PAPER • OPEN ACCESS

Confidence Analysis of Hotspot as Peat Forest Fire Indicator

To cite this article: D A Shofiana and I S Sitanggang 2021 J. Phys.: Conf. Ser. 1751 012035

View the article online for updates and enhancements.



IOP ebooks[™]

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection-download the first chapter of every title for free.

Confidence Analysis of Hotspot as Peat Forest Fire Indicator

D A Shofiana¹, I S Sitanggang²

¹ Department of Computer Science, Universitas Lampung, Bandar Lampung 35141, Indonesia ² Department of Computer Science, IPB University, Bogor 16680, Indonesia

email: dewi.asiah@fmipa.unila.ac.id¹, imas.sitanggang@apps.ipb.ac.id²

Abstract. Indonesia has recorded 14.9 million hectares of peat forest that continue to be deforested due to fire across Sumatra and Kalimantan. To operate a successful firefight, fast detection is a key element. Hotspot that appeared consecutively in more than two days is a strong indicator of fire existence. As the interest in data mining arose, an advanced technique can be implemented toward hotspot dataset into finding solutions. Many previous works have been carried out to mine sequence patterns and succeeded in determining as well as predicting areas with high occurrence of fire. However, none of the studies analyses the outliers, such as several hotspots which confidence decrease significantly in an adjacent interval of time. Confidence determines the quality of hotspot, with a value above 70% strongly indicates that fire spot exist. This study generated sequence patterns using the SPADE algorithm and analyses 21 hotspots considered as outliers using the Landsat-8 image. The result shows that 85.71% of hotspots have decreased confidence due to haze cover.

Keyword: confidence, hotspot, Landsat 8, peatland fire, sequential pattern mining, SPADE.

1. Introduction

Indonesia has approximately 14.9 million hectares of peatland distributed through Papua, Kalimantan, and Sumatra. Peatland forest holds a vital role in our ecosystem, but the fire that occurs periodically threats its preservation. The continuous deforestation triggers various parties to partake in firefighting. Fast detection is a decisive factor for a successful firefight.

One indicator of fire occurrence is the hotspot. It is acquired through remote sensing by satellite, one of which is NASA's Earth Observing System (EOS). Continuous detection of hotspots on a large scale can be a strong indicator of fire. Nevertheless, not every detection is accurate, as the industrial area which has a higher temperature than its surroundings is considered as a false alarm. The hotspot confidence level is the conviction rate of a hotspot being fire spot. [1] stated that confidence is the quality of hotspot, determined as the geometric mean approached by five sub-confidence parameters. Each sub-confidence focuses on one hotspot aspect such as temperature, cloud, haze, and water consistency. These aspects determine whether a spot detected by the moderate resolution imaging spectroradiometer (MODIS) instrument in the EOS is a true hotspot.

Dataset of hotspot itself includes spatial and temporal information. It has time and location attributes collected daily, issuing a vast dimension and size of data. These data need to be processed in some certain procedures; thus; it can produce valuable information referred in decision making. On the other hand, the concern in data mining emerges due to the advance of information technology driving

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution Ð of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd 1

to an exponential growth of scientific and engineering databases [2]. Data mining is considered as an advanced alternative which generates trends and patterns with limited sources that can improve decision making. Classic statistical analysis will not work on such vast data like hotspot dataset. Therefore, data mining tools can be an alternative to analyze and extract high-level information from a dataset [3].

Indeed, several data mining methods have been applied to the forest fire domain. [4] applied Decision Tree, Random Forest, and Logistic Regression to detect fire appearance in the Slovenian forests, which shows that the decision tree produces the best model. [5] applied the DBSCAN Algorithm on the hotspot data in Sumatra in the years of 2002 and 2013. Many other publications had implied the MODIS data analysis using various data mining techniques ([6], [7], [8], [9], [10]). There are several techniques in analyzing as well as extracting information in data mining, among is sequential pattern mining. Sequential pattern mining is a process to extract sequential patterns from a sequence dataset, which its support exceeds the minimum support [11]. In hotspot case, it can obtain information on the longest fire occurrence and how extensive it is based on its location.

