JURNAL CAHAYA MANDALIKA

P-ISSN: 2828-495X
E-ISSN: 2721-4796

APPLICATION OF DATA MINING FOR EVALUATION OF BEST SELLING INTERNET PACKAGE USING K-MEANS CLUSTERING

Devo Fauzan Rahman[1], Viany Utami Tjhin[2]
[1,2]Information System Management Department,
BINUS Graduate Program-Master of Information System Management
1devo.rahman@binus.ac.id, 2vtjhin@binus.edu

Kata Kunci :
Validasi Klaster, Penambangan Data, Indeks Boulder Davies, Pengelompokan K-Means, Penjualan

Data Mining adalah proses untuk mendapatkan informasi yang berguna dari gudang basis data dalam bentuk pengetahuan. Data riwayat transaksi penjualan dapat menjadi informasi untuk pengambilan keputusan bisnis. Salah satu perusahaan telekomunikasi terbesar di Indonesia mengalami kendala penjualan produknya di kota Jambi. Tren penjualan produk tersebut mengalami penurunan tahun lalu. Hal ini disebabkan penetrasi kompetitor yang semakin mengancam eksistensi produk tersebut sebagai produk terlaris perseroan. Beberapa penelitian sebelumnya menggunakan analisis berdasarkan pengelompokan data untuk membentuk informasi. Oleh karena itu, penelitian ini bertujuan untuk mengolah data penjualan menjadi beberapa cluster dengan menggunakan algoritma K-means clustering. Hasil clustering berupa 5 kategori paket yaitu sangat laris (1 data), laris (1 data), laris sedang (1 data), kurang laris (1 data) dan tidak laris (74 data). . Pengelompokan 5 klaster ini didasarkan pada evaluasi Davies Bouldin Index yang menghasilkan nilai 0,035. Hasil clustering ini dapat berguna bagi manajemen perusahaan untuk menyusun strategi pemasaran di masa mendatang.

Keywords :
Cluster Validation, Data Mining, Davies Bouldin Index,K-Means Clustering, Sales

ABSTRACT

Data Mining is a process to obtain useful information from a database warehouse in the form of knowledge. Data transaction history of sales can be informative for a business decision. One of the largest telecommunications companies in Indonesia is experiencing problems selling its products in Jambi city. The sales trend for the products has decreased last year. This is due to competitor penetration which increasingly threatens the product's existence as the company's best-selling product. Several previous studies used analysis based on clustering data to form information. Therefore, this study aims to process sales data into several clusters using the K-means clustering algorithm. The results of the clustering form 5 package categories, namely very in demand (1 data), in demand (1 data), moderate in demand (1 data), less in demand (1 data) and not-in demand (74 data). The grouping of these 5 clusters is based on evaluation by Davies Bouldin Index which yields a value of 0.035. The results of this clustering can be useful for company management to develop marketing strategies in the future.

INTRODUCTION

For decades, this company has led the internet package market in Indonesia. The company's customer penetration has reached 87% as of July 2021, where this figure far exceeds its competitors who market the same product (Frederick & Bhat, 2021).

With the number of competitors slowly increasing, this company must maintain the existence of its product as the preferred broadband internet package for the Indonesian people (Hutajulu et al., 2020). Currently, sales of this company's products are starting to be affected by the onslaught of competitors in various parts of Indonesia, especially in the city of Jambi. It can be seen from the sales trend in Jambi City which tends to continue to decline somewhat during 2021, where previously from January around 1882 packages were sold, while in December only 1037 packages were recorded.

Even though the packages offered by this company are very diverse, starting from those based on the number of services and the amount of bandwidth. For the number of services, there are 2 options, namely Dual Play and Triple Play (Pereira et al., 2013). In the dual-play package itself there are 2 sub-service options, the first is Dual-Play with Telephone and Internet services, then the second is Dual-Play with Internet and Television services. For Triple Play services cover all Telephone, Internet and Television services. All packages offered have various bandwidth options, ranging from 10 Mbps, 20 Mbps, 30 Mbps, 40 Mbps, 50 Mbps, 100 Mbps, 200 Mbps to 300 Mbps.

