

# Implementation of The Agglomerative Hierarchical Clustering Method In Ordering Hijab Products

Tiwy Ardyanti<sup>1)\*</sup>, Mhd Furqan<sup>2)</sup>

<sup>1,2)</sup> Department of Computer Science, Faculty Science and Technology, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

<sup>1)</sup>[tiwyardiyanti29@gmail.com](mailto:tiwyardiyanti29@gmail.com), <sup>2)</sup>[mfurqan@uinsu.ac.id](mailto:mfurqan@uinsu.ac.id)

**Submitted** : Sept 3, 2024 | **Accepted** : Oct 15, 2024 | **Published** : Oct, 24, 2024

**Abstract:** The ever-evolving internet technology has an impact on various sectors, including the hijab business, where the demand for hijab products is increasing through online transactions. This research was conducted at the Kinan Hijab Store in Kota Pinang, North Sumatra, with the aim of optimizing the management of hijab product stock. The problem faced is the imbalance in the stock of hijab products, where some hijab products have excess stock that are less in demand while popular hijab products often experience a shortage of stock. To solve this problem, the Agglomerative Hierarchical Clustering method is used to group hijab products based on sales data, product type, and price. This study uses hijab sales data from May to July 2024. After the clustering process, hijab products are grouped into two categories: "Popular" and "Less Desirable". The "Popular" category includes 190 products, while the "Less Desirable" category includes 983 products. Product stock in the "Popular" category will be increased by 50% of the average sales, while stock in the "Less Desirable" category will be reduced by 25%. The effectiveness of the Agglomerative Hierarchical Clustering (AHC) method in stock planning and management by showing that it improved the inventory allocation based on customer demand patterns. The clustering method categorized hijabs into two main groups: "Popular" and "Less Preferred", based on key sales metrics such as quantity sold, price, and total sales. The implementation of the stock plan is carried out based on the sales pattern of each hijab category. Overall, the application of this method not only helps stores in understanding customer purchasing patterns but also optimizes product availability, which can ultimately increase customer satisfaction.

**Keywords:** Agglomerative Hierarchical; Clustering; Data Mining; Hijab; Sales Stock Management

## INTRODUCTION

Kinan Hijab Store, located on Kota Pinang Street in the Labuhan Batu Selatan District of North Sumatra, is a company engaged in the hijab product business. As a professional company in its field, it strives to provide excellent service that satisfies its customers with high-quality and attractive hijab products. (Sasmita Susanto, Karisma, and Isnaeni 2019). In recent years, hijab wearing has experienced a significant surge, partly due to the dismantling of discriminatory beliefs. Consequently, the community of hijab-wearing women has expanded rapidly, embracing contemporary fashion trends and allowing women to express their personal style while adhering to Islamic principles. Despite the evolving fashion landscape, Muslim women remain committed to dressing in accordance with Islamic guidelines. The demand for hijab products has become a fundamental aspect of the modern Muslim woman's lifestyle (Sofiyah & Zafi, 2020). A common problem is that managing the diversity of hijab products can be a complicated task without an efficient grouping system. Grouping the types of hijab products based on quality and information in the presentation of prices and stocks of each product that are less effective, various problems can arise that have an impact on operations, lack of understanding of customer needs and customer satisfaction. Without proper grouping, businesses may face the problem of overstocking products that are not in demand and understocking products that are in high demand. So it is highly recommended for more accurate performance in grouping hijab product orders. By using the Agglomerative Hierarchical Clustering method, it is possible to group variants of hijab product types based on hijab product order data (Sasmita Susanto et al., 2019).

This method was selected for several reasons. Firstly, Agglomerative Hierarchical Clustering Method does not require a pre-specified number of clusters, unlike algorithms such as K-Means which necessitate a

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

predetermined cluster count. This offers greater flexibility to researchers, particularly when cluster patterns are unknown a priori. By visualizing the dendrogram, the optimal number of clusters can be identified visually. Secondly, this technique is more suitable for small to medium-sized datasets, such as hijab sales data, as it enables hierarchical clustering, starting from individual data units and gradually forming larger clusters. Furthermore, the clustering process can be implemented using various distance metrics like Manhattan Distance, allowing for better adaptation to data characteristics. In the context of hijab inventory management, where sales patterns can be diverse and non-spherical, K-Means becomes less suitable. Additionally, retailers may not know the exact number of clusters beforehand, making AHC a more flexible option.

We have chosen this method for our research because it is more flexible than other methods like K-Means. This method lets us find the best number of clusters without having to guess it beforehand. The dendrogram visualization also helps us see the clustering results more clearly. In addition, studies by (Justitia et al., 2021). and (Widyawati et al., 2020) show that this method gives us more accurate results. The high Silhouette Coefficient scores mean that the groups we found are really good at showing the patterns in the data. This method is also great for handling complex data and can use different ways to measure distance between data points, making it perfect for managing hijab inventory. Seeing this background, the author conducted this research with the aim of creating an Agglomerative Hierarchical Clustering system for clustering data on the type of hijab product order and to help reduce the stock of products that are less in demand and increase the stock of hijab products that are in demand by identifying customer purchase patterns by grouping the data of hijab product types of businesses can better understand customer needs and optimize strategic inventory marketing.

