

# An application of data mining techniques in designing catalogue for a laundry service

*Annisa Uswatun Khasanah\**, *Deliana Ardhitama Erlangga*, and *Ahmad Mustopa Jamil*

Department of Industrial Engineering, Universitas Islam Indonesia

**Abstract.** Catalogues are the media that companies use to promote their products or services. Since catalogue is one of marketing media, the first essential step before designing product catalogue is determining the market target. Besides, it is also important to put some information that appeal to the target market, such as discount or promos by analysing customer pattern preferences in using services or buying product. This study conduct two data mining technique. The first is clustering analysis to segment customer and the second one is association rule mining to discover an interesting pattern about the services that commonly used by the customer at the same service time. Thus, the results will be used as a recommendation to make an attractive marketing strategy to be put in the service catalogue promo for a laundry in Sleman Yogyakarta. The clustering result showed that the biggest customer segment is university student who come 3 until 5 times in a month on weekends, while the association rule result showed that clothes, shoes, and bed sheet have strong relationship. The catalogue design is presented in the end of the paper.

## 1 Introduction

Marketing is an important activity that must be done by a company in order to maintain the business continuity process. According to Stanton [1], the definition of marketing is an overall system of business activities aimed at planning, pricing, promoting and distributing goods or services that satisfy the needs of both existing and potential buyers. Marketing is a company activity which is directly related to the consumer and the main purpose of this activity is to maximize the company's profits. Marketing can be conducted with several ways and the common one that have been used by lot of companies is using catalogue.

Catalogues are the media that companies use to promote their products or services. Catalogues are used to describe in detail the products or services the company offers to consumers. It usually produces in a given period, in which each period issuing different offers. Offers are usually associated with new products or services, product or service promos, discounts, or special events. In designing a catalogue, there are several things that must be considered by a company. Beside catalogue give the consumer detail information about products or services, such as figure, price, detail specification, contact person, and so on, a catalogue should be also contained some information that appeal to consumers. To create an attractive product catalogue, company should know who the customer is, and what the customer wants. If the company can design a good catalogue, it can encourage the customer to look inside the catalogue, and then entice them to buy the products or use the services offered.

One important thing to be succeed in marketing is knowing the market target [2]. Since catalogue is one of marketing media, the first essential step before designing product catalogue is determining the market target. This step can be conducted through customer segmentation. Segmentation is greatly embedded in a marketing strategy because different customer groups will require a different marketing mix [3]. Customer or market segmentation is the division of the market into different customer groups where specific customer groups can be selected as market targets to be achieved with a specific marketing mix [4]. After knowing the consumer group who became the target market then the marketing strategy, in this case is the catalogue contain and design, can be adjusted.

In designing product or service catalogue, it is very important to put some information that appeal to the target market, such as discount or promos. Usually a company will see the pattern of consumer buying in giving promo or discount. By analysing the consumer buying pattern, or in the service industry is about the services that commonly used by the customer, the company can get valuable information to make the right marketing strategy. Examples of discounts or promos given to consumers are usually a discount on a good or service if it is bought or used in conjunction with another item or service.

The purpose of this study is to give suggestion related with the contain and the design of a service catalogue by conducting customer segmentation and analysing the services that commonly used by the customer at the same service time. This kind of analysis can be obtained using data mining techniques. Data mining is the process of automatically discovering useful information in large data repositories [5]. Data mining has been used widely in

\* Corresponding author: [annisa.uswatun@uii.ac.id](mailto:annisa.uswatun@uii.ac.id)

many different scopes of study and one of them is in marketing. Data mining has several popular techniques such as clustering analysis and association rule mining. There are many methods that can be used in clustering analysis, one of the most popular method is K-means algorithm. In this study clustering analysis, using K-means algorithm, is used to conduct the customer segmentation, and association rule is used to discover an interesting pattern about the services that commonly used by the customer at the same service time, then those results will be used as a recommendation to make an attractive marketing strategy to be put in the catalogue. This study was conducted in a laundry in Sleman, Yogyakarta. The rest of this paper is organized as follows, Section 2 present the literature review, Section 3 presents the research method related to this study, Section 4 shows the results and discussion. The concluding is finally made in Section 5

