Available online at www.coins-indocompt.org

COmputing and INformation Systems Journal

E-ISSN: 3109-3248

Vol. 1, No. 3, December, 2025

Optimization of Laying Ducks Feed Composition Using The K-Means Clustering Algorithm Method

Nur Arifah[a], Uky Yudatama[b*], Ardin Primadewi[c]

[a,b,c]*Universitas Muhammadiyah Magelamg, Mertoyudan, Magelang, 56172, Indonesia*

*uky@unimma.ac.id

Abstract

Indonesia has significant potential in duck farming; however, the productivity of local laying ducks remains low due to traditional feed management practices that do not consider nutritional balance. This study aims to determine the optimal feed composition for laying ducks using the K-Means Clustering algorithm. A dataset consisting of 13 types of feed ingredients and five nutritional variables (protein, fat, fiber, calcium, and phosphorus) was analyzed. Prior to clustering, data normalization was performed, and the Elbow Method indicated that three clusters were the optimal grouping (SSE reduction stabilizing at k = 3). The results show that Cluster 1 contains high-protein ingredients (⩾18%), such as fishmeal (49.03%), blood meal (80.31%), soybean meal (46.27%), and fresh snails (48.70%), making it the most optimal cluster to support egg production due to its superior protein and phosphorus composition. Cluster 2 represents low- to medium-protein ingredients (1%–9%), functioning mainly as energy and fiber complements, whereas Cluster 3 consists of high-fiber ingredients (9%–15%) suitable as digestive balancers. Based on these findings, Cluster 1 is recommended as the primary component in formulating duck feed. This study demonstrates the practical value of machine learning in livestock feed optimization and provides a data-driven recommendation that can help farmers improve productivity and efficiency

Keywords: Laying Ducks, Feed Composition, K-Means Clustering, Machine Learning, Nutritional Optimization

1. Introduction

Indonesia is one of the countries with the highest potential for duck production in Southeast Asia, supported by its climatic suitability, abundant natural resources, and long-standing cultural integration of duck farming. Laying ducks contribute significantly to the national egg supply, accounting for approximately 19.35% of total egg production, making them an important source of high-quality animal protein for the population(Fitri Yani et al., 2024).Duck eggs are well known for their high nutritional value, distinct flavor, and superior amino-acid profile compared to several other animal-protein sources, leading to increasing consumer demand across various regions.

In many rural areas, including Secang District in Magelang Regency, duck farming plays an essential role in household income generation. The availability of land, water access, and local feed ingredients make the region highly suitable for laying-duck farming, with some farmers capable of producing hundreds of eggs per day. However, despite this potential, the productivity of Indonesian local ducks remains far below optimal levels. This is primarily due to traditional management practices, especially in feed formulation, which remain highly subjective and inconsistent across farmers.

Feed is a major determinant of poultry productivity, accounting for up to 60–70% of total production costs and directly influencing egg output, growth performance, and livestock health. Yet, farmers in many rural areas formulate feed based merely on availability and peer recommendations, without considering the specific nutritional needs of laying ducks. Such practices often result in imbalanced nutrient composition, deficiencies in essential components such as protein, calcium, and phosphorus, or excessive fiber levels that reduce feed efficiency. These inconsistencies contribute to reduced egg production, poor feed conversion, and in some cases, increased mortality due to prolonged nutritional stress (Rozi et al., n.d.).

To address these challenges, data-driven precision-feeding approaches have gained significant attention globally. Recent developments in artificial intelligence (AI) and machine learning (ML) offer powerful tools to optimize livestock feed formulation by analyzing multi-variable nutritional datasets and identifying patterns that are not easily observed through conventional methods (Zhang et al., 2025a). In particular, unsupervised learning techniques such as K-Means Clustering have been widely applied in agriculture and animal science to classify feed materials, identify nutrient similarities, detect anomalies, and support decision-making in feed optimization (Giantika Utami, 2022).

Machine learning has shown strong potential in several livestock-related domains, such as predicting feed efficiency (Zhang et al., 2025b), optimizing dietary protein levels (Ma et al., 2022a), classifying feed ingredients (Nuningtyas et al., 2023), modeling egg-production curves (Tukiyat et al., 2024), and generating precision-nutrition guidelines for poultry (Naeem et al., 2025a). K-Means clustering, in particular, has been demonstrated as an effective tool for grouping feed ingredients based on chemical composition, metabolizable energy, amino acids, or mineral content (Akintan et al., 2024). These studies prove that AI-assisted feed formulation can significantly enhance productivity, reduce costs, and improve sustainability.

In the context of laying duck farming, the application of machine learning techniques remains limited, especially at the local farmer level. Therefore, integrating K-Means clustering to analyze nutrient content and production performance presents an opportunity to formulate objective, consistent, and data-driven feed recommendations. By clustering feed ingredients based on nutritional similarity, farmers can identify which combinations of protein, fat, fiber, calcium, and phosphorus are most supportive of egg production. This approach minimizes trial-and-error feeding practices and promotes efficient feed utilization.

