Journal of Emerging Information Systems and Business Intelligence
ISSN: 2774-3993
Journal homepage: https://ejournal.unesa.ac.id/index.php/JEISBI/

Clustering of Goat Buyers in West Java with K-Means Algorithm

Faizatul Mukaromah[1], I Kadek Dwi Nuryana[2]
[1,2]*Department of Information Systems, State University of Surabaya, Surabaya, Indonesia*
faizatulm238@gmail.com, dwinuryana@unesa.ac.id

ABSTRACT

The advancement of information technology has encouraged the utilization of data as a strategic resource across various fields, including the livestock sector. This study aims to implement the K-Means algorithm to segment goat buyers in West Java Province based on demographic characteristics such as age, marital status, geographic location, type of goat purchased, and transaction methods. This segmentation is expected to assist business actors in understanding purchasing patterns and designing more targeted marketing and distribution strategies. The study uses 1,250 transaction records and follows the stages of selection, preprocessing, transformation, data mining, and interpretation using the Knowledge Discovery in Databases (KDD) approach. Geographic distances between buyer locations and reference points were calculated using the Haversine formula. To determine the optimal number of clusters, the Elbow Method and Silhouette Score were used, with the best result obtained at a Silhouette score of 0.16 for 3 clusters. Each cluster was analyzed based on modal characteristics such as age, marital status, district, type of goat purchased, number of goats per transaction, purchase purpose, delivery method, payment method, as well as Recency, Frequency, and Monetary (RFM). The results indicate that the K-Means algorithm is effective in grouping goat buyers into relevant and meaningful segments. This information can be used by farmers and stakeholders to improve distribution efficiency, stock optimization, and data-driven marketing strategies. This study also emphasizes the importance of integrating technologies such as Python and Streamlit for interactive visualization and ID-based buyer tracking in advanced analytics.

Keyword: K-Means, market segmentation, goat buyers, data mining, Haversine, West Java.

Article Info:
Article history:
Received July 19, 2025
Revised August 27, 2025
Accepted September 12, 2025

Corresponding Author
Faizatul Mukaromah
State University of Surabaya, Surabaya, Indonesia
Faizatulm238@gmail.com

1. INTRODUCTION

With the rapid advancement of information technology, data has become one of the most valuable resources across various sectors, including business and industry. Data is not merely a collection of numbers or facts, but rather a strategic asset that can provide deep insights into consumer behavior, market trends, and hidden patterns that are not easily identified through conventional means[1]. To extract such hidden knowledge from large and complex datasets, one of the most prominent approaches employed is data mining. Data mining involves a wide range of analytical techniques aimed at discovering patterns and relationships in massive datasets. Among these techniques, clustering is a widely adopted method where data with similar attributes are grouped into specific categories. One of the most popular and effective clustering algorithms is K-Means, which was first introduced by Lloyd in 1957. K-

Means clusters data into k groups based on the proximity of data points to the cluster centroids. Due to its simplicity, efficiency, and versatility, K-Means has been successfully applied in various domains, including marketing analytics, customer segmentation, and technical data grouping.

In the context of Indonesia's livestock sector, particularly goat farming in West Java, the demand for goat products—such as meat for consumption or animals for religious purposes—continues to increase. This growth indicates a significant opportunity for improving market targeting and distribution strategies[2]. However, one of the ongoing challenges faced by goat farmers is the lack of a systematic approach to understanding customer preferences and market behavior. Factors such as buyer demographics, geographical location, transaction frequency, type of goat purchased, and preferred payment and delivery methods all influence consumer decision-making and demand patterns. Without a clear understanding of these characteristics, farmers may struggle to optimize their marketing efforts. In this regard, the application of K-Means clustering to segment goat buyers based on their demographic and transactional data can offer an effective solution. Segmentation enables farmers to better understand market structures, identify key customer segments, and develop targeted strategies that cater to the specific needs of each group.

Python, as a high-level programming language, offers extensive support for implementing the K-Means algorithm through its powerful libraries for data manipulation, clustering, and visualization[3]. Its flexibility and compatibility with various data formats make it an ideal platform for clustering tasks, including customer segmentation. By using Python, this research aims to generate accurate clustering models that not only facilitate data analysis but also provide meaningful insights to support decision-making in goat product distribution. Through the segmentation of goat buyers using K-Means, farmers can divide the market into smaller, more homogeneous groups with similar characteristics. This allows for better inventory planning, pricing strategies, and overall marketing efficiency. Ultimately, this approach can enhance customer satisfaction and reduce the risk of overstock or undersupply by aligning product offerings with consumer needs and preferences.

