

# **Text Mining and Human Computer Interaction (HCI) Approach for Online Purchasing**

Submitted by:

K.A.N.Malkanthi

IM/2011/033

Supervised by:

Dr. Thashika D.Rupasinghe

A dissertation submitted in partial fulfilment of the requirements for the Degree of  
Bachelor of Science in Management and Information Technology

**Department of Industrial Management**

**Faculty of Science**

**University of Kelaniya**

**January 2017**

# DECLARATION

I hereby state that this research report and findings presented in it are my own and it has not been submitted before nor is it currently being submitted for any other academic programme. Where material has been used from other sources due recognition has been given by mentioning the source.

Full name of the student: - Kiringoda Arachchige Nadeeka Malkanthi

Student No: - IM/2011/033

Signature of the student: - .....

Date: - 02/01/2017

Name of the supervisor: - Dr. Thashika D.Rupasinghe

Signature of the supervisor: - .....

Date: - .....

# ACKNOWLEDGEMENT

First and foremost, I would like to offer my sincere gratitude to my research supervisor Dr. Thashika Rupasinghe. who guided me by providing supervision and direction which truly helped the progression and smoothness of the research.

I would be grateful to our research project coordinator Dr. T.K Wijesiriwardhane and all the lecturers of Department of Industrial Management for giving us an immense support, encouragement and guidance to get maximum use of knowledge and capabilities.

I would like to express my deep sense of gratitude to Julian Mcauley, Assistant Professor, Stanford University for his kind support in providing his dataset for my research purpose.

Last but not least special thanks should be given to all of my batch mates for extending their supportive hands of friendship towards the successful drive of the research.

Nadeeka Kiringoda

UG. B.Sc. Management and Information Technology

University of Kelaniya

# ABSTRACT

The internet-based technologies have influenced all parts of human lives within a short time. The internet is used for conducting commercial transactions electronically and it is the base of the concept called e-commerce. Most of the businesses have engaged in utilizing the internet to sell their product and services. Hence, spend millions of dollars to create and maintain their corporate websites. The consumer behavior in online shopping is continuously changing due to the personal characteristics of the shoppers as well as the environmental factors. The e-commerce based transactions are becoming more popular and the number of consumers who interact with the e-commerce sites have been drastically increased along with the reviews they leave after purchases. This makes it difficult for potential customers to read, comprehend, and make sound decisions on individual purchases. Furthermore, makes even difficult task for the corporate entities to track their websites to manage customer opinions. Text mining is the process which explores, evaluates, and interprets data patterns by converting unstructured text data into more meaningful information. In this study, we address the aforementioned issues by proposing a Human Computer Interaction (HCI) enabled SVM classification approach and lexicon based method use to extract features from reviews and keyword vector based method has been developed to categorize the online reviews of e-commerce websites according to HCI factors. HCI factors such as; usability, simplicity, and accessibility are considered along with consumer reviews extracted from the attribute dictionaries such as Stanford parser and past literature. The Study has derived different sentiment scores for each HCI attributes and illustrate the results which are affect to consumer browsing experiences and their buying behavior. So this study has help to from the corporate standpoint for improving HCI attributes to increase and improve good browsing experience while increasing consumers interaction with website

***Keywords: - Human Computer Interaction (HCI), Online Purchasing, Predictive Analytics, Text Mining***

# CONTENTS

DECLARATION .....	ii
ACKNOWLEDGEMENT.....	iii
ABSTRACT.....	iv
LIST OF FIGURES.....	vii
LIST OF TABLES.....	viii
ACRONYMS .....	viii
CHAPTER 1 -INTRODUCTION.....	1
1.1    Introduction .....	1
1.2    Purpose of the Research .....	2
1.3    Rationale for the Research.....	3
1.4    The Scope .....	4
1.5    Objective (s) of Research .....	4
1.6    Research Questions.....	5
1.7    Expected Outcome (s).....	5
CHAPTER 2 -LITERATURE REVIEW .....	6
2.1    Web Consumers Buying Intention in Web-Based E-Commerce .....	6
2.2    Factors which Influence to Online Buying Behavior .....	7
2.3    Predictive Analytics .....	10
2.4    Text Mining and Sentiment Analysis.....	10
CHAPTER 3 -METHODOLOGY .....	18
3.1    Research Approach- Experimental.....	18
3.2    Design of Research.....	18
3.2.1    Data Collection and Preprocessing .....	21
3.2.2    Feature Extraction.....	22
3.2.3    Model Generating .....	23
3.2.4    Measure Consumer Experience with Respect to each HCI Attributes.....	26
3.2.5    Evaluate Final Level of Consumer Experience .....	27
3.2.6    Predictive Dashboard .....	28
3.2.7    Making Prediction .....	28
3.2.8    Model Validation.....	29
3.2.9    Development Environment.....	30
CHAPTER 4-IMPLEMENTATION .....	31
4.1    Developing Environment .....	31
4.2    Data Set.....	32

4.3	Import data into R.....	32
4.4	Pre-Processing.....	33
4.5	Feature Extraction.....	33
4.5.1	Pseudocode.....	34
4.6	Model.....	34
4.6.1	pseudocode.....	36
CHAPTER 05-MODELLING THE CONSUMER BEHAVIOR WITH RESPECT TO HUMAN COMPUTER INTERACTION FACTORS WHICH AFFECT TO WEBSITE .....		38
5.1	MODEL DESIGN .....	38
5.2	MODEL DEVELOPMENT.....	39
5.3	Results.....	40
5.4	Conclusion.....	42
CHAPTER 06-DISCUSSION AND FINDINGS.....		43
6.1	Experiment Design Working Environment.....	43
6.2	Compiling HCI attributes Dictionary .....	44
6.2.1	Word Cloud .....	44
6.2.2	High Frequency words in the testing dataset .....	45
6.2.3	Text mining with SVM classifier .....	46
6.2.4	Text mining with Lexicon based Developed Algorithm.....	49
6.2.5	Comparison SVM and lexicon based comparison and Validation.....	49
6.3	Analyzing the results of the Model .....	50
6.3.1	Test Results .....	50
6.3.2	Analysis and comparison.....	50
6.3.3	Polarity Scores for each Categories .....	51
6.3.4	Negative and positive opinion score comparison.....	52
6.3.5	Satisfaction Scale according to each HCI factors .....	53
CHAPTER 06-CONCLUSION AND FUTURE WORKS .....		55
6.1	Conclusion.....	55
6.2	Limitation of the study.....	57
6.3	Future Works .....	57
REFERENCES .....		58

# LIST OF FIGURES

Figure 1- 1-scope of the research .....	4
Figure 2- 1-Factors influence to online buying behavior .....	8
Figure 2- 2-Process of Text Mining .....	11
Figure 2- 3-Summary of Text Mining Techniques .....	12
Figure 2- 4-Summary of Basic Text retrieving Methods .....	13
Figure 2- 5-Techniques in Sentiment Analysis .....	14
Figure 3- 1-Keyword vector.....	24
Figure 3- 2-Model Development.....	25
Figure 3- 3-Full customer experience related to one attribute .....	26
Figure 3- 4-Final Evaluation on attributes.....	27
Figure 3- 5-Apply Model .....	28
Figure 4- 1-Data Types Which Can Import Into R .....	32
Figure 4- 2-Steps of Pre-Processing .....	33
Figure 4- 3-Steps for Model Development .....	35
Figure 5- 1-Model Design.....	39
Figure 5- 2-Consumer behaviour changing when change the quality of graphics in website A .....	40
Figure 5- 3-Consumer behaviour changing when change the Font size of website A.....	41
Figure 5- 4-Consumer behaviour changing when improve the accuracy of product description of website A .....	41
Figure 6- 1-Features of Computer.....	43
Figure 6- 2-R studio Environment .....	43
Figure 6- 3-Word Cloud.....	44
Figure 6- 4-High Frequency words.....	45
Figure 6- 5-Results of SVM classifier .....	47
Figure 6- 6-Results of SVM classifier .....	48
Figure 6- 7-Results from lexicon based developed algorithm .....	49
Figure 6- 8-Testing data set results in R studio .....	50
Figure 6- 9-Results After export into excel work sheet as csv file .....	50
Figure 6- 10-Polarity Scores for each Categories .....	51
Figure 6- 11-Negative and positive opinion score comparison .....	52
Figure 6- 12-Satisfaction Scale according to each HCI factors.....	53

# LIST OF TABLES

Table 2- 1-Comparison of Text Mining Models exists from Literature .....	16
Table 4- 1-Libraries Used in Working Environment .....	31

# ACRONYMS

ASUM	Aspect and Sentiment Unification Model
HCI	Human Computer Interaction
IAA	Impact-Asymmetry Analysis
IPA	Importance Performance Analysis
IRPA	Impact Range Performance Analysis
JST	Joint Sentiment Topic
LDA	Latent Dirichlet Allocation
PA	Predictive Analytics
SDLA	Sentence Latent Dirichlet Allocation
SVM	Support Vector Machines
TAM	Technology Acceptance Model



# CHAPTER 1 -INTRODUCTION

## 1.1 Introduction

Through the evolution of technology, internet technology appeared during the last quarter of the 20th century. It influenced all parts of human lives frequently in a short time. Moreover, the internet is used for commercial purposes and it is the base of e-commerce concept.

E-commerce is defined in the Oxford dictionary [2016] as commercial transactions conducted electronically on the internet. Furthermore, e-commerce is referred as trading or facilitation of trade in products and services through computer networks. In addition, e-commerce is a process of buying, selling, transferring or exchanging products, services, and information via electronic networks and computers.

Most of the businesses are currently engaged in utilizing the internet to sell products and services. It offers the opportunity to expand business through emerging technologies. So that, businesses have realized the importance of the internet and e-commerce has become more popular in the business context [Miles et al.,2000]. Through the growth of online shopping, corporations spend millions of dollars to create and maintain the corporate websites to face the competition in the business environment [Hausman and Siekpe,2009]. Today, many of the corporations uses multichannel in e-commerce to communicate with the customers to increase the competitive advantage [Dikj et al., 2007]. Different types of people interact with e-commerce websites to fulfill their different purposes according to their wired lifestyle. This is called as online shopping. Consumer behavior in online shopping is continuously changing due to different types of human factors such as gender, age, religion, culture, profession, income status, education level and, etc. These important factors affecting the purchase decision process can be categorized into four dimensions. They are personal dimension, psychological dimension, social dimension, and cultural dimension. PACT (People, Activities, Context, Technology) framework describes the online consumer behavior with Human Computer Interaction.

Usability is easy to use and learn and also it is matched with most requirements close to the user. Human Computer Interaction (HCI) is different from usability and HCI is

a discipline concerned with the study, design, construction and implementation of human-centric interactive computer systems [WWW33]. HCI also affects to designing screens and menus which are easier to use and studies the functionality of humans effectively. By the time computers become more and more pervasive. Therefore, the designers design interfaces and devices which are more interactive with the user.

Predictive Analytics (PA) is the branch of the advanced analytics, which is used to make a prediction about unknown future probabilities and trends. For that Predictive analytics use many techniques from data mining, statistics, modeling, machine learning, and Artificial Intelligence. The above-mentioned techniques analyze historical data in order to recognize the patterns more important in predictive analysis. In an online purchasing, predictive analytics helps to identify and target the right consumer at the right time.

## **1.2 Purpose of the Research**

Nowadays, busy lifestyle of the people has made more popularity on online purchasing. Mainly E-commerce websites help people to buy materials online and those websites can be accessed whole day. Moreover, E-commerce websites give a lot of benefits to the consumer and to the organization. In the consumer side, these websites provide a lot of information about products and services and it has more products and services including cheaper ones. Furthermore, those organizations related to websites delivery products to the consumers and website has made benefits to the organization. It opens the door to reach to the global consumers, with the improvement of the consumer relationship. In addition, it enhances supply chain after reducing the cost.

Online Reading materials purchasing is an easy and time efficient process for the consumers. Furthermore, it is a more trustworthy process.

Consumers have different preferences based on their age, education level, and so many other factors. Since the consumers have different requirements, it is better to have a more appropriate method to interact them with the websites and enhance their revisit level. For that product reviews and consumer feedback are very important data for the businesses. If the businesses have a technical method to analyze these hidden

data, it will be very effective and supportive in order to make useful changes to the business process.

According to the literature, there is a need to have a framework which identifies the attributes of a website which are influenced to HCI. The results of the framework have been displayed through dashboard which integrated HCI factors. Furthermore, through this dashboard can be identified the quality HCI factors which are increasing the revisit level of consumers and vice versa. Therefore, it is helping to managers of the company to make decisions to increase the quality of different factors of the website to give a better experience to the consumers.

### **1.3 Rationale for the Research**

After the industrial revolution, internet plays the most powerful role in organizations to compete in today's rapidly growing digital economy. The value of e-commerce growing rapidly and it is the factor of globalizing the digital economy. Due to the evolution, customer lifestyles have become more complex and it affects the consumer's buying intention [Khaniwale, 2015]. More complex lifestyle is the power of the changing consumer traditional shopping into online. Organization concern to interact each consumer to their websites [Cao et al.,2005; Koufaris,2002]. Through the Technology Acceptance Model (TAM) and IS success model evaluate the website quality and design principles on its effectiveness to increase consumers' interaction with the e-commerce site [Cao et al.,2005]. Moreover, both HCI design and web design have different types of guidelines. Many of them are used in different segmentations in e-commerce. As well as, it supports to interact consumers to websites and change the consumer's mind to revisit the site well [Nielsen,2000; Helander and Khalid,2000]. E-commerce technology has great diversity and it includes from simple point-and-click browsing web pages to 3D shopping malls [Dikj et al.,2007]. The Consumers online buying behavior depends on consumers' attitudes, emotions, and their needs. To predict their behavioral intention, different type of models, theories are used. Furthermore, theory of planned behavior and theory of reasoned action also used to predict consumers online buying intention [Hansen et al., 2004]. Neural networks and logistic regression model used to identify the patterns used in consumer buying behavior after building human pattern recognition function [Chiang et al.,2006]. Siegel [2013] has described and evaluated the power of

prediction through many real-world examples. Later the usage of machine learning into predict future outcomes and effectiveness of decision making has explained after analyzing the historical data.

