A Multimodal Approach of Sentiment Analysis to Predict Customer Emotions Towards a Product

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 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 "A Multimodal Approach of Sentiment Analysis to Predict Customer Emotions Towards a Product" 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:Mostafijur Rahman Akhond

Mostafijur Rahman Akhond
Lecturer
Department of Computer Science and Engineering
BRAC University

Co-supervisor:Md.Golam Rabiul Alam

Md. Golam Rabiul Alam,PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:Mahbubul Alam Majumdar
(Chair)

Mahbubul Alam Majumdar,PhD
Chairperson
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

Begun roughly a decade ago, Sentiment analysis, the science of understanding human emotions, can trace its roots back to the middle of the 19th century. The motive of sentiment analysis is to extract and predict human emotions through facial expressions, speech or even text in some cases. Being inspired by the existing ideas, we propose a multi-modal model in the market that uses both facial cue and speech to forecast customers' sentiments and satisfaction towards a certain product. Our model helps various companies get key insights for specific market regions and their customers, and to gain a competitive advantage over the other. In this study, we estimate product perception of a demography based on emotions that were extracted from customers' facial expressions and speech. Although many researches have been made in the eld, but very few of them are multifaceted, integrated systems, where the different components rely on each other to produce an absolute result. We extract the emotions of people by recording their facial cues and speech patterns as they interact with a specific product of the market, such as a mobile phone in our case. We analyzed their facial expressions using AWS Rekognition. For the textual part, we analyze the sentiments using an algorithm which has a mixture of Tensorflow, Keras, Sequential model and RNN. Finally, we merge the previously obtained emotions from the video section with the textual sentiment to get the features for our predictive model. The model was generated using an algorithm named XGBoost. We have achieved an average accuracy of 81 percent approximately with 0.065 standard deviation by implementing cross validation of k-fold nature with folds of 3 and also 5 different iterations.

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

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 Md.Golam Rabiul Sir and Mostafijur Rahman 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

Chapter 1

Introduction

1.1 Overview

Sentiment Analysis, which is also more properly recognized as opinion mining , describes the use of processing of natural language, analyzing texts,computational linguistics, and also the use biometrics to extract, determine,measure, and research various states and subjective information. In essence, it could be defined as the process of determining the emotional tone behind a cascading flow of sentences or facial cues, which can be used for the understanding of the the hidden emotions, subtle opinions expressed through a certain medium.There are various types of sentiment analysis out there, which is used to extract the perceptions of certain people in certain fields. In fig-1.1 we can see the various classification of Sentiment Analysis [1]-

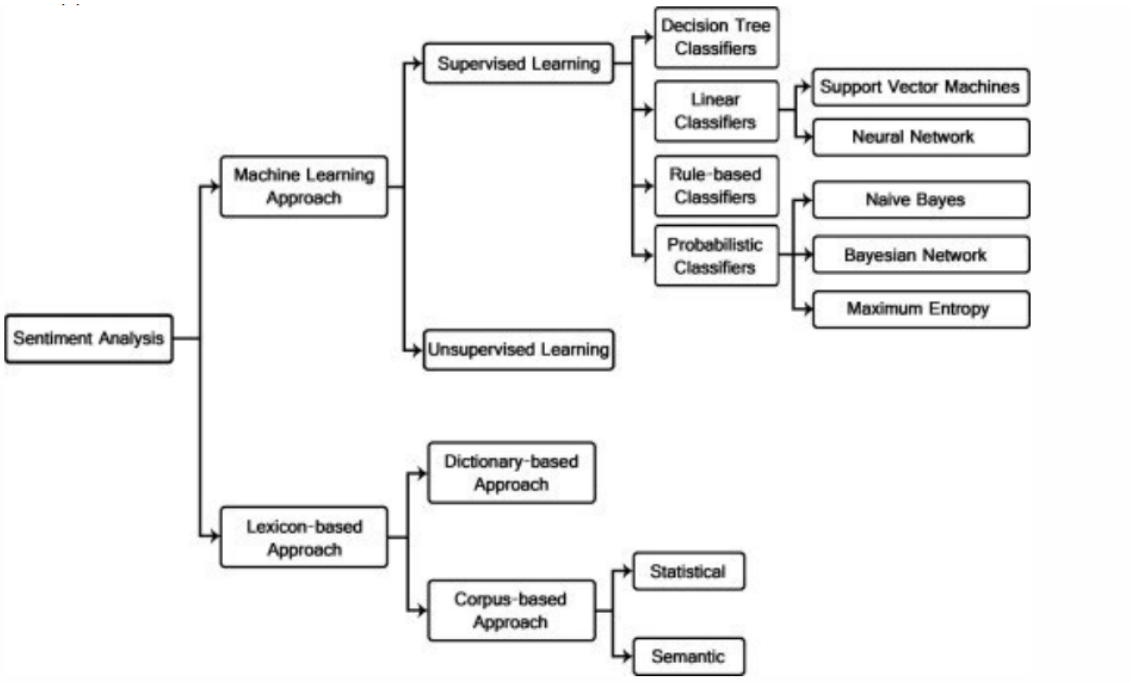


Figure 1.1: Types of sentiment analysis

It is a very common however a powerful categorization tool that dissects an incoming data which ranges from texts, audio, video and identifies if the hidden sentiment

is positive, negative or even neutral.Currently, it is a field which generates huge great interest among various people and also used for many sort of development since it has many powerful application which can be very useful in various situations.Through the recent findings in deep learning sector, the algorithms have gained the ability to analyze sentiments more effectively .Through using various advanced AI techniques and machine learning techniques in creative ways ,it can be an useful tool for doing in-depth researches . It has become a very important tool for understanding thousands and millions of varied conversations with human like accuracy level , the results of which can be implemented in various sectors.Moreover it helps data analysts understand public opinion, perform various research of market, observe brand and product reputation.The age of getting meaningful insights from sentiment analysis has now arrived with the advancement in technology.

1.2 Thoughts behind working in the marketing landscape

In this ever changing world, the need of detecting underlying emotions has become a necessity.By detecting motive behind a human decision which can range from purchasing a specific product, crime motivation, employee to employer relationship etc we can gain key advantages in various fields. Sentiment analysis is such a gateway where we can use the latest technological advancements to extract human emotions. Sentiment analysis, which is the process that is very much automated,which can be used for understanding an opinion from texts,spoken language, or even video feeds.It is a hot topic in the current technological and socio-economical landscape.

In this fast moving earth where about 2.7 quintillion bytes of data is produced every day[2], sentiment analysis has truly become a huge tool for making sense of that raw data. Through the help of such systems, unstructured informations which are gathered for research purposes, could automatically be turned into structured analysis of data that are taken from public opinions about products, services, brands, or any field in which consumers can express their opinions. This filtered and analyzed data can be very useful for marketing analysis, product reviews, product feedback. Customer sentiment can be found in tweets, comments, reviews, surveys or other places where people mention company brand. Sentiment Analysis tries to understand these emotions with application or software, and it is a must-understand for developers and business leaders in a modern business environment.Companies can understand what people think of a companys' own services or products it can also understand what they think about its' rivals as well. The total consumer experience of the demography can be understood very quickly with the help of sentiment analysis. So we can understand that the impact of sentiment analysis in business sectors is impossible to overlook. It can prove to be a major tool 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.It is a must for observing a companys' reputation across various medias. SA allows us to determine users' perception about a company and understand

what procedures we must take to enhance the brands' overall reputation.Sentiment analysis enables businesses to stay on top of their respective fields by focusing on feedback regarding topics such as their products, PR campaigns and other promotional efforts. It also lets businesses judge whether the product in question is an asset or a liability to the goodwill of the company.So we can see that through the use of sentiment analysis, incredible strives could be taken in terms of marketing and understanding the customer demography.

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. It is 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 if there are any areas of improvement that they need to look at for the next iterations of their products.

1.3 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. 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 is saying, customers' sarcasm, ambiguous statements, doubts from facial cues through a multimodal system can be revolutionary when it comes to market research.

1.4 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.

1.5 Research objectives

Our research objective covered various ground which ranged from market research, to customer perception and lastly to fill 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 are 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.

- The customer's mindset is key element when it comes to a successful design and creation of a product and it's subsequent campaign. Through the combination of the above mentioned techniques, we will be hoping to have an in depth insight to a customers' mind.

- We hope to answer some of the problem points of sentiment analysis such as lack of in-depth customer mind analysis rather than a classication of just happy and sad, a new way of extracting emotions from customers besides the previously established 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 and it's issues. This approach will ensure new ways of evolving the product in future iterations, keeping in account, the current mindset of the customers and their relevant reactions and/or feedback on certain aspects or the product as a whole.

- Information that the companies can use in a multitude of ways, from product globalization to localization based on certain demographic responses to the product, to rebranding according to appropriate response and many other use cases in between.

- Understanding the recent market trends directly through the customer's point of view, providing in return an edge over competitors in the same market.

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,[3] , 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.

To truly get a better knowledge, 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 [4] , 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 [5], 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 [6], they had proposed a model which had made significant leaps towards understanding compositionality in tasks such as detecting sentiments with better supervised methods.In here,the authors talked about a Sentiment Treebank which had finely tuned sentiment labels for 215,154 phrases in 11,855 sentences which resided as a parse tree. To address new compositionality challenges, the Recursive Neural Tensor Network was introduced. By training it on a treebank , it outperformed all previous iterations on various different metrics. It improved upon positive/negative classification from 80.03 percent to 85.2 percent. Moreover, this was the only system that could correctly depict the impact of negation . It also was capable of understanding scope at both positive and negative phrases in various tree levels.
Thus it captured the compositional effects with very high accuracy.[6]
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 [7], the author had proposed a system which can identify sentiments of particular features of a multitude of company's mobile sets and rating them as a whole. The author firstly processed the collected data in to a supervised structure. Then they selected the most common specifications from the training data. The system used Naïve Bayes, Support Vector Machine, Logistic Regression, Stochastic Gradient Descent and Random Forest algorithms. All of them were used for performance comparison. The system provided an average polarity of each specifications. It also provided an average polarity of the specific mobile phones based on which a rating would be given. Through this the customers can pick the best according to their desire[7].

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 [8], dealt with clearing "noisey web texts" which can cause significant problems in lexical level but also syntactic levels[8]. In this paper,a random collection of 3516 tweets were used to determine the customers' underlying emotions towards various mobile brands such as Nokia, T-mobile etc. The qualitative portion was conducted by using QDA Miner 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 measure the quantitative sentiment score[8]. 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 [9], focused on opinion mining which is domain-specific. They had built an architecture where in the first part , they collected information from multiple e-commerce websites such as flipcart, snapdeal,amazon etc[9]. The reviews from the customers then were used for feature extraction. After doing this, the reviews were taken through the trainer classifier to find the patterns of emotions inside the text based comments.This pattern can be recognized with techniques which included N-Gram Extraction and part of speech Extraction.From here the collected comments are filtered by removing the irrelevant comments and clarity score is calculated. Through the use of this system, it can give the feedback for a specific product. The emotions ranged from positive , negative and neutral. They hoped that through this system ,it might provide an better solution in opinion acquisition problem. 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 [10], 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 hospitals, healthcare professionals for necessary services in order to give an automated health monitoring system. This paper demonstrated how effective the proposed system could be by taking into account facial cues and speech. 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 [11]. 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 in-

sights to track facial features to detect emotions. In the paper [12], 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 [13], they explored, analyzed various techniques, recent advancements and also future directions in the sector of SA(sentiment analysis). One of the problems it presented were 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 [14], 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 emotion detection . From 47 videos that depicted a certain part , they chose 30-seconds cuts which covered a specific topic. It also transcribed the videos in a manual way[14]. 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[14]. For language, word polarity was considered.Through the use of a software called openEAR speech pauses and tonal pitch were extracted. With these extracted features , Hidden Markov models were implemented for sentiment categorization.The system took this multi modal features as input. This multi modal system exceeded the performance of unimodal systems.

The article [15], 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 to handle requests for video analyzation, 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 [17], 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 [18], the authors have proposed a system called the Multimodal EmotionLines Dataset (MELD) that is an evolved version of EmotionLines[18]. The system contained about 13,000 excerpts from 1433 speeches from the television series named "Friends". Here, every utterance is connected with emotion and sentiment labels.It included audio, visual and textual inputs. They have used seven emotions for the annotation that include joy,anger,disgust etc. 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 features. 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 a process of understanding a function that uses an input to map an output based on various examples. It is a generalization of the concept of knowing from example where a learner is given two different sets of data which includes a test set and a training set. It learns from a set of categorized iterations in the training domain. From there it can recognize unlabeled/uncategorized data in the set, which were used for testing, which has the best accuracy. The goal of the system is to create a guideline, an application that categorize new entries inside the testing domain by analyzing the entries it was previously selected by the programmer. In here, the set that was used for training included n ordered pairs (x1, y1), (x2, y2),(x3, y3)..., (xn, yn), where each xn is a data of measurements for a single entry. y_n is the label used for that entry [7]. The data used for training in this method is another set of k measurement without labels: (xn+1, xn+2,+x3 ..., xn+k). The target here is to make educated predictions that follows a certain trend or path.

Usually, this method works with three particular section:

- Binary categorization: In binary categorization, the algorithm classifies the data in two distinct categories.
- Non-binary categorization:In this classification the algorithm tries to select between more than two versions of classification for a target variable.
- Regression: In this model, it detects continuous data points, 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 results.This 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-

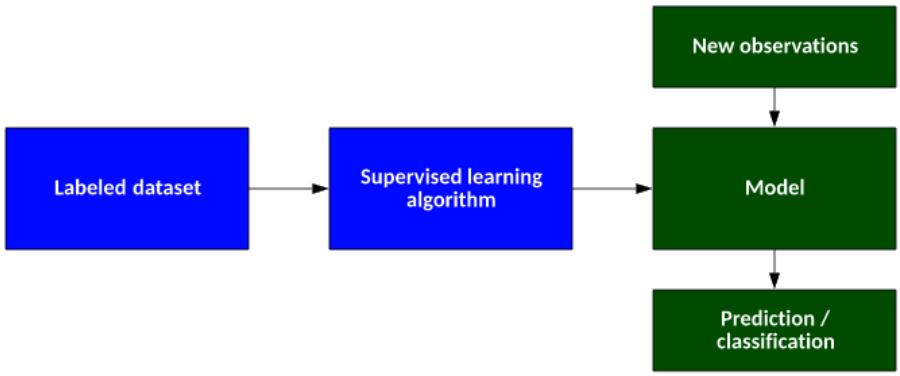


Figure 2.1: Supervised Learning Environment

Supervised learning can be considered as the more business focused section of MachineLearning. This is not as independent as unsupervised learning.Moreover It has a more practical value than other learning such as reinforcement learning, as it only excels in closed game-like system architecture.

