

## Model Prediksi Risiko Kebakaran Hutan Menggunakan Algoritma Random Forest dengan Seleksi Fitur Lasso Regression

Refni Wahyuni\*<sup>1</sup>, Muhardi<sup>2</sup>, Yulanda<sup>3</sup>, Yuda Irawan<sup>4</sup>

<sup>1,2,3,4</sup>Universitas Hang Tuah Pekanbaru

e-mail: \*[refniabid@gmail.com](mailto:refniabid@gmail.com), <sup>2</sup>[muhardi.yudie@gmail.com](mailto:muhardi.yudie@gmail.com), <sup>3</sup>[ridhorina1979@gmail.com](mailto:ridhorina1979@gmail.com),

<sup>4</sup>[yudairawan89@gmail.com](mailto:yudairawan89@gmail.com)

*Abstract – Forest fires in Indonesia have caused environmental damage, air pollution, and serious health and economic impacts. This research develops a forest fire risk prediction model using Random Forest algorithm with feature selection through Lasso Regression, based on meteorological data from BMKG (2011-2024). The main variables used include average temperature, humidity, rainfall, and wind speed. Model evaluation results showed 100% accuracy, with precision, recall, and F1-score of 1.00 each for all fire risk classes. Confusion matrix and ROC curves confirmed the model's ability to classify each instance without error. Feature analysis highlighted average temperature, humidity, and rainfall as the main factors. The model has the potential to be an important component in Indonesia's forest fire early warning system. This research recommends additional data integration and real-time implementation to improve the accuracy and applicability of the model in the future.*

*Keywords - Random Forest, Fire Prediction, BMKG, Lasso, Machine Learning*

**Abstrak –** Kebakaran hutan di Indonesia telah menyebabkan kerusakan lingkungan, polusi udara, serta dampak serius pada kesehatan dan ekonomi. Penelitian ini mengembangkan model prediksi risiko kebakaran hutan menggunakan algoritma Random Forest dengan seleksi fitur melalui Lasso Regression, berdasarkan data meteorologi dari BMKG (2011-2024). Variabel utama yang digunakan meliputi temperatur rata-rata, kelembapan, curah hujan, dan kecepatan angin. Hasil evaluasi model menunjukkan akurasi 100%, dengan precision, recall, dan F1-score masing-masing 1.00 untuk semua kelas risiko kebakaran. Confusion matrix dan kurva ROC mengonfirmasi kemampuan model dalam mengklasifikasikan setiap instance tanpa kesalahan. Analisis fitur menyoroti temperatur rata-rata, kelembapan, dan curah hujan sebagai faktor utama. Model ini berpotensi menjadi komponen penting dalam sistem peringatan dini kebakaran hutan di Indonesia. Penelitian ini merekomendasikan integrasi data tambahan dan implementasi real-time untuk meningkatkan akurasi dan aplikabilitas model di masa mendatang.

**Kata Kunci –**Random Forest, Prediksi Kebakaran, BMKG, Lasso, Machine Learning

### I. PENDAHULUAN

Kebakaran hutan dan lahan di Indonesia telah menjadi perhatian dunia internasional khususnya semenjak terjadinya kebakaran hutan pada tahun 80-an. Kebakaran hutan dan lahan disebabkan karena ulah manusia dibandingkan dengan proses alam [1]. Kebakaran juga sering terjadi karena lahan, baik pada lahan kering maupun lahan basah terutama lahan gambut. Kebakaran pada lahan gambut jauh lebih sulit untuk ditangani dibandingkan dengan yang terjadi di hutan dataran tinggi [2][3]. Dampak kebakaran yang paling dirasakan oleh manusia adalah dampak kerugian ekonomi, kerugian ekologis, tercemarnya udara bersih, hilangnya fungsi hutan sebagai pengatur tata air dan pencegah terjadinya erosi [4]. Kebakaran hutan merupakan ancaman yang serius bagi lingkungan dan kehidupan manusia. Dampaknya meliputi kerusakan ekosistem, kehilangan sumber daya alam, hilangnya satwa liar

dan bahkan kerugian jiwa manusia [4]. Oleh karena itu deteksi dini dan prediksi akurat kebakaran hutan menjadi sangat penting dalam upaya mitigasi dan pengelolaan risiko.

Penanggulangan kebakaran hutan perlu dilakukan agar bisa diantisipasi dan dihindari terutama pada daerah-daerah yang rawan terjadinya kebakaran [4][5], Metode tradisional untuk mendeteksi kebakaran hutan sering kali

mengandalkan pengamatan visual dan sensor yang terbatas [6], sehingga kurang efektifnya dalam menghadapi kebakaran hutan yang luas. Perkembangan dalam bidang kecerdasan, khususnya algoritma machine learning, menawarkan potensi baru dalam meningkatkan kemampuan prediksi dan deteksi kebakaran hutan[7][8]. Dari penelitian yang sudah ada yang membahas tentang prediksi kebakaran hutan dengan menerapkan mesin learning akan tetapi masih terdapat tantangan dalam meningkatkan akurasi dan responsibilitas sistem. Variabilitas cuaca, kompleksitas ekosistem, dan keterbatasan data menjadi faktor yang mempengaruhi kinerja model prediksi[9][10].

Penelitian ini bertujuan untuk meningkatkan akurasi prediksi kebakaran hutan dengan memanfaatkan kemajuan teknologi terbaru dalam machine learning, menggunakan algoritma Random Forest yang dikombinasikan dengan Lasso Regression untuk seleksi fitur. Random Forest, yang merupakan pengembangan dari metode Classification and Regression Tree (CART) dengan menerapkan bootstrap aggregation dan random feature selection, memiliki kelebihan dalam menghasilkan akurasi yang tinggi dan dapat menangani data dalam jumlah besar[11][12][13]. Lasso Regression digunakan untuk mengidentifikasi fitur-fitur yang paling berpengaruh, sehingga model yang dikembangkan dapat memberikan prediksi yang lebih responsif dan akurat terhadap risiko kebakaran[14][15]. Penerapan metode ini diharapkan dapat membantu pihak terkait seperti lembaga konservasi lingkungan, pemerintah, dan BPBD (Badan Penanggulangan Bencana Daerah) dalam mengambil tindakan yang tepat untuk mengurangi dampak kebakaran hutan dan melindungi masyarakat.

