STMIK Amik Riau

JAIA – Journal Of Artificial Intelligence And Applications
Journal Homepage : http://jurnal.sar.ac.id/index.php/JAIA/

Data Mapping System Of Riau Province Fire Potential Using K-Means Clustering Method

Rahmaddeni, Andi Kurnianto
STMIK Amik Riau, Purwodadi Indah, Pekanbaru

Article Info

Received : July 20, 2020
Revised : August 20, 2020
Accepted : August 30, 2020

Keywords :

Data Mining
K-Means Clustering
Hotspots
Visualization of the Mapping

Abstract

According to a report from the Riau Province BLHK states that hotspots in Riau Province are always present every year despite the number of hotspots that have been suppressed (http://dislhk.riau.go.id/). One of the causes is the frequent land clearing occurred as a trigger from a hotspot in Riau Province. There is a need for countermeasures as soon as possible to overcome the problem of hotspots that will cause forest fires. These problems need to be watched out quickly, one of which is to know in advance the hotspots that are likely to emerge based on existing data. Data mining processing is very suitable to be applied in order to produce relevant data to find out the possibility of hotspots. In this study the data grouping was done in the form of a visualization of hotspot mapping using the K-means Clustering method. The parameters used include 3 number of clusters (critical, alert, vigilant), 12 regencies / cities in Riau Province and 3 attributes (hotspots, number of fires, number of events). With the results of the visualization of the mapping using the K-means Clustering method, it is expected to be able to help the relevant parties, namely the Riau Provincial Forest Service in handling early the hotspots that are likely to emerge.

1. Introduction

Riau Province is a province that has the 4th widest forest coverage out of 15 provinces that have the widest forest in Indonesia. As one of the provinces which has a large forest area, the threat of a forest fire disaster will be seen from the point of fire every year. This is according to a report from the Riau Province BLHK stating that hotspots in Riau Province are always present every year despite the number of hotspots that have been suppressed (http://dislhk.riau.go.id/). One of the reasons is land clearing which often occurs as a trigger for hotspots in Riau Province. There is a need for countermeasures as soon as possible to overcome the problem of hotspots that will cause forest fires. The problems above need to be watched out quickly one of them is to know in advance the hotspots that are likely to emerge based on data that existed before. Processing data mining is very necessary in order to produce relevant data to determine the possibility of hotspots.

Rahmaddeni,
Email: rahmaddeni@sar.ac.id.

Data mining is an analysis step towards the process of finding knowledge in a database in the form of data patterns or relationships between data in large-scale data processing. Therefore data mining has an important role in various fields. One study of data mining discusses clustering. Basically, clustering is a method for finding and grouping data that has similar characteristics between one data with another data. The principle of grouping is to maximize the similarity between members of one cluster and minimize the similarity between members of different cluster [1][2][3][4][10].

K-Means is a non-hierarchical data clustering method that attempts to partition existing data into one or more skilled and classic clusters in data grouping [1][5][6]. K-Means clustering lies in the partitioning clustering method often used in data mining [7].

Several studies relating to the K-Means Clustering method include building a clustering system using the K-Means method to determine the access point position based on the user's hotspot position, which concludes that there is ease in grouping with the clustering system [8]. K-Means Clustering Algorithm to group Al Qur'an verses can be concluded that the grouping of Al Quran verse data in Indonesian using these algorithms produces verses with certain keywords in the right group [9]. Furthermore, the use of k-means clustering for the determination of hotspots in Riau Province which concludes that areas prone to hotspots are areas that have multiple hotspots that are located close together and are scattered in many areas, so that these points will potentially cause fires [1].

Based on the existing problems and supported by previous research, the researcher will build a visualization mapping system for areas that have the potential to cause forest fires in Riau Province using the k-means clustering method to help the BLKH of Riau Province handle earlier hotspots that are likely to emerge.

