Customer Attitude Towards E-Commerce using Naive Bayes Algorithm : Machine Learning Approach

B. Naveena Devi[1], Dr. Y. Rama Devi[2], Dr. R. Rajeswara Rao[3]

[1]*Research Scholar, Dept. of CSE, JNTU Hyderabad, India*
[2] *Professor, CBIT, Hyderabad, India.*
[3] *Professor & Head, Dept. of CSE, JNTUKCE, Vizianagaram, A.P*

Abstract : *In the era of the e-commerce transmission and its services provided to the customer over the Internet, the Internet is commonly used by both customers and business to buy and sell their goods and services across the world. This study focuses on the factors influencing customer decision and attitudes towards adopting online shopping in India. This study is mainly emphasizes on how demographic variable (age, income and occupation) & Psychological factors (Motivation, Perception, Learning, Belief and attitudes) affect customer buying behavior towards E-commerce. The research questionnaire and hypotheses were developed on review of the literature. Based on the research objectives, a structured questionnaire with 30variables, mainly with a 5-point Likert scale was used. The survey was carried out on selected 1500 customers using E-Commerce applications in India. For data collection, random sampling was adopted. Data mining Tools and Techniques i.e. WEKA, R programming, R Rattle and R Cmdr. have been used to analyze the data. Multinomial Logistics model is applied for predicting customer buying behaviours. Data mining techniques like Principle Component Analysis and factor analysis is applied for Data Dimension Reduction. Originally there were 30 factors influencing attributes/variables which were reduced to 7 after the PCA technique is applied. This research result represent that customer region is the variable that is highly influencing the customer buying behavior and the E- Commerce. The second reason is the customer income which is also influencing the customer buying behavior and E- Commerce next to the customer Region to some extent. The rest of the demographic variables are not showing significant influence.*

Keywords - *Ecommerce, Customer Perception, Customer attitude, Decision Tree, Ordinal /Multinomial Logistic, Naive Bayes Algorithm, Confusion Matrices, sensitivity, specificity.*

I. INTRODUCTION

The terms “Electronic Commerce”, “Internet marketing” and “on-line shopping” are now commonly used by business executives and consumers throughout the world as businesses are recognizing the potential opportunities for commerce in the on-line business environment (Karakaya and Charlton, 2001). At its core, e-commerce refers to the purchase and sale of goods and/or services via electronic channels such as the Internet. E-commerce was first introduced in the 1960s via an electronic data interchange (EDI) on value-added networks (VANs). The medium grew with the increased availability of Internet access and the advent of popular online sellers in the 1990s and early 2000s. Amazon began operating as a book-shipping business in Jeff Bezos' garage in 1995. EBay, which enables consumers to sell to each other online, introduced online auctions in 1995 and exploded with the 1997 Beanie Babies frenzy. Rising incomes and a greater variety of goods and services that can be bought over the internet is making buying online more attractive and convenient for consumers all over the country.

II. LITERATURE REVIEW

E-commerce is a tool for reducing administrative costs and cycle time, streaming business processes, and improving relationships with both business partners and customers (Charles, 1998). The Internet and electronic commerce were the two most significant development of information during 1990s. There has been a marked increase in the number of consumers who purchase over the Internet, as well as an increase in sales worldwide conducted via electronic commerce. Innovation and electronic commerce relationships have resulted in tremendous changes in market competition among various industries (Blosch, 2000; Hamid & Kassim, 2004). Customer purchasing decisions are influenced by perception, motivation, learning, attitudes and beliefs. The perception is reflected to on how the customers select, organize, and interpret information to form knowledge. The motivation is reflected to the customer’s desire to meet their own needs. Learning is reflected to the customers’ behavior experience arising. Attitudes are reflected to customers’ steadily favourable or

unfavourable assessments, feelings, and inclinations towards object or idea. Finally, Beliefs is reflected to customers' thoughts about a product or service (Kotler & Armstrong, 1997).

III. OBJECTIVES

- This study is to identify the attitude of customer towards E Commerce Business.
- To investigate how demographic variable (age, income and occupation) affect customer buying behavior and E Commerce.
- To examine the factors those affect the customer buying behavior (perception, motivation, learning, attitudes and beliefs) and E Commerce.

