

Predicting Effectiveness of Marketing through Analyzing Emotional Context in Advertisement using Deep Learning

by

Sheikh Mohammad Arafat

16301147

Rifatul Islam

16301186

Ishraque Arefin Rafi

16201002

Md. Rashedul Islam

17301213

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
April 2020

© 2020. Brac University
All rights reserved.

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:

Arafat

Sheikh Mohammad Arafat
16301147



Ishraque Arefin Rafi
16201002

রিফাতুল ইসলাম

Rifatul Islam
16301186



Md. Rashedul Islam
17301213

Approval

The thesis titled “Effectiveness of Marketing Through Analyzing The Emotional Context in Advertisement Using Deep Learning” submitted by

1. Sheikh Mohammad Arafat (16301147)
2. Rifatul Islam (16301186)
3. Ishraque Arefin Rafi (16201002)
4. Md. Rashedul Islam (17301213)

Of Spring, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on April 07, 2020.

Examining Committee:

Supervisor:



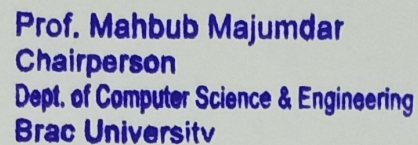
Md. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
Institution

Program Coordinator:



Md. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:



Prof. Mahbub Majumdar
Chairperson
Dept. of Computer Science & Engineering
Brac University

Mahbubul Alam Majumdar
Professor and Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

In this modern age, marketing strategy is becoming a new challenge. Not only the global market but also people's choices are shifting to catch the attention of buyers. Also, based on consumer's choice organizations are bringing changes in their marketing policy to increase the chances of their product selling rate. Basically, to promote their products and grab buyer's attention they are promoting advertisements on every media platform. But they are not aware of the effectiveness of marketing and which emotional states are needed more and which are not needed much. Therefore, we lead this study to recognize a successful advertisement and identify the rate of the emotional states which make good impact in people mind to purchase the product. Using deep learning and supervised machine learning algorithms as well as feature extraction methods for instance, LSTM-RNN, SVM, XGBOOST, Naïve Bayes, Multiple Linear Regression, MFCC, Zero-Crossing Rate, Power Spectral Density, we find out and evaluate the rate of the emotional states to figure out the liking and purchase intent which makes an advertisement successful.

Keywords: Effectiveness of Marketing; Emotional States; Deep Learning; Supervised Machine Learning; LSTM-RNN; MFCC.

Acknowledgement

To begin with, we all are grateful to Almighty Allah for getting this amazing opportunity to learn something new. It has been a very great experience for us where we can face many difficulties and we have overcome this gratefully. By the grace of Allah, we have been able to put our best effort and successfully complete it on time.

Secondly, we would like to convey our gratitude to our respected supervisor Dr. Md. Golam Rabiul Alam for his amazing guidance and tireless contribution throughout the whole phase of our thesis. From the very beginning of the journey to the end he provided us a lot of resources and all kinds of helps and led us to our desired goal. Moreover, he inspired us a lot whenever we lost our track and felt dissolving, to move forward to our goal.

Lastly, we will always be thankful to our parents, friends and all of our well-wishers who have been supported us throughout the journey of our research. We always a lot of inspiration from them and get the strength to move forward beyond all the barriers. Therefore, we also like to acknowledge all the assistance over the internet especially from related works from where we got countless number of resources and learn many things from there.

And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

Declaration	i
Approval	ii
Abstract	iii
Acknowledgment	iv
Table of Contents	v
List of Figures	vii
List of Tables	x
Nomenclature	xi
1 Introduction	1
1.1 Background	1
1.2 Motivation	1
1.3 Types of Advertisement	2
1.4 Problem Statement	3
1.5 Objective and Contribution	4
1.6 Thesis Structure	5
1.7 Work Plan	6
2 Related Work	8
3 Proposed Method	12
3.1 System Model	12
3.2 Feature Extraction	15
3.2.1 MFCC	16
3.2.2 Short-time Energy	17
3.2.3 Zero Crossing Rate	18
3.2.4 Power Spectral Density	19
3.2.5 Spectrogram	20
3.3 Dataset Description	21
3.3.1 Questionnaire Description	21
3.3.2 Textual Dataset	26
3.3.3 Feature Extracted Dataset	34
3.4 Methodology	38

3.4.1	XGBOOST	40
3.4.2	Naïve Bayes	40
3.4.3	Support Vector Machine (SVM)	40
3.4.4	Multiple Linear Regressions (MLR)	42
3.4.5	LSTM-RNN	43
4	Implementation and Result Analysis	48
4.1	Data pre-processing	48
4.2	Combined Decision Making	50
4.3	Results	51
4.3.1	Naive Bayes	51
4.3.2	XGBoost	52
4.3.3	LSTM-RNN	53
4.3.4	Support Vector Machine	59
4.3.5	Multiple Linear Regression	67
4.3.6	Performance Measurement	70
4.3.7	Final Result of Standard Emotional States	74
5	Conclusion and future Work	75
5.1	Conclusion	75
5.2	Future Work	75
	Bibliography	79
	Appendix A Adam Optimiztion Algorithm	80

List of Figures

1.1	Global advertising spending from 2010 to 2019(in billion U.S. dollars)	2
1.2	Workflow of the research	7
3.1	System Model of Finding Success Rate of Emotion	12
3.2	Survey Responses from Calmed to Excited (1 to 5)	13
3.3	Classes in Survey Questions	13
3.4	Data Cleaning (Data Generalization)	14
3.5	Implement LSTM-RNN	14
3.6	Scatter plot of MFCC coefficients value	16
3.7	Line Plot of MFCC coefficients value	17
3.8	Scatter Plot of Short-time Energy	17
3.9	Scatter Area Plot of Short-time Energy	18
3.10	Scatter Plot of Zero Crossing rate	19
3.11	Area plot of Zero Crossing Rate	19
3.12	Scatter Plot of Power Spectral Density	20
3.13	Area Plot of Power Spectral Density	20
3.14	Scatter Plot of Spectrogram	20
3.15	Area Plot of Spectrogram	21
3.16	Image from an Advertisement	21
3.17	Image from an Advertisement	22
3.18	Image from an Advertisement	22
3.19	Five Questions of an Advertisement	23
3.20	First Question from an Advertisement	23
3.21	Second Question from an Advertisement	24
3.22	Third Question from an Advertisement	24
3.23	Fourth Question from an Advertisement	25
3.24	Fifth Question from an Advertisement	25
3.25	Age Classification Among Participants	26
3.26	Gender Classification Among Participants	27
3.27	Participant Classification According to Salary	27
3.28	Time Duration Spent for Advertisement	28
3.29	Most Used Medias for Watching Advertisements	28
3.30	Question Based on Arousal	29
3.31	Question Based on Valence	30
3.32	FQuestion Based on Dominance	30
3.33	Question Based on Liking	30
3.34	Question Based on Purchase	31
3.35	SAM scale of Pleasure level	31

3.36	SAM scale of Motivating level	31
3.37	SAM scale of Excitement level	32
3.38	Survey Response for Valence	32
3.39	Survey Response for Arousal	33
3.40	Survey Response for Dominance	33
3.41	Data Generalization	34
3.42	Zero Crossing Rates for several Advertisements (Scatter Plot)	35
3.43	Power Spectral Density for several Advertisements (Aera Plot)	35
3.44	Short Time Energy for several Advertisements (Scatter with Smooth Line Plot)	36
3.45	Methodology	38
3.46	Precision Formula Representation	42
3.47	Recall Formula Representation	42
3.48	An unrolled RNN	44
3.49	Flow Diagram/Computational Graph of a Canonical LSTM Network with Hidden State Vectors h_t and Input vectors x	44
3.50	Sequential Flowchart of LSTM-RNN Model	46
3.51	LSTM-RNN Model	47
4.1	60 seconds sliced audios	49
4.2	Success rate based on purchase	50
4.3	Area Plot of Naive Bayes Classifier	51
4.4	Column Plot of Naive Bayes Classifier	52
4.5	Area Plot of XGBOOST Classifier	53
4.6	Column Plot of XGBoost Classifier	53
4.7	Area Plot of LSTM-RNN	54
4.8	Column Plot for LSTM-RNN	54
4.9	Arousal Accuracy of LSTM-RNN	55
4.10	Arousal Loss of LSTM-RNN	55
4.11	Valence Accuracy of LSTM-RNN	55
4.12	Valence Loss of LSTM-RNN	56
4.13	Dominance Accuracy of LSTM-RNN	56
4.14	Dominance Loss of LSTM-RNN	56
4.15	Liking Accuracy of LSTM-RNN	57
4.16	Liking Loss of LSTM-RNN	57
4.17	Purchase Accuracy of LSTM-RNN	57
4.18	Purchase Loss of LSTM-RNN	58
4.19	Sample Training of LSTM-RNN	58
4.20	SVM Scores from Textual Dataset in Clustered Column Plot	60
4.21	SVM Scores from Textual Dataset in Area Plot	60
4.22	SVM Scores from Textual and Extracted Dataset in Clustered Col- umn Plot	61
4.23	SVM Scores from Textual and Extracted Dataset in Area Column Plot	61
4.24	Success Rate of Successful Advertisements (Column Plot)	61
4.25	Success Rate of Successful Advertisements (Area Plot)	62
4.26	Success Rate of Unsuccessful Advertisements (Column Plot)	62
4.27	Success Rate of Unsuccessful Advertisements (Area Plot)	63
4.28	SVM of Arousal from Most Successful Advertisements	64

4.29	SVM of Valence from Most Successful Advertisements	64
4.30	SVM of Dominance from Most Successful Advertisements	64
4.31	SVM of Liking from Most Successful Advertisements	64
4.32	SVM of Purchase from Most Successful Advertisements	65
4.33	SVM of Arousal from Unsuccessful Advertisements	66
4.34	SVM of Valence from Unsuccessful Advertisements	66
4.35	SVM of Dominance from Unsuccessful Advertisements	66
4.36	SVM of Liking from Unsuccessful Advertisements	66
4.37	SVM of Purchase from Unsuccessful Advertisements	67
4.38	Residual Error of Liking intent	67
4.39	Residual Error of Purchase intent	68
4.40	Regression Analysis of Liking intent from Successful Advertisements .	68
4.41	Regression Analysis of Liking intent from Unsuccessful Advertisements	69
4.42	Regression Analysis of Purchase intent from Successful Advertisements	69
4.43	Regression Analysis of Purchase intent from Unsuccessful Advertisements	70
4.44	Heatmap of Confusion Matrix of Arousal	72
4.45	Heatmap of Confusion Matrix of Valence	72
4.46	Heatmap of Confusion Matrix of Dominance	72
4.47	Heatmap of Confusion Matrix of Liking	73
4.48	Heatmap of Confusion Matrix of Purchase	73

List of Tables

3.1	Arousal Table	34
3.2	Valance Table	34
3.3	Short Time Energy	36
3.4	Power Spectral Density	37
4.1	Naïve Bayes of Emotional States from Textual or Survey Data	51
4.2	XGBoost of Emotional States from Textual or Survey Data	52
4.3	LSTM-RNN Prediction for Emotional States from Survey or Textual Data	54
4.4	SVM of Emotional States from Textual Dataset	59
4.5	SVM of Emotional States from Textual and Extracted Dataset	59
4.6	SVM of Emotional States from Successful Advertisements	63
4.7	SVM of Liking and Purchase intent from Successful Advertisements	63
4.8	SVM of Emotional States from Unsuccessful Advertisements	65
4.9	SVM of Liking and Purchase intent from UNsuccessful Advertisements	65
4.10	Regression Analysis of Liking intent	67
4.11	Regression Analysis of Purchase intent	67
4.12	Regression Analysis of Liking intent from Most Successful Advertise- ment	68
4.13	Regression Analysis of Purchase intent from Most Successful Adver- tisement	69
4.14	Performance measurement for Arousal	71
4.15	Performance measurement for Valence	71
4.16	Performance measurement for Dominance	71
4.17	Performance measurement for Liking	71
4.18	Performance measurement for Purchase	72
4.19	Standard Emotional States Scores	74
4.20	Standard Emotional States Rate (in percentage)	74

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ϵ Residual

EDA Electro dermal Activity sensor

GSR Galvanic skin response sensor

LSTM – RNN Long Short Term Recurrent Neural Network

MFCC Mel Frequency Cepstral Coefficient

MLR Multiple Linear Regressions

PSD Power Spectral Density

SVM Support vector machine

XGBoost Extreme Gradient Boosting

ZCR Zero Crossing Rate

Chapter 1

Introduction

1.1 Background

In this era of science and technologies advertisement is the key way of connecting the consumers with the companies. Every company has to have their own strategy to reach their goal. There are several sectors a company can have such as HRM, IT, Marketing etc. Marketing is one of the major parts of every company's journey to success. Therefore, if we look around the world's most successful companies like Coca-Cola, McDonald's, and Nike we can see that marketing is one of the main reasons of this level of success[1]. There are several ways of marketing and advertisement in one of the most important among them. Through the advertisement companies are able to let the consumers know about the uniqueness of their product, about the impact of their product in the society or in a specific community. The companies always try to communicate with targeted community about the product and advertisement is the key way to connect them emotionally and logically. Moreover, in this recent time, advertisement can be taken place in different sources as in social media, television, radio etc. As a result, we can easily observe that an advertisement can make a huge impact on the consumer mind about the product and the company. That is why every advertisement has to have proper amount emotional appeal or proper message they are trying to convey. As a result, having proper advertisements with perfect amount of emotional appeal is must for a company's success.

1.2 Motivation

The world's market is accelerating very rapidly with time. Everything is in our hand with the help of technology; people have become more technologically involved rather than human interaction. As a result, they basically judge anything more according to the status on social media or on television or on radio. People's eyes are always on the screen and they actually move one place to another very frequently, because there are millions of contents about many things in social media or other platform. So, people's choices are changing very frequently according to contents they are seeing on the screen. As a result, the rate of purchasing of a product depends on whether the companies can introduce the product perfectly through advertisement or not to the customers. On the other hand, in this time with huge amount of technological resources, people easily lose their interest on a

thing, if that thing does not have any impact emotionally on them. So, making proper advertisement has to be the first choice for any companies' success. To solve the problem, we are analysing the factor of advertisement which are playing important role to make any advertisement successful. We are trying to predict the level of success of an advertisement using audio frequency. Moreover, we are using machine learning algorithms and deep learning algorithms for our work purpose to predict this success rate from the audio of advertisements. We have learnt the advancement of machine learning algorithms and deep learning algorithms with the help of our supervisor. Besides, we are using real time data set which is made by us and by the suggestion of our supervisor we have done survey to learn the real impression of the people about advertisements. As a result, with the combination of real time data and the survey we can predict the success rate of the advertisement.

In the time of digital media, all the companies around the world are investing lots of money for the advertisement purpose. From 2010 to 2019 we have seen that there is a drastically change in the investment on advertisement sector by the companies around the world. In 2010 the global advertisement investment was 399.26 billion U.S dollars and by the time of 2019 which increased to 563.02 U. S billion dollars[39].

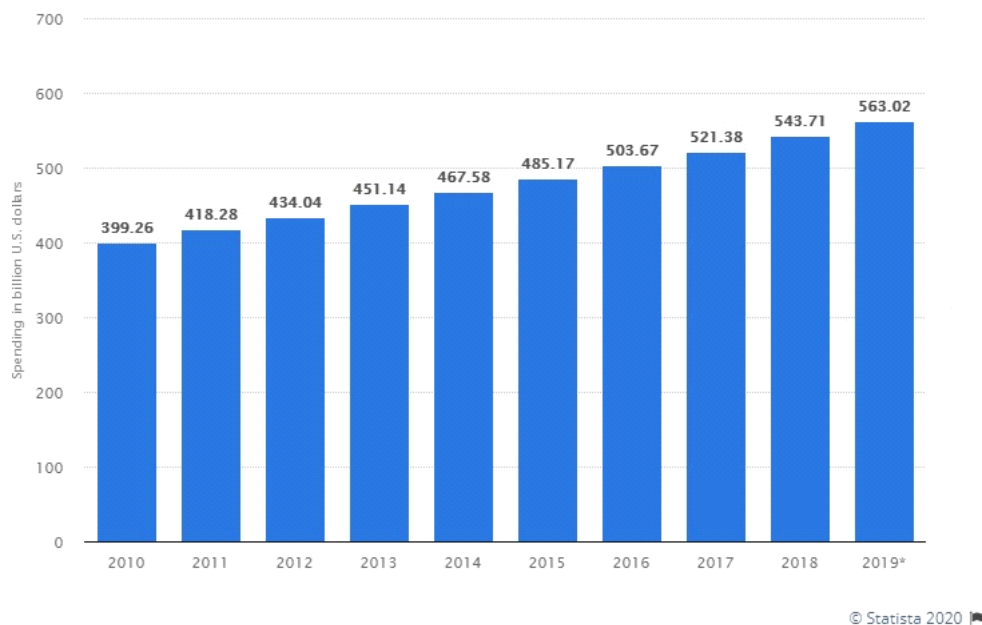


Figure 1.1: Global advertising spending from 2010 to 2019(in billion U.S. dollars)

1.3 Types of Advertisement

Every company spend a lot of money on their marketing for letting the consumers know about the uniqueness about the product, the importance about their product and for proper marketing having a perfect advertisement which has perfect amount of emotional appeal is mandatory. These advertisements are shown in different digital media. In different digital media we have observed that different types of advertisement are taking places. For example, display advertisements, social media

ads, newspaper, outdoor advertising, radio and podcasts, email marketing etc.

The main purpose of any advertisement is to get the attention of the consumer. An advertisement is successful when the consumers get satisfied by having some emotional feeling after watching the advertisement. Different companies are trying to get the attention of the consumers differently which means different advertisement tries to focus on different things according to their targeted customers. By this the companies can get proper attention from their targeted customers and may lead them to their expected profit.

As people now a days are more likely screen oriented, so the purchasing rate of a product depends on the successful advertisement which can properly attract the targeted customers. These advertisements are made for different digital medias as in Facebook, YouTube, television, radio etc. For different media advertisement should be different, because different media platform has different categories. For example, advertisements shown in Facebook and YouTube should be short more like few seconds because, these ads are shown in the middle of any video and people do not like any sort of interruption in their videos. On the other hand, Television advertisements are a bit longer than Facebook advertisements. Moreover, different advertisements try to focus on different things such as some can be sad, some can be happy, some can convey some social messages. However, successful advertisements have to have the proper amount of emotional content which will help to get the full attention of the targeted customers.

1.4 Problem Statement

As it is seen that companies around the world spend a lot of money on their marketing purpose, because it is the key to get success for every company and advertisements are the most important part of marketing. Companies spend a lot of money from their budgets on advertisement purpose. however, many companies can not reach their proper success because their advertisement can not get the proper attention of the customers. Many companies have to bear a huge loss at the end just because of their unsuccessful advertisement.

When companies try to launch a new product, they have to train the customers about their product. They have to let the customers know the importance of their product in their life. Every product come with a new idea, new features, new solution of a problem. So, they have to present it to the customers properly so that the customers can feel the importance of their product and purchase it. Though it is a matter of regret many companies are failed to present it properly to the customer and loose the attention of the customers about the product. As a result, they have to bear a huge loss every year. People always get to know a company by their marketing because people will not go to a company to know about the facilities they are providing, about the new product they are launching. So, if the companies can not show the identity of a product properly to the customers, they will not able to connect the customers and people will not purchase their product and as result now a days many companies are facing a huge loss every year in Bangladesh and many other countries.

Now a days it is the common scenario of many companies that they are facing a huge loss every year just for the lack of successful advertisements. The advertisements they are making for their marketing do not have the proper emotional appeal which may attracts the customers. They do not focus on this emotional appeal, rather than they make their advertisement on their product qualities or on their product specification. However, before making an advertisement a company should think about what kind of advertisements are attracting people and making people watch more than once. Moreover, whether that advertisement contains proper ratio of emotional content the consumers are searching. If the proper amount of emotional contents is missing in an advertisement that advertisement must not be successful and will not able to get the customers attention. As a result, the investment on the advertisement will go in vain and the company will bare the loss. This is why in our paper we are focusing on what type of advertisements people like, what type of emotional contents have to have in an advertisement to be a successful advertisement.

In our paper we will solve the problem of unsuccessful advertisement. First, we will identify the emotional contents people look for in an advertisement and by knowing that we will able to know the emotional ratio of an advertisement should have to attract the customers. By that we can able to let the companies know about how much emotional content should their advertisements have to have to making their advertisement successful.

1.5 Objective and Contribution

Every company has their own marketing policy to achieve their goal. previously traditional marketing policy was used for meeting the organization goal or success, but now this traditional marketing policy has been shifted to strategic marketing policy where contribution of technology is significant. With the help of technology any thing can be predictable now a days if enough data is provided. By using machine learning and deep learning algorithm anything can be predicted with the help of enough useful data. Moreover, many companies around the world using data analysis to predict the steps they should take for the success of their companies. We can see that many companies are investing a lot of money on their advertisement purpose. So, if they can predict the emotional states in the advertisement they are launching, it can make an impact on the customers mind or not, then they can easily reach their goal. That is why our main aim will be predicting the success rate of any advertisement companies are launching using audio frequency of the advertisement and monitoring the emotional appeal in that advertisement. Mainly, we will find out successful and unsuccessful advertisements as well as the accuracy scores of the emotional states/appeals in the successful advertisement. So, any advertise or marketing company can test their advertisement through our algorithm and we can find out the values of the emotional states in the advertisement. Also, we can tell if the advertisement will be successful or not as well as if the people will buy the product or not. Even if the advertisement is not successful one then how much emotional appeal will be needed in the advertisement that can be possible to find

out. Previously there are some works on advertisements, but basically, we are doing it by using audio frequency. We collected real life data by our own and compare with the people's reaction on several advertisements and by that we predicting the success rate by analysing some factors.

With the help of our collected advertisements and after doing the survey and audio extraction and combining both we have got a data set. However, we can not work with these data, because these data are not classified. To classify the data, we are using several machine learning and deep learning algorithms as in SVM, MLR, Naive Bayes, LSTM RNN. By using these algorithms, we have got the classified data by which we can work on. These algorithms show us different rate of purchasing the product and public impression towards it. Positive impression will accelerate the public feelings to buy a product.

According to the objectives we have just discussed, we must have some contributions to the companies around the world which will be done through our thesis. The key contributions of our thesis are as followed-

- Conducting a survey and getting datasets from the combination of survey/textual data and audio extraction data.
- Classifying data using some machine learning and deep learning algorithm as in SVM, MLR, Naive Bayes, LSTM RNN.
- Getting different rate of purchasing the product and public impression using the classified data.
- Getting the result of public feelings about buying a product from positive/negative impression.

1.6 Thesis Structure

Firstly, we are going to find the factors of an advertisement for which we are going to conduct a survey by showing them some advertisements we have chosen randomly. in this survey several question will be asked by which we will able to get the impression of the people about the advertisement. After that we will run some machine learning algorithms to classify the data we got from the survey. On the other had, we will use extract the audio files from the advertisements, because throughout our project we will be working on these extracted audio frequencies. For these extractions we are going to use different audio extraction methods. After that, we run some deep learning algorithm on these extracted audios.

It is clear that we are getting data from two sources, one from the survey and another one from the extracted audio. For working with these data, we have to combine these data from where we can get the main data set, we will be working on. Moreover, we will compare machine learning algorithm with deep learning algorithm to find the success standard. By using this success standard, we can predict the standard emotion rate. By these processes we can be able to tell what content and advertisement should have to be a successful advertisement by which the viewers will be pleased and get the motivation to purchase the product or service.

1.7 Work Plan

Before starting our research, we have made a work plan depending on the things we have to cover throughout our research, because without a proper plan we might be loose our work flow. For building a work plan we had to go through several papers to get some idea about how we are going complete the research on our topic. Moreover, we had to study on the marketing process of several companies and the things they are focusing before launching their advertisements in the market. As a result, keeping all these in our mind we had made a work plan which we have followed throughout our research.

