Altruistic Content Voting System using Crowdsourcing

Miss Menka K. Beldar[1], Prof.M. D. Gayakwad[2], Miss Kavita K. Beldar[3], Prof.M. K. Beldar[4]

[1]*Department of Information Technology,*
Bharati Vidyapeeth (Deemed to be) University College of Engineering, Pune, India
menkabeldar@gmail.com

[2]*Department of Information Technology,*
Bharati Vidyapeeth (Deemed to be) University College of Engineering, Pune, India
Gayakwad.milind@gmail.com

[3]*Department of Information Technology,*
Bharati Vidyapeeth (Deemed to be) University College of Engineering, Pune, India
beldarkavita@gmail.com

[4]*Department of Mechanical Engineering,*
Bharati Vidyapeeth (Deemed to be) University College of Engineering, Pune, India
mkbeldar@bvucoep.edu.in

Abstract--- ***Online reviews provide valuable sources for potential customers to make buying decisions. However, the sheer volume of available reviews as well as the huge variations in the review quality present a big impairment to the effective use of the reviews, as the maximum helpful reviews may be hidden in the large amount of low quality reviews. The goal of this paper is to develop models and algorithms for predicting the content quality of reviews, which provides the basis for discovering the unbiased and biased intension for reviews. We first show that the content quality of a review depends on four important factors: Prosocial Behavior, Community identification, Content Quality and Authenticity of Source. Based on the analysis of those factors, we present a biased and unbiased review intension.***

*Keywords---****Online Reviews, online communities, Content quality, Attitude towards Online Reviews, pro-social behaviour***

I. INTRODUCTION

Currently, a huge amount of reviews is available online. Also presenting a valuable source of information, these informational contents created by users, also called User Generated Contents. powerfully effect the buying decision of customers. In a recent survey revealed that 67.7% of customers are effectively influenced by online reviews when making their buying decisions. More exactly, 54.7% recognized that these reviews were also equally, very or important in their buying decision making. Trusting on online reviews has therefore become a second nature for customers.

Online customer reviews are a form of electronic word of mouth, which is created and distributed by customers who have buying and used products. Online customer reviews are well-defined as a type of published online product information created by users based on personal usage and experience. More precisely, “online customers reviews” have been delineated by other terms such as “electronic word-of-mouth,” “customers-generated information,” “user-generated content,” and “customers feedback.” Recent marketing tendencies, however, have made online customers reviews a distinct class of electronic word-of-mouth,” communications. On the other hand, most of the online customers reviews are created by 12 anonymous individuals, and this anonymity makes it difficult for review writers to be supposed as knowledgeable and trustworthy sources of information. In the case of huge online vendors, product reviewers often display more than a user name. Thus, compared to electronic word of mouth communicated via customers’ social networks such as E-commerce, the source credibility of online customers reviews is difficult to specify based on limited knowledge of reviewers. On the other hand, online customers reviews are considered as more manageable than other forms of customers-generated content in practical terms.

Many websites have combined a voting system that permits users to give a one-click positive or negative vote to the information they read online (i.e., “like/dislike”, or “recommend/not recommend”). The information, collected with their vote counts, and sometimes ranked based on the total voting scores, are apparent to all visitors. The content voting systems divide into two types. The first type is anonymous visitors, which allows to vote information within a site. Typical websites are Youtube.com, Delicious.com, where visitors can without registering or login vote for content by clicking the “thumb up” icon, or vote against content by clicking the “thumb down” icon. The second type is registered members to share information come across anywhere on the Internet to a special site. Websites using this voting system contain Google+, Reddit.com, Facebook.com, and Quara.com.

TABLE 1. USE OF CONTENT VOTING SYSTEM IN VARIOUS FIELDS

	Google Play Store	**Discussion Forums**	**E-commerce**
Objectives	Easily available online to user to download various applications according to their requirements.	Resourceful self-help content, peer-to-peer support is encouraged, sharing, discussing ideas & experiences.	Create social support.
Number of voters	Predicted by number of users who download the applications.	Determined by number of fallowers.	Determined by number of Customers or users.
Motivation voting recipients	Depends on user rating for application, recommend to other user to use that application.	Where you can leave & expect to see response to messages you have left.	Marketing to peers who are interesting in the same product.
Possible mechanism	Install sources for apps including browsing discovery feature and search queries	Creating community & allowing learners to reflect & feedback.	Built trust & loyalty, social bonding.
Example	Android apps and Games, eBooks, TV shows & movies.	Quora, Reddit.	Flipkart, Amazon, Snapdeal.

