Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 3, No. 2, September 2020, pp. 103 – 112
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Clustering Productivity of Rice in Karawang Regency Using The Fuzzy C-Means Method

[1]Suna Mulyani, [2]Betha Nurina Sari, [3]Azhari Ali Ridha
[1,2,3] Informatics Engineering, Faculty of Computer Science, Singaperbangsa Karawang University
Email: [1]suna.16212@student.unsika.ac.id, [2]betha.nurina@staff.unsika.ac.id, [3]azhari.ali@unsika.ac.id

Article Info

Article history:
Received Jul 30th, 2020
Revised Aug 10th, 2020
Accepted Oct 13th, 2020

Keyword:
Data Mining
Clustering
Fuzzy C-Means
Rice Productivity
Silhouette Coefficient

ABSTRACT

Rice is a major food commodity that has a strategic role in the development of community nutrition, agriculture, and the economy in Indonesia. Karawang Regency is known as a city of rice barns which is one of the largest rice-producing and supplying regions in the province of West Java and even Indonesia. The importance of rice as a staple food in Karawang Regency needs to ensure rice productivity remains stable. Data Mining is a data mining technique that produces an output in the form of knowledge. The purpose of this study is to classify the productivity of rice plants to know the area of high rice productivity in Karawang Regency. The data used in this study were 180 data from 30 districts. Data grouping will use the Fuzzy C-Means (FCM) algorithm which is a data clustering technique where the existence of each data point in a cluster is determined by the degree of membership. With Silhouette Coefficient evaluation techniques the results of clustering obtained in 2010, 2011, 2013, 2014, and 2015 show that the results of grouping have a good structure that is above 0.5. Only in 2012 showed that the grouping results had a weak structure of 0.49.

Corresponding Author:
Betha Nurina Sari,
Informatics Engineering Study Program,
Singaperbangsa Karawang University,
St. HS. Ronggo Waluyo, Puseurjaya, East Telukjambe District, Karawang Regency, West Java.
Email: betha.nurina@staff.unsika.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v3i2.10415

1. INTRODUCTION

Rice which has the Latin name Oryza sativa L is a major food commodity that has a strategic role in the development of community nutrition, agriculture, and the economy in Indonesia. Rice is a very important crop after wheat and corn crops, especially for the middle to lower classes of society, because the role of rice occupies the highest position as a staple food intake, the majority of which is needed by the Asian population, especially the Indonesian people.

Based on data obtained from the FAO (Food and Agriculture Organization of the United Nations) organization on rice production shows that international rice production is experiencing an increasing trend every year. And in 2014, Indonesia became the third-largest country with a contribution of 9.54% of rice production or equal to 70,846,465 tons [1]. Rice plants grow in the tropics and subtropics by having hot weather and high humidity and a 4-month rainy season. And in lowland rice requires an altitude of 0-650 masl with temperatures of 22-27 degrees Celsius while in highlands 650-1500 masl with temperatures of 19-23 degrees celcius. Karawang Regency is very suitable for rice planting because Karawang has an average air temperature of 27 degrees Celsius [2].

The agricultural sector as the main support of Indonesian society requires strong and rapid economic growth. This sector is also one of the main components in the government program in alleviating poverty. Indonesian agriculture in the past has achieved good results and made important contributions to the growth of the Indonesian economy. Karawang Regency is one of the rice-producing districts with a high level of productivity in the province of West Java, which we know as the city of rice barns. At present, Karawang

Regency has a land area of 175,327 hectares with details of 97,529 hectares of paddy fields and 77,798 hectares of dry land. Besides that, Karawang Regency can produce around 1.4 million tons of GKP in a year. With such a large amount of rice production, Karawang contributes up to 9% of rice produced by West Java Province [3].

