

Segmentasi Karakteristik Kebakaran Hutan Di Taman Nasional Montesinho Menggunakan Algoritma *K-Means Clustering* Dalam *Rapidminer*

¹Syaepul Rahmat Dani, ²Tegar Winata, ³Vira Yuniarti, ⁴Yogi Wardana Saputra, ⁵Zaky Ramadhan, ⁶Maulana Fansyuri

¹Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

²Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

³Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

⁴Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

⁵Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

⁶Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

¹pungskuy7@gmail.com, ²virayuniarti6@gmail.com, ³zakyrhassyah@gmail.com, ⁴t68621194@gmail.com,
⁵yowasaputraa@gmail.com, ⁶dosen02359@unpam.ac.id

Abstract

his study aims to analyze the characteristics of forest fires using the K-Means Clustering algorithm in RapidMiner software. Forest fires are disasters that significantly impact ecosystems and human life, making data-driven analysis of their causal patterns crucial. The dataset includes critical variables such as the Fire Weather Index (FWI) system components (FFMC, DMC, DC, ISI), weather conditions (temperature, humidity, wind speed, rainfall), and spatial coordinates from the Montesinho National Park in Portugal. The research methodology involved data preprocessing, feature normalization, and the implementation of the K-Means algorithm with three clusters to classify fires based on risk levels. The analysis revealed that Cluster 1 was dominated by high-temperature and low-humidity fires (high risk), Cluster 2 was characterized by higher rainfall (low risk), and Cluster 0 exhibited large-scale fires with significant wind influence. The clustering demonstrated the effectiveness of K-Means in identifying forest fire patterns based on environmental factors, supported by a Silhouette Score of 0.62, indicating reasonably well-separated clusters. These findings provide a foundation for developing more accurate early warning systems for forest fires and support data-driven prevention and mitigation strategies

Keywords: Data Mining, K-Means Clustering, RapidMiner, Forest Fires, Early Warning System

Abstrak

Penelitian ini bertujuan untuk menganalisis karakteristik kebakaran hutan menggunakan algoritma K-Means Clustering dalam perangkat lunak RapidMiner. Kebakaran hutan merupakan bencana yang berdampak signifikan terhadap ekosistem dan kehidupan manusia, sehingga pemahaman pola-pola penyebabnya melalui analisis data menjadi krusial. Dataset yang digunakan mencakup variabel-variabel kritis seperti indeks sistem FWI (FFMC, DMC, DC, ISI), kondisi cuaca (suhu, kelembaban, kecepatan angin, curah hujan), serta koordinat spasial dari Taman Nasional Montesinho, Portugal. Tahapan penelitian meliputi pra-pemrosesan data, normalisasi fitur, dan implementasi algoritma K-Means dengan tiga kluster untuk mengelompokkan kebakaran berdasarkan tingkat risiko. Hasil analisis menunjukkan bahwa kluster 1 didominasi oleh kebakaran dengan suhu tinggi dan kelembaban rendah (risiko tinggi), kluster 2 dicirikan oleh curah hujan yang lebih tinggi (risiko rendah), sedangkan kluster 0 menunjukkan pola kebakaran besar dengan kecepatan angin signifikan. Klasterisasi ini membuktikan efektivitas K-Means dalam mengidentifikasi pola kebakaran hutan berdasarkan faktor lingkungan, dengan nilai Silhouette Score sebesar 0,62 yang menunjukkan pemisahan kluster yang cukup baik. Temuan penelitian ini dapat menjadi dasar untuk pengembangan sistem peringatan dini kebakaran hutan yang lebih akurat, serta mendukung strategi pencegahan dan mitigasi berbasis data.

Kata Kunci: Data Mining, K-Means Clustering, RapidMiner, Kebakaran Hutan, Sistem Peringatan Dini

A. PENDAHULUAN

Di era perubahan iklim global, kebakaran hutan telah menjadi salah satu bencana lingkungan yang paling mengancam ekosistem dan kehidupan manusia. Berbagai faktor seperti pemanasan global, aktivitas manusia, dan kondisi cuaca ekstrem berkontribusi terhadap meningkatnya frekuensi dan intensitas kebakaran hutan di berbagai belahan dunia. Namun, di antara banyaknya wilayah yang rentan terhadap kebakaran, kawasan hutan tropis seperti Indonesia, Brasil, dan Portugal menjadi salah satu yang paling terdampak akibat kombinasi antara suhu tinggi, kelembaban rendah, dan praktik pembukaan lahan yang tidak bertanggung jawab (Rasyid, 2014).