In addition, sales of Internet Packages in the city of Jambi during 2021 are dominated by the Jambi City area, which is as much as 64.3% of Total Sales throughout 2021. The remainder is entirely dominated by areas outside Jambi City (Setiawan & Tomsa, 2022). Because of this, the authors decided to conduct research related to Indihome sales in Jambi City, where this research aims to find out the results of the classification of internet package sales that are most in demand by the community so that companies can find out the internet packages that are wanted by the community, so that they can fulfill the community's wishes in the future.

LITERATURE REVIEW

Researcher aims to analyze sales data by applying the K-Means Clustering algorithm to form sales clusters (Fithri & Wardhana, 2021). The results of data clustering form cluster1, cluster2 and cluster3 with percentage values of 62% (11 data), 8% (56 data) and 30% (25 data), respectively. Cluster validation of K-Means Clustering algorithm with Davies Bouldin Index produces a value of 0.2. The information of sales clustering can be an alternative solution, input for stock management and marketing strategies (Doeringer & Terkla, 1995).

Researcher discusses the application of data mining, using the K-Means Clustering algorithm with the CRISP-DM method (Nama et al., 2019). Implementation using RapidMiner 9.10 which is done by entering sales transaction data with a total of 4 attributes and forming 4 clusters consisting of very in demand, in demand, moderate in demand and less in demand. the second cluster with 944 products, the third cluster with 2 products, and the fourth cluster with 43 products.

Researcher (Chih-Hsuan & Yin, 2020) combines the issue of sales prediction with demand planning. The experimental results show that the presented framework successfully achieves the above purpose and has the potential to be generalized to other components of the industry (Quan et al., 2021).

Makes a forecasting model to predict future demand based on previous data. There are three stages of the method used, namely clustering, Clustering, Multivariate Adaptive Regression Splines, Support Vector Regression (Sharifzadeh et al., 2019).

Aims to categorize products in the Shopee marketplace that have a level of interest from low to high by users. The categorization method used is K-Means Clustering (Rejito et al., 2021). Then the clustering results were also validated with the DBI (Davies Bouldin Index) method and the results were good.

RESEARCH METHODOLOGY

Data Collecting

This study uses internet package sales data in Jambi City throughout 2021 which has a total dataset of 14,827. This study uses a variable consisting of total internet package, bandwidth speed and add on data. The data collection method used is in accordance with the Knowledge Discovery in Database (KDD) stages, which can be found with the following stages :

Data Preprocessing

Before the data mining process is carried out, it is necessary to carry out the preprocessing stage first (Tan et al., 2009). At this stage the data goes through the process of cleaning, integration, selection and transformation first to ensure the dataset used is ready to proceed to the data mining process:

1. Data Cleaning
   Data cleaning is the process of removing noise and inconsistent or irrelevant data. In general, the data obtained, both from a company's database and experimental results, have imperfect entries such as missing data, invalid data or just a typo.
2. Data Integration
   Data integration is the merging of data from various databases into one new database. Not infrequently the data needed for data mining does not only come from one database but also comes from several databases or text files.
3. Data Selection
   The data in the database is often not all used, therefore only data that is suitable for analysis will be retrieved from the database. For example, a case that examines the tendency of people to buy in the case of market basket analysis, does not need to take the customer's name, just the customer id.
4. Data Transformation
   The Transformation stage is the stage of changing the data that has been selected, so that the data is suitable for the data mining process. At this stage, the Attribute Construction and Aggregation process will be carried out (Bhatt & Kankanhalli, 2011).

Clustering Process

This stage is the process of looking for interesting patterns or information in the selected data using certain techniques or methods based on the overall KDD process. The method used in this study is the clustering method and the K-Means algorithm (Velmurugan, 2014). Before using the K-Means algorithm, the optimal number of clusters will be determined first using the Davies Bouldin Index method.

