



# CONSUMER SEGMENTATION KNOWLEDGE ANALYSIS USING DATA SCIENCE APPROACH FOR MARKETING STRATEGY RECOMMENDATIONS

Gigih Forda Nama, Rika Okta Nabella, Puput Budi Wintoro, Yessi Mulyani

Department of Informatics, University of Lampung, Indonesia.

Corresponding email: [gigih@eng.unila.ac.id](mailto:gigih@eng.unila.ac.id)

**Abstract.** Companies that provide services to users should start applying and developing data science methodology for product and service development. This research aims to conduct a data science approach using clustering analysis to find out the consumer segmentation characteristics. Python programming language and RapidMiner were primary tools for analyse those patterns. The K-Means algorithm, Google Data Studio, and RFM (Recency, Frequency, Monetary) model are used for gathering the knowledge of data transaction in the company. Cross Industry Standard Process for Data Mining (CRISP-DM) are used for advanced analysis techniques applied to discovering previously unknown patterns. The dataset collection derived from the company's database. The result shown that 3 clusters had 0.197 scores using the Davies Bouldin Index (DBI) validity test. There is 1090 member of low priority clusters, 73 medium priority clusters, and 969 high priority clusters. We provide several marketing strategy recommendations for each cluster at this company.

3479

## Keywords:

Clustering, K-Means, Model RFM, CRISP-DM, Davies Bouldin Index, Google Data Studio, RapidMiner, data science

## 1. INTRODUCTION

XYZ company is one of the companies engaged in the education sector by selling online course services. The target market is for students, college students, and employees who want to improve skills and expertise integrated with work and business.

Thousands of transaction history data have not been used optimally to find out their patterns. So, it is necessary to manage consumer data with appropriate data science techniques because, so far, the company's marketing activity process is still applying the conventional treatment to their consumers.

Utilization of data science methodology in segmenting consumers based on the Recency, Frequency, and Monetary (RFM)

model, those are variables of last transaction in days, the number of transactions, and the total value of purchases that have been made within a specific time, so that the right clusterization with the K-Means method is to determine the characteristics of consumers in each resulting cluster.

In addition to consumer transaction data that will be analysed to determine data knowledge, other data that support research such as gender and age data, course rating data, affiliate marketing income data, website access trend data, and potential consumer data will also be analysed for data knowledge so that it will be used as a marketing strategy recommendation that the company can do to increase revenue.



## 2. LITERATURE REVIEW

### 2.1. Consumer Segmentation

Segmentation is the process of dividing a heterogeneous market into homogeneous groups. The grouping is based on the demographic, geographic, psychographic and behavioural groups [1].

To reach consumers more effectively, segmentation can produce groups with several criteria. This is done with the strategic objectives of each consumer segment to be more structured and specific. Characteristics that can be used as a reference include [2]:

1. Geographic segmentation will classify consumers into market groups based on the regional scale of their geographical location, such as region, climate, city, or village.
2. Demographic segmentation will classify consumers into market groups based on demographic variables, such as age, family size, family life cycle, gender, occupation, income, religion, social class, and race.
3. Psychographic segmentation will classify consumers into market groups based on lifestyle and personality variables.
4. Behavioural segmentation will classify consumers into market groups based on consumer knowledge, attitudes, and responses given by consumers to a product.

### 2.2. Marketing Strategy

A marketing strategy is a plan that has been systematically and comprehensively prepared that will be used as a guide in carrying out marketing activities to achieve the company's ultimate goal of generating profits or increasing company revenues. Implementing a marketing strategy includes market segmentation, determining the target market, and positioning [3]. This means that there is a relationship that segmentation activities are a step in implementing a marketing strategy.

### 2.3. Data Mining

Data mining is extracting/excavating/finding hidden information from large databases and is helpful for interested parties. Data mining can help data holders analyses and find unexpected

relationships between data, which will help decision-making [4].

The data mining stages used are the Knowledge Discovery Databases or KDD process model, which has nine steps which can be seen in Figure 1 [5].

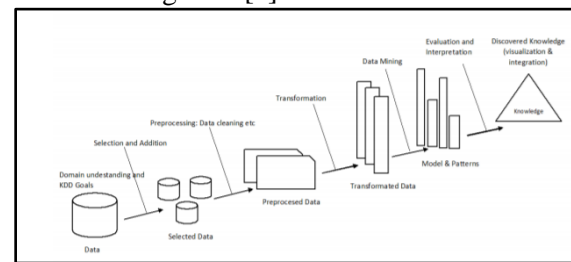


Figure 1 Data Mining Stages

Stages in the Knowledge Discovery Databases process:

1. Domain Understanding and KDD Goals.
2. Selection and Additions.
3. Data cleaning.
4. Data transformation.
5. Data mining.
6. Evaluation and Interpretation.
7. Discovered Knowledge.

