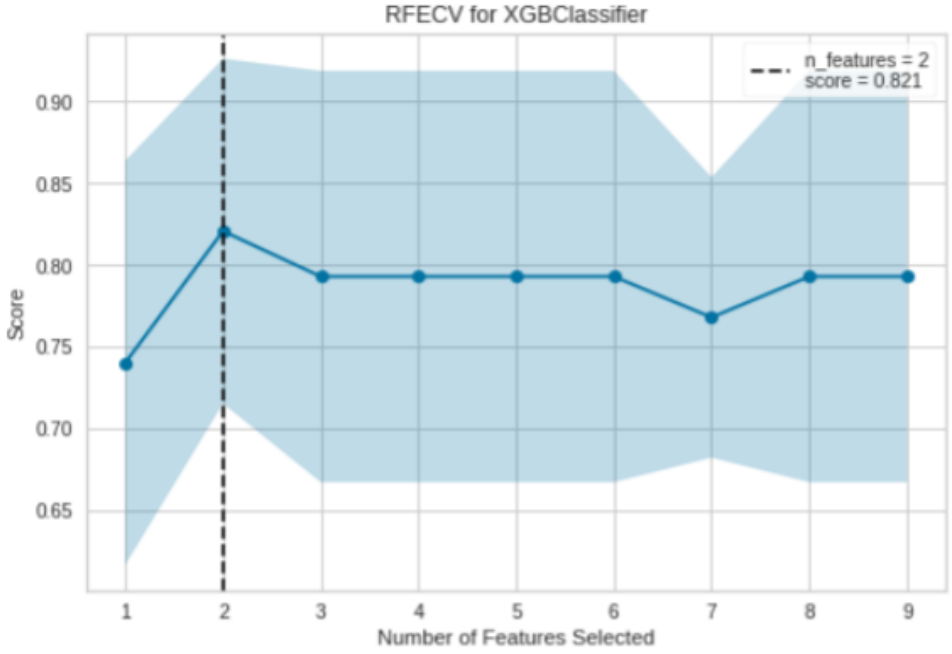

Figure 3.6: RFE Curve

It is visible that with 2 features the accuracy is about 82.1 Percent for XGB classifier. Most significant features are "HAPPY" "textsentiment"

Feature importance plot of RFE for XGB-

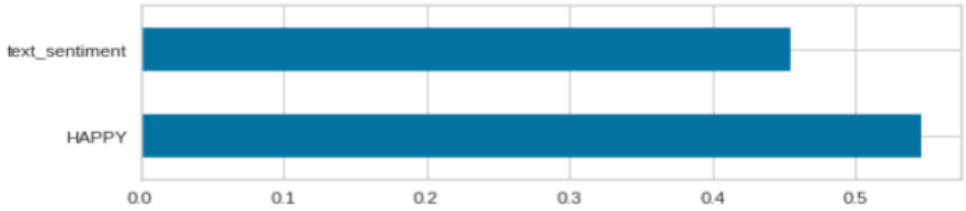

Figure 3.7: RFE Plot

Using Stratified K-fold cross validation with 4 folds and Random Forest classifier as the model the RFE curve is as follows -

