INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY: APPLIED BUSINESS AND EDUCATION RESEARCH
2023, Vol. 4, No. 8, 3046 – 3057
http://dx.doi.org/10.11594/ijmaber.04.08.37

Research Article

Spatio-Temporal Pattern Analysis of Forest Fire in Malang based on Remote Sensing using K-Means Clustering

Annisa Puspa Kirana[1]*, Mungki Astiningrum[1], Candra Bella Vista[1], Adhitya Bhawiyuga[2], Aris Nur Amrozi[1]

[1]Information Technology Department, State Polytechnic of Malang, Indonesia
[2]Faculty of Computer Science, University of Brawijaya Malang, Indonesia

Article history:
Submission May 2023
Revised August 2023
Accepted August 2023

**Corresponding author:*
E-mail:
puspakirana@polinema.ac.id

ABSTRACT

Forest and land fire significantly impact the balance of the environment, such as haze pollution, destruction of ecosystems, the high release of carbon in the air, deterioration of health, and losses in various other fields. Based on these factors, developing an early warning system is essential to prevent forest fires, especially in forest and land areas. One of the data that can be used to monitor areas where there are frequent fires is hotspot data taken from the NASA MODIS Fire satellite. Data mining techniques are carried out to process the hotspot data so that the distribution of hotspot swarms is obtained. The data on the distribution of the clustering of hotspots are used to detect areas that are prone to fire from year to year. This study used the K-Means clustering algorithm. The data used in this study is hotspot data from Malang District, Indonesia. The range of hotspot data from January 2018 to June 2022. We use Silhouette coefficient testing to get the best number of classes in the cluster—this study's most recent application of the K-means clustering method to analyze hotspot distribution in a spatial-temporally. We use hotspot data in Malang's forest and land area using hotspot confidence levels >80%.

Keywords: *Hotspot, Spatial-temporal analysis, K-means, Clustering, Silhouette coefficient*

Introduction

A disaster is an event or series of events that threatens and disrupts people's lives and livelihoods caused by natural and non-natural factors or human factors. One of the disasters that often occurs in Indonesia is fires. Forest and land fires area in 2022 reached 233,173.40 Ha (Ministry of Environment and Forestry of the Republic of Indonesia, 2022). Without the support and innovation of handling, forest and

How to cite:
Kirana, A. P., Astiningrum, M., Vista, C. B., Bhawiyuga, A., & Amrozi, A. N. (2023). Spatio-Temporal Pattern Analysis of Forest Fire in Malang based on Remote Sensing using K-Means Clustering. *International Journal of Multidisciplinary: Applied Business and Education Research*. *4*(8), 3046 – 3057. doi: 10.11594/ijmaber.04.08.37

land fires in the next few years can increase in area. Therefore, early handling must be carried out to find out the occurrence of fires. One way to discover a fire's occurrence is to know the hotspot detected by satellites. Hotspots can be defined as areas with a relatively higher surface temperature than the surrounding area based on a certain temperature threshold monitored by remote sensing satellites. Hotspots indicate land and forest fires in a specific area, so the more Hotspots, the more potential land fire events. Hotspot location coordinates are taken from satellite data extraction. The fires that occurred are still difficult to know early. Limited information on geothermal spots that have the potential for fires to occur leads to a lack of early supervision of impending fires. Given the factors causing the high incidence of fires and these very adverse impacts, it is essential to develop an early warning system to prevent land fires, especially in forest and land areas.

Several previous studies have developed early warning systems by utilizing hotspot data, including research conducted by Sitanggang et al. (2018) which uses algorithms an ST-DBSCAN to predict the appearance of hotspots (Sitanggang et al., 2018). Kirana et al. conducted another study that used Kulldorff's Scan Statistic method to detect hotspot pattern distribution in the Sumatra area (Kirana et al., 2016; Kirana et al., 2015) and Borneo (Sitanggang et al., 2020). In addition to using hotspot distribution data to predict fire-prone areas, several previous studies have also utilized the vegetation health index (VHI) value. Some studies that use the value of VHI to indicate fire-prone areas are in (Sun et al., 2013; Ririd et al., 2020; Ariyanto et al., 2020). This study used the K-Means method for hotspot grouping analysis based on spatial and temporal aspects. In this study, the method silhouette coefficient was also used to produce the most optimal cluster. The K-Means clustering method is a method commonly used in clustering-based segmentation. K-means Clustering is a clustering algorithm widely used to group data based on the nearest centroid to the data (Bianto et al., 2008). Using the K-Means method can make it easier to segment fire points to find out how big the fire is when it occurs. K-means was chosen in research because K-Means is a reasonably simple clustering algorithm that partitions data into several clustering. K-means is relatively easy to implement and has also been widely used.

