Prediction of Forest Fire Occurrence in Peatlands using Machine Learning Approaches

Dedi Rosadi[i]
Department of Mathematics
Gadjah Mada University
Yogyakarta, Indonesia
dedirosadi@gadjahmada.edu

Widyastuti Andriyani
Master of Information Technology
STMIK Akakom
Yogyakarta, Indonesia
Widy.ugm@gmail.com

Deasy Arisanty
Department of Geography Education
Lambung Mangkurat University,
Banjarmasin, Indonesia
deasyrsnt@gmail.com

Dina Agustina
Department of Statistics
State University of Padang
Padang, Indonesia
dinagustina31@gmail.com

2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) | 978-1-7281-8406-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ISRITI51436.2020.9315359

***Abstract*— In this paper we consider the application of various machine learning approaches for prediction of the forest fire occurrence in the peatlands area. Here we consider some classical classification methods, such as support vector machine (SVM), k-Nearest Neighborhood (kNN), Logistic Regression (logreg), Decision Tree (DT) and Naïve Bayes (NB). For comparison purpose, we also consider more advanced algorithms, namely AdaBoost (DT based) approach. It is known that only a little number of similar studies is available for modeling peatlands fire occurrences in Indonesia. To illustrate the method, we consider the method using topographical and meteorological data from South Kalimantan Province. All computations are done using open source software R**

Keywords*— *early warning system, forest fire occurrence, topographical and meteorological data, peat lands fire, machine learning

I. INTRODUCTION

Forest fire is an important environmental world phenomenon and affected the environment, infrastructures, and human life. There are various techniques in literature that could be used in modeling forest fire, which can be categorized into three main methods, namely physics-based method, statistical method, and machine learning method, see e.g. [1]. It is also known that the variables that often used in the literature are basically based on satellite data, infra-red or smoke detectors and various sensors (such as the weather and meteorological data), see e.g. [2].

Among many approaches, machine learning and data mining approach has receive many attention by many researchers. For instances, [2] studied classification algorithm, namely the Multiple Regression (MR), Decision trees (DT), Random Forests (RF), Neural Networks (NN) and Support Vector Machines (SVM) to model the forest fire occurrence prediction using meteorological and forest weather index (FWI) variables. Reference [3] considers hybrid approach between clustering technique, normalization to preprocess the data and apply classification approach to the normalized data. They show that the combination of the Fuzzy C-Means clustering with Cosine distance, Min-Max normalization and Back-Propagation Neural Networks (with one hidden layer) classification method can give a relatively accurate prediction compare to other classification approach (SVM, K-Nearest Neighborhood, DT, and Naive Bayes) and the case without the clustering the data. [4] improve the empirical performance of [2] and [3] by applying what so called Adaptive Boosting (AdaBoost) approach ([5]) to model the same data as [2] and [3]. More recent studies are available, see e.g. [1] and the references therein. In [6], it was considered neural networks classification method using what so called the extreme learning machine (elm) and backpropagation (bp) approach to predict the forest fire occurrence. However, except [6], none of the mentioned result above is applying the method for peatlands area. In this paper, we consider various classical classification methods, such as support vector machine (SVM), Decision Tree (DT), Logistic Regression (logreg), k-Nearest Neighborhood (kNN) and Naïve Bayes (NB). For comparison purpose, we consider more advanced model, namely AdaBoost (DT based) approach. To illustrate the method, we consider the method using topographical and meteorological data from South Kalimantan Province. To the best of our knowledge, the AdaBoost approach has not been considered yet in any of the previous study in forest fire prediction before in Indonesia, either for peatlands or non peatlands study.

The rest of this paper is organized as follows. In this section we already provide a quick introduction to the problem that we considered in this paper. In Section 2, we outline short description of necessary theory related to our considered approach and provide the algorithms. In Section 3, we provide empirical results. Section 4 concludes the results.

II. METHODS

A. Classical Classification Methods

As we already noted, in this paper, we consider various classical machine learning classification methods, such as support vector machine (SVM), k-Nearest Neighborhood (kNN), Logistic Regression (logreg), Decision Tree (DT) and Naïve Bayes (NB) method. Here we only provide a short summary of the description of the methods since all of these methods has been considered in various standard basic machine learning books. See e.g. [7] for further detail.

1) Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm often used for classification. The SVM algorithms will find an optimal hyper plane in an N-1 dimensional space (N denotes the number of features/variables) which will classify the data. This can be obtained by setting an optimization problem for obtaining the maximum margin, i.e. the maximum distance between data points of both classes.
SVM can be of two types namely

a) Linear Separable SVM

Hyper plane or support vector machine formula for linear data:

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + \mathbf{w}_0 \tag{1}$$

with assumption:

$$g(\mathbf{x}) \geq 1, \forall \mathbf{x} \in class\ 1 \tag{2}$$

and $g(\mathbf{x}) \leq -1, \forall \mathbf{x} \in class\ 2$

Calculate the z margin for each support vector

$$\begin{aligned} z_+ &= \frac{|g(\mathbf{x})|}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|}, \forall \mathbf{x} \in class\ 1 \\ z_- &= \frac{|g(\mathbf{x})|}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|}, \forall \mathbf{x} \in class\ 2 \end{aligned} \tag{3}$$

Then the Total Margin z is

$$z = \frac{1}{\|\mathbf{w}\|} + \frac{1}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \tag{4}$$

Where: $\mathbf{w}$ = weight vector
Minimizing $\mathbf{w}$ is a non-linear optimization function used to maximize separation. This minimization of $\mathbf{w}$ can be done using the KKT (Karush-Kuhn-Tucker) condition with the Langrange multiplier λ_i.
Minimizes weight vector to maximum distance by using the formula:

$$\mathbf{w} = \sum_{i=0}^{N} \lambda_i y_i \mathbf{x}_i \tag{6}$$

KKT (Karush-Kuhn-Tucker) condition:

$$\sum_{i=0}^{N} \lambda_i y_i = 0 \tag{7}$$

b) Non-Linearly Separable SVM

Finding non-linear separable SVM can be used kernel functions to map data into higher dimension. The data is mapped to a higher dimension using the kernel trick so that it is linearly separable.
There are 2 kernel functions that are often used in SVM:

1. Radial Basis Function (RBF)

$$k(\mathbf{z}, \mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\boldsymbol{\sigma}^2} \right) \tag{8}$$

   It is usually used when the number of variables is less than the number of observations.

2. Polynomial Kernel

$$k(\mathbf{z}, \mathbf{x}) = (\mathbf{x}^T\mathbf{z} + c)^d \tag{9}$$

   It is usually used when the number of observations is less than the number of variables

See e.g. [8[for further detail.

2) K Nearest Neighbor (KNN)

KNN is a popular supervised machine learning algorithm in the classification problems. Here the data is classified according to k-data which is similar to it. The similarity is defined by the distance between points, which can be calculated using some distances, e.g.:

1. Euclidean

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{10}$$

2. Mahattan

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i| \tag{11}$$

3. Minkowski

$$d(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{n} \left(|x_i - y_i| \right)^q \right)^{1/q} \tag{12}$$

See e.g. [9[for further detail.

3) Naïve Bayes

A Naive Bayes classifier is one of the probabilistic classification methods, derived based on Bayes Theorem, Bayes theorem can be used to calculate posterior probability P(c|x) from P(c), P(x) and P(x|c).
Formula for Bayes theorem:

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \tag{13}$$

$$P(x \mid c) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \tag{14}$$

Where, here we have

P(c|x) is the posterior probability of class (c or target) given predictor (variables x, attributes).

P(c) is the prior probability of certain class.

P(x|c) is the likelihood which is the probability of predictor given certain class.

P(x) is the prior probability of certain predictor.

Steps for Naïve Bayes algorithm:

1. Convert the data set into a frequency table.
2. Calculate the empirical probability and define the Likelihood table.
3. Equation (13) to obtain the posterior probability for each class. We define that the class suitable for the data is the class with the highest posterior probability

See e.g. [10] for further detail.

4) Logistic Regression

Logistic Regression is one of the popular classification method. Logistic Regression can be used to classify the data into two classes or more.
Steps for Logistic Regression in general can be given as

a) Logistic regression hypothesis

The logistic regression classifier is defined as

$$h_\theta(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x} \quad (15)$$

Linear function is basically used as an input to another function such as g in the following relation

$$h_\theta(\mathbf{x}) = g(\boldsymbol{\theta}^T \mathbf{x}) \text{ where } 0 \le h_\theta(\mathbf{x}) \le 1 \quad (16)$$

Function g is the logistic or sigmoid function

$$g(\mathbf{z}) = \frac{1}{1+e^{-\mathbf{z}}} \text{ where } \mathbf{z} = \boldsymbol{\theta}^T \mathbf{x} \quad (17)$$

Sigmoid curve divided class into positive or negative. The output comes under the probability of positive class if it lies between 0 and 1.

b) Logistic regression decision boundary

Define a loss function to measure performance of algorithm using the weights on functions, represented by theta as follows:

$$h_\theta(\mathbf{x}) = g(\boldsymbol{\theta}^T \mathbf{x}) \quad (18)$$

c) Define the logistic regression parameter

$$J(\boldsymbol{\theta}) = \frac{1}{m}(-x^T \log(h) - (1-x)^T \log(1-h)) \quad (19)$$

After defining the loss function our prime goal is to minimize the loss function. The formula that we use is:

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \theta_j} = \frac{1}{m} x^T (g(\boldsymbol{\theta}^T \mathbf{x}) - x) \quad (20)$$

See e.g. [11] for further detail.