Based on these considerations, this research aims to analyze and determine the optimal feed composition for laying ducks using the K-Means Clustering algorithm. The specific objective of this study is to group feed ingredients based on their nutritional characteristics and identify the cluster with the most suitable nutrient balance to support optimal egg production. The findings of this study are expected to assist farmers in adopting more scientific and effective feed-formulation practices, thereby improving productivity and supporting the development of sustainable duck farming systems.

2. Methods

2.1. Research procedure

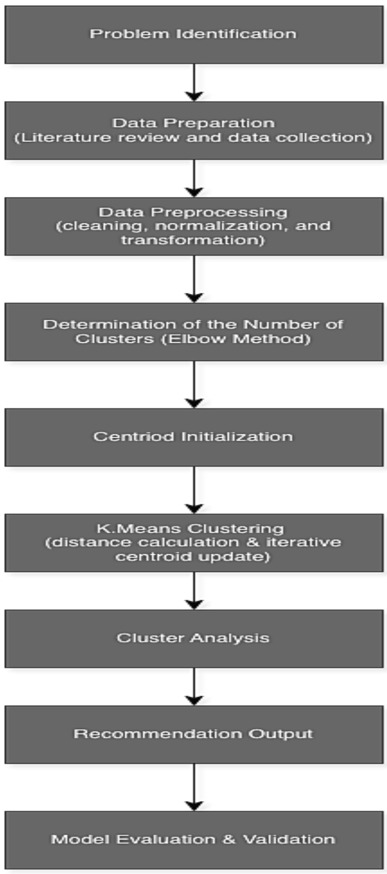

Figure 1. Flowchart of research stages

This study employs a structured, iterative workflow designed to ensure the accuracy, validity, and reproducibility of the clustering analysis for optimizing feed composition for laying ducks. The comprehensive procedure, illustrated in Figure 1, integrates data collection, machine learning techniques, and domain expertise. It consists of nine interconnected stages: (1) problem identification and literature review, (2) data collection and preparation, (3) data preprocessing (cleaning, normalization, outlier assessment), (4) determination of the optimal number of clusters (k) using the Elbow Method, (5) centroid initialization, (6) the iterative K-Means clustering process, (7) cluster analysis and interpretation, (8) formulation of practical recommendations, and (9) model evaluation and validation. The flowchart highlights key decision points and the iterative feedback loop within the K-Means algorithm, ensuring a transparent and systematic research process.

2.2. Data collection and preparation

The study began by identifying key challenges faced by laying-duck farmers in Secang District, Magelang Regency, Indonesia, through direct engagement. The primary problems included inconsistent feeding practices, limited knowledge of precise nutrient requirements, and difficulties in formulating scientifically appropriate, cost-effective feed.

Table 1. Feed ingredient data and sources

Yes	Feed	Source
1	Fresh Snails	Practitioner Interview of Bp. Faisal, Laying Duck Breeder in Purwosari Secang Village
2	Brown Rice	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
3	Soybean Meal	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
4	Coconut Meal	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
5	Fine Bran	Practitioner Interview of Bp. Atok Laying Duck Farmer in Donomuyo Secang Village
6	Corn	Practitioner Interview of Bp. Faisal, Laying Duck Breeder in Purwosari Secang Village
7	Soybean	Recommended Feed for Laying Ducks Page 107 to 108(H. Rahmat Rukmana, 2024)
8	Laying Duck Feed HDIP3 1218	Interview of Practitioner Bp. Atok Laying Duck Breeder in Donomuyo Secang Village, Bp. Faisal Laying Duck Farmer in Purwosari Secang Village, and Bp. Fahrur Laying Duck Farmer in Karang Kajen Secang Village
9	Blood Flour	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
10	Lamtoro Leaf Flour	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
11	Fishmeal	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
12	Cassava Flour	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
13	Fish Bone Meal	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)

Data Sources:

Primary Data: Collected through structured interviews and on-site observations with three experienced laying-duck farmers (n=3) across three villages (Purwosari, Donomuyo, and Karang Kajen). Data included locally used feed types and their perceived nutritional values.

Secondary Data: Sourced from authoritative literature on duck nutrition and standard feed formulation guides (Akintan et al., 2024)These values were cross-referenced with established poultry nutrient requirements (Ma et al., 2022a)to ensure scientific validity.

The final curated dataset comprised 13 feed ingredients, each characterized by five critical nutritional variables: Protein, Fat, Fiber, Calcium, and Phosphorus. To enhance computational precision and avoid rounding errors during analysis, all raw percentage values were multiplied by 100, converting them into dimensionless numerical units (e.g., 48.70% became 4870). This is a common practice in data preparation for clustering algorithms.

2.3. Data preprocessing

Data preprocessing includes cleaning inconsistent values, removing incomplete entries, and preparing the dataset for analysis. Numerical normalization was performed using min–max normalization, which scales variables into a uniform range (0–1) to prevent dominance of variables with larger magnitudes(Hamouda et al., 2025). Preprocessing was conducted to ensure data quality and suitability for clustering.

2.3.1. Data collection and preparation

The dataset was rigorously inspected for missing, erroneous, or inconsistent values. No missing entries were found. All values were verified for logical consistency (e.g., non-negative values for nutrients).

