Smart Techno (Smart Technology, Informatic, and Technopreneurship)
Volume 08, Number 01, 2026 pp. 24-38
E-ISSN : 2541-0679
Open Access: https://lppm.primakara.ac.id/jurnal/index.php/smart-techno/

Clustering of Planted Area, Harvested Area, and Rice Production in Each Village of Jember Regency Using K-Means Clustering and the Davies Bouldin Index

Hestina Restu Astika[1*]
[1]Informatics Engineering, Universitas Muhammadiyah Jember, Jember, Indonesia

ARTICLE INFO
Article history:
Initial submission
07-09-2025
Received in revised form
12-11-2025
Accepted 23-11-2025
Available online 27-02-2026

Keywords:
K-Means Clustering, Davies Bouldin Index, Rice Production, Geographic Information System.

DOI:
ttps://doi.org/10.59356/smart-techno.v7i1

ABSTRACT

Rice (Oryza sativa L.) is a cultivated crop that serves as the primary staple food for the majority of Indonesia's population. East Java Province is one of the regions with the highest rice production in the country. Therefore, increasing rice production is essential to meet national food demands. This study aims to classify villages in Jember Regency based on the variables of planted area, harvested area, and rice production, using data obtained from the official publications of the Jember Regency Central Bureau of Statistics for 2022 and 2023, covering a total of 248 villages. The data were processed using the K-Means Clustering algorithm, followed by determining the optimal number of clusters using the Davies Bouldin Index. The clustering results were visualized in an interactive web-based map through a Geographic Information System. Based on testing cluster counts from 2 to 10, the optimal number of clusters was found to be three, with a Davies Bouldin Index value of 0.605. This study is expected to provide benefits for the Jember Regency Central Bureau of Statistics, the community, and farmers in storing, managing, and disseminating information regarding rice crops in Jember Regency.

1. INTRODUCTION

Rice (Oryza sativa L.) is a highly important cultivated crop, as rice serves as the staple food for the majority of the population in Asia, including Indonesia. East Java Province is one of the largest rice-producing regions in Indonesia, therefore, the province is required to continuously increase its rice production to meet food demands. Jember Regency is among the key rice-producing regions in East Java, with a rice production of 607.37 thousand tons of dry unhusked rice (GKG) and a harvested area of 118.49 thousand hectares in 2022. In 2023, rice production increased to approximately 616.73 thousand tons of GKG with a harvested area of around 120.19 thousand hectares (BPS Jember Regency, 2023).

Optimizing rice production data management requires the application of clustering methods and visualization techniques to support the processing and storage of large-scale data. Based on observations at the Jember Regency Central Bureau of Statistics (BPS), a system that classifies rice production data at the village level is currently unavailable, the available data classification only covers sub-district or regency levels. Therefore, there is a need for a system capable of categorizing the villages in Jember Regency into high, medium, and low production levels. Furthermore, web-based geographic information system visualization is essential to allow the public and farmers to access information on planted area, harvested area, and rice production for each village.

In data mining, clustering is used to group data into several clusters based on similarities. Several clustering algorithms exist, one of which is K-Means Clustering (Wijayanto & Yoka Fathoni, 2021). The K-Means algorithm was selected for this study because it is easy to understand and implement, and it is effective for clustering large volumes of numerical data.

Several previous studies have applied K-Means Clustering and the Davies–Bouldin Index (DBI). A study by Wahidah et al. (2023) analyzed disaster-prone areas in Jember

Regency using K-Means Clustering, clustering sub-districts into high, medium, and low vulnerability categories, and obtaining an optimal cluster number of three with a DBI value of 1.156763. Another study by Ananda et al. (2022) applied K-Means for clustering birth certificate data in Indonesia, resulting in a DBI value of 0.059 with four optimal clusters. Additionally, research conducted by W. Putri & Afdal (2023) examined the use of the K-Means algorithm to cluster disability data in Rokan Hilir Regency, generating three clusters representing low, medium, and high disability levels with a DBI value of 0.063.

Based on previous studies, the K-Means Clustering algorithm was chosen for this research due to its simplicity, ease of implementation, and effectiveness in grouping large numerical datasets based on similarity values such as planted area, harvested area, and rice production data across villages in Jember Regency. To ensure that the clustering results produce an optimal number of clusters, this study applies the Davies Bouldin Index (DBI) to evaluate clustering quality. A smaller DBI value indicates better clustering performance.

In accordance with the explanation above, this research aims to cluster the planted area, harvested area, and rice production of each village in Jember Regency using K-Means Clustering and the Davies Bouldin Index to provide mapping information that can be accessed by the public, farmers, and relevant institutions to support more accurate decision-making in rice production management and improvement based on village-level production categories of high, medium, and low.

2. LITERATURE REVIEW

Data Mining

Data mining is an analytical process aimed at extracting information and identifying patterns from large datasets. This process includes data collection, data retrieval, data analysis, and data processing to produce useful models or knowledge (Amna et al., 2023). In general, there are two main approaches in implementing data mining (Fahada, 2024):

1. Supervised Learning is an approach that uses training data consisting of labeled variables to classify new data based on existing patterns. Techniques included in this approach comprise prediction and classification.
2. Unsupervised Learning is an approach that operates without data labeling, allowing the available data to be grouped into several clusters, such as two or more groups. Techniques that fall under this approach include clustering, estimation, and association.