## 2.1 Clustering Analysis using K-means Algorithm

Clustering analysis is the process of identifying natural groupings or clusters within multidimensional data, based on some similarities, like Euclidean distance [6]. Its main purpose is to group samples with the same statistical characteristics together into same cluster in order to achieve higher similarities within same cluster, also there are more significant differences between different clusters [7]. There are many methods that usually used to deal with clustering problems, however K-means is the most widely used clustering method even though it was proposed over 50 years ago [8].

K-means was firstly proposed by MacQueen in 1967 [9,10]. K-means algorithm is also well known for its efficiency in clustering large data set [11]. K-means algorithm is easy to be implemented and computationally efficient [5,10,12]. The K-means algorithm starts with K cluster centroids, which are initially randomly selected or derived from some a priori information. Each point in data set is then assigned to the closest centroid, and each collection of points assigned to a centroid is a cluster. The centroid of each cluster is then updated based on the point assigned to each cluster. This process is repeated until no point change clusters, or equivalently, until the centroids remain the same [5].

Some scholars have already used K-means algorithm to conduct customer segmentation [13-15]. Hosseini et al. [13] used K-means to classify customer product loyalty under business to business concept in a company in Iran. It used K-means algorithm with K-optimum according to Davies-Bouldin Index. The results show that with this method the company was able to assess customer loyalty with a better marketing strategy compared to most other companies in Iran. Hong & Kim [14] integrated self-organizing map and K-means algorithm to segment customer based on customer's psychographic data for an online store to offer customized marketing. Wei et al. [15] also combined K-means algorithm and self-organizing map (SOM) to segment customer and develop marketing strategy in a hair salon in Taiwan. The results indicated

four customer types including loyal customers, potential customers, new customers and lost customers. The marketing strategy for each customer types was also developed

## 2.2 Association Rule Mining

Association rule mining can be illustrated as follows, suppose that the set  $I=(i_1,i_2,\dots,i_k)$  is the total number of items. The item set X and Y are part of the total set I. If Y happen in one transaction when X happens, as association rule can be expressed as [16]. In market basket analysis, association rule can be can be express as "A customer who buys product X1 and X2 will also buy product Y with probability c%" are found [17]. For example, in a restaurant egg and coffee may be ordered together primarily between 7 am and 11 am, and maybe turkey and pumpkins are sold at the same time in the week before Thanksgiving [18]. In service industry, association rule mining can be express as a customer who use a certain service will also use another service at the same service time.

The strength of an association rule can be measured in terms of its support and confidence. Tan et al. [5] explain that support determines how often a rule is applicable to a given data set. A low support rule is uninteresting from a business perspective and sometimes it will be eliminated. While, confidence determines how frequently items in Y appear in transaction that contain X, the higher the confidence the more likely it is for Y to present in transaction that contain X. The formal definitions of these metrics are

$$\text{Support}, s(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (1)$$

$$\text{Confidence}, c(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2)$$

Some scholars have applied association rule mining. Abdulsalam et al [19] implement association rule mining which is also known as market basket analysis to reveal sales pattern in a supermarket. This study showed rules of purchasing one product would lead to the purchase of another product. This study concluded that apple and chocolate had strong correlation. This information helps the company to create marketing and advertising strategies that outshine the competitors. Verma et al. [20] apply association rule mining to find out the pattern incidents in a steel plant in India. Czibula et al. [21] used association rule to predict defect of a software system in order to conduct continuous improvement. Kumar, & Toshniwal [22] apply data mining techniques to analyse data related with road accident in India. Cluster analysis was used to cluster the accident data, then association rule mining was used to identify the situation when an accident occurred in each cluster. The results can be used to make some preventions for each different categories, thus it can minimize the number of accidents happened.

