



(RESEARCH ARTICLE)



## Classification of forest fire areas using machine learning algorithm

Saruni Dwiasnati <sup>1,\*</sup> and Yudo Devianto <sup>2</sup>

*Faculty of Computer Science, Universitas Mercu Buana, Indonesia.*

World Journal of Advanced Engineering Technology and Sciences, 2021, 03(01), 008-015

Publication history: Received on 19 May 2021; revised on 02 July 2021; accepted on 05 July 2021

Article DOI: <https://doi.org/10.30574/wjaets.2021.3.1.0048>

### Abstract

Forest fires that occur will cause various kinds of problems, both in terms of health, such as smoke that can interfere with the respiratory system, in terms of the economy such as the economic wheel cannot run as usual, in terms of the environment can damage the surrounding environment and the environment that is missed by smoke, and other disasters. Forest fires can also have an impact on the costs that will be incurred to resolve the problems that arise due to forest fires, so research is needed to find out and measure the area affected by forest fires that burned in the range of 1980 - 2019 using a dataset of approximately 10,000. The target in this research is to be able to generate the best percentage scenario and find out the model of using the algorithm used to explore the algorithm in the Machine Learning method for the model for estimating the area of forest fires, namely the Siak Kampar Peninsula in Riau Province. In this study, 7 parameters were used to create a forest and land fire hazard map, namely weather temperature, Burned Area Density, hotspot density, wind speed, land cover type, rainfall, and land use. The seven parameters will be searched for accuracy results using the Classification method with Machine Learning algorithms, including Naïve Bayes, SVM, and K-Nearest Neighbor (K-NN). In this study, comparisons were made to obtain the best algorithm for estimating forest fire areas. By generating each algorithm is 71.72% for the Naïve Bayes algorithm, 75.00% for the SVM algorithm, and 64.71% for the K-NN algorithm.

**Keywords:** Forest Fire; Classification; Machine Learning; Riau

### 1. Introduction

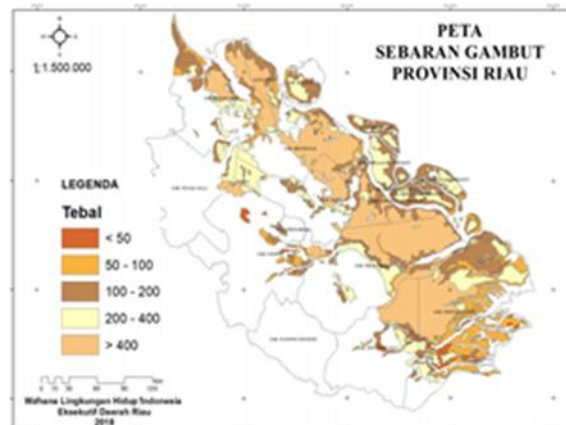
Ecological disasters from year to year always occur in Indonesia. Floods, landslides, droughts, and smog from forest fires and peat swamps lurk and fill the days of residents across the archipelago. As if it has never learned a lesson from every ecological, social and cultural crisis, the country still persists with the paradigm of economic growth based on extractive industries and high-risk development, which threatens the safety and productivity of the people, and the ecological balance. Natural wealth continues to be exploited, exploited without ever calculating the costs of externalities that are forced to be incurred by residents.

Forest and land fires that occur every year in Indonesia are motivated by the increasing need for land use for agriculture and settlements. Forest and land fires are not a new phenomenon in tropical rain forests [1]. Forest and land fires that are becoming increasingly common are very disturbing because they produce emissions, especially in Indonesia as a tropical country. Forest and land fires can occur based on 2 main factors, namely natural factors and human-made factors, natural factors can be in the form of a long dry season so that the plants become dry and human activities in the form of illegal burning to expand the land [2]. The definition of forest and land fires in general according to BNPB is a situation where a building in a place such as a house or settlement, factory, market, building, etc. is hit by a fire that causes casualties and/or losses in either the economic sector or the environmental sector [3]. Based on the definition

\* Corresponding author: Saruni Dwiasnati  
Faculty of Computer Science, Universitas Mercu Buana, Indonesia.

of fire, fire causes losses that can endanger life, property or the environment inhabited by living things. The location and time of fire is difficult to predict because the forest in an area in Indonesia is classified as extensive.

