Utilization of Big Data In E-Commerce Business

Nico Fernando[1], Mery[2], Jessica[3], Johanes Fernandes Andry[4]
Bunda Mulia University
Jl. Lodan Raya Nomor 2 Jakarta Utara, Indonesia
E-mail: nicolai3111998@gmail.com[1], mery.nebpehtire1824@gmail.com[2],
nerumahiro46@gmail.com[3], jandry@bundamulia.ac.id[4]

Abstract

In this day and age with the internet brings revolution to the way all fields work, especially in the field of business. With the internet technology data became big and known as "Big Data". The development of big data has increased significantly so that it can be utilized in various fields, especially in business areas that have been implemented with internet technology. This electronic buying and selling media have a wide range such as from small to large stores that can utilize media or with a site. This makes users always need this technology, so an E-commerce can also be said to be the largest data-producing media. This study addresses the problem of the extent to which big data generated from E-commerce can affect business and provide benefits for business organizations such as expanding the scope of transactions, supporting decision-making, and others. The research method used in compiling this research is to collect data and information and then conduct processing and analysis of the data. So it is expected from the results of utilization of big data in this E-commerce business that has been processed can provide support especially in making decisions with cluster results that have been obtained for example such as to know the most sales patterns in order to be able to add stock to certain goods and determine promotions based on future sales. The study concluded that based on gift shop sales data, the average item purchased by the store's customers is found in items that are included in cluster group 0 so that the gift shop can increase the stock of items contained in cluster group 0.

Keywords: Big Data, Ecommerce, Business

1. Introduction

Today, developments in usage such as the internet, social media, and online commerce have grown rapidly. This growth is not only due to our activities as a millennial community that can not be separated from gadgets or gadgets especially mobile phones that can connect to the internet network because this gadget can practically be accessed anytime and anywhere. Indonesia itself certainly has considerable capabilities in the development of internet-based technology. Activities are done online such as surfing web pages, social networks, online buying, and selling, or just watching videos. One of the things that are on the rise in online buying and selling transactions or what is commonly called E-commerce [1].

Based on population data, Indonesia's population has now reached 262 million people and approximately more than 50 percent (about 143 million people) have been connected to the internet throughout 2017, the above information is data from the Association of Internet Service Providers of Indonesia (APJII). The number of gadget or mobile phone users with soaring development can be seen as an amazing trading capability. Knowing the fact that the implementation of E-commerce technology becomes one of the useful parts to help the success of a company's goods. To shorten and increase sales, pay attention to the rapid growth of information technology that we need an online service such as E-commerce. The use of the internet allows small and medium-sized micro-businesses utilizing E-commerce to conduct sales by targeting international markets, making it highly likely that there are opportunities to penetrate exports [2].

Through an E-commerce application set up in a store or other part of the client can correlate spontaneously even if it does not visit the store. A lot of research focuses on accuracy in the use of E-commerce. In addition, many reviewers saw good results given by E-commerce analogous to its negative results. Big data can be used by anyone in a small, medium, or large business. Some of the functions of Big data that have been felt especially by the business world include to know the opinions of residents about the results issued through social media, help determine decisions correctly, soar company icons in the eyes of clients, and understand market trends and customer will [3].

The absence of valid data will contribute to the development or increase of Big Data. There are already many areas that make use of this big data as well. Based on this, the author performs the processing of gift shop data using the K-Means method. The data used is sales data obtained from 2010. In general, sellers use sales data to obtain information about sales information and stock availability. However, the data has never been used for more in-depth processing due to a large amount of sales data, this prevents the seller from obtaining any information that can actually help the seller to make the most of the information generated from the sales data such as using that data to determine a good business strategy in making a sale. Business strategies are urgently needed for businesses to be able to make a guideline or director of action in selling goods or services in order to achieve their desired goals [4].