2.3 Market Research

Companies which produce and commercialize products can understand more about their current consumer base and target consumers and the reception of their products through researching the market. They can also use marketing research to learn about their company's brands,reputation.Market research through the process of collecting, dissecting, and understanding the underlying information about a market, product or service could give then organization an edge over their peers.
One of the first steps in planning phase of a business is the research of market. Companies use this to tackle various challenges that includes segmentation,product requirements,design.The need for creating a distinct identity for a product is a must for any company. Market research helps companies collect, record present and past data on the performences and public opinions of their current products which can help them build future products as well.

2.4 Amazon Rekognition

AWS Rekognition allows programmers work with Web Services which are provided by amazon to include image and video analysis in their systems. Through the use of AWS Rekognition, various application or systems can identify, record and determine facial cues, still objects in images and videos. Moreover with this we can compare various facial feature.[19].
With the API provided by Rekognition, a high level of advancement and sophistication can be included in our system, model or application with advanced image

categorization and visual object search. This API is based on deep learning. It is a subset of in the machine learning domain. Its main aim is mapping higher level abstractions through the use of various nonlinear neural networks transformation and it's architecture.
The uses of this API include-

- Searchable images library:This API makes not only videos but also images searchable, which enables programmers and creators to discover objects or faces that appear within them.
- User verification through faces: We can discover the user identities by comparison of their images recorded in live feed with their previously stored images that were used for references.
- Opinion detection:We can identify emotions such as happiness, anger and shock from various images of faces. A programmer can conduct analysis in real time of various images and store the values of various emotional aspects.
- Facial recognition: We can record and keep a track of facial images as reference data and allow the search of facial images with the API functions.

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[19]-

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

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, Random Forrest,Naive Bayes, Logistic Regression, Linear Discriminant Analysis, K-nearest Neighbors

2.5.1 XGBoost

XGBoost is based on decision-tree. It is an ensemble Machine Learning algorithm which uses a framework which is based upon gradient boosting. In many times, it could be acceptable to rely on the results of a single model of machine learning. However ensemble learning provides a solution which is systematic enough to integrate the predictive capacity of a multitude of learners. The result is a singular all encompassing model that produces the average output of several different modular system.
The models that come together to form the ensemble, are mostly known as base learners. This could be either be taken from learning algorithm residing in the same domain or by using separate learning algorithms.This is mostly used in decision trees.[20].

In the method known as boosting, the decision trees are created in a sequential manner.This is done in such a way that for the next iterations of each subsequent trees,it reduces the errors that were generated from previous trees.The procedure works like this that for each various version of trees, they learn from their previous versions. Next it updates the residual errors. So, the tree which is created next will learn from an updated iteration of the error[20].
In boosting the base learners are preliminary learners. In here the bias is extremely high.However, every one of these learners that are preliminary adds some very important information for future trend prediction.It allows this technique to create a strong learner by joining various weak learners. The final stronger learner brings down both the bias and also the variance[20].

If compared to bagging methods such as Random Forest(where trees goes to their maximum depth) boosting uses trees with fewer splits. These small no in-depth trees can easily be interpreted. Various parameters which includes gradient boosting learning rate,the number of iterations, and the depths of those iterations.They can be chosen in an optimal way through validation methods such as k-fold cross validation. In the end,it is a must task to systematically choose the criteria which signals termination for boosting[20].
Boosting includes othe following steps[20]:

- The first iteration of the F0 model is defined to determine the y target variable. This model would have a residual which is (y – F0)
- h1, which is a new model, is then fitted to the residuals generated from the previous step.
- After that, h1 along with F0 are integrated together to give F1.This is the boosted updated iteration of F0. Moreover,the MSE(mean squared error) from the new iteration would be much lower than the previous iteration:

$$F_1(x) < -F_0(x) + h_1(x) \tag{2.1}$$

To increase the performance of F1, we could map after F1 residuals .From there a new model called F2 can be generated:

$$F_2(x) < -F_1(x) + h_2(x) \tag{2.2}$$

This can go on for 'k' iterations.This will go on till residuals tottaly have been minimized :

$$F_m(x) < -F_(m-1)(x) + h_m(x) \tag{2.3}$$

Here, the subsequent learners do not disrupt the functions generated from the previously found phases. Instead, they generate information of their own to weigh down error percentage.
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[21]-
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:

$\mathrm{F}_0(x) = argmin \sum_{i=1}^{n} L(y_i, \gamma)$.
(2.4)

2. m=1 to M:

   (a) Measure pseudo-residuals:

$$r_{im} = -\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}_{F(x)=F_{m-1}(x)} \tag{2.5}$$

   For i= 1,...,n

   (b) Fit a preliminary Learner(such as a tree) $h_m(x)$ to pseudo-residuals and train it through the training set $(x_i, r_{im})_{i=1}^n$

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

$$\gamma_m = argmin \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)). \tag{2.6}$$

   (d) Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{2.7}$$

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[22],

$$\frac{\partial L(y, f^{m-1}(x) + f_m(x))}{\partial f_m(x)} = 0 \tag{2.8}$$

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-

$$L(y, f^{m-1}(x)) + f_m(x)) \approx L(y, f^{m-1}(x)) + g_m(x) f_m(x) + \frac{1}{2} h_m(x) f_m(x)^2 \tag{2.9}$$

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:

$$\mathrm{h}_m(x) = \frac{\partial^2 L(Y, f(x))}{\partial f(x)^2}_{f(x)=f^{(m-1)}(x)} \tag{2.10}$$

Then the loss function can be rewritten as :

$$L(f_m) \approx \sum_{i=1}^{n} [g_m(x_i) f_m(x_i) + \frac{1}{2} h_m(x_i) f_m(x_i)^2] + const \propto \sum_{j=1}^{t_m} \sum_{i \in R_{jm}} [g_m(x_i) w_{jm} + \frac{1}{2} h_m(x_i) w_{jm}^2] \tag{2.11}$$

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=

$$L(f_m) \propto \sum_{j=1}^{t_m} [G_{jm} w_{jm} + \frac{1}{2} H_{jm} w_{jm}^2] \tag{2.12}$$

Now, with the fixed already trained structure, for each region, it is straightforward to determine the optimal weight :

$$w_{jm} = -\frac{G_{jm}}{H_{jm}}, j = 1, ..., T_m \tag{2.13}$$

Putting this back to the loss function, we get ,

$$L(f_m) \propto -\frac{1}{2} \sum_{j=1}^{T_m} \frac{G_{jm}^2}{H_{jm}} \tag{2.14}$$

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-

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

(2.15)

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

$$\mathrm{L}(\mathrm{f}_m) \propto \sum_{j=1}^{t_m}[G_{jm}w_{jm} + \frac{1}{2}H_{jm}w_{jm}^2] + \gamma T_m + \frac{1}{2}\lambda\sum_{j=1}^{t_m} w_{jm}^2 + \alpha\sum_{j=1}^{t_m}|w_{jm}| = \sum_{j=1}^{t_m}[G_{jm}w_{jm} + \frac{1}{2}(H_{jm}+\lambda)w_{jm}^2] + \alpha|w_{jm}| + \gamma T_m$$
(2.16)

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

$$\mathrm{W}_{jm} = \begin{cases} -\frac{G_{jm}+\alpha}{H_{jm}+\lambda} & G_{jm} < -\alpha \\ -\frac{G_{jm}-\alpha}{H_{jm}+\lambda} & G_{jm} > \alpha \\ 0 & else \end{cases}$$
(2.17)

The gain of each split is defined correspondingly:

$$Gain = \frac{1}{2}[\frac{T_\alpha(G_{jmL})^2}{H_{jmL}+\lambda} + \frac{T_\alpha(G_{jmR})^2}{H_{jmR}+\lambda} - \frac{T_\alpha(G_{jm})^2}{H_{jm}+\lambda}] - \gamma \tag{2.18}$$

$$\mathrm{T}_\alpha(G) = \begin{cases} G+\alpha & G_{jm} < -\alpha \\ G-\alpha & G_{jm} > \alpha \\ 0 & else \end{cases}$$
(2.19)

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)

SVM which is also known as Support vector machine is very popular machine learning algorithm.It is supervised. This is popular for regression and categorization. Though, it's main task is classifying test.In this method every entry or data can be referred as a plot point in m -dimensional space. Here, the value of each feature in the domain are the plots. m is the feature number in the set. After that, categorization is done. It identifies the specific hyper-plane to distinguish two classes accurately. Below in figure-2.2 it shows the classifications of two classes-

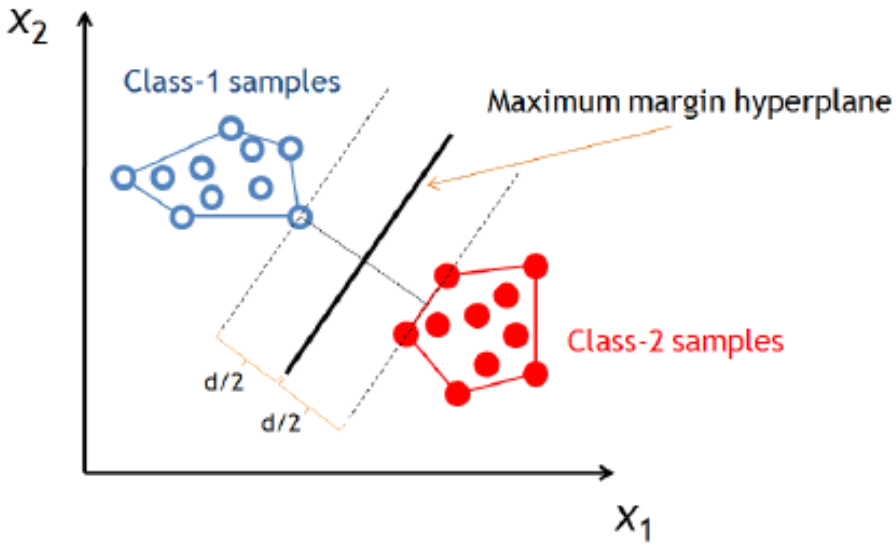


Figure 2.2: SVM Classifications of two classes

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-2.3 below-

Figure 2.3: SVM Classifications using hyper-line

The distance which exists in between the nearest point of data from any set and the line can also be called as the margin. The main target is to select a hyperplane with the highest length of margin between itself and any data point within the training domain.Because of this, it has a higher chance of categorizing new data accurately.To construct a hyper-plane line SVM uses an sequential training algorithm.It can be used to lessen the error percentage.

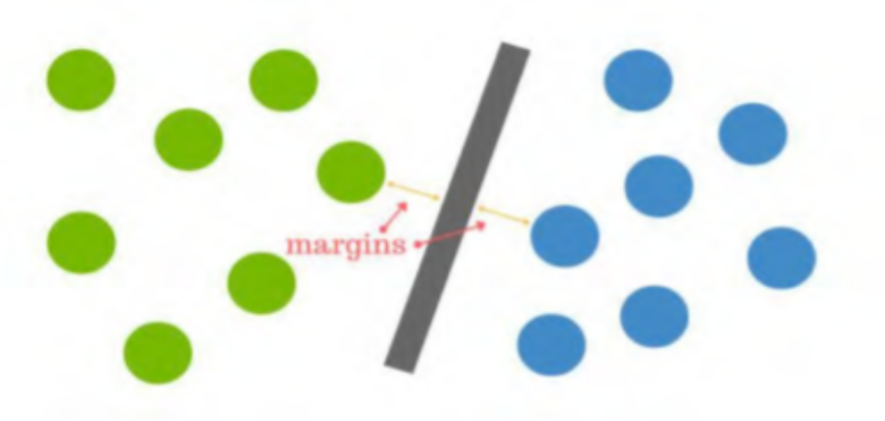


Figure 2.4: Best hyper-plane

From the above figure-2.4, we get an example of a classifier which is linear. However,not all categarization methods are this simple.It is often very complex .Moreover the complex build are imperative in order to create a separation which is optimal. In real world data is rarely as clean as the figure. Datasets will be mixed as the below figure-2.5:

Figure 2.5: SVM Classifications with jumbled data

To categorize the above jumbled dataset, it is a must to get away from 2D view of the dataset.We need to move forward to 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 figure 2.6 below-

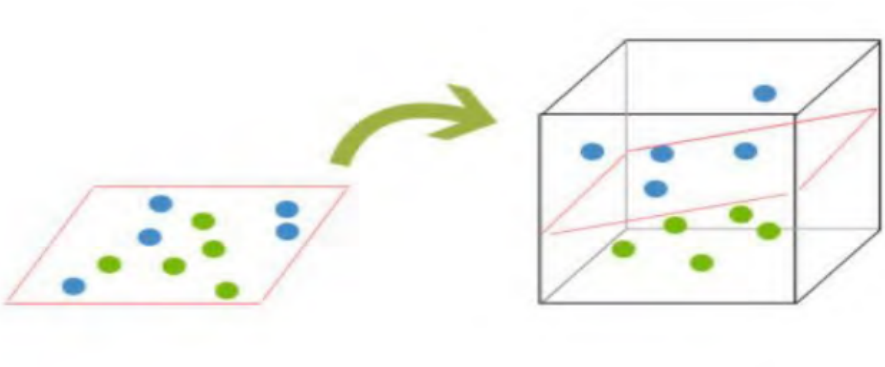


Figure 2.6: SVM Categorization in 3D dimension

In our system, we had used Scikit learn library to implement this algorithm .

2.5.3 Random Forest

This algorithm is also known as random decision forests. This be used to establish predictive systems for regression and categorization challenges.It is a very well known method of ensemble category. It is a simple to use,adaptive ML algorithm which produces a great result most of the time.This is possible even without the neccessity of tuning hyper-parameter.This is also one of the most widely used algorithms for its simple structure[23].

This is an algorithm which can learn in a supervised manner. Just like the way it is called, it generates a forest which is random. The addition of all Decision Trees makes this forest. This is mostly trained with the help of the method known as bagging.Main concept behind this method revolves around the fact that a finely designed integration of various learning models improves the overall result. In simpler terms, this creates a multitude of decision tress.After that they are mixed together to get a more refined results that are more stable. The main decisions that are made when building a random tree are-

1. In which method the leafs are splitted with.
2. The type of determinating factor used as predictor for each leaf.
3. The ways we can inject randomness into the forest through the trees.

. Each tree can be grown in the following way:

1. When the number of different cases inside the training domain is M, then we will sample M cases in a random way from the initial domain of data. The selected sample would be chosen as the set which will be used for training for tree building.
2. If when we have n input variables, a number m¡¡n is selected in such a way that at each respective nodal points we would have n variables that are chosen in a random way through the n.Then the split which is the best out of all the points in the set is used for splitting the nodal point. The value of n is always kept as a constant value during the building of the forest.

