Riau Forest Fire Prediction using Supervised Machine Learning

B S Negara, R Kurniawan, M Z A Nazri, S N H S Abdullah, R W Saputra and A Ismanto

1 Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Jl. HR. Subrantas Km.15, Pekanbaru, 28293, Indonesia
2 Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
5 Riau Data Camp (RDC), Riau, Indonesia
*Corresponding email: rahmadkurniawan@uin-suska.ac.id

Abstract. Forest fire is one of the environmental problems in terms of economically and ecologically detrimental. The number hotspot of forest fires in Indonesia has become increasing dramatically in September 2019 with 16178 hotspots that caused hazardous haze. The hazardous haze has disturbed about 1.04 million and 6.5 million people in Pekanbaru City and Riau Province subsequently. Hence, the development of early warning systems may provide effective strategic information plus accurate prediction results tailored for forest fire prevention and control. Weather data can be used as the main source for analyzing forest fires. This study aims to predict forest fire in Riau, Indonesia. This study has used 1733 weather data in five years (2015-2019). The forest fire prediction models were developed by using two supervised machine-learning techniques namely Decision Tree (DT) and Bayesian Network (BN). The experimental result shows that BN outperforms DT with accuracy rate and RMSE value in pairs of 99.62% and 0.076, and 93.18% and 0.244 subsequently. Despite its low performance, DT able to extract the main factors that caused the forest and land fires efficiently. Thus, it can be concluded that the prediction model using BN has the potential to be used effectively but still has plenty of room for improvement.

Keywords: Bayesian Network, Decision Tree, forest fire, haze, IoT, machine learning, Riau.

1. Introduction
Forest fire is one of the environmental problems in terms of economically and ecologically detrimental [1]. Forest fires in Indonesia have increased dramatically in September 2019 with 16178 hotspots. Forest fires have triggered dangerous haze for Indonesian people. In September 2019, the hazardous haze affected about 1.04 Million people in Pekanbaru City whereas approximately 6.5 Million people in Riau Province. The haze that covered Riau originates from forest fires in Sumatra. In September 2019, there were 1055 hotspots in Riau, 2705 hotspots in Jambi and 1507 hotspots in South Sumatra. The other regions of Indonesia with the highest number of hot spots are Central Kalimantan with 5574 hotspots and West Kalimantan with 2756 hotspots [2]. Pekanbaru is in the spotlight because on September 22nd, the Particulate Concentration Information (PM10) in Pekanbaru exceeds 600 μgram / m³, even though
the Threshold Limit Value of the allowed air pollution concentration is NAV PM10 = 150 µgram / m³ [3].

The haze became an annual disaster for two decades in Riau. Quick and accurate predictions are needed for the prevention and control of forest fires. Providing free masker and dismissed the schools and colleges are not the right solutions for the long term. Decisive and real action such as early warning system is needed to address the problem of forest fires that cause this serious air pollution problem. Weather data can be used as the main source for analyzing forest fires [1, 4–6]. Internet of Things technology using automatic sensors is operated for obtaining real-time temperatures. Weathers data were obtained can be analyzed using machine learning techniques. For the development of early warning systems, accurate prediction results can be used to provide effective information in forest fire prevention and control. Therefore, this study aims to use machine learning techniques for predicting the main factors causing forest fires in Pekanbaru, Riau and predicting based on the new weather data.

