# 6. 3504-12201-2-Proofreading 107-118.pdf

Institut Teknologi Dirgantara Adisutjipto

## Document Details

Submission ID

**trn:oid:::3618:124587904**

Submission Date

**Dec 16, 2025, 10:43 AM GMT+7**

Download Date

**Dec 16, 2025, 11:39 AM GMT+7**

File Name

**6. 3504-12201-2-Proofreading 107-118.pdf**

File Size

**1.5 MB**

**12 Pages**

**4,902 Words**

**28,208 Characters**

# 14% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

## Filtered from the Report

- Bibliography
- Quoted Text

## Match Groups

**51** Not Cited or Quoted 13%
Matches with neither in-text citation nor quotation marks

**4** Missing Quotations 1%
Matches that are still very similar to source material

**0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation

**0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

## Top Sources

10% Internet sources
5% Publications
8% Submitted works (Student Papers)

## Integrity Flags

**0 Integrity Flags for Review**

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

## Match Groups

- **51** Not Cited or Quoted 13%
  Matches with neither in-text citation nor quotation marks
- **4** Missing Quotations 1%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

## Top Sources

10% Internet sources
5% Publications
8% Submitted works (Student Papers)

## Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

| | | | |
|---|---|---|---|
| 1 | Internet | ejournals.itda.ac.id | 3% |
| 2 | Internet | www.mdpi.com | 2% |
| 3 | Internet | www.ijisae.org | 1% |
| 4 | Submitted works | Syntax Corporation on 2025-11-14 | <1% |
| 5 | Internet | www.ijoaia.com | <1% |
| 6 | Internet | journal-isi.org | <1% |
| 7 | Internet | ioinformatic.org | <1% |
| 8 | Publication | Andrian Syahputra, Muhammad Ikhsan. "Data Visualization using Tableau With K... | <1% |
| 9 | Submitted works | KCA University on 2023-07-23 | <1% |
| 10 | Internet | ejournal.sultanpublisher.com | <1% |

11 Internet

web.realinfo.tv <1%

12 Internet

swdc.govt.nz <1%

13 Submitted works

Asia Pacific Instutute of Information Technology on 2025-09-03 <1%

14 Submitted works

University of East Anglia on 2022-01-16 <1%

15 Internet

bexpertindoconsult.com <1%

16 Internet

cms.dryvit.com <1%

17 Submitted works

University of New South Wales on 2023-05-04 <1%

18 Internet

pmc.ncbi.nlm.nih.gov <1%

19 Internet

www.sdiarticle2.org <1%

20 Publication

I Ketut Sardiana, Putu Perdana Kusuma Wiguna, Anak Agung Ayu Wulandira Sawi... <1%

21 Submitted works

Sunway Education Group on 2025-04-22 <1%

22 Submitted works

The University of Manchester on 2022-05-10 <1%

23 Submitted works

University of Lancaster on 2021-12-17 <1%

24 Submitted works

University of Queensland on 2021-03-25 <1%

25 Publication

biao an, Xianglong Luo, fengrong yu, Ting Jin, zhumei luo, guanyou zhu. "Applicat... <1%

26 Internet

ejournal.unitomo.ac.id <1%

27 Submitted works

GL Bajaj Institute of Technology and Management on 2025-12-06 <1%

28 Publication

Haoran Sun, Bingzhen Sun, Xixuan Zhao, Qiang Bao, Xiaoli Chu. "Three-way dyna... <1%

29 Submitted works

University of Nottingham on 2025-04-19 <1%

30 Submitted works

Brunel University on 2025-12-03 <1%

31 Publication

Salih M. Salih, Harith Sadaa Madhan Al-Fahdawy, Khdyar Yaes Khdyer Al-Kubissi, ... <1%

32 Publication

Tianliang Zhang, Xiaoji Lan, Jianhua Feng. "A Progressive Clustering Approach for... <1%

33 Submitted works

University of East London on 2024-08-30 <1%

34 Submitted works

University of Sussex on 2022-10-24 <1%

COMPILER
Vol. 14, No. 2, November 2025, pp. 107-118
ISSN: 2252-2403 (Print), ISSN: 2549-2403 (Online)
Accredited Fourth Grade by Kemenristekdikti, Decree No:10/C/C3/DT.05.00/2025
DOI: 10.28989/compiler.v14i2.3504

# K-Means Clustering on Rice Harvest Data for Planting Season Recommendation in Subak Cepaka, Tabanan

**Ni Made Cahyani Dewi[1,*], Ahmad Tri Hidayat[2]**
[1,2]Informatics Study Program, Yogyakarta University of Technology, Indonesia

**Article Info**

***Article history:***

Received October 6, 2025
Accepted November 30, 2025
Published December 14, 2025

***Keywords:***

K-Means
Subak
Recommendation
Historical Varvest Data

**ABSTRACT**

The Subak farming system in Tabanan Regency, Bali, is vital as a primary rice granary but faces challenges in determining optimal planting patterns. Planting decisions based only on inherited experience often do not match climate conditions, reducing productivity and increasing crop failure risks. This study implements the K-Means Clustering algorithm on five years of historical rice harvest data (2020–2024) to generate accurate planting season recommendations. Monthly data were analyzed and grouped into three categories: rainy, dry, and transitional seasons. The clustering results were integrated into a mobile application that provides farmers with accessible recommendations through an interactive interface and visualization. The effectiveness of the clustering model was evaluated using the Silhouette Score, which indicated good separation and cohesion among clusters, while efficiency was assessed through processing time and algorithm simplicity, confirming that K-Means performed the task with minimal computational cost. This system enables farmers to make data-driven planting decisions, optimize productivity, and support sustainable food security in Bali.



