International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

RESTRUCTURE THE PLATFORM OF BUSINESS INTELLIGENCE BY COMBINING BIG DATA WITH TRADITIONAL BUSINESS INTELLIGENCE

Dr. Priyanka Srivastava
Department of Commerce, Kalinga University, Raipur, Chhattisgarh

Naseem Akhter
Research Scholar, Kalinga University, Raipur, Chhattisgarh.

ABSTRACT

The principle of 'knows your client' has always underpinned e-commerce. The use of big data analytics in e-commerce has made it possible to personalise purchasing for each customer. To micro-segment customers and obtain granular insights, big data derived from transaction or business activity, click stream, video, and speech is analysed. By merging big data and traditional business intelligence, this study reimagines the framework platform for business intelligence. The method of data collecting is the emphasis of this research. Using an enterprise's e-commerce data as an example, this study employs the K-Means method in clustering analysis to group consumers in order to achieve the goal of tailored marketing for various customers.

Keywords: Big Data Analytics; Business Intelligence; Personalisation; Challenges; business

I. INTRODUCTION

The way E-commerce is done has changed as a result of increasing technical advances and technology. Different advances have resulted in different data genres and rules for data transmission and exchange throughout the global investment and user society. Despite several attempts to standardise, data challenges persist, preventing e-business from attaining its full potential. In today's technology-driven economy, dealing with business operations with diverse data sharing kinds, language, and frameworks presents several challenges. The rise of social media, the widespread use of portable devices, and the sheer scale of global corporations are changing the way people study products and make purchases. The rise of quicker internet connection and pervasiveness fuels the expansion of information. Nowadays, almost everything is done online; individuals purchase and sell knowledge over the internet. Ecommerce retailers used the internet to offer items and services, increasing sales and brand awareness. Nonetheless, with the rise of big data and the development of more educated / information-driven means for firms to engage with consumers, there will be more educated / information-driven methods for companies to connect with customers. So that Ecommerce providers / enterprises may harness / analyse the massive amounts of data (Big Data) generated by "Electronic Data Interchange (EDI)" to acquire a better knowledge of customer preferences.

II. BIG DATA ANALYTICS

Big data analytics refers to the use of larger datasets and advanced analytical approaches. As a result, big data analytics is concerned with "big data" and "analytics," as well as how the two intersected to produce business analytics. Variety, Volume, and Velocity are three significant properties of big data. Variety refers to the different levels of complexity (organised and disorganised datasets), velocity refers to the rate at which data is processed, and volume

refers to the amount of data. Because of these characteristics, it is hard to correctly process and analyse enormous volumes of data using traditional technologies. As a result, information may be recovered from data when certain data extraction techniques (such as deep learning and hierarchical clustering) are applied into the computational framework for big data. Today's e-commerce landscape is awash with numerous analytics that are utilised to solve business problems. Big data nowadays permits e-business enterprises to minimise expenses and develop benefits with ease by utilising cost-effective space and capabilities, as well as cutting-edge analysis tools.

Big Data Analytics and E-Commerce

Big data analytics (BDA) enables e-commerce businesses to make better use of data, increase conversion rates, improve decision-making, and encourage customers. BDA may help internet companies by increasing consumer transaction expense efficiency (e.g., customer-seller contact internet), administrative transaction price effectiveness (e.g., Flipkart's system performance suggestion algorithms), and time cost effectiveness (e.g., scanning, negotiating and after purchase tracking). The most difficult issue for BDA is to make corporate benefit from such a massive data influx. BDA is a high-performance corporate management talent that allows them to achieve business criteria such as finding steady and lucrative clients, determining the best pricing, detecting quality problems, and deciding on the lowest stock level possible. E-commerce enterprises are one of the fastest categories of BDA purchasers due to the requirement to stay on top of the sector. The challenge with the BDA architecture is dealing with various forms of data in order to generate valuable feedback in order to maximise transactions. The relevance of data analysis in numerous operational domains of e-business, such as advertising, information technology administration, development and service, and banking, is shown by the sheer volume of educational and industrial work. When clients register in an e-business, a huge amount of client-related data becomes available. This information is of great relevance to decision-makers in the firm. Business intelligence (BI) is a technology-driven strategy for data analysis and presentation that helps managers, administrators, and other international customers make better strategic choices. Maintaining current key applications, such as "Business Intelligence (BI)," may result in improved decision-making, more revenue, more efficient marketplaces, and lower expenses. BI is defined as providing decision-makers with meaningful information insight via the utilisation of a variety of information sources and structured and unstructured data. Market conditions were forecasted, rivals were evaluated, and client-specific market research was conducted using BI. The following is a description of the three-layer architecture of business intelligence in e-commerce:

- **Level 1**

Information: data sets received from the e-commerce platform include customer records, browsing history, feedback, and social networking information. The interference is removed from the findings using the input filtering device.