Remaining sections of this paper is organized as fallows: section II is related work, section III discussed about our proposed research model, section IV gives a conclusion and future work of our research.

Following are the List of abbreviations which are used in this research.

TABLE 2: LIST OF ABBREVIATIONS

Notations	**Description**
PB	Prosocial behavior
CI	Community identification
CQ	Content quality
AS	Authenticity of source
X	Input
W	Weight
Y	Outcome

II. RESEARCH MODEL

Based on an extensive review of the literature, this study proposes the following conceptual model to help understand the factors that influence the perception of quality of information from online reviews. These factors include perceived in formativeness, perceived persuasiveness, source credibility, perceived quantity of reviews, attitude towards online reviews.

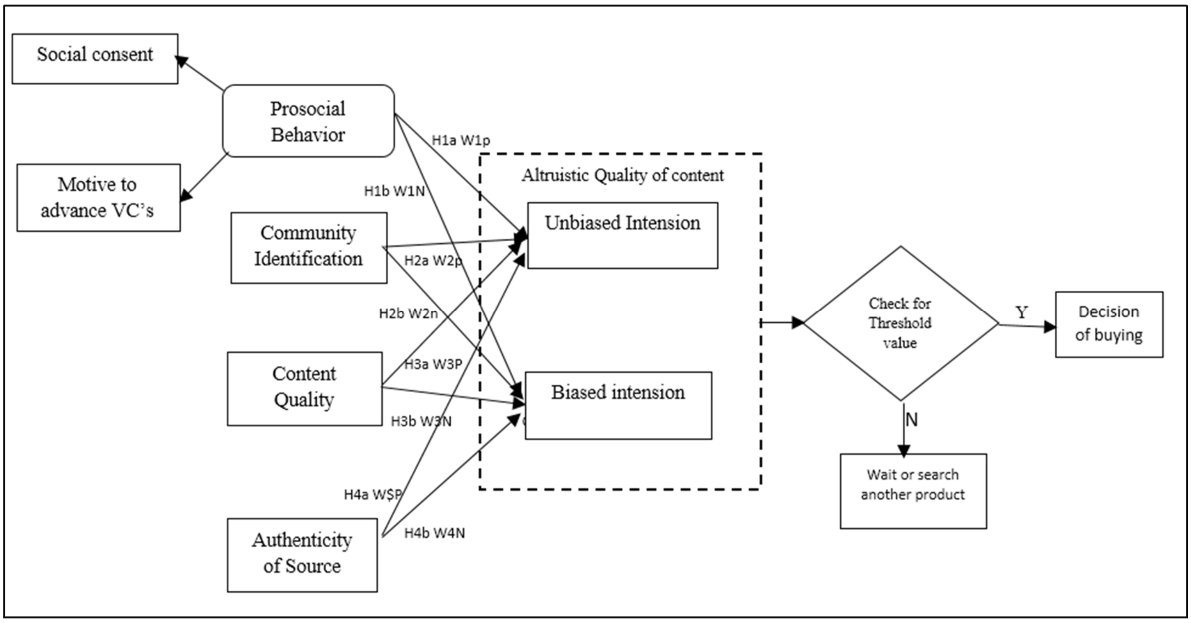

Figure 1. Proposed Architecture of content quality

A. Prosocial behavior

Prosocial behavior is voluntary behavior intended to benefit another, is a social behavior that "benefit other people or society as a whole", such as helping, sharing, donating, co-operating, and volunteering. Prosocial behavior is any action intended to help others. One motivation for prosocial behavior is altruism, or the desire to help others with no expectation of reward. When the motivation for prosocial behavior is to help others without any thought to what you might get in return, it is called altruism. Prosocial behaviors can be performed for a variety of reasons, ranging from selfish and manipulative reasons (e.g., helping get something in return) to moral and other-oriented reasons (e.g., helping because of moral principles or sympathy for another's plight). Therefore, we hypothesize the fallowing.