5) Decision Tree

Decision Tree algorithm is supervised learning algorithms that can be used to solve the classification problem. It can be applied using the following steps:

1. The complete dataset is used for the starting point of the tree (as the root node)
2. The Attribute Selection Measure (ASM) is applied to select the best attribute. This measure may be applied to find the root node and for sub-nodes. Popular measure for ASM are
   a. Information Gain

   DT will split the node with highest information gain (IG) , and it will be split first. IG can be calculated as:

   $$\textit{Information gain} = \textit{Entropy}(S) - [\textit{Weighted Average} * \textit{Entropy}(\textit{each feature})] \quad (21)$$

   The entropy of feature S can be calculated using:

   $$\textit{Entropy}(S) = \sum_{i=1}^{c} -p_i \log_2 p_i \quad (22)$$

   b. Gini Index

   The second ASM often used is called as Gini index, which is defined as:

   $$\textit{Gini Index} = 1 - \sum_j p_j^2 \quad (23)$$

3. The root node is divided into subsets of datasets based on the best attribute and the decision tree node is generated. This step is iteratively applied until the final node/leaf node is reached where the tree cannot be classified more.
4. To avoid over fitting, the size tree may be pruned. The pruning can be done using the methods which is called as Cost Complexity Pruning or Reduced Error Pruning.

See e.g. [12] for further detail.

B. Ensemble Classification: Adaboost Method

Adaptive Boosting or in short AdaBoost, is a machine learning ensemble approach. Using AdaBoost, the performance of the 'weak learners' can be improved by combining them into a weighted sum that is called as a boost classifier, which has the form

$$F_T(x) = \sum_{t=1}^{T} f_t(x) \quad (24)$$

where each f is a weak learner that takes an object x as input and returns the class of the input x. There are various variations of AdaBoost algorithms, the algorithm used in this study is called Stagewise Additive Modeling using a Multi-class Exponential loss function (SAMME) AdaBoost. See [5] for the detail.

III. Results and Discussion

A. Data Description

For the empirical study, we collect topographical and meteorological data from Kalimantan Selatan Province, especially a few days after the occurrence of fire hotspots of peatlands, where the variable "area" is labeled as "1". For classification purpose, we also collect data at the same spots at some other time when there is no fire occurrence, and therefore the area is labeled as "0". The variable collected is the time (of data is collected), the district area, LST (Land surface Temperature), Wind Speed, Humidity, Height and NDVI (normalized vegetation index). Data is collected only for year 2018 and is consisting of 202 cases.

B. Algoritms

The algorithms we apply as follows

Preprocessing steps

1. We use all of the data and split the data into two categories, which is the case of data with the variable area has the value 0 and labeled as "No Burned Area" and the case of data area has the value larger than 0 and labeled as "Burned Area"
2. Normalize all of the seven variables in both of the categories using min-max normalization, which is defined as

   $$v_i' = \frac{v_i - \min A}{\max A}(\text{newmax } A - \text{new min } A) + \text{new min } A \quad (25)$$

   where min A and max A is the minimum and the maximum values of an atribute A, v_i denotes the data value in attribute A (will be mapped into v_i'). Here we use the range [0,1] as the range of [new min A, new max A].

Classification step

3. We randomly split the data into training data (70, 80 and 90 percent) and testing data (30, 20 and 10 percent)
4. We apply the machine learning classification approaches to the training data and the best obtain model are tested into the testing data. Here for the computation we use R ([13])
5. To check the performance of the classification method, we use the accuracy measure, defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (26)$$

where TP, TN, FP, FN denotes the true positive, the true negative, the false positive and the false negative cases, respectively, in the categorical classification data..

For comparison purpose, we implement various machine learning approaches in step 4. The considered approaches are implemented in R, where it was used the function and the packages (in the parentheses) as follows: svm {e1071} [14] , knn {class} [15], glm {stats} [16], ctree {party} [17], naiveBayes {e1071} [14] and boosting {adabag} [18].

