

Analisis Prediksi Kebakaran Hutan dengan Menggunakan Algoritma *Random Forest Classifier*

**Dede Husen¹, Dede Sandi², Sepriadi Bumbungan³,
Kusrini⁴, Kusnawi⁵**

¹Universitas Amikom Yogyakarta

^{2,3}Magister Teknik Informatika

E-mail: *¹dede@students.amikom.ac.id, ²dede.sandi@students.amikom.ac.id,
³sepriadibumbungan@students.amikom.ac.id, ⁴kusnawi@amikom.ac.id,
⁵kusrini@amikom.ac.id

Abstrak

Kebakaran hutan merupakan salah satu bencana yang sangat merugikan di dunia, tak terkecuali di Indonesia. Berdasarkan laporan dari Kementerian Lingkungan Hidup dan Kehutanan total kebakaran hutan dan lahan dalam rentang 2015 – 2019 yang terbakar adalah seluas 1.6 juta (Ha) [1]. Beberapa faktor yang mempengaruhi terjadinya kebakaran hutan diantaranya adalah faktor alam dan manusia. Faktor alam seperti kondisi suhu, kelembapan, kemarau, Elnino, erupsi gunung dan petir, kemudian para peneliti menemukan fakta bahwa aktivitas manusia di hutan seperti pembukaan lahan, eksploitasi kayu, perburuan dan pembakaran memiliki efek kausalitas terhadap terjadinya kebakaran hutan khususnya di daerah yang masih mempunyai hutan yang luas. Beberapa penelitian telah dilakukan untuk mencegah terjadinya kebakaran hutan seperti dengan menggunakan Teknik Data Mining dan Machine Learning yakni dengan melakukan prediksi kapan terjadinya kebakaran hutan berdasarkan kondisi cuaca dan histori laporan kebakaran namun masih belum sempurna. Maka dari itu pada penelitian ini kami mengembangkan konsep sistem prediksi kebakaran hutan yang akan menjadi salah satu acuan kebijakan pemerintah dalam mengeluarkan kebijakan yang bersifat preventif. Dengan melakukan pemodelan menggunakan model Algoritma Random Forest pada data kebakaran hutan dari tahun ketahun di wilayah Indonesia diharapkan dapat membantu pemerintah dalam melakukan pencegahan kebakaran hutan dengan kebijakan hukumnya dan analisis yang ada bisa digunakan oleh Balai Besar Teknologi Modifikasi Cuaca (BBTMC) yang dapat membantu menentukan kapan modifikasi cuaca dilakukan.

Kata Kunci : Kebakaran Hutan, Data Mining, Random Forest

Abstract

Forest fire is one of the most devastating disasters in the world, and Indonesia is no exception. Based on a report from the Ministry of Environment and Forestry, the total forest and land fires in the 2015 – 2019 range that burned were 1.6 million (Ha) [1]. Several factors that influence the occurrence of forest fires include natural and human factors. Natural factors such as temperature conditions, humidity, drought, Elnino, mountain eruptions and lightning, then the researchers found the fact that human activities in the forest such as land clearing, timber exploitation, hunting and burning have a causal effect on the occurrence of forest fires, especially in areas that still have forest fires.. Several studies have been conducted to prevent forest fires, such as by using Data Mining and Machine Learning techniques, namely by predicting when forest fires will occur based on weather conditions and historical reports of fires, but they are still not perfect. Therefore, in this study, we developed the concept of a forest fire prediction system which will become one of the government's policy references in issuing preventive policies. By doing modeling using the Random Forest Algorithm model on forest fire data from year to year in the territory of Indonesia, it is hoped that it can help the government in preventing forest fires with

its legal policies and the existing analysis can be used by the Center for Weather Modification Technology (BBTMC) which can help determine when to modify weather done.

