

Spatial Distribution and Implementation of the K-Means Clustering Method at Hotspots in North Sumatra

Sri Wahyuni

²Badan Meteorologi Klimatologi dan Geofisika, Indonesia

Article Info

Article history:

Received February 1, 2021

Revised February 16, 2021

Accepted March 11, 2021

Keywords:

hotspot
clustering
spatial
k-means
satellites

ABSTRACT

Hotspots are indicators of forest and land fires. Hotspot monitoring can be carried out with the help of remote sensing tools and geographic information systems. Hotspot data is obtained from the MODIS sensors from the TERRA and AQUA satellites which contain information on latitude and longitude coordinates and the level of confidence divided by three levels, namely low, medium and high confidence levels. Based on the spatial results, the number of hotspots in North Sumatra Regency is in February, March, June, July, and August. Districts that are dominant with hotspots are Karo Regency, Labuhan Batu Regency, Mandailing Natal Regency, Padang Lawas Regency and South Tapanuli Regency. Based on the results, the process of applying the k-means clustering method to the weka application, the data obtained is in the form of a clustered group and the results can be made into indicators in determining hotspots in districts in North Sumatra province per month.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Kartika Dewi,
Badan Meteorologi Klimatologi dan Geofisika, Indonesia
Email: sinambela.m@gmail.com

1. INTRODUCTION

Forest and land fires in Indonesia are an important note for the government because they produce haze and other impacts due to forest fires on the world community. Forest fires usually occur during the dry season. For the island of Sumatra, the provinces that experience high potential for forest fires are Riau, Jambi and South Sumatra every year. For the province of North Sumatra, there is the potential for forest fires in several districts. Various efforts have been made by the government to tackle these forest fires [1].

Hotspots can be interpreted as indicators of forest fires. These land fires can be identified using remote sensing technology[2]. One of the remote sensing technologies is the MODIS (Moderate resolution imaging spectroradiometer) sensor from the TERRA and AQUA satellites owned by NASA (National Aeronautics and Space Administration). These satellites record and monitor hotspots around the world in real time 24 hours. The TERRA satellite which carries the MODIS sensor is an environmental observation satellite that can be used for regional surface temperature data extraction [3].

Based on the results of the research that has been done, it can be seen that the analysis of the area of forest and land fires using remote sensing technology can be carried out using two approaches, namely digital and visual interpretation [4]. Hotspots can be validated by direct and indirect

observation [5]. Direct observations are carried out by checking directly into the area where there are many hotspots.

Hotspot data obtained from the MODIS satellites contains information on the coordinates of the location of the area recorded by the satellite. The data of the coordinate location points were analyzed spatially and used the K-Means algorithm. The k-means method is one of the clustering methods which aims to divide the data set into clusters [6]. This method is used because the K-Means algorithm is easy to adapt and can be used for processing large datasets.

2. RESEARCH METHOD

Hotspots are hotspots that are indicated as points that have a higher temperature than the surroundings. In the hotspot data generated from the MODIS sensor, the hotspot is the result of heat detection on a certain pixel scale, for example 1 km x 1 km. Generally, the more hotspots that are detected in an area, the greater the possibility of forest fires in that area [7].

LAPAN (National Institute of Aeronautics and Space) has issued technical guidance on the level of confidence or confidence level of a hot spot to determine actions in dealing with it. Table 1 shows the meaning of the level of confidence in the data taken from the technical guidelines[7].

Table 1. Meaning of Trust Interval in Hotspot Information

Confidence Level (C)	Class
$0\% \leq C < 30\%$	Low
$30\% \leq C < 80\%$	Medium
$80\% \leq C < 100\%$	High

The K-Means algorithm is one of the most frequently used data grouping techniques. K-Means groups similar objects in the same cluster. An overview of the algorithm is as follows [9] determine the value of "K" or the number of clusters desired and determine the centroid or cluster center. The centroid is determined randomly from the existing data objects according to the value "K", if the value "K" is two then the centroid that is determined is also two. Divide each object into the most similar clusters. The most similar clusters are those with the closest centroid. Calculation of the proximity between the object and the centroid can be done using a distance function, such as Euclidean Distance. Recalculate cluster "K" with the average of all objects assigned to each cluster. Repeat the second and third steps until there are no more objects moving clusters [6].