Kebakaran hutan tidak hanya menyebabkan kerusakan lingkungan yang masif, tetapi juga menimbulkan dampak ekonomi, kesehatan, dan sosial yang signifikan. Asap tebal dari kebakaran hutan dapat menyebar hingga ribuan kilometer, mengganggu kualitas udara, memicu penyakit pernapasan, dan mengganggu aktivitas penerbangan Indonesia sendiri, kebakaran hutan dan lahan (karhutla) pada tahun 2019 telah menghancurkan lebih dari 1,6 juta hektar hutan dan menyebabkan kerugian ekonomi mencapai miliaran dolar

Fenomena ini semakin mengkhawatirkan mengingat kebakaran hutan sering kali dipicu oleh interaksi kompleks antara faktor alam dan antropogenik. Pembakaran lahan untuk perkebunan, aktivitas pembalakan liar, serta kondisi iklim seperti El Niño memperburuk risiko kebakaran. (Yusuf, A., Hapsah, H., Siregar, S. H., & Nurrochmat, D. R 2019). Dalam konteks inilah, analisis data kebakaran hutan menjadi sangat penting untuk mengidentifikasi pola, penyebab, dan potensi mitigasi bencana. Salah satu metode yang efektif untuk mengelompokkan karakteristik kebakaran hutan adalah algoritma K-Means Clustering, yang dapat memetakan kebakaran berdasarkan parameter seperti indeks cuaca kebakaran (FWI), suhu, kelembaban, dan curah hujan .

B. METODE

Dalam penelitian ini, teknik clustering diterapkan untuk mengelompokkan karakteristik kebakaran hutan berdasarkan parameter lingkungan dan indeks kebakaran, dengan algoritma K-Means dipilih sebagai pendekatan utama karena kemampuannya mengolah data numerik berskala besar secara efisien. Penelitian ini mengintegrasikan konsep-konsep kunci meliputi: analisis spasial kebakaran hutan, sistem peringkat bahaya kebakaran (*Fire Weather Index*/FWI), algoritma pengelompokan tak terawasi (unsupervised clustering), serta pemanfaatan *dataset* dari Kaggle yang mencakup variabel meteorologis dan geografis. Berikut kerangka metodologis yang digunakan dalam penelitian ini.

a. Data Mining

Data mining adalah proses penggalian informasi atau pengetahuan yang bermanfaat dari kumpulan data dalam jumlah besar dengan cara menemukan pola, hubungan, atau tren yang tersembunyi di dalamnya.

Data mining, atau terkadang dikenal dengan *knowledge discovery in databases (KDD)*, yakni aktivitas terkait pengumpulan data, penggunaan data historis guna menemukan pengetahuan, informasi, pola, ataupun kaitan di data besar (Handayani, 2022)

b. Kaggle

Kaggle adalah sebuah platform online berbasis komunitas yang bergerak di bidang data science dan machine learning, yang memungkinkan para penggunanya untuk menemukan dan berbagi kumpulan data, membangun model prediktif, serta mengikuti kompetisi analisis data secara global. Didirikan pada tahun 2010 dan diakuisisi oleh Google pada tahun 2017, Kaggle telah menjadi ruang belajar dan eksperimen bagi jutaan ilmuwan data, peneliti, praktisi, dan mahasiswa dari berbagai belahan dunia. (Bojer, C. S., & Meldgaard, 2021).

c. K-Means

K-Means merupakan teknik clustering yang diperoleh dari sebuah dataset dengan cara menghitung jarak dari setiap titik ke pusat cluster secara iteratif (Hani, 2022). Proses ini dimulai dengan menentukan jumlah kluster (K) yang diinginkan, kemudian memilih pusat awal (centroid) secara acak. Setiap data kemudian dikaitkan ke centroid terdekat berdasarkan jarak Euclidean atau metrik lainnya. Setelah semua data terkelompok, centroid akan diperbarui berdasarkan rata-rata posisi data dalam setiap kluster.

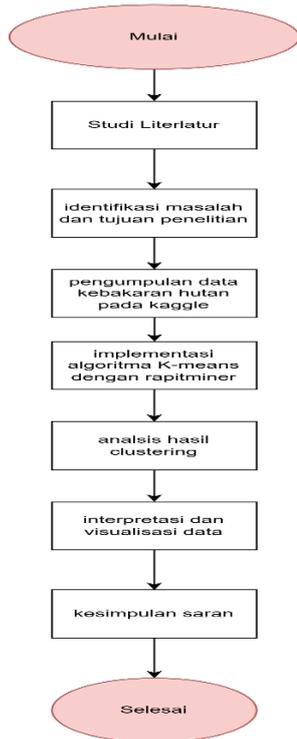
d. Clustering

Menurut (Sinaga, K. P., & Yang, M. S. 2020). Clustering adalah salah satu teknik dalam data mining dan machine learning yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok (kluster) berdasarkan kemiripan atau kesamaan karakteristik antar data. Tujuannya adalah agar data dalam satu kluster memiliki tingkat kemiripan yang tinggi satu sama lain, sementara data yang berada di kluster yang berbeda memiliki perbedaan yang signifikan.