2.3.2. Data collection and preparation

The features (Protein, Fat, Fiber, Calcium, Phosphorus) had vastly different numerical ranges (e.g., Protein: 110 - 8031, Calcium: 0 - 3551). To prevent variables with larger magnitudes from dominating the Euclidean distance calculation and disproportionately influencing the cluster formation, all features were normalized to a common scale using Min-Max scaling(Ma et al., 2022b). This technique scales the data to a fixed range of [0, 1]. The formula is given by:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where:
x is the original value,
x_{min} and x_{max} are the minimum and maximum values of that feature across the entire dataset,
x_{norm} is the normalized value.

2.3.3. Data collection and preparation

Nutrient values were screened for extreme deviations using descriptive statistics (e.g., Interquartile Range - IQR) and visual inspection (e.g., box plots). No significant outliers that could be considered erroneous were removed, as all values represented plausible, real-world nutrient compositions. This step ensured that the clustering would be based on the natural structure of the data without being skewed by data entry errors.

Table 2. Nutrient Composition of Feed Ingredients (Improved Formatting)

No	Feed Type	Protein	Fat	Fiber	Calcium	Phosphorus
1	Fresh Snails	48.70	20.30	0.00	6.92	5.23
2	Brown Rice	9.36	1.62	1.31	0.00	1.82
3	Soybean Meal	46.27	14.98	8.62	0.30	8.20
4	Coconut Meal	21.60	12.10	15.49	1.65	0.21
5	Fine Bran	8.54	9.11	7.60	9.75	5.67
6	Corn	5.50	1.00	10.00	0.20	0.90
7	Soybean	40.40	16.70	3.20	0.22	0.68
8	HDIP3 Feed	17.00	3.00	10.00	2.90	0.45
9	Blood Meal	80.31	0.76	5.07	0.93	7.90
10	Lamtoro Leaf Meal	14.20	10.90	9.40	0.86	0.25
11	Fishmeal	49.03	4.71	5.66	35.51	11.28
12	Cassava Flour	1.10	0.50	0.90	0.84	0.12
13	Fish Bone Meal	9.20	0.44	0.00	7.35	3.45

2.4. Data collection and preparation

The Elbow Method was employed to determine the optimal number of clusters (k) for the K-Means algorithm(Ayu et al., 2019). The algorithm was run for a range of k values (from k=1 to k=10). For each k, the Within-Cluster Sum of Squares (WCSS) or Sum of Squared Errors (SSE) was calculated. The SSE measures the total squared Euclidean distance between each data point and its assigned cluster centroid, representing intra-cluster variance. The optimal k is identified at the "elbow" of the plotted SSE-vs-k curve. This point represents where the marginal gain in explained variance (reduction in SSE) drops significantly, indicating that adding more clusters provides diminishing returns. The analysis concluded that k=3 was the optimal choice for this dataset(Naeem et al., 2025b).

2.5. Data collection and preparation

Initial centroids for the K-Means algorithm were selected using the random initialization method. As the algorithm's convergence can be sensitive to initial starting points, it was run multiple times (n_init = 10) with different random seeds(Javed Mehedi Shamrat et al., 2020). The final result was chosen from the run that yielded the lowest SSE, ensuring a stable and robust clustering solution(Maori, 2023)

2.6. Data collection and preparation

The K-Means clustering algorithm was used to partition the data into k distinct, non-overlapping clusters (Annas & Wahab, 2023). The algorithm aims to minimize the within-cluster variance (SSE). It operates through an iterative refinement process with two main steps(Maori, 2023):

1. Cluster Assignment
   Each data point is assigned to the cluster whose centroid is closest, as measured by the Euclidean distance. The Euclidean distance between a point and a centroid in a p-dimensional space is calculated as:

$$d\left(x_i, C_j\right) = \sqrt{\sum_{k=1}^{p}(x_{ik} - C_{jk})^2} \quad (2)$$

2. Centroid Update
   The centroid of each cluster is recalculated as the mean (arithmetic average) of all data points currently assigned to that cluster:

$$C_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i \quad (3)$$

The algorithm repeated assignment–update cycles until convergence, defined as no change in cluster membership or minimal centroid shift.

2.7. Data collection and preparation

The resulting clusters were analyzed by examining the centroid values (denormalized to their original scale for interpretation) to define the nutritional profile of each group. Validation was conducted on two fronts:

- Internal Validation: The quality of the clustering structure was assessed using the SSE trend across iterations and the Silhouette Score (Alva Mustika & Dhika, 2022)The Silhouette Score measures how similar an object is to its own cluster compared to other clusters, providing insights into cohesion and separation. A score close to 1 indicates well-defined clusters.
- External Validation: The practical and biological relevance of the clusters was evaluated by comparing their nutritional characteristics against established zootechnical standards and literature for laying duck nutrition (Maylawati et al., 2020). This ensures the clusters are not just statistically sound but also meaningful for the intended application.

Hypotheses:

- H_1: The application of K-Means clustering improves nutritional efficiency and egg productivity through optimized feed grouping.
- H_0: K-Means clustering does not significantly improve feed-use efficiency or egg productivity.

