



Multivariate Analysis of Livestock Feed Availability Using K-Means Clustering for Regional Potential Mapping in Pati Regency

Ridwan Potabuga¹, Fajar Husain Asy'ari^{2*}, Ellen Proborini³

¹ Information Technology Study Program, Institut Kesehatan dan Teknologi Graha Medika, Kotamobagu, North Sulawesi, Indonesia

^{2,3} Informatics Study Program, Sekolah Tinggi Teknik Pati, Pati, Central Java, Indonesia

ABSTRACT

Livestock feed availability plays a crucial role in livestock production and regional food security. This study maps the potential of 21 sub-districts in Pati Regency based on livestock feed availability from 2020 to 2024 using multivariate K-Means clustering. Secondary data from the Pati Regency Agriculture Office were analysed as annual feed availability (tons/year), which served as the quantitative basis for assessing livestock carrying capacity. The analytical workflow included missing-value checking, outlier detection, data standardisation, principal component analysis (PCA), and selection of the cluster solution using the Silhouette Coefficient. Although $k = 2$ produced the highest overall silhouette score, it isolated Margoyoso as a single extreme observation; among the non-degenerate solutions ($k \geq 3$), $k = 3$ produced the highest silhouette score (0.4383). The three-cluster solution identified moderate availability (Cluster 1), very high and fluctuating availability (Cluster 2), and high and relatively stable availability (Cluster 3). A sensitivity check excluding Margoyoso indicated that the moderate and high-stable grouping patterns remained broadly unchanged. These results show that K-Means clustering can provide a structured basis for regional feed-availability mapping and support differentiated livestock-development planning.

KEYWORDS: K-Means clustering; livestock feed availability; carrying capacity; regional potential; Pati Regency

*Corresponding author: Fajar Husain Asy'ari, fajarhusain@sttp.ac.id

Article History: Received: 2026-08-12; Accepted: 2026-09-25

1. INTRODUCTION

Livestock feed availability is a fundamental component of agricultural and livestock systems, as it directly influences livestock productivity, production efficiency, and regional food security. Adequate feed resources ensure that livestock populations can be maintained sustainably without causing excessive pressure on land and environmental resources. This concept is commonly referred to as carrying capacity, which reflects the ability of a region to support livestock activities based on its available feed resources. In many agricultural regions, carrying capacity varies considerably because of differences in land-use patterns, crop production systems, climatic conditions, and agricultural management practices. In Pati Regency, Central Java, livestock feed availability from 2020 to 2024 varies markedly among sub-districts, with some areas experiencing rapid growth in feed capacity while others remain relatively stagnant or show only moderate development [1]. These disparities indicate unequal regional potential and highlight the need for systematic, data-driven analysis to support effective livestock development planning [2], [3].

As demands for efficiency, sustainability, and evidence-based policy-making in agriculture continue to increase, regional potential mapping based on livestock feed availability has become increasingly important. Accurate identification of regional characteristics enables governments and stakeholders to design targeted interventions, optimise resource allocation, and minimise inefficiencies in livestock production systems. Traditional descriptive approaches are often insufficient to capture complex patterns within multivariate agricultural data. Data-driven analytical methods are therefore required to classify regions objectively according to their similarities and differences. Multivariate clustering provides a robust solution for grouping regions with similar characteristics, allowing policymakers to understand spatial patterns and prioritise development strategies [4], [5].

Among various clustering methods, the K-Means algorithm is one of the most widely used because of its simplicity, computational efficiency, and suitability for numerical data such as livestock feed carrying-capacity indicators [6], [7]. K-Means groups observations based on centroid distances, which makes it effective for identifying

homogeneous regional clusters when it is applied to standardised quantitative data. To ensure meaningful results, K-Means is often combined with cluster-validity techniques such as the Silhouette Coefficient, which evaluates cluster separation and cohesion [8]. This combination allows researchers to determine the optimal number of clusters and to improve the interpretability of the clustering outcome.

Recent studies have demonstrated the use of clustering in agricultural and food-system analysis. K-Means has been applied in precision agriculture for crop classification, land-suitability assessment, and agricultural resource optimisation [2], [9], while crop-selection and recommendation studies have combined clustering with machine-learning approaches [10]. Recent work also includes federated K-Means for rice cultivar clustering [23] and K-Means with principal component analysis for plantation-crop production in Indonesia [24]. Methodological studies have addressed incomplete, noisy, and high-dimensional agricultural data [11], [12], whereas other applications have examined food distribution and food security [3], [13], livestock characteristics and farm management [14], [15], and regional welfare or agricultural indicators using validity indices such as the Silhouette Coefficient [25], [26]. Together, these studies illustrate the broad use of clustering for heterogeneous agricultural data while underscoring the importance of context-specific validation.

Despite the growing body of literature on clustering applications in agriculture and livestock systems, few studies have specifically classified sub-districts according to livestock feed availability in the context of carrying capacity, particularly at the local administrative level of a regency. Most existing studies emphasise crop production, national or provincial food security, or livestock performance indicators rather than feed-resource availability at the sub-district scale. This study therefore performs multivariate clustering of the sub-districts of Pati Regency based on livestock feed availability across multiple observation periods. The objectives are to identify substantively distinct feed-availability profiles and to provide a regional potential map that can serve as an analytical basis for local policy formulation in livestock development and feed-distribution planning. The findings are expected to help the Pati Regency government design more targeted and sustainable strategies for livestock-sector development while also contributing to the literature on data-driven regional analysis in agriculture [2], [3].