### LITERATURE REVIEW

According to research conducted by (Nellie et al., 2023) After the weight values have been determined, the grouping calculations will be performed using agglomerative hierarchical clustering. Previously, the Manhattan Distance method will be used to calculate the distance. After that, the distance that is closest can be determined. The data will be clustered according to the shortest distance once the distance calculation is complete. Following that, the system will display the grouping as a dendrogram. The data used was updated as of the date of scraping, which is November 25, 2022, and contains a total of 2467 data. The Agglomerative Hierarchical Clustering method yielded the best silhouette coefficient value, 0.5025559374455285, forming 20 clusters.

According to research conducted by (Dhanita & Wasito, 2024). Hierarchical clustering is a method in data science for grouping similar data points into clusters, ultimately forming a hierarchical structure. There are two types of hierarchical clustering: agglomerative clustering and divisive hierarchical clustering. For this study, agglomerative clustering will be used. Agglomerative clustering employs a bottom-up approach. This method starts with each data point as an individual cluster. In each iteration, the algorithm combines the two most similar clusters into a new combined cluster. To select the two clusters to be merged, the Euclidean distance formula will be used. This process is repeated until all data points are contained within one cluster or a predefined stopping criterion is met. The result is a dendrogram with a tree-like structure.

Clustering is an existing approach to data mining. Cluster analysis identifies objects that are related to each other but not related to other groups. The goal of cluster analysis is to minimize the distance (Pane et al., 2024). Data objects can be combined into one or more groups so that objects in one group will have a high degree of similarity with each other (Computer, 2024). Data preprocessing in data mining is a critical process that involves cleaning, transforming, and integrating data to prepare the data for analysis. This study uses the Standard Scaler preprocessing method, which standardizes features by removing the average and scaling the variant units (Ulvi & Ikhsan, 2024). The process of calculating pre-processing on data using the standard scaler formula is as follows:

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

Information :

- $Z_i$  : Standard score or z-score of the  $i$ th data
- $X_i$  : The original data value of the  $i$ th data
- $\mu$  : Average of all data
- $\sigma$  : Standard deviation from the overall data

Hierarchical clustering method that unites  $N$  clusters until they form a single cluster (Abdurrahman, 2019). This method begins by putting each data as a separate cluster (atomic cluster) after which uniting several clusters into an even larger cluster until finally all the data becomes one in a single cluster (Wijaya & Swanjaya, 2021). The data set grouping strategy uses Agglomerative hierarchical clustering, which is when the grouping process uses this technique using the Manhattan Distance and Single Linkage methods (Justitia et al., 2021b).

$$d(C_{(ij)}, C_k) = \min \{d(C_i, C_k), d(C_j, C_k)\} \quad (2)$$

Information :

- $d(C_{(ij)}, C_k)$  : Distance between combined clusters  $C_{(ij)}$  and  $C_k$  cluster
- Min : Minimum value of two distances

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The Manhattan distance formula to calculate distance uses the following equation:

$$D_{man}(x, y) = \sum_{i=1}^n |x_{1i} - x_{2i}| \tag{3}$$

Information :

- D : Distance between x and y
- x : Cluster Data Center
- y : Attribute Center Data

Silhouette coefficient is a method to measure the validation of either a single cluster, a single cluster or an entire cluster. The value of the *i*th data silhouette coefficient is based on two components, namely *a<sub>i</sub>* and *b<sub>i</sub>*.

**METHOD**

This research is a quantitative research with an exploratory approach. This research was conducted at the Kinan Hijab Store in Kota Pinang, North Sumatra, with the aim of optimizing the management of hijab product stock. The total dataset is 1173 data. This study uses hijab sales data from May to July 2024. After the clustering process, hijab products are grouped into two categories: "Popular" and "Less Desirable". The stages of this research can be discussed in Figure 1 which consists of literature study, data collection, needs analysis, data preprocessing, design, algorithm application, results and conclusions.



Figure 1. Stages of research

Figure 1 shows the framework of the research that will be carried out starting from the literature study stage to the application stage.

Literature Studies are carried out to collect knowledge from various literary sources in the form of books, journals and other scientific works related to the topics raised by the authors.

The data collection technique carried out in this study is interviews. The interview was conducted by the researcher with Mrs. Handayani as the owner of the shop. The interview conducted by the author to get information about the hijab shop and to get order data on hijab products used to carry out grouping in the research.

After data collection, the data preprocessing stage is carried out which consists of several stages, namely Data Transformation and Data Selection. The data to be selected is "Qty, Price, Netto, Pay and Amount" on the hijab sales dataset. The technique carried out is to standardize data using a standard scaler.