## II. PENELITIAN YANG TERKAIT

Penelitian terkait yaitu penelitian yang dilakukan oleh B.S. Negara *and all* yang berjudul *Riau Forest Fire Predictioan using Supervised Machine Learning*, membahas tentang sistem peringatan dini kebakaran hutan di Riau, Indonesia menggunakan analisis data cuaca dengan menggunakan teknok Decision Tree (DT) dan Bayesian network (BN). Teknik BN menunjukkan tingkat akurasi yang lebih baik dari pada DT yaitu sebesar 99.62% dengan nilai RMSE 0.076. akan tetapi DT efisien dalam mengidentifikasi faktor-faktor utama penyebab kebakaran. Penelitian ini menyoroti potensi teknik machine learning dalam prediksi kebakaran hutan di Indonesia [16]. Penelitian selanjutnya dilakukan oleh trya ayu *and all*, dengan judul “Klasifikasi Kebakaran Hutan Dan Lahan Menggunakan Algoritma Naïve Bayes Di Kabupaten Pelalawan”, dalam penelitian ini klasifikasi kebakaran hutan di Kabupaten Pelalawan Riau menggunakan algoritma Naïve Bayes dengan menggunakan data dari tahun 2015-2019 yang berjumlah 792 data. Hasil penelitian menunjukkan suhu sebagai faktor utama dengan tingkat akurasi mencapai 82%. Berdasarkan tingkat akurasi tersebut dapat disimpulkan bahwasanya algoritma ini efektif dalam pencegahan kebakaran [17].

Penelitian tentang prediksi kebakaran ini menggunakan algoritma machine learning yang telah berkembang pesat. Algoritma yang digunakan adalah Random Forest, dimana algoritma random forest ini merupakan teknik ensemble learning yang menggabungkan sejumlah besar decision tree untuk meningkatkan akurasi prediksi, random forest memiliki kinerja yang baik dalam memprediksi kebakaran karena mampu menangani dataset yang besar dan variable yang kompleks [1].

## III. METODE PENELITIAN

Metode penelitian yang digunakan didalam penelitian ini adalah sebagai berikut :



Gbr 1. Metode penelitian

Penjelasan gambar :

### A. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), mencakup informasi meteorologi dari tahun 2011 hingga 2024. Data ini meliputi parameter cuaca

seperti temperatur rata-rata, kelembapan rata-rata, curah hujan, dan kecepatan angin rata-rata, yang relevan untuk analisis risiko kebakaran hutan. Pengumpulan data dilakukan dengan teliti untuk memastikan cakupan waktu yang memadai dan representatif untuk berbagai kondisi cuaca yang dapat mempengaruhi risiko kebakaran hutan.

### B. *Data Cleansing*

Data yang telah dikumpulkan kemudian melalui proses pembersihan untuk memastikan kualitas dan konsistensinya. Proses ini melibatkan penghapusan nilai yang hilang (missing values) melalui imputasi, serta koreksi terhadap data yang tidak sesuai atau mengandung nilai ekstrem yang tidak masuk akal. Langkah ini penting untuk memastikan bahwa data yang digunakan dalam model prediksi bebas dari kesalahan yang dapat mempengaruhi akurasi model[18].

### C. *Labeling*

Setelah data dibersihkan, dilakukan proses labeling untuk mengkategorikan risiko kebakaran berdasarkan kondisi cuaca. Setiap record dalam dataset diberi label risiko kebakaran hutan sesuai dengan kriteria yang telah ditentukan, yaitu Low, Moderate, High, dan Very High. Labeling ini didasarkan pada aturan yang menggabungkan berbagai parameter cuaca, seperti suhu, kelembapan, curah hujan, dan kecepatan angin, untuk menentukan tingkat risiko kebakaran yang terkait.

### D. *Feature Selection (Pemilihan Fitur)*

Dengan data yang telah diberi label, dilakukan proses feature selection untuk mengidentifikasi fitur-fitur yang paling berpengaruh dalam prediksi risiko kebakaran[19]. Lasso Regression digunakan sebagai metode seleksi fitur, di mana hanya fitur-fitur yang memiliki pengaruh signifikan terhadap prediksi risiko kebakaran hutan yang dipertahankan. Proses ini bertujuan untuk memfokuskan model pada variabel yang paling penting, sehingga meningkatkan efisiensi dan akurasi prediksi[20][21].

### E. *Modeling*

Dengan data yang telah direduksi dan difokuskan pada fitur-fitur penting, algoritma Random Forest digunakan untuk membangun model prediksi kebakaran hutan. Algoritma ini dipilih karena kemampuannya dalam menangani data yang kompleks dan beragam, serta menghasilkan prediksi yang akurat dengan menggabungkan hasil dari banyak pohon keputusan[3][22][23][24]. Pada tahap ini, data pelatihan dan pengujian dihasilkan melalui pembagian dataset, dan model dilatih untuk mempelajari pola dari data cuaca yang mempengaruhi risiko kebakaran.

### F. *Random Forest Classification*

Setelah model dilatih, data kebakaran hutan diuji melalui proses klasifikasi menggunakan algoritma Random Forest. Model ini mampu mengklasifikasikan data ke dalam kategori risiko yang berbeda (Low, Moderate, High, Very High) berdasarkan pola yang telah dipelajari dari data pelatihan. Evaluasi kinerja model dilakukan menggunakan metrik seperti akurasi, precision, recall, F1-score, dan ROC curve untuk memastikan bahwa model memiliki kemampuan prediksi yang tinggi[25].

