The Study of Big Data Analytics in E-Commerce

Pavithra B [1], Dr. Niranjanmurthy M [2], Kamal Shaker J [3], Martien Sylvester Mani F [3]

Assistant Professor, Dept. of MCA, Jain University, Bangalore, India[2]

Assistant Professor, Dept. of MCA, MSRIT, Bangalore, India[2]

Student, Dept. of MCA, MSRIT, Bangalore, India[3]

Abstract: Big data is a compilation of huge data-sets that cannot be processed using conventional computing techniques. Big-data is not just only a data; slightly it has become a whole theme, which involves a variety of tools, techniques and frameworks. It refers to using complex datasets to drive focus, direction, and decision making within a company or organization. This is achieved by implementing applicable systems for gaining an accurate and deep understanding of the knowledge obtain by analysing the organization's data. In this survey paper we have discussed the different types of data held and its inverse usage for e-commerce and also different ways of providing security and safety for the data when it is used in bulky services, we also have discussed about the issues in big data with respect to e-commerce and how e-commerce can make use of applications over big data in an effective.

Keywords: Click-Stream Data, Web- Analytics, Predictive Analysis, Personalization, Dynamic Analysis.

I. INTRODUCTION

Big Data is a term which is always evolving. It is a large amount of structured and unstructured data that can be mined for information. These data sets are very large and complex that traditional data processing is not capable to process them. Big data is being used in many sectors.We shall see the influence of Big Data Analytics in changing the E-Commerce business, so that the business evaluated in the form of these E-commerce can benefit the most users n organizations from using Big Data, because there will be information of the data collected on day to day bases.

Many big retailers values this data's information and helps them for predicting the user interests and provide their customers relative and interested searches when they shop on their site, so that they attract the customer by giving the required and relevant searches of the products or items. These preferences are all generated from Big Data analytics. Big data consists of two types of data one is structured and the other one is unstructured.

The structured data deals with simple basic and regular data like name and address. Unstructured data deals with numerous data which is retrieved from places like social media and includes videos. This information plays a vital role to e-commerce businesses, which can better serve the customer. E-Commerce businesses and retailers which operates online, to give a better service for this competitive and fast environment, attracts the customers by online advertisements. In accordance with the online approaches are split up with very specific analytics tools for separate data and information's. Decisions will be optimized because of the huge volume of data obtained from multiple sources and in different formats.

These are the focus of Big Data where it gives out superior, data driven outputs.

Utilization of big data in ecommerce is given below
- The most vital role of big data is to provide a better experience for the customer when they make use of the website and also try to satisfy the user needs by giving the relevancy search.
- To predict the user interest and behaviours Predictive analytics is been used so that to provide the required products of user interest and satisfy his demands by giving proper online advertisements based on the predictive data.
- Big data is been used for personalization which in turn personalizes the users information such as mail id and address in order to increase the rate of conversation.
- By making use of real time analysis prices are changed in order to compete with other retailers.

Data retrieved from different devices, sources and application are emerged into big data. The data is retrieved from Social Media such as Facebook; twitter which holds the data posted by millions of people across the world such data is called as social media data. Stock Exchange data which deals with data about the decisions made on buying and selling a share of different companies by the customers. Transport data includes the information about the distance and availability of a vehicle. Search Engine data which retrieve lots of data from different databases. The data thus obtained is basically of 3 types Structured-data, Relational-data and Unstructured-data such as word, pdf, text and Semi-Structured data such as XML data.

II. RELATED WORK

Big Data plays an important role if the business is made online to meet the consumers demand and their preference. This study clearly shows that Data Analytics could be used to assist business oriented organizations to improve their performance and give out the best services for the customer in a reliably way. Data Analytics also provides personalization of the products with respect to the customer's interest and their aspects towards pricing. This could assist with customer retention and acquisition alongside improving brand image and loyalty. In addition to increased sales opportunities for companies, it also provides useful insights regarding process improvement across the seller and buyer networks. Data analytics is also helpful for organizations in capitalizing on their data since costs are saved through self-service options. Finally, it highlights where improvements need to be done in terms of sales performance improvement as discussed in [1].

The author Shahriar Akter and Samuel Fosso Wamba has discussed in the paper [2] the different types of big data that are used in e-commerce. The first type is online transactional data which deals with goods selling and providing various other services online such as amazon, eBay, Expedia or provide services through on going transactions like Netflix, Match.com etc. Where these truncations make use of data which is further broadly classified as (a) transaction or business activity data (b) click-stream data (c) video data and (d) voice data. Using customers browsing and transactional behaviour customers can be tracked by the e-commerce data which plays as a key role in personalizing the offers and their interest pattern.

