Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

Agglomerative Hierarchical Clustering (AHC) Method for Data Mining Sales Product Clustering

Ridha Maya Faza Lubis[1,*], Jen-Peng Huang[2], Pai-Chou Wang[2], Kiki Khoifin[1], Yuli Elvina[1], Dyah Ayu Kusumaningtyas[1]

[1] Department of Business and Management, Southern Taiwan University of Science and Technology, Taiwan
[2] Department of Information Management, Southern Taiwan University of Science and Technology, Taiwan
Email: [1,*]db01g208@stust.edu.tw, [2]jehuang@stust.edu.tw, [3]pwang@stust.edu.tw,
[4]db11g215@stust.edu.tw,[5]da911g211@stust.edu.tw , [6]db01g207@stust.edu.tw
Correspondence Author Email: db01g208@stust.edu.tw
Submitted: **04/06/2023**; Accepted: **29/06/2023**; Published**: 29/06/2023**

Abstract−Supermarkets are Indonesian terms that refer to large stores or supermarkets that offer a variety of daily needs such as food, drinks, cleaning products, household appliances, clothing, and so on. In contrast to stalls or small shops, supermarkets have a larger size and provide a variety of products. Because of this, many people prefer to shop for their daily needs at the supermarket rather than at the nearest shop because the existence of the supermarket makes it easier for consumers to buy various products in one place without having to move to another store. However, sales in supermarkets also pose a problem, namely how to sort or group products that are not selling well so they can be replaced with products that are selling better or reduce the number of suppliers. This is where data mining or data analysis techniques that use business intelligence are needed. The research was conducted to classify the best-selling products in supermarkets using the Agglomerative Hierarchical Clustering (AHC) method, in which alternatives with the same matrix or distance are grouped into certain clusters. In applying the AHC method, the number of clusters formed is 3. There are three different clusters, namely cluster 0, cluster 1, and cluster 2, each with a different alternative group. Each cluster has a different number of products and a different percentage. Cluster 0 is the cluster with the highest number of products and the largest percentage, namely 45% with a total of 9 products, followed by cluster 2, and cluster 1 has the smallest number of products and percentage, namely 0.30% with a total of 6 products and 0 .25% with a total of 5 products. In addition, sales data for several products each month are grouped based on certain price ranges.

Keywords: Data Mining; Clustering; AHC Method; Sales Products

1. INTRODUCTION

Self-service stores or shops are referred to as such in Indonesian and offer a range of daily necessities, such as food, drinks, cleaning supplies, home appliances, clothing, and others. Supermarkets are larger and offer a wider variety of items than tiny shops or market stalls. Today, many mothers choose to shop at the supermarket rather than the neighborhood store for their daily requirements. This is due to the fact that supermarkets allow customers to quickly purchase a variety of goods in one location without having to visit other establishments. As a result, supermarkets facilitate shopping for customers and help them save time.

Supermarket staff undoubtedly filter (group) products that are thought to be less in demand in order to minimize the number of suppliers or they can be replaced by raising the number of stocks of products that are in demand. This is due to the rise in sales at supermarkets. Grouping is simple if the number of products being sold is limited, but in supermarkets, where there are usually a variety of products, it can be challenging to determine which ones are doing well. Based on these issues, data mining, a method of data analysis employing fresh information or business intelligence, is required.

Data mining is a process that examines and identifies information produced from massive volumes of data using artificial intelligence, statistics, and machine learning technologies. Data mining will be very advantageous in the future since it can create patterns that help with decision-making. Large amounts of data are used when conducting data mining, making this technique crucial. Each group of data mining experts addresses a different set of issues. Association, estimate, description, prediction, grouping, and classification are some of these categories [1][2][3]. According to the clustering research purpose, the clustering method is the main emphasis of this study. A data mining technique or approach called clustering is used to categorize data according to how similar they are. This method is used to group related data in a dataset [4][5][6]. The K-Means, SOM, K-Medoids, Fuzzy C-Means, and AHC algorithms can all be used in the clustering process. The best-selling items in a supermarket are grouped in this study using the Agglomerative Hierarchical Clustering (AHC) approach.

An earlier study by Harun Al Rasyid, et al. in 2022 discussed how Nutrition Stores had trouble maintaining product inventories because of the quick turnover of goods, leading to empty items' availability and the necessity of delaying purchases. The Nutrition Store may arrange inventories by sales within a specific timeframe by using transaction and sales data to analyze trends of sales and circulation of goods. Clustering, specifically the K-Means algorithm, is a data mining technique used to determine the degree of sales based on the formation of clusters. The findings revealed that three clusters were constructed from the 41 categories of sales and transaction data collected from April to May 2022, including extremely desirable (3), moderately desirable (2), and less desirable (1). Each cluster had several sold items[7].

According to research done in 2019 by Muhammad Dahria et al, PT. Koko Pelli has difficulty choosing the best product to promote since it bases its decisions solely on sales data, whereas quality and pricing must also be taken

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

into account. Sales and earnings are anticipated to rise as a result of grouping the finest products utilizing data mining technology and the k-means algorithm. Two goods were classified in Cluster 1, nine products were grouped in Cluster 2, and three products were grouped in Cluster 3, according to the findings of this study. Thus, it was anticipated that PT. Koko Pelli would be assisted in choosing the finest products to advertise and boost earnings by employing the best product grouping developed using Data Mining technology [8].