3. After this every tree available is grown to the largest depth possible without any pruning.

Increasing the correlation found between various trees makes the forest error percentage go up. The forest can be improved with the strength of individual trees which reside in it. A tree can be termed as a good classifier if it has a low error rate.So the forest error rate goes down as we increase the individual tree strength. As soon as the forest has been trained enough, this can be implemented to make various educated guess for new unclassified data entries. To do an educated guess for a specific point x, each tree uses the below equation -

$$f_n^j(x) = \frac{1}{N^e(A_n(x))} \sum_{Y_i \in A_n(x) I_i = e} Y_i \tag{2.20}$$

Here An(x) refers to the leaf containing x. $N^e(A_n(x))$ denotes the number of estimation data point it has.
The guess which are made by each individual tree relies on the estimation that is made by it. However, as measurable points are given to the structure, structure values in one singular tree might have the chance to add value to the prediction as estimation points of another different 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

This classification algorithm implements the theorem of Bayes' to categorize the various incoming data.It is called 'naïve' for the reason that it thinks that all the attributes related to a singular data entry are completely independent from others. This classification method implements theory of probability to categorize objects. The main particular concept behind this is that the probable nature of an event occurring might be tuned as new entries are introduced along the way. This model is very simple to build .Moreover,it contains no complex cascading estimation parameters, thus making it very powerful for very big data domains [24]. As mentioned earlier Bayes theorem gives a path 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 -

$$p(c|x) = \frac{p(x|c)p(c)}{P(x)} \tag{2.21}$$

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 distinct variations of Naïve Bayes model which we can see in figure 2.7-

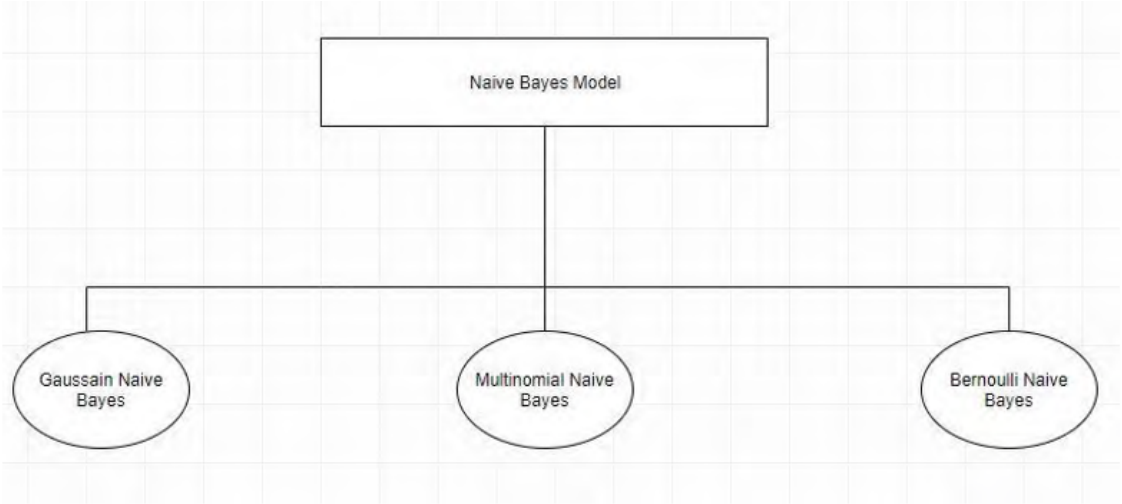


Figure 2.7: different types of Naïve Bayes model

In our research Paper,We use Gaussian naive Bayes.

Every time when handling continuous data, such as ours, a normal educated guess is made that the continuous values correlated with each distinct class are scattered in accordance with a Gaussian distribution.For example, let's say that the dataset used for training includes an attribute which is continuous,x.The first task is to classify the various points data by the class. After doing that we then proceed measure the variance and mean of x . Let μ_k be termed as the mean found for the multitude of data points which are measurable in x.This is connected with the C_k class.After that we let σ_k^2 be the Bessel corrected variance in x.Let's say that after collecting observation value v we then proceed to find the probability distribution of v given a class C_k through the equation given below ,

$$\text{p(x=v}|\ C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\, e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}} \quad (2.22)$$

2.5.5 Logistic Regression

Regression analysis can be termed as the process of using records of data or observations to map and give measurable value to the connection between a target variable that is is also known as a variable which is dependent and a domain of independent variables. This is an important tool for analyzing and mapping data purposes.We can find various types of regression methods available to make predictive decisions about various things. These techniques are mostly driven by three parameters. Parameters are given figures below in figure 2.8-

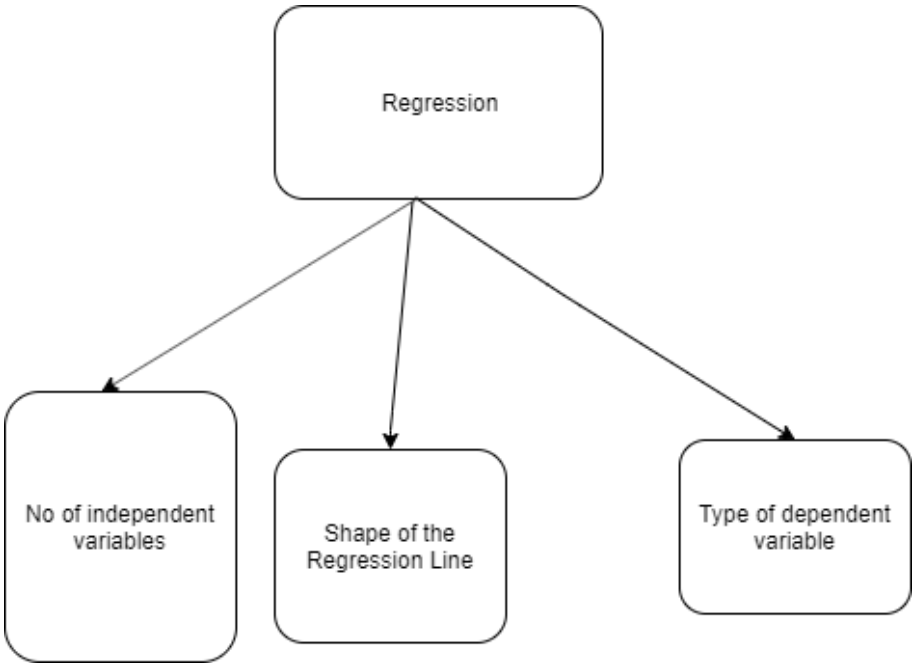


Figure 2.8: Parameters of Regression Analysis

Logistic regression is a type of regression analysis.It is mostly used whenever the dependent variable is of the binary kind. Just like every other regression analysis, this method is also a predictive model. This is implemented to give meaning to each data and describe the connection that exists between one dependent variable which is binary with one or more ratio-level,nominal variables which are independent.This model tries to quantify the connection between categorical dependent variable and one or more variables that are independent by plotting the dependent variables probability scores on a graph. The mathematical equation of this is given below-

$$\log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m \tag{2.23}$$

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

2.5.6 Linear Discriminant Analysis

Linear Discriminant Analysis or LDA is a popular method which is used for reducing dimensionality in the pre-processing phase for trend-categorization and various machine learning applications.
The main target is to deliver a dataset onto a space which has lower dimensionality along with a better separability capability of the classes.Through this it can avoid overfitting.Not only that but also it can deduct computational costs. It is like PCA ,but it focuses on maximizing the seperatibility among known categories.Even beside having the capacity of determining the component axes which can be used to increase the variance of the data,it can also be focused in the axes that maximizes the margin between a multitude of classes . This Approach can explained in 5 steps.

1. At first we measure the m-dimensional mean vectors for the distinctive classes from the domain which includes the data.

2. After that we compute what we know as scatter matrices.

3. Then we determine the eigenvectors(e_1,e_2,e_3,....e_m) and also their assigned eigenvalues(λ_1,λ_2,λ_3,..λ_m) for the previously found scatter matrices.

4. We then sort them by decreasing the eigenvalues and choose n eigenvectors with the highest eigenvalues to generate a $m \times n$ dimensional W.In here each and every column signifies an eigenvector.

5. We then make use of this $m \times n$ matrices to change the sample data onto the new subspace. It can be explained through the matix multiplication $Y = X \times W$ where X is a $k \times m$-dimensional matrix representing the k samples, and y are the transformed $k \times v$ -dimensional samples found 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[25]. 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 consists 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 data into train data and test section.
4. Data Pre-processing to clean and prepare it.
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 methods on the review data, with individual features to produce polar results 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 in Figure 3.1-

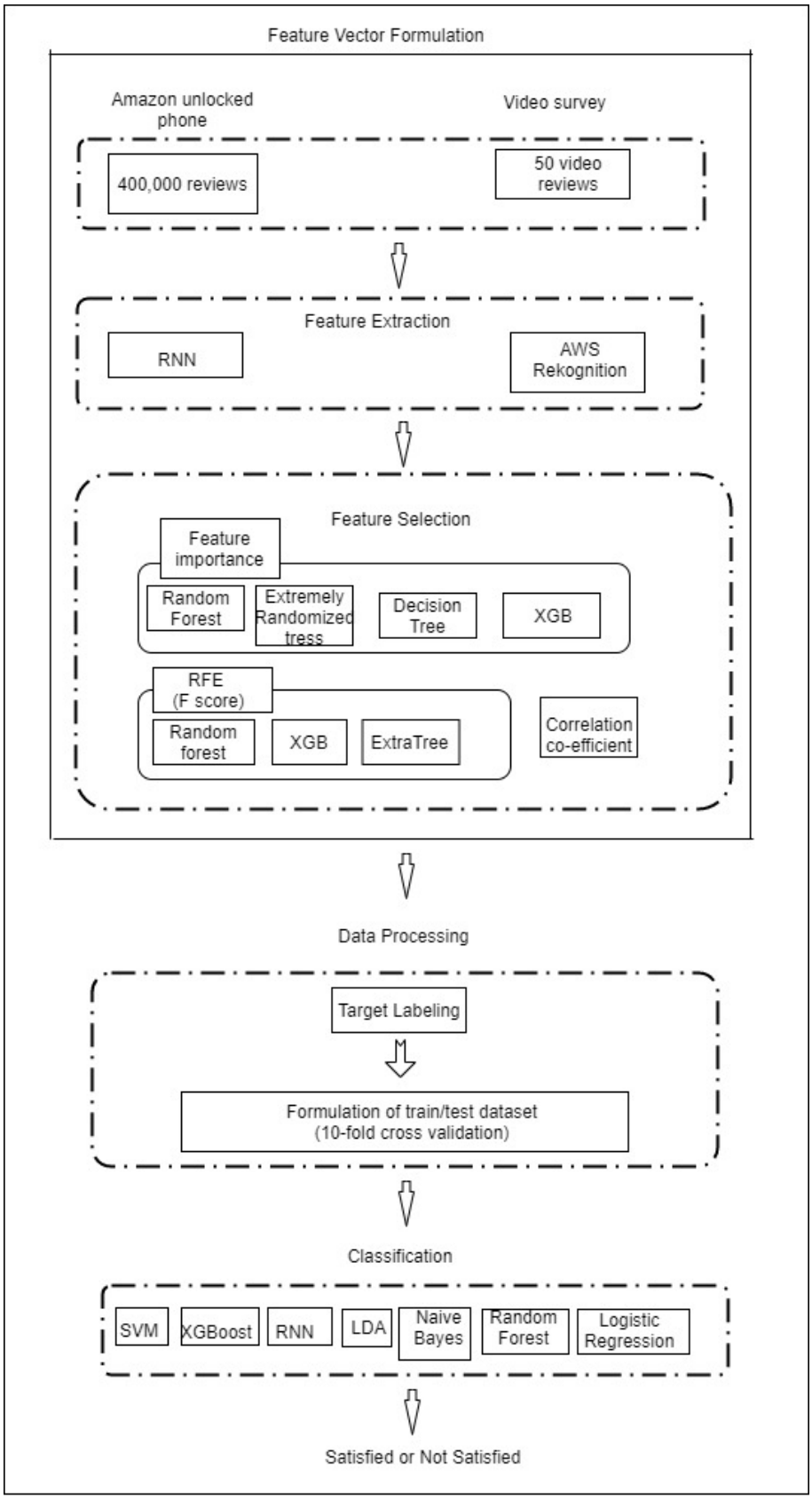


Figure 3.1: Work flow of the proposed system

3.2 Data Collection

In the domain of sentiment analysis, the most technically demanding and one of the most complex task is to find the proper data needed for the system to work.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.

We also had trained our model for textual sentiment analysis using data gained from the review of amazon's unlocked mobile phones. Roughly four hundred thousand data were found available in the dataset, which contained six columns in total. They

were, ‘product name’, ‘company name’, ‘review’,‘rating’, ‘rating votes’. The model name of the mobile phone was provided in the “Product name” column and the name of the phone company was provided in “Company name”. There were also product reviews given by the users which were found in the column “review”. These three columns contained textual data whereas the rest of the columns consisted of numerical data. For example, the user ratings were found in the column “Ratings”. 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 in which the dataset is split into training and testing data. This process is very useful and also needed for machine learning as the primary concept of machine learning depends on training and testing data and finding how accurate the output provided by the machine is.
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

Preprocessing of data is the process in which the data acquired is cleaned up in order for the machine learning algorithm to understand the overall content of the data. It operates by finding any incomplete or irrelevant and in some cases, inaccurate data, and removing them.

