Implementation of Data Mining on Online Shop in Indonesia

O Chouat* and A H Irawan

Departemen of Science, University of Monastir, Tunisia
Departemen Teknik dan Ilmu Komputer, Universitas Komputer Indonesia, Indonesia

* eddysoeryantos@email.unikom.ac.id

Abstract. This research attempts to explore the implementation of Data Mining on online shop in Indonesia. Data Mining is a method of collecting data and finding a certain pattern in order to discover new and useful informations. Data Mining has been used comprehensively and effectively by many institutions and for certain purposes. In online store, Data Mining includes the statistical functions on a wide range of functions designed to support statistical performance analysis on the data such as the purchasing of the customers. The method used in this journal was qualitative research which concerned with Data Mining phenomena happened in online retail in Indonesia. The results of implementing Data Mining on online retail is to help the owner when maximizing stock supply, maximizing marketing strategy with market-share that can be seen from the segment of products purchased by customers. Other than that, there is the convenience of sorting products for both buyers and sellers.

1. Introduction

In this advanced era, a data mining is very useful for large data processing, evolved substantially and now accompanies a text on data mining. [1] Data mining is thus emerging as a class of analytical techniques that go beyond statistics and aim at examining large quantities of data. What is important to keep in mind is the problems associated with data mining are fundamentally statistical in nature; that is, to infer patterns or models from data.[2]

Data Mining is a detailed process of analyzing large amounts of data and picking out the relevant information. It refers to extracting or mining knowledge from large amounts of data. The data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamical. Data mining serves as a helper in decision making because each transaction data is stored into the data base and data mining will find the pattern as what appropriate algorithm used in the search data. In this case the system using apriory algorithm so that the seller can see the percentage of product sales and which products are the most superior in the seller. Pérez et.al, Data mining is a widely used approach for the transformation of data to useful patterns, aiding the comprehensive knowledge of the concrete domain information.[3]

According to Rygielski et.al, technologies such as data warehousing, data mining, and campaign management software have made customer relationship management a new area where firms can gain a competitive advantage. Particularly through data mining—the extraction of hidden predictive information from large databases—organizations can identify valuable customers, predict future behaviors, and enable firms to make proactive, knowledge-driven decisions. The automated, future-oriented analyses made possible by data mining move beyond the analyses of past events typically

provided by history-oriented tools such as decision support systems. Data mining tools answer business questions that in the past were too time-consuming to pursue. [4]

If consumers often access Fashion category, the application will provide recommendations according to the consumer access habits of clothes and clothing. [5] Convenience and product type influence consumer's intention to engage in online shopping. [6] Buiding experiential marketing was positively related to customer satisfaction and customer loyalty, and was a significant predictor of them.[7] Data Mining can also be used to analyze unfamiliar transaction, as well as in a Bank that is applied in the Credit Card. The system will also record the activities of Credit Card users, such as shopping habits with offline or online transactions or shopping habits at supermarkets etc. So if there is an unusual transaction, the system will automatically detect the unfamiliarity to be confirmed further to the consumer whether this transaction is from the card holder itself or not. And if an application does not use Data Mining, then the company cannot arguably maximize opportunities for possible transactions done by consumers. So, the usage of Data Mining helps a lot because Data Mining has advantages in Business aspect, and step away promotion sites for online shops to enhance their profitability [8] then with the effectiveness and analysis about accurate fraud detection. [9]

Kanungo et.al said about presenting a simple and efficient implementation of Lloyd's k-means clustering algorithm, which we call the filtering algorithm. This algorithm is easy to implement, requiring a kd-tree as the only major data structure. We establish the practical efficiency of the filtering algorithm in two ways. First, we present a data-sensitive analysis of the algorithm's running time, which shows that the algorithm runs faster as the separation between clusters increases. Second, we present a number of empirical studies both on synthetically generated data and on real data sets from applications in colour quantization, data compression, and image segmentation. [10, 11] It can be concluded that some of the above references or literature do not explain the phenomenon of data mining, and do not describe how the process of formation in detail. This research attempts to explore the implementation of Data Mining on online shop in Indonesia.

2. Method
The research method used in this paper was Qualitative Research that describes the data mining algorithm apriory that can easily solve the problem by analyzing the usage of data mining.

2.1. Algorithm Analysis
A priori algorithm is an algorithm that is appropriate in this case, but first, we need to know what the value of support is, support value is the value that transaction value containing goods x divided by total transaction multiplied 100 (See Figure 1).

$$\frac{\text{Number of transactions containing A}}{\text{Total Transaction}} * 100\%$$

Figure 1. Searching for more than two supports of values items

2.2. Algorithm Implementation
We can see through the purchasing table to get the representation from particular store for a day (See Tables 1-3).

