International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Identify Influencing Factor for Ecommerce entrepreneursand Start-ups companies with Machine Learning approach

[1]Maruthish Epuri
Sri Krishnadevaraya University,
maruthish.ds@gmail.com

[2] Prof. T Bhaskar Reddy
Sri Krishnadevaraya University
bhaskarreddy.sku@gmail.com

ABSTRACT

The purpose of this research is to provide insights on the Customer behavior in rural areas towards Online shopping in the high growing world. In this study following factors that contribute to the success of Online Marketing research: Public reviews, Transport facility, Introducing Gadgets, new era of Internet facility. The main objective of this study is to investigate the factors influencing the adoption of E-Commerce in rural India. This provides more awareness that defines the customer behavior on purchasing the products through online and the companies marketing strategies of improving their business in rural world. It also creates a platform for understanding the business through various factors which are clearly drafted. This provides such a framework where many small scale, large scale and multinational companies could be followed with versatile and innovative technologies involved. An appropriate use of the technologies and ideologies that emerged from this research, in the context of business improvements and customer satisfaction will assist organizations to properly utilize all resources with the aim of improving online marketing. This may be repeated to find lot more opportunities in order to receive the profits for business/satisfaction to buyers on a regular basis, or after major changes in the high evolving online market.There is no scope of limitation in analyzing the customer behavior of online shopping, but needs regular check to find the limitations which obstacle the business growth.This research papers represents the concept of knowledge giving through business growth of rural marketing.

KEYWORDS:E-Commerce, Online marketing, Factor Analysis, Machine Learning Algorithms Data Reliability, Rural India

1. INTRODUCTION

E-commerce is the buying and selling of goods and services, or the transmitting of funds or data, over an electronic network, majorly the internet. This type of business transactions occur either as business-to-business (b2b), business-to-consumer (b2c), consumer-to-consumer (c2c) or consumer-to-business(c2b). The terms e-commerce and e-business are frequently used interchangeably. Sometimes the term e-tail is also used in reference to transactional processes for online shopping. The term E-commerce was originally conceived to describe the process of conducting business transactions electronically using technology from Electronic Data Interchange (EDI) and Electronic Funds transfer (EFT).

Electronic Commerce (e-commerce) is a means of doing business through networks of computer like offering different goods and service ranging from tour and travels, movies, hotel reservation, matrimonial service, electronic gadget, fashion accessories and groceries. A developing country can be recognized if it introduces e-commerce effectively and efficiently. It will be enhanced its output and gives competitive advantages.

2. LITERATURE REVIEW

India has been experienced remarkable growth in information technology and e-commerce. According to the e-bay census guide 2011, India is home to 3311 e-commerce hubs, 1267 rural hubs, 391 export hubs and 2217 import hubs. Due to the advancement of technology the world has become a global village. People can buy goods with a click of mouse button without moving out of their place. Consumer by couriers and postal services is not very reliable in smaller cities, towns and rural areas. However, many banks in India put the Internet banking facilities. The speed post and courier system were also improved tremendously in recent years. The reason of preferring E-commerce over traditional system both from Purchasers and Consumers end because of some of the following reasons:

Door step delivery, Speed and easy accessibility, Cost efficient, wide variety of product selections, Generated invoices and bills for future reference,

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Quick Services, Less investment costs (Manpower, Labor cost, Infrastructure), Direct Marketing with customers, Less likely product returns, Marketing strategy of with/without Warehouse, Always can be in contact with customers to provide the better information about all types of available products in market and many more.Following are the possible expected constraints: Lack of "touch and feel" factor, Lack of proper courier services in rural areas. Fear of carrying the financial transactions over internet, The lack of trust and the fear of being fooled etc.

Purchasers and Consumers have some challenges to go with ecommerce as follows:

PURCHASER'S END: Weak ICT access and low internet penetration, No or weak electronic payment infrastructure and networks, High delivery costs or delivery not feasible, the inefficient logistics and transportation are some of the key challenges in the delivery of goods in rural areas.