- **Level 2**

Hadoop: The data is gathered and then dumped into HDFS (Hadoop Distributed File System), after which data reduction is applied to the information from HDFS (Hadoop Distributed File System). The mapping phase displays the collection of many other data sets, while the reduction step provides the underlying data result.

- **Level 3**

Data Analysis and Business Intelligence: After performing map reduction, the data is ready to be evaluated. The R tool is used to analyse the data in this stage. R simplifies a variety of mathematical operations and generates a variety of visual displays.

Big Data Analytics Challenges

People with specific expertise, partners, dependable infrastructure, and innovative ideas are all required when using analytics. Despite the design, cleaning, and distillation required to conform with the syntax and semantics of the standard business object model, e-business representatives must be able to synthesise data in a timely way. To know that there is a high throughput between the time data is created and the time data is accessible for use. The key concerns include information confidentiality, security and piracy, as well as legal issues. Verifying the quality and correctness of information is a difficult task. Complexity of combining many traditional data kinds with traditional data for examination by multiple systems. The following are some of the challenges:

- Data recording and filtering
- Information inclusion, grouping, and composition
- Query handling, simulation of results, and analysis
- Interpretation
- Dynamic and changing interactions
- Difficulty of data, difficulty of processing, complexity of method

To overcome the challenges, e-businesses must embrace Big Data, which is driven by a perfect blend of technology, strategic planning, and sociology. The ability to recognise and respond to meaningful data requires real-time processing of massive data sets. When big data is combined with an opportunity, it creates a challenge. The subject under discussion is how to control which information to focus on and which to ignore, because focusing on the wrong data sets would not generate the intended outcomes for the specific business objectives.

III. METHOD OF DATA COLLECTION

In the context of massive data, data gathering methods are divided into three categories: system log collection, network data collection, and data interface collection. The logging subsystem of the device collects log data and can create log messages as needed. REST APIs are commonly used to get data information in commercial data APIs. Web crawler technology is mostly used to collect network data. Its main premise is to imitate a browser accessing a Web server via a unified resource locator URL address, obtaining the Web server's authorization, returning to the original page, and parsing the contents.

Traditional crawler technology may have issues, therefore when the need arises, targeted crawler technology tailored to crawl online resources emerges. Focus crawlers browse web-related links on the Internet carefully to collect the information they require based on crawling goals (using an ecommerce sales theme). The goal of a focused crawler is not to cover all of the web pages. Instead, it focuses on web sites that are relevant to certain subjects and provides data resources for user queries that are topic-oriented.

IV. APPLICATION CASE OF AN ENTERPRISE E-COMMERCE DATA

K-Means algorithm

K-means is a widely used clustering method D entities are divided into N clusters using this

algorithm. This guarantees that the cluster's similarities are as high as feasible while the cluster's similarities are as low as possible. The following is how the K-means algorithm works:

- Randomly select N data points as the centroid;
- Calculate the distance between each data point in the data set and the centroid, and aggregate all the data points in the data set into N clusters;
- Iteratively calculates a new centroid using the N sets of data points calculated in step 2;
- Repeat steps 2-3 until the distance between the final centroid and the previous centroid is small (satisfactory convergence);

The K-MEANS method is based on two key concepts: centroid and distance. Centroid is defined as the centre of a set of data with similarities and may be thought of as a sample or a data point A in a data set. Because the method chooses any item at random as the centroid of the first grouping and depicts the clustering results, centroid selection has a significant influence on clustering outcomes. Of fact, with randomly partitioned data sets, this finding is frequently absurd. The concrete correction of centroid takes numerous rounds to approach the desired clustering results: items with similarities are sorted into groups, all of which have the same centroid. Furthermore, the end result is not always expected due to the unpredictability of the first centroid selection, therefore it takes several iterations to acquire the starting centroid randomly in each iteration until the final clustering results fit the expectations.

In reality, distance is a measure of similarity. Manhattan distance, Euclidean distance, Minkowski distance, Chebyshev distance, and other distance formulae are common. Euclidean distance is the most often used distance formula in clustering analysis because it is obvious and simple to compute, and it is utilised for coordinate migration and alter rotation of object points. Finally, because the distance value stays constant, the object similarity may still be determined by the item's initial similarity. If D (x, y) is the distance between objects a and b, then d (x, y) must meet the three criteria listed below:

- Non-negative: that is, d (x, y) 30 is constant; if and only if x = y, d (x, y) =0
- Symmetry: d (x, y) = d (y, x)
- Triangle inequality: any object a, b, c has d (x, y) +d (y, z) 3d (x, z)

Application Analysis of Electronic Commerce Data in the Enterprise

Independent data is of limited use in the age of big data. The objective is to use data to forecast future trends as well as to uncover hidden information. Many Chinese herbal decoction businesses keep up with the times, and there are equivalent stores selling Chinese herbal decoction on the electronic commerce website, so a big number of clients have acquired consumption records. Customers can be divided into groups based on their consumption patterns, and various groups of consumers can get personalised marketing based on their preferences. Differentiated services for diverse categories of clients are made possible by customer categorization. It also allows businesses to notice slight changes in the market and customers in real time and alter their strategy accordingly.

