

# Prediksi Area Kebakaran Hutan Menggunakan Algoritma *Random Forest*

Liana Trihardianingsih<sup>1\*</sup>, Hanifah Permatasari<sup>2</sup>

<sup>1</sup>Sistem Informasi Fakultas Ilmu Komputer  
Universitas PGRI Jombang  
<sup>\*</sup>liana@upjb.ac.id

<sup>2</sup>Sistem Informasi Fakultas Ilmu Komputer  
Universitas Duta Bangsa Surakarta  
<sup>3</sup>hanifah\_permatasari@udb.ac.id

**Abstrak**— Kebakaran hutan adalah bencana yang memiliki dampak serius bagi lingkungan, merusak sumber daya hutan, mengancam ekosistem secara keseluruhan, termasuk flora, fauna, dan kualitas lingkungan di suatu wilayah, bahkan membahayakan nyawa manusia. Meskipun sering kali dipicu oleh kondisi lingkungan yang kering dan panas, aktivitas manusia seperti pembukaan lahan pertanian juga dapat menjadi penyebabnya. Variabel seperti tingkat kelembaban, kecepatan angin, dan curah hujan berperan penting dalam menentukan seberapa luasnya kebakaran hutan dapat menyebar. Dalam penelitian ini, *Random forest* digunakan untuk memprediksi luas area yang terkena kebakaran. Dengan nilai  $n = 1000$ , model ini menghasilkan RMSE sebesar 0.557, MSE sebesar 0.310, dan R2 Score sebesar 0.844 dengan penggunaan 12 atribut, sedangkan dengan 8 atribut, model menunjukkan RMSE sebesar 0.548, MSE sebesar 0.301, dan R2 Score sebesar 0.849. Hasil analisis juga menunjukkan bahwa atribut yang memiliki pengaruh paling penting dalam prediksi kebakaran hutan adalah temperatur. Hal ini menegaskan pentingnya temperatur dalam mempengaruhi kebakaran hutan dan memperkuat keandalan model dalam memberikan prediksi yang akurat dan informatif untuk mitigasi risiko kebakaran hutan.

**Kata kunci**— Kebakaran, Prediksi, *Random forest*, RMSE, MSE

**Abstract**— Forest fires pose significant threats to the environment, causing extensive damage to forest resources and endangering ecosystems, including flora, fauna, and environmental quality in specific regions, and potentially risking human lives. While typically triggered by dry, hot environmental conditions, human activities such as land clearing for agriculture can also be contributing factors. Variables like humidity levels, wind speed, and rainfall are pivotal in determining the extent of forest fire spread. This study utilizes *Random forest* for predicting fire-affected areas. Using a value of  $n = 1000$ , the model yields an RMSE of 0.557, MSE of 0.310, and R2 Score of 0.844 with 12 attributes, and with 8 attributes, it shows an RMSE of 0.548, MSE of 0.301, and R2 Score of 0.849. Analysis further identifies temperature as the most influential attribute in predicting forest fires, underscoring its critical role and enhancing the model's accuracy and utility in providing precise insights for forest fire risk mitigation.

**Keywords**— Forest Fire, Prediction, *Random forest*, RMSE, MSE.

## I. PENDAHULUAN

Salah satu permasalahan lingkungan yang menimbulkan dampak yang cukup besar untuk lingkungan adalah kebakaran hutan [1]. Sebagai negara tropis yang memiliki luas hutan sebesar 95,6 juta hektar pada tahun 2020 dan merupakan terluas ketiga di dunia, terjadinya kebakaran hutan di Indonesia terutama di wilayah Kalimantan dan Sumatera [2] bukanlah sebuah fenomena yang langka. Kerusakan yang ditimbulkan dari adanya kebakaran hutan tidak hanya mengganggu ekosistem hutan, flora, dan fauna yang ada di sekitar hutan tapi juga dapat menyerang kesehatan warga yang tinggal di sekitar hutan karena akan terpapar secara langsung oleh asap yang dihasilkan ketika kebakaran hutan terjadi.