IV. FACTORS INFLUENCING CONSUMER BUYING BEHAVIOUR TOWARDS E COMMERCE

This study is mainly focus on how demographic variable (age, income and occupation) & Psychological factors (Motivation, Perception, Learning & Beliefs and attitudes) affect customer buying behavior towards E Commerce. There are 4 main types of factors influencing consumer behavior: cultural factors, social factors, personal factors and psychological factors

1. Cultural factors: Cultural factors are coming from the different components related to culture or cultural environment from which the consumer belongs. Culture and societal environment: Culture is crucial when it comes to understanding the needs and behaviour of an individual. Throughout his existence, an individual will be influenced by his family, his friends, his cultural environment or society that will "teach" him values, preferences as well as common behaviours to their own culture.
2. Social factors: Social factors are among the factors influencing consumer behavior significantly. They fall into three categories: reference groups, family and social roles and status.
3. Personal factors: Decisions and buying behavior are obviously also influenced by the characteristics of each consumer.

a. Age and way of life: A consumer does not buy the same products or services at 20 or 70 years. His lifestyle, values, environment, activities, hobbies and consumer habits evolve throughout his life. For example, during his life, a consumer could change his diet from unhealthy products (fast food, ready meals, etc.) to a healthier diet, during mid-life with family before needing to follow a little later a low cholesterol diet to avoid health problems. The factors influencing the buying decision process may also change. The family life cycle of the individual will also have an influence on his values, lifestyles and buying behavior depending whether he's single, in a relationship, in a relationship with kids, etc.. As well as the region of the country and the kind of city where he lives (large city, small town, countryside, etc).

b. Purchasing power and revenue: The purchasing power of an individual will have, of course, a decisive influence on his behavior and purchasing decisions based on his income and his capital. This obviously affects what he can afford, his perspective on money and the level of importance of price in his purchasing decisions. But it also plays a role in the kind of retailers where he goes or the kind of brands he buys. As for social status, some consumers may also look for the "social value" of products they buy in order to show "external indications" of their incomes and their level of purchasing power..

c. Lifestyle: The lifestyle of an individual includes all of its activities, interests, values and opinions. The lifestyle of a consumer will influence on his behavior and purchasing decisions. For example, a consumer with a healthy and balanced lifestyle will prefer to eat organic products and go to specific grocery stores, will do some jogging regularly (and therefore will buy shoes, clothes and specific products), etc.

d. Personality and self-concept: Personality is the set of traits and specific characteristics of each individual. It is the product of the interaction of psychological and physiological characteristics of the individual and results in constant behaviours. It materializes into some traits such as confidence, sociability, autonomy, charisma, ambition, openness to others, shyness, curiosity, adaptability, etc. While the self-concept is the image that the individual has – or would like to have – of him and he conveys to his entourage. These two concepts greatly influence the individual in his choices and his way of being in everyday life. And therefore also his shopping behavior and purchasing habits as consumer. In order to attract more customers, many brands are trying to develop an image and a personality that conveys the traits and values - real or desired – of consumers they are targeting.

4. Psychological factors: Among the factors influencing consumer behavior, psychological factors can be divided into 4 categories: motivation, perception, learning as well as beliefs and attitudes.

a. Motivation: Motivation is what will drive consumers to develop a purchasing behavior. It is the expression of a need is which became pressing enough to lead the consumer to want to satisfy it. It is usually working at a subconscious level and is often difficult to measure. Motivation is directly related to

the need and is expressed in the same type of classification as defined in the stages of the consumer buying decision process.

b. Perception: Perception is the process through which an individual selects, organizes and interprets the information he receives in order to do something that makes sense. The perception of a situation at a given time may decide if and how the person will act. Depending to his experiences, beliefs and personal characteristics, an individual will have a different perception from another. The perception mechanism of an individual is organized around three processes:

i. Selective Attention: The individual focuses only on a few details or stimulus to which he is subjected. The type of information or stimuli to which an individual is more sensitive depends on the person. Consumers will also be much more attentive to stimuli related to a need. For example, a consumer who wishes to buy a new car will pay more attention to car manufacturers' ads, while neglecting those for computers. Lastly, people are more likely to be attentive to stimuli that are new or out of the ordinary. For example, an innovative advertising or a marketing message (Unique Value Proposition) widely different from its Competitors is more likely to be remembered by consumers.

ii. Selective Distortion: In many situations, two people are not going to interpret information or a stimulus in the same way. Each individual will have a different perception based on his experience, state of mind, beliefs and attitudes. Selective distortion leads people to interpret situations in order to make them consistent with their beliefs and values. Selective distortion often benefits to strong and popular brands. Studies have shown that the perception and brand image plays a key role in the way consumers perceived and judged the product.

iii. Selective Retention: People do not retain all the information and stimuli they have been exposed to. Selective retention means what the individual will store and retain from a given situation or a particular stimulus. As for selective distortion, individuals tend to memorize information that will fit with their existing beliefs and perceptions. Selective retention is also what explains why brands and advertisers use so much repetition in their advertising campaigns and why they are so broadcasted. So that the selective retention can help the brand to become a "top of mind" brand in the consumer's mind.

c. Learning: Learning is through action. When we act, we learn. It implies a change in the behavior resulting from the experience. The learning changes the behavior of an individual as he acquires information and experience. Rather, if you had a good experience with the product, you will have much more desire to buy it again next time.

d. Beliefs and attitudes: A belief is a conviction that an individual has on something. Through the experience he acquires, his learning and his external influences (family, friends, etc.), he will develop beliefs that will influence his buying behavior. While an attitude can be defined as a feeling, an assessment of an object or idea and the pre-disposition to act in a certain way toward that object. Attitudes allow the individual to develop a coherent behavior against a class of similar objects or ideas. Beliefs as well as attitudes are generally well-anchored in the individual's mind and are difficult to change.