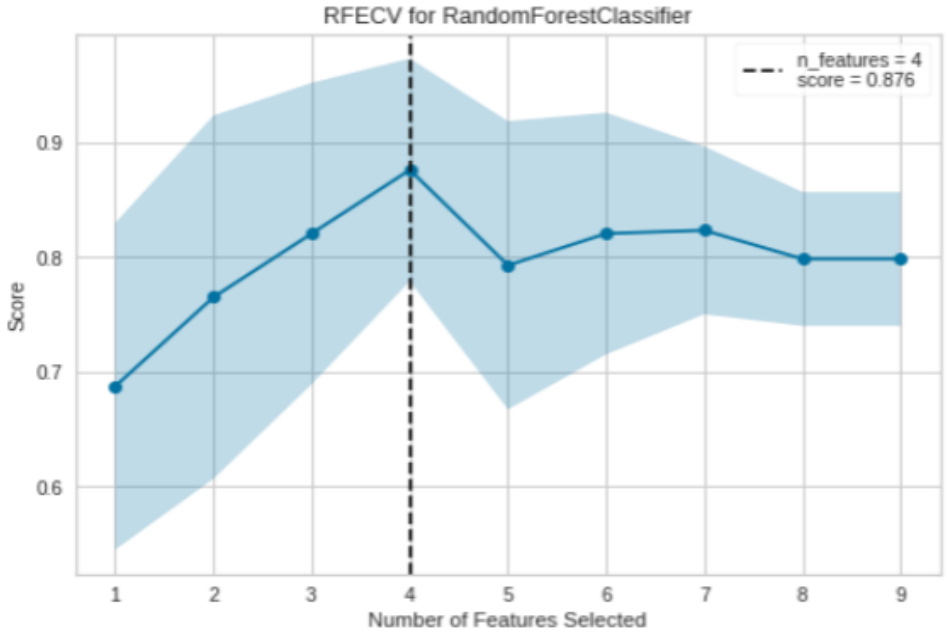

Figure 3.8: RFE Curve with random forest classifier

It is visible that with 4 features the accuracy is about 87.6 percent for RF classifier. Most significant features are CALM, HAPPY, ANGRY, text-sentiment.

Feature importance plot of RFE for Random Forest-

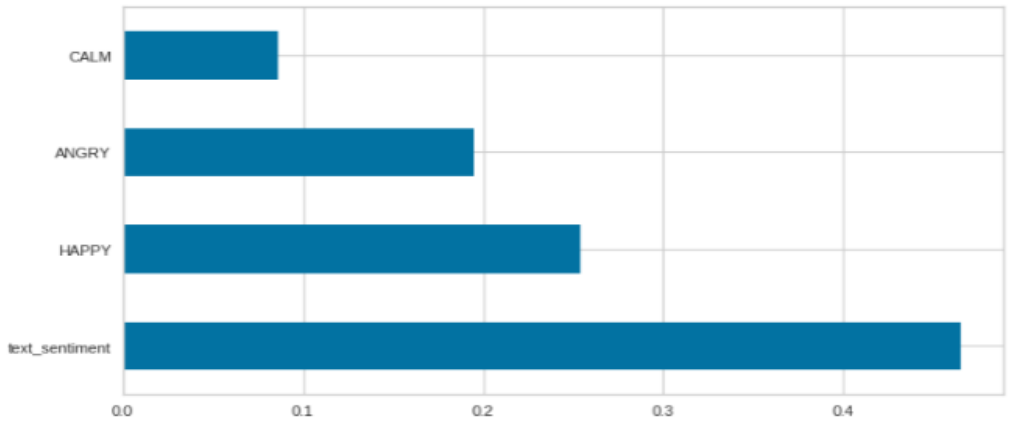

Figure 3.9: RFE plot with random forest classifier

Using Stratified K-fold cross validation with 4 folds and Extra trees classifier as the model the RFE curve is as follows-

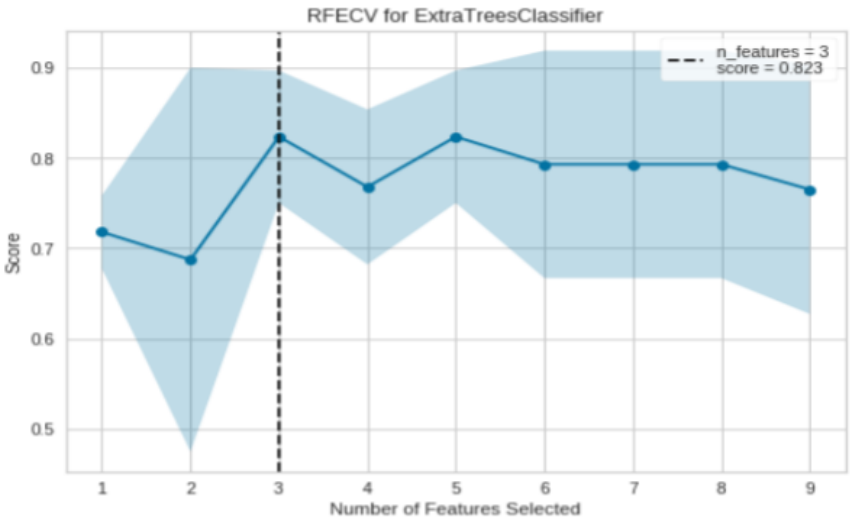

Figure 3.10: RFE plot with Extra Tree classifier

It is visible that with 3 features the accuracy is about 82.3 percent for RF classifier. Most significant features are CALM,HAPPY, text-sentiment.