Along with the growth of the internet, a lot of users interact with the internet through the knowledge of information technology. According to that data collection mechanisms are very important and it becomes more complex and advanced. Customer feedback, product reviews, and product rating are the unstructured data to make a decision on consumer buying intention after handling, analysis and interpreting the meaningful structured data. For that, text mining and sentiment analysis provide different methods [Hu and Liu,2004; Jo and Oh,2011].

According to the previous literature, this study also focuses on HCI, PA, and E-commerce. In addition, it will be focused on online reading materials purchasing by applying Human-computer interaction factors such as age, education level, culture, religion. Moreover, this research based on predictive analytics and text mining to help to organizations to make decisions.

## 1.4 The Scope

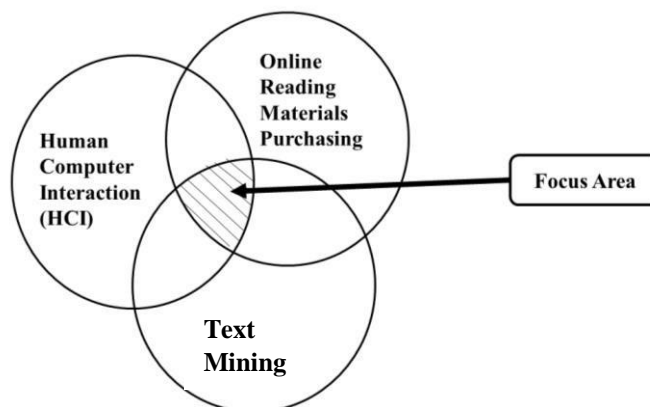


Figure 1- 1-scope of the research

## 1.5 Objective (s) of Research

According to the study, different main and sub objectives can be categorized as follows.

- Identify the background of E-Commerce websites related to online purchasing of reading materials.

1. Identify the relationship between Human Computer Interaction factors such as usefulness, accessibility and simplicity and online purchasing from websites.
  2. Identify the quality of different attributes/features related to HCI of this website to interact the consumers.
- Identify possible suggestion for an E-Commerce website to improve its better for satisfying customers and the organization.
    1. Analyzing historical data related to online purchasing of reading materials for identify the Human Computer Interaction factors which affect the customer buying behavior.
    2. Develop a framework to evaluate the e-commerce website, based on HCI factors and sentiment mining. That means to prioritize the factors which affect “good browsing” experience.
    3. Visualization the results with appropriate tolerance.

## **1.6 Research Questions**

RQ1. What are the Human Computer Interaction factors affect to the online purchasing?

RQ2. What are the methods to capture consumer reviews, feedback from E-Commerce websites?

RQ3. How to develop a framework to evaluate the e-commerce website applying Human Computer Interaction factors?

RQ4. How to use the framework to make decision about website quality and its design?

## **1.7 Expected Outcome (s)**

The expected outcome is to meet the research objectives. Thus, at the end of this research, the deliverables will be

- Identify Human Computer Interaction factors, which is used much more in E-Commerce websites .

- Identify attributes related to HCI of online purchasing website and its relationship with consumer buying behavior.
- Come up with a framework which embedded from Human Computer Interaction (HCI) factors.
- Use that framework to make the best decision to better market the consumers with appropriate tolerance.

## **CHAPTER 2 -LITERATURE REVIEW**

This section reviews previous literature on study of Human Computer Interaction, online purchasing, E-commerce, Predictive Analytics, and text and sentiment analyses

### **2.1 Web Consumers Buying Intention in Web-Based E-Commerce**

All the consumers are now using computers through the knowledge of information technology. In this environment, the traditional consumer also accesses the web to purchase products and services. Therefore, online consumer performs all the functions through online purchasing as same as traditional consumer does. Consumer buying behavior depends on various human computer interaction factors such as accessibility, simplicity, etc. Further consumer purchases through e-commerce websites online or offline. These online consumers have more powerful demand and utilization in their shopping expeditions than offline consumers. This study has concentrated on new customer retention and unplanned purchasing through newly created framework with the combination of technological acceptance model, marketing, and psychology. This framework study about web consumers buying behavior and recognize the patterns. In addition to that, it helps to identify emotional and cognitive responses after evaluating consumers and website factors. Further, this study has used Technological Acceptance model in different ways to evaluate web consumers buying behavior [Koufaris,2002].

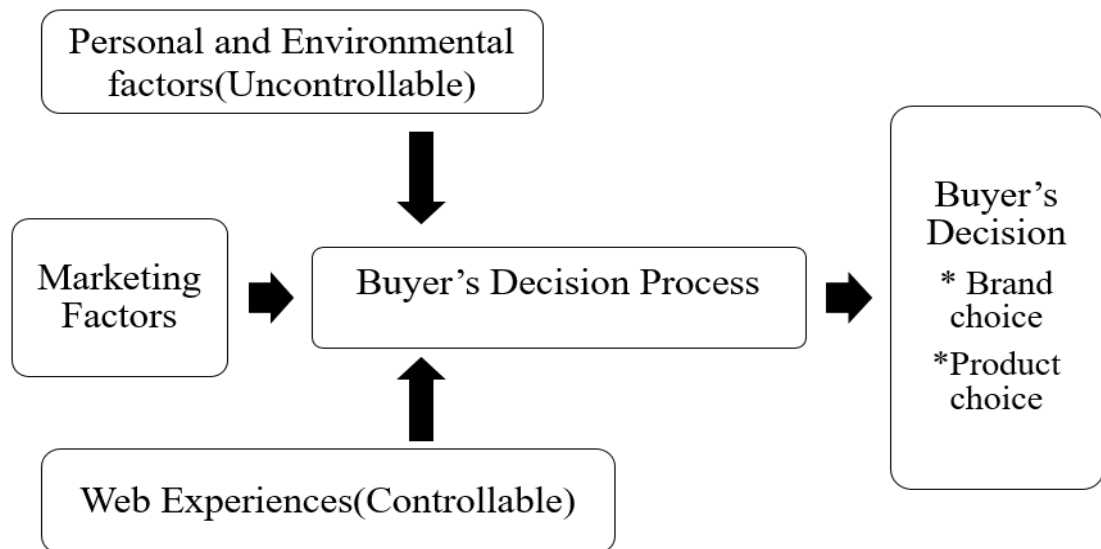
Cao et al. [2005] have argued the effect of website quality for increasing the consumer buying intention from commercial e-commerce websites. This study has basically done to the commercial e-commerce websites using the TAM model, IS success

model with the concepts of trust. TAM model is widely used for diagnosing the problems in e-commerce website before consumers' experiences from it. Further, it determines two factors in the user acceptance website. 1. Perceived usefulness 2. Perceived ease of use.

## **2.2 Factors which Influence to Online Buying Behavior**

Number of factors have influenced on online consumer buying behavior. It is a complex mixture of combination of so many factors such as social, cultural, educational, personality, environmental factors. Among them, some factors are controllable and some are uncontrollable. Furthermore, trust of the website, confidence, appealing website, proper and full information about goods and services are encouraged to make purchasing decision [Kotler,2003].

According to the online purchasing, it' based on three factors. Personal and environmental factors, Marketing factors and web experiences factors. Furthermore, these factors can be revealed as consumer characteristic, Customer concerns in online shopping, and Human computer interaction factors. Customer concern in online shopping which is investigate about human behavior and its attributes which are influencing online shopping behavior. Trust is the main factor and it is a combination of Safety and privacy of information, security and delivery and return on time. Consumer characteristics such as demographic factors, personality, tradition and cultural factors have played a major role and depend on other factors such as computer literacy, technology awareness, past experiences also [Wan, 2000].



*Figure 2- 1-Factors influence to online buying behavior*

Source: [Kotler,P,2003]

Human Computer Interaction is mainly concerned about website design, website interaction, it's display, user's easiness to learn, efficient which are effected on consumer behavior. Furthermore, this is fulfilling through characteristic which are related to the website such as information available on website, visual effects and attractiveness, quality of content, ease of navigation, less time consuming and security assurance of the website [Wan,2000]

Through the revolutionized web and internet e-commerce is continuously growing. Today the e-commerce web environment has differed from early days. As the competition grows between online consumers and shoppers, organizations cannot simply assume that if their website customers will be sited. Therefore, the e-commerce sites have to be designed with the fulfillment of user expectation [Nah and Davis,2002].

Furthermore, it was highlighted the challenges faced by browsing the websites for searching information which split into three major categories. 1.Web usability 2. Interface design 3. Trust. Emotions are the fundamental component of a human. Emotions such as joy, anger, hate, pride, etc., and those emotions change time to time. Therefore, it is identified as Human computer interaction and emotion connect with a strong relationship to change consumer buying behavior [Brave and Nass,2003].



Miles et al. [2000] have studied customer product features reviews from a number of different types of e-commerce websites and proposed a new framework. It evaluates the four types of design dimensions include in e-commerce websites. 1. Front end dimension 2. Criteria management dimension 3. Comparison support dimension 4. Marketplace dimension. Further, this literature has identified e-commerce as a decision support system.

E-commerce has three subsystems 1. Web store 2. Customer 3. Web technology. A customer plays the stronger role. Through that consumer interaction started the transaction of information through the subsystems of e-commerce. Therefore, Human Computer Interaction and web design guidelines are valuable to connect to the subsystems [Helander and Khalid, 2000]. In addition, a model was proposed to illustrate the flow of information and that model dynamically changes with the changes of consumer decision making.

Usability is ease of use and learnability of a human made objects. Christine et al. [2001] have proposed three different types of models after linking the usability concepts with trustworthiness for evaluating web-based e-commerce sites. It finds out e-commerce website design with concern usability factors and identify trustworthiness of the websites for its consumers. In addition, Usability factors can be referred as 1. The ease of learning how to use website 2. The efficiency of the interface design 3. The ease of memorizing how to use interfaces 4. The ease of recovering from errors 4. General satisfaction with the interface [Nielsen 2012].

Kuan et al. [2003] have proposed new framework after evaluating past methods to give a better understanding of user satisfaction and consumer intention of planned purchases in web-based e-commerce. This model concerns on four types of believing used in e-commerce websites. 1. Perceived website usefulness 2. Perceived interface quality 3. Perceived information quality 4. Perceived service quality. Those factors do not only focus on the theoretical environment, but also on the empirical environment on the website to support consumer believes and user expectation. Furthermore, it has identified a strong relationship between user satisfaction and the intention to purchase. With the advent of information technology, e-services distribute through a number of channels. That means multichannel environment. The consumers buying behavior through different types of channels in the environment of multi-channels are continuously changing through the choices of consumers on those channels [Dikj et al., 2007].

Within the web environment, human factors can be categorized into five types. 1. Enjoyment 2. Cognitive outcome 3. User empowerment 4. Visual appearance 5. Organization of information content. Through that number of web, design tools increase the purchasing intention and reuse intention of e-commerce sites [Hausman and Siekpe,2009].

## **2.3 Predictive Analytics**

Predictive analytics is the branch of advanced analytics. It measures future outcomes through analyzing the historical data, using machine learning techniques, and statistical algorithms.

Siegel [2013] has played a major role in predictive analytics to make the power of prediction of future outcomes through analyzing historical data in big businesses.

Hansen et al. [2004] have tested the ability and the power of predicting consumers online buying intention of two consumers buying theories through the study of consumers online buying behavior. According to the two theories, the theory of reasoned action and theory of planned behavior, the theory of planned behavior provides better ability to predict the consumer's buying intention.

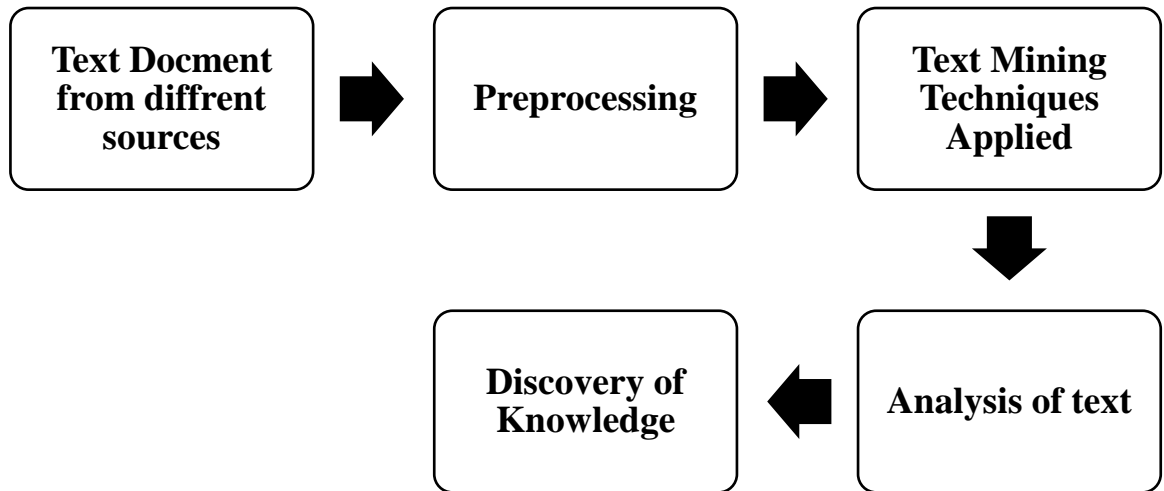
Chiang et al. [2006] have explained and predicted the traditional and web consumers' choices on their future buying intention through the neural network model and logistic regression model.

Hair [2007] has studied about predictive analytics in market segmentation and its' future development trends of marketing area. Furthermore, literature perceived the importance of PA in knowledge creation and its' overall effect on the organization.

## **2.4 Text Mining and Sentiment Analysis**

### **Text Mining**

Text mining is the technology which is used to extract information and discover new knowledge from unstructured and semi-structured data sources like different written resources, and websites. According to the text mining process, it is a step by step process and it started with the collection of various types of documents. Figure 2-2 has illustrated the process of text mining.



