

ANALYSIS OF FACTORS CAUSING THE PRODUCT PURCHASED BY CONSUMERS BASED ON DATA ON THE SALE OF GOODS IN A MARKET PLACE IN INDONESIA

Baihaqi Siregar¹, Erna B Nababan¹, Opim S Sitompul¹

¹Department of Information Technology,
Faculty of Computer Science and Information Technology,
University of Sumatera Utara, Medan 20155, Indonesia
Email: baihaqi@usu.ac.id, ernabrn@usu.ac.id, opim@usu.ac.id

Abstract

Starting from the success of giant web service companies as well as Google and Facebook in managing and utilizing unstructured data in the form of consumer generated media and click stream in a very large volume, a concept known as Big Data then became the center attention in the world of information technology. The fact also shows that more and more organizations in the world, whether private companies or government agencies, have difficulty managing data whose volumes are growing and their types are increasingly complex. They have to organize and analyze these data, and they must find the meaning or value of the ever-expanding and increasingly complex data pile, which is said to have exceeded the capability of conventional data processing applications to process it. The condition of this kind of data is also categorized as Big Data, which is interpreted as a set of data in a very large number of challenges lies in how the data should be stored, how to search in the pile of data, how to distribute it, how to visualize it, and how the data should be analyzed. The long-term goal of IbKIK's proposed program is the establishment of a startup company in the field of analytic data from the world of campus directly. Within the planned three-year period, it is desirable that the company be financially self-sufficient by being a data analytics consultant and also creating a sophisticated and advanced Big Data Analytic application platform product. The advantages possessed when the company started from the academic world is the quantity and quality of human resources as intellectual actors can be selected quickly and accurately. Especially with the synchronization between the curriculum content that is taught with its implementation directly through the program IbKIK become useful products and economic value. From the academic point of view, the desired outcomes are from this program published several journals and proceedings of national and international scale, the publication of textbooks, getting HKI, and publications in the mass media. Also, with the success of this company can produce a derivative company engaged in other areas that are still related as a supporter of the business. The product output of the community service activity that has been done for the first year of the planned three years period is the establishment of a Product Information System Sold on E-Commerce Transactions at Market Place. Also, has been established research unit as the forerunner of the business unit under the auspices of the Faculty of Computer Science and Information Technology University of Sumatera Utara under the name Data Science Research Group.

Keywords: *big data analytic, market place e-commerce transaction, IbKIK program*

1. INTRODUCTION

Niche market for business processing and data analysis for large data scale is still wide open. It can be because the business involves human resources, known as scientific data, highly capable in statistical science, computer

programming and server handling, and understand the

business processes running on the institution from which the data to be analyzed is derived. The slices for business skills can be seen in Figure 1.1.

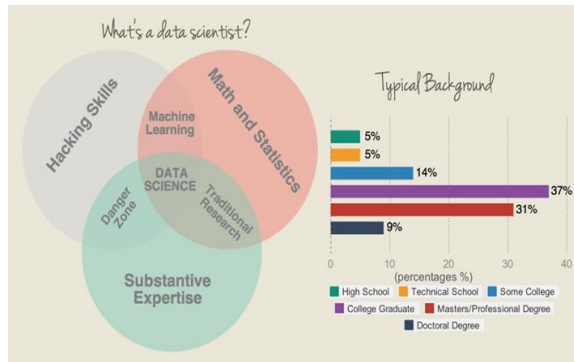


Figure 1.1. Sliced HR Capability as Data Scientist [1]

That is Solution247 and Dattabot. Both of these serve data analysis from multinational companies. But both still focus on the analysis of data that is supervised learning. While through this IbKIK program we plan to plunge in the field of data analysis that is supervised learning, unsupervised learning, and reinforcement learning. Certainly, with the price of products and services more competitive. With target clients in the form of oil and gas and non-oil industry, online and traditional markets, government with e-government data, NGOs, factories, and so forth.

The product that will be generated through this IbKIK program is Big Data Analytic application platform, which through the use of this application dataset as input data can generate predictive analysis which is accurate and provides insight in the form of user-friendly dashboard view from UI and UX side. To produce this product, we use open source infrastructure such as Python and R programming languages, Apache Hadoop, Kafka, Zookeeper, and Spark application servers. As well as PostgreSQL as database engine. As an illustration of how the steps taken along with the objectives to be achieved from the analysis of this data can be seen in Figure 1.2 which represents the use of machine learning using the library learn-bit in Python. In addition to producing products, we will also serve the services of data analysis.

