

DECISION SUPPORT SYSTEM FOR POTENTIAL SALES AREA OF PRODUCT MARKETING USING CLASSIFICATION AND CLUSTERING METHODS

Evasaria M. Sipayung, Cut Fiarni, Randy Tanudjaya

Department of Information System, Harapan Bangsa Institute of Technology (ITHB),
Bandung, Indonesia
evasaria@ithb.ac.id, cutfiarni@ithb.ac.id

ABSTRACT

Increasing use of information system and technology in business especially in marketing, helping organization in many ways. Especially on big data era, where data of transactions already being kept by an organization and need to be analyzed to gain information and knowledge. Unfortunately, the marketing strategy based on product characteristic and transaction patterns are still not largely extracted and implement due to the different characteristics of data and information need to be analyzed first to find the most suitable methods in order to gain the intended knowledge. In this paper we concentrated on analyzing and implementing a decision support system to help gain marketing strategy in the certain sales area based on product's characteristics such as: type, material and accessories. The proposed system will gain pattern analysis of product characteristics and costumer's need as a basic source of knowledge in marketing decision making. This pattern is gained from sales transaction data, so the sales volume for each sales area can be predicted. A systematic model that implements clustering method of data mining technique is proposed to manage the marketing intelligence in order to support marketing strategy and decision. We used clustering method to form groups of data transaction and gained a correlation between potential product and sales area as a base method for business analysis. Then the proposed system will use classification method of data mining technique to create patterns for product correlation as based method for business intelligence for new products. This research concludes a prototype of sales area decision support system with rate of precision 98.7% and a rate of recall 70.27%.

Keywords: Decision support system, marketing strategy, data mining, clustering, classification.

1. INTRODUCTION

Restructuring marketing strategy is usually determined by the director based on the achievement of sales targets given previously. Moreover, in determining the target market, the director gives instructions for making sales based on the target, as well as shipping the same quantity of products in each sales area, which sometimes can lead to a buildup of product. Besides of problem on these product buildup, there is also a problem of shortage of the product in other sales areas. Sometimes the sales of certain product rapidly depleted in certain sales areas, but in another sales area the same products are still piled up. This is because the company does not have a clear knowledge of the needs of certain product in

certain sales areas, nor a prediction system of products characters in the specific sales area.

An immense amount of data transaction of of product can be used to gather market intelligence on consumers and characters of products in each of sales areas. By mining data transaction, companies can analyze top product to gain patterns of its characters and purchasing pattern on each sales areas. By gaining these patterns and knowledge can overcome problem such as build up of products and shortage of product in certain sales areas. In particular, often this works using a clustering method of data mining technique, where the goal is to build a predictive model that can label data item to find it proportional cluster in order to gain knowledge of marketing intelligence

(Sipayung dkk., 2014). This knowledge is fundamental in creating a marketing strategy that is more effective and efficient.

We address this issue in this paper by presenting an integrated framework for knowledge extraction and prediction system, in the context of marketing intelligence by using data mining technique for the proposed decision support system. In this paper the analysis focused on the development of the information into a knowledge of sales areas and the marketing the product, so it can determine the characteristics of the potential products that could be used for prediction of new products on certain sales areas. On the proposed system, we use clustering and classification methods of data mining technique for extracting pattern of data transaction to gain rule for marketing prediction.

2. THEORETICAL BACKGROUND

Data mining is a process of extracting potentially useful information from raw data. A software engine can scan large amounts of data and automatically report interesting patterns without requiring human intervention (Rajagopal, 2011). Companies uses mining tools and techniques to find useful relationships, patterns and anomalies that can help managers make better business decisions. Data mining tools perform analyses that are very valuable for business strategies, scientific research and for companies is to getting to know company's customers better. Managerial insights are no longer the only factor trusted when it comes to decision-making. Data driven decisions can lead to better firm performance (Gancheva, 2013). Data mining process is divided into three phases: pre-processing, data mining and post-processing as shown in Figure 1.