Year	Growth in the number of e-commerce customers worldwide (in millions)	Growth in e-commerce sales per customer worldwide (in US$)	Growth in big data analytics (BDA) market worldwide (in billions)
2011	792.6	1162	7.3
2012	903.6	1243	11.8
2013	1015.8	1318	18.6
2014	1124.3	1399	28.5
2015	1228.5	1459	38.4
2016	1321.4	1513	45.3

Source: Adapted from emarketer (2013) and (Piatetsky. 2014)

Fig 1 : Global growth in e-commerce and big data analytics

Indicative transactional market services are been achieved by top most vendors like Amazon and eBay for their innovative platforms and reliable product recommender systems. Most important firms like Google, Amazon and Facebook lead to the development of web analytics, cloud computing, and social media platforms. When compared to traditional transactional records and the ecommerce

transactional record the e-commerce data are less structured and contains more customer opinion and their behaviour interest which is been discussed in [3].

TABLE I
APPLICATIONS FROM BIG IMPACT TO BIG DATA

Factors	E-Commerce and Market Intelligence	Security and Public Safety
Applications	• Recommender systems • Social media monitoring and analysis • Crowd-sourcing systems	• Crime analysis • Computational criminology • Terrorism informatics • Cyber security
Data	• Search and user logs • Customer transaction records • Customer generated content	*Criminal records *Crime maps *Criminal networks News and web contents *Terrorism incident databases * Viruses, cyber attacks, and botnets
Analytics	• Association rule mining • Database segmentation and clustering • Anomaly detection • Graph mining • Social network analysis • Text and web analytics • Sentiment and affect analysis	• Criminal association rule mining and clustering • Criminal network analysis • Spatial-temporal analysis and visualization • Multilingual text analytics • Sentiment and affect analysis • Cyber attacks analysis and attribution
Impacts	Long-tail marketing, targeted and personalized recommendation, increased sale and customer satisfaction	Improved public safety and security

Generally web mining leads to some issues when it is applied on e-commerce. Some of the issues are discussed below. The vital parameters which are used for implementation in retailer's are automatic timeouts of user session need to be considered as one of the issues because without interactive time systems the data mining intelligent algorithms can't be applied and the user end applications need to be based the organizations constraints

which varies from organization to other organization and the type of users. The other issue is the larger clients are made to lose their shopping carts due to time outs because a standard time session will be fixed in the data mining application oriented algorithm so causes the clients to turn out their sessions in an unmannerly fashion. The other issues is Generating logs for millions of transactions which is again a costly exercise. It may be worth in generating the appropriate logs for particular user's sessions by considering random samples, but these random samples doesn't contain the complete details thus techniques need to be designing so that mandatory information should be recorded and access it in an intelligent way.

Another issue is designing the user interface forms because the user form has to be designed very intelligently that it should The users should be made to enter these values, since it has found that several users left the default values untouched. Mining data at the right level of granularity is essential .Otherwise, the results from the data mining exercise may not be correct, an example is been discussed in [4].

Big data which comprises of structured and unstructured data provides more opportunity for electric oriented companies. In this case data is been collected from the data is collected from customer's internal processes, vendors, markets and business environment. [5] in this paper data mining process for e-commerce with three algorithm is been discussed: association, clustering and prediction. Some of the benefits of data mining to e-commerce is been discussed in terms of customer client relationship, basket analysis, planning the business criteria's and marketing segment. Some of the challenges of data mining is been discussed for making model comprehensible to the business users which in turn supports the business users , slow changing data dimensions, transformation of data and building the models all these are accessible for users. Indicative transactional market services are been achieved by top most vendors like Amazon and eBay for their innovative platforms and reliable product recommender systems. Most important firms like Google, Amazon and Facebook lead to the development of web analytics, cloud computing, and social media platforms. When compared to traditional transactional records and the ecommerce transactional record the e-commerce data are less structured and contains more customer opinion

There has been growing emphasis on big data analytics (BDA) in latest environment. But it's been poorly explored as a concept and works has not been done much on the basis of its experimental improvising the business which will be beneficial for the organization The paper presents an interpretive framework that explores the definitional aspects, distinctive characteristics, types, business value and challenges of BDA in the e-commerce

landscape. The paper also involves discussions regarding future research challenges and opportunities in theory and practice. Data retrieved from different devices, sources and application are emerged into big data. The data is retrieved from Social Media such as Facebook, twitter which holds the data posted by millions of people across the world such data is called as social media data. Stock Exchange data which deals with data about the decisions made on buying and selling a share of different companies by the customers.

