

# The Clustering of Paddy Fields Using Machine Learning Algorithms in the Province of South Sumatera

Suryati<sup>1\*</sup>, Ahmad Sanmorino<sup>1</sup>, Ermatita<sup>2</sup>

<sup>1</sup>Faculty of Computer Science, Universitas Indo Global Mandiri, Palembang, Indonesia

<sup>2</sup>Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia

\*Corresponding author. Email: suryati@uigm.ac.id

## ABSTRACT

In this paper, we were clustering paddy fields in the province of South Sumatera using machine learning algorithms. We chose machine learning algorithms because this algorithm is very relevant to the problem to be studied. The contribution of this study is in the form of result from the clustering of paddy fields which used as input or consideration for the government in determining policies related to the farming area. Policies that can use the results of the clustering of paddy fields include determining the most appropriate area for expanding paddy fields in the province of South Sumatera. The results of clustering using machine learning techniques show the expansion of paddy fields could be done in 13 districts or as much as 76.4% of the total districts in the province of South Sumatera.

**Keywords:** paddy fields, clustering, machine learning, determining policies

## 1. INTRODUCTION

Paddy fields in South Sumatera province are spread in several regencies and cities. The uneven area of paddy fields in South Sumatera province is the reason for this study. Through this study, we try to classify paddy fields in the province of South Sumatera so that they could be taken into consideration for the government in determining policies related to paddy field expansion. We realize that the extent of paddy fields in an area could be influenced by various factors, but with our study, it is expected to be an alternative input for the government as the policymaker.

Table 1 shows the distribution of paddy fields in South Sumatera province is as follows:

**Table 1** The Distribution of paddy fields in the province of South Sumatera [1]

Districts	Rice Field Area (Hectares)			
	Irrigation		Non Irrigation	
	2015	2016	2015	2016
OKU	3.24 4	3.24 4	5.628	5.628
OKI	650	5.65 0	185.34 8	182.67 8
Muara Enim	6.39 5	6.37 7	20.622	21.978
Lahat	15.8 45	16.8 26	1.680	1.652
MURA	13.4	13.3	17.030	17.295

Districts	Rice Field Area (Hectares)			
	Irrigation		Non Irrigation	
	2015	2016	2015	2016
	21	51		
MUBA	0	0	66.810	63.011
Banyuasin	0	0	226.51 8	226.51 8
OKU Selatan	16.0 99	16.1 70	1.941	1.870
OKU Timur	43.5 06	44.2 64	42.114	41.624
Ogan Ilir	0	0	67.627	67.622
Empat Lawang	13.1 05	13.1 05	986	986
Pali	0	0	6.579	8.459
Muratarra	415	500	6.716	6.631
Palembang	0	0	6.189	6.078
Prabumulih	0	0	700	700
Pagar Alam	3.44 0	3.44 0	0	0
L. Linggau	1.63 7	1.66 5	257	257

Could be seen in table 1, there are districts that have very large paddy fields, but on the other hand, there are also areas with very limited paddy fields. A balance of paddy fields between one region and another is needed so that the distribution and sale of rice could be optimized because the delivery time is faster. Another advantage is that rice does not accumulate in warehouses and is not eaten by termites because distribution is done quickly. In other words,

distribution could be done more evenly and at the consumer level, can meet the basic needs of the community. Studies related to paddy fields and rice distributions are still very relevant at this time, as seen in the following chart:

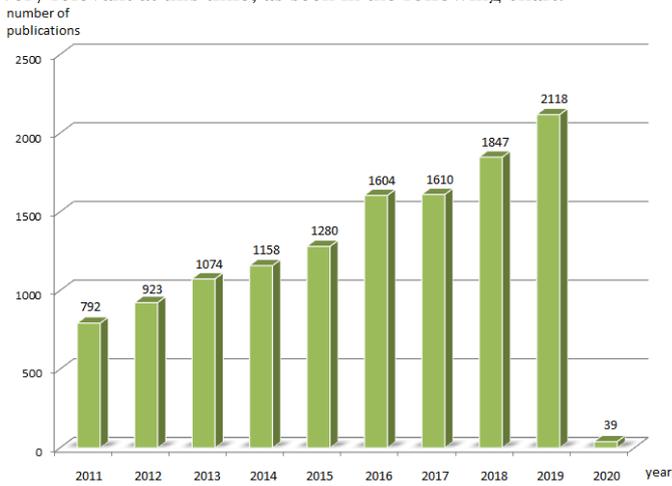


Figure 1 Trend Research related to paddy fields [2]

