e-JURNAL PENYELIDIKAN DAN INOVASI
ISLAMIC INNOVATION CENTER FOR TECHNOLOGY, UIS
https://ejpi.uis.edu.my
e-ISSN 2289-7909

VOL. 12 NO. 5, SPECIAL EDITION, (DECEMBER, 2025): page 1-18
DOI: 10.53840/ejpi.v12i5.309

Received: October 27, 2025 Accepted: December 5, 2025 Published: December 31, 2025

Identification of Potential Forest Fires using The Random Forest Method in Kubu Raya Regency

Identifikasi Potensi Kebakaran Hutan Menggunakan Metode Random Forest di Kabupaten Kubu Raya

Muhamad Anang Pratama[1], Erwin Hermawan[2], Sahid Agustian Hudjimartsu[3]

[1,2,3] Informatics Engineering, Faculty of Technology & Science, Universitas Ibn Khaldun Bogor
Jl. Sholeh Iskandar, RT.01/RW.10, Kedungbadak, Kec. Tanah Sereal, Bogor City, West Java 16162 Indonesia.
Email: [1]p.anang41@gmail.com, [2]rwien.gfm@gmail.com, [3]shudjimartsu@uika-bogor.ac.id

ABSTRACT

Forest fires are a serious threat to the environment, health, and economy, so early detection is an important step to minimize their impact. This study aims to identify the potential for forest fires in Kubu Raya Regency by applying the Random Forest algorithm. Various meteorological and geographical variables—such as temperature, air humidity, wind speed, and vegetation density—were used as input parameters, along with fire hotspot data from the MODIS (Moderate Resolution Imaging Spectroradiometer) satellite. The Random Forest model is trained by building many random decision trees and combining their predictions to produce a more stable and accurate estimate of fire potential. The results of the analysis showed that the model built had an accuracy rate of **92%,** with user accuracy value, manufacturer accuracy, and kappa coefficients indicating the consistency of the model's performance in predicting the distribution of potential fires. These findings confirm that the Random Forest method is effective in detecting potential forest fires in the study area. Overall, this study concludes that the Random Forest model can be used as a reliable early detection system to support efforts to prevent, control, and mitigate the risk of forest fires in Kubu Raya Regency. The implementation of this model is expected to help the government and the community in taking quick and appropriate action to minimize the impact of forest fires in the future.

Keywords: *Forest Fires; Potential; Random Forests; Kubu Raya Regency*

ABSTRAK

Kebakaran hutan merupakan ancaman serius bagi lingkungan, kesehatan, dan ekonomi, sehingga deteksi dini menjadi langkah penting untuk meminimalkan dampaknya. Penelitian ini bertujuan untuk mengidentifikasi potensi kebakaran hutan di Kabupaten Kubu Raya dengan menerapkan algoritme Random Forest. Berbagai variabel meteorologi dan geografis—seperti suhu, kelembaban udara, kecepatan angin, serta kepadatan vegetasi—digunakan sebagai parameter input, bersama data hotspot kebakaran dari satelit MODIS (Moderate Resolution Imaging Spectroradiometer). Model Random Forest dilatih dengan membangun banyak pohon keputusan acak dan menggabungkan prediksinya untuk menghasilkan estimasi potensi kebakaran yang lebih stabil dan akurat.Hasil analisis menunjukkan bahwa model yang dibangun memiliki tingkat akurasi sebesar 92%, dengan nilai akurasi pengguna, akurasi produsen, dan koefisien kappa yang menunjukkan konsistensi performa model dalam memprediksi sebaran potensi kebakaran. Temuan ini mengonfirmasi bahwa metode Random Forest efektif dalam mendeteksi potensi kebakaran hutan di wilayah penelitian. Secara keseluruhan, penelitian ini menyimpulkan bahwa model Random Forest dapat digunakan sebagai sistem deteksi dini yang andal untuk mendukung upaya pencegahan, pengendalian, serta mitigasi risiko kebakaran hutan di Kabupaten Kubu Raya. Implementasi model ini diharapkan membantu pemerintah dan masyarakat dalam mengambil tindakan cepat dan tepat guna meminimalkan dampak kebakaran hutan di masa mendatang.

Kata kunci: *Kebakaran hutan; Potensi; Random Forest; Kabupaten Kubu Raya*

INTRODUCTION

Forest fires occur due to two factors, namely natural factors and human factors. Natural factors include long droughts that make plants dry and flammable. Human factors include illegal burning to expand land (Humam et al., 2020). Forest and peatland fires are an annual problem, especially in Indonesia during the dry season. More than 529,266.64 hectares of land were affected by forest and land fires, according to BNPB's 2018 annual disaster report. Hotspots are located in Sumatra, Kalimantan, Riau, and a small part of Central Java and Sulawesi (Bencana et al., 2023).

Most of West Kalimantan is peatland because it is located on the equator. Peatlands are generally flammable because they can store biomass, litter, and mineral soil (Rachman, Saharjo, & Putri, 2020). Village data compiled by the Ministry of Environment and Forestry of the Republic of Indonesia (MoEF) shows that forest fires that occurred in West Kalimantan in 2020 reached an area of 32,000 hectares and were caused by a number of factors, one of which is human behavior. Human behavior includes converting forests and land to be used for residential land, rice fields, plantations, and mining, as well as engaging in activities outside the forest (Simanjuntak, Kusnandar, & Debataraja, 2022).

Every dry season, land fires often occur in Kubu Raya Regency. As of October 2013, there were 349 hot spots in Kubu Raya Regency, according to data from the Forest and Land Fire Management Unit (UPKHL) of the West Kalimantan Provincial Forestry Service. The condition of Kubu Raya's flammable land makes it vulnerable to forest fires (Jawad, Nurdjali, & Widiastuti, 2015). Although the impact of forest fires is very felt, detecting the source or location of forest fires is not easy. The extent of forests in Indonesia is one of the main obstacles. One of the methods that has long been used to detect the source of forest fires is to utilize satellite imagery data (Indradjad, Purwanto, & Sunarmodo, 2019).