However, it is important to understand, identify and analyse the positive attitudes and beliefs but also the negative ones that consumers can have on a brand or product. To change the brand's marketing message or adjust it's positioning in order to get consumers to change their brand perception.

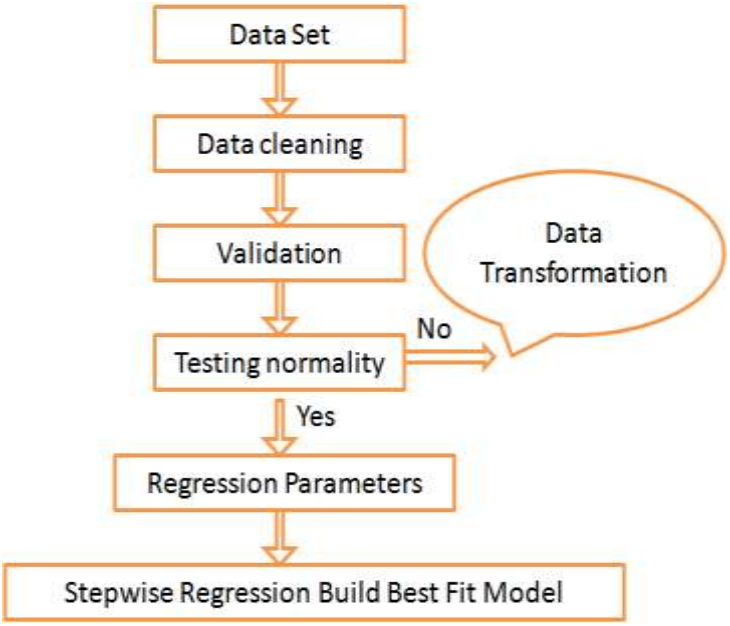
V. METHOD OF STUDY

To measure these objectives a cross sectional descriptive study was designed.

Questionnaire Design: Based on the research objectives, a structured questionnaire with 30variables, mainly with a 5-point Likert scale was used, in which 1 = strongly disagree and 5 = strongly agree.

Data Collection: In this research, the survey method was selected to understand the factor affecting customer attitude towards E-commerce applications in India. For data collection, random sampling was adopted. To ensure all questions being answered in a proper way, questionnaires were completed and screened one-by-one.

Data Accessing and Cleaning: The data collected through survey is saved in csv format. This data is converted into data table format. Data cleaning and validation process is carried out using R Programming language and Weka tool.



Data Flow Chart

VI. Data Analysis

6.1Analysis Tools & Techniques

Data mining Tools and Techniques i.e. WEKA, R programming, R Rattle and R Rcmdr have been used to analyse the data. Multinomial Logistics model is applied for predicting customer buying behaviours. Data mining techniques like Principle Component Analysis and factor analysis is applied for Data Dimension Reduction. Originally there were 30 factors influencing attributes/variables which were reduced to 7 after the PCA technique is applied. And the analysis is carried here afterwards with the reduced 7 variables.

6.2 Data Access & Cleaning

A survey is conducted on 1500 customers. After a pilot study we consider 727 to be a good sample with a sample error of 5%. The data is cleaned. Imputation, Data validation and data reliability, process is performed using R Programming language and Weka Programming Language.

6.3. Data Validation and Quality Check:

Data is validated by performing the reliability test for the data, using Cronbach's Alpha. KMO and Bartlett's tests are performed for the sample adequacy. A good result of 87% is observed indicating the reliability percentage of the data. Also the KMO test for the factor data validation is obtained to be 96%. Tables for the data reliability and validity tests are given below.

VII. Result

7.1: Data Reliability:

The Cronbach's Alpha which is a reliability statistic is obtained. This is found to be 87% which means the data is reliable. The table for this is shown below.

Reliability Statistics

Cronbach's Alpha	N of Items
.873	44

7.2: Sample Validity:

A sample of 727 is adequate for the study is confirmed by the KMO statistic .The table for which is shown below. A good result of 96% is showing the sample validity.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.964
Bartlett's Test of Sphericity	Approx. Chi-Square	62659.454
	df	435
	Sig.	.000

7.3 Naive Bayes Algorithm Model:

Dimension reduction technique like factor analysis is applied to reduce the dimensions of the data without loss of the data. 30 factor variables are reduced to 7. Irrelevant variables are removed and then Logistics Multinomial model is applied .Models are obtained for the 7 demographic variables that are affecting the e commerce. A detailed study is done on each model and conclusions are made about the factors that are influencing consumer buying behavior. All the models are evaluated by the confusion matrix and the respective model diagnostics for each variable which are shown below.