For the videos, Frame to Frame analysis is done using “AWS Rekognition API”. Image and video analysis in our applications were made easy with the help of Amazon Rekognition. The Rekognition API was provided with frame to frame images of the video by us and it automatically identified the people,scenes, objects, activities and also detected any inappropriate content. We used the Amazon Rekognition API to create static emotions of a person as the API is capable of providing highly accurate

facial recognition and reliable facial analysis from the images and videos that we provided it with. The statistic provided a percentage of 9 emotions, they are as follows:

- 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. This is a process which divides large strings of text into more compact strings, or "tokens".Longer portions of texts can be tokenized into small sentences and sentences 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. The feature importance of each feature of the dataset can be acquired by utilizing the feature importance property of the system. Feature importance yields a score in terms of importance or relevance for each feature of the dataset. The higher the score, the more important or relevant is the feature regarding the output variable. To systemize the importance of the different features, we can use the Decision tree models that are based of ensembles. One crucial aspect of understanding as to how our model is making the predictions (therefore making is more explicable) is to have the knowledge of the feature that is giving the most importance. It also helps to get rid of any features that do not interest our model (or confuse it to make a wrong decision).Below in Figure-3.2 is a Feature importance plot for Random Forrest:

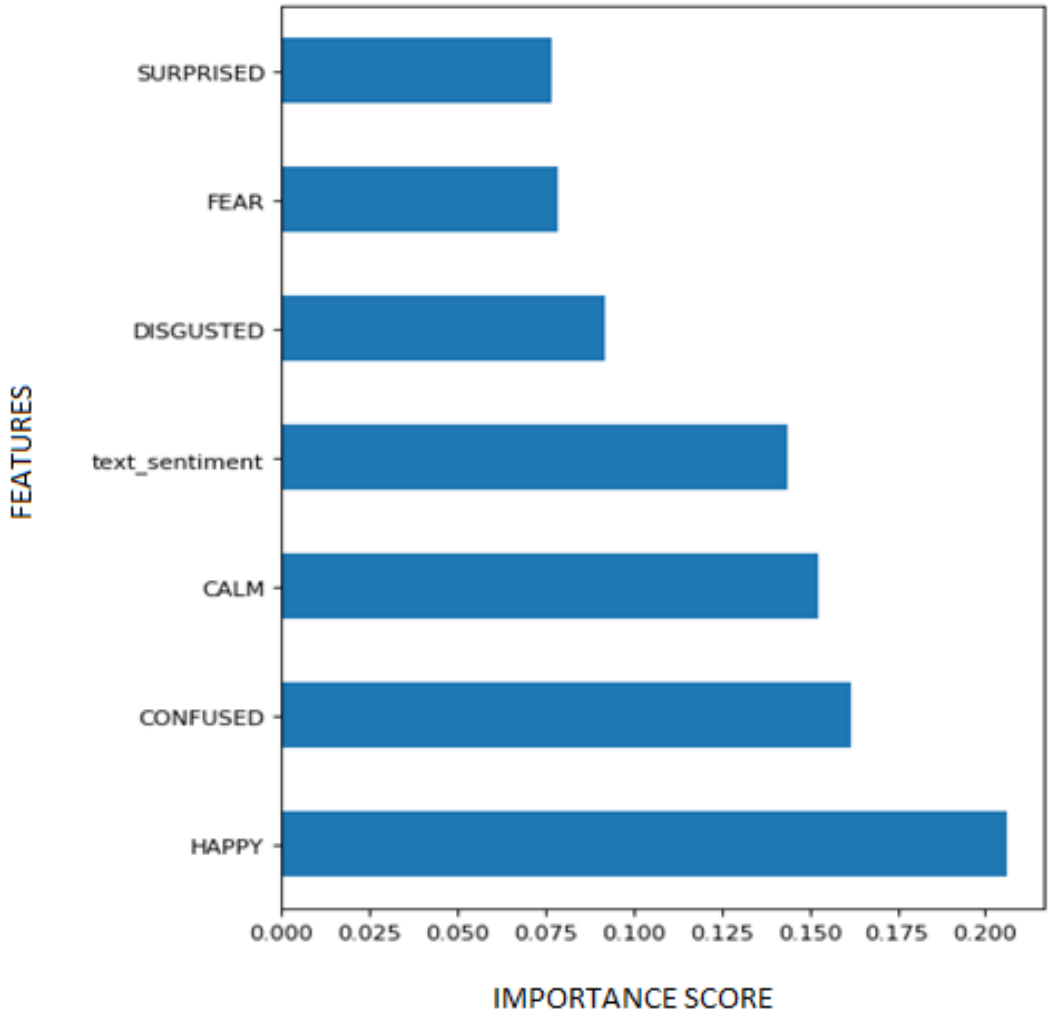


Figure 3.2: Feature importance plot for Random Forrest

Fig3.2 shows that Random forest model gives the most importance to the 4 features which are text-sentiment, CALM, CONFUSED, HAPPY. So these are the most contributing features according to our Random Forest model.

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

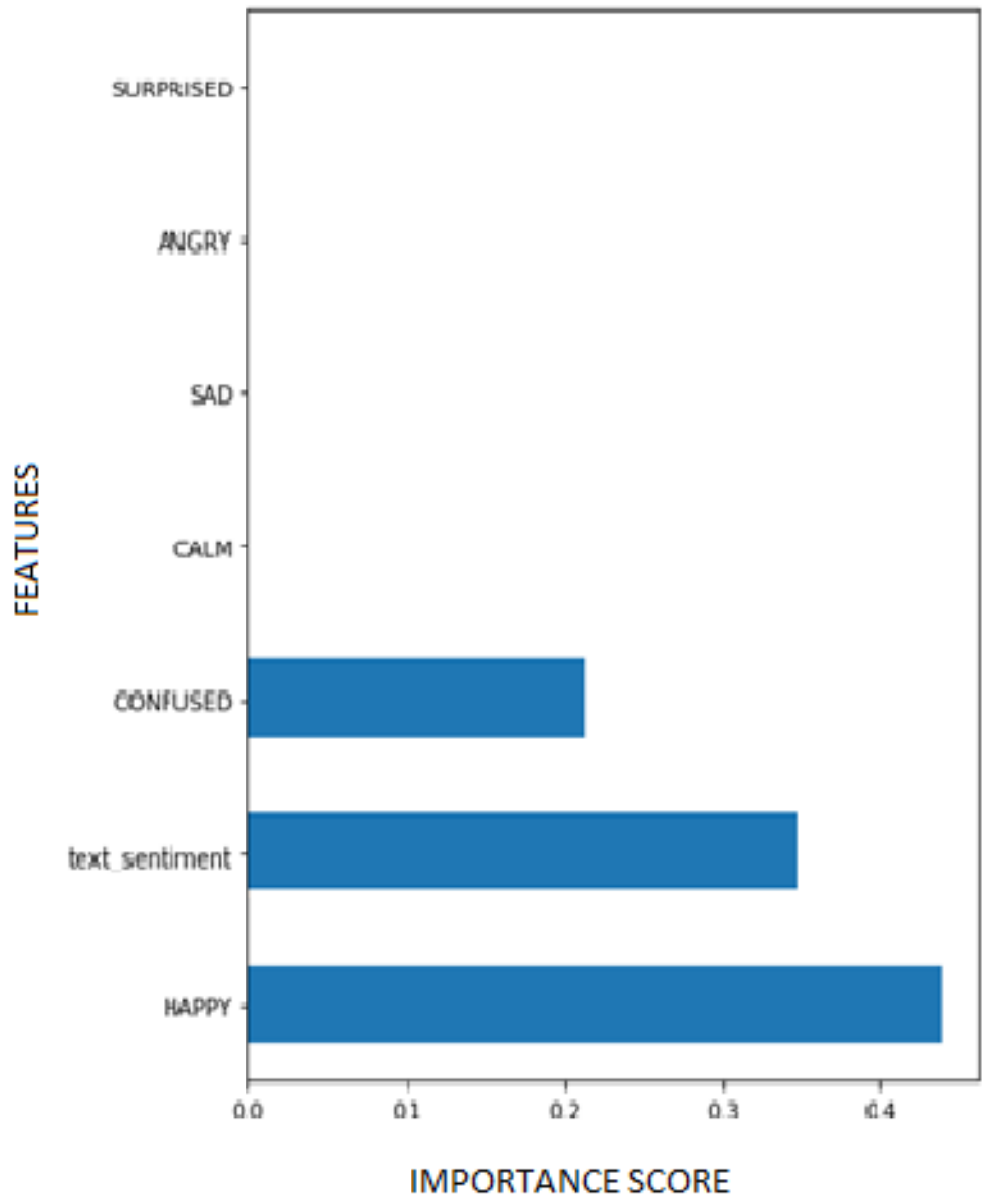


Figure 3.3: Feature importance plot from XGBoost

Fig3.3 shows that XGBoost model gives the most importance to the 4 features which are text-sentiment, CONFUSED, HAPPY. So these are the most contributing features according to our XGBoost model.

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

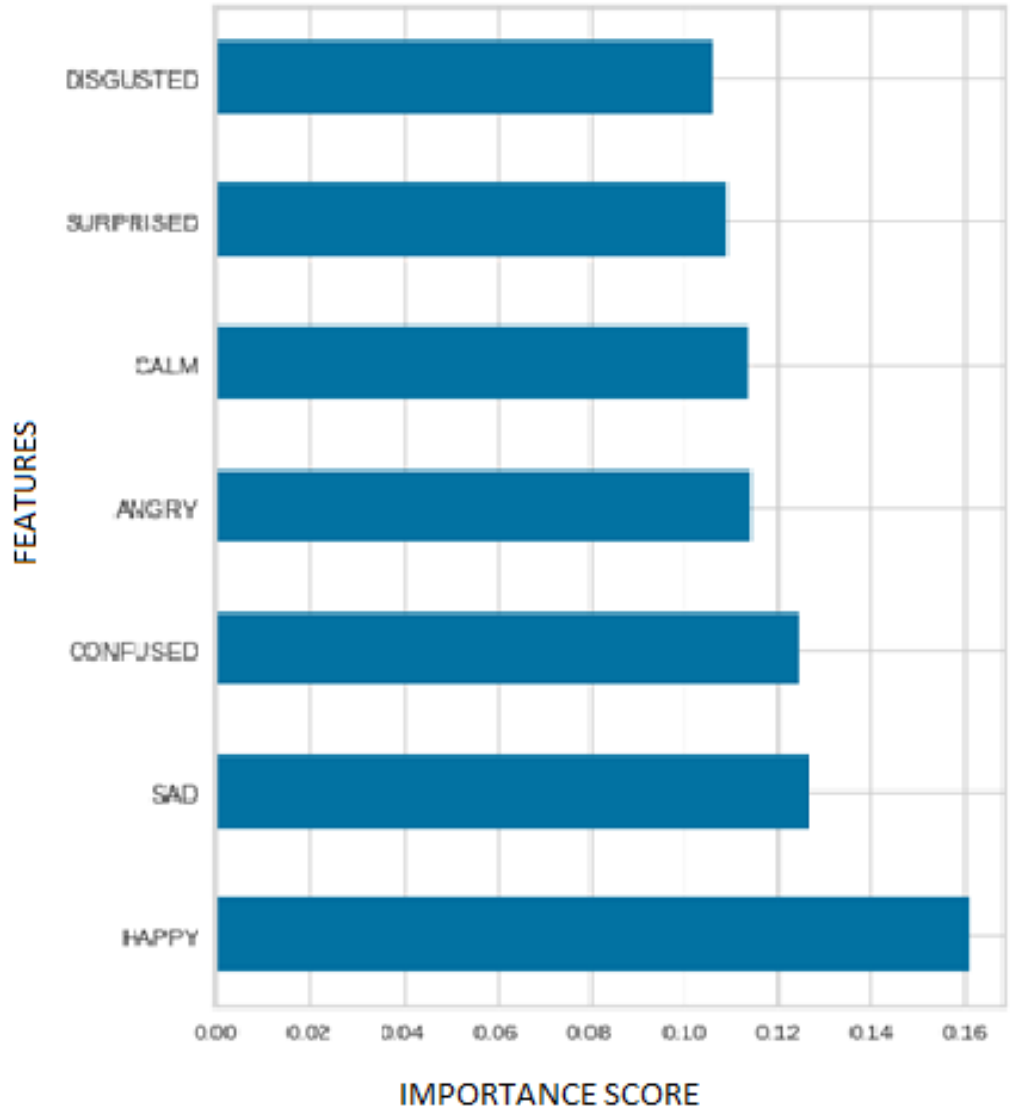


Figure 3.4: Feature importance plot from Extremely Randomized Tree

Fig3.4 shows that Extra Trees model gives the most importance to the 4 features which are text-sentiment, CALM, CONFUSED, HAPPY. So these are the most contributing features according to our Extra Trees model.

An active visualization of a trained decision tree structure will help in performing feature selection. The features at the top of the tree are the ones that our model regards with the most importance in order to perform it's classification. Hence, we can create a model with a viable accuracy score by picking some of the features from the top and discarding the rest.

Using Decision tree we can visualize the data.

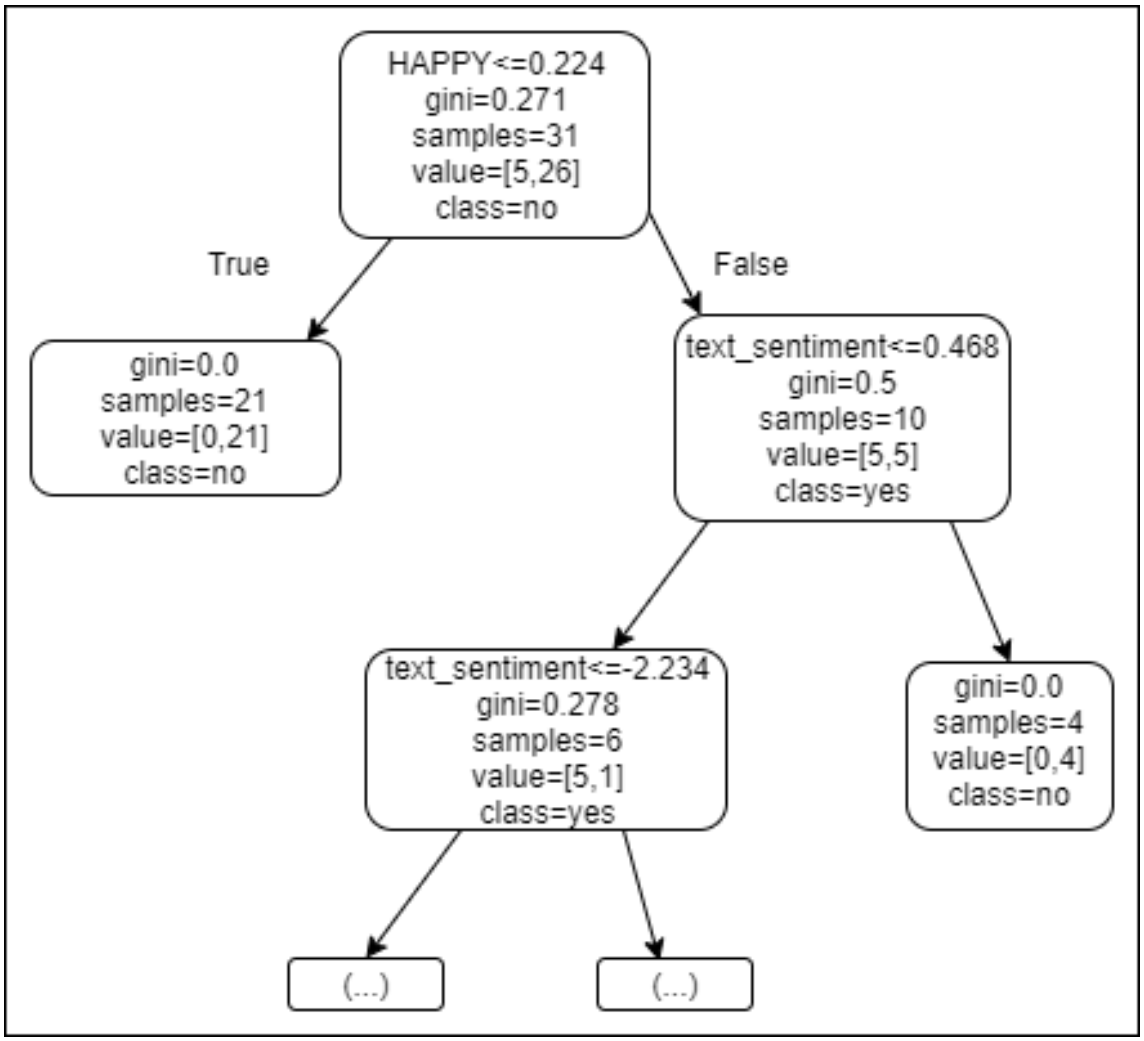


Figure 3.5: Decision Tree visualization

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)

This is a feature selection method which adjusts a system and obtains specified number of features by eradicating the weakest feature(s). Using the models coef or feature importances attributes the features categorized. By repeatedly removing a small number of features per loop, RFE aims to get rid of possible collinearity and dependencies in the system. It is needed for RFE keep a specified number of features, however, the numbers of valid features cannot always be known in advance. In order to find the most favorable number of features cross validation is used with RFE, which scores different feature subsets and then settles on the collection of features with the most scores. The RFECV visualizer demonstrates the number of features in the model, with their corresponding cross-validated test score and variability and also shows the selected number of features.