Clustering

Clustering is an unsupervised learning method that groups data based on similarities in their characteristics (Chaundhry et al., 2023). In clustering, a dataset is divided into several groups so that data within the same cluster share similar characteristics, while data with different characteristics are placed into separate clusters.

K-Means Algorithm

The K-Means Clustering algorithm is a method within the partitioning or non-hierarchical approach that groups data into several mutually exclusive clusters. The K-Means algorithm is widely known for its ease of use in processing large datasets and its ability to handle outliers efficiently (Handayani, 2022). The stages of the data clustering process using K-Means can be seen in the steps below (Amalina et al., 2024).

1. Determining the desired number of clusters (k).
2. Selecting the initial centroids from randomly chosen data points.
3. Calculating the distance of each data point to all centroids using the Euclidean distance formula.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

Description:

$d(x,y)$: Euclidean distance between a data point and a centroid

n : number of attributes
$x_i = (x_1, x_2, \dots, x_i)$: data point values
$y_i = (y_1, y_2, \dots, y_i)$: centroid values

4. Assigning each data point to the cluster whose centroid has the shortest distance.
5. Calculating the average of all data points within each cluster to determine the new centroid using the following formula.

$$\mathrm{C_k} = \frac{1}{\mathrm{n_k}} \sum \mathrm{d_i} \quad (2)$$

Descriptionn:
C_k : new centroid
n_k : number of data points in the cluster
d_i : data points within the cluster

6. Afterward, the distance of each data point to the new centroids is recalculated, and the process continues until no changes occur in the cluster assignments, indicating that the clustering process is complete.

Davies Bouldin Index (DBI)

Davied L. Davies and Donald W. Bouldin introduced the Davies–Bouldin Index (DBI) in 1979 as a metric for evaluating clustering quality by maximizing the distance between clusters while minimizing the distance among points within the same cluster. The smaller the DBI value, or the closer it is to zero, the better the resulting clustering quality. The steps for calculating the Davies Bouldin Index (DBI) are as follows (Fauzi et al., 2023):

1. The first step is determining the SSW (Sum of Squares Within Cluster) value by measuring the closeness among members within a cluster. A smaller value indicates better cohesion, as it reflects a higher level of similarity. SSW is used to measure the cohesion or homogeneity within a cluster and can be calculated using the following formula.

$$SSW = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (3)$$

Descriptionn:
m_i : number of data points in the *i*-th cluster
$d(x_j, c_i)$: distance between a data point and the centroid
x_j : data point within the cluster

2. Next, the SSB (Sum of Squares Between Cluster) value is determined by measuring the degree of separation between clusters, where a larger distance indicates better separation from other clusters. The purpose of this step is to assess the level of separation or heterogeneity among clusters. Separation refers to the differences between one cluster and another, and it is calculated using the following formula.

$$SSB_{ij} = d(c_i . c_j) \quad (4)$$

Descriptionn:
c_i : centroid of the first cluster
c_j : centroid of the second cluster
$d(c_i . c_j)$: distance between the two centroids

3. Calculating the ratio, which is used to evaluate the comparative quality between one cluster and another. This ratio represents cohesion and separation, where the separation value must be greater than the cohesion for the clusters to be considered well-formed. The formula is as follows.

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

Descriptionn:
R_{ij} : ratio between clusters
SSW_i, SSW_j : cohesion values of each cluster
SSB_{ij} : separation value between the two clusters

4. Finally, the Davies Bouldin Index is calculated by determining the average of the maximum ratio values for each cluster using the following formula.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i\neq j}(R_{ij}) \quad (6)$$

Descriptionn:

k : number of clusters

R_{ij} : ratio between the *i*-th and *j*-th clusters

max : the highest ratio value among the compared clusters

Geographic Information System (GIS)

A Geographic Information System (GIS) is a computer-based system used to manage, store, process, analyze, and display data related to the geographical conditions of a particular area (Hua, 2015). This system enables users to analyze and visualize spatial data effectively. A geographic information system consists of computer systems, spatial data, and users. It represents the real world on a computer screen similar to maps on paper, but with greater flexibility and functionality compared to conventional maps (Prasetya et al., 2021).

Leaflet

Leaflet, or Leaflet.js, is a JavaScript library used by application developers to build geographic information system applications. Leaflet works efficiently across both mobile and desktop platforms and can be integrated with various plugins. It features an elegant design and is easy to use. Digital maps are displayed using Leaflet JavaScript, which supports GeoJSON files a data format capable of storing geographic features (T. Putri, 2022).

3. METHOD

Research Stages

In completing this final project, the stages carried out in this research are as follows:

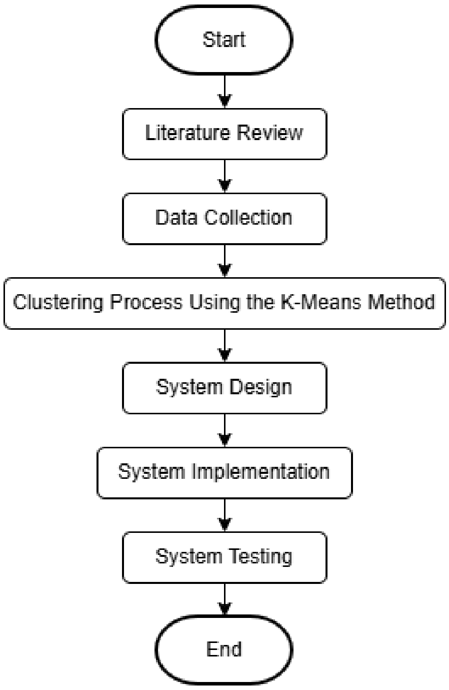

Figure 1. Research Stages

Data Collection

The data used in this study consist of rice crop data from 2022 and 2023, covering the variables of planted area, harvested area, and rice production for all villages in Jember Regency. These data were obtained from the official publications of BPS Jember Regency. The 2022 and 2023 datasets were merged into a single table to serve as the dataset for the clustering process. For the boundary map of Jember Regency, geographic coordinates in the form of latitude and longitude for each village were collected from Google Maps, manually input into Excel, and then converted into GeoJSON format using the online converter available at https://mygeodata.cloud/converter/xls-to-geojson

K-Means Clustering Algorithm Flowchart

The steps of the computation using the K-Means Clustering algorithm can be seen in the figure below.