Several works correlated to sequential pattern mining in the fire domain have been conducted. In 2015, [12] and [13] applied PrefixSpan algorithm and Clospan algorithm to generate sequence patterns from Riau Province's hotspot dataset within the years of 2000-2014. Based on the sequence patterns, it is concluded that a consecutive appearance in more than two days is the minimum interval of hotspot occurrence to be stated as a strong indicator of fire. Another study by [14] conducted research to identify fire spots based on the hotspots resulted from sequential pattern mining using PrefixSpan algorithm, as well as burned area classification using satellite imagery. Furthermore, [7] generate hotspot sequences with an improvement in spatial features. Previous studies [12], [13], and [15] cut the decimal digits of the locations of hotspots from three to two digits, affecting the imprecise of the actual location. In the latest work, they maintain the three digits, which generated more precise hotspots locations, making it more beneficial for the officials in fire detection and extinguishment [7].

Many kinds of researches of hotspot sequences have been conducted but none related to the confidence analysis. This work analyzes the confidence of hotspot sequences approached using sequential pattern mining and validate the hotspot with the Landsat-8 satellite image. Sequential pattern mining has several possible algorithms, such as GSP, PrefixSpan, and SPADE. Regarding the study by [16], in terms of computation and memory efficiency, PrefixSpan is still the best. However, the number of frequent sequences generated is less than the other two algorithms. Therefore, this research applies the SPADE algorithm that can mine more frequent sequences with a faster computation rate than GSP. The analysis of hotspot sequences is carried out so that the level of confidence of a hotspot with a high chance of becoming a fire spot can be known.

2. Materials and Methods

2.1. Data

Hotspot datasets were acquired from the FIRMS NASA. Sumatra and Kalimantan hotspot data from the year 2014-2015 were selected as the study interest, because of its well-known high occurrence of fire within that period. Peatland map provided by Wetland International is used for hotspot selection, and Landsat-8 satellite image acquired from LAPAN is used to verify the analysis result at the end. The attributes of hotspot dataset are presented in Table 1. This research was conducted in four main steps: data preprocessing, sequential pattern mining, confidence analysis, and verification with satellite images.

	Table 1. Attributes of hotspot ^{**}				
No	Attribute	Description			
1	Latitude	Coordinate of hotspot (°)			
2	Longitude	Coordinate of hotspot (°)			
3	Brightness Temperature	Temperature of hotspot in channel 21 and 22 (K)			
4	Scan	Width of satellite image (pixel)			
5	Track	Length of satellite image (pixel)			

ICASMI 2020

Journal of Physics: Conference Series

IOP Publishing

1751 (2021) 012035 doi:10.1088/1742-6596/1751/1/012035

No	Attribute	Description
6	Acq_date	Acquisition date of hotspot
7	Acq_time	Acquisition time of hotspot
8	Satelite	Satelite (Aqua, Tera)
9	Confidence	Hotspot quality (0–100%)
10	Version	5.0 = MODIS NASA-LANCE
		5.1 = MODIS MODAPS-FIRMS
11	Bright_t31	Temperature of hotspot in channel 31 (K)
12	Frp	Fire radiative vigor (MegaWatts)
Note: ^a Source: https://earthdata.nasa.gov		

2.2. Data Preprocessing

To properly generate accurate sequence patterns using SPADE algorithm, essential preprocessing steps are required. Steps held in this study including data selection and transformation. Dataset acquired from FIRMS NASA is divided into four parts: Kalimantan 2014, Kalimantan 2015, Sumatra 2014, and Sumatra 2015. As referring to [7], this work does not cut the decimal digits of hotspot location coordinate. Selection is carried out to set the data within the research area for further in-depth analysis. Six selected attributes of hotspots in this research are longitude, latitude, acquisition date (acq_date), acquisition time (acq_time), brightness temperature, and confidence. Peatland map from Wetland International was utilized to select hotspot within the peatland area. The number of hotspots in each part of the dataset is provided in Table 2.

Table 2. Number of hotspots (before and after selection)

Dataset	Number of hotspots			
	Before Selection	After Selection		
Kalimantan 2014	40,876	32,575		
Kalimantan 2015	75,565	32,491		
Sumatra 2014	51,760	40,631		
Sumatra 2015	62,709	34,497		

Transformation of datasets into a sequential data format was carried out as the preparation for SPADE Algorithm input. Figure 1 shows the format of sequential data input, with the left to right are sequence ID (SID), date_code or event ID (EID), size, and items.