Table 1. Productivity Data for 2010-2016

No.	Year	Harvested Area	Production (Ton)	Productivity (Kw / Ha)
1	2010	194.850	1.364.924	70,05
2	2011	197.005	1.459.406	74,08
3	2012	193.458	1.438.775	74,80
4	2013	195.929	1.481.466	75,61
5	2014	195.285	1.485.298	76,05
6	2015	190.725	1.524.842,55	79,99

(Source: Department of Agriculture, Forestry, Plantation and Animal Husbandry in Karawang Regency)

Based on Table 1.1, rice productivity data for 2006 to 2016 above, shows that from year to year rice productivity has experienced instability both in terms of harvested area, production, and productivity. Recorded in 2015, Karawang Regency was able to record production figures of 1,524,842.55 tons of GKP with a productivity of 79.99 even with a harvest area of 190,725. However, in 2016 rice production declined at 1,458,065.22 tons of GKP with a productivity of 75.39 although the harvested area increased by 193,403 more than the previous year. This proves the imbalance, which can be caused by several factors that affect the productivity of rice. As a city that holds the title of national rice granary, Karawang Regency must continue to increase rice production every year to balance the high population of Karawang which continues to grow. To meet the needs of rice, the Department of Agriculture, Forestry, Plantation, and Animal Husbandry in Karawang Regency must continue to work to optimize the yield of rice agriculture.

A method is needed to classify the yields based on harvested area and rice production in each district. The aim is to find out the cause of the instability of rice production and find out areas with not optimal rice production. So it needs to get attention and effective handling because it relates to the policy-making of aid distribution carried out by the Department of Agriculture, Forestry, Plantation, and Animal Husbandry in Karawang Regency. The policy taken must of course have relevance and be supported by knowledge derived from available data.

In the world of computer science, there is a technique to explore the added value of information that has not been known manually from a database by extracting data patterns to generate new knowledge called data mining. Data mining has its method based on the purpose of utilizing data sets namely estimation, prediction, classification, clustering, and association according to [2]. And one data mining method that can be used to group rice productivity is theclustering method. Clustering is a method of grouping data with the same characteristics into the same 'region' and data with different characteristics to another 'region'.

In a previous study conducted by M. Rosyid Ridlo, et al, entitled "Implementation of the K-Means Algorithm for Mapping Rice Harvest Productivity in Karawang District". From the results of the study stated that, based on the results of the trials and analysis described, rice harvest productivity data in Karawang Regency can be mapped using data mining grouping techniques into 3 groups consisting of harvest productivity exceeding the target, according to the target, and less than the target [4]. Mario also analyzes clustering of rice-productivity data in karawang using another algorithm, namely EM (Expectation-Maximization) clustering [2].

This research will be based on the use of the Fuzzy C-Means (FCM) algorithm. Fuzzy C-Means (FCM) is a data clustering technique in which the existence of each data in a cluster is determined by the membership value. And the basis for researchers using the Fuzzy C-Means (FCM) algorithm is also because in several studies using the Fuzzy C-Means (FCM) algorithm, it explains that Fuzzy C-Means (FCM) is better than the K-Means algorithm. As in the research of Rahman Syarif, et al, with the title "Comparison of K-Means Algorithms with Fuzzy C-Means (FCM) Algorithms in Clustering GPS-Based Modes of Transportation". From the results of the study stated that the grouping of GPS-based transportation modes, the results of tests conducted for 10 times showed that the accuracy of the Fuzzy C-Means algorithm is better than the K-Means algorithm [5]. Ramadhan,et al [6] compared k-means and Fuzzy C-Means for grouping user knowledge modeling data, the result is the FCM method is better. Ramadhan, et al [7] applied clustering of data on flood disasters in Indonesia with FCM. Parlina, et al [8] uses FCM for grouping employee data whose assessment standards are dynamic.

Based on the problems that have been described above, researchers will research to classify the productivity of rice plants in Karawang Regency by utilizing data from the Department of Agriculture, Forestry, Plantation, and Animal Husbandry in Karawang Regency using the Fuzzy C-Means (FCM) method.

2. RESEARCH METHOD

The methodology to be used in this study uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology with the steps according to Figure 1.

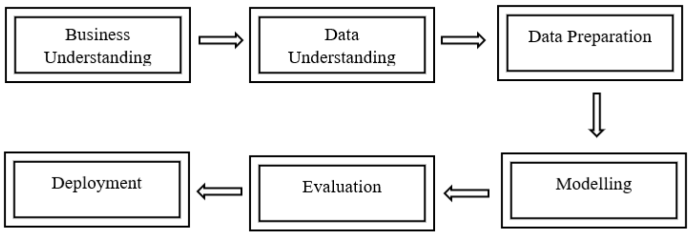

Figure 1. Research Methodology

The Business Understanding phase is needed to understand the substance of data mining activities that will be carried out as well as the needs of the business perspective. Data Comprehension Phase is done by collecting data, describing data, and exploring data. Then do the selection, cleaning, building, and integrating data at the data processing stage. Apply the Fuzzy C-Means algorithm at the Modeling stage. The Evaluation Phase will use the Silhouette Coefficient, which will be the interpretation phase of the results of data mining modeling using Fuzzy C-Means.