Feature importance plot of RFE for Extremely Random Trees-

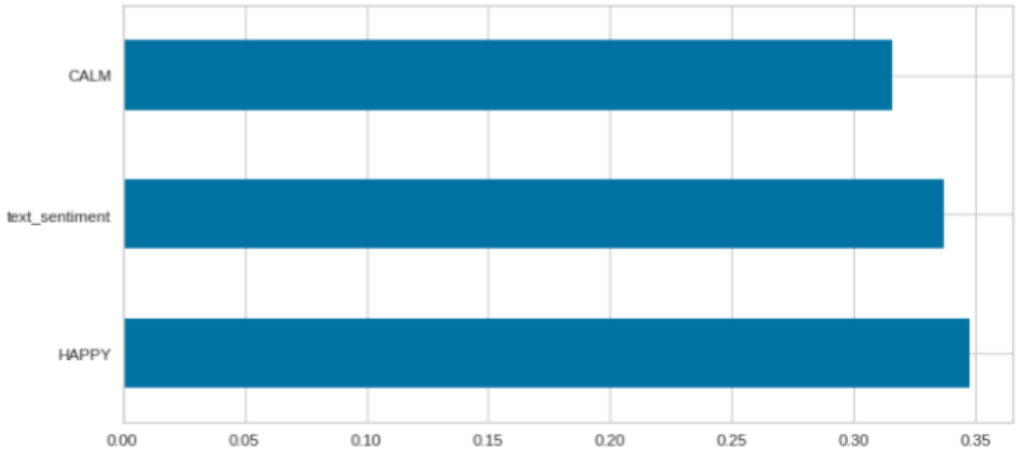

Figure 3.11: RFE plot with Extra Tree classifier

3.5.3 Correlation Matrix Analysis

Another possible method which can be used in order to reduce the number of features in our dataset is to inspect the correlation of our features with our labels.Using Pearson correlation our returned coefficient values will vary between -1 and 1:

- If the correlation between two features is 0 this means that changing any of these two features will not affect the other.

- If the correlation between two features is greater than 0 this means that increasing the values in one feature will make increase also the values in the other feature (the closer the correlation coefficient is to 1 and the stronger is going to be this bond between the two different features).
- If the correlation between two features is less than 0 this means that increasing the values in one feature will make decrease the values in the other feature (the closer the correlation coefficient is to -1 and the stronger is going to be this relationship between the two different features).