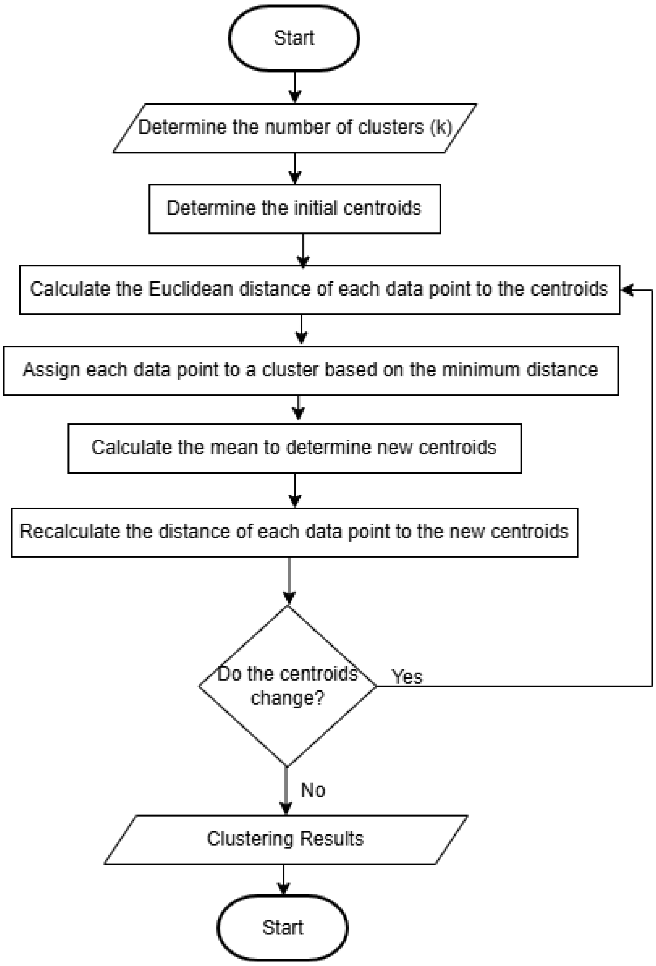

Figure 2. K-Means Clustering Calculation Flowchart

4. RESULT AND DISCUSSION

Implementation of the K-Means Clustering Algorithm

1. **Implementation Using Microsoft Excel**
   The implementation of the K-Means Clustering algorithm using Microsoft Excel requires several stages as follows:
   a) Preparing the dataset

Table 1. Dataset

i-th Data	Sub-district	Village	2022			2023		
			LT	LP	HP	LT	LP	HP
1	Ajung	Ajung	1482	1639	9534	1489	1448	**8915**
2	Ajung	Klompangan	1418	1487	9083	1424	1296	**8202**
3	Ajung	Mangaran	985	1123	6805	987	932	**5822**
4	Ajung	Pancakarya	1045	1106	6511	1052	915	**5651**
5	Ajung	Rowoindah	584	614	3584	583	423	**2473**
6	Ajung	Sukamakmur	1289	1353	8181	1298	1164	**7351**

7	Ajung	Wirowongso	850	879	5100	849	692	**4078**
8	Ambulu	Ambulu	125	214	1324	228	223	**1399**
9	Ambulu	Andongsari	481	563	3487	725	708	**5980**
10	Ambulu	Karanganyar	493	575	3566	567	555	**3482**
...	...	...	...	...	...	...	...	**...**
240	Umbulsari	Umbulrejo	125	101	625,19	65	61	**384**
241	Umbulsari	Umbulsari	136	117	724,23	66	61	**385**
242	Wuluhan	Ampel	900	803	4918	825	532	**3340**
243	Wuluhan	Dukuhdempok	662	591	3619	613	592	**3717**
244	Wuluhan	Glundengan	1046	933	5709	892	927	**5821**
245	Wuluhan	Kesilir	624	557	3411	586	558	**3504**
246	Wuluhan	Lojejer	595	531	3252	556	532	**3340**
247	Wuluhan	Tamansari	883	788	4836	821	790	**4960**
248	Wuluhan	Tanjungrejo	860	767	4697	795	769	**4829**

Description:
LT = Planted Area
LP = Harvested Area
HP = Production Yield

b) Determining the initial centroid values

c) Calculating the distance to each centroid using the Euclidean distance formula in Excel

$$= SQRT((\,LT\ 2022\ first\ data - \ LT\ 2022\ centroid)^2 + \\ (\,LP\ 2022\ first\ data\ - LP\ 2022\ centroid)^2 + \\ (\,HP\ 2022\ first\ data - \ HP\ 2022\ centroid)^2 + \\ (\,LT\ 2023\ first\ data - LT\ 2023\ centroid)^2 + \\ (\,LP\ 2023\ first\ data\ - LP\ 2023\ centroid)^2 + \\ (\,HP\ 2023\ first\ data - HP\ 2023\ centroid)^2)$$

After all distances have been calculated, each village is classified into a cluster based on the shortest distance to the centroid. The results of the distance calculations from each village to the centroids can be seen in the table below.