Proses deforestasi di Indonesia seringkali dilakukan dengan cara membakar hutan untuk membuka lahan dan mengelola lahan [3]. Kegiatan itulah yang menjadi pemicu terbesar terjadinya

kebakaran. Namun, meski kebanyakan kebakaran hutan di Indonesia disebabkan oleh aktivitas manusia, penyebab terjadinya kebakaran juga diduga berasal dari lingkungan yang kering dan panas [4]. Sehingga, diperlukan data data yang dapat dikelola untuk membantu mengantisipasi, mengontrol, dan mencegah terjadinya kebakaran hutan. Data-data yang bisa dimanfaatkan diantaranya adalah tingkat kelembaban, kecepatan angin, dan curah hujan [5].

Salah satu cara untuk mengetahui kondisi kebakaran hutan adalah mempelajari indeks cuaca hutan, seperti Forest Weather Index (FWI). Cara lain yang dapat digunakan adalah dengan mempertimbangkan faktor-faktor seperti suhu dan angin yang mempengaruhi lahan dan kebakaran hutan [6]. Pada penelitian terdahulu yang dilakukan oleh [7], dengan membandingkan tiga algoritma untuk memprediksi area kebakaran menunjukkan bahwa algoritma Linear Regression (LR) memiliki nilai RMSE tertinggi sebesar 1.496, algoritma K-

Nearest Neighbor (K-NN) memiliki nilai RMSE tertinggi kedua sebesar 1.981, dan algoritma Support Vector Machine (SVM) memiliki nilai RMSE tertinggi ketiga sebesar 3.088.

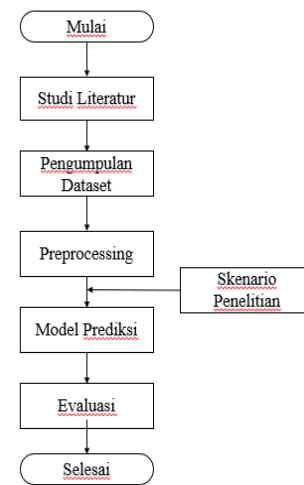
Penelitian lain juga dilakukan [8], dengan menggunakan Naïve Bayes dan KNN untuk menghitung akurasi dari prediksi kebakaran hutan menggunakan 10 variabel. Hasil penelitian menyatakan bahwa Hasil yang diperoleh menunjukkan bahwa metode Naive Bayes memiliki nilai akurasi sebesar 53,33% dan K-NN memiliki nilai akurasi sebesar 62,66%.

Penelitian lain [9] menggunakan metode Decision Tree C4.5 untuk mendeteksi kebakaran hutan. Penelitian ini memanfaatkan beberapa variabel seperti bulan, kecepatan angin, curah hujan, dan kelembaban udara. Dengan menerapkan lima skenario yang mengkombinasikan variabel-variabel tersebut, penelitian ini berhasil memprediksi terjadinya kebakaran hutan dengan akurasi mencapai 92%. Data yang digunakan dalam penelitian ini terdiri dari 80% data latih dan 20% data uji

Berdasarkan pemaparan di atas terkait kondisi kebakaran yang terjadi di Indonesia, peneliti akan memanfaatkan data-data yang diperoleh secara publik terkait Forest Weather Index, untuk memprediksi area kebakaran hutan. Algoritma yang akan digunakan pada penelitian ini adalah Random forest. Random forest merupakan sebuah algoritma yang terdiri dari gabungan banyak pohon keputusan yang dibangun dalam satu dataset dan digunakan untuk proses hitung berulang untuk menghasilkan hasil perhitungan yang maksimal [10]. Sehingga, diharapkan dengan menggunakan algoritma Random forest akan diperoleh nilai error yang lebih rendah dan mampu memprediksi area kebakaran yang lebih akurat.

## II. METODOLOGI PENELITIAN

Setiap penelitian selalu memiliki model penelitian yang berisi prosedur kerja untuk digunakan sebagai penentu alur penelitian. Penelitian ini akan dimulai dengan meninjau literatur, mengumpulkan data, melakukan preprocessing, proses prediksi, dan evaluasi dengan menghitung nilai error. Gambar 1 merupakan alur penelitian yang akan diterapkan pada penelitian ini.



Gambar 1. Alur Penelitian

### A. Pengumpulan Dataset

Dataset yang digunakan pada penelitian ini merupakan dataset area kebakaran yang dapat diakses secara publik pada repositori UCI. Data yang digunakan terdiri dari 13 atribut yang terdiri dari koordinat x, koordinat y, *month*, *day*, FFMC, DMC, DC, ISI, *temp*, RH, *wind*, *rain*, dan *area*, seperti yang ditunjukkan dalam Gambar 2.