Keyword : Forest Fire, Data Mining, Random Forest

1. PENDAHULUAN

Kebakaran hutan dan lahan di Indonesia telah menjadi perhatian dunia internasional khususnya sejak kebakaran hutan yang terjadi pada tahun 80-an [2]. Penyebab kebakaran dan lahan lebih disebabkan oleh ulah manusia dibandingkan proses alam [3]. Berbagai penelitian telah banyak dilakukan di Indonesia salah satu diantaranya adalah yang dilakukan oleh Ariesta Lestari [4] dalam penelitiannya yang berjudul Deteksi Dini Kebakaran Hutan dan lahan di Kalimantan Tengah, dimana hasil penelitian ini menjelaskan faktor-faktor yang mempengaruhi terjadinya kebakaran hutan diantaranya adalah faktor alam dan faktor manusia, faktor alam disini dapat bencana alam, fenomena El-nino, dan rendahnya curah hujan, selain itu faktor manusia juga menjadi salah satu faktor penyebabnya. Dalam penelitian tersebut dijelaskan bahwa sudah banyak berbagai metode *data mining* telah diterapkan untuk mengembangkan model yang dapat melakukan prediksi atas munculnya kebakaran hutan. Tidak hanya melakukan prediksi terhadap kemungkinan terjadinya kebakaran hutan tetapi juga memberikan estimasi besarnya terjadinya kebakaran hutan yang akan terjadi dan mendeteksi tingkat kebenaran *hotspot* yang terdeteksi oleh satelit. *Logistic regression* dan *decision tree* merupakan metode paling sering di implementasikan untuk mengestimasi dan mengklasifikasi kemungkinan terjadinya kebakaran hutan di suatu wilayah.

Kemudian penelitian yang dilakukan oleh Dlamini [5] mencoba membuat konsep diagram menggunakan *Bayesian Network*. Dengan diagram ini dapat terlihat faktor mana saja yang mempunyai pengaruh paling besar dalam kebakaran hutan dan bisa memprediksi perubahan yang akan terjadi jika kondisi dalam faktor tersebut mengalami perubahan. Sebagian besar hasil riset mengenai penanganan kebakaran hutan menggunakan metode data mining mengambil lokasi di Eropa, Amerika atau Australia. Dimana tidak semua hasil penelitian mengenai kebakaran hutan di negara-negara lain dapat dikembangkan dan diaplikasikan di Indonesia. Hal ini dikarenakan karakteristik kebakaran hutan di Indonesia yang cukup unik dan berbeda dibandingkan dengan negara lain. Tidak banyak riset yang mengadopsi fungsi *data mining* untuk menangani masalah kebakaran hutan di Indonesia terutama di Kalimantan Tengah. [6] mengimplementasikan *decision tree* untuk mengidentifikasi keberadaan *hotspot* di Riau. Vayda [7] memberikan argumen bahwa tidak semua *hotspot* yang terdeteksi akan menyebabkan terjadinya kebakaran hutan. Oleh karena itu riset yang mereka lakukan akan fokus pada pengidentifikasian *hotspot* yang berpotensi menyebabkan kebakaran hutan dengan mengakomodasi semua karakteristik kebakaran hutan di Kalimantan Tengah.

Pada penelitian ini penulis membuat pemodelan dengan *Algoritma Random Forest* dimana tingkat akurasi dari algoritma ini cukup tinggi. Hasil model yang dibuat kami harap dapat membantu pemerintah dalam melakukan

tindakan preventive kebakaran hutan dimasa mendatang

dan *date* dengan total 6417 baris data dari tahun 1999 sampai dengan tahun 2018.

2. METODE PENELITIAN

2.1. Tipe Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan pemodelan data data analisis pada laporan jumlah kebakaran dari tahun 1998 sampai tahun 2018 di Indonesia. Yang kemudian dilakukan pemodelan dalam data science dengan menggunakan model *Random Forest Classifier*.

2.2. Metode Pengumpulan Data

Pengumpulan data dilakukan dari berbagai sumber dengan menyusun dataset secara dengan dimana data yang terkumpul adalah data *time series* yang isinya adalah laporan data kebakaran berkala dari tahun 1998 sampai dengan tahun 2017.