2. Research Methods

a. Data Collection

The process of collecting data in this study was carried out by observing and taking data on the distribution of hotspots in Riau Province at the Department of Environment and Forestry (BLKH) of Riau Province.

b. Data Analysis

In the analysis phase the application of the k-means method is used to perform mathematical calculations which include Determination of the center of the cluster, Calculation of distance to the center of the cluster, grouping data, Determination of a new cluster center, and the process of clustering the k-means algorithm. Which method is used to determine the results of the calculation of fire data based on the variable amount of fire data, land area, soil type, and the distance between settlements.

c. System Planning

At this design stage the author will describe the proposed system design form into two forms, namely global design and detailed design. The approach used is an object-oriented approach, where the tools used are UML (ex. Use case diagrams, sequence diagrams, activity diagrams and class diagrams).

d. Implementation

At this stage the steps of the k-means clustering method are applied to the system built using the PHP programming language and the MySQL database. This implementation uses two users namely admin and leader to access the system being built.

e. Testing

The last stage carried out in this method is to test the system that has been built and implemented using the black box method. The goal is to do this stage to see the weaknesses and advantages of the system that has been built.

3. Results and Discussion

a. Use Case Diagram System

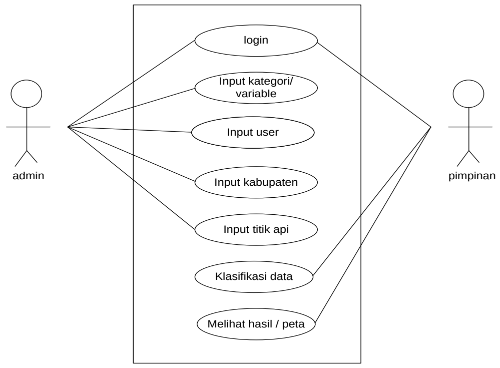


Figure 1. Use Case Diagram

Figure 1 above explains that the system built has 2 actors namely admin and leader. Admin can do the login process, input categories / variables, manage user data, input districts, and input hotspot data. While the leader enters the login process and sees the results of the k-means clustering process data classification and visualization of the hotspot map.

b. Calculation and Implementation of K-Means Clustering Method into the System

Based on the data obtained at the Department of Environment and Forestry (BLKH) of Riau Province, 3 attributes were taken as variables for the application of the K-Means clustering method in determining or labeling fire levels, critical, alert and vigilant. The experiment was carried out using parameters or variables as follows:

a. Number of clusters: 3 (critical, alert, vigilant)
b. Total data: 12 regency
c. Number of attributes: 3 (hotspots, number of fires, and number of events)

Table 1. K-Means Clustering Variable

No	Regency / City	Hotspot	Number of Fires	Number of Events
1	Kuantan Singingi	2170	24.5	3
2	Indragiri Hulu	1919	45.3	8
3	Indragiri Hilir	1378	82	5
4	Pelalawan	3296	162.16	41
5	Siak	682	76.5	9
6	Kampar	548	83.25	35
7	Rokan Hulu	1869	68	10
8	Bengkalis	1826	64	11
9	Rokan Hilir	3198	392	16
10	Kepulauan Meranti	515	236.11	19
11	Pekanbaru	0	12.7	6
12	Dumai	395	122.75	25

Table 1 above illustrates the courtesy variable used in the k-means clustering process for which data is obtained at the Department of Environment and Forestry (BLKH) of Riau Province.

The next step in the process of the K-Means Clustering method is to determine how many iterations are obtained to determine the status of the hotspot. The stages are described below:

1. Iterasi 1

a. Determination of the initial center of the cluster

Based on Table 1 above, random values are determined to be assumed as preliminary data in determining cluster events

Table 2. Random Value of Initial Cluster Center

Total Hotspot	Number of Fires	Number of Events	Cluster
3296	162.16	41	C1
1826	64	11	C2
0	12.7	6	C3

Based on Table 2 above, C1 value is taken from the highest data, C2 value is taken from the middle data, C3 value is taken from the lowest data.

b. Calculation of cluster center distance

To measure the distance between the data and the center of the cluster used Euclidean distance, then the distance matrix will be obtained as follows:

JAIA – Journal Of Artificial Intelligence And Applications
Vol. 1, No. 1, October 2020, pp. 41-47

Euclidean distance formula :