2.8. Recommendation implementation

Based on the interpreted nutritional profiles of the clusters, specific feed composition recommendations were formulated for farmers. The cluster with the most favorable nutrient profile for egg production (e.g., high protein, balanced calcium and phosphorus) was identified as optimal. Practical advice on feed selection and potential mixing ratios was provided, tailored to the availability of local ingredients identified in the study.

2.9. Evaluation and validation

The model's performance and stability were evaluated by analyzing the internal validation metrics (SSE, Silhouette Score) and the external validity of its recommendations (H. Rahmat Rukmana, 2024). The robustness of the clustering was tested by re-running the algorithm multiple times with different initializations and checking for consistent results. The hypotheses were evaluated by determining if the algorithm produced clusters that were both statistically distinct and nutritionally meaningful for improving laying duck productivity.

2.10. Hypothesis

Main Hypothesis (H1) Theapplication of the K-Means Clustering algorithm in optimizing the feed composition of laying ducks can improve the nutritional efficiency and productivity of eggs compared to conventional feed compositions that are not optimized. The K-Means Clustering algorithm can group the feed composition of laying ducks based on optimal nutritional content according to the physiological needs of laying ducks, so as to increase egg productivity and feed use efficiency. Hypothesis Zero (H0) The application of the K-Means Clustering algorithm in optimizing the feed composition of laying ducks does not have a significant influence on the nutritional efficiency and productivity of laying duck eggs. The K- Means Clustering algorithm does not have a significant influence on improving feed use efficiency and egg productivity, which means that the resulting feed composition is not more optimal than conventional feed composition. Experimental Design Stage

The experimental design stage used in this study is presented in the following figure:

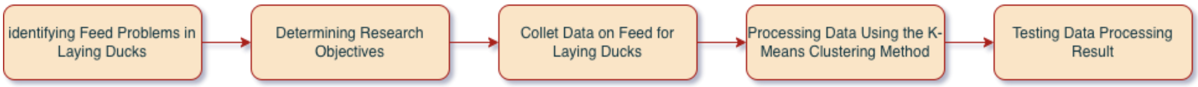

Figure 2. Stage of Experimental Design

Table 3. Feed Data and Its Sources

Yes	Feed	Source
1	Fresh Snails	Practitioner Interview of Bp. Faisal, Laying Duck Breeder in Purwosari Secang Village
2	Brown Rice	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
3	Soybean Meal	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
4	Coconut Meal	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
5	Fine Bran	Practitioner Interview of Bp. Atok Laying Duck Farmer in Donomuyo Secang Village
6	Corn	Practitioner Interview of Bp. Faisal, Laying Duck Breeder in Purwosari Secang Village
7	Soybean	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
8	Laying Duck Feed HDIP3 1218	Interview of Practitioner Bp. Atok Laying Duck Breeder in Donomuyo Secang Village, Bp. Faisal Laying Duck Farmer in Purwosari Secang Village, and Bp. Fahrur Laying Duck Farmer in Karang Kajen Secang Village
9	Blood Flour	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
10	Lamtoro Leaf Flour	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
11	Fishmeal	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
12	Cassava Flour	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)
13	Fish Bone Meal	Recommended Feed for Laying Ducks Page 107 to 108 (H. Rahmat Rukmana, 2024)

The main problem in this study is identifying the optimal feed composition to increase the productivity of laying ducks. The purpose of this study is to use *the K-Means Clustering* method to group data related to feed composition based on its characteristics and identify the most optimal cluster. The primary data source obtained by the researcher was by conducting observations and direct interviews with farmers in Secang District, Magelang Regency. Meanwhile, secondary data was obtained from literature studies and scientific journals. The type of data required in this study is the main nutrient content in the form of Protein, Fat, Fiber, Calcium, and Phosphorus in every 100 grams of each type of feed, with a total of 13 types of feed where 4 types of feed were obtained from direct interviews with farmers in 3 villages in Secang District, and 9 types of e-book feed recommendations that are adjusted to the availability of feed ingredients that are easy to get. Feed data for laying ducks based on observations, interviews and literature studies by adjusting to the availability of feed ingredients that are easy to obtain along with their nutritional content in percent of 100 grams of product multiplied by 100 to facilitate the calculation of K-Means can be seen in Table 3 and Table 4.

Table 4. Nutrient Content Data and Sources

Yes	Feed Type	Protei n % (x100)	Fat% (x100)	Fiber % (x100)	Calciu m % (x100)	Phos phoru s % (x100)
1	Fresh Snails	4870	2030	0	69,2	52,3
2	Brown Rice	936	162	131	0	182
3	Cake Soybean	4627	1498	862	3	820
4	Coconut Meal	2160	1210	1549	165	21
5	Fine Bran	854	911	7,6	975	567
6	Corn	550	10	1000	2	9
7	Soybean	4040	1670	320	22,2	68,2
8	Duck Fee HDIP3 121 Egger	d 8 1700	300	1000	290	45
9	Blood Flour	8031	76	507	93	790
10	Lamtoro Leaf Flour	1420	1090	940	8,6	25
11	Fishmeal	4903	471	566	3551	1128
12	Cassava Flour	110	50	90	8,4	12,5
13	Fish Bone Meal	920	4,4	0	73,5	34,5

In processing feed data using the K-Means Clustering algorithm, there are several steps used, namely: Determine the number of clusters to be formed, namely 3 (Three) clusters, namely Cluster1, Cluster2, Cluster3.