Several techniques on forest fire prediction using machine learning techniques have been developed. Aerial images were used for forest fire prediction based on forest resource surveys [5]. Machine learning mixes informatics and statistical analysis to improve prediction, hence widely used to solve uncertainty problems. Machine learning has many effective algorithms e.g. Artificial Neural Network (ANN), Bayesian Network, Decision Tree (DT), Rough set Theory (RsT), Linear Regression (LR) and Support Vector Machine (SVM) [6–8]. However, there is no algorithm that works best with the 'No Free Lunch' theorem for every problem [8]. To predict forest fires, the ANN algorithm was used. A multilayer perceptron whose structural parameters have been heuristically classified was selected i.e. the number of hidden layers and the total of neurons per layer [9]. Artificial Neural Network has also been utilized to predict forest fires. Weather data for the year 2012 alone was used to predict forest fires in Lebanon [4]. Wind data and WindNinja as wind field simulator were applied for accurate prediction of Mediterranean basin forest fire [10]. The Genetic Algorithm as the calibration technique was applied for the accurate prediction of forest fire propagation [11]. Genetic algorithms have also been used as a real case of study for forest fires [12].

Based on literature reviews, several algorithms and techniques were employed to predict forest fires e.g. Artificial Neural Networks and Genetic Algorithms. The most serious problem for predicting forest fire is input parameter data used. Data significantly influences the accuracy of prediction results. The related research uses a year data namely in 2012 [4]. The previous studies using ANN and GA algorithms. Although the ANN algorithm is complicated and requires a long computational time. Genetic algorithms require several simulations to be performed. The eventual long executions were added to this fact. The prediction has to be made much quicker than it is in real-time. Therefore, all available computing resources need to be exploited. [12]. The uncertainty problem in predictions is also crucial. Predicting forest fires is needed in two schemes such as the main factors affecting forest fires occurred and forest fires prediction. Therefore, this study uses more complete weather data and two algorithms that are known to be accurate e.g. Decision Tree (DT) and Bayesian Network (BN), for predicting forest fires in Riau Indonesia.

Also known as the Belief Network, Bayesian Belief Network, Bayes Nets or called the Probabilistic Network is the Bayesian network. The Bayesian network is a method created by Thomas Bayes in 1763 based on the Bayes theorem. Over the past decade, the Bayesian Network approach has become very popular for a number of smart applications such as machine learning, text processing, bioinformatics processing, medical diagnostics, weather forecasts and other smart device applications [13]. Modeling techniques of the Bayesian Network have several features that make this method useful in many issues of data analysis and management. Even with a small amount of data, the Bayesian network can also show good prediction accuracy [13]. Often useful in integrating a variety of information is the Bayesian Network Approach.

Decision Tree is a simple algorithm but has good accuracy. Decision Tree is recommended to be used for large amounts of data because it does not require a long computational time [8]. In addition, Decision Tree also has a tree diagram that allows researchers to analyze the attributes affecting the class.
2. Material and methods
We have obtained the weather data from IBM Cloud [14] from January 2015 to September 2019 with a total 1733. The data type consists of date, integer, real and polynomial. Figure 1 briefly explains the methodology that has been implemented.

![Diagram of methodology](image)

Figure 1. The methodology of machine learning technique selection to predict forest fire.

2.1. Data pre-processing
To create a good quality dataset, we have pre-processed the data. Pre-processing involves multiple processes such as integration, selection, cleaning, reduction, and transformation. There are no missing values in the weather dataset, but the pre-processing of the data is completed by creating an interval at the hot spot attribute e.g. none, low, mid and high.

2.1.1. Data integration. In this study, data integration is completed to make the data into a whole file. Weather data consists of monthly. Hence, it is necessary to integrate the data into a whole file. Data integration also makes it easy to conduct experiments on each algorithm used.

2.1.2. Data cleaning. This step requires cleaning data from incorrect writing and formatting. Data cleaning refers to finding incomplete, wrong, inaccurate or irrelevant data components. We simply change the incorrect format in weather data to deliver the correct analysis.

2.1.3. Data reduction. This step needs reducing the raw data into a more useful form. We did not reduce the weather data because the data obtained are already in a useful form for analysis.

2.1.4. Data transformation. This stage requires transforming data for changing the scale of measurement of original data into other forms so that the analysis tool can read weather data.