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area	
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0

Gambar 2. Detail Dataset yang Digunakan

### B. Preprocessing

Pada tahap *preprocessing*, atribut-atribut yang tidak diperlukan akan diubah, dihilangkan, dan juga dipertimbangkan kembali untuk menentukan data mana yang akan digunakan untuk pemrosesan selanjutnya. Preprocessing pertama dilakukan dengan mengubah data bulan dan hari yang semula huruf menjadi angka; misalnya, Januari diubah menjadi 1, Februari diubah menjadi 2, dan seterusnya, dan Senin diubah menjadi 1, Selasa diubah menjadi 2, dan seterusnya. Perubahan teks menjadi angka dapat dilihat pada Gambar 3.

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	3	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	10	2	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0
2	7	4	10	6	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0
3	8	6	3	5	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2
4	8	6	3	7	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0

Gambar 3. Preprocessing Awal

### C. Skenario Penelitian

Pada penelitian ini akan terdapat dua skenario penelitian. Skenario pertama, akan dilakukan dengan menggunakan 12 atribut yaitu koordinat x, koordinat y, *month*, *day*, FFMC, DMC, DC, ISI, *temp*, RH, *wind*, dan *rain*. Sedangkan skenario kedua akan dilakukan dengan menggunakan 5 atribut yang terdiri dari koordinat Y, RH, DMC, *temp*, dan *wind*. Perbedaan skenario 1 dan 2 dapat dilihat secara detail pada Gambar 4 dan Gambar 5.

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	
0	7	5	3	5	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0
1	7	4	10	2	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0
2	7	4	10	6	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0
3	8	6	3	5	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2
4	8	6	3	7	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0

Gambar 4. Detail Data Skenario 1

Y	DMC	temp	RH	wind	
0	5	26.2	8.2	51	6.7
1	4	35.4	18.0	33	0.9
2	4	43.7	14.6	33	1.3
3	6	33.3	8.3	97	4.0
4	6	51.3	11.4	99	1.8

Gambar 5. Detail Data Skenario 2

### D. *Random forest*

Metode untuk mengelompokkan data berdasarkan tujuan atau outputnya, yang memungkinkan analisis pola data, saat ini dikenal sebagai klasifikasi. *Random forest* melakukan klasifikasi dengan menggunakan sekumpulan pohon keputusan, di mana hasil akhirnya ditentukan dengan menggabungkan hasil dari pohon-pohon tersebut yang telah dilatih menggunakan data pelatihan, dan melakukan pengujian dengan berbagai atribut secara acak [11].

### E. Evaluasi

Evaluasi adalah proses yang dilakukan untuk menentukan apakah hasil perhitungan data sudah sesuai dengan hipotesis atau justru bertentangan dengannya. Pada penelitian ini, evaluasi model dilakukan dengan menghitung nilai *Root Mean Square Error* (RMSE), *Mean Squared Error* (MSE), dan koefisien determinasi ( $R^2$ ).

## III. HASIL DAN PEMBAHASAN

Penggunaan *Random forest* untuk memprediksi kebakaran hutan melibatkan pembagian data menjadi 80% untuk pelatihan dan 20% untuk pengujian. Tujuannya adalah untuk membantu model mengklasifikasikan data diikuti dengan evaluasi kinerja dengan membandingkan nilai RMSE, MSE, dan  $R^2$ .

Pada proses melatih *Random forest* untuk proses prediksi, kedua skenario akan mendapat nilai  $n$  atau jumlah pohon keputusan yang akan dibentuk oleh *Random forest* dalam melatih model sehingga mampu menghasilkan nilai error terkecil. Nilai  $n$  yang akan diujikan adalah 1000 karena mengingat jumlah data yang cukup banyak. Tabel 1 merupakan hasil pengujian menggunakan *Random forest* pada kedua skenario.