Managing the stock of sales goods is one of the business strategies needed by the seller to transact so that the goods stored by the seller can be distributed to the customers in good condition. However, in conducting business activities in E-commerce often the seller experiences errors in stockpiling goods such as sellers buying too much stock for goods that do not sell very well. This causes the inventory of goods to not match the sale resulting in the goods piling up in the warehouse and can harm the seller because the goods are rarely or difficult to sell. Therefore, this study will be explained about the utilization of Big Data in the E-commerce business represented by gift shop sales data using Rapid Miner. So this research is expected to address the problems experienced by E-commerce sellers in order to utilize big data effectively in stock processing [5].

2. Library overview
2.1 Previous Research

There are several journals that conduct research on the utilization of big data in the E-commerce business one example is "Utilization of Big Data in the Application of Dynamic Pricing" where this study has a case study using Amazon.com written by Mar'atus Sholikhatun Nisa and Yusuf Amrozi. The study discusses how Amazon companies use Big Data to implement dynamic pricing to customize prices or discounts for specific items according to specific customers in real-time. So from the previous research mentioned above, we authors get guidelines and ideas to conduct a study on Big Data that can be utilized to assist shop sellers or businesses in making decisions or in strategizing businesses [6].

2.2 Big Data

The rapid growth of a scientific method that serves to achieve this practical goal gives rise to a new period called the digital period. The arrival of internet technology makes data also undergo significant changes. The data in the past was not very large in number and now the data today is very much obtained from us as a society of internet users. This situation gave birth to a well-known thought called Big Data. According to Thomas, "Big Data is a term that describes the large volume of data - both structured and unstructured - that floods the business on a daily day. But it's not the amount of data that matters. That's what organizations do with important data [7].

Big data can be analyzed for insights that lead to better decisions and strategic business moves". The above terms can then be interpreted in general as a data set that forms the concept of 3V i.e. volume (data that has a large size), variety (data has various variations of extensions or formats), velocity (data in terms of data transfer and data processing), and also has other V additions namely veracity and value. The concept of 3v's can be seen in the image below:
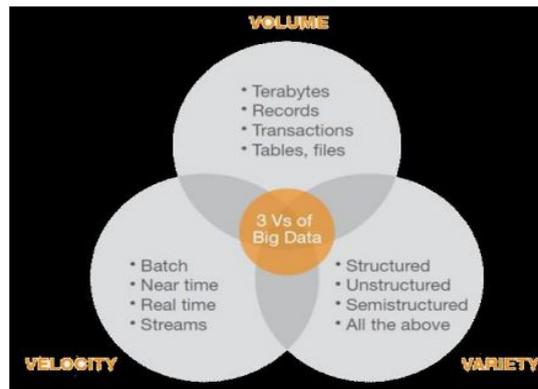
Figure 1. Of 3v's Concept

Big data itself can also be utilized in various fields one of which is the business economy. The business economy itself has more areas, one of which is E-commerce.

2.3 E-commerce

E-commerce is a sales and purchase activity of products or services that are mediated with internet networks. E-commerce itself is a member of the business. E-commerce itself also requires a place to store data or post data commonly referred to as databases, internet-based mail or e-mail, and couriers or means of settlement or repayment of goods that have been purchased on the E-commerce. Meanwhile, E-commerce according to Laudon and Laudon (1998) is: "The process of buying and selling goods electronically by consumers and from company to company through computerized business transactions" [8].

