

Prediksi Aktifitas Gunung Berapi Karangetang di Sitaro Menggunakan Metode K-Nearest Neighbors (KNN)

¹Alfiansyah Hasibuan, ²Djami Olli, ³Bryan Chelvan Walujan, ⁴Boby Putra Situmorang, ⁵Sesilia Derek
^{1,2,3,4,5}Universitas Negeri Manado
¹alfiansyahhasibuan@unima.ac.id

Submit : 08 Sept 25 | Diterima : 16 Sept 2025 | Terbit : 20 Sept 2025

ABSTRAK

Gunung Karangetang di Kabupaten Kepulauan Sitaro, Sulawesi Utara, merupakan salah satu gunung berapi paling aktif di Indonesia yang aktivitasnya berpotensi menimbulkan ancaman bagi masyarakat dan lingkungan. Pemantauan aktivitas vulkanik secara manual sering kali terhambat oleh kondisi cuaca, sehingga diperlukan sistem otomatis yang dapat memberikan informasi real-time dan akurat. Penelitian ini bertujuan untuk melakukan prediksi terhadap aktifitas gunung karangetang dengan tahap melakukan pengklasifikasian dari aktifitas gunung karangetang. Data yang digunakan terdiri dari parameter seismik dan komposisi gas (VA, VB, Tremor, SO₂, dan CO₂/SO₂) yang telah melalui proses normalisasi. Metode penelitian meliputi tahapan pengumpulan data, preprocessing data, pelatihan model KNN, serta evaluasi performa model melalui confusion matrix dan classification report. Hasil pengujian menunjukkan bahwa nilai $k = 1$ memberikan performa terbaik dengan akurasi 85% untuk melakukan prediksi dari aktifitas gunung. Kelas 3.0 memiliki precision, recall, dan f1-score tertinggi (0,95), sedangkan kelas minoritas (0.0 dan 1.0) masih menunjukkan performa rendah dengan f1-score masing-masing 0,29 dan 0,53. Perbedaan antara nilai macro average (precision 0,66; recall 0,64; f1-score 0,63) dengan weighted average (0,85) mengindikasikan ketidakseimbangan data antar kelas. Penelitian ini menyimpulkan bahwa algoritma KNN efektif digunakan untuk klasifikasi dan akurasi prediksi aktivitas vulkanik pada gunung karangetang, terutama pada kelas dominan, namun masih memerlukan perbaikan dalam pengenalan kelas minoritas. Untuk penelitian selanjutnya, disarankan penerapan resampling data, cross-validation, serta eksplorasi algoritma lain seperti Neural Network atau Support Vector Machine guna meningkatkan akurasi dan pemerataan performa antar kelas.

Kata Kunci: Gunung Karangetang, Machine Learning, KNN, Prediksi Aktivitas Vulkanik

PENDAHULUAN

Gunung berapi merupakan salah satu fenomena alam yang paling menakutkan dan memiliki dampak yang signifikan serta dapat mengancam keselamatan hidup manusia dan lingkungan sekitar. Aktivitas vulkanik dapat menyebabkan kerusakan yang signifikan, mulai dari letusan yang memuntahkan lava hingga hujan abu vulkanik yang dapat mempengaruhi kesehatan dan aktivitas manusia. (Mandita & Lintang Ragadanu Arizona, 2023)

Gunung Karangentang adalah salah satu gunung berapi yang paling aktif di Indonesia, yang terletak di Kabupaten Kepulauan Siau Tagulandang Biaro (SITARO), Sulawesi Utara. Gunung ini sering disebut “Api Siau” karena berada di Pulau Siau, yang menjadi pusat Kabupaten Sitaro. Karangentang memiliki ketinggian sekitar 1784 meter di atas permukaan laut dan dikenal memiliki sejarah letusan yang panjang karena sering mengalami erupsi, yang tercatat sejak abad ke-17. Sejarah vulkanik Karangentang tercatat meletus secara signifikan sejak tahun 1675, dengan beberapa letusan besar yang menyebabkan kerusakan dan ancaman bagi masyarakat disekitar wilayah gunung berapi. Dimana letusan-letusan tersebut masing-masing disertai dengan aliran lava, lontaran material vulkanik, dan lahar panas. Aktivitas vukanik gunung ini juga seringkali memaksa untuk dilakukannya evakuasi masyarakat di daerah sekitar gunung dan mengakibatkan kerusakan lahan pertanian, hutan dan infrastruktural local (Adi, 2023).