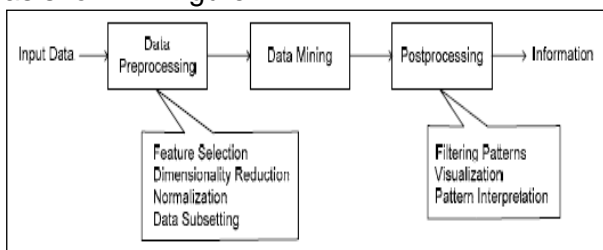


Figure 1. Phases of Data Mining

As illustrate on the Figure 1, in the first stage the input data must undergo a process of data pre-processing, where the data are arranged so it can be ready to be processed by the existing methods of data mining. The techniques of data pre-processing are feature selection, normalization, subsetting the data and dimension reduction. In the second phase, the data is processed using association rules, clustering, or classification method of data mining technique. At the last phase, the data was carried out by data mining techniques, entering the post processing, interpretation of data use visualization, filter patterns, pattern interpretation, and others, to gain information and extracted knowledge.

3. RESEARCH METHOD

The study research is to analyzed the most suitable data maning techniques to be used as algorithm model of the Decision Support System For Potential Sales Area of Product Marketing. The following data that used in this research is taken from CV.XYZ, a toy company located in Bandung, Indonesia and has several distribution stores. The toy company produces more than 200 variations of toys, and it's sold in all Indonesian regions, in 144 stores. The company also has several sales areas, which in each sales area there are several store that divided based on Indonesian regions.

3.1. Problem Analysis

In order to analyze the most suitable data mining techniques, first we will the marketing product and sales areas problems as shown on figure 2 and figure 3. As illustrated in Figure 2, there are 144 stores, the product analysis for product with code "E501246FF3" shown that each store has different sales quantities.

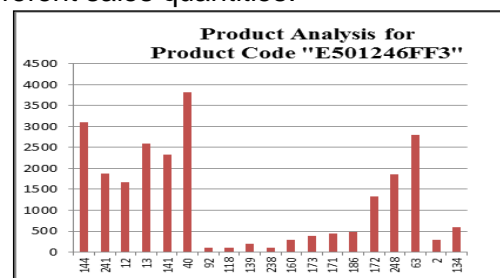


Figure 2. Product Sales Analysis

In addition, product sales also analyzed on the return of products that result in the buildup of the product for the same product code that is shown in Figure 3.

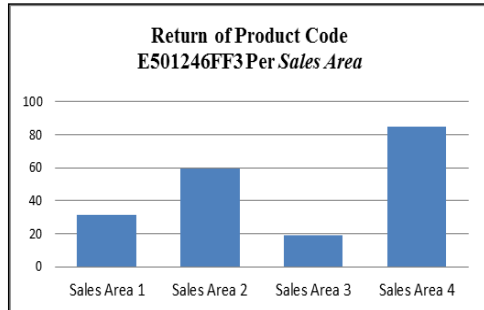


Figure 3. Return of Product per sales area

From the problem analysis, it was apparent that differences in the level of sales and the buildup to the specific product code on a sales area. Given such issues, the company has acquired three datasets that will be used to analyze the most suitable data mining techniques as follows: sales and transaction dataset, products, and stores dataset.

In the analysis of proposed solution system, the main features required are a few key features and support that is needed on the part of decision support system, such as:

- Find Pattern between frequency and quantity of transaction on certain sales area within specified timeline in year.
- Form Grouping of potential products and sales areas
- Analysis of the correlation between characteristics of the product group sales of on certain sales areas.
- Recommendations characteristics of new products that have the marketing potential to a particular sales area.

The results of an analysis of the objectives to be achieved are two strategic applications that are divided into business analytics and business intelligence. Business analytics solutions feature an application that concentrates on making the analysis of the points a, b, c of the key features of the proposed system. While the business intelligence features of the solution application features concentrate on points d of the key features of the proposed system.