### 3 Research Method

The research was conducted in laundry “A” in Sleman Yogyakarta. The data consist of 100 data from customer of the laundry and collected using questioner. This research consists of three steps, the first step is clustering, the second step is association rule mining, and the third step is designing the content of service catalogue that contain an interesting information based on the result of clustering and association rule analysis. Clustering is used to segment the customer and association rule mining is used to know the several services of the laundry that commonly used by the customer at the same service time. The laundry that become the object of this research offer laundry service such as clothes, bed sheets, blankets, shoes, dolls, bag, curtain, prayer kit, and so on. The result from clustering and association rule than will be used to create suitable business strategy for a certain customer segment. The business strategy than will be put in the service catalogue of the laundry. Software SPSS was used to apply K-means clustering and Rapid Miner was used to apply association rule mining. The research step for this study is represented in Figure 1.

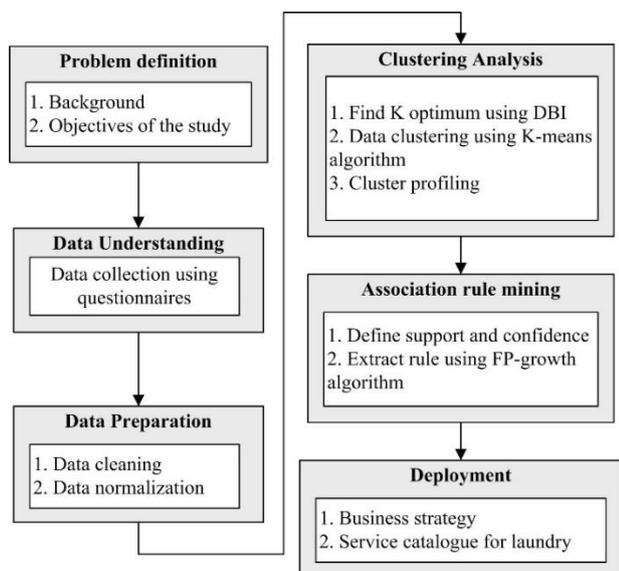


Fig 1. Research step

### 4 Result and Discussion

#### 4.1 Clustering Analysis

In this step, laundry customer was segmented based on several attributes including occupation, arrival frequency, arrival day, arrival time, and laundry weight. Several number of cluster (K) were implemented and Davies Bouldin Index (DBI) was used to evaluated the performance. Lower DBI index indicates better result. Table 1 shows the DBI value for each K.

Based on those result, K=3 has the smallest value of DBI, it means that the best cluster number for this data is 3. Thus for the next step, the data is divided into 3 clusters. The cluster centroid and the Euclidean distance were calculated by the software. The results showed that cluster

1 consist of 59 cluster members, cluster 2 consist of 38 cluster member, and cluster 3 consist of 3 cluster members. Table 2 gives the detail cluster profile.

Table 1. Davies Bouldin Index for K optimum

Number of cluster (K)	DBI
2	1,126
<b>3</b>	<b>1,022</b>
4	1,210
5	1,338

All of the member in cluster 1 are university student and dominated with customer with arrival frequency 3 until 5 times per month, come in weekend (Friday, Saturday, and Sunday), and the weight of the laundry is around 4 until more than 5 kg. Gender is not included as the attribute in this analysis but for additional information cluster 1 is dominated with female. Cluster 2 is also dominated with university student who come in weekdays (mostly on Monday, Wednesday, and Thursday) with weight of the laundry less than 3 kg. For cluster 2 there is no dominated arrival frequency. Half of the customer come 3 until 5 times per month, half of them come once until twice per month. It dominated with male. Meanwhile, cluster 3 is dominated with businessman, and civil servant. There is no dominated arrival frequency, day, and laundry weight for this cluster. In all cluster, there is no significant difference or dominated arrival time (morning, afternoon, evening). Thus, arrival time is not important attribute to be considered in this analysis.

The clustering result shows that the biggest customer segment (more than 50%) is university student who mostly come in weekend, with arrival frequency 3 until 5 times per month and weight of the laundry around 4 until 5 kg. This result is an important pattern than can be a useful information to create the business strategy.