51	1444853600	1	1444853600	
52	1442793600	2	1442793600	1442793600
52	1442966400	1	1442966400	
53	1444780800	1	1444780800	

Figure 1. Sequential Data Input Format

EID is a variable that represents the date of hotspot occurrence, which is expressed in an integer that resulted from the UNIXTIME function. Items column appear appertaining to the value in size column. It expresses one or more EID which implies the sequences of hotspots. For instance, in Figure 1 exist SID '52' with EID '1442966400' indicating that only one hotspot found in an area, so the EID is recorded once in the Items column. Meanwhile, if two hotspots were found such as SID '52' with EID '1442793600', then EID is recorded twice in the Items column.

2.3. Sequential Pattern Mining

Sequential pattern is a pattern presenting a sequence of events. Pattern will be found when the data of events are relatively large, and it occurs several times in a row [17]. Sequential pattern mining is considered as an advanced data mining technique which can extract sequential patterns from a dataset with its support value exceeds the minimum support. Minimum support value is defined by its

1751 (2021) 012035 doi:10.1088/1742-6596/1751/1/012035

user, which acts as a threshold on selecting patterns, where less-interesting patterns can be omitted. Thus, the pattern discovery process will be effective and efficient [11].

A sequential pattern is a pattern discovered from itemset, where every item occurs at nearly the same time [18]. Sequence $\langle a_1, a_2, ..., a_n \rangle$ is encompassed in sequence $\langle b_1, b_2, ..., b_m \rangle$, if $i_1 < i_2 < ... < i_n$ and $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, ..., a_n \subseteq b_{i_n}$. As an example, sequence $\langle (Q)(S, T, V)(T, V) \rangle$ is encompassed in sequence $\langle (P)(Q)(S, T, U, V)(T, V) \rangle$ since $(Q) \subseteq (), (S, T, V) \subseteq (S, T, U, V), (T, I) \subseteq (T, V)$. However, $\langle (P)(Q) \rangle$ is not part of $\langle (P, Q) \rangle$ since $\langle (P)(Q) \rangle$ implies Q happens after P whereas $\langle (P, Q) \rangle$ indicates P occurred concurrently as Q. A maximum sequence is a sequence that is not encompassed in any other sequences [18].

Sequential pattern discovery using equivalent classes (SPADE) is an algorithm to mine frequent sequence patterns by splitting problems into subproblems. Each subproblem is solved independently and then conquered altogether with the combining operation. General steps of SPADE algorithm [11]:

- 1. Calculate the support value of every item.
- 2. Calculate the support value of every k-frequent sequence items, with k, is the maximum length of the sequence which its support exceeds the minimum support.
- 3. Disintegrate a class of all sequences based on its length.
- 4. List all sequences to generate a new sequence.

Sequence pattern is a list that outlines the sequence of items in a dataset which appearance satisfies the minimum support. Hotspots sequence patterns were attained from the previous data transformation step.

2.4. Confidence Analysis

The analysis focuses on the confidence aspect with a purpose to determine its relation with hotspot sequences. Moreover, the hotspot appearance time and other variables which are expected to be related to confidence are also included. In total, there are 6 out of 12 attributes selected: longitude, latitude, brightness_temperature, acq_date, acq_time, and confidence.

Confidence is an attribute representing the hotspot quality [1]. The confidence value ranges from 0 (low) to 100 (high). A heuristic measure of the confidence (C) of each detected fire pixel is determined as the geometric mean of up to five sub-confidence parameters, designated C₁ to C₅. The five parameters represent in terms of brightness temperature (T_4), number of adjacent cloud pixels (N_{ac}), number of adjacent water pixels (N_{aw}), the standardized variables $z_4 = T_4 + \overline{T}_4 / \delta_4$ and $z_{\Delta T} = \Delta T + \overline{\Delta T} / \delta_{\Delta T}$, as well as the ramp function $S(x; \alpha, \beta)$, defined as [1]:

$$S = \begin{cases} 0; & x \leq \alpha \\ & \frac{x \cdot \alpha}{\beta \cdot \alpha}; & \alpha \leq x \leq \beta \\ 1; & x \geq \beta \end{cases}$$

 ΔT is the difference between T₄ and T₁₁ (bright_t31 attribute). Confidence level of a hotspot is a aggregate value of five sub-confidence (C₁ to C₅), each value normalized to be ranged from 0 (low) to 1 (high). At daytime, the confidence of hotspot depends on each sub-confidence value defined as: (1) $C_1 = S(T_4; T^*_4, 360 \text{ K}), (2) C_2 = S(z_4; 3.0, 6), (3) C_3 = S(z_{\Delta T}; 3.5, 6), (4) C_4 = 1-S(N_{ac}; 0, 4), and (5) C_5 = 1-S(N_{aw}; 0, 4)$ [1].