C1a: Higher prosocial behavior leads to stronger Unbiased Intension.
C1b: Higher prosocial behavior leads to stronger Biased Intension

B. Community identification

An online community or also called an internet community is a virtual community whose members interact with each other primarily via the Internet. For many, online communities may feel like home, consisting of a "family of invisible friends". Those who wish to be a part of an online community usually have to become a member via a specific site and necessarily need an internet connection. An online community can act as an information system where members can post, comment on discussions, give advice or collaborate. Commonly, people communicate through social networking sites, chat rooms, forums, e-mail lists and discussion boards. People may also join online communities through video games, blogs and virtual worlds. Therefore, we hypothesize the fallowing.

C2a: higher community identification leads to stronger Unbiased Intension.
C2b: Higher community identification leads to stronger Biased Intension.

C. Information Quality

Information quality should be an important factor for a website It influences the buying decision process of customers, allowing them to locate and select the merchandise that best satisfies their needs. Based on use and gratification theory, individuals may search for information actively from various alternatives (e.g., catalogues, other shopping websites), and they are aware of their needs and select appropriate media to gratify them, which may be goal-directed behaviour. Individuals would like to buy through a specific shopping website, but they may base their preferences and retrieve active search information from various alternatives (e.g., other shopping websites) to match their desires. Therefore, it can be inferred that the provision of high-quality information on a website is important to creating customers trust. Therefore, we hypothesize the fallowing.

C3a: Higher content quality leads to higher stronger Unbiased Intension.
C3b: Higher content quality leads to higher stronger Biased Intension.

D. Authenticity of Source

Internet online operators to share positive or negative views on a part of information with the rest of the community, they need the ability to judge the information they have read or encountered it is the authenticity of source. Authenticity plays an important role in influencing specific motivation and behavior. KSE refers to the confidence in one's ability to share votes that are valuable to others or the community. It can be manifested in the form of individuals believing that their shared votes can help others in the community (e.g., easily find valuable information, avoid wasting time on trivial information, etc.), or make a difference to the community (e.g., enriching quality information). Knowledge sharing research has shown that knowledge sharing self-efficacy, serving as a self-motivating factor, influences knowledge sharing intention and behaviors in virtual communities or organizations. Therefore, we hypothesize the fallowing.

C4a: Higher authenticity of source leads to stronger Unbiased Intention.
C4b: Higher authenticity of source leads to stronger Biased Intention.

III. METHODOLOGY

4.1 Dataset and preprocess

This paper utilizes the review dataset of Cell phones and Accessories, which was collected from one of the e-commerce websites, Amazon. Before processing the dataset for classification, it needs to be pre-processed, for each of processing in the classification phase. The reviews are processed one at a time. The major focus of the pre-processing phase is on removal of the stop words, i.e. the words which do not contribute in the decision and classification of the reviews, for e.g. "is","an","that", "what", "on" etc.

4.2 Analysis of pre-processed reviews

After the pre-processing phase where the reviews are classified into four categories that are Prosocial behavior, Community identification, Content quality and Authenticity of source. We create a set of word with assigning weight to each word based on the intensity of the word used and assigning the word instance value to the word based on the comparative and superlative degree [24]. For Example, the words instance is Good=1, better=2, best=3, refer table 1.

TABLE 3: ASSUMED VALUES TO THE WORDS

Sr. No	Word	Value	Instance value
1	Worst	-1	2
2	Bad	-0.50	1
3	Good	+0.25	1
4	Better	+0.50	2
5	Best	+0.75	3

The Graphical representation of the word and its weight is represented as below**:**