Table 2. Clustering Results of Iteration 1

i-th Data	Village	C1	C2	C3	*Cluster* 1	*Cluster* 2	*Cluster* 3
1	Ajung	0	875.39	4248.01	1	-	**-**
2	Klompangan	875.39	0	3390.74	-	2	**-**
3	Mangaran	4248.01	3390.74	0	-	-	**3**
4	Pancakarya	4554.38	3700.08	352.25	-	-	**3**
5	Rowoindah	8979.44	8123.27	4736.33	-	-	**3**
6	Sukamakmur	2124.36	1267.16	2127.69	-	2	**-**
7	Wirowongso	6709.33	5853.15	2470.42	-	-	**3**
8	Ambulu	11439.28	10599.86	7228.34	-	-	**3**
9	Andongsari	6962.15	6230.46	3423.53	-	-	**3**
10	Karanganyar	8300.14	7462.34	4102.15	-	-	**3**
...	...	...	...	...	...	...	**...**
240	Umbulrejo	12661.09	11816.12	8435.32	-	-	**3**
241	Umbulsari	12587.69	11741.56	8359.21	-	-	**3**
242	Ampel	7396.33	6531.84	3164.96	-	-	**3**
243	Dukuhdempok	8079.43	7245.56	3901.85	-	-	**3**
244	Glundengan	5051.72	4232.91	1118.07	-	-	**3**
245	Kesilir	8383.58	7548.84	4200.44	-	-	**3**
246	Lojejer	8617.9	7782.54	4430.85	-	-	**3**
247	Tamansari	6298.84	5471.93	2188.69	-	-	**3**
248	Tanjungrejo	6494.76	5667.21	2373.92	-	-	**3**
	Total				**9**	**15**	**224**

d) After completing the K-Means calculation using the Euclidean distance formula and determining the new centroids until no further changes occur in cluster assignments, the process is considered complete. This condition indicates that all data points within each cluster have reached stability. Therefore, the data presented in the table below can be regarded as the final result of the clustering process.

Table 3. Final Clustering Results of Iteration 22

i-th Data	Village	C1	C2	C3	*Cluster* 1	*Cluster* 2	*Cluster* 3
1	Ajung	1793.09	6887.88	10772.66	1	-	-
2	Klompangan	1167.36	6055.42	9932.52	1	-	-
3	Mangaran	2729.10	2728.63	6561.76	-	2	-
4	Pancakarya	2996.38	2405.93	6244.24	-	2	-
5	Rowoindah	7371.49	2249.64	1963.27	-	-	3
6	Sukamakmur	915.26	4793.37	8666.56	1	-	-
7	Wirowongso	5119.86	768.28	4129.30	-	2	-
8	Ambulu	9762.25	4557.63	676.77	-	-	3
9	Andongsari	5182.58	1820.88	4544.83	-	2	-
10	Karanganyar	6629.18	1424.41	2475.43	-	2	-
...	...	...	...	...	...	...	...
240	Umbulrejo	10996.43	5788.49	1902.93	-	-	3
241	Umbulsari	10927.03	5719.12	1839.26	-	-	3
242	Ampel	5847.75	1259.11	3595.71	-	2	-
243	Dukuhdempok	6395.49	1191.89	2702.48	-	2	-
244	Glundengan	3357.46	1854.71	5743.02	-	2	-
245	Kesilir	6700.49	1496.26	2397.48	-	2	-
246	Lojejer	6935.54	1730.91	2162.42	-	2	-
247	Tamansari	4608.03	605.97	4490.13	-	2	-
248	Tanjungrejo	4804.11	414.51	4294.03	-	2	-
Total					**40**	**105**	**103**

Based on Table 2, which presents the results of the 22nd clustering iteration, it can be explained that the villages assigned to Cluster 1 include Ajung, Klompangan, Sukamakmur, and others. Meanwhile, the villages classified into Cluster 2 consist of Mangaran, Pancakarya, Wirowongso, and several others. Furthermore, the villages grouped into Cluster 3 include Rowoindah, Umbulrejo, Umbulsari, and additional villages.