Agglomerative Hierarchical Clustering modeling on hijab product order data to group product types into groups based on variables of hijab model types, and prices used in hijab product orders. Modeling the hierarchical clustering method can simplify the characteristics of the most popular hijab products that consumers often buy, to reduce the excess stock of products that are not in demand and the shortage of stock that is in high demand. The parameter used is single linkage and to calculate the distance using Manhattan distance.

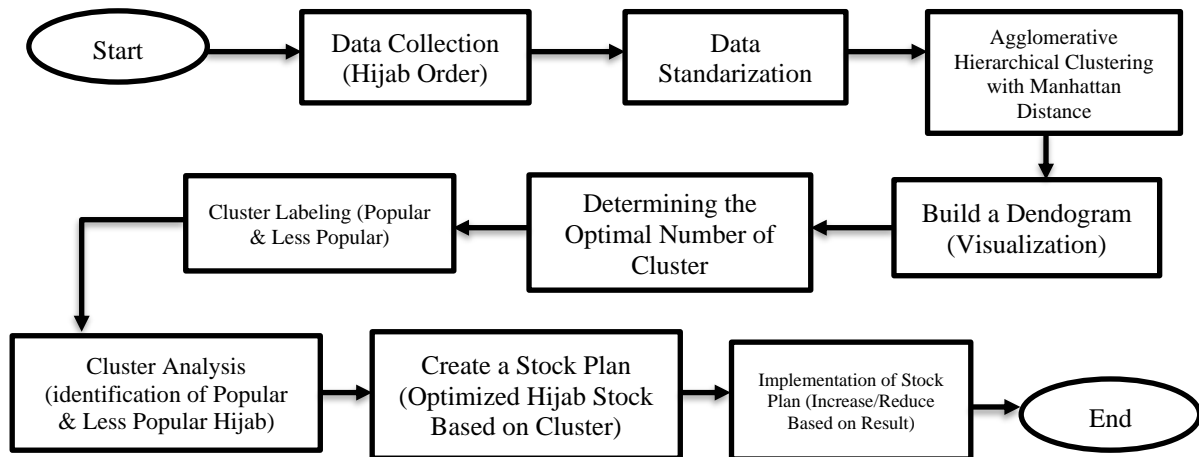


Figure 2. Research Flowchart

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Within the framework of this study, the researcher will conduct tests on the ordering of hijab products used in the clustering model.

In this study, the application began to take data from the hijab shop in Pinang City, Labuhan Batu Selatan. The results of the Agglomerative Hierarchical Clustering model to find out the most popular or in-demand or under-in-demand hijab products. This analysis helps in understanding the needs of customers who contribute to the business decision of hijab products to stock hijab products that are in demand or are not in demand. This research is used as a basis for an effective and efficient marketing inventory strategy by preventing shortages of hijab products in demand and excess stocks of hijab products that are not in demand.

**RESULT**

Hijab sales data is stored in a file with .csv extension, The data taken amounted to 1173 hijab sales data, where there are several types of hijab sold from May to July in 2024. The data will later be given popular and less popular labels which will later be determined through Qty (total types of hijab sold) and using the Agglomerative Hierarchical clustering algorithm to group hijab sales data and increase or decrease stock based on the cluster results obtained.

Table 1 Hijab sales dataset

No	Date	Customer	Sales Number	Qty	Total Items	Barcodes	..	Total
1	1-May-2024	Retail	J110642	3	1	301206401	..	72000
2	1-May-2024	Retail	J110643	1	1	301205771	..	250000
3	1-May-2024	Retail	J110644	1	5	301206432	..	55000
...	....	...	...	...	...	...	..	...
1173	31-July-2024	Retail	J112145	2	1	301205378	...	14000

The next step is to drop data that is not used in the study. Examples of labels used are "Total Item", "Netto", "Pay", "Qty", "Price", "Discount", "Amount" so that it is obtained as shown in the following table.

Table 2 Data used before pre-processing

No.	Qty	Price	Net	Pay	Sum
1	3	24000	72000	72000	72000
2	1	250000	250000	250000	250000
3	1	55000	133000	133000	55000
4	1	7000	133000	133000	7000
...	...	...	...	...	...
1173	1	14000	32000	32000	14000

The process of calculating pre-processing on data such as tables uses a standard scaler formula like the following equation:

$$Z_i = \frac{x_i - \mu}{\sigma}$$

Where initially is to calculate the average of each column with an example of the column "Qty", namely:

$$\mu = \frac{(3 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + \dots + 1)}{1173}$$

$$\mu = 1.378516624040921$$

Next is to calculate the variance (S2):

$$S^2 = \frac{(3 - 1.378516624040921)^2 + (1 - 1.378516624040921)^2 + \dots + (1 - 1.378516624040921)^2}{1173}$$

$$S^2 = 2.281909$$

The last step is to calculate the Standard Scaler:

$$Z_i = \frac{3 - 1.378516624040921}{2.281909} = 0.710582$$

Next, calculate the average of each column with the example of the "Price" column, namely:

$$\mu = \frac{24000 + 250000 + 55000 + 7000 + 20000 + 45000 + 55000 + 17000 \dots + 14000}{1173}$$