The formulas used for k-means are:

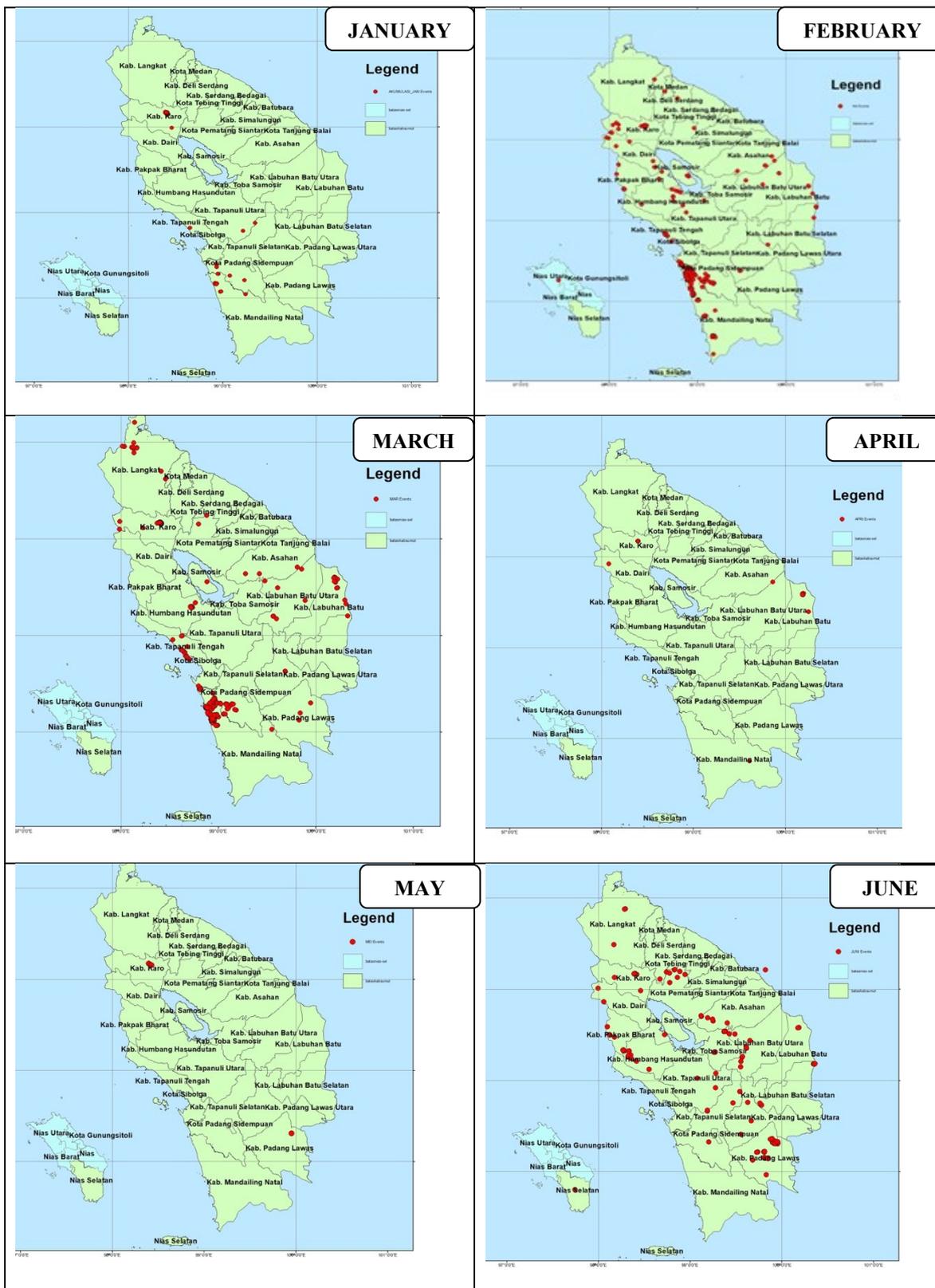
$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2}$$

(1)

Where: d: distance, i: i1, i2, i3,, ip, j: j1, j2, j3,, jeep, k: attribute value, n: dimension data, x: data object.