According to the Wahana Lingkungan Hidup (Walhi) stated that Riau is one of the provinces in Indonesia with quite complex environmental problems, especially the problem of the rate of destruction of forests and peat ecosystems whose impacts are listed as the five biggest contributors to disasters that occurred in Indonesia. Riau Province is one of the many provinces in Indonesia that experience inequality in access and land tenure, especially on peatlands. It is known that the area of peatlands in Riau Province is approximately 50% of the total area of Riau Province which is spread over almost all districts. So that from the total area of Riau Province of  $\pm$  9 million hectares, more than 4 million hectares are peat with varying depths. Invasion and industrial expansion in Riau resulted in a decline in peat quality, which also contributed to climate change [4].



**Figure 1** Map of peat distribution by depth in Riau province

Inequality that occurs due to ecological disasters not only refers to ecological disasters such as forest and peatland fires that cause smog and large carbon releases into the atmosphere, but is also a factor that causes economic decline for the people of Riau as well as social and cultural changes along with that matter. For example, the construction of large canals carried out by the company in the context of activities to support its business causes the peat to become dry and prone to fire. This is certainly contradictory to the peat management that has been carried out for a long time by the community, without knowing the consequences of peat drought, peatland still has economic value for the community by planting plants suitable for peat land, such as coconut.

According to TFCA (National Disaster Management Agency, 2019) Sumatra, the Siak Kampar Peninsula, which is located in Riau Province, still has more than 400,000 hectares of peat swamp forest cover. The peat swamp forest area of the Siak Kampar Peninsula is one of the largest remaining expanses of peat swamp forest in Sumatra. This area is a habitat for the Sumatran Tiger and several endangered species. The peat swamp forest area in Riau contains a potential of around 16.9 billion tons of carbon. While specifically in the Kampar Peninsula area, it has a potential of around 7 billion tons of carbon. Types of Habitat found on the Siak Kampar Peninsula:

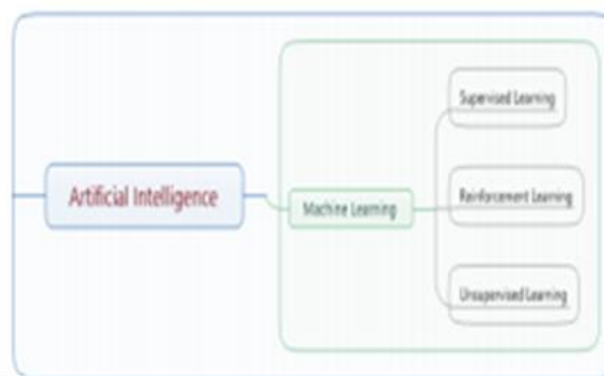
- Lake
- Public/other lakes: shoreline, forest
- River area: forest
- Downstream river area: continuous water
- Swamp without Peat: forest
- Peat Swamp.

Forest/land fires in the Siak Kampar peninsula occur almost every year since massive land clearing activities took place. Based on data on hotspots or hotspots found in 2019-2020 Forest and land fires (Karhutla) still occur in Sumatra. Today, it was observed by satellite that there were 358 hotspots distributed and Riau, which contributed the most to the hotspots, which resulted in the air in Pekanbaru being surrounded by smoke. Morning visibility 5 Km. The same condition also occurred in the districts of Indragiri Hulu, Dumai, and Pelalawan, which were surrounded by smoke. Based on information from the Meteorology, Climatology and Geophysics Agency (BMKG) Pekanbaru Station, the number of hotspots is at a 50% confidence level. Based on the Terra/Aqua satellite at 06.00 WIB, out of 584 hotspots, Riau has the highest number compared to other provinces with 150 locations (<https://news.detik.com/berita>). Based

on the monitoring of the National Oceanic Atmospheric Administration (NOAA) satellite, there are 18 distribution points in the province in community areas, industrial plantation forest areas, oil palm areas, protected forests and peatlands. Fires are supported by external factors such as global warming, long-term drought providing ideal conditions for forest and land fires to occur [5]. If we look more closely, the government's policy in the 1980s to open forest concessions, convert natural forests into plantations, transmigration, irrigation development, and agricultural expansion is thought to have increased the area of forest fires. National policies that encourage land use change increase forest fires [6] [7]. Determining the level of vulnerability of the fire area is one way to prevent forest and land fires.

