Factor Analysis of Ad Liking and Prediction of Purchase Intent through Emotionomics

by

Syeda Tanzina Farhin Toma
16101003
Shahan Jamil Bhuiyan
16101091
Tahmid Dawood
16101105
Chandan Kumar Saha
13101147

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

Department of Computer Science and Engineering
Brac University
December 2019

© 2019. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Syeda Tanzina Farhin Toma
16101003

Shahan Jamil Bhuiyan
16101091

Tahmid Dawood
16101105

Chandan Kumar Saha
13101147

Approval

The thesis/project titled “Factor Analysis of Ad Liking and Prediction of Purchase Intent through Emotionomics” submitted by

1. Syeda Tanzina Farhin Toma (16101003)
2. Shahan Jamil Bhuiyan (16101091)
3. Tahmid Dawood (16101105)
4. Chandan Kumar Saha (13101147)

Of Fall, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on December 26, 2019.

Examining Committee:

Supervisor:
(Member)

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

Program Coordinator:
(Member)

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

Head of Department:
(Chair)

DR.Mahbubul Alam Majumdar
Professor
Department of Computer Science and Engineering
BRAC University

0.1 Ethics Statement

The thesis is carried out in complete compliance with research ethics norms, and the codes and practices set by BRAC University. In our thesis we use the data from primary sources. We collect data from different participants and we use our own dataset for our thesis. we are ensuring we use references and in text citations properly. We the four co authors take full responsibility for the thesis code violations. For solving problems we read different websites, youtube tutorials, and Questionnaire Free tools. We also took help from our university faculty members. Finally, we declare that we give credit every people from whom we took help. We did not make any fraud able means for completing the thesis. Our work is in compliance with the ethics standard set by BRAC university.

Abstract

Marketing strategy is being a new challenge in this modern era. Along with the global market, people's choices are also changing so to grab the focus of buyers, organizations are making changes in their marketing policy based on user's choice to increase the possibility of their product being sold. Advertisements are the way to promote products and they are available on every media platform nowadays. The focus was always to gather public attention through different messages but which factors of advertising were needed more was not fixed. To recognize the dependencies of a successful advertisement and identify the factors which create good impression in people mind we lead this study. Using supervised machine learning algorithms and feature extraction method we find out the factors of ad liking and predict the purchase intent through emotionomics which makes an ad successful.

Keywords: Factor Analysis; Advertisements; EEG; Emotionomics.

Dedication

We would like to dedicate this research to our parents. Without their support we may not able to do our studies. We also want to dedicate the research to our friends who helped us to do better and most importantly the participants who helped us collecting data from survey. Specially the EEG data collection survey was difficult and our participants helped us greatly with there patience. Our supervisor helped us throughout the year. We want to dedicate this to him also.

Acknowledgement

Firstly, we would like to say thanks to our beloved family members whom we will always be indebted to. Secondly, we thank our supervisor Md. Golam Rabiul Alam for all his help and constant guideline. We also thank Dr. Mohammad Zavid Parvez for helping us with hardware support. Thirdly, we would like to thank all the faculty members and stuffs for providing us with such a wonderful study environment where could develop ourselves as well as this research properly.

Table of Contents

List of Figures

List of Tables

Nomenclature

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

α	Alpha
β	Beta
δ	Delta
γ	Gamma
σ	Sigma
$\sum$	Summation
θ	Theta
DT	Decision Tree
EEG	Electroencephalography
LR	Logistic Regression
NB	Naive Bayes
RF	Random Forest
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine

Chapter 1

Introduction

1.1 Background

All of us know that advertisements is one of the means of communication with the users of the products. Everyday different types of products are producing by the organisations for the use of people. These products will not be sold without advertising perfectly[1]. In our Medias, we see different advertisements of different types. Medias are also different types. Printing media, electronic media and social media etc. Some media contains printed copy for advertising, some media contains banners, and poster etc. these are the sorts of advertising process. The renowned marketing processing mostly depend on visual advertisements. It is also known to us that proper advertisement of a product can bring profit for an organisation. An organisation have many departments consisting of Finance, Marketing, Human Resource Management (HRM), Information Technology (IT) etc. All the departments are important for an organisation. The success of an organisation is fully dependent on their contribution. From an US survey we found that mostly 10-12 percent revenue is budgeted for the marketing of a company. A start-up also use their 2-3 percent budget for marketing and advertising[2]. Competitive industries use up to 20 percent for their marketing.

1.2 Motivation

Present global market is changing rapidly. The main reason of changing of the market is the change of human behaviour. Human behaviour is changing more rapidly. The demand of the peoples are increasing proportionate to their behaviour. For this reason different organisations are producing their important goods or useful things. These things are technological gadgets, foods and beverage, drinks, chemicals and aromatic products etc. As peoples demand is increasing, new companies are also arriving with their products to face the demands of the people. Without proper advertisements, the companies will be failed to introduce their products to people. We are looking for analysing the factors of advertisements which are playing the important roles to make an advertisement successful. We try to predict the success level of the advertisement by using machine learning algorithms. Our supervisor help us learn about the advancement of machine learning algorithms. He also suggest us to make a real time data set and inspire us to take a survey to learn the human impression towards any advertisement. He encouraged us that we can predict the success rate of the advertisements.

In 2018, from an US report, there were some top companies they invested billion dollars for the advertising of their companies. Though these companies are renowned, they also paid huge amount for their advertising of the product. Almost 200 advertisers spent 163 billion US dollars for the advertisement of their companies [3].

Table 1.1: The top spenders on advertisement in 2018

Company names	Spent Money on Advertising($ Billions)
Comcast Corp	$6.12
AT&T	5$.36
Amazon	$4.47
Procter & Gamble Co.	$4.30
General Motors Co.	$3.14
Walt Disney Co.	$3.13
Charter Communications	$3.04
Alphabet(Parent company of Google)	$2.96
American Express Co.	$2.80
Verizon communications	$2.68

1.3 Types of Advertisements

In order to have effecting marketing of the products of any organisation, there should be a popular media for advertising. Digital media is one of them. In digital media, we see varieties of advertisements including display advertisements, social media ads, newspaper and magazines, outdoor advertising, radio and podcasts, video ads, email marketing etc. [4].

To get the attention of the audiences these advertisements are designed differently. This will also help the organisations to seek the proper attention. This may lead them get their expected profit.

People now a days are engaging with advertisements. Most importantly new products may gain positive impression by a proper advertisement. These advertisements are used to show in different channels. As it is a part of marketing of a company or a product, advertisements telecast in all the Medias like TV, social media etc. The types of advertisements are varies from media to media. Short videos are more preferable in social Medias like YouTube. Sometimes for delivering social message through advertisements, ad makers try to make advertisement where a whole short film is summarized.

1.4 Problem Statement

We know that the importance of marketing is always important for companies. For start-up companies, the advertisement strategy plays an important role. When they get public focus, they can easily launch their products in markets. This is the main goal of a perfect organization. Every owner of the start-ups has their dreams to see their organizations at a higher level. Without perfect advertisements, sales may crash and the company may need to close. This will be a great loss for them. Mod-

ern economic system, our young generations are comfortable with their own business rather than engaging in others organisations. So, they open new organisations of their own. Some starts from heir student life. They risks their money and time for starting their own start-ups. They are known as Entrepreneurs. Everyone appreciates new ideas and views and entrepreneurs are playing an important role in this field. Though it is a matter of regret that many entrepreneurs failed to make profit. The key reason of the loss is not maintain a proper advertisement policy. They pay all the attention in their works but they failed to make profits. Because their products or their identity is not clear to people. Proper identity of any organisation is fully dependent on its marketing.

Now a days, we learned that many companies also failing to earn their expected profit because of their weak marketing strategies. Advertisement is one of the major part of any companies marketing strategies. As they failed to make the advertisements without delivering proper messages and proper information, they fail to earn profit.

A successful advertisement can take the sales to sky high, on the hand an advertisement with negative impact can pull down the image of an entire organization. This is why how an organization should approach their product advertisement should have a proper guideline. The way advertisement works are quite simple, consumers see ads and thus are persuaded to think, feel or do something differently as a result of the ad experience. But only a few ads actually get noticed. Others are randomly overlooked just like when we look at a fruit we overlook the tree. It is because the tree fails to capture attention so do those advertisements [5]. That is why this paper focuses on analyzing which of the factors make a successful advertisement.

A successful advertisement can remove any negative attitude towards any company or their products. For this reason, we come up with an idea to identify the key factors of the marketing process particularly identifying the factors of advertisements. We will also suggest which factors must be needed to make a proper advertisement. These factors will be identified from the public responses.

1.5 Objectives and Contributions

To achieve the organizational goal, traditional Marketing policy of an organization have been shifted to a strategic approach through a significant contribution of Technology. Nowadays technology can predict almost anything if enough data is analyzed. Some organization or in some cases entire market depends on advertisements and its success. Before a product is launched organizations research the market and come up with a plan to highlight their product to the target market. Our main aim is to predict the success rate of the advertisements and analyzing the factors using machine learning algorithms. Previously, there were some works on advertisements but here we are focusing on the relation between human emotion and advertisement while buying any product. For predicting the success rate and analyzing the factors, we are working on dataset which is done on the basis of a group of people's responses.

The SVM, a machine learning algorithm will help us with text characterization. There will be a survey form with some question about the ads we will be showing to the audiences. The questions will be designed in such a way that they cover almost every aspect of the ads and can capture how the audiences actually felt about them. Then using SVM we will get the pattern from which segment of people prefer which part of the ad and to target that particular segment what might be a suitable approach [6]. Though we are also using different types of algorithms also. Such as, Naive Bayes, Logistic Regression and many different algorithms which will be described later. These algorithms show us the different rates of purchasing the product and the pubic impression towards it. Positive impression will accelerate the public feelings to buy a product.

1.6 Thesis Structure

To identify the factors of an advertisement we primarily will conduct a survey. We will choose some popular ads and some not so popular ads from both television media and printing media. There will also be some posters and banners. We will segment the audience into multiple sectors according to their age, gender and profession as the target market is usually set considering these[7]. Then we will show them some of the advertisements and try to read their reaction using Electroencephalography (EEG) sensor. We will also use a survey form for collecting textual data and analyze them using machine learning algorithms. These algorithms will be resulted on the basis of the impression of the advertisements and the chance of purchasing the product.

The human mind is a very complex system. It can process almost anything but reading it with our present technology is a bit difficult. So, some models and algorithms were introduced to study and analyze it. The inputs are the biosensor observations and outputs are the valence and arousal level induced by the video stimuli to determine the mood of the user[7]. This will help us know what the audiences are feeling while watching our selected ads.

1.7 Workplan

For the research purpose, firstly we set our work plan. Without a proper planning it would not be done. We had to read several research papers, we had to analyze the marketing processes, organizations policies regarding advertisements and finally the significance of human emotion in case of advertisements.

Table 1.2: Workflow of the research

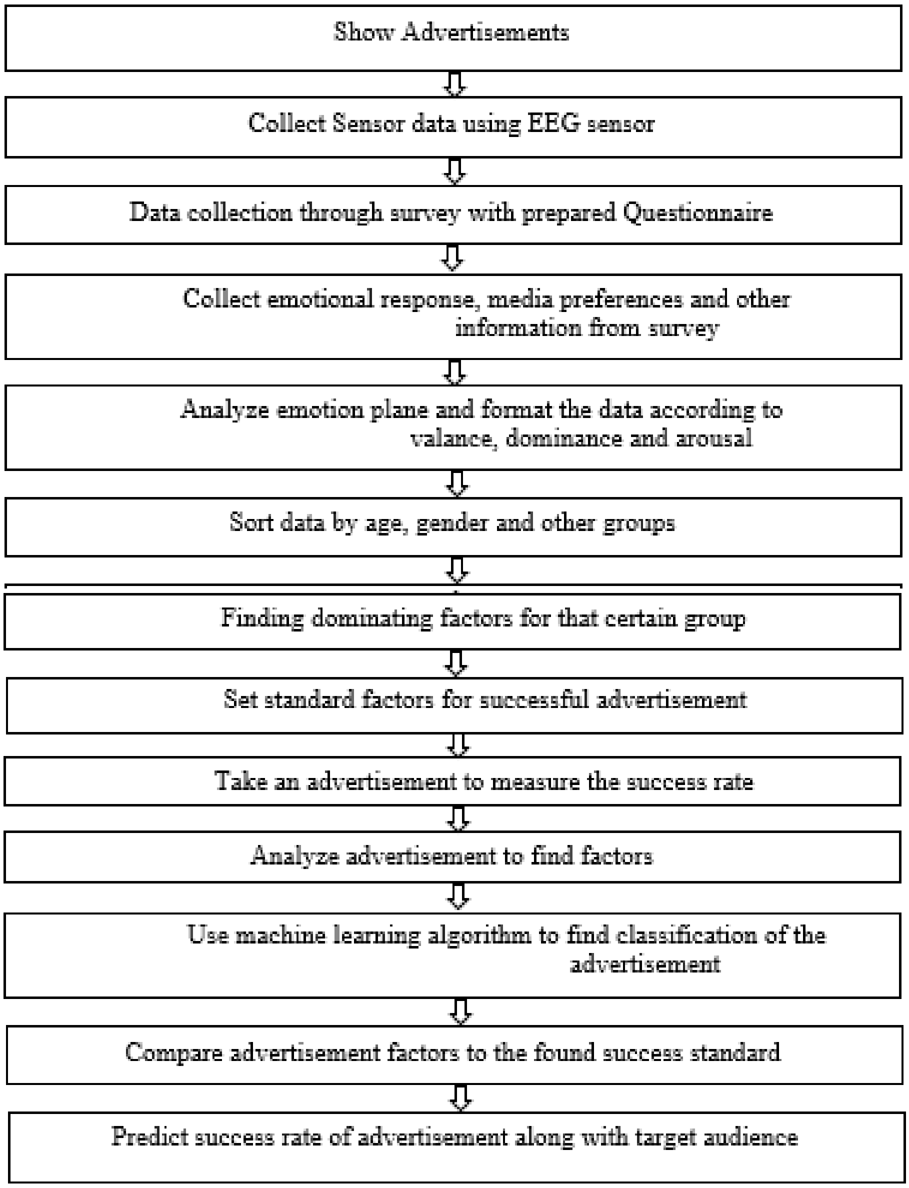


Table 1.2 shows the workflow diagram of the complete system. The system will follow the abovementioned steps to fulfill our purpose of predicting the success rate of any advertisement. First, we will take the advertisements selected for this research and show it a group of people. They will be connected with EEG sensor that will record their reactions to the shown advertisement. We will have the sensory data from this and to perform feature extraction. Then with the prepared questionnaire, we will take another survey to find the preferences of people. From analyzing the collected data we will find an emotional response, media preferences and other information. We will analyze the collected data and extract features, which then we will sort according to valance, and dominance and arousal plan for further classification. We will then take the data and sort it into different groups from which the factors of success will be found. Finding these groups will be done by Machine learning algorithm process. We will find dominating factors of groups that make an advertisement successful.

In the next step, we will take an advertisement of which we will try to predict the success rate. We will analyze that particular advertisement and find the most dominating factors that will have the most impact on its success. Then using SVM, KNN, Logistic Regression, Decision tree etc. algorithm to find the same class to which the advertisement belongs to. That class will have the set of the standard factor for it which are the contributor to its success. Using that algorithm, we will compare the two-factor groups, see the percentage of similar factors they have and based on their matches, our system will predict the success rate of that advertisement along with target audience, preferred media, and other information. We also work on human emotion using EEG sensor. The relation between human emotion and advertisement success predicting is our one of important part of this research. We will also focus on the brain activity signal of human during watching an advertisement. People may have different types of emotion change in different advertisements. We assume that people sometime buy products attached emotionally after watching advertisements. If people attached emotionally, they will have different brain signal or different brain activity will be shown during watching advertisements. This kind of change in emotional state is recorded in a dataset in csv format by using EEG sensor.

Chapter 2

Related Work

2.1 Literature Review

Advertising is a large and profitable industry, and advertisers aim to view their products or services as valuable as well as highly attractive and rewarding. In [8] it is mentioned that, for a multi-billion dollar business like online advertisement, forecasting ad clickings through levels (CTR) is a huge training challenge. In setting up a deployed CTR prediction system, they have presented a wide range of case studies and themes from recent experiments. Those included enhancements in conventional supervised learning focused on an FTRL-Proximal online learning algorithm (with outstanding convergence properties) as well as the use of per-coordinated training speeds. In this paper[8] it has also addressed some of the problems that occur in a real-world system that might seem to be outside the mainstream area of machine learning science. These consist helpful space saving techniques, methods for successful assessment and visualizing, pragmatic methods for presenting condense forecasts for expected chances, methods for testing, and methods for automatic function management. Eventually, they have also listed some ways that, following promising findings elsewhere in the literature, did not turn out to be favorable. The purpose of this paper is to illustrate the close relationship between conceptual advances and functional technology in this industrial setting, and to explain the profundity of problems that occur when conventional machine learning approaches are implemented in a complex dynamic environment. A number of case studies from recent research in setting the applied framework used by Google to forecast ad clicking through supported search advertising rates were discussed in this study. Because this arrangement of concerns is now well known, they have chosen to focus on a variety of topics that have received little attention in a functioning system but are equally valuable. They discuss memory savings issues, performance analysis, prediction condensation, calibration, and functionality management with the same thoroughness that is typically given to the problem of developing an effective learning algorithm. This technique has been checked as mentioned. First, they trained on real data with a ground truth model, but using features very different from normal. Rather, the existing click labels have been eliminated and new labels have been checked using surface-truth model estimates as the real CTRs. This is important because knowing the true labels is required to determine the accuracy of a condensation method,

$$e_t = |\sigma^{-1}(p_t) - \sigma^{-1}(p_t^*)| \tag{2.1}$$

Where Pt* was the real CTR (model of ground-based truth).

In [9] they have worked on significance of clustering of customer in U.S. wine market. Over the last 15 years, the U.S. wine market has been growing steadily. From 2688 in 1999, the number of wineries rose to 8862 in 2016. Approximately 7 percent of all wineries are allocated in the Mid-Atlantic region, including New Jersey, New York, and Pennsylvania. However, as the market is growing, competition has also risen. Now, many foreign wine companies from Europe, South America, and Oceania sell or plan to sell their products to the fast-growing U.S. wine market to catch up to the rushing pace of the business. Promoting the local wine industry in the United States is critical. In this context, this research [9] sought to forecast the Mid-Atlantic wine market segment-based on purchasing behavior, perceptions, and social demographic attributes. A Cluster Review utilized Class 1 Detractors, Class 2 Enthusiasts, Class 3 Positive and Class 4 Advocators to classify the Mid-Atlantic wine industry into four clusters. Here clustering the customer segment was signified to make their advertisement and marketing easier. According to their findings, Class 1, detractors are the most common group to purchase local wine. Neatly, 67.4 percent of Detractors said they had never previously purchased local wine. 2. Enthusiasts and Class 4 Advocators are Mid-Atlantic's target market for local wineries and we should also pay more attention to these two market segments. 74.5 per cent of Class 2 recorded buying wine from the wine region of the Mid-Atlantic. Approximately 60 percent of Class 4 advocates said they had previously bought local wine. Class 4 features are very similar to Class 2 features. Class 3 Neutral's probability of buying local wine is 50/50. In other clusters, they drink and buy wine more often than consumers. Usually, unless they want to grow their reach outside Class 2 and Class 4, it is not encouraged that Mid-Atlantic wineries should enter this market segment. Assessing consumer perception will help producers, wholesalers, and retailers target ultimate consumers and specific segments of the market. Cluster Analysis, a two-way contingency table, and Chi-square independence tests were performed to study the differences between the four market segments by them in this research [9]. The independence test for Chi-square is to test if two of the variables are associated.

In [7] quite similar work has been done but on human resources of hospital management industry. The purpose of this research is to investigate factors influencing management decisions to implement HRIS (Human Resource Information System) in the hospital industry in Bangladesh — an emerging developing country. This paper incorporates two influential acceptance theories— Human-Organization-Technology fit (HOT-fit) and Technology-Organization-Environment (TOE) paradigm to understand this issue. 13 Factors in four dimensions are examined to examine their impact in hospital decisions on HRIS adoption. A total of 550 copies of standardized questionnaires were distributed through the non-probability sampling method to HR executives from 92 private hospitals in Bangladesh. Among its partners, there were 383 available questionnaires indicating a positive response rate of 69.63 per cent. Based on the initial implementation of HRIS, we classify the sample into three core groups, namely adopters, prospectors and laggards. The results obtained suggest 5 of the most critical factors, i.e. IT services, resources for top management, personnel IT skills, potential cost, and competitive pressure. Moreover the technical dimension accompanied by the organizational, human and environmental aspects is the most important aspect of the proposed 4 dimensions. Ultimately, the study found substantial differences between different adoption classes in all aspects. The study results often reveal concrete ideas for increasing the likelihood of introducing HRIS for scientists, clinics, and state. This study has significant implications for understanding the implementation of HRIS in developing countries.

This paper [10] showed us, emotion measurement, advertising leaves on viewers were performed by making ad emotion model based on subjective human opinions as well as objective multimodal features and showing how successful ad emotion modeling can have a positive impact on real-life applications. First, they assembled a sample of 100 advertisements quite carefully; they analyzed the efficacy of this ad data set to elicit emotions through audiences in a coherent manner. At the end of this step, they compare the effective opinion of 5 experts and 14 novice annotators and find that the 2 groups are highly concordant. Second, they are Convolutional Neural Network (CNN)'s utility in encoding emotional audio-visual features. Because the compiled ad dataset is fairly small and inadequate for CNN research, they use domain adaptation to transfer successful knowledge gained for modeling ad emotion from the LIRISACCEDE film dataset. Extensive experiment confirms that the synthesized CNN descriptors outperform popular audio-visual featured proposed especially for valence recognition. Third, they demonstrate how accurate ad emotion encoding can promote the automated incorporation of advertisements into video streaming. Previous works that are discussed in this paper are I Affect Recognition (AR) and (ii) the influence of emotions that are evoked. Here, both material and user-centric approaches include labels to classify stimulus emotion and these labels are collected from reputable annotators whose successful opinions are usually acceptable given the emotional interpretation of human subjectivity. Content-centered AR approach predicts the emotion induced by analyzing audio-visual signals in the stimuli examined and user-centered AR methods predict the reaction evoked by evaluating user physiological changes. They compile a control collection of successful ads in this paper that produce concordant affective opinions from experts and inexperienced consumers. They also synthesize descriptors of emotions based on CNN, which are found to outperform audio-visual characteristics. They also conduct a user analysis to demonstrate how encoding can make the ad-insertion process easier to improve the user experience. Examine how effectively they elicit emotions through audiences to determine the efficacy of the advertisements as control stimuli. They rated all ads on a 5-point scale, which ranged from -2(very unpleasant) to 2 (very pleasant) for valance, 0 (calm) to 4 (highly aroused) for arousal and 0 (boring) to 4 (highly engaging) for engaging.

In [11], this article discusses a case study concerned with the use of clustering approaches to consumer behavior analysis in the food market through data mining.