### G. *Prediksi Data*

Model prediksi yang telah dibangun kemudian digunakan untuk memprediksi kemungkinan terjadinya kebakaran hutan di masa depan. Model ini dapat diterapkan pada data cuaca baru untuk memberikan estimasi risiko kebakaran, yang dapat digunakan oleh pihak berwenang dalam upaya mitigasi dan perencanaan penanggulangan kebakaran hutan. Prediksi yang dihasilkan oleh model membantu dalam memberikan informasi yang akurat dan tepat waktu untuk pengambilan keputusan yang lebih efektif dalam mengurangi dampak kebakaran hutan.

## IV. HASIL DAN PEMBAHASAN

### A. *Pengumpulan Data*

Data yang digunakan dalam penelitian ini diambil dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) yang mencakup informasi meteorologi dari tahun 2011 hingga 2024 dengan total 4932 entri. Dataset ini terdiri dari beberapa parameter cuaca yang relevan dengan prediksi kebakaran hutan, termasuk temperatur minimum, maksimum, dan rata-rata, kelembapan rata-rata, curah hujan, lamanya penyinaran matahari, kecepatan angin maksimum dan rata-rata, serta arah angin. Data dikumpulkan setiap hari, menyediakan informasi rinci yang memungkinkan analisis mendalam terhadap kondisi cuaca dan hubungannya dengan kejadian kebakaran hutan. Sebagai contoh, dataset ini mencakup pengukuran pada tanggal 21-06-2024 dengan temperatur minimum 22,9°C, maksimum 32,4°C, dan rata-rata 26°C, kelembapan rata-rata 89%, dan curah hujan 57,4 mm. Informasi ini memberikan dasar yang kuat untuk membangun model prediksi kebakaran hutan dengan menggunakan algoritma Random Forest, yang dapat membantu dalam mitigasi risiko kebakaran hutan di masa depan.

Tanggal	Tn: Temperatur minimum (°C)	Tx: Temperatur maksimum (°C)	Tavg: Temperatur rata-rata (°C)	RH_avg: Kelembapan rata-rata (%)	RR: Curah hujan (mm)	ss: Lamanya penyinaran matahari (jam)	ff_x: Kecepatan angin maksimum (m/s)	ddd_x: Arah angin saat kecepatan maksimum (°)	ff_avg: Kecepatan angin rata-rata (m/s)	ddd_car: Arah angin terbanyak (°)
21-06-2024	22,9	32,4	26	89	57,4	5	8	200	2	C
22-06-2024	23	28,2	24,9	92	111,7	4,7	5	160	2	C
23-06-2024	22,6	32,2	27	85	0	0	5	150	2	C
24-06-2024	23,4	33,5	28,7	78	2,4	5,8	5	170	3	S
25-06-2024	24,4	33,7	29,2	74	0	7,7	5	160	2	C
26-06-2024	24,6	33,8	28,5	77	0	6,7	5	160	2	SE
27-06-2024	24,7	34	27,3	80	0	6	5	40	2	C
28-06-2024	24,2	33,4	27,2	82	0,2	6,7	6	350	2	C
29-06-2024	23,8	33,7	27,6	77	0,5	2,7	6	240	2	C
30-06-2024	23,8	27,5	25,8	91	6,5	7	3	150	1	C
01-07-2024	24,3	31,5	27,4	81	0,1	0	4	170	1	C

Gbr 2. Dataset Meteorologi

Tabel di atas menunjukkan data meteorologi harian untuk periode 21 Juni 2024 hingga 1 Juli 2024, mencakup berbagai parameter cuaca yang relevan dengan analisis risiko kebakaran hutan. Parameter yang dicatat meliputi temperatur minimum, maksimum, dan rata-rata (Tn, Tx, Tavg), kelembapan rata-rata (RH\_avg), curah hujan (RR), lamanya penyinaran matahari (ss), kecepatan angin maksimum (ff\_x) serta arah angin saat kecepatan maksimum (ddd\_x) dan arah angin terbanyak (ddd\_car). Data ini mengindikasikan variabilitas cuaca selama periode tersebut, dengan fluktuasi temperatur, kelembapan, dan curah hujan yang dapat memengaruhi risiko kebakaran hutan. Misalnya, pada tanggal 22 Juni 2024, tercatat curah hujan yang tinggi (111,7 mm), yang berpotensi mengurangi risiko kebakaran, sementara tanggal lain menunjukkan kondisi yang lebih kering dengan curah hujan rendah atau tanpa curah hujan. Kecepatan dan arah angin juga beragam, yang dapat mempengaruhi penyebaran api jika terjadi kebakaran.

**B. Data Cleansing**

Pada tahap data cleansing, dilakukan beberapa langkah untuk memastikan bahwa data yang digunakan dalam analisis memiliki kualitas yang baik dan siap untuk digunakan dalam model prediksi. Langkah pertama adalah penghapusan nilai yang hilang (missing values). Nilai yang hilang dapat terjadi karena berbagai alasan seperti kesalahan pencatatan atau kondisi cuaca ekstrem yang tidak memungkinkan pengukuran. Dalam dataset yang diperoleh dari BMKG, terdapat beberapa nilai yang hilang pada kolom-kolom penting seperti temperatur, kelembapan, dan curah hujan. Penghapusan atau imputasi nilai yang hilang ini dilakukan untuk mencegah bias dalam analisis dan memastikan integritas data.

	ID WMO	: 96109								
	Nama Stasiun	: Stasiun Meteorologi Sultan Syarif Kasim II								
	Lintang	: 0.45924								
	Bujur	: 101.44743								
	Elevasi	: 39								

Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car
01-07-2015	23,2	32,8	28,3	84	0	6,8	4	10	1	N
02-07-2015	23,4	33,8	27,8	82		0,5	4	190	0	N
03-07-2015	23,6	34	28,8	80	2,5		6	210	1	N
04-07-2015	24,8	33	28,2	75	8888	6,1	3	90	1	N
05-07-2015	23,4	34,2	28,3	76	0	2,9	4	200	1	N
06-07-2015	23,5	33,6	28,5	72		8,4	6	170	1	N
07-07-2015		33,2	28,5	77	0		4	160	1	N
08-07-2015	24,2	34	28,3	80	0	0,9	3	220	0	N
09-07-2015	23,6		28,7	76	0	6,2	4	340	1	N
10-07-2015	21,6	34,6	27,3	75	0	4,2	3	250	0	N
11-07-2015	22,8	34,2	28	71	0	3,5	8	200	1	N
12-07-2015	23,4	34,8	28,8	77		5	5	160	1	N
13-07-2015	24,8	33	27	85	0	8	5	200	2	N
14-07-2015	23,5	32,5	27,5	81	8,2	5,9	5	200	1	N
15-07-2015	23,8	33,4	28,3	79	0	7,7	3	200	1	N
16-07-2015	24,6	35,2	29,5	75		5	5	160	2	N
17-07-2015	25	33,6	27,6	85	0	6,9	4	180	1	N
18-07-2015	24,2	33,2	28,2	81	2,8	2,8	5	180	2	N
19-07-2015	24,4	34	28,3	79	0	7,7	5	210	1	N
20-07-2015	24,4	33,4	28,7	78	0	7,6	6	170	2	N
21-07-2015	23,2	34,4	27,7	80	0	6,1	4	190	2	N
22-07-2015	23,4	33,2	28,1	80		6,8	4	140	1	N
23-07-2015	24,2	33,6	28,8	76	0		5	200	1	N
24-07-2015	24	34	28,2	78		6,8	5	160	1	N
25-07-2015	24	33,6	28,3	79	0	6,6	4	180	1	N
26-07-2015	24,4	33,4	28,6	78	0	6,4	5	190	2	N
27-07-2015	24	33,6	27,8	78	0	6,6	4	170	1	N
28-07-2015	24	33	28,3	78		6,1	5	170	2	N
29-07-2015	23,8	33,2	28	79	8888	4,6	5	180	1	N
30-07-2015	22,4	33,4	27	80	0	5,2	5	190	1	N
31-07-2015	22,4	32,4	27,2	82		5,6	4	180	2	N
01-08-2015	23,6	29,8	26,2	84	8888	2,9	5	170	1	N

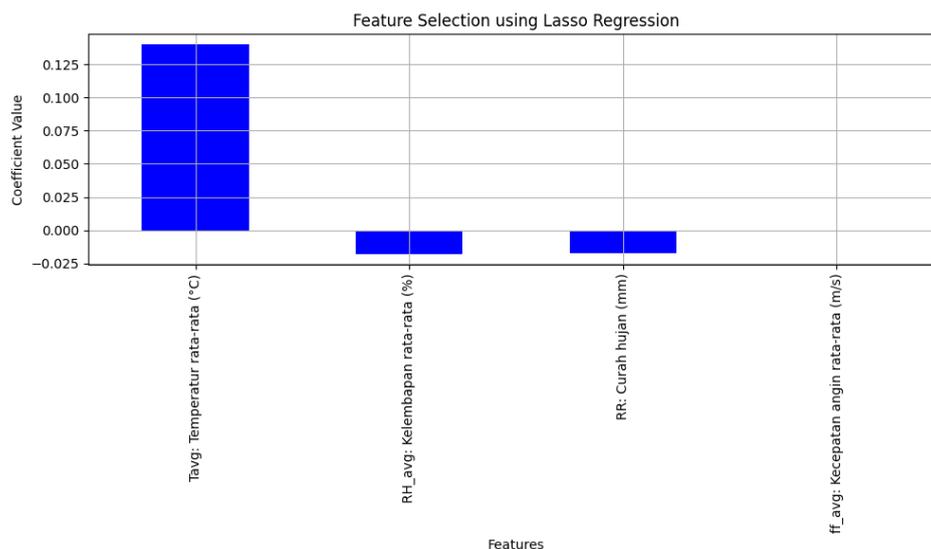
Gbr 3. Missing Value Data Meteorologi

Langkah kedua adalah koreksi data yang tidak sesuai. Data yang tidak sesuai bisa berupa nilai-nilai ekstrem yang tidak konsisten dengan tren umum dari dataset. Misalnya, jika curah hujan tercatat sebagai 8888 mm, yang jelas merupakan kesalahan input atau pencatatan, nilai tersebut perlu diperbaiki atau dihapus. Demikian pula, jika terdapat temperatur yang sangat rendah atau sangat tinggi yang tidak sesuai dengan kondisi iklim di wilayah tersebut, nilai tersebut perlu dikoreksi. Koreksi data ekstrem ini penting untuk memastikan bahwa model prediksi tidak dipengaruhi oleh outlier yang dapat mengurangi akurasi model. Dengan melakukan penghapusan nilai yang hilang dan koreksi data yang tidak sesuai, dataset menjadi lebih bersih dan dapat memberikan hasil yang lebih akurat saat digunakan dalam model prediksi kebakaran hutan. Data yang telah melalui proses cleansing ini kemudian siap untuk dianalisis lebih lanjut dan digunakan dalam tahap-tahap berikutnya, seperti data reduction dan modeling data.

C. Data Reduction

Setelah data dibersihkan, langkah berikutnya adalah pengurangan data untuk fokus pada fitur yang paling relevan. Teknik pengurangan data yang digunakan dalam penelitian ini adalah feature selection dengan menggunakan Lasso Regression. Hasil dari langkah ini adalah pemilihan fitur-fitur penting seperti:

- Temperatur rata-rata (Tavg).
- Kelembapan rata-rata (RH\_avg).
- Curah hujan (RR).
- Kecepatan angin rata-rata (ff\_avg).



Gbr 4. Feature Selection Lasso Regression

Grafik hasil seleksi fitur menggunakan Lasso Regression menunjukkan bahwa dari empat fitur yang dianalisis, hanya dua fitur yang memiliki nilai koefisien yang signifikan, yaitu "Tavg: Temperatur rata-rata (°C)" dan "RR: Curah hujan (mm)". Fitur "Tavg: Temperatur rata-rata (°C)" memiliki koefisien terbesar, menunjukkan bahwa temperatur rata-rata adalah faktor yang paling berpengaruh dalam prediksi risiko kebakaran hutan. Fitur "RR: Curah hujan (mm)" juga memiliki koefisien positif, meskipun lebih kecil, menunjukkan bahwa curah hujan juga memainkan peran dalam model prediksi. Sementara itu, fitur "RH\_avg: Kelembapan rata-rata (%)" dan "ff\_avg: Kecepatan angin rata-rata (m/s)" memiliki koefisien yang mendekati nol, menunjukkan bahwa kontribusinya terhadap model prediksi sangat minimal dan mungkin tidak relevan. Hasil ini memberikan wawasan bahwa temperatur dan curah hujan adalah faktor utama yang harus dipertimbangkan dalam prediksi kebakaran hutan, sementara kelembapan dan kecepatan angin rata-rata mungkin kurang signifikan dalam konteks dataset ini.