Correlation of all the features with our output variable is as follows-

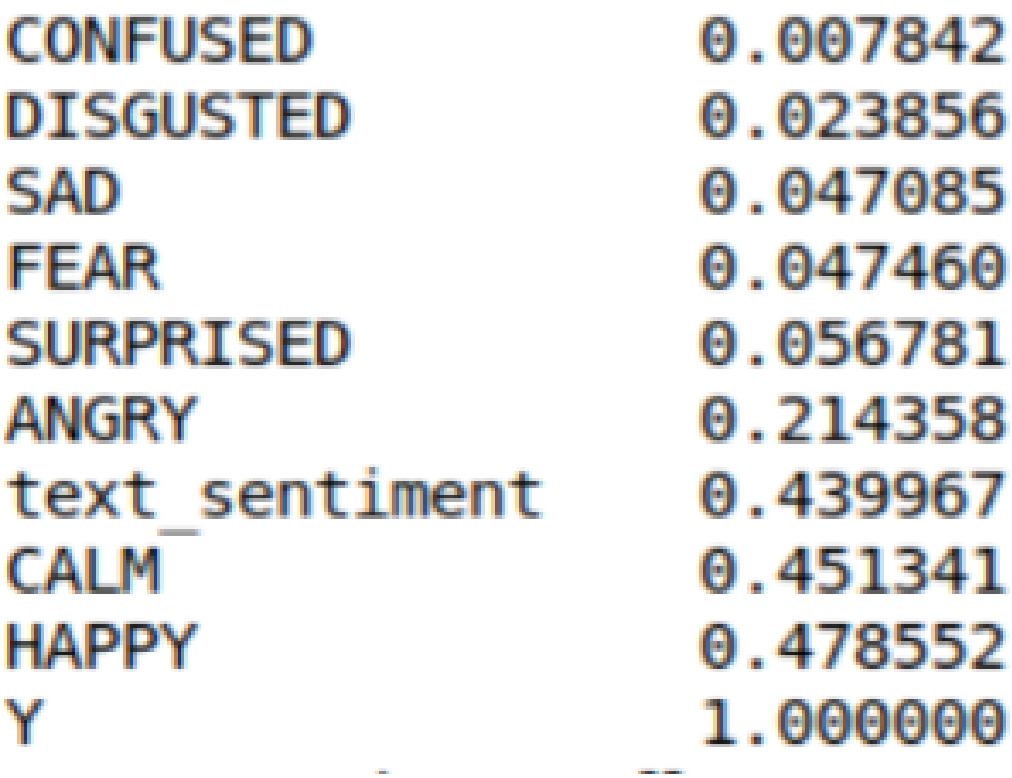

Figure 3.12: Correlation of all the features

The relationship between the different correlated features by creating a Correlation Matrix is shown below-

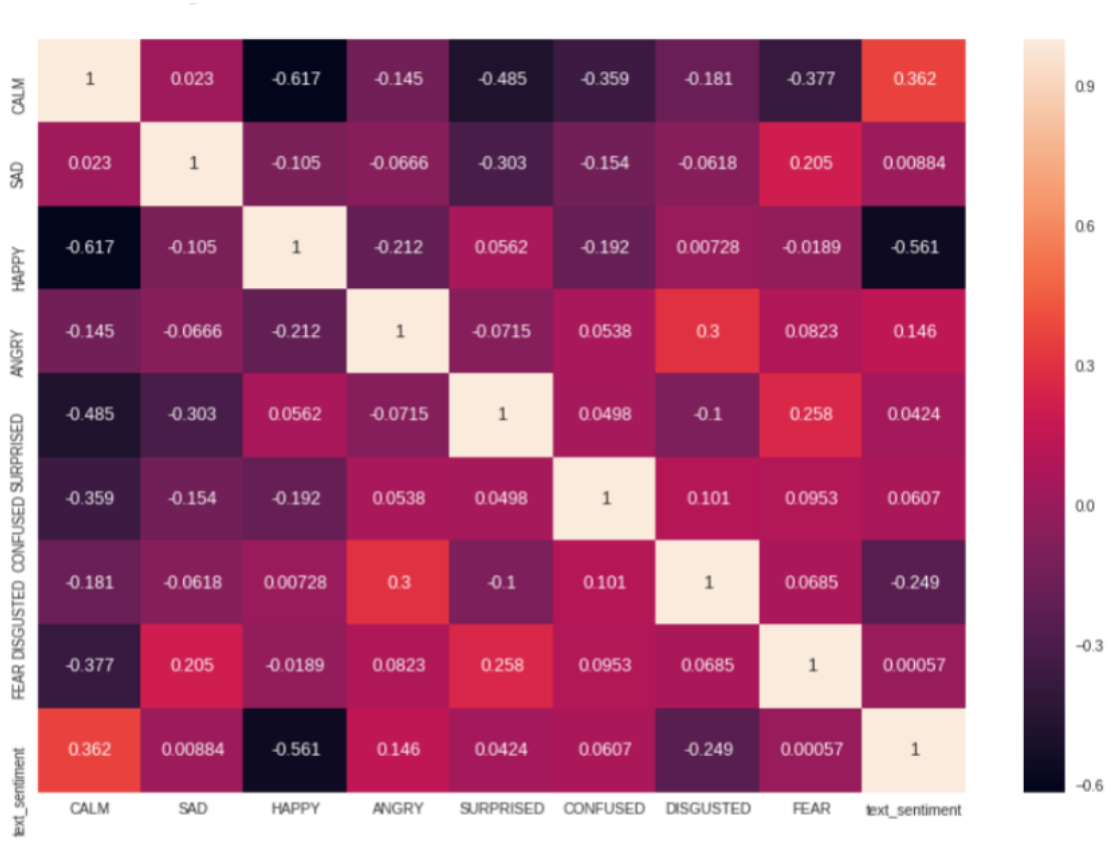

Figure 3.13: Correlation matrix

From this we can see that the most correlated features are 'HAPPY', CALM, text-sentiment, ANGRY