*Figure 2- 2-Process of Text Mining*

Source: [Gaikwad et al.,2014]

Furthermore, text mining is a collection of different types of methods and techniques. Information extraction, summarization, categorization, clustering and information visualization, are the techniques used in the text mining process for filtering, analysis, generating the text. Moreover, it helps to manage a great amount of unstructured information for extracting patterns. Figure 2-3 illustrated the summarization of the text mining techniques.

<b>Information extraction</b>	<ul style="list-style-type: none"> <li>• Identify the key phrases and relationships within the text.</li> <li>• Transform a textual database into more structured databases.</li> <li>• ASUM, JST models are used to extract information</li> </ul>
<b>Categorization</b>	<ul style="list-style-type: none"> <li>• Used in classifying the newly arrived documents</li> <li>• Naïve Bayes model, SVM model used to categorized the text</li> </ul>
<b>Clustering</b>	<ul style="list-style-type: none"> <li>• Find groups of documents with the similar content</li> <li>• K -means clustering used to cluster the text</li> </ul>
<b>Visualization</b>	<ul style="list-style-type: none"> <li>• Improve and simplify the large textual sources into a visual hierarchy.</li> <li>• Improve understandability</li> <li>• Used for government purposes (identity terrorist networks)</li> </ul>
<b>Summarization</b>	<ul style="list-style-type: none"> <li>• Reduce length and detail of documents and point out the important data</li> </ul>

*Figure 2- 3-Summary of Text Mining Techniques*

According to the information retrieval to fulfil user requirements, there are four basic text mining methods.1. Term based method 2. Phrase based Method 3. Concept base Method 4. Pattern taxonomy Method [Gaikwad et al.,2014].

<b>Term Based Method (TBM)</b>	<ul style="list-style-type: none"> <li>• Used to analysis the term in the document.</li> <li>• Efficient the computational performance</li> </ul>
<b>Phrase Based Method (PBM)</b>	<ul style="list-style-type: none"> <li>• Used to analysis the phrases which are less ambiguous and more discriminative than individual terms</li> </ul>
<b>Concept Based Method (CBM)</b>	<ul style="list-style-type: none"> <li>• Analysis sentences and document level effectively and discriminate between non-important terms and meaningful terms.</li> <li>• Useful for natural language processing</li> </ul>
<b>Pattern Taxonomy Method (PTM)</b>	<ul style="list-style-type: none"> <li>• Used to analysis patterns basis.</li> <li>• Pattern can be structuted into taxonomy using in a relationship</li> </ul>

*Figure 2- 4-Summary of Basic Text retrieving Methods*

### **Sentiment Analysis and Opinion mining**

Sentiment analysis is a new kind of text analysis which aims at determining the opinion and subjectivity of reviewers. Furthermore, it has helped to classify the huge number of texts through supervised learning.

Different techniques used to analysis the text to extract the sentiments. From a technical point of view, machine learning, lexicon-based, statistical and rule based approaches have been used.

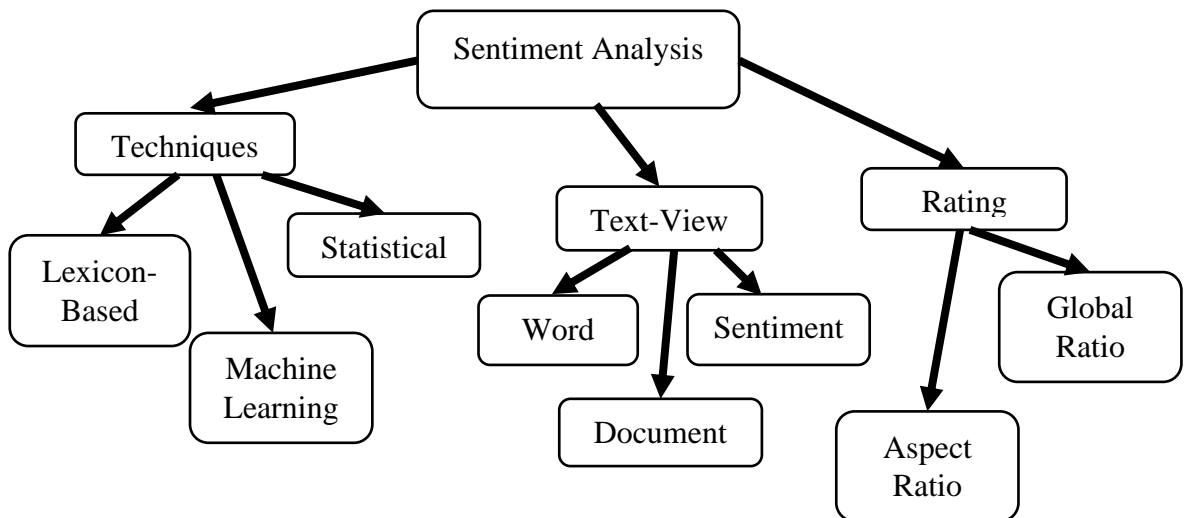
**Machine Learning Methods:** use this method for several algorithms such as a Naïve Bayes classifier to determine the sentiments after training it through known data set.

**Lexicon Based Methods:** use this method through calculating sentiment polarities for each and every semantic orientation of words in reviews.

**Rule Based Method:** use this method through identifying opinion words in a text and considering rules and classify those.

**Statistical Method:** use this method after representing each review as a mixture of latent aspects and rating.

Sentiment Analysis can be done in another aspect also. Such as text view and rating view. Below diagram shows all classification done in sentiment analysis.



*Figure 2- 5-Techniques in Sentiment Analysis*

Source: [ Analis et al,2014]

Not only that, but also a lot of past studies have done on text mining and sentiment analysis. It has introduced different types of models after combining text mining methods and techniques to evaluate reviews and discover new knowledge. Furthermore, these models help to summarize a large number of reviews.

Jo and Oh [2011] have proposed two different models to discover and evaluate different reviews and sentiments, and analysis those sentiments in a rich way. Sentence Latent Dirichlet Allocation (SLDA) model is mainly proposed for discovering and evaluating sentences in the reviews using one single aspect. Aspect and Sentiment Unification Model (ASUM) is the extension of SLDA and it discovered review sentences in different aspects and through this, all customer reviews can be summarized. In addition, ASUM performed a quantitative evaluation of sentiment classification.

Li and Wu [2010] have proposed an algorithm to analyze the emotional polarity automatically for detecting and forecast online forum hotspots. This algorithm

combined with K-means clustering and Support Vector Machines (SVM) classification to cluster analysis. Furthermore, this algorithm based on text mining and sentiment analysis techniques and it is very helpful for the organizations for future predicting approaches.

Lin and He [2009] have also proposed a probabilistic framework for sentiment analysis through Latent Dirichlet Allocation (LDA) and Joint Sentiment Topic (JST) models. According to the literature perceived JST model is more suitable for document-level sentiment classification compares to other existing supervised approaches.

Ye et al. [2009] have proposed three supervised machine learning algorithms for reconstructing the sentiment classification through consumer reviews. Naïve Bayes model, SVM classifies Model and dynamic language N-Gram Model are the three models which had been proposed for mining consumer reviews. SVM classifying model and N-Gram model are better for mining the sentiments, and an overload of consumer reviews.

Archak et al. [2007] have proposed a method after combining text mining techniques to summarize all the customer reviews for individual product features through analyzing consumer reviews. Furthermore, literature has suggested a model called econometric model. It demonstrates the value of consumer reviews on the internet using economic data.

Hu and Liu [2004] have proposed a different set of techniques for mining and summarizing a large number of customer reviews and provide a feature based summary of products sold out. Those proposed techniques based on natural language processing methods /text mining methods and data mining methods.

Hu et al. [2012] have also proposed methods combining sentiment mining techniques to evaluate consumer review and detect the manipulation of review. This method identifies the writing styles of the reviews and it reflects the consumers purchasing decision significantly.

Mostafa [2013] has proposed qualitative and quantitative methods to analyze the consumer reviews on the social media. Early mentioned methods are used to review consumer reviews about well-known brands like Nokia, IBM. According to the literature, qualitative and quantitative methods are better than the traditional methods such as focus group and personal interviews to analysis consumer reviews on well-

known brands. Table 2-1 summarizes the models which are used as a methodology in past literature to fulfil user and organizational requirements.

*Table 2- 1-Comparison of Text Mining Models exists from Literature*

Text mining models	Advantages	Disadvantages
Support vector machines model	<ul style="list-style-type: none"> <li>• Categorize traditional text and work better for text classification.</li> <li>• High accuracy.</li> </ul>	<ul style="list-style-type: none"> <li>• Inefficient to train.</li> <li>• Not suitable for “industry scale” applications and many training applications.</li> <li>• No standard way for dealing with multiclass problems.</li> </ul>
Naïve Bayes model	<ul style="list-style-type: none"> <li>• Easy to implement</li> <li>• Require a small amount of training data to estimate the parameters.</li> <li>• Good results obtained in most domains.</li> </ul>	<ul style="list-style-type: none"> <li>• Low accuracy.</li> <li>• Very simple representation doesn't allow for rich hypotheses</li> </ul>
N-gram based language model	<ul style="list-style-type: none"> <li>• Used for natural language processing</li> <li>• Takes characters (letters, space) as the basic unit</li> <li>• Accuracy is high</li> </ul>	<ul style="list-style-type: none"> <li>• Can't take words as basic units</li> </ul>
ASUM	<ul style="list-style-type: none"> <li>• Discover sent aspects in the reviews in an unsupervised way.</li> <li>• Capture aspects that are closely coupled with a sentiment</li> </ul>	



	<ul style="list-style-type: none"> <li>• Illustrate different sentiments in same aspects</li> </ul>	
Joint sentiment topic model	<ul style="list-style-type: none"> <li>• Detect sentiments in document level and extract a mixture of texts.</li> <li>• More flexibilities and can be adapted to other applications</li> </ul>	<ul style="list-style-type: none"> <li>• Detect sentiments without any word order.</li> </ul>
Logistic regression model	<ul style="list-style-type: none"> <li>• Logistic regression is a pretty well-behaved classification algorithm that can be trained as long as you expect your features to be roughly linear and the problem to be linearly separable. Feature engineering to turn most non-linear features into linear pretty easily. Output can be interpreted as a probability</li> </ul>	<ul style="list-style-type: none"> <li>• Limited expressive power. Hard to implement.</li> </ul>

# CHAPTER 3 -METHODOLOGY

The general research methodology has used for the research is described in this chapter.

## 3.1 Research Approach- Experimental

Five different types of research approaches can be adapted on a research; namely, experiment, survey, archival analysis, case study and history. Survey and experimental researches are basically coming under quantitative approaches. Case study, research, ethnography, action research and grounded theory approach can be taken under qualitative approaches.

As objectives of this study are to measure the consumer experiences in the process of buying, reading materials from an e - commerce website and haven't concern about products. Furthermore, it has measure the final customer satisfaction for all attributes related to HCI. The quantitative research approach is more preferable in this case. The survey research approach is not in line with the research objectives. As this study is to measure and identify the HCI factors which are affect to change consumer interaction with online reading material purchasing e-commerce website. Therefore, the experimental research approach is more preferable for this study. The research has started with a studying an existing literature to find out the approaches, gaps and applicability of existing available literature. Then research questions are formulated.

## 3.2 Design of Research

The ultimate goal is identified the most satisfaction HCI factors and come up with a predictive model to analysis the consumer profile based on their reviews through appropriate tolerance.

This study has contained with main two phases.

- 1.Training and validate the model after generating

2.Extend and used to make predictions.

According to this level first step only have developed and done.

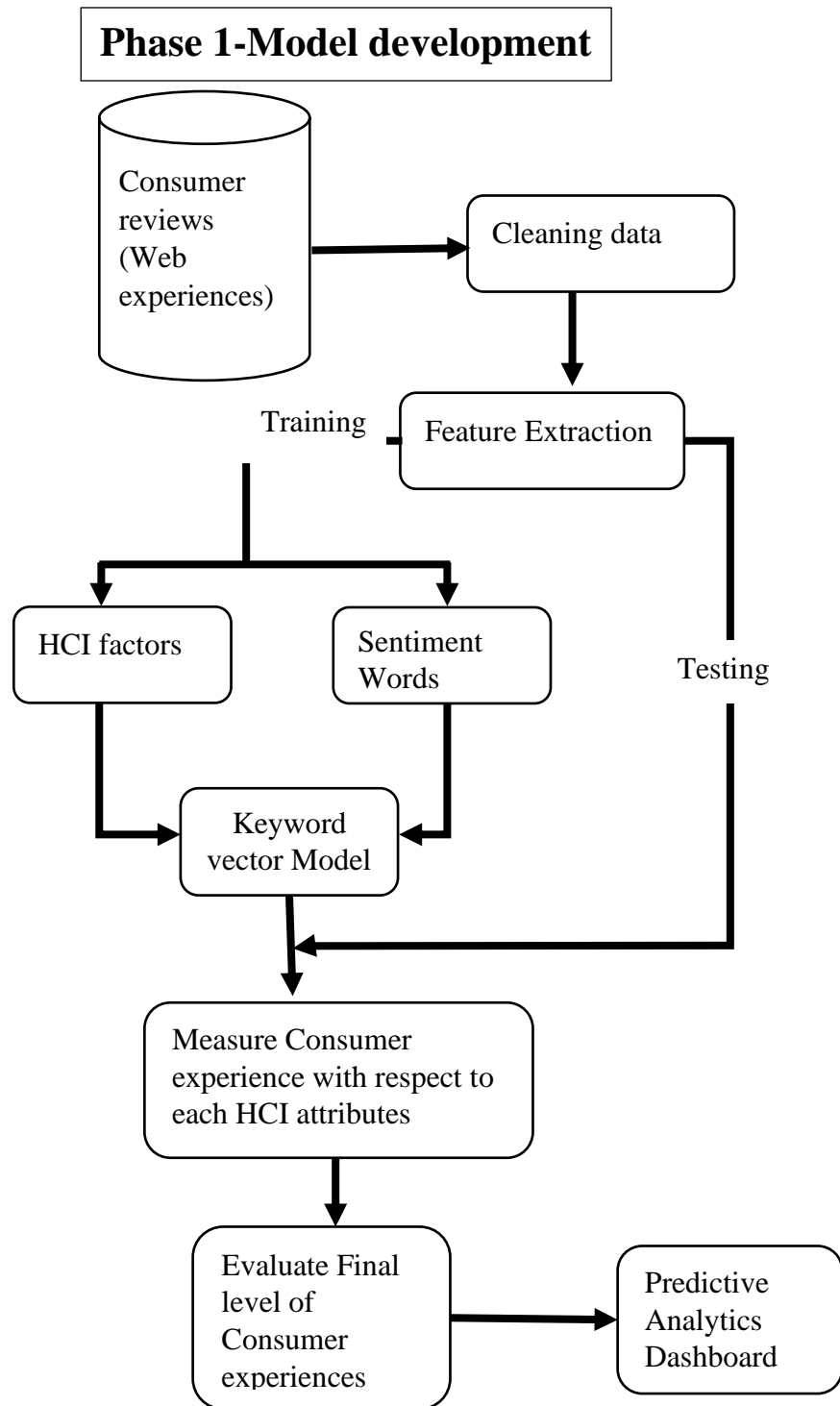


Figure 3- 1-Model Development

### 3.2.1 Data Collection and Preprocessing

Since the study is focused on the consumer's web experience in online reading material purchasing, the secondary data is collected related to it through the internet and prepare for the next phases. The data collected from Data Dives which is enabling users to interact with big data by running queries, creating elements, sharing results, uploading their own data and much more. The dataset is Amazon Product Reviews and all products including title, reviews, price and rating. SNAP at Stanford (<https://snap.stanford.edu/>) is the main source for this data set. Further, from website called sitejabber data also extracted and used.