To determine the initial cluster center point (centroid) randomly, it can be seen in table 4 as follows:

Table 5. Initial Cluster Center Points

Data To	Centroid	Protein % (x100)	Fat % (x100)	Fiber % (x100)	Calcium % (x100)	Phosphoru s % (x100)
1	1	4870	2030	0	69,2	52,3
5	2	854	911	7,6	975	567
8	3	1700	300	1000	290	45

Calculate the distance of the data from the initial centroid using the formula (1) Eucledian Distance and determine the data in the nearest cluster, which is calculated from the center of the cluster. The calculation results of the centroid that have been determined are as follows:

Table 6. Cluster Determination Calculation Results in Iteration 1

Data To	Centroid1	Centroid2	Centroid3	Minimum	*Cluster*
1	0,00	4297,19	3753,75	0,00	1
2	4359,42	1296,84	1208,62	1208,62	3
3	1295,71	4039,66	3271,78	1295,71	1
4	3228,94	2263,88	1165,03	1165,03	3
5	4297,19	0,00	1677,98	0,00	2
6	4873,31	1774,03	1221,00	1221,00	3
7	960,92	3461,34	2808,41	960,92	1
8	3753,75	1677,98	0,00	0,00	3
9	3822,55	7299,56	6400,67	3822,55	1
10	3697,85	1565,07	886,39	886,39	3
11	4004,00	4883,68	4720,57	4004,00	1
12	5156,68	1594,81	1870,58	1594,81	2
13	4439,13	1386,58	1320,14	1320,14	3

From the calculation with Formula (1) above, the results of the dataset calculation in iteration 1 are obtained. Table 3 can be grouped as follows: Cluster1 consists of 5 items, Cluster2 consists of 2 items and Cluster3 consists of 6 items. After obtaining the cluster results in iteration 1, the process will continue in iteration 2 by determining a new cluster center. New cluster determination uses the formula (2). The following is a table of the calculation results:

Table 7. New Centroid Calculation Results

New Centroid	Protein% (x100)	Fat% (x100)	Fiber% (x100)	Calcium% (x100)	Phosphoru s% (x100)
1	5294,2	1149	451	747,68	571,7
2	482,00	480,50	48,80	491,70	289,75
3	1281	462,73	770	89,85	52,75

Followed by the determination of iteration 2 cluster with the following results:

Table 8. Cluster Determination Results in Iteration 2

Data to	Centroid 1	Centroid 2	Centroid 3	Minimum	Cluster
1	1374,63	4678,97	3991,63	1374,63	1
2	4558,66	753,45	801,60	753,45	2
3	1162,79	4404,27	3587,78	1162,79	1
4	3416,92	2403,63	1394,47	1394,47	3
5	4474,41	797,41	1418,62	797,41	2
6	4997,88	1203,91	895,47	895,47	3
7	1625,25	3797,91	3045,80	1625,25	1
8	3798,34	1587,91	543,20	543,20	3
9	3020,07	7600,67	6806,23	3020,07	1
10	4012.13	1532,66	670,12	670,12	3
11	2965,47	5465,75	5127,99	2965,47	1
12	5391,97	797,41	1418,54	797,41	2
13	4624,96	812,98	966,38	812,98	2

From the calculation above, the results of the dataset calculation in iteration 2 are obtained. Table 5 can be grouped as follows: Cluster1 consists of 5 items, Cluster2 consists of 4 items and Cluster3 consists of 4 items. If the results of the cluster are changed, the process will be continued to iteration 3. From the calculation with the new cluster center , the results of the dataset calculation in iteration 3 were obtained. The process will continue to repeat from step c until there are no changes to the cluster members for each group.

2. Results and discussion

3.1. Problem identification

Duck egg farmers in Secang District still use traditional methods to determine feed formulations, which often do not meet the specific nutritional needs of the livestock. Feed composition is determined based solely on ingredient availability, leading to inconsistencies. Variations in feed types can significantly affect egg production. The data used were obtained from interviews with three farmers and literature studies adjusted to the availability of feed ingredients in the region. The data consist of 6 attributes: feed name, protein content, fat content, fiber content, calcium content, and phosphorus content.

The following is a table of feed composition data sources:

Table 9. Feed Composition Data Sources

Subject	Year	Background
Composition Of Feed Needed by Laying Ducks	2024	Protein For Growth, Replacement of Dead Tissue, And Antibody Formation 16-18%. P (Phosphorus) Is Needed in Growth, Increasing Appetite, Production and Egg Weight. Ca (Calcium) For Production and Shell Thickness 3.0% (H. Rahmat Rukmana, 2024)
	2020	The Content in The Ration Including Protein, Fat, Carbohydrates, Vitamins, Minerals and Water Should Be Available in Sufficient Quantities. Protein Functions as The Basic Building Block of All Body Tissues as Well as The Material for Making Eggs and Sperm. Fats Function as The Absorption of Vitamins (A, D, E, K), Provide Essential Fatty Acids, Have an Important Effect on Calcium Absorption and Increase Energy Use Efficiency. Vitamins Function as Helpers (Catalysts) in the Process of Forming or Breaking Down Other Nutrients in the Body, So They Are Only Needed in Small Amounts. Minerals Are Needed to Form the Body's Skeleton (Bones), Aid Digestion and Metabolism in Cells as Well as For the Formation of Eggshells(Slamet Prihatin, 2020)
	2022	The Nutritional Needs of Laying Duck Feed Are as Follows, Crude Protein By 16-18%, Crude Fiber 8%, Calcium (Ca) 3.5-4%, Phosphorus (P) 0.3 -0.9% (drh. Karinadintha Marsya Rachman, 2022)