From the above understanding, it can be concluded that there are three things in E-commerce namely, the process of selling and purchasing through the internet. Then, there are consumers (Buyers) and or businesses. Finally, the use of the internet network through online gadget media to make trade transactions. Therefore, the capability in this case to support the guarantee of E-commerce can be realized through an online purchase method that is in the layer of a protected internet network (encrypted). E-commerce itself has a classification according to the transaction process, according to M. Suyanto (2003) E-commerce can be divided into the 3 most common namely B2B or Business to Business, B2C or Business to Consumer, and C2C or Consumer to Consumer [9].
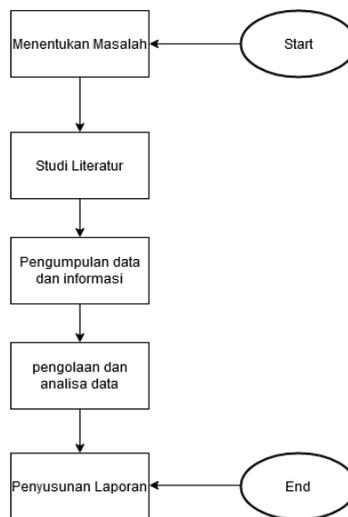
3. Research methods

Figure 2. Flowchart

Based on Figure 2 above, the research flow used in this research / paper is as follows:
1. Determining the Problem, in this stage is a step to establish the problems that exist in the e-commerce gift shop.
2. Literature Study is the next stage by the way the author digs or searches for relevant journals that will be used as a reference for this method.
3. Data collection and information, which is the third step that is the author collects data and information through the Kaggle website. Then, the gift shop e-commerce data was found.
4. Data Processing and Analysis, which is the fourth step after collecting data, is the author of processing the data that has been obtained with a tool named RapidMiner. The processed data will be analyzed.
5. Summing up the results of the study, the last step is the step taken by the author to conclusive the results of the analysis that has been done by the authors on the data that has been processed.

4.Discussion

The data to be retrieved and processed in this discussion comes from a site called Kaggle which contains a transactional data set of gift shops. The gift shop is a business engaged in e-commerce that does not have a physical store so that all transactions made in the gift shop are online using an application connected to the internet connection. The gift shop sells unique gifts for all kinds of events. The average customer of such gift shop is a wholesaler. This data amounts to 2500 rows with 8 attributes. Attributes contained in gift store data are InvoiceNo, StockCode, Description, Quantity, Invoice Date, UnitPrice, Customer, and Country. The purpose of this data is to determine the business strategy for the goods to be sold by the gift shop in the future, to regulate the import of goods, and gain a lot of profit. Here is an example of some of the data to be processed shown in table 1 [10].

Based on table 1 above, the InvoiceNo attribute is data generated based on transactions that have been made by the seller and the buyer. Next stock code attribute is a unique identity that the product has based on the name and type of the product, Furthermore Description is a description of the product name based on their respective StockCode, The fourth attribute is a quantity which is the amount of each product purchased by the customer in one InvoiceNo, Next invoice date attribute is the date of the transaction that has been made by the buyer when the transaction has been completed, Furthermore, the UnitPrice attribute is the unit price on each product using dollar currency units or other currencies based on the buyer's Country, the next attribute is the customer is the identity owned by the buyer on each transaction that has been done [11].

The last one is the country attribute is the country where the customer makes the purchase transaction so that the seller can know where the customer is making the transaction. The data will be processed in the RapidMiner application using the clustering method. One of the clustering methods to be used is the K-Means algorithm because the K-Means algorithm compared to other clustering algorithms can produce more accurate results or calculations. K-Means is used by authors/researchers because it is also a simple concept and is an unsupervised algorithm that has the purpose of finding groups in existing data and divided by groups defined by a variable named K. Here is the formula of K-Means, in Figure 3 [12].

Table 1. Gift Shop E-commerce Data

| Invoice No	Stock Code	Description	Quantity	Invoice Date	Unit Price	Customer	Country
536365	85123A	White Hanging Heart T-Light Holder	6	12/01/2010 8:26	2.55	17850	United Kingdom
536365	71053	White Metal Lantern	6	12/01/2010 8:26	3.39	17850	United Kingdom
536365	84406B	Cream Cupid Hearts Coat Hanger	8	12/01/2010 8:26	2.75	17850	United Kingdom
536365	84029G	Knitted Union Flag Hot Water Bottle	6	12/01/2010 8:26	3.39	17850	United Kingdom
536365	8429E	Red Woolly Hottie White Heart.	6	12/01/2010 8:26	3.39	17850	United Kingdom
536365	22752	Set 7 Babushka Nesting Boxes	2	12/01/2010 8:26	7.65	17850	United Kingdom

