



Artikel Penelitian

Prediksi Kebakaran Hutan Berdasarkan Titik Panas dan Iklim Menggunakan Algoritma *Random Forest*

Aditya Firmansyah^a, Muhammad Farhan Syahidin^b, Yaya Sudarya Triana^{c}*^{a,b,c} Program Studi Sistem Informasi Universitas Mercu Buana, Indonesia**INFORMASI ARTIKEL***Sejarah Artikel:*

Diterima Redaksi: 25 Mei 2024

Revisi Akhir: 12 September 2024

Diterbitkan Online: 13 September 2024

KATA KUNCIKebakaran hutan,
Prediksi,
Klasifikasi,
Random Forest,
*Web Scraping***KORESPONDENSI**E-mail: yaya.sudarya@mercubuana.ac.id***A B S T R A C T**

Kebakaran hutan dan lahan semakin sering terjadi, menyebabkan dampak lingkungan yang menyebar ke luar wilayah kebakaran. Permasalahan yang terjadi salah satunya karena musim kemarau yang panjang di wilayah Kabupaten Ogan Komering Ilir Provinsi Sumatera Selatan yang menjadi faktor utama dalam meningkatnya risiko kebakaran, sebanyak 1.111 titik kebakaran tercatat pada tahun 2023. Permasalahan lainnya juga pada titik panas yang salah mendeteksi kebakaran yang seharusnya tidak kebakaran dan kasus tidak kebakaran yang seharusnya kebakaran, hal tersebut menyebabkan kerugian lingkungan maupun kerugian dana. Oleh karena itu, dibutuhkan model klasifikasi untuk memprediksi kasus kebakaran. Penelitian ini menggunakan gabungan data titik panas dan data iklim sebanyak 4343 data menggunakan metode *Random Forest*. Proses yang dilakukan yaitu studi literatur dan tahapan prediksi yang terdiri dari *web scraping*, *data pre-processing*, *splitting data*, pemodelan, dan evaluasi. Hasil penelitian berupa laporan klasifikasi, *confusion matrix*, dan *feature importance*. Hasil pengujian menunjukkan tingkat akurasi model yang baik sebesar 85.8% yang menunjukkan model menghitung seberapa tepat kinerja yang dilakukan model. Dengan penerapan model menggunakan metode *Random Forest*, model prediksi ini mengidentifikasi kasus kebakaran sangat baik sehingga informasi ini dapat digunakan untuk keputusan manajemen penanggulangan kebakaran dengan tepat dan meminimalisir terjadinya kerugian.

1. PENDAHULUAN

Indonesia mempunyai dua musim yaitu musim hujan dan musim kemarau. Pada saat musim kemarau wilayah Indonesia pada rentang waktu tertentu akan menjadi kering dan panas, sehingga mengakibatkan beberapa wilayah setiap tahunnya mengalami kebakaran hutan dan lahan di daerah tertentu [1], [2]. Selain itu, Indonesia juga termasuk negara dengan hutan terluas di dunia dan salah satu negara yang dijuluki sebagai paru-paru dunia dikarenakan Indonesia menghasilkan oksigen yang besar dan memiliki tutupan lahan yang luas sehingga mampu menyimpan banyak karbon dalam hutan [3].

Hutan dan lahan di dunia khususnya Indonesia sangat perlu diperhatikan dan dijaga karena kejadian kebakaran hutan dan lahan yang sering terjadi dikarenakan kondisi iklim yang semakin

panas dan kering. Dalam beberapa kasus, terjadinya kebakaran hutan dan lahan juga disebabkan oleh manusia yang mencoba membuka lahan baru pertanian atau untuk tujuan lainnya [4], [5]. Wilayah Indonesia memiliki daerah yang tiap tahunnya sangat rawan terjadi kebakaran hutan dan lahan antara lain Provinsi Jambi, Kalimantan Tengah, Riau dan Sumatera Selatan [6].

Salah satunya yang memiliki tingkat rawan tertinggi yaitu wilayah Sumatera Selatan, terdapat wilayah Kabupaten Ogan Komering Ilir yang sangat rawan terhadap kebakaran hutan dan lahan. Banyaknya kebakaran hutan dan lahan mengakibatkan kerugian seperti terganggunya aktivitas masyarakat di sekitar lokasi kebakaran [7]. Kebakaran hutan dan lahan tidak hanya menyebabkan pemanasan global dan menyebabkan kerugian properti manusia juga mengancam keselamatan jiwa. Dengan adanya pemanasan global ini, menyebabkan kenaikan jumlah kasus kebakaran hutan di seluruh dunia maupun di Indonesia [8].

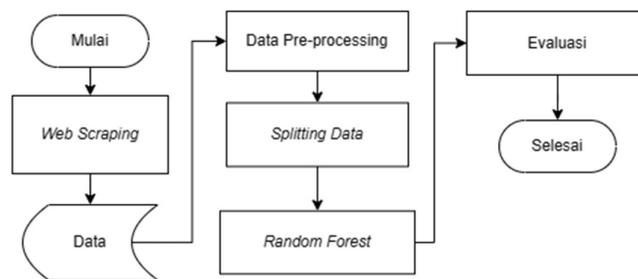
Masalah dari kasus kebakaran hutan dan lahan yang banyak terjadi saat ini mulai beragam, salah satunya yaitu masalah lingkungan yang menimbulkan polusi asap yang tidak hanya berdampak pada lokasi kebakaran saja, melainkan sampai melintasi batas wilayah lokasi kebakaran [9]. Musim kemarau masih menjadi salah satu faktor besar munculnya titik panas atau potensi kebakaran di wilayah Sumatra Selatan Kabupaten Ogan Komering Ilir. Berdasarkan data yang didapat dalam portal web Badan Nasional Penanggulangan Bencana (BNPB) periode Januari 1 Desember 2023 wilayah Kabupaten Ogan Komering Ilir memiliki 1.111 titik kebakaran. Melihat kejadian kebakaran hutan dan lahan yang memberikan dampak negatif semakin besar tiap waktunya, maka dari itu diperlukan suatu usaha peringatan dini untuk kebakaran hutan dan lahan [10].

Berdasarkan uraian diatas, penelitian kebakaran hutan dan lahan ini menggunakan hubungan titik panas dan iklim dengan memahami pola hubungan kedua data tersebut. Penggunaan metode klasifikasi menggunakan algoritma *Random Forest* menghasilkan prediksi besarnya kebakaran yang terjadi yang nantinya dapat membantu pemerintah Kabupaten Ogan Komering Ilir dalam membuat keputusan seberapa besar

penanganan yang harus dilakukan untuk meminimalisir kerugian. Tujuan penelitian ini untuk mengetahui performa dari metode *Random Forest* pada kasus kebakaran menggunakan data titik panas dan data iklim.

2. METODE

Setiap tahapan penelitian ini digambarkan dengan diagram penelitian yang dimulai dari proses pertama yaitu *web scraping* sampai proses pemodelan dengan menerapkan algoritma *Random Forest* dengan menghitung tingkat *accuracy*, *precision*, *recall*, hingga evaluasi hasil. Data yang digunakan bersumber dari *website* Badan Riset dan Inovasi Nasional (BRIN) dan Badan Meteorologi Klimatologi Geologi (BMKG) yang nantinya digabungkan dengan target tingkat kepercayaan yang menjadi acuan terhadap terjadinya kebakaran. Lalu, data iklim digunakan untuk faktor dalam memprediksi saat klasifikasi. Penelitian ini menerapkan algoritma *Random Forest* dengan evaluasi *confusion matrix* dan *feature importance* untuk mendukung saat melakukan klasifikasi. Adapun penerapan pemodelan algoritma dan evaluasi dilakukan dengan menggunakan *Python* pada *Jupyter Notebook*.