2. METHODOLOGY

This study employs a quantitative approach using secondary data obtained from the Pati Regency Agriculture Office. The dataset consists of livestock feed availability (carrying capacity) measured in tons for each sub-district of Pati Regency over the period 2020–2024. A total of 21 sub-districts are included in the analysis. The data are analysed using the K-Means clustering algorithm to classify sub-districts according to similarities in livestock feed availability [6], [16]. The overall research procedure is illustrated in Figure 1.

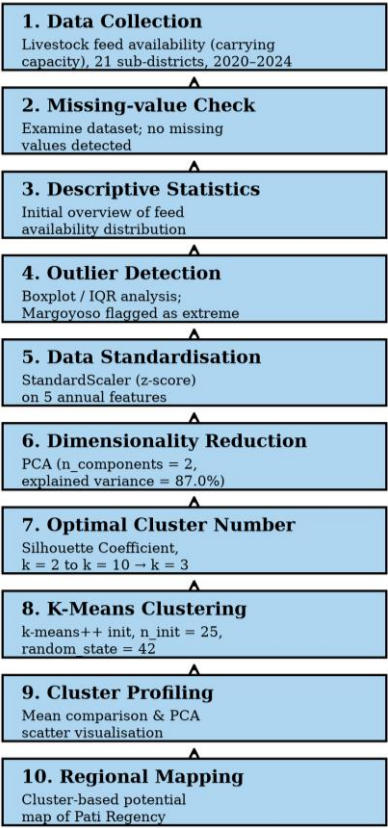


Figure 1. Research workflow of the proposed K-Means clustering methodology

The dataset comprises five annual features for each sub-district. The variables used in the analysis are defined in Table 1.

Table 1. Definition of the variables used in the study

Variable	Description	Unit
X1	Feed availability (annual feed-supply capacity basis for carrying capacity), 2020	tons/year
X2	Feed availability (annual feed-supply capacity basis for carrying capacity), 2021	tons/year
X3	Feed availability (annual feed-supply capacity basis for carrying capacity), 2022	tons/year
X4	Feed availability (annual feed-supply capacity basis for carrying capacity), 2023	tons/year
X5	Feed availability (annual feed-supply capacity basis for carrying capacity), 2024	tons/year

2.1. Research Design

The research procedure follows several sequential stages [4]. First, the livestock feed availability data for each sub-district are collected and prepared for analysis; this stage includes data compilation and formatting to ensure compatibility with the clustering process. Second, the dataset is examined for missing values. When missing values are detected, they are replaced with the mean of the corresponding variable to maintain completeness and consistency. Third, a descriptive statistical analysis is conducted to obtain an initial overview of the distribution of feed availability across sub-districts.

The next stage involves outlier detection to identify extreme values that may influence the clustering result [12]. Detected outliers are retained in the main analysis and evaluated through a sensitivity check because they may represent genuine regional variation rather than data errors. Data standardisation is then performed so that all variables contribute equally to the clustering process. Because the five annual features are strongly collinear, the standardised data are subsequently projected onto the first two principal components, which jointly explain 87.0% of the total variance, before clustering; this step reduces redundancy among the yearly variables and yields a two-dimensional representation of the sub-districts [21]. The number of clusters is assessed using the Silhouette Coefficient [8], and the selected solution is interpreted in relation to the study objective and the effect of extreme observations [17]. Finally, the K-Means algorithm is applied to group the sub-districts, and the mean values of each cluster are calculated to characterise the resulting groups. The sensitivity analysis for the most extreme observation is reported in Section 3.3, followed by regional potential mapping in Section 3.4.

2.2. K-Means Clustering Algorithm

K-Means clustering partitions data points into K clusters by minimizing the distance between each data point and the centroid of its assigned cluster [6], [18]. For a dataset with p features per observation, the distance between a data point x_i and a centroid c_k is calculated using the Euclidean distance:

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^p (x_{ij} - c_{kj})^2} \quad (1)$$

where $d(x_i, c_k)$ is the Euclidean distance, p is the number of features, x_{ij} is the value of feature j for observation i , and c_{kj} is the value of feature j for centroid k . The algorithm minimises the within-cluster sum of squared errors:

$$SSE = \sum_{k=1}^K \sum_{x_i \in C_k} |x_i - c_k|^2 \quad (2)$$

where C_k is the set of observations assigned to cluster k and K is the number of clusters. Each centroid is updated iteratively as the mean of the observations assigned to it:

$$c_k = \frac{1}{|C_k|} \sum (x_i \in C_k, x_i) \quad (3)$$

The initial centroids are selected using the k-means++ strategy to improve convergence and solution quality [19]. Each data point is assigned to the cluster with the nearest centroid, and the centroid positions are updated iteratively until convergence is achieved [7].

