

STUDY OF FACTORS AFFECTING CUSTOMER BEHAVIOUR USING BIG DATA TECHNOLOGY

Prabin Sahoo¹, Dr. Nilay Yajnik²

^{1,2}SVKM's NMIMS, Deemed to be University, Mumbai

¹prabins@hotmail.com, ²nilayy@nmims.edu

ABSTRACT

Big data technology is getting momentum recently. There are several articles, books, blogs and discussion points to various facets of big data technology. The study in this paper focuses on big data as concept, and insights into 3 Vs such as Volume, Velocity and Variety and demonstrates their significance with respect to factors that can be processed using big data for studying customer behaviour for online users.

KEYWORDS

Big data, customer behaviour, customer activities, Volume, Varieties, Quantitative research

1. INTRODUCTION

The internet is the phenomenal innovation of the information system era. With billions of internet users across the globe generate huge amount of data like never before. There are millions of web sites which cater to various demands of its users. E-commerce, search engines, online shopping, banking, trading etc. are all accessible from any parts of the world. This has brought a new dimension to the user behaviour pattern. For example, there are multiple airline sites through which one can book an air ticket to any destination. While these facilities make the consumer life simpler, but creates a competitive environment for the service providers. Since there are multiple options available for customers, it is very difficult to sustain the customer base as with a slightest inconvenience can force customer to switch to a different service provider. Therefore, it is utmost important for the service providers to understand the customer demand, choices, preferences etc. and provide them high quality of services. So the challenge is how to find out these factors? Traditionally, in a buyer and supplier relationship, the customer visits a retail store; the supplier interacts with the customer and may request the buyer to fill a survey form. The information collected can be fed into a database, and can be processed to find out the liking and disliking factors. In such scenario, the supplier depends solely on the response rather the mercy of the consumer if he or she likes to fill the survey form with all integrity. For online shopping, there are sites which ask for filling an online survey, but that is again if the buyer wants to fill it or not. Therefore, a mechanism is required through which information about the customer can be processed to understand the behaviour of customer. Since each activity on the internet is recorded, so when a customer visits a site can be tracked. However, the log generated out of each visit is enormous as it accumulates over a period of time. The logs contain unstructured data which require huge amount of efforts for extraction, parsing and finally loading into a database. The entire process is very time consuming and therefore the traditional system fails. Big data is the technology which solves the problem of processing huge volume of data. It has primarily 3

characteristics such as Volume, Varieties, and Velocity. This paper focuses on these characteristics to find out its significance on the behavioural attributes of online customers. A survey has been conducted from the big data users and a model has been developed to process users' response with established theory.

2. LITERATURE REVIEW

There are theories on studying customer behaviours (Chung-Hoon Park and Young-Gul Kim, 2003); (Susan M. Keaveney, Madhavan Parthasarathy, 2001); (Limayem, Moez ; Dept. of Inf. Syst., City Univ. of Hong Kong, Kowloon, China ; Khalifa, M. ; Frini, A., 2000). These studies discuss about the factors/attributes of online customer.

Information satisfaction (Chung-Hoon Park and Young-Gul Kim, 2003) is closely related to product information quality, service information quality and user interface quality. This theory is useful for the big data research under consideration, as the purpose is to establish the outcome from big data analysis to behavioural factors.

Customer switching behaviour (Susan M. Keaveney, Madhavan Parthasarathy, 2001) theory has observed that online service usages are more with continuers than the switchers. This theory is useful for big data as it can identify the customers who are using frequently the online services. Further the theory also indicates that propensity of risk taking behaviour in buying a new product is more in case of service continuers. This will help in introducing new product/services on the most frequent product/service line of business.

Big data technology is based on 3 key concepts i.e. volume, variety and velocity (Philip Carter, 2011; Ramesh Nair, Andy Narayanan, 2012). It can handle huge data volume with structured, unstructured format. The logs from web are mix of unstructured or semi structured and structured in nature which can be processed using big data technology.

Volume: Volume is the amount of data that can be processed. Though volume represents huge volume in case of big data, but it is a relative term to be defined precisely.

Velocity: The demand to analyse the data in real time (Philip Carter, 2011; Ramesh Nair, Andy Narayanan, 2012), speed of data in and out (Wikipedia, retrieved 2014)

Variety: Data can be of structured, semi structured and unstructured in nature. Structured data can have delimiter to separate various columns. For example, "customer name, customer id, address, item, quantity" represent a structured format where the fields are separated by comma. Unstructured data can be of various types such as email, blog, twitter logs whereas semi structured data can be a combination of both i.e. structured and unstructured.

3. MODEL, CONCEPTS

The proposed model figure 2.0 uses theory, survey (Primary data), statistical model and big data. The theory is based on quality of service/product and the service usages. Quantitative research methodologies have been used in building the model.