But in order the proposed system to be able on making business analytics and business intelligence, data processing methods are needed. Because of the proposed system will find patterns and extracted knowledge from three datasets as explained before, therefore, the most suitable chosen method of proposed system is data mining. Data mining technique that address the needs of the functions shown are clustering for the first 3 of key features and classification for the last key features of the proposed solution system.

3.2. Architecture of Proposed System

In this research in order to develop the integrated framework for knowledge extraction and prediction system in the context of marketing intelligence by using clustering and classification method of data mining technique, we proposed a decision support system for the potential sales area of the product market. The architecture of the proposed system illustrates on figure 4.

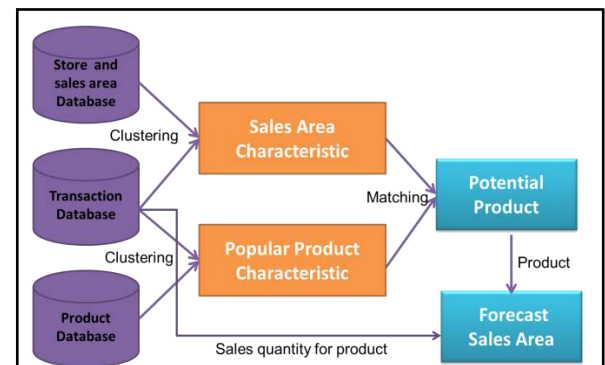


Figure 4. Architecture of Proposed System

Finally, as illustrate from figure 4, knowledge gain from potential product and transaction database to gain pattern in order to forecast product for each sales area. Marketers are interested in knowing how various marketing programs affect future sales of their product (Shaw dkk., 2001). In order to gain knowledge for forecast sales area, the proposed system will use visualization tools to identify trends in sales, character of potential products, and marketing sales area.

3.3. Data Mining Methods

In this section we will explain the data mining technique used as an algorithm

model of the proposed system as shown on the figure 4, which is clustering and classification method. In this phase we generate the cluster to profiling product characteristic and its relationship with transaction on each sales area. While learning relationship between product character, transaction and marketing profile of sales area, marketer gain knowledge of forecast marketing sales area, with this following phase:

1. Frequency and quantity of purchase product of each sales area. This concern with how often does the customer of each store in each sales area buy product and how much the customer spends on a typical transaction. This information helps marketers to gain knowledge of sales area characteristic
2. Product character and transaction on each sales area. This information gain with modelling attribute data with scheme definition as explained on the previous section. This information will help the marketer to gain knowledge of popular product characteristic.

3. Clustering Potential Product.

In this phase clustering was done to store and sales area database also to transaction database and then we use matching approach to form potential product. We used K-Means Algorithm for the clustering methods of data mining. This phase given result 3clusters, in which cluster 1 is the list of products that have sale above average (high revenue), cluster 2 give the list of products with average sale, and cluster 3 is the list of products with sale below average.

4. Identifying Potential Product.

The quantity sales for each product ID is calculated within one year, from the total sales gained will be calculated for each product presentation in one year. Presentation of the results of each product are sorted from largest then calculated cumulative to 70%. Based on data analysis has been conducted, the potential product is a product that gives value to the amount of 70% of total sales and 70% value is the value of the total quantity of products sold from existing sales in the current year. There are 32 potential products from 217 products were

sold in 2012. Potential products details are in Table 1.