7.3.1 Marital Status:

	Test Data	Train Data
Accuracy	0.6781	0.6437
95% CI	(0.5958, 0.7529)	(0.6033, 0.6827)
No Information Rate	0.589	0.5869
P-Value [Acc> NIR]	0.01681	0.002915
Kappa	0.3126	0.2384
Mcnemar's Test P-Value	0.08005	9.93E-05

	Test Data	**Train Data**
Sensitivity	0.8023	0.7801
Specificity	0.5	0.45
PosPred Value	0.697	0.6683
NegPred Value	0.6383	0.5902
Prevalence	0.589	0.5869

For the Marital status group, we can see from the above classification table that, the metrics for the test were Sensitivity (80%), Specificity (50%), Positive predictive value (70%), Negative Predictive value (64%), and Balanced Accuracy (65%). The Accuracy of Test data is 68%. The metrics for the training set were Sensitivity (78%), Specificity (45%), Positive predictive value (67%), Negative Predictive value (59%), and Balanced Accuracy (62%). The Overall Accuracy of Training data is 64%.

7.3.2 Customer Gender:

	Gender Group	
	Test Data	**Train Data**
Accuracy	0.6849	0.6162
95% CI	(0.6029, 0.7592)	(0.5753, 0.6559)
No Information Rate	0.5137	0.5198
P-Value [Acc> NIR]	2.01E-05	1.80E-06
Kappa	0.376	0.22
Mcnemar's Test P-Value	6.87E-05	0.00

	Test Data	**Train Data**
Sensitivity	0.5067	0.4122
Specificity	0.8732	0.8046
PosPred Value	0.8085	0.6609
NegPred	0.6263	0.5971
Prevalence	0.5137	0.4802

For the Gender group, from the above table it can be drawn that, the metrics for the test set were Sensitivity (51%), Specificity (87%), Positive predictive value (81%), Negative Predictive value (63%), and Balanced Accuracy (69%). The Overall Accuracy of Test data is 68%. The metrics for the training set were Sensitivity (41%), Specificity (80%), Positive predictive value (66%), Negative Predictive value (60%), and Balanced Accuracy (61%). The Overall Accuracy of Training data is 62%.

7.3.3 Customer Region:

	Test Data				Training Data			
	Metro	**Town**	**Urban**	**Village**	**Metro**	**Town**	**Urban**	**Village**
Sensitivity	0.9268	0.51852	0.8333	0.9583	0.9277	0.623	0.892	0.9138
Specificity	0.9524	0.98319	1	0.8443	0.9687	0.9847	0.9654	0.8839
PosPred Value	0.8837	0.875	1	0.5476	0.9222	0.9157	0.9181	0.6625
NegPred Value	0.9709	0.9	0.9109	0.9904	0.971	0.9076	0.9537	0.9762
Prevalence	0.2808	0.18493	0.3699	0.1644	0.2857	0.21	0.3029	0.1997
Detection Rate	0.2603	0.09589	0.3082	0.1575	0.2651	0.1308	0.2702	0.1824
Detection	0.2945	0.10959	0.3082	0.2877	0.2874	0.1429	0.2943	0.2754
Balanced Accuracy	0.9396	0.75086	0.9167	0.9013	0.9482	0.8039	0.9287	0.8988

Region group, it is evident from the above table that, under **Metro**, the metrics for the test set were Sensitivity (93%), Specificity (95%), Positive predictive value (88%), Negative Predictive value (97%), and Balanced Accuracy (94%). Under **Town**, the metrics for the test set were Sensitivity (52%), Specificity (98%), Positive predictive value (88%), Negative Predictive value (90%), and Balanced Accuracy (75%). Under **Urban**, the metrics for the test set were Sensitivity (83%), Specificity (100%), Positive predictive value (100%), Negative Predictive value (91%), and Balanced Accuracy (92%). Under **Village**, the metrics for the test set were Sensitivity (96%), Specificity (84%), Positive predictive value (55%), Negative Predictive value (99%), and Balanced Accuracy (90%). The Overall Accuracy of Test data is 82%.

For the Region group, , it is evident from the above table that, under the **Metro**, the metrics for the training set were Sensitivity (93%), Specificity (97%), Positive predictive value (92%), Negative Predictive value (97%), and Balanced Accuracy (95%). Under **Town**, the metrics for the training set were Sensitivity (62%), Specificity (98%), Positive predictive value (92%), Negative Predictive value (91%), and Balanced Accuracy (80%). Under **Urban**, the metrics for the training set were Sensitivity (89%), Specificity (97%), Positive predictive value (92%), Negative Predictive value (95%), and Balanced Accuracy (93%). Under **Village**, the metrics for the training set were Sensitivity (91%), Specificity (88%), Positive predictive value (66%), Negative Predictive value (98%), and Balanced Accuracy (90%). The Overall Accuracy of Training data is 85%.

