MINING FIRE HOTSPOTS OVER NUSA TENGGARA AND BALI ISLANDS

Arie Vatresia[1*], Rendra R. Rais[2], Ferzha P. Utama[3], and Widhia Oktarianti[1]

[1]Informatics Engineering, Engineering Faculty, Universitas Bengkulu,
Jl. W.R. Supratman Kandang Limun Street, Bengkulu Indonesia
[2]Ministry of Environmental and Forestry, Indonesia
[3]Information System, Engineering Faculty, Universitas Bengkulu,
JL. W.R. Supratman Kandang Limun, Bengkulu, Indonesia

Received: 22 February 2022, Revised: 13 April 2022, Accepted: 20 April 2022

MINING FIRE HOTSPOTS OVER NUSA TENGGARA AND BALI ISLANDS. Forest fires are still one of the most common problems in Indonesia. In fact, many of these forest fires origin from human activities, namely fires that are intentionally raised for a purpose such as widening the land to prepare for the planting season in the Nusa Tenggara Island. Forest fire events can be identified by observing hotspot data which are monitored through remote sensing satellites. Hotspot is an area that has a relatively higher surface temperature than the surrounding area based on certain temperature thresholds monitored by remote sensing satellites. The area is represented as a point that has certain coordinates. The actual fires can be monitored by observing the hotspot attribute, namely Confidence, Brightness Temperature and FRP (Fire Radiate Power). To find the similarities of the three mentioned attributes, the clustering process is carried out to make monitoring easier. The objective of this research is to cluster hotspots in the Nusa Tenggara and Bali Islands from year 2013 to 2018 using the K-Means Clustering Method with 28,519 hot spot data. This could be a benefit for the Ministry of Environment and Forestry in Indonesia to identify the priority level of the area to be monitored. By knowing this result, the ministry can use this data for patrol priority management. This research successfully clustered three types of hotspot classes based on the risk of fire with details as follow; High Risk Class contains 12,212 data with ranges of mean values of confidence in the range of 49.3–100%, brightness in the range of 305.1–421.3^{o} K and FRP in the range of 2.5–714.3; Medium Risk contains 12,250 data mean values of confidence with a range of 20.3–74.3%, brightness in the range of 301.06–341.86o K and FRP in the range of 3.6–141.4; and Low Risk contains 4,057 data with a range of mean values of confidence in the range of 0–39.8%, brightness in the range of 300–365.86^{o} K and FRP in the range of 3.5–275.6. All of the clusters were obtained by the implementation of K-Means clustering over the hotspot data and its parameter as mentioned, respectively. The cluster performance showed the confidential value of 88.45% accuracy using 100 hotspot data from 2019.

Keywords: Hotspot, Nusa Tenggara and Bali Islands, Data Mining, K-Means, Clustering

MENAMBANG DATA TITIK KEBAKARAN HUTAN DI KEPULAUAN NUSA TENGGARA DAN BALI. Kebakaran hutan masih menjadi salah satu masalah yang sering terjadi di Indonesia. Padahal, kebakaran hutan tersebut banyak berasal dari ulah manusia, yakni kebakaran yang sengaja dimunculkan untuk tujuan seperti pelebaran lahan untuk persiapan musim tanam di Pulau Nusa Tenggara. Peristiwa kebakaran hutan dapat diidentifikasi dengan mengamati data titik api yang dipantau melalui satelit penginderaan jauh. Hotspot adalah suatu daerah yang memiliki suhu permukaan relatif lebih tinggi dari daerah sekitarnya berdasarkan ambang batas suhu tertentu yang dipantau oleh satelit penginderaan jauh. Area direpresentasikan sebagai titik yang memiliki koordinat tertentu. Kebakaran yang sebenarnya dapat dipantau dengan mengamati atribut hotspot yaitu Confidence, Brightness Temperature dan FRP (Fire Radiate Power). Untuk mengetahui kesamaan dari ketiga atribut tersebut maka dilakukan proses clustering untuk mempermudah monitoring. Penelitian ini bertujuan untuk mengelompokkan hotspot di Pulau Nusa Tenggara dari tahun 2013 hingga 2018 menggunakan Metode K-Means Clustering dengan 28.519 data hot spot. Hal ini dapat menjadi manfaat bagi Kementerian Lingkungan Hidup dan Kehutanan di Indonesia untuk mengidentifikasi tingkat prioritas kawasan yang akan dipantau. Dengan mengetahui hasil ini, kementerian dapat menggunakan data ini untuk manajemen prioritas patroli. Penelitian ini berhasil mengelompokkan tiga jenis cluster optimum hotspot berdasarkan risiko kebakaran dengan rincian sebagai berikut;

---

* Corresponding author: arie.vatresia@unib.ac.id

©2022 IJFR. Open access under CC BY-NC-SA license. doi:10.20886/ijfr.2022.9.1.73-85

High High Risk Class berisi 12212 data dengan rentang nilai mean confidence pada rentang 49.3–100%, brightness pada rentang 305.1–421.3o K dan FRP pada rentang 2.5–714.3; Medium Risk berisi 12250 data dengan rentang nilai mean confidence 20,3–74,3%, brightness pada rentang 301.06–341.86°K dan FRP pada rentang 3.6-141.4; dan Low Risk berisi 4.057 data dengan rentang nilai mean confidence pada rentang 0–39.8%, brightness pada rentang 300–365.86°K dan FRP pada rentang 3.5–275.6. Semua klaster diperoleh dengan mengimplementasikan klasterisasi K-Means atas data hotspot dan parameternya masing-masing. Kinerja cluster menunjukkan nilai kerahasiaan akurasi 88,45% menggunakan 100 data hotspot dari tahun 2019.