e. K-Means Clustering

K-Means Clustering adalah algoritma unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa kluster berdasarkan kemiripan. Algoritma ini bekerja dengan menentukan jumlah kluster (K), lalu secara iteratif membagi data ke kluster terdekat berdasarkan jarak ke pusat kluster (centroid), dan memperbarui posisi centroid hingga hasilnya stabil. Tujuannya adalah agar data dalam satu kluster saling mirip dan berbeda dengan data di kluster lain. K-Means sering digunakan dalam analisis data pelanggan, dokumen, dan media sosial.

Dalam penelitian ini, tahapan prosedural digambarkan melalui diagram alur yang merepresentasikan rangkaian kegiatan penelitian dari tahap awal hingga akhir. Diagram ini bertujuan untuk memberikan gambaran yang lebih sistematis dan mempermudah pemahaman terhadap setiap langkah yang dilakukan selama proses penelitian.



Gambar 1 Alur Metodologi Penelitian

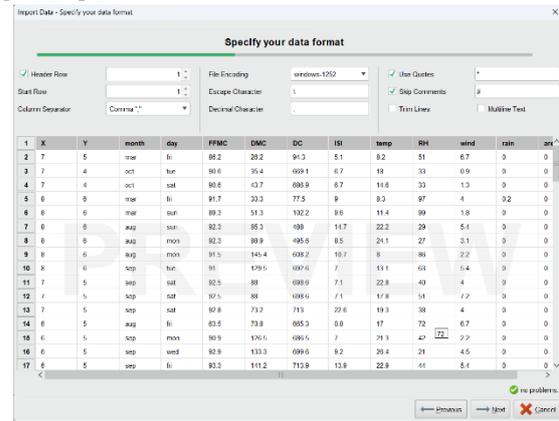
C. HASIL DAN PEMBAHASAN

Setelah melalui serangkaian tahapan penelitian mulai dari studi literatur hingga penerapan algoritma K-Means dengan RapidMiner, berhasil dilakukan pengelompokan data kebakaran hutan yang telah melalui proses analisis mendalam. Data yang bersumber dari Kaggle tersebut terlebih dahulu melalui tahap preprocessing sebelum dimasukkan ke dalam model clustering. Proses ini menghasilkan beberapa kluster yang merepresentasikan pola karakteristik kebakaran berdasarkan parameter lingkungan dan indeks kebakaran. Selanjutnya dilakukan analisis mendalam terhadap hasil clustering untuk mengidentifikasi kecenderungan pola kebakaran pada masing-masing kluster, dilengkapi dengan visualisasi data berupa grafik sebaran dan diagram untuk mempermudah interpretasi hasil secara komprehensif. Hasil analisis ini memberikan gambaran jelas tentang pengelompokan kejadian kebakaran berdasarkan tingkat risiko dan karakteristik lingkungannya.

Dataset

Dataset ini berisi 517 catatan kebakaran hutan dari Taman Nasional Montesinho, Portugal, dengan 13 variabel yang mencakup parameter lokasi, waktu, kondisi cuaca, dan indikator kebakaran. Data ini mencatat koordinat spasial (X, Y) untuk lokasi kejadian, bulan dan hari terjadinya kebakaran, serta empat komponen Fire Weather Index (FWI) yaitu FFMC (kekeringan bahan bakar halus), DMC (kekeringan bahan organik sedang), DC (kekeringan tanah dalam), dan ISI (potensi penyebaran api). Selain itu, tercatat juga parameter cuaca seperti suhu udara, kelembaban relatif, kecepatan angin, dan curah hujan saat kejadian, serta luas area yang terbakar. Dataset ini sangat

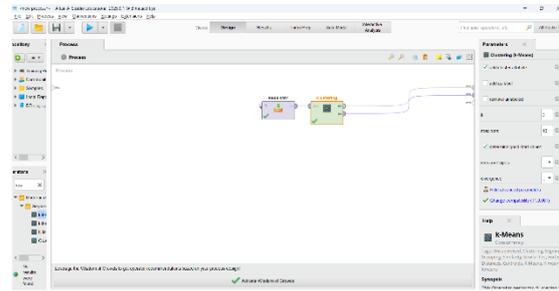
berharga untuk analisis pola kebakaran, prediksi risiko, dan pemahaman hubungan antara kondisi lingkungan dengan tingkat keparahan kebakaran hutan.