SELLER'S END: No proper courier service, Poor infrastructure, more expensive Standard urban-rural connectivity, Low literacy, widespread poverty, low income, and unemployment in rural areas lead to limited purchase capabilities.

Thanks to the emergence of low cost smart phones, however, poor last mile connectivity could act as a deterrent. Keeping control on logistics and on ground fleet management, especially transport and courier companies, are very essential for growth. E-commerce is the word ruling the business since a few years and we have many achievements, some of them as follows: It is a win-win situation for both the consumers and the merchants. Recently, the technology is well developed in the rural years and change the ideology that e-commerce companies may not reach the rural region of India. Moreover, the last few years has witnessed a dramatic change in the Indian Postal service as it is getting collaborated with more than 400 e-commerce websites. Many e-commerce giants revenue coverage from the villages are appreciably notable. Certain unique attributes like COD (cash on delivery) mode of payment and direct imports that lower costs considerably are probably going to bring about a speedy growth in this industry in the upcoming years.

3. KEY SOURCES/PARTNERS OF E-COMMERCE SUCCESS:

Post Offices are traditionally reliable and cheap and so far, We have around 1,55,000 post offices in India and through this Indian postal service Indian postal service, there is much infuses in the reliability and competitiveness when it comes to shipping E- Commerce products to end consumers in the remote villages in rural India.

Affordable cost for common man, Ease use and accessible, retail shopping mobile app can be used anywhere in India for online shopping.

According to the Boston Consulting Group (BCG) report, almost half of all Indian internet users will be from rural areas in 2020. BCG expects the Internet to contribute $200 billion to India's GDP (5 percent of total GDP) by 2020.Internet & Mobile Association of India (IAMAI) estimated that rural India will have 109 million mobile internet users by June 2016, 60.2 per cent higher than figures of June 2015.Another flagship program under National e-Governance Plan to open internet kiosks – Common Service Centers (CSCs) – at 1, 00,000 Gram Panchayats was launched in 2009. Now the Government of India has increased the number to all 2, 50, 000 Panchayats.

The rural regions are already well covered by basic telecommunication services and are now witnessing in increasing penetration of computers and smart phones. Akodara, located in Gujarat's Sabarkantha district is proud to be India's first 'digital village'. The village of roughly 1,200 people has fully adopted by ICICI bank to demonstrate the kind of progress digitization can make on a society. Facilitating the creation of local-language, or vernacular, Internet Strategic investments in internet/broadband penetration in rural India. Need of best Internet Facility, Government of India is looking to install Wi-Fi hotspots at more than 1,000 gram panchayats across India.Facilitating digital transactions and easing payments. Postal Facility: Effective and efficient use of the Indian-Post logistical infrastructure

4. OBJECTIVES

This study focuses on the following research objectives:

- To study the influence of demographic variables on E- Commerce in rural India.
- The main aim of the study is to understand the factors which influence customer expectation to adopt E-commerce in rural India.
-

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

5. HYPOTHESIS

H1: There is a significant relationship between factors influencing demographic variables and adoption of E- Commerce Factors

H2: is there any significant association with psychographic variables vas E commerce Facrtors

6. RESEARCH METHODOLOGY

QUESTIONNAIRE DESIGN: Based on the research objectives, a structured questionnaire with 61 independent variables information contained, mainly with 50 different opinions.

DATA COLLECTION: Customers were considered as population of research interest. The research was carried out on selected 1068 people from rural areas in India as per the sample determination standard formula . For data collection random sampling was adopted. To make sure all questions being answered in a proper way, questionnaires were completed and screened one-by-one.

DATA CLEANING: Data mining Tools and Techniques i.e. R programming, R Rattle, Weka and SPSS 2.0 software's (Regression, Classifying& Clustering) has been used to analyse the data.