Hotspot data was obtained from the National Oceanic Atmospheric and Administration (NOAA) satellite and from sensors attached to the MODIS (Moderate Resolution Imaging Spectroradiometer) sensors found on the Terra and Aqua satellites. In this study, the K-Means method was used for clustering hotspots of forest land in the Malang district area in 2018-2022. The hotspot data source comes from FIRM MODIS Fire and can be downloaded at the http://earthdata.nasa.gov/data/nrt-data/firms website address. The data results from remote sensing from the MODIS (Moderate Resolution Imaging Spectroradiometer) sensor installed on the Terra and Aqua satellites. The distribution of forest land in Malang dataset was obtained from the Ministry of Environment and Forestry of the Republic of Indonesia. The data can be downloaded at http://pkps.menlhk.go.id/piaps.

This study aims to apply the technique of grouping hotspots using K-Means clustering on forest land in the Malang District area. The distribution pattern of hotspot clusters can be used to determine areas that have a high density of hotspots. Areas are known to be used by the authorities to help strengthen the implementation of policies in preventing fires from an early age, for example, by banning burning in forest land locations that are prone to fire. This research also aims to build a system that can group hotspot data containing hotspot information that can cause fires of various statuses, such as high, medium, and weak potential. The system aims to facilitate dissemination by issuing more accurate and informative information about field conditions to the public.

After the clustering process, the distribution pattern will be analyzed based on spatial and temporal aspects of hotspots in the Malang District. The benefits of this study are an early warning system and early detection of forest fires, especially in Malang District, by providing an overview of the pattern of clustering hotspot data on forest land based on its spatial and temporal aspects. Fire-prone areas have been known to be used to compile fire control plans, especially fire prevention, especially in the Malang District area.

Material and Methods

This study aims to apply the technique of grouping hotspots using K-Means clustering on forest land in the Malang District area. After the clustering process, the distribution pattern will be analyzed based on spatial and temporal aspects of hotspots in Malang. The benefits of this study are an early warning system and early detection of forest fires, especially in Malang District, by providing an overview of the pattern of clustering hotspot data on forest land based on its spatial and temporal aspects. Fire-prone areas have been known to be used to compile fire control plans, especially fire prevention, especially in the Malang District area.

Study Areas

Malang District is located in a highland area, with coordinates 112° 17' 10.9" - 112° 57' 0.0" East Longitude and 7° 44" 55.11" - 8° 26' 35.45" South latitude. Figure 1 explains that Malang District's area is 334,787 Ha, consisting of 33 sub-districts spread over urban and rural areas. Malang has varied forests, consisting of lowland rainforests, mountain forests, production forests, protected forests, coastal forests, conservation forests, and mangrove forests (malangkab.go.id, 2018). Altitude variations between 0 to 3000 meters above sea level. The topography varies from beaches to mountains with gentle to steep slopes. The total area of forest areas throughout Malang District is 42,365.2 ha. The status of protected forests and 43,105.1 ha of production forests. The protected forest is spread over several districts, namely Donomulyo, covering an area of 421 ha, Pagak 1,250 ha, Gedangan 787.70 ha, Sumbermanjing 8,275.7 ha, Tirtoyudho 4,153.7 ha, and Ampelgading covering an area of 5,223.1 ha [10].This study clusters hotspot data in forest and land areas in Malang District, Indonesia, from January 2018 - June 2022. Based on data from the detection results of the Terra and Aqua satellites, the data source was obtained by FIRM MODIS Fire.

Figure 1. Malang District Area

Stages of Research

This section describes the stages of research. The steps carried out in the study consisted of six phases: data collection and analysis, data preprocessing, implementation of the K-Means method for determining hotspot clusters using the K-Means method, cluster validation, and clustering visualization. Figure 2 is a flowchart about the stages of research.

