Proceeding Book of ***The 6th ICMEM 2021****, 11-13 August 2021, Bandung, Indonesia*
ISBN: *978-623-92201-2-9*

Evaluation of Indonesia Kids Apparel Customers Behavior Using Segmentation and Association Rule Mining

Agus Winarta* and Nurdin Sobari

Faculty of Economics and Business, Universitas Indonesia, Indonesia

E-mail address: winarta.agus94@gmail.com

Abstract - **The apparel and fashion industry of local brands in Indonesia has been growing rapidly. The increase of e-commerce usage by the apparel and fashion industry has resulted in a stiff competition between local and global fashion brands. Therefore, a good and strategic marketing strategy is needed to maintain the industry growth and sustain the local industry. This research aims to build customer segmentation to help local fashion brands build marketing strategy in e-commerce. The study utilized Machine Learning using Association Rules Mining (ARM) and Consumer segmentation of RFM models. The ARM is one of the most popular techniques to learn a pattern or associations of attributes of customers. Consumer segmentation of RFM models was used to understand about consumer's behaviors. The data was collected from local fashion brands in e-commerce platforms. After data was collected, the data was preprocessed, and then analyzed using the RFM model. After the RFM model concluded, the model was used to associate consumers type, discount and delivery promotion by using ARM to understand the relations about the sales promotions and the consumers type. The segmentation is done by clustering with the k-means algorithm.**

Keywords - **Association Rules Mining, Consumer's Segmentation, RFM**

I. INTRODUCTION

The fashion industry is the biggest industry in the world, and has been playing a key role in development of the world's economy, which is estimated employs more than 300 million people worldwide. The number of SMEs that play part in the Fashion Industry is significant [1]. For the past years, the growth of Small Medium Enterprises (SMEs) has grown and become an important part of the economy of Europe. But with the increase of globalization, the competition in all sectors has increased, especially in the fashion industry [2].

Fashion industry in Indonesia is also considered as important factors for economic growth, there are more than 64 million SMEs in Indonesia, and contribute only 15.8% of national GDP [3]. The fashion industry in Indonesia is estimated to reach more than 250 billion IDR in 2025, an increase of more than 50% from 2019. The opportunity of this industry is also supported by growing income and wages which lead into increase in both supply and demand [1]

Covid-19 pandemic has affected every major industry, including the fashion industry. The pandemic has shifted the traditional retail into online transaction retail. Amid the pandemic, The kids category is one the fastest recovered category, maintaining higher sales more than adult fashion category [4]

There are many ways that can be done to increase the competitiveness of a company to improve profitability, one of the most popular ways is the Customer segmentation. Customer segmentation can be identified by various attributes, such as locations, ages, genders, etc [5]. One of the used technique in segmenting the market is RFM. RFM stands for recency, frequency and monetary. RFM segmentation is one of the attributes of behavioral segmentation, using past transaction of consumers to identify their purchase behavior and response to marketing activities [6], [7]

Since the 21st century, digital marketing and data sciences have evolved and developed in use and profitability [8]. This development is driven by innovations of technology, growth in knowledge, also the change of customer's need and expectation. One of the biggest changes is the emergence of big data and methods to analyze them, such as machine learning [9]. The combination of big data and machine learning algorithms can provide chances to predict and potentially explain consumer's behavior [10]. One of the most popular Machine learning technique to understand and discovers patterns and relations among various attributes [11] By understanding both the segmentation and their patterns in transactions, company can build up effective marketing activites. Several studies has been conducted to improve the customer segmentation method using various way, Guney, et al uses ARM and RFM to understand the pattern of customer profile and the genre of the video on demand (VOD) they find out that using both ARM and RFM establish more detailed customer segmentation [12]

Customer Segmentation can help businesses to customize marketing plans, identify trends, plan product development, advertising campaigns, and deliver relevant products [5]. This research is conducted to identify SMEs customer segmentation in Indonesia, by analyzing consumers behavior using RFM and then further analyzed using Association Rule Mining to understand the patterns of the consumers.

II. METHODOLOGY

This study proposes a combined approach using data mining for clustering and association rule mining to study the consumers behavior of children's apparel in Indonesia. Figure 1 shows the proposed methodology, consisting of four main steps, first, the preprocessing and

ISBN*: 978-623-92201-2-9*

data cleaning, the segmentation clustering using k-means, association rule mining using WEKA, and then customers profiling by their consumption pattern. Waikato Environment for Knowledge Analysis (WEKA) is java based software that can implement many state of the art machine learning and data mining algorithms [13]