Table 1. Potential Product of 2012

No	Product ID	QTY	%	Σ%
1.	E50124FF3	13770	6.42	6.42
2.	E501247FF3	10878	5.07	11.50
3.	E820707F\$KB	9180	4.28	15.78
4.	E820705F\$KB	8280	3.86	19.64
5.	E5016B57FF	7560	3.53	23.17
6.	E5016B59FF	7020	3.27	26.44
7.	E820701F\$KB	7020	3.27	29.72
8.	E613104F	6719	3.13	32.85
9.	E820708F\$KB	6300	2.94	35.79
10.	E521210F6	5684	2.65	38.44
11.	E5216B10F6	5394	2.52	40.95
12.	E501245C\$3KB	5346	2.49	43.45
13.	E501247C\$3KB	5214	2.43	45.88
14.	E501238FF3	4606	2.15	48.03
15.	E701601F	4556	2.13	50.15
16.	E613105F	4028	1.88	52.03
17.	E600201FC	3894	1.82	53.85
18.	E600105C	3360	1.57	55.42
19.	E620201C	3328	1.55	56.97
20.	E5016B57C\$KB	3150	1.47	58.44
21.	E620501H	3036	1.42	59.85
22.	E5016B59C\$KB	2340	1.09	60.94
23.	E800601F	2194	1.02	61.97
24.	E880201H	2180	1.02	62.99
25.	E931301F	2144	1.00	63.99
26.	G932201F\$	2132	0.99	64.98
27.	E840104FC	1876	0.88	65.85
28.	E600300F	1716	0.80	66.66
29.	E701104F	1716	0.80	67.46
30.	E5016B50FF	1710	0.80	68.25
31.	E702201F	1700	0.79	69.05
32.	E502415H6	1632	0.76	69.81
33.	E920404F	1440	0.67	70.48

Product ID E920404F is not included in potential product because it's cumulative is more than 70%.

4. RESULT AND DISCUSSION

In this section we will describe the design and implementation of the proposed system based on the architecture as explained in the previous section. This process includes not only the actual writing programming code and train data set, but also the preparation of requirements, the system design, and confirmation that the proposed marketing intelligent system has met the research objectives.

4.1. Design System

Before starting the system design, analysis is required to determine the attributes of the product. The attribute is determined by the existing code of the product. These codes of product determine which types of products, materials and accessories owned by a certain code. Therefore, the decoding is done based on

the length and type of the product code of the character of the product. Based on the architecture of proposed systems we build a UML use case diagram (figure 5) and activity diagrams (figure 6) to illustrate how the system operates.

Based on the use case as shown on figure 5, there are several features that are available to address issues that have been defined, namely the analysis of potential product features and new product forecast (decision support system). In addition, there are some additional features that are used by systems such as master data management product.