$$\mu = 55664.96164$$

\*name of corresponding author



Next is to calculate the variance ( $S^2$ ):

$$S^2 = \frac{(24000 - 55664.96164) + (55000 - 55664.96164) + \dots + (14000 - 55664.96164)}{1173}$$

$$S^2 = 108412.188$$

The last step is to calculate the Standard Scaler

$$x_i = 24000$$

$$Z_i = \frac{24000 - 55664.96164}{108412.188} = -0.292204$$

Next, calculate the average of each *column* with the example *column* "Netto", namely:

$$\mu = \frac{72000 + 250000 + 133000 + 133000 + 20000 + 45000 + 49500 + \dots + 32000}{1173}$$

$$\mu = 176477.2805$$

Next is to calculate the variance ( $S^2$ ):

$$S^2 = \frac{(72000 - 176477.2805) + (133000 - 176477.2805) + \dots + (32000 - 176477.2805)}{1173}$$

$$S^2 = 318575.1454$$

The last step is to calculate the Standard Scaler

$$x_i = 72000$$

$$Z_i = \frac{72000 - 176477.2805}{318575.1454} = -0.328092$$

Next, calculate the average of each *column* with an example of the "Pay" *column*, namely:

$$\mu = \frac{72000 + 250000 + 133000 + 133000 + 133000 + 133000 + 49500 + \dots + 32000}{1173}$$

$$\mu = 176571.9096$$

Next is to calculate the variance ( $S^2$ ):

$$S^2 = \frac{(72000 - 176571.9096) + (250000 - 176571.9096) + \dots + (133000 - 176571.9096)}{1173}$$

$$S^2 = 318417.7221$$

The last step is to calculate the Standard Scaler

$$x_i = 72000$$

$$Z_i = \frac{72000 - 176571.9096}{318417.7221} = -0.328411$$

Next, calculate the average of each *column* with the example *column* "Number", namely:

$$\mu = \frac{72000 + 250000 + 55000 + 7000 + 20000 + 45000 + 49500 + 17000 + 5000 \dots + 14000}{1173}$$

$$\mu = 60435.35379$$

Next is to calculate the variance ( $S^2$ ):

$$S^2 = \frac{(72000 - 60435.35379) + (250000 - 60435.35379) + \dots + (55000 - 60435.35379)}{1173}$$

$$S^2 = 108559.6714$$

The last step is to calculate the Standard Scaler

$$x_i = 72000$$

$$Z_i = \frac{72000 - 60435.35379}{108559.6714} = 0.106528$$

The Pre-Processing process is carried out to prepare raw data so that it is ready to be used in the analysis or modeling process. In this study, the data transformation process uses a Standard Scaler which is used on the labels "Qty", "Price", "Netto", "Pay" and "Amount".

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Table 3 Data after pre-processing

No.	Qty	Price	Net	Pay	Sum
1	0.710582	-0.292204	-0.328092	-0.328411	0.106528
2	-0.165877	1.793322	0.230885	0.230603	1.746179
3	-0.165877	-0.006136	-0.136532	-0.136839	-0.050068
...	...	...	...	...	...
1173	-0.165877	-0.384484	-0.453704	-0.454032	-0.427740

Before entering the process of the hierarchical Agglomerative clustering algorithm, which is to find the best k value using a silhouette score. Silhouette Score is one of the metrics used in clustering to assess how well the objects in the cluster are separated. The average Silhouette Score for all data points can be used to evaluate the overall quality of the cluster. By trying different k values, we can choose the k value that gives the highest average silhouette score, which will be an indication that the cluster with k has good separation between the clusters and strong cohesion within the cluster. An example is in the image below.

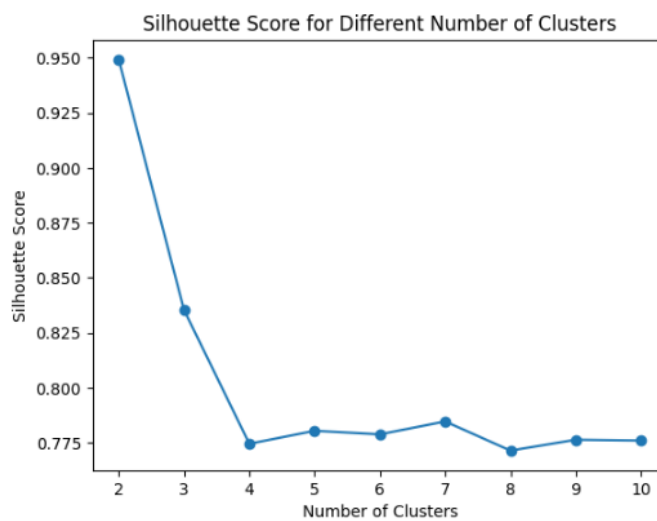


Figure 3. Silhouette Score To Find The Best K

In Figure 3, it is found that the best cluster is at 2 with a score close to 0.950 because a higher Silhouette Score indicates better clustering, where the data points in the cluster have higher similarities (internal consistency) and are more different from the data points in other clusters (external proximity).