The data obtained from the questionnaire survey of the Institute of Marketing and Exchange of Mendel University's Faculty of Business and Economics in Brno was subjected to various types of algorithms for cluster research to identify market segments. The aim of this research is to define the possibilities in the problems of these approaches and explain their suitability or inadequacy to solve these issues. Consumer behavior is the multidisciplinary subject, since there is no self-discipline that can provide a comprehensive view. Experimental psychology focuses on examining the role of the substance in cognition, thinking and memory processes. Clinical psychology explores the role of the consumer in emotional adaptation, microeconomics investigates the role of the product in the distribution of individual and family capital, social psychology analyzes the role of the product in the actions of the buyer as a representative of social groups, and sociology finds a solution to the role of the product in social institutions and group relationships. Customer clustering is most important components in modern and successful marketing leading to improved customer relationship management in this paper, they proposed a method of clustering of customer based on information about the goods and they use a genetic algorithm to enhance the quality of cluster. Data collection and storage focuses on implementing multiple methods whose logical order allows complicated behavioral analysis or product decision making, hence the application of forecasts and professional projections of future development. Weka technology was used as a method for data analysis. Weka is a machine learning software written in Java (Waikato Platform for Information Analysis). The study concluded that cluster analysis can be extended to the amount of such attributes, but not all forms of approaches are suitable for this reason. But, if the findings are important, the researchers need to work with a consumer behavior specialist. It would be valuable for this type of data to check certain methods, such as setting up group rules.

In [12], it is mentioned that business clustering is a critical event in the administration of sales and business interactions. A product seller is classified in traditional marketing based on general characteristics such as the statistical information of consumers and their lifestyle features. Current mass marketing strategies are no longer reactive to the variety of customer needs. This heterogeneity needs to be managed by clustering approaches that position consumers with the same need and related shopping activity in the same clusters. Companies may deliver goods, services and relevant assets to consumers through close relationships using the appropriate clustering. Consumer clustering is one of the components of current and efficient advertising which contributes to better customer relations management (CRM). When choosing the correct variables, clustering is an important issue. Cluster variables have two central variables and product-based parameters. General variables include statistical information (age, race, etc.), lifestyle, and factors based on customer buying patterns. Although there may be information available, it may vary over time. For e.g. by using public variables, employment, marital status, profession and similar cases render the clustering more questionable. In this paper we suggest a system of consumer clustering centered on product information. The selection process was meant to convert and combine information from one or more databases into one set of data. In this article, they have proposed a model of consumer identification focused on factors such as shopping cases and financial information relevant to the experiences of consumers. Clustering and clustering value feature was further identified as one measure of similarity. Genetic algorithms have been used to maintain clustering precision.

In [13] it is portrayed that the emergence of private asset networking peer-to-peer networks has introduced new connectivity possibilities outside control. While observing continued growth, academic literature has recently started to examine individual customer habits towards such new ways of use. Based on studies on the fundamental market motivations for peer-to-peer sharing, this study uses cluster analysis to identify prototypical users. They also identified five main dimensions (concerns, rewards, product-specific aspects, social aspects, and possession-related aspects) based on customer-related data from a large-scale online survey (n=745), they identify five key dimensions (concerns, rewards, product-specific aspects, social side, and ownership aspects). They define 4 types of customers with distinct demographic and attitudinal profiles on these grounds: Social Enthusiasts, Conflicted Materialists, Skeptic Ascetics, and Individualistic Agitators. Based on population differences and contact patterns between these classes, we draw lessons for professionals to tailor their business models and marketing strategies to the particular motivational behaviors of the respective user groups. The paper describes the sample data and the methods used to define the cluster research experiments. They draw on data collected in November 2015 at (blinded for review) within the context of a large scale online survey in Germany. The study examined reasons for and against Internet users utilizing sharing networks. Participants was drawn from a topic pool of the University. Participants was confidential. A series of questions they are required to answer about their personal attitude to P2P platforms. 58 reflective objects for 18 structures were included in the collection (see list in the appendix). Up to five objects reflect each build. Through voicing maximum rejection (1) to maximum agreement (7), each issue was measured on a 7-point Likert scale. Things are shuffled in and between topics to exclude series effects. Step by step clusters represent different outcomes in their clustering process. The first cluster reflects the strongest advantages, product-specific, and social values, while it assigns no significance to the facets of possession and sets the lowest value of issues. The second group was especially distinct from the first cluster in terms of their interests. It's not that this team doesn't see the positives of exchanging, they see many potential drawbacks as well. Each class frequently reflects the highest score for factors relevant to possession. Across two important ways, the third group varies from the others. First of all, issues linked to possession play a much less prominent position for this band, which may indicate a more post-materialistic, less flamboyant style of life. Each group represents the lowest absolute values in the advantages, problems and product-related elements of the three dimensions. Finally, the fourth cluster members vigorously resist the social aspects of sharing. They did not seem to understand the exchange of profits at all (both from a general and product-related point of view) and express concerns about the renting of P2P.

In [14] they talked about bio-medical sector but the training data and decision making was worth studying. Jackknife's cross-validation method was used for identification in this study. It is a widely used method used successfully in the past to verify forecasting accuracy. Data are divided into F folds in the Jackknife sample. F-1 folds engage in training, and test groups belonging to the remaining fold are expected based on F-1 folds learning. This method of sampling is replicated F times and the grade of each specimen is estimated. For identification, a 10-fold cross-validation system has been used in this study. They have figures introducing the cross-validation process of Jackknife, splitting information into 10 folds. Except the last one that can house less than S/10 samples, each fold hosts S/10 samples. Linear, RBF, sigmoid and polynomial SVM kernels were used in this work. After an extensive experimental method, SVM kernels are chosen. In addition to SVM, they have tested several other classifiers such as decision trees, KNN, and PNN, and found that SVM kernels produce superior results in comparison with other classifiers. We have therefore chosen four variants of SVM for improved output.

In [15], for the study of active human systems, they have provided a multi-modal data set. Thirty-two participants registered the electroencephalogram (EEG) and peripheral physiological stimuli as each viewed 40 one-minute music video excerpts. In terms of levels of anticipation, valence, like / dislike, superiority and familiarity, participants graded each video. Frontal face footage has also been reported for 22 of the 32 participants. A novel stimulus collection approach is introduced utilizing retrieval from the last.fm database using active identifiers, image highlight recognition and an interactive analysis system. A detailed review of the scores of the subjects is provided during the study. Correlations were examined between the EEG pulse levels and the scores of the participants. Methods and findings were described using EEG modalities, peripheral physiological stimuli and multimedia content evaluation for single-trial assessment of anticipation, valence and like / dislike scores. Ultimately, the judgment integration of the outcomes of the assessment from the multiple modalities is carried out. The data-set that they made is made available to the public and we invite other scientists to use it to test their own effective methods of state estimation. The dataset provided discusses the idea of classifying the psychological aspects caused by displaying music videos for various users. According to them, the reactions to these triggers (music video clips) have never been investigated, and work in this area focused primarily on samples of photos, music or non-music video. In an adaptive music video-recommender, an emotion-recognizer conditioned by common physiological responses to material, music videos, is best able to achieve his target. To explore the correlations of the individual scores with the EEG signals, the EEG information are widely indexed, down-sampled to 256 Hz, and high-pass filtered using the EEGlab6 toolbox with a 2 Hz cutoff rate. We observed detrimental associations of anticipation in the group of theta, alpha, and gamma. The decline in central alpha power of higher anticipation is in accordance with the results of an inverse relationship between alpha intensity and the general level of enthusiasm. In all the frequency bands examined, Valence demonstrated the greatest associations with EEG measurements and correlates.

Chapter 3

Our Proposed Approach

3.1 System Model

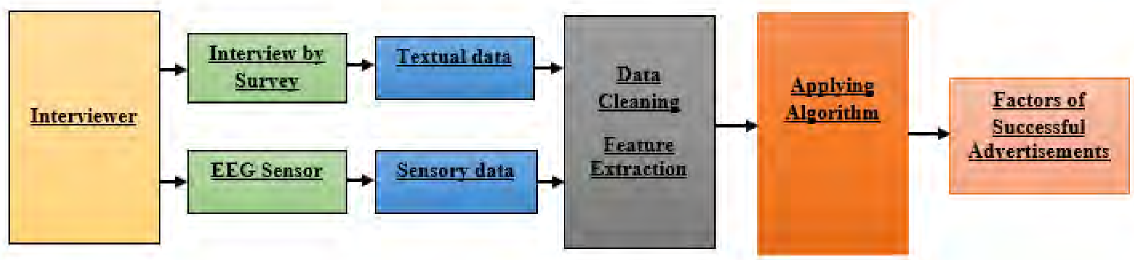


Figure 3.1: System Model of Finding Factors

Figure 3.1 shows that methodology of our research. Primarily, we are collecting data from general people like users and experts. The users are basically differentiate by their gender, age, income, how much time they spend in Medias etc. The experts are of two parts. One is the experts of organizations and another is advertise companies. During survey, we are also gathering the data for detecting human emotion by using sensor. The sensor is EEG sensor. We have two dataset. One is textual data where people can share their views towards any advertisement. Here we can find the factors of the advertisements or which factors are more engaging with the success of any advertisement. We will classify them according to duration, ad status, impression etc.

Another data set is of EEG sensor data. Here, we analyze the purchase of any product or the impression of any advertise is depend on the human emotion or not. We try to find out the human emotion is responsible for purchasing any product. We also take reviews on the advertisements pleasant level, dependency level and excitement level. According to these we will find the impression of that advertisement of a subject. The pleasant, dependency and excitement level represent the stage of valance and arousal of a person. The next phase is to cleaning the data. Here,

we will keep the important data from the survey reports and drop the unnecessary data. The clean process will be done by algorithms like feature extraction algorithm and finding out the relation between output and the factors of advertisement. So, we can easily identify the key factors of advertising the products.

Now, for processing the data we will use a Machine Learning algorithm, SVM, KNN, Random forest etc. For human emotion part we will process the data using these algorithms also. Finally, using these algorithm we will get the result of key factors of the successful advertisements.

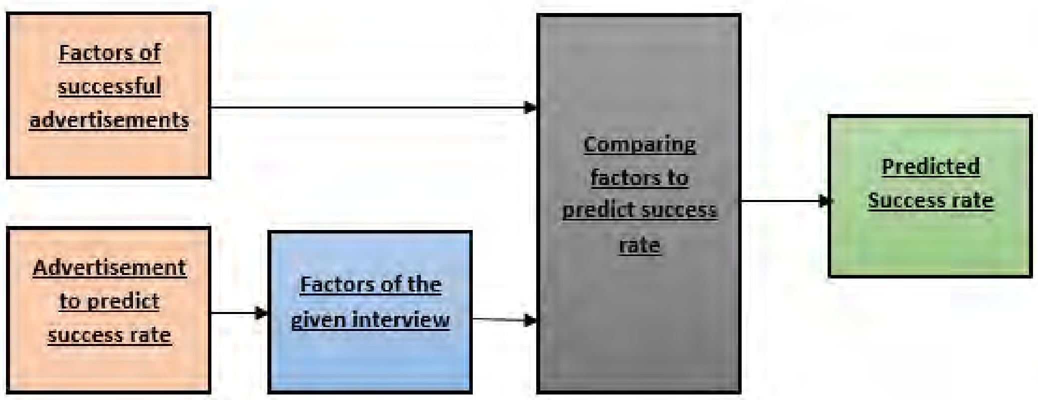


Figure 3.2: System Model of Success Rate Prediction

Figure 3.2 shows, the process of comparing the factors of the advertisements with the factors of standard advertisement. This will help to predict the success rate of advertisements. For predicting the success rate we need both textual data and the human emotion data collected through sensor.

Figure 3.3: Flowchart of the Research

Figure 3.3 shows how the machine learning model work for this research. Among all the results we will choose the best from different datasets based on Impression and Purchased.

Chapter 4

Dataset description

4.1 Questionnaire Description

For making our dataset we have to do a survey. We collect answers from the different types of people. They have to provide us answers for two different dataset. Textual dataset and Emotion dataset from EEG sensor. These answers help us to find out the actual factors of advertisements and their relations between impression of the advertisements and purchase rate of the product after seeing the advertisements.

Figure 4.1: Question Regarding Impression

Figure 4.1 shows a question where participants were asked that what type of impression they had about that advertisement. This is the primary question because if the participant had positive impression about the advertisement, they will think to purchase the product. They also have positive impression about the product also.

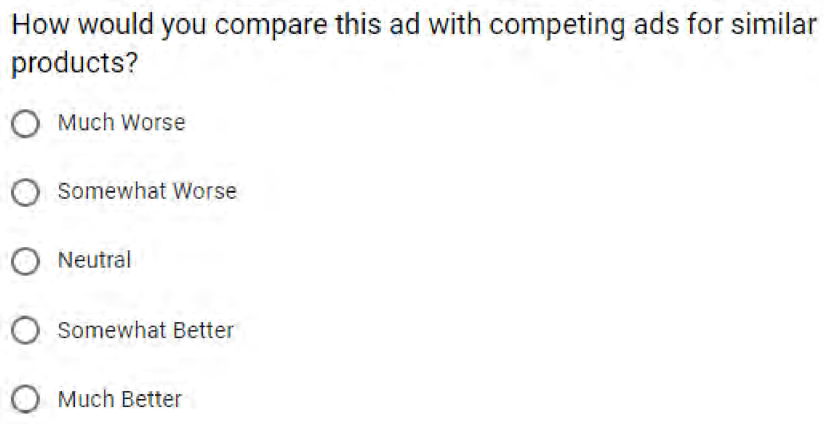


Figure 4.2: Comparison With Similar Ads

Figure 4.2 shows the comparison of the advertisement of the similar products. It will

help to learn in which extend people will have a positive feelings about a product. The differences of the advertisement of the similar product is discussed here.

Figure 4.3: Appealing and Engaging

Figure 4.3 shows the question about appealing and engaging. People are interested in such kind of advertisement is also our concern. If they are not concerned about appealing advertisement they will not purchase it.

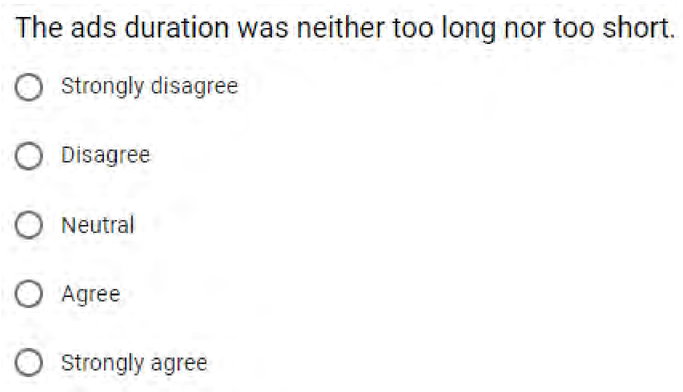


Figure 4.4: Duration of Ad

Figure 4.4 shows the question about duration of the advertisement. We wanted to know that people want long duration advertisement or short duration advertisement.

Figure 4.5: Conveyed Message

Figure 4.5 shows the message of the advertisement is concerning factor for participants or not. We asked them this question to know that they focus on message or the visual representation.

Figure 4.6: Convinced by Ad

Figure 4.6 shows the question about the advertisement is convincing or not. Here the inner meaning of the advertisement is important.

Figure 4.7: See Similar Ads in Future

Figure 4.7 shows the question that the participants want the same type advertise or the same concept advertisement or not.

The advertisement is believable.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Figure 4.8: Believable Ad

Figure 4.8 shows the message of the advertisement is believable or not and this is important to participants or not.

The ads message is relevant to me.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Figure 4.9: Relevant Ad

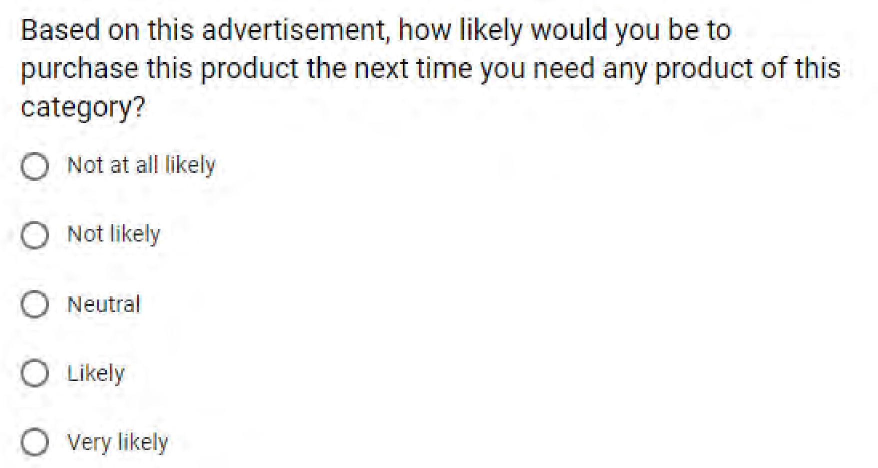
Figure 4.9 shows the message of the advertisement is relevant or not. If it is relevant, people will like it or not.


The benefits described in the ad are believable to me.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Figure 4.10: Believable Benefits in Ad

Figure 4.10 shows the benefits of the product is shown in the advertisement or not. According to this people will measure the advertisement.


Based on this advertisement, how likely would you be to purchase this product the next time you need any product of this category?

- Not at all likely
- Not likely
- Neutral
- Likely
- Very likely

Figure 4.11: Purchased Product

After all these questions Figure 4.11 shows whether people will purchase the product viewing these advertisements or not.

4.1.1 Textual Dataset

For the textual dataset, we divided the questionnaire in three parts. Basic information of the people over whom we do our survey, Advertisement details and SAM scale survey. All these data are kept secured. All the data are collected from specific questions. Textual dataset basically for our primary survey. We get people reaction on particular advertisement. Though we do not get any emotional response from textual data.

Basic Information of the people: The basic information we need for our research is the age, gender, salary, the medium of watching advertisements they preferred and the overall time they watch advertisements in the preferred mediums. We take survey from 170 people. They are from different genders, their age are different, they use different mediums for watching advertisements etc. All have not equal time for watching advertisements. So, everything has to be analyzed for categorized them. These basic information helps us to separate them according to groups. These information are collected through Google form.

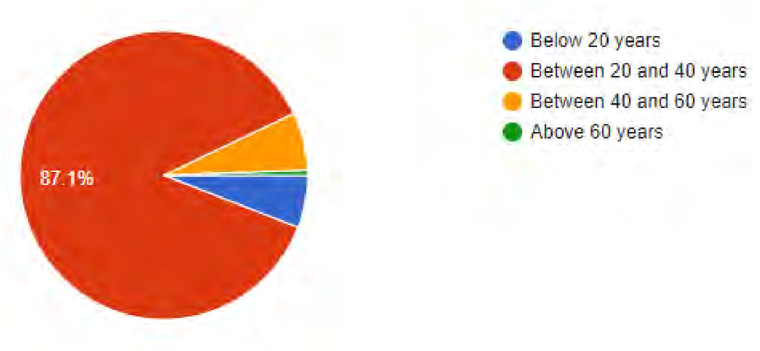


Figure 4.12: Age

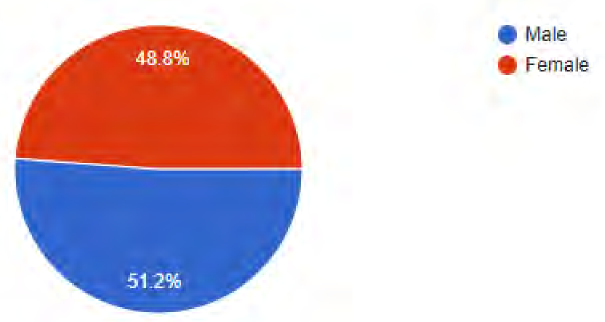


Figure 4.13: Gender

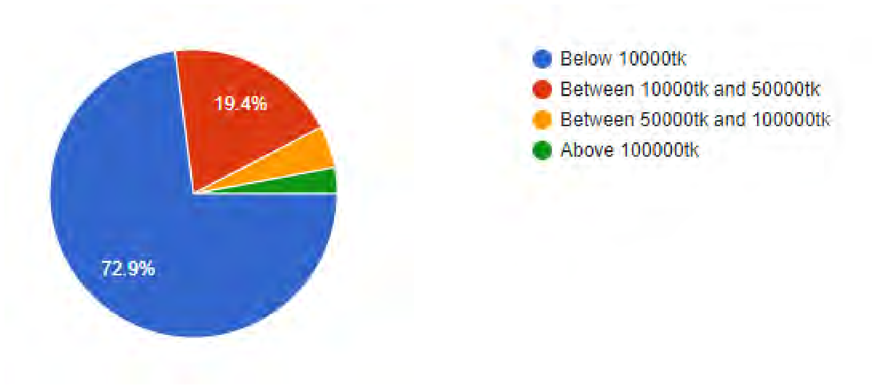


Figure 4.14: Salary

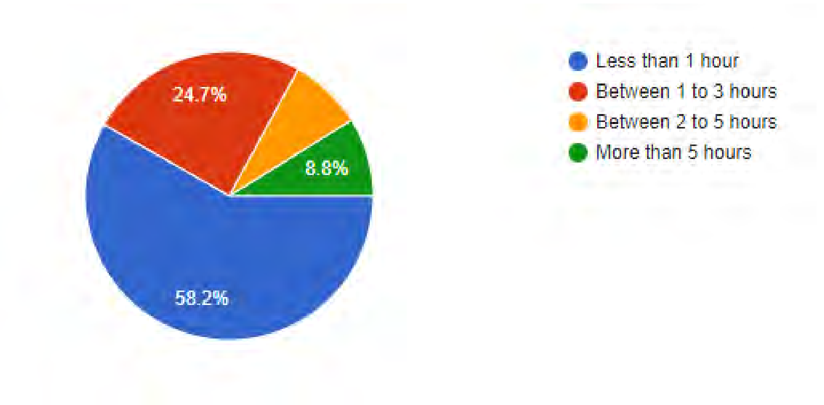


Figure 4.15: Duration spent for advertisement

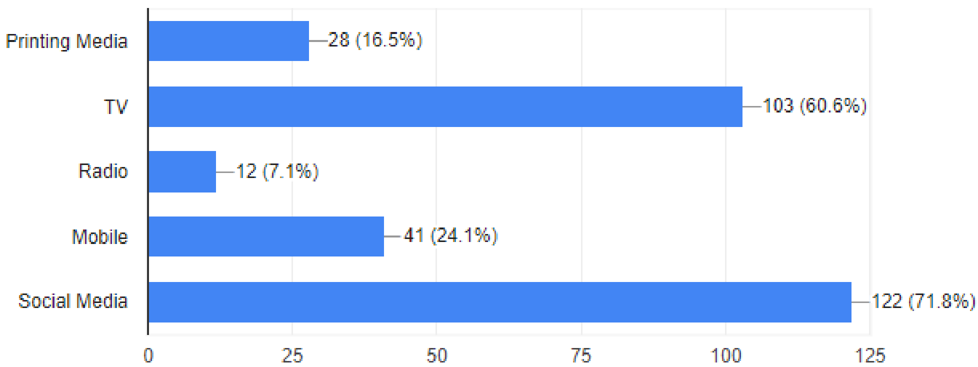


Figure 4.16: Medias participants mostly used for advertisement

In figures above we are seeing that the participants of different classes who participate in the survey. Different age of people participated as well as different salary range people are also participate in the survey. We also notice that the number of male and female participants are also almost same. The duration and the time they pass in watching advertisements are also showed in the pie charts and bar diagrams.