R (Recency) reflects the time span between the customer's most recent transaction to the present moment in the RFM model, which is a frequently used multi-factor customer categorization technique. The number of times

a client cooperates with a company (i.e., purchasing behaviour) during a certain period of time is represented by F (Frequency). M (Monetary) denotes the amount earned by a transaction between customers and businesses over a set period of time. The absolute quantity generated by customers is used by RFM to determine the worth of consumers.

Now we crawl relevant data from a Chinese herbal medicine decoction piece ecommerce website, clean and gather the original data using specific data processing methods, and then analyse the consumer data (2000 pieces). R denotes the amount of time since the last purchase of Chinese herbal medicines. F stands for the frequency with which traditional Chinese medicine is purchased. M is the total amount of money spent on a certain platform. As seen in Table 1, some genuine data is intercepted.

Table 1: Data related to Consumer

Customer serial number	**R**	**F**	**M**
1	25	6	334.19
2	3	5	759.23
3	6	16	1383.41
4	3	11	3280.82
5	15	7	154.68
6	18	6	501.43
7	5	2	1721.89
8	28	2	107.21
9	25	9	973.38
10	2	21	764.54
11	17	2	1251.4
12	26	3	923.31
13	17	11	1011.20
14	32	16	1847.59
15	5	7	1669.52

Various data items have different numerical sizes and units; thus they can't be utilised to directly participate in the computation. For example, the total quantity of product M

purchased by customers is a huge numerical attribute, with a unit of more than 100, and the frequency of purchasing items over a given period of time is frequently minimal, with little influence on the amount of consumption. To make these qualities work, they must be compared to their respective ranges to guarantee that there are no differences in units and values, allowing this standard data to be utilised directly for calculations in the future. The data in this study is processed using the normalised processing approach. Table 2 shows a subset of the processed data from the year 2021.

Table 2: After normalization Consumer-related data

Customer serial number	R	F	M
1	0.000889 179	0.000329254	0.000168388
2	9.8798E-05	0.000274377	0.000382509
3	0.000131728	0.000878009	0.000697001
4	9.8798E-05	0.000603634	0.001652971
5	0.000461057	0.00038413	7.79183E-O6
6	0.00062571	0.000329255	0.000252613
7	0.000164662	0.000109749	0.000867571
8	0.000856251	0.000109748	5.40012E-05
9	0.000691586	0.000493879	0.000490411
10	6.58653E-05	0.00115241	0.000385211
11	0.00049399	0.000109756	0.000630501
12	0.000856251	0.0001 64628	0.000465182
13	0.000559854	0.000603629	0.000509471
14	0.00098798	0.000878009	0.000930891
15	0.000164663	0.00038414	0.000841131

The number of clusters is set to 3, the maximum number of iterations is 3, and the distance function is Euclidean distance using the K-Means method. The findings for each cluster may differ since the initial centroids are random. The detection clustering findings are

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

essentially the same after several trials, therefore this clustering result may be utilised to study the cluster characteristics of cluster users and conduct group targeted marketing. As demonstrated in the consumer group in Figure 1, the K-Means method clustering produced the following image of groups one, two, and three.