Pusat Vulkanologi dan Mitigasi Bencana Geologi (PVMBG) melaporkan bahwa menurut data pemantauan visual yang dilakukan dari bulan Desember 2024 sampai Januari 2025 menunjukkan adanya penurunan aktivitas pada Gunung Karangetang. Seperti tidak adanya longoran yang berasal dari kawah utama maupun dari kawah bagian utara serta tidak ada lagi kawah yang menghasilkan pijaran lava yang terlihat di malam hari. Meskipun begitu kawah utara beberapa kali mengeluarkan asap putih yang menjulang setinggi 400 m dan menyebar ke berbagai arah. Aktivitas gempa juga menurun yang ditandai dengan lima gempa vulkanik per hari serta lima gempa yang menunjukkan emisi gas per hari oleh karena itu ada 11 Januari tingkat kewaspadaan diturunkan menjadi level 2 dengan skala 1-4 dan masyarakat juga disarankan untuk menjauh 1,5 km dari kawah dua (kawah utara) dan kawah utama (kawah selatan) dengan jarak yang diperpanjang hingga 2,5 km disepanjang sisi barat, barat daya, selatan dan tenggara. (Gentur Dwi Teguh Santoso, 2025)

Risiko bencana alam dari aktivitas vulkanik tetap menjadi ancaman, sehingga pemerintah setempat bekerja sama dengan Badan Penanggulangan Bencana Daerah (BPBD) dan Pusat Vulkanologi dan Mitigasi Bencana Geologi (PVMBG) untuk memantau aktivitas gunung dan memberikan peringatan dini. Masalah yang dihadapi BPBD yaitu Kondisi cuaca juga terkadang menghalangi proses pengamatan secara visual. Sehingga BPBD tidak bisa memprediksikan kondisi/aktifitas yang terjadi pada gunung Karangentang. Terkait permasalahan yang telah diuraikan, penelitian ini akan mengembangkan system yang akan digunakan untuk memantau aktifitas gunung pada kondisi cuaca tertentu. Metode KNN yang dikembangkan untuk melakukan prediksi dari aktifitas gunung karangetan menggunakan data latih yang bersumber dari Alat IOT dan data yang telah dimiliki oleh BPBD dengan variable data seismik, dan komposisi gas.

Data ini kemudian dapat dianalisis menggunakan teknik machine learning untuk mengidentifikasi pola-pola yang mungkin mengindikasikan terjadinya erupsi. (Samsugi et al., 2021). Selanjutnya pada Penerapan machine learning yang akan digunakan untuk klasifikasi dan prediksi erupsi gunung berapi untuk meningkatkan akurasi prediksi dengan memanfaatkan algoritma K-NN. Pada Penelitian ini akan menggunakan algoritma *K-Neural Network*, yang dikenal dengan kemampuannya dalam mengidentifikasi pola yang kompleks dan non-linear pada data (Pangestu & Rosyani, 2023). Algoritma ini dapat memproses data dalam jumlah besar dan mengidentifikasi hubungan yang kompleks antara berbagai variabel yang berkontribusi terhadap aktivitas vulkanik

TINJAUAN PUSTAKA

Algoritma K Nearest Neighbors (KNN)