3.2. Data Pre-processing

Feed data were converted into a DataFrame using the Pandas library. Normalization was performed using StandardScaler from Scikit-learn to standardize the scale of all features (mean=0, standard deviation=1), preventing features with larger scales from dominating distance calculations in the K-Means algorithm.

```
# Normalisasi data
scaler = StandardScaler()
scaled_data = scaler.fit_transform(df[['x', 'y', 'z', 'a', 'b']])

# Membuat DataFrame dari hasil normalisasi
df_scaled = pd.DataFrame(scaled_data, columns=['x_scaled', 'y_scaled', 'z_scaled', 'a_scaled', 'b_scaled'])

# Menggabungkan dengan kolom 'Data Ke'
df_final = pd.concat([df['Data Ke'], df_scaled], axis=1)

print("\nHasil Normalisasi (StandardScaler):")
print(df_final)
```

Figure 3. Implementation of Data Normalization

3.3. Centroid Initialization and Clustering Process

Initial centroids were manually determined based on preliminary data analysis (data points 1, 5, and 8). The K-Means algorithm was run with a maximum of 100 iterations. An iterative process was performed until convergence was achieved (centroids did not change between iterations).

```
# Centroid awal
initial_centroids = np.array([
    [4870, 2030, 0, 69.2,52.3],
    [854, 911, 7.6, 975, 567],
    [1700, 300, 1000, 290, 45]
])


max_iter = 100
converged = False
iteration = 0
previous_clusters = None
all_clusters = {}  # Untuk menyimpan cluster tiap iterasi
```

Figure 4. Early Centroid Implementation

3.4. Clustering Result Analysis

Based on the Sum of Squared Errors (SSE) calculation and the Elbow Method, the optimal number of clusters was determined to be 3. The most significant decrease in SSE occurred from k=2 to k=3, after which the decrease became insignificant.

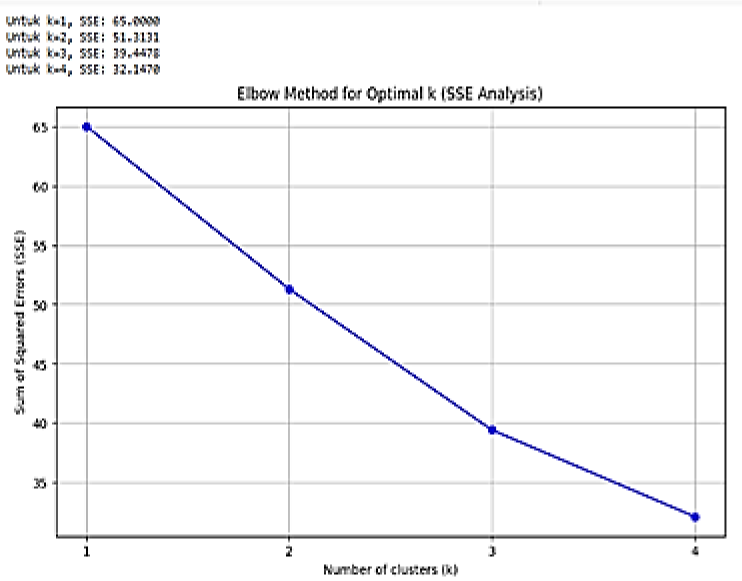

Figure 5. SSE Visualization

From the graph in figure 5 above, it can be seen that for k = 1, the SSE value is 65.0000, for k = 2, the SSE value drops to 51.3131, for k = 3, the SSE value is quite significant to 39.4478, for k = 4, the SSE value becomes 32.1470. The largest decrease in SSE occurred between k = 2 and k = 3. After k = 3, the decrease in SSE still occurs but is no longer significant. The graph begins to form a flatter angle (elbow), which shows that the addition of the number of clusters after k = 3 does not provide a significant advantage in reducing SSE. In this graph, the point occurs at k = 3. Therefore, it can be concluded that the optimal number of clusters for this data is 3.

3.5. Discussion

The resulting clusters have distinct characteristics that align with the nutritional standards for laying ducks.

Table 10. Cluster 1 Optimization of Feed Composition of Laying Ducks

Feed Type	Protein% (x100)	Fat% (x100)	Fiber% (x100)	Calcium % (x100)	Phosphor us % (x100)
Fresh Snails	4870	2030	0	69,2	52,3
Soybean Meal	4627	1498	862	3	820
Soybean	4040	1670	320	22,2	68,2
Blood Flour	8031	76	507	93	790
Fishmeal	4903	471	566	3551	1128

Cluster 1 has high protein (>18%), moderate to high fat, low fiber, and varying mineral content. It is highly suitable as a primary feed to support egg productivity, growth, and duck health (Rachman, 2022). However, it should be mixed with fibrous materials to maintain digestive balance.