2.3. Tools and Test Design

Data processing and clustering are performed in Google Colaboratory (Google Colab) using Python 3.11 and the scikit-learn library (version 1.5.2), which supports data cleaning, statistical computation, clustering, and visualisation [20]. To ensure full reproducibility, all key analytical settings are reported here. Standardisation uses StandardScaler (zero mean and unit variance per feature). Dimensionality reduction uses PCA with $n_components = 2$ and $svd_solver = 'full'$, which retains 87.0% of the total variance ($PC1 = 59.2\%$, $PC2 = 27.8\%$). The K-Means implementation uses the following parameters: $init = 'k-means++'$, $n_init = 25$, $max_iter = 300$, $tol = 1e-4$, $algorithm = 'lloyd'$, and $random_state = 42$. Cluster validity is assessed using the Silhouette Coefficient [8], while cluster characteristics are analysed through mean comparisons and the two-dimensional principal component analysis (PCA) projection [21]. The resulting clusters are interpreted to identify regional patterns of livestock feed availability and to support strategic planning for livestock development in Pati Regency.

3. RESULTS AND DISCUSSION

3.1. Data Description and Preprocessing

The data used in this study were obtained from the Pati Regency Agriculture Office and accessed through the Open Data Pati Regency portal (<https://opendata.patikab.go.id/>) [1]. The dataset contains the livestock feed availability (carrying capacity) of 21 sub-districts in Pati Regency from 2020 to 2024, as presented in Table 2. The data indicate variation in feed availability across sub-districts and observation periods, reflecting differences in regional agricultural potential and feed-resource distribution.

Table 2. Carrying capacity of livestock feed in Pati Regency, 2020–2024 (tons)

ID	Sub-district	2020	2021	2022	2023	2024
1	Sukolilo	24,763.40	19,148.15	19,111.94	19,489.53	21,411.59
2	Kayen	17,600.58	8,980.61	9,256.57	9,256.57	9,164.19
3	Tambakromo	24,948.29	11,223.93	10,912.22	10,912.22	10,821.42
4	Winong	13,141.06	10,657.22	11,057.40	11,057.40	10,918.24
5	Pucakwangi	10,552.71	8,109.21	9,909.16	10,435.07	8,607.74
6	Jaken	5,591.68	10,060.12	7,724.63	7,724.63	7,140.85
7	Batangan	1,695.53	7,896.69	3,004.42	3,004.42	3,050.42
8	Juwana	1,851.96	2,967.55	1,830.65	1,830.65	2,004.60
9	Jakenan	11,550.83	1,680.95	4,933.48	4,933.48	4,676.11
10	Pati	3,649.12	4,485.77	4,814.91	4,814.91	4,627.33
11	Gabus	10,632.56	4,002.28	6,096.51	6,096.51	5,673.61
12	Margorejo	8,629.21	5,523.35	9,242.72	9,242.72	9,960.46
13	Gembong	24,948.29	8,404.15	12,386.45	12,386.45	13,581.26
14	Tlogowungu	2,824.93	11,023.30	17,232.44	17,232.44	23,121.17
15	Wedarijaksa	3,516.22	17,381.66	4,019.72	4,019.72	4,234.81
16	Trangkil	1,331.82	3,695.89	4,835.54	4,835.54	5,679.07
17	Margoyoso	1,590.12	4,390.95	40,547.84	40,547.84	84,999.04
18	Gunungwungkal	2,131.33	6,985.46	14,835.20	14,835.20	11,172.65
19	Cluwak	3,575.77	9,317.87	9,494.49	9,494.49	12,792.87
20	Tayu	2,521.13	9,714.00	4,647.17	4,647.17	5,434.63
21	Dukuhseti	3,126.42	4,363.21	4,962.08	4,962.08	6,499.34
	Total	180,172.96	170,012.32	210,855.54	211,759.04	265,571.40

The initial step in data processing was a missing-value check. Based on the examination of the dataset in Table 2, no missing values were identified, so the data could be used directly without imputation or additional cleaning. The result of the missing-value check is shown in Figure 2, and Figure 3 presents a visual confirmation of the completeness of the dataset.

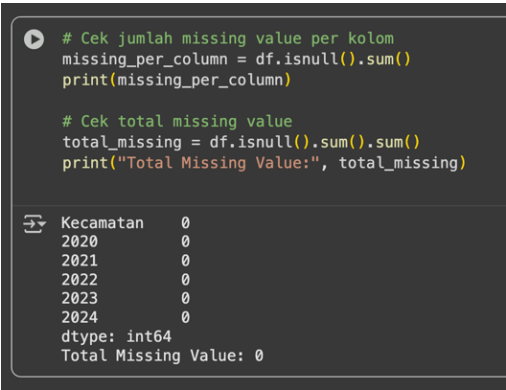


Figure 2. Missing-value check output of the dataset

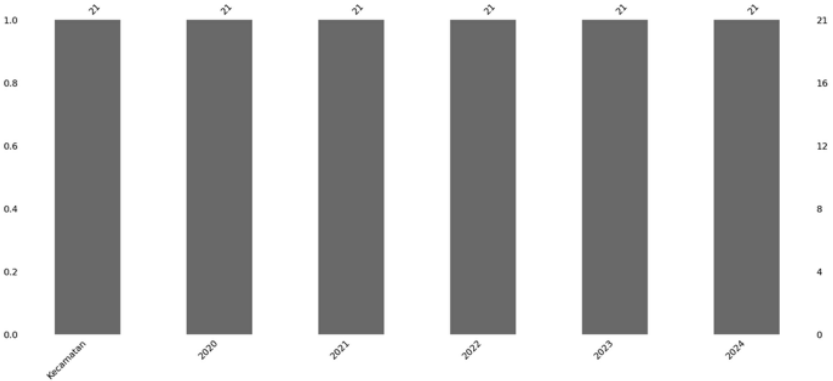


Figure 3. Visual representation confirming the completeness of the dataset

A descriptive statistical analysis was then conducted to provide an initial overview of feed availability across sub-districts. Outlier detection was performed using boxplot and interquartile-range (IQR) analysis to identify