Machine learning can be defined as the application of computers and mathematical algorithms adopted by means of learning that comes from data and produces predictions in the future [8]. The learning process in question is an attempt to acquire intelligence through two stages, including training and testing [9]. The field of machine learning is concerned with the question of how to build computer programs to improve automatically based on experience [10]. Recent research reveals that machine learning is divided into three categories: Supervised Learning, Unsupervised Learning, Reinforcement Learning [11].

The linkage scheme of artificial intelligence and machine learning can be explained in Figure 2



**Figure 2** Schematic of Artificial Intelligence and Machine Learning

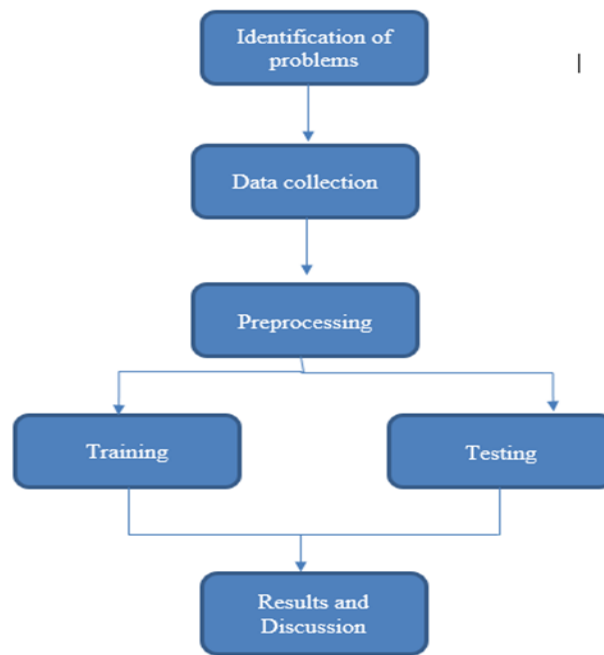
In general, the parameters used to conduct research related to exploration algorithms in determining forest fire areas consist of 7 parameters, namely weather temperature, Burned Area Density, hotspot density, wind speed, land cover type, rainfall, and land use.

In research conducted by [12] revealed that the research conducted by comparison or comparison to get the best algorithm in estimating forest fire areas, the algorithms used were LR (Linear Regression), K-NN (K -Nearest Neighbor) and SVM (Support Vector Machine).

This study describes the exploration of algorithms contained in machine learning to determine the accuracy results and compare which accuracy is the best accuracy result from one algorithm to another. The method used in this study uses Machine Learning with the algorithms used include Naïve Bayes, SVM, and K-Nearest Neighbor (K-NN). The weighted score is calculated from the average score from semesters 1 to 5. In this study, we use K-Nearest Neighbor as an algorithm to classify the data. Finally, the prediction application that we propose will recommend a list of universities for users to register for SNMPTN admission which is useful for schools to predict the acceptance of their students in State universities [13].

## 2. Research Methods

### 2.1. Research Stages



**Figure 3** Research Stages

#### 2.1.1. Problem Identification

Forest fires are increasingly widespread in areas that have lush forests and lands with various ecosystems in them. Sumatra and Kalimantan are areas that have extensive forests and land. Therefore, information is needed to find out which areas are prone to forest fires in Riau Province by comparing the 3 algorithms in the Classification Method with Machine Learning Techniques.

#### 2.1.2. Data Collection

To categorize data on forest fires, various types of data are needed that can be used in data processing to produce a model that is used as a reference to the tools used.

#### 2.1.3. Preprocessing

Documents generally have an arbitrary or unstructured structure. Therefore, we need a process that can change the form of previously unstructured data into structured data. The preprocessing stage has several processes, namely Data Representation, Punctuation Removal, Case Folding, Stemming, Stopwords Removing, and Tokenizing. Furthermore, the data that has undergone preprocessing will be converted into numerical form with the Term Weighting stage. In this study, there are three methods of term weighting used, namely term frequency, inverse document frequency, and term frequency-inverse document frequency. The first step is to combine all the data that has been obtained from the WEB from various sources, then divide the data into Data Training or Data Testing.