7. DATA ANALYSIS

7.1 Data Access & Cleaning

A survey is conducted on 2000 customers. After a pilot study we consider 777 to be a goodsample with a sample error of 5%. The data is cleaned. Imputation, Data validation and datareliability, process is performed using R Programming language and Weka ProgrammingLanguage.

7.2 Analysis Tools & Techniques

Data mining Tools and Techniques i.e. WEKA, Python (3.7), R programming, R Rattle and R Rcmdr havebeen used to analyse the data. We have applied unsupervised machine learning technique for Data Validation and Standardization and Supervised learning techniques such asClassification models are applied for predicting customerbuying behaviours. For data standardization and validation we have applied Principle Component Analysis and factor analysis is applied for Data Dimension Reduction. Originally there were 47variables influencing attributes/variables which were reduced to 10 after the PCA technique is applied, the detailed information is available under section 8. And the analysis is carried here afterwards with the reduced 10 variables.

7.3. Data Validation and Quality Check:

Data is validated by performing the reliability test for the data, using Cronbach's Alpha. KMOand Bartlett's tests are performed for the sample adequacy. A good result of 87% is observedindicating the reliability percentage of the data. Also, the KMO test for the factor data validationis obtained to be 85.5%. Tables for the data reliability and validity tests are given below.

8. RESULT

8.1: Data Reliability:

The Cronbach's Alpha which is a reliability statistic is obtained minimum should be 70% or else increase sample size. This is found to be 76% which means the data is reliable. The table for this is shown below.

Recommended by Nunnally, J. C. (1978). Assessment of Reliability[25]

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.759	.777	50

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

8.2: Data Consistence:

A sample of 1068 is adequate for the study is confirmed by the KMO statistic. Minimum should be 50%, if less than 50% should increase sample size. A good result of 85.5% is showing the sample validity.

Variibles having negative covariation matrix values doesn't have KMO, so as per the sample collection, we have good consistency in the data and we need sufficient consistency for doing PCA analysis.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.855
Bartlett's Test of Sphericity	Approx. Chi-Square	138467.55
	Df	1081
	Sig.	0

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.656	20.545	20.545	9.656	20.545	20.545
2	7.558	16.081	36.625	7.558	16.081	36.625
3	5.997	12.759	49.385	5.997	12.759	49.385
4	5.025	10.692	60.076	5.025	10.692	60.076
5	4.118	8.763	68.839	4.118	8.763	68.839
6	3.743	7.963	76.802	3.743	7.963	76.802
7	3.152	6.706	83.508	3.152	6.706	83.508
8	2.506	5.332	88.840	2.506	5.332	88.840
9	2.110	4.489	93.329	2.110	4.489	93.329
10	1.375	2.925	96.254	1.375	2.925	96.254

8.4: PCA Data Validation for Further Modelling implementation

Originally there were 47 variables influencing attributes/variables which were reduced to 10 after the PCA technique is applied, the detailed information is available under section 8. And the analysis is carried here afterwards with the reduced 10 variables with the recommendation of 96% of Total Variance Explained with minimum 1.375 to 9.656 Eigen values. The below rotated matrix table is given with the grouping variables together based on the grouping we have assigned new Factor names and with the help of new factor variables we are predicting customers buying behaviour

Pricing & Delivery	Q22	.866		
	Q21	.771		
	Q19	.764		
	Q46	.684		
	Q39	-.549		
	Q42	-.594		
	Q60	-.777		
Brand Influence & Security	Q38		-.299	.830
	Q28		-.014	.795
	Q30		.260	.781
	Q48		-.280	.768