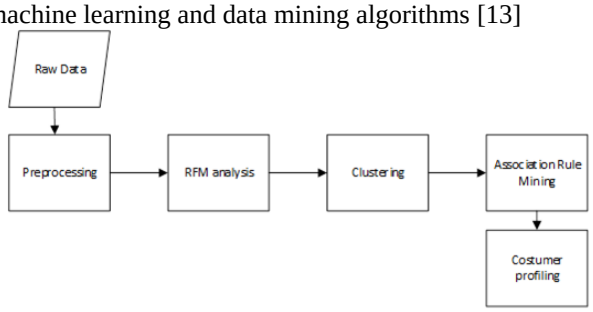

Fig. 1

The raw data was collected from local Indonesia brand "XYZ " from an ecommerce platform in Indonesia. Transactions data was collected from august 2018 into February 2021 in order to do RFM analysis. The transaction data has more than 20 thousand rows. The important information that was extracted was the invoice numbers, transaction date, customers id, province, SKUs, total discount, total shipment fee paid by customers, total expected shipment fee, and total discount.

The data was preprocessed, and cleaned using Microsoft excel, then standardized to be able to be analyzed using r and WEKA. Each of the customers was given their RFM scores respectively based on their purchase history. After the RFM scores conclude, the clustering using k-means cluster is conducted. "Elbow Method" was used to find the most effective number of clusters. Each of the clusters was then labelled as "Loyals", "Most Spender", "At risk", and "Potentials" respectively according to their RFM scores.

The customers were then labeled based on their segment and combined with their purchase history. Next, the data processed is then analyzed using Association rule mining to get the consumption pattern of each consumer segment. The association rule mining uses the apriori algorithm. The customers profile was then built based on the customers segmentation and their consumption pattern.

III. RESULTS

A. Preprocessing Data

The transaction data was collected from local indonesia clothing brand. The data extracted contains 48.262 rows. After the data was extracted, the data is then cleaned up and invalid values were removed. This operation was performed using microsoft excel. The transactions which has more than one items is shown in multiple duplicate rows is also removed. This was done to give accurate data of past purchase history of customers. In the end, total of 20.614 rows was left after the data was cleaned up. After that the RFM analysis was produced for each of customers. Table 1 presents the descriptive analysis of RFM scores.

TABLE 1
RFM SCORES

	Maximum	Minimum	Average
Recency	573	0	171.5
Frequency	16	1	1.26
Monetary	3483372	0	195237

B. Customers Segmentation Results

After the data was cleaned up, the data was standardized and scaled to do cluster analysis. The RFM segmentation analysis was conducted using k-means clustering. The ideal number of clusters was obtained using the "Elbow Method". The results for the optimal number of clusters can be interpreted from figure 2. In this method, the curve declines significantly until k=4, and further increase of clusters didn't show significant change in number of squares. Thus, the optimal number of clusters is 4.

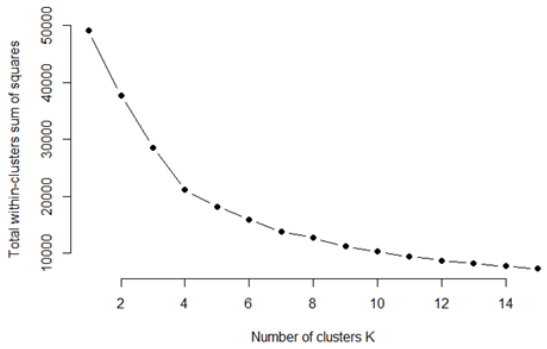

Fig. 2

To explain the customers segmentation, customer profile are created based on the RFM analysis. The label is based on the RFM scores such as most spender, loyals, promising, and at risk. Table 4 represents the customers profile, average RFM scores, and size of each group.

Referring to table 2, we can see that the customers' profile of kids apparel in Indonesia, Customers that are labelled as Promising is those who recently made a purchase, this clusters have the lowest recency, and average monetary. At risk is customers who haven't make a purchase in a long time, and can be considered as lost. Next, customers who are labelled as loyals are those who have make purchases multiple times, and have higher monetary value. Last, customers who is in Most spender category are those who have spent the most money. The customers who are profitable are the one who has high RFM scores, customers who is in loyals and most spenders clusters can be categorized as profitable [14]

ISBN: *978-623-92201-2-9*

TABLE II
RFM SCORES

Label	Avg Recency	Avg Frequency	Avg Monetary	Size
Promising	103	1.14	159769	49%
At Risk	277	1.10	154468	36%
Loyals	143	3.92	215744	5%
Most Spender	142	1.14	510091	10%

C. Association Rule Mining based on Customers Profiling

In the transaction dataset, the discount and shipping promotions is extracted, each of the customers id in the dataset was categorized based on their RFM label. After that, the data was then analyzed using WEKA to get the rules. Apriori rules was set to find the relationship between each clusters and what promotion they use.For each group, there are rules generated according to the minimum confidence level, then the relations between the variables are analyzed. Table 3 shows the association rules generated.