Kata kunci: *Hotspot, Kepulauan Nusa Tenggara, Penambangan Data, K-Means, Penggerombolan*

I. INTRODUCTION

Indonesia has lost more than 70% of the total forest area due to deforestation (van Etten 2018) (Sirat, Setiawan, & Ramdani, 2018) (Hansen et al., 2013; Santika et al., 2017). One of the causes is forest fires caused by hot temperatures on the surface as a trigger for fires. With its form as an archipelago, it is very difficult for MoEF (Ministry of Environment and Forestry) (Langmann and Heil 2004) to monitor in real terms in the field without the assistance of indication areas that need priority protection (Thah and Sitanggang 2016). Forest and land fire incidents in Indonesia occurred on a large scale in 1982-1983, 1991, 1994, 1997-1998, 2006 and 2015 (Albar et al., 2018; Dennis, 1999; Riyanto et al., 2020). Forest and land fires in 2015, which again threatened Indonesia, had occurred over 80% of Sumatra and Kalimantan lands resulting in the areas were covered with thick smoke (Dennis, 1999; Sutomo and van Etten, 2018). The impacts of forest and land fires have not only affected the health, economy and social society of the community and the nation but have also affected other countries. The forest and land fires in 2015 damaged 2.61 million ha of forest and burned land which caused economic loss of up to 221 trillion rupiah (about 15.4 trillion USD). For this reason, serious efforts are needed to overcome this catastrophe. Prevention efforts need to be initiated by knowing the location of the fire potential spots for further analysis over the causes of forest and land fires (Dennis, 1999). In 2016, efforts to prevent and overcome forest and land fire disasters were showing an improvement due to favourable weather with relatively high and even rainfall throughout the year (Chuvieco, 2011). This improvement was continued in 2018, satellite data showed the number of land and forest fires were decreasing (83%) as the number of hotspots were also decreasing significantly (90%). This phenomenon leads to an interesting correspondence between the number of hotspot data and the number of land fire distribution over Indonesia's landscape. This dynamics event is an interesting things to be analysed to see the occurrences of emerging hotspots (Albar et al., 2018; Andrienko et al., 2010). Furthermore, with the existence of satellites and hotspot detection, protection and prevention processes will run better (Phua et al., 2008; Thah and Sitanggang, 2016; Young et al., 2017).

The Ministry of Environment and Forestry (MoEF) has taken the necessary steps in the field to control forest and land fires by mobilizing support for facilities and infrastructure both at the central and regional levels (Manggala Agni, Forest Ranger), and involving various parties, including the Government. Regions, Disaster Mitigation Body, Indonesian army and Indonesian police department. Apart from taking concrete actions in the field, the MoEF should also conduct efforts to analyse hotspot data and the area of forest and land fires using remote sensing technology. Monitoring activities are carried out through analysis of hotspot data obtained from MODIS Aqua-Terra satellite imagery. The data on the distribution and area of forest and land fires were obtained from the on-screen delineation process based on the latest Landsat 8 OLI imagery data guided by hotspot data (Foga et al., 2017; Viewer et al., n.d.). From the data, we can see that the main problem of fire hotspots in Indonesia is because the fire was not carried out intentionally (Dennis, 1999; Tacconi et al., 2007). Hotspot distribution is clustered naturally, so that if a cluster of hotspot location is known, it can be used in the analysis of fire hotspots. This research tried to cluster the hotspot data from 2013-2019 to obtain the group of hotspot occurrences. Focus area of the clusters covered Nusa Tenggara and Bali Islands, including Bali, which are in the eastern part of Indonesia. . This location is also one of the sources of natural wonders in the world where the biodiversity hot spot taking place, yet there is no current analysis regarding to the term of hot spot over this area. Hence, the processing of data mining could help to understand the problem.

Data mining is an activity that includes collecting, using historical data to find order, pattern and relationship in large datasets; one of the very popular is K-Means clustering (Fischer et al., 2020; Li et al., 2018; Pei et al., 2020). K-Means clustering is a non-hierarchical grouping method that aims to group objects so that the distances of each object to the centre of the group in one group are minimum. K-Means is included in unsupervised learning clustering that is not using training data to make predictions or classifications (Sirat et al., 2019; Wu and Peng, 2017). Distribution and grouping of hotspots can be analysed by the attributes of confidence, brightness, temperature and FRP on hotspot data using the K-Means Clustering method. This research proposed the implementation of K-means clustering algorithm over web based application to help the process of hotspot monitoring in this area within 5-years duration. This also helps the Ministry to integrate the technology for conservation activity. This study showed a new approach in the process of forest fire control using computational algorithms that can classify vulnerable areas with 3 different levels according to their vulnerability based on the recommendation of the best clusters. This research also shows the system that was built on a web basis to facilitate monitoring and changes in the location of hot spots that appear in the Nusa Tenggara and Bali areas. This system is also equipped with interactive graphics and images that make it easier to read and interpret the emergence of hotspots in the area. Furthermore, this research would also show the rate of confidential of the clusters. This result can be implemented along the hotspot data to show the accuracy rate of the algorithm performances.