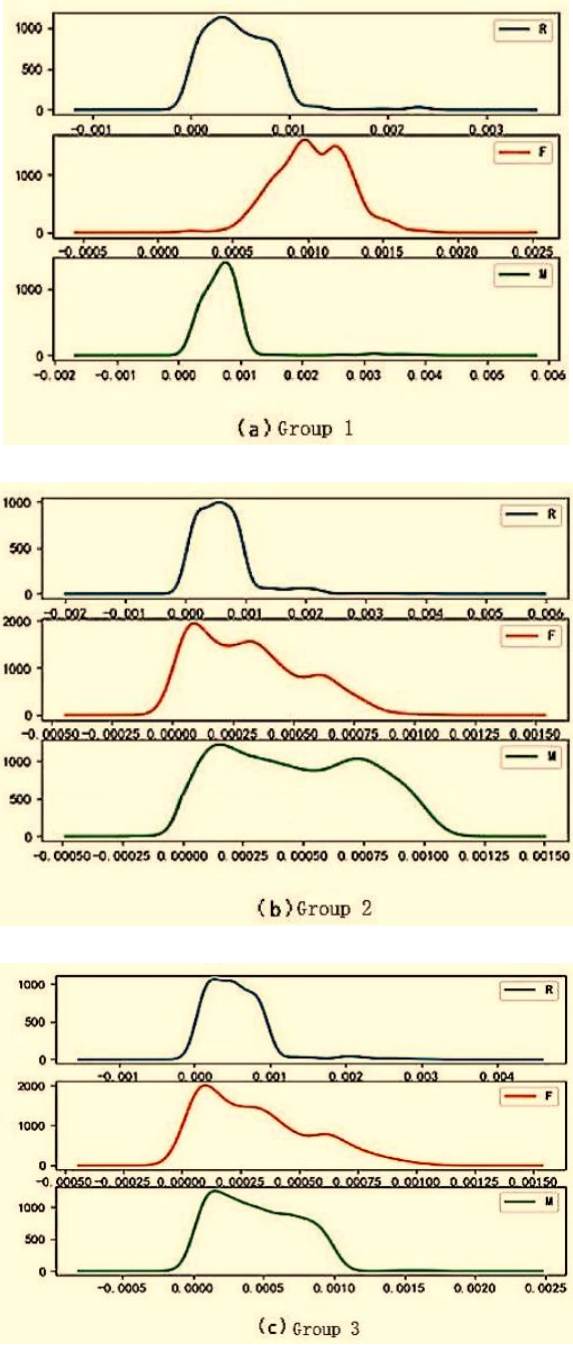

Figure 1. Graphs showing different groups of consumers

Group 1: The time interval (R) was shorter and the total quantity of consumption (M) was higher the previous time these client's spent money. They are not only the best customers, but they are also potential consumers. They provide a significant contribution to the organisation, yet their share is insignificant. Enterprises should prioritise allocating resources to differentiated management and one-to-one marketing in order to increase client

loyalty and satisfaction, as well as optimise their high consumption levels.

Group 2: These clients have a general buy frequency (F), a short time interval (R) between their last e-commerce website consumption, and a moderate total consumption (M). The reasons for the fall in consumption differ, and their consumer value changes are very unknown. As a result, it's critical to stay on top of client information and communicate with them as soon as possible. Enterprises can forecast changes in client consumption behaviour based on recent consumption intervals and frequency, focus on these customers, and implement specialised marketing initiatives to extend their life cycle.

Group 3: These clients' purchase frequency (F) is average, the number of days between consumption (R) is moderate, and the total quantity consumed (M) is low. They are the general public and low-value clients of traditional Chinese medicine decoction businesses, and they can only be purchased when traditional Chinese medicine decoction pieces are on sale.

V. CONCLUSION

The ability of e-commerce enterprises to provide personalised services has been altered by information and communication technology. BDA has shown to be quite useful in the development of business intelligence and in advising organisations on how to make better business decisions. Tracking and intense information searching are required to make the most of big data technologies and to take benefit of it. Great client interactions, better market efficiency, productivity improvements, and digital skills are all advantages of using e-business big data. Because of its power to transform business operations as well as the ability to leverage a wide range of efficient decision-making tools and services, Big Data has gotten a lot of attention. The use of Big Data technologies in the workplace makes it simpler to discover, store, and analyse complicated data.

REFERENCES: -

1. B. N. Silva, M. Diyan, and K. Han, “Big Data Analytics,” in SpringerBriefs in Computer Science, 2019.
2. P. J. Wu and K. C. Lin, “Unstructured big data analytics for retrieving e-commerce logistics knowledge,” Telemat. Informatics, 2018.
3. M.R. Tahiduzzaman. Big data and its impact on digitized supply chain management. IJRDOJournal of Business Management. 2017.
4. Yang Chao. Research on key technologies of BI system based on big data technology [D]. South China University of Technology, 2016.
5. S. Jeble, S. Kumari, Y. Patil. Role of big data and predictive analytics. In: International Journal of Automation and Logistics. IEEE. 2016.
6. Porche, I.R., Wilson, B., Johnson, E., Tierney, S., Saltzman, E. Big data: challenges and opportunities, data flood. RAND Corporation, 2014.
7. U. Edosio, “Big Data Analytics and its Application in E- Commerce,” E-Commerce Technol. Univ. Bradford, Vol. 1, no. July, pp. 0–12, 2014.
8. J. Fan. Challenges of Big Data Analysis. National Science Review. 2013.

9. Y. Riahi. Big Data and Big Data Analytics: Concepts, Types and Technologies. ijre. 2008.