Gambar 1. Tahapan Alur Penelitian

2.1. Web Scraping

Web Scraping merupakan teknik pengambilan data dari internet secara otomatis yang berupa halaman website yang menggunakan HTML atau XHTML [11]. *Web Scraping* melakukan proses pengambilan data tertentu dari internet. Pada penelitian ini *Web Scraping* dilakukan untuk mengambil data titik panas yang berasal dari *website* BRIN dan data iklim yang berasal dari *website* BMKG. Data nantinya akan dimanfaatkan untuk dianalisis lebih lanjut untuk mencapai hasil yang memuaskan [12].

2.2. Data Pre-Processing

Data Pre-Processing adalah proses untuk menghindarkan *data noise* agar dapat dilanjutkan ke tahapan selanjutnya [13][14]. *Data Pre Processing* mempunyai beberapa tahapan seperti pembersihan data, *handling missing value*, dan transformasi data. pembersihan data merupakan teknik yang berguna untuk menangani data yang tidak konsisten agar dapat menghasilkan informasi yang kuat [15]. Lalu, *handling missing value* dengan beberapa cara seperti mengisi *missing value* dengan data atau informasi yang ada [16]. Dalam imputasi sebuah *data time series* dapat dilakukan pengisian nilai kosong menggunakan *forward filling*. *Forward filling* mempunyai cara seperti memasukkan data menggunakan nilai akhir yang tersedia dari data sebelum sebuah

kekosongan data [17][18]. Sedangkan transformasi data merupakan data yang bertipe sama atau data ordinal agar dataset menjadi data yang lebih sederhana [19].

2.3. Splitting Data

Splitting data yaitu teknik membagi antara *data training* dan *data testing*. Teknik ini berpengaruh juga dalam menghasilkan kinerja model klasifikasi pada algoritma *machine learning* [20]. *Data training* adalah data aktual yang benar adanya, sedangkan *data testing* adalah data yang digunakan untuk mengukur kinerja berhasilnya klasifikasi. Pada tahap pengumpulan data dan waktu proses yang lebih efisien dapat dicapai dengan menggabungkan hanya fitur penting. Teknik *feature selection* (FS) diusulkan untuk menentukan atau menggabungkan fitur-fitur yang penting. Pendekatan FS ini dapat menghindari *overfitting*, kinerja model buruk, dan waktu komputasi yang lama, terutama ketika data memiliki dimensi yang tinggi dengan banyak fitur. Setelah dilakukan *feature selection* nantinya akan terdapat dua kelas yaitu data dependen dan independen. Teknik FS mengaitkan pemilihan secara otomatis maupun manual, fitur-fitur yang memiliki dampak terbesar akan memiliki dampak terbesar juga pada variabel target. FS pada dasarnya akan melihat *feature importance score* [21]. *Data training* dan *data testing* nantinya diberikan pada model lalu hasil pembelajaran tersebut untuk memproses dataset baru yang disebut *testing* [22].

2.4. Random Forest

Random Forest yaitu salah satu algoritma dengan metode klasifikasi yang tercipta dari pengembangan metode *Decision Tree*. *Random Forest* masuk ke dalam kategori *Supervised Learning* yang digunakan untuk klasifikasi atau prediksi [23], [24]. Algoritma ini melakukan teknik yang disebut pohon prediksi yang setiap pohon keputusan bergantung pada Banyaknya jumlah pohon akan mempengaruhi hasil akhir dan metode *Random Forest* menggunakan klasifikasi berdasarkan hasil seleksi dari pohon prediksi yang setiap pohonnya bergantung pada nilai acak pada tiap pohon prediksinya [25], [26], [27]. Dalam algoritma *Random Forest* terdapat seleksi fitur yang dapat mengambil fitur terbaik untuk meningkatkan performa terhadap model klasifikasi. Fitur ini berguna pada data dengan parameter yang memiliki data kompleks dan metode ini sangat akurat dalam melakukan prediksi, algoritma ini juga dapat menangani inputan variabel yang banyak [28], [29], [30]. Pada beberapa penelitian tentang *Random Forest* banyak digunakan untuk klasifikasi *data imbalance* karena memiliki performa kinerja yang baik [31].

Pada pohon prediksi tiap-tiap pohonnya memiliki keputusan bergantung dengan nilai random vektor dan dijadikan sebagai sampel acak pada semua pohon. Dalam algoritma *Random Forest* memiliki rumusan yang digunakan dalam penelitian ini sebagai berikut:

$$GI = 1 - \sum (P_i^2) \quad (1)$$

P_i merupakan probabilitas dari sebuah *tuple* untuk diklasifikasikan ke dalam kategori atau kelas tertentu. Setiap pohon keputusan dalam *Random Forest*, Gini Index (*GI*) untuk setiap atribut dihitung dan yang memiliki *GI* terendah diambil sebagai simpul pemisah. Keputusan akhir diambil dengan menggunakan sistem pemungutan suara mayoritas yang menghitung prediksi banyaknya pohon keputusan [32]. Adapun algoritma *Random Forest* memiliki rumusan seperti berikut :

$$l(y) = \operatorname{argmax}_c (\sum I_n n(y) = c) \quad (2)$$

Algoritma dengan rumusan penyelesaian masalah ini mempermudah proses semua data yang tersedia [33].

2.5. Klasifikasi

Klasifikasi merupakan proses penemuan sebuah model atau definisi persamaan karakteristik dalam suatu kelas pada data yang bertujuan untuk memprediksi kelas dari dataset. Klasifikasi ini merupakan salah satu teknik yang berfungsi untuk memprediksi kelas dari suatu objek [34].

2.6. Evaluasi Model

Model yang baik memiliki akurasi, precision, dan recall yang seimbang dan memiliki hasil tinggi, serta waktu proses komputasi yang efisien. Untuk mengetahui keefektifan model, prosedur evaluasi harus dilakukan menggunakan *confusion matrix* dan *feature importance* [35].

2.6.1. Confusion Matrix

Confusion Matrix merupakan tabulasi yang biasanya berfungsi untuk evaluasi kinerja algoritma klasifikasi. *Confusion matrix*

digunakan sebagai evaluasi algoritma *machine learning* yang digunakan untuk menyimpulkan masalah klasifikasi [36]. Gambaran tabulasi dari *Confusion Matrix* sebagai berikut :

Tabel 1. *Confusion Matrix*

Kelas Aktual	Kelas Prediksi	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Confusion matrix menghasilkan perhitungan untuk evaluasi kinerja algoritma klasifikasi setelah mengetahui nilai yang terdapat dalam *confusion matrix* seperti *accuracy*, *recall*, *precision*, dan *F1-score* [37]. Berikut adalah perhitungannya :

Accuracy, nilai perhitungan dari seberapa tepat kinerja yang dilakukan model algoritma klasifikasi. Hasil *accuracy* menunjukkan persentase dari hasil yang benar.

$$\text{accuracy} = \frac{TP+T}{TP+FP+FN+TN} \times 100\% \quad (3)$$

Precision, nilai yang menghitung seberapa tepat kinerja model untuk mengidentifikasi sampel positif dari data terprediksi.

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

Recall, nilai yang menghitung seberapa tepat model untuk mengidentifikasi sampel positif dari data aktual.

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

F1-score, perhitungan dalam mengukur kinerja model klasifikasi.

$$F1 - \text{score} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \times 100\% \quad (6)$$

Pada *confusion matrix* terdapat istilah seperti TP adalah data yang nilainya *positive* lalu diprediksi benar sebagai *positive*. Sedangkan FP adalah data yang bernilai *negative* lalu diprediksi sebagai *positive*. Kemudian FN adalah data yang mempunyai nilai *positive* tetapi diprediksi sebagai *negative*. Terakhir, TN adalah data yang bernilai *negative* dan diprediksi juga benar *negative* [38].