2.2. Classification model
In this study, we have used two algorithms, Decision Tree and Bayesian Network. Let $U = \{ \text{Time, Temperature Max}, \text{Temperature Avg}, \text{Temperature Min}, \text{Dew Point Max}, \text{Dew Point Avg}, \text{Dew Point Min}, \text{Humidity Max}, \text{Humidity Min}, \text{Wind Speed Max}, \text{Wind Speed Min}, \text{Pressure Max}, \text{Pressure Min}, \text{Precipitation Avg}, \text{Hotspot} \}$ be a set of variables. The Bayesian network $B$ is a network structure over a set of variables $U$, which is a Directed Acyclic Graph (DAG) over $U$ and a set of tables of probability $B_P = \{ p(u|pa(u)) | u \in U \}$, which is $p(u)$ is the organization of $u$ in structure [11]. A Bayesian network is a distribution of probabilities $P(U) = \prod_{u \in U} p(u|pa(u))$. The classifier has learned from a sample set of weather data (attributes, hotspot). The training task is to find a suitable Bayesian network with a collection of data $D$ over $U$. Figure 2 demonstrates the likelihood of the Bayesian Network structure for each node.
Figure 2. Bayesian Network’s simple structure to predict forest fire.

Meanwhile, the Decision Tree algorithm has been run as Algorithm 1:

| Input: Weathers (Time, Temperature Max, Temperature Avg, Temperature Min, Dew Point Max, Dew Point Avg, Dew Point Min, Humidity Max, Humidity Min, Wind Speed Max, Wind Speed Min, Pressure Max, Pressure Min, Precipitation Avg, Hotspot) |
| Output: results |
| Method: |
| 1. Begin the Weathers (Time, Temperature Max, Temperature Avg, Temperature Min, Dew Point Max, Dew Point Avg, Dew Point Min, Humidity Max, Humidity Min, Wind Speed Max, Wind Speed Min, Pressure Max, Pressure Min, Precipitation Avg, Hotspot) at the tree root. |
| 2. Divide the training into sub-sets. Subsets should be made to contain Weathers with the same value for an attribute for each sub-set. |
| 3. Repeat step 1 and step 2 on each sub-set until all branches of the tree have leaf nodes (results); End. |

2.3. Testing and evaluation

For classification modeling, each experiment was performed using the data set split factor (training data: test data) from 90 to 10 [8]. From each experiment, these 9 different models are obtained. The parameters to test each modeling’s output as follows: accuracy Root Mean Square Error (RMSE) and confusion matrix.

3. Result and discussion

We have successfully used Decision Tree and Bayesian Network algorithms to predict forest fires in Pekanbaru, Riau. The percentage of weather data was arranged from the data training 90% to 10%. Decision Tree algorithm has produced good accuracy of 93.18% and Root Mean Square Error 0.244. In nine experiments, the best accuracy is in the first experiment when the training data 90%.

Decision Tree has produced good accuracy, but in nine experiments, Bayesian Network is more accurate. Bayesian Network produced the best accuracy of 99.62% and Root Mean Square Error 0.076. In nine experiments, the best accuracy is in the second experiment when the training data 80%. Table 1
shows classification accuracy for the Decision Tree and Bayesian Network algorithm in nine experiments.

Table 1. Classification Accuracy of Decision Tree and Bayesian Network.

| Experiments | Training Dataset Percentage | Decision Tree | Bayesian Network |
|-------------|-----------------------------|---------------|------------------|
|             |                             | Accuracy | RMSE | Accuracy | RMSE |
| 1           | 90%                         | 93.18%   | 0.244 | 99.24%   | 0.090 |
| 2           | 80%                         | 88.35%   | 0.326 | 99.62%   | 0.076 |
| 3           | 70%                         | 88.19%   | 0.324 | 99.50%   | 0.088 |
| 4           | 60%                         | 85.53%   | 0.386 | 99.06%   | 0.115 |
| 5           | 50%                         | 92.17%   | 0.258 | 97.29%   | 0.171 |
| 6           | 40%                         | 86.18%   | 0.344 | 98.24%   | 0.151 |
| 7           | 30%                         | 83.76%   | 0.405 | 97.13%   | 0.185 |
| 8           | 20%                         | 85.79%   | 0.359 | 97.18%   | 0.204 |
| 9           | 10%                         | 84.45%   | 0.365 | 65.47%   | 0.559 |
| **Average** |                             | 87.51%   | 0.335 | 94.75%   | 0.182 |