TABLE 2: GLOBAL GROWTH IN BIG DATA AND DATA ANALYTICS

Year	Growth in the number of e-commerce customers worldwide (in millions)	Growth in e-commerce sales per customer worldwide (in US$)	Growth in big data analytics (BDA) market worldwide (in billions)
2011	792.6	1162	7.3
2012	903.6	1243	11.8
2013	1015.8	1318	18.6
2014	1124.3	1399	28.5
2015	1228.5	1459	38.4
2016	1321.4	1513	45.3

Source: Adapted from emarketer (2013) and (Piatetsky 2014)

The study includes the analytics used in get to know many new things about big data and business analytics, through which we can get many opportunities in learning their applications used in real time environments. The major applications of knowing big data is that sellers learn to predict the buyer's strategy and learn all the optimal strategies to get the good result and best services according to their search interest. As discussed in paper [7] an experimental data is been collected and intelligent algorithm is applied in conducting the negotiation process. There are huge amount of data which is produced and gathered, for which current computer techniques have made possible to store and process the data. Business oriented organization is very much interested in tracking their data through big data analytics to increase their business performance.

There are many big data analytical tools available like predictive analytics, descriptive analytics, and survival analysis now a day. Linear regression, logistic regression, neural networks, and support vector machines are few methods and techniques which are already designed. These methods are applied in various business areas like Retails, Marketing, Detection of frauds, Customer relationship management, Social network analysis etc.[8]

Marketing process can be automated using any of the digital marketing tools which also help in extracting the data for analysis. One of the latest trends in marketing is making use of big data in marketing. (Lee 2007; Chaffey, Smith 2008; DavidavicienC 2012; Wang et al. 2009).

Most of the scientific studies are dealing with the evaluation of the effectiveness of the online marketing or advertising campaigns and the quality, or the effectiveness, of the web sites (Cao, Zhang, Seydel 2005; DavidavicienC, Tolvaišas 2011; PabedinskaitC, Davidavicius 2012); however, little attention has been paid to the identification and description of the target e-spaces and the challenges related to it (PabedinskaitC, Davidavicius 2012; DavidavicienC et al. 2014) Researchers mainly focus on the identification of the online users behaviour (Wright 2006; Sinclair 2007; Li, Li 2008; Kawase, Herder 2011; Zeb, Fasli 2011; PabedinskaitC, ŠliažaitC 2012); They have discussed various motivational determining the method to motivate the users for instance the use of tools for online marketing their browsing skills and behaviour of different generational users are all important part of big data.

In the presence there is no perfect definition for the word Big Data, the most widely used definition is in term of three terms like volume, velocity and variety also referred as 3 –Variety refers to the heterogeneous nature which comprises of made up of structured and unstructured data, Velocity relates the speed at which data is captured, and Volume refers to the size of data.   (Russom, 2011) (Edosio, 2014). Because of these characteristics it is not possible to manage and analyse big data using traditional data base, so using special tools and technologies like hadoop and file systems big data can be managed well. Managing data can involve process like storage, real time analysis ad in accordance to this a special data mining algorithms like clustering algorithm and machine learning algorithm can be used (Fan et al., 2013). The study of data analytics is been categorised into three types, they are as follows:

- Social Media Analytics: considers the social media data which are generally huge in size or large volume of data (Hea et al., 2013)
- Predictive Analytics: Considers to use data which are used for forecasting the consumer behaviour and tends (Mosavi & Vaezipour, 2013).
- Mobile Analytics: This considers data generated from mobile phones, tablets and other (Li & Du, 2012). [10]

At present generation the e-commerce oriented companies should analyse the need of large volume of data and products, customers, transactions and aware that scale matters and take into consideration many other implications related to successful implementation of big data technology.
Big data applications in e-commerce create new opportunities for quick personalization of products and services according to customers' needs, improve the process of integration between partners' value chains and finally, dramatically reduce co sets. [11]

The e-commerce is slowly addressing security issues on their internal networks. There are guidelines for securing systems and networks available for the ecommerce systems are personalized for implement. Training the customers on the security issues is still a growing stage, but will prove to be the most critical element of the e-commerce security architecture. Trojan horse programs have the greatest threat against e-commerce which was launched against client server. These code applications can be installed in a remote computer by which privacy has become a major concern for the consumers to find out theft and other such frauds for the better improvements of the user services and also for the security of their data which will be shared on the big data, this was implemented as in paper [12].

[13] discuss the number of transactions and size of the data involved in that While E-commerce is often considered to be online transaction were the payment will be made as online which involves of small transaction to bigger transaction. The size of the transaction also depends on the economics of business transaction. The delivery costs also sometimes matters, suppose if one item is purchased and shipped overseas it is more costly than buying the goods in bulk or more or less it will be the same cost so the cost measures also plays an vital role in increasing the business firms and organization which will be reflected on both the consumers as well as the businesses organization especially when they deal with online purchases.