extreme values that could influence the clustering result, as shown in Figure 4. Using the $1.5 \times \text{IQR}$ criterion, one outlier is observed in 2021 (Sukolilo), Margoyoso is flagged as an outlier in 2022 and 2023, and in 2024 three sub-districts exceed the upper fence (Sukolilo, Tlogowungu, and Margoyoso), with Margoyoso reaching 84,999.04 tons—the most extreme value in the dataset. These observations correspond to sub-districts with substantially higher feed availability than the others and therefore require careful interpretation. Because extreme observations can dominate distance-based clustering, the sensitivity analysis reported in Section 3.3 examines the effect of excluding Margoyoso from the clustering interpretation.

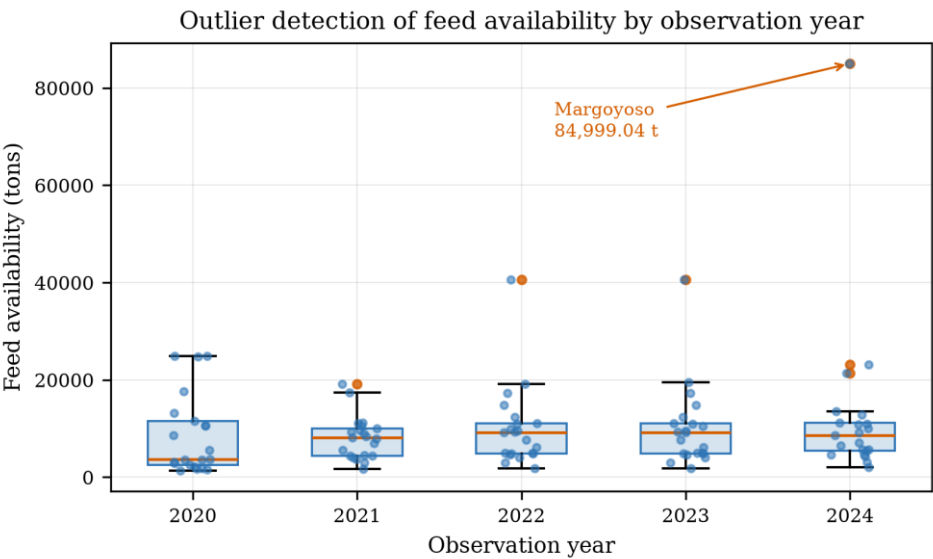


Figure 4. Outlier detection of livestock feed availability across observation years

3.2. K-Means Clustering Results

To determine the number of clusters for interpretation, the Silhouette Coefficient was computed for candidate solutions from $k = 2$ to $k = 10$, as shown in Figure 5. The scores are $k = 2$ (0.7174), $k = 3$ (0.4383), $k = 4$ (0.3832), $k = 5$ (0.4082), $k = 6$ (0.3778), $k = 7$ (0.3668), $k = 8$ (0.3641), $k = 9$ (0.3500), and $k = 10$ (0.3248). The highest overall score occurs at $k = 2$; however, this two-cluster solution isolates Margoyoso from the remaining 20 sub-districts and is therefore substantively uninformative for the study objective of identifying multiple regional profiles. Among the candidate solutions with $k \geq 3$, the silhouette score is highest at $k = 3$ (0.4383). The $k = 3$ solution is therefore retained for interpretation because it has the highest silhouette score among these solutions and yields three substantively distinct feed-availability profiles. Figure 6 presents the average silhouette value of the selected solution.



Figure 5. Silhouette Coefficient scores for $k = 2$ to $k = 10$; the selected solution ($k = 3$) is highlighted

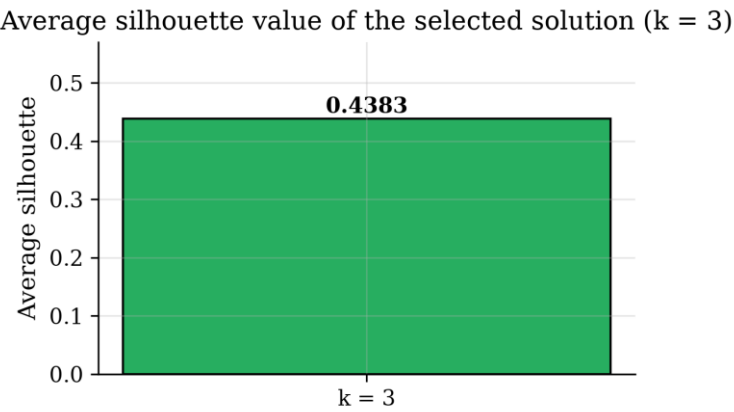


Figure 6. Average silhouette value of the selected clustering solution (k = 3)

The average livestock feed availability of each cluster from 2020 to 2024 is presented in Table 3 and visualised in Figure 7. Cluster 2 shows highly fluctuating feed availability, Cluster 1 maintains a moderate level, and Cluster 3 demonstrates high and relatively stable feed availability throughout the observation period. These patterns highlight distinct regional characteristics in feed-resource distribution.