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^{n}(x-y)^2} \qquad (1)$$

X = cluster center
Y = data
Calculation of the distance from the 1st data to the cluster center is:

$$C1 = \sqrt{(2170-3296)^\wedge 2+(24.5-162.16)^\wedge 2+(3-41)^\wedge 2)} = 1135.019945$$

$$C2 = \sqrt{(2170-1826)^\wedge 2+(24.5-64)^\wedge 2+(3-11)^\wedge 2)} = 346.3527826$$

$$C3 = \sqrt{(2170-0)^\wedge 2+(24.5-12.7)^\wedge 2+(3-6)^\wedge 2)} = 2170.034156$$

Then the distance matrix will be obtained.

Table 3. Distance Matrix

No	c1	c2	c3
1	1135.019945	346.3527826	2170.034156
2	1382.343756	94.90885101	1919.277927
3	1920.011882	448.4016057	1379.74182
4	0	1473.579107	3299.572592
5	2615.598906	1144.070037	684.9842626
6	2749.139281	1278.370276	553.2832028
7	1430.43913	43.19722213	1869.822208
8	1473.579107	0	1826.727317
9	251.1083941	1410.671117	3220.430482
10	2782.070021	1322.273365	561.5211733
11	3299.572592	1826.727317	0
12	2901.311798	1432.273913	410.4838639

c. Data grouping

The distance from the calculation results will be compared and the closest distance or the smallest value between the data and the cluster center will be selected, this distance shows that the data is in one group with the closest cluster center.

Table 4. Cluster Grouping

No	C1	C2	C3	Cluster
1	1135.019945	346.3527826	2170.034156	c2
2	1382.343756	94.90885101	1919.277927	c2
3	1920.011882	448.4016057	1379.74182	c2
4	0	1473.579107	3299.572592	c1
5	2615.598906	1144.070037	684.9842626	c3
6	2749.139281	1278.370276	553.2832028	c3
7	1430.43913	43.19722213	1869.822208	c2
8	1473.579107	0	1826.727317	c2
9	251.1083941	1410.671117	3220.430482	c1
10	2782.070021	1322.273365	561.5211733	c3
11	3299.572592	1826.727317	0	c3
12	2901.311798	1432.273913	410.4838639	c3

d. Determination of the new cluster center

Once it is known that the members of each cluster are then the new cluster center is calculated based on the data of the members of each cluster according to the cluster member center formula. So that the following calculations are obtained :

Table 4. Cluster Initials

No	Regency / City	Hotspot	Number of Fires	Number of Events	Cluster
1	Kuantan Singingi	2170	24.5	3	c2
2	Indragiri Hulu	1919	45.3	8	c2
3	Indragiri Hilir	1378	82	5	c2
4	Pelalawan	3296	162.16	41	c1
5	Siak	682	76.5	9	c3
6	Kampar	548	83.25	35	c3
7	Rokan Hulu	1869	68	10	c2
8	Bengkalis	1826	64	11	c2
9	Rokan Hilir	3198	392	16	c1
10	Kepulauan Meranti	515	236.11	19	c3
11	Pekanbaru	0	12.7	6	c3
12	Dumai	395	122.75	25	c3

Table 4 above describes the new clusters obtained from the member data of each cluster

Table 5. Cluster Calculation Based on Cluster Data

No	C1			C2			C3		
	Total Hotspot	Number of Fires	Total Loss	Total Hotspot	Number of Fires	Total Loss	Total Hotspot	Number of Fires	Total Loss
1	3296	162,16	41	2170	24,5	3	682	76,5	9
2	3198	392	16	1919	45,3	8	548	83,25	35
3				1378	82	5	515	236,11	19
4				1869	68	10	0	12,7	6
5				1826	64	11	395	122,75	25
Total	6494	554,16	57	9162	283,8	37	2140	531,31	94
	3247	277,08	28,5	1832,4	56,76	7,4	428	106,262	18,8

Table 5 above illustrates the calculation of clusters C1, C2 and C3 based on the cluster data obtained in Table 4. The total details of each cluster are as follows:

Total hotspots: C1 = 3247, C2 = 1832,4, C3 = 428. Number of Fires: C1 = 277.08, C2 = 56.76, C3 = 106,262. Total Losses: C1 = 28.5, C2 = 7.4, C3 = 18.8.