C is described as the geometric mean from all sub-confidence, mathematically defined as $C = \sqrt[5]{C_1 C_2 C_3 C_4 C_5}$. The geometric mean is used since it corresponds to the characteristics that are consistent with differences in the scale of variables compared to arithmetic mean [19]. For nighttime, the threshold in the definition of C1 is reduced from 360 K to 320 K. Moreover, C4 and C5 are also not included in the C calculation. The C₁ differs between daytime and nighttime since the threshold of T₄ also differs. A spot is considered a hotspot if $T_4 \ge 360$ K for daytime, whereas for nighttime will decrease to $T_4 \ge 320$ K. For daytime fire pixels detected over water, C5 is likewise omitted from the calculation since the adjacent water pixels only provides information about the fact that the fire pixel itself lies over water [1].

2.5. Verification with Satellite Image

Verification was conducted by overlaying hotspot sequences from the analysis step with Landsat-8 imagery. The overlay process is carried out to see the suitability of the hotspot location from analysis step with the hotspot captured by Landsat-8. To perform verification with Landsat image, a composite image must be created. Band combination with different spectral resolution but the same spatial is called composite image. The combination of particular bands causes the image to have different information [20]. Bands used in this research were 7 (Short Wave Infrared), 5 (Near Infrared), and 4 (Red) that suited for fire detection in peatland area. Band 7 is also very sensitive to radiation emission so it can detect heat source [21].

3. Results and Discussion

Each dataset is processed using the SPADE algorithm, which generated several sequence patterns. Only attractive patterns are chosen for further in-depth analysis.

3.1. Generate and Select Hotspot Sequence Patterns

Sequence patterns were generated using the cspade function available under R within the arulesSequence package. This research aims to generate minimum 2-frequent sequence (or more if exist) for all datasets, in which leads minimum support to be set as 0.02%. It specifies that only patterns with a minimum of two hotspot appearances (2-frequent sequence) were selected. A sequence with two items is named a 2-frequent sequence, three items as a 3-frequent sequence, and so on. Figure 2 shows a snippet of the output generated using SPADE. The number of generated sequences for each dataset is provided in Table 3.

134	<{1448582400}>	0.0002213089
135	<{1448755200}>	0.0002529244
136	<{1448928000}>	0.0009800822
137	<{1449532800}>	0.0003161555
138	<{1450915200}>	0.0004742333
139	<{1451260800}>	0.0002529244
140	<{1444780800},{1444953600}>	0.0006006955
141	<{1444348800},{1444780800}>	0.0002529244
142	<{1442793600},{1442966400}>	0.0003161555
143	<{1442188800},{1442620800}>	0.0002213089
144	<{1442016000},{1442188800}>	0.0002845400
	Figure 2. Snippet of SPADE	E Output

Table 3. Number of frequent sequences generated by SPADE

Dataset	1-frequent sequence	2-frequent sequence			
Kalimantan 2014	286	4			
Kalimantan 2015	139	6			
Sumatra 2014	241	6			
Sumatra 2015	226	12			

Selected patterns were stored in a CSV file to identify the attribute's value. As presented on Figure 2, there is sequence <{1442793600},{1442966400}> with support value of 0.0003161555. EID 1442793600 and 1442966400 sequentially refers to September 21st, 2015 and September 23rd, 2015. The <{1442793600},{1442966400}> sequence implies that on September 21st, 2015, several hotspots are found on some locations that also occurs at the exact location on September 23rd, 2015.

Interesting patterns in this study are the 2-frequent sequences which indicate the continuous appearance of hotspot at the same location. As presented in Table 3, there are 28 interesting 2-frequent sequence patterns from all datasets. From a further identification process to obtain the hotspots data within the sequence, a total of 484 hotspots were found.

3.2. Confidence Analysis

When a hotspot is part of a sequence, it is supposed to be a strong indicator of fire occurrence. The threshold of confidence to be acknowledged as a strong fire indicator is 70%. However, some

hotspots are considered as "outliers", where the confidence level decrease significantly far less than 70% in adjacent time intervals. To analyze the confidence level alteration, this work only selects hotspots within the sequence that happen to have a decrement in confidence level. The number of selected hotspots from 2-frequent sequence at a minimum support of 0.02% for each dataset is given in Table 4. Minimum support is a user-specified threshold on mining the frequent sequences. A sequence is called a frequent sequence if it appears more than the minimum support times in the dataset. There are many hotspots within the sequences where confidence level was less than 70% that occurs at nighttime. Table 5 presents the information of mean, maximum, and minimum value of

confidence and T4 at daytime and nighttime.