Tabel 1. Perbandingan Nilai RMSE, MSE, dan  $R^2$   $n=100$ 

Skenario	Evaluasi		
	RMSE	MSE	$R^2$
Skenario 1	0.557	0.310	0.844
Skenario 2	0.548	0.301	0.849

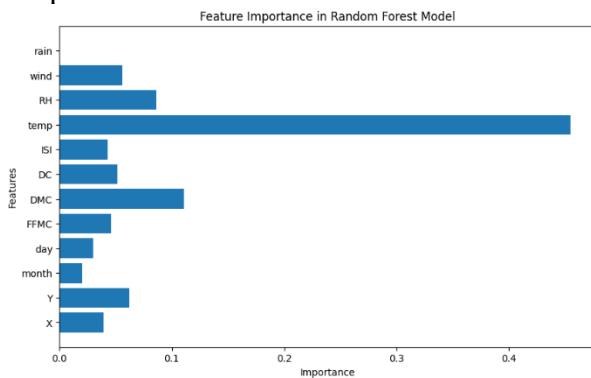
Berdasarkan Tabel 1, hasil evaluasi dari dua skenario, Skenario 1 dan Skenario 2, menunjukkan performa yang solid dalam memprediksi kebakaran hutan menggunakan model *Random forest*. Dalam Skenario 1, ditemukan bahwa model menghasilkan RMSE sebesar 0.557 dan MSE sebesar 0.310, dengan  $R^2$  mencapai 0.844. Hal ini menunjukkan bahwa meskipun terdapat sedikit variabilitas antara prediksi dan nilai aktual, model secara keseluruhan mampu menjelaskan variasi dalam data dengan baik.

Di sisi lain, Skenario 2 menunjukkan peningkatan performa dengan nilai RMSE yang sedikit lebih rendah, yaitu 0.548, MSE sebesar 0.301, dan  $R^2$  mencapai 0.849. Hasil ini menunjukkan bahwa model lebih akurat dalam

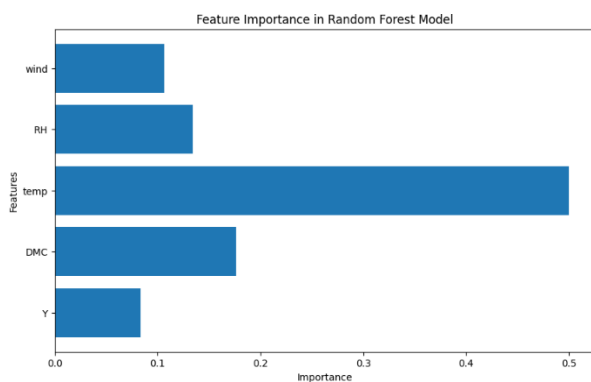
memprediksi kejadian kebakaran hutan pada Skenario 2, dengan kesalahan prediksi yang lebih kecil dan kemampuan yang lebih baik untuk menangkap pola dalam data.

Performa yang lebih tinggi pada Skenario 2 juga mengindikasikan bahwa fitur-fitur yang digunakan dalam pelatihan model mungkin lebih relevan atau bahwa model secara keseluruhan lebih cocok dengan data uji yang digunakan. Dengan demikian, evaluasi ini menunjukkan bahwa *Random forest* efektif dalam aplikasi prediksi kebakaran hutan, dengan kemampuan untuk meningkatkan akurasi prediksi melalui optimisasi fitur dan tuning model yang lebih lanjut.

Untuk lebih mendalami terkait kemampuan *random forest*, berdasarkan proses pelatihan data yang melibatkan 1000 pohon keputusan, selanjutnya dapat dilakukan analisis terkait seberapa besar kontribusi setiap atribut terhadap proses prediksi area kebakaran. Visualisasi yang disajikan dalam Gambar 6 menggambarkan tingkat kontribusi variabel yang digunakan dalam skenario 1, sedangkan Gambar 7 menggambarkan tingkat kontribusi variabel yang digunakan dalam skenario 2 dari penelitian ini.



Gambar 6 Visualisasi Important Features Skenario 1



Gambar 7 Visualisasi Important Features Skenario 2

Berdasarkan pada Gambar 6 dan Gambar 7 terlihat bahwa atribut *temp* memiliki pengaruh paling penting pada proses prediksi area kebakaran dikedua skenario. Selanjutnya terdapat atribut DMC dan RH. Sedangkan atribut *wind* pada Gambar 6 terlihat tidak memiliki pengaruh apa-apa.

#### IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, Hasil penelitian ini menunjukkan bahwa dengan menggunakan karakteristik penyebab kebakaran, penelitian ini dapat memprediksi kebakaran pada suatu area. Hasil pengujian dataset kebakaran hutan UCI, yang dilakukan dengan metode Random forest dengan menggunakan nilai  $n = 1000$ , menghasilkan RMSE 0.557, MSE 0.0,310, dan R2 Score 0.844 pada penggunaan 12 atribut (Skenario 1) dan RMSE 0.548, MSE 0.301, dan R2 Score 0.849 pada penggunaan 8 atribut (Skenario 2). Hasil ini menunjukkan bahwa penelitian ini dapat digunakan untuk memprediksi kebakaran hutan.

Untuk penelitian selanjutnya, kami mengusulkan untuk mencoba lebih banyak kombinasi atribut dan mengimplementasikan kedalam sebuah aplikasi yang memiliki fitur untuk melakukan prediksi berkala secara otomatis dan fitur beberapa lokasi sehingga dapat memaksimalkan peran manusia untuk mencegah kemungkinan terjadinya kebakaran hutan di suatu titik lokasi.

#### REFERENSI

- [1] [1] Z. Mohammed, C. Hanae, and S. Larbi, "Comparative study on machine learning algorithms for early fire forest detection system using geodata," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, p. 5507, Oct. 2020, doi: 10.11591/ijece.v10i5.pp5507-5513.
- [2] M. B. R. Prayoga and R. H. Koestoer, "Improving Forest Fire Mitigation in Indonesia: A Lesson from Canada," *J. Wil. dan Lingkungan*, vol. 9, no. 3, pp. 293–305, Dec. 2021, doi: 10.14710/jwl.9.3.293-305.
- [3] I. A. Riyanto, A. Cahyadi, F. Kurniadhini, H. Bachtiar, D. Apriyana, and B. K. A. Caraka, "Understanding Forest Fire Management in Indonesia from a Global Perspective," *ASEAN J. Sci. Technol. Dev.*, vol. 37, no. 1, Apr. 2020, doi: 10.29037/ajstd.593.
- [4] E. Abdul Kadir, S. Listia Rosa, A. Syukur, M. Othman, and H. Daud, "Forest fire spreading and carbon concentration identification in tropical region Indonesia," *Alexandria Eng. J.*, vol. 61, no. 2, pp. 1551–1561, Feb. 2022, doi: 10.1016/j.aej.2021.06.064.
- [5] D. A. Wood, "Prediction and data mining of burned areas of forest fires: Optimized data matching and mining algorithm provides valuable insight," *Artif. Intell. Agric.*, vol. 5, pp. 24–42, 2021, doi: 10.1016/j.aiaa.2021.01.004.
- [6] D. R. Manalu, M. Zarlis, H. Mawengkang, and O. S. Sitompul, "Forest Fire Prediction in Northern Sumatera using Support Vector Machine Based on the Fire Weather Index," in *Computer Science & Information Technology (CS & IT)*, Dec. 2020, pp. 187–196. doi: 10.5121/csit.2020.101915.

- [7] F. Fitriyani and R. Sanjaya, "KOMPARASI ALGORITMA LR, K-NN DAN SVM UNTUK ESTIMASI AREA KEBAKARAN HUTAN," *Infotronik J. Teknol. Inf. dan Elektron.*, vol. 3, no. 2, p. 103, Dec. 2018, doi: 10.32897/infotronik.2018.3.2.109.
- [8] M. S. Ahsan, Z. Zakaria, Z. Hadi, S. E. A. Kurni, and K. Kusrini, "Forest fire predicting using Naive Bayes and KNN algorithm," *Sinkron*, vol. 7, no. 4, pp. 2340–2347, Oct. 2022, doi: 10.33395/sinkron.v7i4.11609.
- [9] A. P. Mahendra, D. Pradipta, M. R. B. Saputro, and K. Kusrini, "Application of the Decision Tree Method to Forest Fire Detection (Case Study: in Palembang, South Sumatra)," *JTECS J. Sist. Telekomun. Elektron. Sist. Kontrol Power Sist. dan Komput.*, vol. 2, no. 1, p. 75, Feb. 2022, doi: 10.32503/jtecs.v2i1.2196.
- [10] D. Alita and A. R. Isnain, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *J. komputasi*, vol. 8, no. 2, Oct. 2020, doi: 10.23960/komputasi.v8i2.2615.
- [11] P. Rosyani, S. Saprudin, and R. Amalia, "Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO)," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 132, Apr. 2021, doi: 10.26418/justin.v9i2.44120.