#### 2.1.4. Results

The results of this study are models of various algorithms used by using Rapidminer tools. In the Rapidminer tools it produces Accuracy Value, Performance Vector.

## 2.2. Types of research

This study uses a qualitative approach that seeks to study and describe the area of forest fires in the Siak Kampar peninsula in Riau Province. A qualitative approach is also used to describe the Weather Temperature around the forest, Hotspot Density (Spots of Fire), Wind Speed, Rainfall, Burned Area Density, Land Use, and Type of Land Cover. The author performs data processing on the WEB which can be used to obtain the data needed in this study. The parameters

that the author uses to support the analysis of the classification of forest fire areas in Riau Province. The analysis used in this study is classification analysis with several algorithms contained in Machine Learning Techniques, such as Naïve Bayes, SVM, and K-NN using Rapidminer Tools to be able to generate models for each algorithm used.

### 3. Results and discussion

Machine learning is a branch or application of artificial intelligence. This branch of science focuses on creating systems or algorithms that continually learn from data and improve their accuracy over time without any particular programming. In machine learning applications, algorithms or sequences of statistical processes are trained to find certain patterns and features in large amounts of data. It aims to make a decision or prediction based on these data. An example of the use of machine learning is a digital assistant that we can use on a smartphone to execute a command. In addition, machine learning applications can also be felt when advertisements on the internet recommend products that match our interests.

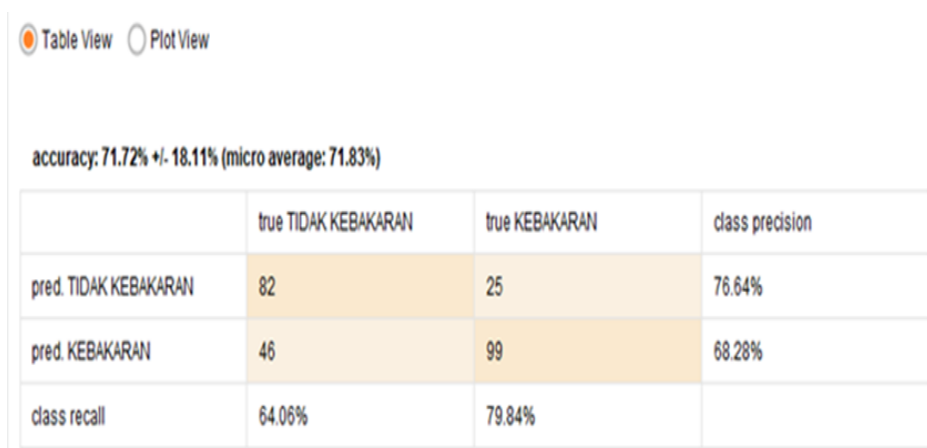
In the classification, there are categorical target variables. For example, the classification of income can be separated into three categories, namely high income, medium income, low income. Other examples of classification in business and research are:

- Determine whether a credit card transaction is a fraudulent transaction or not.
- Estimating whether a mortgage application by a customer is a good or bad credit.
- Diagnosing a patient's illness to find out what category the disease belongs to.

The data used in this study uses secondary data from the website of the National Disaster Management Agency (BNPB) with the web address <http://www.bnpb.go.id> and the data used is data on forest fire areas, namely the Semenanjung Siak Kampar in Riau Province in 2015. 1994 – 2001. The raw data obtained from the website is around 1000s for all the variables used, but only 300 are used to be processed to produce accuracy values. This study aims to determine which accuracy value is higher, between the Naïve Bayes algorithm, SVM, and K-NN.