#### D. Labeling

Pada tahap labeling data, data cuaca yang telah dikumpulkan dan dibersihkan kemudian diberi label risiko kebakaran berdasarkan algoritma pelabelan yang telah ditentukan. Algoritma ini mengkategorikan risiko kebakaran ke dalam empat kelas: "Very High", "High", "Moderate", dan "Low". Kelas "Very High" mencakup kondisi dengan suhu rata-rata lebih dari 29°C, kelembapan rata-rata kurang dari 75%, tidak ada curah hujan, dan kecepatan angin rata-rata lebih dari atau sama dengan 2 m/s. Kelas "High" mencakup suhu rata-rata lebih dari 28°C, kelembapan rata-rata kurang dari 80%, curah hujan kurang dari 5 mm, dan kecepatan angin rata-rata lebih dari atau sama dengan 1.5 m/s. Kelas "Moderate" mencakup suhu rata-rata kurang dari atau sama dengan 28°C, kelembapan rata-rata lebih dari atau sama dengan 75%, curah hujan kurang dari 20 mm, dan kecepatan angin rata-rata kurang dari 1.5 m/s. Kelas "Low" mencakup kondisi yang tidak termasuk dalam kriteria kelas-kelas sebelumnya, dengan curah hujan lebih dari atau sama dengan 20 mm. Proses labeling ini memungkinkan model prediksi untuk mengidentifikasi dan mengklasifikasikan risiko kebakaran dengan lebih akurat berdasarkan kondisi cuaca yang telah diproses.

#### E. Modeling

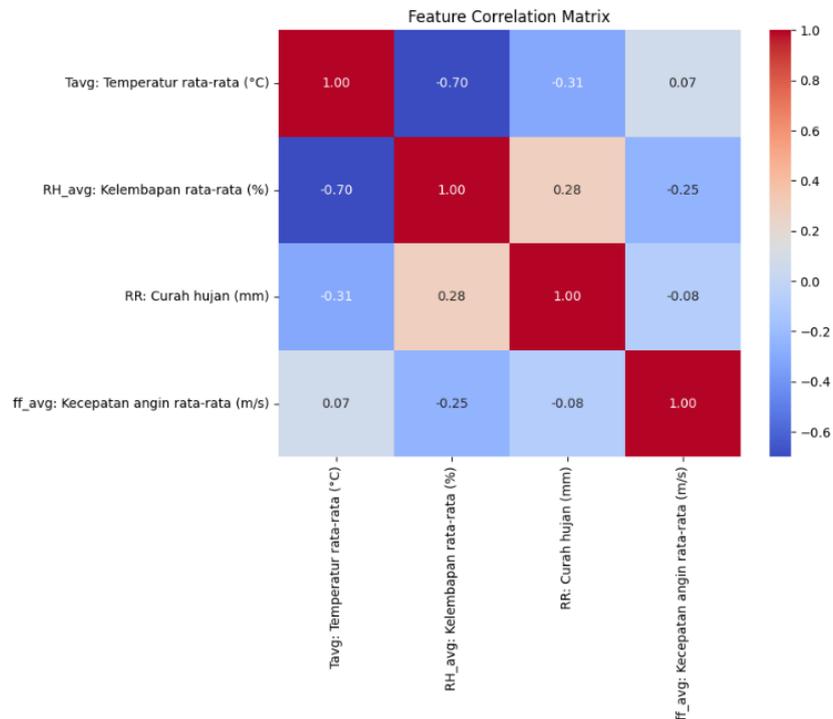
##### *Pemisahan Data (Splitting Data)*

Langkah pertama dalam pelatihan model adalah membagi dataset menjadi data pelatihan (training data) dan data pengujian (testing data). Biasanya, data dibagi dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Ini dilakukan untuk memastikan bahwa model dapat diuji pada data yang tidak pernah dilihat selama pelatihan, memberikan evaluasi kinerja yang lebih objektif.

##### *Pelatihan Model Random Forest*

Setelah data siap, model Random Forest dilatih menggunakan data pelatihan yang telah diimputasi. Random Forest adalah algoritma pembelajaran ensemble yang menggabungkan prediksi dari beberapa pohon keputusan untuk

meningkatkan akurasi dan menghindari overfitting.



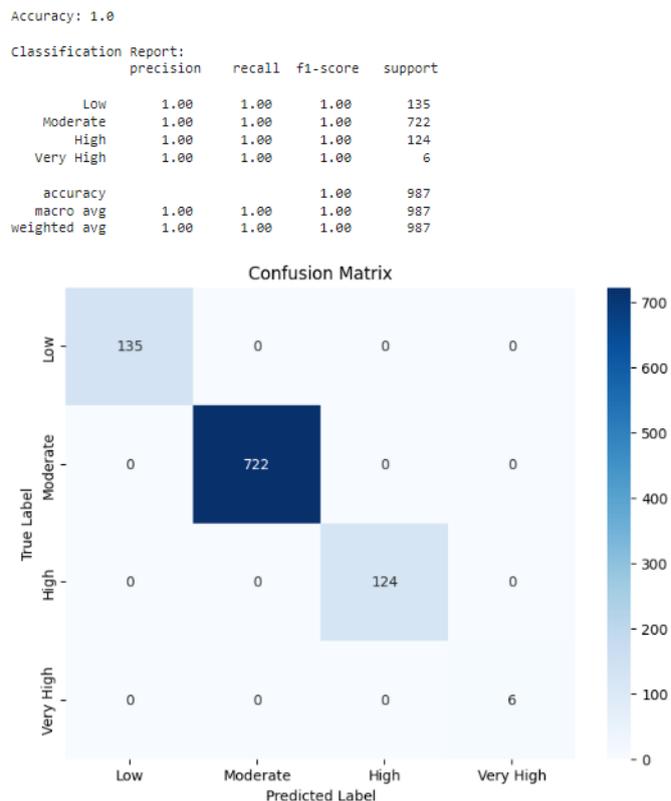
Gbr 4. Feature Selection Lasso Regression