### 2.4. Clustering

Clustering is a formal study in partitioning or grouping data that does not have a categorical label. Clustering is unsupervised learning, meaning that clustering does not have a stage for data training. This is different from the classification, which is supervised understanding. Clustering is usually used to group data with similarities to data objects and vice versa, minimizing similarities to other clusters [6].

### 2.5. K-Means

The K-Means algorithm is a non-hierarchical clustering method and can be used on data that does not have a label. K-Means tries to divide or partition data into clusters to form one or more clusters, with each having the same data characteristics. In contrast, clusters have different data characteristics. The primary purpose of clustering is to minimize the various features in a cluster and maximize the variation between clusters [7].

The following are the steps in grouping data using the K-Means algorithm [8]:

1. Determine the number of clusters
2. Initial initialization and cluster center are done randomly
3. Calculate the distance from each object to the center of the cluster (centroid)

based on the similarity or dissimilarity of the data using the Euclidean Distance method. This Euclidean Distance serves to get the value of the new cluster center (centroid) until the value of the cluster center (centroid) remains with the following formula:

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_j)^2}$$

Description :

$d_{(x,y)}$  = size dissimilarity

$x_i$  =  $(x_1, x_2, x_3, \dots, x_i)$  data variable

$y_j$  =  $(y_1, y_2, y_3, \dots, y_j)$ , variable at the center point

The closest distance from one data to one cluster will determine which data belongs to which cluster [9].

4. Calculate the new cluster center with new membership in each cluster by calculating the average object in the cluster.
5. Recalculate the distance of each object with the new cluster center, until the cluster does not change, then the clustering process is complete

## 2.6. Model RFM

The RFM model stands for Recency, Frequency, and Monetary; companies widely use this model in segmenting customers as one of the marketing efforts that can be done. The RFM model will use past consumer information data to evaluate and predict consumer behavior and values [10].

The RFM model involves three variables: Recency, Frequency, and Monetary, which measure consumer transaction history [11]. Those are :

- a. *Recency*  
*Recency* is the distance from when the consumer last made a transaction to the current time or how recently the consumer made a transaction.
- b. *Frequency*  
*Frequency* is the number of transactions made by consumers in a certain period or how often consumers make transactions.
- c. *Monetary*  
*Monetary* is the total value of transactions made by consumers.

## 2.7. Determination of The Number of Clusters

Determination of the number of clusters using the Davies Bouldin Index, this value will indicate the optimal number of clusters. The Davies Bouldin Index (DBI) was introduced by David L. Davies and Donald W. Boildin in 1979. DBI is a measure used to determine the best number of clusters after the clusterization process is complete [12]. If the DBI value obtained is getting smaller and closer to 0 and not harmful ( $\geq 0$ ), the better the cluster obtained from the K-Means Cluster used [13].

## 2.8. RapidMiner

RapidMiner is a tool created by Dr. Markus Hofmann of the Institute of Technology Blanchardstown and Raif Klinkenberg of rapid-i.com. RapidMiner is designed with a GUI (Graphical User Interface) display to make it easier for users to apply RapidMiner tools. In addition, to its open-source nature, the Java language is under the GNU Public License. It can run on any operating system, which is an advantage of the RapidMiner tools [14].

## 2.9. JupyterLab

Jupyterlab allows for documenting and reproducibility workflows in data science where the data is large and complex. So with jupyterlab, researchers can record the steps in the research process, including data sources, transformation, and finally, answering questions from management or clients [15].

## 2.10. Google Data Studio

Google Data Studio is a program used to visualize data so that it is easy to use and can present complex data attractively. Google Data Studio was launched for free in 2016. Available for free and for anyone with a Google account, Google Data Studio is also a potential solution in data visualization and is full-featured [16].

## 2.11. Google Sheet

Google sheet is a software equivalent to Microsoft Excel, developed in 2005 by the company 2 Web Technologies [17]. Google Sheets has a maximum limit of 256 columns and 20,000 rows per sheet. Meanwhile, the uploaded file converted to google Sheets format must not be larger than 20 MB [18].



## 2.12. CRISP-DM

Cross-Industry Standard Process for Data Mining or CRISP-DM is used in the industrial world to solve various business problems and relationships with data mining.