Table 3. Average feed availability of each cluster, 2020–2024 (tons)

Year	Cluster 1	Cluster 2	Cluster 3
2020	5,141.54	1,590.12	15,963.25
2021	6,061.72	4,390.95	12,402.72
2022	6,640.84	40,547.84	11,996.68
2023	6,681.30	40,547.84	12,050.62
2024	6,716.90	84,999.04	13,321.81

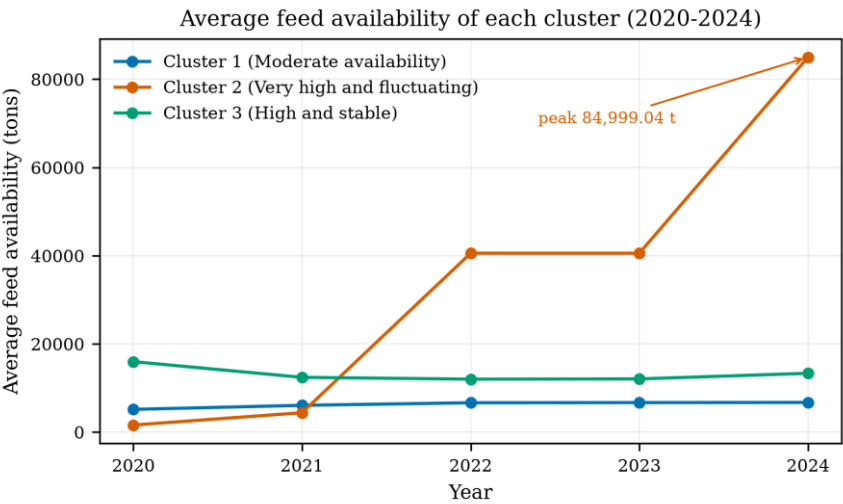


Figure 7. Average feed availability of each cluster, 2020–2024]

A further visualisation using two-dimensional principal component analysis (PCA) is shown in Figure 8, in which each point represents a sub-district and is coloured according to its cluster membership. The PCA projection shows separation among the three profiles: the moderate cluster forms the largest group, the very-high-and-fluctuating cluster is represented solely by Margoyoso, and the high-and-stable cluster forms a separate group between the moderate observations and the extreme Margoyoso observation in the two-dimensional space [21].

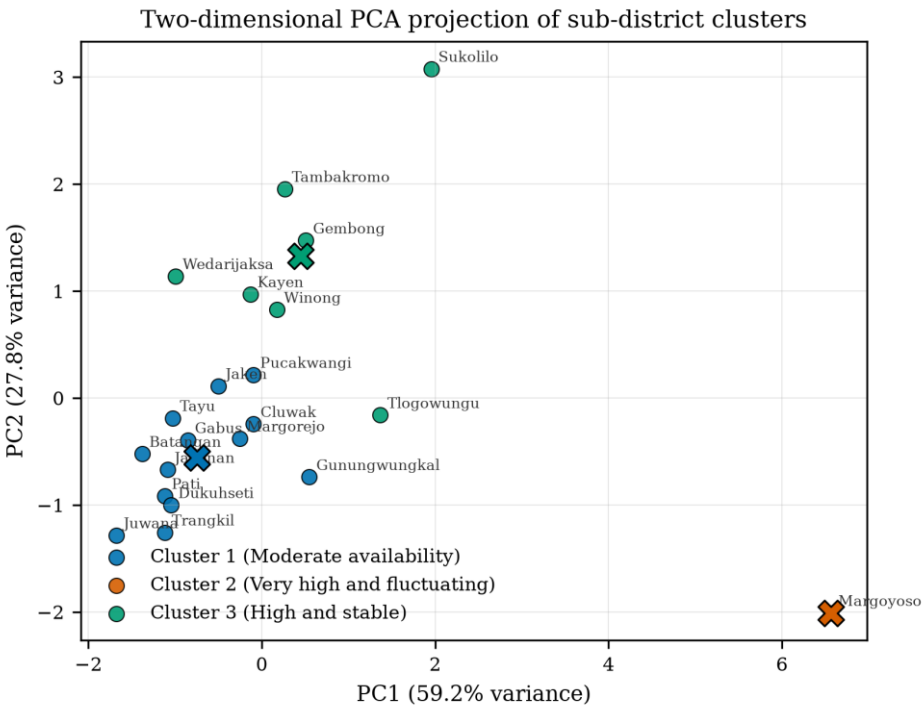


Figure 8. Two-dimensional PCA projection of sub-district clusters based on feed availability

The final cluster membership of each sub-district is summarised in Table 4. Cluster 1 contains 13 sub-districts (Pucakwangi, Jaken, Batangan, Juwana, Jakenan, Pati, Gabus, Margorejo, Trangkil, Gunungwungkal, Cluwak, Tayu, and Dukuhseti) with moderate feed availability. Cluster 2 consists only of Margoyoso, which is characterised by very high and fluctuating feed availability. Cluster 3 contains seven sub-districts (Sukolilo, Kayen, Tambakromo, Winong, Gembong, Tlogowungu, and Wedarijaksa) with high and relatively stable feed availability. The distribution of sub-districts across the three clusters is shown in Figure 9.