To begin with, our workflow started with collecting some advertisements randomly by which we are going to conduct our overall research. After that, we used our collected advertisements into two ways. In first way, we conducted a survey with the video of the advertisement we chose earlier. We had shown these advertisements to the people and collected some data from them, with the help of some self prepared questions. We did the survey randomly, so that we can get the perfect data we are looking for. Moreover, by using those data we collected the purchase and liking rate of the people after seeing the advertisement. Therefore, we have collected the emotional statement according to the Arousal, Valance and Dominance which will help us to determine the emotional content an advertisement have. After that, using these data we used machine learning algorithm to classify the data and find the successful advertisements.

This was the first way of our workflow. On the other way, we extracted the audio from the advertisements, because we are using audio frequency for our thesis. After that, we extracted audio with different audio extraction features. By using these extracted audios, we used deep learning algorithm to find the classification of the data.

After completing the two ways of our work flow using the advertisement we had been collected, we moved to our final part. In the final part we compared machine learning algorithm which we got from the survey with deep learning algorithm which we got from the extracted audio. By comparing that we had found the success status of advertisements which will help us to find the successful advertisements. Finally, after finding the success standard we have predicted the standard emotion rate of a successful advertisement should have by which we can easily say that how much emotional content an advertisement should have to be an successful advertisement which will make an impact on the mind of people who will be seeing it and by following that companies will able to make the desired successful advertisement.

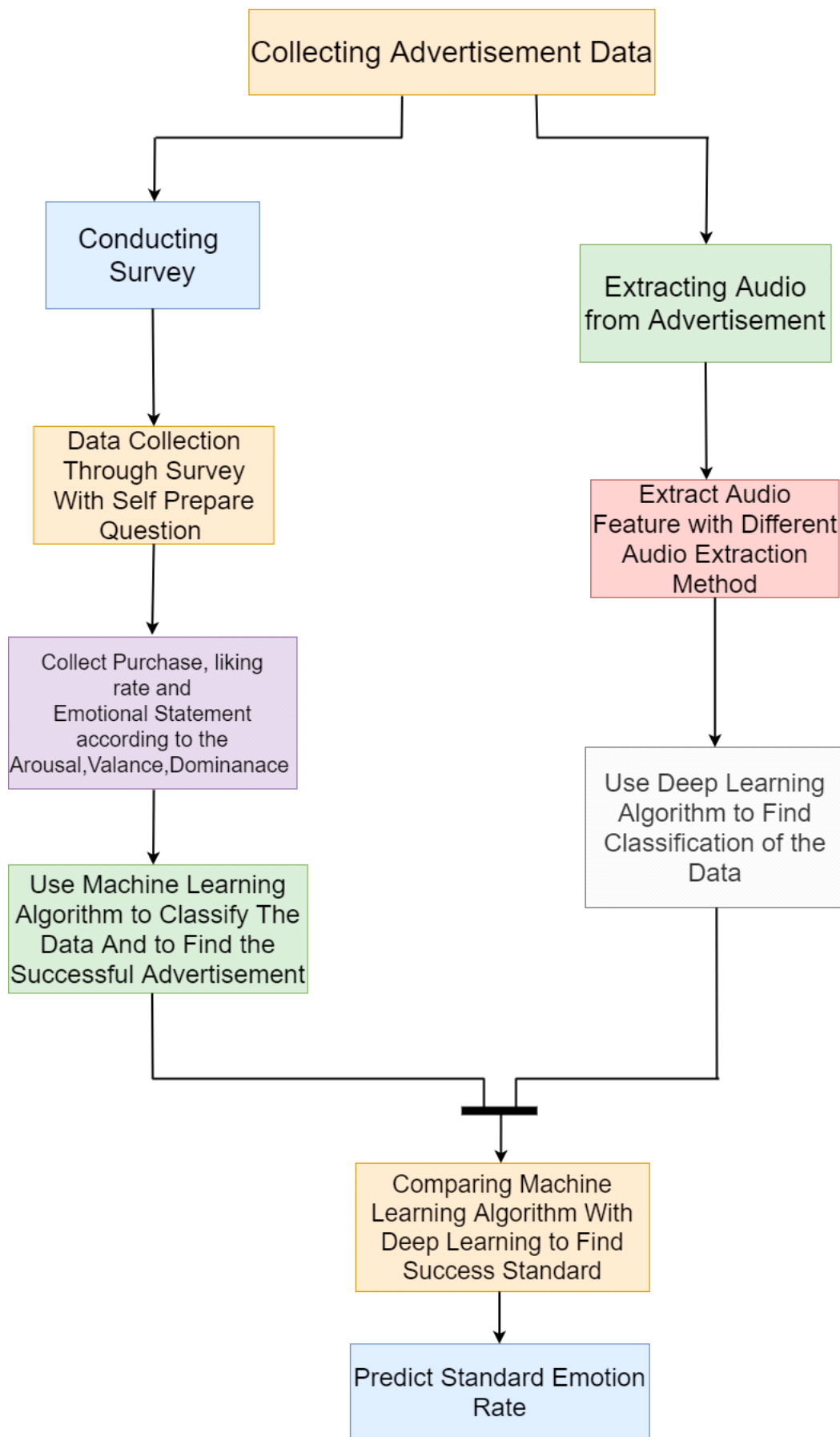


Figure 1.2: Workflow of the research

Chapter 2

Related Work

Anybody who wants to introduce their products in market then advertisement is the best way to promote their product. Through advertisement advertiser promote their product by applying different kind of emotional appeal. These emotional appeals set the marketing impacts of their product.

In this[24] paper they showed us, choosing the right advertising appeal like rational or emotional appeal becomes crucially important when creating effective advertising campaigns. According to this paper, consumers fix their preferences of the product on the basis of elements such as liking, feelings and emotions which are introduced by the advertisement rather than the product or brand attribute information.

In their[28]paper, Lee and Tashev found more efficient high-level features that are stable in terms of long-range contextual effects by implementing a recurrent neural network (RNN), a strong sequential-data learning model. In addition, they also introduced a new learning method for identification of speech feelings, resolving the complexity of emotional marks. All frames within the utterance are assigned to the same mark in standard algorithms. Because the marks are annotated for the whole utterance, however, all frames in the utterance do not inherently bear the same feeling and cannot be mapped to the same symbol. Therefore, taking the emotional state as a random variable is rational, and we suggest a corresponding learning algorithm that can evaluate the value of each frame internally using the expectation-maximization (EM) algorithm in conjunction with effective dynamic programming.

Different acoustic properties of speech are studied and some of the classifier techniques are analyzed in which it is useful to further study new methods of emotion detection. This article investigated the estimation of the next reactions from emotional vocal expressions based on emotional perception, using different classifier groups.

Lee and Tashev[28]suggested the recurrent neural network (RNN)- ELM model, which used the long emotional contextual effect. For several years in the field of speech emotion detection such models have been known as the state-of-the-art models. Nonetheless, people's understanding of speech emotion is limited[28]. Using priori information alone it is hard to remove abundant functionality. The acoustic

features are therefore not adequately reflective of emotional details.

In[33]this paper, it is said that human interaction is a temporally dynamic event which can be inferred from both audio and video feature sequences. On that paper they investigate the long short term memory recurrent neural network (LSTM-RNN) based encoding method for category emotion recognition in the video. From this paper it is understood that LSTM-RNN is able to integrate knowledge of how emotion emerges from discrete frames across a long spectrum of successive frames and emotional cues. LSTM-RNN is a kind of recurrent neural network which has the ability to learn long term dynamic while avoiding the vanishing and exploding gradient problems and this thing is said on that paper. They also used yaafe toolbox to extract the audio features.

This[22]paper, describe the neural network as a human brain in two aspects: getting knowledge from external environment through learning process and storing knowledge through connection strength among nerve cells. Moreover this paper, describe that RNN uses its internal memory to handle any input timing sequence, which make it easier to handle handwriting recognition and speech recognition. It also includes that LSTM is Long Short Term Memory, which can selectively store and discard the information in the hidden layer of neural networks. In other word, it can decide the information storage time in neuron. The advent of LSTM well pointed out the gradient appearing problem and encouraged Recurrent Neural Network growth.

This has been established in this paper that to represent, consider a sine wave of some fixed frequency then PSD plot will consolidate just a single ghostly part that is available at that chosen recurrence. In basic words, power range of whenever area signal $x(t)$, assists with deciding the dispersion of difference of information $x(t)$ over recurrence space in type of ghostly segments into which the genuine sign can be deteriorated[21]. It is widely used over the world for extracting features from audio files. Precise estimation of the PSD would be the key supporter of the achievement of signal partition. Indeed, some early investigations have just taken this approach and have prevailing with regards to assessing the PSD of confused signals, for example, amplifiers' inside clamor what's more, resonance[32]. Acoustic sound contains a lot of data that fluctuates in numerous various ways. It could be separation to receiver, change of acoustic scene, diverse account gear, and so on. The discrete time-signal is spoken to with its adequacy along tests in time. The sound is a case of two classes from the Urban Dictionary dataset. Crude signs can be testing contributions for a classifier; the vector is of high dimensionality and perceptually comparative signs are not really neighbors in the vector space[37].

In the paper[35], occasion spectrogram is mapped onto algorithmic spectrogram. Square astute spectrogram highlight extraction from sound files named squares are considered for include extraction. Distinctive new highlights, for example, focal minutes, highlights dependent on Singular Value Decomposition (SVD) and modify ghostly flux, root mean square (rms) vitality, Renyi entropy are figured and removed from each square. The vigor of these highlights is tried in various commotion conditions. Proposed spectrogram highlights show significant improvement over MFCCs, particularly during boisterous conditions.

This [20] paper, it has described that Support vector machine (SVM) has lot of outstanding ability, especially in classification problems. Its basic design philosophy is to maximize the classification boundaries and its basic purpose is to maximize the hyper-plane. This paper also included that SVM is a machine learning method based on statistical learning. It is based on the concept of the induction theory of Structural Risk Minimization (SRM) which aims to minimize a relation to the generalization error instead of minimizing the mean square error [2]. In a case that is linearly non-separable but nonlinear (better) separable, the SVM replaces the internal product: x, y by a kernel function $K(x;y)$, and then creates an optimal separating hyperplane in the mapped space [5]. In these papers [7][16] it is clearly notified that among many methods, a consensus seems to have been formed on the usage of Support Vector Machines (SVM) due to their versatility, computational performance, the capacity to manage high-dimensional data and the function selection profits [23].

It is said that feature extraction is an integral part of Automatic speech recognition system [28]. This paper also added that feature extraction is a technique to remove the changeability of the input speech signal while retaining the imperative characteristics of the speech. The speech signal is converted into useful parametric-representation, which can be further, analyzed and classified. Moreover, this paper describe the Mel Frequency Cepstral coefficient (MFCC) as a cepstral domain based analysis technique for speech recognition that matches the human auditory system and the merit of this feature extraction method is it is very efficient and accurate with low complexity. Mel-Frequency Cepstral Coefficients (MFCC) feature extraction methodology is a leading approach for speech feature extraction and current analysis aims to spot performance enhancements [14]. From the audio files that has been sliced victimization mfcc the total dataset are often ready. MFCC is an audio feature extraction technique that extracts speaker specific parameters from the speech [6]. It offers US some constant values that is terribly helpful. MFCC algorithmic program makes use of Mel-frequency filter bank at the side of many alternative signal process operations. Matrix of MFCC options obtained from our implementation of MFCC algorithmic program has range of rows up to range of input frames and it is employed in feature recognition stage [18]. Choice of options and size of the information plays vital role for recognition theme. The steps towards building of an emotion recognition system area unit, an emotional speech corpora is chosen or enforced then emotion specific options area unit extracted from those speeches and eventually a classification model is employed to acknowledge the emotions. The most challenge of feeling recognition from speech is that every speech is of various length, currently MFCC feature extraction methodology works in a very window methodology meaning it set a 25ms frame over the speech signal and calculate thirteen cepstral constant from every frame those area unit used as features [11]. What is more, it is aforesaid that feature extraction is an integral part of Automatic speech recognition system [28]. This paper conjointly said that feature extraction may be a technique to get rid of the changeableness of the input speech signal whereas retentive the imperative characteristics of the speech. The speech signal is born-again into helpful parametric- illustration, which might be additional, analyzed and classified. Moreover, this paper describe the Mel Frequency Cepstral constant (MFCC) as a cepstral domain based mostly analysis technique for speech

recognition that matches the human sensory system and therefore the advantage of this feature extraction methodology is it's terribly economical and correct with low complexness.

It is demonstrated that XGBoost classifier uses extreme gradientboosting[25]which has been demonstrated to be successful in a wide variety of tasks[29], ranging from suggesting jobs to supporting neural networks by balancing imports of features[34]. Now to understand XGBoost we will consider examples that will give a very clear view to all. In the paper[25]they concluded that the key factor behind XGBoost's performance is its scalability in all scenarios. In distributed or memory-limited environments the device runs more than ten times faster than current common implementations on a single computer and scales into billions of instances. XGBoost's scalability is attributed to many major programs and algorithmic optimizations.

This paper showed that many hierarchical algorithms use energy to make a decision, in addition to other metrics, but they decided to concentrate on energy-based techniques[13]. It is talking to the edge's all out force range[9]. Short-term energy is used in a variety of problems surrounding audio production. It provides the basis for distinguishing between spoken dialogue and unvoiced fragments of speech for speech patterns. In the case of very high quality recording, the properties of short-term energy are used to separate audio from the silence[17].

Regression models think about the yield as the after effect of a parametric capacity, with the highlights playing the job of factors[31]. In that paper they also discovered that Persuaded by ongoing examination in full of feeling demonstrating for music and content using regression models for full of feeling grouping of nonexclusive sound. In particular, they examined the utilization of Multiple Linear Regression (MLR) and Multiple Quadratic Regression without the collaboration terms (MQR). Multiple linear regression (MLR), often referred to simply as multiple regression, is a mathematical method that uses many explanatory variables to forecast the outcome of the answer variable[31]. The purpose of multiple regression toward the mean (MLR) is to model the causative relationship between the informative (independent) variables and therefore the response (dependent) variable.

In this paper[15]it has been stated that zero- rate (ZCR) is one of the most important acoustic features widely used in voice activity detection, voiced/unvoiced speech classification, image processing for speech classification, optics, biomedical engineering, radar and fluid mechanics. Naive Bayes is a collection of controlled strategies in machine learning, used for classification. The crux of this system of classification is the Bayes Theorem[19].

Chapter 3

Proposed Method

3.1 System Model

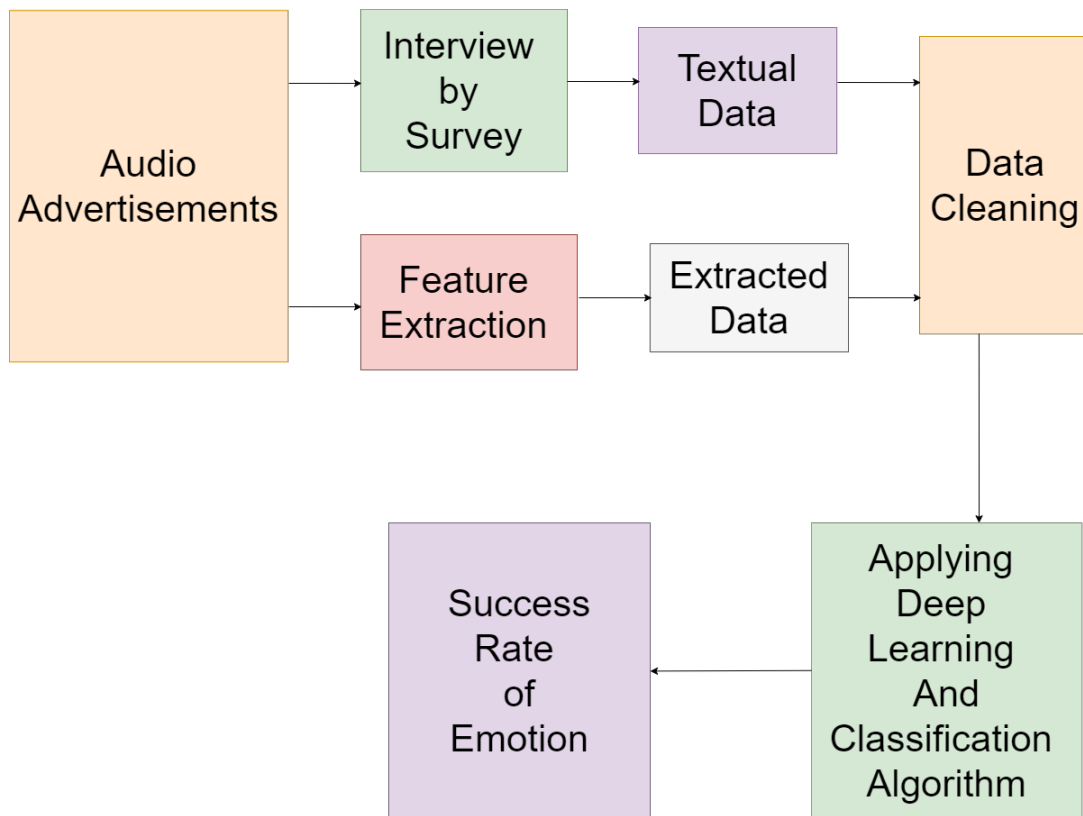


Figure 3.1: System Model of Finding Success Rate of Emotion

Figure 3.1 shows the methodology of our research. Firstly, we are conducting a survey for general people such as users and experts to determine successful advertisements. To clarify, successful advertisements mean those advertisements which have maximum emotional impact on the users that the users or customers will surely buy the products after watching the advertisements. The users are generally differentiated by their age, gender, income, how much time they spend in medias etc. Additionally, the experts are two of two parts. One is the experts of advertise companies and another is similar organizations. During the survey we collected 50 responses in total. Basically, during the survey we show advertisements videos to

users and experts and ask several questions to them to get their proper feedback. From the survey we get the dataset which is textual data where people can share their feedback towards any advertisement. In addition, from the dataset we can get the emotional statements according to Arousal, Valence, Dominance, Liking and Purchase. In detail, we used five classes for a specific question. For example, the question in Figure 3.2 has five classes which are 1 to 5 that tells how much excited

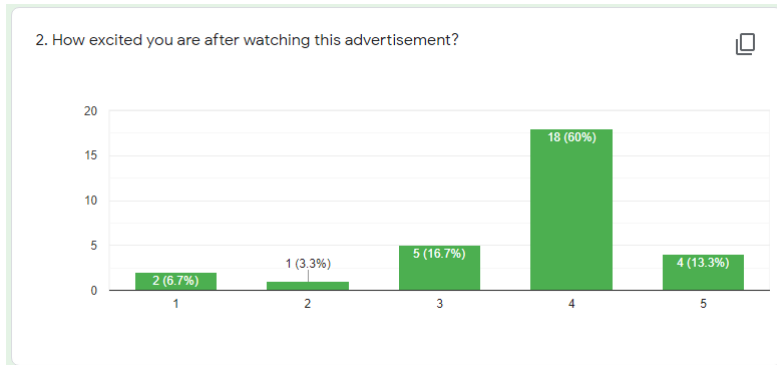


Figure 3.2: Survey Responses from Calmed to Excited (1 to 5)

a user after watching a particular advertisement. Figure 3.2 shows a specific response from the survey which tells that for a particular advertisement most of the users gave 4 in the feedback of Calmed to Excited. Therefore, we can find out the advertisements which are most successful and have maximum emotional impacts.

On the other hand, we get few more datasets from the Audio Feature Extraction Methods. Primarily, we get the WAV files from the advertisement audios. Since WAV files are more suitable for the Audio Feature Extraction Methods we did not use the MP3 files. Then we use several Audio Feature Extraction Methods on the 100 WAV files we got from the advertisements. In particular, we use MFCC (Mel-frequency cepstral coefficients), Short Time Energy, Zero Crossing Rate, Power Spectral Density and Spectrogram Analysis. From the Feature Extractions Methods we get extracted data which are coefficient values.

For the data cleaning, we generalize our dataset to get the proper result. Even the dataset we got from the survey we needed to generalize that also. To illustrate, in the survey we got feedback in five classes for example, from calmed to excited level we have five classes like 1 to 5.

2. How excited you are after watching this advertisement? *

1 2 3 4 5

Calmed ○ ○ ○ ○ ○ Excited

Figure 3.3: Classes in Survey Questions

We can see in Fig 3.3 that a user can give his feedback through five classes from 1 to 5 that how much he/she is calmed or excited after watching the advertisement. Fig 3.4 shows that how we generalize the dataset from five classes to three classes. Actually, to get better result we generalize the classes from five to three in the dataset.

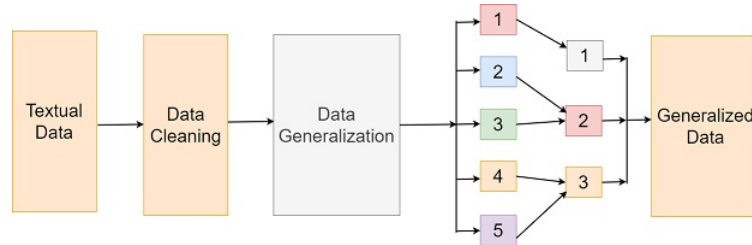


Figure 3.4: Data Cleaning (Data Generalization)

Now, since we got our datasets from Feature Extractions and Survey we use the datasets to implement Deep Learning to predict the success rates of different emotions. So, we implement LSTM-RNN (Long Short-Term Memory Recurrent Neural Network) using the datasets we got from Survey and Feature Extraction Methods. Actually, we use the generalized dataset we got from the Survey dataset to implement LSTM-RNN.

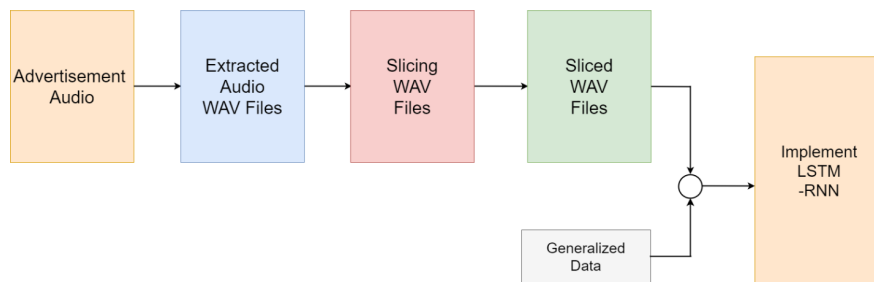


Figure 3.5: Implement LSTM-RNN

Also, Fig 3.5 shows that we slice the 100 WAV files to 1000 WAV files and use all the Feature Extraction Methods on the 1000 WAV files. Therefore, we get larger dataset and we use the larger dataset from the sliced audio WAV files to implement LSTM-RNN. So, we use the generalized dataset and larger dataset from sliced audios to implement LSTM-RNN. Finally, LSTM-RNN will predict the success rate of various emotions from the successful advertisements.

Similarly, we use one of the major classifying and predicting algorithms which is SVM (Support Vector Machines). Basically, SVM classify our datasets and tells us proper emotion rate in an individual advertisement. SVM do not use the larger dataset that we got from sliced audios but it uses the generalized datasets we got from the Survey. To illustrate, SVM takes all the emotional statements like Arousal, Valence, Dominance, Liking and Purchase as input and predict a particular one maybe Arousal or Valence or Dominance.

Moreover, alongside LSTM-RNN, SVM and other prediction and classification algorithm there are also other methods which have been implemented to interpret Performance Measurement.

To evaluate the performance of our research Precision, Recall and F1 Score have been implemented on the algorithm. Also, Confusion matrix was executed to visualize the balance of performance of the algorithms in the research.

Naive Bayes classifier has also been used to classify our dataset more specifically arousal, valence, dominance, purchase and liking. This classifier gives us the classified score of these emotional states which are related to arousal, valence and dominance. Moreover, it also gives us the result of purchase rate and liking rate which are related to purchase and liking of our dataset.