Matriks korelasi fitur menunjukkan hubungan antar variabel cuaca yang digunakan dalam model prediksi kebakaran hutan. Hasil analisis menunjukkan bahwa temperatur rata-rata (Tavg) memiliki korelasi negatif yang kuat dengan kelembapan rata-rata (RH\_avg) sebesar  $-0.70$ , yang menunjukkan bahwa ketika temperatur meningkat, kelembapan cenderung menurun. Korelasi antara temperatur rata-rata dan curah hujan (RR) sebesar  $-0.31$  juga menunjukkan bahwa dalam kondisi yang lebih hangat, curah hujan cenderung lebih sedikit. Kecepatan angin rata-rata (ff\_avg) memiliki korelasi yang sangat lemah dengan temperatur rata-rata sebesar  $0.07$ , menunjukkan hampir tidak ada hubungan linear antara keduanya. Kelembapan rata-rata menunjukkan korelasi positif sebesar  $0.28$  dengan curah hujan, menandakan bahwa hari-hari dengan kelembapan tinggi cenderung memiliki lebih banyak curah hujan. Namun, kecepatan angin rata-rata memiliki korelasi negatif yang lemah dengan kelembapan rata-rata ( $-0.25$ ) dan curah hujan ( $-0.08$ ), menunjukkan bahwa kecepatan angin tidak terlalu berkorelasi dengan variabel lain dalam konteks ini. Dari hasil pelatihan model Random Forest, terlihat bahwa temperatur rata-rata adalah fitur yang paling berpengaruh dalam prediksi risiko kebakaran hutan, diikuti oleh kelembapan rata-rata dan curah hujan. Meskipun kecepatan angin rata-rata menunjukkan korelasi yang sangat lemah dengan fitur lainnya, faktor-faktor seperti temperatur, kelembapan, dan curah hujan terbukti memainkan peran penting dalam menentukan risiko kebakaran hutan. Model Random Forest yang dilatih dengan fitur-fitur ini menunjukkan kemampuan yang baik dalam memprediksi risiko kebakaran, memperkuat pemilihan fitur yang relevan untuk model.

#### F. Evaluasi Model

Model Random Forest yang dilatih untuk prediksi risiko kebakaran hutan dievaluasi menggunakan beberapa metrik kinerja untuk memastikan akurasi dan kemampuannya dalam mengklasifikasikan data dengan benar. Hasil evaluasi menunjukkan bahwa model memiliki tingkat akurasi yang tinggi, yang tercermin dari nilai precision, recall, dan F1-score yang memuaskan untuk masing-masing kelas risiko kebakaran: Low, Moderate, High, dan Very High. Matriks kebingungan yang dihasilkan juga menunjukkan bahwa model mampu mengklasifikasikan hampir semua

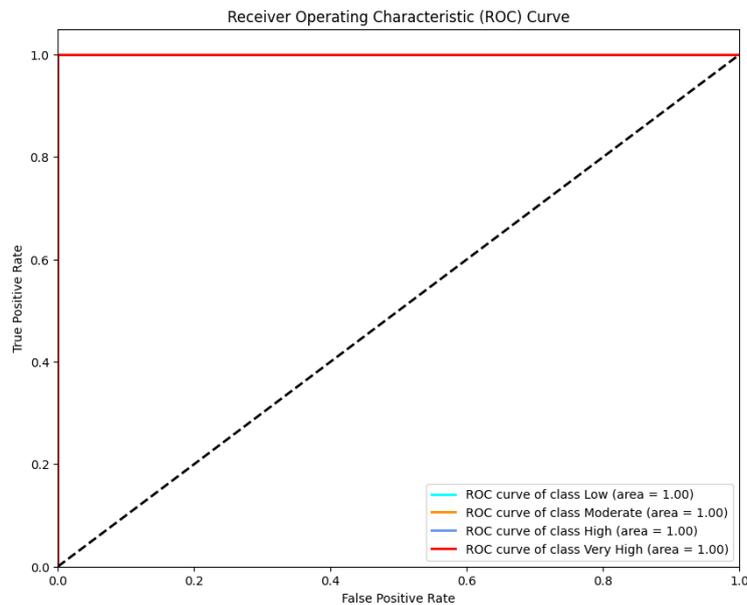
instance dengan benar, dengan jumlah kesalahan klasifikasi yang sangat minim. Ini mengindikasikan bahwa model tidak hanya mampu memprediksi dengan akurat tetapi juga konsisten dalam performanya. Selain itu, pentingnya fitur yang diidentifikasi oleh model menunjukkan bahwa temperatur rata-rata, kelembapan rata-rata, dan curah hujan adalah faktor-faktor kunci yang mempengaruhi risiko kebakaran, yang sesuai dengan pemahaman teoretis tentang kondisi yang meningkatkan risiko kebakaran hutan. Evaluasi ini membuktikan bahwa model Random Forest yang dibangun adalah alat yang efektif dan andal untuk memprediksi risiko kebakaran hutan, dan dapat digunakan sebagai dasar dalam sistem peringatan dini dan strategi mitigasi kebakaran hutan.