By using remote sensing, data can be collected without much fieldwork. The results can be faster and cheaper with the help of technologies such as satellites, and can be an overview and tool for rapid policy-making to stop forest and land fires. One of the field survey methods requires a large cost, so the data analysis process uses satellite data because this method is fast, precise, and accurate, so the process does not take long (Endrawati et al., 2016).

The phenomenon of forest fire patterns in peatlands can be studied and modeled using random forest methods with the help of modern technologies such as machine learning. The Analytic Hierarchy

Process (AHP) method is also used to predict fire spots. The AHP method is also used to select the best options in decision-making and the Fuzzy Analytic Hierarchy Process (FAHP) method. However, the randomized forest method can study data samples that do not correlate with each other, making them less effective. In addition, the results have a low error rate (Melgivari, Thamrin, & Roslinda, 2019). (Kusuma et al., 2021). In this identification, this parameter is used. This research will focus on identifying the potential for forest fires in Kubu Raya Regency with remote sensing and random forest machine learning. The results of the research are expected to be a guideline for the community in an effort to overcome forest fires in the peat forest area of Kubu Raya Regen

PROBLEM STATEMENT

Based on the above background, the problems faced by the author can be formulated as follows:

1. How to identify potential forest fires in Kubu Raya Regency?
2. What are the results of the accuracy of identifying potential wildfires using the random forest method?

LITERATURE REVIEWS

Table 1: Literature Review

No.	Author, Year	Heading	Result	Relevance to Research
1.	Muhammad Idrus Fachruddin (2015)	Comparison of Random Forest Classification Methods and Supporting Vector Engines for Epilepsy Detection Using Electroencephalographic (EEG) Recording Data. [10]	Random forest is a classification method that consists of a combination of classification trees (CARTs) that are independent of each other. Classification predictions are obtained through the voting process (the largest number) of the classification tree that is formed. Random forest is a development of the ensemble method first developed by Leo Breiman (2001) and used to improve classification accuracy.	Explaining the Random Forest Method
2.	Tamas Faiz Dicelebica, Aji Ali Akbar, and Dian Rahayu Jati (2022)	Identification and prevention of forest and peatland fire disaster-prone areas based on geographic information systems in West Kalimantan. [11] .	Identifying fire potential over the past five years, West Kalimantan has become an area where forest and peatland fires are frequent. This is because West Kalimantan Province, which is located on the island of Kalimantan, is	Explaining about wildfire identification

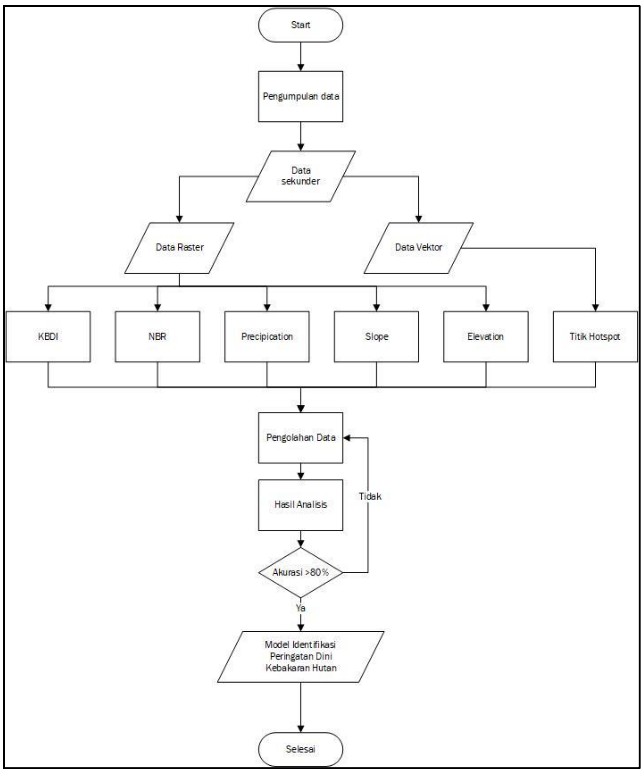
			an island that has a type of peatland other than the islands of Sumatra and Papua. This province is an area that is crossed by the equator and most of its territory is peatland. In general, peat areas have flammable characteristics, the ability to store biomass, waste, and mineral soils.	
3.	Trya Ayu Pratiwia, Muhammad Irsyada, Rahmad Kurniawana (2021)	Classification of Forest and Land Fires Using Algorithms Naif Bayes (Case Study: Riau Province). [12] .	Data Mining is the process of extracting implicit and useless data into information from large amounts of data. One of the parts of the Data Mining is a classification method. One of the applications of classification for forest fires.	Explaining the Classification Method
4.	Abdul Jawad, Bachrun Nurdjali, Tri Widiastuti (2015)	Zoning of Forest and Land Fire-Prone Areas in Kubu Raya Regency, West Kalimantan Province. [5]	Kubu Raya Regency has become a subscription to land fires every time it enters the dry season. According to data from the Forest and Land Fire Management Unit (UPKHL) of the West Kalimantan Provincial Forestry Service until October 2013, the number of hotspots in Kubu Raya Regency was 349. Kubu Raya is indeed vulnerable to forest fires, because the land condition is dominated by peatland so it is easy to burn.	Explaining fire-prone zoning in Kubu Raya
5.	Bambang Hero Saharjo and Dimas Adi Nugraha (2022)	The Effect of Rainfall on the Decline of Hotspots in Indonesia in 2019-2020. [13]	Climate has a major influence on forest and land fires, especially rainfall. Indirectly, rainfall can affect the occurrence of forest and land fires, although it is not a determinant of fire. According to Nurhayati et al. (2014), hotspot data can be combined	Explaining the influence of climate on wildfires

	with data such as rainfall so that a model of the relationship between the amount of rainfall and the number of hotspot detections in the area can be found. In other words, the relationship between hotspots and rainfall conditions can be an indicator of the occurrence of forest and land fires.