This data is used to calculate the second iteration.

2. Iteration 2

The steps taken in iteration 2 are the same as those taken in iteration 1. The steps are as follows:

a. Determination of the initial center of the cluster

The center of the second cluster is taken from the results of iteration 1 calculations in Table 5.

b. Calculation of cluster center distance

Calculate the distance from data 1 (Table 4) to the cluster center in step 1.

c. Data grouping

d. The next step is to compare the cluster results in the first iteration with the second iteration. If the value of the cluster member has a convergent value, the iteration process is stopped and the cluster results are valid cluster results.

Table 6. Cluster Comparison Between Iteration 1 and Iteration 2

No	Cluster 1 pada Iterasi 1	Cluster 2 pada Iterasi 2
1	c2	c2
2	c2	c2
3	c2	c2
4	c1	c1
5	c3	c3
6	c3	c3
7	c2	c3
8	c2	c3
9	c1	c1
10	c3	c3
11	c3	c3
12	c3	c3

Because there are still differences between the first and second clusters, a third iteration is carried out and the determination of the new third cluster.

Table 7. Cluster Calculation Based on Cluster Data

No	C1			C2			C3		
	Total Hotspot	Number of Fires	Total Loss	Total Hotspot	Number of Fires	Total Loss	Total Hotspot	Number of Fires	Total Loss
1	3296	162,16	41	2170	24,5	3	682	76,5	9
2	3198	392	16	1919	45,3	8	548	83,25	35
3				1378	82	5	1869	68	10
4							1826	64	11
5							515	236,11	19
6							0	12,7	6
7							395	122,75	25
Total	6494	554,16	57	5467	151,8	16	5835	663,31	115
	3247	277,08	28,5	1822,33	50,6	5,333333	833,57143	94,758571	16,42857

Table 7 above illustrates the calculation of clusters C1, C2 and C3. The total details of each cluster are as follows:

Total hotspots: C1 = 3247, C2 = 1822,33, C3 = 833,57. Number of Fires: C1 = 277.08, C2 = 50,6, C3 = 94,75. Total Losses: C1 = 1822,33, C2 = 833,57, C3 = 16,42.

This data is used to calculate the third iteration.

3. Iteration 3

The steps taken in iteration 2 are the same as those taken in iteration 1 and 2. The steps are as follows:

a. Determination of the initial center of the cluster

The center of the third cluster is taken from the results of iteration 1 calculations in Table 7.

b. Calculation of cluster center distance

c. Data grouping

d. The next step is to compare the cluster results in the first iteration with the third iteration.

Table 8. Cluster Comparison Between Iteration 1 and Iteration 3

No	Cluster 1 pada Iterasi 1	Cluster 3 pada Iterasi 3
1	c2	c2
2	c2	c2
3	c2	c2
4	c1	c1
5	c3	c3
6	c3	c3

JAIA – Journal Of Artificial Intelligence And Applications
Vol. 1, No. 1, October 2020, pp. 41-47

7	c2	c2
8	c2	c2
9	c1	c1
10	c3	c3
11	c3	c3
12	c3	c3

If the value of the cluster member has a convergent value (as in Table 8) then the iteration process is stopped and the cluster results are valid cluster results.

Table 9. The Final Result of K-Means Clustering Method Calculation

No	Regency / City	Hotspot	Number of Fires	Number of Events	Cluster Group	Cluster Status
1	Kuantan Singingi	2170	24.5	3	c2	Alert
2	Indragiri Hulu	1919	45.3	8	c2	Alert
3	Indragiri Hilir	1378	82	5	c2	Alert
4	Pelalawan	3296	162.16	41	c1	Critical
5	Siak	682	76.5	9	c3	Vigilant
6	Kampar	548	83.25	35	c3	Vigilant
7	Rokan Hulu	1869	68	10	c2	Alert
8	Bengkalis	1826	64	11	c2	Alert
9	Rokan Hilir	3198	392	16	c1	Critical
10	Kepulauan Meranti	515	236.11	19	c3	Vigilant
11	Pekanbaru	0	12.7	6	c3	Vigilant
12	Dumai	395	122.75	25	c3	Vigilant