***Corresponding Author:***

Ni Made Cahyani Dewi,
Department of Computer Science,
Yogyakarta Technology University,
Jl. Siliwangi, Jombor Lor, Sendangadi, Kec. Mlati, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55285, Indonesia.
Email: cahyanidewi86@gmail.com

## 1. INTRODUCTION

The Subak irrigation system in Bali is not only a traditional mechanism for distributing water to rice fields but also reflects the philosophy of *Tri Hita Karana*, emphasizing harmony between humans, nature, and God. This system has been recognized by UNESCO as a world cultural heritage due to its unique governance rooted in social and spiritual values. Tabanan Regency, known as the rice granary of Bali, continues to face challenges in determining optimal planting patterns because of unpredictable climate change and the limited use of technology in agriculture [1].

Traditionally, most Subak farmers still rely on inherited knowledge to determine planting seasons. Such approaches often fail to match current environmental conditions, leading to reduced productivity, water imbalance, and even crop failure. These challenges highlight the need for data-driven and technology-based decision-making tools to provide more accurate planting recommendations [2], [3].

Clustering has emerged as one of the most widely used techniques in agricultural data analysis. Among various clustering algorithms, K-Means is popular due to its simplicity, efficiency, and effectiveness in detecting seasonal patterns from historical data. Compared to Hierarchical Clustering, which is easy to visualize but less efficient for large datasets, or DBSCAN, which can identify complex patterns but is sensitive to parameter selection, K-Means is more suitable for medium-sized datasets with relatively clear seasonal trends. Another alternative, Fuzzy C-Means, offers flexibility in cluster membership but requires higher computational resources [4].

ISSN: 2252-3839 (Print)-2549 2403 (Online)

Several international studies have demonstrated the effectiveness of K-Means in the agricultural domain. Javadi et al. [5] applied K-Means for delineating land management zones, while Swain et al. [6] integrated K-Means with ensemble learning to support crop selection. Shawon et al. [7] emphasized the role of machine learning, including clustering, in yield prediction, and Guevara-Viejó and Martínez [8] applied K-Means to assess agricultural production systems.

In Indonesia, K-Means has been employed for rice production mapping, regional clustering, and food productivity analysis. Examples include studies on new student data grouping [9], clustering of agricultural results [10], rice productivity mapping [11], [12], regional rice production segmentation [13], and predictive modeling for rice harvests [14], [15], [16], However, most studies remain limited to spatial or statistical analysis and have not integrated clustering results into mobile-based recommendation systems accessible to farmers [17].

In this context, the present study applies the K-Means Clustering algorithm to historical rice harvest data from Tabanan Regency (2020–2024), obtained from the Central Bureau of Statistics (BPS Bali). Previous research has primarily focused on clustering for agricultural mapping and productivity analysis but has not yet addressed the integration of clustering results into mobile-based decision support systems accessible to local farmers. This gap limits the practical impact of data analysis on real farming decisions.

Therefore, this study aims to bridge that gap by developing a system that not only performs data-driven clustering but also transforms the results into actionable planting season recommendations categorized into rainy, dry, and transitional seasons. The clustering outcomes are integrated into a mobile application featuring an interactive interface and visual analytics to ensure accessibility and usability for farmers. The desired outcome of this research is to provide an effective and efficient decision-support tool that helps farmers determine optimal planting schedules, enhance productivity, and strengthen the digital transformation and sustainability of the Subak agricultural system in Bali.

## 2. RESEARCH METHOD

This study adopts the CRISP-DM (Cross Industry Standard Process for Data Mining) framework as the research methodology. The research process consists of six main phases: problem identification (business understanding), data collection (data understanding), data preparation, modeling, evaluation, and implementation (deployment). In general, the research stages are illustrated in Figure 1, while the detailed explanation of each stage is presented in the following subsections.

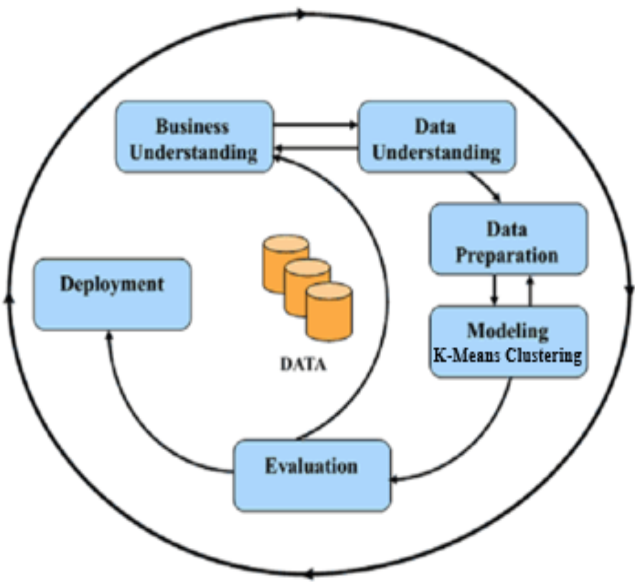



Figure 1. Research Stages based on the CRISP-DM Framework [18]

### 2.1. Business Understanding

Tabanan Regency is recognized as the rice granary of Bali through the traditional Subak farming system that has been preserved for generations. The main challenge faced by farmers lies in determining the optimal planting season, which still relies heavily on inherited traditional knowledge. Such an approach often fails to align with current climate variations, leading to inefficiencies and potential yield reduction.

This study applies the K-Means Clustering algorithm to analyze five years (2020–2024) of historical rice harvest data from the Central Bureau of Statistics (BPS Bali). Although the dataset spans only five years, each record represents aggregated monthly productivity data, providing 12 temporal features that capture distinct seasonal cycles. This temporal granularity ensures sufficient variability to produce meaningful clustering results.

The outcomes of this analysis serve as the foundation for developing a mobile-based planting season recommendation system designed for Subak farmers. The system aims to provide accessible, data-driven insights that support better decision-making, thereby strengthening agricultural resilience and promoting digital transformation in traditional Balinese agriculture.

### 2.2. Data Understanding

The dataset used in this study consists of historical monthly rice harvest data in Tabanan Regency for the period 2020–2024, obtained from the Central Bureau of Statistics (BPS) of Bali Province. The data is stored in CSV (Comma-Separated Values) format, which allows efficient processing using Python. The main variable analyzed is the harvested rice area (hectares) per month. This dataset was selected because it reflects seasonal patterns, namely rainy, dry, and transitional seasons. By analyzing this data, the study aims to discover recurring patterns that appear annually and provide support for planting decision-making.