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

	Q18		.201	.734								
	Q40		.127	.524								
Life Style & Accessibility	Q52		-.092	-.163	.904							
	Q23		.259	.026	.856							
Decision making	Q50		.050	-.117	-.153	.952						
	Q51		.117	-.103	-.099	.948						
	Q43		-.382	.016	.073	.766						
	Q59		.129	-.475	.450	.661						
	Q33		-.521	-.054	.051	.612						
Online Shopping	Q55		-.149	-.193	-.227	-.066	.887					
	Q36		.016	.226	.038	.029	.816					
	Q54		-.172	-.079	.041	.160	.780					
	Q53		-.294	-.270	-.014	.104	.669					
Discounts & Comfort	Q26		.096	.233	.190	.056	.180	.846				
	Q32		-.359	.091	-.082	-.295	.140	.812				
	Q16		.152	-.250	.011	.066	.059	.674				
	Q20		.365	.363	-.082	-.099	.076	.573				
	Q15		-.115	.538	.002	-.082	.014	.564				
	Q56		-.049	.244	-.053	.040	.451	.463				
Websites design	Q57		-.016	.142	-.157	-.072	.077	.095	.911			
	Q58		.346	-.003	.047	.387	.113	.321	.706			
Safety measures	Q17		.170	-.024	-.184	-.003	-.010	.322	-.075	.884		
	Q61		-.207	-.109	.211	.116	-.310	-.125	-.199	.851		
Time Saving	Q34		-.255	.166	-.063	-.019	.096	-.024	-.097	-.249	.896	
	Q35		-.223	.119	.159	.382	.185	.117	.134	.439	.694	
	Q29		.370	.278	-.028	.339	-.563	.051	-.205	-.174	.486	
Proactive intimation	Q41		-.291	-.030	.501	.013	-.116	.072	.037	.007	.035	.783

Factors Data Summarization and Visualization Summarization

	Mean	sd	se(mean)	IQR	Cv	skewness	Kurtosis
Pricing & Delivery	0.34	0.89	0.03	1.42	2.66	-0.01	-1.00
Brand Influence & Security	0.15	0.90	0.03	1.21	6.02	-0.57	-0.56
Life Style & Accessibility	-0.19	1.02	0.03	1.36	NA	0.70	-0.08
Decision Making	-0.17	0.84	0.03	0.89	NA	-0.91	2.96
Online Shopping	0.19	1.04	0.03	0.56	5.43	10.74	170.36
Discounts and Comfort	-0.02	1.10	0.03	1.63	NA	1.71	6.32
Websites and Design	0.29	0.91	0.03	1.35	3.11	-0.42	0.19
Safety Measures	0.15	0.93	0.03	1.18	6.31	0.30	-0.05
Time Saving	-0.02	1.07	0.03	1.08	NA	0.71	-0.15
Proactive Intimation	0.03	0.99	0.03	0.99	29.46	-0.14	2.37

8.5 Insights: data Quality Check through summarized techniques

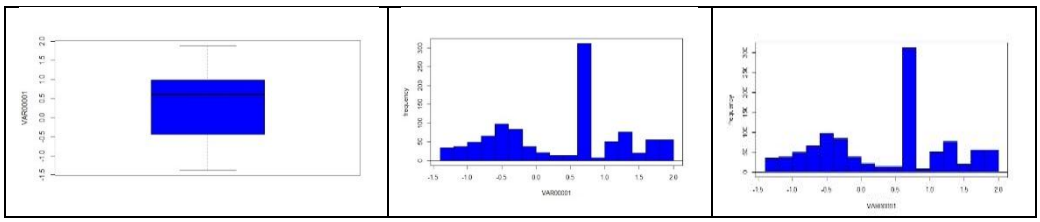
- Pricing & Delivery factor is affecting by Gender. Both male and female are more concern about delivery cost and price of