Another machine learning algorithm or classifier that has been used in our research is XGBoost. This classifier also gives us the classified score of different emotional states which are related to arousal, valence and dominance. Purchase rate and liking rate have also been classified by this Xgboost classifier which is actually indicating whether the people will buy the product or not after watching the advertisements.

Furthermore, we implement MLR (Multiple Linear Regression) to get accurate score of success rates of emotion. MLR, also known as Multiple Regression uses explanatory variables to predict a responsible variable. Since we are using more than one explanatory variables, we implement Multiple Linear Regression. Here, Arousal, Valence and Dominance are our explanatory variables and we predict either Liking or Purchase.

Moreover, alongside LSTM-RNN, SVM and other prediction algorithm there are other methods which have been implemented on the datasets to interpret Performance Measurement. Precision, Recall and F1 Score have been implemented to figure out and evaluate the performance of our research. These also help to understand the Confusion Matrix of the datasets.

Finally, we compare the predicted scores we found from LSTM-RNN, accuracy scores from SVM as well as variance scores from MLR and we get the maximum score which is our final success rate of emotion.

3.2 Feature Extraction

We had to work with audio data to get the emotions and that is why feature extraction from the audio was important. To extract the audio feature, we use five different audio extraction methods. We have used zero crossing rate, power spectral density, Mel frequency cepstral coefficient(MFCC), energy and spectrogram. To do these feature extractions first we had to extract the data from our collected advertisement. After extracting the audios from advertisement, we have used the .wav file of those audios to do these feature extractions.

3.2.1 MFCC

Firstly, we have tried to work with the same duration audio file to maintain a standard and that is why we have cut down each audio file into sixty seconds. Then we have performed the Mel frequency cepstral coefficient (MFCC) feature extraction on these sixty seconds audio files. From Mel frequency cepstral coefficient (MFCC) we have tried to get the coefficient values. To do this operation, first we had to check the shape of Mel frequency cepstral coefficient (MFCC) for our audio files. We have got the shape of (20, 2584) for each audio file from Mel frequency cepstral coefficient (MFCC). The shape we have got that means we have twenty coefficients and for each coefficient we have two thousand five hundred eighty-four columns. As we took the same duration audio files to do this operation that is why we have got the same shape for each of our audio file. Here as we are getting twenty coefficients and each of this coefficient contain two thousand five hundred eighty-four columns. As we are getting huge number of columns therefore, we have taken the mean value for each coefficient. Thus, we have got twenty coefficient values for each of our audio file.

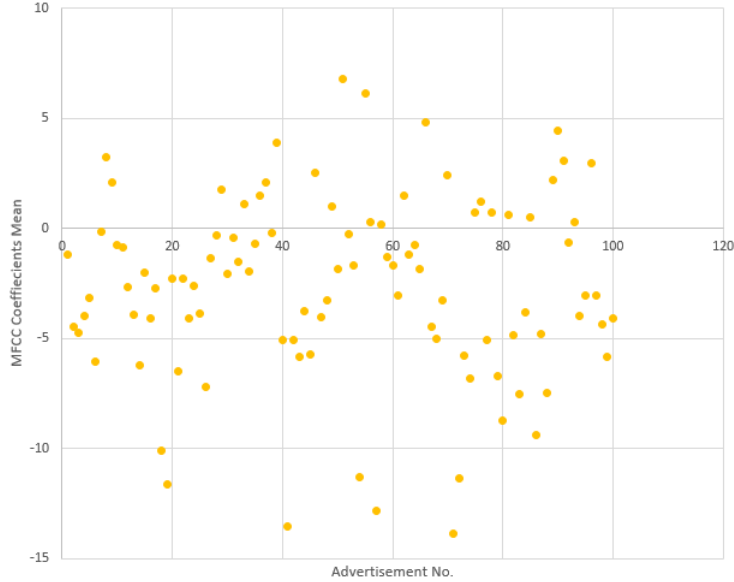


Figure 3.6: Scatter plot of MFCC coefficients value

In the figure 3.6 and 3.7, the MFCC coefficients values of advertisements audios are showed. For each advertisement's audio there are 20 coefficient values. We have taken all the audio of same duration therefore for every MFCC feature extraction there are 20 coefficients for every audio. To compute MFCC –

$$nMFCC = \frac{audiolength \times samplingrate}{windows} \approx \frac{audiolength \times samplingrate}{hop} \quad (3.1)$$

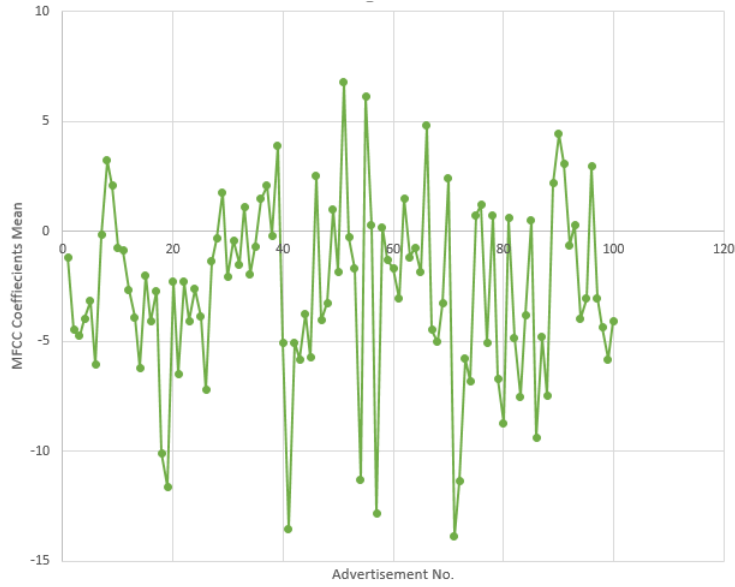


Figure 3.7: Line Plot of MFCC coefficients value

3.2.2 Short-time Energy

Short-time energy is another feature extraction method that we have used to extract data from our audios. Basically Short-time energy of a signal represents the total magnitude of a signal and for audio signals it represents how loud the signal is. For Short-time energy feature extraction first we had to cut down our audio files into sixty seconds. After that we load our audio files by using the .wav extension of the audio files. Then we have performed the energy feature extraction on those sixty seconds audio files. We have tried to get the sample rate of our audio files to do this feature extraction method. We have got the sample rate of 22050 for each of the audio files of our advertisement. Then we use the hop length and frame length of our audio to get the value of energy from our audio files. To[8]calculate Short Time Energy –

$$STE = \frac{1}{T-1} \sum_{t=1}^{T-1} (|St| \times |St|) \quad (3.2)$$

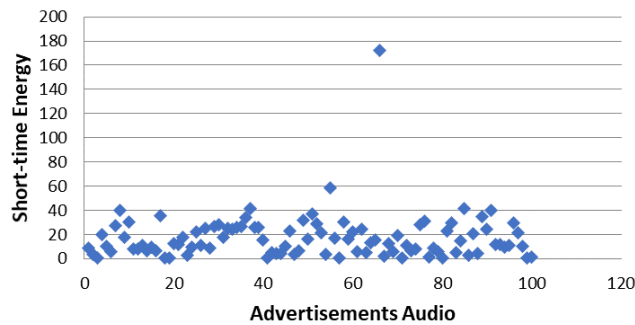


Figure 3.8: Scatter Plot of Short-time Energy

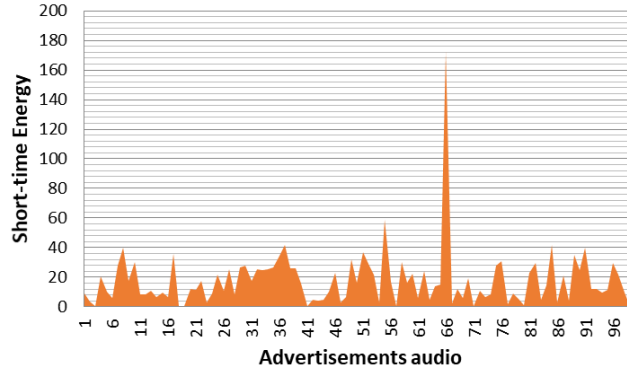


Figure 3.9: Scatter Area Plot of Short-time Energy

In the figure 3.8 And 3.9, we have showed the short-time energy values of our audio files.

3.2.3 Zero Crossing Rate

We have also used zero crossing rate feature extraction method for our audio files to extract data. Actually, how many times a signal crosses the horizontal axis that is represented by the zero-crossing rate. The zero crossover rate implies a shift in the performance of the audio frame from a positive value to a negative value, and vice versa. It gives a positive value when there is a lot of noise in the audio signal and gives a negative value when the audio frame is silent. It is defined as the number of time-domain zero-crossings within a given area of signal, divided by that region's number of samples[4]. A zero convergence is said to exist in the case of discrete-time signals, whenever consecutive samples have separate algebraic signatures. The rate at which zero crossings occur is a simple measure of the signal's frequency material. Zero-crossing rate measures of the occurrence of times that the value of speech signals passes in a given time interval/frame. Speech signals are broadband signals, and thus analysis of the average zero-crossing rate is much less accurate[12]. We have computed the presence of zero crossing rate for each of our audio file. That means we have computed how many times the signal of an audio file has changed the horizontal axis. For this process we have cut down the audio files of our advertisements into sixty seconds. We have computed the zero-crossing rate for each of this sixty second audio files. Therefore, from this feature extraction we are getting the presence of zero crossing from zero second to sixty seconds of each audio sample. We have summed up this each second zero crossing up to sixty seconds for each audio file. Thus, we are extracting data from our audio by this zero-crossing rate feature extraction method. To[30]Calculate Zero Crossing Rate –

$$ZCR = \frac{1}{T-1} \sum_{t=1}^{T-1} F(S_t \times (S_t - 1) < 0) \quad (3.3)$$

where, S is a signal of length T.

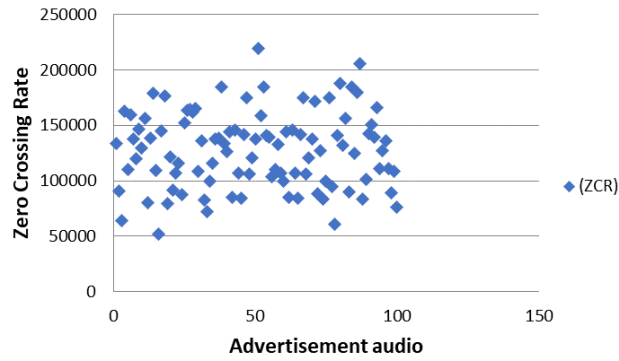


Figure 3.10: Scatter Plot of Zero Crossing rate

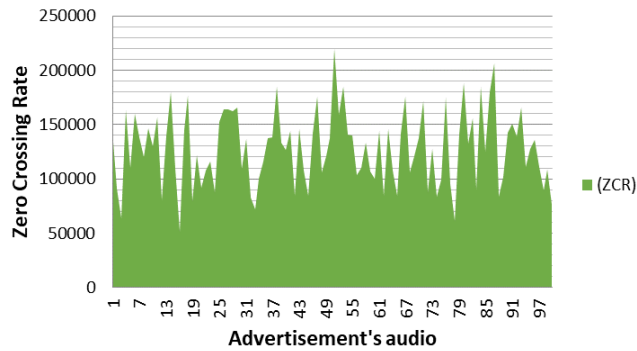


Figure 3.11: Area plot of Zero Crossing Rate

3.2.4 Power Spectral Density

Power spectral density (PSD) is another feature extraction method that has been used to extract feature from audio files. Power spectral density (PSD) stands for the measurement of signal's power content on the basis of frequency. Power spectral density (PSD) is typically used to characterize the signals. To run the power spectral density (PSD) features extraction first thing that has to be done is cut down the audio files into sixty seconds of the advertisements. The cut down of audio files is needed to be done because we need the same duration for each audio file so that we can get same power spectral density (PSD) shape for our audio files. After that the mean has done of power spectral density (PSD) as because the power spectral density (PSD) has given the value of (129) that means for each audio power spectral density (PSD) has one hundred twenty-nine values which has to be reduced. Therefore, to reduce the value of power spectral density (PSD) the mean has taken place so that we can get one power spectral density (PSD) value for each of our audio file.

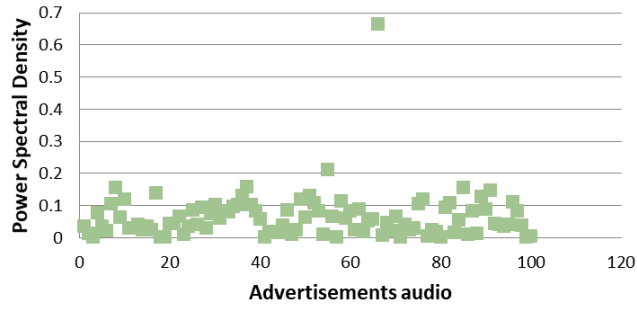


Figure 3.12: Scatter Plot of Power Spectral Density

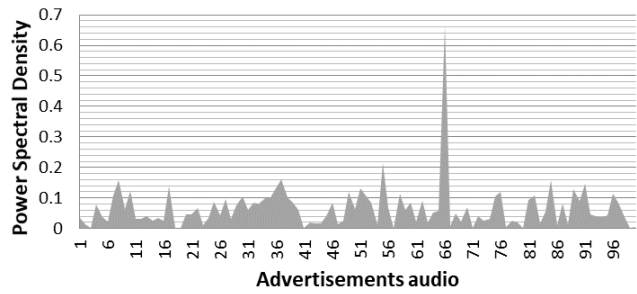


Figure 3.13: Area Plot of Power Spectral Density

3.2.5 Spectrogram

Spectrogram feature extraction method has also been used to extract the feature from our advertisement audios. Spectrogram visually represents the spectrum of an audio frequency. This spectrogram can be varying with time. The first thing that has been needed to cut down the advertisement audios into sixty seconds. After doing that the next step is to take the signal of our audio to run this feature extraction method. As all of our audio has same duration therefore all of our audio will have same spectrogram shape. The shape of each audio file for spectrogram is (129, 5906). Therefore, for each audio there will huge number of spectrogram value. As we needed one spectrogram value for each of our audio file that is why the mean operation of spectrogram was needed. Thus, there is one spectrogram value for each advertisement audio file.

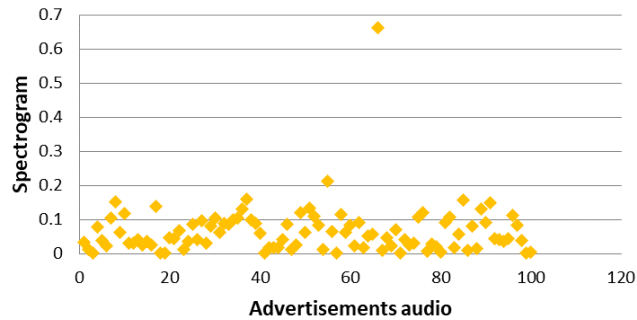


Figure 3.14: Scatter Plot of Spectrogram

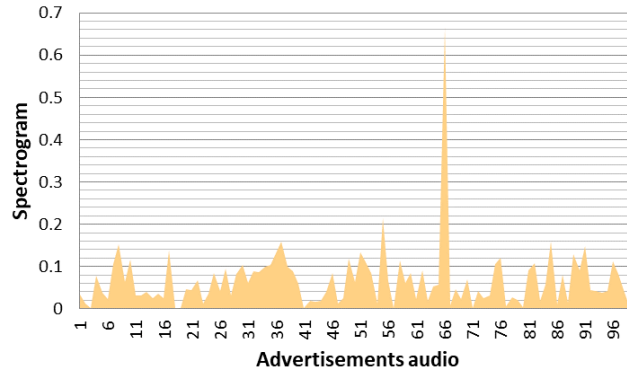


Figure 3.15: Area Plot of Spectrogram

3.3 Dataset Description

3.3.1 Questionnaire Description

First of all, we conduct a survey to get proper dataset to implement different algorithms. From the survey we collect answers and feedback from various types of people. So, from the survey we get one dataset consisting of different emotion statements like as Arousal, Valence, Dominance, etc. Also, we got another dataset from the audio WAV files. Since we implement various Audio Feature Extraction Methods on the WAV files we got a proper dataset.

During the Survey we asked few questions to the users for every advertisement. Before asking the questions we showed them the Advertisements videos.



Figure 3.16: Image from an Advertisement

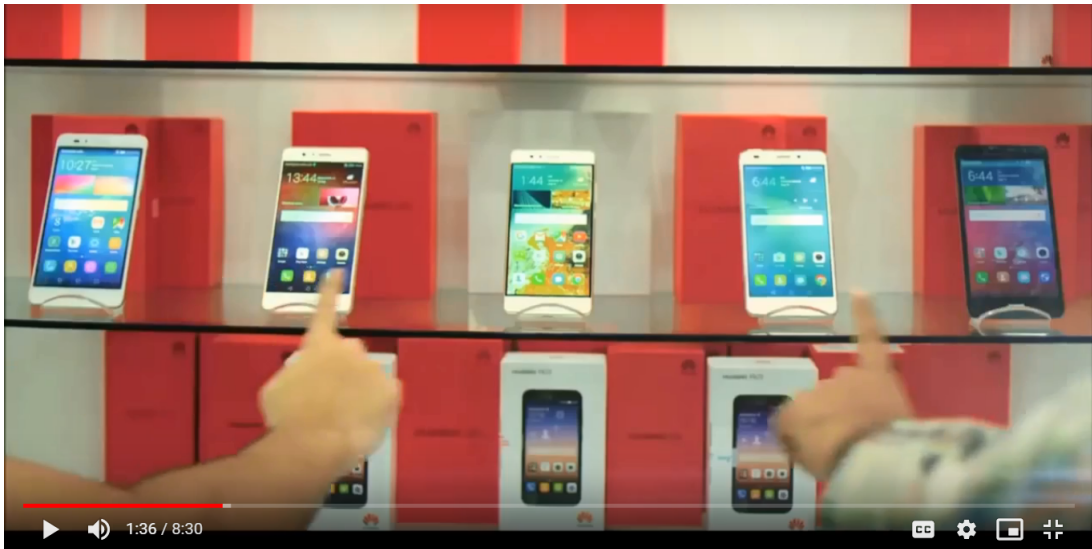


Figure 3.17: Image from an Advertisement



Figure 3.18: Image from an Advertisement

Fig 3.16, 3.17 and 3.18 shows the images from few advertisements we showed to users before the survey.

Advertisement 30 ✕ ⋮

Watch advertisements and read the question carefully and chose your answer carefully

1. If you give the scale of 1 to 5 then how much rating you will give this advertise on the basis of your feelings? *

1 2 3 4 5
 unpleasant pleasant

2. How excited you are after watching this advertisement? *

1 2 3 4 5
 Calmed Excited

3. Are you motivated enough to buy the product after watching the advertisement? *

1 2 3 4 5
 Low Motivated Highly Motivated

4. Do you like the advertisement or not? *

1 2 3 4 5
 Unlike Like

5. Will you buy the product or not? *

1 2 3
 No Yes

Figure 3.19: Five Questions of an Advertisement

After showing them the advertisements, we asked them few questions for every advertisement. But the questions were same for every advertisement. To illustrate, we asked various types pf questions to know how they feel after watching the advertisements.

1. If you give the scale of 1 to 5 then how much rating you will give this advertise on the basis of your feelings? *

1 2 3 4 5
 unpleasant pleasant

Figure 3.20: First Question from an Advertisement

Figure 3.20 shows a question where participants were asked to scale about how they felt about the advertise on the basis of Unpleasant to Pleasant level. We set up five classes on the basis of Unpleasant to Pleasant level. This is the primary question because if the participant feels pleasant about the advertisement, they will think to buy the product. They will not only have a positive impression on the advertisement but also will have a better perception about the product.

⋮

2. How excited you are after watching this advertisement? *

1 2 3 4 5

Calmed ○ ○ ○ ○ ○ Excited

Figure 3.21: Second Question from an Advertisement

In the Fig 3.21 we can see a question where participants were asked to give their feedback about how excited they were while watching the advertisement. In detail, the feedback is made on the basis of their excitement level from 1 to 5. So, if they are calmed that means they give the feedback 1 or 2, again if they are excited then they give feedback 4 or 5 and if they are not sure calmed or excited what they feel then they give their feedback to 3. Therefore, from these five classes of feedbacks we get our dataset and get the proper view of how much a participant want to purchase the product or not.

⋮

3. Are you motivated enough to buy the product after watching the advertisement? *

1 2 3 4 5

Low Motivated ○ ○ ○ ○ ○ Highly Motivated

Figure 3.22: Third Question from an Advertisement

In the Fig 3.22 we can see another question where participants were asked if they feel highly motivated or not while watching the advertisement. To clarify, the participants are giving their feedbacks if they feel enough motivated to purchase the product after watching the advertisement or not. In detail, the feedback is made on the basis of their being motivated or not on the level from 1 to 5. So, if they are not enough motivated to purchase the product that means they give the feedback 1 or 2, again if they are highly motivated to purchase the product then they give feedback 4 or 5 and if they are not sure that they feel motivated or not then they give their feedback to 3. Therefore, from these five classes of feedbacks we get our dataset and we will get the absolute view if a participant wants to purchase the product or not.

4. Do you like the advertisement or not? *

1 2 3 4 5

Unlike Like

Figure 3.23: Fourth Question from an Advertisement

Figure 3.23 shows a question where participants were asked if they like the advertisement enough to purchase the product or not. In detail, the feedback is made on the basis of their feeling of liking or not on the level from 1 to 5. So, if they do not like the advertisement to purchase the product that means they give the feedback 1 or 2, again if they like the advertisement enough to purchase the product then they give feedback 4 or 5 and if they are not sure that they like the advertisement or not then they give their feedback to 3. Therefore, from these five classes of feedbacks we get our dataset and get the absolute view if a participant wants to purchase the product or not.

5. Will you buy the product or not? *

1 2 3

No Yes

Figure 3.24: Fifth Question from an Advertisement

Figure 3.24 shows a question where participants were asked if they really want to purchase the product after watching the advertisement or not. So, if they do not like the advertisement and do not want to buy the product that means they give the feedback 1 or 2, again if they like the advertisement and want to purchase the product then they give feedback 4 or 5 and if they are not sure that if they want to purchase the product or not then they give their feedback to 3. Therefore, from these five classes of feedbacks we get our dataset and get the absolute view if a participant wants to purchase the product or not.

On the other hand, except the Survey another dataset was taken from the Audio Feature Extractions. Basically, the Audio Feature Extraction Methods were implemented on the audio WAV files and the audio WAV files were taken from the Advertisements which were selected for the Survey.

3.3.2 Textual Dataset

For the textual dataset, the questionnaire was actually divided in three parts and they are Basic information of the participants over whom the survey was conducted, Advertisement details and SAM scale survey. Also, all the data are collected from specific question and are kept secured. Actually, from the primary survey the textual dataset and participant's reaction were collected but the emotional responses were not found from textual data. The emotional responses were found from the classification and prediction algorithms which were used on the textual dataset.

Basic Information of the Participants

Generally, the basic information which are necessary for the research are age, gender, salary, the medium of watching advertisements they preferred and the overall time they watch advertisements in the preferred mediums. In addition, from the Survey the response was collected from 50 people. They are from different genders, different age, they use different mediums to watch advertisements etc. Also, the time for watching advertisements for all of them do not match. These basic information helps to separate them according to groups. Furthermore, the information is collected through Google Form.