Big data which comprises of structured and unstructured data provides more opportunity for electric oriented companies. In this case data is been collected from the data is collected from customer's internal processes, vendors, markets and business environment. [5] in this paper data mining process for e-commerce with three algorithm is been discussed: association, clustering and prediction. Some of the benefits of data mining to e-commerce is been discussed in terms of customer client relationship, basket analysis, planning the business criteria's and marketing segment. Some of the challenges of data mining is been discussed for making model comprehensible to the business users which in turn supports the business users , slow changing data dimensions, transformation of data and building the models all these are accessible for users. Indicative transactional market services are been achieved by top most vendors like Amazon and eBay for their innovative platforms. The recent technology deals with popular techniques and more applications which support greater quantities of data to keep them growing and safe. the unstructured. Unstructured social media data, distributed in multiple computer systems, act as sources of both relevant and irrelevant information. References of an e-commerce site in social media applications act like advertising and may generate traffic and improve revenue.

Both traffic and revenue are measured by the Analytics Application in order to evaluate the impact factor of each medium. Social media is becoming popular day by day and allow user to express their opinions towards the shared internet as the goods are dealt on the opinion before purchasing of goods, based on seeing the review the customer opinion keeps changing for every item there will be a feedback or review of the comments so as to purchase the category oriented products, in the same way the corporate want to monitor the peoples interest and their area of interest so as invest more on that particular item. .

This paper reviewed about world are realizing that e-commerce is not just buying and selling over Internet, rather it is improve the efficiency to compete with other giants in the market. Their opinions on specific topic are inevitably dependent on many social effects such as user preference on topics, peer influence, user profile information.[16]

III. BIG DATA FOR E-COMMERCE

Online buyers and sellers are making use of big data for better shopping, selling the goods and gaining good customer relationship, giving better services and customer satisfaction in order to generate more sales. This is how big data benefits e-commerce companies:

- Distribute something more Valuable.
- Distribute More Personalized Interactions.
- Build Accurate Predictions.
- Decrease the shopping cart rejection Rate.
- Supply Customers with a better experience.

Example of Big data Survey in Online shopping when it comes to online grocery sales, Amazon is already No. 1, bagging a 22% share in 2015 vs. Wal-Mart's 13%. This way company can find their status to work on particular store to grow with the competition and also can easily find and compare with others online stores.

Fig 2: Amazon Gross merchandise volume

A. Big Data Challenges
The major challenges associated with big data are as follows:
- Capture related information's
- Storage space
- Searching of related information's
- Sharing
- Transfer
- Analysis
- Presentation

Big data are used to give more accurate analysis, which leads to give out the best decision, compete with other organization, cost reduction, risk reduction and effective customer services.

B. Application of Big Data Analytics in E-Commerce
Big data is widely used across the word for various reasons its greatly changing the e-commerce landscape
In E-Commerce, most of their data would fall under two categories – structured and unstructured. Their structured data is their regular data . The more valuable piece of data which is not captured but is extremely important is the unstructured data

C. How can E-commerce utilise big data
- Builds a better customer relationship and gains their experience from big data, the customer gets an expertise opinions according to their interest and this data can be used to satisfy the customer
- Predictive analytics is used to predict the customer choice and his area of
- Personalisation involves using Big Data to personalise emails and increase conversion rates.
- By taking the expertise opinion Pricing can be changed constantly to keep up the competition

IV. CONCLUSION

As the Big data is been used in many sectors ,it widely influences E-commerce services and plays a major role in business making decisions. The usage of big data has enormously grown in e-commerce, many big retailers values this data's information and helps them for predicting the user interests and provide their customers relative and interested searches when they shop on their site, so that they attract the customer by giving the required and relevant searches of the products or items.
Using the related information from this survey paper the researchers can come up relevant and challenging systems to extending the benefits of big data usage toward ecommerce for both the customers as well as the retailers. In this competitive and fast environment customers mostly go with the online advertisements or through search engines by decreasing the inefficiency of the real time markets. Companies can find their status to work on

particular product to grow with the competition and also can easily find and compare with others online stores. In our survey paper we have concentrated on the types of data which are taken as input and output with the relevant processing steps. There are numerous technologies available at the core of big data processing. The discussion of various papers states that Hadoop server as one of the best technique for providing scalable and inexpensive platform for processing the data. The researchers say that online retailers make uses of big data for better shopping experience, gain customer satisfaction and generate more sales. Big data have a positive impact on outcomes and productivity. Area such as record linkage, graph analytics and machine learning has shown that it's been critical help in the applications. HPCC Systems open source offering provides a single platform that is easy to install manage and code. The built-in analytics libraries for Machine Learning make it easy for users to analyze Big Data.