2. METHODS

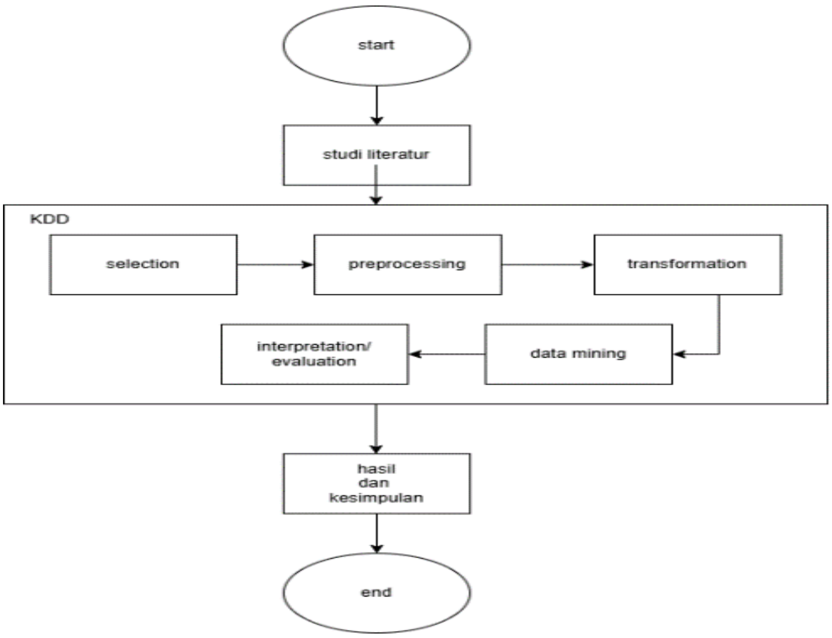

Figure 1. Research Method Flowchart

This study applies a quantitative exploratory approach to segment goat buyers in West Java Province based on their demographic characteristics and transactional behavior. The dataset includes information such as buyer ID, age, marital status, sub-district, type and number of goats purchased, purchase purpose, delivery method, payment method, and geographic coordinates. In the initial stage, data selection was performed by removing descriptive columns such as name, phone number, and address to maintain analysis relevance and ensure data efficiency. The preprocessing stage included handling missing values, calculating geographic distance using the Haversine formula[4], and computing mode values for several categorical attributes. The data was then transformed by calculating Recency, Frequency, and Monetary (RFM) metrics, applying label encoding to categorical variables, and normalizing numerical features using the StandardScaler. The data mining process was conducted using the K-Means algorithm to cluster customers into homogeneous groups. The optimal number of clusters was determined using the Elbow Method and validated through the Silhouette Score. The clustering results were visualized using Principal Component Analysis (PCA) to aid interpretation[5]. Each cluster was analyzed descriptively to assign appropriate segment labels based on purchasing behavior patterns. This study employed Python programming language along with libraries such as Pandas, Scikit-learn, and Matplotlib to support the data analysis and visualization process, as commonly practiced in customer segmentation studies based on data mining.

3. RESULTS AND DISCUSSION

3.1 Data Selection

```
Data columns (total 22 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   id                           2881 non-null   int64
 1   Umur                         2881 non-null   int64
 2   Status                       2881 non-null   object
 3   kecamatan                    2881 non-null   object
 4   latitude                     2881 non-null   int64
 5   longitude                    2881 non-null   int64
 6   jenis_kambing_yang_dibeli    2881 non-null   object
 7   jumlah_kambing_per_transaksi 2881 non-null   int64
 8   tujuan_pembelian             2881 non-null   object
 9   latitude_ref                 2881 non-null   float64
 10  longitude_ref                2881 non-null   float64
 11  metode_pengiriman            2881 non-null   object
 12  metode_pembayaran            2881 non-null   object
 13  tanggal_transaksi            2881 non-null   datetime64[ns]
```

Figure 2. Selected Variables

Data selection analysis indicated that only numerical columns were used for K-Means clustering, while non-numeric fields such as name, phone number, and address were excluded to ensure computational efficiency. This exclusion is essential, as the K-Means algorithm requires numeric data for calculating distances, such as those used in the Haversine formula. Textual fields like names and phone numbers do not provide meaningful quantitative input for clustering processes. Furthermore, removing these non-numeric columns helps reduce potential noise in the dataset and allows the model to focus on variables that better reflect the demographic characteristics of the buyers—attributes that are more significant for identifying segmentation patterns. This step also minimizes computational complexity and contributes to producing more accurate clustering results.

ISSN: 2774-3993

3.2 Preprocessing Data

	id	latitude	longitude	latitude_ref	longitude_ref	jarak_km
0	1	-6.881526	107.642692	-7202563	107879217	7812.990875
1	1	-6.881526	107.642692	-7202563	107879217	7812.990875
2	2	-6.854912	107.548712	-7202563	107879217	7822.566581
3	2	-6.854912	107.548712	-7202563	107879217	7822.566581
4	3	-6.881526	107.642692	-7202563	107879217	7812.990875

Figure 3. Distance Calculation Results Using Haversine Formula

The distance between locations was calculated using the Haversine formula, which accounts for the curvature of the Earth and uses a standard Earth radius of 6,371 km to ensure accuracy. For example, the coordinate pair (–6.881526, 107.642691) and the reference point (–7.202563, 107.879217) yields a distance of 7,812.990875 km. Meanwhile, the pair (–6.854912, 107.548712) with the same reference point results in a distance of 7,822.566581 km. This significant variation in distance reflects the spatial differences between the analyzed locations, which likely represent different cities or regions.

	id	jenis_kambing_yang_dibeli
0	1	Boer
1	2	Gibas
2	3	Boer
3	4	Boer
4	5	Boer
...	...	...
1245	1246	Kacang
1246	1247	Kacang
1247	1248	Kacang
1248	1249	Kacang
1249	1250	Boer

1250 rows × 2 columns

Figure 4. Mode of goat purchase type for each ID.