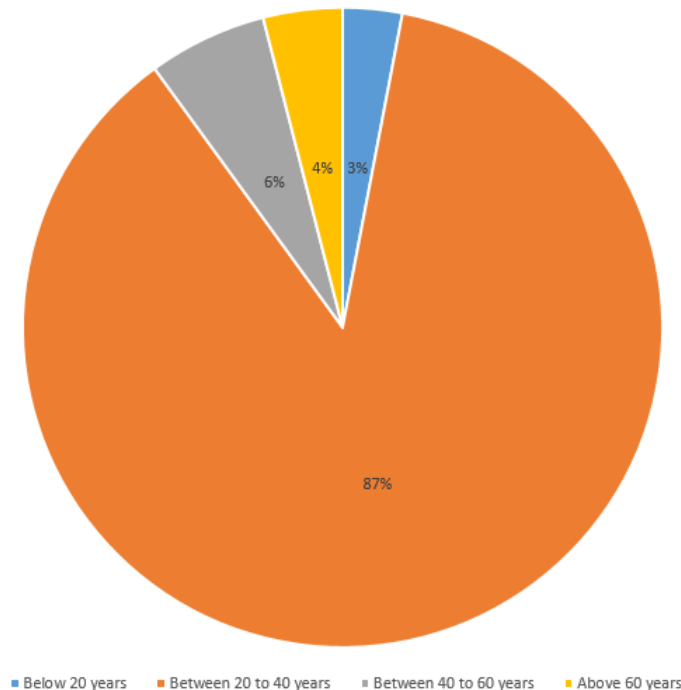


Figure 3.25: Age Classification Among Participants

Fig 3.25 shows that 87% participants are those who are between 20 and 40 years old. Also, 6% are those who are between 40 and 60 years old as well as 4% are above 60 years old. Moreover, 2% are below 20 years old.

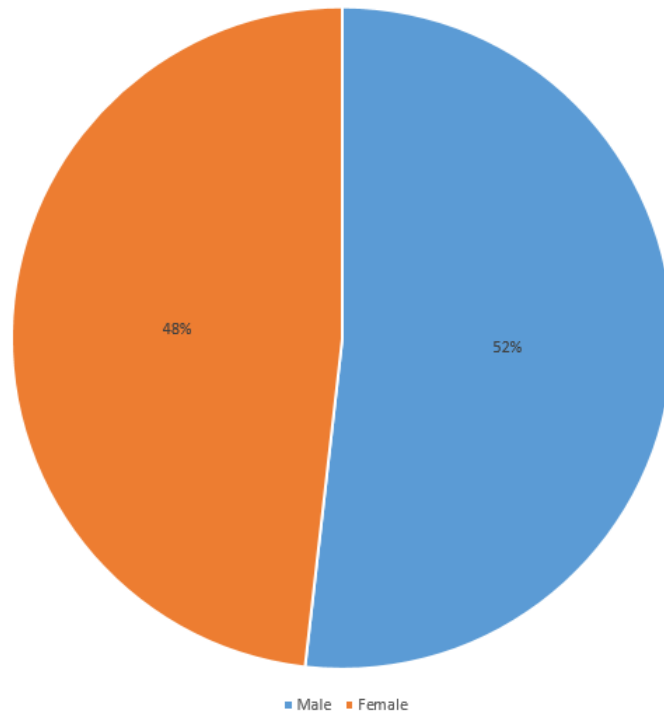


Figure 3.26: Gender Classification Among Participants

Fig 3.26 shows that 52% participants are male and rest 48% are female.

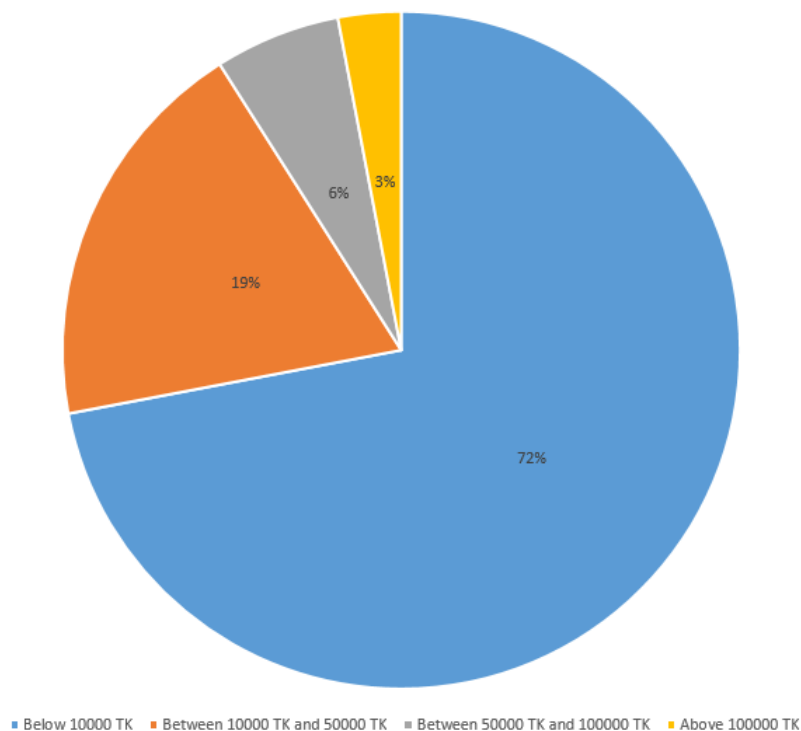


Figure 3.27: Participant Classification According to Salary

Fig 3.27 shows that 72% participants get salary below 10000 TK and 19% are them who get salary between 10000 TK and 50000 TK and. Similarly, 6% participants earn between 50000 TK and 100000 TK. Also, 3% get salary above 100000 TK

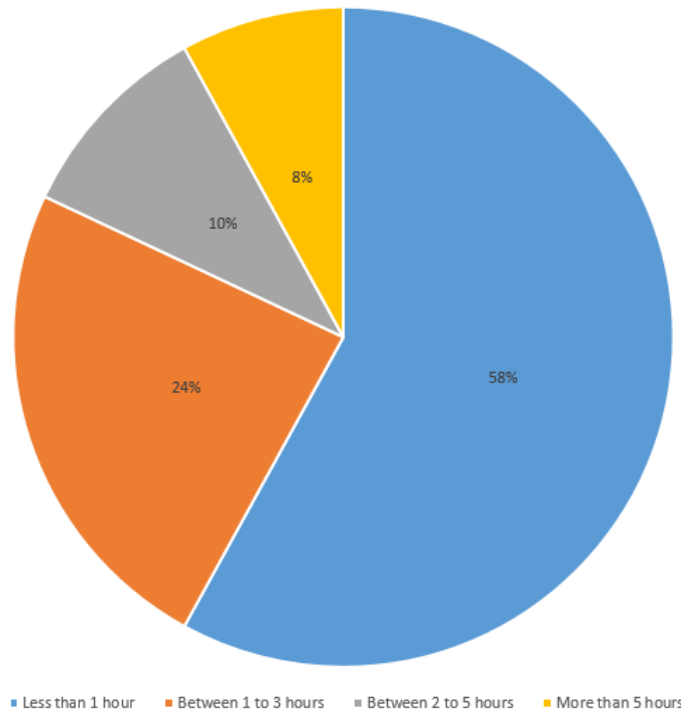


Figure 3.28: Time Duration Spent for Advertisement

Fig 3.28 tells that 58% participants spend less than 1 hour in watching advertisements and 24% of them spend between 1 to 3 hours in watching advertisements. Also, 10% of them spend between 2 to 5 hours and 8% of them spend more than 5 hours in watching advertisements.

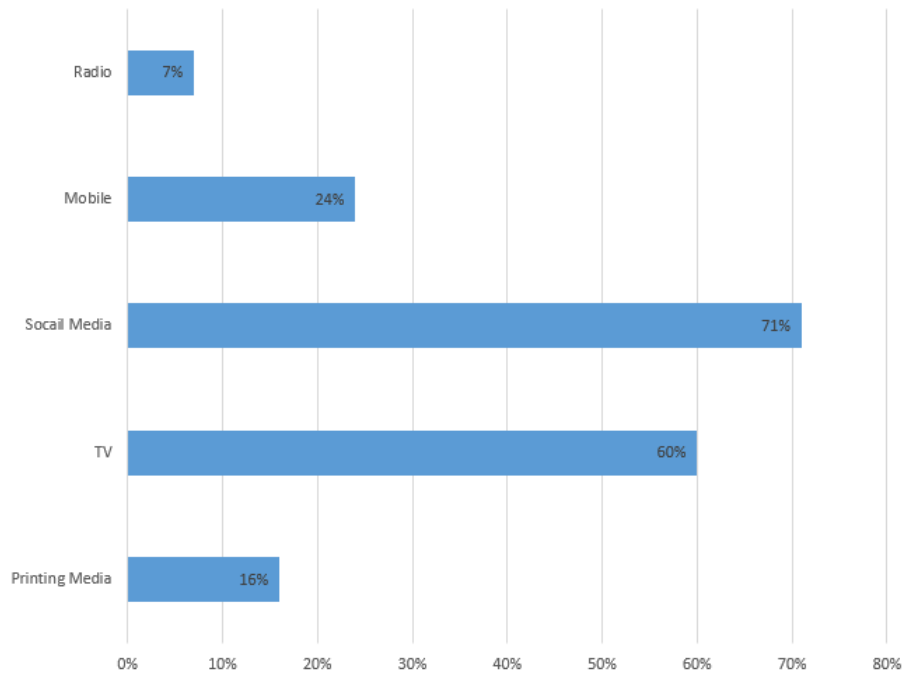


Figure 3.29: Most Used Medias for Watching Advertisements

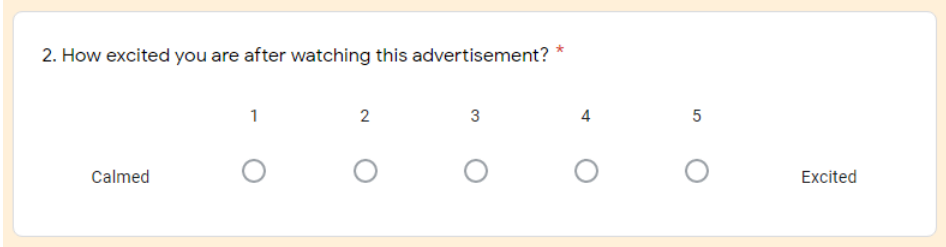
Fig 3.29 refers that 71% participants watch advertisements through social medias and 60% of them watch advertisements through Television. Moreover, 24% of them watch advertisements while using their mobile phones. Also, 16 % watch advertisements through printing media and 7% listen to audio advertisements in Radio.

Advertisement Information

The advertisement information is for identifying the features which are necessary to find out which advertisements are successful and which are not. Generally, the participants had to watch an advertisement and they had to answer the questions. According to the people's reviews the advertisements were selected and also both popular and common advertisements were in the list. For analyzing the emotional statements in the advertisements, the questions were classified in three scales. Arousal, Valence and Dominance. Furthermore, along with these scales Liking and Purchase intent are analyzed too. Every question is answered on the basis of five classes 1 to 5. Strongly Disagree, Disagree, Neutral, agree and Strongly Agree. Arousal, Valence, Dominance, Liking and Purchase get analyzed according to these questions.

I. Arousal

Arousal is basically the measurement of calm and excitement. The participants are asked in our questionnaire that the advertisements are relevant or not and the way of selling product as well as the message of the advertisements are believable or not. The response is classified in five sections. Strongly disagree, disagree, neutral, agree and strongly agree.



2. How excited you are after watching this advertisement? *

1 2 3 4 5

Calmed ○ ○ ○ ○ ○ Excited

Figure 3.30: Question Based on Arousal

II. Valence

Valence is mainly positive or negative effects in any situation. It can be pleasant situation or unpleasant maybe. In the Survey a question was asked to measure the valence.

1. If you give the scale of 1 to 5 then how much rating you will give this advertise on the basis of your feelings? *

1 2 3 4 5

unpleasant pleasant

Figure 3.31: Question Based on Valence

III. Dominance

The dominant nature of the emotion or the controlling of the emotion is called Dominance. Emotion can be classified in two types. Submissive and Dominant. For instance, fear is known as submissive emotion whereas anger is dominant emotion.

3. Are you motivated enough to buy the product after watching the advertisement? *

1 2 3 4 5

Low Motivated Highly Motivated

Figure 3.32: FQuestion Based on Dominance

IV. Liking

Here, liking is the personal liking intent of the participants. To illustrate, after watching the advertisement if they like the advertisement or not. This information helps us to find out if the participant or user will purchase the product or not. Also, this can tell if the advertisement is successful or not.

4. Do you like the advertisement or not? *

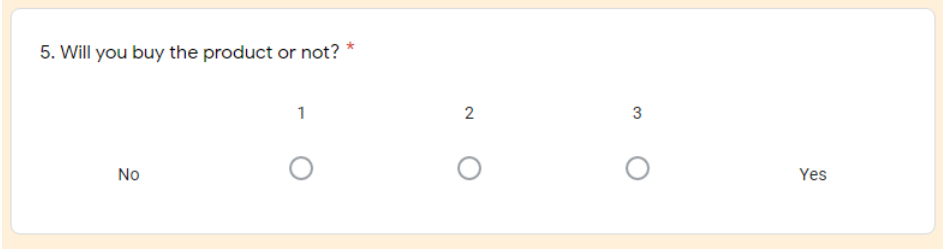
1 2 3 4 5

Unlike Like

Figure 3.33: Question Based on Liking

V. Purchase

Here, purchase is the participant's or user's will to buy the product or not after watching the advertisement.



5. Will you buy the product or not? *

No 1 2 3 Yes

Figure 3.34: Question Based on Purchase

SAM Scale Survey

Basically, the SAM scale survey is to measure the excitement, pleasure and motivating level of the information.

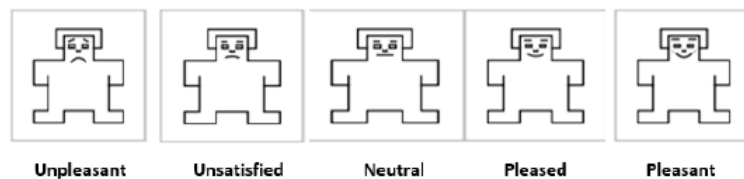


Figure 3.35: SAM scale of Pleasure level

Generally, pleasure level of any information can be classified from Unpleasant to Pleasant. Also, there are Unsatisfied, Neutral and Pleased. People may purchase any product if the advertisement seem to them pleasant. Otherwise, if they are unsatisfied, they may not achieve any positive attraction from the product. Therefore, this product will not bring any profit for the organization.

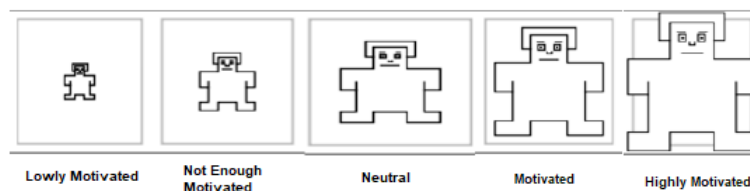


Figure 3.36: SAM scale of Motivating level

Motivating level is the measurement of how people are being motivated to buy any product after watching an advertisement. Usually, people buy products for their own benefits. So, if they are not enough motivated to buy the product after watching the advertisement this product will not bring any profit to the organization. Generally,

motivating level of any information can be classified from Lowly Motivated to Highly Motivated. Also, there are Not Enough Motivated, Neutral and Motivated.

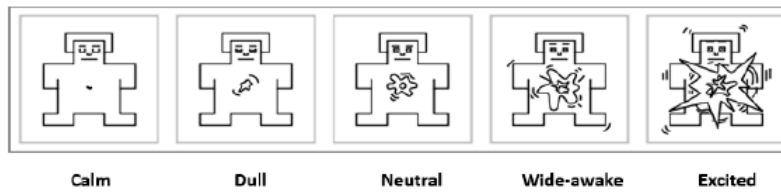


Figure 3.37: SAM scale of Excitement level

The excitement level of any advertisement also helps to evaluate the purchase rate of buying any product or what kind of impression is created by the product in the customer’s mind after seeing the advertisement. Here, calm is known as lower excitement and excited is known as higher excitement.

Fig 3.35, 3.36 and 3.37 shows different types of SAM scale. This helps to identify the pleasure, excitement and motivating level properly which can predict the successful advertisements.

Response from Survey

Primarily, the Survey was conducted and we found a dataset from the Survey which is the Textual Dataset. In a calm environment without any background noise the participants watched the advertisements first then they started the Survey and answered few questions for every advertisement. Also, they gave their proper feedback in the survey for instance, if they like the advertisement or not.

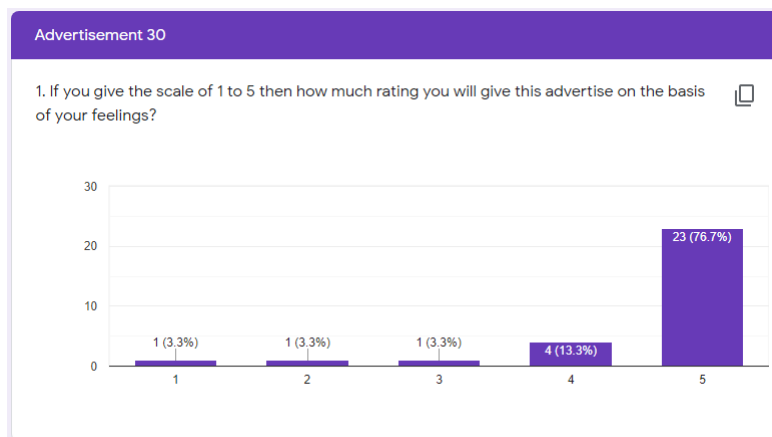


Figure 3.38: Survey Response for Valence

Here, in Fig 3.38 we can see a response of a particular survey question. The participants were asked if they can give the advertisement their feedback on the basis of pleasure then in the scale of 1 to 5 how much rating they want to give. This

question was for analyzing Valence. So, 76.7% of them gave 5 in the rating. To illustrate, they pretty much pleasant about the advertisement.

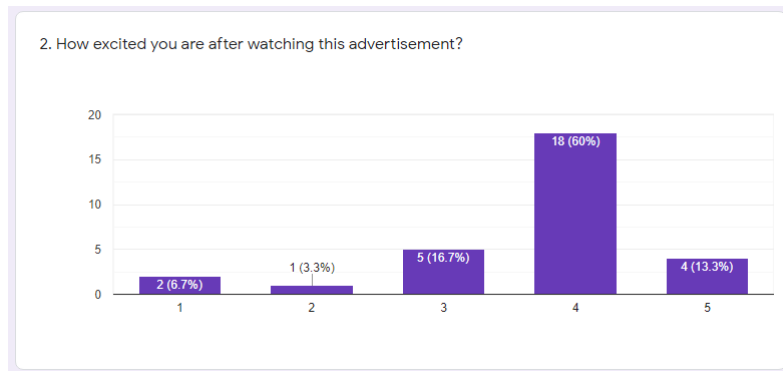


Figure 3.39: Survey Response for Arousal

Here, in Fig 3.39 we can see a response of a particular survey question. The participants were asked if they were excited or calm while watching the advertisements. This question was for analyzing Arousal. So, 60% of them gave 4 in the rating. To illustrate, they were pretty much wide-awake while watching the advertisement.



Figure 3.40: Survey Response for Dominance

Here, in Fig 3.40 we can see a response of a particular survey question. The participants were asked if they were highly motivated or not while watching the advertisements. This question was for analyzing Dominance. So, 66.7% of them gave 4 in the rating. To illustrate, they were motivated while watching the advertisement.

Fig 3.38, 3.39 and 3.40 shows the responses from the survey and from these responses we got the Textual Dataset. In textual dataset there are different values from the emotional statements. For instance, 1 to 5 scaling values found for Arousal, Valence, Dominance and Liking as well as Purchase.

After scaling 1 to 5 we generalized the dataset values from 1 to 3 to properly implement classifying and prediction algorithms later.

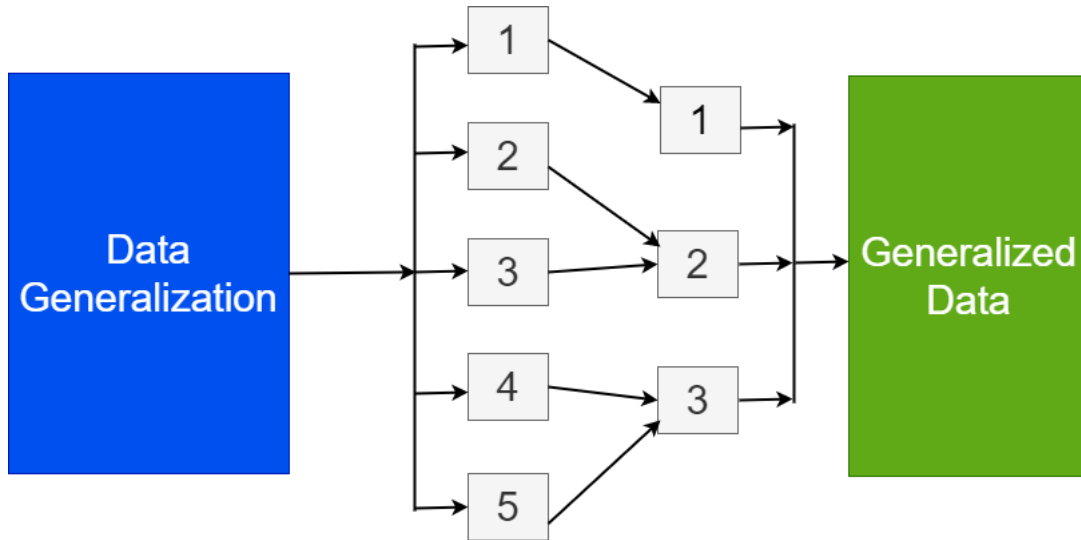


Figure 3.41: Data Generalization

In Fig 3.41 we can see the dataset was in scaled 1 to 5 value and after generalizing the dataset contains 1 to 3 value

Audio	1	2	3	4	5	6	7	8	9
Arousal	3	3	3	3	3	3	2	3	3

Table 3.1: Arousal Table

Audio	1	2	3	4	5	6	7	8	9
Arousal	3	3	3	3	3	3	2	3	3

Table 3.2: Valance Table

3.3.3 Feature Extracted Dataset

Alongside the textual dataset we also got another dataset from the Audio WAV files. Firstly, we implemented Audio Feature Extraction Methods on the Audio WAV files. MFCC (Mel-frequency cepstral coefficients), Short Time Energy, Zero Crossing Rate, Power Spectral Density and Spectrogram Analysis were implemented on the WAV files and we found a proper dataset with coefficient values.

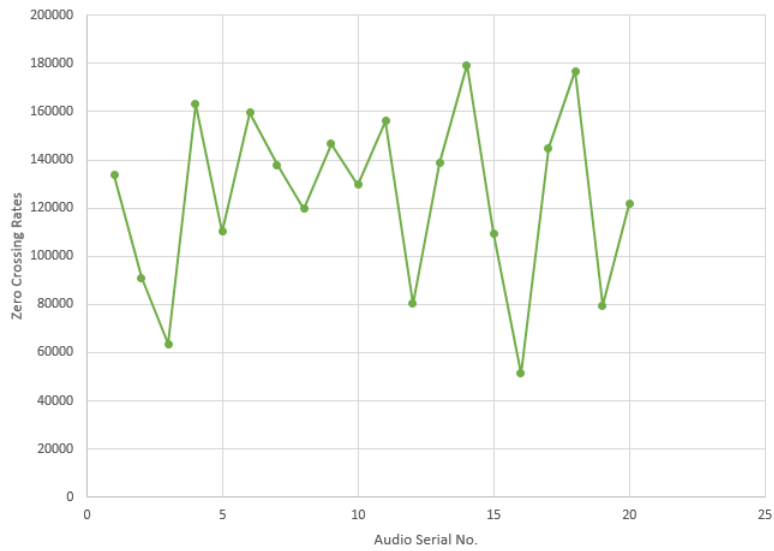


Figure 3.42: Zero Crossing Rates for several Advertisements (Scatter Plot)

Fig 3.42 shows the dataset imported from implementing Zero Crossing Rates on the Audios WAV files.