2.1. Data Mining

According to Turban et al, stated that data mining is a term used to describe the discovery of knowledge in a database. Data Mining is a process that uses statistics, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases [9].

2.2. Clustering

Clustering is a technique or method for grouping data. According to Tan, 2006 clustering is a process to group data into several clusters or groups so that data in one cluster has a maximum level of resemblance, and data between clusters has a minimum similarity [10].

2.3. Fuzzy C-Means

Fuzzy C-Means (FCM) is a data clustering technique in which the existence of each data point in a cluster is determined by the degree of membership.

According to Kusumadewi and Purnomo, the Fuzzy C-Means (FCM) algorithm is as follows [11]:

1. Determine the data to be in cluster X, in the form of a matrix of size n x m (n = number of sample data, m = attribute of each data). Xij = i-th sample data (i = 1,2, ..., n), j-attribute (j = 1,2, ..., m).
2. Determine if:
   a. Number of clusters (c)
   b. Rank (w)
   c. Maximum iteration (MaxIter)
   d. The smallest expected error (ζ)
   e. Initial objective function (p0 = 0)
   f. Initial iteration (t = 1).
3. Generate random numbers μik, i = 1,2, ..., n; k = 1,2, ..., c; as elements of the initial partition matrix U. Calculate the sum of each column (attribute):

$$Q_i = \sum_{k=1}^{c} \mu_{ik} \tag{1}$$

Then count:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \tag{2}$$

4. Calculate the center of the k-th cluster: Vkj, with k = 1,2, ..., c; and j = 1,2, ..., m.

$$V_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^{w} * X_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^{w}} \tag{3}$$

5. Calculate the objective function on the t-iteration, Pt:
The objective function is used as a looping condition to get the right cluster center. To obtain the tendency of data to enter which cluster in the final step. For the initial iteration, the value of t = 1.

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}(\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right] (\mu ik)^{w}) \tag{4}$$

6. Calculate partition matrix changes:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k}^{c}\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}} \tag{5}$$

7. Check stop condition::
   a. | Pt - Pt-1 | <ζ) or (t> MaxIter) then stop;
   b. If not, the iteration is increased t = t + 1, repeat step 4.
      Where:

V_{kj}	=	K-center cluster value on the jth variable
μ_{ik}	=	Elements of the U partition matrix or k-data membership function (k = 1,2,3, ..., n) on the I cluster (i = 1,2,3, ..., c)
X_{ij}	=	I data on the jth variable
w	=	Rank (weighting, w∈ [1, ∞)
n	=	Amount of data processed
i	=	Iteration
P	=	Objective function
m	=	Many variables

2.4. Sillhouette Coefficient

Silhouette Coefficient is a validation method used to test the quality of cluster results. After grouping, the next step is to evaluate the results of the grouping using cluster validation. Cluster validation is done to measure how well the grouping results are obtained [12]. The interpretation of silhouette coefficient values [13] can be seen in Table 2.

Table 2 Interpretation of silhouette coefficient

Interval Silhouette Coefficient	Interpretation
0.71-1.0	The strong structure has been discovered
0.51-0.70	The reasonable structure has been found
0.26-0.50	The weak structure may mock
<0.25	Not found the substantial structure

2.5. RStudio

RStudio programming tool or Integrated Development Environment (IDE) R language which has a better interface than RGUI. RStudio has 2 license versions, namely Open Source Edition and Commercial Edition.

3. RESULT AND ANALYSIS

The results of the research carried out were analyzing the Fuzzy C-Means algorithm in classifying the productivity of rice plants in Karawang regency viewed from the evaluation using the Silhouette Coefficient.

3.1. Business Understanding

The business objective based on the background in this study is to find out the results of grouping related to the level of rice productivity in Karawang Regency based on the characteristics of each district. The results of an accurate grouping can be used as a basis for decision making and policy for related agencies to increase rice productivity.

3.2. Data Understanding

The data understanding stage is the stage of collecting initial data and studying data and then analyzing what can be done with these data. Data collected consisted of sub-district data, seed average rainfall, average rainy day, planting area, production, harvested area, paddy area, paddy area, plant pest organisms (stem borer, rat, brown planthopper, and snail golden apple) and productivity. Table 2 is a description of the data from the data obtained.