As part of the preprocessing stage, a mode calculation was performed to determine the most frequently purchased goat type for each individual customer ID. This step aimed to identify specific buyer preferences by analyzing the transaction history and extracting the dominant goat type associated with each buyer. For instance, based on a sample of calculated results, customers with IDs 01, 03, 04, and 05 most frequently purchased the “Boer” goat. This indicates a consistent and strong preference among certain customers, suggesting a specific behavioral pattern in their purchasing habits. By identifying the mode of goat types, this process supports the creation of detailed customer profiles, enables more targeted marketing strategies (e.g., promotions focused on Boer-type goats), and contributes to improved stock optimization and production planning based on clearly identified market demand.

	id	total_jumlah_kambing
0	1	6
1	2	8
2	3	5
3	4	3
4	5	20
...	...	...
1245	1246	10
1246	1247	12
1247	1248	3
1248	1249	15
1249	1250	4

1250 rows × 2 columns

Figure 5. Total Number of Goats Purchased by Each ID

To understand the purchase volume of each customer, a data aggregation was performed to calculate the total number of goats purchased by each customer ID throughout their transaction history. This process involved summing the quantity of goats from all transactions per customer into a single total value. From the sample results presented, it was observed that some early customer IDs, such as 03 and 1248, had a total of three goats purchased. This data suggests that these customers tend to make small-scale purchases within the observed period. Such information is highly valuable for distinguishing between low-volume buyers and wholesale or high-frequency customers, enabling more accurate segmentation and targeted inventory and marketing strategies.

	id	tujuan_pembelian
0	1	Kurban
1	2	Aqiqah
2	3	Dijual kembali
3	4	Kurban
4	5	Kurban
...	...	...
1245	1246	Kurban
1246	1247	Aqiqah
1247	1248	Aqiqah
1248	1249	Aqiqah
1249	1250	Kurban

1250 rows × 2 columns

Figure 6. Mode of Purchase Purpose for Each ID

To understand the purpose behind each purchase and identify dominant customer tendencies, a mode calculation was performed on the purchase purpose for each customer ID. This step identified the most frequent category of purchase intent across each buyer’s transaction history. Based on the sample results, it was found that for several early customer IDs—such as 1, 4, and 5—the most common purchase purpose was "For Qurban." Segmenting customers based on their primary purchasing intentions allows businesses to tailor communication strategies, product offerings, and inventory management to more effectively meet end-consumer demands.

	id	metode_pengiriman
0	1	diantar
1	2	pick up
2	3	pick up
3	4	pick up
4	5	diantar
...	...	...
1245	1246	diantar
1246	1247	pick up
1247	1248	pick up
1248	1249	diantar
1249	1250	pick up

1250 rows × 2 columns

Figure 7. Mode of Delivery Method for Each ID

This process identified the most frequently used delivery method in each customer's transaction history. As shown in Figure 7, which displays a sample of the calculated results, several early customer IDs—such as A01, 2, 3, and 4—most frequently selected the "pick up" method. This indicates a strong tendency among these customers to prefer self-collection over alternative options such as home delivery.

	id	metode_pembayaran
0	1	transfer
1	2	transfer
2	3	tunai
3	4	tunai
4	5	tunai
...	...	...
1245	1246	tunai
1246	1247	tunai
1247	1248	transfer
1248	1249	transfer
1249	1250	tunai

1250 rows × 2 columns

Figure 8. Mode of Payment Method for Each ID

To analyze the financial preferences of each customer and identify the most frequently used payment methods, the mode of payment method was calculated for each customer ID. This process identified the most common payment method category appearing in the transaction history of each individual. As shown in Figure 8, which displays a portion of the calculation results, it can be observed that for some of the initial customers, such as IDs 3, 4, and 5, the most frequently used payment method is cash. This indicates a strong tendency among these customers to choose cash as their preferred payment method.