2.6.2. Feature Importance

Pada banyaknya studi kasus yang menggunakan *Random Forest*, *Random Forest* juga mempunyai metode untuk mengukur *feature importance* sebagai fitur tambahan yang berguna [39], [40]. Metode ini digunakan untuk mengukur *variable importance* yang dapat dilihat dari *overall variable importance*, *permutation variable importance*, dan *tree interpreter variable importance*. Penting mengetahui fitur keseluruhan dalam model pohon keputusan, seperti *Random Forest*, dihitung berdasarkan seberapa besar sebuah fitur dapat mengurangi ketidakhomogenan (*impurity*) pada setiap simpul pohon. Ketidakhomogenan, yang mengukur seberapa homogen data di suatu simpul, diharapkan berkurang saat kita turun lebih dalam ke tingkat pohon yang lebih spesifik. Pentingnya suatu fitur ditentukan dari besarnya penurunan ketidakhomogenan ini, dan dihitung dengan memberikan bobot sesuai dengan probabilitas mencapai simpul yang bersangkutan. Dengan demikian, pentingnya *variable* dapat diukur secara

objektif, memberikan wawasan tentang *variable* mana yang paling berkontribusi terhadap keputusan yang diambil oleh model [41].

3. HASIL

3.1. Pengumpulan Data

Penelitian ini menggunakan dua data yang didapatkan dari website BRIN dan BMKG. Data yang diambil antara lain yaitu data iklim (curah hujan, sinar matahari, kelembapan, temperatur) dan data titik panas (id, tanggal (WIB), waktu (WIB), lintang, bujur, tingkat kepercayaan, satelit, radius kemungkinan, kecamatan, kabupaten, provinsi, tipe) dari periode Januari 2023 sampai Desember 2023 (12 bulan). Kedua data tersebut memiliki peran yaitu data titik panas sebagai data dependen dan data iklim sebagai data independen atau menjadi faktor yang akan mempengaruhi data titik panas. Data tersebut akan dilakukan penggabungan data menjadi satu untuk mempermudah dalam pembuatan pemodelan.

3.1.1. Deskripsi Data Titik Panas

Data titik panas akan digunakan sebagai fitur dan menjadi kelas target sebagai variabel dependen. Berikut adalah deskripsi atribut dari data titik panas yang digunakan.

Tabel 2. Deskripsi Data Titik Panas

Kode Atribut	Nama Atribut	Keterangan
X0	id	no identifikasi unik
X1	tanggal (WIB)	tanggal titik panas terjadi
X2	waktu (WIB)	waktu terjadinya titik panas dalam zona Waktu Indonesia Barat (WIB)
X3	lintang	koordinat geografis
X4	bujur	koordinat geografis
X5	tingkat kepercayaan	tingkat kepercayaan
X6	satelit	terjadinya kebakaran
X7	radius kemungkinan	luas daerah yang terpengaruh titik panas
X8	kecamatan	lokasi terjadi titik panas
X9	kabupaten	lokasi terjadi titik panas
X10	provinsi	lokasi terjadi titik panas
X11	tipe	tanda titik panas

3.1.2. Deskripsi Data Iklim

Data iklim akan digunakan sebagai fitur dan menjadi kelas pengaruh sebagai variabel independen. Berikut adalah deskripsi atribut dari data iklim dapat dilihat pada Tabel 3.

Tabel 3. Deskripsi Data Iklim

Kode Atribut	Nama Atribut	Keterangan
Y0	tanggal	tanggal pengamatan
Y1	curah hujan	intensitas hujan
Y2	sinar matahari	jangka waktu sinar matahari
Y3	kelembapan	kadar air udara

Y4 temperatur suhu udara

3.2. Data Pre-Processing

Pada tahapan ini, pengolahan data dilakukan untuk kedua data yaitu data titik panas dan data iklim. Proses pengolahan data seperti pembersihan data untuk memilih atribut penting dan membuang atribut yang tidak dipakai, menangani nilai hilang dengan cara imputasi nilai menggunakan *forward fill*, penggabungan data untuk memudahkan proses pembuatan model menggunakan algoritma *Random Forest*. Terakhir dilakukan *labelling data* tingkat kepercayaan menjadi kelas kebakaran dan tidak kebakaran dalam bentuk biner.

3.2.1. Data Titik Panas

Proses awal yaitu melakukan *load data* titik panas dari data yang sudah diambil dari website BRIN hotspot yang didalamnya terdapat atribut yang dapat dilihat pada Gambar 2.

```
import pandas as pd

# Membaca file data titik panas
df1 = pd.read_excel("titikpanas-tahun2023-januari-desember-used.xlsx")

# Menampilkan data
df1
```

id	tanggal (WIB)	waktu (WIB)	lintang	bujur	tingkat kepercayaan	satelit	radius kemungkinan	kecamatan	kabupaten	provinsi	tipe	
0	3758874	2023-01-07	13.30.22	-3.272473	105.342690	8	noaa20	1125	Tulang Selapan	Ogan Komering ltr	Sumatera Selatan	Cluster
1	3759266	2023-01-13	13.17.50	-3.869538	105.918377	8	noaa20	1280	Pematang Panggang	Ogan Komering ltr	Sumatera Selatan	Cluster
2	3759398	2023-01-13	13.17.50	-3.349391	105.913872	8	noaa20	1125	Tulang Selapan	Ogan Komering ltr	Sumatera Selatan	Cluster
3	3760206	2023-01-28	13.36.58	-3.283174	105.918852	8	noaa20	1125	Pampangan	Ogan Komering ltr	Sumatera Selatan	Cluster
4	3760207	2023-01-28	13.36.58	-3.241103	105.163788	8	noaa20	1125	Pampangan	Ogan Komering ltr	Sumatera Selatan	Cluster
...
4339	4039839	2023-12-29	13.24.49	-3.494742	104.928200	8	noaa20	1125	Pedamaran	Ogan Komering ltr	Sumatera Selatan	Cluster
4340	4038882	2023-12-30	13.37.36	-3.458313	105.490181	8	noaa20	1125	Cengal	Ogan Komering ltr	Sumatera Selatan	Cluster
4341	4039026	2023-12-27	13.43.08	-3.185282	105.030762	8	snpp	1125	Pampangan	Ogan Komering ltr	Sumatera Selatan	Cluster
4342	4038740	2023-12-31	01.31.41	-4.203616	104.863403	8	snpp	1280	Mesuji	Ogan Komering ltr	Sumatera Selatan	Cluster
4343	4038868	2023-12-31	13.18.39	-3.635983	105.747276	8	noaa20	1125	Cengal	Ogan Komering ltr	Sumatera Selatan	Cluster

Gambar 2. Load Data Titik Panas

Setelah *load data* dilakukan penghapusan atribut yang tidak digunakan untuk mengurangi kesalahan pada saat pemodelan nantinya, terdapat beberapa atribut yang dihapus pada data titik panas. Atribut tersebut dihapus karena tidak digunakan pada penelitian ini. Atribut yang dipakai pada data titik panas adalah tanggal (WIB) dan tingkat kepercayaan. Penghapusan atribut dapat dilihat pada Gambar 3.

```
# Menghapus kolom satelit dan kecamatan
df1 = df1.drop(['id', 'waktu (WIB)', 'lintang', 'bujur', 'satelit', 'radius kemungkinan', 'kecamatan', 'kabupaten', 'provinsi', 'tipe'], axis=1)
df1.head()
```

tanggal (WIB)	tingkat kepercayaan	
0	2023-01-07	8
1	2023-01-13	8
2	2023-01-13	8
3	2023-01-28	8
4	2023-01-28	8

Gambar 3. Penghapusan Atribut Data Titik Panas

Selanjutnya merupakan tahapan *formatting data*. Pada data titik panas memiliki kolom tanggal (WIB) yang dimana tipe data kolom tersebut diubah menjadi datetime dan format tanggal menjadi tahun-bulan-tanggal. Selain itu nama kolom tanggal (WIB) diubah menjadi tanggal. pada Gambar 4 menunjukkan *formatting data* kolom tanggal.

```
# Ubah format tanggal menjadi datetime object
df1["tanggal (WIB)"] = pd.to_datetime(df1["tanggal (WIB)"], format="%Y-%m-%d")

# mengubah nama "tanggal (WIB)" menjadi tanggal
df1.rename(columns={'tanggal (WIB)': 'tanggal'}, inplace=True)

df1.head()
```

	tanggal	tingkat kepercayaan
0	2023-01-07	8
1	2023-01-13	8
2	2023-01-13	8
3	2023-01-28	8
4	2023-01-28	8