### 2.3. Data Preparation

The data preparation stage is performed using the Python programming language. The process begins by importing a CSV dataset using the pandas library. The process includes:

a) The data preparation stage was conducted using the Python programming language. The process began by importing the CSV dataset using the Pandas library. Several steps were performed, including:
   Data cleaning, addressing metadata and empty rows in the CSV file.
b) Data selection, retaining only the monthly harvested area columns (January–December) while removing the annual total column, which was irrelevant to clustering.
c) Data transformation, restructuring the dataset so that each row represents one year and each column represents one month.
d) Normalization, standardizing the values to a uniform scale. Normalization was performed using the *StandardScaler* method, formulated as:

$$x' = \frac{x-\mu}{\sigma} \tag{1}$$

where x′ is the normalized value, x is the original value, μ is the mean, and σ is the standard deviation. The result of this stage is a dataset ready for use in K-Means Clustering analysis.

### 2.4. Modeling

The modeling process was conducted using the K-Means clustering algorithm implemented in Python. The optimal number of clusters (k) was determined using the Elbow Method and Silhouette Analysis to avoid subjective assumptions. The initial analysis indicated the highest Silhouette Coefficient at k=2. However, considering the agricultural context, k=3 was selected to represent three planting season categories: rainy, dry, and transitional seasons. The K-Means algorithm follows these steps:

a) Define the number of clusters kkk.
b) Initialize the centroids randomly.
c) Calculate the distance between data point xix_ixi and centroid ckc_kck using Euclidean Distance:

$$d(xi, ck) = \sum_{j=1}^{n} (xij - ckj)2 \tag{2}$$

d) Assign each data point to the nearest cluster.
e) Update centroids as the mean of data points within the cluster:

$$Ck = \frac{1}{nk}\sum xi \tag{3}$$

f) Repeat the process until centroids converge or the maximum iteration limit is reached.

where x′ is the normalized value, x is the original value, μ is the mean, and σ is the standard deviation. The result of this stage is a dataset ready for use in K-Means Clustering analysis.

### 2.5. Evaluation

The evaluation was carried out using the Silhouette Coefficient, which measures clustering quality by comparing the cohesion of data points within the same cluster and the separation between clusters. The coefficient is calculated as follows:

$$S(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{4}$$

Where:

a) a(i): the average distance of data point iii to other points in the same cluster,
b) b(i): the average distance of data point iii to the nearest cluster.

Values closer to 1 indicate better clustering quality. This evaluation ensures that the clustering results accurately represent the planting season patterns of rice in Tabanan.

### 2.6. Data System Implementation

The final stage of this study was the implementation of clustering results into a mobile-based application. The cluster outputs generated in Python were exported and integrated into the system. The application was designed with a simple and responsive interface, enabling access through both computers and smartphones. The main features of the application include:

a) Visualization of historical rice harvest data in graphical form.
b) Visualization of clustering results for planting seasons (rainy, dry, transitional).
c) Planting recommendations based on the identified clusters.

This implementation ensures that the analysis results are not limited to academic contributions but can be practically applied by Subak farmers to make more accurate planting decisions. Consequently, the system is expected to improve agricultural productivity and strengthen food security in Tabanan Regency.

## 3. RESULTS AND ANALYSIS

This study processed historical monthly rice harvest data from Tabanan Regency over the past five years (2020–2024), obtained from the Central Bureau of Statistics (BPS) of Bali Province. The dataset was used to identify seasonal patterns, which later became the basis for the recommendation system.

### 3.1. Data Preparation Results

The data preparation stage formed a critical foundation for this research. Initially, five separate CSV files (2020–2024) were consolidated into a single dataset to construct a comprehensive historical record. Data selection was restricted to Tabanan Regency, widely recognized as the rice granary of Bali.

Subsequently, the dataset was transposed so that each row represented one year and each column represented one month, as shown in Table 1. This transformation facilitated the detection of recurring monthly patterns across years. The dataset was then standardized using the StandardScaler method to ensure that each feature (month) contributed equally to the clustering process. The standardized results are presented in Table 2.

Table 1. Monthly Rice Harvest Area in Tabanan Regency (2020–2024)

| Th | Jan | Feb | March | April | Mei | June | July | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2020 | 1496.29 | 390.83 | 1170.78 | 2452.15 | 1538.29 | 3358.15 | 3532.3 | 3063.84 | 854.9 | 2713.14 | 2735.76 | 1963.15 |
| 2021 | 2436.99 | 2222.48 | 2806.22 | 3241.14 | 2786.79 | 2816.93 | 2059.12 | 1153.44 | 2375.05 | 2371.19 | 2278.91 | 2663.48 |
| 2022 | 1717.38 | 1985.69 | 3167.34 | 3989.82 | 2871.13 | 3230.25 | 2233.88 | 1263.34 | 979.09 | 1537.58 | 2786.63 | 3277.34 |
| 2023 | 2133 | 1828.8 | 2865.24 | 2504.05 | 2414.14 | 4616.63 | 2523.16 | 1718.33 | 1249.6 | 1262.4 | 1921.16 | 3153.85 |
| 2024 | 2497.81 | 1568.72 | 1214.3 | 2038.77 | 3310.43 | 4491.42 | 2233.86 | 1329.51 | 1001.96 | 2784.88 | 1971.93 | 2462.79 |

*Source: BPS Bali, processed*

Table 2. Data Standardization Results

| Th | Jan | Feb | March | April | Mei | June | July | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2020 | -0.62 | -2.01 | -0.99 | 0.05 | -0.58 | 0.81 | 1.01 | 0.58 | -1.13 | 0.44 | 0.47 | -0.21 |
| 2021 | 0.04 | -0.01 | 0.44 | 0.77 | 0.46 | 0.51 | -0.05 | -0.97 | 0.06 | 0.02 | -0.01 | 0.38 |
| 2022 | -0.4 | -0.16 | 0.69 | 1.25 | 0.49 | 0.77 | 0.01 | -0.94 | -1.04 | -0.58 | 0.49 | 0.8 |
| 2023 | -0.02 | -0.26 | 0.49 | 0.16 | 0.09 | 1.7 | 0.2 | -0.38 | -0.8 | -0.87 | -0.19 | 0.73 |
| 2024 | 0.12 | -0.53 | -0.96 | -0.05 | 0.83 | 1.54 | 0.02 | -0.76 | -1.02 | 0.46 | -0.18 | 0.29 |

### 3.2. Clustering Analysis with K-Means

To determine the optimal number of clusters (k), both the Elbow Method and Silhouette Analysis were applied. The results are summarized in Table 3.