From the rules generated, there are several relationships that can be concluded. The most spender customers, they like to make purchases when they are given discounts. The customers who is at risk, or lost and promising, is not interested in discount given, and more interested on the sales promotions given in shipping fees. Free shipping is one of the most efficient ways to attract customers, but this strategy is not sustainable [15]. From this perspective, we need to understand that to get customers retention, more attention is needed more than promotions.

IV. DISCUSSION

From the RFM analysis, the customers segmentation can be concluded into four segments. Each of the segments represents unique traits of RFM scores and can be explained as in in table 4. This customers segmentation can help company to understand their consumers behavior and focus on the profitable consumers.

By Combining the customers segmentation and the pattern found with ARM, the company can choose and focus on building marketing campaign that suit the right consumers with the right promotions. Customers who is in promising segment like to make purchase using both discount and shipping promotions, while those in at risk segment only use the shipping fees promotions, for the loyals and most spender segment, there is no pattern found, indicating that there are other factors that affect the consumers behavior in this segment. The customers who is in promising is align with another research conduct by wang and bae, 2020, they found out that free shipping is one of the best marketing strategy to attain customers [15]

TABLE 3
RULES FOUND BY ARM

Antecedent	Consequent	Confidence	Lift
RFM PROFILE=at_risk 6475	DISCOUNT=NO 4255	0.66	1.21
RFM PROFILE=Promising DISCOUNT=YES 4632	SHIPPING FEES=YES 2915	0.63	1.1
RFM PROFILE=Promising 9167	SHIPPING FEES=YES 5570	0.61	1.06
RFM PROFILE=at_risk 6475	SHIPPING FEES=YES 3649	0.56	0.99
Provinsi=JAWA BARAT 6930	DISCOUNT=NO 3889	0.56	0.99
Provinsi=JAWA BARAT 6930	SHIPPING FEES=YES 3678	0.55	1.02
Provinsi=DKI JAKARTA 4672	SHIPPING FEES=NO 2564	0.55	1.28
Provinsi=JAWA BARAT SHIPPING FEES=YES 3678	DISCOUNT=NO 2091	0.55	1.02
Provinsi=DKI JAKARTA 4672	DISCOUNT=NO 2524	0.54	0.99
DISCOUNT=YES SHIPPING FEES=YES 5486	RFM PROFILE=Promising	0.53	1.19

TABLE 4
CUSTOMERS SEGMENTATION

Segment	Explanation
Promising	New customers acquired by company
At Risk	Old customers who is failed to convert into profitable segment
Loyals	Profitable customers who frequently make a purchase, very important to company
Most Spender	Profitable customers who make purchase with the most monetary value

This can conclude that the customers who purchases Kids apparel in Indonesia has four segments, with each segments has their own characteristics. Interestingly, customers who is in West Java and Jakarta differ in the preferred promotions. Customers in Jakarta are not interested in discounts while Customers in West Java are not interested in discounts. The shipping fee for Jakarta and West Java is not that different, only Rp 1000 to 2000. This can be explained by customers prefer the values and experience more than paying shipping fees, those who has bad experience using cheap shipping will prefer to pay

more for better services [15]. Customers in Jakarta prefer to make purchases even without shipping promotions, this shows that beside promotions, customers experiences and values need to be maximized to retain more customers. From the table, we can see that the at risk and promising segment is dominating most of the consumers. This majority is more interesting to use sales promotions such as discount or free shipping, align with the past research conducted that the sales promotions such as discount and free shipping would attract new customers, but in order to keep customer retention, superior customer experience is needed [15], [16]

It is important for company to make proper marketing plans based on each of their customers segment, so the plan will be suitable for their customers' targets. While the free shipping can be used to attract new customers, it is not suitable for customers retention. This study shows that customers with higher RFM scores is more interested in using price promotions than shipping promotions. This also applies to customers in Jakarta, is more interested in promotions price more than in West Java. We do not find the loyal customers has specific pattern. Some study shows that The customers who had bad experiences in low shipping fees prefer to pay more in order to get better services. In order to retain more customers, company may need to pay extra attention by giving more values in services better than promotions.

V. CONCLUSION

In the past few years, ecommerce in Indonesia has grown significantly, while SMEs in Indonesia also face challenges from big brands. To accommodate the fierce competition in the market, companies need to understand their customers better to make efficient and effective marketing plans. The development of big data analysis in marketing has become more common to be used by companies to understand the customers behavior. This paper has shown a different segmentation by various customers based on their RFM clusters prefer different taste in transactions. We combine the customers clustering RFM analysis and Association Rule Mining (ARM) to get more insights on customers segmentation

This study has several limitations, first, the data is collected from one company, data collection from more company can draw better understanding on general consumers segmentation in indonesia. Second, the values or experience of customers is not calculated in this paper, which may indicate that there are other factors that might also be an important factors that influence the customers behaviors. The conclusions of this paper show that the segmentation provided by combined methods of big data analysis such as clustering and association rule mining can explain the customers behavior better and providing better insights for companies.