Gambar 4. Formatting Data Titik Panas

3.2.2. Data Iklim

Data iklim yang sudah diambil dari *website* BMKG dilakukan tahapan awal yaitu *load data* iklim untuk dilakukan pengolahan. Atribut pada data iklim yaitu tanggal, curah hujan, sinar matahari, kelembapan, dan temperatur. Proses *load data* iklim dilakukan pada Gambar 5.

```
import pandas as pd
# Baca file
df2 = pd.read_excel("iklim-full-tahun2023.xlsx")

# Tampilkan data dari file
df2
```

	tanggal	curah_hujan	sinar_matahari	kelembapan	temperatur
0	2023-01-01	0.0	6.6	81.0	27.8
1	2023-01-02	0.9	5.3	80.0	27.4
2	2023-01-03	15.0	3.4	89.0	25.4
3	2023-01-04	0.0	0.0	87.0	25.4
4	2023-01-05	NaN	0.0	83.0	27.0
...
360	2023-12-27	NaN	4.3	78.0	30.1
361	2023-12-28	NaN	4.0	81.0	29.0
362	2023-12-29	4.5	2.2	75.0	31.1
363	2023-12-30	62.6	5.6	81.0	29.3
364	2023-12-31	61.7	6.4	80.0	29.2

Gambar 5. Load Data Iklim

Setelah *load data*, dilakukan pengecekan terlebih dahulu pada data iklim dengan melihat deskripsi data yang akan menampilkan jumlah data kosong yang terdapat dalam data iklim. Proses ini untuk mengetahui nilai kosong yang nantinya akan dibersihkan karena dapat mengganggu pemodelan. Pada Gambar 6 memperlihatkan nilai *null* pada atribut curah hujan, sinar matahari, kelembapan, dan temperature.

```
# Cek anomali data
df2.isnull().sum()
```

tanggal	0
curah_hujan	118
sinar_matahari	2
kelembapan	1
temperatur	1
dtype: int64	

Gambar 6. Deskripsi Data Iklim

Proses selanjutnya adalah mengisi nilai kosong pada data iklim menggunakan metode *forward fill* menjadi pilihan yang dipilih karena pada data iklim seringkali terdapat pola alami dan memiliki nilai yang sama pada data sebelumnya, sehingga dengan menggunakan metode *forward fill* nilai alami data iklim dapat dipertahankan. Proses mengisi nilai kosong menggunakan metode *forward fill* diperlihatkan pada Gambar 7.

```
# Mengisi nilai hilang menggunakan metode Forward Fill
df2.fillna(method='ffill', inplace=True)
```

```
df2
```

	tanggal	curah_hujan	sinar_matahari	kelembapan	temperatur
0	2023-01-01	0.0	6.6	81.0	27.8
1	2023-01-02	0.9	5.3	80.0	27.4
2	2023-01-03	15.0	3.4	89.0	25.4
3	2023-01-04	0.0	0.0	87.0	25.4
4	2023-01-05	0.0	0.0	83.0	27.0
...
360	2023-12-27	36.9	4.3	78.0	30.1
361	2023-12-28	36.9	4.0	81.0	29.0
362	2023-12-29	4.5	2.2	75.0	31.1
363	2023-12-30	62.6	5.6	81.0	29.3
364	2023-12-31	61.7	6.4	80.0	29.2

Gambar 7. Imputasi Nilai Data Iklim

3.2.3. Penggabungan Data Titik Panas dan Data Iklim

Tahap selanjutnya adalah menggabungkan data titik panas dan data iklim. Atribut tanggal yang terdapat data titik panas dan data iklim akan membuat data iklim tergantung pada banyaknya data titik panas sesuai tanggalnya. Pada Gambar 8 merupakan proses penggabungan data titik panas dan data iklim.

```
df1["tanggal"] = pd.to_datetime(df1["tanggal"])
df2["tanggal"] = pd.to_datetime(df2["tanggal"])

datamerge = pd.merge(df1, df2, on='tanggal', how='left')

datamerge
```

	tanggal	tingkat kepercayaan	curah_hujan	sinar_matahari	kelembapan	temperatur
0	2023-01-07	8	34.5	2.1	85.0	27.4
1	2023-01-13	8	1.8	2.5	81.0	27.9
2	2023-01-13	8	1.8	2.5	81.0	27.9
3	2023-01-28	8	1.0	1.9	86.0	26.7
4	2023-01-28	8	1.0	1.9	86.0	26.7
...
4339	2023-12-20	8	0.6	1.8	76.0	30.4
4340	2023-12-30	8	62.6	5.6	81.0	29.3
4341	2023-12-27	8	36.9	4.3	78.0	30.1
4342	2023-12-31	8	61.7	6.4	80.0	29.2
4343	2023-12-31	8	61.7	6.4	80.0	29.2

Gambar 8. Penggabungan Data Titik Panas dan Data Iklim

3.2.4. Data Gabungan

Setelah melakukan penggabungan pada data titik panas dan data iklim yaitu melakukan *load data* pada data gabungan untuk dilakukan pengolahan lebih lanjut. *Load data* dilakukan seperti pada Gambar 9.

```
# Labeling data tingkat kepercayaan menjadi biner
# Nilai >= 8 dianggap sebagai kebakaran (1), selain itu tidak (0)
data['kebakaran'] = (data['tingkat kepercayaan'] >= 8).astype(int)

# Menampilkan beberapa baris data setelah transformasi
data.head()
```

	tanggal	tingkat kepercayaan	satelit	kecamatan	curah_hujan	sinar_matahari	kelembapan	temperatur	kebakaran
0	2023-01-07	8	noaa20	Tulang Selapan	34.5	2.1	85	27.4	1
1	2023-01-13	8	noaa20	Pematang Panggang	1.8	2.5	81	27.9	1
2	2023-01-13	8	noaa20	Tulang Selapan	1.8	2.5	81	27.9	1
3	2023-01-28	8	noaa20	Pampangan	1.0	1.9	86	26.7	1
4	2023-01-28	8	noaa20	Pampangan	1.0	1.9	86	26.7	1

Gambar 9. Load Data Gabungan

3.2.5. Labelling Data

Tahapan setelah *load data* gabungan yaitu menentukan kelas kebakaran dan tidak kebakaran dari kolom tingkat kepercayaan. Menambahkan kolom kebakaran dari nilai kolom tingkat kepercayaan yang akan digunakan sebagai variabel y. Jika terdapat nilai pada kolom tingkat kepercayaan kurang dari 7 maka akan diubah menjadi nol (0) dan jika nilai lebih dari 8 maka akan diubah menjadi satu (1). Pada Gambar 10 yaitu proses dan hasil *labelling data*.

```
# Labeling data tingkat kepercayaan menjadi biner
# Nilai >= 8 dianggap sebagai kebakaran (1), selain itu tidak (0)
data['kebakaran'] = (data['tingkat kepercayaan'] >= 8).astype(int)

# Menampilkan beberapa baris data setelah transformasi
data.head()
```

	tanggal	tingkat kepercayaan	satelit	kecamatan	curah_hujan	sinar_matahari	kelembapan	temperatur	kebakaran
0	2023-01-07	8	noaa20	Tulang Selapan	34.5	2.1	85	27.4	1
1	2023-01-13	8	noaa20	Pematang Panggang	1.8	2.5	81	27.9	1
2	2023-01-13	8	noaa20	Tulang Selapan	1.8	2.5	81	27.9	1
3	2023-01-28	8	noaa20	Pampangan	1.0	1.9	86	26.7	1
4	2023-01-28	8	noaa20	Pampangan	1.0	1.9	86	26.7	1