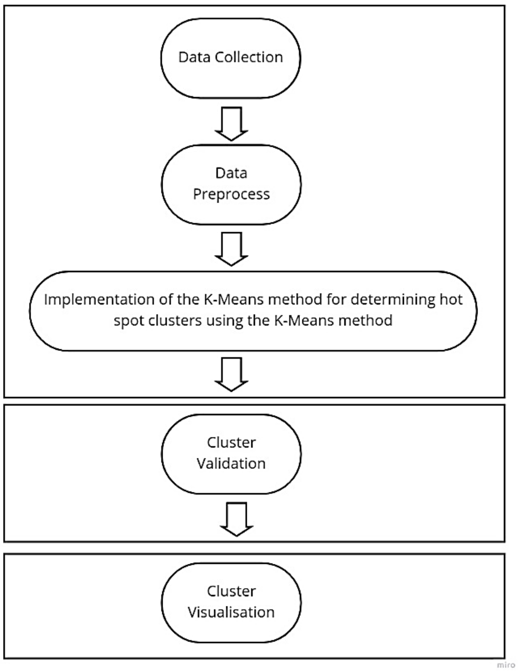

Figure 2 Stages of research

a) Data Collection

The data used as the research object is data on the distribution of hotspots that function as indicators of forest and land fires. The hotspot data studied location is in Malang District with the last 4-year interval, starting from 2018 to 2022. The data has latitude and longitude attributes representing spatial aspects and date attributes representing temporal aspects. The data on the distribution of hotspots studied is in Malang from 2018 to 2022, sourced from FIRM MODIS Fire. The information on the distribution of forests and land in Malang from the Indonesian State Forestry Company. Data on the distribution of hotspots can be downloaded at the website address http://earthdata.nasa.gov/data/nrt-data/firms. Elements of the location dimensions are the coordinates of the longitude and latitude of the hotspot that occurs. Meanwhile, the element of the time dimension is the interval on which the hotspot takes place, which includes annual, monthly, and daily intervals. The hotspot data that has been obtained is then analyzed to obtain the appropriate and appropriate attributes. In the data set, data fields are selected that are needed to speed up data calculations. The required data fields are latitude, longitude, and date.

b) Data Preprocessing

At this stage, data extraction is carried out, data selection of hotspots for forest and non-forest areas, and selection of hotspot distribution locations. At the extraction stage, a selection of the desired attributes and records is carried out. This needs to be done because not all data elements are helpful in decision-making. Data extraction is performed on the DBF file. Hotspot data is in CSV (Comma Separated Values) format. The hotspot data is then represented as a shapefile with an SHP extension to facilitate spatial analysis of the data. After that, a selection of the location of the distribution of hotspots is carried out per district/district. After that, a selection of hotspots is carried out based on the confidence level. Geothermal hotspots or hotspots have a confidence level that indicates that hotspots monitored from distant sensing satellites are actually the actual occurrence of wildfires in the region. The higher the confidence interval, the higher the potential that the hotspot is a land or forest fire. According to LAPAN, the MODIS Active Fire Product User's Guide divides three classes of trust levels shown in Table 1

Table 1. Hotspot Trust Level Values

Trust Level	Class	Action
0% ≤ C ≥ 30%	Low	Noteworthy
30% ≤ C ≥ 80%	Medium	Alert
80% ≤ C ≥ 100%	High	It needs to be addressed immediately

c) K-Means Clustering

James MacQueen introduced the K-Means method in 1976. This method is one of the most commonly used non-hierarchical methods. The K-Means clustering algorithm is known for its simple ability to group big data and outliers quickly. The K-means stage is shown in Figure 3. The step-by-step in general is (MacQueen, 1967):

1. Determining the number of cluster K
2. Determining the centroid
3. Has the centroid changed? If yes, count the distance of the data from the centroid if it is not completed.
4. Group data by distance Nearby.
   The formula calculates the distance (Euclidean):

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$$

d(x,y) = distance
(q_1, p_1) = the coordinate of the first point
(q_2, p_2) = the coordinate of the second point.

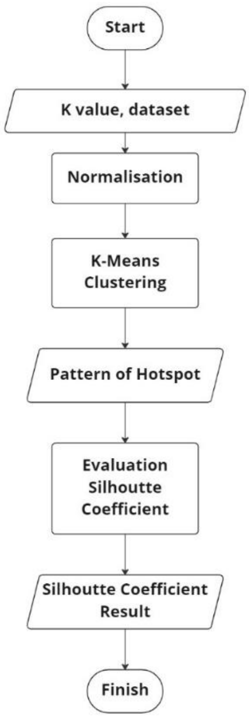

Figure 3 K-Means Clustering Algorithm

d) Silhouette coefficient

The silhouette coefficient is used to measure an algorithm's effectiveness and effectiveness (Rousseeuw, 1987). The silhouette value evaluates an object's cohesion with its particular cluster in comparison to other clusters. A high number on the silhouette implies that the object is well-matched to its own cluster and weakly matched to nearby clusters. The silhouette has a range of 1 to +1. When the majority of items have high values, the clustering configuration is suitable. Numerous points with low or negative values indicate that there may be too many or too few clusters in the clustering design. The Silhouette coefficient formula is:

$$S(i) = \frac{b_{(i)} - a_{(i)}}{max\{a_{(i)}, b_{(i)}\}}$$

Where
S(i) = the silhouette coefficient of the data point i.
a(i) = the average distance between *i* and all the other data points in the cluster to which *i* belongs.
b(i) = the average distance from *i* to all clusters *i* does not belong.