Furthermore, a dendrogram is constructed to describe hierarchical relationships between objects or data in clustering analysis, especially in the context of Hierarchical Clustering. Dendrograms show how clusters are formed and combined at each step in the clustering process.

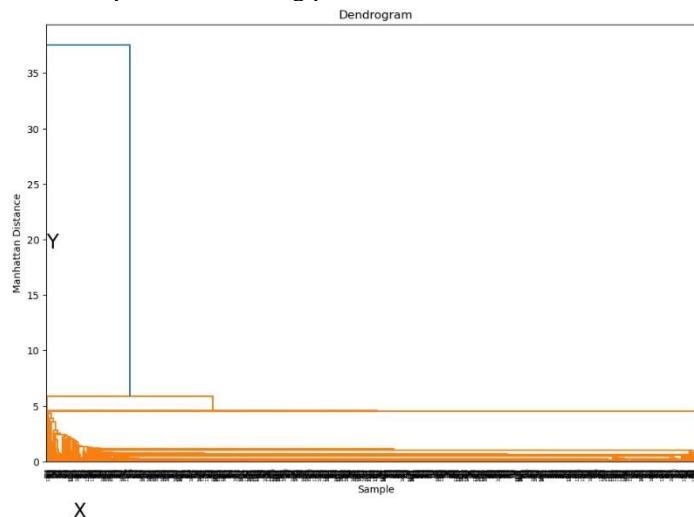


Figure 4. Dendrogram

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Dendrograms show the hierarchical structure of data grouping, with the X-axis indicating each sample, and the Y-axis indicating the distance or difference between clusters. In this dendrogram, we can see one very long vertical line on the left, at a distance of about 35, which indicates that two large clusters with significant differences are combined at that stage of grouping. At the bottom of the dendrogram, we can see a shorter vertical line that joins clusters with smaller distances, indicating that the data are more similar to each other.

The agglomerative process with single linked parameters and the process of determining the category of hijab sold is categorized as popular or less in demand. The process of determining the category is focused on Qty (Number of hijabs), where if  $> -0.165538283705652$  it is categorized as a "Popular" hijab. On the other hand, if  $<$  or equal to  $-0.165538283705652$ , then the hijab is categorized as "Less Desirable".

Table 4 Clustering Results

No.	Qty	Price	Net	Pay	Sum	Cluster	Category
1	0.710582	-0.292204	-0.328092	-0.328411	0.106528	1	Popular
2	-0.165877	1.793322	0.230885	0.230603	1.746179	1	Less in Demand
3	-0.165877	-0.006136	-0.136532	-0.136839	-0.050068	1	Less in Demand
...	...	...	...	...	...	...	...
1173	-0.165877	-0.384484	-0.453704	-0.454032	-0.427740	1	Less in Demand

Using the equation of Manhattan distance.

$$D_{man}(P1, P2) = |0.710582 - (-0.165877)| + |-0.292204 - 1.793322| + |-0.328092 - 0.230885| + |-0.328411 - 0.230603| + |0.106528 - 1.746179|$$

$$= 0.876459 + 2.085526 + 0.558977 + 0.559014 + 1.639651$$

$$= 5.72$$

$$D_{man}(P1, P3) = |0.710582 - (-0.165877)| + |-0.292204 - (-0.006136)| + |-0.328092 - (-0.136532)| + |-0.328411 - (-0.136839)| + |0.106528 - (-0.050068)|$$

$$= 0.876459 + 0.286068 + 0.191560 + 0.191572 + 0.156596$$

$$= 1.70$$

$$D_{man}(P1, P4) = |0.710582 - (-0.165877)| + |-0.292204 - (-0.006136)| + |-0.328092 - (-0.136532)| + |-0.328411 - (-0.136839)| + |0.106528 - (-0.492221)|$$

$$= 0.876459 + 0.156876 + 0.191560 + 0.191572 + 0.598749$$

$$= 2.02$$

So we get a table of Manhattan distance results as follows:

Table 5 Manhattan Matrix distances

	P1	P2	P3	P4	P5
P1	0	5.72	1.70	2.02	1.75
P2	5.72	0	4.33	5.22	5.33
P3	1.70	4.33	0	0.89	1.0
P4	2.02	5.22	0.89	0	0.59
P5	1.75	5.33	1.0	0.59	0

The next step is to find the smallest distance in the distance matrix. The smallest distance is 0.59. Then combine P4 and P5 into a cluster (C1) and recalculate the distance between the new cluster (P4-P5) and all other points or clusters.

$$\text{Distance between } C1 \text{ and } P1 : D(C1, P1) = \frac{D(P4,P1)+D(P5,P1)}{2} = \frac{2.02+1.75}{2} = 1.88$$

$$\text{Distance between } C1 \text{ and } P2 : D(C1, P2) = \frac{D(P4,P2)+D(P5,P2)}{2} = \frac{5.22+5.33}{2} = 5.27$$

$$\text{Distance between } C1 \text{ and } P3 : D(C1, P3) = \frac{D(P4,P3)+D(P5,P3)}{2} = \frac{0.89+1.0}{2} = 0.94$$

Table 6 New distance matrix

	P1	P2	P3	C1 (P4-P5)
P1	0	5.72	1.70	1.88
P2	5.72	0	4.33	5.27
P3	1.70	4.33	0	0.94
C1 (P4-P5)	1.88	5.27	0.94	0

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The smallest distance is currently 0.94 between P3 and C1. Then combine P3 with C1 where the result is a new cluster C2 that combines P3, P4 and P5. The next step is to calculate the average distance.