Table 2 shows the distribution of types of publications related to paddy fields in the Science Direct database [2]:

Table 2 The Distribution of types of publications related to paddy fields in the ScienceDirent database

No	Article type	Number of Publication
1	Review articles	1.314
2	Research articles	16.607
3	Encyclopedia	339
4	Book chapters	1.807
5	Conference abstracts	251
6	Book reviews	115
7	Case reports	21
8	Conference info	31
9	Correspondence	35
10	Other	1.526

Based on the trends and distribution of publications (Table 2), it could be understood that research related to paddy fields is still a hot issue in the last decade, at least until 2020. Some researches that are quite interesting and related to the research topics we do include discussion on the classification of paddy fields for various needs [3][4][5]. Then other researchers also conducted research on the metal content in polluted rice fields [6], there are also studies that study the Archaeal ether lipid composition in ancient and modern paddy soils [7]. Research related to paddy fields takes a significant part in the agriculture and animal husbandry sectors. Our research is our second research related to agriculture and animal husbandry [8]. Utilization of technology such as smartphones can also be done in agriculture [9], it does not rule out the research we are doing this can utilize research that we have done before [10].

## 2. METHODOLOGY

The research steps we undertake are as follows:

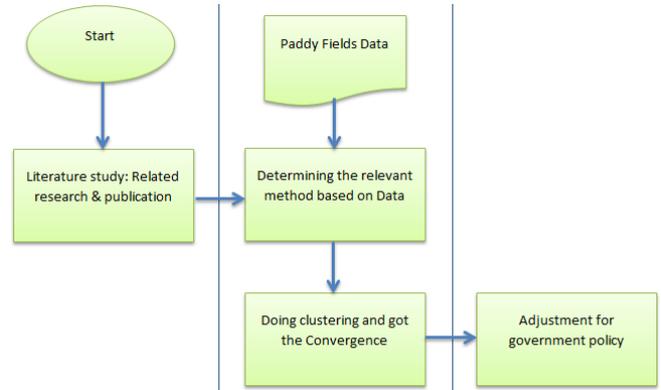


Figure 2 The Research Steps

First, we conducted a study of related research literature and publications that had been carried out by other researchers. Then based on rice field data (secondary data) obtained from BPS South Sumatera province [2], we determined the machine learning method that is relevant for classifying paddy fields. Clustering is done iteratively until it reaches a convergent condition, i.e. every member of each cluster does not change anymore. The last step is to adjust the results of clustering would be considered by the government in determining policy.

The use of clustering algorithms is still very relevant for solving various problems. This could be seen from several recent studies that use clustering algorithms for various needs [11][12]. The clustering method is a machine learning technique that could be used to group data, especially past data(related to data mining) that could be used for future solutions [13][14]. The steps of the clustering method that will be used in this study are as follows [15][16]:

- a) Define the dataset to cluster
- b) Determine the number of classes/clusters needed
- c) Determine the vector that will be used to cluster
- d) Determine the cluster centroid randomly from the dataset
- e) Calculate each difference in the value of the dataset against centroid using the distance formula
- f) Determine the clusters for each dataset based on the calculation done in step 5
- g) Perform steps 1 to 6 repeatedly (n iteration) until it reaches a convergent condition.

## 3. RESULTS AND DISCUSSION

Based on the data in Table 1, we perform clustering using the data mining algorithm and the results could be seen in the following Table 3:

**Table 3** The Final cluster of paddy fields in South Sumatera province

Districts	Vector	Cluster ID	Cluster Centroid
OKU	3244, 5628	0	6223.16, 5961.16
OKI	5650, 182678	1	2825, 204598
Muara Enim	6377, 21978	0	6223.16, 5961.16
Lahat	16826, 1652	0	6223.16, 5961.16
Mura	13351, 17295	0	6223.16, 5961.16
Muba	0, 63011	2	14754.66, 57419
Banyuasin	0, 226518	1	2825, 204598
OKU Selatan	16170, 1870	0	6223.16, 5961.16
OKU Timur	44264, 41624	2	14754.66, 57419
Ogan Ilir	0, 67622	2	14754.66, 57419
Empat Lawang	13105, 986	0	6223.16, 5961.16
Pali	0, 8459	0	6223.16, 5961.16
Muratar	500, 6631	0	6223.16, 5961.16
Palembang	0, 6078	0	6223.16, 5961.16
Prabumulih	0, 700	0	6223.16, 5961.16
Pagar Alam	3440, 0	0	6223.16, 5961.16
Lubuk Linggau	1665, 257	0	6223.16, 5961.16