Advertisement Information: The advertisement information are for basically identifying the features which are necessary for making an advertisement successful. Here, the participants had to watch an advertisement and they had to answer the questions. We select the advertisements according to the people reviews. For analyzing the advertisement part, we classify the questions into three scales. Dominance, Valance and Arousal. These are the parts of Liker Scale. Each question bears any of the scales. All the questions are answered in five different types. Strongly Disagree, Disagree. Neutral, Agree and Strongly Agree. We analyze the valance, arousal and dominance according to these questions

Valance: Valance is positive or negative effects of any situation. The valance is marked as how positive the news or visual representation. It can be negative also. In our questionnaire some questions are asked to measure the valance.

The participants are asked that they would like to buy or purchase the product after seeing the advertisement. They were given five possible answers. Not at all likely, Not likely, Neutral, Likely, Very Likely. The lower mark is marked equal to 1 and higher is marked equal to five. Neutral is 3. The impression of the advertisement is also part of valance.

Dominance: Dominance is the representation of the dominant nature of the emotion or the controlling of the emotion. Emotion can be classified two types. Submissive and Dominant. For example, anger is dominant whereas fear is known as submissive emotion.

In our questionnaire the participants are asked that the message of the advertisements are understandable or not and the message conveyed in the advertisement is clear or not. These questions help us to measure the dominance of the information.

Arousal: Arousal is the measurement of excitement and calm of the information is. In our questionnaire the participants are asked that the advertisements are relevant or not and the message of the advertisements are believable or not. The answers are classified in five sections. Strongly disagree, disagree, neutral, agree and strongly agree.

SAM Scale Survey: The SAM scale survey is to measure the excitement, pleasure and the dependency level of the information.

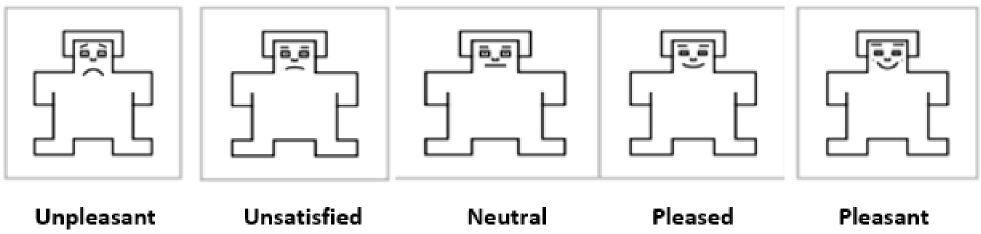


Figure 4.17: SAM scale of Pleasure level

Pleasure level of any information can be classified from Unpleasant to Pleasant. There are also Unsatisfied, Neutral and Pleased. These levels can be evaluated as what people are admiring from the advertisement. People may buy any product if the advertisement seem to them pleasant. If they are unsatisfied, they may not gain any positive attraction from the product. This product will not bring any profit for the organization at the end.

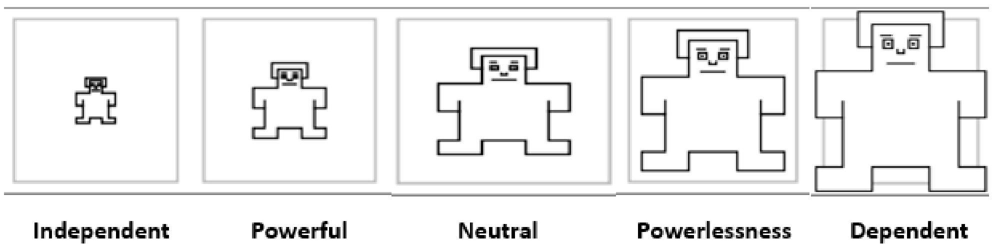


Figure 4.18: SAM scale of Dependency level

Dependency level is the measurement of how people are dependent on buying any product after watching an advertisement. People buy products for their own benefits. Sometime they purchase any product without seeing any advertisement or any type of marketing. They purchase that for their own purpose. When they see the advertisement and decide to buy the product, it will create the dependency of the product with the advertisement. More involvement of the advertisement, the product will get more public value. The higher dependency is named as Dependent and the lower is Independent. Less than neutral is Power and more than neutral is Powerlessness.

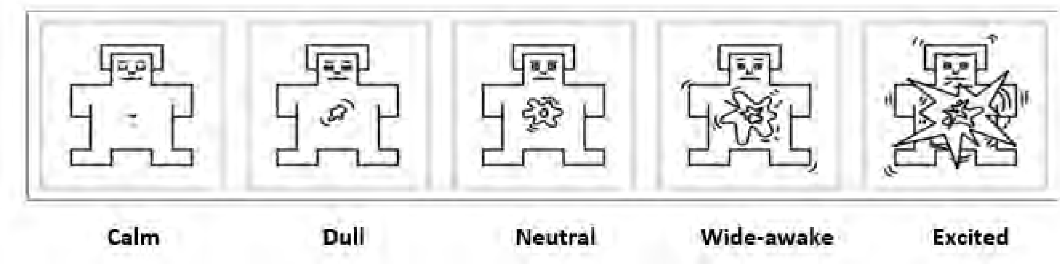


Figure 4.19: SAM scale of Excitement level

The excitement level of any advertisement is also helps to evaluate the purchase rate of busying any product or what is the impression of the product create in people mind by seeing the advertisement. Lower excitement is known as clam and higher excitement is considered as excited. Less than neutral excitement content is leveled as dull and more than neutral is considered as wide-awake. There are some advertisement might be of less excited but get good impression. This is the exceptional case. It depends on what is the content of the advertisement and what is the message the organizations want to deliver by this visual presentation.

Figure 4.17, 4.18 and 4.19 showing different types of SAM scale. It will help to identify the pleasure, dependency and excitement level properly and easily.

Emotionomics: Emotion in favor of logic and performance has been overlooked for far too long. Breakthrough in neuroscience have been shown that people are decision-makers mainly emotional. This idea has not yet been embraced by many businesses, dismissing emotion in favor of logic and performance. There were even fewer people acting on it. Emotionomics looks a motivation in terms of both consumer and workplace business opportunities. The emotional value of a product will make a difference in today's highly competitive marketplace where many goods look alike. In fact, there is no question that an organization with an emotionally involved workforce will gain competitive advantage. According to Dan Hill's "Emotion: Leveraging emotion for business success" is based on observations learned from facial coding, the only realistic way to measure and track emotional responses from customers and employees. It shows how passion for business success can be manipulated by branding, product design, advertisement, revenue, customer satisfaction and workforce management. Emotion counts, and this book will not only help move closet to customers and employees, but will also help move ahead of competitors [16].

EEG Sensor: The electroencephalogram (EEG) reads brain's electrical activity from human. In 1929, the first recordings were done by Hans Berger. But evidence of similar research work was found done on animals as early as 1870. The collected waveforms reflect an idea to mimic the activity of the brain's superficial layer, the cortex. Because of the electromagnetic anomaly from the human-brain structures below the cortex, this action came to live. Signals generated by the nerve cells are called motion potentials. They move through a bridge called the synapse from one brain-cell to another. Synapse connects the dendrites of each brain cell to another's. Different molecules called neurotransmitters help overcome the difference with the signals that might have been changed or lost. There are two forms of neurotransmitters. One supports the possible movement in traveling to the following neuron or in other words the destination cell, and the other stops moving it to some other nerve cell working as a checker. Simply speaking, the brain has to go through a lot of hardship to a balance among each of these neurotransmitters in the brain. EEG test measurement is quite low as it is written in micro Volts (V) with high scale frequencies up to about 30 Hertz (Hz) [17].

Table 4.1: Ranges of different waves

Different Types of Waves	Ranges of Frequencies (Hz)
14 Hz – 30 Hz	B wave
7 Hz – 13 Hz	α wave
4 Hz – 7 Hz	θ wave
Up to 4 Hz	γ wave

Beta Waves (frequency vary from 14 Hz to about 30 Hz)

Beta waves are most closely related to being attentive or active, aware and warning. Low-amplitude beta waves are correlated with power focus, or a distracted or nervous state of mind. Beta waves are also correlated with voluntary actions (movement inhibition and tactile response feedback). The warnings are often referred to as EEG beta waves when analyzed by an EEG unit[18].

Alpha Waves (frequency vary from 7 Hz to 13 Hz)

Alpha waves are often aligned with a realm of mind that is comfortable, quiet and coherent. It is possible to observe alpha waves in the occipital and posterior brain areas. Alpha waves can be generated by shutting one's eyes and relaxing, and during intensive analytical strategies such as meditation, mental calculation, and problem solving, they are never present. Alpha waves range in amplitude from 9 to 11 Hz in most people. These are often referred to as EEG alpha waves when measured by an EEG device [18].

Theta Waves (frequency vary from 4 Hz up to 7 Hz)

Brain activity is referred to as Theta activity within a frequency range between 4 and 7 Hz. The detected theta rhythm in EEG size is discovered regularly in younger adults, mainly over the temporal areas and at some point of hyperventilation. Theta pastime with amplitude greater than approximately 30 millivolts (mV) is considered less frequently in older people, preventing somnolence at some point. These are often referred to as EEG theta waves when measured with the help of an EEG device [18].

Delta Waves (frequency range up to 4 Hz)

Delta activity is mainly found in infants. Delta waves are associated in older subjects with deep sleep ranges. Delta waves in patients with absence seizures have been documented inter-racially (between seizures) that contain brief, surprising attention lapses. Delta waves are characterized by excessive waves with low frequency (about 3 Hz). During wakefulness, delta rhythms can exist— they are sensitive to eye-opening and can also be enhanced by hyperventilation. These are often referred to as EEG delta waves when calculated using an EEG unit [18].

Small metallic discs referred to as electrodes are positioned on the scalp in distinct positions. These positions are identified by using the records who measures the head the use of the International 10/20 System. This relies on taking measurements between positive constant points on the head. The electrodes are then placed at points that are 10 percent and 20 percent of these distances. Each electrode website is labeled with a letter and a number. The letter refers to the vicinity of talent underlying the electrode. There are a fantastic range of electrodes that can be used. The majority are small discs of stainless steel, tin, gold or silver blanketed with a silver chloride coating. These commonly have a lead attached. Alternative strategies consist of a cap in which the electrodes are already embedded. When neurons are activated, nearby currents are produced. EEG measures the current that flows for the duration of the excitations of the dendrites of many pyramidal neurons in the cerebral cortex. Potential differences are triggered by summed postsynaptic potentials from pyramidal cells that create dipoles between soma and apical dendrites.

Conventional analog instruments carry with it electronic equipment, a meter and a piece of writing device. A meter may be a coil of wire within a force field. The signaling from the electronic equipment passes through the wire inflicting the coil to oscillate. A pen mounted on the meter moves up and down on every occasion the coil moves. The pen attracts the trace onto paper moving below it. The electronic equipment output is controlled by high and low-frequency filters and sensitivity controls. The high and low-frequency filter values can set the window inside that the graph activity is recorded. This is often called the information measure. The sensitivity controls the scale of the activity displayed. For example for instance, as an example a sensitivity of ten V/mm implies that a symptom with an amplitude of one hundred V can manufacture a one cm vertical deflection. The speed at the paper moves on also will have an effect on the looks of the waveforms. A digital electroencephalogram system converts the wave into a series of numerical values. This method is understood as Analogue-to-Digital conversion (ADC). The values keep within the storage device, manipulated so re-displayed as wave-forms on a visual display unit. The speed at the wave information is sampled so as to convert it into a numerical format is understood because of the rate. A second issue that affects the accuracy of the wave is sampling skew. Sampling skew happens once all channels aren't sampled at the same time. Several digital electroencephalogram systems sample channel one initial, then sample channel a pair of, then channel three, etc. The break between sampling of every channel is understood as sampling skew. To cut back the sampling skew, some digital systems use burst mode sampling. This will increase the speed between consecutive channels sampling so as to cut back the quantity of sampling skew.

The demonstration is a third issue concerning the accuracy of the optical electroencephalogram wave. A monitor display's accuracy depends on how many points or pixels the area unit can be produced. Due to the screen resolution, the number of pixels that can be obtained is increased. The resolution of the screen is delineated in numbers representing the pixels within the horizontal and vertical axis.

EEG signals that are digitized is manipulated to vary the paste-up 'on-line' at the time of recording or 'off-line' when the recording is completed. This 'remonstrating' is accomplished by recording all electroencephalogram channels with a typical reference conductor. Notwithstanding the paste-up accustomed show the information whereas it's being recorded, information is kept into the pc memory in common reference model. This enables the information to be displayed victimization totally different montages at a later time. Since digital systems store the analog signal as numerical values, remonstrating could be an easy subtraction method which ends in cancellation of the common reference.

Figure 4.20: EEG Sensor

Figure 4.20 shows the EEG sensor which we used in our research for survey. It helps to find out the change of human emotion. The change will be found in visual graphs. There are different channels in this sensor which will process different types of data. These data are actually the graphical representation of human reaction.

For collecting data we make a survey procedure. 28 people was present in this survey. We connected the device with a Bluetooth from EEG sensor to laptop for operating the sensor. There was an application for operating the sensor. The name of the software is EmotivPro. We can see the output here. At first, we check the connectivity of the sensor. We tried to connect 100 percent the channels. Because of sometime technical difficulties and connection problem we got 95-96 percent connectivity.

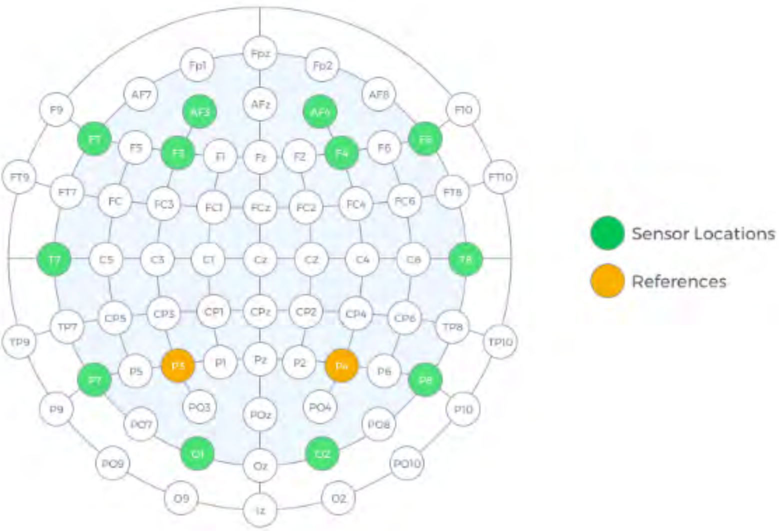


Figure 4.21: EEG Sensor Channel and reference

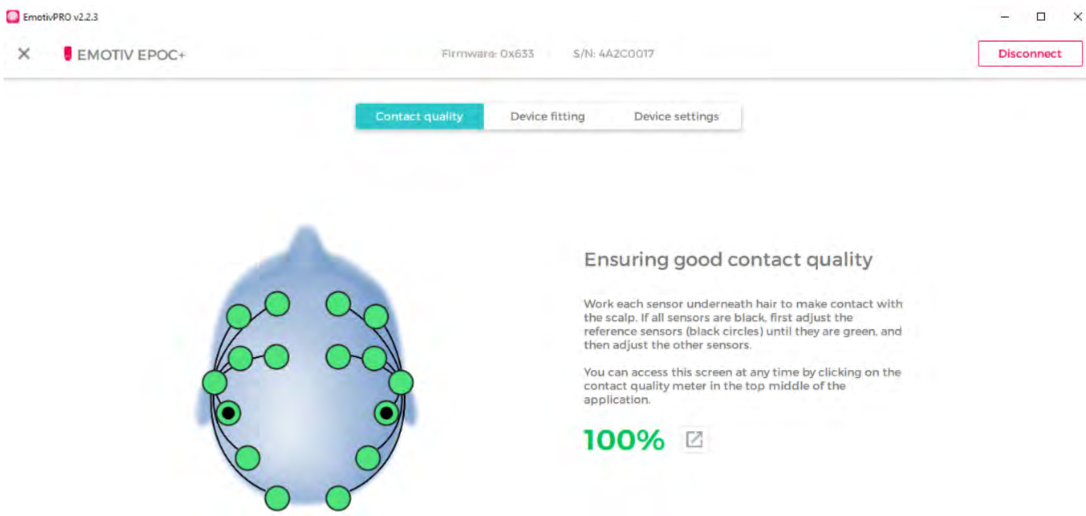


Figure 4.22: Connectivity test of EEG sensor

Figure 4.22 shows the connectivity check of the EEG sensor in EmotivPro version 2.2.3 application. As we get 100 percent connectivity, all the channels are showing green. There are two points which are black color surrounded by green circle. These two points are considered as reference point. We had to set the reference point first. Then we set other channels.

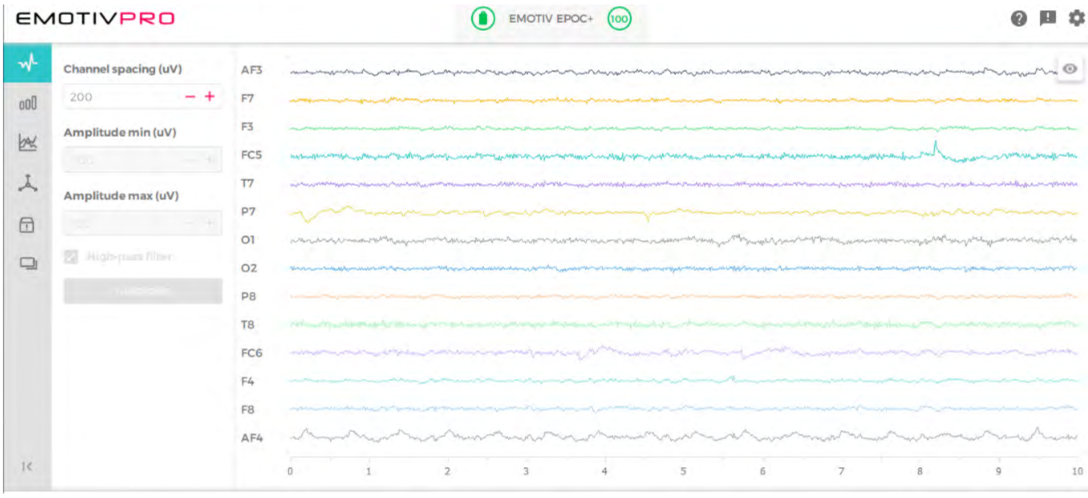


Figure 4.23: EEG Channels signal

Figure 4.23 shows the signals of different channels of EEG sensor. The signals fluctuate according to the human reaction. As we showed 10 videos to the participants, they had different emotional stage for different advertisement. All the reactions are recorded by this sensor. The sensor signal was converted into numerical value. These value are the measurement of different emotional state of human.

4.1.2 EEG Sensor Dataset

For making the EEG sensor dataset we had to make two datasets. One is brain activity signals dataset by using EEG sensor and another one is the textual dataset of recording the purchasing, impression and SAM scale survey. In this survey 30 participants were present. But 2 participants could not finish the survey. One participant had connectivity problem and another participant could not finish because of running out of charge. So we extracted both participants dataset. Then we worked with 28 participants' brain activity data.

We showed them 10 videos. There was an interval period in between two videos. The interval time was between 50 to 60 seconds. In this interval period the participants had to answer five questions regarding that particular video. The significance of that interval period was to remove the previous emotion or brain activity the participant had from earlier advertisement. This is also called as Neutral period.

In each advertisement the participants were asked five specific questions. So, they had to answer 50 questions for 10 advertisements. We also get sensor channel data from the EEG sensor. We are working on basically arousal and valance. The sensor data processed 4 types of wave's alpha, beta, gamma and theta.

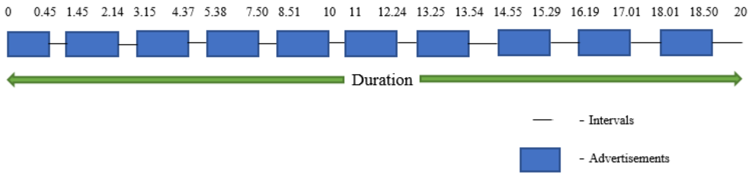


Figure 4.24: Mapping of advertisements

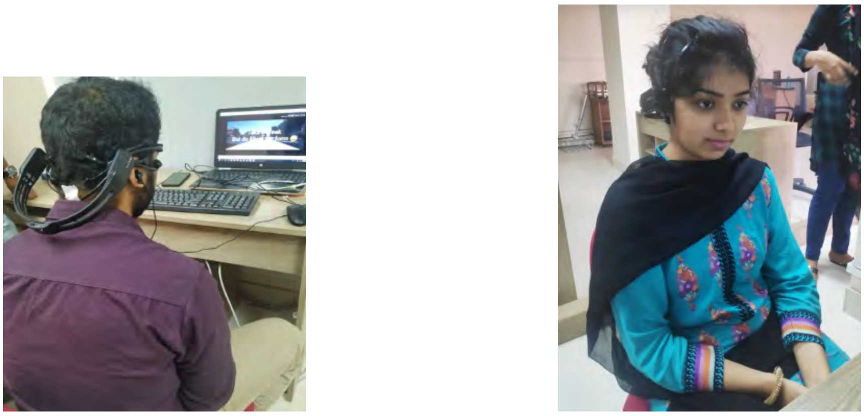
In data collection of EEG sensor, we mapped the total video. In initial phase when the sensor started to record data it take some steps to keep the participant in neutral stage of their emotion. The participants had to open their eyes 15 seconds, then off their eyes for 15 seconds. And 5 seconds for starting the video and recording start. So 35 seconds is considered as preparation phase.

After the preparation phase the first advertisement started. Its duration is 45 second. After watching the advertisement the participant had to answer 5 questions according to the advertisement in the neutral period. The brain activity signals recorded in this neutral period is out of our thesis purpose. We extract that interval part. Then the second advertisement started. Same procedure will be followed rest of the part.

4.2 Dataset Description

4.2.1 Data Collection Envioronment

As we have two types of dataset, we collect textual dataset from online and offline survey. For EEG sensor data collection, we collect data in a silent room and participants felt very comfortable.



(a) Subject 1 (b) Subject 2

Figure 4.25: EEG Sensor Data Collection

Fig. 7 (a) and (b) shows the data collection environment of EEG sensor.

4.2.2 Textual Dataset Description

There was no dataset available for our research as it one of the very first researches in this sector. So we had to create this dataset from scratch. So we created some fields to gather the information we seek. We thought of the questions that might arise against what general people may feel while watching an advertisement. Primarily we conveyed a survey on the questions we made and created the dataset based upon that. We took the name, sex, and income range to identify which criteria of people are interested in which type of advertising and products. We kept particular fields to extract the exact feelings viewers have while watching an advertisement. Questions like "have you seen this advertisement before?", "how many times have you seen the advertisement of the product" focuses on if viewers liked the ad in general. "Based on the advertisement, how likely would you be the next time you need a product of this category", "if the ad was visually appealing and engaging?", "if the duration of the ad was too long or too short?", these questions highlight if the ad was proper structurally. "If the ad's message was understandable?", "if the ad's message was relevant?" are the questions to check the transparency of the actual message it was hoping to give. And the pleasance, dependency and excitement level on SAM scale determines how the viewer's felt emotionally while the ads played. The overall goal of this was to find out the attachment between viewers or consumers and advertisements and extract what we need for our research.