Our survey paper helps identifying various applications of big data into e-commerce so that we can know the importance of big data, it makes better understanding of Usage of big data and its components. We also discuss by the survey made on the issues related to E-Commerce if big data is not in related to that, so that researchers can work on the issues related to big data and extend their work on that. The ultimate challenge of Big data analysis is to generate business value from their explosion of big data. We also have discussed the major challenges associated with big data so that we extend our research in finding solution to one of the challenges associated with that. Researchers can get information about the issues regarding big data and major challenges associated with that. So they can get concise information about big data which helps them in extending their research work on big data related to e-commerce.

REFERENCES

[1] Ahmad Ghandour-"Big Data Driven E-Commerce Architecture",International Journal of Economics, Commerce and Management,ISSN 2348 0386 Vol. III, Issue 5, pp:940-947,May 2015.

[2] ShahriarAkter and Samuel FossoWamba-"Big data analytics in e-commerce: A systematic review and agenda for future research",Electronic Markets 26 173-194, 2016.

[3] HsinchunChen,Roger H. L. Chiang and Veda C. Storey-"Business Intelligence And Analytics:From Big Data To Big Impact",MIS Quarterly Vol. 36 No. 4, pp. 1165-1188,December 2012.

[4] Arti, SunitaChoudhary and G.N Purohit-"Role of Web Mining in E-Commerce",International Journal of Advanced Research in Computer and Communication Engineering,ISSN (Online) : 2278-1021:ISSN (Print) : 2319-5940 Vol. 4, Issue 1, pp:-251-253,January 2015

[5] Mustapha Ismail, Mohammed Mansur Ibrahim, Zayyan Mahmoud Sanusiand MuesserNat,"Data Mining in Electronic Commerce: Benefits and Challenges, "International .Communications, Network and System Scien-ces,pp:501-509,December 2015.

[6] ShahriarAkter and Samuel FossoWamba,"Big data analytics in E-commerce: a systematic review and agenda for future research",pp:173–194,March 2016

[7] Hasan Al-Sakran,"B2C E-Commerce Fact-Based Negotiation Using Big Data Analytics and Agent-Based Technologies "International Journal of Advanced Computer Science and Applications, Vol. 5, No. 12,pp: 30-37,2014

[8] O. Liu, W.K. Chong, K.L. Man and C.O. Chan, "The Application of Big Data Analytics in Business World" Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol II, March 16 - 18, 2016

[9] ArnoldinaPabedinskait-, Vida Davidavicien- and PauliusMilišauskas,"Big Data Driven E-Commerce Marketing"8th International Scientific Conference "Business and Management 2014",ISBN online 978-609-457-651-5,pp:645-654,May 15–16, 2014.

[10] UyoyoZinoEdosio,"Big Data Analytics and its Application in Ecommerce "Research Gate, Conference Paper ,April 2014.

[11] G. Ilieva, T. Yankova and S. Klisarova,"Big Data Based System Model Of Electronic Commerce" Trakia Journal of Sciences, Vol. 13, Suppl. 1, pp 407-413, 2015

[12] Niranjanamurthy M, DR. Dharmendra Chahar-"The study of E-Commerce Security Issues and Solutions" ISSN (Online) : 2278-1021 International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 7, July 2013

[13] Niranjanamurthy M. et. al.-"Analysis of E-Commerce and M-Commerce: Advantages, Limitations and Security issues" ISSN (Online) : 2278-1021 International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 6, June 2013

[14] Radoslav Fasuga, Jiří Rýdel, Eduard Kubanda -"Gloffer — Framework for the implementation of portal solutions for E-commerce" ICETA-2015 INSPEC Accession Number: 16284512, IEEE 26-27 Nov. 2015

[15] Constantine J. Aivalis, Kleanthis Gatziolis, Anthony C. Boucouvalas-"Evolving analytics for e-commerce applications: Utilizing big data and social media extensions" 2016 International Conference on Telecommunications and Multimedia (TEMU), DOI: 10.1109/TEMU.2016.7551938 IEEE 25-27 July 2016

[16] Nitu Kumari, Shailendra Narayan Singh-"Sentiment analysis on E-commerce application by using opinion mining"6th International Conference on Cloud System and Big Data Engineering (Confluence), ISBN: 978-1-4673-8203-8 IEEE 14-15 Jan. 2016