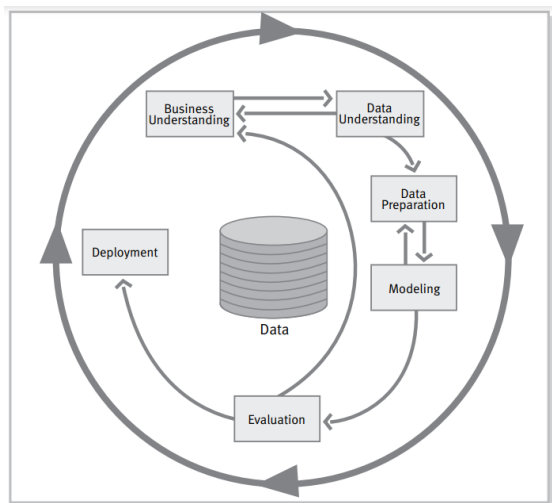


Figure 2 CRISP-DM Method Stages

There are six phases in the CRISP-DM method [19] :

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

## 2.13. Related works

Several studies related to this research, those are, research conducted by Wiratama Ahsani Taqwim, Nanang Yudi Setiawan, and Fitra A. Bachtiar entitled "Analysis of Customer Segmentation with RFM Model at PT. Arthamas Citra Mandiri Using the Fuzzy C-Means Clustering Method" in 2019 [20]. Further research conducted by Yulison Herry Chrisnanto and Ade Kaniangsih entitled "Clustering of Customer Equity Based on Recency Frequency Monetary (RFM) Using K-Means Clustering" in 2019 [21]. Further research conducted by Christina Deni Rumiarti and Indra Budi entitled "Customer Segmentation in Customer Relationship Management in Retail Companies: A Case Study of PT Gramedia Asri Media" in 2017 [22]. Some of reference on system information development using several open source application found on works [23][24][25].

## 3. RESEARCH METHODOLOGY

### 3.1. Preparation tools and materials

The tools and materials used in this research are computer equipment, JupyterLab software, RapidMiner, Google Sheet, and Google Data Studio. While the material used in this study is the company's consumer sales transaction data from January 1, 2021, to November 25, 2021, and other data will be analyzed as a form of marketing strategy recommendations.

### 3.2. Research Stages

The stages carried out in this research use the Cross-Industry Standard Process for Data Mining method, which consists transparently of the following steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment.

## 4. RESULT

### 4.1. Business Understanding

- a. Determine Business Objectives  
This phase aims to design a data mining model that can be used to cluster consumers based on the RFM model from the company's transaction data and other data knowledge owned by the company to produce a data mining model that can be a recommendation for marketing strategies to companies so that marketing becomes more efficient. Right on target.
- b. Assess Situation  
Marketing activities in the company are still not focused on priority consumers, and knowledge of user data in the company database is still unknown to the company.
- c. Determine Data Mining Goals  
The data mining model created aims to perform clusters in knowing the characteristics of consumer behavior in each cluster. In addition, by using data mining, it is hoped that data will provide knowledge of previously hidden data into valuable knowledge and help determine business decisions.
- d. Produce Project Plan  
Based on this business understanding, a project plan was obtained, namely conducting consumer segmentation analysis using the k-means clustering method based on the RFM model and other data knowledge as a marketing

3482



strategy recommendation for the company.

#### 4.2. Data Understanding

##### a. Collect Initial Data

The data used is transaction data from January 1, 2021, to November 25, 2021, and other data in the company database that will be analyzed as a form of marketing strategy recommendations.

##### b. Describe Data

Table 1 Database Table Description

Tabel	Keterangan
Users	User data information
Users_detail	Detailed information of users
Course	Detailed information on course data
Category	Detailed information regarding course category data
Payment	Detailed information regarding course payments
Payment_balance	Detailed information regarding sales commission and balance withdrawal
Rating	Detailed information regarding course ratings
Ci_sessions	Detailed information regarding website access times

#### 4.3. Data Preparation

Several stages need to be carried out in the data preparation stage, including selecting data, integrating data, cleaning data, and formatting data.

users_id	email	first_name	last_name	no_hp	recency	frequency	monetary	filter_no_hp
2					14	4	0 No	
3					93	2	0 No	
4					98	1	0 No	
5					319	1	0 No	
6					312	1	0 No	
7					304	2	0 No	
8					305	2	150000 Yes	
9					305	1	25000 Yes	
10					303	2	0 Yes	
11					303	1	0 No	
12					280	2	150000 Yes	
13					303	1	250000 Yes	
14					303	1	150000 Yes	
15					302	1	250000 Yes	
16					117	1	0 No	
17					302	1	200000 Yes	
18					302	1	25000 Yes	
19					302	1	200000 Yes	

Figure 3 Consumer Segmentation Data

Figure 3 results from consumer segmentation data carried out at the data preparation stage using two database tables: the users and. This payment data will be used to segment consumers to know which clusters are high priority, medium priority, and low priority.