These are some general data found from survey. We made the survey on 170 different people.

After the survey we started to make a dataset. Considering our research and our main object of the research we build the dataset. It contains fourteen major features which is considered as the main features of the advertisements.

Mainly we tried to solve two problems. One is finding out the purchase rate analyzing the factors. Another is finding out the features of creating positive impression of advertisements among people. Though all the advertisements must not have these factors which lead them to become a successful advertisement.

Impression is a factor which indicates the positive and negative view of people to an advertisement. It also helps to measure the reaction of any product of people. If people gain positive attitude from the advertisement of the product people will have positive impression. Negative otherwise. In our research it is one of the key factor of advertisements.

Ad status means the difference among the advertisements of same product. We may know about people will buy the same product of another company by watching another advertisement or not. The comparison among the advertisements of same product. This factor is very important while purchasing a product after viewing the advertisement.

Appealing and Engaging is the reflection of how much it creates appeal in human mind for buying the product. Engaging indicates the visual representation of the advertisement is capable of catching public emotion or not. Positive appeal of the products may play an important role to purchase the product and also may create a good impression. People is attracted to these advertisements which are more engage with them, with their regular life. So, this factor is also important of advertisement.

Duration means how much time is allocated for viewing the advertisement. The duration of any advertisement depends on the subject of the ad. In case of advertisement of products people like short videos with a message. On the other hand, in short documentary or social message type advertisements sometimes make lengthy advertisements. This type of advertisement actually is the summarization of a large but effective social message. So, duration of the advertisements is an important factor in predicting success rate of advertisement.

Conveyed Message is the message which is delivered by the advertisement about the product. This is the most important part of an advertisement. If the message of the product is not cleared to the people to the customers, they will not be inspired to purchase the product. When new product is launched by any organization people should not be able to know about the product at first. If the product is represented properly with a proper message, people will have their attention towards the product. Most importantly among all the factors, conveyed message is the most valuable part of an advertisement. A proper message should be delivered by an advertisement.

Convinced means people are ready to buy this product or not. If they are not convinced they will not buy the product. So the advertisement should have to be convincible. The ad makers try to deliver proper message, proper visual effects, and proper usefulness in the advertisement for grabbing people attention. If people are convinced, they will be convinced to buy the products. So, an advertisement should need to have convincible power.

Preferred means people would like to watch such advertisements future or not. An advertisement is already exist and it has many factors. If people like the factors of the advertisement and if they want to watch such type of advertisement in future, the factors present in the advertisement will be considered as important factors.

Understandable means the message delivered by the advertisement is perfect or not. Some advertisement may show unrealistic scenario which is not easily believable or the things are not actually present in real world. So people will not get proper info from the advertisement. The message should be clear and understandable is the key factor of an advertisement. The main motto of an advertisement is to aware people about the advertisement. If people will not understand the message, they will not buy the product.

Relevant is how much relatable the advertisement with the product. The advertisement should be relevant to the product. For example, you are watching an advertise of a mobile but it is viewing other things rather than mobile it will not relevant. We intend to buy any product only when we get relevant and proper message from the advertisement of the product. So the advertisement should have to be relevant to the product.

Believable benefits means the message delivered in the advertisement is believable or not. If any advertisement is viewing the advantage of using any product more than the products capability, it will not be a believable advertisement. Proper advertisement has always show actual benefits of a product. We are looking for those advertisement always where we find the proper message with good visual representation. So this factor is also help to find out to predict the success rate of advertisement.

Pleasure is the measurement of the advertisement is pleasant or unpleasant. It is an important factor of an advertisement. If people may not find any pleasant message from the advertisement, the advertisement will not get proper attention.

Dependency is the purchasing of the product is how much dependent on the advertisement. More or less we buy the products by watching the advertisement. Because this is the most common way of spreading the info of any product to the people.

Excitement is how much excited people feel while watching the advertisement. Actually the excitement level is depend on the advertisement's visual presentation and the inner meaning. Greater presentation will gain perfect people attention.

Purchased is the feature which indicates we will buy the product or not. This feature is dependent on above factors. If the factors are important and the factors are present in any advertisement, people will purchase the product. People purchase products for their own benefits. Without proper advertisement positive impression of any product will not be created. People may not be able to know the products details by their own. When the organizations stared to do marketing of their product in different Medias, people get the info about the product. If the organizations maintain these factors in their advertisements, people will have positive impression of the product and they will be insisted to buy this product. Proper advertisement can grab people attention towards the product. The purchase rate of the product will increase.

As we mentioned in our questionnaire description we used liker scale for answering our questions related to advertisement and SAM scale survey. There were Strongly Disagree, Disagree, Neutral, Agree and Strongly Agree. We evaluate the answers 1 as strongly disagree, 2 as disagree, 3 as neutral, 4 as agree and 5 as strongly agree.

Considering the evaluation, we make a dataset where all the values will be 1 to 5. Our training set for each dataset will be Purchased and Impression. In each dataset we will work on both of these columns. This is one of our dataset. Our dataset is where all the values are evaluated 1 to 5 without the Purchased column. Here the column Purchased is either 0 or 1 and the neutral value 3 is considered as 1. Here 1 is interested in purchasing and 0 is not interested. Again, there is another dataset we worked on the neutral value of Purchased column will be considered as 0. The same goes for Impression column for once the neutral impression value will be 0 and another time the neutral value will be 1. We also work on a dataset where all the values are 0 and 1. Where, 0 is negative response and 1 is positive response. In this case we also considered first the neutral value is 0 and again the neutral value 1. In this case 0 is also considered as negative impression or negative purchase intent and 1 is positive impression or purchase intent.

We divide the dataset among these section because we want to analyze the each and every condition regarding this dataset. That may help us to get perfect accuracy or preferable factors of advertisements.

4.2.3 EEG Sensor Dataset Description

The EEG sensor dataset is made of the values of sensor data and the reviews we got from participants such as impression of the advertisements and the purchase rate according to the advertisements.

As we mentioned the participants need to watch 10 videos and they had to give review according to that. There were 30 participants in this survey. 2 participants could not complete the survey because of technical disturbance and internet connection. Among 28 participants 4 participants were female and 24 were male.

We calculate the values of mean, median, entropy of each advertisements of each participants. We also work on 28 channels of the sensor. So we will get 3 values from each channel like mean, median and entropy. Then we relate the impression and purchase with the sensors channels values.

4.3 Feature Selection

Among all those features we will find out the most important features which will make an advertisement perfect. We know that our features are impression, ad status, conveyed message, preferred, appealing and engaging, pleasure, dependency, excitement, purchased, understandable, duration, believable benefits, relevant and convinced.

We will find out the important features using Chi square algorithm. This method is used to select the important factors of an advertisement. We will find out the important features for considering two features. Impression and purchased. Predicting these two feature will be our main aim of this research. Not all the features are equally dependent on both impression and purchased.

As we found the features which are primary features should be included in an advertisement. After finding these features we will operate a feature extraction algorithm. We did the same for the emotional dataset. There we asked 5 questions for each video. One participant had to watch ten videos. So one participant answered five questions for each video. There we find also find five important features. Impression, purchased, pleasure, dependency and excitement.

In the sensor value part, we also extract features. There were some features which were not related to the out thesis purpose. We extract them and kept the features of different channels signals. These signals are basically brain activity representation. The channels are basically keep record of the data. By analyzing these data we will find our expected output.

4.4 Methodology

The data we used in our research is collected by a survey. The survey data is collected by Google form. All the data are sorted in an excel file. It helps to represent the statistical data. We also use the Pie chart or Bar diagram for statistical assessment. For the survey, we set up a questionnaire. These questions are related to people choice, people need and people behavior.

After collecting the feedbacks, we select some specific features for predicting the highest rate of purchasing specific products. So, among a lot of features, we work on some specific features of our dataset. It is the part of our pre-processing of our dataset before applying classifier algorithms. Before preprocessing of data, we cleaned the data. Some data are unnecessary as they are not relevant to our research.

Firstly, we conduct a survey for data collection. Then split the data. Here the preprocessing stage started. We cleaned the data and remove the unnecessary data. Then converts all the string type data to numerical data.

After splitting the data, we use Cronbach's alpha and Shapiro Wilk test for questionnaire analysis. It also determines the reliability of dataset. In time of analyzing dataset, we found every features are not important to determine our factors. So, we use Chi square and Random Forest standardization methods for feature extraction. So that, we can easily identify which features are appropriate for us in our research. These are the phases of our data processing.

We also use Chi Square method to analyze the importance features of the dataset. As we are predicting the impression of the advertisement and the purchase rate of the product, the Chi square method is used the features vs the training set. When we train Impression the others features relation are compared among them according to impression. Same goes for the feature Purchase also.

The next phase was training and labeling data. This is useful for making dataset more accurate. It will help to gain good accuracy score. Better accuracy indicates better performance of the dataset. Finally, to get accuracy and result we use four classification algorithm. These are Logistic Regression, KNN, SVM and Random Forest. Each of the classifier shows different accuracy score and confusion matrix.

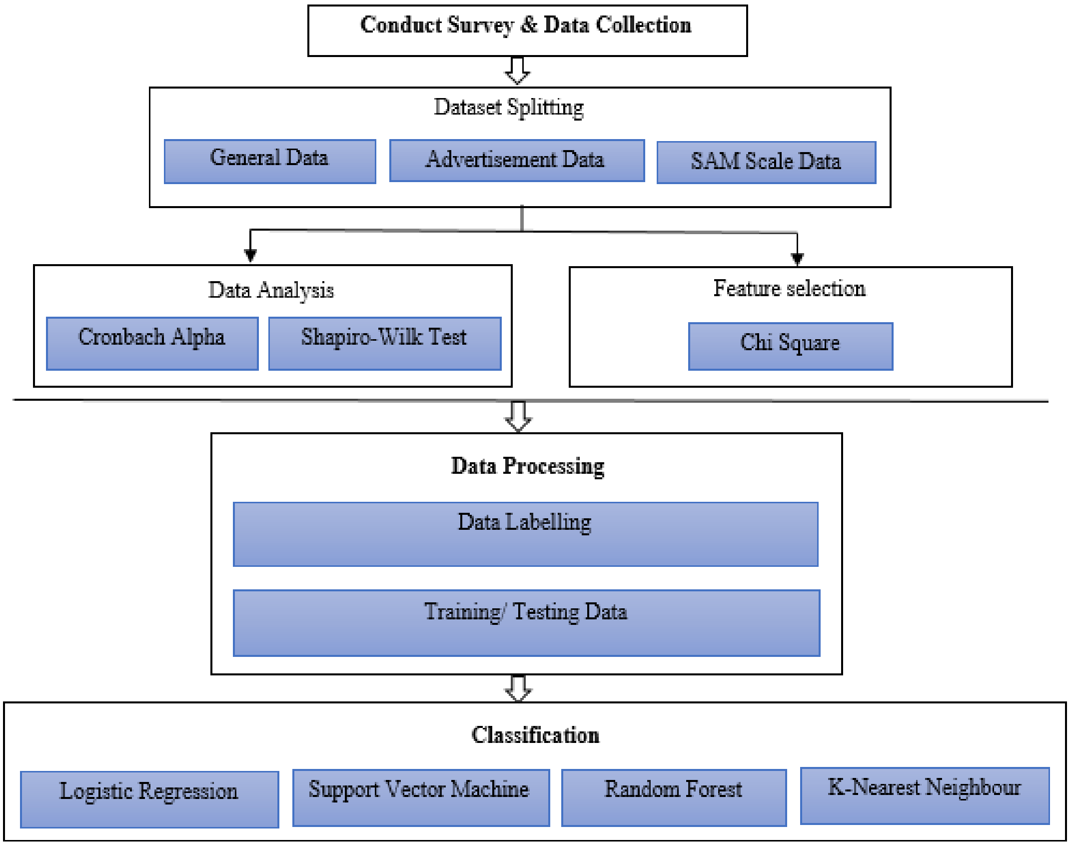


Figure 4.26: Working methodology

Figure 4.25 is the working methodology of our research. Here we describe the algorithms we use for questionnaire analysis and feature extraction for our research. We also discuss about the classification algorithms in the diagram.

4.5 Dataset Visualization

QQ PLOT: The Q-Q map, or quantile-quantile plot, is a statistical tool that helps us decide whether a set of data is derived from a logical mathematical distribution such as regular or exponential.For example, we can use a common Q-Q plot to test the hypothesis if we run a statistical analysis assuming that our dependent variable is normally distributed. It's only a visual test, not an airtight proof, so it's a bit subjective. But it lets us see how valid the theory is, and if not, how flawed the assumption is, and what data points contribute to the violation. A Q-Q plot is a scatter plot developed by plotting two sets of quantiles against each other. If both sets of quantiles come from the same distribution, the points should be seen forming an almost straight line. Here's a standard Q-Q plot instance when both quantile sets come from normal distributions.

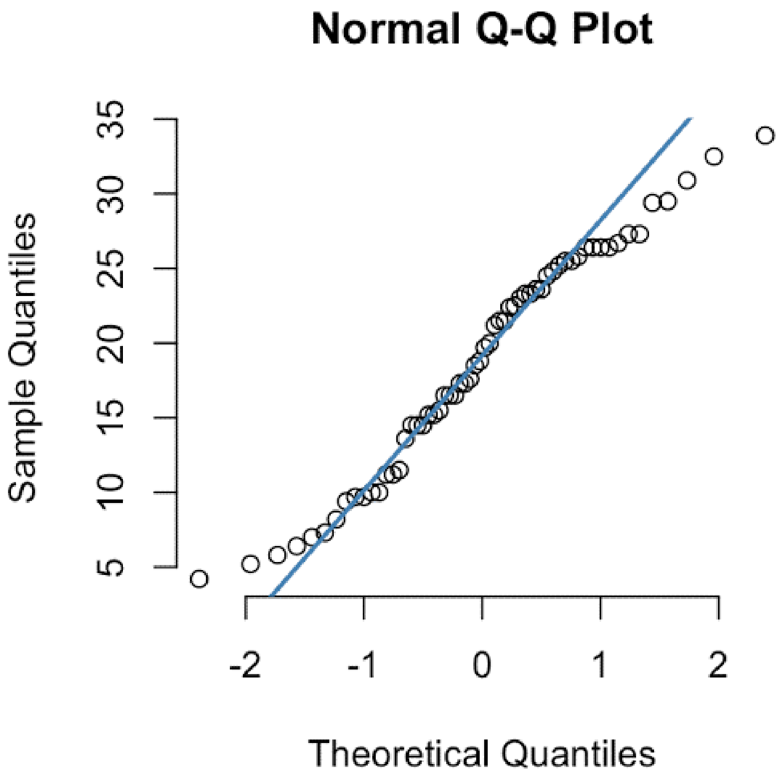


Figure 4.27: QQ Plot

Therefore, quantiles are essentially all ascending ordered data, with different data points being marked as the point below which a certain percentage of the information drops. Nevertheless, there are many methods to measure quantiles. Q-Q plots take sample data, arrange it in ascending order, and then plot it against the hypothetical distribution measured quantiles. To match the size of sample data, the amount of quantiles is chosen. Although generic Q-Q plots are the most commonly used in practice due to so many statistical methods which presume normality, Q-Q plots can be produced for any distribution [19].

Heatmap: A heat map is a representation of two-dimensional data in which color values are represented. The graphical heat map provides an instant visual overview of the details. More comprehensive heat maps allow viewers to understand complex data sets. There may be a variety of methods for displaying heat maps, but they all have one thing in common. They use color to express data-value interactions that would be much harder to understand if they were shown numerically in a graph. [20].

Line Plot: A line plot is a graphical representation of data along a number line with Xs or dots registered above the data representing the number of instances in the data collection that a response occurs. The frequency is expressed by Xs or lines. There will be an outlier in a line plot. The outlier is a statistic that is much bigger or much less than the other data set figures. A line plot is a horizontal line that is the same length x-axis. To offer the reader a summary of what is being shown, it is essential for a line to plot to have a title and an x-axis symbol. There must also be legends with line plots that illustrate what is being calculated [21]

Histogram: Histograms are data distribution graphs designed to show data centering, dispersion (spread) and form (relative frequency). Histograms can provide a visual display of large amounts of data in a tabular or spreadsheet type that is difficult to understand. These are used to explain how a process's performance corresponds to customer expectations (targets and specifications) and to better answer the question if they can satisfy customer requirements [22].

Horizontal Bar Graph: A horizontal bar graph is a rectangular bar graph with lengths and heights equal to the values they represent. They reflect the data horizontally graphically. This indicates the types of data being measured on one side of the graph. The other axis reflects the values of each type of results. The length of each bar is equivalent to the size of the type of results, and all bars extend from left to right [23].

Pair Plot: Pair plot builds relationships in a dataset pairwise. This feature generates an Axes grid so that each data vector is spread across a single row in the y-axis and across a single column in the x-axis by default. The diagonal axes are treated differently and a map is drawn to display the univariate distribution of data in that column for the component. A selection of variables can also be shown or various variables can be illustrated on the rows and columns. Although this plot alone can be helpful in a study, painting the statistics dependent on categorical variables will make it more important [24].

Scatter Plot: For two separate numerical variables, a scatter plot uses dots to represent values. The horizontal and vertical axis location of each dot shows values for an individual data point. To observe the relationship between variables, the scatter plot is used. The primary uses of scatter plots are to track and present relationships between two numeric variables. The dots in a scatter plot not only show data points values, but also trends when the data as a whole is taken. Depending on how near points cluster sets together, we can classify data points into classes. Scatter plots can also reveal if there are any unforeseen data gaps and outsiders. The dots in a scatter plot not only show data points values but also trends when the data as whole is taken [25].

Line charts: Line charts are suitable over time to show patterns. A simple example would be how a certain company's stock value grows on the stock market over rime. It does not need to be time along the x-axis, however. Any data that is consistent with the variable on the X-axis can be plotted as a function. Line chart emphasize time flow and shift rate rather than quantity. If you have continuous data you want to show through a diagram then a line diagram is good option. This graph is particularly effective in trying to identify a trend or pattern in your data such as seasonal effects and significant changes over time [26].