2.3. Teknik Analisa Data

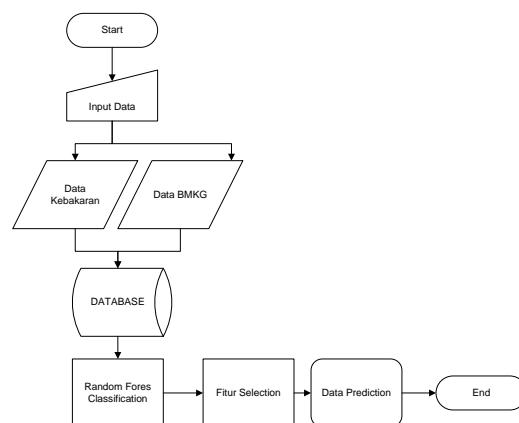
Analisa data adalah proses telaah dan pencarian makna dari data yang diperoleh untuk menemukan jawaban dari masalah penelitian. Analisis data yang dilakukan meliputi *Data Cleansing* (Pembersihan Data), *Data Reduction* (Pengurangan Data), *Data Transformation*, *Data Integration*. Tahapan – tahapan tersebut adalah salah satu tahapan penting dalam pemodelan data.

Setelah data dapat di Analisa maka selanjutnya adalah tahapan menganalisisnya dengan menggunakan Algoritma *Random Forest*. Algoritma ini merupakan sebuah *Classifier* dengan *ensemble method* yang terdiri dari beberapa pohon keputusan (*decision tree*) yang membentuk suatu hutan “forest” secara acak (*random*).

Data yang di uji mempunyai 5 fitur diantaranya fitur *year, state, month, number,*

2.4. Desain Model

Berikut desain model / flowchart aplikasi yang akan kami rancang



Gambar 1. Flowchart Alur Aplikasi

Flowchart diatas adalah tahapan dari mulai inputan data hingga ketahapan prediksi data, dimana data yang diolah adalah data dari BMKG dan data dari berbagai pelaporan kebakaran di Indonesia.

2.5. Tahapan Pengujian

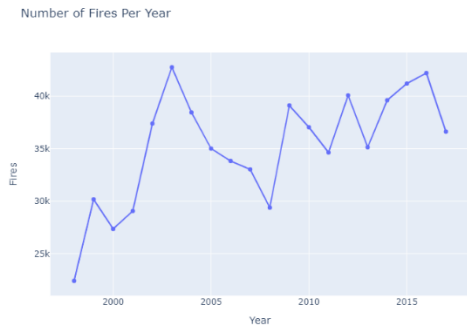
Pada tahapan ini kami menggunakan Jupiter notebook (pyhton) untuk tahapan pemodelan dan pengujianya. Dimulai dari pemilihan library-library yang akan di butuhkan, selanjutnya kami membuat visualisasi data serta melakukan perbandingan pengujian dengan 2 algoritma yaitu Algoritma *K-Neighbors Classifier* dan *Random Forest Classifie*.

3. HASIL DAN PEMBAHASAN

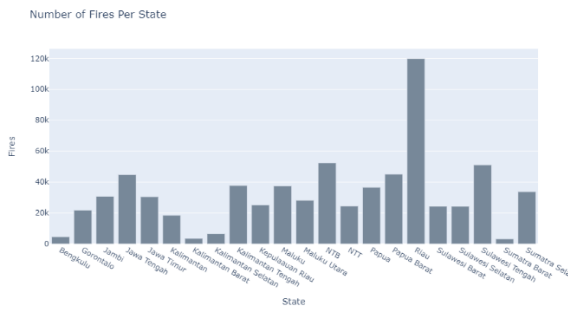
Berikut kami paparkan hasil dari penelitian kami melalui beberapa poin-poin penjelasan berikut :