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

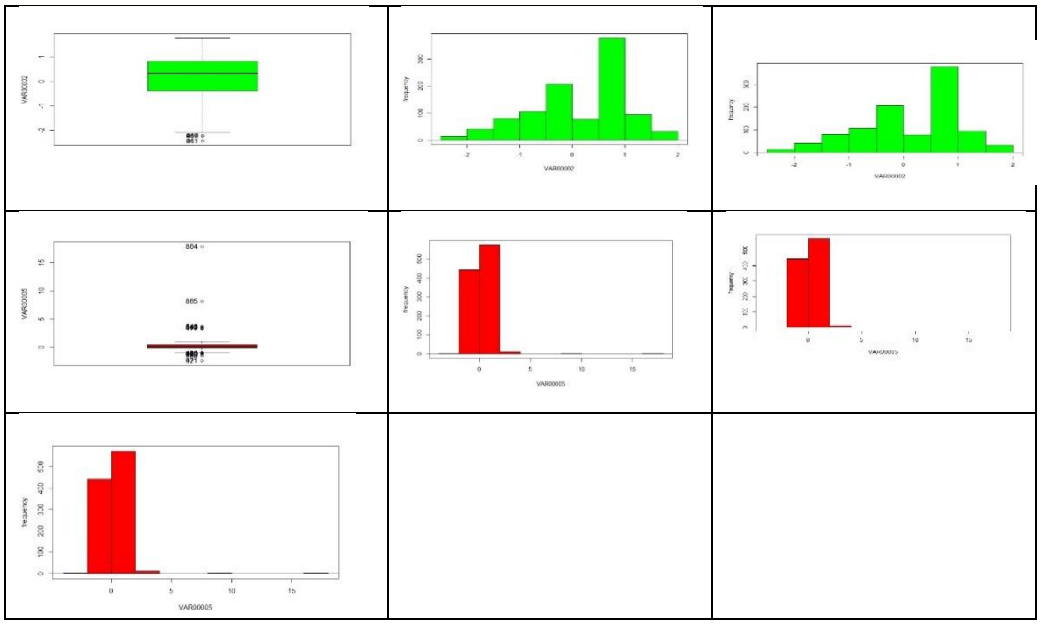
the products. Marital Status affects this factor as well. Different age group people are reacting in different ways about price and delivery factor so that there is association between age group,Pricing& Delivery factor. Education qualification and monthly family income are also having association with Pricing and Delivery Factor. Customer Region and Occupation are having strong association with Pricing and delivery factor.

- Brand Influence & Security factor has association with Gender, Age group, educational status, Occupation and highly associated with Monthly family income. This means as the remaining variables influencing this factor but as this factor having strong association means Brand Influence & Security factor is mostly depends on how much the person is earning per month. And there is no significance for this factor with Marital Status and Customer region.
- Life Style & Accessibility factor has high association with Age group, Educational qualification and Monthly income. And there is no association between Life Style & Accessibility and Customer Region.
- Decision making is associated with all demographic variables but it is highly associated with Occupation and Customer Region.
- The factor Online Shopping has less association with Gender and very high association with Occupation and Customer Region. It seems like as most of the doctors are encouraging them to buy the medicines online as some medicines are not available in rural areas. Also the people are attracted for offers and they are buying if any of their friends suggests. There is no Association between Online Shopping and Age group.
- Every demographic variable has association with Discounts & Comfort factor. But in these variables, Gender is having the lowest association, Monthly income and Occupation has the highest Association. Here monthly earnings and Occupation of the candidate effects Discounts & Comforts factor.
- Websites Design factor has significance with each and every demographic variable where Occupation has the highest association and marital status having the lowest association.
- Every Demographic variable concerning with the safety measures so that all are having Association with Safety measures factor. Here Customer Region having strong association.
- Time saving factor has strong association with Monthly family income, Occupation and Customer Region and this factor has weak association with Gender. Age group has no association with Time saving factor.
- Proactive intimation factor has strong association with Customer region and weak association with Gender

9. Visualization



International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



10.CONCLUSION

This research study revealed that adoption of E-commerce, followed by Factor **Name** variables strongly affect the overall expectation of E-commerce consumers in rural India. In order to promote customer satisfaction, it is inevitable for E-commerce service provider to give due emphasis to all the above-mentioned factors. It is well known that E-commerce in rural India has potential of exploring consumer. The researcher would like to add that a proactive and creative approach by E-commerce provider in rural India, for example, providing consumer education on E-commerce and friendly customer service will help improve the consumer confidence, and eventually overall customer satisfaction levels in rural India. A developing country can be recognized if it introduces e-commerce effectively and efficiently. It will be enhanced its output and gives competitive advantages. Shows it has no Outliers, so will go with further analysis with the supervised learning algorithms for customer behaviour predictions