In 2021, Banu Harli Trimulya Suandi As and Lisna Zahrotun conducted research on the admissions procedure at the Faculty of Industrial Technology, Ahmad Dahlan University, which was confronted with several issues, including the extremely high number of new students entering, the low number of students graduating on time, and the imbalance in the ratio of lecturers and excellent students. This issue affects inefficient teaching and learning techniques as well as the overuse of university infrastructure. Therefore, a hierarchical clustering method was used in the research to recommend prospective students who satisfied certain criteria for admission to that faculty. This approach consists of several processes, including data loading, cleaning, transformation, Euclidean distance, agglomerative hierarchical clustering, and testing with the silhouette coefficient. The findings revealed that 158 student records were recommended, all of whom were from Java and had math test scores that satisfied specific requirements for each study program. The Silhouette Coefficient produces very good test results, with values for each study program of 0.868, 0.883, 0.879, and 0.873[9].

2. RESEARCH METHODOLOGY

2.1 Research Stages

A set of actions known as research stages will be taken to finish this study. To help readers better comprehend the research's contents, these phases are organized in a structured way. The research will be conducted in the steps listed below:

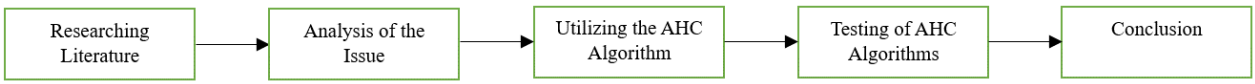


Figure 1. Research Stages

The processes involved in carrying out this research were evident from the stages of the research depicted in Figure 1, which included:

a. Analyzing the Literature
   Literature review becomes a crucial component of an effective research process early on. When doing literature research, sources of information are gathered, examined, judged, and synthesized from a variety of sources, including books, journals, articles, reports, and other materials pertinent to the topic or issue being investigated. The goal of a literature review is to get a greater grasp of the subject or issue being investigated and to spot knowledge gaps that need to be addressed by more research.
b. Problem-Solving
   issue analysis is a methodical procedure for comprehending and locating the source of an issue. The goal of problem analysis is to identify the underlying cause of the issue to find a suitable and efficient solution. The problem analysis method comprises gathering, analyzing, and interpreting data to determine the elements contributing to the issue and how they affect the system in question.
c. Utilizing the AHC Algorithm
   The researcher's next action, after performing the analysis, is to apply the algorithm to the data that has been gathered. Data is processed in a particular way using algorithms. The AHC approach is the algorithm employed in this work to resolve the issue.
d. Testing of AHC Algorithms
   The AHC algorithm has to be tested after being put into practice. Using RapidMiner, algorithm testing is done. The test is deemed successful if the findings match those obtained via its application.
e. Conclusion
   The process of gathering a summary of research or analytical findings.

2.2 Data Mining

The goal of the field of data mining, which emerged from databases, is to discover new knowledge. This knowledge is produced by a process of large-scale data analysis to find hidden and useful data in a database. Finding patterns, correlations, and trends that can be used in decision-making requires the application of artificial intelligence, mathematics, statistics, machine learning, and database approaches[10][11][12]. In addition to helping people learn new things, data mining can also help people identify and create patterns that will be helpful in the future. These tendencies may have gone unnoticed in the past. Large databases are utilized in the data mining process, also known as a data mining process, to gather fresh data[1][13][5] [4][14]. Several categories of data mining techniques, such as clustering, prediction, classification, estimation, and association, can be utilized to assist in problem-solving in research[15][16][17][18][19].

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

2.3 Clustering

One method for grouping objects based on similarities is called clustering. Depending on how many clusters are wanted, similar things will be placed together into one cluster, while distinct objects will be divided into several clusters. Several methods, including K-Means, SOM, K-Medoids, Fuzzy C-Means, and AHC, can be employed in clustering procedures[20][21].

2.4 AHC Method

Data mining techniques for creating clusters include AHC (Agglomerative Hierarchical Clustering). Each starting object is treated as a single atomic cluster or cluster throughout the cluster-building process utilizing the AHC method. These objects are then joined with other objects that are nearby to form a larger cluster. Up until a particular set of requirements are met, this merging procedure is repeated. The Euclidean Distance is used to determine the separation between two things [21][22][23].

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \quad (1)$$

The subsequent clustering procedure can create clusters using agglomerative techniques [24][25][26]:

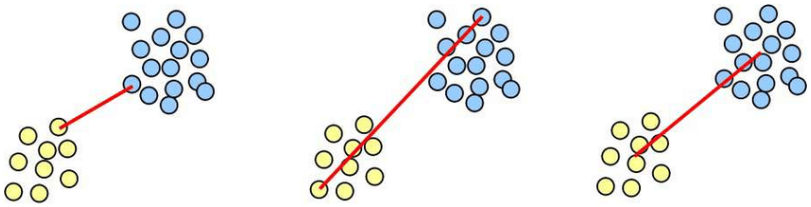


Figure 2. Mode of Data Grouping

The distinctions between the three ways (methods) of grouping data, namely Single Linkage, Complete Linkage, and Average Linkage, can be understood from Figure 2. The nearest data in each of the two clusters are selected using Single Linkage based on their shortest distance. Complete Linkage selects the distance that separates the two clusters' farthest-removed data points. The average distance between each pair of data from the two clusters is then chosen by Average Linkage[27].