Table 4. Sub-district membership by cluster in Pati Regency

Cluster	Category	Sub-districts
1	Moderate availability	Pucakwangi, Jaken, Batangan, Juwana, Jakenan, Pati, Gabus, Margorejo, Trangkil, Gunungwungkal, Cluwak, Tayu, Dukuhseti
2	Very high and fluctuating	Margoyoso
3	High and stable	Sukolilo, Kayen, Tambakromo, Winong, Gembong, Tlogowungu, Wedarijaksa

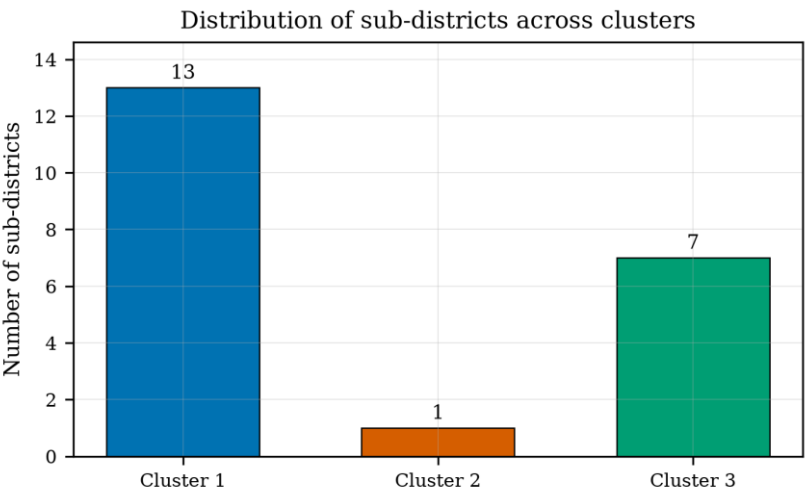


Figure 9. Distribution of sub-districts across the three clusters

3.3. Sensitivity Analysis

Because Margoyoso is the only member of Cluster 2 and has the most extreme values in 2022–2024, a sensitivity check was conducted by excluding this sub-district and reassessing the remaining grouping pattern. The comparison indicates that the moderate and high-stable groups remain substantively unchanged after the exclusion, showing that the distinction between these two broader regional profiles is not solely produced by Margoyoso. The single-member very-high-and-fluctuating cluster should therefore be interpreted as an outlier-driven profile rather than as evidence of a broadly shared regional pattern.

3.4. Cluster Profiling and Regional Potential Mapping

To provide a clearer spatial interpretation of the clustering result, the distribution of livestock feed availability across sub-districts is further visualised as a regional cluster map in Figure 10. The map shows that the sub-districts with moderate availability are concentrated in the central and eastern parts of the regency, the high and stable group lies mainly in the western and northern areas, and Margoyoso forms a distinct high-fluctuation zone.

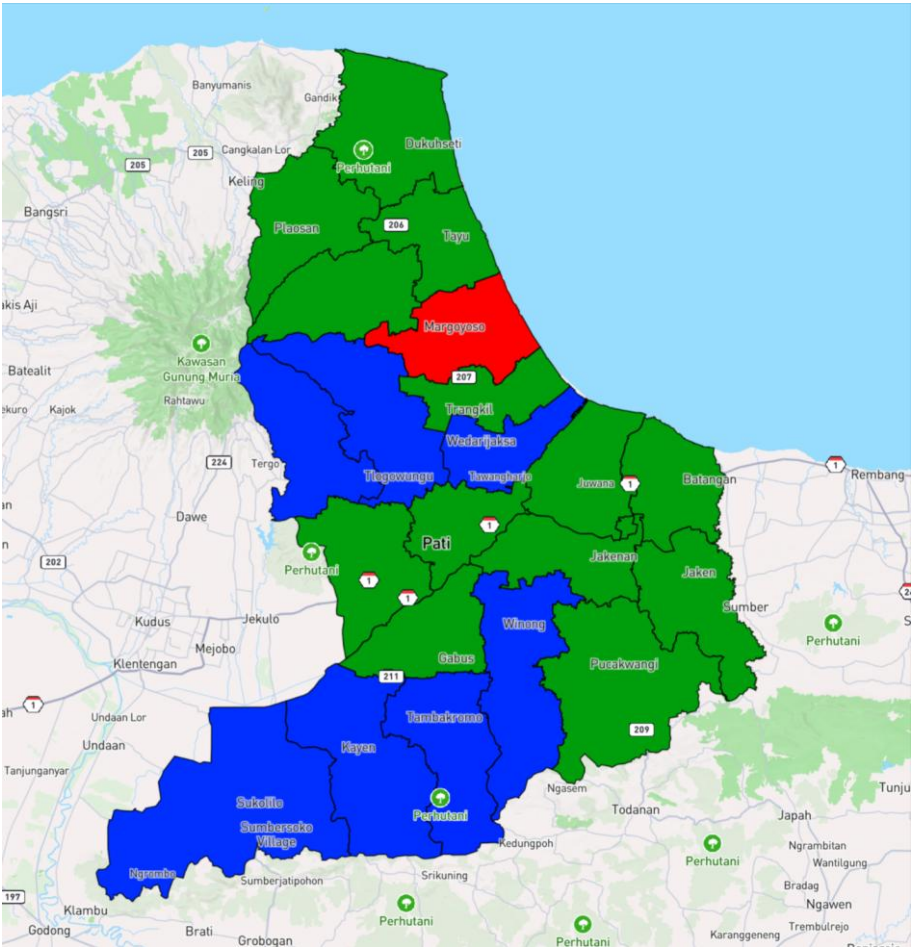


Figure 10. Regional cluster map of livestock feed availability (carrying capacity) in Pati Regency (k = 3)