ISSN: 2774-3993

3.3 Data Transformation

	id_baru	tanggal_transaksi	recency
0	A01	2024-11-23	205
1	A01	2024-11-12	205
2	A02	2024-03-31	201
3	A02	2024-11-27	201
4	A03	2024-03-24	449
5	A04	2024-04-02	440
6	A05	2024-05-16	222
7	A05	2024-11-06	222
8	A05	2024-03-17	222
9	A05	2024-09-08	222

Figure 9. Recency Calculation Results for Each ID

The recency calculation was performed by setting June 16, 2025, as the reference date. After ensuring that the transaction_date column was in the correct datetime format, the data was grouped by id_baru to identify the most recent transaction date for each customer. Recency was then calculated as the number of days between the reference date and the customer's last transaction date. The resulting recency values were successfully merged back into the main dataset using id_baru as the join key, thereby enabling further analysis of customer transaction activity.

	id_baru	frekuensi
0	A01	2
2	A02	2
4	A03	1
5	A04	1
6	A05	4

Figure 10. Frequency Calculation for Each ID

Following the frequency calculation, a new column named frequency was created to represent the number of unique purchase activities per customer. This process involved grouping the data by id_baru and counting the size of each group, which reflects the unique occurrences of each customer ID within the transaction dataset. The use of drop_duplicates prior to this step ensured that the resulting frequency accurately represented distinct transactions, eliminating irrelevant duplicates. For instance, customer ID A05 was recorded with a frequency of 4, indicating four separate transaction events—potentially suggesting a higher level of engagement.

	id_baru	monetary_total
0	A01	12000000
2	A02	16000000
4	A03	10000000
5	A04	6000000
6	A05	40000000

Figure 11. Total Monetary Calculation for Each ID

The total monetary value per customer was calculated by multiplying the jumlah_kambing_per_transaksi (number of goats per transaction) by the assumed unit price of IDR 2,000,000, resulting in a new column named monetary for each transaction record. The data was then grouped by id_baru, and the values were aggregated to obtain the total monetary value for each customer. For example, customer ID A05 recorded a monetary value of IDR 40,000,000, while ID A04 had a value of IDR 6,000,000—reflecting fewer transactions and a lower quantity of goats purchased. The aggregated results were subsequently merged back into the main dataset using id_baru as the key, with duplicates removed to ensure accuracy and consistency in the final dataset.

Umur	Status	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency
58	sudah menikah	cibeunying	Boer	3	Kurban	diantar	transfer	7812.990875	205
21	belum menikah	cimahi utara	Gibas	4	Aqiqah	pick up	transfer	7822.566581	201
51	belum menikah	cibeunying	Boer	5	Dijual kembali	pick up	tunai	7812.990875	449
66	sudah menikah	karangpawitan	Boer	3	Kurban	pick up	tunai	7763.811280	440
21	belum menikah	karangpawitan	Boer	5	Kurban	diantar	tunai	7763.811280	222

Figure 12. Results of Dropping Unnecessary Columns

Unnecessary columns were removed from the data frame to streamline the dataset. Columns such as latitude, longitude, latitude_ref, longitude_ref, transaction_date, and monetary were dropped, as their relevant information had already been summarized into derived features. Specifically, geographic coordinates were consolidated into the jarak_km (distance in kilometers) feature, transaction_date was only required for calculating recency, and the monetary field had been aggregated into monetary_total. As shown in Figure 13, which displays the data frame after column removal, only key variables such as id_baru, jarak_km, and selected behavioral and demographic features remain. This cleanup process enhances computational efficiency, reduces data complexity, and ensures that only the most relevant and meaningful attributes are retained as input for the machine learning algorithm—ultimately improving the focus and accuracy of the clustering analysis.

	id_baru	Umur	Status	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	tanggal_transaksi	ja
0	A01	58	sudah menikah	cibeunying	Boer	3	Kurban	diantar	transfer	2024-11-23	7812.
2	A02	21	belum menikah	cimahi utara	Gibas	4	Aqiqah	pick up	transfer	2024-03-31	7822.
4	A03	51	belum menikah	cibeunying	Boer	5	Dijual kembali	pick up	tunai	2024-03-24	7812.
5	A04	66	sudah menikah	karangpawitan	Boer	3	Kurban	pick up	tunai	2024-04-02	7763.
6	A05	21	belum menikah	karangpawitan	Boer	5	Kurban	diantar	tunai	2024-05-16	7763.

Figure 13. Assignment of ID Numbers Using New_ID

Customer identifiers were replaced with a new, more consistent, and manageable ID format. This process involved generating sequential new IDs (e.g., A01, A02, and so on), mapping the old IDs to the new ones, removing the original ID column, and rearranging the columns so that the new id_baru appears as the leading column. As shown in Figure 14, which presents a portion of the data frame after this transformation, the id_baru column has been successfully created and positioned at the far left, replacing the original customer ID. The presence of id_baru is crucial, as it provides a standardized and unique identifier for each customer throughout the dataset—facilitating subsequent data operations such as grouping, feature merging, and clustering analysis.

```
: 'kecamatan':
Angka: {'bojong': 0, 'cibeunying': 1, 'cimahi': 2, 'cimahi selatan': 3, 'cimahi tengah': 4, 'cimahi utara': 5, 'karangpawitan': 6}
colab.research.google.com – To exit full screen, press Esc

: 'jenis_kambing_yang_dibeli':
Angka: {'Boer': 0, 'Domba': 1, 'Etawa': 2, 'Gibas': 3, 'Kacang': 4}

: 'tujuan_pembelian':
Angka: {'Aqiqah': 0, 'Dijual kembali': 1, 'Dikonsumsi': 2, 'Kurban': 3}

: 'Status':
Angka: {'belum menikah': 0, 'sudah menikah': 1}

: 'metode_pengiriman':
Angka: {'diantar': 0, 'pick up': 1}

: 'metode_pembayaran':
Angka: {'transfer': 0, 'tunai': 1}
```