METHODOLOGY

Below are the steps taken to support the research process so that the research carried out can run more structured and systematic. As shown in Figure 1.

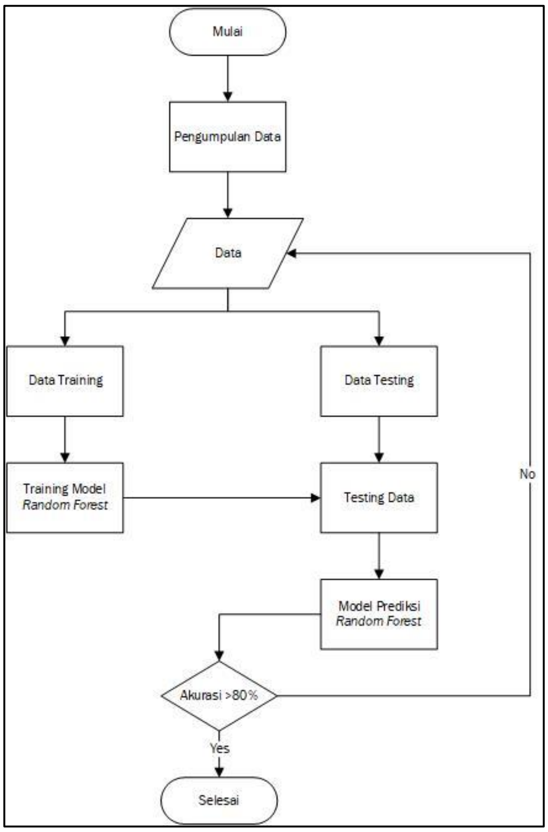
Figure 1: Research Flow Diagram



The flow diagram in the image depicts a series of research stages that are systematically arranged to support the process of identifying potential forest fires. The research started from the secondary data collection stage which consisted of two main types of data, namely raster data and vector data. Raster data includes important indices such as the Keetch–Byram Drought Index (KBDI), Normalized Burn Ratio (NBR), as well as precipitation data that are used to look at drought conditions, vegetation health, and weather dynamics. Meanwhile, the vector data used included slope, elevation, and hotspot points obtained from satellite imagery as an early indicator of the presence of hot spots in the research area. All of the data is then combined and processed through the data processing stage to obtain the basic information needed in model analysis.

At the analysis stage, the model is run to produce an output in the form of the level of forest fire potential based on a combination of various parameters that have been inputted. The results of the analysis were then evaluated based on the accuracy value. If the accuracy value obtained has not met the minimum limit, which is less than 80%, then improvements are made through parameter readjustment, data cleaning, or model optimization until more accurate results are obtained. However, when the accuracy of the model has reached or exceeded 80%, the results are considered adequate and are used to develop a forest fire early warning identification model. Overall, this flow chart shows that each stage of research is designed in a structured and interconnected manner, from data collection, processing, model evaluation, to the preparation of an identification system. These stages ensure that the research process not only runs systematically and coherently, but also produces a valid, reliable, and implementable model in supporting efforts to mitigate and prevent forest fires in the Kubu Raya Regency area.

Figure 2: Random Forest Streams



In the data analysis stage using the Random Forest method at RStudio, the process begins with data preparation which includes importing all research variables into the RStudio work environment. The data that has been collected then goes through a cleaning process, such as handling lost values, checking outliers, and normalization or transformation if necessary to maintain the consistency of the data structure. After the data is ready to be used, the next step is to divide the dataset into two parts, namely training data (training set) and test data (test set). The general proportion of the distribution is 70% for training and 30% for testing, but this proportion is flexible and can be adjusted based on research needs and data characteristics. The Random Forest model is then built using training data by calling *the randomForest()* function or through *a caret* package for a more controlled training process. At this stage, a number of important parameters such as the number of trees (ntree), the number of random variables selected on each tree separation (mtry), and the depth of the tree can be adjusted to optimize model performance. Once the model is successfully formed, the model is used to predict the test data to assess the accuracy of the prediction in mapping the potential for wildfires.

The predicted results are then compared with the actual values to calculate the evaluation metrics. Some of the evaluation indicators used include accuracy, kappa, sensitivity, specificity, and confusion matrix to evaluate the performance of the model more comprehensively. If the evaluation results show that the accuracy of the model does not reach the predetermined minimum limit, which is 80%, then the model needs to be corrected through a process of parameter tuning, adjustment of input variables, or reselection of a more representative subset of data. Conversely, if the model's accuracy value reaches more than 80%, the model is considered to have met the feasibility standards and is subsequently stored as the final model. In its implementation, RStudio provides various supporting packages such as *randomForest*, *caret*, *rpart*, and *e1071* that facilitate the process of developing, tuning, and evaluating the model. The use of these packages allows the study to produce more accurate, stable, and reliable prediction models to support the identification of potential wildfires

FINDINGS AND DISCUSSIONS

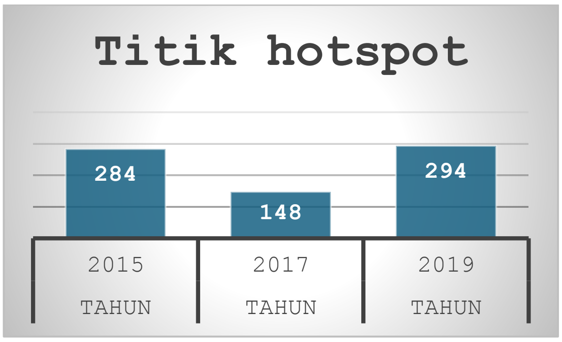
Data Needs Analysis

In the process of analyzing data needs, this study discusses the identification of potential forest fires using random forest methods in Kubu Raya Regency. The next stages are as follows.