Gambar 10. Labelling Data

3.3. Pengembangan Model

Tahapan awal dalam proses pembuatan model yaitu proses pembagian dataset menjadi data uji dan data latih, terdapat variabel x (curah hujan, sinar matahari, kelembapan, temperatur) dan variabel y (kebakaran). Dalam pembagian data tersebut juga dapat dilihat dimensi atau ukurannya. Tahap pertama pemodelan dapat dilihat pada Gambar 11.

```
from sklearn.model_selection import train_test_split

# Memilih fitur dan Label
X = data[['curah_hujan', 'sinar_matahari', 'kelembapan', 'temperatur']]
y = data['kebakaran']

# Membagi data menjadi set pelatihan dan pengujian
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Menampilkan ukuran dari setiap set
(X_train.shape, y_train.shape), (X_test.shape, y_test.shape)

(((3040, 4), (3040,)), ((1304, 4), (1304,)))
```

Gambar 11. Splitting Data

Sebelum lanjut ke tahapan pemodelan menggunakan algoritma *Random Forest* dilakukan proses import sklearn, metrics, seaborn, matplotlib, confusion_matrix, classification_report, accuracy_score, dan , algoritma *Random Forest* menggunakan *RandomForestClassifier*. Pada Gambar 12 memperlihatkan kumpulan library yang digunakan pada penelitian ini.

```
import sklearn
from sklearn import metrics
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score

import seaborn as sns
import matplotlib.pyplot as plt
```

Gambar 12. Import Library pada Jupyter Notebook

3.3.1. Model Random Forest

Selanjutnya yaitu tahapan pembuatan model *Random Forest*, melakukan proses pelatihan pada model pada *training set* dengan label yang sesuai menggunakan *rf_model.fit*. Setelah itu, melakukan prediksi variabel target pada data pengujian atau *testing*. Terlihat proses pembuatan model *Random Forest* pada Gambar 13.

```
# Membuat dan melatih model Random Forest
rf_model = RandomForestClassifier(n_estimators=300, max_depth=10, criterion='gini', random_state=42)
rf_model.fit(X_train, y_train)

# Membuat prediksi pada set pengujian
y_pred = rf_model.predict(X_test)

# Menghitung akurasi dan hasil klasifikasi
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(y_test, y_pred)

# Menampilkan hasil
print("\nAkurasi model Random Forest:")
print(accuracy)
print("\nLaporan klasifikasi:")
print(classification_rep)
```

Gambar 13. Pembuatan Model Random Forest

Hasil akurasi sebesar 85.8% dari pemodelan *Random Forest* dengan data uji 30% ditunjukkan pada Gambar 14.

```
# Menampilkan hasil
print("\nAkurasi model Random Forest:")
print(accuracy)
```

Akurasi model Random Forest:
0.8588957055214724

Gambar 14. Hasil Akurasi *Random Forest* pada Data Uji 30%

Hasil akurasi sebesar 85.9% dari pemodelan *Random Forest* dengan data uji 20% ditunjukkan pada Gambar 15.

```
# Menampilkan hasil
print("\nAkurasi model Random Forest:")
print(accuracy)
```

Akurasi model Random Forest:
0.859608745684695

Gambar 15. Hasil Akurasi *Random Forest* pada Data Uji 20%

Hasil akurasi sebesar 86.4% dari pemodelan *Random Forest* dengan data uji 10% ditunjukkan pada Gambar 16.

```
# Menampilkan hasil
print("\nAkurasi model Random Forest:")
print(accuracy)
```

Akurasi model Random Forest:
0.864367816091954

Gambar 16. Hasil Akurasi *Random Forest* pada Data Uji 10%

3.4. Evaluasi

Setelah melakukan pemodelan *Random Forest* yaitu melakukan evaluasi menggunakan *confusion matrix* dalam bentuk heatmap akan memperlihatkan prediksi benar dan salah dengan cara melihat TP, TN, FP, FN dan laporan klasifikasi yang memberikan wawasan lebih dalam terhadap kinerja model wawasan tersebut yaitu *precision*, *recall*, dan *f1-score*. Pada Gambar 17 merupakan proses pembuatan *confusion matrix*.

```
# Membuat confusion matrix dari hasil terakhir
cm = confusion_matrix(y_test, y_pred)

# Plot confusion matrix
plt.figure(figsize=(6, 4))
sns.heatmap(cm, annot=True, fmt='g', cmap='Blues', xticklabels=['Tidak Kebakaran', 'Kebakaran'],
            yticklabels=['Tidak Kebakaran', 'Kebakaran'])
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.title('Confusion Matrix')
plt.show()

classification_rep = classification_report(y_test, y_pred)
print("\nLaporan klasifikasi:")
print(classification_rep)
```

Gambar 17. Pembuatan *Confusion Matrix*

Proses untuk menampilkan fitur penting dari model *Random Forest* yang nantinya akan diberikan bobot untuk setiap fitur pada model yang telah dilatih dapat dilihat pada Gambar 18.

```
# Mendapatkan fitur penting dari model Random Forest
feature_importances = rf_model.feature_importances_

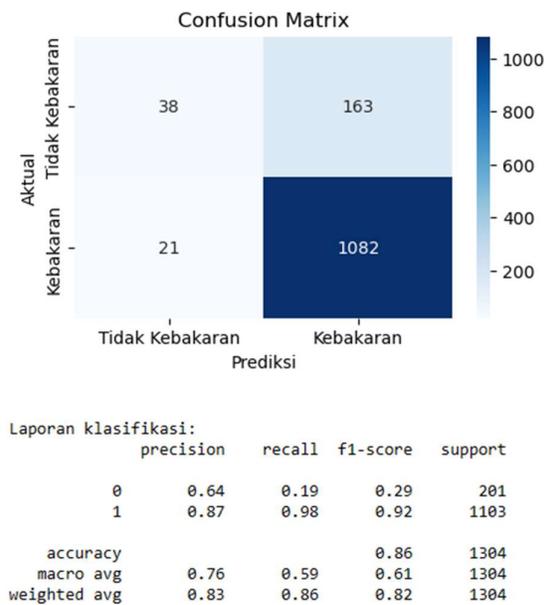
# Membuat dataframe untuk fitur penting
feature_importance_df = pd.DataFrame({
    'Feature': X.columns,
    'Importance': feature_importances
}).sort_values(by='Importance', ascending=False)

feature_importance_df
```

Gambar 18. Pembuatan *Feature Importance*

3.4.1. *Confusion Matrix*

Hasil model *Random Forest* pada data uji 30% dapat dilihat pada Gambar 19 yang mendapatkan nilai *true positive* sebanyak 38, *false negative* sebanyak 163, *false positive* sebanyak 21, dan *true negative* sebanyak 1082. Dalam *Classification Report* model memperoleh akurasi sebesar 86%, dengan kelas tidak kebakaran (0) memperoleh *precision* 64%; *recall* 19%; *f1-score* 29%, dan kelas kebakaran (1) memperoleh *precision* 87%; *recall* 98%; *f1-score* 92%.

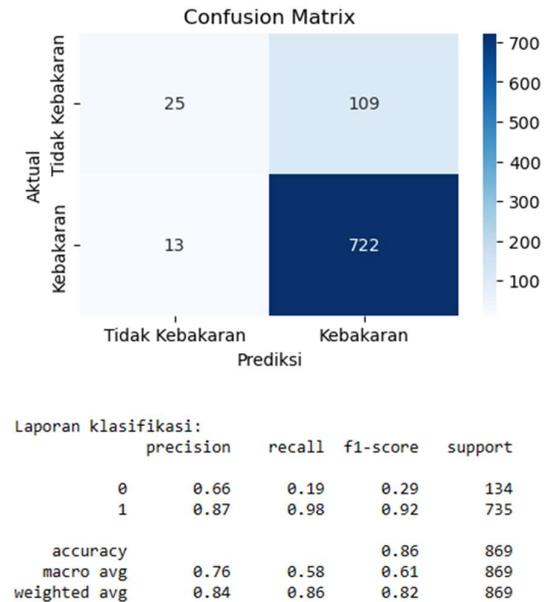


Gambar 19. Hasil *Confusion Matrix* pada Data Uji 30%

Hasil model *Random Forest* pada data uji 20% dapat dilihat pada Gambar 20 yang menghasilkan nilai *true positive* sebanyak 25, *false negative* sebanyak 109, *false positive* sebanyak 13, dan *true*