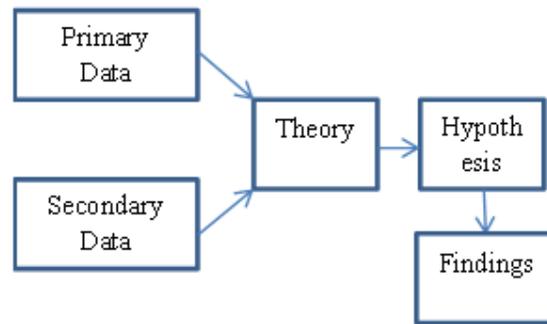


Figure 1. Quantitative research to establish Big Data factors for behaviour prediction

3.1. Quantitative research

Quantitative research is required to establish numerically that big data is useful in predicting the behaviour of customers. Survey has been conducted with the big data users to gather big data usages for e-commerce through questionnaire sessions. Participants were selected from Big Data users from e-commerce industries. Using statistical analysis hypotheses have been evaluated using IBM SPSS©.

3.2. Dependent Variable vs. Independent Variable

In table 1.0 the dependent and independent variables have been shown. The goal is to evaluate the impact of independent variables on the dependent variables.

Table 1. Independent vs. Dependent variables

Independent Variable	Dependent Variable
Transaction information	i) Service Usages
Feedback	
Browsing information	ii) Quality of Service

3.3. Theory

3.3.1 Quality of product/service

Quality of service or product (Valarie A. Zeithaml, A. Parasuraman, Arvind Malhotra, 2002) has been discussed as an important aspect in the web based services. For example, the transaction on a particular service is less because of less demand of the service or may be that the web service quality is not up to the mark. Therefore, to understand the root cause it is important to analyse the quality of service. If the quality is good, then the numbers of transactions decide the demand of the service or the product under study. Quality of service is further supported by the theory of quality of product information (Chung-Hoon Park and Young-Gul Kim, 2003).

3.3.2 Service Usages

Online service usages (Susan M. Keaveney, Madhavan Parthasarathy, 2001) theory indicates that the online service usage is one of the factors that determine the customer switching behaviour.

This is crucial in favour of online usage and therefore use of big data in studying customer behaviour. With higher usages of online services, high volume of logs will be produced which can be processed using big data to find out the customer access pattern on a service or product.

3.4 Model

In this model key factors such as blogs, surfing patterns, and transaction information are being fed into big data. The data format is unstructured and the volume is huge which cannot be processed through traditional data processing tools. Using Big Data these factors are tested against the theory of i) Quality of service and ii) Service usages.

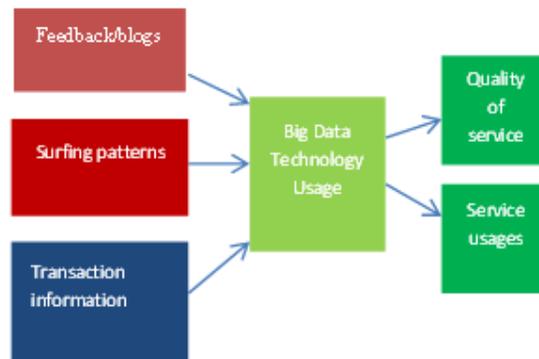


Figure 2. Model for studying key factors of customer behaviour using big data

3.4 Concepts

3.4.1 Feedback through blogs:

The social media has enabled a new medium through which customers express their opinion about a product or service. If a service/product does not meet the user expectations, users do not hesitate to express it over blogs. The data out of blogs are unstructured and the volume is huge.

Surfing patterns:

Understanding surfing patterns is one of the major factor for improving the quality of service. In big data context, big data should be able to identify the demands of a particular service. For example, if a customer is searching a travel destination and how frequently and how many customers are looking for the particular destination can predict the preference of customer. Service demand is discussed (A. Dan et al., 2004).

3.4.2 Transactions:

Online transactions are very common. Customers buy and sell using online web sites. Enormous amount of data is generated out of these transactions. Big data has the capabilities to process these data.

3.4.3 One Sample T Test:

One sample t test has been chosen in this model as the full population information is not available and to make sure the sample selected comes from a particular population. Big Data is evolving

and it is new in the industry. It is having its own class. Therefore one sample T test is justifiable to determine that the sample is selected from a population of known mean (μ).

3.4.4 Customer behaviour prediction:

Using big data technology if the quality of service and service usages are identified, then determining the behavioural aspect of customer can be established.