Data Processing Analysis

At this stage, spatial data is processed in the form of hotspot data in Kubu Raya Regency. Google Earth Engine to download images, *Rstudio* to perform spatial analysis and *ArcGIS* to display hotspot points that have previously been analyzed in *Rstudio,* here are hotspot data from 2015, 2017, 2019 in Figure 3.

Figure 3: Diagram Hotspot Points in Kubu Raya Regency



Satellite Image Data Processing

At this stage, the researcher processes satellite image data using the Google Earth Engine (GEE) platform to retrieve the required data.

a. *Normal Burn Ratio* (NBR)

Normalized Burn Ratio (NBR) is an index designed to identify burned areas, a high NBR value can indicate good vegetation while a low NBR value indicates vacant land and burned areas. Here is the NBR source code to process in Google Earth Engine.

Load the Landsat image.

```
var img = ee. Gambar('MODIS/006/MOD09GA/2012_03_09');

var img = ee. Image Collection ("LANDSAT/LC08/C02/T1_TOA")

  .filterDate('2019-01-01', '2019-12-31')

  .median();
```

```
Use normalizedDifference(A, B) to calculate (A - B) / (A + B)
var nbr = img.normalizedDifference(['B5', 'B7']);
Create a palette: list of hex strings.
var palette = ['FFFFFF', 'CE7E45', 'DF923D', 'F1B555', 'FCD163', '99B718',
           '74A901', '66A000', '529400', '3E8601', '207401', '056201',
           '004C00', '023B01', '012E01', '011D01', '011301'];
Center the map
Map.centerObject(Fortress)
Display the input image and the NBR it comes from.
Map.addLayer(img.select(['B6', 'B5', 'B4']),
   {profit: [0.1, 0.1, 0.1]});
  Map.addLayer(nbr.clip(Kubu Raya), {min: 0, maks: 1, palet: palet},
```

The above source code is used to load the Landsat imagery, calculating the Normal Burn Ratio (NBR) index in the vulnerable period from January 1, 2019 to December 31, 2019.

b. *Ravine*

Precipication is the process of dropping all material poured from the atmosphere onto the earth's surface in liquid (rain) or solid (snow) form. Rainwater that seeps into the soil as groundwater is called percolation. Here is a script for finding rainfall to process in Google Earth Engine.

```
var dataset = EE ImageCollection('UCSB-CHG/CHIRPS/PENTAD')
            .filter(ee. Filter.date('2019-01-01', '2019-12-30'))
            .means ()
            .klip(Kubu Raya)
var precipitation = dataset.select('presipitasi');
var presipitasiVis = {min: 0.0,
                    max: 112.0,
 palet: ['001137', '0aab1e', 'e7eb05', 'ff4a2d', 'e90000'],
};
```

The above code is used, to visualize rainfall data in the Kubu Raya area over a certain time span, i.e. from January 1, 2019 to December 31, 2019, and display the image on the map.

c. *Keetch-Byram Drought Index* (KBDI)

The Keetch-Byram Drought Index has been commonly used as an indicator of fire potential due to its sensitivity in responding to hydrological drought, following processing scripts in Google Earth Engine.

```
var collection = ee. ImageCollection('UTOKYO/WTLAB/KBDI/v1')
  .filterBounds(Kubu Raya)
  .select('KBDI')
  .filterDate('01-01-2019', '30-12-2019')
  .means ()
  .klip(Kubu Raya)
var bandViz = {min: 0,
                          Max: 800,
  Pole: ['001A4D', '003CB3', '80AAF', '336600', 'CCCC00', 'CC9900', 'CC6600',
    '660033']};
Map.addLayer(collection, bandViz, 'Keetch-Byram Drought Index');
```

The code above is to visualize the *Keetch-Byram Drought Index* (KBDI) in the Kubu Raya area from January 1, 2019 to December 30, 2019.

d. Altitude

Elevasi (height) is the height of an object from a certain point, here to find out the height of Kubu Raya Regency. Here is a script to find the height for processing in Google Earth Engine.

```
Band: ["Elevation"]
gamma: 1
Max: 251
min: -2
Opacity: 1


var Elevasi_Kubu Raya = SRTM.clip(Kubu Raya)
var Slope_Kubu Raya_degree = ee. Terrain.slope(Elevasi_Kubu Raya)// Lowering the Slope or slope in degrees
var Slope_Kubu Raya_percent = ee. Terrain.slope(Elevasi_Kubu Raya)// Lowering the Slope or slope in percent
.divide(180).multiply(Math.PI).tan().multiply(100).rename('Slope_percent')// slope to percent conversion function
```

The code above is the use of Google Earth Engine (GEE), to visualize the altitude or altitude (masl) in the Kubu Raya area.

d. Slope

A slope is a slope or sloped surface (forming an angle with a flat plane). Usually the slope of the opening (surface) of an open mine. Here is the Slope processing script in Google Earth Engine.

Band: ["Slope"]

Gamma: 1

Max: 7.986710548400879

Opacity: 1

var Slope_Kubu Raya_degree = ee. Terrain.slope(Elevasi_Kubu Raya)// Lowering the Slope or slope in degrees

var Slope_Kubu Raya_percent = ee. Terrain.slope(Elevasi_Kubu Raya)// Lowering the Slope or slope in percent

.divide(180).multiply(Math.PI).tan().multiply(100).rename('Slope_percent') // tilt to percent conversion function

Map.addLayer(Slope_Kubu Raya_degree,slope_vis,"slope_Kubu Raya")

The code above is the use of Google Earth Engine *(GEE),* to visualize the slope or slope of the *Kubu Raya* area.

Spatial Data Processing

In this spatial processing, using spatial analysis Arc Map refers to the investigation of entities by examining, evaluating, and modeling spatial data characteristics such as location, attributes, and relationships that reveal the geometric or geographic characteristics of the data. The first stage of spatial analysis in *Arc Map* is to enter the Landsat image data and the selected host data, then select the arctoolbox-extraction-extract multi value to point menu, the command is to find the image value to a point.