Collect Initial Data

The data collected consisted of sub-district data, average seed rainfall, average rainy day, planting area, production, harvest area, rice field area, rice field area, plant pests (stem borer, rat, brown planthopper, and snail, mulberry) and productivity in the Department of Agriculture, Forestry, Plantation and Animal Husbandry, Karawang Regency, the data obtained from 2010-2015 below are some of preliminary data on Table 2.

Table 2. Dataset

YEARS	SUB-DISTRICT	SEED	AVERAGE RAINFALL	AVERAGE RAIN DAY	PLANTING AREA	PRODUCTION	HARVESTED AREA	AREA OF RAW RICE FIELDS	RICE FIELD AREA	STEM BENDERS	RAT	BROWN STEM PLANT HOPPER	GOLDEN APPLE SNAIL	PRODUCTIVITY
2010	Karawang Barat	113.55	102.8	11.2	4542	32327	4533	2243	2243	36	30	24	0	7.13
2010	Karawang Timur	96.1	102.8	11.2	3844	22927	3845	1847	1847	46	35	12	0	5.96
2010	Majalaya	112.4	84.7	4.7	4496	31165	4446	2233	2233	0	76	99	0	7.01
2010	Klari	135.85	200.7	13.2	5434	38495	5434	2392	2392	133	175	342	70	7.08
2010	Telukjambe Barat	106.4	221.9	10.8	4256	27403	4198	2108	2108	0	0	119	0	6.53
2010	Telukjambe Timur	69.275	221.9	10.8	2771	13218	2064	935	935	0	0	535	0	6.4
2010	Ciampel	53.45	159.3	10.3	2138	10797	1507	852	852	101	83	377	29	7.16
2010	Pangkalan	121.725	178.3633	8.94	4869	34425	4604	2341	2341	103	74	244	70	7.48
2010	Tegalwaru	101.65	178.3633	8.94	4066	26638	3837	1912	1912	147	92	152	77	6.94
2010	Rengasdengklok	100	140.4	5	4016	28848	4817	2026	2026	77	44	37	0	5.99
2010	Jayakerta	181.05	183.3	0	7242	48639	7142	3571	3571	40	145	0	0	6.81
2010	Kutawaluya	218.6	140.4	5	8744	63079	8744	4372	4372	108	78	74	0	7.21
2010	Batujaya	246.55	184.2	0	9862	68416	9862	4931	4931	100.4	124	441	37	6.94
2010	Tirtajaya	282.9	175.8	0	11316	76736	11316	5658	5658	222	247	36	0	6.78
2010	Pakisjaya	154.55	178.3633	0	6182	40820	5980	3166	3166	87	211	103	41	6.83
2010	Pedes	257.8	186.2	7.1	10312	71886	10312	5156	5156	100.4	173	180	74	6.97
2010	Cilebar	270.825	140.4	5	10833	72346	10833	5417	5417	132	36	54	69	6.68
2010	Cibuaya	197.3	135.9	5.3	7892	56926	7892	3946	3946	25	143	544	0	7.21
2010	Cikampek	32.05	336.9	17	1282	8474	1282	641	641	6	39	74	10	6.61
2010	Purwasari	80.55	296.1	17	3222	19037	3222	1611	1611	45	31	377	0	5.91
2010	Tirtamulya	126.05	188.1	16.1	5042	36769	5062	2521	2521	100	279	644	4	7.26
2010	Jatisari	205.2	221.5	11.7	8208	52203	8096	4104	4104	59	181	364.0333	0	6.45
2010	Banyusari	190	168.7	8.1	7600	45564	7600	3814	3814	121	102	641	29	6
2010	Kota Baru	80.175	324.8	13.9	3207	23182	2972	1466	1466	146	192	391	0	7.8
2010	Cilamaya Kulon	228.5	162.3	9	9140	60863	9494	4570	4570	73	200	364.0333	166	6.41
2010	Cilamaya Wetan	269.2	178.3633	0	10768	44535	9486	5321	5321	121	156	439	115	6.99
2010	Telagasari	198.45	77.9	0	7838	45497.47	7738	3919	3919	31	94	364.0333	0	6.99
2010	Lemahabang	189.9	172.4	0	7596	53571	7596	3798	3798	64	128	808	0	7.05
2010	Rawamerta	209.7	176.1	9.1	8388	58469	8382	4191	4191	85	91	170	151	6.98
2010	Tempuran	313.1	145	6.8	12524	80430	12554	6467	6467	100	84	235	158	6.41

Describe Data

This stage is the stage to familiarize yourself with the data that has been collected and try to find initial insights about what information can be obtained from the data that has been collected. The following is a description of the attributes of the initial data obtained, which can be seen in Tabel 2.