The Davies-Bouldin Index (DBI) is a cluster validation method for quantitative evaluation of clustering results. This measurement aims to maximize the intercluster distance between one cluster and another. In this study, DBI will be used to detect outliers in each cluster that is formed. The stages of DBI in this study are as follows :

1. Find the Sum of Square Between clusters (SSB) is the equation used to determine the separation between clusters
2. Find Sum of Square Within cluster (SSW) is an equation used to determine the cohesion matrix in a cluster-i
3. the ratio measurement (Rij) is carried out to determine the comparative value between the i-th cluster and the j-th cluster. A good cluster is a cluster that has the smallest possible cohesion value and the greatest possible separation.
4. The ratio value obtained is used to find the Davies Bouldin Index (DBI) value from the following equation:

$$\mathbf{DBI} = \frac{1}{N}\sum_{i=1}^{N} \max_{i \neq j}(R_{i,j})$$

From this equation, the smaller the DBI value obtained (non-negative>= 0), the better/valid the cluster obtained.

After finishing getting the optimal number of clusters to be used, proceed with clustering with the K-Means algorithm. The stages of the K-Means algorithm in this study are as follows:

1. Determine the number of clusters, in this study used 5 clusters namely Cluster 0, Cluster 1, Cluster 2, Cluster 3 and Cluster 4.
2. Determine the initial center point (centroid) of each cluster.
3. Perform data distance calculations with centroids using the Euclidean Distance equation.
4. After all the data is placed into the closest cluster, then recalculate the new cluster center based on the average members in the cluster.
5. After getting a new center point for each cluster, do it again from step three until the center point for each cluster doesn't change anymore and no data moves from one cluster to another.

Pattern Evaluation

To identify interesting patterns into the knowledge based found. In this stage the results of data mining techniques in the form of typical patterns and predictive models are evaluated to assess whether the existing hypothesis is indeed achieved. If it turns out that the results obtained do not match the hypothesis, there are several alternatives that can be taken, such as making feedback to improve the data mining process, trying other data mining methods that are more suitable, or accepting this result as an unexpected result that may be useful.

RESULTS AND DISCUSSION

Data Collecting

The main data source used in this study is sales data for the entire Jambi City area with a total of 14,827 transactions throughout 2021. The data received comes from the googlesheets database belonging to the Consumer Unit in Jambi City Region. After that, the data is converted into Microsoft Excel, making it easier to clean and filter data.

Table 1. Internet Package Sales Dataset

NO	ORDER ID	REGIONAL	WITEL	DATEL	STO	TYPE TRANSAKSI	STATUS RESUME	STATUS MESSAGE
1	510909101	1	JAMBI	INNER	JMB	NEW SALES	Completed (PS)	Completed
2	510918737	1	JAMBI	INNER	JMB	NEW SALES	Completed (PS)	Completed
3	510958311	1	JAMBI	INNER	JMB	NEW SALES	Completed (PS)	Completed
4	510958665	1	JAMBI	INNER	JMB	NEW SALES	Completed (PS)	Completed
5	510959403	1	JAMBI	INNER	MND	NEW SALES	Completed (PS)	Completed
6	510961801	1	JAMBI	INNER	MND	NEW SALES	Completed (PS)	Completed
7	510961832	1	JAMBI	OUTER	SRJ	NEW SALES	Completed (PS)	Completed
8	510962368	1	JAMBI	INNER	JMB	NEW SALES	Completed (PS)	Completed
9	510963093	1	JAMBI	OUTER	MBN	NEW SALES	Completed (PS)	Completed
10	510964570	1	JAMBI	INNER	JMB	NEW SALES	Completed (PS)	Completed
11	510964588	1	JAMBI	OUTER	SRJ	NEW SALES	Completed (PS)	Completed
12	510964670	1	JAMBI	OUTER	MAB	NEW SALES	Completed (PS)	Completed

Data Preprocessing

The next stage after data collection is through Data Preprocessing to prepare data so that it is ready for processing. In this preprocessing stage, data needs to be integrated because the raw data is still separated between transaction data every month. After the data is integrated, the data is transferred to Microsoft Excel for the preprocessing stage. Microsoft Excel was chosen because of its ease of use and compatibility with RapidMiner tools which will be used for the data mining process.