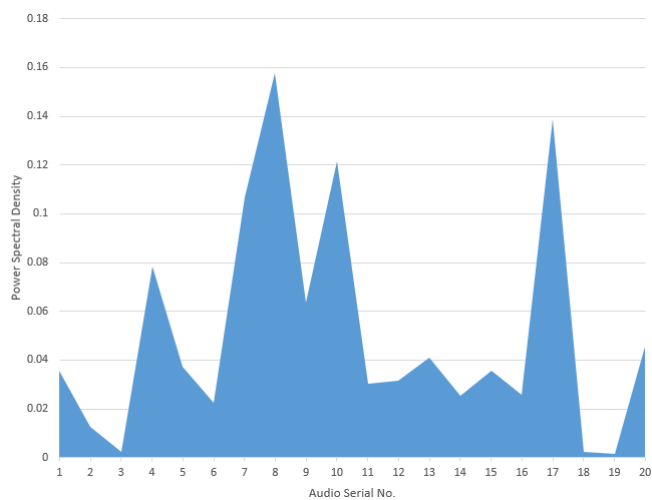


Figure 3.43: Power Spectral Density for several Advertisements (Area Plot)

In Fig 3.43 we can see the coefficient values of Power Spectral Density from the dataset.

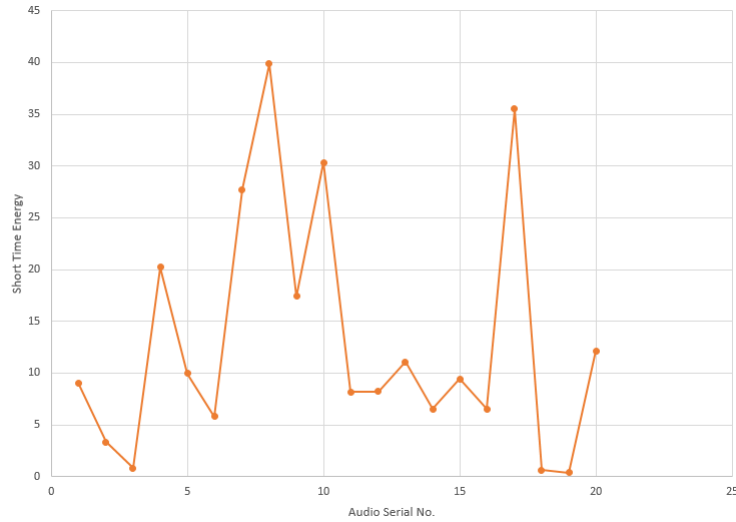


Figure 3.44: Short Time Energy for several Advertisements (Scatter with Smooth Line Plot)

In Fig 3.44 we can see the coefficient values of Short Time Energy from the dataset. Since Short Time Energy was implemented on the Audio WAV files the we got the coefficient values of Short Time Energy.

Audio	Short Time Energy
1	9.04786827
2	3.3407651
3	0.80651416
4	20.2259369
5	9.9921401
6	5.80404097
7	27.7008885
8	39.8767616
9	17.4127872
10	30.3229674

Table 3.3: Short Time Energy

Here, Table 3.3 shows the values from the dataset. This dataset was collected from the Audio Feature Extractions. On the selected Audio WAV files Short Time Energy was implemented and from there we found the values in fig 3.8.

Audio	Power Spectral Density
1	0.03574637
2	0.0127264
3	0.00228985
4	0.07848948
5	0.03726017
6	0.02250676
7	0.10656762
8	0.15784803
9	0.06339972

Table 3.4: Power Spectral Density

Here, Table 3.4 shows the values from the dataset. This dataset was collected from the Audio Feature Extractions. On the selected Audio WAV files Power Spectral Density was implemented and from there we found the values in fig 3.12.

3.4 Methodology

Our research work has been started by collecting the advertisements. After that, several processes has done to collect data and to classify that data. Both machine learning algorithm and deep learning algorithm has been used throughout our research. In this chapter there will be a description of our method or process that has been used in the research.

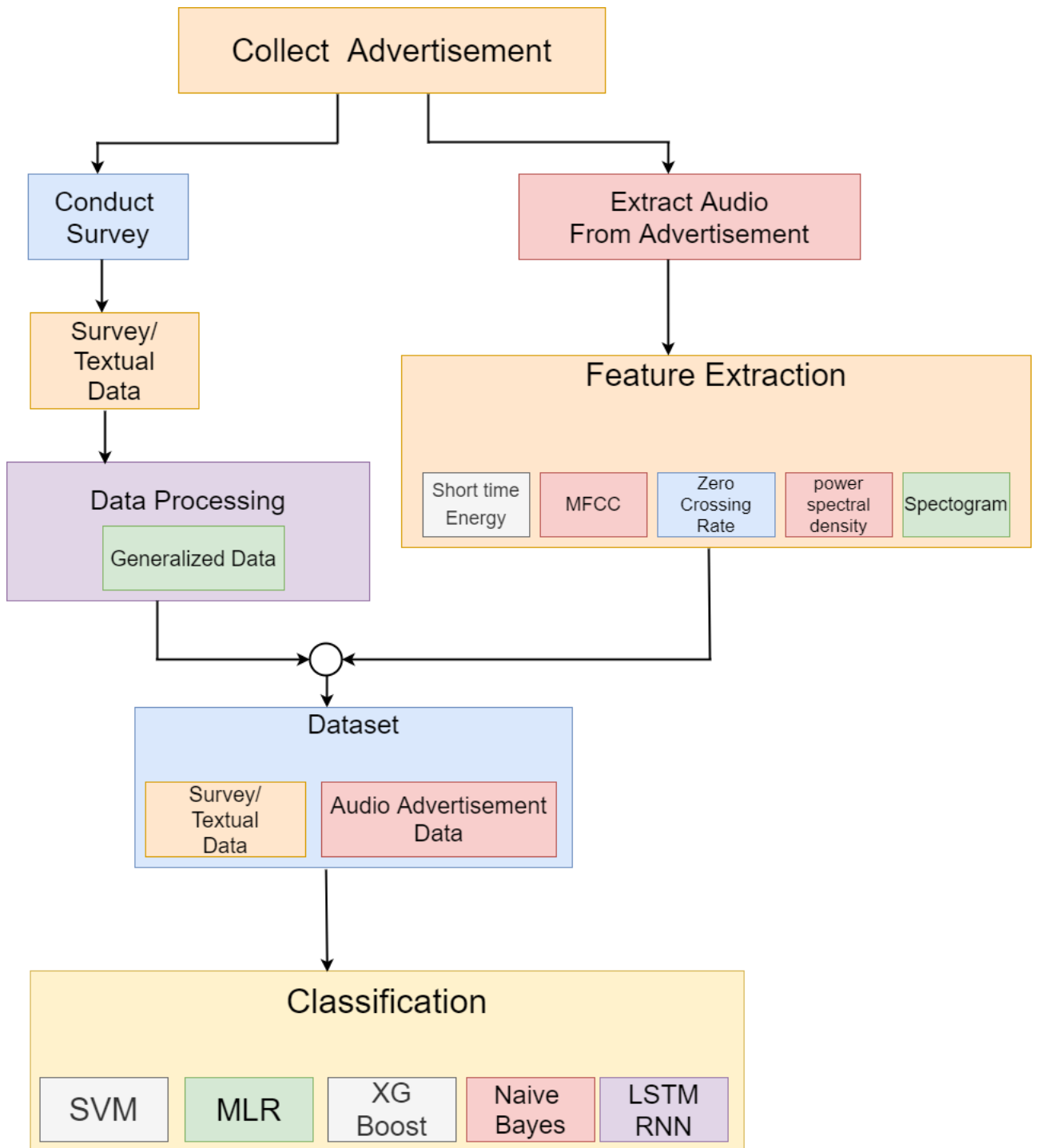


Figure 3.45: Methodology

Acquiring Dataset

The data has been used in our research is collected from the advertisement's audio and by conducting survey. The advertisement's audio data is collected from audio's feature extraction. Moreover, from the Google form the survey data has been collected. For conducting this survey we have set up some questions. The questions are related to people's emotion and their choice about the advertisement's product. Then both of this audio feature extraction data and survey data has been sorted into an excel file.

Firstly, after collecting the advertisement then we have split or extract the audios from the advertisements. Then we have to cut down the audios into sixty seconds as we wanted to keep all our audios into same duration. When the splitting or extracting audio from advertisement and cut down of audios were done then the feature extraction part has begun. Audio feature has been extracted by different feature extraction method. Five different extraction methods we have used to extract features from our audios. Short-time energy, Mel frequency cepstral coefficient (MFCC), power spectral density (PSD), zero Crossing ate and spectrogram are the feature extraction methods that we have used. Thus from the audio feature extraction we have collected the advertisement's audio data.

Along with the collecting advertisement's audio feature extraction data we have also conducted the survey to get the people's reaction or emotions and their thought about the product whether they will purchase the product or not. To get these data we have set up questions which are related to people's emotion and their choice. For each question there was scale of one to five. The scaling was done for our betterment so that we can easily understand about people's reaction about the product. Not only this but also we have labelled the scales with different emotions which helps us to know the emotions of people about the advertisement. Moreover, by these questions, scaling and labelling, we also came up to know the chance of purchasing the product after watching the advertisement.

The next step was generalized the survey data. As it was scaled from one to five, we had tried to reduce the scaling from one to three so that we can get more accurate emotion from the survey. The survey data has been generalised in such a way that we took the scaling of four and five as three, two and three as two and one was generalized as one.

After that, both these survey data and audio feature extraction data have been sorted together into an excel file. We have kept these data together for our further classification process. Now for each advertisement's audio we are getting audio feature extraction data and arousal, valence, dominance, purchase rate, liking rate which are basically sorted according to people's emotion and choice after watching the advertisement. In the classification process our target will be the survey data on the basis of our advertisement's audio feature extraction data. That is why both survey data and audio feature extraction data have been kept together into an excel file.

Data Classification Through Machine Learning

In the classification process we have used different kind of machine learning algorithms to classify the data properly. Support Vector Machine (SVM), XGBOOST, Naïve Bayes and Multiple Linear Regression have been used as classifier in ours research.

3.4.1 XGBOOST

In XGBOOST classifier we have classified the arousal, valence, dominance, purchasing rate and liking rate which we have got from the survey according to the people's emotional statement and choice about the product after watching the advertisement. This classification was done according to the audio feature extraction data and on the basis of arousal, valence, dominance, purchase rate and liking rate that means we have targeted the survey data to get the output according to our audio feature extraction data. We have split our data for test and train purpose. Both testing and training part have contained audio extraction feature data along with the people's emotion as arousal, valence, dominance and people's choice about the product as purchase rate and liking rate. The classification was taken place separately for arousal, valence, dominance, purchase rate and liking rate. That means at first we have classified the data for arousal. After that the classification has taken place for valence and then dominance. Later on the purchase rate and liking rate have been classified.

3.4.2 Naïve Bayes

For Naive Bayes classifier the same process has done. In the Naive Bayes classifier we have split the data for test and train. As input we have taken audio extraction feature data and for output the survey data has been taken. Here, we have also classified the data separately that means classification of arousal, valence, dominance, purchase rate and liking rate was taken place for once at a time. In this paper[19] they showed the Bayes theorem as follows -

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) \times P(B)}{P(B)} \quad (3.4)$$

The Naive Bayes algorithm assumes that all attributes are a priori equally essential and all attributes are statistically independent given the target class.

3.4.3 Support Vector Machine (SVM)

Support-Vector machines (SVM) are supervised learning models with a related learning algorithm. Supervised learning is a learning model built to make prediction, given an unforeseen input instance. Basically, Support-Vector machines analyse data used for the study of classification and regression. In other words, provided that the training data are labelled, an optimal hyper plane is generated by an algorithm that categorizes new instances. Primarily, the Audio Feature Extraction methods were implemented on the Audio WAV files and Extracted dataset was imported from there. Also, the textual dataset was grabbed from the Survey. Basically,

there were two times SVM were implemented on the datasets. In the first time, the SVM was implemented on both Extracted dataset and Textual dataset. The data was divided into train and test part. Extracted dataset has been used as input and Textual dataset like Arousal, Valence, etc. have been used as output for predicting emotional rates. For the second time, the SVM was implemented only on Textual dataset and this time input was Arousal, Valence, Dominance, Liking and Purchase. On the other hand, output was one of the emotional statements. In detail, either Arousal or Valence was output. Similarly, either Liking or purchase was also output but only one of the emotional statements can be in the output and the emotional statement we used in output must be removed from input.

Now to make the successful classification on SVM the focus should be on four aspects that are given below:

Before going into those we need to know some definitions:

- True Positives: The situations in which we estimated YES and the real production were both YES.
- True Negatives: The situations in which we estimated NO and the real production were both NO.
- False Positives: The situations in which we estimated YES and the real production were both NO.
- False Negatives: The situations in which we estimated NO and the real production were both YES.

Confusion Matrix

A confusion matrix is a description of the effects of the forecast on the issue of classification. The number of accurate and incorrect predictions is listed by count values and broken down by gender. That is the secret to the matrix of uncertainty. The confusion matrix reveals how uncertain the classification algorithm is when it makes predictions. To calculate confusion matrix-

$$Accuracy\ of\ matrix = \frac{Total\ Positives + False\ Negatives}{Total\ Number\ of\ Samples} \quad (3.5)$$

Precision

Precision speaks about how accurate / precise the prediction is out of those expected to be positive, how many of them are currently positive. Precision is a reasonable indicator to assess when the cost of False Positive is high. To calculate precision –

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3.6)$$

So the formula represents as –

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

Figure 3.46: Precision Formula Representation

Recall

Recall simply measures how many of the Real Positives the model catches by marking it as positive (True Positive). Applying the same definition, realization came that Recall is the product parameter that we use to pick our better choice when there is a relatively higher cost associated with False Negative. To calculate recall –

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (3.7)$$

So the formula actually represents

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

Figure 3.47: Recall Formula Representation

F1 Score

F1 Score could be a safer metric to use because we try to find a balance between Precision and Recall AND an unequal distribution of the class (a high number of Real Negatives). To Calculate -

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.8)$$

In the papers[26][10]it has been stated that these methods have been accurate for different real life scenario.

3.4.4 Multiple Linear Regressions (MLR)

Multiple Linear Regression (MLR) also known as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The MLR was implemented only on the textual dataset and the Arousal, Valence and Dominance were put into the input. On the other hand, Liking and purchase were put into output. Actually, one emotional statement either liking

or purchase can be in the output; also both Liking and Purchase must be removed from the input. To calculate the MLR the given equation below can be observed -

$$Y_0 = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon \quad (3.9)$$

here

- Y_i = Dependent variable
- x_i = Explanatory variable
- β_0 = y-intercept(Constant term)
- β_p = Slope coefficients for each variables
- ϵ = Residual

Residual, ϵ , is the longitudinal difference between a regression line and a sample point. Every single data point has a residual. If they are above the regression line then they are positive and negative if they are below the regression line. If the line of regression eventually goes through the point, then the residual at that point is zero. This has been used widely to determine MLR[9].

Data Prediction through Deep Learning

To overcome the earlier shallow network that can prevent the efficient training we can use deep learning and not only that but also we can use deep learning for hierarchical multi-dimensional training data[38]. Therefore to get efficient training and accuracy we have used long short term recurrent neural network (LSTM-RNN) model in our research.

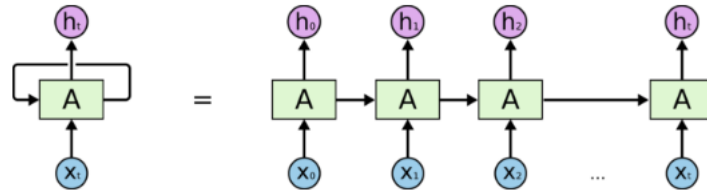
3.4.5 LSTM-RNN

An recurrent neural system (RNN) is a class of counterfeit neural systems where associations between hubs structure a coordinated chart along a transient arrangement. This permits it to display transient unique conduct. Derived from feed forward neural systems, RNNs can utilize their inner state (memory) to process variable length groupings of information sources. RNNs can experience the ill effects of evaporating/detonating angles during preparing. Numerous varieties have been created to address this. Long momentary memory (LSTM) uses a gating component and memory cells to relieve the data stream also, reduce inclination issues[3]. A profound neural system is a neural system with many stacked layers[27].

Unlike feed forward neural networks, RNNs are able to process input sequences using their internal state (memory). This makes them specific to activities such as unsegmented recognition, linked handwriting, or voice recognition. Both inputs are distinct one from another in other neural networks. In RNN, however, all inputs are connected to one another. To Calculate RNN –

Current state formula is:

$$h_t = \tan h \times ((W_{hh} \times h_t - 1) + (W_{xh} \times x_t)) \quad (3.10)$$



An unrolled recurrent neural network.

Figure 3.48: An unrolled RNN

Here, W is weight, h is that one hidden vector, W_{hh} denotes the weight when it was in previous hidden state, W_{hx} is the weight when it is in current input state, \tanh is the activation function.

For output:

$$y_t = W_{hy} \times h_t \tag{3.11}$$

Here, Y_t is the output state. W_y denotes the weight at the output state.

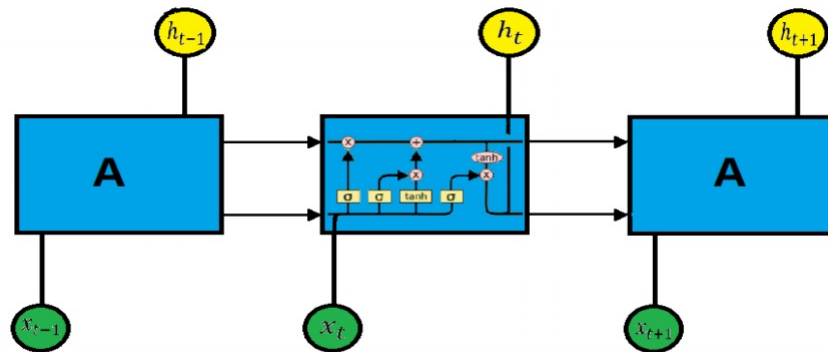


Figure 3.49: Flow Diagram/Computational Graph of a Canonical LSTM Network with Hidden State Vectors h_t and Input vectors x

The most widely recognized LSTM is made out of three doors and a memory cell. Much the same as a customary RNN, LSTMs ceaselessly update a concealed state on each time emphasize LSTM's increasingly powerful conduct, notwithstanding, originates from the utilization of a memory cell that influences the yield of each concealed state. These memory cells give the system the opportunity to figure out what data to monitor, include, or erase for guaranteed task. These entryways can be deciphered as channels cautiously plying inward portrayals of groupings for maximal order power[36]. To determine whether lstm is working fine or not calculations should be –

$$O_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \tag{3.12}$$

$$h_t = O_t \times \tanh(c_t) \tag{3.13}$$

For our long short term recurrent neural network (LSTM-RNN) model the first thing that has been done is to cut down the each audio into six seconds which was previously sixty second audios. Thus the each audio has been sliced down into ten audios of same duration. Now for these ten audios we will again extract the features with five different extraction methods. Each advertisement's audio now turned into ten audios of same duration and they are containing five different audio feature extraction data. However, the arousal, valence, dominance, purchase rate and liking rate will be same for these ten audios. After that, we have loaded our dataset and converted the data-set into an array. In the next step we have normalized our data-set between (0, 1). Now our data-set is in 2 dimension shape. We have reshaped our data-set into 3 dimensional array as our long short term recurrent neural network (LSTM-RNN) model expect multiple samples of one or more time steps and one or more features. We have done this as because we have used this data directly into our long short term recurrent neural network (LSTM-RNN) model. We have used the sizes to reshape the array to specify the number of samples by our total rows and columns as time steps and fix the number of features as 1.

After that, we have split our data into train and test part. According to the audio feature extracted data part we have targeted arousal, valence, dominance, liking rate and purchase rate have been used for prediction.

In LSTM input layer we have used our column as time steps which was 24 and the number of features at a time which was 1 as input. In our long short term recurrent neural network (LSTM-RNN) model we have used two LSTM layers of 1024 units, 3 Dropout layers to avoid the over fitting nature and 2 dense layers.

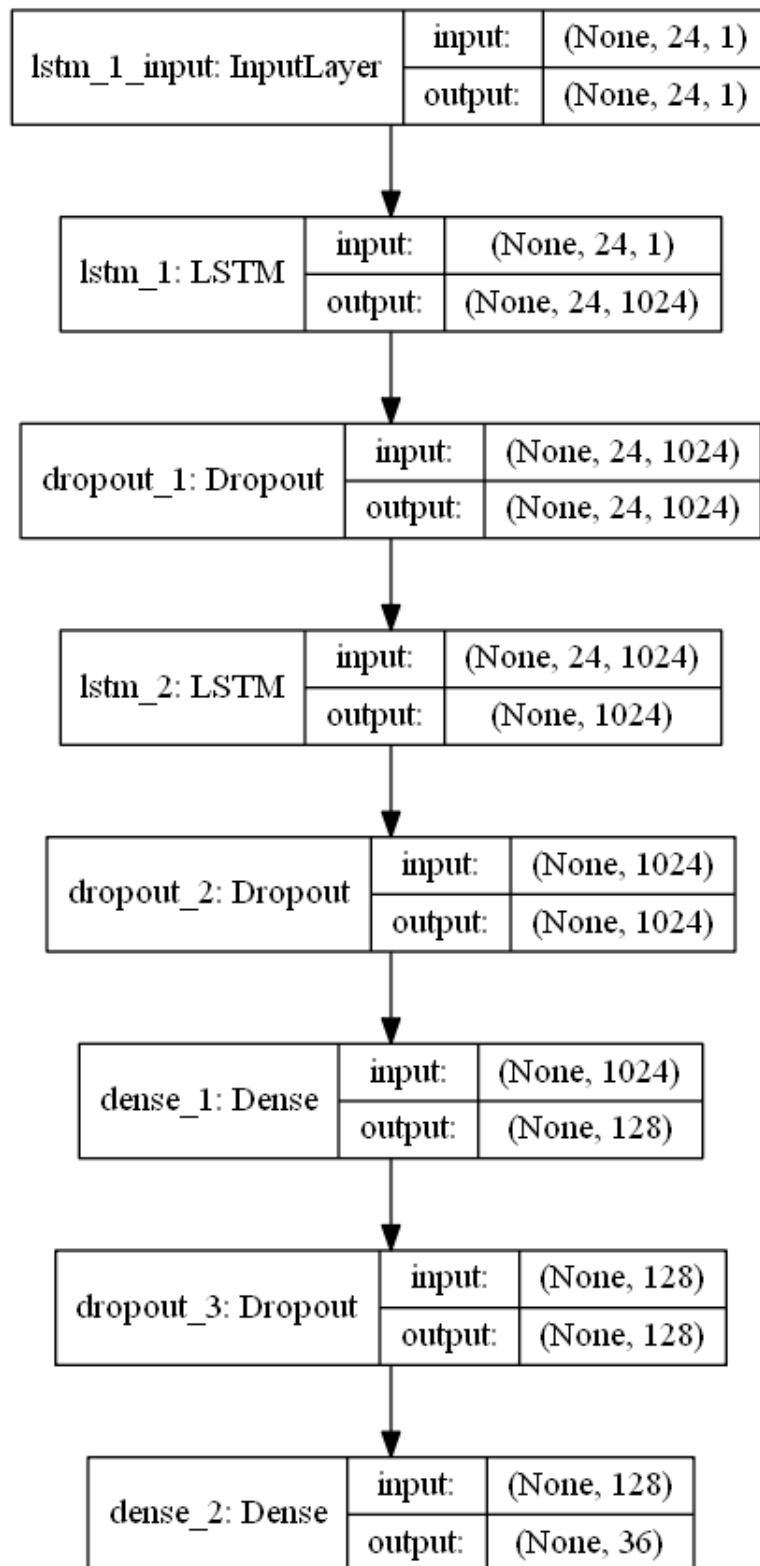


Figure 3.50: Sequential Flowchart of LSTM-RNN Model

In our long short term recurrent neural network (LSTM-RNN) model we have used Adam optimizer with the learning rate 1e-5, decay 1e-6 and the batch size 32 to train the network for accuracy. Sequential model has been used in our long short term recurrent neural network (LSTM-RNN) model. Therefore in our input layer we have used the ReLU activation to ensure the linearity. All these process have been implemented for 300 steps training.