Table 3. Comparison of Inertia and Silhouette Coefficient for Various k

| Number of Clusters (k) | Inertia Value | Silhouette Coefficient Value |
|---|---|---|
| 2 | 36.8491 | 0.2639 |
| 3 | 26.6142 | 0.2341 |
| 4 | 17.0925 | 0.2631 |
| 5 | 11.8598 | 0.2432 |
| 6 | 9.0032 | 0.2111 |

The analysis shows that the highest Silhouette Coefficient (0.2639) occurred at k=2. Nevertheless, considering the agricultural context of Bali where cropping patterns are traditionally divided into three main seasons, this study adopted k=3 to represent the rainy, dry, and transitional periods. The average harvest characteristics for each cluster are shown in Table 4.

Table 4. Average Harvest Area per Cluster, k=3

| Cluster | Average Harvested Area (Hectares) | Season Category |
|---|---|---|
| Cluster 0 | 2275.69 | Transitional |
| Cluster 1 | 1630.93 | Dry Season |
| Cluster 2 | 2724.84 | Rainy Season |

Cluster 2 corresponds to the rainy season with the highest average harvested area (2724.84 ha), Cluster 1 represents the dry season with the lowest productivity (1630.93 ha), and Cluster 0 represents the transitional season with intermediate values (2275.69 ha).

### 3.3. Cluster Quality Evaluation

Cluster evaluation was conducted using both the Elbow Method and Silhouette Analysis to determine the optimal number of clusters ($k$). Figure 2 shows that inertia decreases sharply up to $k = 3$ and then stabilizes, indicating the formation of an "elbow" around $k = 3–4$. Meanwhile, Figure 3 illustrates that the highest Silhouette Coefficient was achieved at $k = 2$. Although $k = 2$ provided slightly better statistical separation, $k = 3$ was selected to maintain consistency with the local agricultural context, which naturally recognizes three distinct planting seasons: rainy, dry, and transitional.

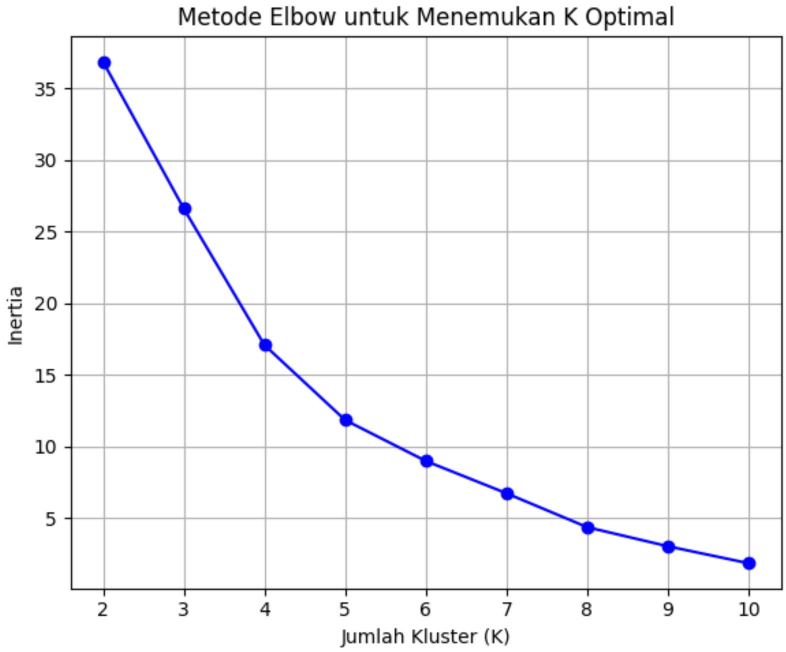
The Silhouette Coefficient of 0.2341 for $k = 3$ indicates moderate overlap between clusters, yet the positive value confirms that most data points are more similar to their own cluster than to others.

This choice represents a balanced compromise between statistical validity and contextual interpretability, producing actionable insights that are both analytically sound and meaningful to Subak farmers in planning seasonal activities.

Additionally, comparative testing using DBSCAN and Fuzzy C-Means was conceptually evaluated. DBSCAN failed to form stable clusters due to the small dataset size and continuous temporal nature, while Fuzzy C-Means produced overlapping memberships that reduced interpretability. Consequently, K-Means was confirmed as the most appropriate algorithm for this dataset and research objective.