ACKNOWLEDGMENT

We thank the company that supported this research by providing the transactional data of their company so that this research can be conducted.

REFERENCES

[1] G. Gonda, E. Gorgenyi-Hegyes, R. J. Nathan, and M. Fekete-Farkas, “Competitive Factors of Fashion Retail Sector with Special Focus on SMEs,” *Economies*, vol. 8, no. 4, p. 95, Nov. 2020, doi: 10.3390/economies8040095.

[2] P. Gazzola, E. Pavione, R. Pezzetti, and D. Grechi, “Trends in the Fashion Industry. The Perception of Sustainability and Circular Economy: A Gender/Generation Quantitative Approach,” *Sustainability*, vol. 12, no. 7, p. 2809, Apr. 2020, doi: 10.3390/su12072809.

[3] H. Hernita, B. Surya, I. Perwira, H. Abubakar, and M. Idris, “Economic Business Sustainability and Strengthening Human Resource Capacity Based on Increasing the Productivity of Small and Medium Enterprises (SMEs) in Makassar City, Indonesia,” *Sustainability*, vol. 13, no. 6, p. 3177, Mar. 2021, doi: 10.3390/su13063177.

[4] Euromonitor International, “Childrenswear in Indonesia,” p. 7, 2021.

[5] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, “RFM ranking – An effective approach to customer segmentation,” *J. King Saud Univ. - Comput. Inf. Sci.*, p. S1319157818304178, Sep. 2018, doi: 10.1016/j.jksuci.2018.09.004.

[6] E. Turban, J. Whiteside, D. King, and J. Outland, “Introduction to Electronic Commerce and Social Commerce | Efraim Turban | Springer,” 2017. https://www.springer.com/gp/book/9783319500904 (accessed Apr. 19, 2021).

[7] F. Huseynov and S. Ö. Yıldırım, “Behavioural segmentation analysis of online consumer audience in Turkey by using real e-commerce transaction data,” p. 17.

[8] M. T. P. M. B. Tiago and J. M. C. Veríssimo, “Digital marketing and social media: Why bother?,” *Bus. Horiz.*, vol. 57, no. 6, pp. 703–708, Nov. 2014, doi: 10.1016/j.bushor.2014.07.002.

[9] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, “Internet of things: Vision, applications and research challenges,” *Ad Hoc Netw.*, vol. 10, no. 7, pp. 1497–1516, Sep. 2012, doi: 10.1016/j.adhoc.2012.02.016.

[10] J. F. Hair and M. Sarstedt, “Data, measurement, and causal inferences in machine learning: opportunities and challenges for marketing,” *J. Mark. Theory Pract.*, vol. 29, no. 1, pp. 65–77, Jan. 2021, doi: 10.1080/10696679.2020.1860683.

[11] M. Abdel-Basset, M. Mohamed, F. Smarandache, and V. Chang, “Neutrosophic Association Rule Mining Algorithm for Big Data Analysis,” *Symmetry*, vol. 10, no. 4, p. 106, Apr. 2018, doi: 10.3390/sym10040106.

[12] S. Guney, S. Peker, and C. Turhan, “A Combined Approach for Customer Profiling in Video on Demand Services Using Clustering and Association Rule Mining,” *IEEE Access*, vol. 8, pp. 84326–84335, 2020, doi: 10.1109/ACCESS.2020.2992064.

[13] I. H. Witten, E. Frank, L. Trigg, M. Hall, G. Holmes, and S. J. Cunningham, “Weka: Practical machine learning tools and techniques with Java implementations,” 1999.

[14] R. A. Daoud, A. Amine, B. Bouikhalene, and R. Lbibb, “Combining RFM model and clustering techniques for customer value analysis of a company selling online,” in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, Nov. 2015, pp. 1–6. doi: 10.1109/AICCSA.2015.7507238.

[15] L. Wang and S. Bae, “How to avoid the free shipping pitfall? Changing consumer attitudes from the perspective of information interaction,” *Electron. Commer. Res. Appl.*, vol. 42, p. 100996, Jul. 2020, doi: 10.1016/j.elerap.2020.100996.

***ISBN**: 978-623-92201-2-9*

[16] R. R. Doern and C. F. Fey, "E-commerce developments and strategies for value creation: The case of Russia," *J. World Bus.*, vol. 41, no. 4, pp. 315–327, Dec. 2006, doi: 10.1016/j.jwb.2006.08.001.