2. Implementation Using the RapidMiner Application

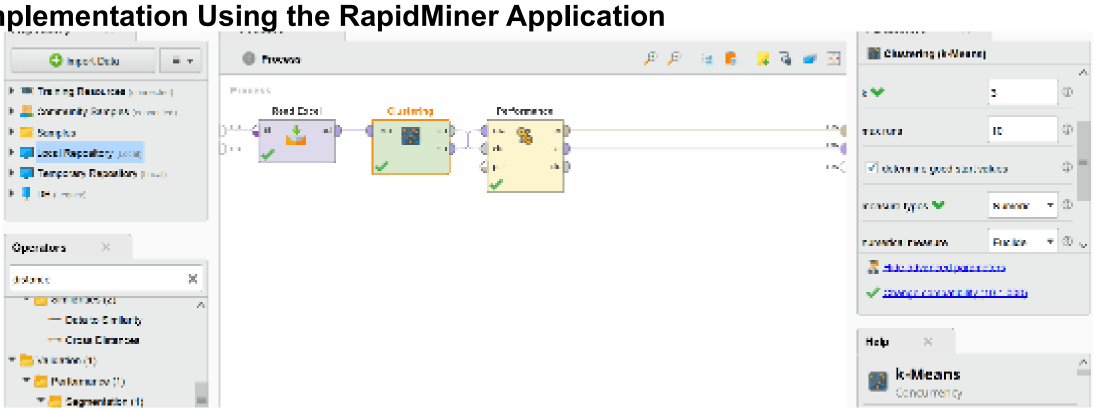

Figure 3. RapidMiner Workflow Steps

In Figure 3, the Read Excel operator is used to import the dataset in Excel format, which is then connected to the Clustering operator to perform clustering on the uploaded dataset using the K-Means algorithm. The output is subsequently connected to the Performance operator, which evaluates the clustering results and determines the optimal number of clusters using the Davies Bouldin Index (DBI) based on Euclidean distances.

In this study, the optimal number of clusters was determined using RapidMiner through DBI validation. The best cluster configuration is obtained from the clustering results that yield the lowest DBI value. Table 4 presents the DBI values tested for 2 to 10 clusters. The highlighted row in the table represents the optimal cluster configuration with the smallest DBI value.

Table 4. Davies–Bouldin Index (DBI) Values

Number of Clusters	**DBI Value**
2	0,608
3	**0,605**
4	0,633
5	0,606
6	0,621
7	0,63
8	0,706
9	0,629
10	0,618

Based on Table 4, the optimal number of clusters (k) is 3, with the lowest DBI value of 0.605. A lower DBI value indicates better clustering quality. The results of the clustering process using RapidMiner with the optimal number of clusters (k = 3) are presented in the figure below.

Row No.	Desa/Kelur...	cluster	Luas Tanam...	Luas Panen ...	Produksi (T...	Luas Tanam...	Luas Panen ...	Produksi (T...
1	Ajung	cluster_2	1482	1639	9534	1489	1448	8915
2	Klompangan	cluster_2	1418	1487	9083	1424	1296	8202
3	Mangaran	cluster_0	985	1123	6805	987	932	5822
4	Pancakarya	cluster_0	1045	1106	6511	1052	915	5851
5	Rowoindah	cluster_1	584	614	3584	583	423	2473
6	Sukamakmur	cluster_2	1289	1353	8181	1288	1164	7351
7	Wirowongso	cluster_0	860	879	6100	849	692	4078
8	Ambulu	cluster_1	125	214	1324	228	223	1369
9	Andongsari	cluster_0	481	563	3487	725	708	6980
10	Karanganyar	cluster_0	493	575	3568	567	555	3482
11	Pontang	cluster_0	385	475	2965	572	559	3507
12	Sabrang	cluster_0	583	664	4116	635	621	3886
13	Sumberejo	cluster_0	874	947	5072	875	953	5980

Figure 4. Clustering Results Using RapidMiner

Based on the clustering results shown in Figure 4.2, the data have been grouped into three clusters. The number of data points in each cluster, as well as the total number of data points, is presented in the figure below.

Cluster Model

```
Cluster 0: 105 items
Cluster 1: 103 items
Cluster 2: 40 items
Total number of items: 248
```

Figure 5. Number of Data Points per Cluster

Each cluster comprises villages that share similar characteristics, from Cluster 0 to Cluster 2. This demonstrates that the K-Means Clustering algorithm successfully grouped the villages into three distinct clusters. The distribution of data within each cluster is shown in the figure below.

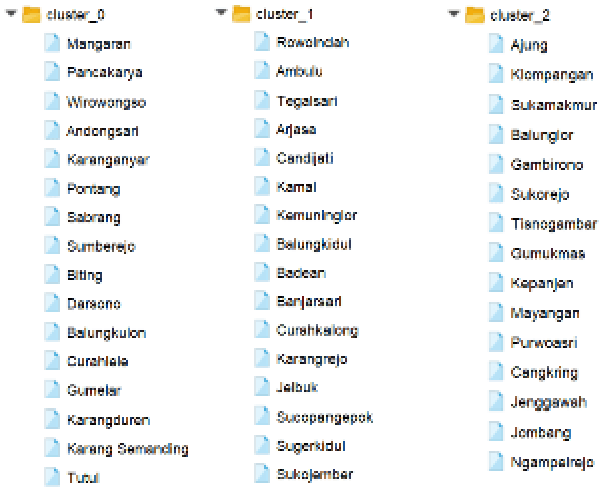

Figure 6. Cluster Distribution

3. Implementation Using Python

After the planted area, harvested area, and rice production data from 248 villages in Jember Regency for the years 2022 and 2023 were processed using RapidMiner, the next step was to implement the K-Means Clustering algorithm using the Python programming language. The implementation was carried out using the Visual Studio Code application equipped with the Jupyter extension and was saved in the Jupyter Notebook (.ipynb) format. The implementation using Python is presented as follows.

```python
dataset = pd.read_excel("data_padi.xlsx")

features = [
    'Luas Tanam (Ha.) 2022', 'Luas Panen (Ha.) 2022', 'Produksi (Ton) 2022',
    'Luas Tanam (Ha.) 2023', 'Luas Panen (Ha.) 2023', 'Produksi (Ton) 2023'
]
X = dataset[features].values

dbi_scores = []
k_values = range(2, 11)

for k in k_values:
    init_centroids = X[:k]  # centroid awal
    kmeans = KMeans(n_clusters=k, init=init_centroids, n_init=1, random_state=42)
    labels = kmeans.fit_predict(X)
    dbi = davies_bouldin_score(X, labels)
    dbi_scores.append(dbi)
    print(f'k={k}, DBI={dbi:.4f}')
```