II. MATERIAL AND METHOD

A. Hotspot

Hotspot is an area that has higher temperature than other areas which can be detected by satellite. The area is represented in a point that has certain coordinates. Satellites known to detect hotspots are NOAA-18/ AVHRR, Terra/Aqua MODIS (Wijedasa et al., 2012) and remote sensing satellite data (Thah and Sitanggang, 2016; Wijedasa et al., 2012). Hotspot's data has attributes to detect fire hotspots. The attributes used in the research are described as **a) Location** at a hotspot using latitude, and longitude as indicating the place where the location of a hotspot occurs; **b) Confidence** or confidence level of hotspot where quality scaled from 0% to 100%. Confidence shows the probabilities of fire occurrences in the field monitored by the

Table 1. Cluster Definition (Thah and Sitanggang, 2016)

Confidence (C)	Class	Action
$0\% \leq C < 30\%$	Low	Need to have attention
$30\% \leq C < 80\%$	Medium	Watchfull
$80\% \leq C \leq 100\%$	High	Immediate countermeasures

satellite image (Broich et al., 2011). The higher the confidence interval, the higher the potential that the hotspot is actually a forest or land fire. The following levels of confidence are shown in Table 1; **c) Brightness temperature,** a descriptive measure of radiation emission in the form of temperature emitted in parts of the Earth's atmosphere (Obregon et al., 2014; Zhu and Woodcock, 2014). Brightness temperature is a basic feature in remote sensing images that are detected in a specific location and measured in units of Kelvin measurement. This research used the temperature that was already captured by the satellite and defined in data columns in each hotspot data to be clustered using K-Means algorithm; **d) FRP** (Fire Radiative Power), describes the radiation power of fire pixels which is then integrated into MW (MegaWatts). FRP provides information on the output of heat radiation from a fire detected. The amount of radiant heat energy is released per unit time (FRP) which is thought to be related to the level of fuel consumed. This research used the value of FRP that was already captured by the satellite for each column in hotspot data.

Table 1. Cluster Definition (Thah and Sitanggang, 2016)

Confidence (C)	Class	Action
0% ≤ C < 30%	Low	Need to have attention
30% ≤ C < 80%	Medium	Watchfull
80% ≤ C ≤ 100%	High	Immediate countermeasures

B. Data Mining

DData mining is an activity of extracting or mining knowledge from large-sized/large amounts of data, this information will later be very useful for development. The purpose of data mining is to specify patterns that must be found in data mining tasks. In general, the purpose of data mining can be grouped into 2, namely to be able to understand more about the behaviour of the observed data, or often referred to as Descriptions, and to be able to estimate conditions that will occur in the future or called Prediction (Phua and Batcha, 2020; Wu et al., 2016).

C. K-Means Clustering

Clustering is an effort to group records, observations, or group them into classes that have similar objects. Clustering is different from classification which has no target variables in clustering. Clustering does not try to classify, estimate, or predict the value of the target variable. However, clustering algorithms try to divide the entire data into groups that have similarities (homogeneous), where the similarity of records in a group will be of maximum value, while similarities with records in other groups will be of minimal value (Atluri et al., 2018). K-Means is one method of non-hierarchical clustering data that attempts to partition existing data into one or more clusters/groups (Compieta et al., 2007). K-Means is included in unsupervised learning clustering that does not use training data or training data to make predictions or classifications. Based on the mathematical model, this algorithm does not have a variable target. One purpose of this algorithm is to group objects that are almost the same in a particular area (Tork, 2012; Wu, 2012). The clustering process begins by identifying data to be clustered. At the beginning of the iteration, the center of each cluster is set freely. Then the distance between the data with each cluster center is calculated. Euclidean formula can be used to calculate the distance of the i data in the center of the cluster (Ren et al., 2020; Wu, 2012) as in equation (1):

$$d(x, y) = \sqrt{\sum_{j=1}^{d} \{x_j - y_j\}^2} \quad \text{.............................. (1)}$$

where:
$d(x,y)$ = distance between x-data to y-data center
xj = data *object*
yj = data *centroid*
P = Number of Attributes

The new central cluster value can be calculated by finding the average value of the data that is a member of the cluster, using the formula in equation (2) (Ren et al., 2020).

$$Vxy = \frac{\sum_{k=1}^{Ny} Xkx}{Ny} \quad \text{...} (2)$$

where:
Vxy= x-cluster data for column y
Xkx = the k-data for the x-column
Ny = number of y cluster members

D. Data Normalization

Normalization is a transformation process to change the value of data and is used to equalize the scale of data attributes into a smaller specific range such as -1 to 1 or 0 to 1. Min-Max Normalization is a normalization technique by performing linear transformations on the original data attributes to produce a range of the same value in equation (3) (Young et al., 2017).

$$v' = \frac{v - Min_A}{Max_A - Min_A} \quad \text{...} (3)$$

where:
v '= normalized v data
v = data v
MinA = minimum value in column A
MaxA = maximum value in column A

E. Data Collection Method

Satellites known to detect hotspots are NOAA Satellite, Terra/Aqua MODIS, and remote sensing satellite data. All of data and material in this study was collected from NASA's Fire Information for Resource Management System (NASA FIRMS) Modis Catalog (https://firms.modaps.eosdis.nasa.gov/) for the area of Nusa Tenggara and Bali Islands (2013 to 2018). Currently, this data is still the most effective in monitoring quickly land and forest fires for a large area. Current remote sensing satellite technology allows monitoring of land and forest fires in near real time (Endrawati, 2018; Langner and Siegert, 2009; Wijedasa et al., 2012).

F. Data Processing Method

Data processing aims to convert raw data from measurement results into finer data so as to provide direction for further assessment. This stage is done to remove hotspot attributes that are not needed by using Microsoft Excel software. The hotspot attributes used are latitude, longitude, acq_date, Confidence, Brightness Temperature, and Fire Radiative Power. After the data is clean, the data in excel format (.xml) is transformed into the format in the database (.csv). The next clean data will be input into the MySQL database and then proceed with making the program using the PHP programming language and CodeIgniter Fremework using the K-Means Clustering method. The flow diagram of the K-Means Clustering method is described in Figure 1.