Using Stratified K-fold cross validation with 4 folds and XGBoost classifier as the model the RFE curve is shown in figure 3.6 -

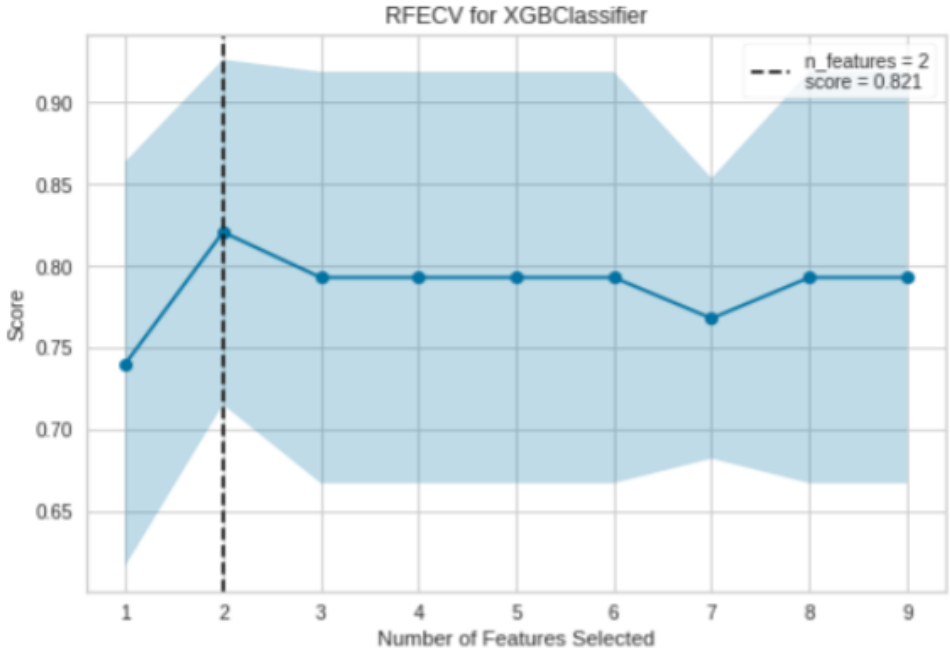


Figure 3.6: RFE Curve

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 is shown in Figure 3.7-

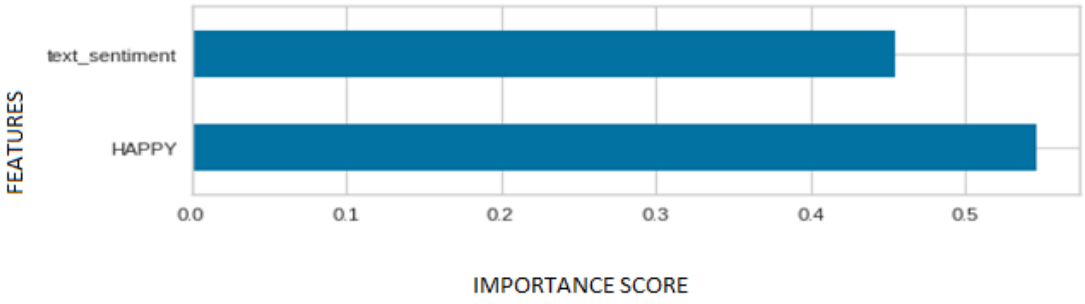


Figure 3.7: RFE Plot

Fig3.7 shows that text-sentiment and HAPPY are the two most important feature detected by RFE with XGBoost model.

Using Stratified K-fold cross validation with 4 folds and Random Forest classifier as the model the RFE curve is shown in Figure 3.8 -

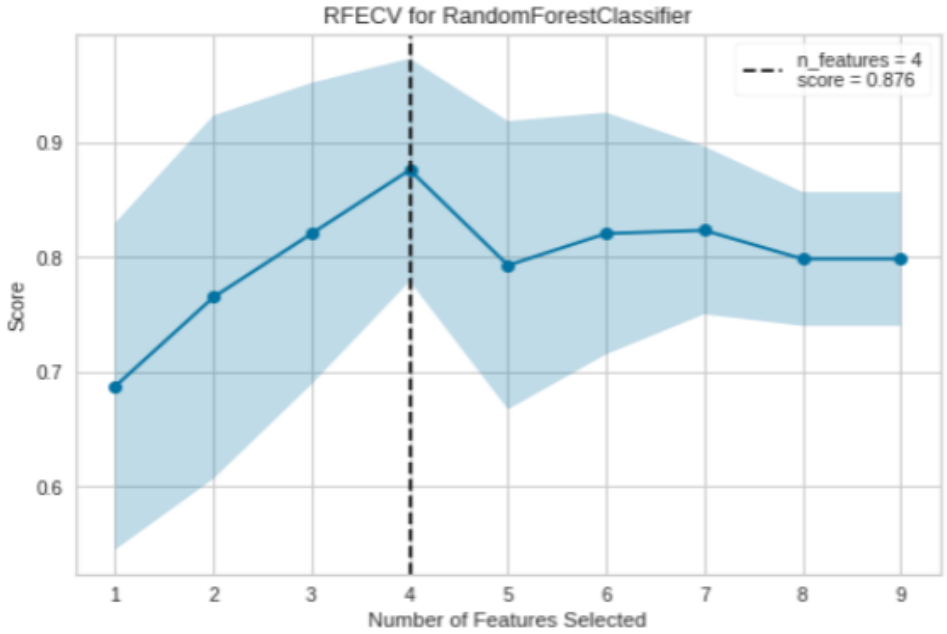


Figure 3.8: RFE Curve with random forest classifier

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 is shown in Figure 3.9-

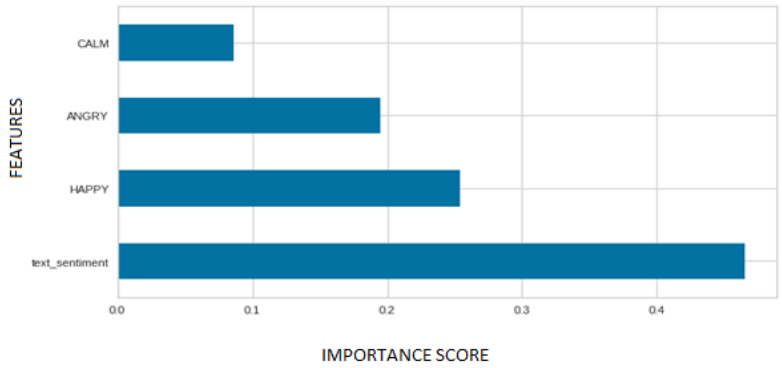


Figure 3.9: RFE feature importance plot for Random Forest classifier

Fig3.9 shows that CALM, ANGRY, text-sentiment and HAPPY are the most important feature detected by RFE with Random Forest model.

Using Stratified K-fold cross validation with 4 folds and Extra trees classifier as the model, the RFE curve is shown in Fig 3.10-

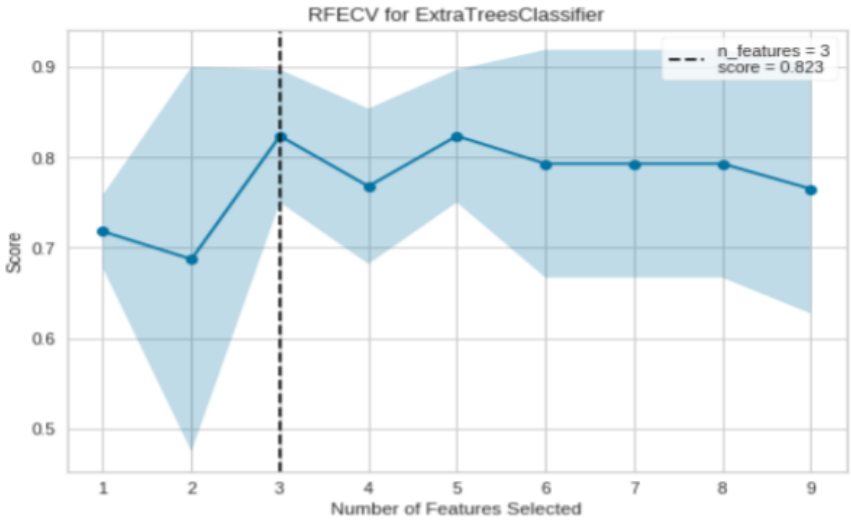


Figure 3.10: RFE curve with Extra Tree classifier

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 is shown in figure 3.11-

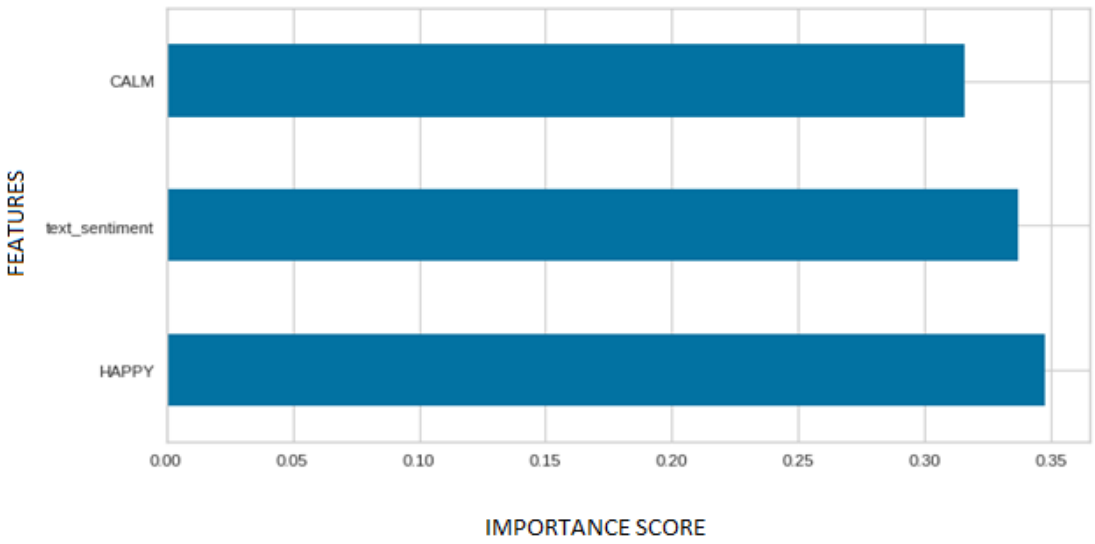


Figure 3.11: RFE feature importance plot with Extra Tree classifier

Fig3.11 shows that CALM, text-sentiment and HAPPY are the most important feature detected by RFE with Extra Tree classifier model.

3.5.3 Correlation Matrix Analysis

one of the method possible that is okay to be implemented for the purpose of minimizing the count of specifications in our domain of data is to observe between connection of the specifications associated labels.By making use of Pearson correlation ,the resultant numbers will range in between -1 and 1:

- When the connection that exists in two attributes are 0 then it would mean that transforming any of these two features would not have any sort of impact on other.
- When the correlation that exists in two features is greater than 0 then it would mean when turning up the values in a singular attribute,it would also improve the results in the different attributes. Here if we are close to the value of 1 then it would suggest that the stronger the bond is going to be.
- When the correlation that exists in two features is in the negative side of 0 then that would mean that increasinng the results in a singular attribute would take down the results of other existing attribute.