3. RESULTS AND DISCUSSION

Products in a store can be categorized according to how well-liked they are utilizing data grouping methods with the AHC algorithm. Objects or data points with the same distance are grouped in the AHC algorithm. After first being grouped as a separate group, each data point is then gradually added to other groups until all of the data points have been added to one group. In this study, 20 sales data from the four months of September, October, November, and December are used as a sample. A list of the sample data used in this investigation is shown in Table 1.

Table 1. Sales Data

Number	Item Code	Selling Price			
		September	October	November	December
1	KB11001	20	34	32	40
2	KB11002	18	15	23	49
3	KB11003	13	13	28	55
4	KB11004	23	43	46	38
5	KB11005	10	51	30	30
6	KB11006	14	37	44	46
7	KB11007	9	29	33	33
8	KB11008	36	27	38	28
9	KB11009	32	48	32	51
10	KB11010	12	35	16	43
11	KB11011	24	50	31	58
12	KB11012	8	34	18	69
13	KB11013	38	25	10	30
14	KB11014	29	38	32	66
15	KB11015	13	32	41	53
16	KB11016	56	46	23	34
17	KB11017	58	43	21	52

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

18	KB11018	35	25	12	67
19	KB11019	20	33	36	29
20	KB11001	30	28	40	44

3.1 Utilizing the AHC technique

a. Apply the following Euclidean formula to determine the matrix distance.

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

$$d(\text{KB11001}, \text{KB11002}) = \sqrt{(20-18)^2 + (34-15)^2 + (32-23)^2 + (40-49)^2}$$

$$d(\text{KB11001}, \text{KB11003}) = \sqrt{(20-13)^2 + (34-13)^2 + (32-28)^2 + (40-55)^2}$$

$$d(\text{KB11001}, \text{KB11004}) = \sqrt{(20-23)^2 + (34-43)^2 + (32-46)^2 + (40-38)^2}$$

$$d(\text{KB11001}, \text{KB11005}) = \sqrt{(20-10)^2 + (34-51)^2 + (32-30)^2 + (40-30)^2}$$

$$d(\text{KB11001}, \text{KB11006}) = \sqrt{(20-14)^2 + (34-37)^2 + (32-44)^2 + (40-46)^2}$$

$$d(\text{KB11001}, \text{KB11007}) = \sqrt{(20-9)^2 + (34-29)^2 + (32-33)^2 + (40-33)^2}$$

$$d(\text{KB11001}, \text{KB11008}) = \sqrt{(20-36)^2 + (34-27)^2 + (32-38)^2 + (40-28)^2}$$

$$d(\text{KB11001}, \text{KB11009}) = \sqrt{(20-32)^2 + (34-48)^2 + (32-32)^2 + (40-51)^2}$$

$$d(\text{KB11001}, \text{KB11010}) = \sqrt{(20-12)^2 + (34-35)^2 + (32-16)^2 + (40-43)^2}$$

Apply the aforementioned calculations to the Euclidean KB11019 and KB11020. The outcome of computing the calculated Euclidean matrix is as follows:

Table 2. Euclidean Distance Matrix

Code	KB11001	KB11002	KB11003	KB11004	KB11005	KB11006	...	KB11019	KB11020
KB11001	0						...		
KB11002	22,96	0					...		
KB11003	27,04	9,487	0				...		
KB11004	17,03	38,2	40,16	0			...		
KB11005	22,2	42,07	45,63	23,52	0		...		
KB11006	15	30,82	30,23	13,6	25,77	0	...		
KB11007	14	25,16	27,95	24,21	22,43	19,47	...		
KB11008	22,02	33,67	39,42	24,27	36,33	30,72	...		
KB11009	21,47	37,01	40,22	21,7	30,63	24,78	...		
KB11010	18,17	22,83	27,8	33,32	25	28,3	...		
KB11011	24,43	37,5	38,83	25,98	31,34	24,12	...		
KB11012	34,37	29,77	27,6	45,29	44,25	35,36	...		
KB11013	31,45	32,09	41,45	43,69	43,13	46,17	...		
KB11014	27,8	31,94	31,91	32,26	42,78	27,75	...		
KB11015	17,41	25,57	23,11	21,7	31,94	9,17	...		
KB11016	39,46	51,28	58,34	40,53	46,97	49,3	...		
KB11017	42,31	48,96	54,62	45,23	54,16	50,37	...		
KB11018	37,88	28,88	32,06	49,65	54,72	45,28	...		
KB11019	11,75	29,95	34,48	17,03	21,47	20,12	...	0	
KB11020	14,7	25,04	27,91	18,6	35	18,89	...	19,13	0

b. By joining the two clusters with the shortest distance (minimum value over the entire matrix), the single-linked mode is created. The following table shows that the minimum value is 9.17 (KB11006, KB11015).

Table 3. Matrix Minimum Value

Code	KB11001	KB11002	KB11003	KB11004	KB11005	KB11006	...	KB11019	KB11020
KB11001	0						...		
KB11002	22,96	0					...		
KB11003	27,04	9,487	0				...		
KB11004	17,03	38,2	40,16	0			...		
KB11005	22,2	42,07	45,63	23,52	0		...		
KB11006	15	30,82	30,23	13,6	25,77	0	...		
KB11007	14	25,16	27,95	24,21	22,43	19,47	...		
KB11008	22,02	33,67	39,42	24,27	36,33	30,72	...		
KB11009	21,47	37,01	40,22	21,7	30,63	24,78	...		