Through the query running in the data dives, data relevant to the online reading materials purchasing have been selected from Amazon Product Reviews dataset. It also includes the title, review summary, review, customer rating, etc.

Since the visualization of the data is needed to understand about the distribution of data in the data set. Hence, the dataset has been explored and visualized using scatter plots and histograms. Through this, it can be identified outliers in the dataset. Therefore, data has been cleansed by leaving out outliers and the all tuples which contain the N/A values and unknown values. Because, inconsistent data can be laid waste with the results. After cleaning the dataset, it has been prepared in a proper way. Therefore, all data has been included Book Title, Book category, Reviewers Profile Name, review summary, review score and review text.

Source of Dataset: <https://www.datadives.com/>.

SQL query for that:

```
SELECT
    product_price,
    product_productId ,
    product_title,
    review_helpfulness,
    review_profileName,
```

```
review_score,  
  
review_summary,  
  
review_text,  
  
review_userId,  
  
review_time  
  
FROM  
  
ds_90.tb_10804  
  
WHERE review_summary CONTAINS 'book'
```

### 3.2.2 Feature Extraction

Keywords are set of significant words in a document that give a high-level description of the content for investigating readers and are useful tools for many purposes. Furthermore, keyword extraction is necessary for many purposes such as text categorization. Two types of approaches are in feature extraction. One is being the keyword extraction with prior dictionary and other one is being the keyword extraction from text without prior dictionary. For that, set of techniques has to be followed.

**Remove Punctuation:** All punctuation marks which are in the reviews have to be removed. It has included special characters also.

**Remove Numbers:** All numbers which are not given any meaning to the text have to be removed.

**Filter Noisy data:** All noisy words filtered into its correct word such as happpppy into happy.

**Remove stop words:** Stop words are words which can exclude from any description. Articles (a, an, the), Prepositions (in, for, from, with, within etc.) and conjunctions (and, or, but etc.) are included in this list. These words don't contain information about the items.

**Stemming:** All words which mean one word have reduced into its stem form. Within an example, experience, experiences-all these words can be reduced to its stem form as experience. This technique is very useful to determine the similarity and distance according to consumers presenting style to the same feature.

**Compile two dictionaries:** A dictionary of attributes and a dictionary of sentiment words have been compiled based on the collected review data. The dictionary of attributes comprises the important HCI factor related attributes/features which are extracted from reviews using Stanford parser. Also, sentiment words which describe the positive and negative experience related to the attributes have extracted from the reviews again using the Stanford parser. Through these dictionaries.

### 3.2.3 Model Generating

**Construct keyword vector to consumer reviews.**

After extracting keywords and compiling two dictionaries, consumer reviews are constructed into keyword vectors with reference to the two dictionaries. Hence, to apply this keyword vector for all reviews and get measured from its model has developed.

In the consumer review, attribute which measure HCI factors and sentiment words appear in the same sentence have been considered as customer opinion. In this situation, the sentiment score is counted for the corresponding attribute.

For an example, following review was written by a consumer in amazon website related their experiences from it.

“Graphics are perfect. Colors are also good.”

According to this example, keyword vector has been constructed with reference to the dictionaries.

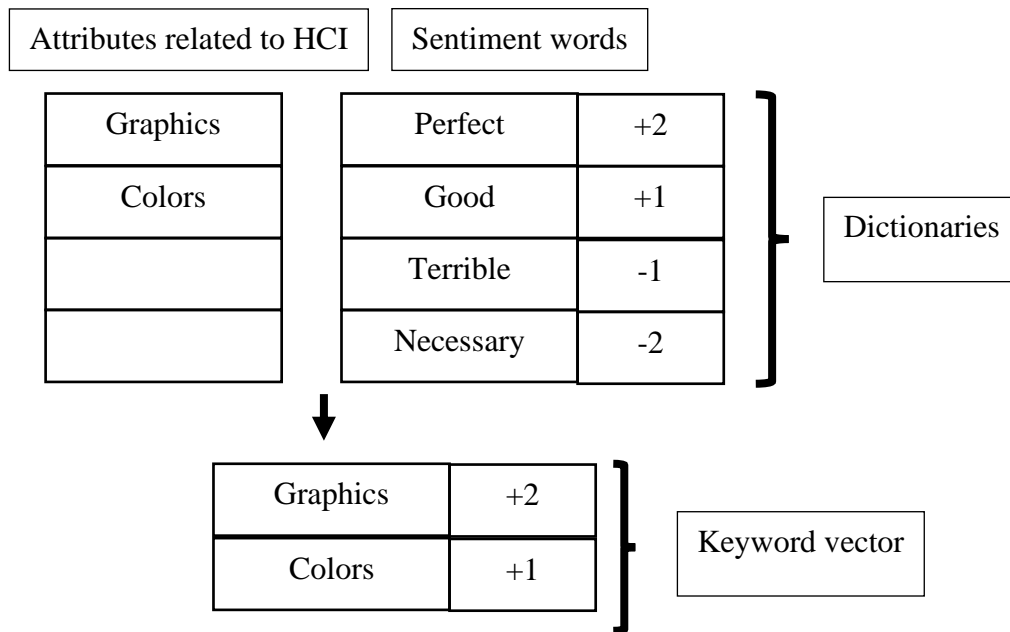


Figure 3- 2-Keyword vector

The above keyword vector can be constructed for one review. Therefore, a model is developed to support multiple reviews. Key word vector is the basic structure for the model. Furthermore, it has been a multidimensional array of the different attributes. In other hand this model has been developed without multidimensional array and only key word vector. category wise data store in same keyword vector.



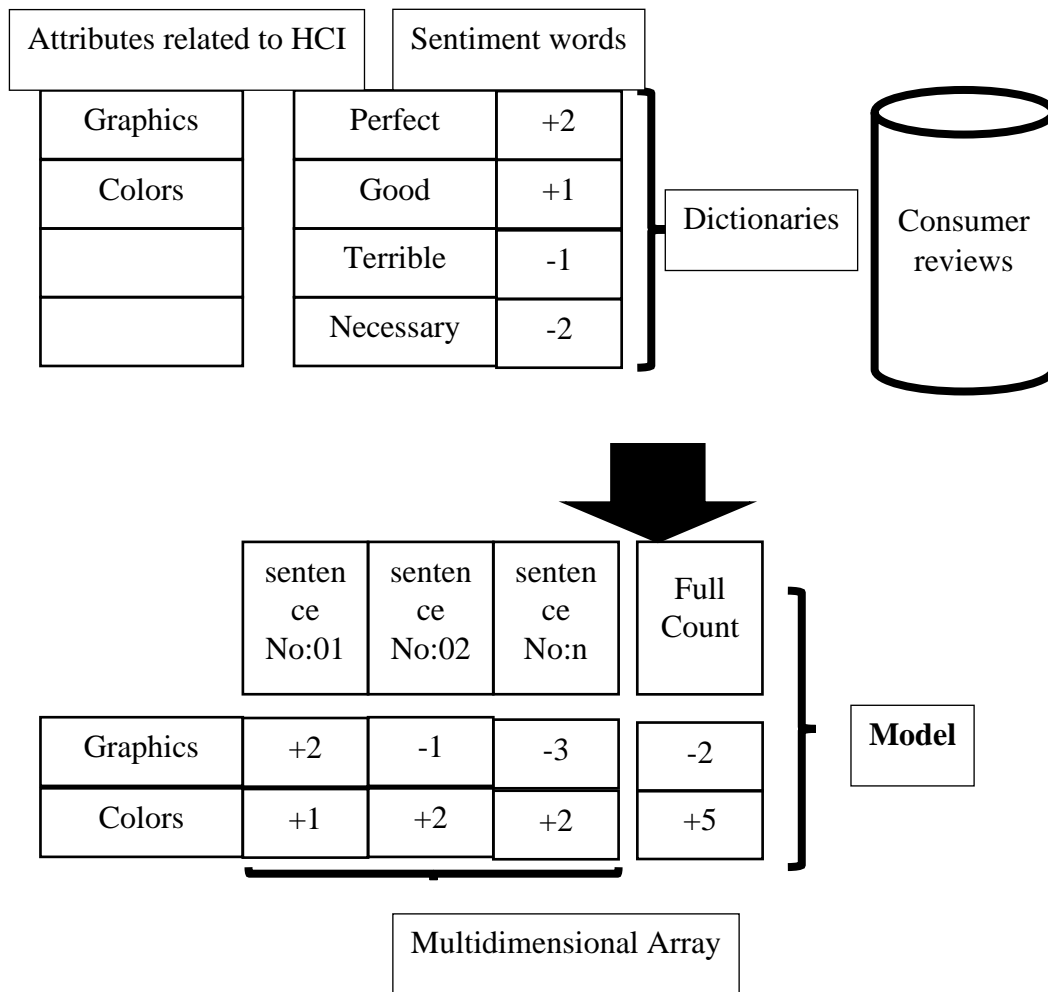


Figure 3- 3-Model Development

This is the structure of the model. According to the two dictionaries, ratings for attributes are recorded for all reviews in a multidimensional array. That means, using this keyword vector is extended to every customer review that attributes appear in.

### 3.2.4 Measure Consumer Experience with Respect to each HCI Attributes

Through the model, all sentiment words of attributes which is detected from individual customer review has been extended to all reviews appear on the web site. Through this polarity for sentiment words have been counted. Thus, through this study the level of consumer experiences with respect to Human computer interaction factors which is embedded in the website are measured by summing up the polarity scores from consumer reviews which is recorded in a multidimensional array.

$$CE = \sum_{i=0}^n (p_1 + p_2 + p_3 + \dots + p_n)$$

CE: Total sentiment polarity for attribute related to HCI

P<sub>i</sub>: Polarity of Sentiment word

According to the model which describes by Figure 3-4 The figure has calculated its all attributes total sentiment polarity using this function. According to that, it has clearly sum up Graphics have -2 polarity rating and Colors have +5 polarity rating.

	Review No:01	Review No:02	Review No:n	Full Count
Graphics	+2	-1	-3	-2
Colors	+1	+2	+2	+5

Figure 3- 4-Full customer experience related to one attribute

### 3.2.5 Evaluate Final Level of Consumer Experience

To formulate effective and efficient service-improvement strategies and identify the service attributes which are impacting on overall customer satisfaction. For that results of different attributes are arrange into ascending order. Furthermore, this result is help to identify different factors different effect to the consumers buying behavior. Therefore, it has helped to make different decisions.

According to the example, through this step has evaluate final level customer experiences and identified best attributes through their weighted polarity.

	Review No:01	Review No:02	Review No:n	Full Count
Graphics	+2	-1	-3	-2
Colors	+1	+2	+2	+5

Result

→

Colors > Graphics

*Figure 3- 5-Final Evaluation on attributes*

In here Graphics has -2 polarity and colors have +5 polarity. Therefore, that model has evaluate website Colors are better than Graphics. Through these steps, model has been tested and validated for future prediction.

Below steps are identified as steps of methodology. But in this stage below steps have not developed. So that, those are open the future direction paths.

### 3.2.6 Predictive Dashboard

This dashboard has displayed all attributes related to HCI which are extracted from the reviews. Furthermore, all attributes are displayed in descending order and it clearly shows the better attributes and worst attributes which are affect to change interaction with the online reading material purchasing e-commerce website. In other words, this dashboard helped to make managerial decisions to change the quality of attributes to increase consumer interaction with websites.

According to the example, the model has been identified colors are perfect, but the graphics are terrible. Therefore, dashboard which is displaying those are clearly showing that. Hence, Dashboard is helped to managers to increase the quality of graphics.