1. Textual dataset where all the values of the features are classified 1 to 5.

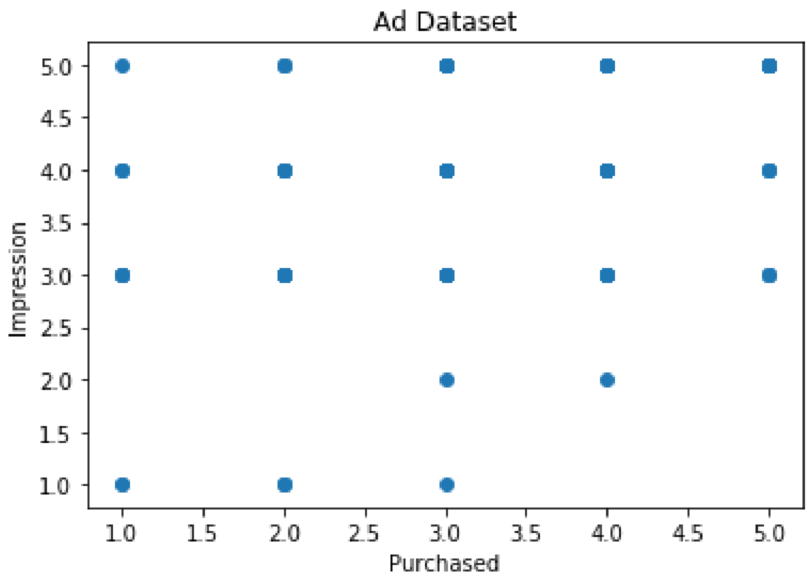


Figure 4.28: Scatter Plot

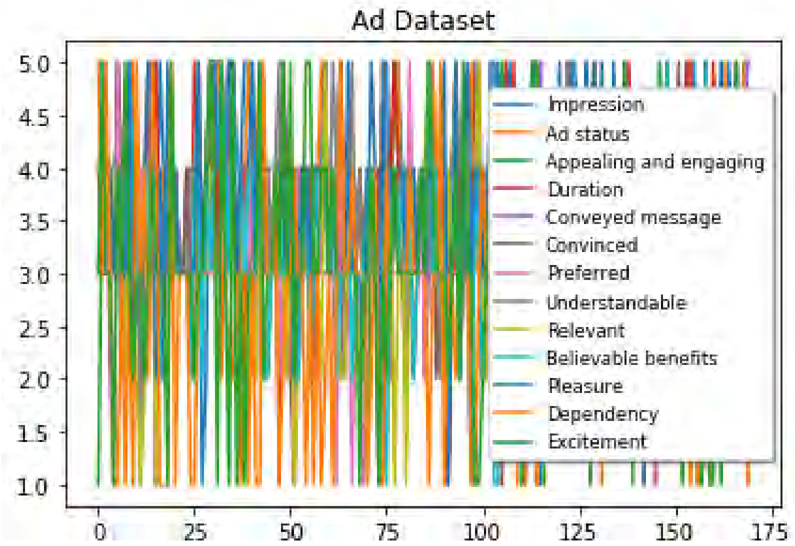


Figure 4.29: Line Plot

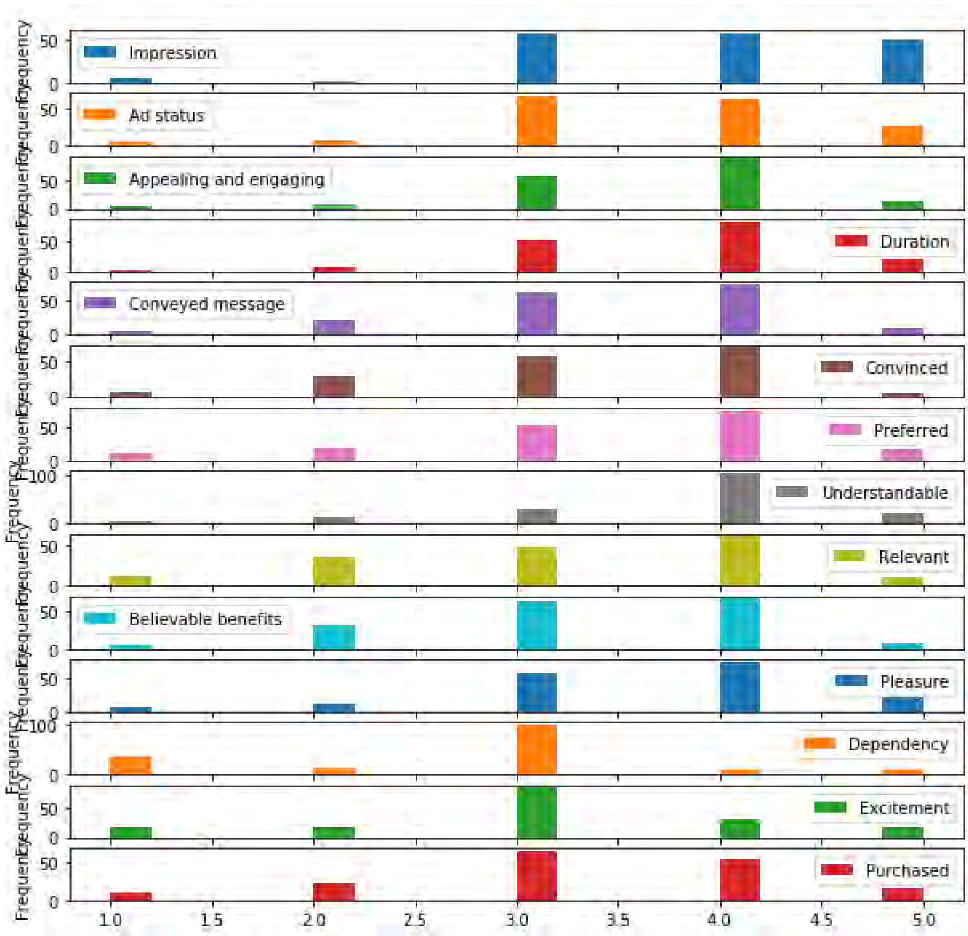


Figure 4.30: Histogram

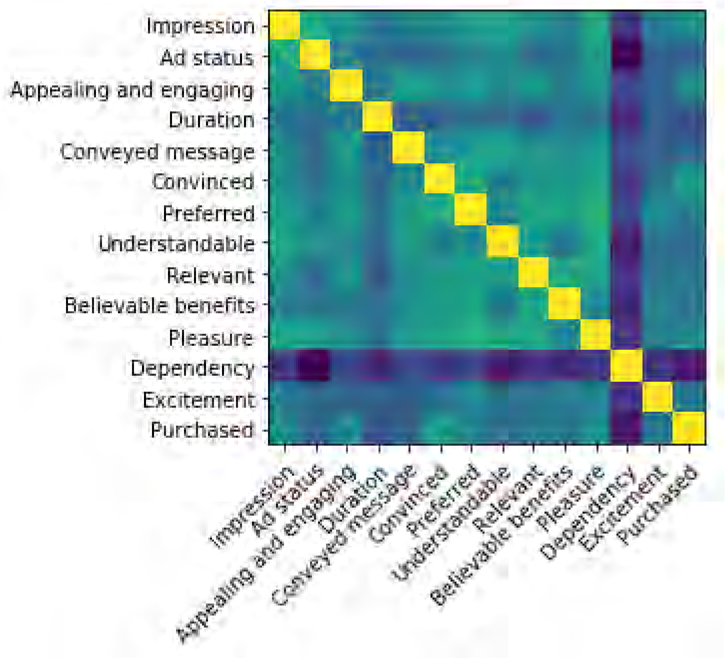


Figure 4.31: Heat Map

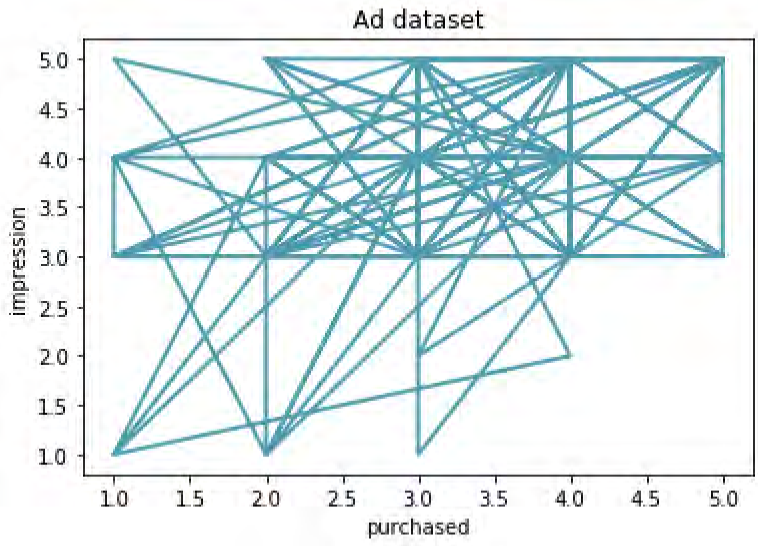


Figure 4.32: Line Chart

Textual dataset where all the values of the features are classified 1 to 5. Based on Purchased.

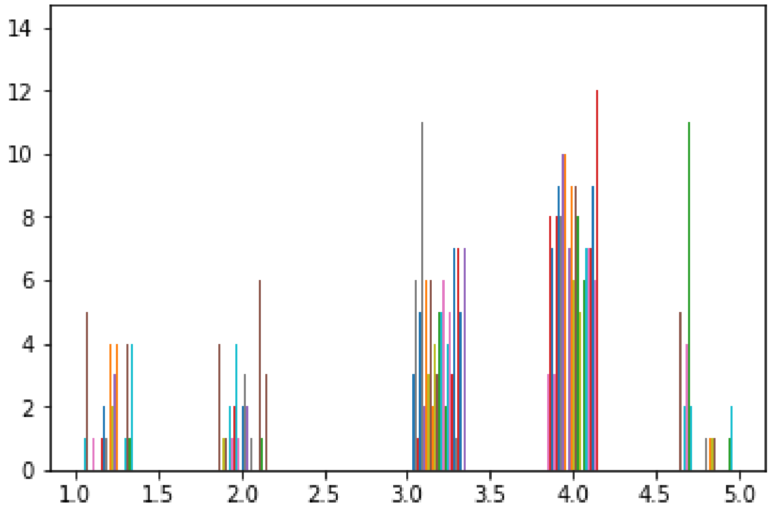


Figure 4.33: Histogram for Purchased

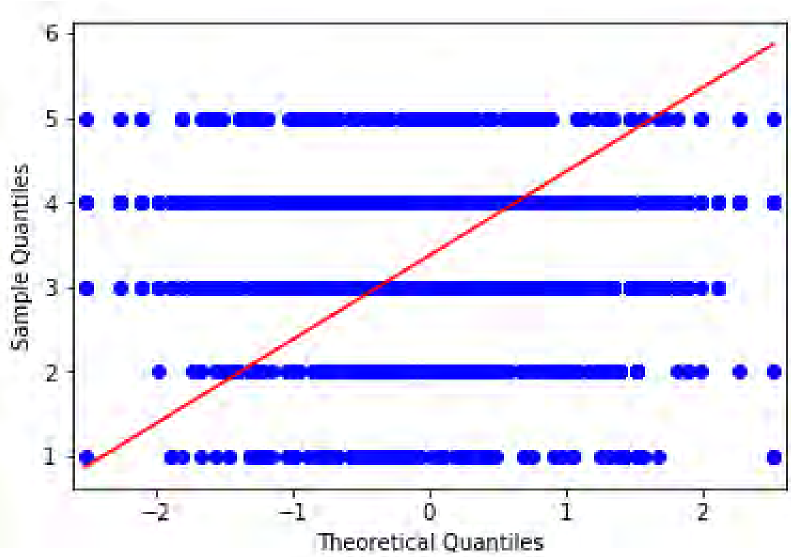


Figure 4.34: QQ Plot for Purchased

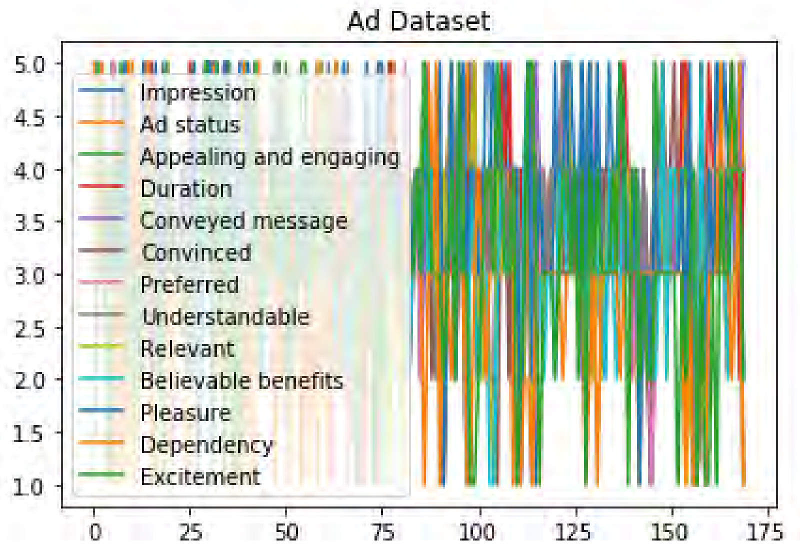


Figure 4.35: Line Plot for Purchased

Textual dataset where all the values of the features are classified 1 to 5. Based on Impression.

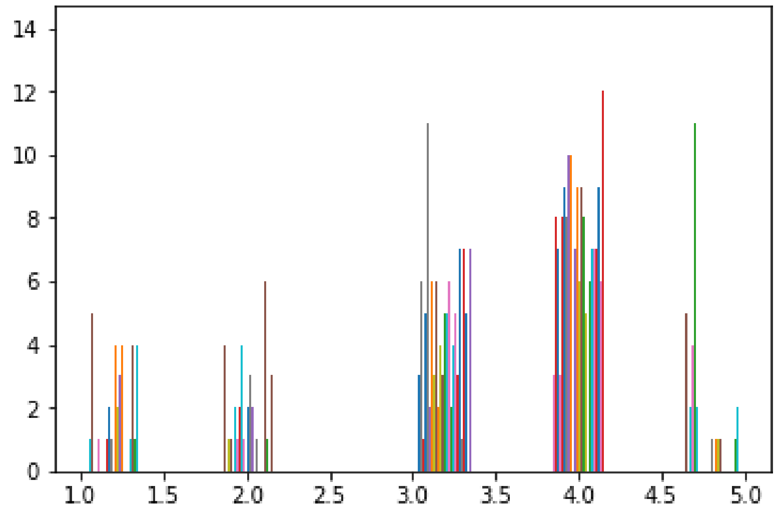


Figure 4.36: Histogram for Impression

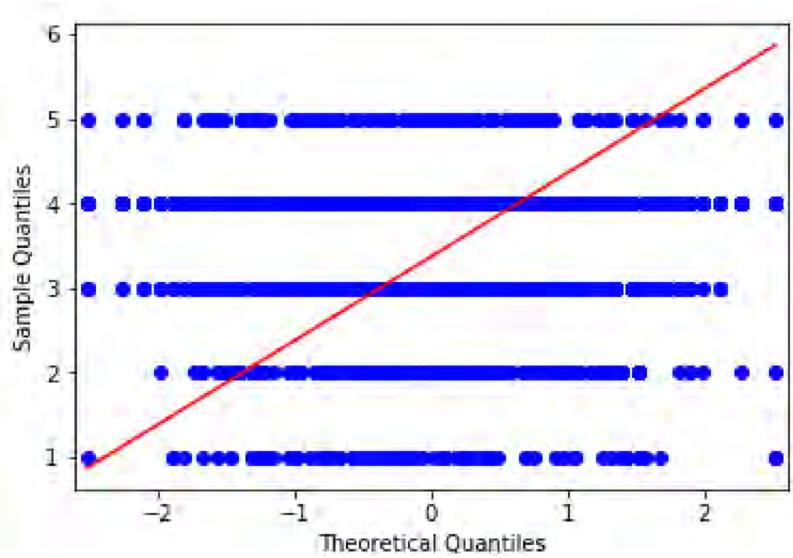


Figure 4.37: QQ Plot for Impression

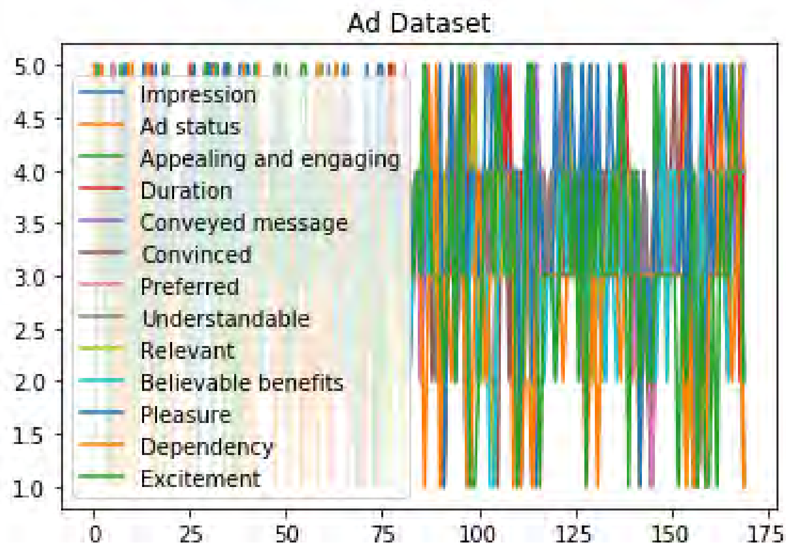


Figure 4.38: Line Plot for Impression

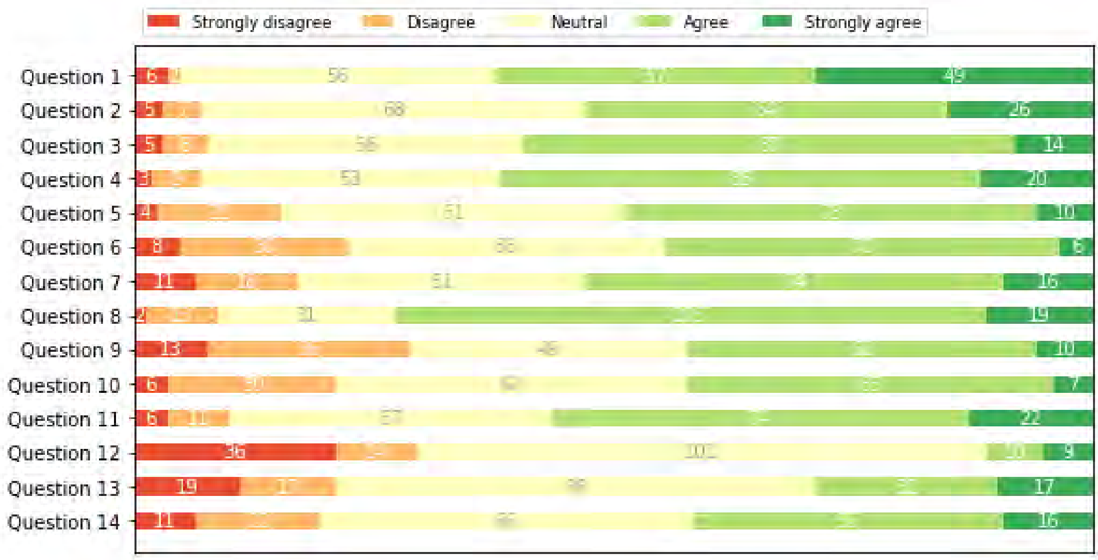


Figure 4.39: Horizontal Bar Graph for Impression

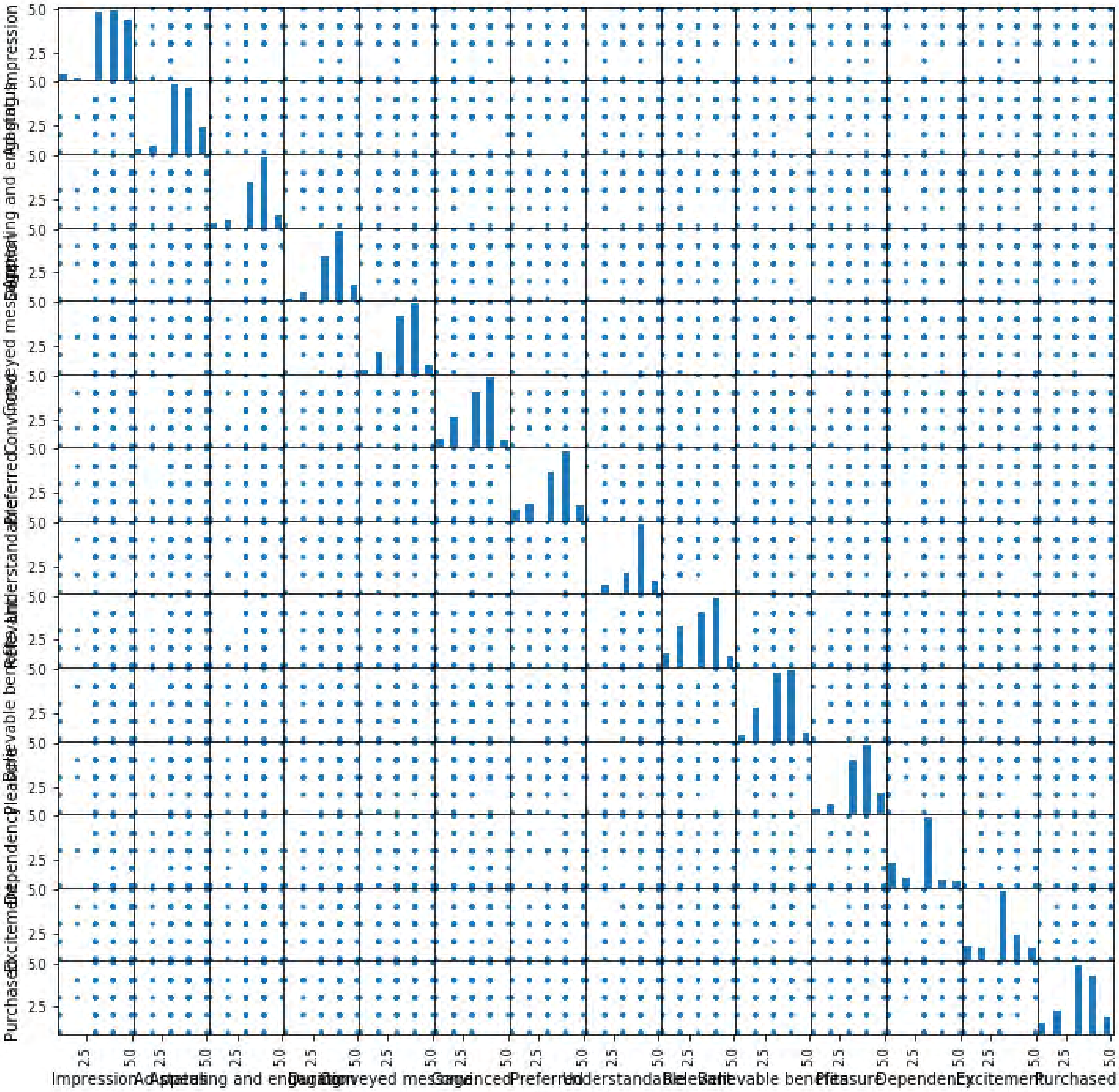


Figure 4.40: Pair Plot for Impression

Chapter 5

Algorithms

We use several algorithms for our research. Those are explained in this section.

5.1 Support-Vector Machine

Support-Vector machines are supervised learning models with a related learning algorithm that analyzes data used for the study of classification and regression. In other words, provided that the training data (supervised learning) are labeled, an optimal hyperplane is generated by an algorithm that categorizes new instances. In two dimensional spaces, this hyper plane is a line that divides the plane into two parts.It was mostly used for the problem of classification. In this algorithm, each data element is plotted as a point in n-dimensional space (where n is the number of features you have) with each feature being the value of a particular coordinate. Then we perform classification by finding the hyper plane that differentiate the two classes very well [27]. The further are the points from the location of the hyper plane, the more surely it can be said to have been properly classified.

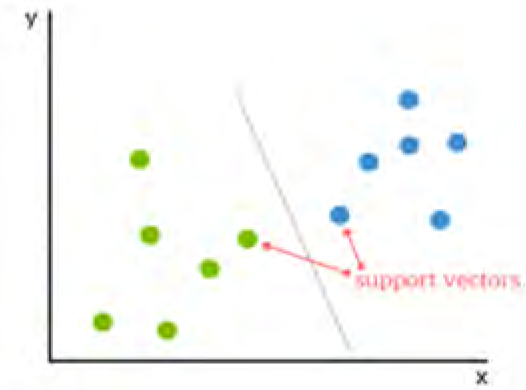


Figure 5.1: Support Vector Machine

SVM is searching for the nearest points. SVM then draws a line that connects these points. It draws this line of connection by subtracting the vector. The vector support machine then declares the best line of separation to be the line that bisects and is perpendicular to the line of connection. SVM is better because when you get a new sample you will have already made a line that keeps B and A as far away from each other as possible and so it is less likely that one will spillover across the line into the other's territory [28]. For cleaner data sets it works pretty well and efficiently. SVM is also more reliable because it uses a subset of training data points, but it is not suitable for wider and more complex data sets that could overlap groups.

5.2 Random Forest

The ability to distinguish findings correctly is extremely valuable in various business purposes, such as predicting whether a single consumer may buy an item or determining whether a specific borrower is defaulting or not. Random forests or random decision forests are an ensemble learning method for classification, regression. Random forest algorithm is a supervised classification algorithm. As the name suggests, this algorithm creates the forest with several trees. These trees can be thought as some yes/no questions about the data that eventually leads to a predicted class. This model is an interpretative model as it works just as real-life scenario, bunch of questions are asked and based on the answers the decision is made.

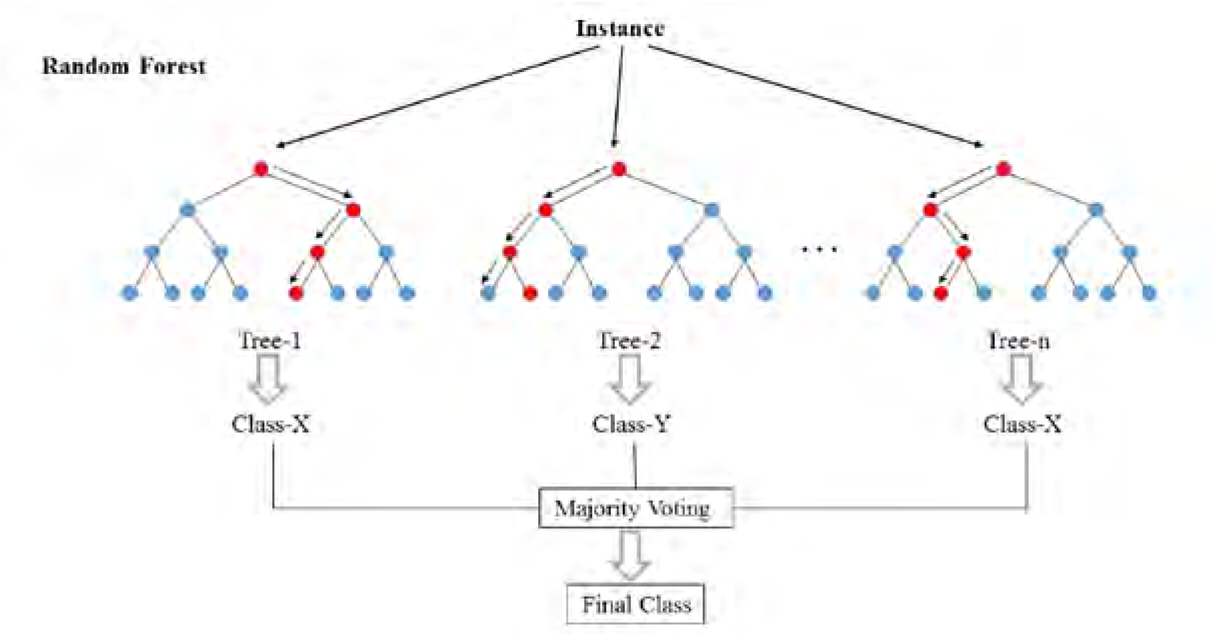


Figure 5.2: Random Forest Algorithm

Random forest algorithm's core concept is simple but very fundamental as it relies on crowd knowledge. In general, the more trees in the forest the more robust the forest looks. Likewise, the higher the number of trees in the forest gives the results of high accuracy in the random forest classification. The definition of a decision tree is a structure based on rules. Given the targets and features of the training data set, some set of rules will be developed for the decision tree algorithm. The same set rules used to perform the prediction on the test the dataset. Suppose a dataset has features. Instead of trying all features every time we make a new decision node, we only try a subset of the features. We do this primarily to inject randomness that makes individual trees more unique and reduces the correlation between trees which improves the forest's performance overall [29]. Some benefits of random forest are like The Random Forest classifier is capable of handling missed values and the Random Forest classifier can be based on categorical values.