Decision Tree explained that were four main factors that caused the forest and land fires, namely temperature, humidity, wind speed, and rainfall. Forest and land fires often occur in Pekanbaru Riau when:

- The temperature min is above 23.9 °C
- The humidity max is below 97%
- Wind speed max is above 10 km/h
- Precipitation average is below 0.02 (in)

In addition, there were two additional factors caused forest and land fires in Riau such as:

- The pressure max is above 29.84 Hg.
- The maximum dew point max is below 23.3 °C.

Based on the confusion matrix of the Bayesian Network algorithm table is obtained that the best class precision is in the *none, low and high* class 100% and the best class recall is in the *none, low and mid-* class 100%. Table 2 shows the confusion matrix of the Bayesian Network algorithm.

Table 2. The Bayesian Network's confusion matrix.

| BN            | True None | True Low | True Mid | True High | Class Precision |
|---------------|-----------|----------|----------|-----------|-----------------|
| Pred. None    | 55        | 0        | 0        | 0         | 100%            |
| Pred. Low     | 0         | 159      | 0        | 0         | 100%            |
| Pred. Mid     | 0         | 0        | 29       | 1         | 96.67%          |
| Pred. High    | 0         | 0        | 0        | 22        | 100%            |
| **Class Recall** | **100%** | **100%** | **100%** | **95.65%** |                 |

Next experiments have a detailed comparison between Decision Tree and Bayesian Network in terms of accuracy for each training data allocation. Bayesian Network is more accurate in eight experiments. Whereas Decision Tree is only accurate in the last experiment. Bayesian Network has a higher average accuracy of 94.75% than Decision Tree 87.51%. Figure 3. (a) shows the comparative accuracy of Decision Tree and Bayesian networks in training data allocation.
Root Mean Square Error (RMSE) is an alternative evaluation to measure the accuracy of a model's prediction results. We conclude that how many error rates generated by a Decision Tree and Bayesian Network algorithm using RMSE. Bayesian Network obtained high error in the last experiment. Figure 3. (b) shows the RMSE rate of Decision Tree and Bayesian Network based on training data allocation. Based on the tests and evaluations stage, we concluded that the Bayesian Network algorithm is more accurate than the Decision Tree algorithm. Hence, we can use this model to predict forest fires in new data. We have also attempted to make predictions using the new data. The new data were obtained from already weather forecasting results. Based on the results for 2020, will be there forest and land fires, which are April - September, especially in August and September.

4. Conclusion
We have successfully predicted forest fires in Riau Indonesia using Decision Tree and Bayesian Network algorithms. The algorithms provide good accuracy and have a low error rate. In accordance with the ‘No Free Lunch Theorem,’ this study has shown that each algorithm has its own advantages. The Decision Tree algorithm has a lower accuracy than the BN algorithm. However, it has advantages in explaining the relationship of each attribute of weather data that affects the hotspot level using the tree diagram. Whereas the Bayesian Network has advantages in terms of accuracy. Thus, it can be concluded that the prediction model using Bayesian Network has the potential to be used effectively but still has plenty of room for improvement.

This study is still a basic experimental stage. We still need additional experiments such as using Deep Learning and forecasting algorithm. We also need data from other IoT sensors to produce more data that are accurate. Environmental experts are also needed to interpret predictive data.