4. HYPOTHESIS

H. Big Data enables studying Quality of services by its ability to process blogs which is unstructured in nature.

- H1. Big Data enables studying Quality of services by enabling processing of huge volume of data.
- H2. Big Data enables studying Quality of services by enabling processing of huge volume of data in less time.
- H3. Big Data enables studying Quality of services by enabling processing of varieties of data

H. Big Data enables studying Service Usages by its ability to process browsing (surfing) pattern

- H1. Big Data enables studying service usages by enabling processing of huge volume of data.
- H2. Big Data enables studying service usages by enabling processing of huge volume of data in less time.
- H3. Big Data enables studying service usages by enabling processing of varieties of data

H. Big Data enables studying by its ability to process transactional patterns

- H1. Big Data enables studying transactional patterns by enabling processing of huge volume of data.
- H2. Big Data enables transactional patterns by enabling processing of huge volume of data in less time.
- H3. Big Data enables transactional patterns by enabling processing of varieties of data

5. SURVEY

QUESTIONNAIRE:

- a) I use big data for processing blogs to study customers' feedback on product/services.
 - 1) Strongly Disagree 2) Disagree 3) somewhat agree 4) Agree 5) Strongly agree
- b) I use big data because it helps in processing huge volume of data which is not possible using traditional system
 - 1) Strongly Disagree 2) Disagree 3) somewhat agree 4) Agree 5) Strongly agree
- c) I use big data because it helps in processing data faster than traditional approach.
 - 1) Strongly Disagree 2) Disagree 3) somewhat agree 4) Agree 5) Strongly agree
- d) I use big data because it helps in processing varieties of data such unstructured, structured, semi structured.
 - 1) Strongly Disagree 2) Disagree 3) somewhat agree 4) Agree 5) Strongly agree
- e) I use big data to study quality of services
 - 1) Strongly Disagree 2) Disagree 3) somewhat agree 4) Agree 5) Strongly agree

- f) I use big data to study service usages
 1) Strongly Disagree 2) Disagree 3) somewhat agree 4) Agree 5) Strongly agree
- g) I use big data to study customers' browsing behaviour
 1) Strongly Disagree 2) Disagree 3) somewhat agree 4) Agree 5) Strongly agree
- h) I use big data to study transactional data which provides insights about service usages.
 1) Strongly Disagree 2) Disagree 3) somewhat agree 4) Agree 5) Strongly agree

6. FINDINGS

6.1 BLOGS -> QUALITY OF SERVICE

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
blogs	48	4.48	.545	.079

One-Sample Test

Test Value = 4						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
blogs	6.087	47	.000	.479	.32	.64

Figure 3. One sample Test

Figure 3 shows that blogs processing has mean of 4.4 which implies that more respondents use big data to process blogs to study customers' feedback on product/services, this is further supported statistically that with 95% confidence interval blogs are processed by big data. Therefore the null hypothesis that big data is used to process blogs is accepted.

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 blogs & qos	48	.377	.008

Figure 4. Paired sample correlations

blogs * qos Crosstabulation

Count		qos			Total
		2	4	5	
blogs	3	0	1	0	1
	4	1	16	6	23
	5	0	9	15	24
Total		1	26	21	48

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal Gamma	.676	.164	3.076	.002
N of Valid Cases	48			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 5. Paired sample correlation and Gamma Testing

Figure 4 and figure 5 show that blogs and qos are related at a level of significance 0.05 using paired sample correlations and Gamma testing. This implies that with 95% confidence interval blogs analysis determines quality of services. This is because customers write their opinion, experience on a product or service using social media through blogs.

6.2 BROWSING PATTERNS -> SERVICE USAGES**Symmetric Measures**

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal Gamma	.045	.265	.169	.866
N of Valid Cases	48			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 6. Gamma Testing on browsing pattern vs. service usages

Figure 6 shows that with 95% confidence interval browsing patterns do not influence service usages. Therefore the null hypothesis such as browsing pattern is significantly related to service usages is rejected.

6.3 TRANSCATIONS -> SERVICE USAGES**Symmetric Measures**

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal Gamma	.353	.222	1.529	.126
N of Valid Cases	47			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 7. Gamma Testing on transaction information vs. service usages

Figure 7 shows that with 95% confidence interval transaction information do not influence service usages. Therefore the null hypothesis such as transactional data is significantly related to service usages is rejected.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
transactions	47	3.81	.900	.131

One-Sample Test

	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
transactions	-1.458	46	.152	-.191	-.46	.07

Figure 8. One sample test of transaction

Figure 8 shows that with 95% confidence interval transaction data have no significance for big data processing. The mean of 3.8 indicates transactions are somewhat processed using big data which implies that transactional data used occasionally.