5.3 Logistic Regression

Logistic regression is a statistical method of assessing a sample in which a hypothesis is evaluated using one or more independent variables. Using a dichotomous element the effect is measured. Thanks to an independent set of variables it is used to evaluate a binary outcome. This algorithm is an interpretation of the prognosis. It predicts the likelihood that by fitting data into a logit function an event will occur. This algorithm uses the sigmoid function to calculate the relationship between a categorical dependent variable and one or more independent variable. Dependent variable here means varying target class and independent variables are attributes that we use to identify the target class. To predict which class a data belongs to, a threshold may be set. The approximate probability obtained is categorized as per this threshold into groups. Logistic regression is a very effective method for estimating likelihood.

For example, if value 0.5, then consider the item worth purchasing or else- as not purchasing. The boundary of the decision may be linear or non-linear. In order to obtain complex decision boundaries, polynomial order can be increased [19].

There are two steps behind the math of logistic function.

1. Classifying inputs to be in class zero or one
2. Defining a boundary values for the odds

The threshold value can be chosen as per the problem is being solved. Threshold values are important factors and to choose one effectively confusion matrix is an effective mean. Four key parts of confusion matrix,

1. TN(True Negative)
2. TF(True Positive)
3. FF(False Positive)
4. FN(False Negative)

And from them we find, Accuracy = (TP+TN)/total No of Classified Item = (TP+TN)/ (TP+TN+FP+FN) Precision = TP/ (TP+FP) Recall = TP/ (TP+FN) f1-score = 2*((precision*recall)/(precision+recall)) [30]

5.4 Chi-Square

The Chi Square algorithm is regularly utilized for testing connections between categorical variables. The null hypothesis of the Chi-Square test is that no relationship exists on the categorical factors in the populace; they are not dependent.

The Chi-Square measurement is most ordinarily used to assess Tests of Independence when utilizing a crosstabulation (otherwise called a bivariate table). Crosstabulation displays the disseminations of two absolute factors at the same time, with the convergences of the classes of the factors showing up in the cells of the table. The Test of Independence surveys whether an affiliation exists between the two factors by contrasting the watched sample of reactions in the cells to the example that would be normal if the factors were genuinely free of one another. Computing the Chi-Square measurement and contrasting it against a critical value from the Chi-Square dissemination enables the observer to survey whether the observed cell tallies are altogether not the same as the normal cell checks [31]. The calculation of the

Chi-Square statistic is quite straight-forward and intuitive:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \tag{5.1}$$

where fo = the observed frequency (the observed counts in the cells) and fe = the expected frequency if NO relationship existed between the variables

As delineated in the equation, the Chi-Square measurement depends on the distinction between what is really seen in the information and what might be normal if there was genuinely no connection between the factors.

There are various significant contemplations when utilizing the Chi-Square measurement to assess a cross tabulation. On account of how the Chi-Square worth is determined, it is very delicate to test size – when the sample size is excessively enormous (500), practically any little contrast will show up measurably noteworthy. It is likewise delicate to the conveyance inside the cells, and SPSS gives an admonition message if cells have less than 5 cases. This can be tended to by continually utilizing unmitigated factors with a set number of classes [31].

5.5 K-Nearest Neighbor

K-Nearest Neighbor (K-NN) is waiting closely for issues. K-NN is a straightforward algorithm that stores every possible case and classifies new cases according to a measure of comparability. It classifies new cases that are dependent on a similarity of cases that have recently been made available. A case is characterized by a major vote of its neighbors, the case being allocated by a distance function K to the most common class of its K-NN. It is best to make the ideal incentive to select k by first evaluating the details. This considers the distance from that point in the training data whenever the model is with a test data. Then it will locate the point's nearest K leaders. Use the "Euclidean distance test" to calculate the distance. A high K value (K=¿1) is usually more reliable as it reduces overall noise. The main objective is to use a medium containing different data points grouped into classes or groups to predict a sample point output. Unlike almost any other algorithm, K-NN works due to its deeply rooted mathematical theories. The initial steps are to transform the data points into feature vectors or their mathematical significance when implementing K-NN. Then the algorithm works by calculating the difference between these points ' mathematical values[32].

We also use Cronbach's alpha and Shapiro-Wilk test for data analysis.

5.6 Cronbach's Alpha

Cronbach's alpha is a simple measure used to evaluate the efficiency of a composite score's reliability or internal consistency. The alpha of Cronbach gives us a basic method of determining whether or not a score is good. For example, the employee of an organization is given a job satisfaction survey. High reliability means the job satisfaction of this butt and low reliability means something else is calculated. The alpha of Cronbach also gives you a value between 0 and 1 but you can also get a negative number. A negative number indicates that your data is wrong. The general rule of thumb is that the alpha of a Cronbach of.70 and above is better,.80 and above is better and.90 and above is better. [33].

$$\alpha = \frac{N.\bar{c}}{\bar{v}+(N-1).\bar{c}} \tag{5.2}$$

Where, N =number of items
v =average variance
c =average covariance between item-pairs

Table 5.1: Alpha values for Cronbach's Alpha

Cronbach's alpha	Internal consistency
$\alpha \geq 0.9$	Excellent
$0.9 \geq \alpha \geq 0.8$	Good
$0.8 \geq \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

5.7 Shapiro-Wilk

The Shapiro-Wilk test is a way to tell if a random sample comes from a normal distribution. The test gives you a w value; small values shows that your sample is not usually distributed [34]. If the value is less than or equal 0.05 then the test rejects the hypothesis of normality. Failing in the normality test permits you to state with 95 percent confirmation that the data does not fit the normal distribution. Passing the normality test only allows you to state no significant departure from normality was found [35].

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{5.3}$$

Where, X= the ordered random sample values and a= are constants generated from the covariance, variance and means of the sample (size n) from a normally distributed sample.

5.8 Principal Component Analysis

PCA's goal is to reduce dimensions. Reduction of dimension is accomplished by shaping basis vectors. A linear combination of basis vectors can be used to reconstruct any sample from a data set. The basis vector must be orthogonal to each other. It implies that 0 is the dot product of any two basis vectors.

Principal Component Analysis Algorithm Steps

1. Finding mean vector.
2. Assemble every data samples in a mean adjusted matrix.
3. Creating the covariance matrix.
4. Computing the Eigen vectors and Eigen values.
5. Computing the basis vectors.
6. Representing each sample as a linear combination of basis vectors.

First of all, assume we have a set of images that we want to perform Principal Component Analysis (PCA). Assume that each picture isx per y pixel. We can view each object as a vector of size x*y without losing information. Therefore, we can accept an image as a point in x*y dimensional space. Because images typically consist of hundreds of pixels inx and hundreds of pixels in y dimensions, an object is transformed into a point in a very high dimension space. Use PCA, this dimensionality can be significantly reduced. One of the reasons for reducing dimensions is so we can focus on those dimensions where there is a large disparity between items in our database, i.e. high variability. PCA allows us to mathematically the the dimensions so that high variance dimensions (for a given dataset) are selected in the decreased array of dimensions. [36].

5.9 Naïve Bayes

A Naive Bayes classification is an algorithm which uses the theorem of Bayes to classify objects. The key insight in Bayes ' theorem is that as new data are added, the probability of an occurrence can be modified. Classifiers of Naive Bayes assume a strong or naive independence between data point attributes. These classifiers are widely used for machine learning as they are easy to implement. What makes a naive Bayes classifier naive is its assumption that all data point attributes are independent of each other [37]. A classifier that classifies vegetables into carrots and radishes would realize that carrots are orange, triangular and of a certain length, but would not take all of these aspects at once. After all, radishes are also triangular. A naive Bayes classifier is not a single algorithm, but a family of algorithms for machine learning that makes use of statistical independence. These algorithms are comparatively easy to write and operate more effectively than more complex algorithms from Bayes [37]. This algorithm is relatively easy to understand and build. Even with small data sets it can be trained with ease. Unnecessary features are not an issue to it. But it considers every feature to be independent which might not be true in every aspects.

5.10 Adaboost

Boosting is an ensemble technique that tries to create a strong classifier from a number of soft classifiers. AdaBoost is Freund and Schapire's first functional boosting algorithm introduced in 1996, short for "Adaptive Boosting." It focuses on classification concerns and attempts to turn a collection of soft classifiers into a powerful one. AdaBoost is defined as a stepwise approximation procedure to match an additive regression model.AdaBoost minimizes the loss function. Weak models are introduced sequentially, trained using weighted data from training. The process goes on until a pre-set number of weak learners (a client parameter) is generated or no further progress can be made on the learning data set. When done, there's a pool of soft students left to each with a stage price. Calculation of the weighted average of the poor classifiers makes predictions. For a new input example, every poor learner calculates an expected value as either+ 1.0 or -1.0. The stage output of each slow learner weights the anticipated values. The estimation of the ensemble system is taken as the number of weighted predictions. If the sum is positive, the first or second class is anticipated [38].

5.11 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a probabilistic model of the generative corpus. The basic idea is to represent documents as random mixtures on latent topics, where word distribution characterizes each subject. For each document w in a corpus D, LDA presumes the following generative cycle

1. Choose N Poisson ().
2. Choose Dir().
3. for each of the N words wn:

   Choose a topic zn Multinomial ().

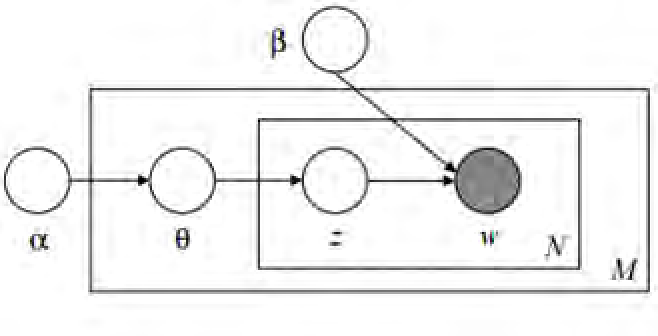
   Choose a word wn from p(wn —zn,), a multinomial probability conditioned on the topic zn.

The basic model includes many simplifying premises, some of which we delete in the following sections. s. First, the dimensionality k of the Dirichlet distribution (and thus the dimensionality of the topic variable z) is assumed known and fixed. Second, the word probabilities are parameterized by a k V matrix where i j = p(wj = 1—zi = 1), What we are considering for now as a set amount to be determined. Finally, the presumption of Poisson is not crucial to anything that follows, and more practical distributions of length of document can be used if appropriate. In particular, notice that N is independent of all other variables producing information (and z). It is therefore an ancillary parameter and in the subsequent development we must generally ignore its randomness. A k-dimensional Dirichlet random variable can take values in the (k 1)-simplex (a k-vector lies in the (k1)-simplex if i 0, k i=1 i = 1), and has the following probability density on this simplex [39].

$$p(\theta|\alpha) = \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\theta_1^{\alpha_1-1}\cdots\theta_k^{\alpha_k-1}, \tag{5.4}$$

Where the parameter is a k-vector with components i ¿ 0, and where (x) is the Gamma function. The Dirichlet is a handy representation on the simplex — it is in the exponential group, has ample finite-dimensional statistics, and is compatible with the multinomial distribution. Such properties should facilitate the design of algorithms for inference and parameter estimation for LDA. Given the parameters and , the joint distribution of a topic mixture , a set of N topics z, and a set of N words w is given by [36]:

$$p(\theta,\mathbf{z},\mathbf{w}|\alpha,\beta) = p(\theta|\alpha)\prod_{n=1}^{N}p(z_n|\theta)p(w_n|z_n,\beta). \tag{5.5}$$



Above figure is the graphical model representation of LDA. The boxes are “plates” representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document.

Where p(zn —) is simply i for the unique i such that zi n = 1. We obtain the marginal distribution of a document by averaging over and summing over z [18]:

$$p(\mathbf{w}|\alpha,\beta) = \int p(\theta|\alpha)\left(\prod_{n=1}^{N}\sum_{z_n} p(z_n|\theta)p(w_n|z_n,\beta)\right)d\theta. \tag{5.6}$$

Finally, we obtain the chance of a corpus taking the sum of the marginal probabilities of individual documents [18]:

$$p(D|\alpha,\beta) = \prod_{d=1}^{M}\int p(\theta_d|\alpha)\left(\prod_{n=1}^{N_d}\sum_{z_{dn}} p(z_{dn}|\theta_d)p(w_{dn}|z_{dn},\beta)\right)d\theta_d. \tag{5.7}$$

The LDA model is shown in Figure 1 as a probabilistic graphic model. As the figure shows, the LDA representation has three grades. The parameters and are parameters of the corpus level, which are expected to be sampled once during the cycle of a corpus being produced. The variables d are document-level variables, checked once per file. Eventually, the Zdn and Wdn variables are word-level variables and in each paper they are sampled once for each document [36].

Chapter 6

Result analysis and Data Visualization

ROC Curve: In machine learning, quality assurance is an important role. So when it comes to a classification problem, we can rely on an AUC-ROC Curve. The ROC (Receiver Operating Characteristics) AUC (Area under the Curve) curve is used to test or simulate the multi-classification problem's quality. It is one of the most important metrics of any type of classification to measure the quality. It is also known as AUROC (the receiver area's operating characteristics). AUC-ROC curve is an estimate of the efficiency of the measurement at different thresholds. ROC is a probability curve and AUC is a separability vector degree. It means how much of a prototype can be distinguished between classes. The higher the AUC, the more 0s as 0s and 1s as 1s as predicted is the better the prototype. The higher the AUC, on the other hand, the simpler the prototype is to discern patients with disease from patients without disease. An exceptional design has AUC close to 1 which means a good measurement of the separability. [40].

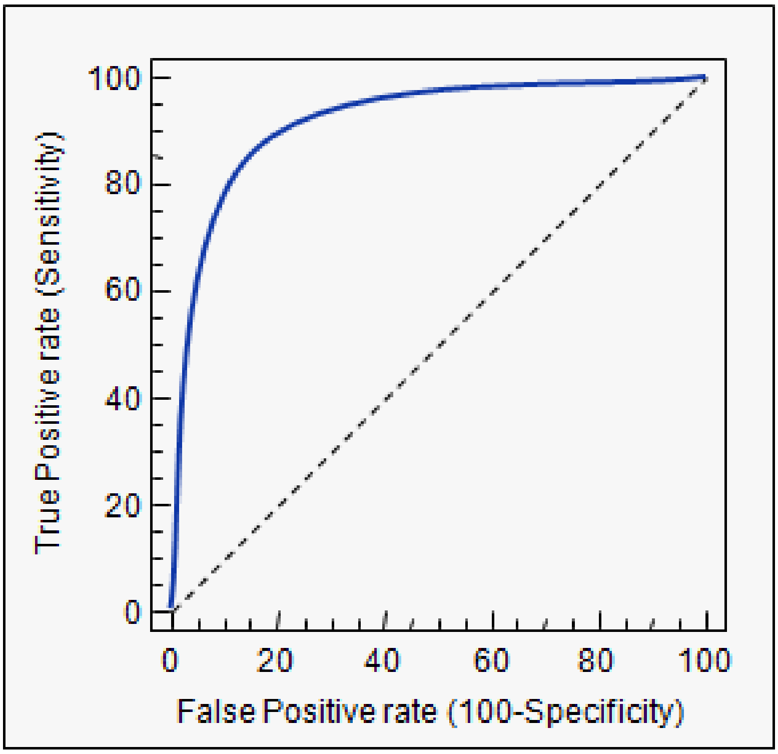


Figure 6.1: ROC Curve

A poor model has AUC close to 0, which implies it has the worst separability score. This actually means that the outcome is being reciprocated. This expects 0s to be 1s and 1s to be 0s. And when AUC is 0.5, this implies the template does not have any ability to distinguish groups at all. Sensitivity and precision are equally related. So when we increase tolerance, there is a reduction in accuracy and vice versa.

When the Sensitivity increased, specificity will decrease. Similarly, Sensitivity decrease and Specificity Increase.

We get more positive values as we lower the threshold so it increases the tolerance and reduces the specificity. Likewise, we get more negative values as we through the limit, so we get more precision and less tolerance.

TPR and FPR is proportionate. If TPR increase, FPR increase. And TPR decrease FPR decrease.

As we note, the precision of FPR is 1. So, once we increase TPR, FPR will also rise and vice versa [40].

6.1 Results of Textual Dataset

We have constructed our dataset in 3 parts. First we have taken the values as input for a 1 - 5 scale as discussed before and tested that for different classification algorithm based on impression and purchase intent. Then we generalized our classifiers only by converting their values to 0-1 (Negative and Positive) from the scale of various levels of that. For this generalization we consider the value of 1 and 2 as 0 as negative and 4, 5 as 1. The neutral value 3 is considered positive and negative both cases separately. In case of that we have divided it twice first, counting the neutral value 3 as positive, a value as 1 and second considering the neutral value negative as 0. All the dataset was set in that way to find the classification better and to see whether we could classify the ad as successful or not. Later we have generalized the entire dataset and taken all the values from a scale of 1 -5 to a scale of 0-1 to just present the data set in a form of yes or no in terms of successful ads. That was also done by considering all the neutrals as positive and negative two different times. Figure 1 describes the entire process of our dataset variation.
After forming out dataset we test the dataset for normality using the following methods -

6.1.1 Shapiro Wilk Test:

The Shapiro Wilk test is done to find whether a dataset comes from a normal distribution [41]. We have tested our dataset to see the statistics and P value and also whether they belong to a normal distribution. The Shapiro Wilk test was

performed on all the separate variations of our dataset which provided us with various answers. The statistics and p-value of these variations show the normality of our dataset. This table shows the values of this test for various datasets.

Table 6.1: Result from Shapiro Wilk Test

Dataset	Statistics	P-Value
Scale of 1 – 5	0.886	0.000
All Generalized (Neutral as 0)	0.636	0.000
All Generalized (Neutral as 1)	0.437	0.000
Impression Generalized (Neutral as 0)	0.886	0.000
Impression Generalized (Neutral as 1)	0.884	0.000
Purchased Generalized (Neutral as 0)	0.880	0.000
Purchased Generalized (Neutral as 1)	0.888	0.000

In the above table we see the statistics value which is one of the indicators of the Gaussian distribution shows the values at a range of 0.88. When all the values are generalized the values drops due to less numbers of scale. P value is actually the easier measure of dataset normality. Even though the datasets show different statistics value the p value is the same for all of them. Our reference p-value was 0.05 as seen above the p value our datasets show that our value is 0.000 which is lower than the set threshold. Thus rejecting the null hypothesis showing that the dataset does not belong to a gaussian distribution.

6.1.2 Cronbach's Alpha Test

To measure the internal consistency of a set of data the cronbach's alpha method is used, to see how closely the features are related as a set. It shows the reliability of different dimensions [42]. We measure the reliability of the features of our dataset by the cronbach's value. The higher the value, the more reliable that feature is. We have run Cronbach's Alpha algorithm based on the two different classifiers and it shows us different features which are important for that dataset. In the different datasets different features come up with reliability, which is explained in the table below.

Table 6.2: Result from Cronbach's Alpha Test (Purchased)

Dataset	Value >= 0.7	Value >= 0.6	Value >= 0.5
Scale of 1 – 5	0	2	3
All Generalized (Neutral as 0)	0	0	6
All Generalized (Neutral as 1)	0	1	1
Impression Generalized (Neutral as 0)	0	1	6
Impression Generalized (Neutral as 1)	1	0	6
Purchased Generalized (Neutral as 0)	0	0	1
Purchased Generalized (Neutral as 1)	0	0	0

Table 6.3: Features from Cronbach's Alpha Test (Purchased)

Dataset	Features with Value >= 0.7	Features with Value >= 0.6	Features with Value >= 0.5
Scale of 1 – 5		Convinced, Pleasure	Impression, Appealing and engaging , Preferred
All Generalized (Neutral as 0)			Impression, Appealing and engaging, Convinced, Preferred, Pleasure, Excitement
All Generalized (Neutral as 1)		Convinced	Preferred, Pleasure
Impression Generalized (Neutral as 0)	Convinced		Appealing and engaging, Preferred, Relevant, Believable benefits, Pleasure, Dependency
Impression Generalized (Neutral as 1)	Convinced		Appealing and engaging, Preferred, Relevant, Believable benefits, Pleasure, Excitement
Purchased Generalized (Neutral as 0)			Appealing and engaging
Purchased Generalized (Neutral as 1)			

Table 6.3 shows the features found from the Cronbach's alpha test. Features shown in the tables have an alpha value equal or greater than 0.5. Among the 14 features in the dataset following features are the ones those are dependable when finding successful advertisement. The table shows values from different datasets based on Purchase intent. We run the same algorithm based on Impression of the advertisement, considering that the success of the advertisements depend on the impression it creates on the user. Impression based algorithm produces different set of important features. Table below shows the number of important features based on Impression.

We run the same test now basing on Impression as the classifiers. Basing the success of an advertisement on Impression wields significantly better result.

Table 6.4: Result from Cronbach's Alpha Test (Impression)

Dataset	Value >= 0.7	Value >= 0.6	Value >= 0.5
Scale of 1 – 5	2	6	3
All Generalized (Neutral as 0)	1	1	9
All Generalized (Neutral as 1)	0	1	2
Impression Generalized (Neutral as 0)	0	0	3
Impression Generalized (Neutral as 1)	0	0	0
Purchased Generalized (Neutral as 0)	1	6	3
Purchased Generalized (Neutral as 1)	2	5	3

Table 6.5: Features from Cronbach's Alpha Test (Impression)

Dataset	Features with Value >= 0.7	Features with Value >= 0.6	Features with Value >= 0.5
Scale of 1 – 5	1.Preferred 2.Pleasure	1.Ad status 2.Appealing and engaging 3.Convinced 4.Understandable 5.Relevant 6.Purchased	1.Conveyed message 2.Believable benefits 3.Excitement
All Generalized (Neutral as 0)	1.Pleasure	1.Preferred	1.Ad status 2. Appealing and engaging 3. Conveyed message 4. Convinced 5.Understandable 6. Relevant 7. Believable benefits 8. Excitement 9. Purchased
All Generalized (Neutral as 1)		Appealing and engaging	Ad status, Understandable
Impression Generalized (Neutral as 0)			Ad status, Preferred, Pleasure
Impression Generalized (Neutral as 1)			
Purchased Generalized (Neutral as 0)	Preferred	Ad status, Appealing and engaging, Convinced, Understandable, Relevant, Dependency	Conveyed message, Believable benefits, Excitement
Purchased Generalized (Neutral as 1)	Preferred, Pleasure	Ad status, Appealing and engaging, Convinced, Understandable, Relevant	Conveyed message, Believable benefits, Excitement

Comparing the alpha values found when using both classifiers the conclusion can be drawn that when we find feature reliability using Impression as our classifiers we see significant values showing more importance or importance than while basing them on Purchased.