3. RESULTS AND DISCUSSION

In this study, using hotspots or hotspots to provide solutions to existing problems using spatial distribution and clustering using the k-means algorithm.



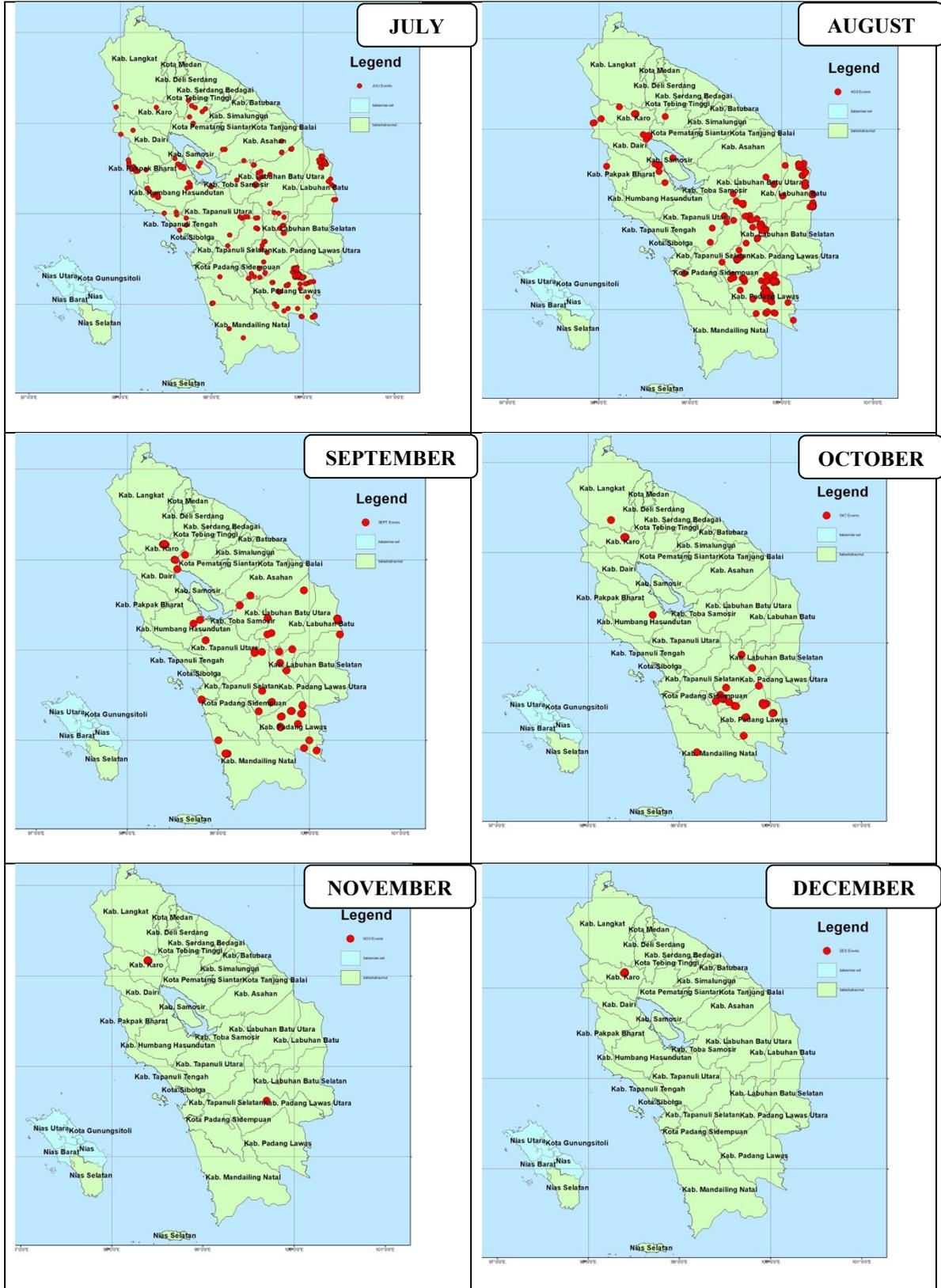


Figure 1. Distribution of cumulative monthly spatial hotspots in 2014 - 2020. The spatial results of hotspots using the Arcgis Map application can be seen that many hotspots in the districts of North Sumatra province occur in February, March, June, July, and August. In terms of climatological factors and environmental factors, it affects the number of hotspots each month. From climatological factors, the number of hotspots in North Sumatra is influenced by the reduced amount of rainfall that month.

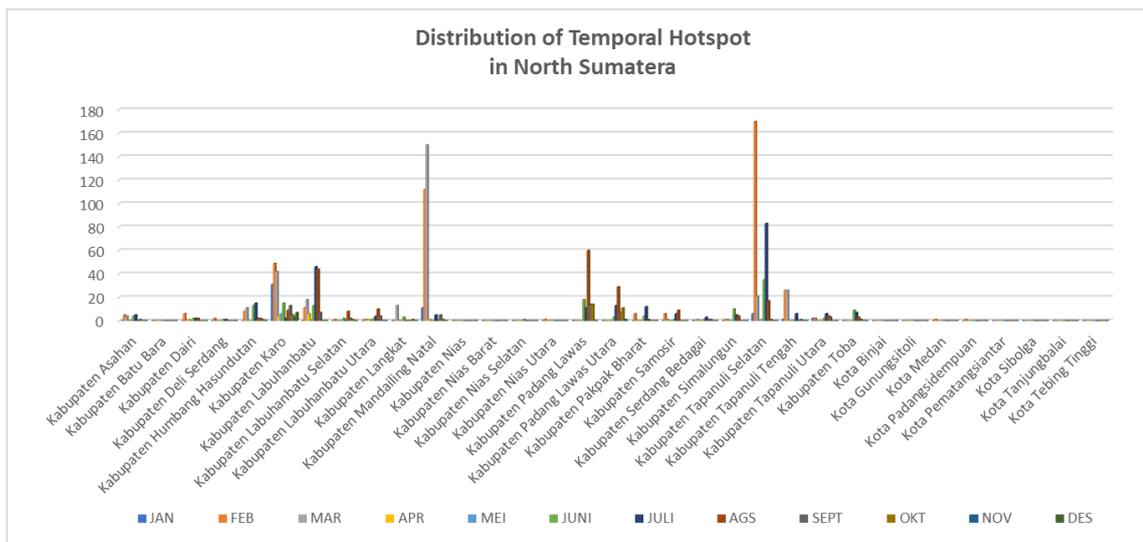


Figure 2. Graphics of monthly hotspot time series for 2014 – 2020

The distribution of temporal hotspots in the districts of North Sumatra province, there are several regencies where there are dominant hotspots per month. From Figure 2 it can be seen that there are 5 districts that have many hotspots per month, such as Karo Regency, Labuhan Batu Regency, Mandailing Natal Regency, Padang Lawas Regency and South Tapanuli Regency. The number of hotspots in the province of North Sumatra in the period 2014 to 2020, the highest peak occurred in 2014 in February.

After doing the mapping using the arcgis map, then using the k-means algorithm clustering application using the weka application. Using the Weka application by using sample data then creating the tested data and it will produce predictions of hotspots that have the possibility of occurring in the district.