Figure 4: Multi-Value Attribute to the Result of the Dot Extract in 2015

TITIK2015

FID	Shape *	LATITUDE	LONGITUDE	ACQ DATE	Tahun	Bulan	Elevation	Slope	KBDI	NBR	Precipita	Keterangan	Musim
45	Point ZM	-0,478634	109,222527	21/03/2015	2015	Maret	11	5,119183	164,773	0,507801	29,958	Terbakar	Hujan
46	Point ZM	-0,479161	109,219048	21/03/2015	2015	Maret	5	2,289336	164,773	0,579548	29,958	Terbakar	Hujan
47	Point ZM	-0,475829	109,218559	21/03/2015	2015	Maret	14	1,618833	164,773	0,484276	29,958	Terbakar	Hujan
48	Point ZM	-0,760418	109,543945	24/03/2015	2015	Maret	1	9,157589	160,045	0,448031	33,499	Terbakar	Hujan
49	Point ZM	-0,553513	109,313278	24/03/2015	2015	Maret	10	0	156,689	0,348766	29,4374	Terbakar	Hujan
50	Point ZM	-0,849723	109,484856	30/05/2015	2015	Mei	3	3,620014	190,566	0,519014	31,3597	Terbakar	Kemarau
51	Point ZM	-0,761778	109,262962	20/06/2015	2015	Juni	16	3,61976	157,134	0,589013	28,4196	Terbakar	Kemarau
52	Point ZM	-0,468933	109,318001	22/06/2015	2015	Juni	9	4,578669	155,585	0,538117	29,978	Terbakar	Kemarau
53	Point ZM	-0,524304	109,460861	24/06/2015	2015	Juni	22	1,618777	145,868	0,333026	32,3696	Terbakar	Kemarau
54	Point ZM	-0,713658	109,347183	28/06/2015	2015	Juni	5	0	164,451	0,530008	30,1166	Terbakar	Kemarau
55	Point ZM	-0,835284	109,320442	30/06/2015	2015	Juni	10	8,093885	171,132	0,33182	28,556	Terbakar	Kemarau
56	Point ZM	-0,712878	109,344612	30/06/2015	2015	Juni	6	3,61992	164,451	0,305979	-9999	Terbakar	Kemarau
57	Point ZM	-0,483964	109,236542	30/06/2015	2015	Juni	14	0	162,423	0,593434	29,958	Terbakar	Kemarau
58	Point ZM	-0,815564	109,500076	02/07/2015	2015	Juli	12	3,619769	178,182	0,63942	32,7343	Terbakar	Kemarau

0 (0 out of 8597 Selected)

Random Forest Classification with *Rstudio*

At this stage, the researcher analysed this data using *Rstudio* software and using the R programming language.

Packages on *rstudio*

At this stage, the researcher needs some of the *packages* needed for analysis using random forest classification. The *packages* needed are *raster packages* that serve for geographic data analysis and modelling, random *forest packages* that serve to develop random forests with n adjustable number of trees, a set of functions that help streamline each process that creates predictive models and *package* SF functions to read and write data Geometry Engine Open Source *(*Geos*)* for geometric operations.

```
Library #package

Library (raster)

library(randomForest)

Library (Caret)
```

Separation Process

The process of separation in a random forest occurs during the formation of each decision tree in the *ensemble*. A random forest combines many decision trees built from a randomly drawn subset of data. The *separation* process is an important step in the formation of a decision tree, where the tree divides the data set into two or more. This is the source code in the separation process.

```
ind <- sample(2, nrow(data), replace =
TRUE, prob = c(0.7, 0.3))

training <- data[ind==1,]

Test <- data[ind==2,]
```

Random Forest Modeling

Random forest modelling is a powerful machine learning method and is often used for classification and regression tasks. The randomForest() function of the "*randomForest*" package is used to create a random forest model.

```
#membuat Random Forest Model

random_forest <- randomForest(as.factor(Description)~.,data = train,ntree

= 150, importance = T)
```

In the process of creating a random forest model, the variable "Description" is used as Y while the remaining variable is used as X, and "data = training" is also used to create the model, "ntree = 150" is the number of trees that will be created by the *random forest.*

Export files

At this stage, the researcher exports *a raster file*, this file contains the results of a random forest model, to export the file using the source code *writeRaster* which functions to export files that have previously been processed in raster form. In the image is an *export file* of the source code generated from a random forest model.

#export results

predictors5 <- stack(list.files(file.path("E:/bahan skipsi/kebakaran"), pattern='tif$', full.names=TRUE))

KBRN<- predict(predictor5, rf)

IKH <-writeRaster(KBRN, file name="RF2015.tif")

2015 Prediction Results

Figure 5: Prediction results

```
Confusion Matrix and Statistics

                Reference
Prediction       Terbakar Tidak Terbakar
  Terbakar            105             18
  Tidak Terbakar        0            110
```

a) *True Positive* (TP): 105 (Model predicts "Burn" and is correct)

b) *False Positive* (FP): 18 (Model predicts "Burn" but is wrong)

c) *False Negative* (FN): 0 (Model predicts "Not Burned" but is false)

d) *True Negative* (TN): 110 (Model predicts "Not Burned" and is correct)

The results of the 2015 Model Evaluation can be seen from Table 2.

Table 2: Classification results in 2015

Class	Burn	Not Burn	Entire	User Accuracy
Burn	105	18	123	85.37%
Not Burn	0	110	110	100%
Entire	105	128	233	
Producer Accuracy	100%	85.94%	Overall Accuracy	92.27%
			Kappa	84.63%

Evaluation results from Table 2 above is the classification model used to identify 12 acres: "Burned" and "Not Burned". Based on the data, the model showed an overall accuracy of 92.27%. For the "Burned" class, the model achieved a user accuracy of 85.37% and a manufacturer's accuracy of 100%, meanwhile, for the "Not Burned" class, the user accuracy reached 100%., and the manufacturer's accuracy of 85.94%, the Kappa value was 84.63%.