1






2

Figure 2. Elbow Method Graph for Determining the Number of Clusters

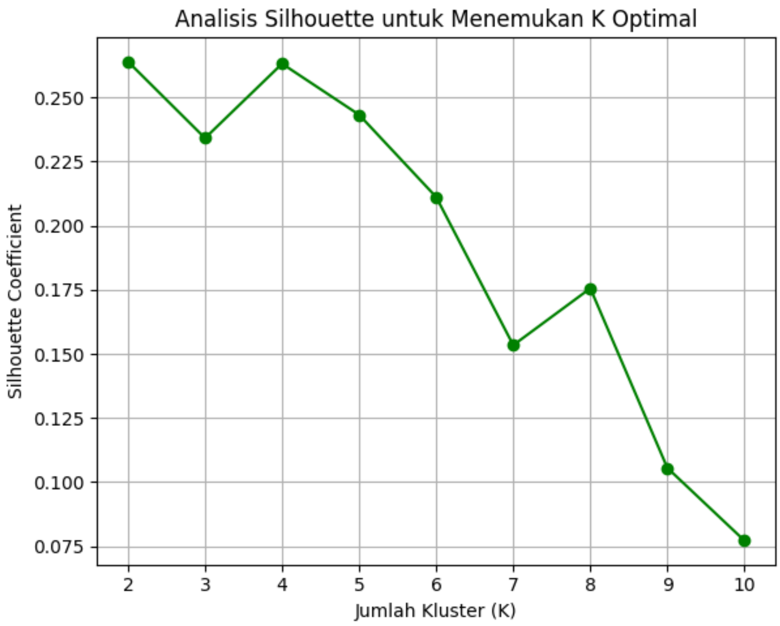



Figure 3. Silhouette Analysis

### 3.4. Visualization of Clustering Results

To enhance interpretability, the clustering results were visualized using Principal Component Analysis (PCA), which reduced the 12-dimensional dataset (monthly features) into two principal components. The visualization is presented in Figure 4.

23

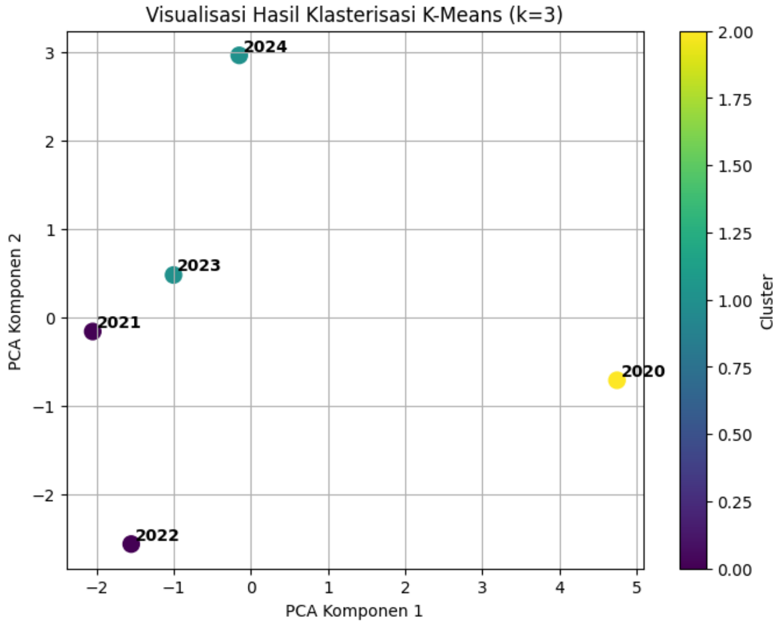



Figure 4. Visualization of K-Means Clustering Results, k=3

Each point in Figure 4 represents a yearly dataset (2020–2024), while colors indicate cluster membership. Year labels were added to clarify the distribution across periods. The visualization reveals that:

a) 2020 forms a distinct cluster associated with high productivity, representing the rainy season.
b) 2022 is grouped separately, reflecting low productivity typical of the dry season.
c) 2021, 2023, and 2024 form a neighboring cluster, indicating transitional conditions between rainfall and drought periods.

These results demonstrate that, despite not achieving the maximum Silhouette Coefficient, the clustering process successfully captures meaningful seasonal dynamics consistent with local agricultural experience. Hence, the model not only fulfills technical criteria but also provides practical value for strategic agricultural planning and sustainable rice cultivation in Tabanan Regency.

### 3.5. System Implementation and Testing

The clustering results were implemented into a mobile-based application designed to provide farmers with planting season recommendations. The interface was developed to be simple and responsive, ensuring ease of use for Subak farmers. The application begins with a *Get Started* page (Figure 5), followed by user registration (Figure 6) and login (Figure 7). Once authenticated, users are directed to the dashboard (Figure 8), which presents several features including harvest statistics (Figure 9).

To obtain planting recommendations, farmers can access the "Planting Pattern" feature, where inputs such as rice variety, season, and land area are required. Based on the given input, the system processes the data and provides recommendations (Figure 10). For instance, the system suggested Cluster 2 (rainy season) with the label "Highly Suitable," which corresponds with clustering results showing the highest average harvest area in this season. Additional features include rice crop maintenance information (Figure 11) and Subak knowledge resources (Figure 12).