Overall, the clustering results reveal distinct regional patterns of livestock feed availability across the sub-districts of Pati Regency and provide an analytical basis for formulating targeted livestock development strategies, optimizing feed-resource allocation, and supporting evidence-based policy decisions at the local level [3], [14]. The cluster structure is consistent with recent empirical studies that applied K-Means with validity-index-based cluster selection to regional data in Indonesia: Sari et al. [25] used PCA followed by K-Means for district typology, while Tacharri et al. [26] used K-Means with silhouette evaluation to identify priority areas in Central Java. The concentration of feed resources in a small number of regions also parallels spatial concentration reported in livestock carrying-capacity assessments in other settings [27], [28]. Likewise, the temporal instability of the Margoyoso profile indicates that feed-supply capacity can change markedly across observation periods, reinforcing the need for periodic re-assessment [29]. The combined K-Means and PCA workflow also follows a strategy used for plantation-crop production clustering, in which PCA was applied to collinear annual production variables before K-Means partitioning [24]. Taken together, these comparisons indicate that data-driven clustering can support agricultural and rural planning when regional heterogeneity is high [9], [15].

4. CONCLUSION

This study applied K-Means clustering to analyse livestock feed availability (carrying capacity) across the sub-districts of Pati Regency during 2020–2024. Three clusters were retained because k = 3 had the highest silhouette score among the non-degenerate solutions considered for interpretation (0.4383) and yielded distinct regional profiles: moderate availability, very high and fluctuating availability, and high but relatively stable availability. The clustering results show clear spatial heterogeneity in feed resources across the regency. The sensitivity analysis further indicates that Margoyoso's extreme values drive the separation of the single-member very-high-and-fluctuating cluster, whereas the moderate and high-stable grouping patterns remain broadly unchanged when Margoyoso is excluded.

The findings indicate that multivariate K-Means clustering can provide a structured approach for mapping regional potential based on livestock feed availability. The resulting classification offers an analytical framework for

differentiated livestock-development planning. Sub-districts with moderate feed availability may require targeted interventions to improve feed capacity and production efficiency; areas with high and relatively stable availability may support sustainable livestock development; and areas with very high but fluctuating availability require adaptive resource management to address temporal variability.

From a practical perspective, this study provides a data-driven basis for local governments and stakeholders to design differentiated livestock-development strategies. The cluster-based mapping can support feed-resource allocation, regional planning, and the identification of sub-districts that require different management responses.

Future research may extend this study by incorporating additional variables such as livestock population size, land-use patterns, climatic factors, or socioeconomic indicators to provide a more comprehensive assessment of regional livestock capacity. Alternative or hybrid clustering methods, such as density-based or fuzzy clustering approaches [22], [11], or robust partitioning methods such as K-Medoids, whose relative performance against K-Means has been evaluated using silhouette scores [30], could also be explored under a common evaluation protocol. Such extensions would allow researchers to assess the robustness of regional classifications and further examine the role of data analytics in sustainable livestock development and agricultural planning.

ACKNOWLEDGMENT

The authors thank the Pati Regency Agriculture Office for providing the open data used in this study.

DECLARATIONS

Artificial Intelligence Use

During the preparation of this work, the authors used generative artificial-intelligence tools (a large-language-model assistant) to assist with English language editing and with debugging of the clustering code. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Author Contributions

All authors contributed equally to the conceptualisation, methodology, software implementation, validation, formal analysis, data curation, preparation of the original draft, and review and editing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Ethics Approval / Ethical Clearance

Not applicable. This study used publicly available secondary aggregate data and did not involve human participants or animals.

Data Availability Statement

The dataset is available from the Open Data Pati Regency portal at <https://opendata.patikab.go.id/> and from the Pati Regency Agriculture Office.

REFERENCES

- [1] Pati Regency Agriculture Office, "Open data on livestock carrying capacity in Pati Regency (2020–2024)," Open Data Pati Regency, 2024. [Online]. Available: <https://opendata.patikab.go.id/>
- [2] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, art. no. 2674, Aug. 2018, doi: [10.3390/s18082674](https://doi.org/10.3390/s18082674).
- [3] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," *Comput. Electron. Agric.*, vol. 143, pp. 23–37, Dec. 2017, doi: [10.1016/j.compag.2017.09.037](https://doi.org/10.1016/j.compag.2017.09.037).
- [4] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999, doi: [10.1145/331499.331504](https://doi.org/10.1145/331499.331504).
- [5] X. Wu et al., "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, Jan. 2008, doi: [10.1007/s10115-007-0114-2](https://doi.org/10.1007/s10115-007-0114-2).
- [6] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, Berkeley, CA, USA, 1967, pp. 281–297.
- [7] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002, doi: [10.1109/TPAMI.2002.1017616](https://doi.org/10.1109/TPAMI.2002.1017616).
- [8] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987, doi: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- [9] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agric.*, vol. 147, pp. 70–90, Apr. 2018, doi: [10.1016/j.compag.2018.02.016](https://doi.org/10.1016/j.compag.2018.02.016).
- [10] K. P. Swain, S. R. Nayak, V. Ravi, S. Mishra, T. J. Alahmadi, P. Singh, and M. Diwakar, "Empowering crop selection with ensemble learning and K-means clustering: A modern agricultural perspective," *The Open Agriculture Journal*, vol. 18, no. 1, 2024, doi: [10.2174/0118743315291367240207093403](https://doi.org/10.2174/0118743315291367240207093403).
- [11] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: [10.1016/j.patrec.2009.09.011](https://doi.org/10.1016/j.patrec.2009.09.011).