3.1. Visualisasi Data

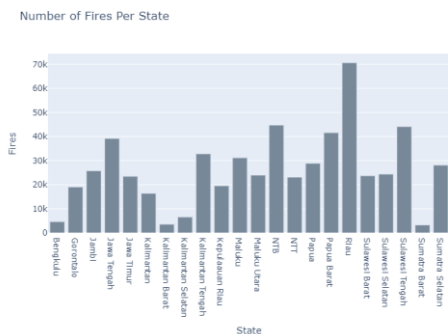
Untuk memahami data apa yang akan kita olah maka sebaiknya kita memahami dulu data-data yang ada dengan memvisualisasikannya kedalam diagram, statistik ataupun lainnya. Berikut kami tampilkan beberapa visualisasi data yang kami olah :



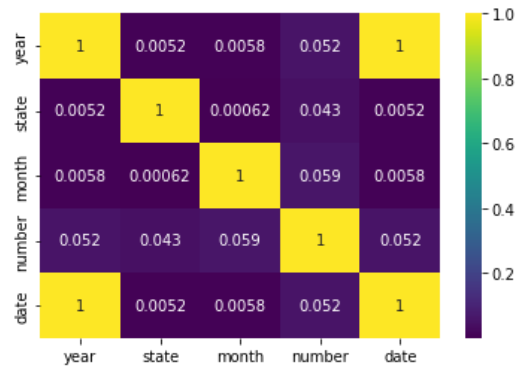
Gambar 2. Laporan kebakaran hutan dari tahun ke tahun



Gambar 3. Grafik Laporan Kebaran Hutan Perbulan



Gambar 4. Grafik Laporan Berdasarkan Daerah



Gambar 5. Heatmap (korelasi antar fitur data)

3.2. Model Pengujian

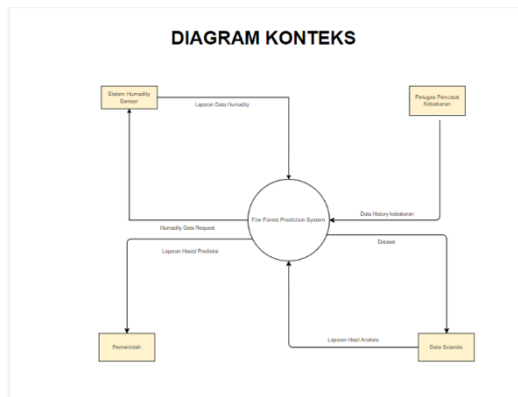
Pada pengujian ini kami menggunakan *jupyter notebook* untuk membantu mengeksekusi perintah-perintah python di browser. kami melakukan pengujian dengan komposisi data training sebanyak 70% dan data testing sebanyak 30% dengan hasil pengukuran performance model dengan *confusion matrix* dengan hasil pada Algoritma *K-NeighborsClassifier* mempunyai tingkat akurasi 0.47 (47%) Nilai *Precision* 0.26, nilai *Recall* 0.26 dan nilai *F-1 Score* sebesar 0.26. sementara itu pengujian yang sama dengan menggunakan *Random Forest Classifier* memiliki nilai *Accuracy* sebesar 1.0 (100%), nilai *Precision* 1.0, nilai *Recall* 1.0 dan *F-1 Score* sebesar 1.0. Berikut kami sajikan dalam bentuk tabel hasil pengujian *Confusion Matrix* dari kedua Algoritma tersebut :

<i>Confusion Matrix</i>	Random Forest	KNN
<i>Accuracy</i>	1.0	0.47
<i>Precision</i>	1.0	0.26
<i>Recall</i>	1.0	0.26
<i>F-1 Score</i>	1.0	0.26

Table 1. Score Confusion Matrix

Dari table tersebut kita dapat menyimpulkan bahwa Algoritma *Random forest* merupakan model yang mempunyai tingkat keakurasi yang tinggi sampai dengan 100%

3.3. Desain Perancangan Sistem



Gambar 5. Diagram Konteks

Gambar diatas merupakan desain rancangan system yang akan dibangun dimana sistem tersebut menggambarkan alur dari aliran data. Pada diagram tersebut terdapat 4 entitas yang berperan yaitu Pemerintah, Data Scienties, Sistem Humadity Sensor, petugas Pencatat kebakaran.