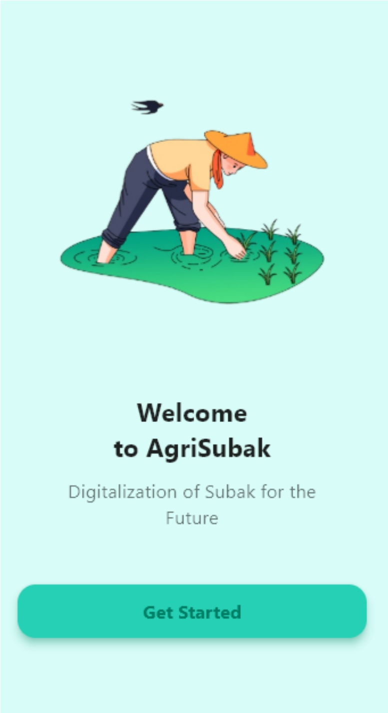



Figure 5. Get Started Page

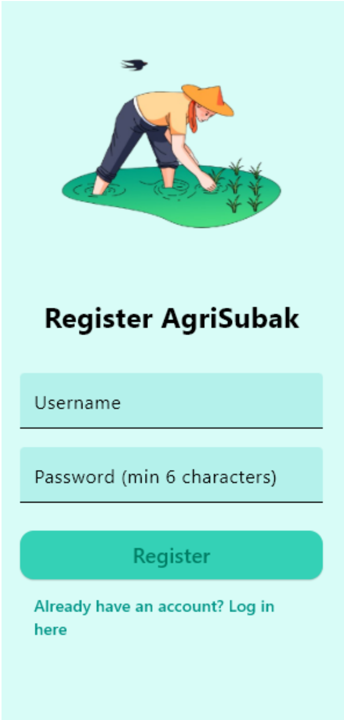



Figure 6. Registration Page

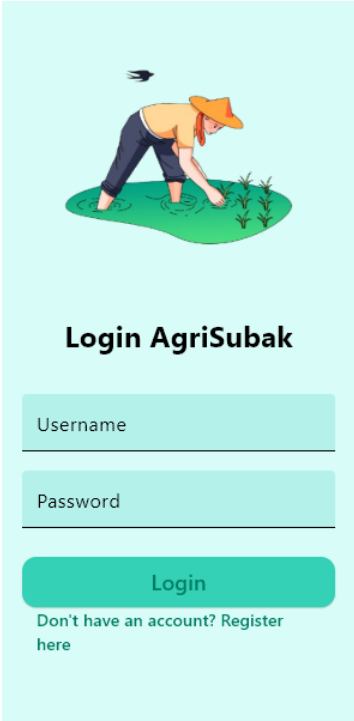



Figure 7. Login Page

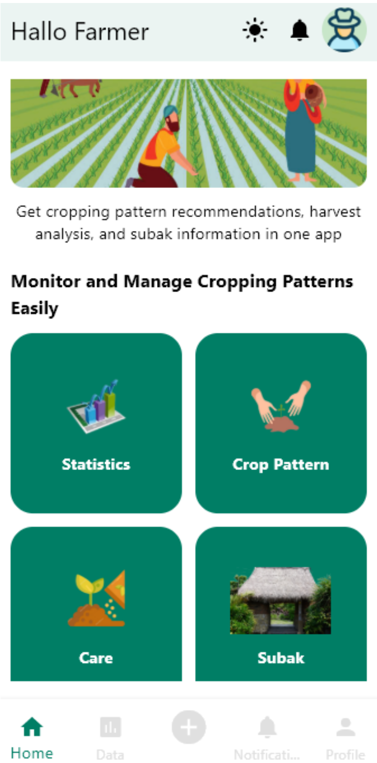



Figure 8. Dashboard Page

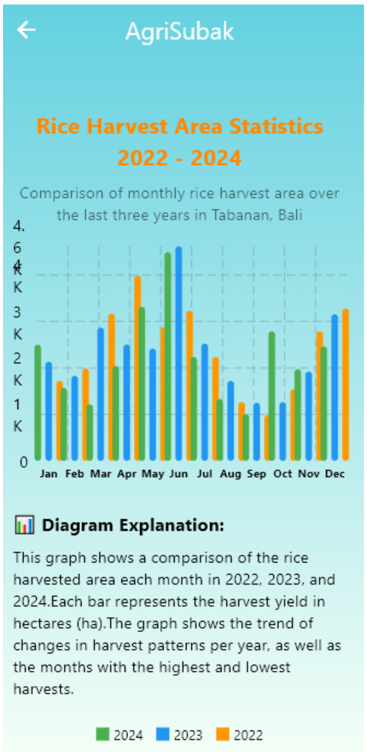



Figure 9. Statistics Page

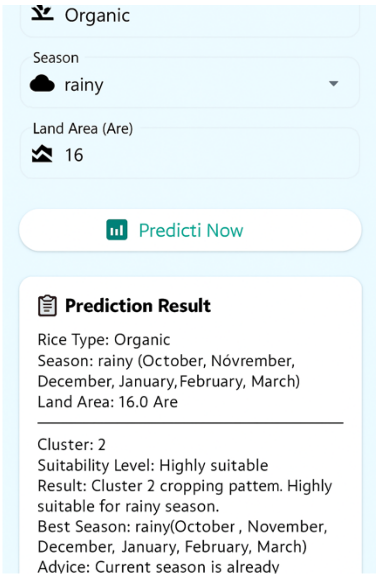



Figure 10. Planting Pattern Page

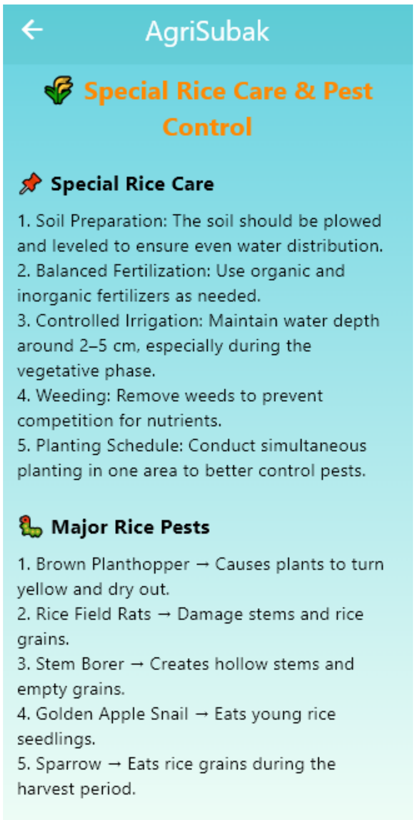


Figure 11. Rice Care Information Page

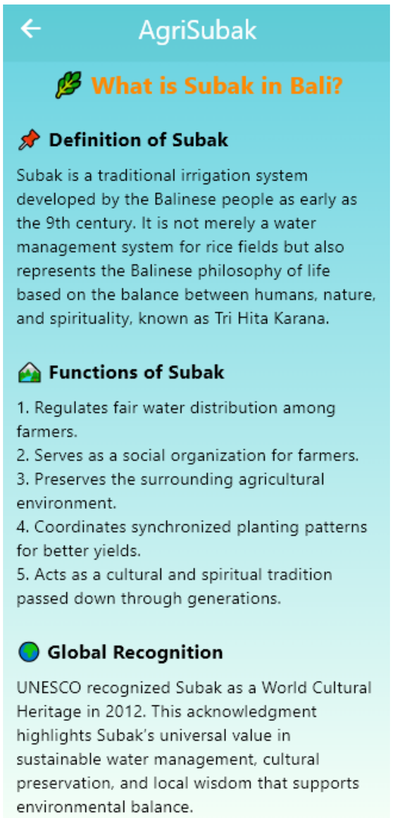


Figure 12. Subak Information Page

To ensure functionality, black-box testing was conducted, focusing on input-output validation of each feature without analyzing the internal code structure. A summary of the testing results is provided in Table 5.