Table 1. Percentage of transactions

No	Item Set
1	PANASONIC – TH- 32A402G, PANASONIC – AS610G SERIES
2	ASUS – ZENFONE 4 NEXT G, PANASONIC – AS610G SERIES
3	SAMSUNG –GALAXY NOTE 3
4	ADVAN – VANDROID S3A
5	ADVAN – VANDROID S3A, SAMSUNG – GALAXY NOTE 3
6	ADVAN – VANDROID S3A
7	ADVAN – VANDROID S3A
8	ASUS – FENOPAD 7 (TAM), ADVAN – VANDOID S3A
9	ASUS – FENOPAD 7 (TAM), ADVAN – VANDOID S3A
10	ASUS – FENOPAD 7 (TAM), ADVAN – VANDOID S3A

Table 2. Item Combination

Item	Support	Perhitungan
Advan – Vandroid s3a	63%	7/11 * 100
Samsung – Galaxy Note 3	45 %	4/11 & 100

Table 3. The purchase of galaxy note 3

Item	Confidence
If we purchase Advan – Vandroid s3a we will automatically purchase Samsung Galaxy note 3	33%
If we purchase Samsung – Galaxy note 3, we will automatically purchase Advan – Vandroid s3a	60 %

3. Results and Discussion
To implement the design above, we need several adequate software and hardware supports that are appropriate for system needs.
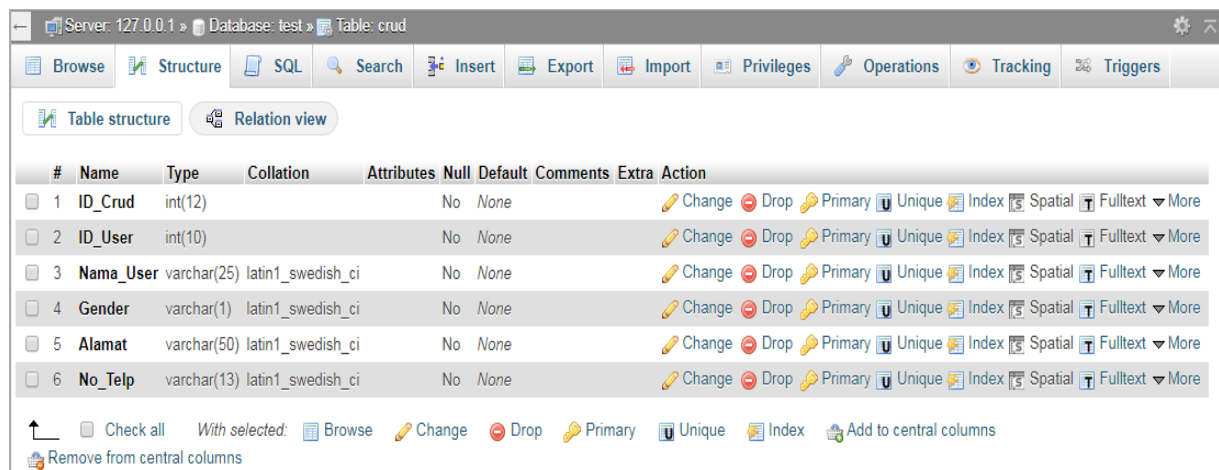
3.1. Making the program
Xampp is one the complete applications serving as software web server, it provides Apache, tomcat, mySQL. (See Figure 2).
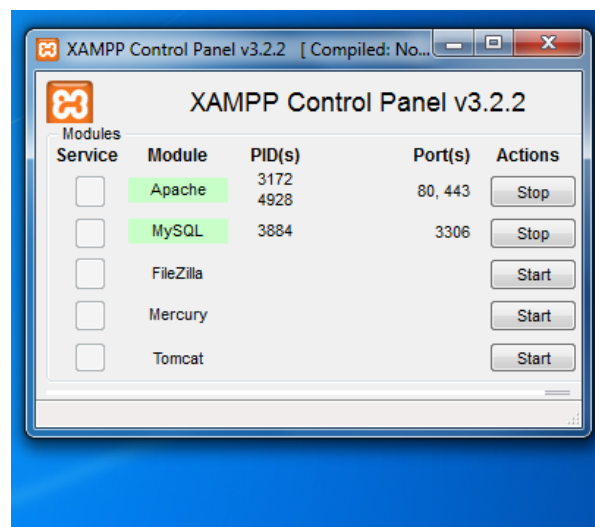
Figure 2. Xampp Application

3.2. Goods Order Report Page

In database design, we should also pay attention to its data structure to be more effective and efficient when adjusting the requirement with data type. Data structure is very important in computer systems because we can arrange and organize data in computer storage media in order to be used efficiently. The following is an example description of this aplication database design. (See Figure 3).