Table 3. Data Description

Attribute	Data Type	Information	Unit
Year	String	Data used from years 2010-2015	
sub-district	String	Districts include: Karawang Barat, Karawang Timur, Majalaya, Klari, Telukjambe Barat, Telukjambe Timur, Ciampel, Pangkalan, Tegalwaru, Rengasdengklok, Jayakerta, Kutawaluya, Batujaya, Tirtajaya, Pakisjaya, Pedes, Cilebar, Cibuaya, Cikampek, Purwasari, Tirtamulya	

Attribute	Data Type	Information	Unit
		, Jatisari, Banyusari, Kota Baru, Cilamaya Kulon, Cilamaya Wetan, Telagasari, Lemahabang, Rawamerta, Tempuran.	
Seed	Numeric	The number of seeds to be planted.	Ton
Average Rainfall	Numeric	Data that includes the average amount of rain in each district.	Hh
Average Rainy Day	Numeric	The average number of rainy days in each district.	(rainy day)
Planted Area	Numeric	Amount of land to be planted with rice.	Mm (millimeter)
Production	Numeric	Amount produced during harvest.	Ha (hectare)
Harvested Area	Numeric	The area of plants collected after the plants are old enough.	Ton
Area of Raw Rice Fields	Numeric	The amount of land area that can be planted with rice.	Ha (hectare)
Rice Field Area	Numeric	Amount of land planted with rice.	Ha (Hectare)
Plant Disturbing Organisms	Numeric	OPT includes: stem borer, rat, brown stem plant hopper, golden apple snail	
Productivity	Numeric	The amount of rice productivity produced	Ha (hectare)

Explore Data

At this stage is data exploration. Data obtained from the Department of Agriculture, Forestry, Plantation, and Animal Husbandry of Karawang Regency were explored using diagrams and tables, traced to find out all the possibilities for each attribute. Figure 3 is an exploration of rice productivity data per the year 2010 to 2015.

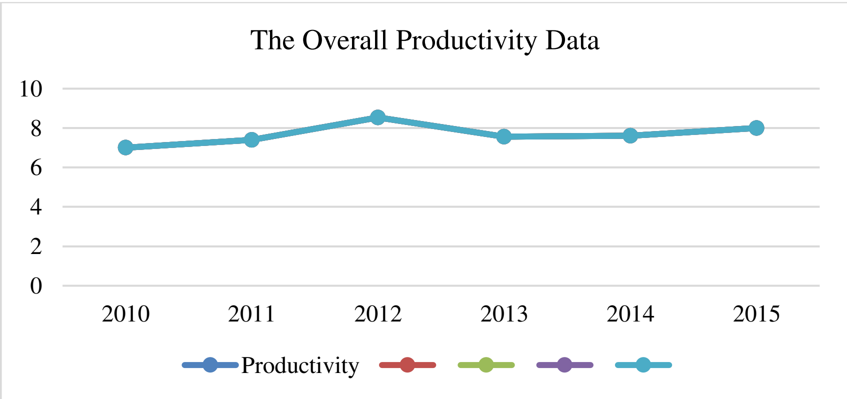

Figure 2 Explore Data

Figure 2 shows the overall productivity data per year from 2010 to 2015 showing the instability of rice productivity in Karawang Regency, where an increase in rice productivity occurred in 2012 and a decrease occurred in 2013.

Verify Data Quality

At this stage, data evaluation and data content checking will be carried out. Data from previous research results have no missing data (Missing Value) so that the data can be processed according to the needs or knowledge to be generated.

3.3. Data Preparation

At the data preparation stage, the variables to be used and the next stage will be determined. The data that will be used is the data from 2010 to 2015. The selection of data that is suitable for grouping needs is 14 attributes, namely districts, seeds, average rainy days, average rainfall, planting area, production, harvested area, standard area rice fields, paddy fields, pests (stem borer, rats, brown plant hopper and mulberry snails) and productivity. This research only determines which groups produce the best rice. Before using the Fuzzy C-Means algorithm, data that has not been processed before need to go through several steps, one of which is data preparation. There are several stages, namely selection, cleaning, building, and integrating data.