Figure 3. of the K-Mean Formula.
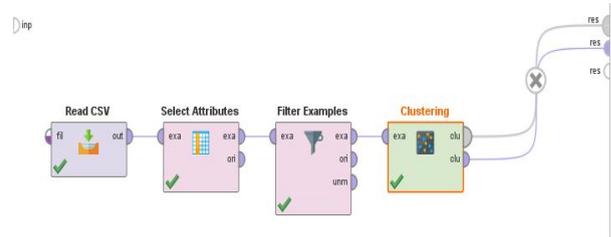
Figure 4. Gift Shop Data Processing Process using Rapid Miner

Table 2. Data grouping results by cluster

Row No.	CustomerID	StockCode	cluster	Quantity	UnitPrice
1905	14606	21643	cluster_0	3	1.250
1906	14606	72741	cluster_0	27	1.450
1907	14606	21620	cluster_0	1	3.750
1908	14606	37449	cluster_0	1	9.950
1909	14606	22260	cluster_0	5	0.850
1910	14606	22136	cluster_0	1	1650
1911	14606	47591B	cluster_0	1	1.650
1912	14606	21041	cluster_0	1	2.950
1913	14606	21035	cluster_0	1	2.950
1914	14606	85064	cluster_0	2	5.450
1915	14606	22121	cluster_0	2	5.950
1916	14606	21488	cluster_0	1	3.950
1917	14606	21484	cluster_0	2	3.450
1918	14606	22357	cluster_0	1	4.250
1919	14606	21985	cluster_0	4	0.850

ExampleSet (1.919 examples, 3 special attributes, 2 regular attributes)

Figure 4 is the process of processing data at a gift shop. First, the data is inserted into the RapidMiner application by using the Read CSV tool to insert a CSV-extension file if you want to insert an XLS-extension file using Excel's Read tools. The next step of data is cleared by selecting the attributes to be processed using the Select Attributes tools and cleaning [13. invalid data such as missing values using the Filter Examples tools. The attributes to be used in the clustering method are CustomerID, Quantity, and UnitPrice. Then, the data will be modeled using the K-Means algorithm. The result is in table 2 [14].

Table 2 is a view of attributes that have been grouped by cluster by using the K-Means algorithm. The grouping resulted in 1919 rows of sample data divided into two clusters namely Cluster 0 and Cluster 1 [15]. The result of the division of these clusters is generated based on UnitPrice with a range of 0.200 to 10 dollars for clusters 0 and 11 to 15 dollars for cluster 1. Here is the result of the total amount of data on the grouping of each cluster:

Figure 5. Number of data of each cluster

Figure 5 is the amount of data from each cluster that has been grouped using the K-Means algorithm using two parameters. From figure 5 above it can be known from the 1919 sample data that the data included in cluster group 0 amounts to 1906 data and cluster group 1 amount to 13 data. From the K-Means process using two parameters, cluster 0 indicates that there are 1906 items sold while in cluster 1 there are 13 items sold so it can be known that the most items sold are in cluster grouping 0. From the results obtained, it is known that the average item purchased by gift shop customers is found on items grouped on cluster 0 so that the gift shop must increase the stock of items contained in cluster 0.

5. Conclusion

Big Data technology is a very important feature for various fields in the world, especially in the business field. Big Data has a lot of data that can be used as information for various fields but it must be processed properly using the correct means and tools in order to get effective information. utilization of big data in the E-commerce business is big data provides many benefits to the E-commerce business such as being able to give the right decision for development in the E-commerce business and predict the number and price of products to be sold in the future Based on the test results that have been obtained, it can be summed up as follows: Designing using the K-Means algorithm with rapidminer software application was successfully implemented. The result of the K-means clustering data mining algorithm method with parameter 2 can be used for the clustering process in this e-commerce sales transaction data to know the most sales patterns and determine the promo based on future sales.