Figure 3. Structure Database Display

3.3. Program Results

3.3.1. Program making. The following display is an example for the data mining algorithm apriory program and serves just as an overview. (See Figure 4).
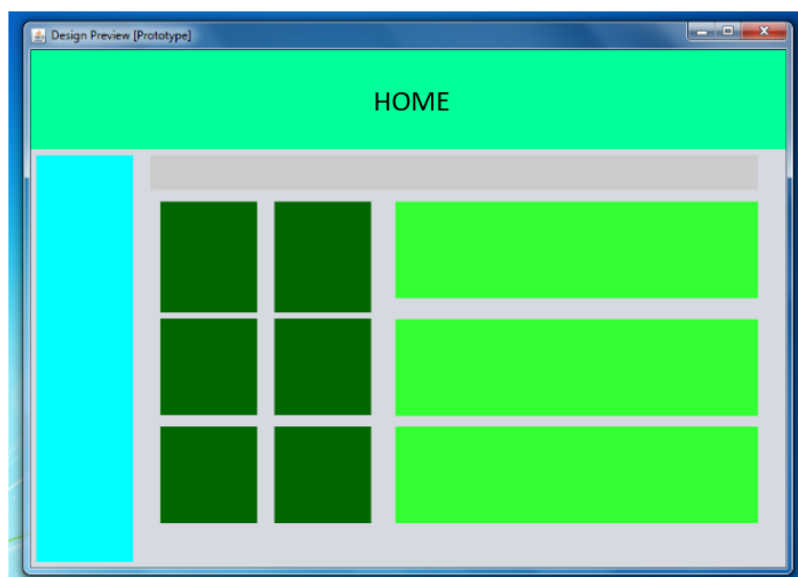

Figure 4. Xampp application

3.3.2. The algorithm. It works in the program by making support and confidence for a decision making material. Because in this algorithm, item support and confidence are very essential and cannot be

eliminated. So, two things are the items support and confidence must always include. Unlike the database scan algorithm that reads all the data in a new database (See Figure 5).
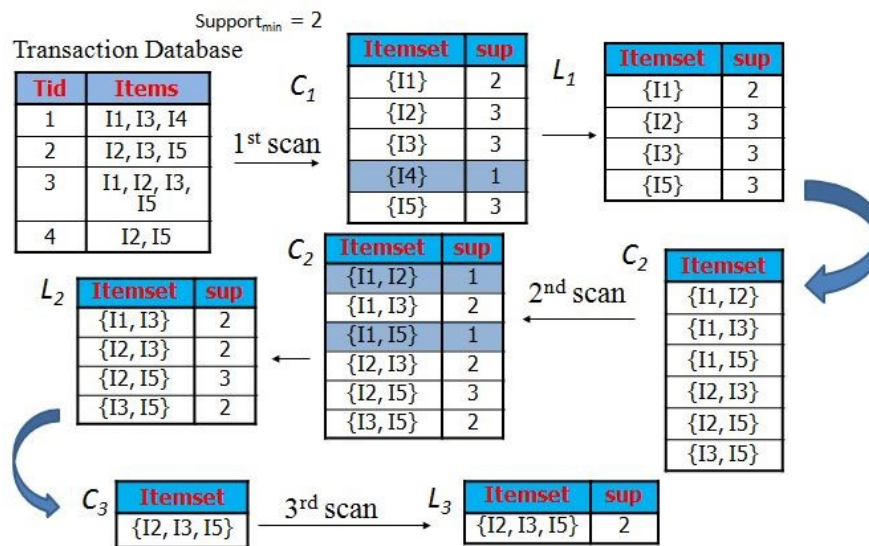

Figure 5. Algorithm Display

The program display for sellers featuring any product whose sales are significant and high. (See Figure 6).