#### 4.2 Association rule mining

This step was conducted to reveal the customer pattern in using services offered by the laundry. To get the information, the customer was asked related with what kind of services that they used by filling the questioner. The data was transform into binary form, 0 for not using the service and 1 for using the service. Then the binary data was put into Rapid Miner software. Minimum support used was 5% and minimum confidence used was 5%. There were several rules generated, however only 8 rules had lift ratio bigger than one. Lift ratio measure the correlation between item set and it can use to measure the importance of the rule. Greater lift ratio indicates stronger associations, the highest lift which means highest correlation [23,24]. From the 8 rules, there are 2 rules which have high support and confidence value, as follows

- bed sheet, shoes ⇒ clothes (s:10%, c:91%)
- clothes ⇒ shoes (s:10%, c:90%)

Table 2. Cluster profile

Cluster	Occupation	Arrival frequency	Arrival day	Arrival time	Laundry weight
Cluster 1 (59 member)	University Students (100%)	More than 3 times a month (71%)	Weekend: Friday, Saturday, Sunday (83%)	There is no domination	Minimum 4 kg (55%)
Cluster 2 (38 member)	University Students (100%)	less than 3 times a month (58%)	Weekdays: Monday, Wednesday, Thursday (53%)	There is no domination	1 kg (32%) 2-3 kg (68%)
Cluster 3 (3 member)	Businessman and civil servant	There is no domination	There is no domination	There is no domination	

The first rule means 10% of all transaction contain bed sheet, shoes and clothes and 91% of the transaction have the item bed sheet and shoes also contain clothes. While, the second rule means 10% of the transaction contain clothes and shoes, and 90% of the transaction have item clothes also contain shoes. This result shows that the tree items clothes, shoes, and bed sheet have strong relationship.

### 4.3 Service catalogue design

This step is conducted to design the content of laundry “A” catalogue service that contain an interesting information based on the result of the clustering and association rule analysis. Catalogue should inform the detail information about the service offered and also some interesting information which can attract the customer to use the service. This study is not designing the whole catalogue service, but only the attractive part of the catalogue that contain promos.

From the previous analysis, it can be known some important information. Not all the information is used on designing the contain of the catalogue, but only the most attractive one. Clustering analysis show that the biggest customer segment (more than 50%) is university student who mostly come on weekends with weight of the laundry 4 until 5 kg for each customer. The attractive information here is the weekend; it means that this laundry will be very busy on the weekend. To create an attractive business strategy and to reduce workload on weekend this laundry can give a discount for student who come on weekdays. By analysing the data, it is known that customer rarely come on Tuesday. So, this laundry can give discount for student who come on Tuesday with minimum laundry weight 4 kg.