users_id	email	first_name	last_name	no_hp	users	users_dateadded	status	transaction	filter_no_hp	duplicate
113					0	2021-01-03	0 Hanya Login	No	FALSE	
117					0	2021-01-11	0 Hanya Login	No	FALSE	
120					0	2021-01-12	0 Hanya Login	No	FALSE	
128					0	2021-01-13	0 Hanya Login	No	FALSE	
128					0	2021-01-14	0 Hanya Login	No	FALSE	
136					0	2021-01-19	0 Hanya Login	No	FALSE	
138					0	2021-01-22	0 Hanya Login	No	FALSE	
147					0	2021-01-27	0 Hanya Login	No	FALSE	
162					0	2021-01-28	0 Hanya Login	No	FALSE	
172					0	2021-02-02	0 Hanya Login	No	FALSE	
173					0	2021-02-03	0 Hanya Login	No	FALSE	
175					0	2021-02-07	0 Hanya Login	No	FALSE	
176					0	2021-02-08	0 Hanya Login	No	FALSE	
177					0	2021-02-08	0 Hanya Login	No	FALSE	
180					0	2021-02-13	0 Hanya Login	No	FALSE	
181					0	2021-02-13	0 Hanya Login	No	FALSE	
182					0	2021-02-13	0 Hanya Login	No	FALSE	
183					0	2021-02-14	0 Hanya Login	No	FALSE	

Figure 4 Potential Consumer Data

Figure 4 shows potential consumer data carried out at the data preparation stage using two database tables: the users and the payment. This data will be used to identify consumers who have the potential to buy the company's products in the future.

3483

users_id	email	first_name	last_name	no_hp	courses	category	payment_type	duplicate
1					CorelDraw Dasar Pemula Hingga Desain	Free	TRUE	
3					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
4					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
5					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
6					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
7					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
8					Kelas Trading Si Koko Saham Akuntansi & Ki	Free	TRUE	
9					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
10					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
11					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
12					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
13					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
14					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
15					Tutorial Web Project Manajemen Pemasaran	Free	TRUE	
16					Kelas Trading Si Koko Saham Akuntansi & Ki	Free	TRUE	
17					[PROMO!!!] Menulis Surat Lari Personal Deve bank transfer	Free	TRUE	
18					[PROMO!!!] Menulis Surat Lari Personal Deve bank transfer	Free	TRUE	
19					Kelas Trading Si Koko Saham Akuntansi & Ki	Free	TRUE	

Figure 5 Category, Course, and Payment Type Data

Figure 5 shows the results of category data, courses, and types of payments that have been made at the data preparation stage using four database tables, namely the users, payment, courses, and category tables. This data will determine the category, the most sold course, and the most used payment type.



	A	B	C	D	E
1	id	jns_kel	avg	usia	
2	349	Laki-laki	27	26 - 45	
3	350	Laki-laki	46	46 - 65	
4	351	Laki-laki	25	12 - 25	
5	352	Laki-laki	26	26 - 45	
6	353	Perempuan	28	26 - 45	
7	355	Laki-laki	26	26 - 45	
8	356	Perempuan	23	12 - 25	
9	357	Laki-laki	26	26 - 45	
10	358	Perempuan	20	12 - 25	
11	359	Perempuan	26	26 - 45	
12	360	Laki-laki	24	12 - 25	
13	361	Perempuan	32	26 - 45	
14	362	Perempuan	23	12 - 25	
15	366	Laki-laki	45	26 - 45	
16	367	Laki-laki	25	12 - 25	
17	368	Laki-laki	24	12 - 25	
18	369	Laki-laki	31	26 - 45	
19	370	Perempuan	29	26 - 45	

Figure 6 Age and Gender Distribution Data

Figure 6 shows age and gender distribution data carried out at the data preparation stage using two database tables: the users and users\_detail tables. This data will be used to determine most website users' a.

	A	B	C	D	E	F
1	id	email	first_name	last_name	income	
2	850				389902	
3	735				380238	
4	713				378196	
5	746				325317	
6	710				315461	
7	633				306653	
8	650				277810	
9	651				265794	
10	793				250846	
11	640				211999	
12	810				179332	
13	814				156500	
14	636				93500	
15	842				45083	
16	711				35000	
17	803				31500	
18	741				31000	
19	725				30137	

Figure 8 Top Data Affiliate Marketing Income

Figure 8 shows top affiliate marketing income data carried out at the data preparation stage using two database tables: users and payment\_balance.ge and gender range. This data will determine which users have the largest total commission based on the spread of course referral codes.