Figure 6. Display Product Sales Considerations Purchase Form

3.3.3. Purchase Form. The customer order form which is available to customers who want to buy products from seller (See Figure 7).
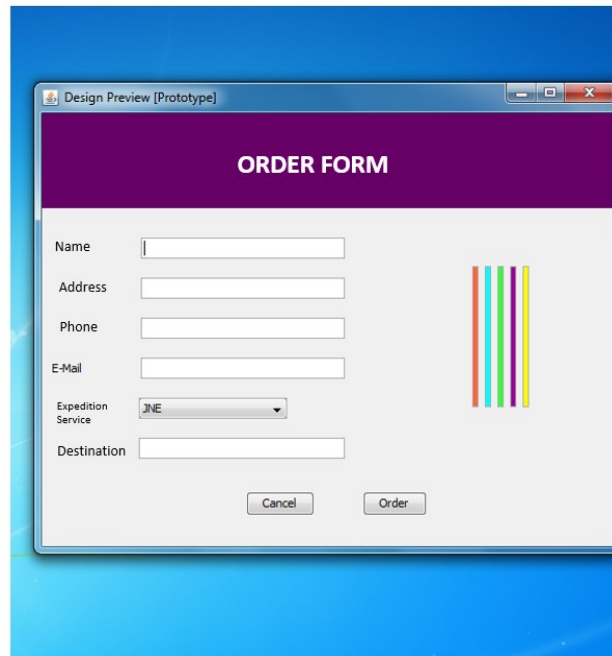

Figure 7. Purchase Form

3.3.4. Store Registration Form. Form to register an online store that no longer needs to buy a very expensive plot, especially for the size of retail that in fact the limitations of capital and this is a solution for retailers (See Figure 8).
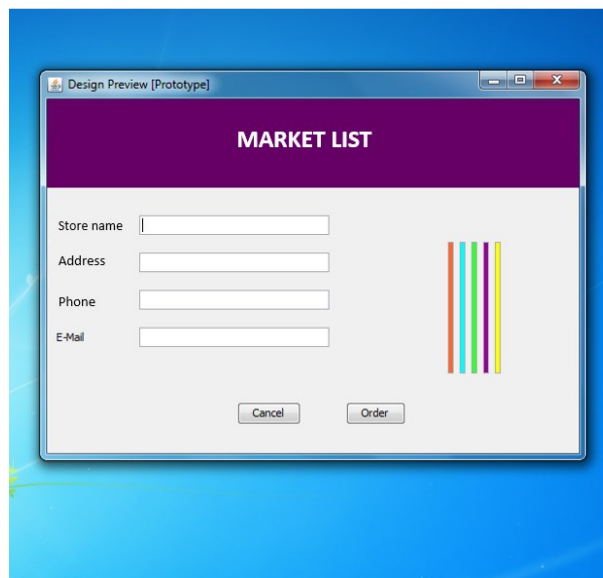

Figure 8. Report Page

4. Conclusion

The conclusion of the implementation of Data Mining, The existence of this system can help make decisions for the owner in order to maximize in stock provision, maximizing in marketing strategy with market share that can be seen from the segment of products purchased by consumers. Then there is the convenience of sorting products for both buyers and sellers.

References

[1]   Hirji K K 2001 Exploring data mining implementation. *Communications of the ACM,* **44**(7), pp. 87-93.

[2]   Hirji K K 2001 Exploring data mining implementation. *Communications of the ACM,* **44**(7), pp. 87-93.

[3]   Pérez M S, Sánchez A, Robles V, Herrero P, and Peña J M 2007 Design and implementation of a data mining grid-aware architecture. *Future Generation Computer Systems,* **23**(1), pp. 42-47.

[4]   Rygielski C, Wang J C, and Yen D C 2002 Data mining techniques for customer relationship management. *Technology in society,* **24**(4), pp. 483-502.

[5]   Chen,M S, Han J, and Yu P S 1996 Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering,* **8**(6), pp. 866-883.

[6]   Chiang K P, and Dholakia R R 2003 Factors driving consumer intention to shop online: an empirical investigation. *Journal of Consumer psychology,* **13**(1-2), pp. 177-183.

[7]   Wu M Y, and Tseng L H 2014 Customer satisfaction and loyalty in an online shop: an experiential marketing perspective. *International Journal of Business and Management,* **10**(1), pp. 104-114.

[8]   Chan C C H, Cheng C B, and Hsien W C 2011 Pricing and promotion strategies of an online shop based on customer segmentation and multiple objective decision making. *Expert Systems with Applications,* **38**(12), pp. 14585-14591.

[9]   Oentaryo R J, Lim E P, Finegold M, Lo D, Zhu F, Phua C, and Perera K S 2014 Detecting click fraud in online advertising: a data mining approach. *Journal of Machine Learning Research,* **15**(1), pp. 99-140.

[10]   Kanungo T, Mount D M, Netanyahu N S, Piatko C D, Silverman R., and Wu A Y 2002 An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence,* **24**(7), pp. 881-892.

[11]   Bansal, D., & Bhambhu, L. (2013). Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women. *International Journal of Advanced Research in Computer Science and Software Engineering*, **3**(9), pp. 1-10