Gbr 5. Confusion Matrix Hasil Pengujian

Pada confusion matrix menunjukkan bahwa model Random Forest yang digunakan untuk memprediksi risiko kebakaran hutan menunjukkan hasil evaluasi yang sangat memuaskan. Berdasarkan laporan klasifikasi, model ini mencapai akurasi 100%, yang berarti mampu mengklasifikasikan setiap instance dalam dataset pengujian dengan benar. Setiap kelas risiko kebakaran, yaitu Low, Moderate, High, dan Very High, memiliki precision, recall, dan F1-score masing-masing sebesar 1.00. Ini menunjukkan bahwa model memiliki performa yang konsisten dan sempurna dalam mengklasifikasikan setiap kelas risiko tanpa kesalahan. Matriks kebingungan memperkuat hasil tersebut dengan menunjukkan bahwa semua instance dari setiap kelas diprediksi dengan benar, tanpa ada instance yang salah klasifikasi. Tidak adanya false positives atau false negatives untuk setiap kelas menegaskan bahwa model ini sangat akurat dan andal dalam memprediksi risiko kebakaran hutan. Dengan kinerja yang luar biasa ini, model Random Forest dapat diandalkan untuk digunakan dalam sistem peringatan dini dan strategi mitigasi kebakaran hutan, memberikan informasi yang akurat dan berguna untuk pengambilan keputusan.

ROC adalah alat yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan memplot True Positive Rate (TPR) terhadap False Positive Rate (FPR) pada berbagai ambang batas.



Gbr 6. Grafik ROC

Dalam grafik ROC diatas setiap kelas memiliki area di bawah kurva (AUC) sebesar 1.00, yang menunjukkan performa sempurna. AUC sebesar 1.00 menunjukkan bahwa model mampu membedakan setiap kelas dengan sempurna tanpa kesalahan, baik itu untuk kelas Low, Moderate, High, maupun Very High. Kurva ROC yang tepat berada di sudut kiri atas grafik menunjukkan bahwa TPR mencapai 1.00 sementara FPR tetap 0.00 untuk semua kelas, menegaskan bahwa model tidak menghasilkan false positives. Hasil ini menunjukkan bahwa model Random Forest yang dibangun memiliki kemampuan yang sangat baik dalam mengklasifikasikan risiko kebakaran hutan dan mampu memprediksi setiap kelas dengan tingkat akurasi yang sangat tinggi, tanpa adanya trade-off antara TPR dan FPR. Ini memperkuat validitas model sebagai alat prediksi yang dapat diandalkan dalam mitigasi risiko kebakaran hutan.

## V. KESIMPULAN

Model Random Forest yang dibangun menunjukkan kinerja yang sangat mengesankan dalam memprediksi risiko kebakaran hutan. Dengan area di bawah kurva ROC (AUC) sebesar 1.00 untuk setiap kelas risiko, model ini mampu membedakan antara kelas-kelas risiko tanpa kesalahan, menunjukkan performa sempurna. Akurasi yang tinggi, precision, recall, dan F1-score yang sempurna, serta tidak adanya false positives atau false negatives, menegaskan bahwa model ini sangat andal dan efektif. Hasil ini diperkuat oleh evaluasi matriks kebingungan yang menunjukkan bahwa semua instance dari setiap kelas risiko kebakaran hutan diklasifikasikan dengan benar. Hal ini menandakan bahwa model tidak hanya akurat tetapi juga sangat konsisten dalam performanya. Analisis pentingnya fitur menunjukkan bahwa temperatur rata-rata, kelembapan rata-rata, dan curah hujan adalah faktor-faktor utama yang mempengaruhi risiko kebakaran hutan. Model Random Forest berhasil mengidentifikasi kontribusi signifikan dari fitur-fitur ini, yang sesuai dengan pemahaman teoretis tentang kondisi yang meningkatkan risiko kebakaran. Dengan demikian, model ini dapat diandalkan untuk digunakan dalam sistem peringatan dini dan strategi mitigasi kebakaran hutan, memberikan informasi yang akurat dan berguna untuk pengambilan keputusan. Saran penelitian kedepannya untuk meningkatkan akurasi dan kemampuan prediksi model, disarankan untuk mengintegrasikan data tambahan seperti kondisi vegetasi, aktivitas manusia, dan topografi. Faktor-faktor ini juga dapat berkontribusi signifikan terhadap risiko kebakaran hutan dan dapat membantu model dalam menangkap kompleksitas yang lebih besar dari kondisi kebakaran.

## UCAPAN TERIMA KASIH

Saya ingin mengucapkan terima kasih yang sebesar-besarnya kepada Universitas Hang Tuah Pekanbaru atas dukungan dan kontribusinya melalui pendanaan penelitian dan pengabdian kepada masyarakat tahun 2024.