Penelitian yang dilakukan oleh F.Mandita, Lintang Ragadanu Arizona, 2023, dengan judul Klasifikasi Status Erupsi Gunung Berapi dengan menggunakan Analisa Template Matching dan Algoritma K Nearest Neighbors. Pada penelitian tersebut hasil yang didapatkan yaitu template matching dan model KNN terbukti efektif untuk klasifikasi sinyal seismik dengan akurasi tinggi menggunakan dataset Mseed berjumlah 2023. Percobaan dengan variasi K pada model KNN menunjukkan akurasi terbaik pada K=1, yaitu 0.91 untuk data mayoritas dan 0.87 untuk keseluruhan data, meskipun terdapat penurunan akurasi seiring meningkatnya nilai K. Penelitian juga menemukan bahwa model KNN lebih unggul dalam klasifikasi sinyal pada dataset yang seimbang. Selain itu, pengujian rasio data 90:10 menunjukkan bahwa K=1 memberikan akurasi tertinggi karena data training yang lebih besar memungkinkan model mengenali pola dengan lebih baik. Namun, penggunaan cross-validation dengan 5 fold menghasilkan akurasi yang lebih rendah dibandingkan pengujian tanpa cross-validation dalam konteks klasifikasi sinyal seismik.(Mandita & Lintang Ragadanu Arizona, 2023)

Penelitian yang dilakukan oleh Anisa, Fitria Wijaya, M Abyan Zhafran Erpindo, Raymond Dini, Peringatan Bencana, Mitigasi 2025, yang berjudul Penerapan Teknologi Internet Of Things Dalam Sistem Peringatan Dini Tsunami Untuk Penguatan Strategi Mitigasi Bencana, hasil yang didapatkan pada penelitian ini yaitu Penggunaan teknologi Internet of Things (IoT) dalam sistem peringatan dini tsunami meningkatkan akurasi prediksi, kecepatan pengambilan keputusan, dan kesiapsiagaan masyarakat melalui pengumpulan data real-time yang dianalisis menggunakan AI dan big data. Inovasi seperti integrasi satelit dan jaringan sensor nirkabel bawah air (UWSN)

memperluas cakupan pemantauan, mengurangi dampak sosial-ekonomi bencana, serta mempercepat respons mitigasi. IoT diprediksi menjadi elemen penting dalam strategi mitigasi bencana di masa depan seiring perkembangan teknologi.(Anisa et al., 2025)

Prediksi

Prediksi pada dasarnya merupakan suatu proses memperkirakan nilai atau kejadian di masa mendatang dengan memanfaatkan informasi yang tersedia pada masa lalu maupun saat ini. Kegiatan ini tidak hanya terbatas pada ranah ilmiah, tetapi juga menjadi bagian penting dalam pengambilan keputusan sehari-hari, baik pada bidang ekonomi, kesehatan, hingga pertanian. Menurut Han, Kamber, dan Pei (2012), prediksi dalam ilmu data dapat dilakukan melalui pemodelan statistik maupun algoritma *machine learning*, yang bertujuan untuk meningkatkan akurasi perkiraan berdasarkan pola data historis.

Klasifikasi

Klasifikasi merupakan salah satu metode utama dalam analisis data yang bertujuan untuk mengelompokkan objek atau data ke dalam kategori tertentu berdasarkan karakteristik atau fitur yang dimilikinya. Secara umum, klasifikasi termasuk dalam ranah *supervised learning*, yaitu pendekatan pembelajaran mesin di mana model dibangun dengan menggunakan data latih yang sudah dilabeli. Menurut Han, Kamber, dan Pei (2012), klasifikasi merupakan proses menemukan model atau fungsi yang menjelaskan dan membedakan kelas data, sehingga model tersebut dapat digunakan untuk memprediksi kelas dari data baru yang tidak diketahui labelnya.