Dataset	Time	С	onf < 70%	$Conf \ge 70\%$	Number of hotspots
Kalimantan 2014	06.00	_	10	12	
	18.00				17
	18.00	_	9	11	42
	06.00				
Kalimantan 2015	06.00	_	11	13	
	18.00				120
	18.00	_	35	61	120
	06.00				
Sumatra 2014	06.00	_	11	23	
	18.00				114
	18.00	_	23	57	114
	06.00				
Sumatra 2015	06.00	_	7	19	
	18.00				200
	18.00	_	30	152	208
	06.00				

Table 4. Number of selected hotspots with a decrement of confidence (conf)

Table 5. Statistics of Confidence and T_4							
		Mean Maximum Minimum					
Dataset	Time	Conf	T_4	Conf	T_4	Conf	T_4
		(%)	(K)	(%)	(K)	(%)	(K)
Kalimantan 2014	06.00 - 18.00	69.54	335.38	100	364.80	21	311.10
	18.00 - 06.00	68.80	317.66	100	371.10	13	305.20
Kalimantan 2015	06.00 - 18.00	71.29	336.28	100	364.70	27	317.10
	18.00 - 06.00	73.23	317.23	100	366.50	8	302.30
Sumatra 2014	06.00 - 18.00	73.91	338.71	100	384.50	19	315.70
	18.00 - 06.00	80.64	329.04	100	392.30	8	305.30
Sumatra 2015	06.00 - 18.00	77.93	349.90	100	430.30	25	313.20
	18.00 - 06.00	88.97	343.41	100	474.80	13	301.70

Appertaining to Table 5, some hotspots have significantly low confidence and T4 included in the sequence patterns. From 484 hotspots found from the interesting sequences, 58 hotspots are expected to have a significant decrease of confidence level. An example of what refers to a decrease in confidence level is illustrated in Table 6.

Table 6. Example of Confidence Decrement					
Longitude	Latitude	Acq_date	Confidence	Brightness (T4)	
114.504	-2.338	14-Oct-2015	77	323.7	
114.504	-2.338	15-Oct-2015	8	303	
114.504	-2.338	16-Oct-2015	74	313.8	

ICASMI 2020		IOP Publishing
Journal of Physics: Conference Series	1751 (2021) 012035	doi:10.1088/1742-6596/1751/1/012035

As mentioned in [15], a hotspot is a strong indicator of fire if the confidence level reaches a minimum of 70%. However, as we trace the data, some of these hotspots previously have high confidence and T4, but then decrease significantly in adjacent time intervals. As referring to [1], several factors are affecting MODIS accuracies, such as cloud or haze cover, trees, and the location or angle of the satellite when detecting hotspots. Therefore, it can be assumed that the significant alteration of confidence level in the same location is likely caused by one of the factors above.

3.3. Verification with Satellite Image

To discover the causative factors of confidence level decrement, a verification process is conducted. Satellite images are needed in this process to see the environmental condition that may affect the selected hotspots. Landsat-8 images collected from LAPAN website were used, with path/row and date of images were based on the selected hotspots information. Table 7 presents the information of images needed in this work.

Table 7.1 Selected Landsat-8 Images ^a				
Dataset	Path/Row	Date		
Kalimantan 2014	119/61	10-October-2014		
	121/61	8-October-2014		
Kalimantan 2015	118/62	20-September-2015		
	119/62	11-September-2015, 13-October-2015		
Sumatra 2014	126/59	17-March-2014		
	127/59	20-February-2014		
Sumatra 2015	123/62	23-September-2015, 25-October-2015		
	123/63	25-October-2015		
	125/61	5-September-2015, 23-October-2015		
$N_{abc} = \frac{a}{c} \mathbf{C}_{abc} = \frac{b}{c} \frac{b}{c} \frac{d}{dc} = \frac{b}{c} \frac{d}{dc} \frac{d}{dc$				

Note: ^aSource: https://landsat-catalog.lapan.go.id/

Each Landsat-8 image was processed in QuantumGIS to form a composite image. Composite image (or also known as the multispectral image) was saved as .tiff format consisting of three layers representing three selected bands: band 7, 5, and 4. Figure 3 shows an example of a composite image. The composite images were then overlaid with the selected hotspots (hotspots from the confidence analysis step), and there are three possible cases, as shown in Figure 4.