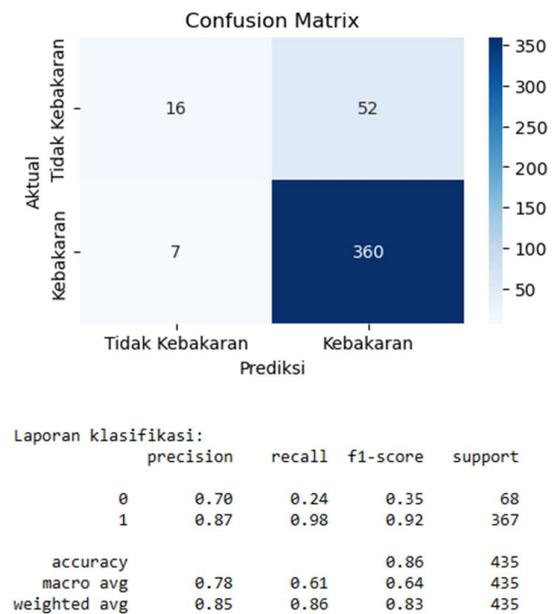
<https://doi.org/10.25077/TEKNOSI.v10i2.2024.145-155>

negative sebanyak 722. Dalam *Classification Report* model memperoleh akurasi sebesar 86%, dengan kelas tidak kebakaran (0) memperoleh *precision* 66%; *recall* 19%; *f1-score* 29%, dan kelas kebakaran (1) memperoleh *precision* 87%; *recall* 98%; *f1-score* 92%.



Gambar 20. Hasil *Confusion Matrix* pada Data Uji 20%

Hasil model *Random Forest* pada data uji 10% dapat dilihat pada Gambar 21 yang menghasilkan *true positive* sebanyak 16, *false negative* sebanyak 52, *false positive* sebanyak 7, dan *true negative* sebanyak 360. Dalam *Classification Report* model memperoleh akurasi sebesar 86%, dengan kelas tidak kebakaran (0) memperoleh *precision* 70%; *recall* 24%; *f1-score* 35%, dan kelas kebakaran (1) memperoleh *precision* 87%; *recall* 98%; *f1-score* 92%.



Gambar 21. Hasil *Confusion Matrix* pada Data Uji 10%

3.4.2. Feature Importance

Hasil *Feature Importance* dari model *Random Forest* yang sudah dilakukan berdasarkan variabel faktor yaitu berasal dari data iklim. Berdasarkan hasil *feature importance* bahwa fitur terpenting yang pertama yaitu sinar matahari 0.477475, temperatur 0.257086, kelembapan 0.200452, curah hujan 0.064987. Pada Gambar 22 memperlihatkan hasil *Feature Importance*.

Feature	Importance
sinar_matahari	0.477475
temperatur	0.257086
kelembapan	0.200452
curah_hujan	0.064987

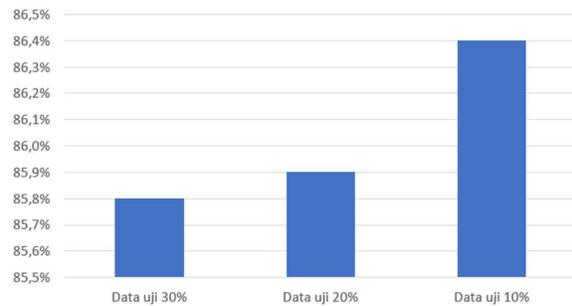
Gambar 22. Hasil *Feature Importance* Model *Random Forest*

4. PEMBAHASAN

4.1. Menentukan Hasil Model Terbaik

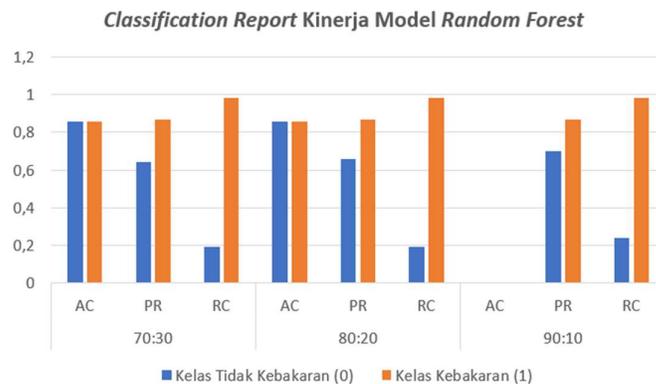
Berdasarkan penerapan model dan *splitting data* yang berbeda hasil penelitian ini menunjukkan bahwa kemampuan algoritma mengalami hasil yang stabil. Model *Random Forest* pada data uji 30% menghasilkan nilai akurasi sebesar 85.8%, pada data uji 20% menghasilkan nilai akurasi sebesar 85.9%. Sedangkan, pada data uji 10% mendapatkan nilai akurasi sebesar 86.4%. Grafik peningkatan akurasi tertinggi didapatkan pada data uji 10%.

Perbandingan akurasi tiap *splitting data* yang berbeda diperlihatkan pada Gambar 23.



Gambar 23. Grafik Perbandingan *Splitting Data* Model *Random Forest*

Selanjutnya, berdasarkan *Classification Report* evaluasi kinerja algoritma klasifikasi *Random Forest* pada rasio pembagian data yang berbeda (70:30, 80:20, 90:10). Parameter evaluasi yang diukur melibatkan accuracy (AC), precision (PR), dan recall (RC). Pertama, dalam pembagian data dengan rasio 70:30, *Random Forest* menunjukkan kemampuan yang baik dalam mengklasifikasikan data dengan akurasi sebesar 85.8%, presisi 87%, dan recall 98%. Kedua, pada pembagian data dengan rasio 80:20, *Random Forest* tetap mempertahankan performa yang unggul dengan akurasi 85.9%, presisi 87%, dan recall 98%. Ketiga, pada pembagian data dengan rasio 90:10, mencapai akurasi sebesar 86.4%, presisi 87%, dan recall 98%. Grafik hasil *Classification Report* dapat dilihat pada Gambar 24.



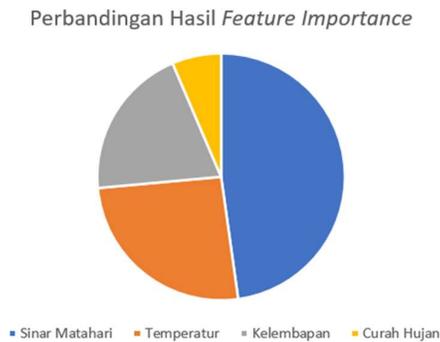
Gambar 24. Grafik *Classification Report*

Hasil pengklasifikasian data mengenai titik panas dan kondisi iklim untuk kategori kebakaran menunjukkan konsistensi yang sangat baik dalam performa pada berbagai pembagian data, termasuk rasio 70:30, 80:20, dan 90:10. Dengan tingkat akurasi sebesar 85.8% untuk rasio 70:30, 85.9% untuk rasio 80:20, dan 86.4% untuk rasio 90:10, model *Random Forest* menunjukkan kemampuan yang sangat baik dalam mengklasifikasikan data. Selain itu, dengan presisi mencapai 87% pada semua pembagian data, menandakan bahwa sebagian besar prediksi positif yang dibuat oleh model tersebut benar, dan recall yang tinggi sebesar 98% menunjukkan kemampuan model dalam mengidentifikasi sebagian besar kasus positif yang sebenarnya.

Sementara itu, hasil klasifikasi data untuk kategori bukan kebakaran menunjukkan performa yang memuaskan pada berbagai pembagian data. Meskipun recall cenderung rendah (antara 19% hingga 24%) pada semua rasio pembagian data, hal ini menunjukkan bahwa model mungkin mengalami kesulitan dalam mengidentifikasi sebagian besar kasus bukan kebakaran yang sebenarnya karena ketidakseimbangan antara kelas kebakaran dan bukan kebakaran. Namun, tingkat akurasi yang stabil (antara 85.8% hingga 86.4%) menunjukkan kemampuan model dalam melakukan klasifikasi secara keseluruhan. Meskipun tingkat presisi bervariasi (antara 64% hingga 70%), ini menandakan bahwa sebagian besar dari prediksi positif model benar untuk kategori bukan kebakaran.