Correlation of all the features with our output variable is shown in Figure 3.12-

FEATURE NAMES	CORRELATION COEFFICIENT
CONFUSED	0.007842
DISGUSTED	0.023856
SAD	0.047085
FEAR	0.047460
SURPRISED	0.056781
ANGRY	0.214358
text_sentiment	0.439967
CALM	0.451341
HAPPY	0.478552
LABEL	1.000000

Figure 3.12: Correlation coefficient of all the features with the output

Fig3.12 description- shows that the most correlated features with our output are ANGRY, CALM, HAPPY, text$_s$*entiment*

The relationship between the different correlated features by creating a Correlation Matrix is shown below in Fig 3.13-

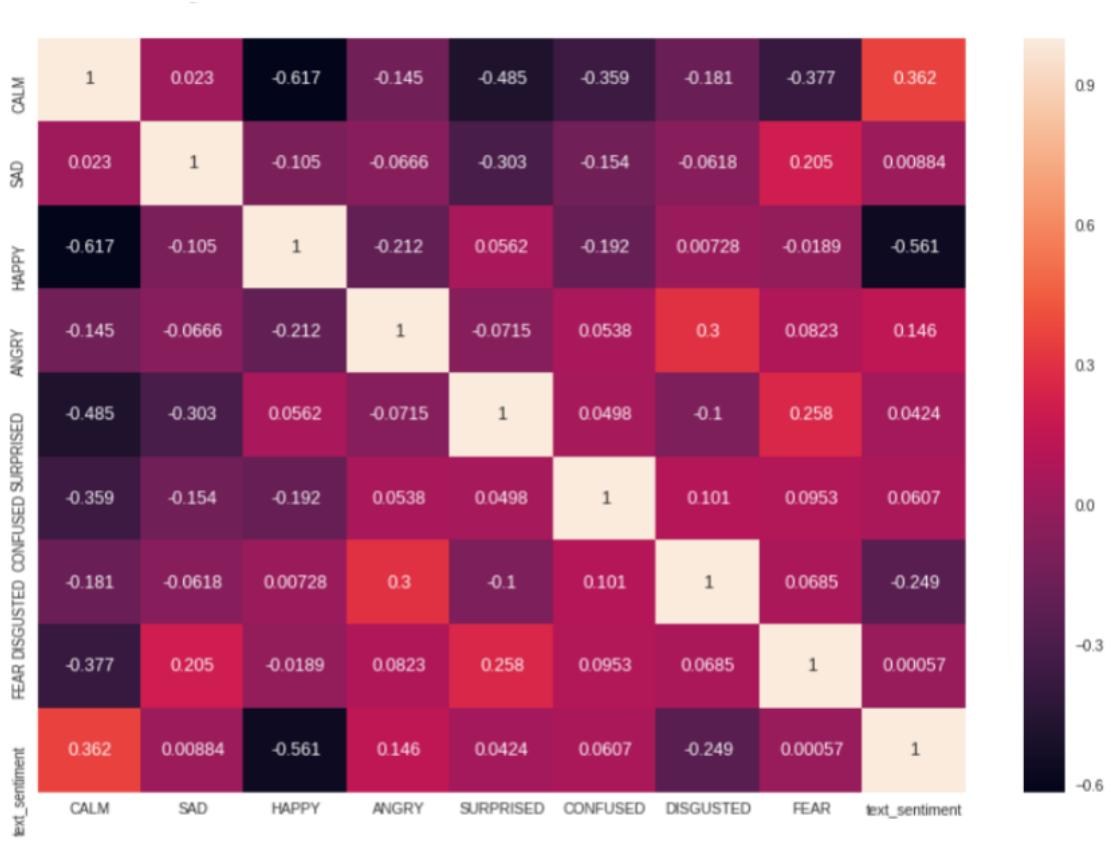


Figure 3.13: Correlation matrix between the feature sets

From this we can see that the most correlated features are:- HAPPY, CALM, text-sentiment, ANGRY
Based on the results obtained by feature selection methods we choose to test our model using the following sets of features -

- Feature Set 1(all features), S1 = 'CALM', 'SAD', 'HAPPY', 'ANGRY', 'SURPRISED', 'CONFUSED', 'DISGUSTED', 'FEAR', 'text-sentiment'
- Feature Set 2, S2 = 'HAPPY', 'CONFUSED', 'text-sentiment', 'CALM'
- Feature Set 3, S3 ='HAPPY','text-sentiment','SAD'
- Feature Set 4, S4 ='HAPPY','text-sentiment','FEAR'
- Feature Set 5, S5 = HAPPY','text-sentiment'
- Feature Set 6, S6 ='CALM','HAPPY','text-sentiment'

Chapter 4

Results and Analysis

4.1 Results and Analysis

Most usually used phrases that are used while evaluating performance of a version are Accuracy, Precision, Recall, FI-score, ROC, AUC and Confusion Matrix. A confusion matrix is a table which is designed to describe the performance of a classification version on a fixed of test facts in which the authentic values are known. In a type task, the precision for a class is the range of proper positives divided by the whole quantity of factors classified as belonging to the fine class . Recall is described because the number of valid positives divided by the whole number of elements that clearly belong to the advantageous class.All the measures except AUC is calculated by using four parameters which are as fol1ows[26]-

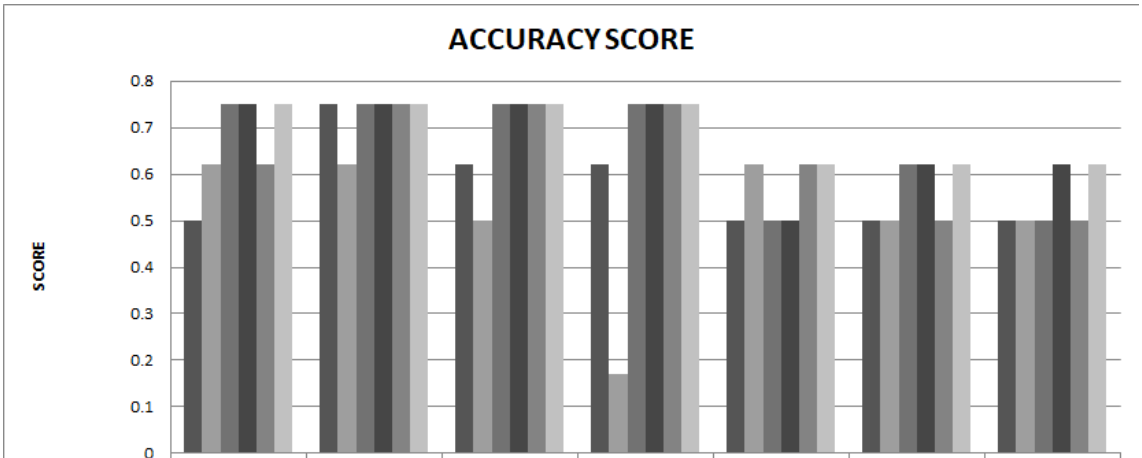
- True Positives (TP) - These can be termed as correctly guessed positive data that means that the value of actual class is yes and also value associated with predicted class is also yes.
- True Negatives (TN) - These can be termed as correctly guessed negative values that means that the value of actual class is no and also value associated with the predicted class is also no.
- False positives and false negatives, these values occur when our actual class contradicts with the predicted class.
- False Positives (FP) – This happens when actual class is no and predicted class is yes.
- False Negatives (FN) – This happens when actual class is yes but class of the predicted type is no. These four parameters are important for calculating these metrics.

4.1.1 Accuracy

This in the end, can be termed as the maximum intuitive overall measure of how models performed.This is definitely a proportion of efficaciously guessed commentary in the overall research. Many might have possibly thought, given a better accuracy when compared to other version is best. In truth, this is a terrific metric however for a specific function which is when we had got data domains that are symmetric .This

is when the results of incorrect positives and incorrect negatives can be considered nearly similar. So, we ought to search in different criteria in order to assess the effectiveness of our system. The equation is - Accuracy $= TP + TN/TP + FP + FN + TN$

For our model, The obtained accuracy for each algorithms using each set of features, the generated graphs are given below in figure 4.1-



	RANDOM FOREST	XGBOOST	LOGISTIC REGRESSION	LDA	KNN	GAUSSIAN NAÏVE BAYES	SVM
FEATURE SET 1	0.5	0.75	0.62	0.62	0.5	0.5	0.5
FEATURE SET 2	0.62	0.62	0.5	0.17	0.62	0.5	0.5
FEATURE SET 3	0.75	0.75	0.75	0.75	0.5	0.62	0.5
FEATURE SET 4	0.75	0.75	0.75	0.75	0.5	0.62	0.62
FEATURE SET 5	0.62	0.75	0.75	0.75	0.62	0.5	0.5
FEATURE SET 6	0.75	0.75	0.75	0.75	0.62	0.62	0.62

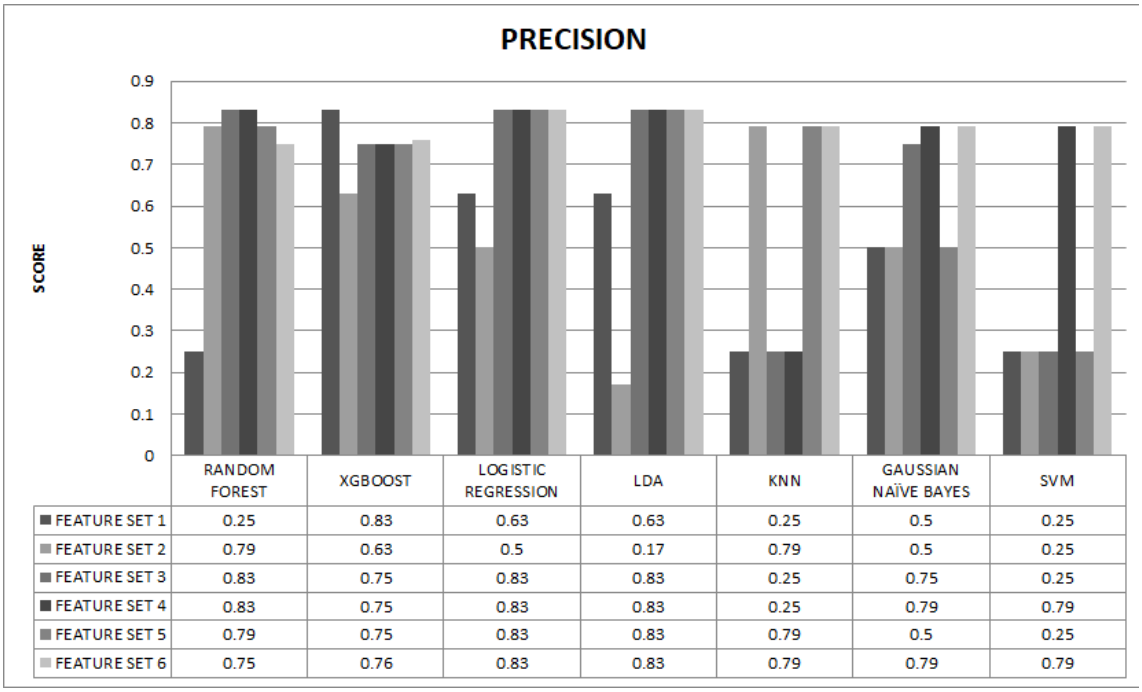
Figure 4.1: Accuracy Score using different algorithms for all feature sets

Fig4.1 shows a comparative graph of the accuracy scores on all the 6 feature sets. From the graph we can see that XGBoost has the best accuracy score over all the feature sets compared to other algorithms.

4.1.2 Precision

Precision is termed as the ratio of accurately estimated true observations to the wholly estimated true observations. A high precision score results in a low false positive rate. We had got 0.788 precision which is pretty good.The equation is -Precision $= TP/TP + FP$

For our model, The obtained precision for each algorithms using each set of features, the generated graphs are given below in Figure 4.2-



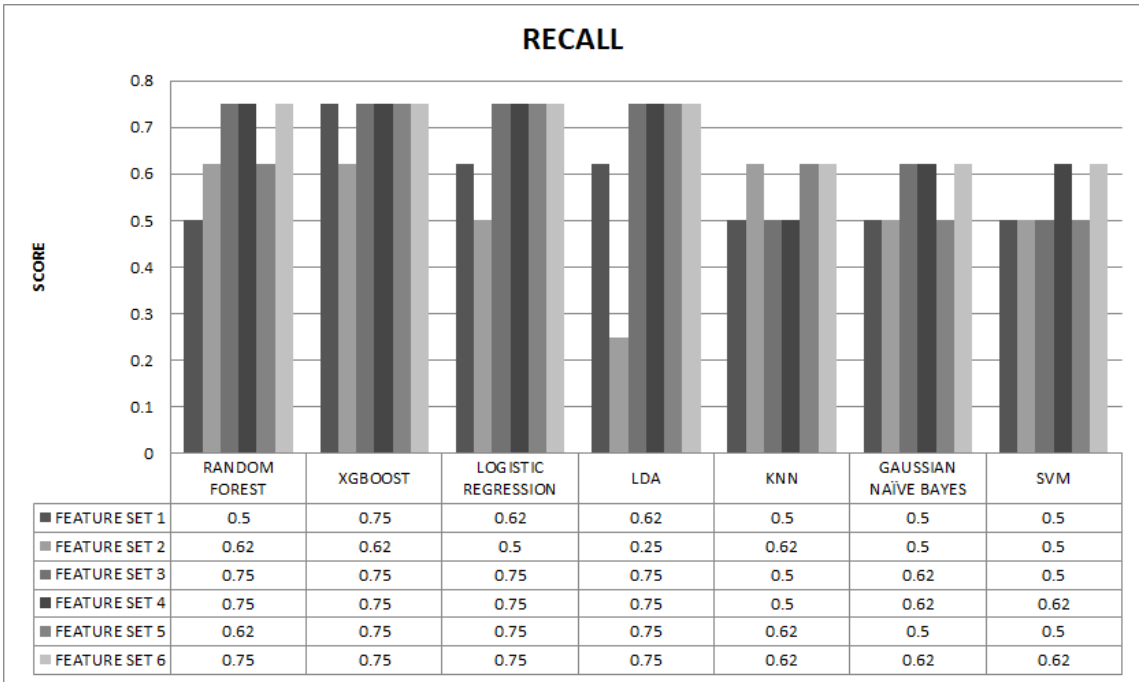
	RANDOM FOREST	XGBOOST	LOGISTIC REGRESSION	LDA	KNN	GAUSSIAN NAÏVE BAYES	SVM
FEATURE SET 1	0.25	0.83	0.63	0.63	0.25	0.5	0.25
FEATURE SET 2	0.79	0.63	0.5	0.17	0.79	0.5	0.25
FEATURE SET 3	0.83	0.75	0.83	0.83	0.25	0.75	0.25
FEATURE SET 4	0.83	0.75	0.83	0.83	0.25	0.79	0.79
FEATURE SET 5	0.79	0.75	0.83	0.83	0.79	0.5	0.25
FEATURE SET 6	0.75	0.76	0.83	0.83	0.79	0.79	0.79

Figure 4.2: Precision scores on all the feature sets for different algorithms

Fig4.2 shows a comparative graph of the precision scores on all the 6 feature sets. From the graph we can see that Random Forest has the best precision score over all the feature sets compared to other algorithms.