Table 7 New spacing of cluster C2

	P1	P2	C2 (P3, P4, P5)
P1	0	5.72	1.823
P2	5.72	0	4.96
C2 (P3, P4, P5)	1.823	4.96	0

In table 7, the smallest distance between cluster 2 and the other points (P1 and P2) is 1.823 between C2 and P1. Next is to combine P1 with C2 which consists of P3, P4, P5. This new cluster is C3 (P1,P3,P4,P5). Calculating the new distance:

$$D(C3, P2) = \frac{D(P1, P2) + D(P3, P2) + D(P4, P2) + D(P5, P2)}{4} = \frac{5.72 + 4.33 + 5.22 + 5.33}{4} = \frac{20.6}{4} = 5.15$$

Table 8  
New distance C3

	P2	C3 (P1,P3,P4,P5)
P2	0	5.15
C3 (P1,P3,P4,P5)	5.15	0

In the last stage, since there are only two clusters left (P2 and C3), we immediately merge the two. The end result is a single cluster that includes all points (P1,P2,P3,P4,P5). Cluster result analysis includes an evaluation process resulting from the Agglomerative Hierarchical clustering algorithm. The purpose of this analysis is to understand the characteristics of the data, for example from the results of the "Popular" and "Less Popular" categories.

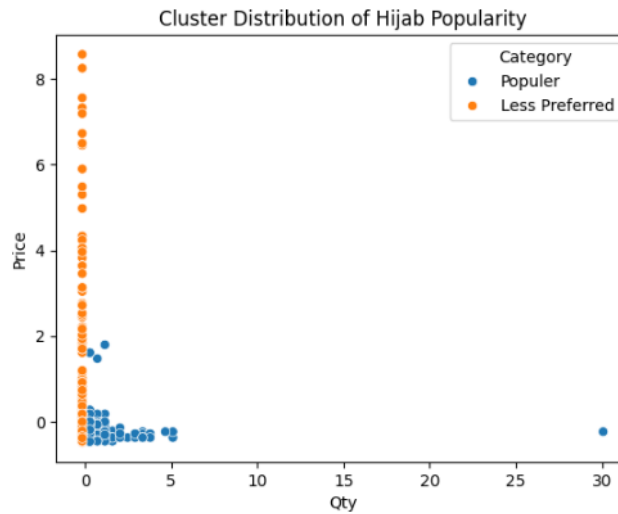


Figure 5. Hijab Distribution Plot By Qty And Price

Figure 5 shows a large number of hijabs in the "Less Desirable" category that have low prices (especially close to 0) and low quantities (below 5). Hijabs in the "Populer" category tend to have higher quantities, although most of them also have relatively low prices. "Populer" data tends to be grouped in low to medium quantities (between 0-5) and has a realistically low price. There is one data point that shows a very high quantity (more than 25), which falls into the "Populer" category.

Hijabs in the "Less Desirable" (orange) category are spread over a wider price range, but most remain in low quantities. There are some items with higher prices (above 8), but the quantity remains low. More popular hijabs tend to be available in larger quantities, although they are sold at relatively low prices, while less in-demand hijabs, although prices can vary, tend not to be sold in large quantities. The number of popular hijab products is 190 and the number of less popular hijab products is 983.

Product stock allocation is the process of determining the distribution of the number of goods or products where the purpose of this stock allocation is to ensure that products are available in the right place, in the right quantity, and at the right time to meet customer demand efficiently. The stock allocation of hijab products is obtained in the table below.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Table 9  
Product stock allocation

No	Qty	...	Total	Cluster	Category	Average Sales	Determining the Stock
1	0.710582	...	0.106528	1	Popular	0	0
2	0.165877	...	1.746179	1	Less in Demand	0	0
3	0.165877	...	0.050068	1	Less in Demand	0.600880	0.450660
...	...	...	...	...	...	...	...
1173	0.165877	...	0.427740	1	Less in Demand	-0.360189	-270142

From the existing program, it is found that for the "Popular" category, the stock is determined by multiplying the "Average sales" by 1.5 because it can increase the stock by 50%. And in the "Less Desirable" category, the stock is determined by multiplying the "Average sales" by 0.75 because it can reduce the stock by 25%.

Stock plan implementation is an operational process carried out to adjust the number of product stocks based on demand analysis, market trends, and business strategies. The following is the stock plan for the sale of hijab products obtained.