Figure 3. Composite Image (band 7-5-4) of Landsat-8



Figure 4. Three possible cases found from the overlaying process

There are 58 out of 484 hotspots found from 28 sequence patterns with a decrement of confidence level, but only 21 hotspots can be verified. Landsat-8 satellite only comes across the same area once every 16 days, resulting in limited available images for a certain date. Thus, this work only verifies hotspots that have Landsat-8 imagery within the range of 3 days. Figure 4 shows that hotspots locations are not always within the haze. The angle of imagery taken by satellite varies was expected to be the main cause. Figure 4 (a) shows there are clouds and (or) haze above the hotspot, Figure 4 (b) shows there are no haze or cloud, and Figure 4 (c) shows there are cloud and (or) haze near the hotspot.

If the first case is found, it can be determined that the cover of haze or cloud was the cause of confidence decrement. Cloud or haze cover will affect the brightness temperature captured by the satellite decrease significantly, which also impacts the attributes building up the confidence level. If the second case is found, it cannot be determined whether cloud or haze affects the decrement. Other factors such as the angle of imagery taken or cover of trees are expected to be the cause. Last, if the third case is found, then further analysis, such as distance measurement, is needed.

One pixel of the MODIS image represents a hotspot within a radius of 1.113 km. If there is more than one hotspot within the range of 1.113 km, it will be detected as a single hotspot [22]. This implies that a hotspot can cause haze within the radius of 1.113 km. In other words, if there is haze detected more than 1.113 km from a hotspot, it is uncertain whether the hotspot caused it or not. Figure 5 shows the distance measurement between a hotspot and the nearest haze.



Figure 5. Distance measurement between a hotspot and nearest haze

From 21 hotspots to be verified, 6 hotspots are covered with haze, 12 hotspots are within the radius of 1.113 km, and the other 3 does not fill both. The result shows that 18 out of 21 hotspots which confidence decrease, or about 85.71%, are most likely caused by haze cover. It can be concluded that most likely at the location of the hotspots, a fire exists, even though confidence

was low. In addition, three unknown hotspots cause a decrease in the confidence value. Other possible factors that may be present are tree cover. Verification of these factors cannot be done since Landsat-8 imagery does not provide information related to tree height.

Moreover, [23] stated that if a fire exists in an area, there is a small probability for trees to cause the decrement of confidence level since the tree might have been burnt and died. The cause of the alleged decrease in confidence is indeed cloud cover or haze. Verification results do not indicate the presence of haze nor clouds, possibly because there are weather factors such as winds that move the haze and clouds, which is not included in this study. Besides, the satellite image used in this work is Landsat-8, and hotspot datasets were collected from Satellite Aqua and Tera owned by FIRMS NASA. The date and time of hotspots and image data are not precisely the same as the satellite differs. When Satellite Aqua and Tera collects data and detects a hotspot, there is a high possibility that haze or cloud have moved and covered the area as Landsat-8 satellite come across.

4. Conclusion

Confidence determines the quality of hotspot, where a hotspot with a value above 70% refers to a fire spot. This study found 484 hotspots from the sequence patterns from the SPADE algorithm, where 58 of them fall on a significant decrease in confidence. From 58 hotspots, only 21 hotspots are eligible to be verified by Landsat-8 satellite image. Verification results show that 85.71% of hotspots have decreased confidence due to haze cover. It can be concluded that even though the confidence level of a hotspot is low, a fire still can exist. A high confidence level of a hotspot does imply it to be a fire spot; however, it does not consistently apply vice versa. This study shows that a fire exists even though the confidence level is low due to cover by haze.