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

KB11010	18,17	22,83	27,8	33,32	25	28,3	...		
KB11011	24,43	37,5	38,83	25,98	31,34	24,12	...		
KB11012	34,37	29,77	27,6	45,29	44,25	35,36	...		
KB11013	31,45	32,09	41,45	43,69	43,13	46,17	...		
KB11014	27,8	31,94	31,91	32,26	42,78	27,75	...		
KB11015	17,41	25,57	23,11	21,7	31,94	**9,17**	...		
KB11016	39,46	51,28	58,34	40,53	46,97	49,3	...		
KB11017	42,31	48,96	54,62	45,23	54,16	50,37	...		
KB11018	37,88	28,88	32,06	49,65	54,72	45,28	...		
KB11019	11,75	29,95	34,48	17,03	21,47	20,12	...	0	
KB11020	14,7	25,04	27,91	18,6	35	18,89	...	19,13	0

In the following matrix update procedure, the codes KB11006 and KB11015 are removed and replaced with the combined codes that have been received, namely KB11006, and KB11015, based on the information in Table 3 which shows that the first cluster produced is KB11006, KB11015.

c. Analyze the matrices that have been mixed with other matrices to find the minimum value.

$$D(\text{KB11006 KB11015}, \text{KB11001}) = \min\{D(\text{KB11006}, \text{KB11001}), D(\text{KB11015}, \text{KB11001})\} = \min\{15; 17{,}41\} = 15$$

$$D(\text{KB11006 KB11015}, \text{KB11002}) = \min\{D(\text{KB11006}, \text{KB11002}), D(\text{KB11015}, \text{KB11002})\} = \min\{30{,}82; 25{,}57\} = 25{,}57$$

$$(\text{KB11006 KB11015}, \text{KB11003}) = \min\{D(\text{KB11006}, \text{KB11003}), D(\text{KB11015}, \text{KB11003})\} = \min\{30{,}23; 23{,}11\} = 23{,}11$$

$$(\text{KB11006 KB11015}, \text{KB11004}) = \min\{D(\text{KB11006}, \text{KB11004}), D(\text{KB11015}, \text{KB11004})\} = \min\{13{,}6; 21{,}7\} = 13{,}6$$

$$(\text{KB11006 KB11015}, \text{KB11005}) = \min\{D(\text{KB11006}, \text{KB11005}), D(\text{KB11015}, \text{KB11005})\} = \min\{25{,}77; 31{,}94\} = 25{,}77$$

$$(\text{KB11006 KB11015}, \text{KB11007}) = \min\{D(\text{KB11006}, \text{KB11007}), D(\text{KB11015}, \text{KB11007})\} = \min\{19{,}47; 22{,}11\} = 19{,}47$$

$$(\text{KB11006 KB11015}, \text{KB11008}) = \min\{D(\text{KB11006}, \text{KB11008}), D(\text{KB11015}, \text{KB11008})\} = \min\{30{,}72; 34{,}47\} = 30{,}72$$

$$(\text{KB11006 KB11015}, \text{KB11009}) = \min\{D(\text{KB11006}, \text{KB11009}), D(\text{KB11015}, \text{KB11009})\} = \min\{24{,}78; 26{,}50\} = 24{,}78$$

$$(\text{KB11006 KB11015}, \text{KB11010}) = \min\{D(\text{KB11006}, \text{KB11010}), D(\text{KB11015}, \text{KB11010})\} = \min\{28{,}3; 27{,}11\} = 27{,}11$$

Perform the aforementioned calculations by locating the cluster's minimum value for KB11006 and KB11015; there is no need to search for the minimum value for other clusters. The outcomes of the aforementioned settlement are shown in the following updated matrix table.

Table 4. First Cluster Formation Matrix

Code	KB1106, KB11015	KB11001	KB11002	KB11003	KB11004	KB11005	KB11007	...	KB11019	KB11020
KB1106, KB11015	0							...		
KB11001	15	0						...		
KB11002	25,57	22,96	0					...		
KB11003	23,11	27,04	9,49	0				...		
KB11004	13,6	17,03	38,2	40,16	0			...		
KB11005	25,77	22,2	42,07	45,63	23,52	0		...		
KB11007	19,47	14	25,16	27,95	24,21	22,43	0	...		
KB11008	30,72	22,02	33,67	39,42	24,27	36,33	30,72	...		
KB11009	24,78	21,47	37,01	40,22	21,7	30,63	24,78	...		
KB11010	27,11	18,17	22,83	27,8	33,32	25	28,3	...		
KB11011	23,87	24,43	37,5	38,83	25,98	31,34	24,12	...		
KB11012	28,53	34,37	29,77	27,6	45,29	44,25	35,36	...		
KB11013	46,17	31,45	32,09	41,45	43,69	43,13	46,17	...		
KB11014	23,28	27,8	31,94	31,91	32,26	42,78	27,75	...		
KB11016	49,3	39,46	51,28	58,34	40,53	46,97	49,3	...		
KB11017	50,37	42,31	48,96	54,62	45,23	54,16	50,37	...		
KB11018	39,62	37,88	28,88	32,06	49,65	54,72	45,28	...		

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

KB11019	20,12	11,75	29,95	34,48	17,03	21,47	20,12	...	0	
KB11020	18,89	14,7	25,04	27,91	18,6	35	18,89	...	37,87	0

By determining the shortest distance between two clusters to create a new cluster, KB11006 KB11015, based on Table 4, a new cluster has been obtained. In the following cluster search procedure, the least value from table matrix 3 is sought after. The subsequent minimum value, associated with the second cluster (KB11002 KB11003), is 9.49. Implementing step 3 and going back to step 2 will result in the product code grouping based on the desired cluster, which can then be used to search for cluster values associated with new clusters. Using the software rapid miner, you may view the results in further detail.