This clustering is based on the area of paddy fields by using 2 vectors namely irrigation and non-irrigation rice fields. Could be seen based on the results of the clustering of paddy fields in South Sumatera province divided into 3 groups. The first group had a percentage of 76.4%, the second group was 11.7%, and the third group was 11.7%. The first group members are Ogan Komering Ulu, Muara Enim, Lahat, Musi Rawas, Ogan Komering Ulu Selatan, Empat Lawang, Pali, North Musi Rawas, Palembang, Prabumulih, Pagar Alam, and Lubuk Linggau. The second group consists of Ogan Komering Ilir and Banyuasin. While the third group, also only has 2 members, Ogan Komering Ulu Timur and Musi Banyuasin. The third group is the district that has the most extensive rice fields compared to other districts, while districts in the first group could be used as alternatives for future expansion of paddy fields.

The results of this clustering could be input or consideration for the government in determining policies specifically

related to the expansion or expansion of paddy fields in the province of South Sumatera.

#### 4. CONCLUSION

The clustering results showed that most paddy fields were in 4 districts namely Ogan Komering Ilir, Banyuasin, East Ogan Komering Ulu and Musi Banyuasin. So if the government wants to expand paddy fields can study and choose from 13 other districts in the province of South Sumatera. We hope that the results of our study could be taken into consideration for the government in determining policies specifically related to the expansion of paddy fields in the province of South Sumatera.

#### ACKNOWLEDGMENT

We would like to thank you Universitas Indo Global Mandiri for supporting this work.

#### REFERENCES

- [1] Badan Pusat Statistik Provinsi Sumatera Selatan, "Luas Lahan Sawah Menurut Kabupaten/Kota dan Jenis Pengairan di Provinsi Sumatera Selatan, 2015-2016," Badan Pusat Statistik Provinsi Sumatera Selatan, 2016.
- [2] Elsevier, ScienceDirect: Journals & Books, Access on September 13, 2019.
- [3] B.S. Anami, N.N. Malvade, S. Palaiah, "Classification of yield affecting biotic and abiotic paddy crop stresses using field images," Information Processing in Agriculture, 2019.
- [4] P. Junzheng, J. Changying, "Fuzzy clustering of paddy field soils in South China," Journal of Terramechanics, Volume 35, Issue 1, 1998.
- [5] S. Ramesh, D. Vydeki, "Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm," Information Processing in Agriculture, 2019.
- [6] H. Song, et al., "Metal distribution and biological diversity of crusts in paddy fields polluted with different levels of cadmium," Ecotoxicology and Environmental Safety, Volume 184, 2019.
- [7] F. Zheng, et al., "Changes in archaeal ether lipid composition in response to agriculture alternation in ancient and modern paddy soils," Organic Geochemistry, 2019.

- [8] A. Sanmorino, Isabella, "The design a system of retention and control on broiler farms based on the flow of data," 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Yogyakarta, September 2017.
- [9] H. Fang, Y. Ye, W. Liu, S. Wei, L. Ma, "Continuous estimation of canopy leaf area index (LAI) and clumping index over broadleaf crop fields: An investigation of the PASTIS-57 instrument and smartphone applications," *Agricultural and Forest Meteorology*, Volumes 253–254, 2018.
- [10] A. Sanmorino, R.M. Fajri, "The Design of Notification System on Android Smartphone for Academic Announcement," *International Journal of Interactive Mobile Technologies (IJIM)*, 2018.
- [11] R. Wang, et al., "Grid density overlapping hierarchical algorithm for clustering of carbonate reservoir rock types: A case from Mishrif Formation of West Qurna-1 oilfield, Iraq," *Journal of Petroleum Science and Engineering*, Volume 182, 2019.
- [12] E. Zhu, Y. Zhang, P. Wen, F. Liu, "Fast and stable clustering analysis based on Grid-mapping K-means algorithm and new clustering validity index," *Neurocomputing*, Volume 363, 2019.
- [13] C.d.S. Garcia, et al., "Process mining techniques and applications – A systematic mapping study," *Expert Systems with Applications*, Volume 133, 2019.
- [14] K. Gibert, "Which method to use? An assessment of data mining methods in Environmental Data Science," *Environmental Modelling & Software*, Volume 110, 2018.
- [15] C. Zhu, C.U. Idemudia, W. Feng, "Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques," *Informatics in Medicine Unlocked*, 2019.
- [16] K. Tian, J. Li, J. Zeng, A. Evans, L. Zhang, "Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm," *Computers and Electronics in Agriculture*, 2019.