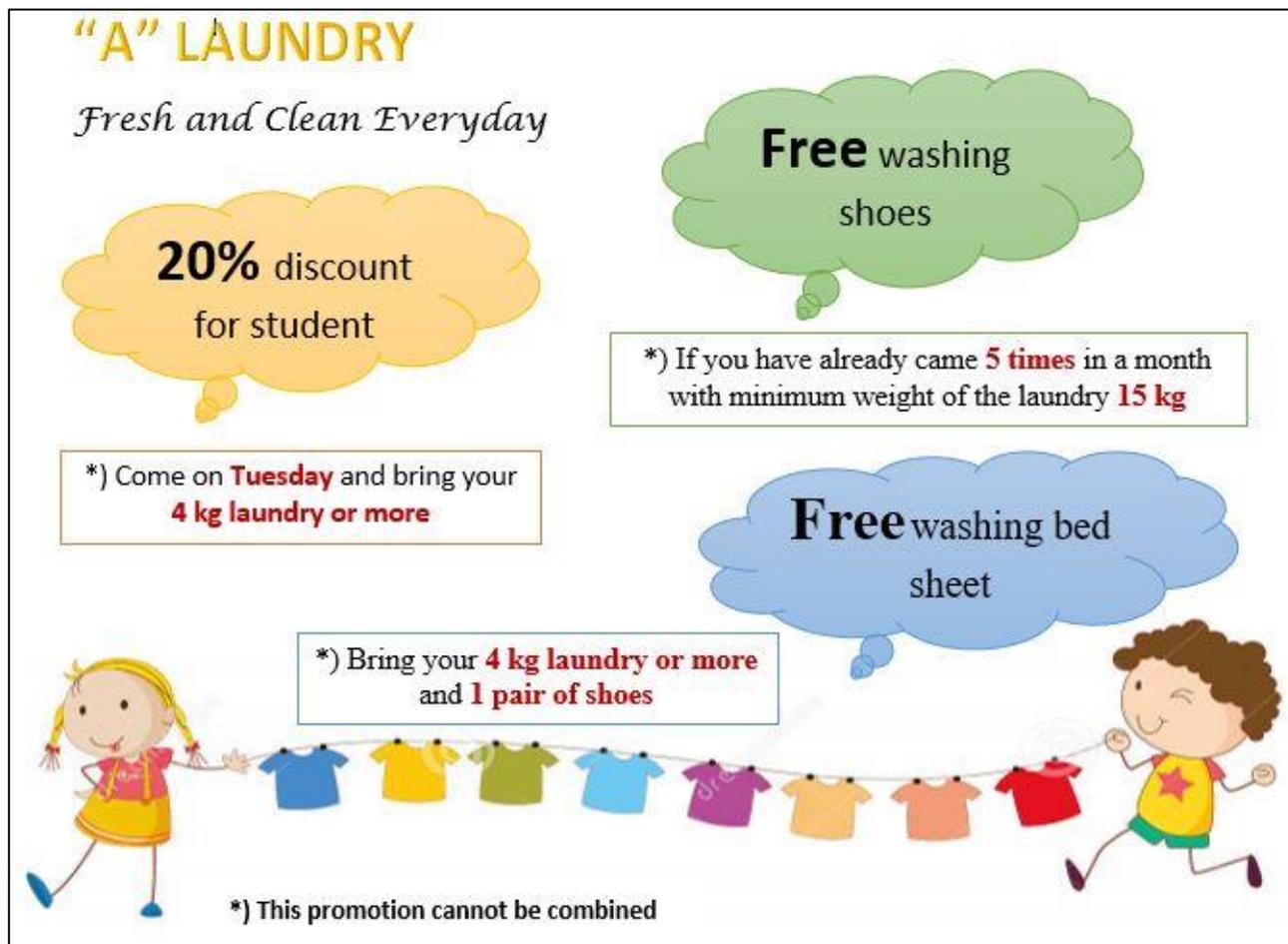
From the association rule mining, it is known that clothes, shoes, and bed sheet has strong relationship. To create an attractive business strategy, by combining the clustering result, this laundry can make a packet promo for this three items. For example, free washing shoes for customer who have already came to wash the clothes 5 times in a month with minimum weight of the laundry 15 kg or free washing bed sheet for customer who wash clothes minimum 5 kg and one pair of shoes. The illustration of the catalogue service promo is shown in Figure 2.

## 5 Conclusion

From the clustering result it can be concluded that the biggest customer segment is university student who mostly come 3 until 5 times in a month on weekends and with weight of the laundry 4 until 5 kg. From the association rule it can be concluded that clothes, shoes, and bed sheet have strong relationship. These important results then used to create service catalogue promo as illustrated in Figure 2. For further research it is better to conduct preliminary study to determine the attributes in order to get a good clustering result. It is better to have more than 100 respondents so the information about customer preferences in using services or products will be better. This kind of analysis can be developed in other service industry or product and can help the owner of a business to create a suitable business strategy for the target market which can beat the competitors.