19

16

Table 5. Black-box Testing Results

| No | Page | Expected Outcome | Status |
|---|---|---|---|
| 1 | Get Started Page | Displays the application's initial screen and allows navigation to the registration page | Passed |
| 2 | Account Registration | The system successfully creates a new account with valid user data | Passed |
| 3 | Login Page | The system accepts valid credentials and redirects the user to the dashboard | Passed |
| 4 | Dashboard | Presents a summary of information and navigation to main features | Passed |
| 5 | Statistics Page | Displays graphical visualization of historical rice harvest data. | Passed |
| 6 | Planting Pattern Input | The system accepts inputs (rice variety, season, land area) and processes them | Passed |
| 7 | Recommendation Result Page | Shows the assigned cluster and planting recommendations based on processed input | Passed |
| 8 | Rice Care Information Page | Provides textual and visual information about rice crop care | Passed |
| 9 | Subak Information Page | Provides educational content about Subak and its cultural significance | Passed |

The testing confirmed that all key features—registration, login, dashboard navigation, prediction input, recommendation output, and statistical visualization—operated correctly and consistently with the design specifications. These results demonstrate that the system is reliable for practical use in supporting Subak farmers' decision-making.

### 3.6. Discussion

The implementation of the K-Means algorithm on historical rice harvest data in Tabanan Regency demonstrated its effectiveness in identifying seasonal patterns rainy, dry, and transitional seasons as summarized in Table 4 and visualized in Figure 4. Despite the inherent variability of agricultural data, the

ISSN: 2252-3839 (Print)-2549 2403 (Online)

clustering process successfully revealed recurring seasonal structures. These results are aligned with prior studies that confirmed the suitability of K-Means for agricultural management and seasonal prediction, [19].

The main contribution of this study lies not only in the analytical application of clustering but also in the integration of results into a mobile-based recommendation system. This integration bridges the gap between data-driven analysis and practical usability, allowing Subak farmers to complement traditional knowledge with modern, algorithmic insights [20].

From a technical perspective, the clustering process offers a foundation for dynamic decision-support systems that can adapt to changing environmental conditions. From a practical perspective, the resulting application provides visualization and recommendation features that enhance farmers' decision-making processes. These contributions collectively improve agricultural productivity, reinforce local food security, and promote sustainable management of the Subak system. Overall, this research demonstrates that data mining techniques, when contextualized within local socio-agricultural frameworks, can deliver solutions that are both technically robust and culturally relevant [21], [22].

### 3.7. System Architecture and Usability Evaluation

The proposed system architecture integrates the clustering analysis results into a mobile-based recommendation platform to deliver practical insights for Subak farmers. The system follows a four-layer data processing pipeline:

a) Data Acquisition Layer: Collects and integrates historical rice harvest data (2020–2024) from official government sources. The data are cleaned, standardized, and formatted to ensure analytical consistency.
b) Data Processing and Clustering Layer: Normalized data are processed using the *K-Means* algorithm to identify seasonal patterns rainy, dry, and transitional. *Principal Component Analysis (PCA)* is then applied to visualize and simplify the clustering outcomes.
c) Recommendation Generation Layer: Translates cluster results into actionable planting guidance. High-productivity clusters correspond to rainy seasons, while low-yield clusters represent dry conditions. These recommendations form the basis for adaptive planting strategies.
d) Mobile Application Delivery Layer: Presents the analysis through an accessible mobile interface that displays seasonal insights and recommendations. The design emphasizes simplicity, cultural alignment, and usability for Subak farmers.

Usability was validated through black-box testing, confirming that all features functioned as intended and system responses were stable. Informal feedback from Subak representatives indicated that the application was easy to use, informative, and beneficial for planting decisions. Overall, the system effectively bridges analytical modeling and real-world application, promoting data-driven decision-making, sustainable rice production, and the digital transformation of the traditional Subak system.

## 4. CONCLUSION

The application of the K-Means Clustering algorithm to historical rice harvest data from Tabanan Regency (2020–2024) successfully identified planting season patterns. The Elbow Method and Silhouette Analysis indicated that the statistically optimal number of clusters was $k = 2$, with a Silhouette Coefficient of 0.2639. However, $k = 3$ was chosen to reflect the three main agricultural cycles recognized in the Subak system rainy, dry, and transitional seasons ensuring that the results remained contextually meaningful and practically applicable [23].

The originality of this study lies in the integration of clustering outputs into a mobile-based application that provides data-driven planting recommendations accessible to local farmers [24]. The system underwent black-box testing, confirming that all core features, including data visualization, cluster-based recommendations, and navigation interfaces, functioned as intended. This demonstrates that the system is both academically valid and operationally feasible for end users [25].

In conclusion, this study makes a tangible contribution to the digital transformation of agriculture in Bali, leveraging historical harvest data to produce actionable insights for crop scheduling. The findings are expected to help Subak farmers optimize planting schedules, increase productivity, and strengthen long-term food security, all while preserving the traditional wisdom embedded within Bali's Subak culture [26].