3.2 Method Testing for AHC

Through the use of the rapid miner program, the AHC method is tested to achieve the final findings (grouping) rapidly and to prevent mistakes or oversights in the manual grouping process. The rapid miner cluster formation procedure is as follows:

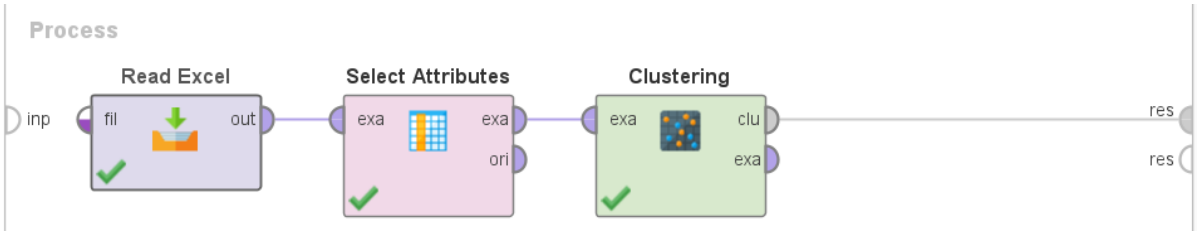


Figure 3. File Input Dataset Process

The procedures for utilizing RapidMiner are shown in Figure 3. The attribute select tool is used to choose the attributes to be utilized in the file once the data file to be processed has been entered (input). If the initial data criteria have been established earlier, the use of the chosen attribute is not particularly significant. However, the chosen property can be utilized to streamline the procedure if the file contains a lot of extraneous data. This study employs the single-linked technique and an agglomerative clustering tool to determine the distance between the nearest data, which is determined by computing the hierarchical distance between two alternatives. The dendrogram that results from running the rapid miner procedure is shown below.

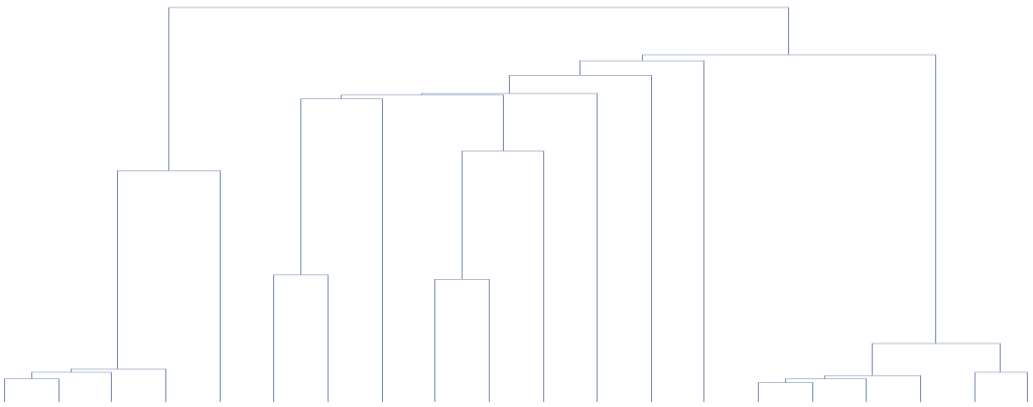


Figure 4. AHC Dendrogram

Figure 4's AHC dendrogram results demonstrate that the formation of a hierarchy is based on the proximity or distance between alternatives so that new clusters can be formed. Because this study will form three clusters, it is necessary to add flatten clustering to the input to determine the number of clusters to be formed, as well as the input process after adding flatten clustering tools.

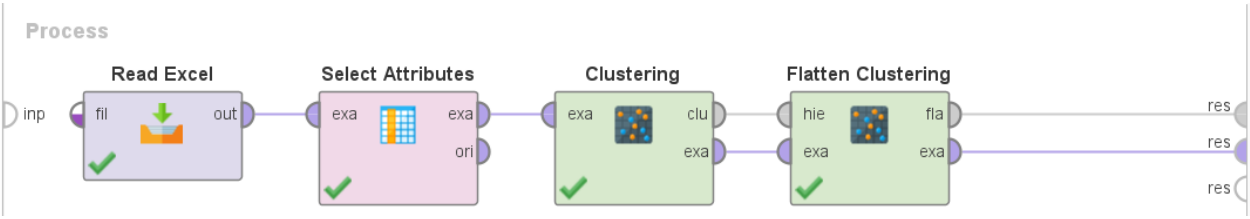
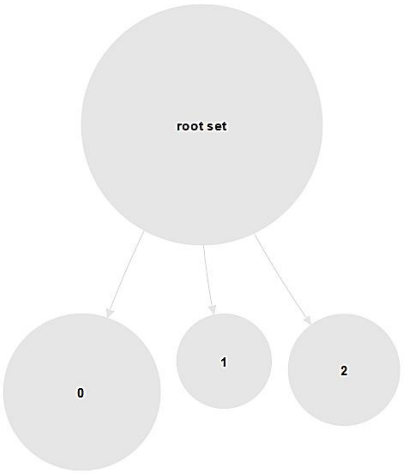


Figure 5. File Input Dataset Process

The input procedure is shown in full in Figure 5, starting with the entry of data from an Excel file, followed by the selection of the attributes to be used, the application of the method (agglomerative hierarchical clustering), the

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

selection of the single linked clustering mode, and the entry of flatten clustering to ascertain the number of clusters formed (3 clusters). The rootset trees that result from the use of the rapidminer application are shown below.