7.3.4 Customer Income Group:

	Test Data			
	Above 80,000	**Bet 30,000 to 40,000**	**Bet 40,000 to**	**Bet 60,000**
Sensitivity	0.9111	0.7826	0.875	0.8704
Specificity	0.9307	0.9675	0.9672	0.9565
PosPred Value	0.8542	0.8182	0.84	0.9216
NegPred Value	0.9592	0.9597	0.9752	0.9263
Prevalence	0.3082	0.1575	0.1644	0.3699
Detection Rate	0.2808	0.1233	0.1438	0.3219
Detection Prevalence	0.3288	0.1507	0.1712	0.3493
Balanced Accuracy	0.9209	0.875	0.9211	0.9134
	Train Data			
	Above 80,000	**Bet 30,000 to 40,000**	**Bet 40,000 to**	**Bet 60,000 to**
Sensitivity	0.8477	0.8879	0.7941	0.92
Specificity	0.9714	0.9536	0.9582	0.9409
PosPred Value	0.9382	0.812	0.802	0.8703
NegPred Value	0.9256	0.9741	0.9563	0.9646
Prevalence	0.3391	0.1842	0.1756	0.3012
Detection Rate	0.2874	0.1635	0.1394	0.2771
Detection Prevalence	0.3064	0.2014	0.1738	0.3184
Balanced Accuracy	0.9095	0.9207	0.8762	0.9304

Income group, it can be concluded from the above table that, for the income group (Above 80,000), the metrics for the test set were Sensitivity (91%), Specificity (93%), Positive predictive value (85%), Negative Predictive value (96%) and Balanced Accuracy (92%). For Income group (Bet 30,000 to 40,000) the metrics for the test set were Sensitivity (78%), Specificity (97%), Positive predictive value (82%), Negative Predictive value (96%), and Balanced Accuracy (88%). For Income group (Bet 40,000 to 60,000) the metrics for the test set were Sensitivity (88%), Specificity (97%), Positive predictive value (84%), Negative Predictive value (98%), and Balanced Accuracy (92%). For Income group (Bet 60,000 to 80,000) the metrics for the test set were Sensitivity (87%), Specificity (96%), Positive predictive value (92%), Negative Predictive value (93%), and Balanced Accuracy (91%).The Overall Accuracy of Test data is 87%.

Under Income group, can be concluded from the above table that, For Income group (Above 80,000), the metrics for the training set were Sensitivity (85%), Specificity (97%), Positive predictive value (94%), Negative Predictive value (93%), and Balanced Accuracy (91%). For Income group (Bet 30,000 to 40,000), the metrics for the training set were Sensitivity (89%), Specificity (95%), Positive predictive value (81%), Negative Predictive value (97%), and Balanced Accuracy (92%). For Income group (Bet 40,000 to 60,000), the metrics for the training set were Sensitivity (79%), Specificity (96%), Positive predictive value (80%), Negative Predictive value (96%), and Balanced Accuracy (88%). For Income group (Bet 60,000 to 80,000), the metrics for the training set were Sensitivity (92%), Specificity (94%), Positive predictive value (87%), Negative Predictive value (96%), and Balanced Accuracy (93%). The Overall Accuracy of Training data is 87%.

7.3.5 Customer Occupation:

	Test Data			
	Employed	**Home Maker**	**Professional**	**Self Employee**
Sensitivity	0.451	0.8148	0.7143	0.4545
Specificity	0.8526	0.7983	0.8468	0.9469
PosPred Value	0.6216	0.4783	0.5952	0.7143
NegPred Value	0.7431	0.95	0.9038	0.856
Prevalence	0.3493	0.1849	0.2397	0.226
Balanced Accuracy	0.6518	0.8066	0.7806	0.7007

	Train Data			
	Employed	**Home Maker**	**Professional**	**Self Employee**
Sensitivity	0.5222	0.9	0.7456	0.5966
Specificity	0.918	0.8635	0.8301	0.9372
PosPred Value	0.7737	0.5473	0.6429	0.71
NegPred Value	0.7815	0.9792	0.8883	0.9002
Prevalence	0.3494	0.1549	0.2909	0.2048
Balanced Accuracy	0.7201	0.8818	0.7878	0.7669

Occupation group the above table delivers that, For Employment group (Employed), the metrics for the test set were Sensitivity (45%), Specificity (85%), Positive predictive value (62%), Negative Predictive value (74%), and Balanced Accuracy (65%). For Employment group (Home Maker), the metrics for the test set were Sensitivity (81%), Specificity (80%), Positive predictive value (48%), and Negative Predictive value (95%), Balanced Accuracy (81%). For Employment group (Professional), the metrics for the test set were Sensitivity (71%), Specificity (85%), Positive predictive value (60%), Negative Predictive value (90%), and Balanced Accuracy (78%). For Employment group (Self Employed), the metrics for the test set were Sensitivity (45%), Specificity (95%), Positive predictive value (71%), and Negative Predictive value (86%), Balanced Accuracy (70%).