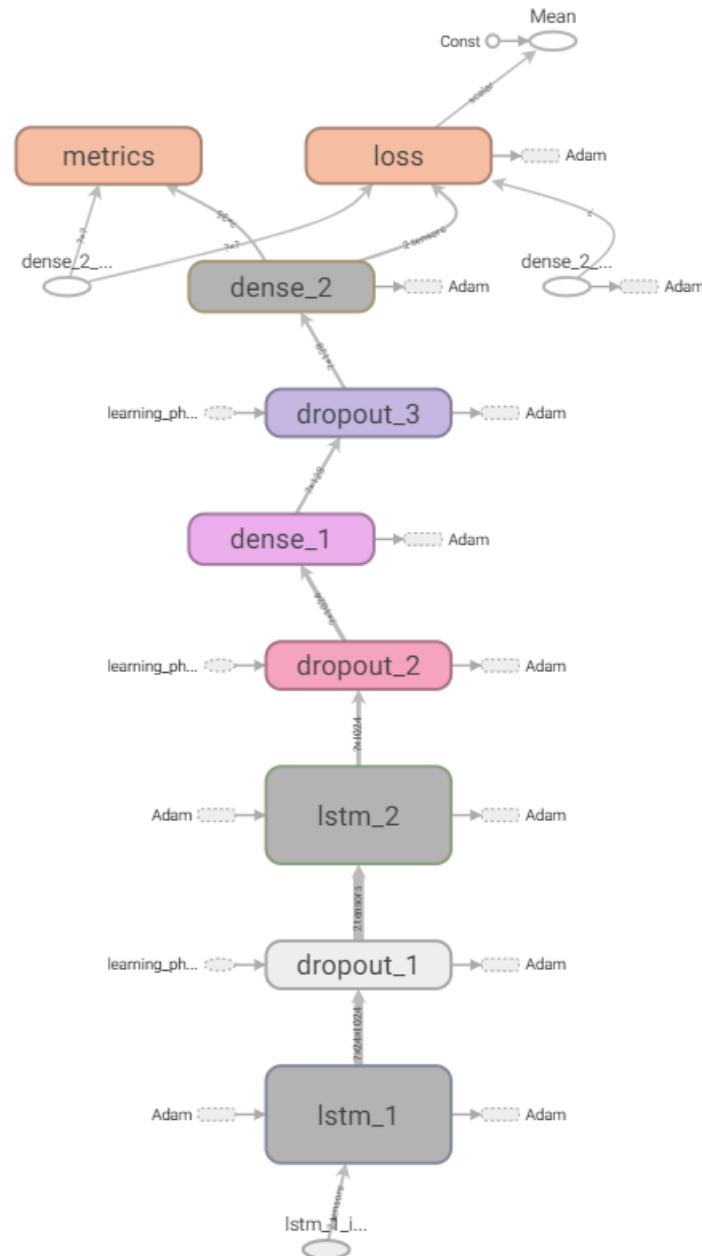


Figure 3.51: LSTM-RNN Model

Chapter 4

Implementation and Result Analysis

This chapter describes the implementation of our proposed model to find out the successful and unsuccessful advertisements and four emotions of those successful and unsuccessful advertisements. These four emotions we are getting from the survey data-set, more specifically from arousal, valence which is basically reflect the people's emotion after watching the advertisement. Moreover, whether the people will buy the product or not and liking the product or not after watching the advertisement we can know about that from our survey's dominance, purchase rate and liking rate after implementing our proposed model. The model follows four stages: collect data from audio feature extraction method and from survey data, data pre-processing for input, data classification and finally the result output, performance measurement. In our feature extraction part, we have used five different feature extraction methods: MFCC, Zero Crossing Rate, Short-time Energy, Power spectral density and Spectrogram. For classification part we use both machine learning algorithm and deep learning algorithm. For machine learning SVM, Multiple linear regressions, XGBoost, Naive Bayes have been used and in deep learning part long-short term recurrent neural network (LSTM-RNN) has been used.

4.1 Data pre-processing

The dataset of our research is consisting of audio feature extraction data and survey data. The audio feature extraction data we are getting from audios by extracting their feature with five different feature extraction methods and survey data we are getting by conducting survey with self-prepare questions. For audio feature extraction data first, we have to collect our audio from the advertisements. Then cut all the audios into sixty seconds so that all of our audio is in same duration. Then the features will be extracted from audios by MFCC, zero crossing rates, power spectral density, short-time energy and spectrogram.

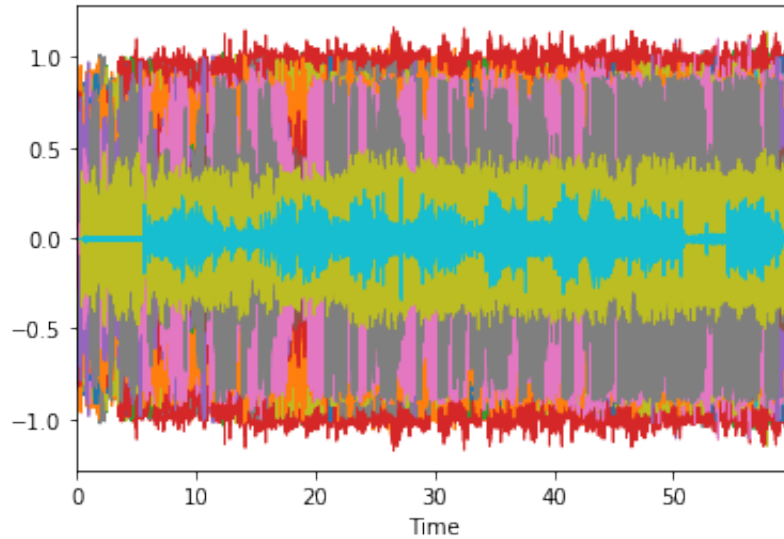


Figure 4.1: 60 seconds sliced audios

The questions of our survey data are related to people's emotion and their choice after watching the advertisements. From there we will gate our arousal, valence, dominance which is related to people's emotion and purchase rate and liking rate which is related to the people's choice whether they will buy the product or not after watching the advertisement. In survey data we have scaled our questions from one to five and labelled them with different emotions. Then we have generalized our data where we have rescaled the data from one to three. Four and five was rescaled as three, two and three was rescaled as two and one was unchanged.

4.2 Combined Decision Making

Form figure 4.2, we are getting our successful and unsuccessful advertisements on the basis of people's purchase intent. From the survey conduct we have got the people's purchase intent of the product after watching the advertisements. After that we have classified these purchase intents by different classifier. In the figure 4.2, the advertisements those are above 80% will be counted as most successful advertisements and the advertisements those are below 50% will be counted as most unsuccessful advertisements. For our betterment to implement and result analysis we are using here four most successful advertisements those are 6,13,27,28 and four most unsuccessful advertisements those are 1,3,22,36 from the figure 4.2.

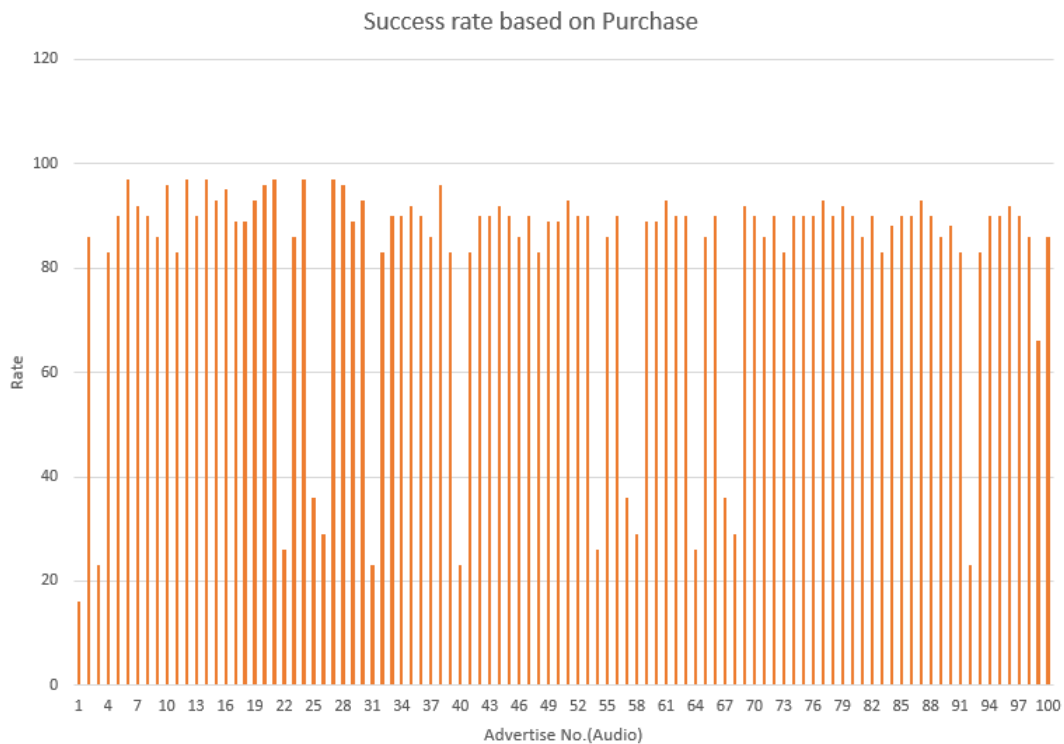


Figure 4.2: Success rate based on purchase

4.3 Results

4.3.1 Naive Bayes

Naïve Bayes classifier has been used to classify our survey data more specifically arousal, valence, dominance, purchase rate and liking rate which is related to people's emotion and choice about the product. We have classified these data on the basis of our audio feature extraction data. In the table 4.1, Naïve Bayes score for different emotional states has been showed which we have collected from survey or textual data.

Emotional State	Naive Bayes Score
Arousal	0.80
Valence	0.83
Dominance	0.86
Liking	0.83
Purchase	0.87

Table 4.1: Naïve Bayes of Emotional States from Textual or Survey Data

The score we are getting from Naïve Bayes for arousal which is actually representing pleasant and unpleasant rate because we have labelled the question for our arousal at the time of survey with these two emotions. Like arousal the labelling for our valence was clam and excitement. For dominance we have used highly motivated or low motivated as our label. Moreover, the purchase rate and liking rate are showing people's intents whether the buy the product or not and liking the advertisement or not after watching the advertisement.

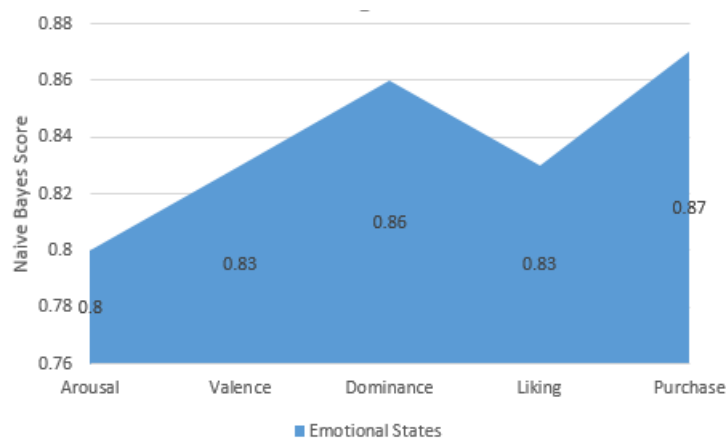


Figure 4.3: Area Plot of Naive Bayes Classifier

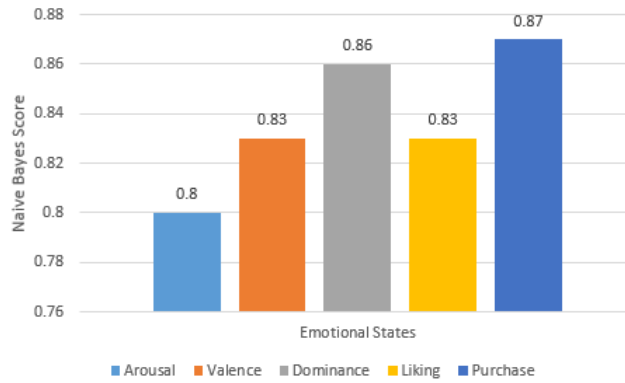


Figure 4.4: Column Plot of Naive Bayes Classifier

4.3.2 XGBoost

XGBoostclassifier has also been used to classify our survey or textual data. From this classification we are getting the score of our arousal, valence, dominance, purchase rate and liking rate states which we collected from our survey on the basis of people's emotion and choice about the product after watching the advertisement. To get the emotional state we have labelled these arousal, valence and dominance states. Moreover, purchase rate and liking rate is reflecting people's choice whether they are buying the product or not and whether they are liking the product or not after watching the advertisement. In the table 4.2, XGBoost score for different emotional states has been showed which we have collected from survey or textual data.

Emotional State	XGBoost Score
Arousal	0.83
Valence	0.83
Dominance	0.80
Liking	0.83
Purchase	0.85

Table 4.2: XGBoost of Emotional States from Textual or Survey Data

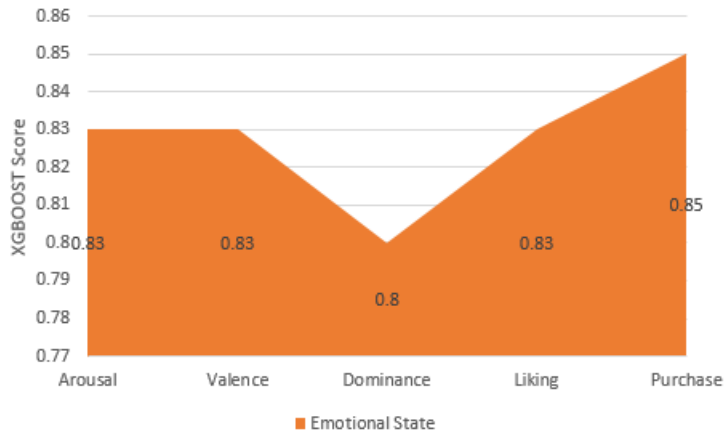


Figure 4.5: Area Plot of XGBOOST Classifier

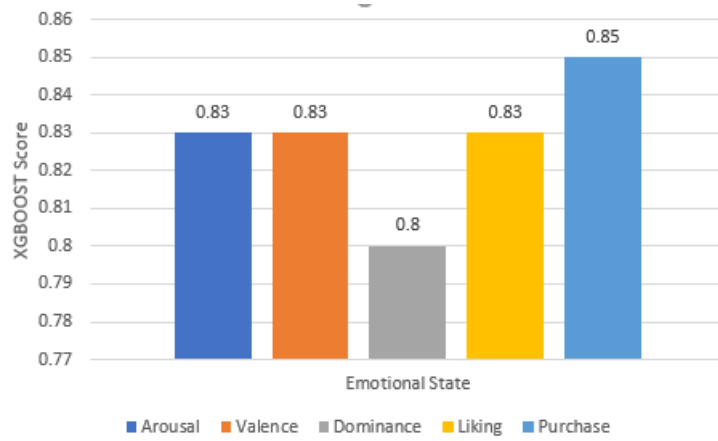


Figure 4.6: Column Plot of XGBoost Classifier

4.3.3 LSTM-RNN

In our research we have used LSTM-RNN model as our deep learning method. LSTM-RNN model has been used to predict our arousal, dominance, valence, purchase rate and liking rate states. To run this deep learning method we have sliced our audios into 6 seconds so that we can get ten same length audios for each advertisement. Then we have reshape our data form two dimensional to three dimensional and after that we have normalized our data between (0, 1) for input. From table 4.3, we are getting our LSTM-RNN model prediction for different emotional state which we are getting from our survey or textual data.

Emotional State	LSTM-RNN
Arousal	0.77
Valence	0.81
Dominance	0.83
Liking	0.80
Purchase	0.78

Table 4.3: LSTM-RNN Prediction for Emotional States from Survey or Textual Data

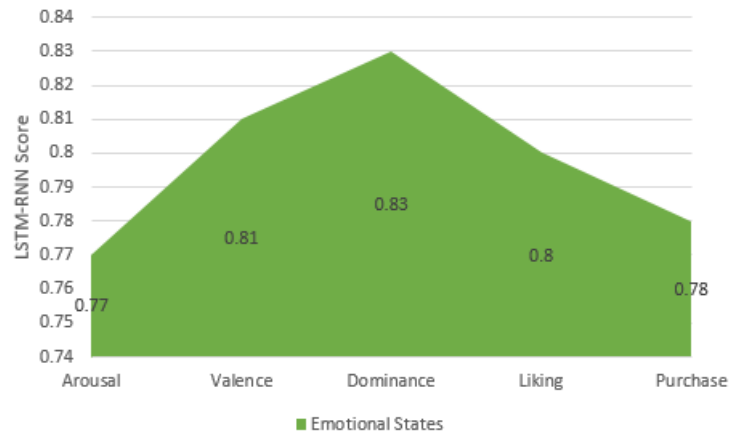


Figure 4.7: Area Plot of LSTM-RNN

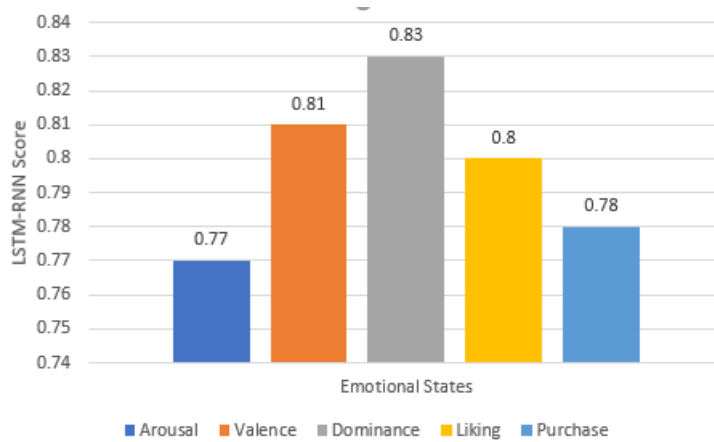


Figure 4.8: Column Plot for LSTM-RNN

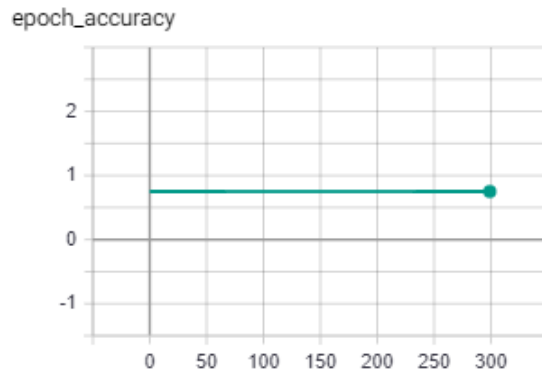


Figure 4.9: Arousal Accuracy of LSTM-RNN



Figure 4.10: Arousal Loss of LSTM-RNN

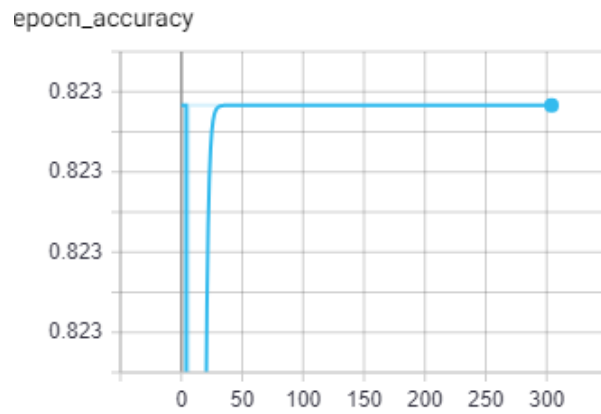


Figure 4.11: Valence Accuracy of LSTM-RNN

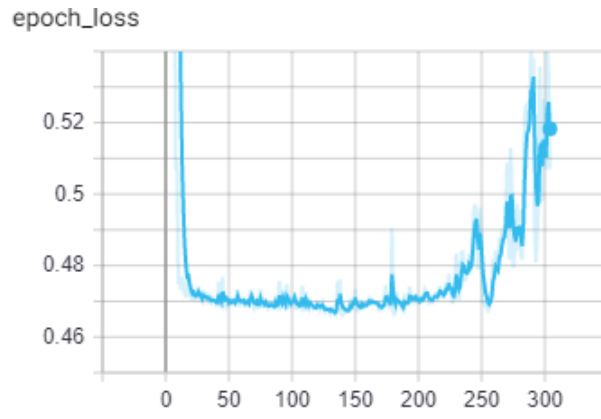


Figure 4.12: Valence Loss of LSTM-RNN

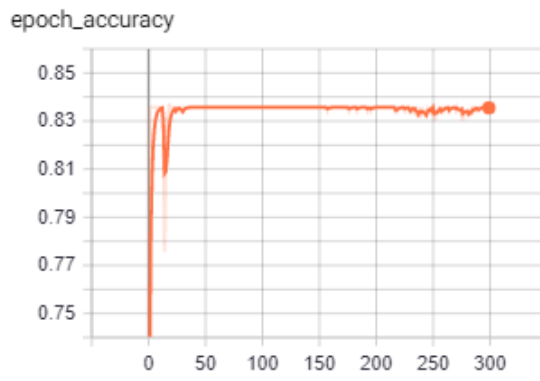


Figure 4.13: Dominance Accuracy of LSTM-RNN

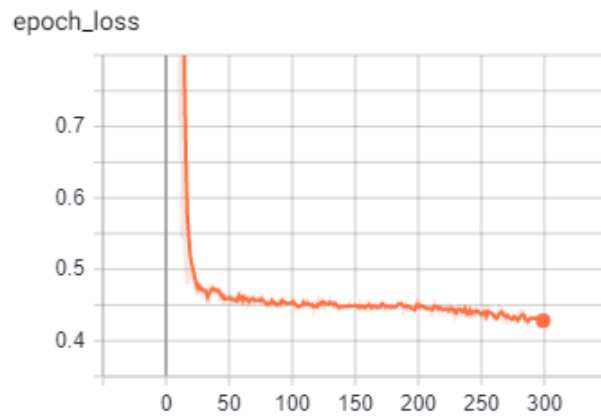


Figure 4.14: Dominance Loss of LSTM-RNN

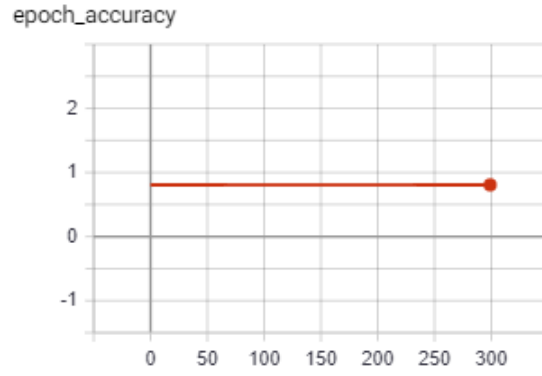


Figure 4.15: Liking Accuracy of LSTM-RNN

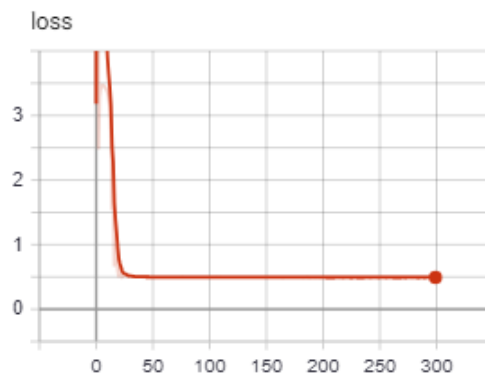


Figure 4.16: Liking Loss of LSTM-RNN

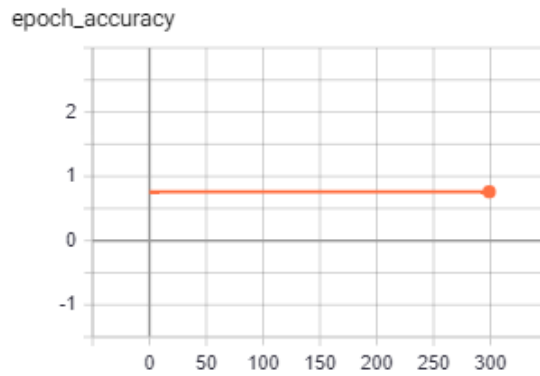


Figure 4.17: Purchase Accuracy of LSTM-RNN

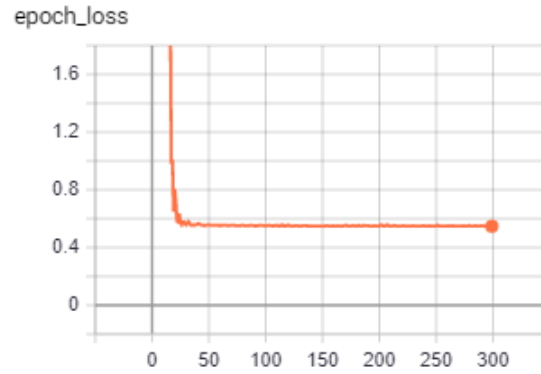


Figure 4.18: Purchase Loss of LSTM-RNN

Model: "sequential_1"