Chapter 4

Results

4.1 Results and Analysis

We applied XGBOOST algorithm to build our model and tested it using k-fold validation technique with 3 folds and 5 iterations. The results obtained using hyper parameter values such as `max_depth`=4, `learning_rate`=0.01, `n_estimators`=250, `min_child_weight`=1 are as follows:

Score from Iteration-1	0.8461538461538461
Score from Iteration-2	0.8399728163438401
Score from Iteration-3	0.8501528461528421
Score from Iteration-4	0.8421738431732411
Score from Iteration-5	0.6923076923076923
Average score	0.8153846153846154

Score from iteration-1	.846153
Score from iteration-2	.839972
Score from iteration-3	.850152846
Score from iteration-4	.8421784317
Score from iteration-5	.6923076923
Average score	.81538461

Standard Deviation of the scores obtained is 0.06153846153846154. These were the best results achieved after hyper parameter tuning by changing the parameter values randomly and this was the minimum deviation value that was obtained using the aforementioned parameter values which indicates that our model parameters were to some extent optimized. Other scores are :

- AUC - 0.75
- precision score - 0.75

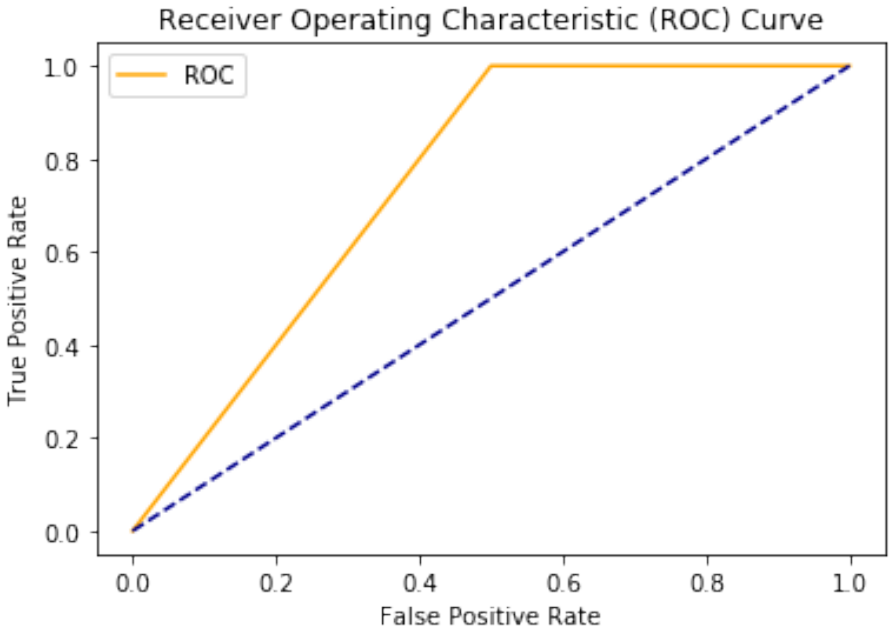

Figure 4.1: Receiver Operating Characteristic(ROC) Curve

- recall score - 0.75
- f1-score - 0.75

However, when using kfold validation for further validation, we got an increased percentage of 81 Percent. Which shows high accuracy.,
Our ROC curve for the model is as follows-

The ROC curve indicates that our model is moderately skillful. It can be further optimized using hyperparameter tuning and other methods of feature extraction and selection.

The confusion matrix for our model is-

It shows that our classifier was able to classify all the true negatives correctly although it was not very successful incase of true positives and false positives. This can also be further improved using other prediction models and increasing the size of dataset.