Table 11. Cluster 2 Optimization of Feed Composition of Laying Ducks

Feed Type	Protein % (X100)	Fat % (X100)	Fiber % (X100)	Calcium % (X100)	Phosphor Us % (X100)
Brown Rice	936	162	131	0	182
Fine Bran	854	911	7,6	975	567
Corn	550	10	1000	2	9
Cassava Flour	110	50	90	8,4	12,5
Fish Bone Meal	920	4,4	0	73,5	34,5

Cluster 2 has low to moderate nutrient levels. It serves as a source of energy and fiber but must be combined with high-protein feed (Cluster 1) to meet balanced nutritional requirements.

Table 12. Cluster 3 Optimization of Feed Composition of Laying Ducks

Feed Type	Protein % (x100)	Fat % (x100)	Fiber % (x100)	Calcium % (x100)	Phosphor us % (x100)
Coconut Meal	2160	1210	1549	165	21
Laying Duck Feed HDIP3 1218	1700	300	1000	290	45
Lamtoro Leaf Flour	1420	1090	940	8,6	25

Cluster 3 has very high fiber (>9%), moderate protein, and low mineral content. It functions as a digestive regulator and should be mixed with other feeds to balance nutrition.
Recommendations for Farmers:

- Use a combination of feeds from Cluster 1 as the primary protein source.
- Add feeds from Cluster 2 for energy and fiber.
- Use feeds from Cluster 3 as digestive regulators if needed.
- Avoid using feeds solely from Cluster 2 or 3, as they do not meet the nutritional standards for laying ducks.

3.6. GUI Implementation

A Gradio-based interface was developed to help farmers easily upload data, run clustering, and obtain optimal feed recommendations. Export features to Excel and PDF are provided for documentation. Clustering with K-Means successfully grouped feeds based on nutritional similarities. Cluster 1 represents optimal feed, Cluster 2 serves as a supplement, and Cluster 3 acts as a digestive regulator. These recommendations help farmers formulate efficient and nutritionally appropriate feed formulations for their livestock.

3. Conclusion

Based on the analysis of thirteen distinct feed types utilizing the K-Means Clustering algorithm, this study successfully identifies an optimal feed composition for laying ducks, designated as Cluster 1. This cluster is distinguished by its nutritional profile, which most closely aligns with the established dietary requirements for maximizing laying productivity, particularly in its balanced levels of critical components: protein, fiber, fat, calcium, and phosphorus. The determination of three as the optimal number of clusters, validated by the Elbow Method, effectively segments the feed data into distinct qualitative tiers, providing a clear, data-driven rationale for selecting the most advantageous formulation. Thus, this research delivers a concrete answer to its central question, demonstrating that data mining techniques can objectively identify superior feed compositions from a complex nutritional dataset.

When contextualized within the broader field of agricultural informatics, this study aligns with and extends previous research that applies computational methods to precision livestock farming. While earlier studies have often focused on larger livestock or different poultry, this work specifically bridges the gap for duck farming, a vital sector in many agricultural economies. The practical implication for farmers is the potential for a decision-support system that reduces reliance on generic, off-the-shelf feed products. By adopting a cluster-based recommendation, farmers can optimize feed efficiency, potentially lowering costs and enhancing egg production rates. For the field of agricultural information technology, this research underscores the significant role of simple yet powerful algorithms like K-Means in transforming raw data into actionable agricultural intelligence, making advanced analytics accessible even for smaller-scale farming operations.

Despite its promising findings, this study is not without limitations. The primary constraint is the relatively small dataset of only thirteen feed types, which may limit the generalizability and robustness of the clusters. Furthermore, the nutritional variables were restricted to macronutrients and minerals, omitting other crucial micronutrients such as essential amino acids (e.g., lysine, methionine) and vitamins, which are vital for comprehensive duck health and productivity. Future research should therefore prioritize expanding the dataset significantly and incorporating these additional variables to build a more holistic and accurate model. The subsequent direction would involve developing a predictive model that can dynamically recommend feed formulations based on desired output metrics, ultimately leading to controlled field trials to empirically validate the algorithm's recommendations against real-world laying performance.

4. References

Akintan, O., Gebremedhin, K. G., & Uyeh, D. D. (2024). Animal Feed Formulation Connecting Technologies to Build a Resilient and Sustainable System. In Animals (Vol. 14, Issue 10). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/ani14101497

Alva Mustika, F., & Dhika, H. (2022). Sistem Rekomendasi Film Menggunakan Metode K-Means Clustering Berbasis Web. Jurnal SIMETRIS, 13(2).

Annas, M., & Wahab, S. N. (2023). Data Mining Methods: K-Means Clustering Algorithms. International Journal of Cyber and IT Service Management, 3(1), 40–47. https://doi.org/10.34306/ijcitsm.v3i1.122

Ayu, D., Dewi, I. C., & Pramita, K. (2019). Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. JURNAL MATRIX, 9(3), 102.

drh. Karinadintha Marsya Rachman. (2022, June 2). Pakan Bebek Petelur Per 100 Ekor: Jenis, Kebutuhan Nutrisi, Takaran dan Komposisi Pembuatan.