2017 Prediction Results

Figure 6: Prediction results

```
Confusion Matrix and Statistics

                  Reference
Prediction        Terbakar Tidak Terbakar
  Terbakar              44             22
  Tidak Terbakar         0            130
```

a) *True Positive* (TP): 44 (Model predicts "Burn" and is correct)
b) *False Positive* (FP): 22 (Model predicts "Burn" but is wrong)
c) *False Negative* (FN): 0 (Model predicts "Not Burned" but is false)
d) *True Negative* (TN): 130 (Model predicts "Not Burning" and true)

The results of the 2017 Model Evaluation can be seen from Table 3**.**

Table 3: Classification Results in 2015

Class	Burn	Not Burn	Entire	User Accuracy
Burn	44	22	66	66.67%
Not Burn	0	130	110	100%
Entire	44	152	196	
Producer Accuracy	100%	85.53%	Overall Accuracy	88.78%
			Kappa	72.63%

The results of the evaluation from Table 3 above are the classification model used to identify the following categories: "Burned" and "Not Burned". Based on the data, the model showed an overall accuracy of 88.78%. For the "Burned" class, the model achieved a user accuracy of 66.67% and a manufacturer's accuracy of 100%., meanwhile, for the "Not Burned" class, the user accuracy reached 100%, and the manufacturer's accuracy of 85.53%, the Kappa value was 72.63%.

2019 Prediction Results

Figure 7: Prediction results

```
Confusion Matrix and Statistics

                  Reference
Prediction        Terbakar Tidak Terbakar
  Terbakar             106             13
  Tidak Terbakar         0             62
```

a) *True Positive* (TP): 106 (Model predicts "Burn" and is correct)
b) *False Positive* (FP): 13 (Model predicts "Burn" but is wrong)
c) *False Negative* (FN): 0 (Model predicts "Not Burned" but is false)
d) *True Negative* (TN): 62 (Model predicts "Not Burned" and true)

Table 4: 2019 Classification Results

Class	Burn	Not Burn	Entire	User Accuracy
Burn	106	13	123	89.08%
Not Burn	0	62	110	100%
Entire	106	75	181	
Producer Accuracy	100%	82.67%	Overall Accuracy	92.82%
			Kappa	84.82%

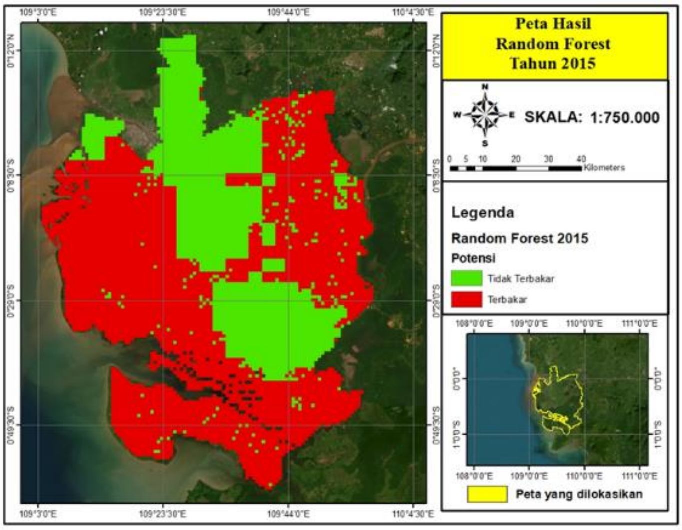
The results of the evaluation from Table 4 of the classification models used to identify two classes: "Burned" and "Not Burned". Based on the data, the model showed an overall accuracy of 92.82%. For the "Burned" class, the model achieved a user accuracy of 89.08% and a manufacturer's accuracy of 100%, meanwhile, for the "Not Burned" class, the accuracy of the user reached 100%, and the accuracy of the manufacturer was 82.67%, the Kappa value was 84.82%.

Modelling Results

At this stage the modelling results to check and evaluate whether there is a specific object or information. It involves data collection, analysis, and assessment to verify information. In the validation of hotspots, this aims to check and confirm whether the detected points as fire points are really real fires.

a. 2015 model year

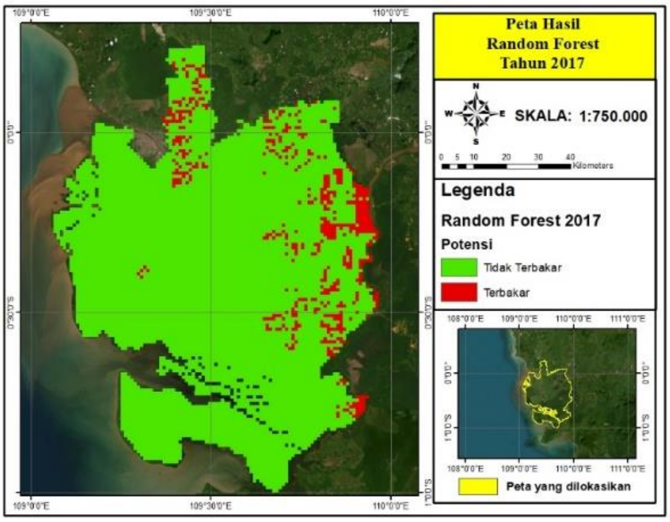
Figure 8: Random Forest Model 2015



The map shown is the result of Random Forest's model for predicting potential fires in 2015, based on data from 2015. The Random Forest model applied generates a fire prediction map that shows high- and low-risk areas. With high accuracy and the ability to detect areas with fire potential precisely.