baru	Umur	Status	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency	frekuensi	mon
A01	58	1	1	0	3	3	0	0	7812.990875	205	2	
A02	21	0	5	3	4	0	1	0	7822.566581	201	2	
A03	51	0	1	0	5	1	1	1	7812.990875	449	1	
A04	66	1	6	0	3	3	1	1	7763.811280	440	1	
A05	21	0	6	0	5	3	0	1	7763.811280	222	4	

Figure 14. Application of Label Encoding

Categorical data transformation was carried out using label encoding. At this stage, label encoding was applied to identified categorical variables such as kecamatan, jenis_kambing_yang_dibeli, tujuan_pembelian, status, metode_pengiriman, and metode_pembayaran. This method converts each unique value in a categorical column into a discrete numerical representation (e.g., 0, 1, 2, etc.) within the same column. For example, if the tujuan_pembelian column contains categories such as "Aqiqah," "Konsumsi," and "Dijual," label encoding will assign values such as 0 for "Aqiqah," 1 for "Dijual," and 2 for "Konsumsi," respectively.

Umur	Status	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency
0.926617	0.959230	-1.004232	-1.218946	0.002824	1.316625	-0.993620	-1.014505	0.298604	-0.864458
-1.426947	-1.042502	0.966396	0.639959	0.708946	-1.198578	1.006421	-1.014505	0.750708	-0.905048
0.481348	-1.042502	-1.004232	-1.218946	1.415068	-0.360177	1.006421	0.985702	0.298604	1.611568
1.435496	0.959230	1.459053	-1.218946	0.002824	1.316625	1.006421	0.985702	-2.023345	1.520239
-1.426947	-1.042502	1.459053	-1.218946	1.415068	1.316625	-0.993620	0.985702	-2.023345	-0.691948

Figure 15. Application of StandardScaler

After converting the categorical variables, the next step was to normalize the numeric columns. For instance, the age values are typically in the tens range; if not normalized, machine learning models may become biased toward features with larger absolute values. The normalization was performed using the StandardScaler, which transforms each value into a standardized scale with a mean of 0 and a standard deviation of 1. This step ensures that all features contribute equally to the distance calculations in clustering algorithms such as K-Means, thereby improving model fairness and accuracy.

3.4 Data Mining

After conducting tests using the silhouette score method for various cluster counts, a comparison of the resulting silhouette score values can be seen in Table 1. For the case of three clusters (k = 3), the silhouette score achieved was 0.16, representing the highest score among the tested cluster counts. This suggests that clustering with k = 3 provides relatively good separation between clusters and strong cohesion within each cluster. In general, a higher silhouette score (with a maximum of 1) indicates better clustering quality. Therefore, the optimal number of clusters for this study was determined to be three. Based on this result, the customer segmentation process was carried out to gain deeper insights into the characteristics of each buyer segment. Once the optimal number of clusters was established, the clustering process was implemented using the K-Means algorithm with k = 3. Each customer data point was then assigned a label according to

ISSN: 2774-3993

the segmentation results. This label was stored in a new column named "Cluster" in the main dataset. The addition of this column enables more accessible and structured analysis of each cluster's characteristics. With cluster information available for each record, further descriptive analysis can be conducted to explore differences in purchasing behavior across customer segments.

Table 1. Table of Silhouette Score Test Results

Jumlah Klaster (K)	Nilai Silhoutte Score
2	0,14
3	0,16
4	0,13
5	0,12
6	0,12
7	0,12
8	0,11
9	0,11