5. Data Cleaning

   This data cleaning process must go through several processes, such as filling in missing values and resolving found data inconsistencies. Data that has passed the Data Cleaning Stages will be stored in a new dataset using Microsoft Excel.

Table 2. Data Cleaning Process

No	Problem	Solution
1	The Service Type attribute does not match the contents of the Package Name. Some values in the Service Type attribute are written as 3P even though the value in the Package Name attribute writes "Usee" which indicates Television service but does not write "Minutes" which indicates telephone service, and vice versa	Make improvements by writing the formula =IF(IFERROR(AND(SEARCH("USEE",Z2),SEARCH("Minute",Z2))),"3P","2P") so that the Service Type value is more accurate
2	Speed attribute does not match Package Name. Some values in the service attribute write down the speed of the service that doesn't match the Package Name. For example writing in the speed attribute "10 Mbps", but in the actual Package Name attribute it is "20 Mbps".	Make improvements by writing the formula =if(regexmatch(Z2,"10 Mbps"),"10 Mbps", if(regexmatch(Z2,"20 Mbps"),"20 Mbps", if(regexmatch(Z2,"30 Mbps")),"30 Mbps", if(regexmatch(Z2,"40 Mbps"),"40 Mbps", if(regexmatch(Z2,"50 Mbps"),"50 Mbps", if(regexmatch(Z2,"100 Mbps "),"100 Mbps", if(regexmatch(Z2,"200 Mbps"),"200 Mbps", if(regexmatch(Z2,"300 Mbps"),"300 Mbps","N/A")))))))) so that the Speed value is more accurate.

6. Data Integration

When doing Data Integration, researchers must combine sales transaction data that is still separate into different data. Then do Data Integration to produce datasets that are ready to be processed. The dataset used is a combination of 12 tables, namely January Sales Data, February Sales Data, March Sales Data, April Sales Data, May Sales Data, June Sales Data, July Sales Data, August Sales Data, September Sales Data, October Sales Data, November Sales Data, December Sales Data.

7. Data Selection

Next, the researcher chose 5 attributes used, namely STO, Package Name, Service Type, Speed, and NPER. These five attributes were chosen because they have an influence on the next stage in the data mining process.

Because the limitation of the problem in this study is Jambi City sales data for 2021, researchers must first filter the area based on the STO Attributes. The categorization of the Jambi City area is in accordance with that contained in Table 2.2. so we need to do filtering on these attributes specifically for the JMB, JMI, KOA, MND, PAP STO areas.

Table 3. Selection Results

Number	STO	Package Name	Service Type	Speed	Closed Date
1	JMB	20 Mbps, 300 Menit, E-Learning, Cloud Storage 16GB, seamless	2P	20 Mbps	202101
2	MND	20 Mbps, 300 Menit, Add on Gamer, Langit Music Cloud, seamless	2P	20 Mbps	202101
3	MND	20 Mbps, Usee Entry	2P	20 Mbps	202101
4	JMB	20 Mbps, Usee New Basic, Iflix, Catchplay	2P	20 Mbps	202101
5	JMB	20 Mbps, 300 Menit, E-Learning, Cloud Storage 16GB, seamless	2P	20 Mbps	202101

8. Data Transformation

At this stage, the process of changing the data is carried out, so that the data can be processed using the K-Means Clustering algorithm. The purpose of this Data Transformation is how to change attribute data that has gone through the Selection and Preprocessing process into a valid Data Package. The Data Transformation Process that will be carried out is as follows:

a. Added the Bundling attribute

This Bundling attribute functions to find out what bundling services have been ordered by customers. There are 3 types of service bundling, namely 2P (Telephone and Internet), 2P (Internet and TV) and 3P (Telephone, Internet and TV).