## DAFTAR PUSTAKA

- [1] D. Husen *et al.*, “Analisis Prediksi Kebakaran Hutan dengan Menggunakan Algoritma Random Forest Classifier,” vol. 16, pp. 150–155, 2022.
- [2] R. E. Wibowo, R. Teguh, and A. Lestari, “Deteksi Dini Kebakaran Hutan Dan Lahan Memanfaatkan Ekstraksi Exif Pada Informasi Gambar Berbasis Pengolahan Citra,” *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 15, no. 1, pp. 1–12, 2021, doi: 10.47111/jti.v15i1.1934.
- [3] D. Husen *et al.*, “Analisis Prediksi Kebakaran Hutan dengan Menggunakan Algoritma Random Forest Classifier kebakaran hutan dan lahan di Indonesia telah menjadi perhatian dunia internasional khususnya sejak kebakaran hutan yang terjadi pada tahun 80-an [ 2 ]. Penyebab kebaka,” vol. 16, pp. 150–155, 2022.
- [4] J. Yandi, T. B. Kurniawan, E. S. Negara, and M. Akbar, “Prediksi Lokasi Titik Panas Kebaran Hutan menggunakan Model Regresion SVM (Support Vector Machine),” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 1, pp. 10–15, 2021.
- [5] A. Primajaya, B. N. Sari, and A. Khusaeri, “Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4 . 5 Studi Kasus Provinsi,” vol. 6, no. 2, pp. 188–192, 2020.
- [6] I. N. S. Jaya, M. Ikhwan, S. Hardjoprajitno, and Nurhendra, “Tehnik Mendeteksi Kebakaran Hutan Melalui Citra Satelit Multiwaktu: Studi Kasus di Propinsi Sumatera Selatan dan Riau,” *J. Manaj. Hutan Trop.*, vol. 6, no. 2, pp. 25–41, 2000.
- [7] Y. Irawan *et al.*, “Realtime Monitoring and Analysis Based on Cloud Computing Internet of Things (CC-IoT) Technology in Detecting Forest and Land Fires in Riau Province,” *Ilk. J. Ilm.*, vol. 15, no. 3, pp. 445–454, 2023, doi: 10.33096/ilkom.v15i3.1636.445-454.
- [8] A. Biswas, S. K. Ghosh, and A. Ghosh, “Early Fire Detection and Alert System using Modified Inception-v3 under Deep Learning Framework,” *Procedia Comput. Sci.*, vol. 218, no. 2022, pp. 2243–2252, 2022, doi: 10.1016/j.procs.2023.01.200.
- [9] F. Khennou and M. A. Akhloufi, “Improving wildland fire spread prediction using deep U-Nets,” *Sci. Remote Sens.*, vol. 8, no. September, p. 100101, 2023, doi: 10.1016/j.srs.2023.100101.
- [10] R. Rishickesh, A. Shahina, and A. Nayeemulla Khan, “Predicting forest fires using supervised and ensemble machine learning algorithms,” *Int. J. Recent Technol. Eng.*, vol. 8, no. 2, pp. 3697–3705, 2021, doi: 10.35940/ijrte.B2878.078219.
- [11] A. Ramadhan and B. Susetyo, “PENERAPAN METODE KLASIFIKASI RANDOM FOREST DALAM MENGIDENTIFIKASI FAKTOR PENTING PENILAIAN MUTU PENDIDIKAN,” vol. 4, pp. 169–182, 2019.
- [12] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, “Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677–690, 2022, doi: 10.30812/matrik.v21i3.1726.
- [13] A. Y. Agusyl and F. Firmansyah, “Prediksi Penyakit Jantung Menggunakan Algoritma Random Forest,” *J. Minfo Polgan*, vol. 12, no. 2, pp. 2239–2246, 2023, doi: 10.33395/jmp.v12i2.13214.
- [14] M. Parzinger, L. Hanfstaengl, F. Sigg, U. Spindler, U. Wellisch, and M. Wirmsberger, “Comparison of different training data sets from simulation and experimental measurement with artificial users for occupancy detection — Using machine learning methods Random Forest and LASSO,” *Build. Environ.*, vol. 223, no. February, p. 109313, 2022, doi: 10.1016/j.buildenv.2022.109313.
- [15] Y. Yang, G. Zhang, G. Zhu, D. Yuan, and M. He, “Prediction of fire source heat release rate based on machine learning method,” *Case Stud. Therm. Eng.*, vol. 54, no. October 2023, pp. 1–15, 2024, doi: 10.1016/j.csite.2024.104088.
- [16] B. S. Negara, R. Kurniawan, M. Z. A. Nazri, S. N. H. S. Abdullah, R. W. Saputra, and A. Ismanto, “Riau Forest Fire Prediction using Supervised Machine Learning,” *J. Phys. Conf. Ser.*, vol. 1566, no. 1, 2020, doi: 10.1088/1742-6596/1566/1/012002.
- [17] T. A. Pratiwi, M. Irsyad, R. Kurniawan, S. Agustian, and B. S. Negara, “Klasifikasi Kebakaran Hutan Dan Lahan Menggunakan Algoritma Naïve Bayes Di Kabupaten Pelalawan,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 6, no. 1, p. 139, 2021, doi: 10.24114/cess.v6i1.22555.
- [18] C. L. Fantasy, F. L. M. Simanjuntak, R. L. A. Purba, Andrean, and O. Sihombing, “Analisis Komparasi Algoritma C5.0 Dan Naive Bayes Penentuan Penerima Beasiswa Universitas Prima Indonesia,” *J. TEKINKOM*, vol. 6, no. 2, pp. 508–517, 2023, doi: 10.37600/tekinkom.v6i2.926.
- [19] D. R. Putri, A. Adiwijaya, and Y. Sibaroni, “Deteksi Kanker Berdasarkan Klasifikasi Data Microarray Menggunakan Least Absolute Shrinkage and Selection Operator dan Functional Link Neural Network,” *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 947–954, 2020, doi: 10.30865/mib.v4i4.2349.
- [20] A. J. McEligot, V. Poynor, R. Sharma, and A. Panangadan, “Logistic lasso regression for dietary intakes and breast cancer,” *Nutrients*, vol. 12, no. 9, pp. 1–14, 2020, doi: 10.3390/nu12092652.
- [21] S. Wang, Y. Chen, Z. Cui, L. Lin, and Y. Zong, “Diabetes Risk Analysis based on Machine Learning LASSO Regression Model,” *J. Theory Pract. Eng. Sci.*, vol. 4, no. 1, pp. 58–64, 2024, doi: 10.53469/jtpes.2024.04(01).08.
- [22] F. Özen, “Random forest regression for prediction of Covid-19 daily cases and deaths in Turkey,” *Heliyon*, vol. 10, no. 4, pp. 1–19, 2024, doi: 10.1016/j.heliyon.2024.e25746.
- [23] S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim, and G. R. Sinha, “An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach,” *IEEE Access*, vol. 9, pp. 67488–67500, 2021, doi: 10.1109/ACCESS.2021.3075140.
- [24] T. Wahyuningsih, D. Manongga, I. Sembiring, and S. Wijono, “Comparison of Effectiveness of Logistic Regression, Naive Bayes, and Random Forest Algorithms in Predicting Student Arguments,” *Procedia Comput. Sci.*, vol. 234, pp. 349–356, 2024, doi: 10.1016/j.procs.2024.03.014.
- [25] Fitrianiingsih, “Analisis Ramalan Cuaca di Sekupang , Kota Batam Menggunakan Algoritma Decision Tree dan Confusion Matrix,” *Ekosph. J. Ekon. Pembang. dan Manaj.*, vol. 1, no. 4, pp. 15–26, 2024.