The data processing was started by determining the optimum cluster numbers. This step occupied the Elbow method (Liu and Deng, 2021; Purnima et al., 2014; Shi et al., 2021) that resulted the optimum number of clusters in the data was 3. After determining the number of clusters, the process was continued to select the centroid of the data for each clusters available, the data will be grouped based on the nearest distance from the centroid

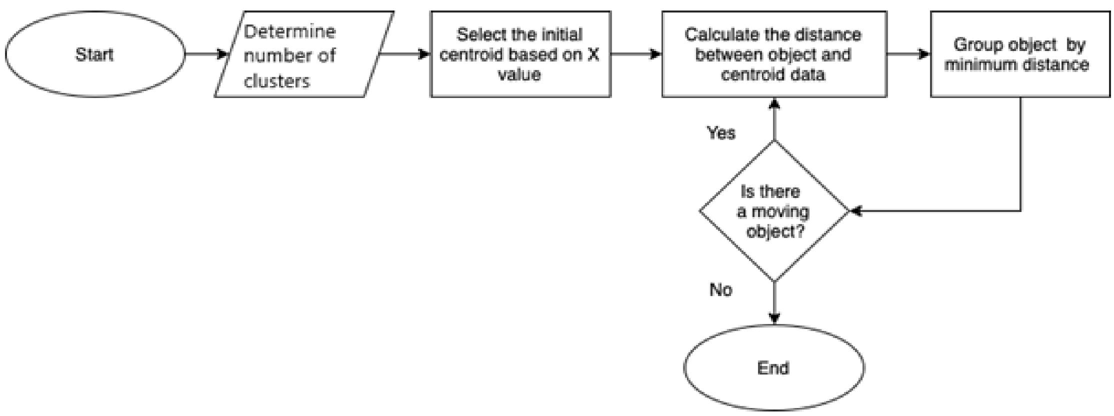

Figure 1. Flow Diagram of K-Means Clustering Method

as previously stated. If there was still moving objects, then the progress would repeat the calculation to the nearest distance until all of the data were clustered.

G. Calculation of K-Means Clustering Algorithm

In the process of calculating vulnerability using the K-Means attribute algorithm that is used 3 attributes of hostpots, i.e. confidence, brightness and FRP with total hostpot data of 28,519. Following are the steps for calculating K-Means Clustering:

1) *Enter the hotspot attribute data into the table in the database*

The hotspot attribute data is entered into the database displayed in the form of a table consisting of latitude, longitude, area, date, confidence, brightness, and FRP based on 28,519 data.

2) *Determine many classes and initial centroids*

This research used 3 classes of Medium Clusters for nominal hotspots, High Clusters for high hotspots and Low Clusters for low hotspots. The number of clusters were determined by Elbow method to see the optimum clusters of the data (Liu and Deng, 2021; Purnima et al., 2014). Clusters were performed by determining the centroid value based on hotspot's monthly data and set the data on certain cluster based on the nearest distance. The selection of the initial centroid value was done so that the value of the cluster each month will be the same. The following samples from the initial centroid in January to April 2013 is described in Table 2:

3) *Data Normalization: using equation (3)*

Data normalization is the process of scaling attribute values from data so that it can fall in a certain range. In this study using min-max normalization, standardizing data by placing data in the range 0 to 1, the smallest value as 0, and the largest value as 1.

4) *Calculate Cluster distance : using equation (1)*

The next step is to calculate the distance between the data object and the initial centroid (Liu and Deng, 2021; Wu, 2012). The data were grouped on the clusters based on the distance calculation process. After doing the calculation, the value of the distance was obtained for each cluster. Finally, the results of the cluster one and the others are compared, then for the value of the distance the results of the smallest cluster is chosen as a member of the cluster of the group.

5) *Calculating the new centroid (cluster center): equation (2)*

After the cluster value is obtained, then calculate the new centroid value instead of the initial centroid value for the next iteration. The

Table 2. Table of sample values from the initial centroid

Month	Cluster	Confidence	Brightness	FRP
January	K1	46	311.6	8.3
	K2	25	311.6	14.5
	K3	92	315.8	15.7
February	K1	48	306.7	10
	K2	15	305.1	6.1
	K3	78	312.2	11.3
March	K1	41	312.2	4.4
	K2	0	328.1	21.5
	K3	97	318.8	24.4
April	K1	47	315.7	18
	K2	22	313.6	10.7
	K3	100	324.9	49.2

process of finding a new centroid by summing the value of the data object included in a group is then divided into a lot of data from the group.

6) *Calculate the cluster distance value again: equation (1)*

After obtaining the new centroid value, we will return to the initial calculation using formula 1 in grouping the data based on the results of calculations in the next iteration. The clustering process will continue to be carried out until the results of the last iteration grouping are the same as the results of grouping the previous iterations which in this case the data object has not changed position in the previous iteration.

III. RESULT AND DISCUSSION

1. System Development

The system was built using the Data Flow Diagram and was developed using PHP and CodeIgneter framework. The hotspot classification page is a page that contains the results of the implementation of the clustering process and displays it into a map (Figure 2) and the data will be in the form of colored coordinate points based on the level of vulnerability.