b. Added Add On attribute

This Add On attribute functions to see what additional services are ordered by Customers outside of the main Bundling Package.

c. Recent attribute data changes

After adding these new attributes, it's time to re-select the data to be used in the next stage. STO attribute data, Package Name, and Service Type are deleted leaving only Speed, Bundling and Add On attributes. Then the Closed Order Date attribute is simplified to Month to make it easier to understand.

d. Merge attributes

After the process of changing the attribute data, then the attribute data that was made before is merged. The combined attributes are Speed, Bundling, and Add On. The combination of these three attributes will represent the types of internet packages that have been offered to customers.

e. Adding Sales Attributes Every Month

Added a new attribute to the dataset, namely Sales, which was obtained from the Month attribute. After the Package data has been combined into one, it is continued by adding the Total Sales attribute every month to support the search for the best-selling packages.

Here are the results of the final data that will be used for the next clustering process:

Table 4. Final Dataset After Preprocessing

PACKAGE	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Paket 20 Mbps Bundling Internet + Telepon Add On Music & Gamer	331	264	226	194	129	174	244	196	171	145	99	98
Paket 20 Mbps Bundling Internet + Telepon Add On Pendidikan	1	1		1		1		1	2	1		
Paket 20 Mbps Bundling Internet + Telepon Add On Pendidikan & IOT	24	10	11	8	6	43	27	10	7	4	3	4
Paket 20 Mbps Bundling Internet + Telepon Tanpa Add On				2						10	8	6
Paket 20 Mbps Bundling Internet + TV Add On Movies	29	28	29	21	10	36	26	21	8	4	1	3
Paket 20 Mbps Bundling Internet + TV Add On Movies & Music	25	18	17	14	30	29	15	10	8	17	18	11
Paket 20 Mbps Bundling Internet + TV Tanpa Add On	19	19	25	12	3	6	14	5	7		1	
Paket 10 Mbps Bundling Internet + Telepon + TV Add On Pendidikan												5
Paket 10 Mbps Bundling Internet + Telepon + TV Tanpa Add On	470	333	282	211	140	235	256	252	268	109	11	38
…	…	…	…	…	…	…	…	…	…	…	…	…
Paket 300 Mbps Bundling Internet + TV Tanpa Add On											2	

Data Mining Process

a. Finding Optimal Number Of Clusters with Davies Bouldin Index

Before going through the Clustering Process, it is necessary to search for the optimal K value first with the Davies Bouldin Index method where we need to find the minimum Davies Bouldin Index value among other clusters.

From the results of calculating the Davies Bouldin Index value using the RapidMiner tool, the researcher can see that the lowest point of the graph is when the position of the number of clusters is at point 5 with the smallest Davies Bouldin Index value of 0.035. Therefore, the researcher decided to use a total of 5 clusters because it could represent the optimal number of k clusters.

b. K-Means Clustering Process

From previous dataset, the clustering process is then carried out using the K-Means method using RapidMiner tools. After the clustering process is executed, the output of the RapidMiner tools produces produces 5 model clusters. Each cluster has members, including Cluster 0 has 74 items, Cluster 1 has 1 item, Cluster 2 has 1 item, Cluster 3 has 1 item, and Cluster 4 has 1 item.