4.1.3 Recall

Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.We have got recall of 0.631 which is good for this model as it's above 0.5. The equation is -Recall = $TP/TP + FN$
For our model, The obtained recall for each algorithms using each set of features, the generated graphs are given below in Figure 4.3-



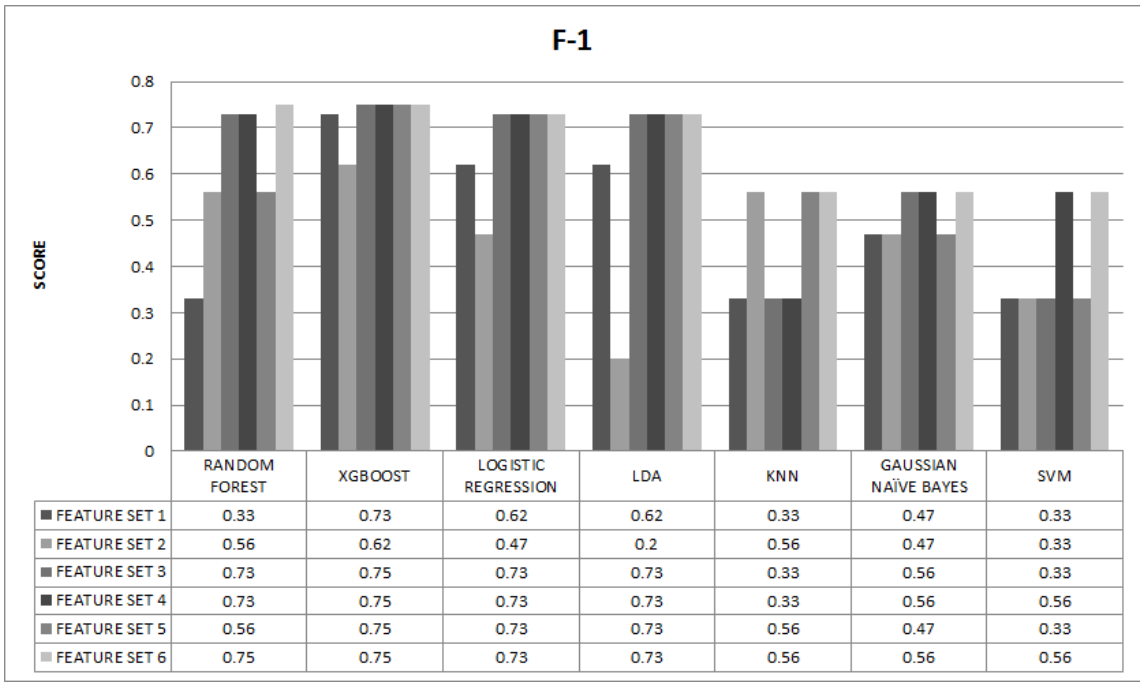
	RANDOM FOREST	XGBOOST	LOGISTIC REGRESSION	LDA	KNN	GAUSSIAN NAÏVE BAYES	SVM
FEATURE SET 1	0.5	0.75	0.62	0.62	0.5	0.5	0.5
FEATURE SET 2	0.62	0.62	0.5	0.25	0.62	0.5	0.5
FEATURE SET 3	0.75	0.75	0.75	0.75	0.5	0.62	0.5
FEATURE SET 4	0.75	0.75	0.75	0.75	0.5	0.62	0.62
FEATURE SET 5	0.62	0.75	0.75	0.75	0.62	0.5	0.5
FEATURE SET 6	0.75	0.75	0.75	0.75	0.62	0.62	0.62

Figure 4.3: Recall scores on all the feature sets for different algorithms

Fig4.3 shows a comparative graph of the recall scores on all the 6 feature sets. From the graph we can see that XGBoost has the best recall score over all the feature sets compared to other algorithms.

4.1.4 F1 Score

Precision and Recall give out F1 Score by taking their weighted average. Therefore, this score takes both false positives and false negatives into account. F1 is more beneficial than accuracy if we have an asymmetric class distribution. However, it is not as easy to grasp as accuracy.Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall. In our case, FI score is 0.701. F1 Score = 2*(Recall1 * Precision) / (Recall + Precision)
For our model, The obtained F-1 Score for each algorithms using each set of features, the generated graphs are given below in Figure 4.4-



	RANDOM FOREST	XGBOOST	LOGISTIC REGRESSION	LDA	KNN	GAUSSIAN NAÏVE BAYES	SVM
FEATURE SET 1	0.33	0.73	0.62	0.62	0.33	0.47	0.33
FEATURE SET 2	0.56	0.62	0.47	0.2	0.56	0.47	0.33
FEATURE SET 3	0.73	0.75	0.73	0.73	0.33	0.56	0.33
FEATURE SET 4	0.73	0.75	0.73	0.73	0.33	0.56	0.56
FEATURE SET 5	0.56	0.75	0.73	0.73	0.56	0.47	0.33
FEATURE SET 6	0.75	0.75	0.73	0.73	0.56	0.56	0.56

Figure 4.4: F-1 scores on all the feature sets for different algorithms

Fig4.4 shows a comparative graph of the F-1 scores on all the 6 feature sets. From the graph we can see that XGBoost has the best F-1 score over all the feature sets compared to other algorithms.

4.1.5 AUC- ROC

The probability that a classifier will rank an arbitrarily selected positive instance more than a arbitrarily selected negative example is represented by the area under the curve (AUC). It estimates the classifying model's expertise in ranking a set of patterns according to the extent to which they belong to the positive class without really assigning ample to their respective classes. The complete accuracy also relies on the capacity of the model to categorize the pattern and on the capacity to choose a baseline in the category used to assign patterns to the positive class if it I above the baseline and to the negative if it is beneath. Thus, the classifier with the higher AUROC statistic (assuming everything to be equal) is likely to have a higher overall accuracy as the ranking of patterns (which AUROC measures is advantageous to both AUROC and overall accuracy. Nevertheless, if one model categorizes patterns well but choses the baseline badly, then it would result in a high AUROC with a poor overall accuracy.

ROC (Receiver Operating Characteristics) is the curve drawn by connecting the dots of x-axis = FPR (False Positive Rates) and y-axis = TPR (True Positive Rates) for different threshold values. It means we choose different threshold values for our model and calculates TPR and FPR for them and the draw the ROC curve and calculate the area under the curve. AUC (Area Under Curve) is the area under the ROC curve. We use this curve because - It tells us how good our model is about seperating the two classes. For our case, classes are satisfied or not satisfied. It helps us about choosing the best threshold. By analyzing all our results we can conclude

that Feature Set 6 has the best overall performance on all the algorithms. So for 6 the ROC curve for each algorithm is