4.2. Menentukan Fitur Terbaik

Terakhir yaitu membandingkan nilai fitur variabel X atau atribut dari data iklim yang dijadikan faktor untuk variabel Y. Perbandingan fitur terbaik dapat dilihat pada Gambar 25.



Gambar 25. Grafik *Feature Importance*

Dalam memprediksi kelas kebakaran dan tidak kebakaran memiliki fitur terpenting, urutan pertama yaitu fitur sinar matahari dengan nilai 0.477475, kedua yaitu fitur temperatur dengan nilai 0.257086, ketiga yaitu fitur kelembapan dengan nilai 0.200452, dan yang terakhir yaitu fitur curah hujan dengan nilai 0.064987. Berdasarkan hasil yang didapatkan terdapat fitur sinar matahari yang paling penting dalam memprediksi kelas kebakaran dan tidak kebakaran dengan nilai yang paling tinggi.

5. KESIMPULAN

Klasifikasi menggunakan algoritma *Random Forest* untuk memprediksi data titik panas berdasarkan faktor data iklim yang digunakan dalam penelitian ini memperoleh nilai akurasi 85,8%, presisi 87%, dan *recall* 98% dengan persentase *splitting data* 70:30 yang menunjukkan model menghitung seberapa tepat kinerja yang dilakukan model. Evaluasi lain juga didapatkan menggunakan *feature importance* dengan nilai paling tinggi yaitu sinar matahari dengan nilai 0.477475, kedua yaitu temperatur 0.257086, ketiga yaitu kelembapan dengan nilai 0.200452, keempat yaitu curah hujan dengan nilai 0.064987. Dengan penerapan model menggunakan metode *Random Forest*, Hasil penelitian dan evaluasi dari kasus kebakaran sangat baik sehingga informasi ini dapat digunakan untuk mengetahui faktor iklim apa yang menyebabkan kebakaran sehingga dapat dengan tepat melakukan solusi, keputusan manajemen penanggulangan kebakaran dengan tepat dan meminimalisir terjadinya kerugian.

DAFTAR PUSTAKA

- [1] S. Dian, H. Permana, G. Saputra, B. Arifitama, W. Caesarendra, and R. Rahim, "Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 7, pp. 4345–4357, 2022, doi: [10.1016/j.jksuci.2021.04.013](https://doi.org/10.1016/j.jksuci.2021.04.013).
- [2] A. Fitriansyah, "Sistem Informasi Pusat Data Dampak Kebakaran Hutan dan Lahan Berbasis Mobile Web di Provinsi Riau," *TEKNOSI*, vol. 03, no. 01, pp. 35–42, 2017.
- [3] Fitriyani, "Implementasi Forward Selection dan Bagging untuk Prediksi Kebakaran Hutan Menggunakan Algoritma Naïve Bayes," *J. Nas. Teknol. dan Sist. Inf.*, vol. 01, pp. 1–8, 2022, doi: [10.25077/TEKNOSI.v8i1.2021.1-8](https://doi.org/10.25077/TEKNOSI.v8i1.2021.1-8).
- [4] E. Abdul, S. Listia, A. Syukur, and M. Othman, "Forest fire spreading and carbon concentration identification in tropical region Indonesia," *Alexandria Eng. J.*, vol. 61, no. 2, pp. 1551–1561, 2022, doi: [10.1016/j.aej.2021.06.064](https://doi.org/10.1016/j.aej.2021.06.064).
- [5] Paloma Ruiz-Benito, Giorgio Vacchiano, Emily R. Lines, Christopher P.O. Reyer, Sophia Ratcliffe, Xavier Morin, Florian Hartig, Annikki Mäkelä, Rasoul Yousefpour, Jimena E. Chaves, Alicia Palacios-Orueta, Marta Benito-Garzón, Cesar Morales-Molino, J. Julio Camarero, Alistair S. Jump, Jens Kattge, Aleks Lehtonen, Andreas Ibrom, Harry J.F. Owen, Miguel A. Zavala, "Available and missing data to model impact of climate change on European forests," *Ecol. Modell.*, vol. 416, no. October 2019, p. 108870, 2020, doi: [10.1016/j.ecolmodel.2019.108870](https://doi.org/10.1016/j.ecolmodel.2019.108870).
- [6] T. A. Pratiwi, M. Irsyad, and R. Kurniawan, "Klasifikasi Kebakaran Hutan dan Lahan Menggunakan Algoritma Naïve Bayes (Studi Kasus: Provinsi Riau)," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 101, 2021, doi: [10.26418/justin.v9i2.42823](https://doi.org/10.26418/justin.v9i2.42823).
- [7] R. O. Mardiyanto, F. Fitriani, R. J. Purnomo, K. Kusriani, and D. Maulina, "Pemetaan Lokasi Kebakaran Hutan Dan Lahan Di Ntb Dengan Menggunakan Algoritma Naive Bayes," *Tek. Teknol. Inf. dan Multimed.*, vol. 2, no. 2, pp. 69–75, 2021, doi: [10.46764/teknimedia.v2i2.44](https://doi.org/10.46764/teknimedia.v2i2.44).
- [8] Liqing Si, Lifu Shu, Mingyu Wang, Fengjun Zhao, Feng Chen, Weike Li, Wei Li, "Study on forest fire danger prediction in plateau mountainous forest area," *Nat. Hazards Res.*, vol. 2, no. 1, pp. 25–32, 2022, doi: [10.1016/j.nhres.2022.01.002](https://doi.org/10.1016/j.nhres.2022.01.002).
- [9] Srikanta Sannigrahi, Francesco Pilla, Bidroha Basu, Arunima Sarkar Basu, Konika Sarkar, Suman Chakraborti, Pawan Kumar Joshi, Qi Zhang, Ying Wang, Sandeep Bhatt, Anand Bhatt, Shouvik Jha, Saskia Keesstra, P.S. Roy, "Examining the effects of forest fire on terrestrial carbon emission and ecosystem production in India using remote sensing approaches," *Sci. Total Environ.*, vol. 725, no. March, p. 138331, 2020, doi: [10.1016/j.scitotenv.2020.138331](https://doi.org/10.1016/j.scitotenv.2020.138331).
- [10] S. M. Rezfiko Agdialta, Dodo Gunawan, Amsari Mudzakir Setiawan, "Pemantauan Tingkat Bahaya Kebakaran Hutan Dan Lahan Dengan Indeks Kekeringan Keetch-Byram Drought Index Selama Periode El-Niño Di Provinsi Sumatera Selatan," *J. Penelit. Kehutan. Sumatrana*, vol. 2, no. 1, pp. 1–13, 2020.
- [11] Nurdin, M. Hutomi, M. Qamal, and Bustami, "Sistem Pengecekan Toko Online Asli atau Dropship pada Shopee," *J. Rekayasa Sist. dan Teknol. Inf.*, vol. 1, no. 10, pp. 1117–1123, 2021.
- [12] S. P. Salim Satriajati, Satria Bagus Panuntun, "Implementasi Web Scraping Dalam Pengumpulan