5. References

- [1] Giglio L, Justice CO, Schroeder W. 2016. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sensing of Environment*. 178:31-41.
- [2] Fayyad U, Piatetsky-Shapiro G, Smyth P. 1996. *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- [3] Hand D, Mannila H, Smyth P. 2001. Principles of Data Mining. MIT Press, Cambridge, MA.
- [4] Stojanova D, Panov P, Kobler A, Dzeroski S, Taskova K. 2006. *Learning to Predict Forest Fires with Different Data Mining Techniques*. In Mladenic D and Grobelnik M, editors, 9th International multiconference Information Society (IS 2006), Ljubljana, Slovenia.
- [5] Usman M, Sitanggang IS. 2015. Hotspot distribution analyses based on peat characteristics using density-based spatial clustering. *Procedia Environmental Sciences*. 24, pp. 132-140.
- [6] Nurpratami ID, Sitanggang IS. 2015. Classification rules for hotspot occurrences using spatial entropy-based decision tree algorithm. *Procedia Environmental Sciences*. pp. 120-126.
- [7] Sitanggang IS, Shofiana DA, Sihombing BSK. 2018. Hotspot Sequence Patterns with an Improvement in Spatial Feature. *International Journal of Engineering and Manufacturing* (*IJEM*). Vol.8, 6:13-25. DOI: 10.5815/ijem.2018.06.02.) chapter 6 pp 173–214
- [8] Pramatasari A, Permatasari NKF, Pramudiyasari T, Manessa MDM, Supriatna. 2020. Spatial Distribution Patterns Analysis of Hotspot in Central Kalimantan using FIMRS MODIS Data. *JGLITrop*, Vol. 4, No. 1, pp. 24-34.
- [9] Albar I, Jaya INS, Saharjo BH, Kuncahyo B, Vadrevu KP. 2018. Spatio-Temporal Analysis of Land and Forest Fires in Indonesia Using MODIS Active Fire Dataset BT – Land Atmospheric Research Applications in South and Southeast Asia. In KP Vadrevu, T Ohara, & C Justice (Eds.) (pp. 105–127). Cham: Springer International Publishing.
- [10] Davies D, Ederer G, Olsina O, Wong M, Cechini M, Boller R. 2019. NASA's Fire Information for Resource Management System (FIRMS): Near Real-Time Global Fire Monitoring using Data from MODIS and VIIRS. Retrieved from https://firms.modaps.eosdis.nasa.gov/
- [11] Zaki MJ. 2001. SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning*. 42:31-60.

- [12] Nurulhaq NZ, Sitanggang IS. 2015. Sequential pattern mining on hotspot data in Riau Province using the PrefixSpan algorithm. In: *International Conference on Adaptive and Intelligent Agro-industry (ICAIA) 2015*. 3:257-260.
- [13] Agustina T, Sitanggang IS. 2015. Sequential patterns for hotspot occurences based weather data using Clospan algorithm. In: *International Conference on Adaptive and Intellegent Agroindustry (ICAIA) 2015*. 3:301-305.
- [14] Sitanggang IS, Istiqomah N, Syaufina L. 2018. Fire Spot Identification Based on Hotspot Sequential Pattern and Burned Area. *BIOTROPIA*. Vol 25, 3:155-163. DOI: 10.11598/brb.2018.25.3.676.
- [15] Abriantini G, Sitanggang IS, Trisminingsih R. 2017. Hotspot sequential pattern visualization in peatland of Sumatera and Kalimantan using Shiny framework. In: *IOP Conference Series 2017: Earth and Environmental Science*. 54:1-7.
- [16] Verma M, Mehta D. 2014. Sequential pattern mining: A comparison between GSP, SPADE, and PrefixSpan. *International Journal of Engineering Development and Research (IJEDR)*. 3016-3036.
- [17] Han J, Kamber M, Pei J. 2012. *Data Mining: Concepts and Techniques, 3rd ed.* Morgan Kaufmann: US.
- [18] Zhao Q, Bhowmick SS. 2003. Sequential pattern mining: A survey. *ITechnical Report CAIS Nanyang Technological University Singapore*. 1-26.
- [19] Manikandan S. 2011. Measures of central tendency: The mean. J Pharmacol Pharmacother. 2:140-142
- [20] Melnikova. 2014. *Landsat Image Processing and Classification in IDRISI SELVA*. Kyrgyz State University of Construction: Rusia, RU.
- [21] Elvidge CD. 2014. *Indonesia Peatland Fire Detection with Nighttime Landsat*. NOAA-NESDIS Nation Geophysical Data Center: Colorado, US.
- [22] Giglio L, Descloitres J, Justice CO, Kaufman Y. 2003. An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment*. 87:273-282.
- [23] Syaufina L. 2008. Kebakaran Hutan dan Lahan di Indonesia: Perilaku Api, Penyebab, dan Dampak Kebakaran. [Forest and Land Fires in Indonesia: Fire Behavior, Causes and Impact of Fire]. Bayumedia Publishing: Malang, ID.