Gambar 6. Rootset tree

Figure 6 shows that cluster 0 has the greatest number of items, whereas clusters 1 and 2 have nearly the same amount of products. The following picture provides information on the number of products in each cluster.

Cluster Model

```
Cluster 0: 9 items
Cluster 1: 5 items
Cluster 2: 6 items
Total number of items: 20
```

Figure 7. Data Cluster Model

Figure 7 shows that there are 9 alternatives organized into cluster 0 (alternatives), 5 alternatives in cluster 1, and 6 alternatives in cluster 2. The following table contains grouping information:

Table 5. Clusters for each alternative

Number	Item Code	Selling Price				Cluster
		September	October	November	December	
1	KB11001	57	43	21	51	cluster_1
2	KB11002	18	15	23	49	cluster_0
3	KB11003	13	13	28	55	cluster_0
4	KB11004	37	25	10	30	cluster_2
5	KB11005	55	43	21	52	cluster_1
6	KB11006	14	37	44	46	cluster_0
7	KB11007	32	25	12	30	cluster_2
8	KB11008	59	42	22	53	cluster_1
9	KB11009	32	25	10	30	cluster_2
10	KB11010	12	35	16	43	cluster_0
11	KB11011	36	25	11	31	cluster_2
12	KB11012	8	34	18	69	cluster_0
13	KB11013	38	25	10	30	cluster_2
14	KB11014	29	38	32	66	cluster_0
15	KB11015	13	32	41	53	cluster_0
16	KB11016	56	46	23	34	cluster_1
17	KB11017	58	43	21	52	cluster_1
18	KB11018	35	25	12	67	cluster_0
19	KB11019	38	26	11	30	cluster_2
20	KB11001	30	28	40	44	cluster_0

The alternatives are sorted into each cluster according to the data points that have the same matrix or the same distance, as can be seen in Table 5. Cluster 1 has options 1, 5, 8, 16, and 17. Cluster 0 includes options 2, 3, 6, 10, 12,

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

14, 15, 18, and 20, whereas Cluster 2 includes options 4, 7, 9, 11, and 13. As observed in the next figure, scatter clustering is as follows.

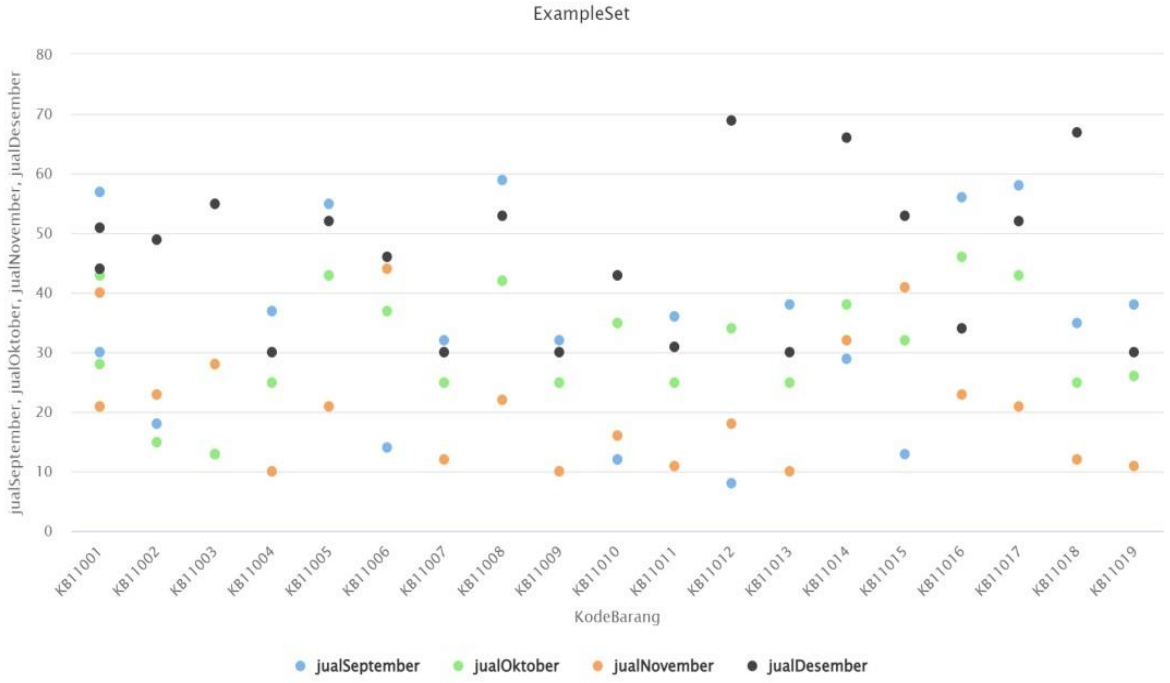


Figure 8. Scatter Clustering Single Link

According to Figure 8, the blue alternative represents sales in September, the green alternative represents sales in October, the orange alternative represents sales in November, and the black alternative represents sales in December. Figure 9 displays the range of each criterion to the number of each choice.