3.5 Interpretation

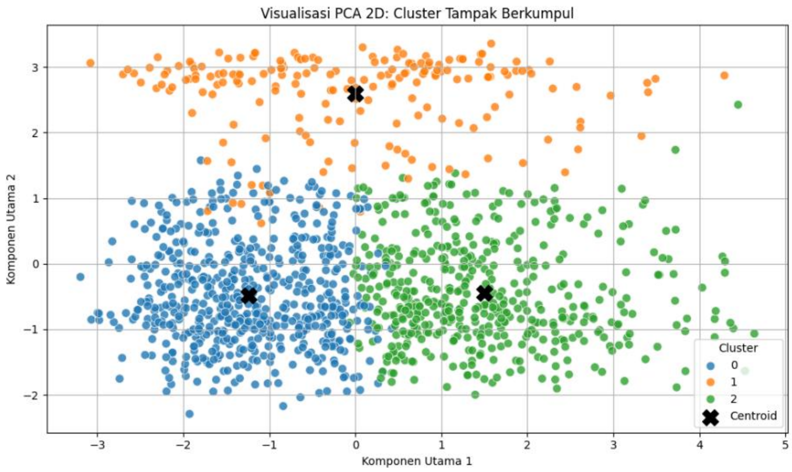

Figure 16. PCA Visualization with the Three Defined Clusters

Figure 16 shows a scatter plot visualization of the K-Means clustering results in a two-dimensional space reduced using Principal Component Analysis (PCA). Each point represents an individual customer, color-coded according to its assigned cluster, while the cluster centroids are marked with black ‘X’ symbols. The visualization reveals that the customer data has been successfully divided into three relatively distinct clusters. Cluster 0 (blue) displays a dense grouping of points concentrated in the lower-left area of the plot, centered around its corresponding centroid. Cluster 1 (orange) forms a separate pattern, occupying the upper-central to right region, with data points mostly clustered around the orange centroid, although some points are spread more broadly. Meanwhile, Cluster 2 (green) is situated in the lower-right region of the plot and, while it may slightly overlap with the other clusters, still forms a distinctive group surrounding its green centroid. This visual separation supports the validity of the clustering results and serves as a foundation for deeper interpretation of each segment’s behavioral characteristics.

us	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency	frekuensi	monetary_total	Clus
.0	0.0	4.0	1.0	0.0	0.0	1.0	7790.44198	191.0	1.0	4000000.0	

	Umur	Status	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency
count	579.000000	579.000000	579.000000	579.000000	579.000000	579.000000	579.000000	579.000000	579.000000	579.000000
mean	53.580311	0.530225	2.341969	2.193437	2.818653	1.188256	0.497409	0.500864	7813.619377	335.246978
std	11.581292	0.499517	1.724187	1.569792	1.427582	1.104004	0.500426	0.500432	12.005695	103.290866
min	17.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	7790.441980	167.000000
25%	45.500000	0.000000	1.000000	1.000000	2.000000	0.000000	0.000000	0.000000	7812.990875	248.000000
50%	56.000000	1.000000	2.000000	2.000000	3.000000	1.000000	0.000000	1.000000	7820.558336	327.000000
75%	63.000000	1.000000	4.000000	4.000000	4.000000	2.000000	1.000000	1.000000	7820.945869	418.000000
max	70.000000	1.000000	5.000000	4.000000	5.000000	3.000000	1.000000	1.000000	7822.566581	532.000000

Figure 17. Descriptive Statistics Results for Cluster 0

The descriptive statistics results for Cluster 0 reveal that the majority of individuals in this group are, on average, 53 years old and predominantly married. The youngest recorded buyer in this cluster is 17 years old. Most customers in Cluster 0 originate from the Bojong area. The most frequently purchased goat type is the Kacang breed (based on original, non-encoded values), with an average of 3 goats per transaction. The primary purpose of purchase is for aqiqah ceremonies. The preferred payment method is cash, and the delivery method is home delivery. In terms of RFM analysis, the recency is 335 days (approximately 11 months), the frequency is 1 purchase, and the monetary value is approximately IDR 7,000,000.00. Based on these characteristics, Cluster 0 is labeled as the Traditional Aqiqah Segment.

Status	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency	frekuensi	monetary_total
1.0	6.0	0.0	3.0	3.0	0.0	0.0	7763.81128	185.0	1.0	8000000.0

	Umur	Status	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency
count	192.000000	192.000000	192.0	192.000000	192.000000	192.000000	192.000000	192.000000	1.920000e+02	192.000000
mean	43.223958	0.515625	6.0	0.541667	2.927083	2.640625	0.473958	0.468750	7.763811e+03	284.057292
std	16.116618	0.501062	0.0	1.227236	1.374752	0.844456	0.500627	0.500327	3.191554e-11	91.212437
min	17.000000	0.000000	6.0	0.000000	1.000000	0.000000	0.000000	0.000000	7.763811e+03	168.000000
25%	29.000000	0.000000	6.0	0.000000	2.000000	3.000000	0.000000	0.000000	7.763811e+03	208.750000
50%	42.000000	1.000000	6.0	0.000000	3.000000	3.000000	0.000000	0.000000	7.763811e+03	262.000000
75%	57.250000	1.000000	6.0	0.000000	4.000000	3.000000	1.000000	1.000000	7.763811e+03	337.250000
max	70.000000	1.000000	6.0	4.000000	5.000000	3.000000	1.000000	1.000000	7.763811e+03	529.000000

Figure 18. Descriptive Statistics Results for Cluster 1

Figure 18 illustrates the characteristics of Cluster 1, where the average customer age is 43 years, and the majority are married. Based on the original (pre-encoded) values, most customers in this cluster are from the Karangpawitan sub-district, and the most frequently purchased goat type is Boer, with an average of three goats per transaction. The primary purchase purpose is for Qurban, with delivery typically handled via home delivery, and payment made through bank transfer. In terms of RFM metrics, the average recency is 284 days (approximately 9 months), the purchase frequency is 2 transactions per customer, and the total monetary value spent is IDR 10,000,000. Based on these characteristics, Cluster 1 is labeled as the Modern Qurban Segment, reflecting its relatively mature, transaction-oriented buyer profile with moderate purchasing power and a preference for convenient services.