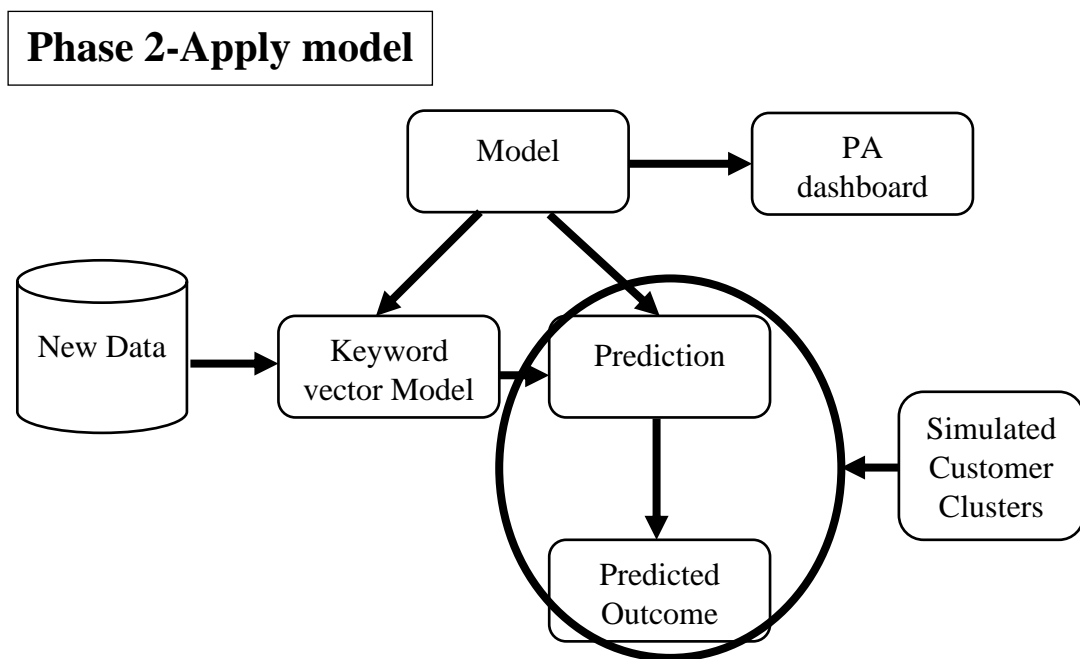


Figure 3- 6-Apply Model

### 3.2.7 Making Prediction

Humans are interacted with the outside world through information being received and sent as input and output channels. Such as visual channel, auditory channel, haptic channel, and movements. Furthermore, emotions are influenced to human capabilities.

Therefore, different types of human interaction with online reading material purchasing e-commerce websites according to their differing of input and output channels such as color blindness.

Developed model has been measured the final customer satisfaction of attributes. Furthermore, all the attributes related to HCI factors have been evaluated. Those measured attributes are recorded and display in a descending order through the dashboard.

According to that, this model is helped to identify the clusters of customers according to their reviews. The customer clusters have been identified through agent based simulation. The agent based model which developed show the consumer behavior changing with respect to the main HCI factors.

### **3.2.8 Model Validation**

Validation is to ensure that the development process improvement framework meets the objectives, research requirement intended outcomes. The intention of the validation is to assure that the framework is effective since it address the right problem in the right way.

For that newly created framework initially validate through after the steps in the model development process. After that, using testing data set it has been validated. For that compares the results from test data set and training data set.

### **3.2.9 Development Environment**

When selecting the tools and technologies for the implementation of the model has considered about the matching all requirements of the algorithm with tool libraries and other packages. Large dataset has been used for predictive analytics, machine learning is the better way to go. Therefore, tool technology has been compatibility with machine learning algorithm requirements. The Naïve Bayes classification algorithm is the main algorithm used in model development. Weka, Sci-kit, R, and octave are the environment which can be used as the development environment. Among them, R and Octave are the best.

#### **Why use R...?**

R is an open source software which provides full access to algorithms and their implementation. It is speedy and given ability to fix bugs and extend the software. Furthermore, it has an active community for allowing researchers to explore and expand the methods used for analyses data. Moreover, it is quite easy to see what the function is actually doing.

# CHAPTER 4-IMPLEMENTATION

Implementation focus on development of lexicon based sentiment analysis for Human Computer Interaction Perspective from online Purchasing.

## 4.1 Developing Environment

R studio is the development environment and have vast array of packages to support different implementation processes. According to this implementation different packages are required for success the model implementation. For that, first all needed libraries have been installed and after loaded it into the working space.

*Table 4- 1-Libraries Used in Working Environment*

Package	Category	Description	Sample Use	Author
rio	Data import Data Export	rio has a good idea: Pull a lot of separate data-reading packages into one, so you just need to remember 2 functions: import and export. CRAN.	export(sortedvec,"modelresults.csv")	Thomas J.Leeper & others
plyr	Data wrangling	While dplyr is my go-to package for wrangling data frames, the older plyr package still comes in handy when working with other types of R data such as lists. CRAN.	Llply(mylist,myfunction)	Hadley Wickham
stringr	Data wrangling	Numerous functions for text manipulation. Some are similar to existing base R functions but in a more standard format, including working with	str_pad(myzipcodevector, 5, "left", "0")	Hadley wickham

		regular expressions. Some of my favorites: str_pad and str_trim. CRAN.		
ggplot2	Data Visualization	Powerful, flexible and well-thought-out dataviz package following 'grammar of graphics' syntax to create static graphics, but be prepared for a steep learning curve. CRAN.	ggplot(sortedvec, aes(x = attributes, y = score ,fill=sign ...	Hadley Wickham
RColorBrewer	Data visualization	Not a designer? RColorBrewer helps you select color pallettes for your visualizations. CRAN		Erich Neuwirth

## 4.2 Data Set

Data set is a small set which is extracted from Stanford repository and [www.sitejabber.com](http://www.sitejabber.com)

## 4.3 Import data into R

The data which is import into R can come in all sorts for formats, flat files, statistical software files, databases and web URLs.



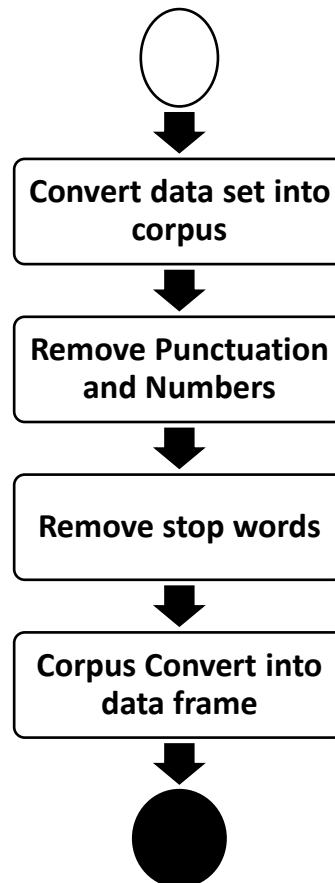
Figure 4- 1-Data Types Which Can Import Into R



Main Consumer review based dataset is text file and it categorized as flat files and it has been import into R for model development and analysis the data.

## 4.4 Pre-Processing

After loading the dataset into working space, it has been preprocess and cleared. To fulfill that, dataset must be converted into the corpus. Corpus is a collection of text documents and it is related to 'tm' package in r.



*Figure 4- 2-Steps of Pre-Processing*

After that, Implementation has been started and it is mainly follow three steps.

## 4.5 Feature Extraction

Feature extracted from unstructured text is the first step for implementing the model which compares the HCI related attributes which embedded in websites. Furthermore, this one is a major challenge being tackled by the field of Natural Language Processing and data mining. Therefore, step by step process has been followed here to overcome the challenge. Cleaned dataset has been followed step by step process to

feature extraction. Train and testing algorithm to identify the features in the dataset. SVM classifier algorithm and developed lexicon based algorithm have been tested.

According to the literature, a number of attributes of websites related to HCI has been systematically reviewed and compile as a pre-defined dictionary.

### 4.5.1 Pseudocode

#### **Input**

*Reading data from HCI related attributes dictionary (Pre-defined from previous literature)*

*Reading data from consumer reviews related to online purchasing website*

#### **Initialize**

*Get output of HCI related to online purchasing related dictionary*

#### **Feature Extraction**

*#split reviews into sentences*

*#spilt sentences into words*

*#compare and get intersection of words and compile HCI attributes dictionary*

*#Get TDM*

*#Get word cloud*

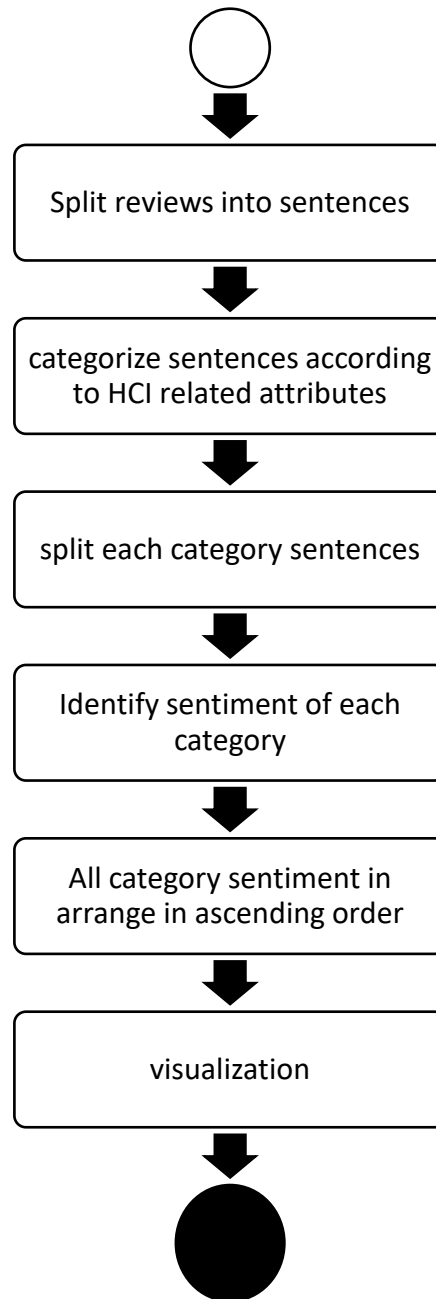
*#visualize*

## 4.6 Model

A model is developed to identify the best and worst attributes related of HCI. After compiling the HCI related attributes dictionary and import it as dataset to the working phase, model development has been started. Furthermore, model has been created from the model development process.

In this section, model development is described in details. In the first step, the dataset has been initialized as data frame and reviews inside the data frame are split into sentences. After that, sentences categories to different categorizes based on lexical dictionaries which contains attributes related to HCI. Then after all sentences in each

category has been split into bag of words. Afterward, bag of words has been further processed for extraction to classify as positive, negative or neutral opinions. Finally, calculate final sentiment for each category and visualize it using plot. Through that decision can be made. overall process of the sentiment analysis can be breakdown as below.



*Figure 4- 3-Steps for Model Development*

## 4.6.1 pseudocode

### Input

*Read dictionary- HCI related attribute dictionary*

*Read dictionary- sentiment data dictionary*

*Read dataset- data set which contains reviews*

### Initialization

*Create different variable to store scores, data*

*Get the full score of sentiment according to each category*

*###pseudocode1*

### Model

*Loop1{*

*#select sentences which contains each HCI factors/categorize sentences according to HCI factors*

*Loop 2{*

*# use this to mining the text and identify each category and its sentiments*

*Function () {*

*#Score for HCI factors and get final score of each category*

*# visualize results*

*}*

*}*

*}*

Furthermore, the pseudocode can be written as in this way to overcome the complexity of above way and reuse the code through functions.so development has been done based on this.

## **#Pseudocode 2**

### **Model**

*Function1 () {*

*# split reviews and calculate scores according to each category.*

*}*

*Function 2 () {*

*For loop2 {*

*#call function 1() to get negative and positive score and overall score for each HCI related attributes*

*}*

*}*

### **Main body**

*For loop1 {*

*# call for function 2 () to categorize reviews sentences according to HCI related attributes*

*}*

# **CHAPTER 05-MODELLING THE CONSUMER BEHAVIOR WITH RESPECT TO HUMAN COMPUTER INTERACTION FACTORS WHICH AFFECT TO WEBSITE**

Lot of Human Computer Interaction Factors are affected to consumers to change their buying behavior and also interact with the websites. Therefore, E-Commerce Websites also affected from this. In this section, describe about how consumers interact with websites to get the best browsing experience with respect to HCI factors. According to the literature, three main HCI factors have. Usability, Accessibility and simplicity are the main HCI factors and Those factors are representing website through different attributes like graphics, photography, product description, product comparisons. In here, simulation agent based model has been developed to only three factors related to HCI which are embedded in E-Commerce site and they are Font Size, Graphics/Photography, and Product Description.

## **5.1 MODEL DESIGN**

A Lot of factors affect to change human mind when accessing the websites. In this section step by step describe the model design. Furthermore, model design has only three attributes related to HCI and online purchasing website. These attributes are extracted from the consumer reviews about online purchasing websites. Quality of Graphics, Colors of website, Font sizes, product description, product comparison are the extracted attributes. According to those attributes, three attributes are used to model and show how consumer behavior change according to those. Furthermore, not only those three factors, but also other environmental factors are considered.

The model has been designed for identifying what factors affect to change consumer behavior and this model is a prototype to solve this problem and identify major factors affect to consumers. This model has been designed through Agent based modelling other than simulation modelling, such as discrete -event simulation or system dynamics. In the model people are identified as agents. Moreover, it has used three

online purchasing websites which are competitive to reading material selling online. Furthermore, the model is considered the number of agents and their situation, that means rush of people and people who read entire product reviews. Changing website different HCI related attributes and identify the human interaction changing behavior. Furthermore, this model has been used to identify most powerful website and it also help for organization to identify what factors are most affected and how those factors are improved to better experience the consumers also.