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 24, 1024)	4202496
dropout_1 (Dropout)	(None, 24, 1024)	0
lstm_2 (LSTM)	(None, 1024)	8392704
dropout_2 (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 128)	131200
dropout_3 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 36)	4644

Total params: 12,731,044
 Trainable params: 12,731,044
 Non-trainable params: 0

Train on 700 samples, validate on 300 samples

```

Epoch 1/300
700/700 [=====] - 27s 38ms/step - loss: 18.1243 - accuracy: 0.7200 - val_loss: 3.3984 - val_accuracy: 0.8333
Epoch 2/300
700/700 [=====] - 25s 35ms/step - loss: 1.9040 - accuracy: 0.5857 - val_loss: 0.7159 - val_accuracy: 0.4567
Epoch 3/300
700/700 [=====] - 25s 35ms/step - loss: 0.6259 - accuracy: 0.7171 - val_loss: 0.4828 - val_accuracy: 0.8333
Epoch 4/300
700/700 [=====] - 25s 36ms/step - loss: 0.6126 - accuracy: 0.7386 - val_loss: 0.4508 - val_accuracy: 0.8333
Epoch 5/300
700/700 [=====] - 25s 36ms/step - loss: 0.6162 - accuracy: 0.7386 - val_loss: 0.4520 - val_accuracy: 0.8333
Epoch 6/300
700/700 [=====] - 26s 37ms/step - loss: 0.5825 - accuracy: 0.7514 - val_loss: 0.4528 - val_accuracy: 0.8333
Epoch 7/300
700/700 [=====] - 26s 37ms/step - loss: 0.5891 - accuracy: 0.7557 - val_loss: 0.4556 - val_accuracy: 0.8333
Epoch 8/300
700/700 [=====] - 25s 36ms/step - loss: 0.5859 - accuracy: 0.7557 - val_loss: 0.4500 - val_accuracy: 0.8333
Epoch 9/300
700/700 [=====] - 25s 36ms/step - loss: 0.5833 - accuracy: 0.7543 - val_loss: 0.4607 - val_accuracy: 0.8333
Epoch 10/300
700/700 [=====] - 25s 36ms/step - loss: 0.5738 - accuracy: 0.7557 - val_loss: 0.4499 - val_accuracy: 0.8333
Epoch 11/300
700/700 [=====] - 25s 36ms/step - loss: 0.5699 - accuracy: 0.7571 - val_loss: 0.4515 - val_accuracy: 0.8333
Epoch 12/300
700/700 [=====] - 25s 36ms/step - loss: 0.5788 - accuracy: 0.7571 - val_loss: 0.4631 - val_accuracy: 0.8333

```

Figure 4.19: Sample Training of LSTM-RNN

4.3.4 Support Vector Machine

After implementing SVM (Support-Vector Machine) on the textual dataset accuracy score of different emotion states were found. There are different accuracy values found for Arousal, Valence, Dominance, Liking and Purchase. SVM was implemented two times, firstly, SVM was implemented on the Textual dataset only which was imported from the Survey response. Secondly, SVM was implemented on both datasets Textual one and Extracted dataset also.

Emotional State	SVM Score
Arousal	0.80
Valence	0.90
Dominance	0.85
Liking	0.90
Purchase	0.95

Table 4.4: SVM of Emotional States from Textual Dataset

Emotional State	SVM Score
Arousal	0.82
Valence	0.85
Dominance	0.86
Liking	0.90
Purchase	0.84

Table 4.5: SVM of Emotional States from Textual and Extracted Dataset

Table 4.4 and 4.5 shows the SVM accuracy scores of different emotional states. Arousal, Valence and Dominance are the emotional states and Liking and Purchase are intents for instance if they really like the advertisement and product or not. Arousal has 0.80 accuracy score, Valence has 0.90, Dominance has 0.85, Liking intent has 0.90 and Purchase intent has 0.95 accuracy score. All the emotional states confirm that purchase intent of the user or customer has 95% accuracy. Therefore, participants liked most of the advertisements that they want to purchase the product pretty much. To illustrate, their purchase intent to buy the product is almost 95%.

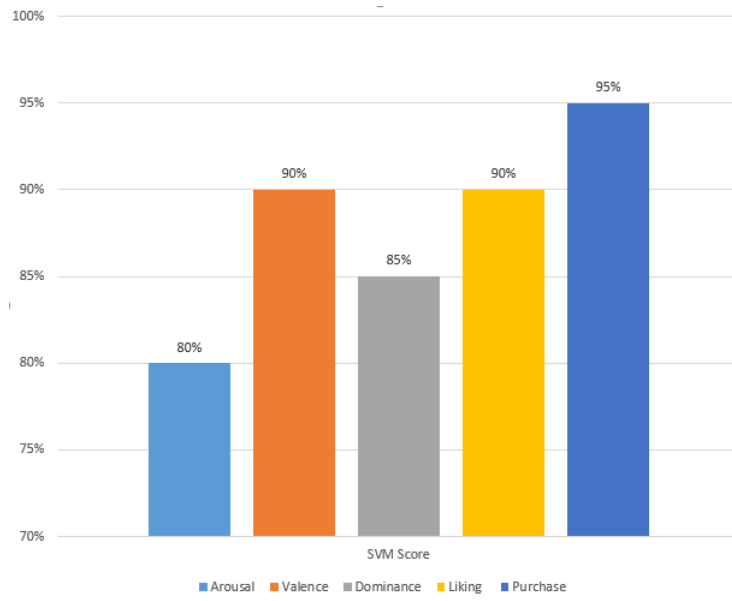


Figure 4.20: SVM Scores from Textual Dataset in Clustered Column Plot

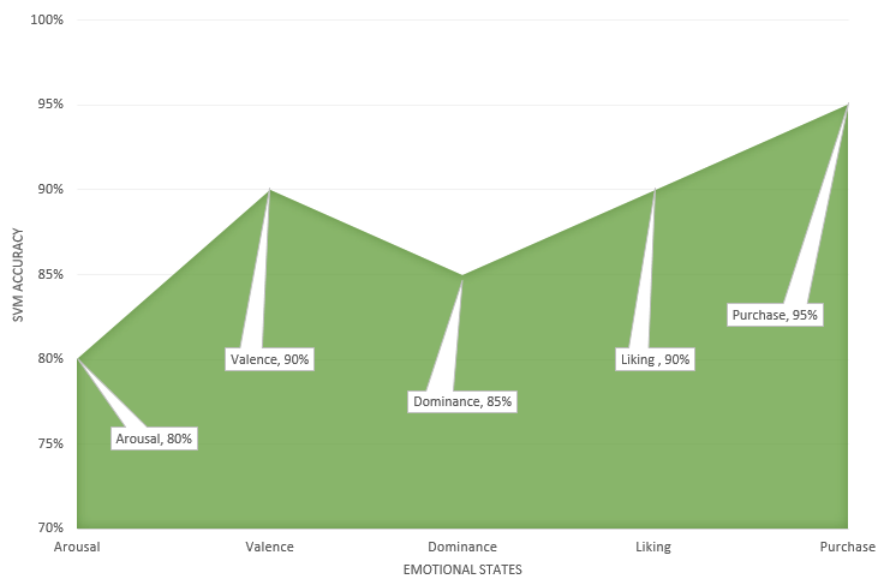


Figure 4.21: SVM Scores from Textual Dataset in Area Plot

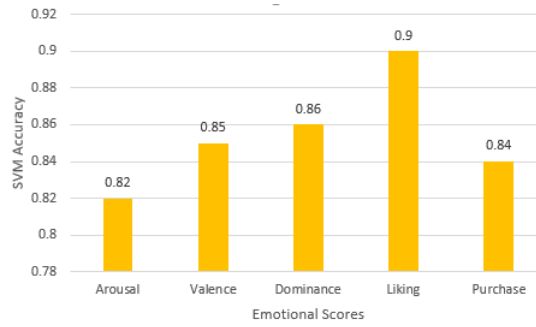


Figure 4.22: SVM Scores from Textual and Extracted Dataset in Clustered Column Plot

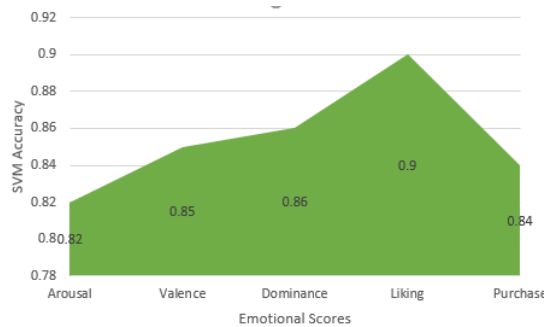


Figure 4.23: SVM Scores from Textual and Extracted Dataset in Area Column Plot

Fig 4.20,4.21, 4.22 and 4.23 shows total purchase intent is 95% on all the advertisements. Also, it tells that Arousal is up to 80%, Valence is 90% and Dominance is 85%. These are all the emotional states whereas the Liking intent is 90%.

From fig 4.24 Successful advertisements and are found based on the Purchase intent. Here, the plot tells that Advertisement 06, 13, 27 and 28 are most successful as well as Advertisement 01, 03, 22 and 26 are most unsuccessful advertisement.

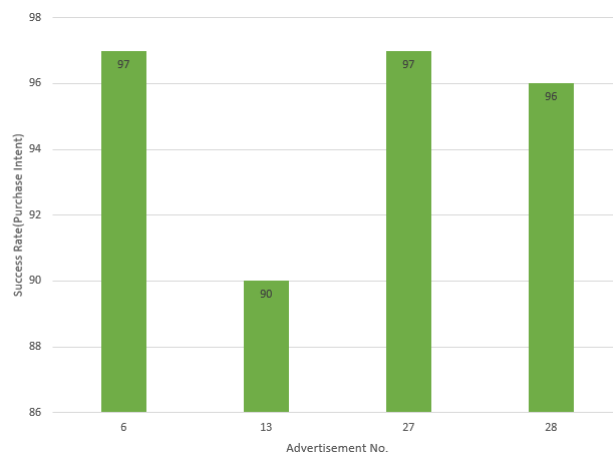


Figure 4.24: Success Rate of Successful Advertisements (Column Plot)

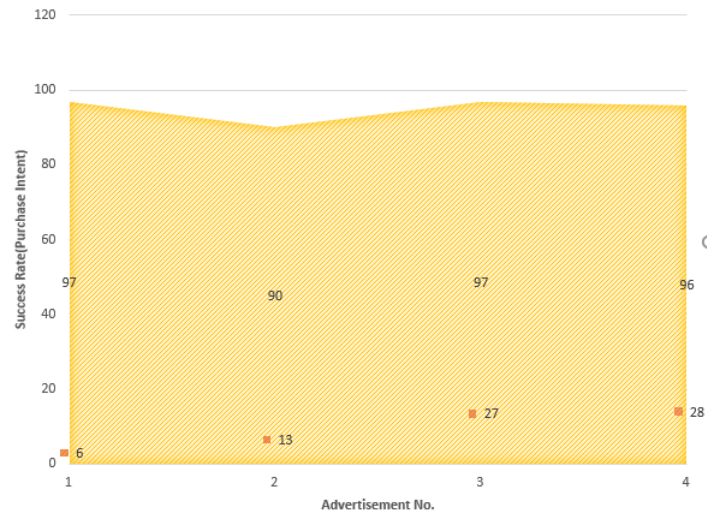


Figure 4.25: Success Rate of Successful Advertisements (Area Plot)

Here, fig 4.24 and 4.25 tells that these four most successful advertisements have Success Rate of 90% or more than 90% based on the purchase intent of the customers or users. Also, along with these four advertisements other successful advertisements have Success Rate above 80%.

Also, fig 4.26 and 4.27 tells that these four most unsuccessful advertisements have Success Rate of 29% or below 29% based on the purchase intent of the customers or users. Also, along with these four advertisements other successful advertisements have Success Rate below 40%.

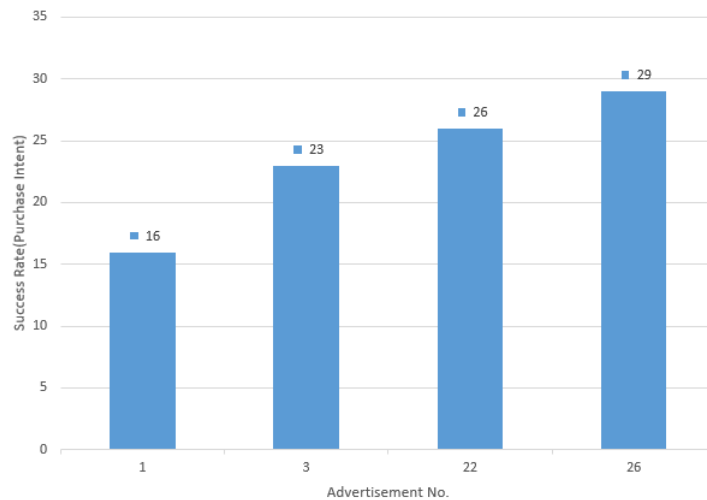


Figure 4.26: Success Rate of Unsuccessful Advertisements (Column Plot)

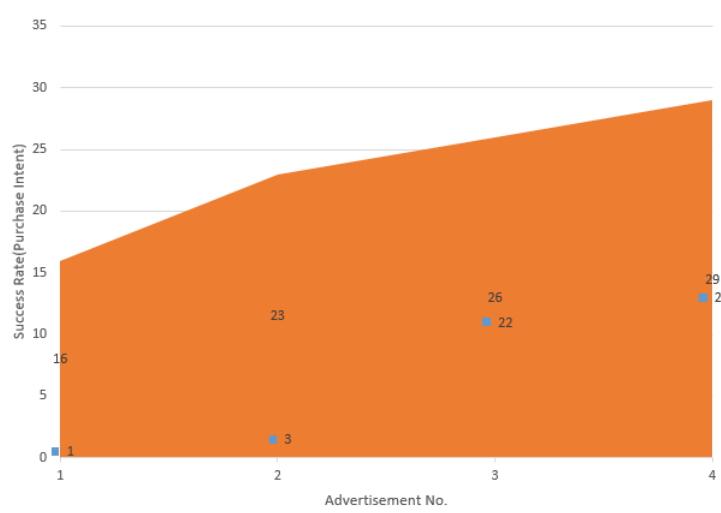


Figure 4.27: Success Rate of Unsuccessful Advertisements (Area Plot)

Advertise Number	Arousal	Valence	Dominance
6	0.77	0.87	0.87
13	0.8	0.78	0.7
27	0.88	0.87	0.75
28	0.89	0.88	0.79

Table 4.6: SVM of Emotional States from Successful Advertisements

Table 4.6 refers the SVM accuracy scores of emotional statements from the most Successful four advertisements. Here, all the scores of Arousal, Valence and Dominance are 0.7 or more than 0.7. In detail, all the emotional states have 70% or more than 70% SVM accuracy score.

Advertise Number	Liking	Purchase
6	0.88	0.78
13	0.7	0.72
27	0.87	0.88
28	0.88	0.78

Table 4.7: SVM of Liking and Purchase intent from Successful Advertisements

Table 4.7 refers the SVM accuracy scores of liking and purchase intent from the most Successful four advertisements. Here, all the scores of Liking and Purchase are more than 0.7. In detail, all the emotional states have 70% or more than 70% SVM accuracy score.

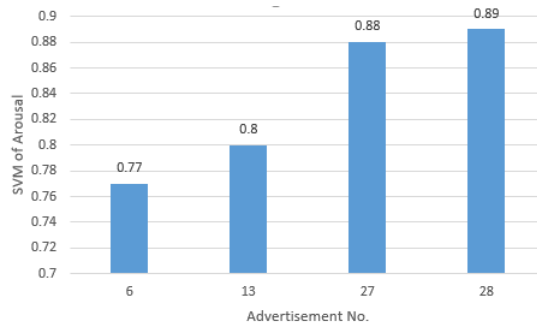


Figure 4.28: SVM of Arousal from Most Successful Advertisements

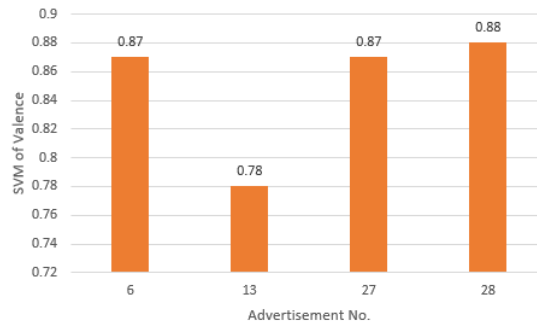


Figure 4.29: SVM of Valence from Most Successful Advertisements

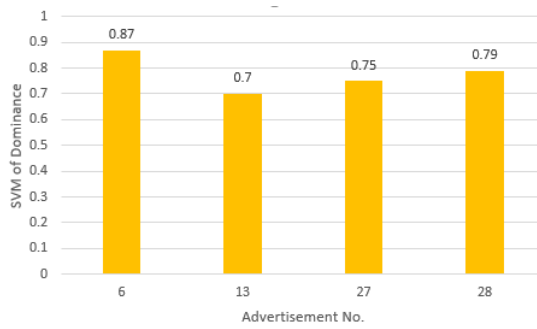


Figure 4.30: SVM of Dominance from Most Successful Advertisements

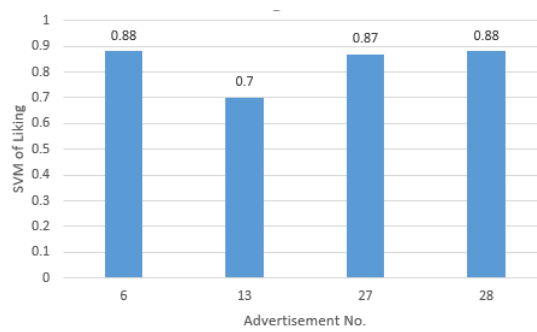


Figure 4.31: SVM of Liking from Most Successful Advertisements

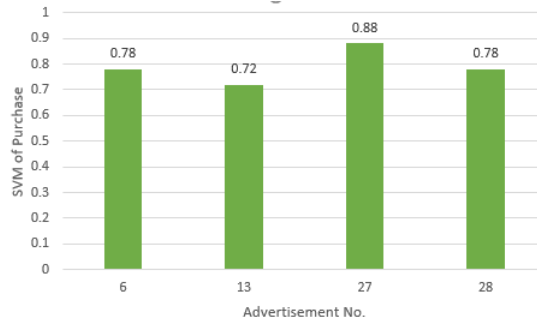


Figure 4.32: SVM of Purchase from Most Successful Advertisements

Advertise Number	Arousal	Valence	Dominance
1	0.32	0.4	0.48
3	0.39	0.42	0.43
22	0.47	0.33	0.4
26	0.330	0.29	0.42

Table 4.8: SVM of Emotional States from Unsuccessful Advertisements

Advertise Number	Liking	Purchase
1	0.4	0.42
3	0.45	0.44
22	0.45	0.42
26	0.4	0.43

Table 4.9: SVM of Liking and Purchase intent from Unsuccessful Advertisements

Table 4.8 refers the SVM accuracy scores of emotional statements from the unsuccessful four advertisements. Here, all the scores of Arousal, Valence and Dominance are less than 0.5. In detail, all the emotional states have 50% or less than 50% SVM accuracy score. Also, table 4.9 refers the SVM accuracy scores of liking and purchase intent from the unsuccessful four advertisements. Here, all the scores of Liking and Purchase are less than 0.5. In detail, all the emotional states have 50% or less than 50% SVM accuracy score.

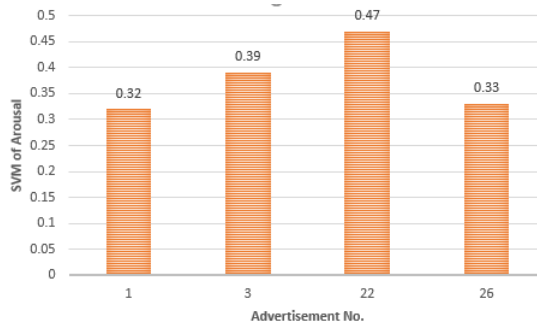


Figure 4.33: SVM of Arousal from Unsuccessful Advertisements

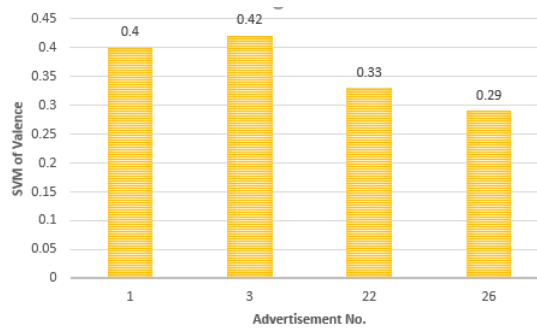


Figure 4.34: SVM of Valence from Unsuccessful Advertisements

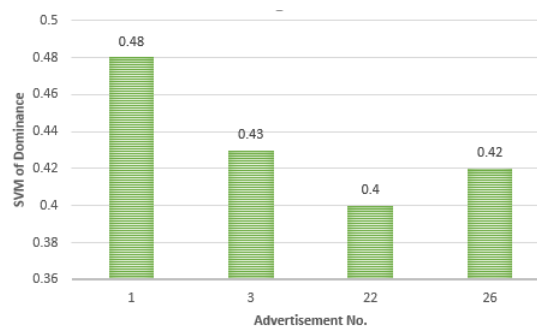


Figure 4.35: SVM of Dominance from Unsuccessful Advertisements

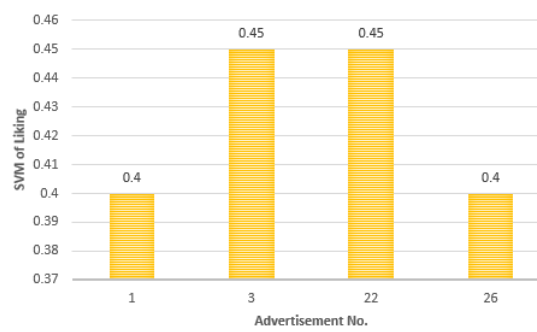


Figure 4.36: SVM of Liking from Unsuccessful Advertisements

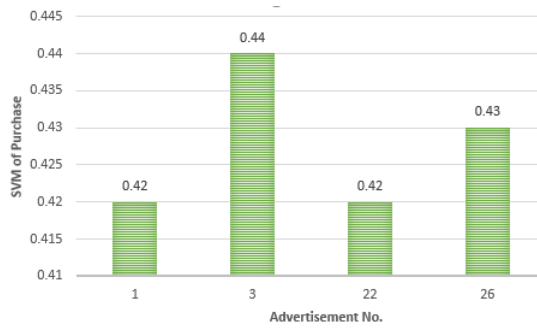


Figure 4.37: SVM of Purchase from Unsuccessful Advertisements

4.3.5 Multiple Linear Regression

Alongside implementing SVM also Multiple Linear Regression was implemented on the datasets where variance score of the Liking intent and Purchase intent was found. Also, residual error plot was imported from the dataset.