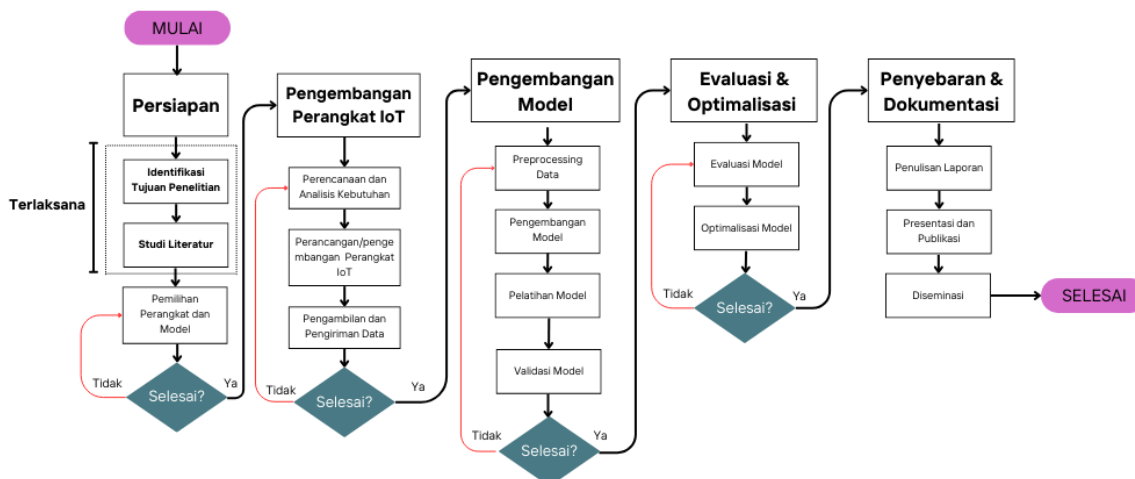
Penerapan dari klasifikasi sering kali digunakan pada berbagai bidang, mulai dari pengenalan pola, diagnosis medis, analisis sentimen, hingga deteksi penipuan. Kecanggihan metode klasifikasi terletak pada kemampuannya untuk mengenali pola kompleks dalam data. Menurut James, Witten, Hastie, dan Tibshirani (2013), beberapa algoritma populer dalam klasifikasi meliputi *k-nearest neighbors (KNN)*, *decision tree*, *naïve Bayes*, *support vector machine (SVM)*, hingga jaringan saraf tiruan (*artificial neural networks*). Masing-masing metode memiliki kelebihan dan keterbatasan, tergantung pada karakteristik data yang digunakan.

METODE PENELITIAN

Metode Penelitian

Penelitian yang berjudul Prediksi Aktifitas Gunung Berapi Karangetang Di Sitaro Menggunakan Metode K-Nearest Neighbors (KNN), penelitian ini untuk melakukan prediksi terhadap aktifitas Gunung Berapi Karangetang Di Sitaro menggunakan metode algoritma K Nearest Neighbors (KNN).

Tahapan Penelitian



Gambar 2. Tahapan penelitian

Teknik Pengumpulan dan Analisis Data

1. Teknik Pengumpulan Data

Data sekunder diperoleh dari laporan-laporan penelitian terdahulu atau laporan dari post pemantauan gunung berapi dan Data primer data yang diambil secara langsung seperti, data dari sensor-sensor yang dipasang seperti, sensor gas untuk menganalisis intensitas gas berbahaya yang muncul saat terjadi aktivitas vulkanik, dan seismic/getaran yang berfungsi untuk mendeteksi pergerakan tanah akibat aktivitas vulkanik yang dimana ini akan menjadi indikator awal terjadinya erupsi data ini juga akan membantu dalam memahami pola gempa yang berkaitan dengan aktivitas vulkanik disekitar gunung berapi

2. Analisis Data

Pada penelitian ini, Teknik yang digunakan untuk menganalisis data adalah Metode Machine learning dengan menggunakan algoritma K-Neural Network (KNN). *K-Neural Network* adalah salah satu algoritma dalam machine learning yang terinspirasi dari cara kerja otak manusia. Algoritma ini terdiri dari lapisan-lapisan yang saling terhubung dan mampu memproses data dalam jumlah besar serta mengenali pola yang kompleks.

Tahapan algoritma K-Nearest Neighbors (KNN)

- a) Melakukan normalisasi data menggunakan metode *Min-Max*

Normalisasi *Min-Max* digunakan untuk mengubah nilai data menjadi skala baru (dengan nilai antara 0 sampai 1).

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

- b) Menentukan parameter *K* terbaik

Pada aktifitas ini, Pilih jumlah tetangga terdekat yang akan digunakan, Pemilihan parameter *k* sangat penting karena terlalu kecil bisa rawan *noise*, terlalu besar bisa mengaburkan pola. Untuk menentukan nilai *k* terbaik dapat menggunakan pendekatan dengan menggunakan rumus heuristic.

$$K = \sqrt{n} \quad (2)$$

Setelah mendapatkan nilai *K* dari rumus tersebut, selanjutnya lakukanlah penyesuaian dengan angka ganjil. Selanjutnya untuk menentukan akurasi dari nilai *K* terbaik, dapat menggunakan metode *Cross Validation*, yaitu untuk mengukur performa model dengan membagi dataset menjadi beberapa bagian.