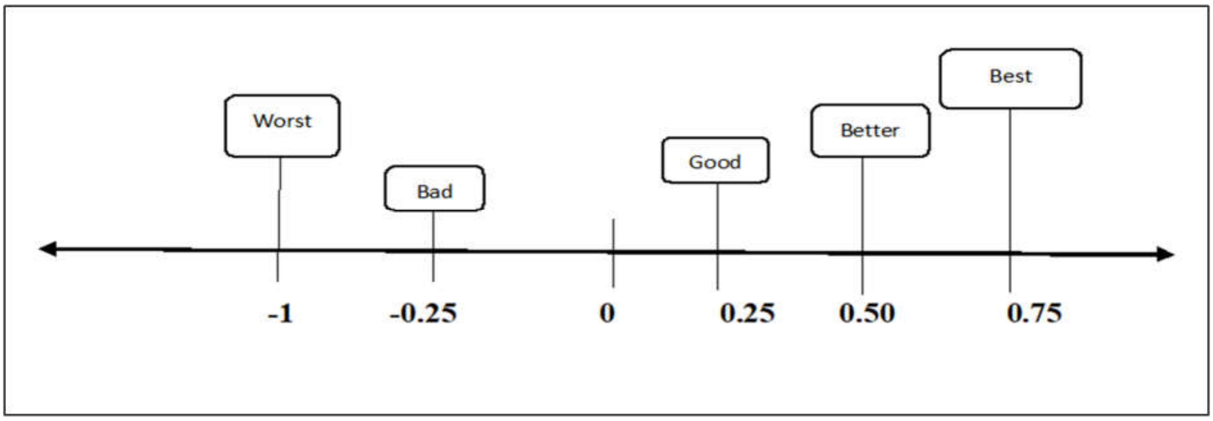

Figure 2. Graphical representation of the word and its weight.

After assigning the values to the words we have assumed the values and thresholds to the parameters that are prosocial behavior, community identification, content quality, authenticity of source which is shown below table.

TABLE 4: ASSUMED VALUES & THRESHOLDS FOR PARAMETERS

Parameters	Value(X)	Threshold
Prosocial behavior	0.46	50%>
Community identification	0.21	25%>
Content quality	0.35	40%>
Authenticity of source	0.29	35%>

4.3 Processing example:

Now, we considered more than 1 reviews and processed it and calculate the unbiased and biased intention behind that reviews.
Review 1: "Best smart phone I ever seen xiaomi is the best among all. Excellent battery life, operating speed".

Review 2: "An average product. Nothing to glorify it, comparable with other brands in the same category, nothing outstanding".

Review 3: "Good product, fast delivery, god packing Redmi note 4 is great mobile. Good lasting battery. Easy to carry".

1. Analysis of pre-processing of reviews

In the pre-processing step we will removal the stop words, i.e. the words which do not contribute in the decision and classification of the reviews, for e.g. "is","an","that", "what", "on" etc.

Review 1: "Best smart phone ever seen xiaomi best among Excellent battery life, Excellent operating speed".

Review 2: "average product. Nothing glorify, comparable other brands same category, nothing outstanding".

Review 3: "Good product, fast delivery, god packing Redmi note 4 great mobile. Good lasting battery. Easy carry".

Based on the assumed values we have calculated the Prosocial behavior, Community identification, Content quality and Authenticity of source in that reviews. The below figure shows that variables x1, x2, x3 are the inputs and w1, w2, w3 are the weights; the sum of the inputs by their weights triggers a certain activation function, which gets compared to a firing threshold which defines an output y. It is represented as,

$Y=\sum_{i=1} wi * xi$

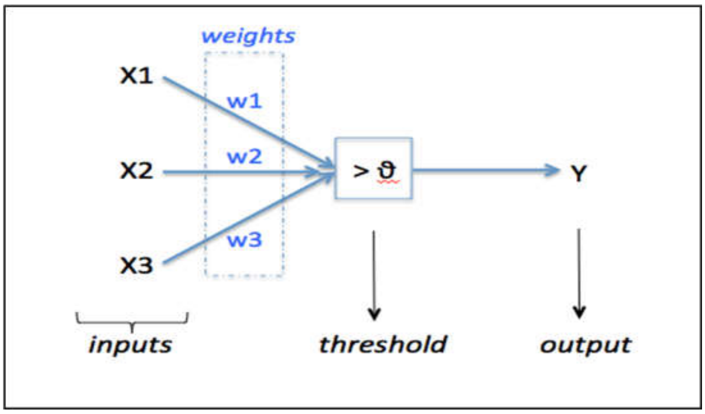

Figure 3 Relationship of input, weight, threshold and output

Using this formula, we have calculated the Prosocial behavior as shown in below table:

TABLE 5 PROSOCIAL BEHAVIOR

Words	Weight(W)	Instances(I1)	W1=W*I1	Y1=W1*X1
Best	0.75	3	2.25	1.035
Best	0.75	3	2.25	1.035
good	0.25	1	0.25	0.115
Nothing	-0.50	1	--0.50	-0.23
Outstanding	0.75	2	1.5	0.69
Good	0.25	1	0.25	0.115
Fast	0.50	1	0.50	0.23
Good	0.25	1	0.25	0.115
great	0.25	2	0.50	0.23

Similarly, we have calculated the community identification as shown in below table:

TABLE 6: COMMUNITY IDENTIFICATION

Words	Weight(W)	Instances(I2)	W1=W*I2	Y2=W1*X2
Best	0.75	3	2.25	0.472
Good	0.25	1	0.25	0.052
Good	0.25	1	0.25	0.052
Good	0.25	1	0.25	0.052

Similarly, for Content quality

TABLE 7: CONTENT QUALITY

Words	Weight(W)	Instances(I3)	W1=W*I3	Y3=W1*X3
Best	0.75	3	2.25	0.787
Best	0.75	3	2.25	0.787
good	0.25	1	0.25	0.087
good	0.25	1	0.25	0.087
Average	0.45	1	0.45	0.157
nothing	-0.50	1	-0.45	-0.157
outstanding	0.50	1	0.50	0.175
Good	0.25	1	0.25	0.087
Good	0.25	1	0.25	0.087
Great	0.75	2	1.5	0.525
Easy	0.50	1	0.50	0.175

Similarly, we have calculated the Authenticity of source in that reviews,

TABLE 8: AUTHENTICITY OF SOURCE

Words	Weight(W)	Instances(I4)	W1=W*I4	Y4=W1*X4
Best	0.75	3	2.25	0.652
Good	0.25	1	0.25	0.072
Good	0.25	1	0.25	0.072
Average	0.45	1	0.45	0.130
Fast	0.50	1	0.50	0.145
Easy	0.50	1	0.50	0.145

$$y=\sum_{i=1} wi * xi$$

here, y= (W1*X1) + (W2*X2) + (W3*X3) + (W4*X4)

y=3.335+0.628+2.797+1.216=7.976

Now, we will calculate the prosocial behavior in percentage and compare it to our given threshold. By using the fallowing formula as,

$$Y=\frac{\sum_{i=1} wixi}{max(\,wixi)*no.of\ instances(I)}$$

Here, y=7.976, max(wi*xi) =1.035, I1=15

$$Y=\frac{7.976}{1.035*15}=0.51 \approx 51\%$$

Here, y=7.976, max(wi*xi) =1.035, I2=6

$$Y=\frac{7.976}{1.035*6}=1.284 \approx 128\%$$

Here, y=7.976, max(wi*xi) =1.035, I3=16

$$Y=\frac{7.976}{1.035*16}=0.48 \approx 48\%$$

Here, y=7.976, max(wi*xi) =1.035, I1=8

$$Y=\frac{7.976}{1.035*8}=0.963 \approx 96\%$$

TABLE 9: PERCENTAGE RESULT OF FOUR PARAMETERS

Parameters	Wi*Xi	$y=\sum_{i=1} wi * xi$	$Y=\frac{\sum_{i=1} wixi}{max(\,wixi)*no.of\ instances}$
Prosocial behavior	**3.335**	**7.976**	**51.37%**
Community identification	**0.628**		**128%**
Content quality	**2.797**		**48.16%**
Authenticity of source	**1.216**		**96.32%**