Fitri Yani, N., Elisia, R., & annisa. (2024). Overview Of Production and Development Challenges Of The Layer Ducks Farming Industry In Indonesia And Globally. Jurnal Tropicalanimal, 2. https://doi.org/10.24036/jeta.v2i2.55

Giantika Utami, D. (2022). Penerapan Algoritma K-Means untuk Pengelompokan Produksi Telur Ayam Ras Petelur di Indonesia. 3(1). www.bps.go.id.

H. Rahmat Rukmana. (2024). Panduan Lengkap Ternak Itik Petelur & Pedaging Secara Intensif (seno wibowo, Ed.). Lily Publisher.

Hamouda, R. E., Youssef, I. M., Gharib, H. B., El-Menawey, M. A., Youssif, M. A. M., Osman, M. A., Abdel-Aziz, Y. A., Rudayni, H. A., Allam, A. A., Alawam, A. S., Khafaga, A. F., Abd El-Hack, M. E., & Abdel-Halim, A. A. (2025). Effects of an enhanced feeding model on productivity and sustainability of broilers and hybrid chickens under Egyptian small-scale family systems. Poultry Science, 104(11). https://doi.org/10.1016/j.psj.2025.105845

Javed Mehedi Shamrat, F. M., Tasnim, Z., Mahmud, I., Jahan, N., & Islam Nobel, N. (2020). Application Of K-Means Clustering Algorithm To Determine The Density Of Demand Of Different Kinds Of Jobs. International Journal ff Scientific & Technology Research, 9, 2. www.ijstr.org

Ma, P., Zhang, Z., Li, Y., Yu, N., Sheng, J., Küçük McGinty, H., Wang, Q., & Ahuja, J. K. (2022a). Deep learning accurately predicts food categories and nutrients based on ingredient 1 statements 2.

Ma, P., Zhang, Z., Li, Y., Yu, N., Sheng, J., Küçük McGinty, H., Wang, Q., & Ahuja, J. K. (2022b). Deep learning accurately predicts food categories and nutrients based on ingredient 1 statements 2.

Maori, N. A. (2023). Metode Elbow Dalam Optimasi Jumlah Cluster pada K-Means Clustering. Jurnal SIMETRIS, 14.

Maylawati, D. S., Priatna, T., Sugilar, H., & Ramdhani, M. A. (2020). Data science for digital culture improvement in higher education using K-means clustering and text analytics. International Journal of Electrical and Computer Engineering, 10(5), 4569–4580. https://doi.org/10.11591/IJECE.V10I5.PP4569-4580

Naeem, M., Jia, Z., Wang, J., Poudel, S., Manjankattil, S., Adhikari, Y., Bailey, M., & Bourassa, D. (2025a). Advancements in machine learning applications in poultry farming: a literature review. Journal of Applied Poultry Research, 100602. https://doi.org/10.1016/j.japr.2025.100602

Naeem, M., Jia, Z., Wang, J., Poudel, S., Manjankattil, S., Adhikari, Y., Bailey, M., & Bourassa, D. (2025b). Advancements in machine learning applications in poultry farming: a literature review. Journal of Applied Poultry Research, 100602. https://doi.org/10.1016/j.japr.2025.100602

Nuningtyas, Y. F., Natsir, M. H., Widyastuti, E. S., Sjofjan, O., Susilo, A., Widiati, A. S., & Lestari, S. P. (2023). Laying Hens Growth Performance in the Peak Production Phase Offered Bio-Herbal as a Feed Additive. Jurnal Ilmu-Ilmu Peternakan, 33(1), 25–32. https://doi.org/10.21776/ub.jiip.2023.033.01.04

Rozi, I. F., Firdausi, A. T., Rahmadhany, T. R., Studi, P., Informatika, T., Informasi, J. T., & Malang, P. N. (n.d.). JIP (Jurnal Informatika Polinema) Penentuan Bahan Makanan Untuk Itik Petelur Menggunakan Algoritma Genetika.

Slamet Prihatin. (2020, December 19). ITIK Mojosari. https://bukanrobot.home.blog/2020/12/19/itik-mojosari/?utm_source=chatgpt.com

Tukiyat, Anggai, S., & Agnia Bilqisti. (2024). Analysis of Broiler Chicken Production Success Classification Using K-Nearest Neighbors And Naive Bayes Methods at PT. Jandela Jaga Kaloka (Jajaka). Digitus : Journal of Computer Science Applications, 2(4), 158–182. https://doi.org/10.61978/digitus.v2i4.396

Zhang, S., Lai, C., Zhao, J., & Wang, J. (2025a). Big Data and AI-Powered Modeling: A Pathway to Sustainable Precision Animal Nutrition. In Advanced Science. John Wiley and Sons Inc. https://doi.org/10.1002/advs.202507564

Zhang, S., Lai, C., Zhao, J., & Wang, J. (2025b). Big Data and AI-Powered Modeling: A Pathway to Sustainable Precision Animal Nutrition. In Advanced Science. John Wiley and Sons Inc. https://doi.org/10.1002/advs.202507564