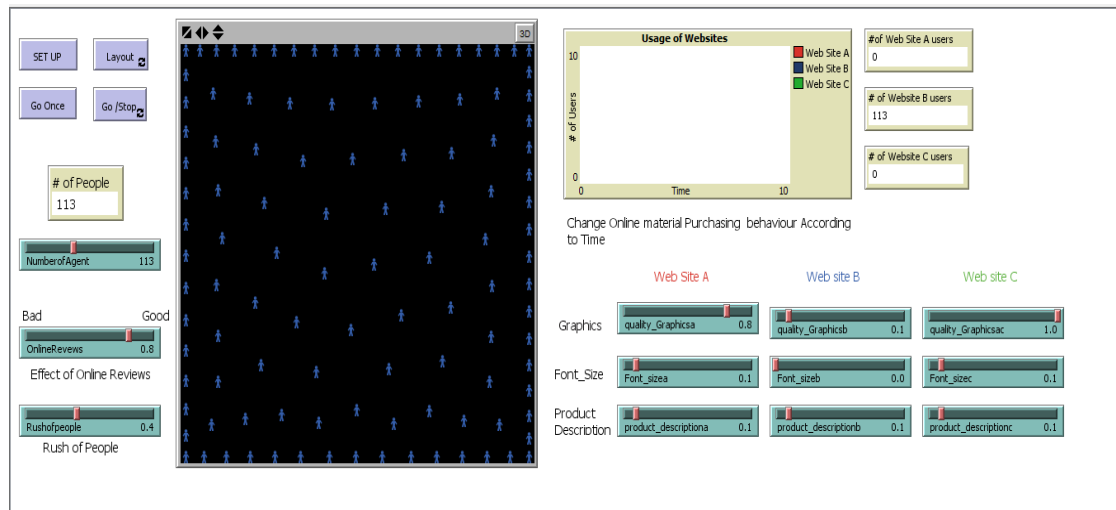


Figure 5- 1-Model Design

## 5.2 MODEL DEVELOPMENT

Development of model has been done through net logo software and it is a simulation development. Through this designed model has become reality. So, designed prototype has been developed little by little with programming. Different interface item has been used to design the interface. In addition to that back end of the model has been programmed and different local and global variables are initialized to store different values.

### NET LOGO

This is a general; purpose software or specially designed software with the toolkit. Furthermore, it is support for multi agent environment.

## 5.3 Results

For that results based on considering below assumptions.

- Preferences to the all three websites are same and behaviour changing randomly when changing the attributes related to HCI

### Consumer behaviour changing when change the quality of graphics in website A

When increasing the quality of graphics of website, A, lot of consumers are preference the website A. Therefore, usage of website A has been increased. Below figure 5-2 showing the results of it

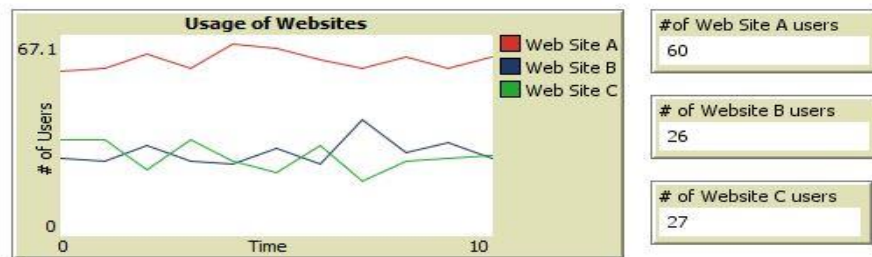


Figure 5- 2-Consumer behaviour changing when change the quality of graphics in website A

### Consumer behaviour changing when change the Font size of website A

When changing the font size of text of website, A, A lot of consumers are preference the website A. Therefore, usage of website A has been increased. Below figure 5-3 showing the results of it





Change Online material Purchasing behaviour According to Time



Figure 5- 3-Consumer behaviour changing when change the Font size of website A

### Consumer behaviour changing when improve the accuracy of product description of website A

When increasing the accuracy of the product description of products of website, A, lot of consumers are preference the website A. Therefore, usage of website A has been increased. Below figure 5-4 showing the results of it



Change Online material Purchasing behaviour According to Time

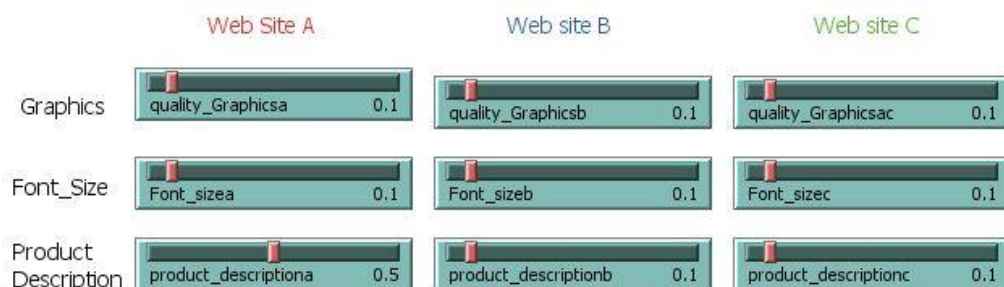


Figure 5- 4-Consumer behaviour changing when improve the accuracy of product description of website A

## **5.4 Conclusion**

Interaction with the website of users are depend upon different HCI related factors which are embedded in online purchasing websites. According to this chapter mainly focus on three websites which can be categorized as e-commerce websites and consumer preferences when changing its different attributes. Furthermore, behaviour of consumers is changing not only according to those factors. But also to their different norms and preferences. Therefore, usability of the website and products on websites is affected to change consumer product choices and access the websites. Moreover, past experiences and environmental factors related to online purchasing also affected to change the consumer's mindset. According to this chapter, it gives a final conclusion as HCI related attributes are influencing to the online purchasing in a different way.

# CHAPTER 06-DISCUSSION AND FINDINGS

## 6.1 Experiment Design Working Environment

All experiments are designed in r studio 3.3.1 version and those experiments done using a computer with windows. Those are the features which have in used computer.

[View basic information about your computer](#)

Windows edition

Windows 10 Education  
© 2016 Microsoft Corporation. All rights reserved.

System

Processor: Intel(R) Core(TM) i3-3217U CPU @ 1.80GHz 1.80 GHz  
Installed memory (RAM): 4.00 GB (3.88 GB usable)  
System type: 64-bit Operating System, x64-based processor  
Pen and Touch: No Pen or Touch Input is available for this Display




Figure 6- 1-Features of Computer

Furthermore, r studio has different features and it is the working environment. Following features have r studio.

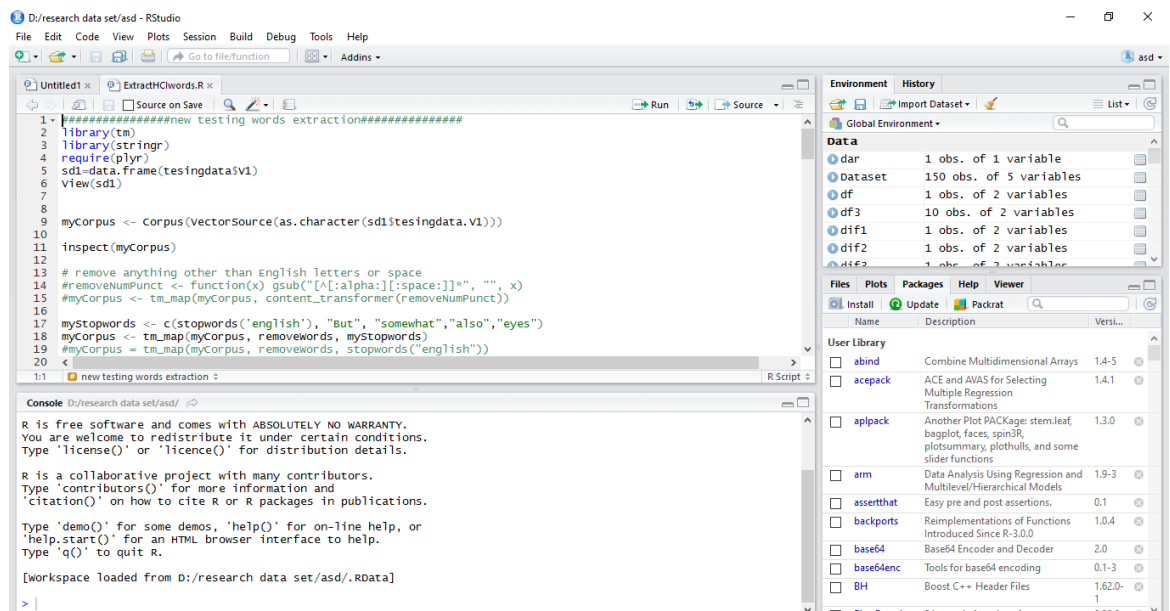


Figure 6- 2-R studio Environment

## 6.2 Compiling HCI attributes Dictionary

Information extraction from unstructured text is one of the challenges being tackled by current research in the field of Natural Language Processing and Data Mining. In the field of information extraction from consumer reviews, most of the work has focused on finding the values for a set of predefined attributes. Recently, there has been growing interest in the automated learning of the attributes themselves, and then finding the associated values.

Through the literature, HCI related attributes dictionary has been created. According to that dictionary, HCI attributes which are related online purchasing has been identified and compile a new dictionary from it. For that, step by step process has been followed. After the completing this step by step process dictionary has been compiled and finalized.

The testing data set has been loaded into the workspace as corpus and removing noisy data. The first step of the corpus converts corpus into TDM and identify high frequency words.

### 6.2.1 Word Cloud

Below figure 6.3 has shown the word cloud for the reviews which is after cleaning the reviews.



Figure 6- 3-Word Cloud

## 6.2.2 High Frequency words in the testing dataset

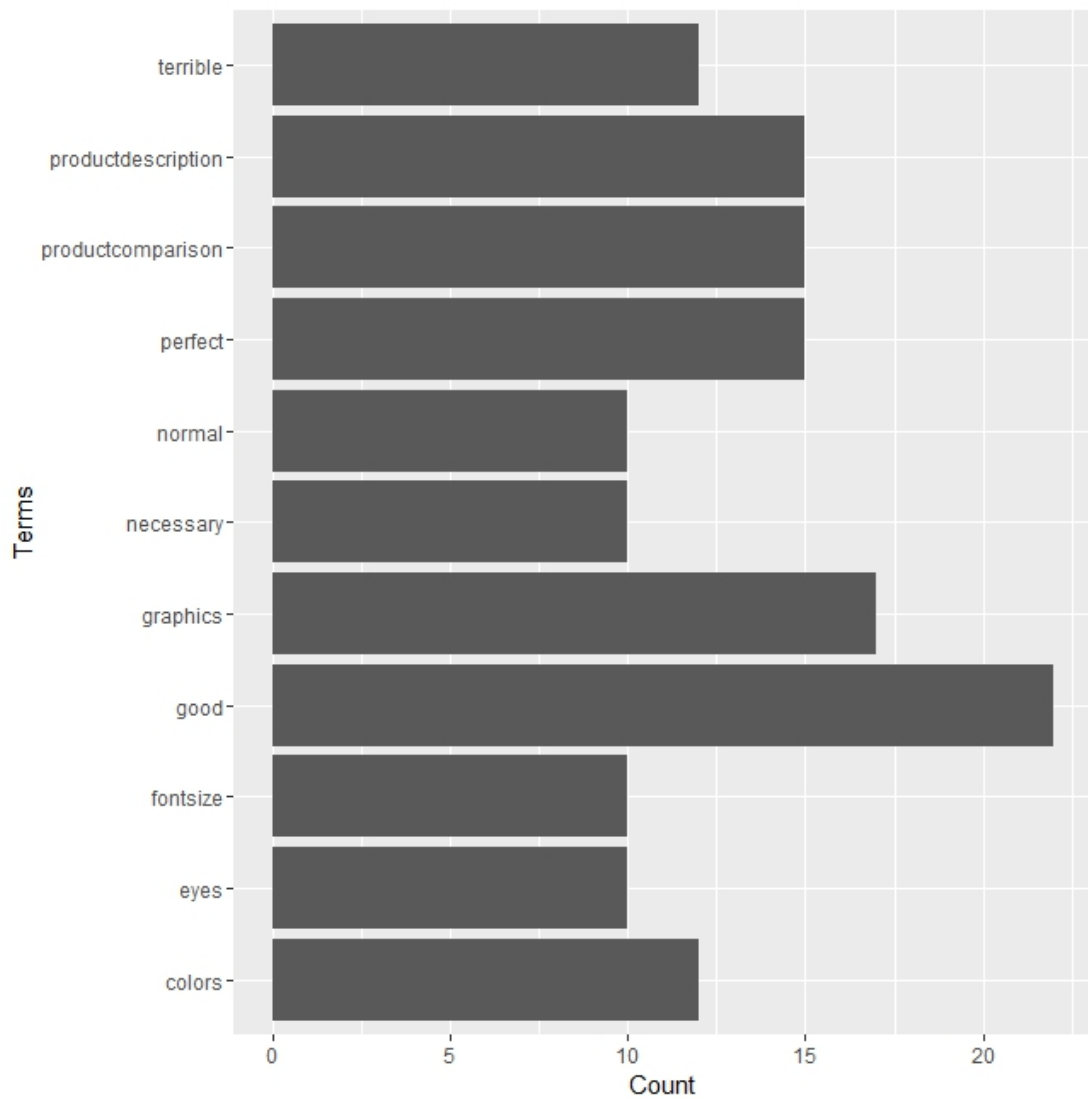


Figure 6- 4-High Frequency words

Figure 6.4 shows the high frequency words of the consumer reviews dataset which are related to online purchasing website. Through that, can be identified the what words are in reviews and can compare with predefined dictionary.