Rumus Error / Akurasi pada Fold ke-*i*

Akurasi pada fold ke - *i*

$$Acc_i = \frac{1}{n_i} \sum_{j=i}^{n_i} I(y_i = \hat{y}_j) \quad (3)$$

Dengan fungsi indicator

$$I(\text{condition}) = \begin{cases} 1 & \text{Jika kondisi benar} \\ 0 & \text{Jika salah} \end{cases} \quad (4)$$

Mencari Error pada fold ke- i

$$Err_i = 1 - Acc_i \quad (5)$$

Selanjutnya, Rumus Rata-rata Cross Validation

Setelah semua fold telah dihitung, kemudian untuk menentukan akurasi rata-rata menggunakan formula berikut

$$Acc_{cv} = \frac{1}{k} \sum_{i=1}^k Acc_i \quad (6)$$

c) Hitung jarak antara data uji dengan semua data latih

Menggunakan rumus jarak, rumus jarak yang digunakan pada penelitian ini Adalah rumus Euclidean Distance.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2 + (x_2 - y_2)^2} \quad (7)$$

d) Urutkan data latih dimulai dari jarak paling kecil

Data latih dengan jarak paling kecil ke data uji dianggap paling mirip

e) Tentukan nilai k tetangga terdekat (data latih dengan jarak terkecil)

f) Terapkan aturan Keputusan

Untuk menentukan kelas yang paling sering muncul dari kkk tetangga (*majority voting*), menggunakan rumus klasifikasi

$$\hat{y} = \arg \max_{c \in C} \sum_{i=1}^k 1(y_i = c) \quad (8)$$

g) Melihat hasil prediksi

HASIL DAN PEMBAHASAN

Hasil Penelitian

Normalisasi

Dalam konteks dataset gas vulkanik dan aktivitas seismik ini, normalisasi menjamin bahwa parameter seperti VA, VB, Tremor, SO₂, dan CO₂/SO₂ dapat dibandingkan secara adil dan konsisten, sekaligus memberikan dasar yang lebih solid untuk analisis lanjutan.

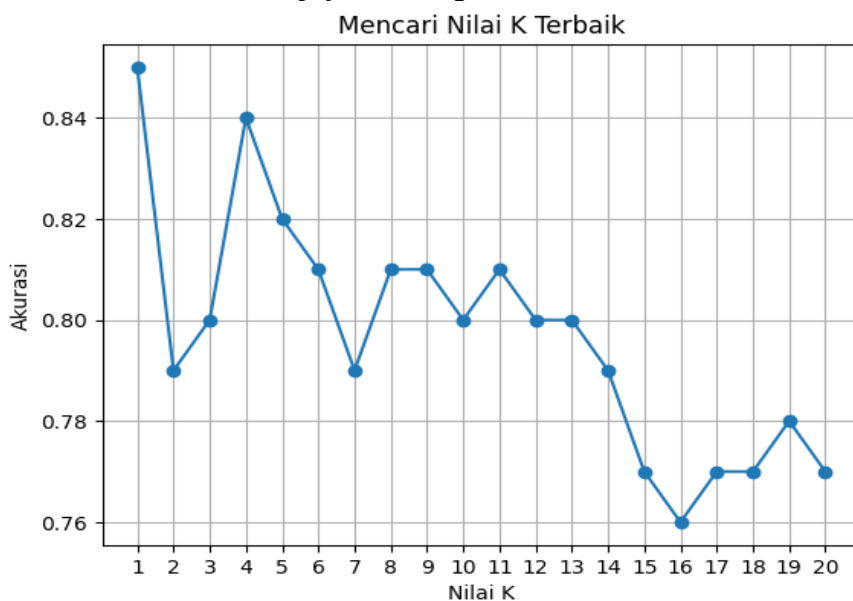
Table 1 Hasil Normalisasi dari dataset Seismik dan gas