Table 10  
Implementation of product stock sales plan

No.	Qty	...	Cluster	Category	Average Sales	Determining the Stock	Stock Plan
1	0.710582	...	1	Popular	0	0	0
2	-0.165877	...	1	Less in Demand	0	0	0
3	-0.165877	...	1	Less in Demand	0.600880	0.450660	0.338
...	...	...	...	...	...	...	...
1173	-0.165877	...	1	Less in Demand	-0.360189	-270142	-0.2026

In the table above, if the product category is 'Popular', the stock will be increased by 50%. Conversely, if a product is categorized as 'Less Desirable', the stock will be reduced by 25%. For other categories, the stock will not be changed. For example, for a product with the Popular category, there is no change in stock because the initial stock value is zero. Meanwhile, products in the Less Desirable category experienced a 25% reduction in stock according to the logic applied to this function. These results are displayed in the table to provide visualization of the implementation of the stock allocation plan by product category.

### DISCUSSION

The research focused on the implementation of the Agglomerative Hierarchical Clustering (AHC) method in managing hijab product orders at Kinan Hijab Store in Kota Pinang. The primary objective was to analyze sales data and group hijab variants based on customer preferences to enhance inventory management. The study produced significant findings that highlight the effectiveness of AHC in optimizing stock allocation and improving business performance. The study collected a total of 1,173 sales transactions from May to July 2024. This dataset encompassed vital information such as the quantity sold, price, and total sales amount for each hijab product. Prior to applying the clustering algorithm, the data underwent preprocessing using the Standard Scaler technique. This normalization step ensured that all variables were standardized, allowing for a more accurate comparison during the clustering process. The AHC method was chosen due to its capability to form clusters without requiring the number of clusters to be predetermined. This flexibility is crucial in scenarios where the distribution of data is unknown. The algorithm begins with each data point as its own cluster and progressively merges them based on similarity until all data points are consolidated into a single hierarchical structure. The resulting dendrogram effectively visualizes the relationships among different hijab products, highlighting which items are more similar and allowing for informed decision-making regarding stock management. The clustering analysis resulted in two main categories of hijab products: "Popular" and "Less Preferred." Products classified as popular were identified based on higher sales volumes, while less preferred items showed lower sales figures. This classification facilitated a better understanding of customer purchasing behavior, enabling the store to manage its inventory more effectively. By identifying popular items, Kinan Hijab Store can ensure adequate stock levels for products in high demand while reducing the stock of less popular items, minimizing potential losses from unsold inventory.

The findings from the clustering analysis provided actionable insights for stock allocation. By understanding the purchasing patterns of customers, Kinan Hijab Store was able to avoid common inventory pitfalls, such as overstocking slow-moving items or understocking popular products. The research indicated that with improved stock management, the store could significantly enhance its operational efficiency, reduce carrying costs, and increase overall profitability. The analysis underscored the importance of aligning inventory levels with actual customer demand, ultimately leading to a more streamlined and responsive retail operation. In addition to improving stock management, the implementation of AHC had a substantial impact on the store's marketing strategies. Armed with accurate data on customer preferences, Kinan Hijab Store could tailor its promotional

\*name of corresponding author



efforts to focus on trending products that resonate with customers. This data-driven approach enables more effective marketing campaigns, enhances customer engagement, and strengthens brand loyalty. The research demonstrates that businesses leveraging data analytics can respond more adeptly to market dynamics, ensuring that their offerings align with consumer expectations.

### CONCLUSION

Based on the results of the study, hijab products are included in two categories: "Popular" and "Less Desirable". The clustering process is used on the sales data of "Kinan Hijab". This study uses the Agglomerative Hierarchical Clustering algorithm for data consisting of 1,173 sales records from May to July 2024. The amount, Netto, Qty, and Price used are processed and standardized before clustering is carried out. The classification results show two main groups. Hijab products in the "Popular" category have lower prices and a higher number of sales, although some of them have higher prices. On the other hand, hijab products in the "Less Desirable" category have a smaller number of sales and lower prices. About 190 hijab products fall into the "Popular" category, while most fall into the "Less Desirable" category. And the result of the calculation of the quality of the group with a silhouette score of 0.949 because when clustering is carried out with 2 clusters, it is possible that the data object is naturally divided into two very different groups. Thus, objects in one cluster are very close to each other and far away from objects in other clusters, resulting in a very high silhouette score. The recommendation from the results of this study is that stock allocation is more focused on products that have lower prices and higher quantities, because data trends show that products with these characteristics tend to be more popular among consumers.