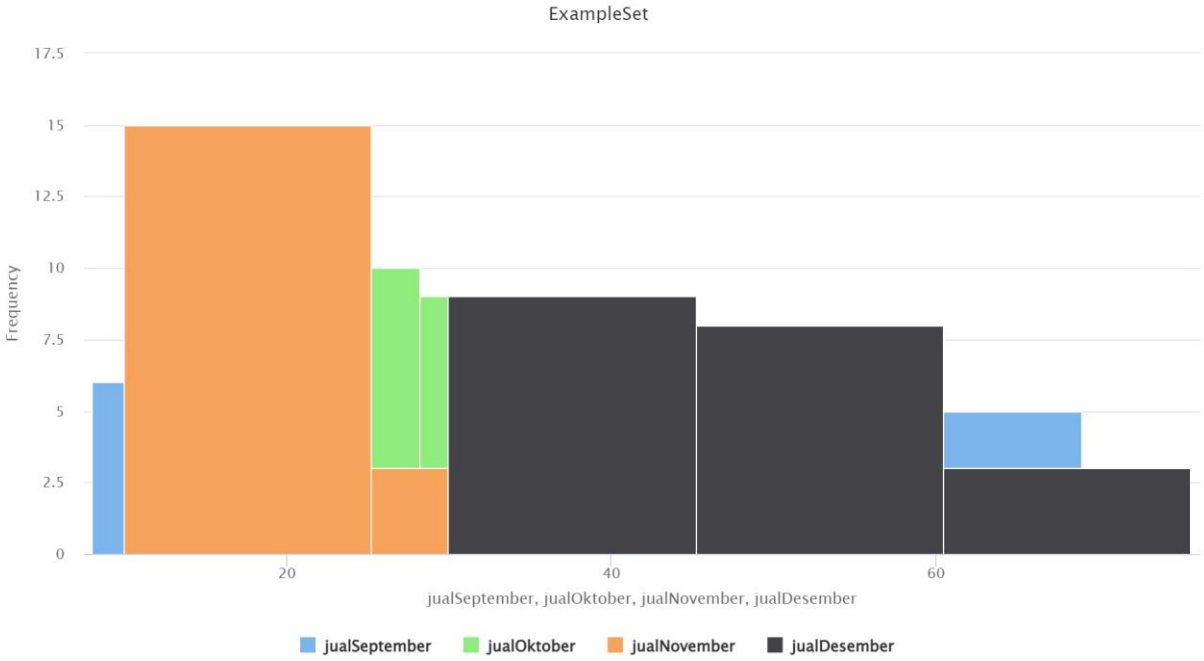


Figure 9. Histogram Clustering

Figure 9 shows that the range of September sales was 8 to 23.25 with six products, 23.25 to 38.5 with nine products, and 53.75 to 69 with five products. Sales in October ranged from 13 to 28.25 for ten products, 28.25 to 43.5 for nine products, and 43.5 to 58.75 for just one. Sales for November ranged from 10 to 25.25 for as many as 15 items, 25.25-40.5 for as many as 3 items, and 40.5-55.75 for as many as 2 items. Sales in December were from $30 to $45.25, $45.25-26, and $50.5-$60.75 with just 3 products, respectively. Figure 10 below displays each cluster's proportion.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

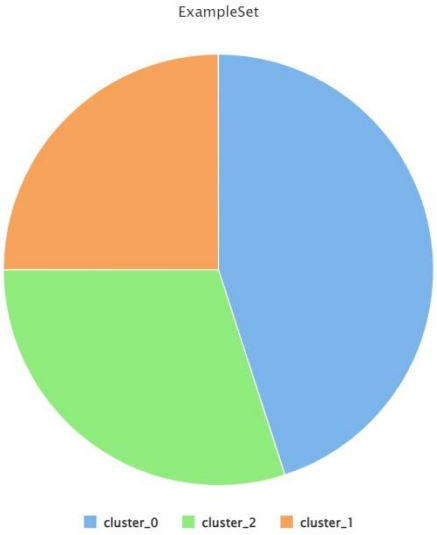


Figure 10. Cluster Visualization

According to Figure 10, the blue portion, or cluster 0, which has a percentage of 45% and a total of 9 products, is the largest portion. The orange portion, or cluster 1, has a large placement of 5 products with a percentage of 0.25%, and the green portion, or cluster 2, has 6 products with a percentage of 0.30%.

4. CONCLUSION

According to this study, options with the same distance or matrix will be sorted into certain clusters. In the application of the AHC approach, three clusters must be produced. Cluster 0, Cluster 1, and Cluster 2 have all been discovered as distinct clusters, each with its alternate groups. Additionally, there are sales statistics for several products each month that are classified according to different price points. Each cluster has a unique proportion and a varied number of goods. Cluster 0 has the most products and the greatest percentage, at 45%, followed by Cluster 2 and Cluster 1. Cluster 1 has the fewest products and the lowest percentage, at 0.30% and 0.25%, respectively.