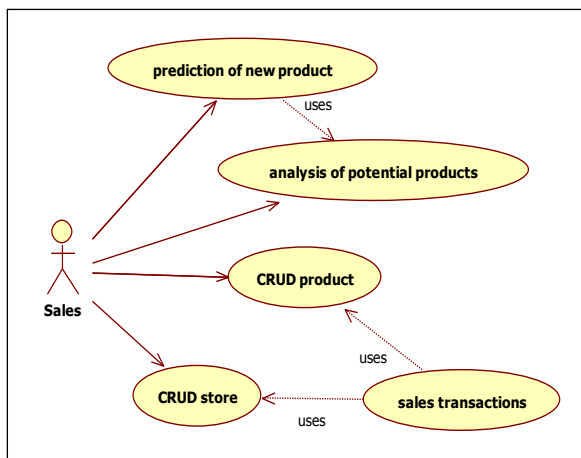


Figure 5. Use case Diagram

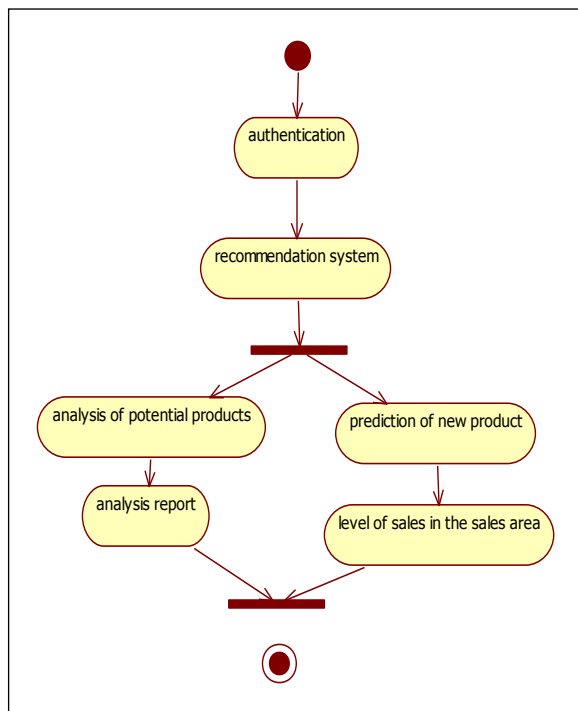


Figure 6. Activity Diagram

4.2. Implementation

The proposed marketing intelligence system developed based on architecture as described in Figure 8. The basic functional requirement of the proposed system are:

1. Authentification of user
2. Input data for product, store and transaction
3. Range time of marketing analysis for product potential
4. Visualize Frequency and Quantity of Sales Product
5. Marketing Intelligence

The next step of system development is interface design of proposed system. Interface design is the process of defining how the system will interact with external entities pleasantly. For this proposed system, the external entities are marketer analysis, sales representative and administrator. Figure 7 shows the user interface for the home page of the system. This home page have identification, help and contact features. Identification is important for market intelligence because it gives marketing trend and forecasting, that lead to strategic decision.

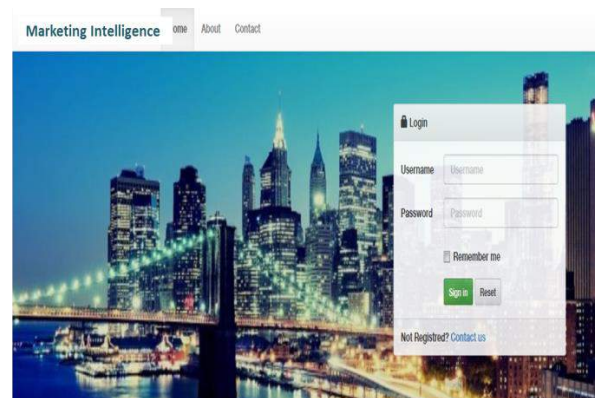


Figure 7. Home Page of Marketing Intelligence System

Figure 8 shows the range time of marketing analysis for product potential. This feature also visualizes the product analysis base on sales performance in each sales area. Marketing intelligence as illustrate on Figure 9 shows marketing intelligence on characters of the product. This feature will analyze not only attributes of the product, but also a potential product in their

relationship with sales area. This feature also used as forecasting potential product of sales area, and can be used as strategic marketing decision.

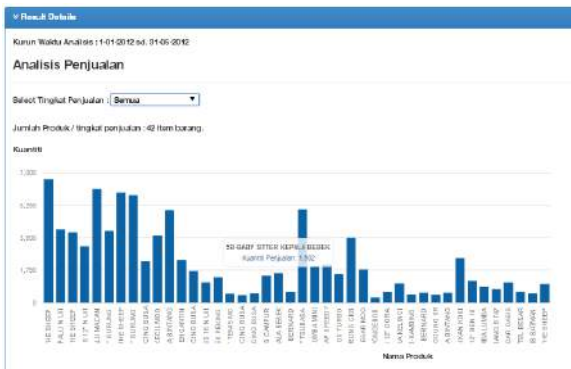


Figure 8. Visualize Frequency and Quantity of Sales Product

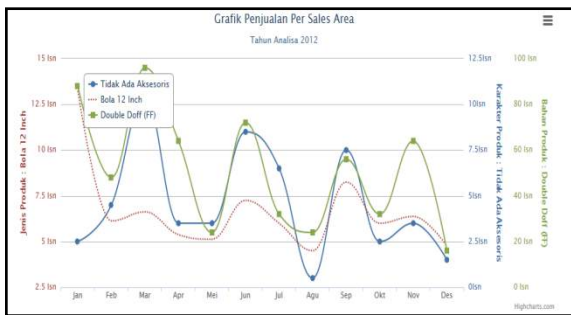


Figure 9. Marketing Intelligence

4.3. The Testing of Proposed System

The results of the implementation of the system run on the Java programming language using Java version 1.7 and supported by a MySQL database version 5. Potential products based on the clustering using WEKA tool. There are 3 clusters, total product per cluster instances: cluster 0-17 item product (53%), cluster 1-6 item product (19%), and cluster 2-9 item product (28%) Based on the results of the implementation and testing of the system, the amount of testing data that being used are 184 items.