On this page, there are categories of choices that are used to display the results of clustering. The categories are **(1) Date Category**: a choice category for displaying data based on the selected date or time range; **(2) Regional Categories**: The contents of this category are 4 major islands in the Nusa Tenggara and Bali Islands, namely Bali Island, West Nusa Tenggara Island, East Nusa Tenggara Island and Southwest Maluku Island; **(3) Cluster Category**: This research clustered the hotspot data by its risk level that were defined as Low Risk Cluster, Medium Risk Cluster and High Risk Clusters. The clustering page shows the distribution of hotspot in Nusa Tenggara and Bali Islands, including Bali based on 3 clusters as criteria. The page also allowed one to modify the result based on date, area, and criteria. This would help the users to fit the needs of each criterion. In addition, on this page there is a choice of regional categories where one can choose the area to be displayed, namely the 4 large islands in the Nusa Tenggara Islands, namely Bali Island, West Nusa Tenggara Island, East Nusa Tenggara Island and Southwest Maluku. In addition to category choices, this mapping page displays the amount of data for each existing cluster and there is detailed hotspot information when the user presses one of the data on the map.

The clustering results page displays a hotspot data table with attributes of latitude, longitude, area, date, confidence, brightness,

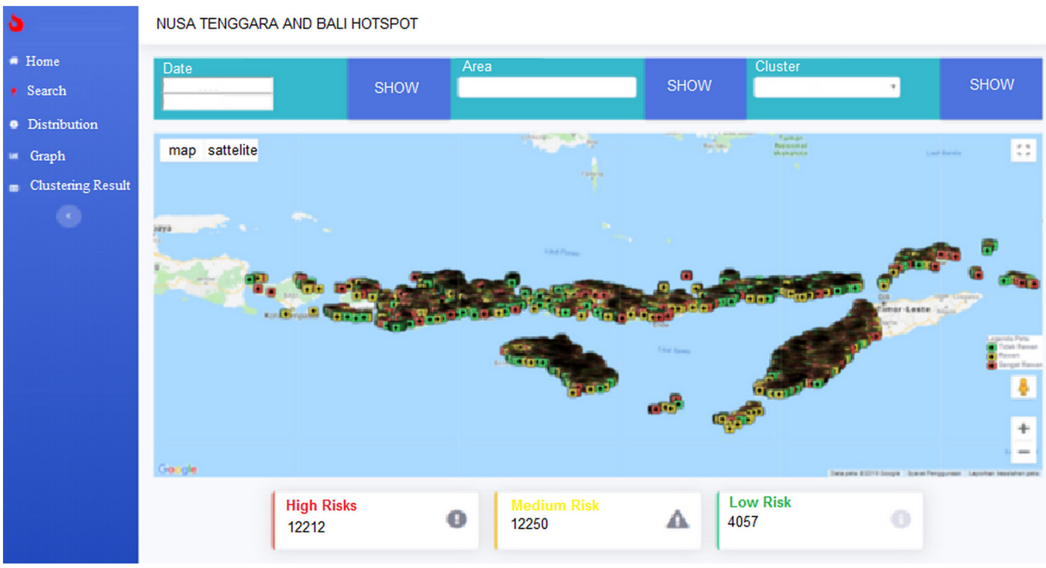

Figure 2. Hotspot classification page

FRP, C1 value, C2 value, C3 value and cluster. The colors in columns C1, C2, and C3 represent the classes in the cluster results. Green colour represents low risks, yellow colour represents medium risks and red colour represents very high risks. This colour determination is based on the smallest C value in the data. Figure 3 is the page containing instructions for carrying out the clustering process by generating data that will be clustered by year with one clustering process using monthly data every year. On this page, one can also perform the order to print the clustering result table data by entering the existing print categories, region and year.

In the website system, the description contains a display of data from the clustering process that has been implemented into a map where each coordinate point represents a data. This description aims to see the distribution of hotspot that occurs in a range of months in each year. Hotspot distribution can be selected certain years by moving the slider in the month section. In Figure 4, it can be seen that the average number of hotspots has showed the lowest number in January and the highest numbers were in October each year. The visualization of hotspot occurrence over Nusa Tenggara islands and its distribution of the data in 2013 are shown in Figure 4.