REFERENCES

[1] G. Gunadi and D. I. Sensuse, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth) :," Telematika, vol. 4, no. 1, pp. 118–132, 2012.
[2] B. D. Mudzakkir, "Pengelompokan Data Penjualan Produk Pada Pt Advanta Seeds Indonesia Menggunakan Metode K-Means," J. Mhs. Tek. Inform., vol. 2, no. 2, pp. 34–40, 2018.
[3] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," J. Nas. Teknol. Dan Sist. Inf., vol. 5, no. 1, pp. 17–24, 2019.
[4] S. Al Syahdan and A. Sindar, "Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota," J. Nas. Komputasi dan Teknol. Inf., vol. 1, no. 2, 2018, doi: 10.32672/jnkti.v1i2.771.
[5] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," J. Teknol. dan Sist. Inf., vol. 2, no. 2, p. 100, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI
[6] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," J. Media Inform. Budidarma, vol. 4, no. 1, pp. 51–58, 2020.
[7] H. Al Rasyid, B. F. K. Soebari, and D. S. Y. Kartika, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PENJUALAN PRODUK PADA ONLINE SHOP TOKO GIZI," in Prosiding Seminar Nasional Teknologi dan Sistem Informasi, 2022, vol. 2, no. 1, pp. 242–248.
[8] M. Dahria, R. Gunawan, and Z. Lubis, "Implementasi K-Means Untuk Pengelompokan Produk Terbaik PT. Koko Pelli," in Seminar Nasional Sains dan Teknologi Informasi (SENSASI), 2019, vol. 2, no. 1.
[9] B. H. T. S. As and L. Zahrotun, "Penerapan Penerapan Data Mining dalam Mengelompokkan Data Riwayat Akademik Sebelum Kuliah dan Data Kelulusan Mahasiswa menggunakan Metode Agglomerative Hierarchical Clustering (AHC)," J. Teknol. Informasi, Komputer, dan Apl., vol. 3, no. 1, pp. 62–71, 2021.
[10] A. Z. Siregar, "Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru," Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen), vol. 2, no. 3, pp. 133–137, 2021, [Online]. Available: https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/73
[11] S. S. S, A. T. Purba, V. Marudut, M. Siregar, T. Komputer, and P. B. Indonesia, "SISTEM PENDUKUNG KEPUTUSAN KELAYAKAN PEMBERIAN PINJAMAN," vol. 3, pp. 25–30, 2020, doi: 10.37600/tekinkom.v3i1.131.
[12] B. S. Pranata and D. P. Utomo, "Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 285−294
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3569

Motor (Study Kasus Bengkel Sinar Service)," Bull. Inf. Technol., vol. 1, no. 2, pp. 83–91, 2020.
[13] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," J. Informatics Manag. Inf. Technol., vol. 1, no. 2, pp. 79–84, 2021, [Online]. Available: https://hostjournals.com/
[14] A. Wanto et al., Data Mining: Algoritma dan Implementasi. Yayasan kita menulis, 2020.
[15] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," J. Media Inform. Budidarma, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.
[16] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," J. Tekno Insentif, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
[17] H. Maulidiya and A. Jananto, "Asosiasi Data Mining Menggunakan Algoritma Apriori dan FP-Growth sebagai Dasar Pertimbangan Penentuan Paket Sembako," Proceeding SENDIU 2020, vol. 6, pp. 36–42, 2020.
[18] A. S. L. T. T. H. Hafizah, "Data Mining Estimasi Biaya Produksi Ikan Kembung Rebus Dengan Regresi Linier Berganda," J. Sist. Inf. Triguna Dharma (JURSI TGD), no. Vol 1, No 6 (2022): EDISI NOVEMBER 2022, pp. 888–897, 2022, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsi/article/view/5732/1938
[19] Y. L. Nainel, E. Buulolo, and I. Lubis, "Penerapan Data Mining Untuk Estimasi Penjualan Obat Berdasarkan Pengaruh Brand Image Dengan Algoritma Expectation Maximization (Studi Kasus: PT. Pyridam Farma Tbk)," JURIKOM (Jurnal Ris. Komputer), vol. 7, no. 2, p. 214, 2020, doi: 10.30865/jurikom.v7i2.2097.
[20] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," TIN Terap. Inform. Nusant., vol. 2, no. 4, pp. 191–197, 2021.
[21] B. Harli Trimulya Suandi As and L. Zahrotun, "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN DATA RIWAYAT AKADEMIK SEBELUM KULIAH DAN DATA KELULUSAN MAHASISWA MENGGUNAKAN METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING (Implementation Of Data Mining In Grouping Academic History Data Before Students And Stud," J. Teknol. Informasi, Komput. dan Apl., vol. 3, no. 1, pp. 62–71, 2021, [Online]. Available: http://jtika.if.unram.ac.id/index.php/JTIKA/
[22] R. A. Setyawan and R. M. Fadilla, "Klasterisasi media pembelajaran daring di era pandemi COVID-19 menggunakan metode Agglomerative," Inf. Interaktif, vol. 5, no. 3, 2020, [Online]. Available: http://www.e-journal.janabadra.ac.id/index.php/informasiinteraktif/article/view/1305%0Ahttps://www.e-journal.janabadra.ac.id/index.php/informasiinteraktif/article/download/1305/890
[23] Marjiyono, "Penerapan Algoritma Ahc Algorithm Dalam Aplikasi Ppembagian Kelas Siswa Baru," Semin. Nas. Teknol. Inf. dan Multimed. 2015, pp. 6–8, 2015.
[24] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," Atmos. Pollut. Res., vol. 11, no. 1, pp. 40–56, 2020.
[25] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-IID data," in 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1–9.
[26] K. Zeng, M. Ning, Y. Wang, and Y. Guo, "Hierarchical clustering with hard-batch triplet loss for person re-identification," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13657–13665.
[27] N. K. Zuhal, "Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering," Pros. Semin. Nas. Teknol. dan Sains, vol. 1, pp. 200–205, 2022, [Online]. Available: https://jurnal.dharmawangsa.ac.id/index.php/djtechno/article/view/966/867