#### 3.1. Naïve Bayes

In the results of the above experiment using the Naïve Bayes algorithm using Rapidminer tools with a total of 252 data testing data resulting in an accuracy value of 71.72%.



	true TIDAK KEBAKARAN	true KEBAKARAN	class precision
pred. TIDAK KEBAKARAN	82	25	76.64%
pred. KEBAKARAN	46	99	68.28%
class recall	64.06%	79.84%	

**Figure 4** Accuracy of Naïve Bayes Algorithm

Figure 5 is the result of performance vector data testing using the Naïve Bayes algorithm which consists of 252 data records, 82 data are classified as areas that are not fire potential but in fact the area is in the category of no fire, 46 data is classified as fire but is included in the fire category, 25 the data is classified as Fire but in reality it is included in the No Fire category and 99 data classified as Fire is in fact actually included in the Fire area category.

```

PerformanceVector:
accuracy: 71.72% +/- 18.11% (micro average: 71.83%)
ConfusionMatrix:
True:  TIDAK KEBAKARAN KEBAKARAN
TIDAK KEBAKARAN:    82    25
KEBAKARAN:    46    99
precision: 68.28% (positive class: KEBAKARAN)
ConfusionMatrix:
True:  TIDAK KEBAKARAN KEBAKARAN
TIDAK KEBAKARAN:    82    25
KEBAKARAN:    46    99
recall: 79.84% (positive class: KEBAKARAN)
ConfusionMatrix:
True:  TIDAK KEBAKARAN KEBAKARAN
TIDAK KEBAKARAN:    82    25
KEBAKARAN:    46    99
AUC (optimistic): 0.578 +/- 0.322 (micro average: 0.578) (positive class: KEBAKARAN)
AUC: 0.469 +/- 0.266 (micro average: 0.469) (positive class: KEBAKARAN)
AUC (pessimistic): 0.359 +/- 0.228 (micro average: 0.359) (positive class: KEBAKARAN)
    
```

**Figure 5** Performance Vector Algorithm Naïve Bayes

accuracy: 75.00%

	true TIDAK KEBAKARAN	true KEBAKARAN	class precision
pred. TIDAK KEBAKARAN	76	11	87.36%
pred. KEBAKARAN	52	76 113	68.48%
class recall	59.38%	91.13%	

**Figure 6** Accuracy of SVM algorithm algoritma

In the results of the above experiment using the Support Vector Machine algorithm using Rapidminer tools with the amount of testing data as much as 252 data resulting in an accuracy value of 75.00%.

### PerformanceVector

```

PerformanceVector:
accuracy: 75.00%
ConfusionMatrix:
True:  TIDAK KEBAKARAN KEBAKARAN
TIDAK KEBAKARAN:    76    11
KEBAKARAN:    52    113
precision: 68.48% (positive class: KEBAKARAN)
ConfusionMatrix:
True:  TIDAK KEBAKARAN KEBAKARAN
TIDAK KEBAKARAN:    76    11
KEBAKARAN:    52    113
recall: 91.13% (positive class: KEBAKARAN)
ConfusionMatrix:
True:  TIDAK KEBAKARAN KEBAKARAN
TIDAK KEBAKARAN:    76    11
KEBAKARAN:    52    113
AUC (optimistic): 0.822 (positive class: KEBAKARAN)
AUC: 0.810 (positive class: KEBAKARAN)
AUC (pessimistic): 0.798 (positive class: KEBAKARAN)
    
```

**Figure 7** Performance Vector SVM algorithm



Figure 7 is the result of performance vector data testing using the SVM algorithm which consists of 252 data records, 76 data are classified as non-potential fires, but in fact the area is in the No Fire category, 52 data is classified as No Fire but is in the Fire category, 11 data classified as Fire but in reality it is in the No Fire category and 113 data classified as Fire in reality actually fall into the category of Fire area.

#### 5.4 K-NN

accuracy: 64.71%

	true TIDAK KEBAKARAN	true KEBAKARAN	class precision
pred. TIDAK KEBAKARAN	12	4	75.00%
pred. KEBAKARAN	14	21	60.00%
class recall	46.15%	84.00%	

**Figure 8** Accuracy of K-NN. Algorithm

In the results of the above experiment using the K-Nearest Neighbor algorithm using Rapidminer tools with a total of 252 data testing data resulting in an accuracy value of 64.71%.