Figure 7. DBI Code Snippet

Figure 8 illustrates the process of calculating the Davies Bouldin Index (DBI) value to determine the optimal number of clusters (k) using k values ranging from 2 to 10. The clustering process was performed by initializing the initial centroids based on the first k rows of the dataset presented in Table 4.1. The DBI value was calculated using the davies_bouldin_score function from the sklearn.metrics library.

```
k=2, DBI=0.6079
k=3, DBI=0.6053
k=4, DBI=0.6213
k=5, DBI=0.6722
k=6, DBI=0.6429
k=7, DBI=0.6502
k=8, DBI=0.7032
k=9, DBI=0.7332
k=10, DBI=0.7531
```

Figure 8. DBI Values Using Python

Based on Figure 8, it can be observed that the lowest DBI value was obtained at k = 3 with a DBI value of 0.6053. This indicates that the most optimal number of clusters for the dataset is three clusters.

```python
optimal_index = np.argmin(dbi_scores)
optimal_k = k_values[optimal_index]

init_centroids = X[:optimal_k]
kmeans_final = KMeans(n_clusters=optimal_k, init=init_centroids, n_init=1, random_state=42)
labels = kmeans_final.fit_predict(X)

dataset['Cluster'] = labels

for i in range(optimal_k):
    villages_list = dataset[dataset['Cluster'] == i]['Desa/Kelurahan'].tolist()
    print(f"\nCluster {i} (Total {len(villages_list)}):")
    for village in villages_list:
        print(f"- {village}")
```

Figure 9. K-Means Implementation Code Snippet

Figure 9 shows the implementation of the K-Means Clustering algorithm based on the optimal number of clusters (k) obtained from the previous DBI calculation. The optimal k value is stored in the variable optimal_k and is then used to initialize the initial centroids from the first k rows of the dataset. The clustering process is performed using the K-Means algorithm from the sklearn.cluster library. The clustering results with three clusters are presented in the following figure.

Cluster 0 (Total 40):	Cluster 1 (Total 105):	Cluster 2 (Total 103):
- Ajung	- Mangaran	- Rowoindah
- Klompangan	- Pancakarya	- Ambulu
- Sukamakmur	- Wirowongso	- Tegalsari
- Balunglor	- Andongsari	- Arjasa
- Gambirono	- Karanganyar	- Candijati
- Sukorejo	- Pontang	- Kamal
- Tisnogambar	- Sabrang	- Kemuninglor
- Gumukmas	- Sumberejo	- Balungkidul
- Kepanjen	- Biting	- Badean
- Mayangan	- Darsono	- Banjarsari
- Purwoasri	- Balungkulon	- Curahkalong
- Cangkring	- Curahlele	- Karangrejo
- Jenggawah	- Gumelar	- Jelbuk
- Jombang	- Karangduren	- Sucopangepok
- Ngampelrejo	- Karang Semanding	- Sugerkidul

Figure 10. Clustering Results Using Python

System Implementation

1. Software Implementation

The software specifications used in this study to implement the system are as follows:

a. Windows 11 64-bit Operating System
b. Google Chrome
c. XAMPP version 3.3.0
d. Visual Studio Code
e. RapidMiner Studio version 10.1

2. User Interface Implementation

This stage is a continuation of the system design that has been previously defined.

a. Home Page

The home page is the first page displayed when the application is launched. On this page, there is a login menu that allows the admin to access the admin dashboard. In addition, users can select the map menu to view the clustering map and the data menu to view tables of planting area, harvested area, and rice production.

Figure 11. Home Page

b. Map Page

This page displays the clustering output generated from the calculations performed by the system in the form of a web-based geographic information system (GIS) that can be accessed by users.

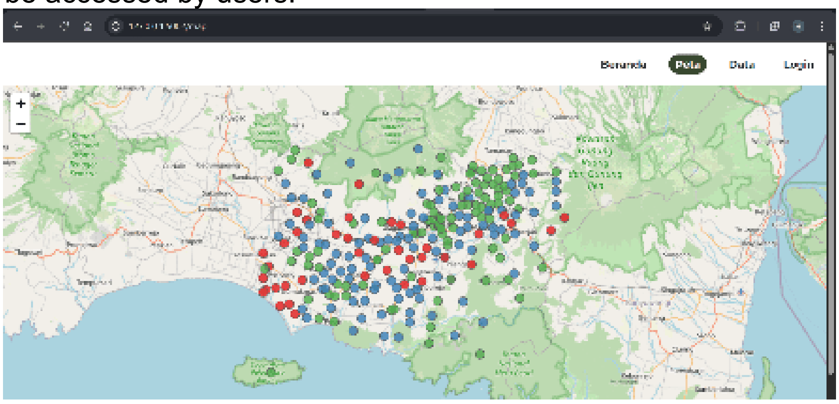

Figure 12. Map Page

c. Data Page

This page displays a table of data on planting area, harvested area, and rice production for all villages in Jember Regency, which can be accessed by users.