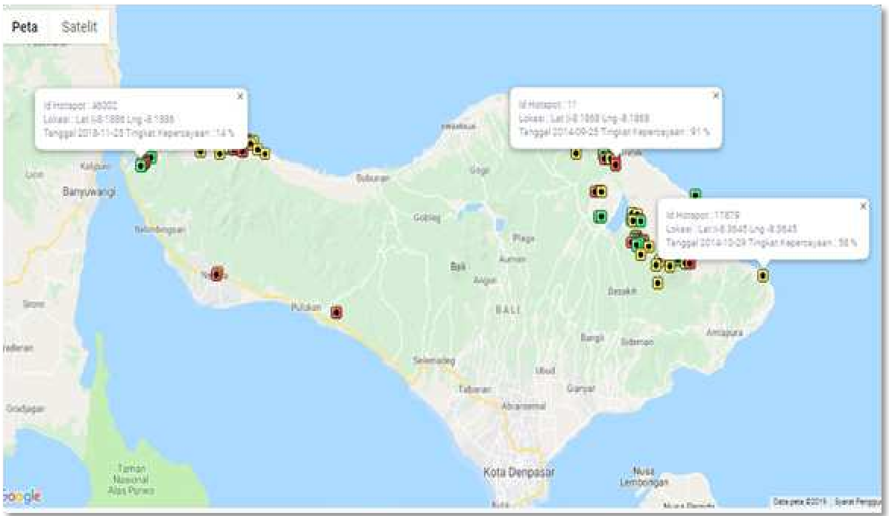

Figure 3. Detailed view of hotspot data

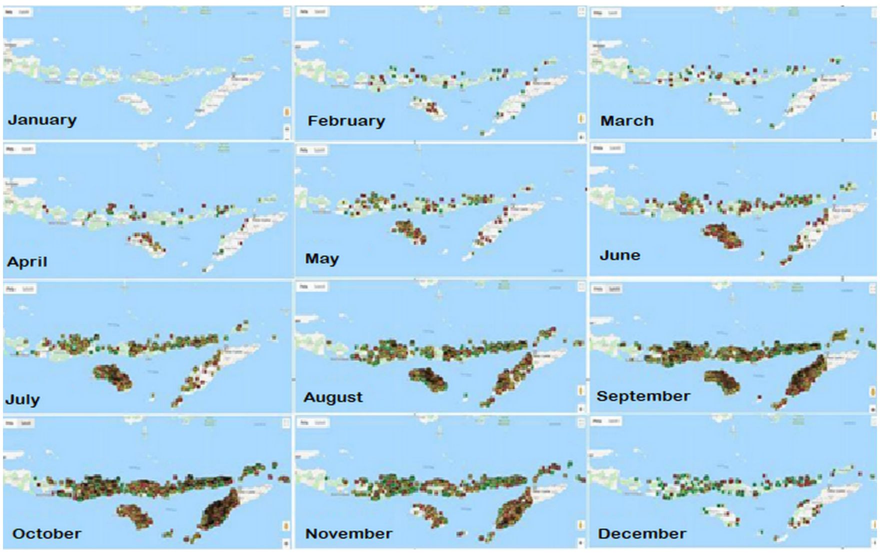

Figure 4. Monthly Hotspot Distribution in 2013

2. Clusters Distribution

From the results of the monthly data clustering process from 2013 to 2018 it was found that the results of this study obtained 3 types of Cluster classes with details of high risks reaches as many as 12,212 data with a range of average values of confidence in the range of 49.3 – 100%, brightness in the range of 305.1 - 421.3 oK and FRP in the range of 2.5–714.3, medium risks reaches as many as 12,250 data with a range of mean values of confidence in 20.3–74.3%, brigthness in the range of 301.06 - 341.86oK and FRP in the range of 3.6 - 141.4 and low risks reaches 4,057 data with a range of average values of confidence in the range of 0 - 39.8%, brightness in the range of 300–365.86 °K and FRP in the range of 3.5–275.6. Figure 5 shows the monthly clustering results over a year. From Figure 5 we found that the high risks cluster represented in red are higher than the medium risk cluster represented in yellow colour, and the low risks cluster represented by color green. The clustering results showed that the lowest number of hotspot data occurred in February and the highest number occurred in October.

In 2014, the results of hotspot showed 7,234 data comprised of 45 data for Bali, 4,883 data for West Nusa Tenggara, 2,078 data for East Nusa Tenggara and 273 data for Maluku. Low risks cluster showed 1,163 data results, medium risks cluster 3,263 data results and high risks 2,808 data results.

The description and interpretation of the clusters are described in Table 3. From Table 3, it is found that the brightness and FRP values are related where the brightness value is high, the FRP is also high in each cluster group but the value is not constant every month. This is because brightness and FRP have no definite range of values such as confidence. Furthermore, from Table 3 an average range of all years will be sought so that the differences in the values of each cluster are shown in Table 3.

In Table 3 it can be seen that the clear differentiator of each cluster is seen in the value of confidence. In cluster 1, confidence clusters was in the normal range of hotspots which meant this cluster belong to the medium risk class. The confidence value in cluster 2 range was included in the low risk hotspot class, which means that the area in this class has the lowest risk from fire burn. Cluster 3 was included in

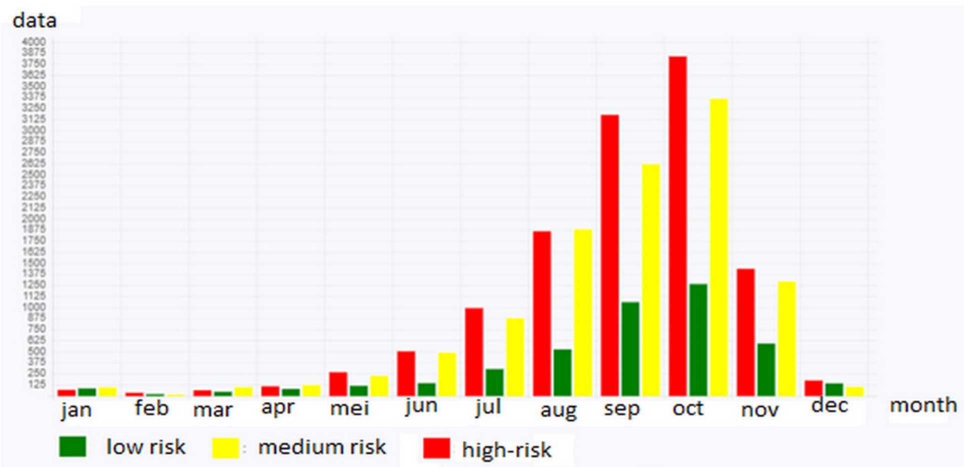

Figure 5. Graph of Clustering Result in 2013

Table 3. Average Clustering Results

Cluster	Confidence	Brightness	FRP	Status
Cluster 1	20.3-74.3	301.06-341.86	3.6-141.4	Medium
Cluster 2	0-39.8	300-365.86	3.5-275.6	Low
Cluster 3	49.3-100	305.1-421.3	2.5-714.3	High

the high risk hotspot class which means the class was very vulnerable to burn with fire. The distribution of the criteria were changed over time and affected the hotspot to be grouped in different clusters. The distribution of the dynamic change of criteria is described in Figure 6. It describes the results of clustering of confidence, brightness and FRP for each year.