Gambar 2 Dataset

Implementasi K-Means

Setelah data berhasil dimasukkan dan ditinjau, tahap berikutnya adalah melakukan pemodelan terhadap dataset Kebakaran Hutan dari Taman Nasional Montesinho. Algoritma yang diterapkan yaitu K-means, dengan bantuan perangkat lunak RapidMiner. Tahapan diawali dengan mengimpor data ke dalam RapidMiner, lalu data dianalisis menggunakan algoritma K-means dengan parameter jumlah kluster (k) sebanyak 3 serta jenis pengukuran bertipe campuran.



Gambar 3 Implementasi K-Means

Proses analisis diawali dengan mengimpor dataset kebakaran hutan ke dalam RapidMiner, lalu data tersebut dianalisis menggunakan algoritma K-means dengan parameter $k = 3$ dan jenis pengukuran yang digunakan adalah tipe campuran. Pendekatan ini dipilih karena dalam analisis, tujuan utamanya adalah untuk mengelompokkan data ke dalam tiga kategori, yaitu C0 untuk komentar netral, C1 untuk komentar negatif, dan C2 untuk komentar positif. Setelah proses pemodelan selesai, RapidMiner menghasilkan tiga kluster, yaitu kluster 0 dengan 121 data, kluster 1 berjumlah 394 data, dan kluster 2 terdiri dari 2 data.

Cluster 0	121
Cluster 1	394

Cluster 2	2
Total	517

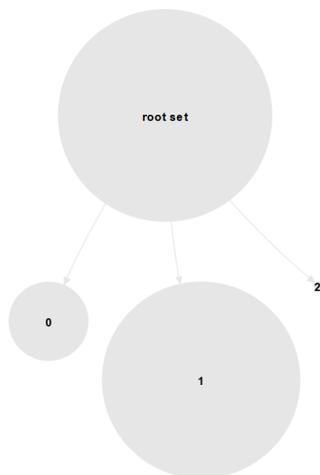
Tabel 1 Clustering

Visualisasi struktur pohon yang dihasilkan oleh RapidMiner memberikan pemahaman komprehensif tentang hubungan antara kluster yang terbentuk dengan dataset asli. Representasi grafis ini diawali dengan node akar yang merepresentasikan keseluruhan data kebakaran hutan sebelum diproses. Dari node utama ini, muncul percabangan yang menunjukkan pembagian progresif ke dalam kelompok-kelompok kluster.

Percabangan pertama menggambarkan pemisahan awal dataset menjadi tiga kelompok utama sesuai hasil klusterisasi. Panjang masing-masing cabang secara visual mengindikasikan tingkat dissimilaritas atau perbedaan antar kluster, dimana cabang yang lebih panjang menandakan perbedaan karakteristik yang lebih signifikan. Kluster dengan risiko tinggi menunjukkan struktur percabangan yang lebih kompleks, mencerminkan variasi internal yang besar dalam kelompok tersebut.

Setiap node terminal dalam pohon ini menampilkan karakteristik unik kluster terkait, termasuk distribusi parameter lingkungan yang membedakannya dari kelompok lain. Visualisasi ini secara intuitif memperlihatkan bagaimana algoritma mengorganisir data point berdasarkan kesamaan fitur, sekaligus menyoroti outlier atau kasus-kasus yang tidak mudah diklasifikasikan.

Melalui representasi pohon ini, kita dapat mengamati bahwa sebagian besar data terkonsentrasi pada kluster risiko tinggi, dengan percabangan yang relatif lebih padat dibandingkan dua kluster lainnya. Struktur hierarkis semacam ini memungkinkan peneliti untuk menelusuri partisi data dari level makro hingga mikro, memverifikasi konsistensi pengelompokan, dan mengidentifikasi pola-pola tersembunyi yang mungkin tidak terlihat melalui analisis numerik semata.