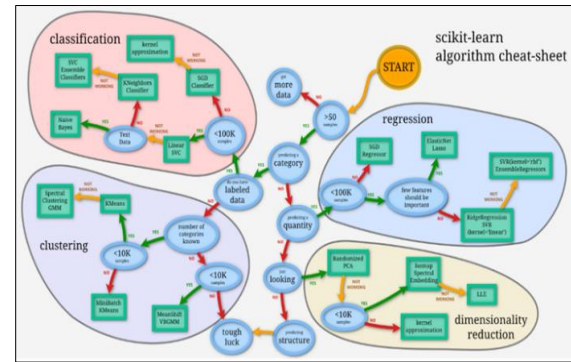


Figure 1.2. Python scikit-learn [2]

The impact and benefits of these products and services for the socio-economic community is through accurate data analysis gained the right direction of targets for policymakers, such as at the level of heads of government and commercial enterprise management. Also for the academic world where it can be implemented directly science that has been set in the curriculum, so that the students get skilled and understand the concept especially in the field of data processing.

According to IDC survey results, there will be an increase in revenue from the Big Data Analytic business from \$ 130.1 billion in 2016 to \$ 203 billion by 2020 [3]. This shows that business opportunities in this field are still very large, so it is a pity if there is no academic from the campus who tried to make it into professional commercial business products. The target market is industry, government with the implementation of e-government, factory, online and traditional market, health, social, economy, culture, geospatial, and others.

2. MATERIAL AND METHOD

The success of the big data analysis process requires the dataset as the main source of raw material. This dataset can be obtained from clients who need insight from the data they have and can be obtained from open data that is widely available in cyberspace.

2.1. Material

The current production capital is several e-books and online learning media resources in machine learning, deep learning, database, artificial intelligence, probability and statistics, and some guidance of R and Python programming language for science data. There is also a PC in

Baihaqi Siregar, et al. Analysis of Factors Causing The Product Purchased By Consumers Based On Data...

the faculty lab as a small-scale data processing and analysis engine.

For further development we need a much more powerful but inexpensive computer processor resource on purchasing and operational costs. Cloud computing technology is the best choice we will use with that consideration. We will also subscribe to some e-learning and participate in bootcamp related to science data.

2.2. Method

Stages through which to gain insight is first to get raw data using crawling techniques from the desired market place. The raw data obtained is transferred to the local database, then performed pre-processing steps such as data extraction, normalization data, and data cleansing. Then the data go into the data exploration stage, which consists of querying process, information discovery, and reporting. If any information is lacking, it will be re-accessed to the local database for additional information. But if it is felt enough information is desired, then the next stage is to the data presentation, the data is displayed as visual information that can be useful for decision making. Overall, all stages of process can be seen in the general architecture in Figure 2.1.

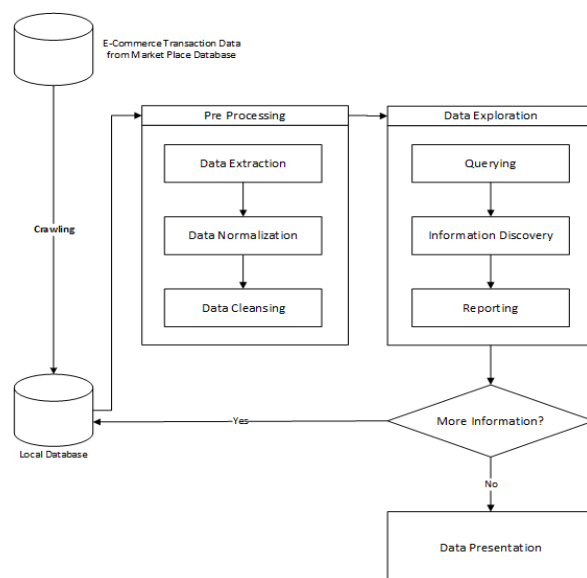


Figure 2.1. General Architecture

3. RESULTS

In the rampant development of technology today, product sales are very easy to do. Of course, by applying the most obvious trend seen from the digital industry, e-commerce.

The growth of e-commerce sites in Indonesia also increasingly many considering the increasing percentage of the number of online transactions in Indonesia. With the number of internet users reaching the number 82 million people or about 30% of the total population in Indonesia, the e-commerce market into a gold mine is very tempting for some people who can see the potential for the future. In Nielsen's report entitled Indonesia Ocean of Opportunities, sales of e-Commerce Indonesia in 2015 reached US \$ 1.68 billion. This figure is only 1.2 percent of total domestic retail sales of US \$ 145.83 billion. When compared with e-commerce sales in countries in Southeast Asia, Indonesia is the largest [4]. With e-commerce sites available, business owners do not need to have a concrete store to start their business, put up billboards along the way for marketing, or even meet their customers directly to make transactions. All that can be done simply by registering their businesses into the various bidding options of the ecommerce marketplace available today. The marketplace or online stores are available in Indonesia such as Tokopedia, BukaLapak, Lazada, Blanja, OLX, and many more.