Select Data

Selection of suitable data for the needs of grouping is 14 attributes, namely sub-district, seeds, average rainy days, average rainfall, planting area, production, harvested area, raw area of rice fields, rice field area, pest cocoa sticks, and mulberry snails) and productivity. For this research, only determine which groups produce rice with the best productivity.

Clean Data

1. Missing Value

Below is a description of the seed variables, average rainy days, average rainfall, planting area, production, harvested area, rice field area, rice field area, and pest pests using RStudio tools to check Missing Value. And in Figure 3 shows that there is no missing data (missing value).

```
      SUB-DISTRICT               SEED    AVERAGE RAINFALL      AVERAGE RAINY DAY
                 0                  0                   0                      0
      PLANTED AREA         PRODUCTION     HARVESTED AREA AREA OF RAW RICE FIELDS
                 0                  0                   0                      0
   RICE FIELD AREA        STEM BORDER                    RAT BROWN STEM PLANT HOPPER
                 0                  0                   0                      0
GOLDEN APPLE SNAIL       PRODUCTIVITY
                 0                  0
```

Figure 3 Missing Value

2. Outlier

Outlier handling is carried out on the data on an annual basis so that it can be seen more clearly in the relationship between regions including good and/or bad clusters each year. So that the related agencies can find out whether the performance results have been carried out well or not in dealing with areas that have decreased rice productivity.

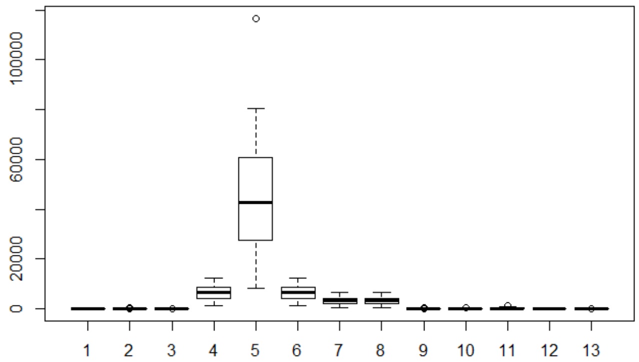

Figure 4 Check Outlier

Therefore, to minimize or overcome these outliers, the value that becomes the outlier will be replaced with the mean. The following is the boxplot on the data after handling outliers.

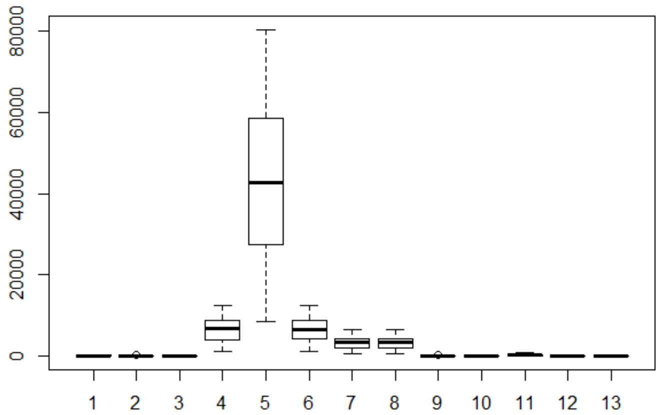

Figure 5 Outlier Handling

Figure 5 shows the boxplot of the results after handling the variables that have the Outlier value. It can be seen in the variables, the average rainy day, production, OPT (rats, brown planthopper), and net productivity of the Outliers. While the average rainfall and pest (stem borer) have been reduced considerably and the value outside the Outlier is not too far from the concentration of data.

Integrate Data

Because the data obtained is only in one source, so there is no need to re-integrate the data.

Format Data

The data obtained from the Department of Agriculture is in the .xlsx format, so it needs to be converted into a .csv format. This is to simplify the modeling process that will be carried out at RStudio.