### **6.2.3 Text mining with SVM classifier**

 results

Filter		
	SVM_LABEL	SVM_PROB
1	pagetitles	0.06062950
2	Colors	0.05882353
3	pagetitles	0.06061107
4	Colors	0.05882353
5	pagetitles	0.06062950
6	Colors	0.05882353
7	Colors	0.06063283
8	Colors	0.05882353
9	Colors	0.05882353
10	Colors	0.06063050
11	Colors	0.05882353
12	pagetitles	0.06062950
13	Colors	0.05882353
14	pagetitles	0.06061107
15	Colors	0.05882353
16	pagetitles	0.06062950
17	Colors	0.05882353
18	Colors	0.06063283
19	Colors	0.05882353
20	Colors	0.05882353
21	Colors	0.06063050
22	Colors	0.05882353
23	pagetitles	0.06062950
24	Colors	0.05882353
25	pagetitles	0.06061107
26	Colors	0.05882353
27	pagetitles	0.06062950

Figure 6- 5-Results of SVM classifier

 results

Filter		
	SVM_LABEL	SVM_PROB
27	pagetitles	0.06062950
28	Colors	0.05882353
29	Colors	0.06063283
30	Colors	0.05882353
31	Colors	0.05882353
32	Colors	0.06063050
33	Colors	0.05882353
34	pagetitles	0.06062950
35	Colors	0.05882353
36	pagetitles	0.06061107
37	Colors	0.05882353
38	pagetitles	0.06062950
39	Colors	0.05882353
40	Colors	0.06063283
41	Colors	0.05882353
42	Colors	0.05882353
43	Colors	0.06063050
44	Colors	0.05882353
45	pagetitles	0.06062950
46	Colors	0.05882353
47	pagetitles	0.06061107
48	Colors	0.05882353
49	pagetitles	0.06062950
50	Colors	0.05882353
51	Colors	0.06063283
52	Colors	0.05882353
53	Colors	0.05882353

Figure 6- 6-Results of SVM classifier

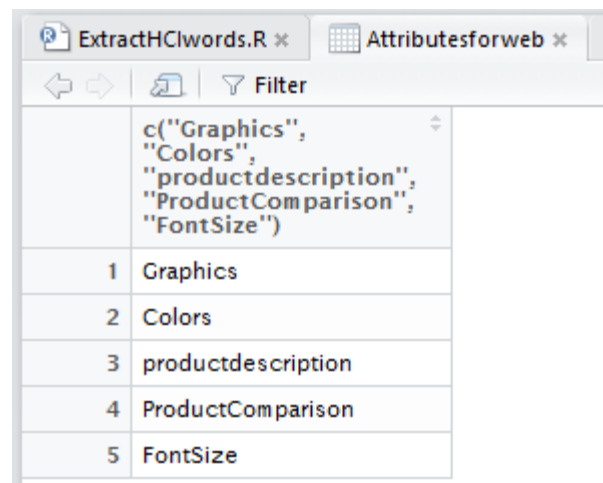
According to the testing data set has been used to identify features and SVM classifier



used for it. After train the SVM classifier has been used to extract the features from testing dataset. But, results show some of the features only extracted through this classifier.

## 6.2.4 Text mining with Lexicon based Developed Algorithm

So, lexicon based algorithm has been used, and compiled attributes related to HCI dictionary.



	<code>c("Graphics", "Colors", "productdescription", "ProductComparison", "FontSize")</code>
1	Graphics
2	Colors
3	productdescription
4	ProductComparison
5	FontSize

Figure 6- 7-Results from lexicon based developed algorithm

Compare bag of words model of testing data with bag of words model of training dataset and this dictionary has been compiled through its intersection.

## 6.2.5 Comparison SVM and lexicon based comparison and Validation

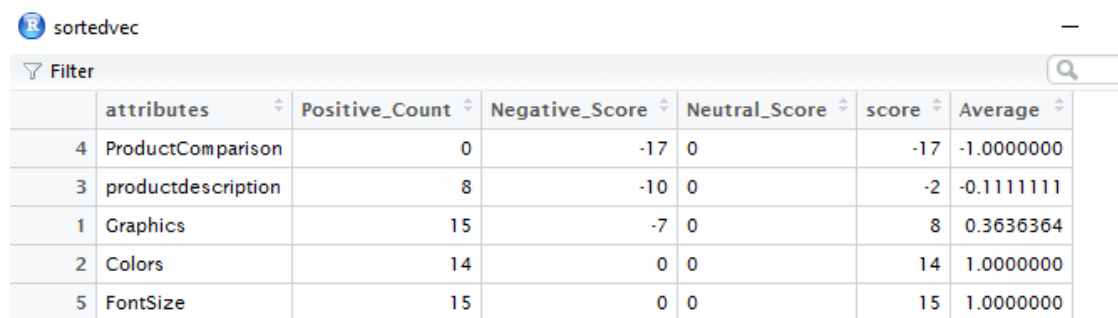
SVM classifier is compiling some attributes, but this new text mining method has identified more attributes. According to the Termdocumentmatrix and high frequency word cloud this compile dictionary has been validated

## 6.3 Analyzing the results of the Model

Developed model has been tested for data set related consumer reviews about online purchasing website. Through the model, polarity scores for each category has been calculated. Furthermore, the model has been calculated negative, positive polarities for each category. Those categories are the bag of sentences which relate to attributes related to HCI which are embedded in online purchasing website.

### 6.3.1 Test Results

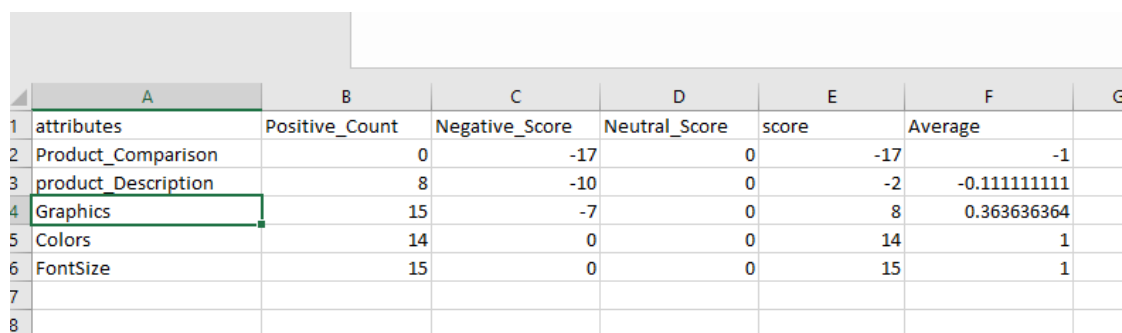
The test result of testing data set is shown below. Figure 6-8 is the result gain from the R studio and Figure 6-9 has been shown the results which is in csv format imported from R studio to excel worksheet.



The screenshot shows the 'sortedvec' window in R Studio. It contains a table with 7 columns: attributes, Positive\_Count, Negative\_Score, Neutral\_Score, score, and Average. The data is sorted by the 'score' column in descending order. The attributes listed are ProductComparison, productdescription, Graphics, Colors, and FontSize.

	attributes	Positive_Count	Negative_Score	Neutral_Score	score	Average
4	ProductComparison	0	-17	0	-17	-1.0000000
3	productdescription	8	-10	0	-2	-0.1111111
1	Graphics	15	-7	0	8	0.3636364
2	Colors	14	0	0	14	1.0000000
5	FontSize	15	0	0	15	1.0000000

Figure 6- 8-Testing data set results in R studio



The screenshot shows an Excel worksheet with the same data as Figure 6-8. The columns are labeled A through G, corresponding to the attributes, Positive\_Count, Negative\_Score, Neutral\_Score, score, and Average. The data is sorted by the 'score' column in descending order. The attribute 'Graphics' is highlighted with a green border.

	A	B	C	D	E	F	G
	attributes	Positive_Count	Negative_Score	Neutral_Score	score	Average	
1	Product_Comparison	0	-17	0	-17	-1	
2	product_Description	8	-10	0	-2	-0.11111111	
3	Graphics	15	-7	0	8	0.36363636	
4	Colors	14	0	0	14	1	
5	FontSize	15	0	0	15	1	
6							
7							
8							

Figure 6- 9-Results After export into excel work sheet as csv file

### 6.3.2 Analysis and comparison

In this section the results from the model has been discussed and presented. Furthermore, Analysis the data and identify how consumer's buying behavior in

buying process and highlight attributes which affect for this buying behavior while shopping online. Furthermore, quality of graphics/photography, colors of the website, font sizes of text, accuracy of product description and comparison of similar products are the main attributes have been mentioned in the reviews and in here identify how consumers behave against those importance factors.

### 6.3.3 Polarity Scores for each Categories

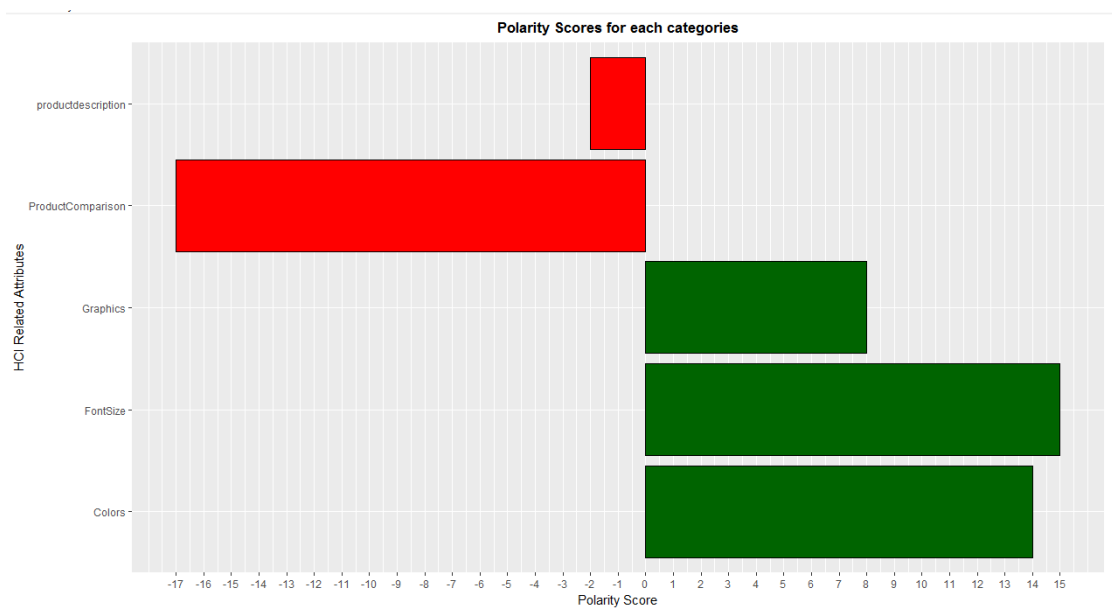


Figure 6- 10-Polarity Scores for each Categories

Figure 6.10 shows the polarity scores for each category and negative polarity categories are shown in red color and positive score category is shown in green color. According to the feature extraction step has been identified five HCI related attributes and those five attributes have mentioned as categories. The x axis is displayed polarity scores and y axis is displayed each categories/attribute.

According to that graph, colors, font size and graphics have positive polarity scores while product comparison and product description have negative polarity. Therefore, this shows the online reading material purchasing website's quality of graphics/photography, font sizes of text and colors of website are good and product comparison and product description of the website are not sufficient to increase the interaction between consumers with the online purchasing website.

According to that results, it has been shown product description and product comparison details want to improve with the accuracy and sufficiently to interact consumer with the website. Furthermore, graphics, colors and font size are no need to change, but improve and continue the consumer interaction with the website need to continue those three attributes in a quality manner.

But this diagram is not clearly shows the consumer opinion. Because this diagram shows the combination of polarity for a collection of different negative and positive opinion sentences. Finally, this diagram has been helped to make prediction about effect of website quality on interaction process of human with website.

### 6.3.4 Negative and positive opinion score comparison

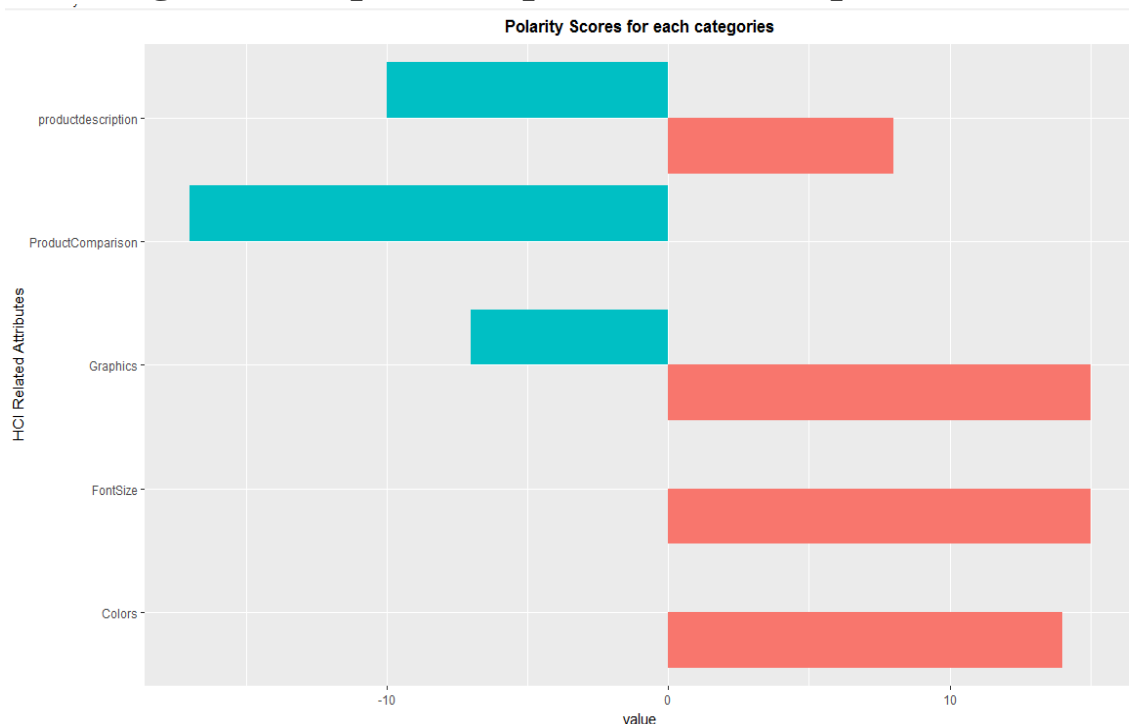


Figure 6- 11-Negative and positive opinion score comparison

This figure 6-11 shows the negative and positive opinion polarity scores for each an attribute which is calculated from number of reviews. In the graph, red color bar is being positive opinions and blue ones are negative opinion polarities.

According to the graph, it shows that colors and font size attributes have no any negative opinions. That means those two attributes have good quality and they help to increase and continue consumer interaction with online reading materials website. Furthermore, product description has not any positive opinions inside the bag of reviews. Therefore, it has been effect to decrease the consumer interaction with the

website. Moreover, it must be improved and give a better product comparison to get better knowledge and idea about products and help to increase consumer interaction level with the website. According to the graphics attributes, it has the higher number of positive opinions rather than number of negative opinions. Therefore, it has been recommended graphics have good polarity, but it must be improved to continue the consumer's interaction with the website. Moreover, product description has positive and negative opinion reviews and scores of those opinions are more close. But there is more positive value rather than negative values. Therefore, small negative value has for this product description attribute. That means according to this website product description not much bad. So, affect for improve this website is not much difficult. That means improvements of product description easy rather than improved the product comparison.