- Berita Kriminal Pada Masa Pandemi COVID-19 (Implementation of Web Scraping in Criminal News Collection during Covid-19 Pandemic),” pp. 300–308, 2020.
- [13] S. Kumari, D. Kumar, and M. Mittal, “An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier,” *Int. J. Cogn. Comput. Eng.*, vol. 2, no. January, pp. 40–46, 2021, doi: [10.1016/j.ijcce.2021.01.001](https://doi.org/10.1016/j.ijcce.2021.01.001).
- [14] F. Alghifari and D. Juardi, “Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes,” *J. Ilm. Inform.*, 2021.
- [15] R. F. P. D. Ikhsan Romli, “Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit ISPA,” *Indones. J. Bus. Intell.*, vol. 4, no. 1, pp. 10–15, 2021, doi: [10.21927/ijubi.v4i1.1727](https://doi.org/10.21927/ijubi.v4i1.1727).
- [16] J. Melvin and A. Soraya, “Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit,” *J. Ris. Rumpun Mat. dan Ilmu Pengetah. Alam*, vol. 2, no. 2, 2023, doi: [10.55606/jurrimipa.v2i2.1336](https://doi.org/10.55606/jurrimipa.v2i2.1336).
- [17] Minh Nguyen, Tong He, Lijun An, Daniel C. Alexander, Jiashi Feng, B.T. Thomas Yeo, “Predicting Alzheimer’s disease progression using deep recurrent neural,” *Neuroimage*, vol. 222, no. August, p. 117203, 2020, doi: [10.1016/j.neuroimage.2020.117203](https://doi.org/10.1016/j.neuroimage.2020.117203).
- [18] A. T. Tunkiel, D. Sui, and T. Wiktorski, “Impact of data pre-processing techniques on recurrent neural network performance in context of real-time drilling logs in an automated prediction framework,” *J. Pet. Sci. Eng.*, vol. 208, no. PE, p. 109760, 2022, doi: [10.1016/j.petrol.2021.109760](https://doi.org/10.1016/j.petrol.2021.109760).
- [19] S. W. Yohana Tri Widayati, Yani Prihati, “Analisis dan Komparasi Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Loyalitas Pelanggan MNC Play Kota Semarang,” *TRANSFORMATIKA*, vol. 18, no. 2, pp. 161–172, 2021.
- [20] R. Oktafiani, A. Hermawan, and D. Avianto, “Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning,” *J. Sains dan Inform.*, vol. 9, no. April, pp. 19–28, 2023, doi: [10.34128/jsi.v9i1.622](https://doi.org/10.34128/jsi.v9i1.622).
- [21] M. D. Anasanti, K. Hilyati, and A. Novtariyani, “Exploring feature selection techniques on Classification Algorithms for Predicting Type 2 Diabetes at Early Stage,” *J. Rekayasa Sist. dan Teknol. Inf.*, vol. 5, no. 158, pp. 832–839, 2022.
- [22] B. N. Azmi, A. Hermawan, and D. Avianto, “Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver,” *J. Teknol. Inf. dan Multimed.*, vol. 4, no. 4, pp. 281–290, 2023, doi: [10.35746/jtim.v4i4.298](https://doi.org/10.35746/jtim.v4i4.298)
- [23] Jussi Pirneskoskia, Joonas Tamminenb, Antti Kallonenb, Jouni Nurmi, Markku Kuusimaa, Klaus T. Olkkolad, Sanna Hoppu, “Random forest machine learning method outperforms predicting one-day mortality : A retrospective study,” *Resusc. Plus*, vol. 4, no. May, p. 100046, 2020, doi: [10.1016/j.resplu.2020.100046](https://doi.org/10.1016/j.resplu.2020.100046).
- [24] Cafer Mert Yeşilkanat Science, “Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm,” *Chaos , Solitons Fractals*, vol. 140, 2020, doi: [10.1016/j.chaos.2020.110210](https://doi.org/10.1016/j.chaos.2020.110210).
- [25] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, “Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah,” *J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020.
- [26] A. M. Tamrizal and A. Yaqin, “Perbandingan Algoritma Naïve Bayes , K-Nearest Neighbors dan Random Forest untuk Klasifikasi Sentimen Terhadap BPJS Kesehatan pada Media Twitter,” *J. Telekomun. dan Komput.*, vol. 12, no. 1, pp. 1–10, 2022, doi: [10.22441/incomtech.v12i1.13642](https://doi.org/10.22441/incomtech.v12i1.13642).
- [27] R. B. Kurniabudi, Deris Stiawan, Darmawijoyo, Mohd Yazid Bin Idris, Sarjon Defit, Yaya Sudarya Triana, “Improvement of Attack Detection Performance on the Internet of Things Network with PSO-Search and Random Forest,” *J. Comput. Sci.*, 2022, doi: [10.1016/j.jocs.2022.101833](https://doi.org/10.1016/j.jocs.2022.101833).
- [28] M. D. P. Gde Agung Brahmata Suryanegara, Adiwijaya, “Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi,” *J. Rekayasa Sist. dan Teknol. Inf.*, vol. 1, no. 10, pp. 114–122, 2021.
- [29] W. Aezwani, W. Abu, N. Laila, N. B. Josdi, M. B. Man, and Y. Sudarya, “An Evaluation of Artificial Neural Networks and Random Forests for Heart Disease Prediction,” *J. Human Univerity (Natural Sci.)*, vol. 49, 2022, doi: [10.55463/issn.1674-2974.49.2.4](https://doi.org/10.55463/issn.1674-2974.49.2.4).
- [30] R. Matsuo, T. Yamazaki, M. Suzuki, H. Toyama, and K. Araki, “A random forest algorithm-based approach to capture latent decision variables and their cutoff values,” *J. Biomed. Inform.*, vol. 110, no. August, p. 103548, 2020, doi: [10.1016/j.jbi.2020.103548](https://doi.org/10.1016/j.jbi.2020.103548).
- [31] W. H. Yoga Religia, Agung Nugroho, “Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing,” *J. Rekayasa Sist. dan Teknol. Inf.*, vol. 1, no. 10, pp. 187–192, 2021.
- [32] S. Sreejith, H. K. Nehemiah, and A. Kannan, “A clinical decision support system for polycystic ovarian syndrome using red deer algorithm and random forest classifier,” *Healthc. Anal.*, vol. 2, no. August, p. 100102, 2022, doi: [10.1016/j.health.2022.100102](https://doi.org/10.1016/j.health.2022.100102).
- [33] Z. Susanti, P. Sirait, and E. S. Panjaitan, “Peningkatan Kinerja Random Forest Melalui Seleksi Fitur Secara Pca Untuk Mendeteksi Penyakit Diabetes Tahap Awal,” vol. 4, no. 3, pp. 51–56, 2023.
- [34] S. Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, “Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random,” *Sains dan Teknol. Inf.*, 2022, doi: [10.33372/stn.v8i2.885](https://doi.org/10.33372/stn.v8i2.885).

- [35] M. Akhas Rahmadayan, "Seleksi Fitur pada Supervised Learning : Klasifikasi Prestasi Belajar Mahasiswa Saat dan Pasca Pandemi COVID-19," *J. Nas. Teknol. dan Sist. Inf.*, vol. 01, pp. 21–32, 2023, doi: [TEKNOSI.v9i1.2023.21-32](https://doi.org/10.35580/TEKNOSI.v9i1.2023.21-32).
- [36] S. Amaliah and M. Nusrang, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng," *J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 2, pp. 121–127, 2022, doi: [10.35580/variansiunm31](https://doi.org/10.35580/variansiunm31).
- [37] R. R. Salam, M. F. Jamil, and Y. Ibrahim, "Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine," *Indones. J. Mach. Learn. Comput. Sci. J.*, vol. 3, no. April, pp. 27–35, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/article/view/590>
- [38] C. Wahyu Surya Gemilang, Purwantoro, "Analisis Sentimen Pengguna Instagram Pada Calon Presiden 2024 Menggunakan Algoritma Support Vector Machine," *J. Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2849–2855, 2024.
- [39] C. Aldrich and L. Auret, "Fault detection and diagnosis with random forest feature extraction and variable importance methods," *IFAC Proc. Vol.*, vol. 43, no. 9, pp. 79–86, 2014, doi: [10.3182/20100802-3-ZA-2014.00020](https://doi.org/10.3182/20100802-3-ZA-2014.00020).
- [40] Y. Zhao, W. Zhu, P. Wei, P. Fang, X. Zhang, and N. Yan, "Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period," *Ecol. Indic.*, vol. 135, p. 108529, 2022, doi: [10.1016/j.ecolind.2021.108529](https://doi.org/10.1016/j.ecolind.2021.108529).
- [41] H. Alsagri and M. Ykhlef, "Quantifying Feature Importance for Detecting Depression using Random Forest," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 5, 2020.