b. 2017 model year

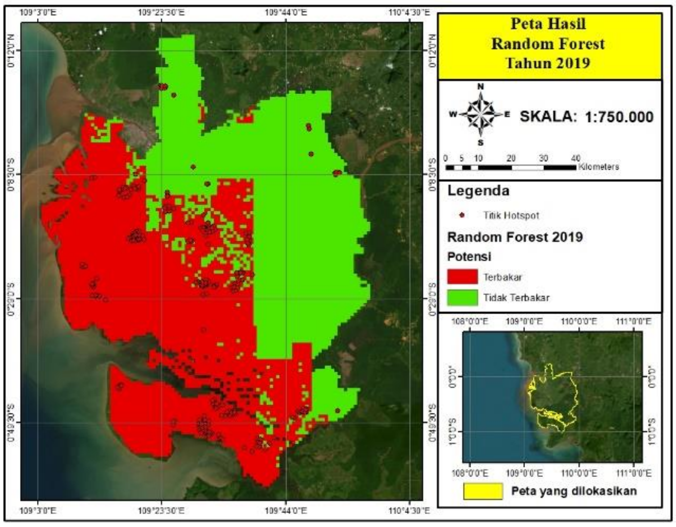
Figure 9: Random Forest Model 2017



The map shown is the result of the Random Forest model to predict potential fires in 2017 The Random Forest model applied produced a fire prediction map showing high and low risk areas. With high accuracy and the ability to detect areas with fire potential precisely.

c. 2019 model year

Figure 10: Random Forest Model 2019



The map shown is the result of the Random Forest model to predict potential fires in 2019. The applied Random Forest model produces a fire prediction map showing high and low risk areas. With high accuracy and the ability to accurately detect areas with potential burns

Limitations of Future Research & Recommendations.

In the data analysis stage using the Random Forest method at RStudio, the process begins with data preparation which includes importing all research variables into the RStudio work environment. The data that has been collected then goes through a cleaning process, including handling lost values, checking outliers, and normalization or transformation if necessary to ensure data quality and consistency. After the data is declared ready, the dataset is divided into two main parts, namely training data (training set) and test data (test set). The proportion of the distribution is generally 70% for training and 30% for testing, but this distribution is flexible and can be adjusted according to the research objectives and data characteristics. The Random Forest model is then built using training data through *the randomForest()* function or *caret* package to obtain a more structured training process and allow for optimal parameter setting. At this stage, several important parameters such as the number of trees (ntree), the number of random variables selected on each separation node (mtry), and the depth of the tree can be adjusted to achieve the best performance. Once the model is formed, the model is used to predict the test data, and the prediction results are then compared to the actual values to measure the model's performance.

Performance evaluation is carried out using various metrics such as accuracy, kappa coefficient, confusion matrix, sensitivity, and specificity. If the accuracy value has not reached the set minimum limit, which is 80%, then a process of parameter tuning, adjusting input variables, or reselecting a more representative subset of data is carried out to improve model performance. However, if the accuracy value has exceeded 80%, the model is considered feasible and is kept as a final model ready for use in the process of identifying potential wildfires. RStudio provides a number of supporting packages such as *randomForest*, *caret*, *rpart*, and *e1071* which are very helpful in the process of developing, validating, and evaluating the model comprehensively. Although the Random Forest model shows good performance, the study has some limitations. First, the accuracy of the model is highly dependent on the quality and resolution of the input data, so the limitation of satellite data or the presence of noise can affect the prediction results. Second, this study only uses certain variables such as KBDI, NBR, rainfall, slope slope, elevation, and hotspots, so that the potential for fires influenced by socio-economic factors or human activities has not been fully accommodated. Third, the model has not been tested for real-time data or streaming-based early warning systems, so its application is still limited to static spatial analysis. In addition, the scope of the research area is limited to one district, making generalization of results to other regions need to be done carefully.

For future research, several recommendations may be considered. First, the integration of higher-resolution data such as Sentinel-2 or Landsat 9 can improve the accuracy of the model. Second, adding additional variables such as population density, distance to roads, land use patterns, and human activities can provide a more comprehensive picture of fire risk. Third, the use of advanced machine learning techniques such as Gradient Boosting, XGBoost, or Deep Learning can be explored to compare its performance with Random Forest. In addition, the development of a real-time early warning system that integrates daily weather data and actual hotspots can make the model more responsive to field conditions. Further research is also recommended to expand the scope of the research area and conduct field validation (ground truthing) to improve the reliability of the model results.

CONCLUSION

Based on the results and discussion of this study, the following conclusions can be drawn. The results of the study show the identification and analysis of forest fire predictions in Kubu Raya Regency in 2015, 2017, and 2019. Based on the results of the evaluation, it can be concluded that the classification model shows good performance in identifying two classes, namely "Burned" and "Not Burned". The overall accuracy of the model varies between 88.78% and 92.82%, with the best performance recorded in 2019. For the "Burned" class, the model achieved the highest user accuracy of 89.08% in 2019, while the manufacturer's accuracy was always 100%, indicating the success of the model in identifying all cases that were truly "Burned".

On the other hand, for the "Not Burned" class, the user's accuracy is always 100% across all tables, indicating that the model is very accurate in classifying these cases. However, the manufacturer's accuracy for the "Non-Burning" class varies, with the highest yield being 85.94% in 2015 and the lowest of 82.67% in 2019. A consistent Kappa value of 84.82% across all tables indicates that the model has excellent and stable performance, exceeding random predictions.