6.2 KNN Algorithm

KNN classifies any given dataset which in our case is the success of given advertisements which are classified based on two different classifiers - purchased and impression. As discussed before Purchased represent the purchase intent of user and Impression represents how the ad affected people. Based on these two success of an ad in tried to be classified. The results are shown below in Tables show the accuracy level of these classification.

Table 6.6: Accuracy from KNN Algorithm (Purchased Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.62	0.62	0.62
All Generalized (Neutral as 0)	0.79	0.88	0.82
All Generalized (Neutral as 1)	0.94	0.97	0.91
Impression Generalized (Neutral as 0)	0.62	0.59	0.62
Impression Generalized (Neutral as 1)	0.56	0.59	0.62
Purchased Generalized (Neutral as 0)	0.91	0.82	0.88
Purchased Generalized (Neutral as 1)	0.88	0.88	0.97

The above table shows the level of accuracy while correctly finding the purchase intent of a consumer. We see that K-NN scores at 97 percent in accuracy.

Table 6.7: Accuracy from KNN Algorithm (Impression Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.74	0.71	0.68
All Generalized (Neutral as 0)	0.85	0.91	0.85
All Generalized (Neutral as 1)	0.97	0.97	1.0
Impression Generalized (Neutral as 0)	0.85	0.82	0.85
Impression Generalized (Neutral as 1)	1.0	0.97	1.0
Purchased Generalized (Neutral as 0)	0.71	0.67	0.62
Purchased Generalized (Neutral as 1)	0.59	0.62	0.71

Following these tables we can see that in different dataset success can be various in various dataset and feature value. KNN detects success at a score of 97 percent.

6.3 Support Vector Machine

Table 6.8: Accuracy from SVM Algorithm (Purchased Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.56	0.59	0.56
All Generalized (Neutral as 0)	0.88	0.85	0.88
All Generalized (Neutral as 1)	0.91	0.94	0.91
Impression Generalized (Neutral as 0)	0.59	0.59	0.59
Impression Generalized (Neutral as 1)	0.62	0.62	0.65
Purchased Generalized (Neutral as 0)	0.79	0.85	0.79
Purchased Generalized (Neutral as 1)	0.91	0.91	0.94

In the above table where it shows the accuracy of SVM in detectecting the purchase intent, it was 94 percent accurate.

Table 6.9: Accuracy from SVM Algorithm (Impression Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.67	0.67	0.68
All Generalized (Neutral as 0)	0.91	0.88	0.91
All Generalized (Neutral as 1)	0.97	0.97	0.97
Impression Generalized (Neutral as 0)	0.82	0.82	0.76
Impression Generalized (Neutral as 1)	1.0	0.97	0.97
Purchased Generalized (Neutral as 0)	0.71	0.65	0.71
Purchased Generalized (Neutral as 1)	0.71	0.65	0.68

Following the above table we can see that in different dataset success can be various in various dataset and feature value where SVM detects success at a score of 97 percent.

6.4 Decision Tree

Table 6.10: Accuracy from DT Algorithm (Purchased Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.56	0.50	0.47
All Generalized (Neutral as 0)	0.79	0.76	0.79
All Generalized (Neutral as 1)	0.91	0.88	0.85
Impression Generalized (Neutral as 0)	0.59	0.50	0.50
Impression Generalized (Neutral as 1)	0.50	0.56	0.47
Purchased Generalized (Neutral as 0)	0.76	0.74	0.74
Purchased Generalized (Neutral as 1)	0.79	0.85	0.82

Following the above table we can see that in different dataset success can be various in various dataset and feature value where Decision Tree detects purchase intent at a score of 91 percent.

Table 6.11: Accuracy from DT Algorithm (Impression Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.57	0.65	0.56
All Generalized (Neutral as 0)	0.74	0.79	0.88
All Generalized (Neutral as 1)	0.99	0.97	0.97
Impression Generalized (Neutral as 0)	0.81	0.88	0.76
Impression Generalized (Neutral as 1)	0.96	0.94	0.97
Purchased Generalized (Neutral as 0)	0.68	0.65	0.62
Purchased Generalized (Neutral as 1)	0.62	0.68	0.59

Following the above tables we can see that in different dataset success can be various in various dataset and feature value where DT detects successfully detects Impression at a score of 99 percent.

6.5 Logistic Regression

Table 6.12: Accuracy from LR Algorithm (Purchased Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.46	0.56	0.56
All Generalized (Neutral as 0)	0.78	0.79	0.79
All Generalized (Neutral as 1)	0.87	0.91	0.88
Impression Generalized (Neutral as 0)	0.49	0.50	0.53
Impression Generalized (Neutral as 1)	0.46	0.50	0.56
Purchased Generalized (Neutral as 0)	0.78	0.71	0.76
Purchased Generalized (Neutral as 1)	0.84	0.91	0.91

Following the above tables we can see that in different dataset success can be various in various dataset and feature value where LR detects successfully detects purchase intent at a score of 91 percent.

Table 6.13: Accuracy from LR Algorithm (Impression Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.49	0.62	0.62
All Generalized (Neutral as 0)	0.84	0.82	0.85
All Generalized (Neutral as 1)	0.97	1.0	0.97
Impression Generalized (Neutral as 0)	0.75	0.85	0.76
Impression Generalized (Neutral as 1)	0.97	0.97	0.97
Purchased Generalized (Neutral as 0)	0.50	0.62	0.62
Purchased Generalized (Neutral as 1)	0.47	0.59	0.65

Following the above table we can see that in different dataset success can be various in various dataset and feature value where LR detects success at a score of 100 percent on Impression.

6.6 Naive Bayes

Table 6.14: Accuracy from NB Algorithm (Purchased Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.47	0.44	0.50
All Generalized (Neutral as 0)	0.74	0.79	0.79
All Generalized (Neutral as 1)	0.81	0.85	0.85
Impression Generalized (Neutral as 0)	0.46	0.47	0.56
Impression Generalized (Neutral as 1)	0.49	0.29	0.41
Purchased Generalized (Neutral as 0)	0.75	0.79	0.74
Purchased Generalized (Neutral as 1)	0.84	0.85	0.85

Following the above tables we can see that in different dataset success can be various in various dataset and feature value where Naive Bayes detects successfully detects purchase intent at a score of 85 percent.

Table 6.15: Accuracy from NB Algorithm (Impression Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.60	0.65	0.68
All Generalized (Neutral as 0)	0.84	0.91	0.88
All Generalized (Neutral as 1)	0.93	0.91	0.94
Impression Generalized (Neutral as 0)	0.82	0.82	0.85
Impression Generalized (Neutral as 1)	0.97	0.97	0.94
Purchased Generalized (Neutral as 0)	0.53	0.53	0.59
Purchased Generalized (Neutral as 1)	0.56	0.59	0.65

Following the above table we can see that in different dataset success can be various in various dataset and feature value where NB detects success at a score of 97 percent based on Impression.

6.7 Random Forest

Table 6.16: Accuracy from RF Algorithm (Purchased Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.62	0.62	0.59
All Generalized (Neutral as 0)	0.74	0.82	0.82
All Generalized (Neutral as 1)	0.88	0.88	0.91
Impression Generalized (Neutral as 0)	0.68	0.53	0.59
Impression Generalized (Neutral as 1)	0.68	0.59	0.68
Purchased Generalized (Neutral as 0)	0.82	0.79	0.79
Purchased Generalized (Neutral as 1)	0.88	0.88	0.88

Following the above table we can see that in different dataset success can be various in various dataset and feature value where NB detects success at a score of 91 percent based on Purchased.

Table 6.17: Accuracy from RF Algorithm (Impression Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.71	0.74	0.71
All Generalized (Neutral as 0)	0.76	0.88	0.88
All Generalized (Neutral as 1)	1.0	0.94	0.97
Impression Generalized (Neutral as 0)	0.85	0.88	0.85
Impression Generalized (Neutral as 1)	0.97	0.97	0.97
Purchased Generalized (Neutral as 0)	0.68	0.68	0.76
Purchased Generalized (Neutral as 1)	0.68	0.68	0.71

Following the above table we can see that in different dataset success can be various in various dataset and feature value where RF detects success at a score of 100 percent based on Impression.

6.8 Adaboost

Table 6.18: Accuracy from Adaboost Algorithm (Purchased Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.51	0.50	0.53
All Generalized (Neutral as 0)	0.72	0.79	0.82
All Generalized (Neutral as 1)	0.81	0.82	0.79
Impression Generalized (Neutral as 0)	0.54	0.62	0.62
Impression Generalized (Neutral as 1)	0.53	0.56	0.65
Purchased Generalized (Neutral as 0)	0.78	0.85	0.88
Purchased Generalized (Neutral as 1)	0.86	0.85	0.82

In the above table we can see that in different dataset success can be various in various dataset and feature value where Adaboost detects success at a score of 86 percent based on Purchased.

Table 6.19: Accuracy from Adaboost Algorithm (Impression Based)

Dataset	Accuracy	Accuracy With Standard Scaling	Accuracy With Robust Scaling
Scale of 1 – 5	0.60	0.71	0.64
All Generalized (Neutral as 0)	0.82	0.94	0.85
All Generalized (Neutral as 1)	0.95	0.88	0.88
Impression Generalized (Neutral as 0)	0.78	0.88	0.82
Impression Generalized (Neutral as 1)	0.98	0.97	0.97
Purchased Generalized (Neutral as 0)	0.59	0.62	0.71
Purchased Generalized (Neutral as 1)	0.60	0.65	0.65

Following the above table we can see that in different dataset success can be various in various dataset and feature value where Adaboost detects success at a score of 98 percent based on Impression.

6.9 ROC Curve

Textual dataset where all the values of the features are classified 1 to 5.

Figure 6.2: ROC Curve based on Purchased

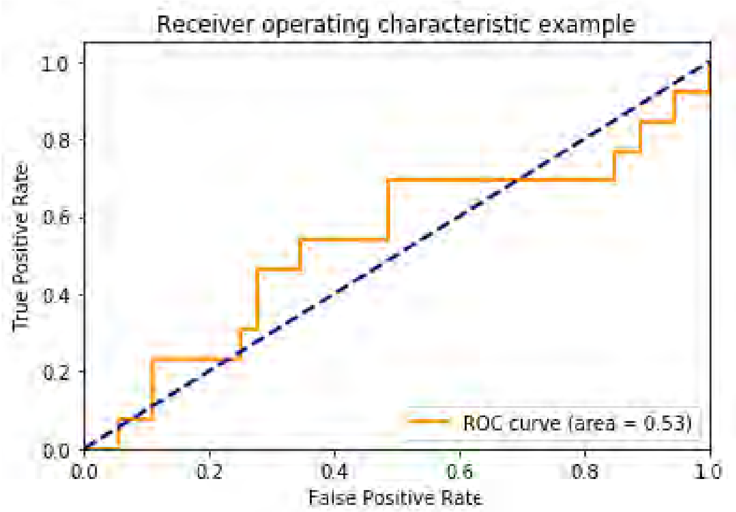


Figure 6.3: ROC Curve based on Purchased (multiclass)

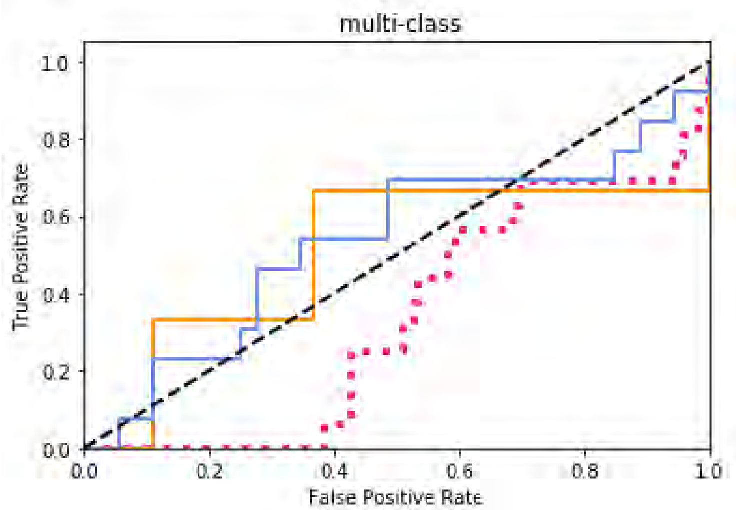


Figure 6.4: ROC Curve based on Impression

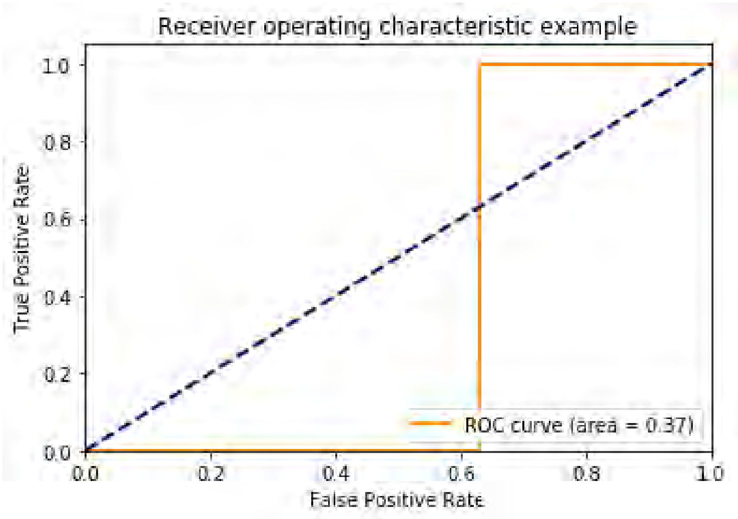
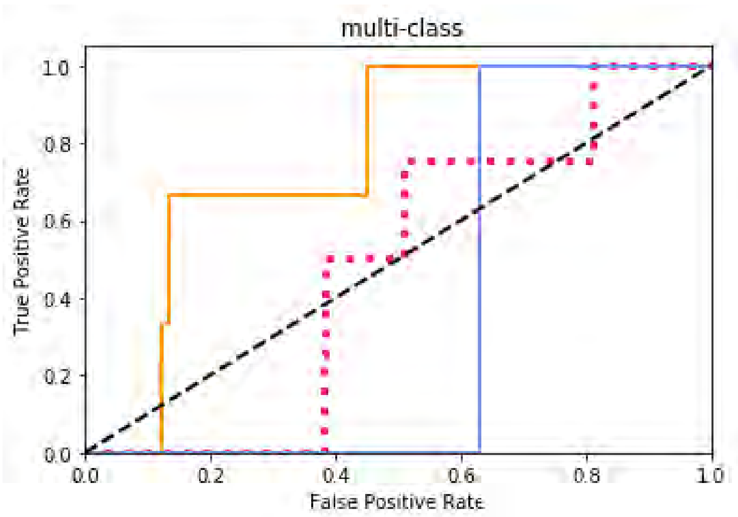


Figure 6.5: ROC Curve based on Impression (multiclass)

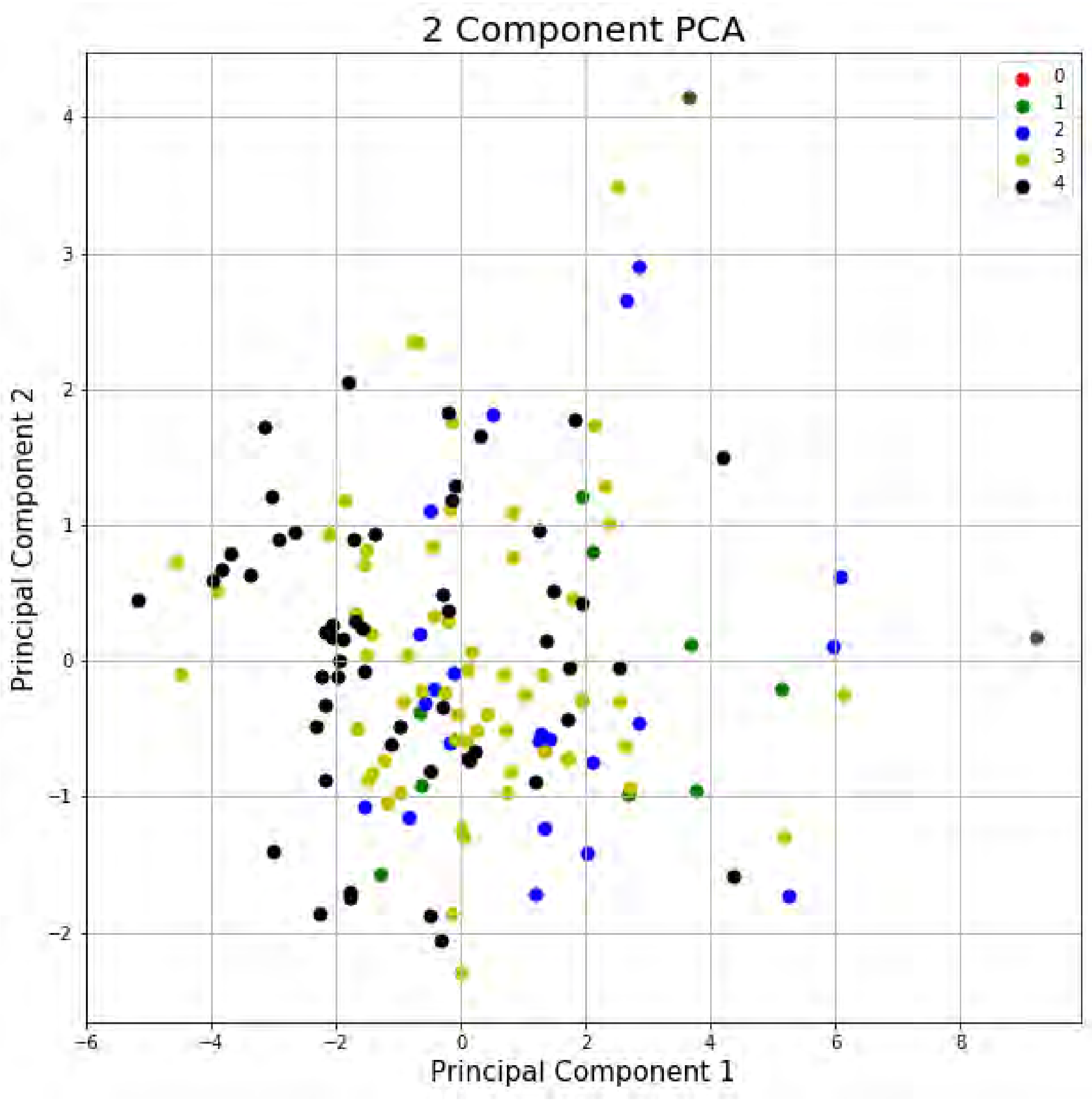


Figures above are the ROC curve of the dataset where all the values are 1 to 5. The TPR and FPR values of the curve is shown here. We also find the curve for multiclass in both for "Purchased" and "Impression".

6.10 PCA

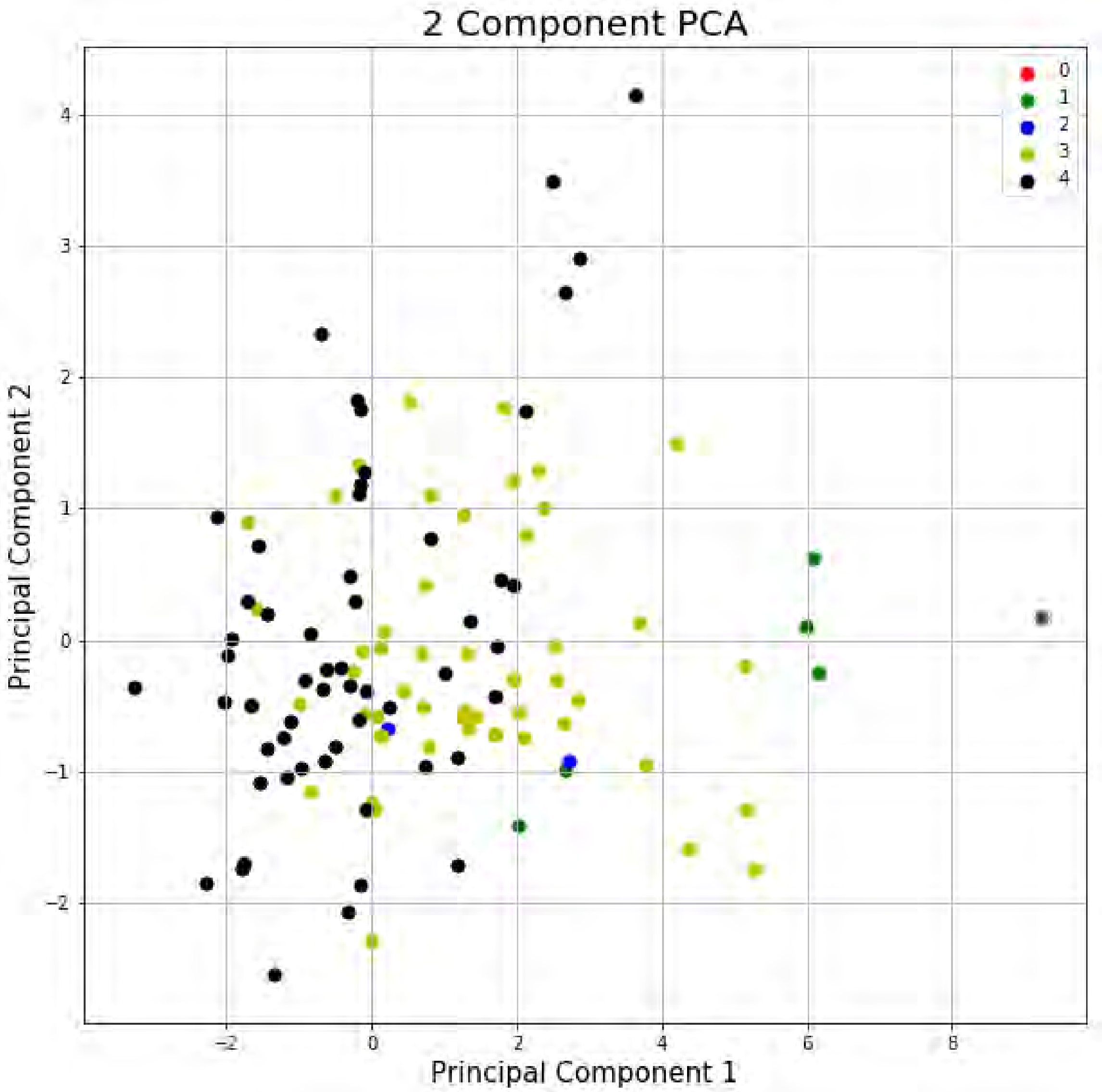
PCA on scale of 1 to 5 based on Purchased

Figure 6.6: Result of PCA based on Purchased



Variance ratio of Purchased [0.40868346 0.09195861]