### REFERENCES

- Abdurrahman, G. (2019). Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, 4(1), 13. <https://doi.org/10.32528/justindo.v4i1.2418>.
- Computer, J. (2024). *Perbandingan Metode Agglomerative Hierarchical Clustering Dan Metode Kmedoids Dalam Pengelompokan Data Titik PanaS*. 2(1), 31–38.
- Desma Aipina, & Harry Witriyono. (2022). Pemanfaatan Framework Laravel Dan Framework Bootstrap Pada Pembangunan Aplikasi Penjualan Hijab Berbasis Web. *Jurnal Media Infotama*, 18(1), 36–42.
- Dhanita Listra Prashanti Dewi, M., & Wasito, I. (2024). Evaluation of Cluster Models for Creating Profiles of Home Buyers. *Jurnal Dan Penelitian Teknik Informatika*, 8(4). <https://doi.org/10.33395/sinkron.v8i4.13888>
- Harjoseputro, Y., Albertus Ari Kristanto, & Joseph Eric Samodra. (2020). Golang and NSG Implementation in REST API Based Third-Party Sandbox System. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(4), 745–750. <https://doi.org/10.29207/resti.v4i4.2218>
- Ikhsan, M., Armansyah, A., & Tamba, A. A. (2022). Implementasi Jaringan Syaraf Tiruan Backpropagation Pada Klasifikasi Grade Teh Hitam. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(2), 387. <https://doi.org/10.30865/json.v4i2.5312>
- Justitia, R. P., Hidayat, N., & Santoso, E. (2021). Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus: RichDjoe Barbershop Malang). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(3), 1048–1054.
- Mhd Furqan, Armansyah, N. (2022). Disease in Corn Leaf Using Gabor Wavelet and K-Means Clustering Algorithm. *Jurnal Mantik*, 5(36), 2152–2156.
- Nellie, Vanesa, Viny Christanti Mawardi, & Novario Jaya Perdana. (2023). *Jurnal Ilmu Komputer dan Sistem Informasi Implementasi Metode Agglomerative Hierarchical Clustering Untuk Sistem Rekomendasi Film*.
- Orisa, M., & Faisol, A. (2021). Analisis Algoritma Partitioning Around Medoid untuk Penentuan Klusterisasi. *Jurnal Teknologi Informasi Dan Terapan*, 8(2), 86–90. <https://doi.org/10.25047/jtit.v8i2.258>
- Pane, P. P., Nasution, Y. R., & Furqan, M. (2024). *Implementasi Data Mining dengan K-Means Clustering untuk Memprediksi Pengadaan Obat*. 5(2), 286–296. <https://doi.org/10.47065/josyc.v5i2.4920>
- Rahman, S., Sembiring, A., Siregar, D., Khair, H., Gusti Prahmana, I., Puspadini, R., & Zen, M. (2023). Python : Dasar Dan Pemrograman Berorientasi Objek. In *Penerbit Tahta Media*.
- Sasmita Susanto, E., Karisma, Y., & Isnaeni, S. (2019). Sistem Informasi Penjualan Pada Toko Jilbab Rjs Kabupaten Sumbawa Berbasis Web. *Jurnal Informatika, Teknologi Dan Sains*, 1(2), 97–103. <https://doi.org/10.51401/jinteks.v1i2.414>
- Sofiyah, A., & Zafi, A. A. (2020). Hijab Bagi Wanita Muslimah Di Era Modern. *Jurnal Pengembangan Masyarakat*, 13(1), 89–102.
- Sudrajat, R. A. (2023). *Perancangan Sistem Informasi Manajemen Toko ( Point Of Sales ) Berbasis Website di UMKM Hijab Nafisa*. 1(2).
- Syahara, U., Kurniawati, E., Suhana, M. P., Anggraini, R., Yandri, F., Kelautan, I., Perikanan, D., Maritim Raja, U., & Haji, A. (2024). INSOLOGI: Jurnal Sains dan Teknologi Penerapan Metode AHC (Agglomerative Hierarchical Clustering) untuk Klasifikasi Habitat Bentik di Desa pengudang, Kabupaten Bintan. *Media Cetak*, 3(3), 306–314. <https://doi.org/10.55123/insologi.v3i3.3547>
- Ulvi, H. A., & Ikhsan, M. (2024). Comparison of K-Means and K-Medoids Clustering Algorithms for Export and

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Import Grouping of Goods in Indonesia. *Jurnal Dan Penelitian Teknik Informatika*, 8(3), 1641–1655. <https://doi.org/10.33395/sinkron.v8i3.13815>
- Widyawati, W., Saptomo, W. L. Y., & Utami, Y. R. W. (2020). Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan. *Jurnal Ilmiah SINUS*, 18(1), 75. <https://doi.org/10.30646/sinus.v18i1.448>
- Wijaya, K. A., & Swanjaya, D. (2021). Integrasi Metode Agglomerative Hierarchical Clustering dan Backpropagation Pada Model Peramalan Penjualan. *Seminar Nasional Inovasi Teknologi, 1996*, 132–141.
- Xia, S., Peng, D., Meng, D., Zhang, C., Wang, G., Giem, E., Wei, W., & Chen, Z. (2022). Ball k-Means: Fast Adaptive Clustering With No Bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1), 87–99. <https://doi.org/10.1109/TPAMI.2020.3008694>
- Yulianti, D. I., Hermanto, T. I., & Defriani, M. (2023). RESOLUSI : Rekayasa Teknik Informatika dan Informasi Analisis Clustering Donor Darah dengan Metode Agglomerative Hierarchical Clustering. *Media Online*, 3(6), 441.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.