ACKNOWLEDGEMENT

Deep gratitude, with sincere sincerity, for the moral and material support that has made this writing possible, is conveyed by the researcher to: My beloved parents, dear Dad and Mother, for the prayers, encouragement, motivation, and boundless love and support over the years. Dr. Ir. Muhammad Nanang Prayudianto, M.Sc. as the Dean of the Faculty of Engineering & Science, Ibnu Khaldun University Bogor, Fitrah Satrya Fajar Kusumah, S.Komp., M.Kom., as the Head of the Computer Engineering Study Program, Faculty of Engineering & Science, Ibnu Khaldun University Bogor, Dr. Erwin Hermawan, S.Si., M.Sc. The main supervisor who has given advice and input to the author, Sahid Agustian Hudjimatsu, S.T., M.Kom, supervisors who have given advice and input to the author, Colleagues from the Geospatial Information Technology group who have helped and motivated the author in completing this final report, Friends of Geospatial Information Technology 2017 who have helped and provided motivation and encouragement to the author in completing this final report, friends who have helped and encouraged the author in completing this final report. The author realizes that in the preparation of this report it is still far from perfect, considering the limitations of knowledge and ability, so what the author conveys is not without shortcomings.

REFERENCES

A. Humam, M. Hidayat, A. Nurrochman, A. I. Anestatia, A. Yuliantina, and S. P. Aji, "Identification of Forest and Land Fire Vulnerability Areas Using Geographic Information Systems and Remote Sensing in the West Tanjung Jabung Area, Jambi Province," *J. Geoscience and Remote Sens.*, vol. 1, no. 1, pp. 32–42, 2020, doi: 10.23960/jgrs.2020.v1i1.14.

A. Indradjad, J. Purwanto, and W. Sunarmodo, "Analysis of Hotspot Point Accuracy Level from S-NPP VIIRS and Terra/Aqua MODIS in Fire Incidents," *J. Remote Sensing and Processing. Digital Image Data.*, vol. 16, no. 1, pp. 1–8, 2019, [Online]. Available: http://dx.doi.org/10.30536/j.pjpdcd.2019.v16.a3053

A. Jawad, B. Nurdjali, and T. Widiastuti, "Zoning of Forest and Land Fire-Prone Areas in Kubu Raya Regency, West Kalimantan Province," vol. 3, pp. 88–97, 2015.

A. Pandu *et al.*, "Analysis of Twitter Sentiment After the Announcement of the Results of the 2019 Presidential Election Using the Lexicon Analysis Method," *J. Tek. Inform.*, vol. 15, no. 1, pp. 33–44, 2020.

A. R. Kusuma, F. M. Shodiq, M. F. Hazim, and D. P. Laksono, "Results of Peatland Fire Pattern Study through Sentinel-2 Satellite Imagery with Machine Learning Random Forest Method Implementation: A Literature Review," *JGISE J. Geospatial Inf. Sci. Eng.*, vol. 4, no. 2, p. 81, 2021, doi: 10.22146/jgise.60828.

A. Rachman, B. H. Saharjo, and E. I. K. Putri, "Forest and Land Fire Prevention Strategies in the Forest Management Units of Kubu Raya, South Ketapang, and North Ketapang in West Kalimantan Province," *J. Ilmu Pertan. Indonesian.*, vol. 25, no. 2, pp. 213–223, 2020, doi: 10.18343/jipi.25.2.213.

A. S. Nurfika Maulina Larasati, Sawitri Subiyanto, "Analysis of Land Use and Utilization (P2T) Using the Geographic Information System of Banyumanik District in 2016," *J. Geod. Undip*, vol. 5, no. 4, pp. 132–139, 2017, [Online]. Available: https://media.neliti.com/media/publications/202086-analisis-penggunaan-dan-pemanfaatan-tana.pdf

A. S. Nurfika Maulina Larasati, Sawitri Subiyanto, "Analysis of Land Use and Utilization (P2T) Using the Geographic Information System of Banyumanik Regency in 2016," *Journal of Undip Geodesy*, vol. 5, no. 4, pp. 132–139, 2017

D. Handayani and A. Setiyadi, "Remote Sensing of Remote Sensing," *May Edition*, vol. 7, no. 2, pp. 113–120, 2003.

Endrawati, J. Purwanto, S. Nugroho, and R. A. S, "Identification of Former Forest and Land Fire Areas

Using Semi-Automated Analysis of Satellite Images," *Semin. Nas. Geomatics 2017 Nov. Technology. Provision of Geospatial Inf. for Developers. Sustainability*, pp. 273–282, 2016.

F. D. Julianto, D. P. D. Putri, and H. H. Safi'i, "Analysis of Vegetation Change with Sentinel-2 Data Using Google Earth Engine," *Indonesian Journal of Remote Sensing*, vol. 02, no. 02, pp. 13–18, 2020.

M. S. Simanjuntak, D. Kusnandar, and N. N. Debataraja, "Mapping Prone to Forest Fires in West Kalimantan in 2020," *Bimaster Bul. Ilm. Matt. Stat. and Ter.*, vol. 11, no. 5, pp. 777–784, 2022, [Online]. Available: https://jurnal.untan.ac.id/index.php/jbmstr/article/view/59892

P. Utami, "Priority Extinguishing Locations of Forest and Land Fires Using Analytic and Promethee Hierarchy Processes," *Techno Xplore J. Computer Science and Techno. Inf.*, vol. 8, no. 1, pp. 40–48, 2023, doi: 10.36805/technoxplore.v8i1.5169.

R. Bencana, K. Hutan, G. Studi, K. K. Raya, and P. K. Barat, "Disaster Risk of Forest and Peatland Fires (Case Study: Kubu Raya Regency, West Kalimantan Province)," vol. 12, no. 3, pp. 160–170, 2023.

R. Melgivari, E. Thamrin, and E. Roslinda, "Community Participation in Forest and Land Fire Management in Limbung Village, Sungai Raya District, Kubu Raya Regency," *J. Sustainable Forest*, vol. 7, no. 3, pp. 1130–1139, 2019, doi: 10.26418/jhl.v7i3.37267.

T. A. Pratiwi, M. Irsyad, and R. Kurniawan, "Classification of Forest and Land Fires Using the Naif Bayes Algorithm (Case Study: Riau Province)," *Journal of Information Systems and Technology (Justin),* vol. 9, no. 2, p. 101, Apr. 2021, doi: 10.26418/justin.v9i2.42823.