Results and Discussion

This study will discuss the analysis of hotspot data in Malang District from 2018 to 2022, and the analysis of cluster formation in Malang District based on spatial and temporal aspects.

a) Hotspot Data on Forest and Land areas in Malang District

Figure 4 summarizes hotspot data for four periods, namely Period 1 (June 2018-May 2019), Period 2 (June 2019-May 2020), Period 3 (June 2020-May 2021), and Period 4 (June 2021-May 2022). Figure 5 shows us the number of hotspot occurrences in each region of the Malang district from 2018 to 2022. Based on Figure 4, it can be seen that there was a significant increase in fires in 2019, with 207 hotspots. Forest fires caused an increase in fires in Malang District in September 2019. Based on Figure 5, the highest number of the hotspot are in Tamansari and Ngadas areas. The largest fire occurred in Bromo Tengger Semeru National Park, where the affected area reached 93.3 hectares (Rousseeuw, 1987). The number of hotspots has been increasing for several years; from 2021 to 2022, hotspots increased by 59.37%.

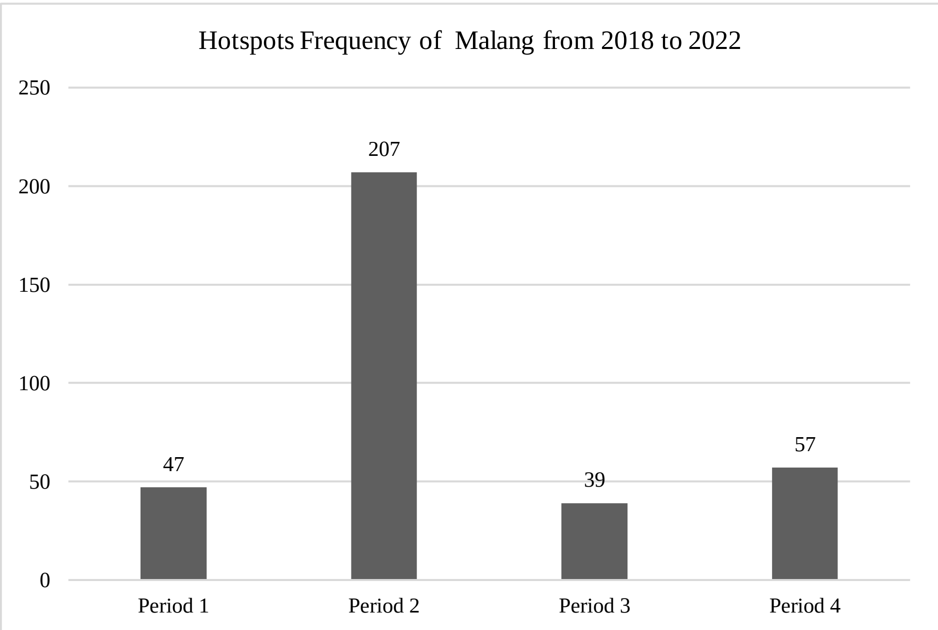

Figure 4 Graph of the hotspot in the forest and divinity area of Malang District from 2018 to 2022

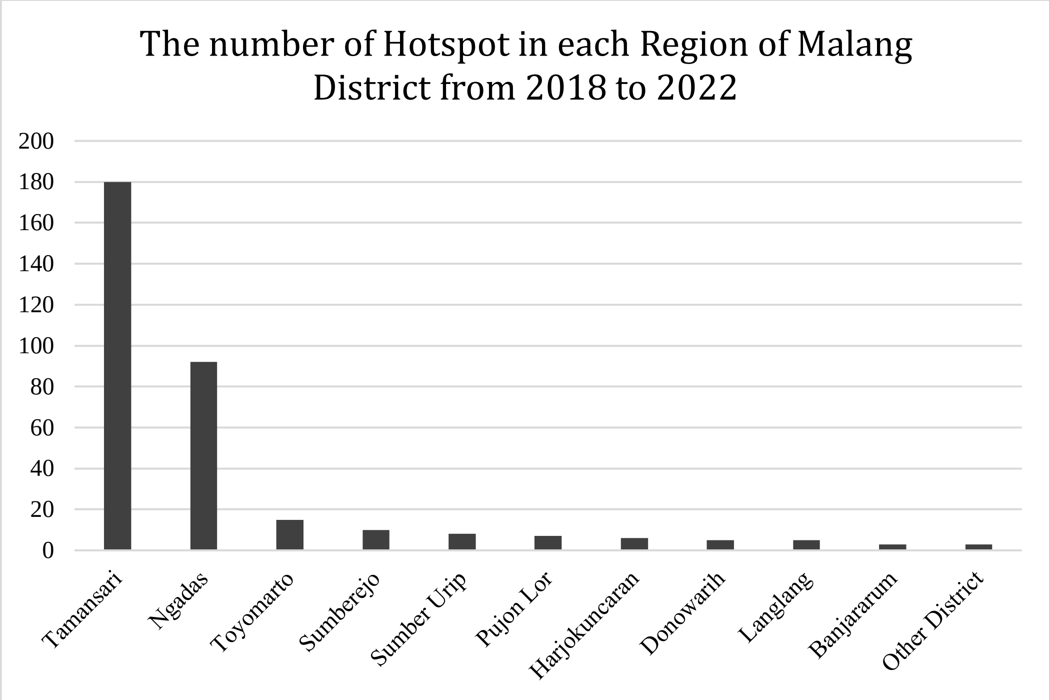

Figure 5 Graph of the hotspot in the forest and divinity area of Malang District from 2018 to 2022