3.4. Modelling

Data will be clustered per year to see more clearly the linkages between regions that are included in the good cluster and / or bad cluster each year. So that the related department can find out whether the performance results have been implemented well or not in handling areas that have decreased rice productivity. Following the process of clustering rice productivity data using the Fuzzy C-Means algorithm using RStudio tools can be seen in Figure 6.

```
#Modeling Fuzzy C-Means
res.fcm <- fcm(x, centers = 3)
as.data.frame(res.fcm$u)[1:6,]
res.fcm$v
summary(res.fcm)
```

Figure 6. Source Code Fuzzy C-Means Clustering

Figure 7 is the result of the clustering process using the Fuzzy C-Means algorithm. Cluster 1 is marked in blue, cluster 2 is marked in yellow, and cluster 3 is marked in gray.

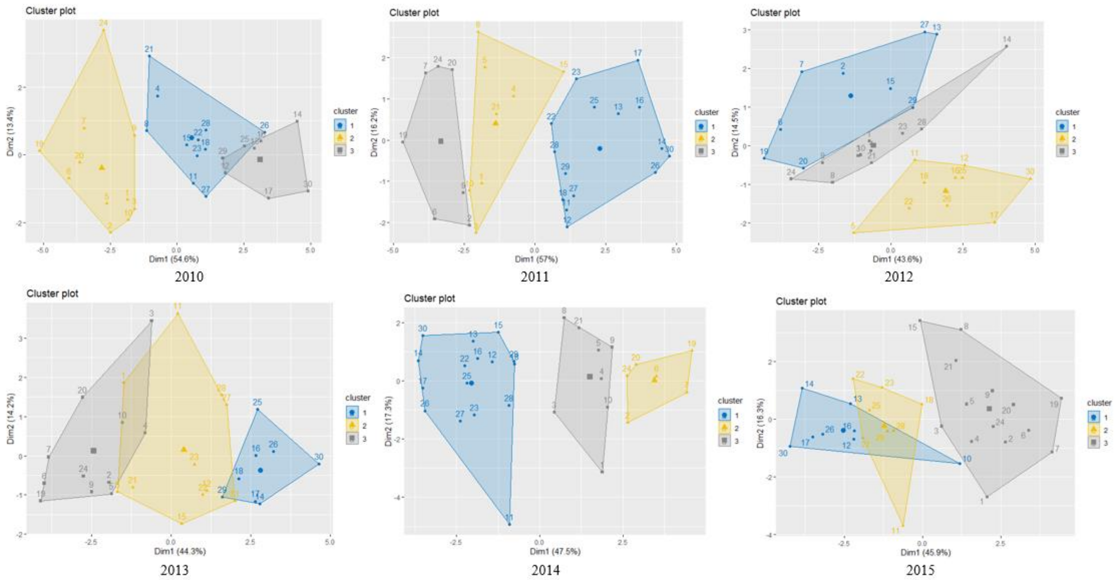

Figure 7. Clustering Results

3.5. Evaluation

This stage is carried out to test the suitability of the results of the modeling phase with the business understanding to be achieved. The evaluation uses Silhouette Coefficient values to see the quality and strength of the cluster. Below is the result of the Silhouette Coefficient value using RStudio tools.

Table 4. Silhouette Coefficient Value

Years	*Cluster*	Cluster Members	*Average Silhouette Coefficient*
2010	1	11	0.5
	2	11	
	3	8	

Years	*Cluster*	Cluster Members	*Average Silhouette Coefficient*
	1	15	
2011	2	8	0.57
	3	7	
	1	9	
2012	2	10	0.49
	3	11	
	1	8	
2013	2	11	0.54
	3	11	
	1	16	
2014	2	6	0.6
	3	8	
	1	8	
2015	2	8	0.55
	3	14	

Table 3 shows that 5 out of 6 data per year have a good interpretation of the results of the grouping, namely in 2010, 2011, 2013, 2014, and 2015. And the 2012 data is the only one that has the results of a grouping with a weak interpretation.

3.6. Deployment

At the deployment stage, a report will be made on the results of data mining activities and presented in the form of a description to be easily understood.

4. CONCLUSION

Fuzzy C-Means algorithm can be implemented for the process of grouping rice productivity areas in Karawang Regency with RStudio tools. The results of the evaluation of the Fuzzy C-Means algorithm for clustering rice productivity in Karawang Regency can be known the average value of the Silhouette Coefficient on the results of the cluster data for 2010-2015 that 5 out of 6 years namely in 2010, 2011, 2013, 2014 and 2015 shows that the grouping results have a good structure above 0.50. Only in 2012 showed that the results of the grouping had a weak structure with an average Silhouette Coefficient value of 0.49.