| No | VA | VB | Tremor | SO ₂ | CO ₂ /SO ₂ |
|-----|-------|-------|--------|-----------------|----------------------------------|
| 1 | 0.958 | 0.088 | 0.556 | 0.103 | 0.324 |
| 2 | 0.510 | 0.000 | 1.000 | 0.388 | 0.167 |
| 3 | 0.781 | 0.141 | 0.944 | 0.088 | 0.000 |
| 4 | 0.844 | 0.053 | 1.000 | 0.814 | 0.117 |
| 5 | 0.885 | 0.767 | 0.000 | 0.851 | 1.000 |
| 6 | 0.000 | 0.383 | 0.833 | 0.143 | 0.700 |
| 7 | 0.688 | 0.947 | 0.778 | 0.360 | 0.756 |
| 8 | 0.927 | 0.449 | 0.222 | 0.364 | 0.220 |
| 9 | 0.417 | 1.000 | 0.444 | 0.285 | 0.595 |
| 10 | 0.719 | 0.136 | 0.000 | 0.226 | 0.586 |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| 490 | 0.265 | 0.840 | 0.722 | 0.813 | 0.271 |
| 491 | 0.735 | 0.996 | 0.556 | 0.330 | 0.286 |
| 492 | 0.610 | 0.906 | 0.111 | 0.562 | 0.467 |
| 493 | 1.000 | 0.755 | 0.556 | 0.999 | 0.000 |
| 494 | 0.124 | 0.336 | 0.778 | 0.867 | 0.084 |
| 495 | 0.496 | 0.834 | 0.722 | 0.421 | 0.724 |
| 496 | 0.000 | 0.647 | 0.944 | 0.138 | 0.651 |
| 497 | 0.789 | 0.136 | 0.278 | 0.000 | 0.064 |
| 498 | 0.248 | 0.000 | 0.000 | 0.773 | 0.480 |
| 499 | 0.575 | 0.332 | 0.889 | 0.030 | 0.472 |

Berdasarkan hasil normalisasi, terlihat bahwa setiap parameter pada dataset, yaitu VA, VB, Tremor, SO₂, dan CO₂/SO₂, telah berada dalam rentang nilai 0–1 sehingga memudahkan analisis komparatif. Secara umum, variabel VA dan VB menunjukkan distribusi nilai yang cukup bervariasi, di mana sebagian besar data berada pada kisaran menengah hingga tinggi, namun tetap terdapat beberapa titik dengan nilai minimum (0), yang menandakan adanya kondisi aktivitas gas yang sangat rendah. Variabel Tremor cenderung memiliki sebaran yang lebih merata, dengan nilai 0 hingga 1 yang menunjukkan adanya variasi intensitas getaran dari kondisi tenang hingga sangat kuat. Untuk parameter SO₂, nilai relatif tinggi pada beberapa data (mendekati 1) dapat menggambarkan adanya periode peningkatan emisi sulfur dioksida yang signifikan. Sementara itu, rasio CO₂/SO₂ menunjukkan fluktuasi yang cukup tajam antar data, di mana terdapat kombinasi nilai rendah hingga sangat tinggi, yang bisa merefleksikan perubahan komposisi gas dalam sistem vulkanik. Dengan distribusi yang kini seragam antar variabel, pola hubungan antar parameter dapat dianalisis lebih jelas, misalnya kecenderungan CO₂/SO₂ tinggi yang sering beriringan dengan nilai SO₂ rendah, atau korelasi antara VA, VB, dan Tremor dalam memetakan potensi peningkatan aktivitas.

Menentukan parameter *K* terbaik

Untuk menentukan nilai *k* terbaik, biasanya dilakukan pendekatan eksperimental dengan menggunakan metode evaluasi seperti cross-validation. Dalam proses ini, data dibagi ke dalam beberapa subset, kemudian model diuji pada berbagai nilai *k*.