3484

	A	B	C	D	E
1	title	date_addded	rating	review	
2		2021-01-29	5		
3		2021-02-22	5		
4		2021-08-04	5		
5		2021-08-04	5	Materinya sangat mudah dipahami	
6		2021-08-07	5	Sangat membantu dan menjadikan ilmu baru bagi si	
7		2021-08-09	5		
8		2021-08-09	5		
9		2021-08-09	5		
10		2021-08-09	5	Penjelasan Baik	
11		2021-08-13	5	Lengkap dan informatif, penyampaiannya pun mena	
12		2021-08-14	5		
13		2021-08-16	5		
14		2021-08-16	5	Sangat membantu	
15		2021-08-19	5	thankyou vocasia sudah menyediakan platform bagi	
16		2021-08-20	5	naise	
17		2021-08-20	5	Penjelasannya singkat namun mudah di pahami. bel	
18				gege	

Figure 7 Course Top Rating Data

Figure 7 shows the results of the top rating course data carried out at the data preparation stage using four database tables, namely users, payment, course, and rating. This data will be used to determine which courses have the top rating.

	A	B	C	D
1	jam	menit	detik	
2	0	4	9	
3	0	9	12	
4	0	5	0	
5	0	9	15	
6	0	17	12	
7	0	34	56	
8	0	19	11	
9	0	2	43	
10	0	2	45	
11	0	5	35	
12	0	44	7	
13	0	50	56	
14	0	17	9	
15	1	1	25	
16	1	7	56	
17	1	24	31	
18	1	52	37	
19	1	56	41	

Figure 9 Data Trend Akses Website

Figure 9 is the result of trending website access data that has been carried out at the data preparation stage using one database table, namely ci\_sessions. This data will determine the trend of accessing most websites every hour.



#### 4.4. Modelling

The clustering modeling in this study uses the K-Means algorithm with the tools contained in the RapidMiner 9.9 software.

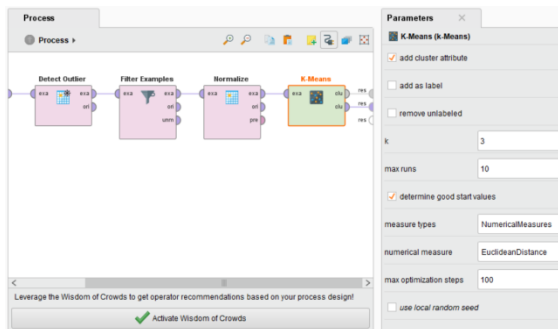


Figure 10 K-Means Algorithm Modeling Structure

Thus, the resulting cluster model is as follows:

Cluster	Konsumen	Centroid		
		R	F	M
1	1090	0.272	0.025	0.000
2	73	0.846	0.036	0.039
3	969	0.078	0.055	0.259
<b>Total</b>	<b>2132</b>			

#### 4.5. Evaluation

At this stage, the last fo to determine the number of clusters generated optimally using the Davies Bouldin Index (DBI).

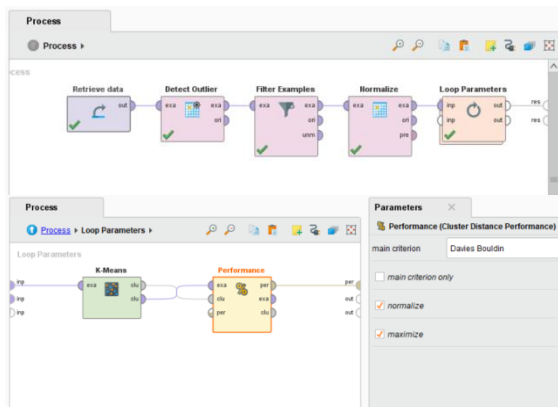


Figure 11 Evaluation Structure of the Davies Bouldin Index

This results in the Davies Bouldin Index value in each cluster as follows:

Table 2 Davies Bouldin Index Score Results

Jumlah Cluster	Davies Bouldin Index (DBI)
2	0.240
<b>3</b>	<b>0.197</b>

4	0.262
5	0.203
6	0.277
7	0.201
8	0.222
9	0.223
10	0.207

Based on table 2, it is known that 3 clusters are the optimal number of clusters because the Davies Bouldin Index value is close to 0.

#### 4.6. Deployment

The modeling results are made in data visualization at the last stage in the CRISP-DM method. In performing visualizations to make it faster, more accessible for readers to understand, and provide convenience in sharing visualization reports, we use the implementation of Google Data Studio.