ACKNOWLEDGEMENTS

The author would like to thank the Department of Agriculture, Forestry, Plantation, and Animal Husbandry in Karawang Regency for providing data and facilitating this research.

REFERENCES

[1]. Yanuarti, Astri Ridha., And Mudya Dewi Afsari. Commodity Profile of Staple Goods and Important Commodities of Rice Commodities. Ministry of Trade of the Republic of Indonesia.

[2]. Wibowo, Mario Tri. Clustering Analysis of Rice Productivity Data in Karawang using EM (Expectation-Maximization) Clustering. Essay. Karawang: Singaperbangsa Karawang University. 2019.

[3]. Karawang Regency Forestry Plantation and Animal Husbandry Department, 2015

[4]. Ridlo, M. Rosyid., Defiyanti, Sofi, Primajaya, Aji. Implementation of K-Means Algorithm for Mapping Rice Harvest Productivity in Karawang Regency. Department of Electrical Engineering and Information Technology, UGM UGM. 2017; 426-433.

[5]. Syarif, Rahman., Furqon, Muhammad Tanzil., Adinugroho, Sigit. Comparison of K-Means Algorithms with Fuzzy C Means (FCM) Algorithms in GPS-Based Transportation Clustering Modes. Journal of Information Technology and Computer Science Development. 2018; 2 (10); 4107-4115.

[6]. Ramadhan, Aditya., Efendi, Zuliar., Mustakim. Comparison of K-Means and Fuzzy C-Means for Grouping User Knowledge Modeling Data. National Seminar on Information Technology, Communication and Industry (SNTIKI). 2017; 219-226.

[7]. Ramadhan, Aditya., Mustakim., Handinata, Rizki., Implementation of the Fuzzy C-Means Algorithm for Grouping Flood Disaster Areas. National Seminar on Information, Communication, and Industry Technology (SNTIKI) .2019; 171-177.

[8]. Parlina, Iin., Prof.Herman Mawengkang, Dr. Syahril Efendi, S.Si, M.IT. Performance Analysis of Fuzzy Tsukamoto clustering algorithm with Fuzzy C-Means. National Journal of Information and Network Technology. 2017; 1 (2); 90-94.

[9]. Nofriansyah, Dicky. Concept of Data Mining Vs Decision Support System. Yogyakarta: Deepublish. 2014.

[10]. Irwansyah, E, and Muhammad Faisal. Advanced Clustering Theories and Applications. Yogyakarta: Deepublish. 2015

[11]. Rizal, Annas Syaiful and Hakin, R.B Fajriya. K-Means Cluster and Fuzzy C-Means Cluster Methods. Proceedings of the National Seminar on Mathematics and Mathematics Education at UMS. 2015; 643-657.

[12]. Nurhayati, Dedeh. Clustering of the Rice Productivity Region in Karwang District using the K-Medoids Algorithm. Karawang: Singaperbangsa Karawang University. 2018.

[13]. Struyf, Anja, Mia Hubert, dan Peter J.Rousseeuw. Clustering in an Object-Oriented Environment. Journal of Statistical Software. 1997; 1(4) : 1-30.

BIBLIOGRAPHY OF AUTHORS

Suna Mulyani, born in Karawang on August 01, 1998. Graduated S1 from the Informatics Engineering study program, Singaperbangsa Karawang University in 2020. Now the writer live in Karawang and the author has an interest in data science.

Betha Nurina Sari M.Kom., born in Kediri on October 23, 1989, Graduated S1 from the Computer Science study program, Brawijaya University in 2012. In 2013, the author continued her Masters in Computer Science at the University of Indonesia and graduated in 2015. Now The writer is a lecturer in Informatics Engineering, Faculty of Computer Science, Singaperbangsa Karawang University, and is a member of the data mining research group.

Azhari Ali Ridha, S.Kom., M.M.SI born in Karawang on September 15, 1980 . Graduated S1 from Computer System study program, Gunadarma University in 2003. The author also continued his Masters degree in Information System Software study program at Gunadarma University and graduated in 2019. The writer is a lecturer in Informatics Engineering, Faculty of Computer Science, Singaperbangsa Karawang University. The author teaches courses in Organization and Computer Architecture, Human Computer Interaction, and Computer applications.