This research clustered over 5 years of hotspot's occurrences over the area and found the distribution of hotspots as described in Figure 7. From the Figure we can see that the highest occurrence of the hotspot over Nusa Tenggara islands were in year 2014 while the lowest occurrence was in 2016 that dropped over 250% of hotspot occurrence over the area. In Figure 7 we can also see that the low risk has occurred very low every year and was dominated by the high risk over the year.

C. Validation

This research includes the process to measure the performance of K-Means to see how the algorithm worked over the hotspots clustering. The performance of K-Means clustering algorithm was increasing as the number of data processed increased. The average of the hotspots running time was 0.321 s for 4,500 hotspot data. Furthermore, this research used the data of hotspot occurrence in 2019 data and mapped the model with the algorithm to validate the result of clusters. Over 100 data were used to measure the accuracy of the clustering method. This research showed that the performance of K-Means for hotspot data were valid at a rate of 88.45%.

IV. CONCLUSION

This research has succeeded in mining hotspot data on the Nusa Tenggara Islands in 2013 using the K-means clustering algorithm method using 3 hotspot attributes (confidence, brightness and FRP). The results of the study obtained 3 types of hotspot classes. Clustering results attributes were generated in each cluster group. Determination of the name of this cluster by looking at the ranking of the results of the 3 clusters and looking at the value of the attribute range. Cluster 1 is categorized as medium risks (12,250 data) with an average value range of confidence in the range of 20.3 – 74.3%, brigthness in the range of 301.06 – 341.86oK and FRP in the range of 3.6–141.4. Cluster 2 belongs to the low risks (4,057 data) with an average value range the average of

Figure 6. (a) Annual Confidence Chart; (b) Annual Brightness Chart; (c) Annual FRP Chart

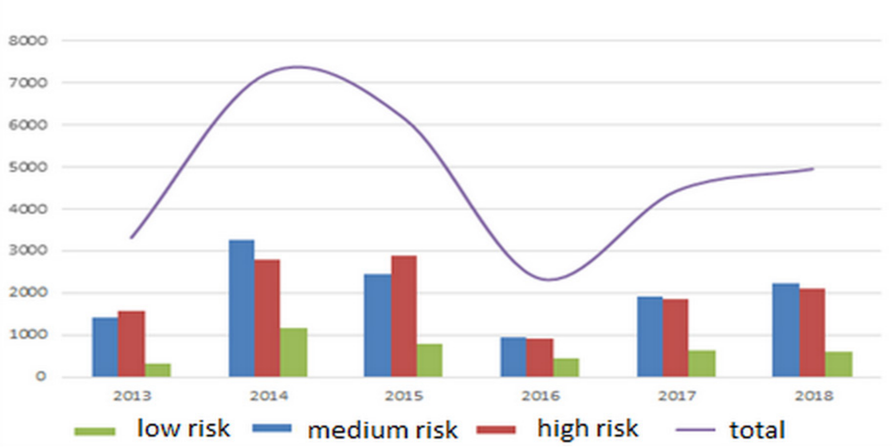

Figure 7. Annual Hotspot Occurrences Comparison

confidence in the range of 0 - 39.8, brightness in the range of 300–365.86 and FRP in the range of 3.5–275.6. Cluster 3 is included in the high risk (12,212 data) with an average value range of confidence in the range of 49, 3–100, brightness in the range of 305.1–421.3 and FRP in the range of 2.5–714.3. The results of this study show that the Nusa Tenggara Islands hotspot from 2013-2018 has an average confidence value range above the average set by LAPAN and the brightness and FRP values in each cluster have a different range each month and the highest range. is in October which month has also a lot of data. The more hotspot data, the greater the brightness and FRP range. This system has succeeded in visualizing the distribution of hotspots, it is found that the distribution of hotspots occurs the most in October and least in February in every year from 2013– 2018. .

ACKNOWLEDGEMENT

This research was supported by the Ministry of Environmental and Forestry that also provide available data that makes this research successfully implemented and helping us to expand the data for this paper. Las but not least, we would like to thank LPPM Universitas Bengkulu for making this research available.

REFERENCES

Albar, I., Jaya, I.N.S., Saharjo, B.H., Kuncahyo, B., Vadrevu, K.P. (2018). Spatio-temporal analysis of land and forest fires in Indonesia using MODIS Active fire dataset. Springer Remote Sensing/Photogrammetry. Springer, Cham. doi://10.1007/978-3-319-67474-2_6.

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S.I., Jern, M., Kraak, M.J., Schumann, H., Tominski, C., 2010. Space, time and visual analytics. *International Journal of Geographical Information Science*. *24*, 1577–1600. doi://10.1080/13658816.2010.508043.

Atluri, G., Karpatne, A., Kumar, V., 2018. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys*. *51*(4) Article 83, 1-41. doi://10.1145/3161602.

Broich, M., Hansen, M.C., Potapov, P., Adusei, B., Lindquist, E., & Stehman, S. V. (2011). Time-series analysis of multi-resolution optical imagery for quantifying forest cover loss in Sumatra and Kalimantan, Indonesia. *International Journal of Applied Earth Observation and Geoinformation*. *13*, 277–291. doi://10.1016/j.jag.2010.11.004.

Compieta, P., Di Martino, S., Bertolotto, M., Ferruci, F., & Kechadi, T. (2007). Exploratory spatio-temporal data mining and visualization. *Journal of Visual Languages and Computing*. *18*(3), 255-279. doi://10.1016/j.jvlc.2007.02.006.