6.4 BROWSING PATTERNS -> QUALITY OF SERVICE/PRODUCT**Symmetric Measures**

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal	Gamma	1.000	.000	20.993	.000
N of Valid Cases		48			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 9. Gamma test Browsing pattern vs. quality of service/product

Figure 9 shows that with 95% confidence interval browsing pattern is significant for big data processing to determine quality of service/product.

6.5 VOLUME, VELOCITY, VARIETIES**One-Sample Test**

	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
volume	-3.517	47	.001	-.417	-.66	-.18
velocity	-.405	47	.688	-.042	-.25	.17
varieties	.000	47	1.000	.000	-.21	.21

Figure 10. One-Sample Test. (volume, velocity, varieties)

Figure 10 shows that volume is having significance in big data processing with 95% confidence interval. Varieties and Velocity are not statistically significant with 95% confidence interval as big data is mostly used for unstructured data processing and velocity from users' perspective is

not that significant probably due to the fact that big data mostly used in batch processing which is slow compared to online data processing, basically batch processing would take hours vs. online processing would take seconds.

7. CONCLUSIONS

Big data plays a significant role in processing online data from internet. Data collected from internet source, social media logs are huge which is limited by the processing power of traditional tools. From the statistical findings it is evident that big data is meant for processing huge volume of data thus capable of eliminating the limitation of traditional tools. Further unstructured data from internet and social media logs such as blogs, browsing patterns found to be mostly used for big data processing. Velocity, Varieties were not found to be as important as volume. This may be because Velocity is implicit to processing huge volume of data. For example, processing 100 TB data can be processed using traditional tool in several days that implicitly tells that huge volume is a challenge because the processing time is more. Logs from social media, internet surfing are primarily unstructured in nature and there is evidence that blogs, browsing behaviour are significant in big data processing, therefore Varieties has not been significantly felt. But since big data is a data processing framework, varieties of data format can be processed in this. For online customers, big data can process to find quality of services/product from blogs, and browsing patterns. Transactional data was not found that significant for online users using big data.

ACKNOWLEDGEMENT

My sincere thanks to Dr. Nilay Yajnik, Head of Information Technology, SVKM's NMIMS, MUMBAI for his guidance.

REFERENCES

- [1] Chung-Hoon Park and Young-Gul Kim (2003), "Identifying key factors affecting consumer purchase behavior in an online shopping context", International Journal of Retail & Distribution Management, Volume 31, Number 1, 2003 . pp. 16-29 # MCB UP Limited. ISSN 0959-0552
- [2] Susan M. Keaveney, Madhavan Parthasarathy (2001) "Customer Switching Behavior in Online Services" ,Academy of Marketing Science, 29, 4; ProQuest CentralLIMAYEM MOEZ , KONG KOWLOON, KHALIFA, M. , FRINI, A. (2000), WHAT MAKES CONSUMERS BUY FROM INTERNET? A LONGITUDINAL STUDY OF ONLINE SHOPPING, SYSTEMS, MAN AND CYBERNETICS, PART A: SYSTEMS AND HUMANS, IEEE TRANSACTIONS ON (VOLUME:30 , ISSUE: 4)
- [3] Valarie A. Zeithaml,A. Parasuraman,Arvind Malhotra (2002), "Service Quality Delivery Through Web Sites: A Critical" Review of Extant Knowledge, Journal of the Academy of Marketing Science. Volume 30, No. 4, pages 358-371
- [4] Philip Carter. (2011), "Big Data analytics: Future architectures, Skills and roadmaps for the CIO", <http://www.sas.com/resources/asset/BigDataAnalytics-FutureArchitectures-Skills-RoadmapsfortheCIO.pdf>, retrieved 2012.
- [5] Ramesh Nair,Andy Narayanan (2012), "Benefitting from Big Data Leveraging Unstructured Data Capabilities for Competitive Advantage", Data and TechnologyPerspective,booz&co.
- [6] T. Ramayah, Jasman J. Ma'ruf, Muhamad Jantan, Osman Mohamad (2002), "TECHNOLOGY ACCEPTANCE MODEL: IS IT APPLICABLE TO USERS AND NON USERS OF INTERNET BANKING", The proceedings of The International Seminar, Indonesia-Malaysia, 14-15th October 2002
- [7] Sara Abbaspour Asadollah, Thiam Kian Chiew, "Web Service Response Time Monitoring: Architecture and Vaidation", Q.Zhou(Ed.):ICTMF 2011, CCIS 164, © Springer-Verlag Berlin Heidelberg 2011
- [8] A. Dan et al., "Web services on demand: WSLA-driven Automated management", © 2004, IBM SYSTEMS JOURNAL, VOL 43, NO 1, 2004
- [9] Wikipedia., "Big Data", http://en.wikipedia.org/wiki/Big_data, retrieved July 2014