Table 9 above explains that in the data grouping that has been done, there are 3 cluster group statuses where c1 is the critical cluster status (Pelalawan and Rokan Hilir), c2 is the alert cluster status (Kuantan Singingi, Indragiri Hulu, and Indragiri Hilir, Rokan. Hulu, and Bengkalis), and c3 is vigilant cluster status (Siak, Kampar, Meranti Islands, Pekanbaru, and Dumai).

c. Result Visualization

The following describes the implementation of the method into the system being built and the results of the visualization in the form of a map

a. Determine the Initial Cluster Center

```
$cektemp = mysql_fetch_array(mysql_query("SELECT * FROM temp"));
if($i==0){
    $iterasiawalcluster1trx = "1100";//total titik api
    $iterasiawalcluster2trx = "500";
    $iterasiawalcluster3trx = "0";

    $iterasiawalcluster1dtg = "7";//jumlah kasus
    $iterasiawalcluster2dtg = "6";
    $iterasiawalcluster3dtg = "5";
}else{
```

Figure 2. Determination of the Initial Cluster Center from the K-Means Clustering Method Step

If a cluster loop occurs, the next iteration will be automatically retrieved from the data obtained in the first iteration

b. Input data into the system by applying the k-means clustering method formula to the program being built.

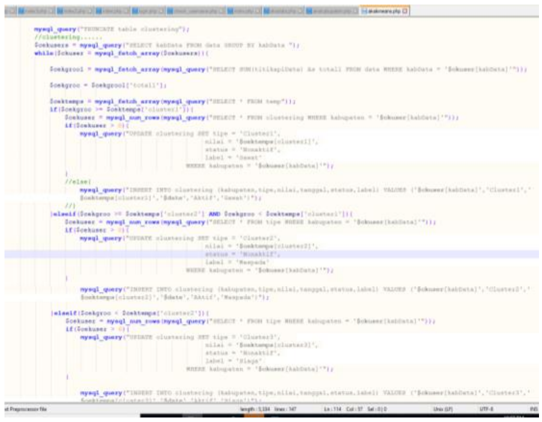
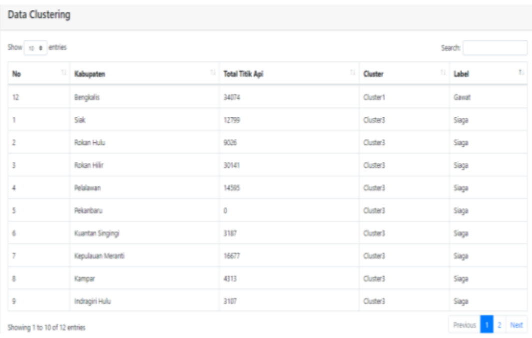


Figure 3. Steps of the K-Means Clustering Method into the Built System

c. If the k-means process has been completed and no longer looping occurs, the status of the regency / city will be obtained whether it is critical, alert, or vigilant, which will be used later in the map visualization.



Data Clustering

Show 10 entries Search:

No	Kabupaten	Total Titik Api	Cluster	Label
12	Bengkalis	34074	Cluster1	Gawat
1	Siak	12799	Cluster3	Siaga
2	Rokan Hulu	9026	Cluster3	Siaga
3	Rokan Hilir	30141	Cluster3	Siaga
4	Pelalawan	14595	Cluster3	Siaga
5	Pekanbaru	0	Cluster3	Siaga
6	Kuantan Singingi	3187	Cluster3	Siaga
7	Kepulauan Meranti	16677	Cluster3	Siaga
8	Kampar	4313	Cluster3	Siaga
9	Indragiri Hulu	3107	Cluster3	Siaga

Showing 1 to 10 of 12 entries Previous 1 2 Next

Figure 4. The Final Result of the K-Means Clustering Process

d. The result of visualization of the process of applying the k-means clustering method is in the form of map visualization.