Dennis, R. (1999). *A review of fire projects in Indonesia (1982-199)*. Bogor: CIFOR. doi://10.17528/cifor/000564.

Endrawati, E., (2018). Identifikasi areal bekas kebakaran hutan dan lahan menggunakan analisis semi otomatis citra satelit landsat. *Seminar Nasional Geomatika 2*, 273. doi://10.24895/sng.2017.2-0.420.

Fischer, C., Pardos, Z.A., Baker, R.S., Williams, J.J., Smyth, P.,....Warschauer, M. (2020). Mining big data in. Mining big data in education: Affordances and challenges. *Review of Research in Education. 44*(1), 130-160. doi://10.3102/0091732X20903304.

Foga, S.C., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., Schmidt, G.L.,....,Laue, B.(2017).. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment. 194* (June), 379–390. doi://10.1016/j.rse.2017.03.026.

Langner, A., & Siegert, F. (2009). Spatiotemporal fire occurrence in Borneo over a period of 10 years. *Global Change Biology. 15*(1), 48–62. doi://10.1111/j.1365-2486.2008.01828.x.

Li, Z., Liu, Q., Tang, J., & Deng, M.(2018). An adaptive method for clustering spatio-temporal events. *Transactions in GIS*, *22*, 323-347. doi://10.1111/tgis.12312.

Liu, F., & Deng, Y. (2021). Determine the number of unknown targets in open world based on elbow method. I*EEE Transactions on Fuzzy Systems*. *2*9 (5), 986-995. doi://10.1109/TFUZZ.2020.2966182.

Obregon, A., Gehrig-Downie, C., Gradstein, S.R., & Bendix, J. (2014). The potential distribution of tropical lowland cloud forest as revealed by a novel MODIS-based fog/low stratus night-time detection scheme. *Remote sensing of environment. 155*, 312–324. doi://10.1016/j.rse.2014.09.005.

Pei, T., Song, C., Guo, S., Shu, H., Liu, Y., Du, Y., Ma, T., & Zhou, C. (2020). Big geodata mining: Objective, connotations and research issues. *Journal of Geographical Sciences. 30*, 251-266. doi://10.1007/s11442-020-1726-7.

Phua, E.J., & Batcha, N.K. (2020). Comparative analysis of ensemble algorithms'prediction accuracies in education data mining. *Journal of Critical Review. 7*(3), 37-40 doi://10.31838/jcr.07.03.06.

Purnima, B., Arvind, K., Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and K-means in WSN. *International Journal of Computer Applications. 105*(9),17-24. doi:// 10.5120/18405-9674.

Ren, Y., Wang, N., Li, M., & Xu, Z. (2020). Deep density-based image clustering. *Knowledge-Based Systems. 19*7 (7). doi://10.1016/j.knosys.2020.105841.

Riyanto, I.A., Cahyadi, A., Kurniadhini, F., Bachtiar, H., Apriyana, D., & Caraka, B.K.A. (2020). Understanding forest fire management in indonesia from a global perspective. *ASEAN Journal on Science & Technology for Development. 37*(1). doi://10.29037/ajstd.593.

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP Journal on Wireless Communications and Networking. 31*. doi://10.1186/s13638-021-01910-w.

Sirat, E.F., Setiawan, B.D., & Ramdani, F. (2019). Comparative analysis of K-means and isodata algorithms for clustering of fire point data in Sumatra region. The *4th International Symposium on Geoinformatics, ISyG 2018*. doi://10.1109/ISYG.2018.8611879.

Sutomo, & van Etten, E. (2018). Spatial and temporal patterns of fires in tropical savannas of Indonesia. *Singapore Journal of Tropical Geography. 39*, 281-299. doi://10.1111/sjtg.12243.

Tacconi, L., Moore, P. F., & Kaimowitz, D. (2007). Fires in tropical forests - What is really the problem? Lessons from Indonesia. *Mitigation and Adaptation Strategies for Global Change. 12*(1)55-66.doi://10.1007/s11027-006-9040-y.

Thah, P. H., & Sitanggang, I.S. (2016). Contextual outlier detection on hotspot data in Riau Province using K-means algorithm. *Procedia Environmental Sciences. 3*3, 258-268. doi://10.1016/j.proenv.2016.03.077.

Tork, H.F. (2012). Spatio-temporal clustering methods classification. Proceeding of the 7th Doctoral Symposium in Informatic Engineering.

Wijedasa, L.S., Sloan, S., Michelakis, D. G., & Clements, G.R. (2012). Overcoming limitations with landsat imagery for mapping of peat swamp forests in sundaland. *Remote Sensing. 4*, 2595–2618. doi://10.3390/rs4092595.

Wu, C., Liu, G., Zhang, X., He, Z., & Zhang, Z. (2016). Discussion on geological science big data and its applications. Kexue Tongbao/ *Chinese Science Bulletin*. 61. doi://10.1360/N972015-01035.

Wu, J. (2012). *Advances in K-means clustering: A data mining thinking*. Springer Theses: recognizing outstanding Ph.D. Research.

Wu, W., & Peng, M. (2017). A data mining approach combining K-means clustering with bagging neural network for short-term wind power forecasting. *IEEE Internet of Things Journal*. *4*(4), 979-986. doi://10.1109/JIOT.2017.2677578.

Young, N.E., Anderson, R.S., Chignell, S.M., Vorster, A.G., Lawrence, R., & Evangelista, P.H. (2017). A survival guide to Landsat preprocessing. *Ecology*, *9*8, 920–932. doi://10.1002/ecy.1730.

Zhu, Z., & Woodcock, C.E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing Environment*, *144*, 152–171. doi://10.1016/j.rse.2014.01.011.