atus	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency	frekuensi	monetary_total	C
1.0	5.0	4.0	5.0	0.0	1.0	1.0	7822.566581	174.0	3.0	24000000.0	

	Umur	Status	kecamatan	jenis_kambing_yang_dibeli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency
count	479.000000	479.000000	479.000000	479.000000	479.000000	479.000000	479.000000	479.000000	479.000000	479.000000
mean	31.250522	0.511482	2.693111	2.265136	3.237996	1.235908	0.505219	0.530271	7815.439584	238.179541
std	10.337195	0.500391	1.717427	1.504059	1.387985	1.124417	0.500495	0.499605	11.200612	63.125474
min	17.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	7763.811280	168.000000
25%	23.000000	0.000000	1.000000	1.000000	2.000000	0.000000	0.000000	0.000000	7812.990875	189.000000
50%	29.000000	1.000000	3.000000	2.000000	3.000000	1.000000	1.000000	1.000000	7820.612998	218.000000
75%	39.000000	1.000000	4.000000	4.000000	4.000000	2.000000	1.000000	1.000000	7820.945869	266.500000
max	59.000000	1.000000	6.000000	4.000000	5.000000	3.000000	1.000000	1.000000	7822.566581	486.000000

Figure 19. Descriptive Statistics Results for Cluster 2

Figure 19 presents the descriptive statistics of Cluster 2, revealing that the average buyer is 31 years old and predominantly married. Most customers in this cluster are from the Cimahi Utara sub-district, and the most commonly purchased goat type is Kacang, with an average of five goats per transaction. The primary purchase purpose is for Aqiqah, with payments typically made in cash and the preferred delivery method being pick-up. Based on the RFM metrics, the average recency is 238 days (approximately 8 months), the purchase frequency is three transactions per customer, and the total monetary value spent is IDR 20,000,000. Considering these characteristics, this cluster is labeled as the Young Aqiqah Segment, representing younger families with moderate purchasing frequency and a preference for direct pick-up and traditional payment methods.

beli	jumlah_kambing_per_transaksi	tujuan_pembelian	metode_pengiriman	metode_pembayaran	jarak_km	recency	frekuensi	monetary_total	Cluster	segment
8946	0.002824	1.316625	-0.993620	-1.014505	0.298604	-0.864458	-0.244735	-0.165114	0	Aqiqah Tradisional
9959	0.708946	-1.198578	1.006421	-1.014505	0.750708	-0.905048	-0.244735	0.221750	2	Aqiqah Muda
8946	1.415068	-0.360177	1.006421	0.985702	0.298604	1.611568	-1.047670	-0.358546	0	Aqiqah Tradisional
8946	0.002824	1.316625	1.006421	0.985702	-2.023345	1.520239	-1.047670	-0.745410	1	Kurban Modern
8946	1.415068	1.316625	-0.993620	0.985702	-2.023345	-0.691948	1.361136	2.542935	1	Kurban Modern

Figure 20. Application of Customer Segmentation Results

The segmentation results using the K-Means algorithm with three clusters produced a varied distribution of customers across the groups. Each cluster was assigned a segment label based on the average characteristics derived from RFM values. For example, customer ID A01 belongs to Cluster 0, which is labeled the Traditional Aqiqah Segment. The second customer falls under Cluster 2 and is categorized as part of the Young Aqiqah Segment. The third buyer also belongs to Cluster 0. The fourth customer is grouped into Cluster 1, indicating they are part of the Modern Qurban Segment. Similarly, the fifth customer is also in Cluster 1, reinforcing the consistency of the Modern Qurban Segment. These examples reflect the meaningful differentiation achieved through clustering and support the segmentation model's ability to group customers based on their transaction behavior and demographics.

```
Jumlah ID per Cluster:
Cluster
0    579
1    192
2    479
Name: count, dtype: int64
Jumlah segment per Cluster:
segment
Aqiqah Muda           479
Aqiqah Tradisional    579
Kurban Modern         192
Name: count, dtype: int64
```

Figure 21. Number of Customer IDs per Cluster and Segment

Based on the results of the clustering analysis, three distinct customer segments were identified. Cluster 0 contains the largest number of buyers, with 579 unique customer IDs, and is categorized as the Traditional Aqiqah Segment. Cluster 1 represents Modern Qurban Buyers, consisting of 192 customers. Cluster 2 includes 479 customer IDs and is identified as the Young Aqiqah Segment. The following is a summary of each cluster and its recommended strategic actions:

Cluster 0 – Traditional Aqiqah

Comprising 579 customer IDs, this segment is characterized by buyers whose main purchase purpose is aqiqah, with a strong preference for cash payment. Descriptive statistics also show that this group typically chooses the "delivery" method. Therefore, a recommended strategy for this segment is to offer free delivery for purchases of more than two goats, or special cash payment discounts to encourage larger transactions.