Electronic trading transactions in the market place produce abundant and interesting data to analyze. Various types of data collected increasingly booming called as Big Data. This is certainly influenced by the growing business, consumer, and transactions are increasing. This online sale becomes interesting to be studied in the form of statistics based on existing data. Not only processed into a more informative form, the resulting data can be used as a source to find sales problems as well as improve performance to improve the quality of product sales better. From the data collected can also be applied method of learning system that can predict the needs in the future.

Speaking of e-commerce data, as for the highest selling products online in favor of fashion and accessories products, with a percentage of 37.6%. In addition, according to data from the last three months on the website of Kominfo, the highest percentage is also occupied

Baihaqi Siregar. et al. Analysis of Factors Causing The Product Purchased By Consumers Based On Data...

by the type of goods / services in the form of clothing that is equal to 73.80%, followed by 27.50% by category cosmetic and medicine [5]. With that highest percentage, we can be sure that this product data is more than any other product data. This draws its own attention in the processing and utilization of available data. From the available data we can analyze the strengths and weaknesses of sellers selling their products. In addition, learning of data can help in making decisions in the future.

The category of fashion or clothing in e-commerce services to be the most desirable given the percentage of sales of these products are fairly high from other categories. With the highest percentage calculation, we can be sure that this product data more than other product data. This draws its own attention in the processing and utilization of available data. From the available data we can analyze the strengths and weaknesses of product sales. In addition, learning of data can help in making decisions in the future.

Figure 3.1 shows an informative graph which states the product association and reputation with the fashion price by the seller residing in Medan city obtained from one of the biggest market place in Indonesia. Data was obtained until October 2017.

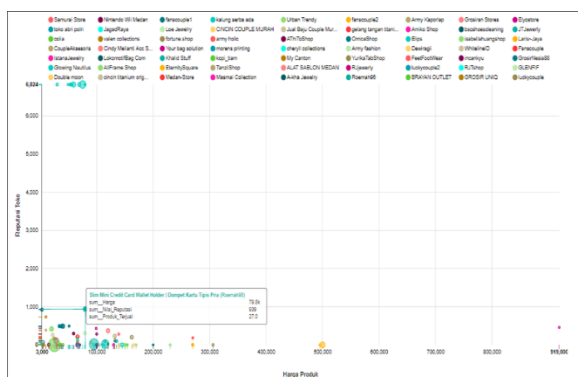


Figure 3.1. Visualization Data 100 Fashion Products with the highest Sales in Medan City

This monitored market place provides various types of data displayed on the product page and the seller. Some of the data is input and part of the data is the assessment and calculation results. Examples of input data such as seller name, product name, price, product photo, and more. As for some examples of data assessment results and calculations such as the number of visitors, the reputation of the seller, the product rating, the number of products sold, percentage of transactions, and others.

Figure 3.1 is a type of visualization bubble chart associated association five types of fashion sales data group in the city of Medan. The information displayed is limited to sampling 100 fashion products that have the highest product value sold. The five types of data groups presented in the graph, among others:

3.1. Seller's Name

The name of the seller is the name of the business of a group of sellers or entrepreneurs who market their products. The seller's name is symbolized by a circle that has an overall size equation, is at the top of the graphic display and consists of various color variations followed by each seller's name, as shown in Figure 3.2.



Figure 3.2. The Legend that Shows the Seller's Name

3.2. Product Name

Product name is a list of fashion products sold from various stores. Product names are also symbolized by circles of varying colors and sizes in the graph. The product circle color is matched to the circle color of the store name, where the store sells the product in question. One shop as a seller or entrepreneur can market several products. Each product circle is located based on the cartesian coordinate point of X Axis and Y Axis according to the value of the support the product has, as shown in Figure 3.3.