##### a. Consumer Segmentation Data Analysis

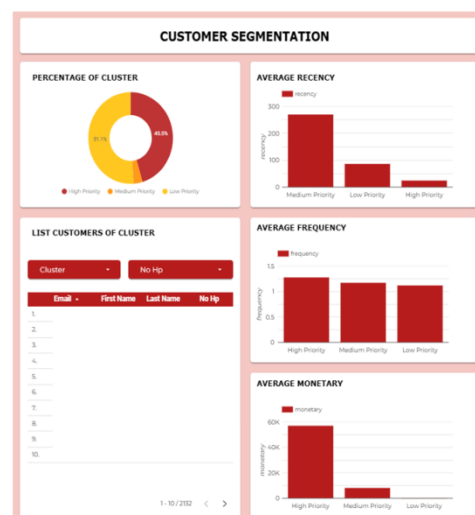


Figure 12 Consumer Segmentation Data Visualization

In the high priority cluster, the strategic recommendation in this cluster is to optimize the category of courses and similar courses that are sold the most in the high priority cluster by the product innovation division because consumers in this cluster are consumers who make paid transactions, so it is hoped that users in this cluster will be interested in buying and selling. Return to the transaction. In addition, the provision of discounted course price voucher codes is a form of company appreciation to loyal consumers.

Furthermore, for the medium priority cluster, the strategy recommendation in this



cluster is to provide free promo courses to increase awareness because the average trend for the last time to make a transaction has been going on for a long time, where the average recency was 270 days ago.

In the low priority cluster, the strategy recommendation in this cluster is to offer the most sold courses to increase transactions in this cluster because the Monetary value of this cluster is IDR 0, meaning this cluster is a consumer who only makes free courses transactions.

Meanwhile, recommendations for pricing new products and product prices when discounting will be made can be in IDR 56,969 or below that price. This is because the average ability of the Monetary value in the high priority consumer cluster is IDR 56,969.

b. Potential Consumer Data Analysis

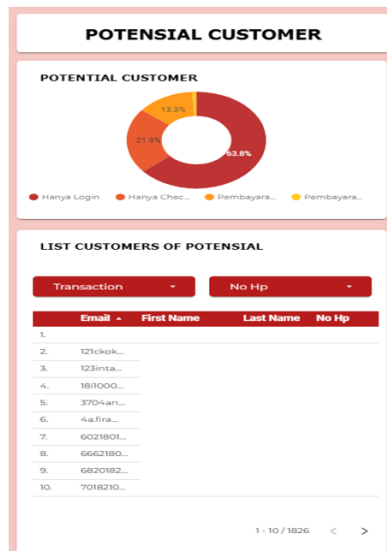


Figure 13 Data Visualization of Potential Consumers

The recommended strategy that can be done for potential consumers is for checkout-only transactions. Email Direct Marketing and Whatsapp messages need med model to be evaluated to remind users to continue the checkout process based on the selected course or the course added to the cart. For users who only log in to the website, it is necessary to offer free courses so that they are expected to be interested in re-opening the website and even making paid transactions on other types of courses.

Meanwhile, for users with expired payment transactions and canceled payments, re-offering can be made to users with courses whose transactions have expired or have

exceeded the payment time and are balanced. So, users are expected to be interested again in making payment transactions.

c. Data Analysis Category, Course, and Payment Type

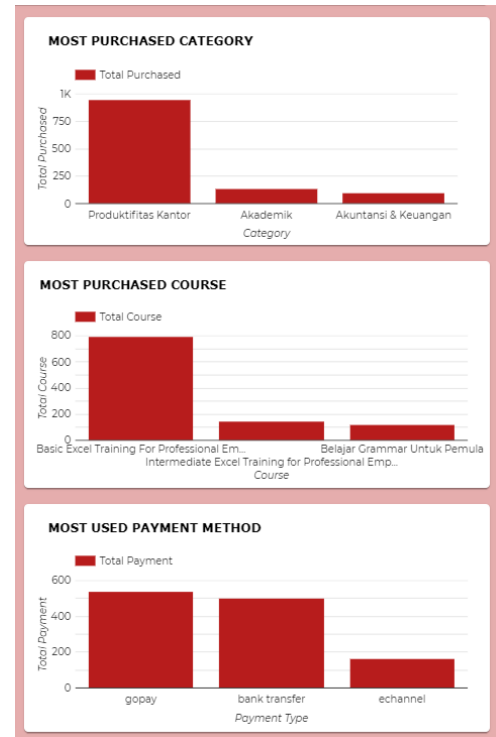


Figure 14 Category, Course, and Payment Type Data Visualization

Strategy recommendations based on knowledge of course categories and widely sold courses are similar to the Basic Excel Training for Professional Employee course and completing other studies in the Office Productivity category. At the same time, the recommendation for payment types is to increase the variety of payment E-Wallet or digital wallets such as ShopeePay, Ovo, and Dana.