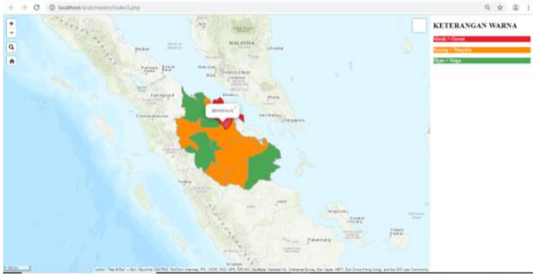


Figure 5. Visualization of Results in the form of a Map

In this view the admin and leadership can see the spatial map of fire data that has high potential for fires in the areas entered based on the data obtained, so that leaders can see the status of the area and take action to solve problems and prevent forest fires.

4. Conclusion

After completing a series of stages towards the development of a data mapping system for potential fire areas in Riau Province using the K-means clustering method, several conclusions can be drawn, including the following:

a. With this system, the environment and forestry agencies in particular are faster in processing data, making it easier to find out information on areas in Riau Province that have high potential for forest fires so that forest fire prevention efforts can be quickly and appropriately handled.

b. In this system, fire data is stored in a database, making it easier to find data if needed.

c. In general, the system being built has been able to run and can produce data that is in accordance with the K-means clustering method calculations.

d. From the test case test, it can be concluded that all processes can run with the expected results. In that sense, the k-means clustering process in segmenting fire data is successful.

5. Reference

[1] S. Sukamto, I. D. Id, and T. R. Angraini, “Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means,” *JUITA J. Inform.*, vol. 6, no. 2, p. 137, 2018, doi: 10.30595/juita.v6i2.3172.

[2] R. Goejantoro, “Perbandingan Pengelompokan K-Means dan K-Medoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus : Data Titik Panas Di Indonesia Pada 28 April 2018) Comparison,” vol. 10, no. April 2018, pp. 143–152, 2019.

[3] N. A. Khairani and E. Sutoyo, “Application of K-Means Clustering Algorithm for Determination of Fire-Prone Areas Utilizing Hotspots in West Kalimantan Province,” *Int. J. Adv. Data Inf. Syst.*, vol. 1, no. 1, pp. 9–16, 2020, doi: 10.25008/ijadis.v1i1.13.

[4] M. Mustofa, “Penerapan Algoritma K-Means Clustering pada Karakter Permainan Multiplayer Online Battle Arena,” *J. Inform.*, vol. 6, no. 2, pp. 246–254, 2019, doi: 10.31311/ji.v6i2.6096.

[5] A. P. Windarto, “Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method,” *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, 2017, doi: 10.29099/ijair.v1i2.17.

[6] J. Qi, Y. Yu, L. Wang, J. Liu, and Y. Wang, “An effective and efficient hierarchical K-means clustering algorithm,” *Int. J. Distrib. Sens. Networks*, vol. 13, no. 8, pp. 1–17, 2017, doi: 10.1177/1550147717728627.

[7] S. Naeem and A. Wumaier, “Study and Implementing K-mean Clustering Algorithm on English Text and Techniques to Find the Optimal Value of K,” *Int. J. Comput. Appl.*, vol. 182, no. 31, pp. 7–14, 2018, doi: 10.5120/ijca2018918234.

[8] A. Fauzan, A. Y. Badharudin, F. Wibowo, J. Raya, and D. Purwokerto, “Sistem Klasterisasi Menggunakan Metode K-Means dalam Menentukan Posisi Access Point Berdasarkan Posisi Pengguna Hotspot di Universitas Muhammadiyah Purwokerto (Clustering System Using K-Means Method in Determining Access Point Position at Muhammadiyah Un,” *Juita*, vol. III, pp. 25–29, 2014.

[9] M. Robani and A. Widodo, “Algoritma K-Means Clustering Untuk

Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia," *J. Sist. Inf. Bisnis*, vol. 6, no. 2, p. 164, 2016, doi: 10.21456/vol6iss2pp164-176.

[10] J. Han, M. Kamber, and J. Pei, Data Mining, Concepts and Techniques. 2012.