b) Distribution of Hotspot Data based on Spatial Aspects

This study used several variations in the period in the Clustering process. The period studied was the last four years, from June 2018 to June 2022. Based on the following spatial aspects, areas prone to fire with the highest frequency of hotspots can be seen in Figure 6. The villages with the highest frequency of fires are Tamansari (49%), Ngadas (25%), and Toyomarto (4%).

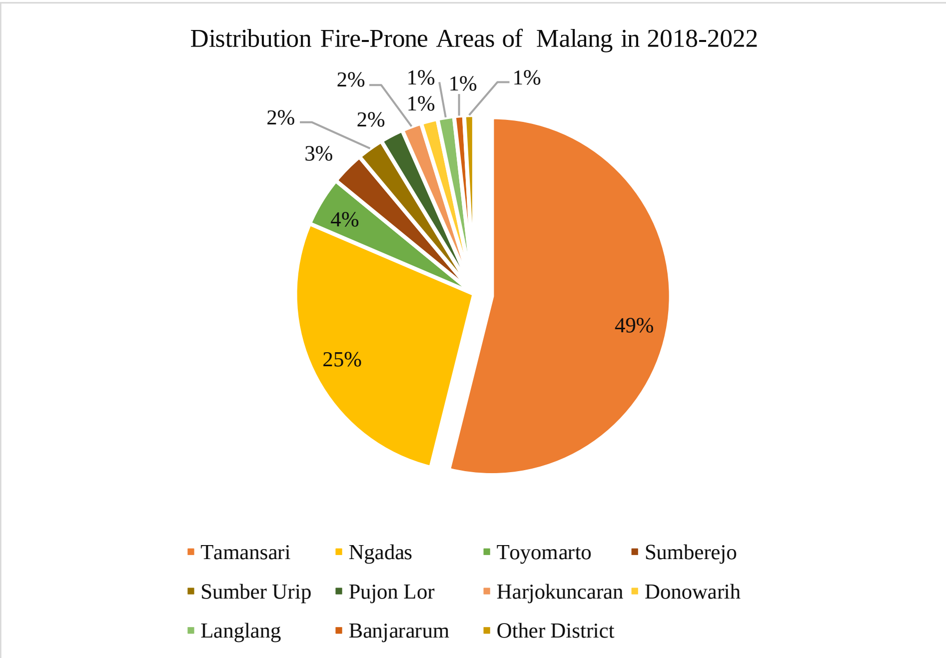

Figure 6 Graph of the pattern of fire-prone villages in Malang District from 2018 to 2022

c) Distribution of Hotspot Data based on Temporal Aspects

Figure 6 is a temporal graph of the fire-prone month in Malang District. It can be seen that the months of January and December are the months with the lowest rate of fires. That month is Indonesia's rainy season, with a higher rainfall level than other months. September and October are the months with the highest rates of fire events. One of the factors causing the high number of fires in the month is the dry season.