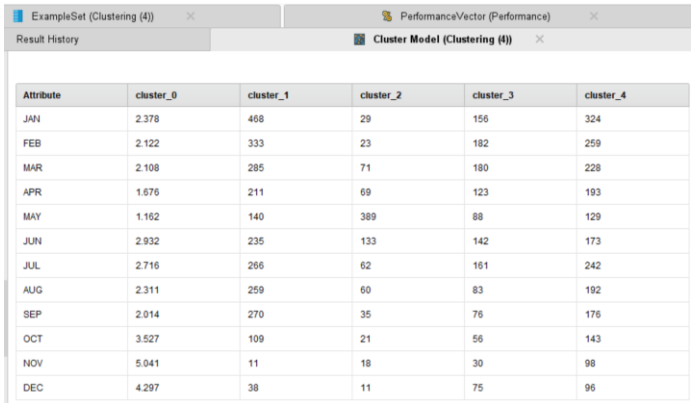
From the number of members of each cluster, the researcher can see which members belong to each cluster. The following is the data of members who enter each cluster :

Table 5. Members of Each Cluster

No	Cluster	Member Counts	Member's Name
1	Cluster_0	74	• 10 Mbps Bundling Internet + Telephone + TV Add On Education • 10 Mbps Bundling Internet + Telephone + TV Tanpa Add On • 10 Mbps Bundling Internet + Telephone Add On Music & Gamer • 10 Mbps Bundling Internet + Telephone Add On Education • 10 Mbps Bundling Internet + Telephone Without Add On • …….
2	Cluster_1	1	Paket 20 Mbps Bundling Internet + Telephone Add On Music & Gamer
3	Cluster_2	1	Paket 20 Mbps Bundling Internet + TV Tanpa Add On
4	Cluster_3	1	Paket 20 Mbps Bundling Internet + Telephone Add On Education
5	Cluster_4	1	Paket 20 Mbps Bundling Internet + TV Add On Movies

c. Analysis of K-Means Clustering Results

After knowing the members of each cluster, the next step is to analyze the characteristics of the cluster members. The first thing to do is to calculate the average total sales of each month's attribute for each cluster first.


ExampleSet (Clustering (4)) | PerformanceVector (Performance)
Result History | Cluster Model (Clustering (4))

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
JAN	2.378	468	29	156	324
FEB	2.122	333	23	182	259
MAR	2.108	285	71	180	228
APR	1.676	211	69	123	193
MAY	1.162	140	389	88	129
JUN	2.932	235	133	142	173
JUL	2.716	266	62	161	242
AUG	2.311	259	60	83	192
SEP	2.014	270	35	76	176
OCT	3.527	109	21	56	143
NOV	5.041	11	18	30	98
DEC	4.297	38	11	75	96

Figure 1. Final Results of the Average Sales of Each Cluster Every Month

From the results shown in Figure 5 it can be seen that the results of each cluster compared to the total sales each month. Looking on the average value of each clusters, the researcher can see the characteristics of each cluster on internet package sales in Jambi City every month:

Table 6. Sales Characteristics of Each Cluster

No	Cluster Name	Characteristics
1	Cluster_0	Has no dominance at all on sales in Jambi City.
2	Cluster_1	Has the dominance of sales in Jambi City almost every month, namely in January, February, March, April, June, July, August and September.
3	Cluster_2	Has dominance of sales in Jambi City only in May. After that sales can be said to be not so significant until so on.
4	Cluster_3	Has a pretty good contribution in sales but unfortunately never dominates every month
5	Cluster_4	Sales dominate Jambi City in November and December and in majority have the second best package sales after cluster_1.

From the results of these characteristics, the researcher must first look at the center point of each cluster to ensure the position of each cluster relative to one another. The trick is to calculate the average of the total monthly sales value of each cluster which is obtained from Figure 5.

From calculating the average total sales value for each cluster, then the author maps it into a plot graph to see the position of each cluster to be like in the Figure 7.

Figure 2. Cluster Plotting

Based on the results of cluster plotting in Figure 7. researchers can see the position of each centroid cluster from the lowest position to the highest. From this, researchers can conclude which clusters are in the best-selling package category to those that are not selling. So that the writer can conclude the categorization of each cluster as shown in Table 8.