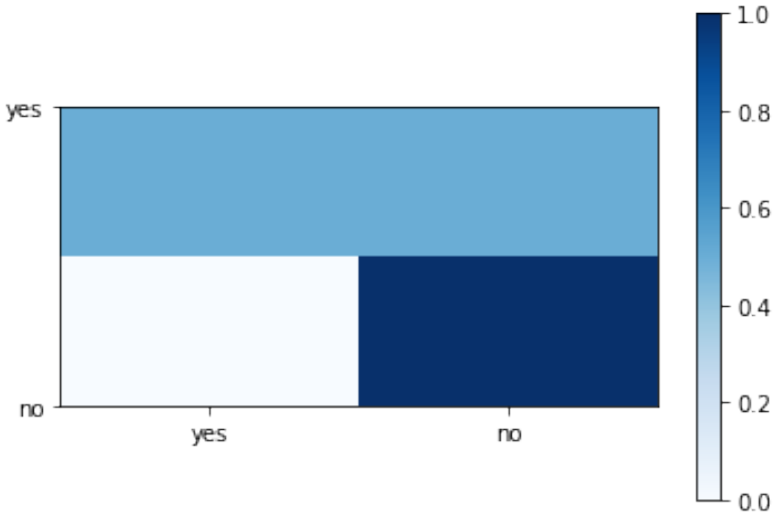

Figure 4.2: Confusion Matrix

Bibliography

[1] D. Osimo and F. Mureddu, "Research challenge on opinion mining and sentiment analysis", *Universite de Paris-Sud, Laboratoire LIMSI-CNRS, Bâtiment*, vol. 508, 2012.

[2] V. Shankar, A. Venkatesh, C. Hofacker, and P. Naik, "Mobile marketing in the retailing environment: Current insights and future research avenues", *Journal of interactive marketing*, vol. 24, no. 2, pp. 111–120, 2010.

[3] D. A. R.N. Devendra Kumar, "Facial expression recognition system "sentiment analysis"", *Journal of Advance Research in Dynamical and Control Systems*, 2017.

[4] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank", in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1631–1642.

[5] A. Kafi, M. Alam, S. Ashikul, S. B. Hossain, and S. B. Awal, "Feature based mobile phone rating using sentiment analysis and machine learning approaches", PhD thesis, 2018.

[6] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments", *Expert Systems with Applications*, vol. 40, no. 10, pp. 4241–4251, 2013.

[7] S. N. Manke and N. Shivale, "A review on: Opinion mining and sentiment analysis based on natural language processing", *International Journal of Computer Applications*, vol. 109, no. 4, 2015.

[8] M. S. Hossain and G. Muhammad, "Cloud-assisted speech and face recognition framework for health monitoring", *Mobile Networks and Applications*, vol. 20, no. 3, pp. 391–399, 2015.

[9] M. H. R. Pereira, F. L. C. Pádua, A. C. M. Pereira, F. Benevenuto, and D. H. Dalip, "Fusing audio, textual, and visual features for sentiment analysis of news videos", in *Tenth International AAAI Conference on Web and Social Media*, 2016.

[10] H. Hatem, Z. Beiji, and R. Majeed, "Human facial features detection and tracking in images and video", *Journal of Computational and Theoretical Nanoscience*, vol. 12, no. 11, pp. 4242–4249, 2015.

[11] S ChandraKala and C Sindhu, "Opinion mining and sentiment classification: A survey", *ICTACT journal on soft computing*, vol. 3, no. 1, pp. 420–425, 2012.

[12] L.-P. Morency, R. Mihalcea, and P. Doshi, "Towards multimodal sentiment analysis: Harvesting opinions from the web", in *Proceedings of the 13th international conference on multimodal interfaces*, ACM, 2011, pp. 169–176.

[13] V. Radhakrishnan, C. Joseph, and K Chandrasekaran, "Sentiment extraction from naturalistic video", *Procedia computer science*, vol. 143, pp. 626–634, 2018.

[14] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully automatic facial action recognition in spontaneous behavior", in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, IEEE, 2006, pp. 223–230.

[15] M. S. Akhtar, D. S. Chauhan, D. Ghosal, S. Poria, A. Ekbal, and P. Bhattacharyya, "Multi-task learning for multi-modal emotion recognition and sentiment analysis", *arXiv preprint arXiv:1905.05812*, 2019.

[16] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, "Meld: A multimodal multi-party dataset for emotion recognition in conversations", *arXiv preprint arXiv:1810.02508*, 2018.

[17] S.-B. Kim, K.-S. Han, H.-C. Rim, and S. H. Myaeng, "Some effective techniques for naive bayes text classification", *IEEE transactions on knowledge and data engineering*, vol. 18, no. 11, pp. 1457–1466, 2006.