$$\begin{aligned}
 \text{Recall [e, f]} &= \frac{ce}{ce + te} \times 100\% & (1) \\
 &= \frac{130}{185} \times 100\% = \mathbf{70.27\%}
 \end{aligned}$$

$$\begin{aligned}
 \text{Precision [f, g]} &= \frac{ce}{ce + fe} \times 100\% & (2) \\
 &= \frac{182}{185} \times 100\% = \mathbf{98.37\%}
 \end{aligned}$$

Where ce is the number of entities extracted correctly, te is the number of true entities not extracted, and fe is the number of false entities extracted

5. CONCLUSION

This study utilize data mining technique to gain knowledge of potential sales area and potential product based on characters of product. In the proposed solution system the attributes of product character are : the type of products, materials, and accessories. These attribute and transaction dataset implement on the decision support system that uses a data mining clustering method to obtain: correlation of products with a sales area, so that the known products frequently purchased by a sales area in a certain timeline and grouping of potential products and sales areas, so that could gain knowledge and form group of potential product and product character in a particular area. Then the proposed system will used classification method on Recommendations characteristics of new products that have the marketing potential to a particular sales area. The level of accuracy of the proposed decision support system that analyzes the attributes of the product type, materials and accessories using the test sensitivity and specificity produces precision accuracy rate of 98.37% and recall of 70.27% accuracy rate.

6. REFERENCES

- (a) Sipayung E.M., Fiarni Cut, Tanudjaya R., 2014. "modeling data mining dynamic code attributes with scheme definition technique", *Proceeding electrical engineering, computer science and informatics 2014*.
- (b) Rajagopal S., 2011. "customer data clustering using data mining technique", *International Journal of Database Management Systems (IJDMSS) Vol.3, No.4, November 2011*
- (c) Gancheva, V. 2013. Market Basket Analysis of Beauty Products. *Thesis on Erasmus University Rotterdam*,
- (d) Shaw, M.J, Subramaniam C. ,Tan G.W., and Welge M.E., 2001. "knowledge management and data mining for

marketing”, Elsevier, Decision Support System 31, 127-137.

- (e) B. Adelberg. 1998. NoDoSe: A tool for semi-automatically extracting structured and semi-structured data from text documents. *In ACM International Conference on Management of Data (SIGMOD)*.
- (f) Hoda Waguih. 2013. A Data Mining Approach for the Detection of Denial of Service Attack. *IAES International Journal of Artificial Intelligence 2 (2)*, 99-106.
- (g) D. Angluin. 1978. On the complexity of minimum inference of regular sets. *Information and Control 39 (3)*, 337–350.

AUTHOR BIOGRAPHIES

Evasaria Magdalena Sipayung is a lecturer in Department of Information System, Institut Teknologi Harapan Bangsa (ITHB), Bandung. She received her Master of Information Technology from Institut Teknologi Bandung in 2007. Her research

interests are in the area of Software Engineering and Database. Her email address is <evasaria@ithb.ac.id>

Cut Fiarni is a lecturer in Department of Information System, Institut Teknologi Harapan Bangsa (ITHB), Bandung. She received her Master of Information Technology from Institut Teknologi Bandung (ITB) in 2003. Her research interests are in the area of Data Mining, Decision Analysis, Intelligence System and IT Governance. Her email address is <cutfiarni@ithb.ac.id>

Randy Tanudjaya received his Bachelor Degree of Information System from Institut Teknologi Harapan Bangsa (ITHB) in 2014. His research interests are in the area of Data Mining, Customer Relationship management, and Database Management. His email address is <randy.tanudjaja@live.com>