According to this diagram, it gives clear idea about five factors which are embedded in online reading material purchasing website. What want to improve and what factors need to continue its quality to interact consumers with the websites.

### 6.3.5 Satisfaction Scale according to each HCI factors

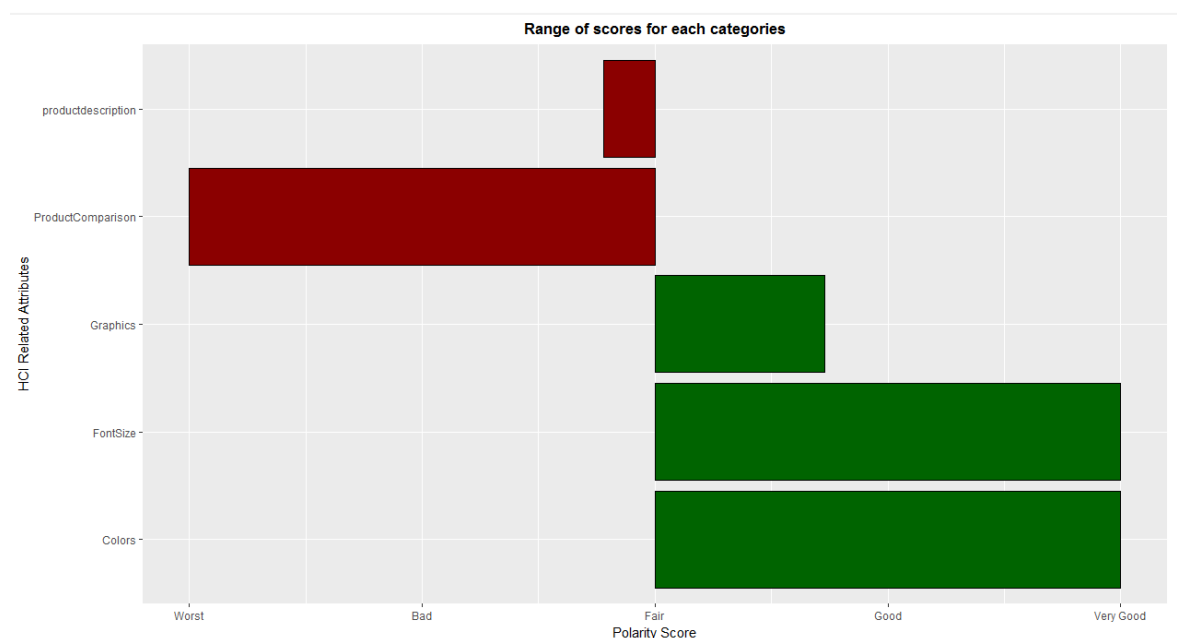


Figure 6- 12-Satisfaction Scale according to each HCI factors

Figure 6-12 show the different rating satisfaction according to HCI related attributes.

Satisfaction from website is core outcome measure for online reading material purchasing website. So, Satisfaction of the consumer is depending on quality of website HCI factors. Through the results from the model, scale of consumer satisfaction about different factors has been recorded.

According to the graph which is shown as figure 6-12 has been identified colors of the website and font size of the website have best quality and those factors which are embedded to online reading material purchasing website have been helped to increase of consumer satisfaction level on website. Furthermore, according to the results product comparison have lowest quality. That means while consumers buying process product comparison give negative mindset for consumers and it not very much help to consumers when buying the reading materials online.

Moreover, graphics level of satisfaction close to good scale type. This results come from the model and through this graph shows, graphics are not much help to satisfy consumers about website when buying reading materials online. In addition to that, product description has satisfaction level closer to level of bad. That means it is not much help to dissatisfy to the website while in buying process.

According to the results of the model has been identified different factors are help to change over the consumer buying behavior from online purchasing reading material website. Among them some factors are help to increase satisfaction level while increasing the revenue of website. Further, some factors are decrease the satisfaction level of consumers about website and those factors are influence to decrease the revenue of the website.

# CHAPTER 06-CONCLUSION AND FUTURE WORKS

In this chapter, has been identify the conclusion of the research, identified limitation and future works.

## 6.1 Conclusion

This study was accomplished to determine the attributes related to HCI which are affect to change consumer behavior in online purchasing the reading materials. with respect to attributes related to HCI. Furthermore, different theories are applied in this area and different theories named as goal oriented online buyer and consumer characteristics, online consumer behavior, factor predicting online shopping and consumer mindset in online shopping.

Furthermore, this study adds new knowledge to the HCI literature and sentiment analysis literature in several ways. Initially feature extraction in sentiment analysis is a challenging area and it has many feature selection methods /techniques. They are NLP based methods, Machine learning or clustering based methods, statistical methods, hybrid methods. According to those methods lexica based dictionary approach is hybrid method which is used in this research to extract the attributes related to websites. This feature extraction has started after preprocess the reviews and end with compiling dictionary which have attributes related to HCI. Moreover, compiled dictionary has attributes related to HCI which are reviews online purchasing websites.

According to the development process of model and develop model has been used to classify the reviews. In this study, a lexicon based approach is used to classify reviews

and identify the HCI attributes which affect to different level of consumers to get good browsing experiences. The dataset has been used to split into sentences after preprocessed it. Furthermore, categorized bag of sentences according to features related to HCI. Sense of each individual category are extracted after split each category into word of bag and compare with sentiment dictionary. Sense of all categories are arranged into one data frame. This sense is the polarity of those category and positive, negative or neutral. This type of sentiment classification performs better than the document level semantic orientation.

Human Computer Interaction area has growing literature and also it is diverse area. HCI consist three main factors called as simplicity, accessibility and usability. When come to the web design, designers and developers must concern those HCI factors while developing different attributes to interact users to the system/website. Therefore, online purchasing related e commerce sites also follow those design styles. Because, today there are large number of organizations are done their business through online. Therefore, competition of selling products through online is higher. So, organization have need to watch and study about their websites always. Questionnaires based studies have done for identify the product and mining the product related data. Furthermore, this type of questionnaires based study are not strength the literature. Therefore, this study has been done based on text mining. For that, text are the different reviews in the website. Among them this study based on reviews related website different features.

There is no any past study about HCI based text mining approach to the online purchasing website. Therefore, this study identified the factors which are affect to increase and decrease the consumer interaction with the online purchasing website. Moreover, this study has been given to the organizations to improve and continue their different attributes of the website. Therefore, this study is build a better approach to the HCI area to the website design and continue it in online purchasing.

Text mining is also growing area and in this study text mining has combined with lexicon based dictionary and extract the Human Computer Interaction attributes which are effect to online purchasing buying behavior of consumers. Moreover, this study has identify the factors which are given good browsing experiences to the user.



Furthermore, this study has open different paths to growing the HCI. Mainly this is open path to further develop and apply model to identify the human factors, like their different visual incompatibilities.

## **6.2 Limitation of the study**

Dataset which is used to identify factors which affect to good browsing experiences, has huge number of sentences which can't be used to this process. Furthermore, difficult to find data to development this model.

## **6.3 Future Works**

Develop predictive analytics dashboard with respective to this study and it is embedded different direction to organization to make decision about their website easy. So, this dashboard has designed to show results of the model in different ways. So, this is the first future direction according to this study

Furthermore, in this study have another future direction and it is being the applying model to identify the consumers. That means this future direction has been considered the different input and output channels such as visual channel, auditory channel, other movements. Furthermore, it must be considered the features related to HCI which embedded to online purchasing material website. After considering all those factors, model can be used to identify human about their different type of input output channel. That means this future work has been covered, all the aspects of human and it direct to study to identify consumer's visionary and other different type of differentiation of channels.

# REFERENCES

- [Analis,2014] Collomb, A., Costea, C., Joyeux, D., Hasan, O. and Brunie, L., A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation,2014.
- [Archak et al.,2007] Archak, N., Ghose, A., and Ipeirotis, P., Show me the money! *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '07*, 2007.
- [Bacon,2003] Bacon, D.R., A comparison of approaches to importance-performance analysis, *International Journal of Market Research*, Vol. 45 No. 1, pp. 55-71,2003.
- [Brave and Nass,2003] Brave, S., and Nass, C., Emotion in human–computer interaction. *Human-Computer Interaction*, pp.53-68.,2003.
- [Cao et al.,2005] Cao,M. Zhang,Q. and Seydel,J., B2C e-commerce web site quality: an empirical examination, *Industrial Management and Data Systems*, vol. 105, no. 5, pp. 645-661, 2005.
- [Chiang et al.,2006] Chiang,W., Zhang,W. and Zhou,L.,Predicting and explaining patronage behavior toward web and

traditional stores using neural networks: a comparative analysis with logistic regression, *Decision Support Systems*, vol. 41, no. 2, pp. 514-531, 2006.

- [Christine et al.,2001] Christine,M.R, Dewit,M. and Aubert,B.,The impact of interface usability on trust in Web retailers, *Internet Research*, vol. 11, no. 5, pp. 388-398, 2001.
- [Dijk et al.,2007] Dijk, G.V., Minocha,S. and Laing,A.,Consumers, channels and communication: Online and offline communication in service consumption, *Interacting with Computers*, vol. 19, no. 1, pp. 7-19, 2007.
- [Gaikwad et al.,2014] Gaikwad,S.V., Chaugule,A. and Patil,P., Text Mining Methods and Techniques,*International Journal of Computer Applications*, vol. 85, no. 17, pp. 42-45, 2014.
- [Hair,2000] Hair J., Knowledge creation in marketing: the role of predictive analytics, *European Business Review*, vol. 19, no. 4, pp. 303-315, 2007.
- [Hansen et al., 2004] Hansen,T., J.Jensen,J.M. and Solgaard,H.S., Predicting online grocery buying intention: a comparison of the theory of reasoned action and the theory of planned behavior, *International Journal of Information Management*, vol. 24, no. 6, pp. 539-550, 2004.
- [Hausman and Hausman,A. and Siekpe,J., The effect of web

- Siekpe,2009] interface features on consumer online purchase intentions, *Journal of Business Research*, vol. 62, no. 1, pp. 5-13, 2009.
- [Helander and Khalid,2000] Helander,M. and Khalid,H., Modeling the customer in electronic commerce, *Applied Ergonomics*, vol. 31, no. 6, pp. 609-619, 2000.
- [Hu and Liu,2004] Hu,M. and Liu,B., Mining and summarizing customer reviews, *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, 2004.
- [Hu et al.,2012] Hu,N., Bose,I., Koh,N. and Liu,L.,Manipulation of online reviews: An analysis of ratings, readability, and sentiments, *Decision Support Systems*, vol. 52, no. 3, pp. 674-684, 2012.
- [Jo and Oh,2011] Jo,Y. and Oh,A., Aspect and sentiment unification model for online review analysis,*Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*, 2011.
- [Josip and Darko,2008] Josip,M. and Darko,P., Prioritizing improvement of service attributes using impact range-performance analysis and impact asymmetry analysis, *Managing Service Quality*,vol.18,no.6,pp.559-576,2008.
- [Kotler,2003] Kotler, P., *Marketing Management*, 11th ed., Prentice-Hall International Editions, Englewood Cliffs, NJ,2003.
- [Koufaris,2002] Koufaris,M.,Applying the Technology Acceptance

- Model and Flow Theory to Online Consumer Behavior, *Information Systems Research*, vol. 13, no. 2, pp. 205-223, 2002.
- [Kuan et al.,2003] Kuan, H.H., Vathanophas V, Bock, G.W., The impact of usability on the intention of planned purchases in e-commerce service websites, in *Proceedings of the 7th Pacific Asia Conference on Information Systems*, Adelaide, South Australia, pp.369-392, Dec 2003.
- [Li and Wu ,2010] Li, N. and Wu,D., Using text mining and sentiment analysis for online forums hotspot detection and forecast, *Decision Support Systems*, vol. 48, no. 2, pp. 354-368, 2010.
- [Lin and He ,2009] Lin,C. and He,Y., Joint sentiment/topic model for sentiment analysis, *Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09*, 2009.
- [Martilla and James,1977] Martilla, J.A. and James, J.C., Importance-performance analysis, *Journal of Marketing*, Vol. 41 No. 1, pp. 77-99,1977.
- [Mostafa,2013] Mostafa,M., More than words: Social networks' text mining for consumer brand sentiments, *Expert Systems with Applications*, vol. 40, no. 10, pp. 4241-4251, 2013.
- [Miles et al.,2000] MILES, G., HOWES, A. and DAVIES, A., A framework for understanding human factors in web-based electronic commerce, *International Journal of Human-Computer Studies*, vol. 52, no. 1, pp. 131-163,

2000.

- [Nah and Davis,2002] Nah, F. F. H., and Davis, S., HCI research issues in e-commerce, *Journal of Electronic Commerce Research*, vol.3, no.3, pp.98-113,2002.
- [Nielsen,2000] Nielsen, J., Why You Only Need to Test with 5 Users, *Nngroup.com*,  
<https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>, 2000.
- [Nielsen,2012] Nielsen, J., Usability 101: Introduction to Usability, *Nngroup.com*,  
<https://www.nngroup.com/articles/usability-101-introduction-to-usability/>, 2012.
- [Oh,2001] Oh, H., Revisiting importance-performance analysis, *Tourism Management*, Vol. 22, pp. 617-27,2001.
- [Siegel,2013] E. Siegel, *Predictive analytics*. Hoboken, New Jersey: John Wiley and Sons, Inc., 2013.
- [Ye et al.,2009] Ye,Q., Zhang,Z. and Law,R.,Sentiment classification of online reviews to travel destinations by supervised machine learning approaches, *Expert Systems with Applications*, vol. 36, no. 3, pp. 6527-6535, 2009.
- [Wan,2000] Wan, H.A., Opportunities to enhance a commercial Web site, *Information and Management*, Vol. 38, pp. 15-21,2000.
- WWW33 <http://www.webopedia.com/TERM/H/HCI.html>,  
accessed on 12/05/2016.