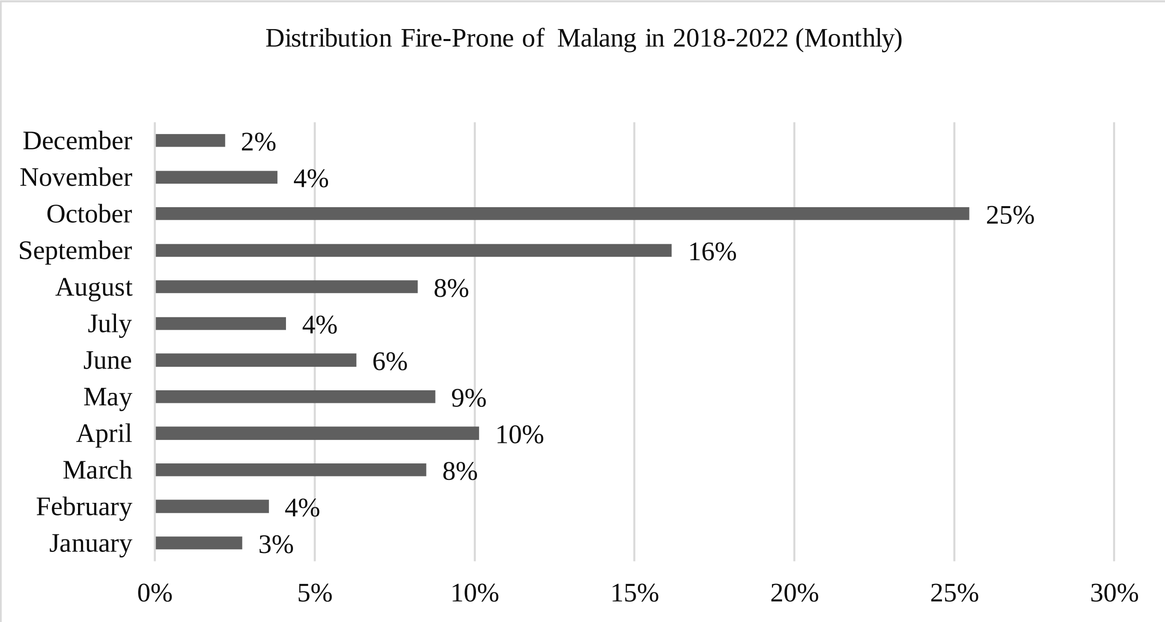

Figure 6 Graph of distribution fire-prone in Malang from 2018 to 2022

c) Cluster Results and Analysis of hotspot distribution using K-Means in 2018-2022

The hotspot cluster from 2018 to 2022 consists of 3 clusters. In Table 1, it can be seen that cluster 1 is an area with a high level of fire density, cluster 2 has a medium density, and cluster 3 has a low hotspot density. The areas with the highest compliance are Tamansari, Pronojiwo, Oro – Oro Ombo, Supiturang, Tumpang, and several other regions, which can be seen in Table 1.

Table 1 Hotspot Trust Level Values

Cluster	District			Pusat Cluster	Radius
Cluster 1 (High)	– Tamansari – Pronojiwo – Oro oro Ombo – Supiturang – Tumpang	– Pasrujambe – Mororejo – Kayukebek – Taji	– Ngadas – Argosari – Ngadirejo – Sumber Putih	Latitude: -8.000754 Longitude: 112.933871	18.65 km
Cluster 2 (Medium)	– Pesanggrahan – Pujon Lor – Bendosari – Sumber Urip	– Bulukerto – Bumiaji – Cendono – Kalimanis	– Tegalgondo – Kalisongo – Giripurno – Jedong	Latitude: -8.034280 Longitude: 112.594422	12.93 km
Cluster 3 (Low)	– Harjokuncaran – Sitiarjo – Ringin Kembar – Jambangan	– Gedangan – Jogomulyan – Tirtoyudo	– Dampit – Pamotan – Gedangan	Latitude: -8.364649 Longitude: 112.753377	24,57 km

Figure 7 illustrates the distribution of hotspot clusters in the Malang area. K-Means produced 3 clusters for hotspots in the Malang district. Cluster 1 is a high-density fire cluster marked with a blue circle. Cluster 2, which includes areas with a density of medium fire points, is marked in red. Cluster 3 has a low density of fire points marked with a purple color image. The center of the fire is widely spread in the North and East regions of Malang District. This is because the forest area is mainly scattered in the region. Based on records from the East Java Provincial Forestry Service, from 2014 to 2019, the condition of the forests in Malang Regency experienced a decrease in the total area that was not significant, where Malang Regency has five types of forest. Namely production forests, protected forests, nature reserves, and national parks.