d. Data Analysis of Age and Gender Distribution





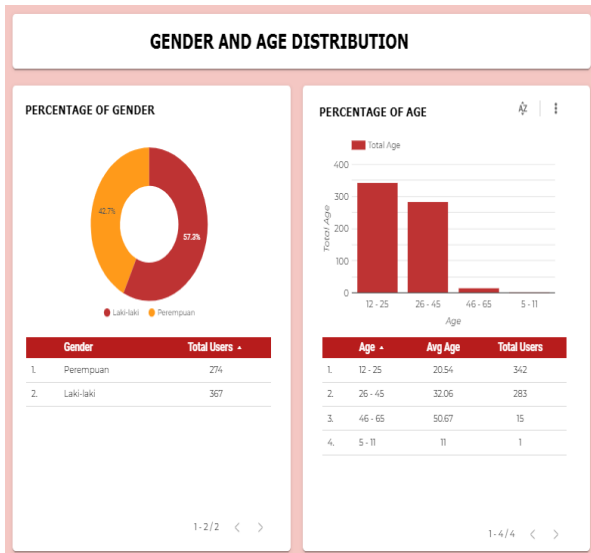


Figure 15 Visualization of Age and Gender Distribution Data

The recommended strategy is based on age knowledge that most users are teenagers with an average age of 21 years and adults with an average of 32 years to maximize the creation of courses to improve students' and employees' skills. In addition, offering classes to students and employees is one of the proper steps to increase company revenue.

Meanwhile, based on gender, there is no significant difference between the number of female and male gender. Besides, the number of users who input information about age and gender is less than the total number of users, making gender and age analysis not optimal.

e. Course Top Rating Data Analysis



Figure 16 Top rated Course Data Visualization

The recommended strategy based on the knowledge of the top 3 rating courses is to establish good cooperation with the three partners or instructors and create new courses with these partners or instructors. Because the three partners and instructors received good reviews from users who had purchased the course, it can be said that some users quickly understood the explanation given. The three partners are Hundred Sciences, Andrizar, and Andy Sanjaya.

f. Top Data Analysis of Affiliate Marketing Income



Figure 17 Affiliate Marketing Revenue Top Data Visualization

Strategy recommendations based on the knowledge of the top 3 affiliate marketing incomes, namely giving additional bonuses to the three users as a form of appreciation for



being an affiliate marketing that has the most significant income among other affiliate marketing, which means providing the most crucial income to the company. In addition, a form of appreciation can be given in the form of an E-Certificate that can support a career path in the marketing field.

#### g. Website Access Trend Data Analysis

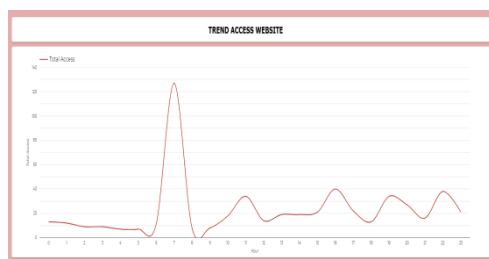


Figure 18 Website Access Trend Data Visualization

Strategy recommendations based on knowledge of website access trend data, namely recommendations for uploading the latest course content at 07.00 am. In addition, it is recommended that the website infrastructure monitoring process be more intensive during busy times and ensure no disturbances occur while users access the website.

However, this recommendation needs to be reviewed further on how the request relates to external data such as increasing internet access trends at certain times because the data used in the analysis of website access trends is incomplete, only certain users have records to display their access on the website, so the results obtained not optimal.

## 5. CONCLUSION

Based on research conducted using RapidMiner software with the K-Means method and the RFM model, it can help determine the segmentation of consumers owned and resulted that 3 clusters have a value of 0.197 using the Davies Bouldin Index validity test. Where the low priority cluster is 1090 consumers, the medium priority cluster is 73 consumers, and the high priority cluster is 969 consumers.

Utilization of other data knowledge, namely data on potential consumers, gender and age, best course ratings, most significant affiliate marketing income, and website access

trends, can also be used in determining company marketing strategies and business decisions, including making the category of the most sold courses, giving additional bonuses and E-Certificate for affiliate marketing that has the most significant income.

The analysis results are visualized with column charts, donut charts, time series charts, tables, and control drop-down lists using the application of Google Data Studio to make it easier for users to read charts, export data, and process data dissemination.