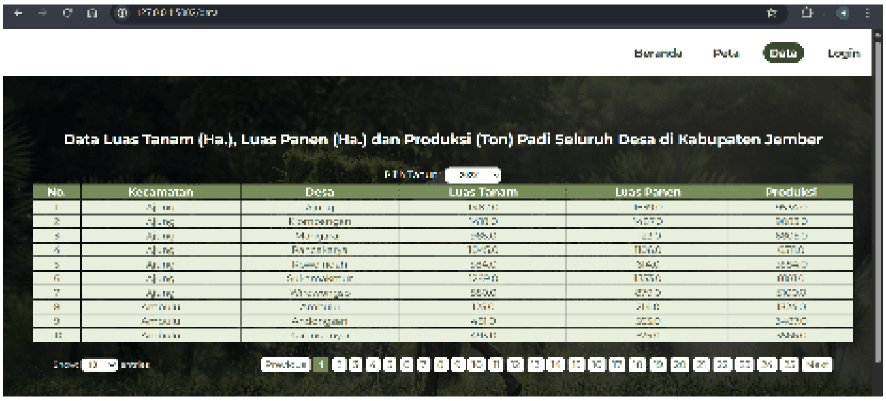

Figure 13. Data Page

d. Login Page

The login page is used by the admin to access the dashboard page. On this page, the admin is required to enter an email and password, then click the "Login" button.

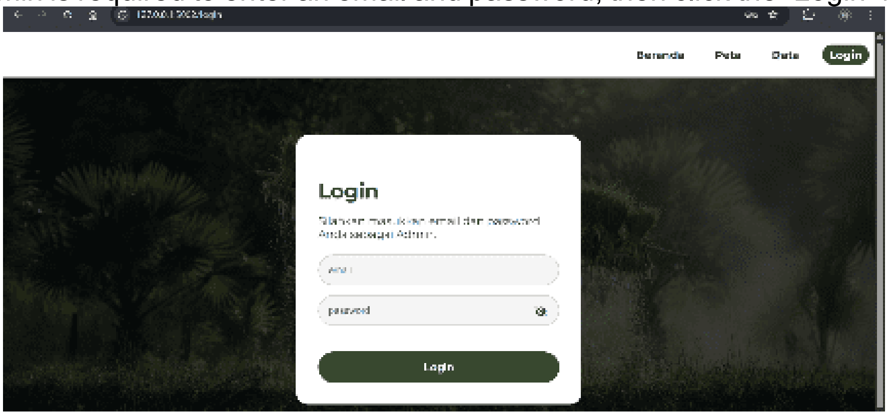

Figure 14. Login Page

e. Clustering Page

This page is used to perform clustering calculations using the K-Means Clustering algorithm. On this page, the admin can select the year to display the data that will be used for the calculation process. Then, the admin clicks the "Display Data" button to display the selected data.

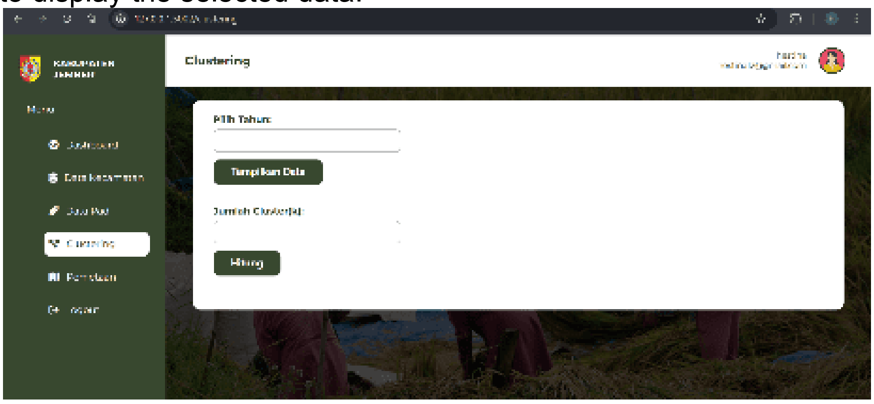

Figure 15. Clustering Page

After the data are displayed, the admin can enter the number of clusters (k) to be used, then click the "Calculate" button to proceed to the centroid page. On the centroid page, the admin inputs the initial centroid values before performing the calculation. After that, the admin clicks the "Calculate" button to proceed to the DBI page, which displays a table of DBI values for k ranging from 2 to 10.

After determining the optimal number of clusters, the admin can click the “Calculate Clustering” button to proceed to the clustering results page, which presents a table of the clustering calculation results. Subsequently, the admin can click “View Clustering Map” to access the clustering map page.

System Testing

1. System Testing for Users

Table 5. System Testing for Users

No.	Tested Feature	Test Scenario	Expected Output	Test Result
1.	Home Page	The user opens the main page of the application through the provided link.	The home page is displayed	Successful
2.	Map	The user clicks the “Map” menu	The clustering map is displayed	Successful
3.	Data	The user clicks the “Data” menu.	The rice data table is displayed.	Successful

2. System Testing for Admin

Table 6. System Testing for Admin

No.	Tested Feature	Test Scenario	Expected Output	Test Result
1.	Login	The admin enters a registered email and password, then clicks the “Login” button.	Redirected to the dashboard page.	Successful
2.	District Data	The admin clicks the “District Data” menu.	Regency data are displayed and can be managed (input, update, delete).	Successful
3.	District Data Form	The admin clicks “Add Data” to input district and village data to be used for the calculation.	The system displays the added district and village data.	Successful
4.	Rice Data	The admin clicks the “Rice Data” menu.	A table of planting area, harvested area, and rice production data per year is displayed and can be managed (input, update, delete).	Successful
5.	Rice Data Form	The admin clicks “Add Data” to input planting	The system displays the	Successful