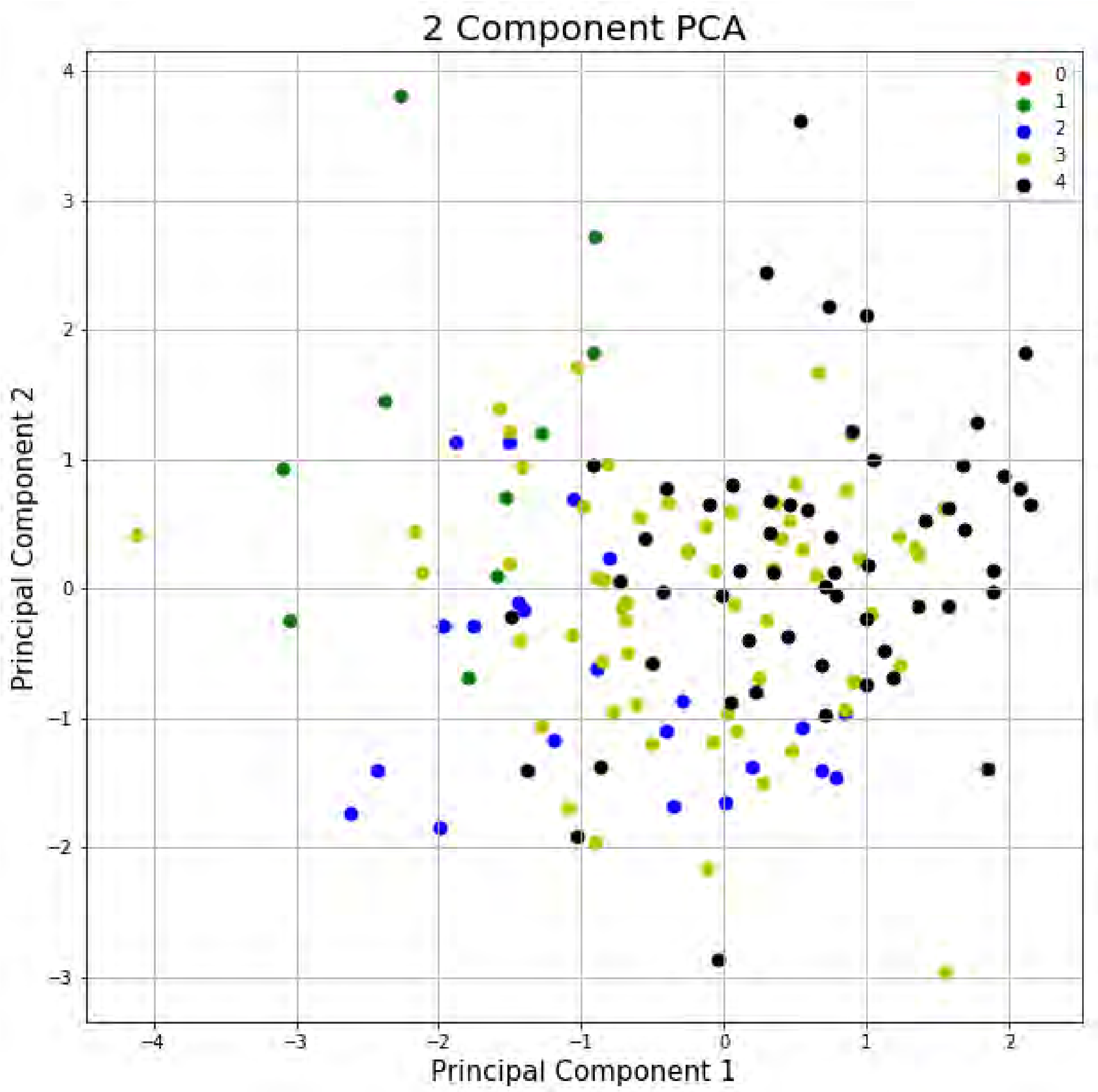
Figure 6.7: Result of PCA based on Impression



Impression: Variance ratio [0.40868346 0.09195861]

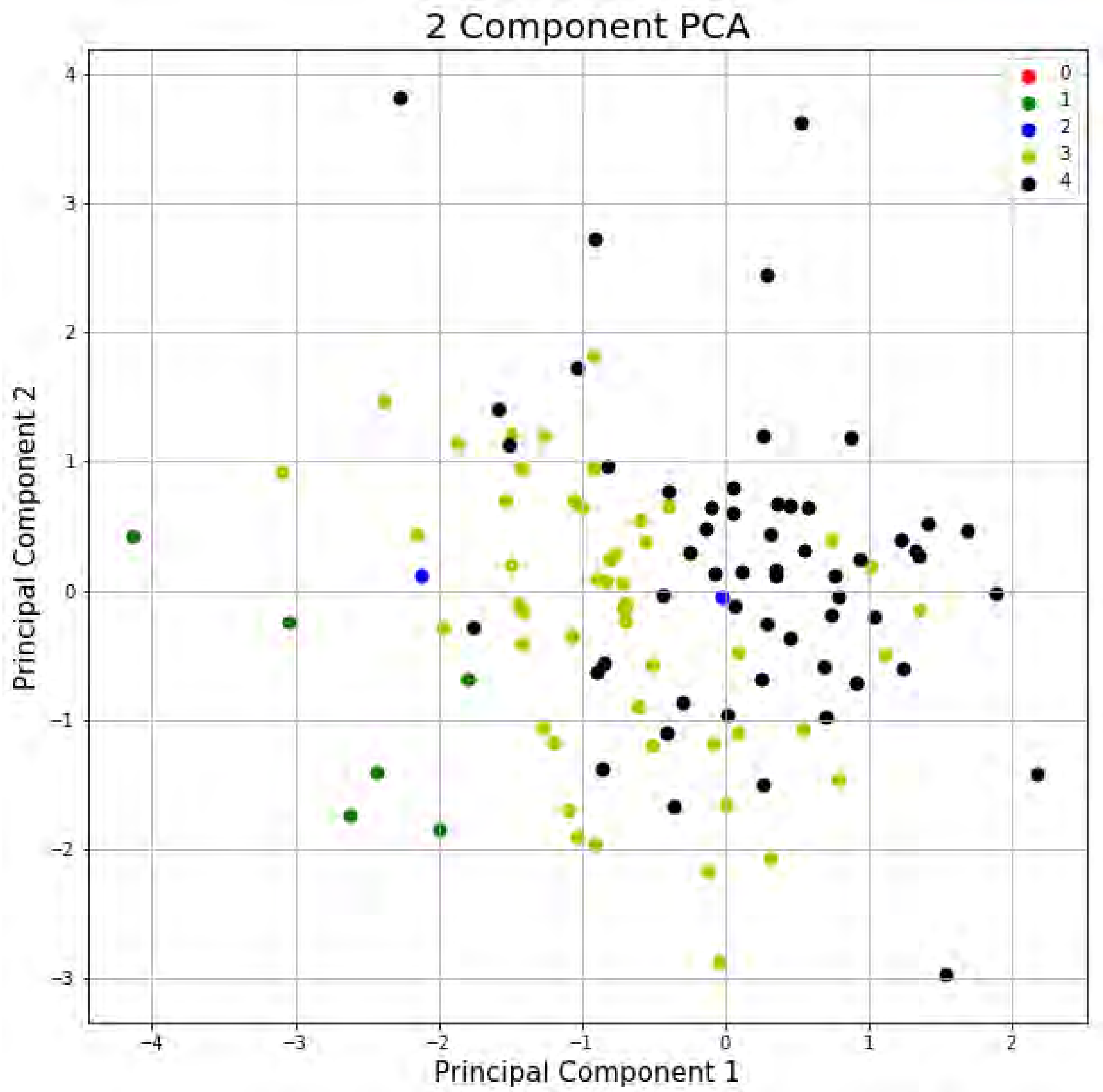
6.11 LDA

Figure 6.8: Result of LDA based on Purchased



LDA based on Purchased Variance ratio [0.65288319 0.18768757]

Figure 6.9: Result of LDA based on Impression



Impression Variance ratio [0.65288319 0.18768757]

6.12 Chi-Square

Figure 6.10: Result of features on Purchased

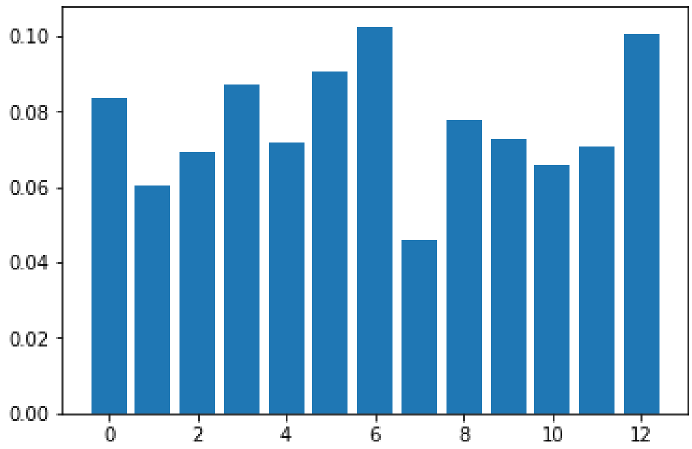


Figure 6.11: Result of important features based on Purchased

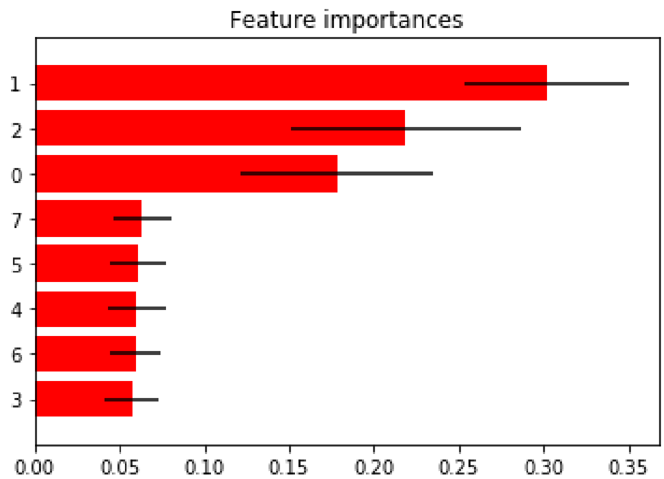


Figure 6.10 shows the features according to their importance. Figure 6.11 is after feature extraction the most important 8 features based on Purchase Intent.

Important Features Purchased based- [“Impression”,”convinced”,”Purchased”,”preferred”, ”duration”,”dependency”,”conveyed message”, ”appealing and engaging”]

Figure 6.12: Result of features on Impression

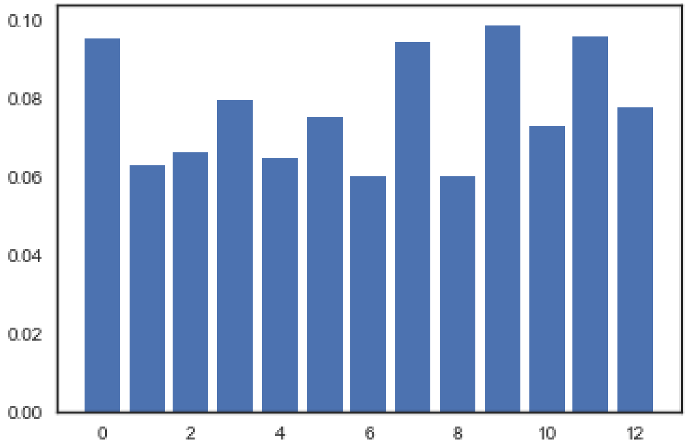


Figure 6.13: Result of important features based on Impression

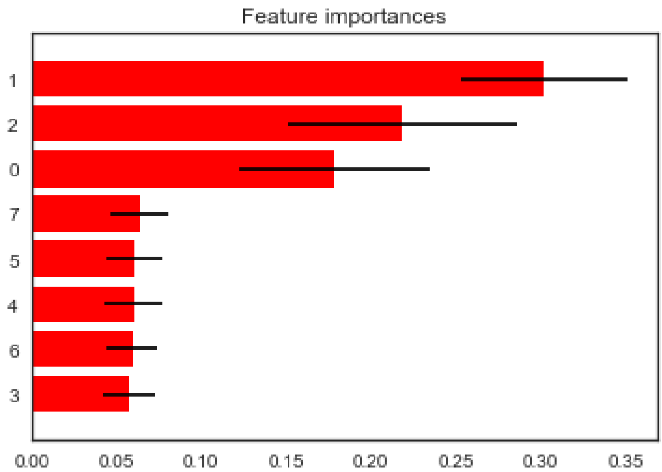


Figure 6.12 shows the features according to their importance. Figure 6.13 is after feature extraction the most important 8 features based on Impression.

Important Features Impression based- [“believable”,”convinced”,”Impression”,”excitement”, ”Purchased”,”understandable”,”duration”,”relevant”]

People Reaction of advertisements: We asked participants during survey which specific parts they actually like or dislike of any advertisements. The factors they like are: Duration of the advertisements like short advertisements, advertisement model (actor-actress), Social message, humor etc. Some people like the advertisement based on the price of the product and decide to buy the product. People mostly dislike unrealistic advertisements. They look for messages. They also dislike long duration advertisements.

6.13 EEG Sensor Result

We've collected the emotional response of people while watching advertisements. Running the algorithms based on EEG channel values we find the following result.

Table 6.20: Shapiro Wilk Test Result on EEG Dataset

Dataset	Statistics	P-Value
EEG Dataset	0.592	0.000

The Statistics value of our created EEG dataset is 0.592 and P value is 0.

Table 6.21: Algorithm Result on EEG Dataset

Algorithm	Accuracy Based on Purchased	Accuracy Based on Impression
SVM	0.52	0.57
Logistic Regression	0.45	0.46
Random Forest	0.46	0.52
Adaboost	0.46	0.51

From above table we can see the highest value on Purchase we get 52 percent in SVM and 57 percent based on Impression. Showing that the purchase intent of an user does depend on their emotion.

Chapter 7

Conclusion and future work

7.1 Conclusion

The main aim of our research is to predicting the success rate of an advertisement by analyzing its factor with group of data and human emotion. We are using machine learning algorithm for detecting the factors of advertisement and predicting the purchase and impression rate of the advertisement according to the factors. We hope that, by this research, it will be helpful for the organizations for making advertisements. The entrepreneurs will also be helpful if they maintain these factors while marketing their organizations by advertisements. The main aim of our research is to ensure the companies should make the advertisements which will help them to create a positive impression to the consumers of their product by watching the advertisements.

In our research, there is ample scope for future amplification for further improvement. We thought of a few of them and willing to apply them in the future. One such improvement is we will take the exact impression of the viewers of the advertisements via brain signal using sensors. It will directly tell us what the viewer actually thought of the advertisement. It will bring down the percentage of the error to a very minimum amount. It is for further improvement. But for immediate updates what we can do is to apply deep learning algorithm and get much better results. Applying deep learning will allow us to get more appropriate numbers that will make the standard scale that we are willing to provide much more efficient. But now we are working mostly on increasing machine learning algorithm efficiency.

7.2 Future Work

We are interested in this field after our thesis also. In this thesis we only focused on the advertisements of food and beverage. In future we want to work on the advertisements of different fields such as technological market like mobile, computer, IOT devices etc. we also want to work on the advertisements of household products which are daily necessary for us. There are also advertisements on car, toiletries, properties which are also portrait on different Medias by visual representation. We also want to focus on these fields of advertisement for analyzing the factors which helps the most profitable companies to earn money.

In our thesis we used machine learning algorithm. In future we will use deep learning algorithm for getting better accuracy of predicting the success rate of advertisement based on purchase rate and impression of the advertisement. We know that deep learning algorithm helps to learn about more accuracy. We will try to use RNN algorithm here for analyzing human emotion. As we already discussed that human emotion is also responsible for purchasing any product based on the context of advertisement.

In our thesis we worked on brain activity of human emotion. However, we also want to work on the facial expression of human while watching an advertisement. We will relate the facial expression with the message provided in advertisements for analyzing factors of advertisement which are responsible for making an advertisement successful. Here, we will use CNN algorithm for getting better accuracy.

We will also try to use more sensor like Electro dermal Activity sensor (EDA) or Galvanic skin response sensor (GSR) and Photoplethysmogram sensor (PPG) for our research. These sensors will also help us to analyzing human behavior and reactions for any message or advertisements.

These will help us to make more efficient advertisements in future.

Bibliography

[1] *What is advertising? definition of advertising, advertising meaning - the economic times*, https://economictimes.indiatimes.com/definition/advertising, (Accessed on 12/24/2019).

[2] *Smallbusiness.chron.com*, https://smallbusiness.chron.com/percentage-gross-revenue-should-used-marketing-20advertising-55928.html, (Accessed on 12/24/2019).

[3] *These 10 companies spent the most on advertising in 2018 - business insider*, https://www.businessinsider.com/10-biggest-advertising-spenders-in-the-us-2015-7, (Accessed on 12/24/2019).

[4] *10 kinds of advertising — bizfluent*, https://bizfluent.com/info-7736409-10-kinds-advertising.html, (Accessed on 12/24/2019).

[5] *Our methods aren't keeping pace with consumers' advertising experiences — articles — quirks.com*, https://www.quirks.com/articles/our-methods-aren-t-keeping-pace-with-consumers-advertising-experiences, (Accessed on 12/24/2019).

[6] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines", *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[7] M. G. R. Alam, A. K. M. Masum, L.-S. Beh, and C. S. Hong, "Critical factors influencing decision to adopt human resource information system (hris) in hospitals", *PloS one*, vol. 11, no. 8, e0160366, 2016.

[8] H. B. McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, *et al.*, "Ad click prediction: A view from the trenches", in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2013, pp. 1222–1230.

[9] R. Govindasamy, S. Arumugam, J. Zhuang, K. M. Kelley, and I. Vellangany, "Cluster analysis of wine market segmentation–a consumer based study in the mid-atlantic usa", *Economic Affairs*, vol. 63, no. 1, pp. 151–157, 2018.

[10] A. Shukla, S. S. Gullapuram, H. Katti, K. Yadati, M. Kankanhalli, and R. Subramanian, "Affect recognition in ads with application to computational advertising", in *Proceedings of the 25th ACM international conference on Multimedia*, ACM, 2017, pp. 1148–1156.

[11] P. Turcinek, J. Stastny, and A. Motycka, "Usage of cluster analysis in consumer behavior research", in *Proceedings of the 12th WSEAS International Conference on Applied Informatics and Communications (AIC '12)*, 2012, pp. 172–177.

[12] E. P. Bafghi, “Clustering of customers based on shopping behavior and employing genetic algorithms”, *Engineering, Technology & Applied Science Research*, vol. 7, no. 1, pp. 1420–1424, 2016.

[13] F. Neunhoeffer and T. Teubner, “Between enthusiasm and refusal: A cluster analysis on consumer types and attitudes towards peer-to-peer sharing”, *Journal of Consumer Behaviour*, vol. 17, no. 2, pp. 221–236, 2018.

[14] S. Rathore, M. Hussain, and A. Khan, “Gecc: Gene expression based ensemble classification of colon samples”, *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 11, no. 6, pp. 1131–1145, 2014.

[15] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “Deap: A database for emotion analysis; using physiological signals”, *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

[16] *Emotionomics: Leveraging emotions for business success ebook by dan hill - 9780749455736 — rakuten kobo*, https://www.kobo.com/us/en/ebook/emotionomics-leveraging-emotions-for-business-success-1, (Accessed on 12/25/2019).

[17] *An introduction to eeg*, https://www.ebme.co.uk/articles/clinical-engineering/introduction-to-eeg, (Accessed on 12/24/2019).

[18] *Emotiv epoc+ 14-channel wireless eeg headset — emotiv*, https://www.emotiv.com/epoc/, (Accessed on 12/24/2019).

[19] *Understanding q-q plots — university of virginia library research data services + sciences*, https://data.library.virginia.edu/understanding-q-q-plots/, (Accessed on 12/24/2019).

[20] *What is heat map (heatmap)? - definition from whatis.com*, https://searchbusinessanalytics.techtarget.com/definition/heat-map, (Accessed on 12/24/2019).

[21] *What is a line plot in math? - definition & examples - video & lesson transcript — study.com*, https://study.com/academy/lesson/what-is-a-line-plot-in-math-definition-examples.html, (Accessed on 12/24/2019).

[22] *Histogram tutorial*, https://www.moresteam.com/toolbox/histogram.cfm, (Accessed on 12/24/2019).

[23] *What is horizontal bar graph? - definition, facts & example*, https://www.splashlearn.com/math-vocabulary/geometry/horizontal-bar-graph, (Accessed on 12/24/2019).

[24] *Seaborn.pairplot — seaborn 0.9.0 documentation*, https://seaborn.pydata.org/generated/seaborn.pairplot.html, (Accessed on 12/24/2019).

[25] *A complete guide to scatter plots — tutorial by chartio*, https://chartio.com/learn/charts/what-is-a-scatter-plot/, (Accessed on 12/25/2019).

[26] *Dochttps*, https://dochttps//www.chartblocks.com/en/support/faqs/faq/when-to-use-a-line-charts.tibco.com/pub/spotfire/7.0.1/doc/html/line/line_what_is_a_line_chart.htm, (Accessed on 12/25/2019).

[27] *Understanding support vector machines algorithm (along with code)*, https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/, (Accessed on 12/24/2019).

[28] *Classification - how does a support vector machine (svm) work? - cross validated*, https://stats.stackexchange.com/questions/23391/how-does-a-support-vector-machine-svm-work, (Accessed on 12/24/2019).

[29] *How the random forest algorithm works in machine learning*, https://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/, (Accessed on 12/24/2019).

[30] *Logistic regression for dummies: A detailed explanation*, https://towardsdatascience.com/logistic-regression-for-dummies-a-detailed-explanation-9597f76edf46?gi=2858d5fdc888, (Accessed on 12/24/2019).

[31] *Using chi-square statistic in research - statistics solutions*, https://www.statisticssolutions.com/using-chi-square-statistic-in-research/, (Accessed on 12/24/2019).

[32] *K-nearest neighbors (knn) algorithm for machine learning*, https://medium.com/capital-one-tech/k-nearest-neighbors-knn-algorithm-for-machine-learning-e883219c8f26, (Accessed on 12/24/2019).

[33] *Cronbach's alpha - statistics solutions*, https://www.statisticssolutions.com/cronbachs-alpha/, (Accessed on 12/24/2019).

[34] *Shapiro-wilk test: What it is and how to run it - statistics how to*, https://www.statisticshowto.datasciencecentral.com/shapiro-wilk-test/, (Accessed on 12/24/2019).

[35] *Shapiro-wilks normality test*, https://variation.com/wp-content/distribution_analyzer_help/hs141.htm, (Accessed on 12/24/2019).

[36] *Principalcomponentanalysis_tutorial.pdf*, http://kiwi.bridgeport.edu/cpeg540/PrincipalComponentAnalysis_Tutorial.pdf, (Accessed on 12/24/2019).

[37] *What is naive bayes? - definition from techopedia*, https://www.techopedia.com/definition/32335/naive-bayes, (Accessed on 12/24/2019).

[38] *Boosting and adaboost for machine learning*, https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/, (Accessed on 12/24/2019).

[39] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation", *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[40] *Understanding auc - roc curve - towards data science*, https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5, (Accessed on 12/24/2019).

[41] *Medium*, https://medium.com/@rrfd/testing-for-normality-applications-with-20python-20b6bf06ed646a9, (Accessed on 12/24/2019).

[42] *What does cronbach's alpha mean? — spss faq*, https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/, (Accessed on 12/25/2019).