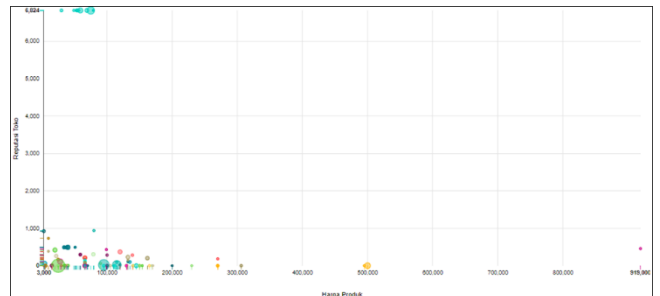


Figure 3.3. The Legend that Shows the Seller's Name

3.3. The Value of Store Reputation

The store reputation value is the value given by the buyer to the seller after the buyer accepts the order. The store reputation value is displayed in the form of store badges, which can be seen on the store page and any pages with the seller's name. In Figure 3.1, store reputation values are

represented by Y (vertical) coordinates. The conclusions obtained from Figure 3.1. is the influence of store reputation value is not a major requirement of consumers to decide whether to buy or not to buy products from the store. The graphic information of Figure 3.1 shows the highest number of sold products occupied by the product from a reputable store located at the point 0. In addition, most product sales are also in the store's reputation range between 0 - 100.

3.4. Product Price

In Figure 3.1, the price of the product is represented by X (horizontal) coordinates. From the observation on the graph in Figure 4.4 it is known that the purchase of most fashion products is a product with a price range between Rp.3,000 – Rp.200,000.

3.5. Product Value Sold

The number of sold products is presented in the size of each product name circle. The higher the value, the wider the size of the product circle. From the data value of the sold product can be obtained influence of store reputation value and price of each product. The highest number of products sold certainly becomes the main benchmark in the analysis of sales of a product. In the graphic in Figure 3.1 there is also a tooltip that appears when selecting product circles. The information shown is the five data groups mentioned above, ie product name, seller name (in parentheses), product price, store reputation value, and number of products sold, as shown in Figure 3.4.

Slim Mini Credit Card Wallet Holder Dompot Kartu Tipis Pria (Roemah95)			
sum__Harga			79.8k
sum__Nilai_Reputasi			939
sum__Produk_Terjual			27.0

Figure 3.4. Products' Tooltips

In addition to presenting data in a more interesting and informative way, the insight of graphical output as shown in Figure 3.1 also presents the ease of capturing statistical information. Some of the information obtained such as the highest-selling or most-sought product, the reputation of the store's reputation and the price of the product on sales, and the price range or reputation range in product sales.

4. CONCLUSION

From the results of analysis conducted on electronic shopping transactions through one of the market place in Indonesia obtained the following conclusions:

1. The data of the seller is the seller or shop domiciled in Medan;
2. Most purchased products are in the fashion category;
3. There is a phenomenon which shows that buyers buy goods not only on the cheap price;
4. There are buyers also considering high reputation store value factors and mid-value items (not too cheap and not too expensive);
5. However, most buyers prefer to buy items that are quantitatively sold out even though they are expensive;
6. Which means buyers have more confidence to shop for sellers or stores that historically have successfully sold similar products than sellers or other stores, even though they are more expensive.

The final result obtained from the multi-year community service activity of IbKIK scheme is the development of a product of Big Data Analytic in the form of application of Product Information System Sold on E-Commerce Transaction in Market Place. This application can be used by government agencies and business people who mainly rely on sales in the realm of cyberspace. The raw data used in the form of tabulation and analysis data used are supervised learning.

For the second year it is planned to use more complex raw data and use unsupervised learning analysis. Data Science Research Group under the supervision of lecturers at the Faculty of Computer Science and Information Technology will serve as a vehicle for learning and development of Big Data Analytic system will be driven by students of the University of Sumatera Utara in accordance with the competence of expertise.

5. REFERENCES

K. Willems, "Become A Data Scientist in 8 Steps: Infographic," 10 November 2014. [Online]. Available: <https://www.datacamp.com/community/tutorials/how-to-become-a-data-scientist>.

scikit-learn.org, "Documentation of scikit-learn 0.19.1," [Online]. Available: http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html.

International Data Corporation, "Double-Digit Growth Forecast for the Worldwide Big Data and Business Analytics Market Through 2020 Led by Banking and Manufacturing Investments, According to IDC," 03 October 2016. [Online]. Available: <https://www.idc.com/getdoc.jsp?containerId=prUS41826116>.

Katadata Indonesia, "Berapa Penjualan e-Commerce Indonesia?," 05 September 2017. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2017/09/05/berapa-penjualan-e-commerce-indonesia>.

Ditjen Aptika Kominfo, "Data & Statistik Kementerian Komunikasi dan Informatika RI," [Online]. Available: <https://statistik.kominfo.go.id/site/data?idtree=430&iddoc=1526>.