IV. DISCUSSION

The summary of the empirical results is given in Table 1. Here we consider several training and testing sample sizes for checking the performance of considered algorithms. The performance of each method in data training and data testing are given in the table. Here we can see that the accuracy of AdaBoost method is outperform the other approaches considered in the study in the in-sample data. However, for out-sample study, since in this study we use the boosting over DT approach, it is only make improvement over the weaker learning method (i.e. DT method), amd it is outperformed by kNN classification approach. In general, this study show that the machine learning approach, either the classical or the more advanced and recent approach can be used for fire occurrence detection of peatlands.

V. CONCLUSION

One of the key successes that help the forest fire firefighting is the early warning systems of fire detection. This study show that the machine learning approach, either the classical or the more advanced and recent approach can be used for fire occurrence detection of peatlands. With the combination of real time data collection based on Internet of Things approach, the considered method will have their importance in their future application.

Acknowledgement
The authors acknowledge the receive of research funding PD 2020 from Deputy of Reseach and Development, Ministry of Research and Technology of Republic of Indonesia.

TABLE I. SUMMARY OF THE PERFORMANCE CLASSIFICATION METHODS

Algorithms	Ratio data testing and training	Accuracy Training	Accuracy Testing
SVM	9:1	91,76%	95,00%
	8:2	90,74%	92,5%
	7:3	90.78%	91,8%
kNN	9:1 (k=3)	-	100%
	8:2 (k=3)	-	95,00%
	7:3 (k=7)	-	91,80%
Logistic Regression (logreg)	9:1	76,37%	90,00%
	8:2	75,92%	85,00%
	7:3	74,46%	83,60%
Decision Tree (DT)	9:1	91,00%	95,00%
	8:2	91,00%	92,00%
	7:3	91,00%	90,00%
Naïve Bayes (NB)	9:1	82,40%	90,00%
	8:2	82,1%	87,5%
	7:3	83,00%	86,9%
Adaboost (DT Based)	9:1	100%	95,00%
	8:2	100%	92,50%
	7:3	100%	91,80%

REFERENCES

[1] D. T. Buia, Q.-T. Buib, Q.-P. Nguyenc, B. Pradhand, H. Nampak and P. Trinh, "A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area", Agricultural and Forest Meteorology, vol. 233, pp. 32-44, February 2017

[2] P. Cortez and A. Morais, "A Data Mining Approach to Predict Forest Fires usingMeteorological Data", In: Neves J, Santos M F, Machado J (eds.) EPIA 2007, pp.512-523, 2007

[3] G. F. Shidik and K. Mustofa, "Predicting size of forest fire using hybrid model", Proceeding ICT EurAsia Conference 2014, pp 316-327, 2014

[4] D. Rosadi and W. Andriyani, "Prediction of forest fire using ensemble method" Presented in ICMSE 2020, Semarang, Indonesia, October 6th 2020,

[5] J. Zhu, H. Zou, S. Rosset and T. Hastie, "Multi-class AdaBoost", Statistics and Its Interface, vol. 2, pp.349–360, 2009

[6] D. Rosadi W. Andriyani and D. Arisanty, "Prediction of forest fire occurrence using extreme learning machine approach", Presented in ICMSDS 2020, Bogor Indonesia, 11th – 12th November 2020

[7] T. Hastie, R. Tibshirani and J.Friedman. "The Elements of Statistical Learning", 2nd Eds , New York: Springer, 2009

[8] A. Christmann, I. Steinwart, "Support Vector Machines", New York: Springer-Verlag, 2008

[9] Z. Zhang, "Introduction to machine learning: K-nearest neighbors", Annals of Translational Medicine, vol.4 , pp. 218-218, June 2016

[10] P. Cichosz, "Data Mining Algorithms: Explained Using R", NY: John Wiley & Sons, 2015

[11] G. James, D. Witten, T. Hastie and R. Tibshirani, "An Introduction to Statistical Learning: with Applications in R". NY: Springer, 2013.

[12] L. Rokach and O. Maimoon, "Data Mining with Decision Trees: Theory and Applications, SIngapore: World Scientific, 2008

[13] R Core Team, "R: A language and environment for statistical computing", Vienna: R Foundation for Statistical, 2020

[14] D. Meyer, "Support Vector Machines* The Interface tolibsvmin package e1071", Vienna: FH Technikum Wien, Austria, 2020

[15] B. D. Ripley, W. Venables, "Package 'class'", Vienna: CRAN, 2020

[16] R Core Team,"The R Stats Package ", Vienna: CRAN, 2020

[17] T. Hothorn, K. Hornik and A. Zeileis, "ctree: Conditional Inference Trees", Vienna: CRAN, 2020

[18] E. Alfaro," Package 'adabag'", Vienna: CRAN, 2020

[i] Corresponding author