## REFERENCE

- [1] B. R. T. Putri, *Manajemen Pemasaran*, Denpasar: Universitas Udayana, 2017.
- [2] H. M. Ritonga, M. E. Fikri, N. Siregar, R. R. Agustin and R. Hidayat, *Manajemen Pemasaran Konsep dan Strategi*, Medan: Universitas Pembangunan Panca Budi, 2018.
- [3] H. Wijoyo, A. Ariyanto, F. Wongso, I. Indrawan, M. M. F. Akbar, N. Anggraini, S. S. and W. S. , *Strategi Pemasaran UMKM di Masa Pandemi*, Solok: InsanCendekiaMandiri, 2021.
- [4] H. Kurniawan, S. Defit and S. , "Data Mining Menggunakan Metode K-Means Clustering untuk Menentukan Besaran Uang Kuliah Tunggal," *Journal of Applied Computer Science and Technology*, vol. 1, no. 2, pp. 80 - 89, 2020.
- [5] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Simetris*, vol. 10, no. 2, pp. 421 - 432, 2019.
- [6] B.E. Adiana, I. Soesanti and A. E. Permasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *JUTEI*, vol. 2, no. 1, pp. 23 - 32, 2018.
- [7] Y. Agusta, "K-Means - Penerapan, Permasalahan dan Metode Terkait," *Jurnal Sistem dan Informatika*, vol. 3, pp. 47-60, 2007.
- [8] M. Wahyudi, M. R. Saragih and S. , *Data Mining : Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*, Yayasan Kita Menulils, 2020.
- [9] A. Bastian, H. Sujadi and G. Febrianto, "Penerapan Algoritma K-Means Clustering Analisis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 26-32, 2018.
- [10] V. Kumar, *Managing Customers for Profit: Strategies To Increase Profits And Build Loyalty*, 2008: Pearson Education, India.

- [11] R. C. Blattberg, B.-D. Kim and S. A. Neslin, *Database Marketing: Analyzing and Managing Customers*, Netherlands: Springer New York, 2008.
- [12] A. A. Az-zahra, A. F. Marsaoly, I. P. Lestyani, R. Salsabila and W. O. Z. Madjida, "Penerapan Algoritma K-Modes Clustering dengan Validasi Davies Bouldin Index pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta," *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 9, no. 1, pp. 24 -36, 2021.
- [13] I. D. Id, T. R. Angraini and Sukamto, "Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means," *JUITA*, vol. 7, no. 2, pp. 137 - 148, 2018.
- [14] S. Hendrian, "Algoritma Klasifikasi Data Mining untuk Memprediksi Siswa dalam Memperoleh Bantuan Dana Pendidikan," *Faktor Exacta*, vol. 11, no. 3, pp. 266 - 274, 2018.
- [15] W. R. Paczkowski, *Deep Data Analytics for New Product Development*, United Kingdom: Taylor & Francis, 2020.
- [16] G. Snipes, "Google Data Studio," *Journal of Librarianship and Scholarly Communication*, vol. 6, no. 1, 2018.
- [17] Y. Kurniawan, *Cara Mudah Menguasai Google Docs*, Jakarta: Elex Media Komputindo, 2010.
- [18] C. Rafter, *Guide To Cleaning Data with Excel and Google Sheets*, DSM Media, 2019.
- [19] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, *CRISP-DM 1.0, The CRISP-DM consortium*, 2020.
- [20] W. A. Taqwim, N. Y. Setiawan and F. A. Bachtiar, "Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt. Arthamas," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 2, pp. 1986-1993, 2019.
- [21] Y. H. Chrisnanto and A. Kaniaingsih, "Pengelompokan Ekuitas Pelanggan Berbasis Recency Frequency Monetary (RFM) Menggunakan K-Means Clustering," *Seminar Nasional Teknologi Informasi dan Komunikasi*, pp. 45-53, 2019.
- [22] C. D. Rumiarti and I. Budi, "Segmentasi Pelanggan pada Customer Relationship Management di Perusahaan Ritel : Studi Kasus PT Gamedia Asri Media," *Jurnal Sistem Informasi*, vol. 13, no. 1, pp. 1-10, 2017.
- [23] Merdeka, N., & Nama, G. F. (2021). Assessment of Information Technology Governance Implementation Based on COBIT Framework 5 Focus on DSS 02 Subdomain (Deliver, Service, And Support) (Case Study of Bank XXX Branch X). *Journal of Engineering and Scientific Research*, 3(2). <https://doi.org/10.23960/jesr.v3i2.91>.
- [24] Nama, G.F., Muludi, K. (2018). Implementation of two-factor authentication (2FA) to enhance the security of academic information system. *Journal of Engineering and Applied Sciences*, 13(8), pp. 2209–2220.
- [25] Indra, M.R., Purwasih, N., Batubara, M.A.M., Nama, G.F., Kasai, N. Design and Implementation of Real-Time Reading System of Atmospheric Corrosion Sensor on Metal Material Using Internet of Things (IoT) Technology. *Proceedings - ICCTEIE 2021: 2021 International Conference on Converging Technology in Electrical and Information Engineering: Converging Technology for Sustainable Society*, 2021, pp. 25–30.