Table 7. Best Selling Package Categorization

No	Cluster Name	Package Name	Category
1	Cluster_1	20 Mbps Bundling Internet + Telephone Add On Music & Gamer	Very In Demand
2	Cluster_4	20 Mbps Bundling Internet + TV Add On Movies	In Demand
3	Cluster_3	20 Mbps Bundling Internet + Telephone Add On Education	Moderate In Demand
4	Cluster_2	20 Mbps Bundling Internet + TV Without Add On	Less In Demand
5	Cluster_0	• 10 Mbps Bundling Internet + Telephone + TV Add On Education • 10 Mbps Bundling Internet + Telephone + TV Without Add On • 10 Mbps Bundling Internet + Telephone Add On Music & Gamer • 10 Mbps Bundling Internet + Telephone Add On Education • 10 Mbps Bundling Internet + Telephone Without Add On • ……..	Not In Demand

Pattern Evaluation

Based on the analysis results from Table 4.8. researchers can classify packets from each cluster.

1. Cluster 0

This cluster is included in the Not In Demand category with an average sales value of 2,690. In this case, researchers can see that the services beyond the 20 Mbps speed, especially those that are larger, are still not much needed by the majority of customers in Jambi City.

2. Cluster 1

This cluster is included in the Very In Demand Package category with an average sales value of 218.75. The researchers can conclude that the majority of customer needs in the city of Jambi for internet speed are sufficient with just 20 Mbps. Then there are telephone and internet bundling packages where researchers can conclude that the majority of Jambi customers do not yet require IPTV services, they only need internet access and a telephone. Then with Add On Music and

Gamer, it can be said that customers in the city of Jambi are interested in accessing music and gaming services.

3. Cluster 2

This cluster falls into the Less In Demand category with an average sales value of 76.75. If you look at the contents of the package, it's the same as Cluster 1, the majority of these customers tend to subscribe speed package of 20 Mbps, but here there is no telephone bundling, only internet and TV. And what's different is that this package has no add-ons, so customers of this package are not at all interested in the additional services have provided.

4. Cluster 3

This cluster is included in the package category which is quite in demand with an average sales value of 112,667. If you look at the contents of the package, it's the same as Cluster 1. What's different is that this package has an Education add-on, where this package tends to be intended for customers who need access to online learning because of this add-on, which contains E-Learning packages which are educational digital content applications and Cloud Storage which functions as virtual storage as archival support the online learning.

5. Cluster 4

This cluster is included in the In Demand category with an average centroid of 187.75. If you look at the contents of the package, they also subscribe to a 20 Mbps speed service, and just like Cluster 2, this package only has internet and television bundling. And this package, according to researchers, is also intended for customers who like to consume entertainment on television media, such as movies and series.

Recommendation For Management

Here are some suggestions for the management of the company regarding the marketing strategy in Jambi City based on the results of the analysis that the researchers previously described:

a. Researchers suggest combining the three Add On Packages (Very in Demand, In Demand and Moderate in Demand) into a new package to add value to the Package itself, so that the scope of customer needs becomes wider.
   1) with lots of additional add-ons auto on will require greater service speed. So the researcher also suggests upgrading the service speed for the new package to 30 Mbps to better accommodate the additional add ons of the package.
   2) the researchers also suggest that in the new package to remove the telephone bundling option, so that only Internet and television bundling packages are according to needs.
b. So, the researcher suggests making the latest package offered to customers, namely the 30 Mbps Internet and Television Bundling package with Add On Music, Gamers, Education and Movies to better accommodate all the needs of potential customers in Jambi.
c. The researcher also suggests creating an existing package upgrade program with an additional, more affordable cost compared to buying a new package specifically for the following package categories : Very in Demand, In Demand and Moderate in Demand. This program offer is intended to increase customer engagement for existing customers as well so that they continue to subscribe to the product.