## PerformanceVector

```

PerformanceVector:
accuracy: 64.71%
ConfusionMatrix:
True:   TIDAK KEBAKARAN KEBAKARAN
TIDAK KEBAKARAN:      12      4
KEBAKARAN:           14      21
    
```

**Figure 9** Performance Vector K-NN. algorithm

Figure 9 is the result of performance vector data testing using the K-NN algorithm which consists of 252 data records, 12 data are classified as areas with no potential fires but in fact the area is included in the No Fire category, 14 data is classified as No Fire but is included in the Fire category, 4 data classified as Fire but in reality they are in the No Fire category and 21 data classified as Fire in reality actually fall into the Fire area category.

## 4. Conclusion

Based on the discussion of the research results that have been discussed in the previous chapter, in the study of Classification of Forest Fire Areas using the Machine Learning algorithm, the following conclusions can be drawn:

- This study was successful in classifying forest fire areas in the Siak Kampar Peninsula district from 1994 to 2001 using the Naïve Bayes, SVM, and K-NN algorithms.
- The attributes used for classification consist of weather temperature, wind speed, type of land cover, rainfall, and land use which produces the highest accuracy value is 75.00% for the SVM algorithm.
- The classification method using the Naïve Bayes algorithm, SVM, and K-NN can be used to predict the Forest Fire Area so that it can find out information about the forest fire area and can be used to measure the size of the forest fire area.

## Compliance with ethical standards

### *Acknowledgments*

Praise and gratitude to God Almighty for all His graces and gifts so that this internal research can be completed. We do not forget to thank the big family of Mercu Buana University, the Faculty of Computer Science, and in particular the PUSLIT of Mercu Buana University for providing research funding support. The theme chosen in this internal research activity is "Classification of Forest Fire Areas using Machine Learning Algorithms". The author realizes that this report is still far from perfect, therefore constructive criticism and suggestions are highly expected better. Hopefully this report can be useful for readers. Thank you.

### *Disclosure of conflict of interest*

All authors declare that they have no conflict of interest.

---

## References

- [1] Hoscilo A, Page SE, Tansey KJ, Rieley JO. Effect of repeated fires on land-cover change on peatland in southern Central Kalimantan, Indonesia, from 1973 – 2005. *International Journal of Wildland Fire*. 2011; 578-588.
- [2] Rasyid F. Permasalahan dan Dampak Kebakaran Hutan. *Jurnal Lingkar Widyaiswara*. 2014; 1(4): 47-59.
- [3] <http://tfcasumatera.org/bentang-alam-semenanjung-kampar-kerumutan-senepis/>.
- [4] <https://www.walhi.or.id/wp-content/uploads/2019/01/Refleksi-2018-Harapan-2019-Walhi-Riau.pdf>.
- [5] Syarif A, Fitria. Increasing the Role as Well As Communities in The Construction of Forest Fire Through the Strengthening of Individual Institutions in Muaro Jambi District. *Jurnal Karya Abdi Masyarakat*. 2019; 3(2): 204-210.
- [6] Sastry N. Forest fires, air pollution, and mortality in Southeast Asia. *Demography*. 2002; 39(1): 1–23.
- [7] Stole F, Chomitz KM, Lambin EF, dan Tomich TP. Land use and vegetation fires in Jambi Province, Sumatera, Indonesia. *Forest Ecology and Management*. 2003; 179: 277–292.
- [8] Goldberg DE, Holland JH. Genetic algorithms and machine learning. *Machine Learning*. 1988; 3(2): 95–99.
- [9] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. *Neurocomputing*. 2006; 70(1–3): 489–501.
- [10] Mitchell TM. *Machine learning*. In McGraw Hill Series in Computer Science. 1997.
- [11] Somvanshi M, Chavan P. A review of machine learning techniques using decision tree and support vector machine. 2016 International Conference on Computing Communication Control and Automation (ICCUBEA). 2016; 1–7.
- [12] Fitriyani dan Rangga Sanjaya. Komparasi Algoritma LR, K-NN dan SVM untuk Estimasi area Kebakaran Hutan. *Jurnal Infotronik*. 2018; 3(2): 103-110.
- [13] Adi Tri Wibowo dan Devi Fitriyah. A K-Nearest Algorithm Based Application to Predict SNMPTN Acceptance for High School Students in Indonesia. *International Research Journal of Computer Science (IRJCS)*. ISSN: 2393-9842. Issue 01. Volume 5. (January 2018).