## REFERENCES

[1] UNESCO, "Cultural Landscape of Bali Province: The Subak System as a Manifestation of the Tri Hita Karana Philosophy." 2012. [Online]. Available: https://whc.unesco.org/en/list/1194/
[2] V. Sitokonstantinou and others, "Fuzzy clustering for the within-season estimation of cotton phenology using remote sensing time series," *Remote Sens.*, vol. 15, no. 12, p. 3115, 2023, doi: 10.3390/rs15123115.
[3] I. Sanela, A. Nazir, and others, "Penerapan metode clustering dengan K-Means untuk memetakan

potensi tanaman padi di Sumatera," *J. Syst. Informatics*, vol. 7, no. 2, pp. 88–95, 2023, doi: 10.47065/josyc.v5i1.4523.
[4] A. Mubarok, A. A. Syahputra, A. T. Permana, L. Sholiah, and Tarwoto, "Implementasi data mining untuk clustering lowongan pekerjaan menggunakan metode algoritma K-Means," *J. Teknol. Inf. dan Komun.*, vol. 9, no. 2, pp. 162–171, 2022, doi: 10.35870/jtik.v9i2.3438.
[5] H. J. et al., "Clustering and smoothing pipeline for management zone delineation," *Sensors*, vol. 22, no. 9, pp. 1–18, 2022, doi: 10.3390/s22020645.
[6] K. P. Swain, S. Tripathy, and R. Nayak, "Empowering crop selection with ensemble learning and K-Means clustering: A modern agricultural perspective," *Open Agric.*, vol. 9, no. 1, pp. 122–132, 2024.
[7] S. M. Shawon and others, "Machine learning in agriculture: Crop yield prediction – A comprehensive review," *Artif. Intell. Agric.*, vol. 15, pp. 74–95, 2024, doi: 10.1016/j.atech.2024.100718.
[8] F. Guevara-Viejo and S. Martínez, "Application of K-Means clustering to assess agricultural production systems," *Agronomy*, vol. 11, no. 6, p. 1123, 2021, doi: 10.3390/agronomy11061123.
[9] K. Setiawan and Y. Y. A. Saputry, "Clustering data calon siswa baru menggunakan metode K-Means di Pusat Pengembangan Anak Fajar Baru Cengkareng," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 1, pp. 75–83, 2024, doi: 10.35870/jtik.v8i1.1426.
[10] H. Pratiwi and E. Handayani, "Implementasi algoritma K-Means untuk pengelompokan data hasil pertanian," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 4, no. 2, pp. 87–95, 2022.
[11] S. Wijayanto and M. Y. Fathoni, "Pengelompokkan produktivitas tanaman padi di Jawa Tengah menggunakan algoritma K-Means," *J. Nas. Teknol. Inf.*, vol. 5, no. 2, pp. 134–145, 2021.
[12] E. A. P. Putri, "K-Means clustering untuk pengelompokan daerah penghasil padi di Indonesia berdasarkan luas panen, produksi, dan produktivitas padi tahun 2022," *Mikrotik J.*, vol. 4, no. 1, pp. 55–65, 2024.
[13] R. H. Cokro and others, "Segmentasi provinsi penghasil padi di Indonesia berdasarkan karakteristik produksi tahun 2024 menggunakan clustering K-Means," *Kohesi J. Ilmu Komput.*, vol. 8, no. 1, pp. 50–60, 2025.
[14] R. Farismana, "Penerapan K-Means clustering untuk pemetaan produktivitas padi dan prediksi panen di Kabupaten Indramayu," *J. Ilmu Sos. dan Manaj. Rev.*, vol. 5, no. 2, pp. 100–110, 2024, doi: 10.52362/jisamar.v8i3.1572.
[15] Chamidah, "Klasterisasi provinsi di Indonesia berdasarkan produktivitas komoditas pangan menggunakan algoritma K-Means," *J. Sains Terap.*, vol. 8, no. 2, pp. 101–110, 2022.
[16] S. N. Sofyan, D. Kurniawan, and A. Rachmawati, "Implementasi data mining untuk clustering produktivitas bawang merah menggunakan metode K-Means," *Jatilima J. Teknol. Inf. dan Multimed.*, vol. 4, no. 1, pp. 45–53, 2025.
[17] R. Hesananda, "Penerapan klasterisasi K-Means terhadap produktivitas padi di Pulau Sumatera sebagai strategi pendukung ketahanan pangan," *J. Informatics Adv. Comput.*, vol. 6, no. 1, pp. 54–63, 2025, doi: 10.35814/ve9jhz24.
[18] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," in *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 2000, pp. 29–39. [Online]. Available: https://cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf
[19] L.-Y. Wu and others, "Research on Using K-Means Clustering to Explore High-Risk Products with Ethylene Oxide Residues and Their Manufacturers in Taiwan," *Foods*, vol. 13, no. 16, p. 2510, 2024, doi: 10.3390/foods13162510.
[20] B. Zhang and others, "Research on Automatic Alignment for Corn Harvesting Based on Euclidean Clustering and K-Means Clustering," *Agriculture*, vol. 14, no. 11, p. 2071, 2024, doi: 10.3390/agriculture14112071.
[21] D. Santoso and others, "Cloud-based framework for agricultural data sharing and analytics," *J. Cloud Comput.*, vol. 12, pp. 1–14, 2023, doi: 10.1186/s13677-023-00412-6.
[22] F. Guevara-Viejó and others, "Application of K-Means Clustering Algorithm to Commercial Parameters of Pleurotus spp. Cultivated on Representative Agricultural Wastes from Province of Guayas," *J. Fungi*, vol. 7, no. 7, p. 537, 2021, doi: 10.3390/jof7070537.
[23] M. K. Shukla and P. Sharma, "Fuzzy K-Means and Principal Component Analysis for Classifying Soil Properties for Efficient Farm Management and Maintaining Soil Health," *Sustainability*, vol. 15, no. 17, p. 13144, 2023, doi: 10.3390/su151713144.
[24] A. Filintas, N. Gougoulias, N. Kourgialas, and E. Hatzichristou, "Management Soil Zones, Irrigation, and Fertigation Effects on Yield and Oil Content of Coriandrum sativum L. Using Precision Agriculture with Fuzzy k-Means Clustering," *Sustainability*, vol. 15, no. 18, p. 13524, 2023, doi: 10.3390/su151813524.

[25] S. Zhang, J. Guo, and Z. Wang, "Combing K-means Clustering and Local Weighted Maximum Discriminant Projections for Weed Species Recognition," *Front. Comput. Sci.*, vol. 1, p. 4, 2019, doi: 10.3389/fcomp.2019.00004.

[26] M. Nikhar and L. P. Thakre, "Smart Agricultural Farm Enhancement with K-Means Learning," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 8, pp. 166–169, 2020.