Under the Occupation group, the above table delivers that, For Employment group (Employed), the metrics for the training set were Sensitivity (52%), Specificity (92%), Positive predictive value (77%), Negative Predictive value (78%), and Balanced Accuracy (72%). For Employment group (Home Maker), the metrics for the training set were Sensitivity (90%), Specificity (86%), Positive predictive value (55%), Negative Predictive value (98%) and Balanced Accuracy (88%). For Employment group (Professional), the metrics for the training set were Sensitivity (75%), Specificity (83%), Positive predictive value (64%), Negative Predictive value (89%), and Balanced Accuracy (79%). For Employment group (Self Employed), the metrics for the training set were

Sensitivity (60%), Specificity (94%), Positive predictive value (71%), and Negative Predictive value (90%), Balanced Accuracy (77%). The Overall Accuracy of Training data is 66%.

7.3.6 Education:

	Test Data			
	Graduation	**Intermediate/10+2**	**Post-Graduation**	**Professional Degree**
Sensitivity	0.34375	0.6875	0.8696	1
Specificity	0.90351	0.8246	0.97	0.9727
PosPred Value	0.5	0.5238	0.9302	0.9231
NegPred Value	0.83065	0.9038	0.9417	1
Prevalence	0.21918	0.2192	0.3151	0.2466
Detection Rate	0.07534	0.1507	0.274	0.2466
Detection Prevalence	0.15068	0.2877	0.2945	0.2671
Balanced Accuracy	0.62363	0.756	0.9198	0.9864

	Train Data			
	Graduation	**Intermediate/10+2**	**Post-Graduation**	**Professional Degree**
Sensitivity	0.39189	0.8254	0.8954	0.9805
Specificity	0.94226	0.8462	0.9533	0.9625
PosPred Value	0.6988	0.5977	0.8726	0.9042
NegPred Value	0.81928	0.9459	0.9623	0.9928
Prevalence	0.25473	0.2169	0.2633	0.2651
Detection Rate	0.09983	0.179	0.2358	0.2599
Detection Prevalence	0.14286	0.2995	0.2702	0.2874
Balanced Accuracy	0.66708	0.8358	0.9243	0.9715

Education group, For Education groups (Graduation), the metrics for the test set were Sensitivity (34%), Specificity (90%), Positive predictive value (50%), Negative Predictive value (83%) and Balanced Accuracy (62%). For Education group (Intermediate/10+2), the metrics for the test set were Sensitivity (69%), Specificity (82%), Positive predictive value (52%), Negative Predictive value (90%), Balanced Accuracy (76%). For Education group (Post Graduation), the metrics for the test set were Sensitivity (87%), Specificity (97%), Positive predictive value (93%), Negative Predictive value (94%) and Balanced Accuracy (92%). For Education group (Professional Degree), the metrics for the test set were Sensitivity (100%), Specificity (97%), Positive predictive value (92%), Negative Predictive value (100%) and Balanced Accuracy (99%). The Overall Accuracy of Test data is 75%.

Under the Education group, For Education group (Graduation), the metrics for the training set were Sensitivity (39%), Specificity (94%), Positive predictive value (70%), Negative Predictive value (82%) and Balanced Accuracy (67%). For Education group (Intermediate/10+2), the metrics for the training set were Sensitivity (83%), Specificity (85%), Positive predictive value (60%), Negative Predictive value (95%) and Balanced Accuracy (84%). For Education group (Post Graduation), the), the metrics for the training set were Sensitivity (90%), Specificity (95%), Positive predictive value (87%), Negative Predictive value (96%) and Balanced Accuracy (92%). For Education group (Professional Degree), the metrics for the training set were Sensitivity (98%), Specificity (96%), Positive predictive value (90%), Negative Predictive value (99%) and Balanced Accuracy (97%).