Cluster 1 – Modern Qurban

With 192 customer IDs, this segment includes buyers whose primary purpose is qurban, and who tend to use bank transfers instead of cash payments. To encourage repeat purchases and build loyalty, it is recommended to offer a 3% discount promotion for customers who register for membership.

Cluster 2 – Young Aqiqah

This segment includes 479 customer IDs, with buyers generally younger in age and purchasing goats primarily for aqiqah. The customers in this cluster are considered more active, as they typically use the pick-up delivery method. To retain and reward loyalty, a suggested strategy is to provide a 5% discount on purchases of at least three Kacang goats, as this is the most commonly chosen breed in this group.

```
Masukkan ID Pelanggan (contoh: A01 - A1250): A110
  Tracking Pesanan - ID: A110

  Nama              : ir. tania manullang, m.ak
  Umur              : 32
  Status            : belum menikah
  Kecamatan         : cimahi selatan
  Lokasi            : (-6.9049, 107.526207)
  Jenis Kambing     : Etawa
  Jumlah Kambing    : 4
  Tujuan Pembelian  : Dikonsumsi
  Pengiriman        : pick up
  Pembayaran        : tunai
  Tanggal Transaksi : 11/30/2024 0:00
  Cluster           : 2
  Segment           : Aqiqah Muda
  Rekomendasi       : Potongan 5% khusus untuk Pembelian kambing kacang minimal 3.
```

Figure 22. Tracking Customer

The customer tracking feature is designed to display detailed information based on a specific customer ID, including name, age, marital status, complete address, and details of the most recent transaction. The displayed data includes the type and quantity of goats purchased, purchase purpose, delivery and payment methods, and the transaction date. The system also shows the segmentation results derived from the K-Means clustering algorithm, which classifies customers into clusters based on descriptive statistical features. For example, a customer with ID A110 belongs to Cluster 2 and is labeled under the “Young Aqiqah” segment. Based on this segment, the system provides relevant recommendations, such as special offers tailored to that customer group. This feature enables the company to deliver more personalized services and implement more targeted promotional strategies.

CONCLUSION

Based on the results of this research, it can be concluded that the implementation of the K-Means algorithm was successful in segmenting goat buyers in West Java Province. The segmentation process utilized demographic and behavioral data, which were analyzed using descriptive statistics, resulting in the identification of distinct customer segments: Traditional Aqiqah, Modern Qurban, and Young Aqiqah. The determination of the optimal number of clusters was supported by the Elbow Method and Silhouette Score, ensuring the effectiveness and accuracy of the clustering process. In addition, the development of a buyer tracking feature based on customer ID and clustering output allowed for the display of detailed customer information, such as location, transaction history, segment assignment, and personalized recommendations. This functionality enhances the precision of marketing and distribution strategies by aligning them with customer-specific characteristics. The findings of this study demonstrate the potential of K-Means clustering not only as a tool for customer segmentation but also as a foundation for personalized marketing in the livestock sector. For future research, the segmentation approach could be further developed by incorporating additional variables or exploring hybrid clustering methods. The buyer tracking system may also be expanded into a real-time application to support dynamic segmentation and adaptive promotional strategies based on live customer behavior.

REFERENCES

[1] T. Hartati, O. Nurdiawan, and E. Wiyandi, “Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari,” *Jurnal Sains Teknologi Transportasi Maritim*, vol. 3, 2021.

[2] Y. Pratiwi and N. Suarna, “PENERAPAN METODE K-MEANS CLUSTERING PADA POPULASI AYAM RAS PETELUR BERDASARKAN PROVINSI,” 2023. [Online]. Available: https://www.bps.go.id

[3] D. D. R. Agustina Inggit, “IMPLEMENTASI K-MEANS CLUSTERING UNTUK PENGELOMPOKAN TINGKAT KEMISKINAN DI PULAU JAWA MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON,” vol. 1, Feb. 2024.

[4] A. Wibowo, A. R. Handoko, and P. Korespondensi, “SEGMENTASI PELANGGAN RITEL PRODUK FARMASI OBAT MENGGUNAKAN METODE DATA MINING KLASTERISASI DENGAN ANALISIS RECENCY FREQUENCY MONETARY (RFM) TERMODIFIKASI SEGMENTATION OF CUSTOMERS OF DRUG PHARMACEUTICAL PRODUCT RETAIL USING CLASTERIZATION MINING DATA METHOD USING MODIFIED MONETARY RECENCY FREQUENCY (RFM) ANALYSIS,” vol. 7, no. 3, 2020, doi: 10.25126/jtiik.202072925.

[5] G. Enzellina and D. Suhaedi, “Penggunaan Metode Principal Component Analysis dalam Menentukan Faktor Dominan,” *Jurnal Riset Matematika*, pp. 101–110, Dec. 2022, doi: 10.29313/jrm.v2i2.1192.