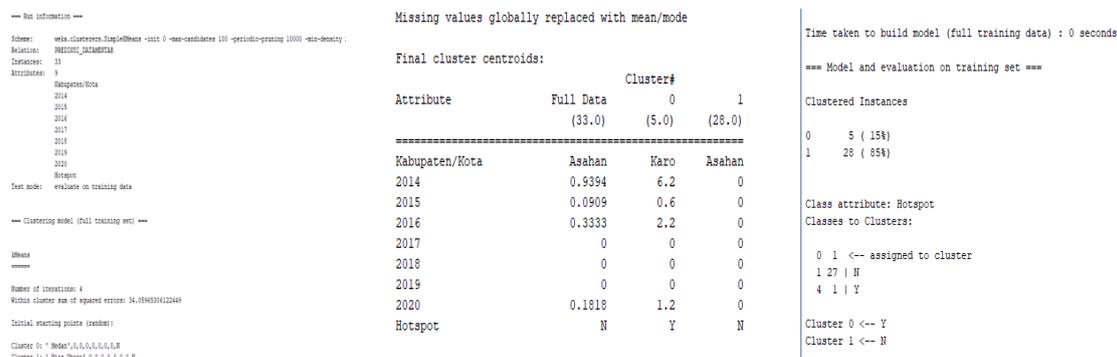


Figure 3. Implementation of the K-Means Clustering Method

In the implementation results using the k-mean clustering algorithm in weka, in January there are several districts that are predicted to have hotspots, namely Karo, Mandailing Natal and South Tapanuli districts.

4. CONCLUSION

Based on the analysis of that can be drawn from this research, namely, among others, based on the results of the spatial distribution of hotspots that are processed using the arcgis map, it can produce data on the number of distribution of hotspots that are dominant in districts in North Sumatra province. The dominant districts to watch out for in the dry season are Karo, Labuhan Batu, South Labuhan Batu, Mandailing Natal, South Tapanuli and Padang Lawas. From the results (output) obtained from the process of applying the k-means clustering method to the weka application, the

data obtained is in the form of a clustered group and the results can be made into indicators in determining hotspots in districts in North Sumatra province per month.

REFERENCES

- [1] N. A. Indrawan and H. A. Adrianto, "Spatio - Temporal Clustering Hotspot di Sumatera Selatan Tahun 2002 - 2003 Menggunakan Algoritme ST - DBSCAN dan Bahasa Pemrograman R Spatio - Temporal Clustering Hotspot in South Sumatera From 2002 - 2003 Using ST - DBSCAN Algorithm and R Programming Lan," vol. 3, pp. 112–121, 2003.
- [2] T. Handayani, A. J. Santoso, and Y. Dwiandiyanta, "Pemanfaatan Data Terra Modis untuk Identifikasi Titik Api Pada Kebakaran Hutan Gambut (Studi Kasus Kota Dumai Provinsi Riau)," *Semin. Nas. Teknol. Inf. dan Komun.*, 2014.
- [3] I. Prasasti, K. A. Sambodo, and I. Carolita, "Pengkajian Pemanfaatan Data TERRA-MODIS untuk Ekstraksi Data Suhu Permukaan Lahan (SPL) Berdasarkan Beberapa Algoritma," *Penginderaan Jauh*, 2007.
- [4] Suwarsono, Rokhmatuloh, and T. Waryono, "Pengembangan Model Identifikasi Daerah Bekas Kebakaran Hutan Dan Lahan (Burned Area) Menggunakan Citra Modis Di Kalimantan (Model Development of Burned Area Identification Using Modis Imagery in Kalimantan)," *J. Penginderaan Jauh*, vol. 10, no. 2, pp. 93–112, 2013.
- [5] I. A. Csiszar, J. T. Morisette, and L. Giglio, "Validation of active fire detection from moderate-resolution satellite sensors: The MODIS example in Northern Eurasia," *IEEE Trans. Geosci. Remote Sens.*, 2006, doi: 10.1109/TGRS.2006.875941.
- [6] R. A. Soelaiman Isye; Gosno, Eric Budiman, "Implementasi KD-Tree K-Means Clustering untuk Klasterisasi Dokumen," *J. Tek. ITS*, vol. 2, no. Vol 2, No 2 (2013), pp. A432–A437, 2013, [Online]. Available: <http://ejurnal.its.ac.id/index.php/teknik/article/view/3872>.
- [7] LAPAN, *PANDUAN TEKNIS (V.01) INFORMASI TITIK PANAS (HOTSPOT) KEBAKARAN HUTAN/LAHAN*. 2016.