Graphical representation of results

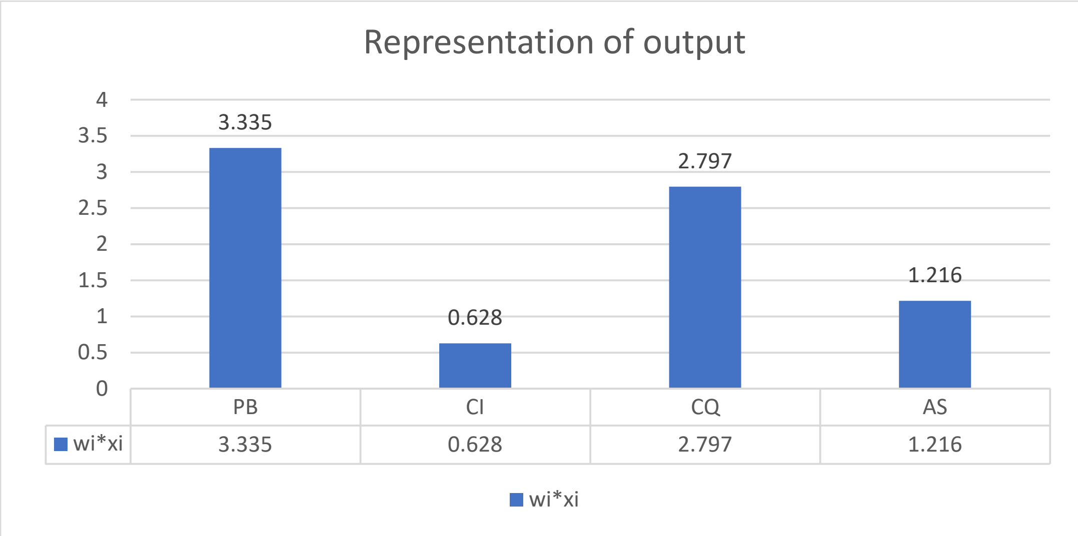

Figure 4. Representation of weight and parameters

Representation of content quality in percentage with four categories:

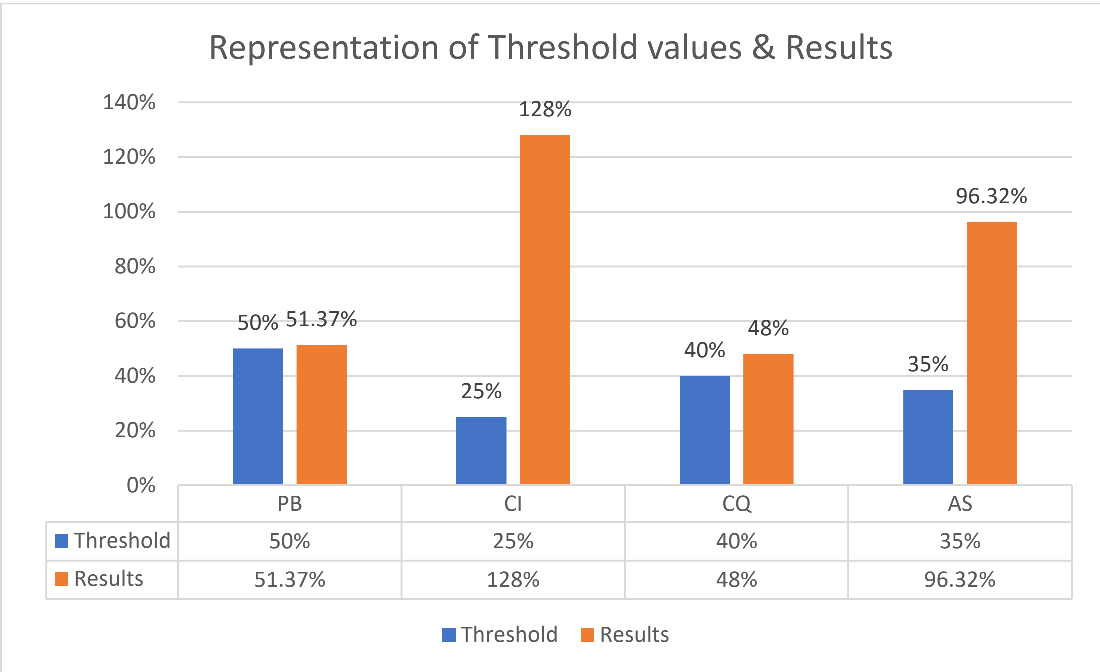

Figure5.Representation of threshold and results

4.4 Empirical Model

The following model was used to test the hypothesis

UI=H1a W1P + H2aW2P+ H3aW3P +H4aW4P

BI=H1b W1N+ H2bW2N+ H3bW3N +H4bW4N

Where,
UI = Unbiased Intention
BI = Biased Intention
W = Weight

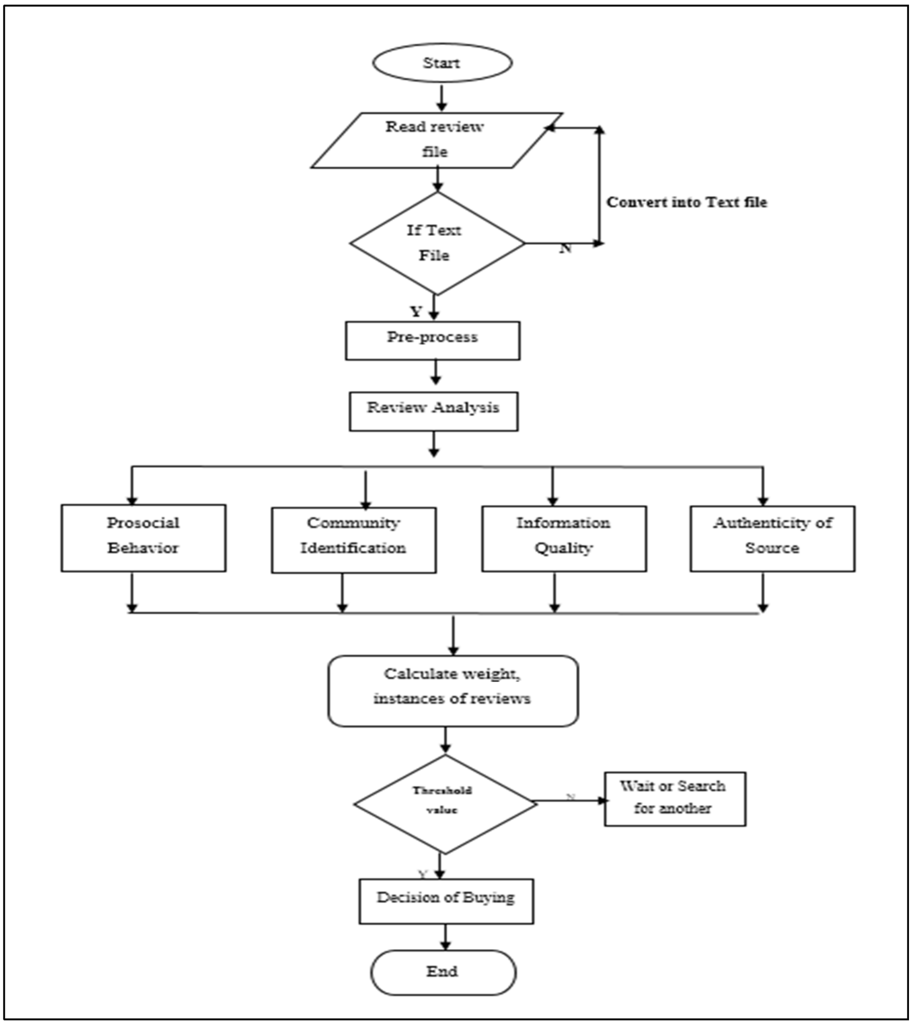

Figure 6. Flowchart of system

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we have considered the important problem of predicting the quality of reviews. We provided a detailed analysis of the online product reviews. Our study in this paper has focused on the Cell phone and Accessories reviews, but our approach is general enough to be easily adapted to other domains as well. This Research presents that classify and analysis of a review into four categories and calculate the Unbiased and biased intention. The main aim of this research is to Calculate the information quality which is helpful to customers, they make decision whether he/she should buy or not buy the products. This can helpful to people to buying valuable product and spend their money on quality products. For future work, we plan to study the related ranking problem, i.e., how do we rank the reviews based on the quality of review? Also, we worked on e-commerce product reviews in future it will be worked on other type of information e, g cartoons, videos etc.

References

[1]. M. Hsu, T. Jub, C. Yen, and C. Chang, "Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations," Int. J. Hum. -Comput. Stud., vol. 65, pp. 153–169, 2007.

[2]. H. Lin, "Knowledge sharing and firm innovation capability: An empirical study," Int. J. Manpower, vol. 28, pp. 315–332, 2007.

[3]. *K. Elwalda, Abdulaziz and Lu, "The Influence of Online Customer Reviews on Buying Intention: the Role of Non-numerical Factors," 2013.*

[4]. *F. Zhu, "Impact of Online Customers Reviews on Sales□: The Moderating Role of Product and Customers Marshall School of Business HKUST Business School and Impact of Online Customers Reviews on Sales□: The Moderating Role of Product and Customers Characteristics," 2009.*

[5]. *S. M. Mudambi and D. Schuff, "What Makes a Helpful Online Review? a Study of Customer Reviews on Amazon.Com 1," vol. 34, no. 1, pp. 185–200, 2010.*

[6]. *R. Keller, Kevin Lane and Staelin, "Effects of Quality and Quantity on Decision Effectiveness," J. Consum. Res., vol. Vol. 14, N, no. September 1987.*

[7]. *Bandura, "Self-efficacy: Tow*ard a unifying theory of behavioral change," *Psychol. Rev.*, vol. 84, pp. 191–215, 1977.