References
[1] Satoh, K., Song Weiguo, and Yang, K. T., 2004, “A Study of Forest Fire Danger Prediction System in Japan,” Proceedings. 15th International Workshop on Database and Expert Systems Applications, 2004., IEEE, pp. 598–602.
[2] Ministry of Environment and Forestry, 2019, “SiPongi - Karhutla Monitoring Sistem,” Data Matrik Titik Panas TERRA/AQUA ≥ 80% Tahun 2019 [Online]. Available: http://sipongi.menlhk.go.id/hotspot/matrik_tahunan?satelit=LPN-MODIS&thn=2019. [Accessed: 15-Oct-2019].
[3] Indonesian Agency for Meteorological, Climatological, and G., 2019, “Informasi Konsentrasi Partikulat (PM10) Pekanbaru | BMKG,” Inf. Konsentrasi Partikulat Kota Pekanbaru [Online]. Available: https://www.bmkg.go.id/kualitas-udara/informasi-partikulat-pm10.bmkg?Lokasi=PEKANBARU. [Accessed: 15-Oct-2019].

[4] Hamadeh, N., Hilal, A., Daya, B., and Chauvet, P., 2015, “Studying the Factors Affecting the Risk of Forest Fire Occurrence and Applying Neural Networks for Prediction,” 2015 SAI Intelligent Systems Conference (IntelliSys), IEEE, pp. 522–526.

[5] Gao, X., Fei, X., and Xie, H., 2011, “Forest Fire Risk Zone Evaluation Based on High Spatial Resolution RS Image in Liangyungang Huaguo Mountain Scenic Spot,” Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, IEEE, pp. 593–596.

[6] Aakash, R. S., Nishanth, M., Rajageethan, R., Rao, R., and Ezhlarasie, R., 2018, “Data Mining Approach to Predict Forest Fire Using Fog Computing,” 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, pp. 1582–1587.

[7] Kavitha S, Varuna S, and Ramya R, 2016, “A Comparative Analysis on Linear Regression and Support Vector Regression,” 2016 Online International Conference on Green Engineering and Technologies (IC-GET), IEEE, pp. 1–5.

[8] Kurniawan, R., Nazri, M. Z. A., and Irsyad, M., 2015, “On Machine Learning Technique Selection for Classification,” 2015 International Conference on Electrical Engineering and Informatics (ICEEI), IEEE, pp. 540–545.

[9] Safi, Y., Bouroumi, A., and Bouroumi, A., 2011, “A Neural Network Approach for Predicting Forest Fires,” 2011 International Conference on Multimedia Computing and Systems, IEEE, pp. 1–5.

[10] Sanjuan, G., Brun, C., Margalef, T., and Cortés, A., 2014, “Wind Field Uncertainty in Forest Fire Propagation Prediction,” Procedia Comput. Sci., 29, pp. 1535–1545.

[11] Artés, T., Cencerrado, A., Cortés, A., and Margalef, T., 2013, “Relieving the Effects of Uncertainty in Forest Fire Spread Prediction by Hybrid MPI-OpenMP Parallel Strategies,” Procedia Comput. Sci., 18, pp. 2278–2287.

[12] Cencerrado, A., Cortés, A., and Margalef, T., 2012, “Genetic Algorithm Characterization for the Quality Assessment of Forest Fire Spread Prediction,” Procedia Comput. Sci., 9, pp. 312–320.

[13] Akbarizan, Kurniawan, R., Nazri, M. Z. A., Abdullah, S. N. H. S., Murhayati, S., and Nurcahaya, 2018, “Using Bayesian Network for Determining The Recipient of Zakat in BAZNAS Pekanbaru,” 2018 2nd International Conference on Electrical Engineering and Informatics (Icon EEI), IEEE, pp. 12–17.

[14] IBM Cloud, 2019, “Pekanbaru, Indonesia History | Weather Underground” [Online]. Available: https://www.wunderground.com/history/monthly/idpekabarawibb/ [Accessed: 15-Oct-2019].