4. KESIMPULAN

Berdasarkan hasil pengujian kami, penulis dapat menyimpulkan beberapa hal sebagai berikut :

1. Berdasarkan analisis prediksi menggunakan *Algoritma Random Forest Classifier* mempunyai tingkat akurasi yang tinggi dibanding dengan menggunakan *Algoritma K-Neighbors*
2. Dengan adanya sistem yang dapat memprediksi deteksi kebakaran hutan diharapkan dapat meningkatkan upaya pencegahan dengan memadukan antara teknologi dan kebijakan hukum yang dapat melindungi hutan kita.
3. Faktor kebakaran hutan di sisi manusia mempunyai andil yang cukup tinggi, dikarenakan masih lemahnya tindakan hukum baik Tindakan preventive maupun Tindakan pelanggaran.

5. SARAN

Penulis menyadari bahwa masih banyak kekurangan dalam penelitian ini, untuk penelitian lebih lanjut diperlukan data yang lebih banyak lagi untuk meningkatkan akurasi prediksi, kemudian dalam pengembangan sistem diharapkan kedepannya dapat di terapkan secara nyata baik dalam bentuk prototype ataupun bentuk pengujian lainya yang lebih nyata dilapangan serta perlunya andil pemerintah dan Lembaga terkait untuk mendukung kelangsungan penelitian ini.

DAFTAR PUSTAKA

- [1] Endrawati, "Analisa Data Luas Areal Kebakaran Hutan dan Lahan Tahun 2019," dalam *Seminar Nasional Geomatika*, Jakarta, 2019.
- [2] P. A. W. Z. D. d. A. B. Erna S. Adiningsih1, "OVEMENT OF LAND AND FOREST FIRE HAZARD MAPPING METHOD FOR SUMATERA AND KALIMANTAN BASED ON REMOTE SENSING DATA," dalam *Pertemuan Ilmiah Tahunan MAPIN XIV*, Surabaya, 2005.
- [3] W. I. S. B. S. d. L. S. Adinugroho, "Panduan Pengendalian Kebakaran Hutan dan Lahan Gambut. Proyek Climate Change, Forests and Peatlands in Indonesia. Wetlands International-Indonesia Programme dan Wildlife Habitat Canada.," Wetlands International - Indonesia Program, Bogor, 2005.
- [4] D. G. R. P. N. T. Ariesta Lestari1, "DETEKSI DINI KEBAKARAN HUTAN DAN LAHAN DI KALIMANTAN TENGAH," *Jurnal Dialog Penanggulangan Bencana*, vol. II, no. 6, pp. 128-134, 2015.
- [5] W. Dlamini, "Application of Bayesian Networks for fire risk mapping using

- GIS and Remote Sensing Data,”
GeoJournal, vol. III, pp. 283-296, 2011.
- [6] I. S. Y. R. M. N. A. A. N. Sitanggang,
“Predictive models for hotspots
occurrence using decision tree
algorithms and logistic regression.”
Journal of Applied Sciences, vol. II, no.
2, pp. 252 - 261, 2013.
- [7] Vayda, “Finding Causes of the 1997 -
1998 Indonesian forest fires problem
and possibilities WWF Indonesia Forest
Fires Project.” World Wide Fund for
Nature (WWF) Indonesia, Jakarta ,
1999.
- [8] E. P. W. .. D. d. A. B. Adiningsih,
“Improvement of Land Forest Fire
Hazard Mapping Methodd for Sumatera
and Kalimantan Based on Remote
Sensing Data.” dalam *Pemanfaatan
Efektif Penginderaan Jauh untuk
Peningkatan Kesejahteraan Bangsa.
Prosiding Pertemuan Ilmiah Tahunan
(PIT) MAPIN XIV.*, Surabaya, 2005.