- [12] C. C. Aggarwal, "Outlier analysis," in *Data Mining: The Textbook*. Cham, Switzerland: Springer, 2015, pp. 237–263, doi: [10.1007/978-3-319-14142-8_8](https://doi.org/10.1007/978-3-319-14142-8_8).
- [13] Y. Liu, X. Ma, L. Shu, G. P. Hancke, and A. M. Abu-Mahfouz, "From Industry 4.0 to Agriculture 4.0: Current status, enabling technologies, and research challenges," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4322–4334, Jun. 2021, doi: [10.1109/TII.2020.3003910](https://doi.org/10.1109/TII.2020.3003910).
- [14] A. Mucherino, P. J. Papajorgji, and P. M. Pardalos, *Data Mining in Agriculture*. New York, NY, USA: Springer, 2009.
- [15] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, "Enhancing smart farming through the applications of Agriculture 4.0 technologies," *Int. J. Intell. Netw.*, vol. 3, pp. 150–164, 2022, doi: [10.1016/j.ijin.2022.09.004](https://doi.org/10.1016/j.ijin.2022.09.004).
- [16] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.
- [17] G. Hamerly and C. Elkan, "Learning the k in k-means," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, 2003, pp. 281–288.
- [18] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982, doi: [10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489).
- [19] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, New Orleans, LA, USA, 2007, pp. 1027–1035.
- [20] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [21] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer, 2002.
- [22] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, Portland, OR, USA, 1996, pp. 226–231.
- [23] A. K. Gaur, J. A. Valan, H. Verma, and H. Kalita, "Rice cultivar clustering using federated K-means: focusing on advancing agriculture 4.0 applications," *Genetic Resources and Crop Evolution*, vol. 72, no. 5, pp. 5549–5570, Dec. 2024, doi: [10.1007/s10722-024-02277-9](https://doi.org/10.1007/s10722-024-02277-9).
- [24] A. G. Valdano and Styawati, "Analysis and clustering of plantation crop production in Indonesia using K-Means clustering and principal component analysis methods," *SISTEMASI*, vol. 15, no. 8, p. 3000, Aug. 2026, doi: [10.32520/stmsi.v15i8.6534](https://doi.org/10.32520/stmsi.v15i8.6534).
- [25] J. R. Sari, F. Fahira, L. Zahra, A. Fitrianto, and K. Alifviansyah, "Identification of latent dimensions of digital readiness and typology of districts/cities in Indonesia using PCA and K-Means clustering," *Journal of Applied Informatics and Computing*, vol. 9, no. 6, pp. 2937–2949, Dec. 2025, doi: [10.30871/jaic.v9i6.11487](https://doi.org/10.30871/jaic.v9i6.11487).
- [26] C. Tacharri, A. Rohmani, and A. Fahmi, "Strategic clustering of poverty areas in Central Java using K-Means and silhouette evaluation," *Sinkron*, vol. 9, no. 2, pp. 895–904, May 2025, doi: [10.33395/sinkron.v9i2.14734](https://doi.org/10.33395/sinkron.v9i2.14734).
- [27] M. Lessmann, A. Kanellopoulos, J. Kros, F. Orsi, and M. Bakker, "A spatially explicit assessment on the carrying capacity of livestock under minimum feed imports and artificial fertilizer use in Dutch agriculture," *Agricultural Systems*, vol. 220, art. 104092, Oct. 2024, doi: [10.1016/j.agsy.2024.104092](https://doi.org/10.1016/j.agsy.2024.104092).
- [28] Y. Zhao, W. Song, Z. Wu, and Y. Bai, "Spatial and temporal analysis of reasonable livestock carrying capacity and ecological carrying capacity of grasslands in pastoral areas of China," *Chinese Science Bulletin*, vol. 69, Aug. 2024, doi: [10.1360/tb-2024-0346](https://doi.org/10.1360/tb-2024-0346).
- [29] C. Duku, G. T. Diro, T. Demissie, and S. Dawit, "Climate change impacts livestock carrying capacity in East Africa," *Regional Environmental Change*, vol. 25, no. 3, Aug. 2025, doi: [10.1007/s10113-025-02440-7](https://doi.org/10.1007/s10113-025-02440-7).
- [30] M. Arifin, I. W. D. Prastya, and J. R. Budiani, "Evaluating K-Means and K-Medoids using silhouette score for Eysenck personality-based clustering of prospective students," *Journal of Applied Informatics and Computing*, vol. 10, no. 2, pp. 1818–1827, Apr. 2026, doi: [10.30871/jaic.v10i2.12338](https://doi.org/10.30871/jaic.v10i2.12338).