11. REFERENCES

1. Akbar M. M. and Noorjahan P. (2009), "Impact of Service Quality, Trust and customer Satisfaction on Customers Loyalty". ABAC Journal Vol. 29, No. 1 (January-April 2009, pp.24-38).
2. 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.
3. Barnes, S.J., &Corbitt, B. (2003). Mobile banking: concept and potential. International Journal of Mobile Communications, 1 (3), 273-288.
4. B. B. Flynn, R. G. Schroeder and S. Sakakibara, A framework for quality management research and an associated measurement instrument, J. Operations Management 11 (1994) 339–366.
5. Bauer, H.H., Hammerschmidt, M. and Falk, T. (2005). "Measuring the quality of e-banking portals." International Journal of Bank Marketing, Vol. 23, No. 2, pp. 153-
6. Doll, et al (1995), "Measuring Customer Information Satisfaction for Website Marketing" Journal of Internet Banking and Commerce, Vol.9, No.3, pp.33-40.
7. Gan, C., Clemes, M., Limsounchai, V., &Weng, A. (2006). A logic analysis of electronic banking in New Zealand. International Journal of Bank Marketing, 24 (6), 360-383.
8. G. S. Dangayach and S. G. Deshmukh, Evidence of manufacturing strategies in Indian industry: A survey, Int. J. Production Economics 83 (2003) 279–98.
9. H. Quazi, Implementation of an environmental management system: The experience of companies operating in Singapore, Industrial Management & Data Systems 99(7) (1999) 302–31 1.

10. J. Nunnally, Psychometric Theory (McGraw-Hill, New York, 1967).
11. K. Muttar, an Investigation of the Validity of Objective and Subjective Measures of Organizational Climate (University Microfilms International, Ann Arbor, 1985).
12. Kolodinsky, J. M., Hogarth, J. M., &Hilgert, M. A. (2004). The adoption of electronic banking technologies by US consumers. International Journal of Bank Marketing, 22 (4), 238-259.
13. Liao, Z. and Cheung, M.T. (2002), "Internet-based E-Banking and Consumer Attitudes: An Empirical Study". Information and Management, Vol. 39, pp. 283–
14. N. Oliver, R. Delbridge, D. Jones and J. Lowe, World class manufacturing: Furtherevidence in the lean production debate, British Journal of Management 5(Specialissue) (1994) S53–S63.
15. P. Chandra and T. Sastry, Competitiveness of Indian Manufacturing, Vikalpa 23(3) (1998) 25–36.
16. R. H. Hayes and S. C. Wheelwright, Restoring Our Competitive Edge: Competingthrough Manufacturing (Wiley, New York, 1984).R. J. Schonberger, World Class Manufacturing (Free Press, New York, 1986).
17. 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.
18. Sathye, M. (1999). Adoption of internet banking by Australian consumers: an empirical investigation. International Journal of Bank Marketing, 17 (7), 324-334.
19. Servon Q. and Kaestner B. (2008), "Internet banking: where does India stand?" Journal of Contemporary Management, Vol. 2, No. 1, pp. 38-58.
20. Shil N and Das B, (2008), "A study of Customer Satisfaction with Regard to Banking: An empirical of QFD", The Icfain Journal of Management Research, Vol. VII, No. 8, pp. 7-26.
21. S. L. Ahire, D. Y. Golhar and M. A. Waller, Development and validation of TQM
22. implementation construct, Decision Sciences 27(1) (1996) 23–56.
23. 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.
24. 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.
25. Nunnally, J. C. (1978). Assessment of Reliability. In: Psychometric Theory (2nd ed.). New York: McGraw-Hill. pages 245-246