CONCLUSIONS AND SUGGESTIONS

Conclusions

After carrying out the data mining process with the k-Means clustering algorithm on sales data in the Jambi City area, it can be concluded that the classification of Best Selling Internet Packages in Jambi City is divided into 5 clusters with the following details:

a. Very in Demand category, includes 1 Package, namely the 20 Mbps Speed Package, with Internet + Phone Bundling, and Add On Music & Gamer.
b. In Demand category includes 1 Package, namely the 20 Mbps Speed Package, with Internet and Television Bundling, and Add On Movies.
c. Moderate In Demand category, has 1 Package, which is a 20 Mbps Speed Package, with Internet and Telephone Bundling, as well as an Education Add On.
d. Less In Demand category, which is a 20 Mbps Speed Package, with Internet and Television Bundling without additional Add Ons.
e. Not In Demand category with the highest number of packages, namely 74 packages, apart from the packages previously mentioned.

Suggestions

The researcher suggests to the management of the company to carry out a strategy of merging the three packages which fall into the Very In Demand, In Demand and Moderate In Demand categories to be combined so that the penetration of customers in Jambi city becomes wider in scope to suit the needs of the community.

In addition, the researchers also suggested that existing customers who had previously subscribed to the package to offer the combined package as a form of customer engagement with existing customers so that they would continue to subscribe to the products in the future.

REFERENCES

Bhatt, C. A., & Kankanhalli, M. S. (2011). Multimedia Data Mining: State Of The Art And Challenges. *Multimedia Tools And Applications*, *51*, 35–76.

Doeringer, P. B., & Terkla, D. G. (1995). Business Strategy And Cross-Industry Clusters. *Economic Development Quarterly*, *9*(3), 225–237.

Fithri, F. A., & Wardhana, S. (2021). The CLUSTER ANALYSIS OF SALES TRANSACTION DATA USING K-MEANS CLUSTERING AT TOKO USAHA MANDIRI. *Jurnal PILAR Nusa Mandiri*, *17*(2), 113–118.

Frederick, D. P., & Bhat, G. (2021). Review On Customer Perception Towards Online Food Delivery Services. *International Journal Of Creative Thoughts (IJCRT)*, *9*(7), B301–B314.

Hutajulu, S., Dhewanto, W., & Prasetio, E. A. (2020). Two Scenarios For 5G Deployment In Indonesia. In *Technological Forecasting And Social Change* (Vol. 160, P. 120221). Elsevier.

Nama, G. F., Lukmanul, H., & Junaidi, J. (2019). Implementation Of K-Means Technique In Data Mining To Cluster Researchers Google Scholar Profile. *International Journal Of Engineering And Advanced Technology (IJEAT)*, *9*(1).

Pereira, P., Ribeiro, T., & Vareda, J. (2013). Delineating Markets For Bundles With Consumer Level Data: The Case Of Triple-Play. *International Journal Of Industrial Organization*, *31*(6), 760–773.

Quan, H., Li, S., Zeng, C., Wei, H., & Hu, J. (2021). Big Data Driven Product Design: A Survey. *Arxiv Preprint Arxiv:2109.11424*.

Rejito, J., Atthariq, A., & Abdullah, A. S. (2021). Application Of Text Mining Employing K-Means Algorithms For Clustering Tweets Of Tokopedia. *Journal Of Physics: Conference Series*, *1722*(1), 12019.

Setiawan, K. M. P., & Tomsa, D. (2022). *Politics In Contemporary Indonesia: Institutional Change, Policy Challenges And Democratic Decline*. Routledge.

Sharifzadeh, M., Sikinioti-Lock, A., & Shah, N. (2019). Machine-Learning Methods For Integrated Renewable Power Generation: A Comparative Study Of Artificial Neural Networks, Support Vector Regression, And Gaussian Process Regression. *Renewable And Sustainable Energy Reviews*, *108*, 513–538.

Tan, K. C., Teoh, E. J., Yu, Q., & Goh, K. C. (2009). A Hybrid Evolutionary Algorithm For Attribute Selection In Data Mining. *Expert Systems With Applications*, *36*(4), 8616–8630.

Velmurugan, T. (2014). Performance Based Analysis Between K-Means And Fuzzy C-Means Clustering Algorithms For Connection Oriented Telecommunication Data. *Applied Soft Computing*, *19*, 134–146.