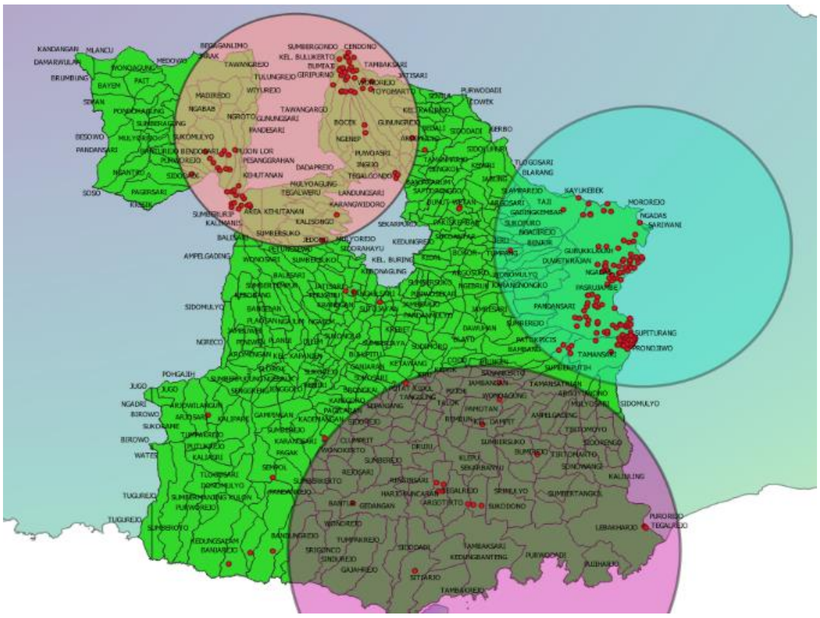

Figure 7 Distribution of hotspot clusters in Malang District from 2018 to 2022

Figure 8 Cluster 1: Area with the highest hotspot density in Malang District from 2018 to 2022

Figure 8 shows us cluster 1. Cluster 1 illustrates that the area with the highest fire points can be seen in Supiturang, Oro-Oro Ombo, and Pronojiwo. In addition, fires were also seen spreading a lot in the Ngadas and Argosari areas. Cluster 1 radius area reach 18.65 km. With the center of the cluster in Latitude -8.000754 and Longitude: 112.933871. From 2018 to 2019, several fires were in protected forest areas and pine forests in cluster 1 regions. This fire has burned 300ha of forest area and was caused by deforestation by burning land.

Figure 9 shows us cluster 2, with medium hotspot density; the fire area is most widely spread in forestry areas around Sumberurip and the Toyomarto. The radius of cluster 2 reaches 12.93 km. Figure 3 shows us cluster 3. Cluster 3 covers the area where the most fires are the Argotirto area, Harjokuncaran, and Ringinkembar. The radius of cluster 3 is 24.57 km.

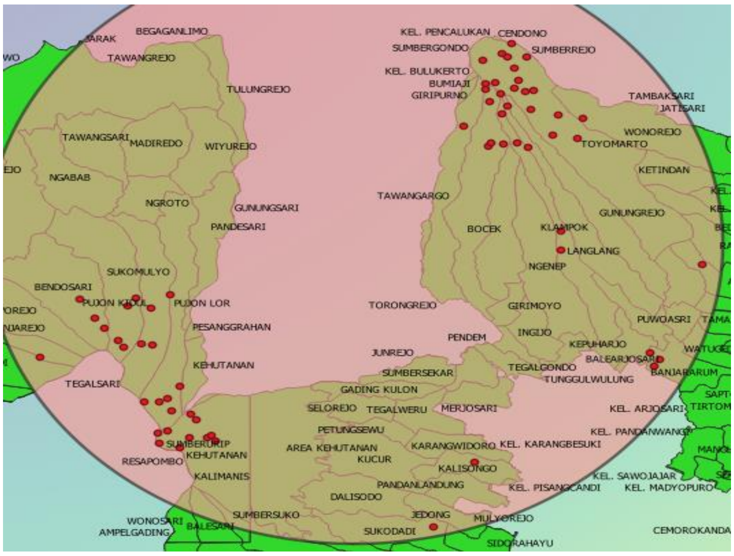

Figure 9 Cluster 2: Area with medium hotspot density in Malang District from 2018 to 2022

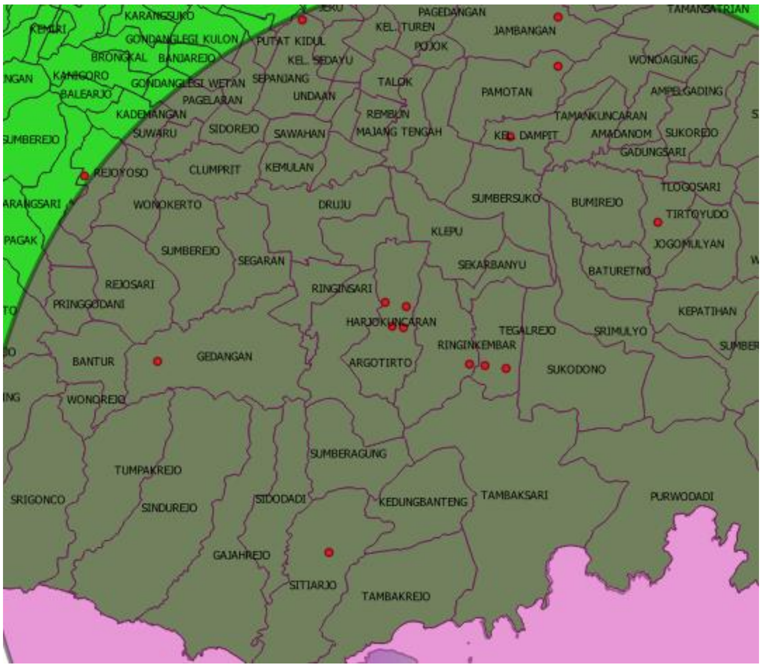

Figure 10 Cluster 2: Area with low hotspot density in Malang District from 2018 to 2022