7.3.7 Age Group:

	Test Data				
	26-35	**36-45**	**46-55**	**Above 56**	**Below 25**
Sensitivity	0.46429	0.18605	0.2439	0.6364	1
Specificity	0.88983	0.85437	0.99048	0.9115	0.627586
PosPred Value	0.5	0.34783	0.90909	0.6774	0.018182
NegPred Value	0.875	0.71545	0.77037	0.8957	1
Prevalence	0.19178	0.29452	0.28082	0.226	0.006849
Detection Rate	0.08904	0.05479	0.06849	0.1438	0.006849
Detection Prevalence	0.17808	0.15753	0.07534	0.2123	0.376712
Balanced Accuracy	0.67706	0.52021	0.61719	0.7739	0.813793

	Train Data				
	26-35	**36-45**	**46-55**	**Above 56**	**Below 25**
Sensitivity	0.48246	0.29688	0.40426	0.688	0.88889
Specificity	0.92291	0.87404	0.96818	0.9364	0.66783
PosPred Value	0.6044	0.53774	0.80282	0.7478	0.0404
NegPred Value	0.87959	0.71579	0.83529	0.9163	0.99739
Prevalence	0.19621	0.33046	0.24269	0.2151	0.01549
Detection Rate	0.09466	0.09811	0.09811	0.148	0.01377
Detection Prevalence	0.15663	0.18244	0.1222	0.1979	0.34079
Balanced Accuracy	0.70268	0.58546	0.68622	0.8122	0.77836

Under the Age group, For Age group (26-35), the metrics for the test set were Sensitivity (46%), Specificity (89%), Positive predictive value (50%), Negative Predictive value (88%) and Balanced Accuracy (68%). For Age group (36-45), the metrics for the test set were Sensitivity (19%), Specificity (85%), Positive predictive value (35%), Negative Predictive value (72%) and Balanced Accuracy (52%). For Age group (46-55), the metrics for the test set were Sensitivity (24%), Specificity (99%), Positive predictive value (91%), and Negative Predictive value (77%), Balanced Accuracy (62%). For Age group (Above 56), the metrics for the test set were Sensitivity (64%), Specificity (91%), Positive predictive value (68%), Negative Predictive value (90%) and Balanced Accuracy (77%). For Age group (Below 25), the metrics for the test set were Sensitivity (100%), Specificity (63%), Positive predictive value (2%), Negative Predictive value (100%) and Balanced Accuracy (81%). The Overall Accuracy of Test data is 36%.

Under the Age group, For Age group (26-35), the metrics for the training set were Sensitivity (48%), Specificity (92%), Positive predictive value (60%), Negative Predictive value (88%) and Balanced Accuracy (70%). For Age group (36-45), the metrics for the training set were Sensitivity (30%), Specificity (87%), Positive predictive value (54%), Negative Predictive value (72%), and Balanced Accuracy (59%). For Age group (46-55), the metrics for the training set were Sensitivity (40%), Specificity (97%), Positive predictive value (80%), Negative Predictive value (84%) and Balanced Accuracy (69%). For Age group (Above 56), the metrics for the training set were Sensitivity (69%), Specificity (94%), Positive predictive value (75%), Negative Predictive value (92%), and Balanced Accuracy (81%). For Age group (Below 25), the metrics for the training set were Sensitivity (89%), Specificity (67%), Positive predictive value (4%), Negative Predictive value (100%), and Balanced Accuracy (78%).The Overall Accuracy of Training data is 45%.

This research on the survey data provides an insight of how the different demographic variables are affecting the customer buying behavior and e- commerce. A close study of the factors influencing is also examined. A multinomial logistic regression model is applied. All the demographic variables are processed through the model. From the model validating techniques and the diagnostics, it is observed that customer region is the variable that is highly influencing the customer buying behavior and the E- Commerce. The second reason is the customer income which is also influencing the customer buying behavior and E- Commerce next to the customer Region to some extent. The rest of the demographic variables are not showing significant influence regarding the survey data.

VIII. LIMITATION

A limitation of this study lies in the sample being from one place. An additional, more diverse, sample should be examined. Also, this study did not seek to address influences on trust and risk. Many additional factors could be proposed as influences on trust and risk and should be studied in the future. In particular, one's experience level or number of years of involvement in e-commerce could be used to explain one's attitude toward buying and/or selling. This study also measures perceptions and not true behaviours. Therefore, future studies should seek to gather data regarding actual consumer behaviours in e-commerce. Finally, given the combination of trust and attitude in the attitude toward selling model, further research is needed to examine why a consumer's attitude and trust might be combined in certain situations or environments.

IX. CONCLUSION

The results of this study will provide insight information of customer attitude towards E-Commerce in India. The result identifies major problems of Customer behaviour and their direct impact on the sale and production. Although this study is based on Indian scenario of business, besides this the results derived from study carried out in the field are equally applicable in other developing countries. This study is to mainly focus on how demographic variables (age, income and occupation) affect customer attitude towards E Commerce. This study helps to analyse the pattern of online buying (types of goods, e-commerce experience and hours use on internet) influence customer attitude towards online shopping. This research focuses on how purchase perception (product perception, customer service and consumer risk) influence customer attitude towards E Commerce. This research has identified that there are a number of factors determining customer's intent to repurchase within the e-commerce, specifically these factors are customer experience with an e-brand and beliefs concerning the importance of convenience, trust and security when purchasing. From the model validating techniques and the diagnostics, it has been observed that customer region is the variable that is highly influencing the customer buying behaviour and the E- Commerce. The second reason is the customer income which is also influencing the customer buying behaviour and E- Commerce next to the customer Region to some extent. The rest of the demographic variables are not showing significant influence regarding the survey data.