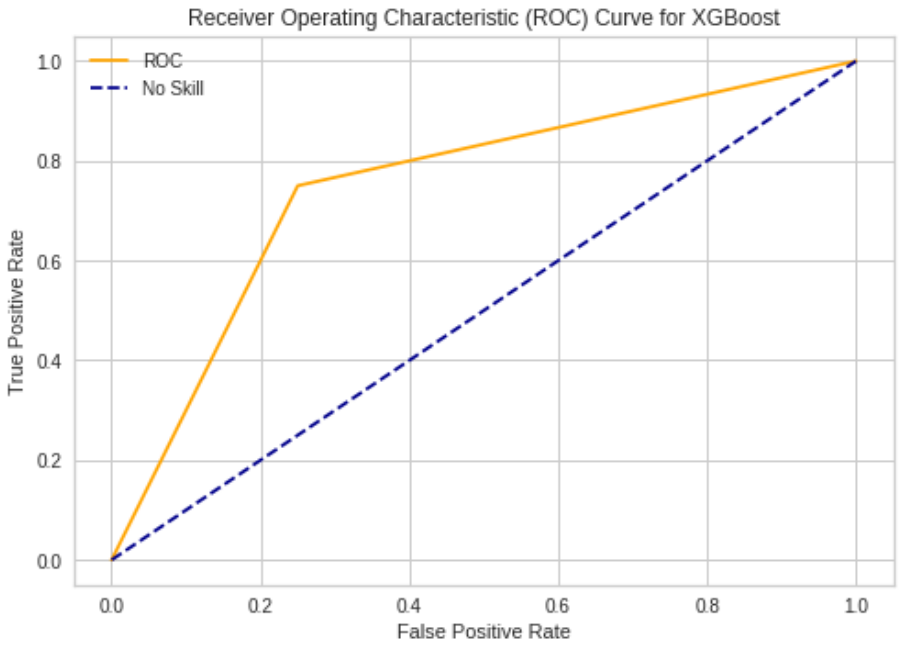


Figure 4.5: ROC Curve for XGBoost

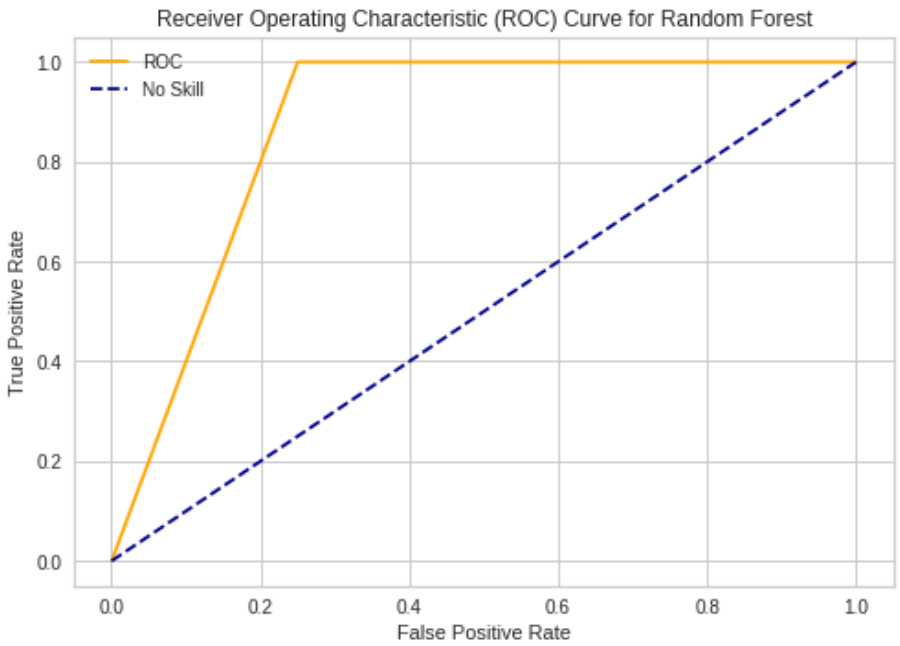


Figure 4.6: ROC Curve for Random Forrest

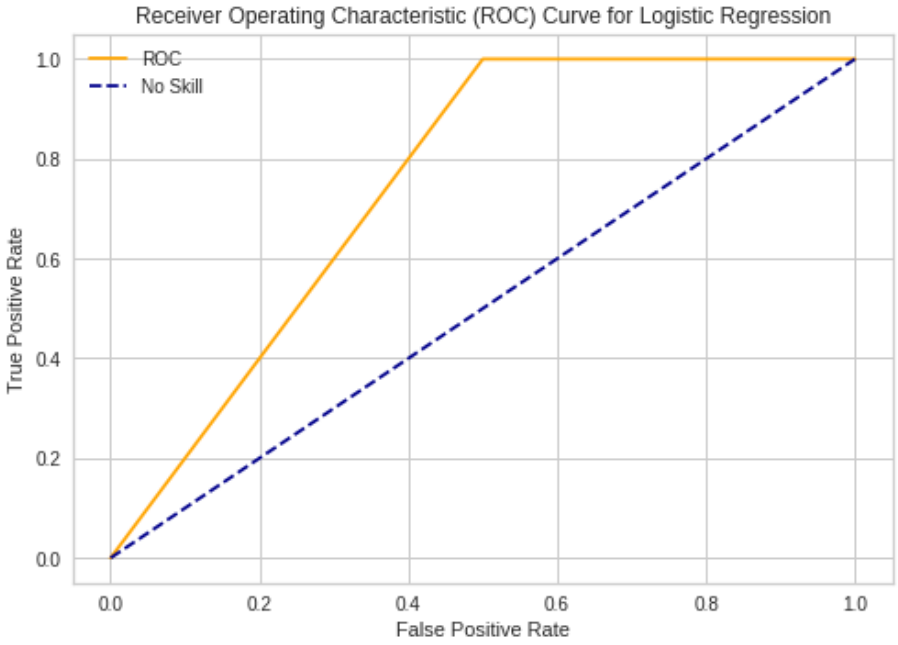


Figure 4.7: ROC Curve for Logistic Regression

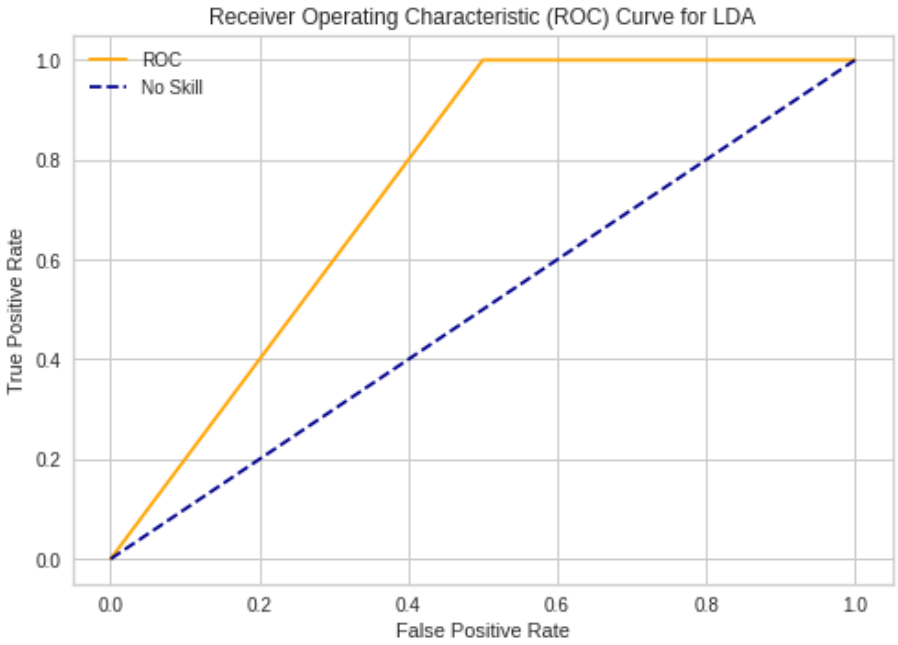


Figure 4.8: ROC Curve for LDA

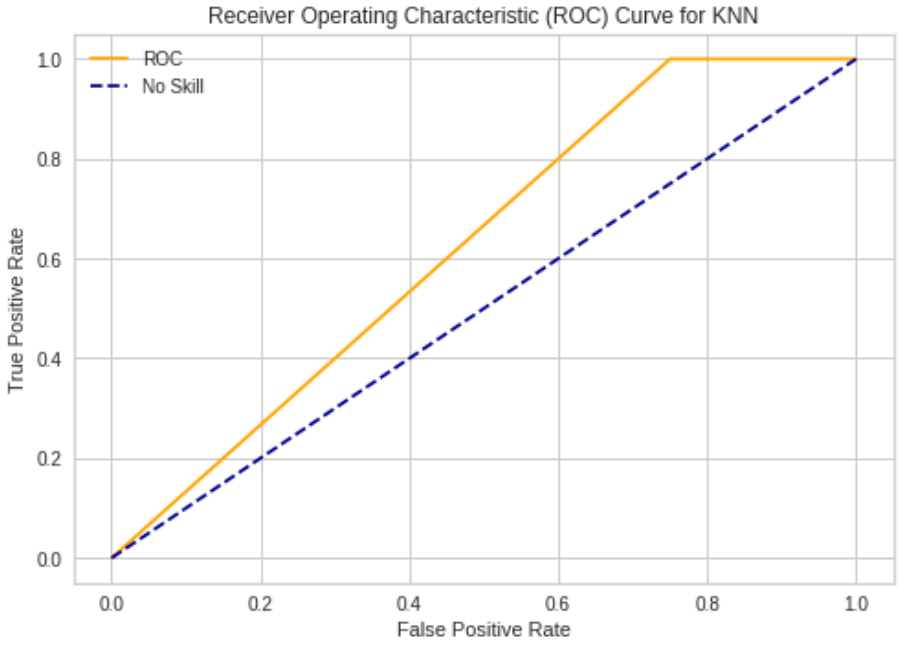


Figure 4.9: ROC Curve for KNN

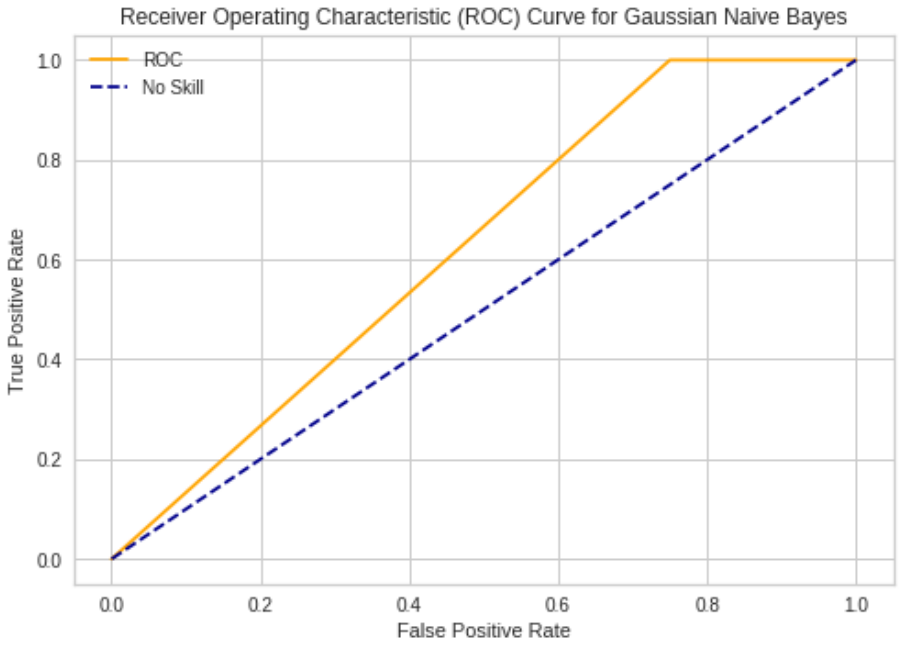


Figure 4.10: ROC Curve for naive bayes

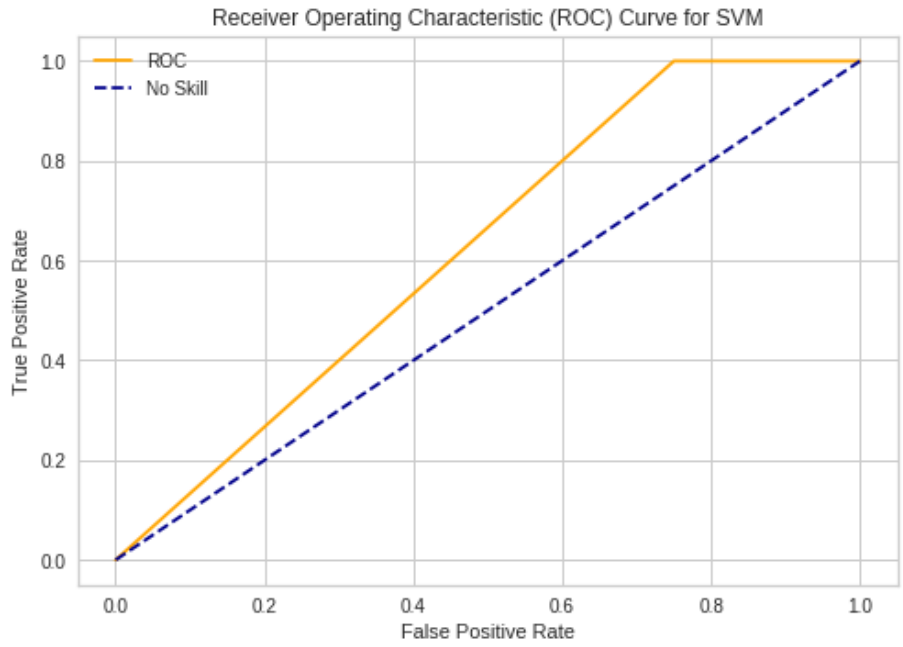


Figure 4.11: ROC Curve for SVM

Based on the performance metrics we can conclude that reduced feature set S6 gives the best overall result on all algorithms. And XGboost outperforms all other algorithms. Thus we can conclude that for our case XGBoost has the best accuracy over all aspects using the feature set 6 which contains 3 most signoficant features i.e "Happy", "CALM", "text-sentiment".

4.1.6 k-Fold Cross-Validation

In the intricate and massive field of machine learning, this method is used for re-sampling and to evaluate different ML algorithms for a constraint sample of data. It is mainly used to estimate expertise of a particular model on hidden data. Cross-validation uses a limited test set so that it can determine the various existing model's predicted performance in a comprehensive manner when this is put into use to have making educated guess on various data never being put into use in the time of the training of those particular systems model. The phases are as follows:

Rearrange the domain of data arbitrarily.

Divide the domain of data in total of k sections.

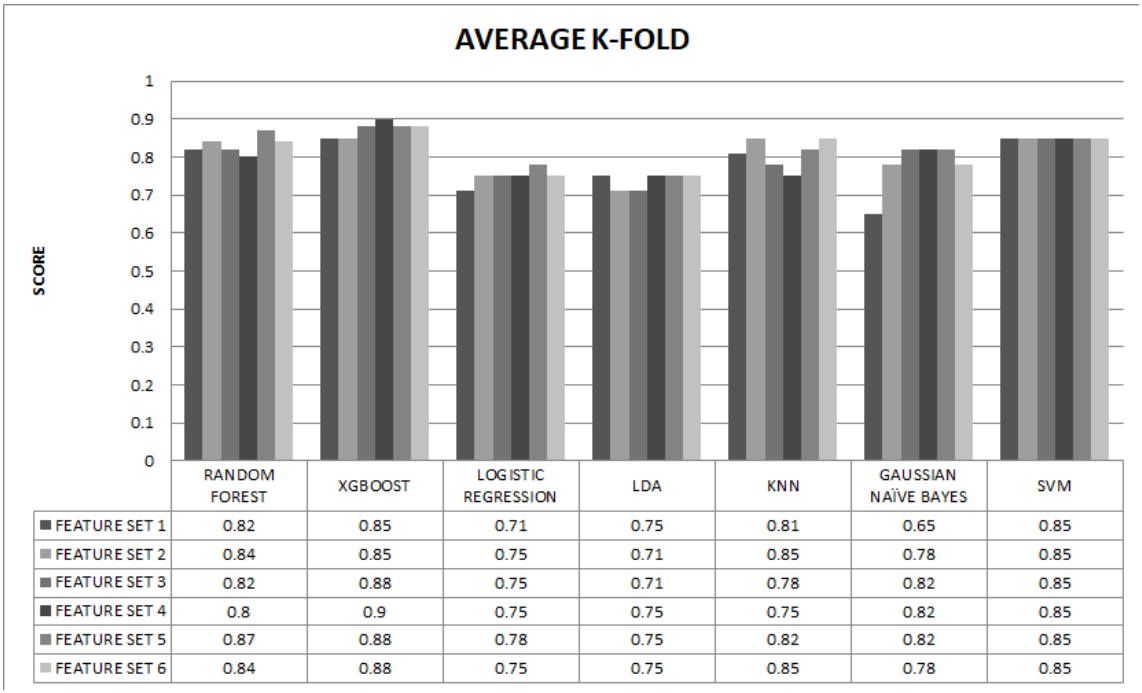
In every separate section:

Select the section as a the chosen data domain for training

Select the left over sections as a data domain for training.

Adjust a model upon the set which is used for training.After that evaluate them upon test set

Reserve the resultant score. After that model need to be rejected
In our case we choose the value of k as 10 and evaluate the mean score and the standard deviation for each model using all of the 6 feature sets one by one and generate the following Figure 4.12-

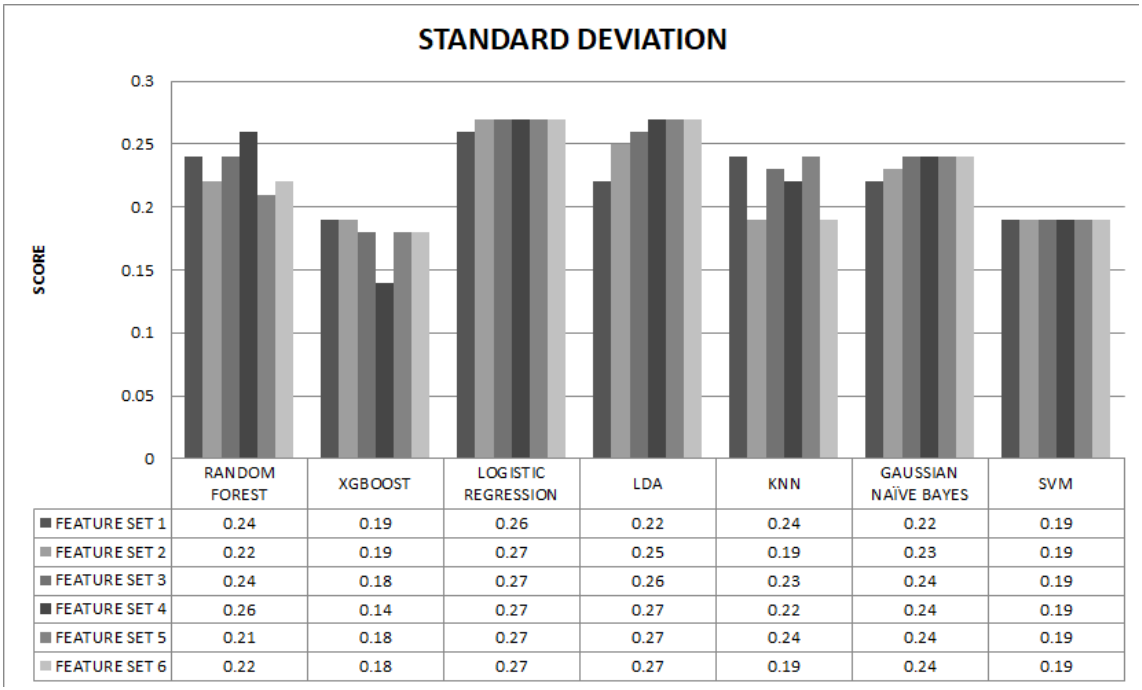


	RANDOM FOREST	XGBOOST	LOGISTIC REGRESSION	LDA	KNN	GAUSSIAN NAÏVE BAYES	SVM
FEATURE SET 1	0.82	0.85	0.71	0.75	0.81	0.65	0.85
FEATURE SET 2	0.84	0.85	0.75	0.71	0.85	0.78	0.85
FEATURE SET 3	0.82	0.88	0.75	0.71	0.78	0.82	0.85
FEATURE SET 4	0.8	0.9	0.75	0.75	0.75	0.82	0.85
FEATURE SET 5	0.87	0.88	0.78	0.75	0.82	0.82	0.85
FEATURE SET 6	0.84	0.88	0.75	0.75	0.85	0.78	0.85

Figure 4.12: Average K-Fold scores for 10 fold cross validation over all the feature sets using different algorithms

The figure shows a comparative graph of the average K-Fold scores on all the 6 feature sets. From the graph we can see that XGBoost has the best average K-Fold score over all the feature sets compared to other algorithms.

The standard deviation of the K-Fold Score can be shown in the below Figure 4.13-



	RANDOM FOREST	XGBOOST	LOGISTIC REGRESSION	LDA	KNN	GAUSSIAN NAÏVE BAYES	SVM
FEATURE SET 1	0.24	0.19	0.26	0.22	0.24	0.22	0.19
FEATURE SET 2	0.22	0.19	0.27	0.25	0.19	0.23	0.19
FEATURE SET 3	0.24	0.18	0.27	0.26	0.23	0.24	0.19
FEATURE SET 4	0.26	0.14	0.27	0.27	0.22	0.24	0.19
FEATURE SET 5	0.21	0.18	0.27	0.27	0.24	0.24	0.19
FEATURE SET 6	0.22	0.18	0.27	0.27	0.19	0.24	0.19

Figure 4.13: Standard deviation of the K-Fold score

The figure shows a comparative graph of the K-Fold score standard deviation on all the 6 feature sets. From the graph we can see that XGBoost has the minimum standard deviation over all the feature sets compared to other algorithms.

Chapter 5

Conclusions and Future works

5.1 Conclusions

In todays' world, The role of sentiment mining has huge implications in understanding consumer's emotions in an automated way .A lot of people can use this in depth knowledge to make improve of their own brand of product or services in an effective manner. This can be a field of interest for both researchers and entrepreneurs as this can provide a new edge in terms of strategy.
In regards to that, we can definitely assure a different approach to analyze sentiments through our whole idea and its implementation. Our project was able to fuse 2 different modals, i.e. text and video, and give a result based on these 2 which are text and video. This ensured us a greater accuracy than any other proposed models available in the market which were based on text or video only and or even both.

5.2 Future work

However,With the advancements of our system in mind for the future, we would like to integrate neuro-analysis along with our previously used modals to make our system even more comprehensive and accurate in the nearby future. Using an EEG(Electrocortigraphy), we want to take signals from the brain, and the emotions its portraying. Then we can integrate to our results from the multi modal system to create a more robust system, which can prove to be a great equalizer in the field of market research

Bibliography

[1] S. Symeonidis, *5 things you need to know about sentiment analysis and classification*. [Online]. Available: https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html.

[2] Monkeylearn, *Sentiment analysis*. [Online]. Available: https://monkeylearn.com/sentiment-analysis/.

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

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

[19] Amazon, *Amazon rekognition*. [Online]. Available: https://aws.amazon.com/rekognition/.

[20] G. Blog, *An end-to-end guide to understand the math behind xgboost*, 2018. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/.

[21] A. Samudrala, *Unveiling mathematics behind xgboost*, 2018. [Online]. Available: https://www.kdnuggets.com/2018/08/unveiling-mathematics-behind-xgboost.html.

[22] *Boosting algorithm: Xgboost*, 2017. [Online]. Available: https://towardsdatascience.com/boosting-algorithm-xgboost-4d9ec0207d.

[23] R. Rafeek and R Remya, "Detecting contextual word polarity using aspect based sentiment analysis and logistic regression", in *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, IEEE, 2017, pp. 102–107.

[24] 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.

[25] N. S. Chauhan, *Classifying heart disease using k-nearest neighbors*. [Online]. Available: https://towardsdatascience.com/implement-k-nearest-neighbors-classification-algorithm-c99be8f14052.

[26] R. Joshi, *Accuracy, precision, recall f1 score: Interpretation of performance measures*, 2016. [Online]. Available: https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures.

[27] J. Brownlee, *A gentle introduction to k-fold cross-validation*, 2018. [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/.