Conclusion

Based on the results of testing and analysis of hotspot clustering using the K-Means algorithm with the silhouette coefficient test method, this research has succeeded in building a system to analyze the distribution of hotspot clusters in Malang in 2018-2022. The data used as the research object is data on the distribution of hotspots that function as indicators of forest and land fires. The hotspot data studied location is in Malang District with the last 4-year interval, starting from 2018 to 2022. The K-Means method can be implemented to cluster hotspot data with the results of the cluster formed can be monitored for potential fires by looking at the average confidence value of each cluster. Based on the spatial aspects, cluster areas with the highest frequency of hotspots are Tamansari (49%), Ngadas (25%), and Toyomarto (4%). From temporal aspects, the months of January and December are the months with the lowest rate of fires. That month is Indonesia's rainy season, with a higher rainfall level than other months. September and October are the months with the highest rates of fire events. One of the factors causing the high number of fires in the month is the dry season. This study detected 3 clusters for the hotspot area. The area with the highest fire point density can be seen in Supiturang, Oro-Oro Ombo, and Pronojiwo.

References

A. P. Kirana, I. S. Sitanggang, and L. Syaufina, “Hotspot Pattern Distribution in Peat Land Area in Sumatera Based on Spatio Temporal Clustering,” *Procedia Environ. Sci.*, vol. 33, pp. 635–645, 2016, doi: 10.1016/j.proenv.2016.03.118.

A. P. Kirana, I. S. Sitanggang, and L. Syaufina, “Poisson Clustering Process on Hotspot in Peatland Area using Kulldorff’s Scan Statistics Method,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 13, no. 4, p. 1376, Dec. 2015, doi: 10.12928/telkomnika.v13i4.2272.

A. Kirana, I. Sitanggang, L. Syaufina, and A. Bhawiyuga, "Spatial and Temporal Clustering Analysis of Hotspot Pattern Distribution of Critical Land in Kalimantan, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 528, no. 1, p. 012042, Jul. 2020, doi: 10.1088/1755-1315/528/1/012042.

A. P. Kirana, R. Ariyanto, A. R. T. H. Ririd, and E. L. Amalia, "Agricultural drought monitoring based on vegetation health index in East Java Indonesia using MODIS Satellite Data," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 732, no. 1, p. 012063, Jan. 2020, doi: 10.1088/1757-899X/732/1/012063.

A. Kirana, A. Ririd, R. Ariyanto, E. Amalia, and A. Bhawiyuga, "Spatiotemporal Pattern Distribution of Drought Area using MODIS Vegetation Health Index. Case Study: Paddy Field in East Java, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 528, no. 1, p. 012056, Jul. 2020, doi: 10.1088/1755-1315/528/1/012056.

Farid Assifa, “Kebakaran Gunung Semeru Capai 93,3 Hektare,” *kompas*, 2019. https://regional.kompas.com/read/2019/09/29/13364011/kebakaran-gunung-semeru-capai-933-hektare.

H. Sun, X. Zhao, Y. Chen, A. Gong, and J. Yang, "A new agricultural drought monitoring index combining MODIS NDWI and day–night land surface temperatures: a case study in China," *Int. J. Remote Sens.*, vol. 34, no. 24, pp. 8986–9001, Dec. 2013, doi: 10.1080/01431161.2013.860659.

I. S. Sitanggang, A. A. N. Risal, and L. Syaufina, “Incremental Clustering on Hotspot Data as Forest and Land Fires Indicator in Sumatra,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 187, p. 012043, Nov. 2018, doi: 10.1088/1755-1315/187/1/012043.

J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Univ. Calif. Press*, vol. 1, pp. 281–297, 1967, [Online]. Available: https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992.

malangkab.go.id, “Kabupaten Malang Satu Data,” 2018. [Online]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjFlPe-zcL5AhU71XMBHXULBToQFnoECAgQAQ&url=http%3A%2F%2Fmalangkab.go.id%2Fuploads%2Fdokumen%2Fmalangkab-Kondisi%2520Geografis.pdf&usg=AOvVaw11Hr5mNPHeNDwUU8IvScCg.

M. A. Bianto, S. Rahayu, and M. Huda, “Perancangan Sistem Pendeteksi Plagiarisme Terhadap Topik Penelitian Menggunakan Metode K-Means Clustering Dan Model Bayesian,” *Semin. Nas. Teknol. Inf. dan Multimed.*, pp. 19–24, 2018.

Ministry of Environment and Forestry of the Republic of Indonesia, "Forest Fire," 2022. https://dataalam.menlhk.go.id/karhutla/terbaru.

P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.