No.	Tested Feature	Test Scenario	Expected Output	Test Result
		area, harvested area, and rice production data to be used for the calculation.	added rice data.	
6.	Form Clustering	Admin memilih tahun dan input jumlah cluster (k) yang akan digunakan untuk perhitungan.	The system displays the data table to be used for the calculation.	Successful
7.	Centroid Form	The admin inputs the initial centroid values to be used for the calculation, then clicks “Calculate DBI”.	The system displays the DBI value table and the optimal number of clusters (k).	Successful
8.	Clustering Process	The admin clicks “Calculate and View Results”.	The clustering result data table is displayed.	Successful
9.	Mapping	The admin clicks “Visualize in Map Form”.	The clustering result map is displayed.	Successful

5. CONCLUSION

Conclusion

Based on the results of the study on the clustering of planting area, harvested area, and rice production in each village in Jember Regency using the K-Means Clustering algorithm and the Davies Bouldin Index, several conclusions can be drawn as follows:

1. The K-Means Clustering algorithm was successfully applied to cluster 248 village data points using the variables of planting area, harvested area, and rice production for the years 2022 and 2023 into three clusters.
2. The K-Means Clustering algorithm can be effectively implemented for the process of clustering villages based on the variables of planting area, harvested area, and rice production through the stages of K-Means Clustering computation.

Recommendations

The system and algorithms used in this study still have several limitations that can be considered for further development in future research.

1. This study only used rice crop data in Jember Regency for the years 2022 and 2023. Therefore, future studies are expected to utilize data from a longer time span or include other types of crops in order to obtain a broader and more comprehensive overview.
2. The determination of the initial centroid points in this study was carried out sequentially from the first row of the data. It is recommended that future research further develop this process by applying other methods for initializing centroids to enable a comparative evaluation of which method yields better performance.

6. REFERENCES

Amalina, T., Pramana, D. B. A., & Sari, B. N. (2024). Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmu Komputer dan Sistem Informasi*, *3*(1), 1–10. https://doi.org/10.70340/jirsi.v3i1.88

Amna, S, W., Sudipa, I. G. I., Putra, T. A. E., Wahidin, A. J., Syukrilla, W. A., Wardhani, A. K., Heryana, N., Indriyani, T., & Santoso, L. W. (2023). Data Mining. In *PT GLOBAL EKSEKUTIF TEKNOLOGI*.

https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
Ananda, D., Rohimah, S., Susilo, B., Wulan, D., & Mustofa, A. (2022). Implementasi K-Means Dalam Pengelompokkan Data Akta Kelahiran di Indonesia: Implementation of K-Means in Grouping Birth Certificate Data in Indonesia. *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, 66–71. https://journal.irpi.or.id/index.php/sentimas/article/view/247%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/247/135
BPS Kabupaten Jember. (2023). *Luas Panen dan Produksi Padi di Indonesia 2023 (Angka Tetap)*. Berita Resmi Statistik. https://www.bps.go.id/pressrelease/2023/10/16/2037
Chaundhry, M., Shafi, I., Mahnoor, Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023). A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry*, *15*(9). https://doi.org/10.3390/sym15091679
Fahada, A. (2024). Pemanfaatan WEB GIS Untuk Pemetaan Dan Klasterisasi Jenis Hasil Perikanan Tangkap Menggunakan Metode K-Means Clustering. In *Doctoral dissertation, Universitas Malikussaleh*. https://rama.unimal.ac.id/id/eprint/591/%0Ahttps://rama.unimal.ac.id/id/eprint/591/5/AMALIA FAHADA_190170178_Pemanfaatan WEB GIS Untuk Pemetaan Dan Klasterisasi Jenis Hasil Perikanan Tangkap Menggunakan Metode K-Means Clustering.pdf
Fauzi, I. F., Gito Resmi, M., & Hermanto, T. I. (2023). *Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit*. *7*(2), 2598–8069.
Handayani, F. (2022). Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar. *Jurnal Teknologi dan Informasi*, *12*(1), 46–63. https://doi.org/10.34010/jati.v12i1.6733
Hua, A. K. (2015). *Sistem Informasi Geografi (GIS): Pengenalan kepada perspektif komputer Geographic Information System (GIS): Introduction to the computer perspective*. *1*(1), 24–31.
Prasetya, Y. Y., Faisol, A., & Vendyansah, N. (2021). Sistem Informasi Geografis Hasil Produksi Padi Di Kabupaten Malang Menggunakan Metode K-Means Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *5*(2), 806–814. https://doi.org/10.36040/jati.v5i2.3788
Putri, T. (2022). *SISTEM INFORMASI GEOGRAFIS PEMETAAN REKLAME DI KOTA MEDAN BERBASIS WEB*.
Putri, W., & Afdal, M. (2023). Penerapan Algoritma K-Means Untuk Pengelompokan Data Penyandang Disabilitas Di Kabupaten Rokan Hilir. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, *3*(1), 30–38. https://doi.org/10.57152/ijirse.v3i1.526
Wahidah, N., Juwita, O., & Arifin, F. N. (2023). Pengelompokkan Daerah Rawan Bencana di Kabupaten Jember Menggunakan Metode K-Means Clustering. *INFORMAL: Informatics Journal*, *8*(1), 22. https://doi.org/10.19184/isj.v8i1.29542
Wijayanto, S., & Yoka Fathoni, M. (2021). Pengelompokkan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means. *Jupiter*, *13*(2), 212–219.