Coefficients			Variance Score
C1	C2	C3	
0.3	0.17	0.44	0.81

Table 4.10: Regression Analysis of Liking intent

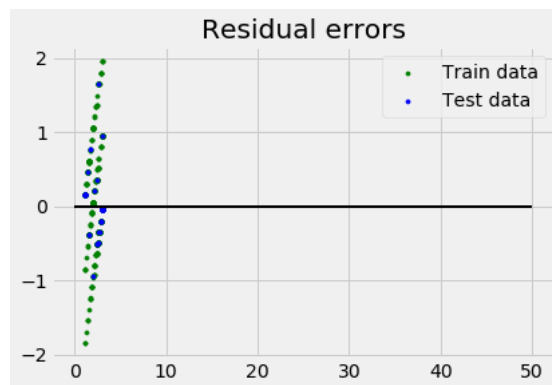


Figure 4.38: Residual Error of Liking intent

Coefficients			Variance Score
C1	C2	C3	
-0.03	0.26	0.57	0.79

Table 4.11: Regression Analysis of Purchase intent

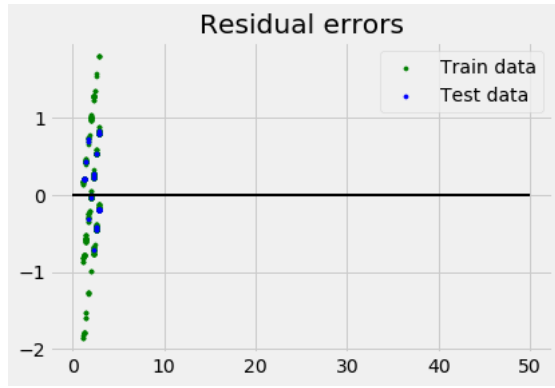


Figure 4.39: Residual Error of Purchase intent

Advertise No.	C1	C2	C3	Variance Score
13	0.71	-0.29	0.66	0.97
27	-0.01	0.3	0.51	0.79
28	1	-0.12	-0.2	0.98
17	1.12	0.25	-0.38	0.59
6	0.59	0.41	0.07	0.87
14	-0.005	0.51	0.44	0.58
15	0.37	0.36	0.33	0.57
24	0.33	-0.06	-0.14	0.53

Table 4.12: Regression Analysis of Liking intent from Most Successful Advertisement

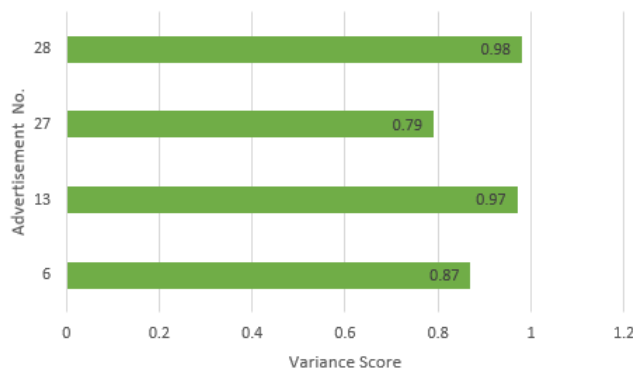


Figure 4.40: Regression Analysis of Liking intent from Successful Advertisements

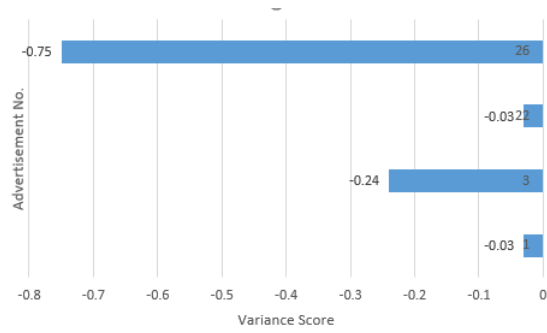


Figure 4.41: Regression Analysis of Liking intent from Unsuccessful Advertisements

Advertise No.	C1	C2	C3	Variance Score
13	0.08	0.58	0.28	0.58
27	0.15	0.13	0.51	0.51
28	-0.5	0.2	0.41	0.56
17	0.32	0.2	-0.05	0.51
6	0.18	0.56	0.37	0.78
14	-0.22	0.14	0.43	0.61
15	-0.3	0.07	0.99	0.95
24	0.5	-0.05	0.47	0.71

Table 4.13: Regression Analysis of Purchase intent from Most Successful Advertisement

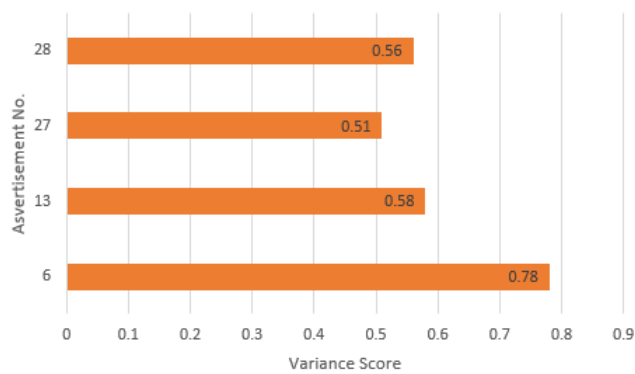


Figure 4.42: Regression Analysis of Purchase intent from Successful Advertisements

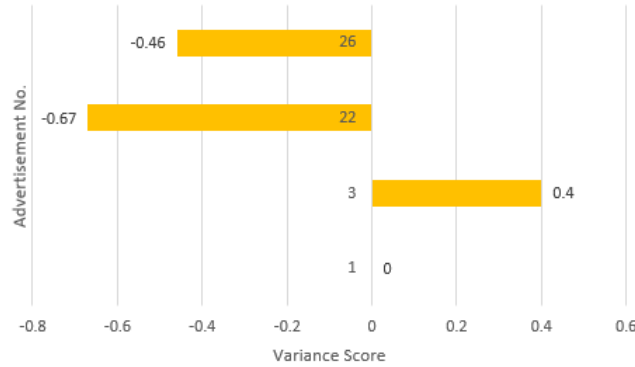


Figure 4.43: Regression Analysis of Purchase intent from Unsuccessful Advertisements

4.3.6 Performance Measurement

Alongside LSTM-RNN, SVM and other prediction algorithm there are other methods which have been implemented on the datasets to interpret Performance Measurement. Precision, Recall and F1 Score have been implemented to figure out and evaluate the performance of our research. These also help to understand the Confusion Matrix of the datasets.

I. Confusion Matrix

In the field of machine learning and specially the problem statistical classification, a confusion matrix, also known as an error matrix. Actually, confusion matrix was implemented to describe the performance of our classification models. It allows the visualization of the performance of our prediction algorithms.

II. Precision

Precision is the ratio of the correctly predicted positive observations to the total predicted positive observations. High precision relates to the low false positive rate. Actually, precision was implemented to find out how precise or accurate the research model is out of those predictive positive, how many are actual positive.

III. Recall

Recall is the ratio of correctly predicted positive observations to the all observations actuals class. Recall is the model metric we use to select our best model when there is high cost/value associated with False Negative.

IV. F1 Score

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false negatives into account. Basically, F1 Score was implemented on our model to make a balance between Precision and Recall.

Generally, to measure the performance of our classification and prediction algorithms over the textual and extracted datasets Precision, Recall and F1 Score was implemented. Also, Confusion matrix was found to describe the performance and visualize the classification and prediction algorithms. Since the response from the Survey questions was generalized in three classes/scale for every scale Precision, Recall and F1 score was found and this was for all the emotional states for instance, Arousal, Valence and Dominance as well as Liking intent and Purchase intent.

Class/Scale	Precision	Recall	F1 Score
1	1	0.5	0.67
2	0.57	0.35	0.44
3	0.79	1	0.88

Table 4.14: Performance measurement for Arousal

Class/Scale	Precision	Recall	F1 Score
1	0.86	0.68	0.76
2	0.53	0.38	0.44
3	0.9	1	0.95

Table 4.15: Performance measurement for Valence

Class/Scale	Precision	Recall	F1 Score
1	1	0.5	0.67
2	0.9	0.7	0.56
3	0.84	1	0.91

Table 4.16: Performance measurement for Dominance

Class/Scale	Precision	Recall	F1 Score
1	1	1	1
2	1	0.33	0.5
3	0.89	1	0.94

Table 4.17: Performance measurement for Liking

Class/Scale	Precision	Recall	F1 Score
1	1	1	1
2	0.77	0.9	0.57
3	0.95	1	0.97

Table 4.18: Performance measurement for Purchase

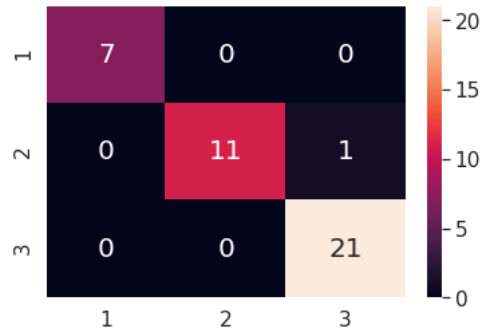


Figure 4.44: Heatmap of Confusion Matrix of Arousal

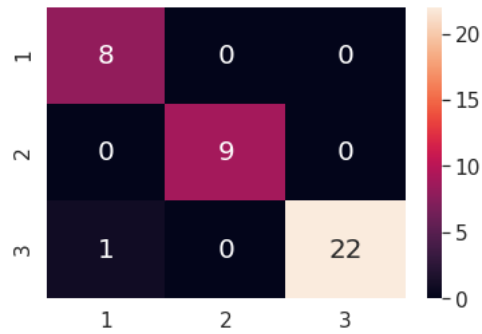


Figure 4.45: Heatmap of Confusion Matrix of Valence

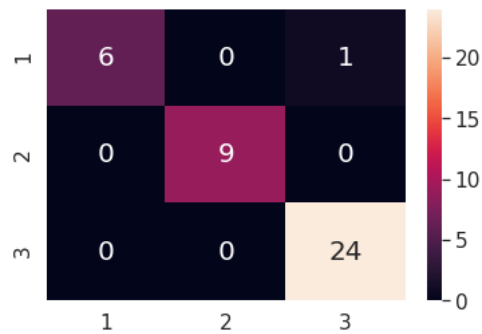


Figure 4.46: Heatmap of Confusion Matrix of Dominance

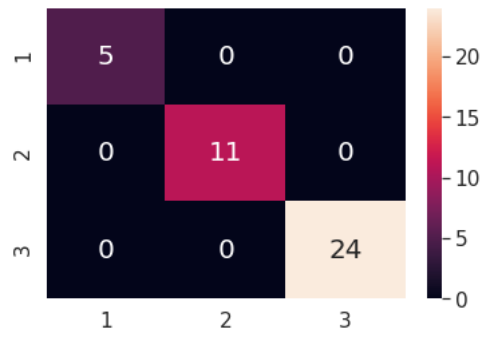


Figure 4.47: Heatmap of Confusion Matrix of Liking

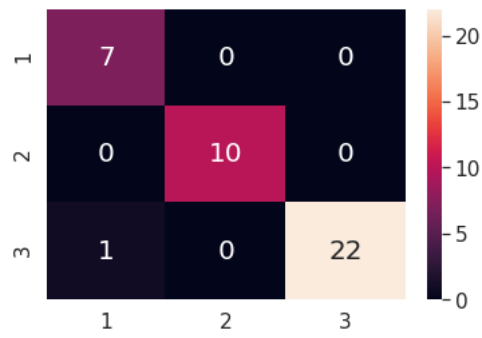


Figure 4.48: Heatmap of Confusion Matrix of Purchase

4.3.7 Final Result of Standard Emotional States

Finally, after implementing all the supervised machine learning and Deep learning algorithm we found the standard value for emotional states to make an advertisement successful. All the classification and prediction models have the values of emotional states which are very similar to each other. Therefore, from all the prediction and classification models the maximum value for the emotional states have been collected.

Emotional State	Arousal	Valence	Dominance	Liking	Purchase
Final	0.82	0.9	0.86	0.9	0.95

Table 4.19: Standard Emotional States Scores

Table 4.19 shows that Arousal has the accuracy score 0.82, Valence has 0.9, Dominance has 0.86, Liking has 0.9 and Purchase has 0.95. In detail, since Arousal is the measurement of the excitement and calm the value 0.82 for arousal means for a particular advertisement people feel 82% excited while watching the advertisement so they sure will buy the product and the advertisement will be successful. Therefore, for any particular advertisement if the Arousal is 82% or more than that, the Valence is 90% or more than that and the Dominance is 86% or more than that then the advertisement will surely be most successful advertisement and it is 95% guaranteed that people will purchase the product. Also, same goes for the other emotional states and liking intent.

Emotional State	Happiness	Sadness	Excitement	Calmness
Standard Rate	90%	10%	82%	18%

Table 4.20: Standard Emotional States Rate (in percentage)

Chapter 5

Conclusion and future Work

5.1 Conclusion

The main purpose of our thesis was to make a good marketing strategy which can be used by every company to reach their desired success. We can see that many companies around the world invest a lot of money every year on advertisements purpose. However, in the end they cannot be succeed all the time, cannot get to desired result they look for. Throughout our research, we tried to focus on this problem and tried to make a solution for them. By our thesis, we can say that what type of emotional content and advertisement should have for being a successful advertisement in the market. So, companies can easily get to know about what type of emotional contact in which rate should have been presented in the advertisements. By that they can easily implement it in their advertisement and make a successful advertisement for the market.

In our thesis, our main aim was to give the companies a solution to make a perfect advertisement for their successful marketing purpose. On the other hand, we can have some improvement in our field which we can implement later. In future, the further improvement can be we can take impressions of the viewers while they are watching the advertisements using brain sensors, so that we can tell whether they are linking this advertisement or not, which will help us to get the more accurate data and moreover, we do not have to do the survey we did before. By this our works will be more efficient and more perfect and more successful. Therefore, we will be trying to use neural networks to add in our works by which we can get the impressions of the people about and advertisement which will help us to get more accurate results and by that we can reduce a lot more error in our work.

5.2 Future Work

While working in our field we had face come difficulties which we are planning to improve in future to get the perfect result and to make our works better. In our project we have worked on audio frequency of the advertisement, but in future we are planning to use video of advertisements also by which we are going to get more data to work with which will help us to get the more accurate results than we are getting now. Moreover, we have worked on the advertisements of televisions and radio, but in future we are planning to implements more advertisements from different

source as in Facebook advertisements, YouTube advertisements banners and many more. There are also advertisements in airplanes, in shopping malls also. We will be working on those also the make the ultimate change in marketing for a product.

In our thesis we used machine learning algorithm and deep learning algorithm, but in future we are planning to implements more things. We are planning to implement neural networks which will help us to get the impressions of the viewers about any advertisements. Moreover, we are planning to implements our project in IOT so that it can be easily accessible to the people all around the world and can be controlled from remote places.

In our thesis we worked on audio frequency and people's reaction on advertisements which is conducted by survey, but in future we are planning to implement human emotion using brain activity, which will help us to find the actual overview about any advertisements people watch. Moreover, we will work on facial expression of people while watching the advertisements which will help us to get more accurate result, because the more data we will be working the more accurate result we will get. Therefore, we are planning to add more machine learning algorithms and deep learning algorithms which will help us to get more accurate data and we will be able to reduce the error.

Finally, we did not try any sensors for our work, but in future, we are planning to try to add some sensors with our work as in ECG sensor which will help us to get the change of heart beat during the time of watching the advertisements which will help us to find what changes happens when people like the advertisements and when they do not. More over we will add some other sensors like Electro dermal Activity sensor (EDA) or Galvanic skin response sensor (GSR) for our research. These sensors will also help us to analyse. the human reaction and behaviour while watching the advertisements.

Finally, with all of these we hope we will able to make more successful advertisement in the markets which can be used by all companies around the world.

Bibliography

- [1] H. Fleishman, “Businesses with brilliant global marketing strategies”, : <https://blog.hubspot.com/marketing/global-marketing-and-international-business>, 13.
- [2] V. Cherkassky, “The nature of statistical learning theory”, *IEEE Transactions on Neural Networks*, vol. 8, no. 6, pp. 1564–1564, 1997.
- [3] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [4] F. Gouyon, F. Pachet, O. Delerue, *et al.*, “On the use of zero-crossing rate for an application of classification of percussive sounds”, in *Proceedings of the COST G-6 conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, 2000, p. 26.
- [5] S. Z. Li and G.-d. Guo, “Content-based audio classification and retrieval using svm learning”, in *First IEEE Pacific-Rim Conference on Multimedia, Invited Talk*, Australia, 2000.
- [6] S. Molau, M. Pitz, R. Schluter, and H. Ney, “Computing mel-frequency cepstral coefficients on the power spectrum”, in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, IEEE, vol. 1, 2001, pp. 73–76.
- [7] G. Guo and S. Z. Li, “Content-based audio classification and retrieval by support vector machines”, *IEEE transactions on Neural Networks*, vol. 14, no. 1, pp. 209–215, 2003.
- [8] B. Moore, “An introduction to the psychology of hearing, academic”, *San Diego*, 2003.
- [9] C. Goutte and E. Gaussier, “A probabilistic interpretation of precision, recall and f-score, with implication for evaluation”, in *European Conference on Information Retrieval*, Springer, 2005, pp. 345–359.
- [10] A. H. Omar, “Audio segmentation and classification”, Master’s thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark, 2005.
- [11] W. Han, C.-F. Chan, C.-S. Choy, and K.-P. Pun, “An efficient mfcc extraction method in speech recognition”, in *2006 IEEE international symposium on circuits and systems*, IEEE, 2006, 4–pp.
- [12] R. Bachu, S. Kopparthi, B. Adapa, and B. Barkana, “Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal”, in *American Society for Engineering Education (ASEE) Zone Conference Proceedings*, 2008, pp. 1–7.

- [13] K. Sakhnov, E. Verteletskaya, and B. Simak, "Approach for energy-based voice detector with adaptive scaling factor.", *IAENG International Journal of Computer Science*, vol. 36, no. 4, 2009.
- [14] M. A. Hossan, S. Memon, and M. A. Gregory, "A novel approach for mfcc feature extraction", in *2010 4th International Conference on Signal Processing and Communication Systems*, IEEE, 2010, pp. 1–5.
- [15] P. Kathirvel, M. S. Manikandan, S. Senthilkumar, and K. Soman, "Noise robust zerocrossing rate computation for audio signal classification", in *3rd International Conference on Trendz in Information Sciences & Computing (TISC2011)*, IEEE, 2011, pp. 65–69.
- [16] Z. Qi, Y. Tian, and Y. Shi, "Robust twin support vector machine for pattern classification", *Pattern Recognition*, vol. 46, no. 1, pp. 305–316, 2013.
- [17] M. Thorogood and P. Pasquier, "Impress: A machine learning approach to soundscape affect classification for a music performance environment.", in *NIME*, 2013, pp. 256–260.
- [18] S. C. Joshi and D. A. N. Cheeran, "Matlab based feature extraction using mel frequency cepstrum coefficients for automatic speech recognition", 2014.
- [19] S. Zheng, "Naive bayes classifier: A mapreduce approach", 2014.
- [20] J. Lee and I. Tashev, "High-level feature representation using recurrent neural network for speech emotion recognition", in *Sixteenth annual conference of the international speech communication association*, 2015.
- [21] J. Saini and R. Mehra, "Power spectral density analysis of speech signal using window techniques", *International Journal of Computer Applications*, vol. 131, no. 14, pp. 33–36, 2015.
- [22] Y. Yang, J. Li, and Y. Yang, "The research of the fast svm classifier method", in *2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, IEEE, 2015, pp. 121–124.
- [23] L. Bahatti, O. Bouattane, M. E. Echhibat, and M. H. Zaggaf, "An efficient audio classification approach based on support vector machines", (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, 2016.
- [24] L. Chao, J. Tao, M. Yang, Y. Li, and Z. Wen, "Long short term memory recurrent neural network based encoding method for emotion recognition in video", in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2016, pp. 2752–2756.
- [25] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system", in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [26] R. M. George and J. A. Mathew, "Emotion classification using machine learning and data preprocessing approach on tulu speech data", 2016.
- [27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

- [28] K. Gupta and D. Gupta, “An analysis on lpc, rasta and mfcc techniques in automatic speech recognition system”, in *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)*, IEEE, 2016, pp. 493–497.
- [29] A. Pacuk, P. Sankowski, K. Wegrzycki, A. Witkowski, and P. Wygocki, “Recsys challenge 2016: Job recommendations based on preselection of offers and gradient boosting”, in *Proceedings of the Recommender Systems Challenge*, 2016, pp. 1–4.
- [30] A. Azzouni and G. Pujolle, “A long short-term memory recurrent neural network framework for network traffic matrix prediction”, *arXiv preprint arXiv:1705.05690*, 2017.
- [31] M. C. Darji, “Audio signal processing: A review of audio signal classification features”, *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 2, pp. 227–230, 2017.
- [32] Y. Hioka and K. Niwa, “Estimating power spectral density for spatial audio signal separation: An effective approach for practical applications”, *Acoustical Science and Technology*, vol. 38, no. 4, pp. 175–184, 2017.
- [33] Z. Shi, M. Shi, and C. Li, “The prediction of character based on recurrent neural network language model”, in *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, IEEE, 2017, pp. 613–616.
- [34] H. Zheng, J. Yuan, and L. Chen, “Short-term load forecasting using emd-lstm neural networks with a xgboost algorithm for feature importance evaluation”, *Energies*, vol. 10, no. 8, p. 1168, 2017.
- [35] M. Mulimani and S. G. Koolagudi, “Acoustic event classification using spectrogram features”, in *TENCON 2018-2018 IEEE Region 10 Conference*, IEEE, 2018, pp. 1460–1464.
- [36] S. B.-R. Sabatier, “A long short-term memory neural network for improved twins’ voice differentiation”, 2018.
- [37] T. A. B. Gombos, “Acoustic recognition with deep learning; experimenting with data augmentation and neural networks”, Master’s thesis, 2019.
- [38] A. Shrestha and A. Mahmood, “Review of deep learning algorithms and architectures”, *IEEE Access*, vol. 7, pp. 53 040–53 065, 2019.
- [39] A. Guttmann, *Global advertising spending 2019*, Jan. 2020. [Online]. Available: <https://www.statista.com/statistics/236943/global-advertising-spending/>.

ADAM OPTIMIZATION ALGORITHM

Require: Objective function $f(\theta)$ with parameters θ , initial parameter vector θ_0 , stepsize α , exponential decay rates β_1, β_2 for estimating the moments and θ_t for resulting parameters in scheme.

Algorithm: Adam optimizing Algorithm

```
1: procedure ADAM ( $f, \theta_0, \alpha, \beta_1, \beta_2$ )
2:  $m_0 \leftarrow 0$  #Initialize 1st moment vector
3:  $u_0 \leftarrow 0$  #Initialize the exponentially weighted
                    infinity norm
4:  $t \leftarrow 0$  #Initialize timestep
5: #Begin Optimizing Procedure
6: while  $\theta_t$  has not converged do
7:  $t \leftarrow t + 1$  #Update timestep
8:  $g_t \leftarrow \nabla f(\theta_{t-1})$  #Compute gradients of objective
                                   at timestep t
9:  $u_t \leftarrow \max(\beta_2 \times (u_{t-1}), |g_t|)$  #Update the exponentially
                                                weighted infinity norm
10:  $\theta_t \leftarrow (\theta_{t-1} - \alpha / (1 - \beta_1^t)) \times (m_t / u_t)$  #Update parameter
11: end while
12: return  $\theta_t$ 
```