ACKNOWLEDGEMENT

The authors thank to the Managing Editor and the anonymous reviewer for his suggestion which have led to an improvement in both quality and clarity of the paper.

REFERENCES

[1] Akinci, S., Aksoy, S., &Atilgan, E. (2004). Adoption of internet banking among sophisticated consumer segments in an advanced developing country. International Journal of Bank Marketing, 22 (3), 212-232.
[2] Bakos, J.Y. (1997), "Reducing buyer search costs: implications for electronic marketplaces", Management Science, 43(12), pp. 1676-1692.
[3] Chellappa, K.R and Pavlou A.P. (2002), "Perceived information security, financial liability and consumer trust in electronic commerce transactions", Logistics Information Management, 15(5), pp. 358-368.
[4] Chiles, T.H. and McMackin, J.F. (1996), "Integrating variable risk preferences, trust, and transaction cost economics". Academy of Management Review, 21, 73-99.
[5] Chung-Hoon P., Y.-G. Kim, (2003), "Identifying key factors affecting consumer purchase behavior in an online shopping context", International Journal of Retail and Distribution Management, 31(1), pp. 16-29.
[6] Clarke, K. (2001), "What price on loyalty when a brand switch is just a click away?" Qualitative Market Research, 4(3), pp. 160-168.
[7] Conner M. and Armitage J.C. (1998), "Extending the theory of planned behavior: A review and avenues for further research", Journal of Applied Social Psychology, Vol. 28, pp. 1430.
[8] Constantinides, E. (2004)., "Influencing the online consumer's behavior: the Web experience". Internet Research. Vol. 14, No. 2, pp. 111-126.
[9] Currall, S. and Judge, T. (1995), "Measuring Trust Between Organizational Boundary Role Persons," Organizational Behavior and Human Decision Processes (64), p. 151-170.
[10] Cuneyt, K. Gautam, B. (2004). The impacts of quickness, price, payment risk, and delivery issues on on-line shopping, Journal of Socio-Economics, Vol.33, PP.241–251.
[11] Dawson, J. (2000). Retailing at Century End: Some Challenges for Management and Research. International Review of Retail Distribution and Consumer Research vol (2): 119-124.
[12] Dekimpe, M.G., Steenkamp, J.-B.E.M. and Mellens, M. (1997), "Decline and variability in brand loyalty'', International Journal of Research in Marketing, Vol. 14, pp. 405-420.
[13] Dillon, T. D., and Reif, H. L. (2004). Factors influencing consumer's e-commerce commodity purchases. Information Technology, Journal of Learning and Performance, 22(2), 1-12. Hasan, H., and Rahim, S. A. (2004). Factors affecting online purchasing behavior. Journal of Communication, 24, 1-19. Jarvenpaa, S. L., & Todd, P. A. (1996-97). Consumer reactions to electronic shopping on the World Wide Web. International Journal of Electronic Commerce, 1(2), 59-88.
[14] Kesh, S., Ramanujan, S. and Nerur, S. (2002) "A framework for analyzing ecommerce security", Information Management & Computer Security, 10 (4)pp. 149-458.
[15] Liao, Z. and Cheung, M.T. (2002), "Internet-based E-Banking and Consumer Attitudes: An Empirical Study". Information and Management, Vol. 39, pp. 283–295.
[16] Rotchanakitumnuai, S., &Speece, M. (2003). Barriers to internet banking adoption: a qualitative study among corporate consumers in Thailand. International Journal of Bank Marketing, 21 (6/ 7), 312-323.

[17] Roehl, W.S. and Fesenmaier, D.R. 1992. Risk Perceptions and pleasure travel: An exploratory analysis. Journal of Travel Research, 30(4): 17–26.
[18] Schlosser, A.E. (2003a). Experiencing products in the virtual world: the role of goal and imagery in influencing attitudes. Journal of Consumer Research, Vol. 30, pp. 184-198.
[19] Tan, M. &Teo, T. S. H. 2000. Factors influencing the adoption of Internet banking. Journal of the Association for Information Systems. Vol. 1 (5), pp. 1-42.
[20] Wang, Y., Wang, Y., Lin, H., & Tang, T. (2003). Determinants of user acceptance of internet banking: an empirical study. International Journal of Bank Marketing, 14 (5), 501-519.