References

[1]　Achjari, D., "Potensi Manfaat dan Problem di E-commerce". Jurnal Ekonomi dan Bisnis Indonesia, 2000, vol. 15(3), pp. 388-395.

[2]　Cordiaz, M., dan Prahasto, T., "Analisis Faktor Kredibilitas Website E-commerce Indonesia Studi Pada Online Purchashing". Jurnal Sistem Informasi Bisnis, 2013.

[3]　Haerulah, E., dan Ismiyatih, S., "Aplikasi E-commerce Penjualan Souvenir Pernikahan Pada Toko "XYZ"". Jurnal PROSISKO, 2017, vol. 4(1).

[4]　Febriyanto, E., & Naufal, R. S. (2019). Attitude Competency Assessment in the 2013 Curriculum Based On Elementary School Prototyping Methods. IAIC Transactions on Sustainable Digital Innovation, 1(1), 87-96

[5]　Irmawati, D., "Pemanfaatan E-commerce Dalam Dunia Bisnis". Jurnal Ilmiah Orasi Bisnis, 2011, vol. 6.

[6]　Khairani, A., dan Irwansyah, "Pemanfaatan Big Data Untuk Percepatan Proses Underwriting Sebagai Strategi Komunikasi Marketing Terpadu Perusahaan Asuransi Jiwa". Jurnal NOMOSLECA, 2018, vol. 4(2), pp. 785-793.

[7]　Maryama, S., "Penerapan E-commerce Dalam Meningkatkan Daya Saing Usaha". Jurnal Liquidity, 2013, vol. 2(1), pp. 73-79.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8]　Megantara, F., dan Warnars Harco Leslie, S., "Implementasi Big Data untuk Pencarian Pattern Data Gudang Pada PT. Bank Mandiri (Persero) Tbk". Jurnal SISFOTEK GLOBAL, 2016, vol. 6(2), pp. 61-68.

[9]　Nuryanti., "Peran E-commerce Untuk Meningkatkan Daya Saing Usaha Kecil dan Menengah (UKM)". Jurnal Ekonomi, 2013, vol. 21(4).

[10]　Pujianto, A., Mulyati, A., dan Novaria, R. "Pemanfaatan Big Data Dan Perlindungan Privasi Konsumen Di Era Ekonomi Digital". Jurnal BIJAK, 2018, vol. 15(2), pp. 127 – 137.

[11]　Septri, W., dan Krisnadi, I., "Pemanfaatan Big Data Dalam Perkembangan Ekonomi Dan Bisnis Di Indonesia". Jurnal Manajemen dan Bisnis ICT, 2019.

[12]　Sirait Emyana Ruth, E., "Implementasi Teknologi Big Data di Lembaga Pemerintahan Indonesia". Jurnal Penelitian Pos dan Informatika, 2016, vol. 6(2), pp. 113-136.

[13]　Sudaryana I, K., Sanjaya, H., Tjong, R., "Analisis Website Wiki Versaillus Dengan Menggunakan Metode PIECES". Journal of Business and Audit Information Systems, 2019, Vol 2(2), pp. 38-46.

[14]　Tu, Britney. "Exploration of Fundamental Clustering Algorithms: K-means and DBSCAN". Dalam https://medium.com/sfu-cspmp/exploration-of-fundamental-clustering-algorithms-k-means-and-dbscan-eac733 2cf74b　(Diakses pada 13 Oktober 2020).

[15]　Febriyanto, E., Naufal, R. S., & budiarty, frizca. (2019). Attitude Competency Assessment in the 2013 curriculum based on elementary school Prototyping methods. IAIC Transactions on Sustainable Digital Innovation (ITSDI), 1(1), 87-96. https://doi.org/10.34306/itsdi.v1i1.6