## References

1. W. J. Stanton, Prinsip Pemasaran, Erlangga, Jakarta, (2001)
2. J. Chance, 8 Keys to a Strong Marketing Strategy, Online access (July 28 2017), URL: [https://www.businessknowhow.com/marketing/block\\_s.htm](https://www.businessknowhow.com/marketing/block_s.htm),(2017)
3. M. S. Kahreh, M. Tive, A. Babania, M. Hesani, Analyzing the applications of customer lifetime value (CLV) based on benefit segmentation for the banking sector. *Procedia-Social and Behavioral Sciences*, 109, 590-594, (2014)
4. P. Kotler, Marketing Management: Analysis, Planning, and Control, 4th edition, Prentice-Hall, Englewood Cliffs, N.J, 195, (1980)
5. P.N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Pearson Education, MA, (2006)
6. A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, *ACM Computer Survey*, 31 (3), 264-323, (1999)
7. J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2nd edition, Morgan Kaufmann, (2006)
8. A.K. Jain, Data clustering: 50 years beyond K-means. 19th International Conference in Pattern Recognition (ICPR), 31 (8), 651-666, (2010)



**Fig 2.** Service catalogue promo  
(Source of the kids holding the tip of the wire:[25])

9. Z. Cai, W. Gong, C.X. Ling, H. Zhang, A clustering-based differential evolution for global optimization, *Applied Soft Computing*, 11, 1363-1379, (2010)
10. W. Kwedlo, A clustering method combining differential evolution with the K-means algorithm, *Pattern Recognition Letters*, 32, 1613-1621, (2011)
11. Z. Huang, Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values, *Data Mining and Knowledge Discovery*, 2, 283-304, (1998)
12. S. Das, A. Abraham, A. Konar, Automatic Clustering Using an Improved Differential Evolution Algorithm. *IEEE Transaction on System, Man, and Cybernetics*, 38 (1), (2008)
13. S. M. S. Hosseini, A. Maleki, M. R. Gholamian, Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, *Expert Systems with Applications*, 37(7), 5259-5264, (2010)
14. T. Hong, E. Kim, Segmenting customers in online stores based on factors that affect the customer's intention to purchase, *Expert Systems with Applications*, 39(2), 2127-2131, (2012)
15. J. T. Wei, M. C. Lee, H. K. Chen, H. H. Wu, Customer relationship management in the hairdressing industry: An application of data mining techniques *Expert Systems with Applications*, 40(18), 7513-7518, (2013)
16. H. K. Han, H. S. Kim, S. Y. Sohn, Sequential association rules for forecasting failure patterns of aircrafts in Korean airforce. *Expert systems with applications*, 36(2), 1129-1133, (2009)
17. J. Hipp, U. Güntzer, G. Nakhaeizadeh, Algorithms for association rule mining—a general survey and comparison, *ACM sigkdd explorations newsletter*, 2(1), 58-64, (2000)
18. Y. Li, P. Ning, X. S. Wang, S. Jajodia, Discovering calendar-based temporal association rules, *Data & Knowledge Engineering*, 44(2), 193-218, (2003)
19. S. O. Abdulsalam, K. S. Adewole, A. G. Akintola, M. A. Hambali, Data Mining in Market Basket Transaction: An Association Rule Mining Approach. *International Journal of Applied Information Systems (IJAIS)*, 7(10), 15-20, (2014)
20. A. Verma, S. D. Khan, J. Maiti, O. B. Krishna, Identifying patterns of safety related incidents in a

- steel plant using association rule mining of incident investigation reports, *Safety science*, 70, 89-98, (2014)
21. G. Czibula, Z. Marian, I. G. Czibula, Software defect prediction using relational association rule mining, *Information Sciences*, 264, 260-278, (2014)
  22. S. Kumar, D. Toshniwal, Analysing road accident data using association rule mining. In *IEEE International Conference on Computing, Communication and Security (ICCCS)*, 1-6, (2015)
  23. M. Hahsler, B. Grün, K. Hornik, Introduction to arules—mining association rules and frequent item sets. *SIGKDD Explor*, 2(4), (2007)
  24. A. El-Halees, Mining students data to analyze e-Learning behavior: A Case Study. (2009).
  25. Dannyphoto80, Illustration of the kids holding the tips of the wire for the hanging clothes on a white (royalty free), Online access (August 8th 2017), URL: <https://www.dreamstime.com/royalty-free-stock-photos-kids-holding-tips-wire-hanging-clothes-illustration-white-background-image32709718>, (2017)