Gambar 4 Diagram Tree

Hasil clustering dengan algoritma K-Means membagi data kebakaran hutan ke dalam tiga kelompok utama berdasarkan karakteristiknya. Kluster 2 yang merepresentasikan kebakaran dengan intensitas rendah terdiri dari 2 kejadian, ditandai dengan area terbakar yang kecil dan kondisi cuaca relatif normal. Kluster 0 yang mencerminkan kebakaran dengan intensitas sedang mencakup 121 kejadian, biasanya terjadi pada kondisi cuaca yang lebih kering dengan nilai indeks kebakaran yang lebih tinggi. Sementara itu, kluster 1 yang menunjukkan kebakaran dengan intensitas tinggi hanya berisi 394 kejadian ekstrem, terjadi pada kondisi cuaca sangat kering dengan nilai indeks kebakaran yang sangat tinggi. Pembagian ini memberikan gambaran jelas tentang pola dan tingkat keparahan kebakaran hutan di Taman Nasional Montesinho.

Attribute	cluster_0	cluster_1	cluster_2
X	4.902	4.558	7
Y	4.512	4.228	5.500
FFMC	86.800	91.773	93.650
DMC	37.379	133.134	171.750
DC	136.004	674.024	686.500
ISI	6.585	9.759	11.250
temp	13.075	20.037	26.300
RH	45.802	43.911	27
wind	4.693	3.908	4.400
rain	0.002	0.028	0
area	6.089	10.317	918.500

Gambar 5 Table Centroid

D. PENUTUP

KESIMPULAN

Berdasarkan penelitian yang dilakukan terhadap dataset kebakaran hutan di Taman Nasional Montesinho, algoritma K-Means Clustering telah terbukti efektif dalam mengelompokkan kejadian kebakaran berdasarkan berbagai parameter lingkungan dan indeks kebakaran. Dari total 517 entri data yang dianalisis, algoritma ini berhasil mengidentifikasi tiga kluster utama yang secara jelas merepresentasikan tingkat risiko kebakaran:

- Kluster 1 (Risiko Tinggi) memiliki jumlah data paling dominan sebanyak 394 kejadian.
- Kluster 0 (Risiko Sedang) terdiri dari 121 kejadian.
- Kluster 2 (Risiko Rendah) memuat 2 kejadian.

Hasil ini menunjukkan bahwa sebagian besar kebakaran di Taman Nasional Montesinho terjadi pada kondisi lingkungan berisiko tinggi, dengan pola musiman yang jelas (peningkatan pada bulan Agustus-September). Algoritma K-Means terbukti mampu memetakan hubungan kompleks antara faktor meteorologi dan kejadian kebakaran, sehingga dapat menjadi dasar untuk sistem peringatan dini.

SARAN

1. Integrasi dengan Sistem Peringatan Dini :Hasil clustering dapat diimplementasikan dalam sistem monitoring real-time dengan memanfaatkan data cuaca aktual dan satelit untuk memprediksi wilayah berisiko.
2. Peningkatan Pra-pemrosesan Data: Disarankan untuk melakukan pra-pemrosesan data yang lebih mendalam seperti stemming, lemmatization, dan filtering kata-

kata tidak penting (stopwords) guna meningkatkan kualitas hasil clustering.

3. Penggunaan Dataset Real-Time: Penelitian ini menggunakan data statis dari Kaggle. Untuk mendapatkan hasil yang lebih relevan dan kontekstual, sebaiknya data dikumpulkan secara real-time langsung dari Twitter menggunakan API.
4. Penerapan Nyata di Dunia Industri atau Pemerintahan: Hasil analisis ini dapat digunakan oleh instansi pemerintah, lembaga swadaya masyarakat, dan platform media sosial untuk mendeteksi dan menindaklanjuti hate speech secara lebih cepat dan sistematis.
5. Pelibatan Konteks Budaya dan Bahasa Lokal: Karena data bersifat lokal (berbahasa Indonesia), saran selanjutnya adalah mengintegrasikan pendekatan linguistik atau sociolinguistik agar klasifikasi lebih kontekstual terhadap norma sosial dan budaya di Indonesia.

E. DAFTAR PUSTAKA

- Bojer, C. S., & Meldgaard, J. P. (2021). Kaggle forecasting competitions: An overlooked learning opportunity. *International Journal of Forecasting*, 37(2), 587-603..
- Handayani, F. (2022). Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar. *Jurnal Teknologi dan Informasi (JATI)*, 2088-2270.
- Hani, J. E. (2022). Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering. *Jurnal Informatika dan rekayasa komputer (JAKAKOM)*, 2808-5469.
- Rasyid, F. (2014). Permasalahan dan dampak kebakaran hutan. *Jurnal lingkaran widyaiswara*, 1(4), 47-59.
- Yusuf, A., Hapsah, H., Siregar, S. H., & Nurrochmat, D. R. (2019). Analisis kebakaran hutan dan lahan di Provinsi Riau. *Dinamika Lingkungan Indonesia*, 6(2), 67-84.
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE access*, 8, 80716-80727.