Gambar 2. Grafik AKurasi untuk Mencari Nilai K Terbaik

Hasil evaluasi kemudian dibandingkan untuk melihat nilai *k* yang menghasilkan tingkat akurasi atau performa tertinggi. Dengan cara ini, diperoleh nilai *k* yang optimal sesuai dengan karakteristik data yang digunakan. Berdasarkan hasil pengujian algoritma K-Nearest Neighbors (KNN) dengan variasi nilai K antara 1 sampai 20, diperoleh bahwa nilai K = 1 memberikan akurasi tertinggi dibandingkan dengan nilai K lainnya. Hal ini menunjukkan bahwa pada dataset yang digunakan, pengklasifikasian paling optimal terjadi ketika setiap data uji hanya dibandingkan dengan tetangga terdekat tunggal dalam data latih.

Hasil Perhitungan menggunakan K-Nearest Neighbors (KNN)

Tabel 2. Confusion Matrix:

| | | | |
|-----|---|----|------|
| [[1 | 2 | 0 | 0] |
| [3 | 5 | 4 | 0] |
| [0 | 0 | 17 | 3] |
| [0 | 0 | 3 | 62]] |

Berdasarkan confusion matrix, terlihat bahwa kelas 3.0 memiliki performa terbaik dengan 62 prediksi benar dari 65 data uji (recall 95%). Hal ini juga tercermin pada nilai precision, recall, dan f1-score yang mencapai 0.95, menunjukkan konsistensi dan keandalan model dalam mengenali kelas ini. Kinerja yang sangat baik pada kelas 3.0 diduga karena jumlah datanya yang dominan, sehingga pola distribusinya dapat lebih mudah dipelajari oleh KNN.

Tabel 3. Classification Report:

| | precision | recall | f1-score | support |
|-----|-----------|--------|----------|---------|
| 0,0 | 0,25 | 0,33 | 0,29 | 3 |
| 1,0 | 0,71 | 0,42 | 0,53 | 12 |
| 2,0 | 0,71 | 0,85 | 0,77 | 20 |
| 3,0 | 0,95 | 0,95 | 0,95 | 65 |

Berdasarkan hasil evaluasi yang di tampilkan di tabel 3 menggunakan algoritma K-Nearest Neighbors (KNN) dengan nilai $k = 1$, diperoleh hasil classification report sebagai berikut:

1. Kelas 0.0
Nilai precision sebesar 0,25, recall 0,33, dan f1-score 0,29 dengan jumlah data uji sebanyak 3. Hal ini menunjukkan bahwa model masih kurang mampu mengenali kelas 0.0 secara optimal. Rendahnya nilai precision dan recall menandakan bahwa sebagian besar data kelas ini salah terklasifikasi ke kelas lain. Kondisi ini dapat disebabkan oleh jumlah data yang sedikit sehingga representasi pola kurang kuat.
2. Kelas 1.0
Pada kelas ini, nilai precision cukup tinggi yaitu 0,71, tetapi recall hanya 0,42 sehingga f1-score menjadi 0,53. Dari total 12 data uji, model hanya berhasil mengklasifikasikan sebagian data dengan benar. Hal ini menandakan bahwa meskipun prediksi yang masuk ke kelas 1.0 cukup akurat, banyak data kelas 1.0 justru diprediksi sebagai kelas lain.
3. Kelas 2.0
Hasil evaluasi menunjukkan nilai precision 0,71, recall 0,85, dan f1-score 0,77 dengan total 20 data uji. Nilai recall yang tinggi menunjukkan bahwa sebagian besar data kelas 2.0 berhasil dikenali dengan baik oleh model, meskipun masih terdapat beberapa kesalahan prediksi ke kelas lain.
4. Kelas 3.0
Kinerja terbaik ditunjukkan oleh kelas 3.0, dengan nilai precision 0,95, recall 0,95, dan f1-score 0,95 pada total 65 data uji. Hal ini berarti mayoritas data pada kelas ini dapat diklasifikasikan dengan benar oleh model. Tingginya performa ini kemungkinan karena jumlah data kelas 3.0 jauh lebih banyak dibanding kelas lain, sehingga pola distribusi datanya lebih mudah dipelajari oleh algoritma KNN.

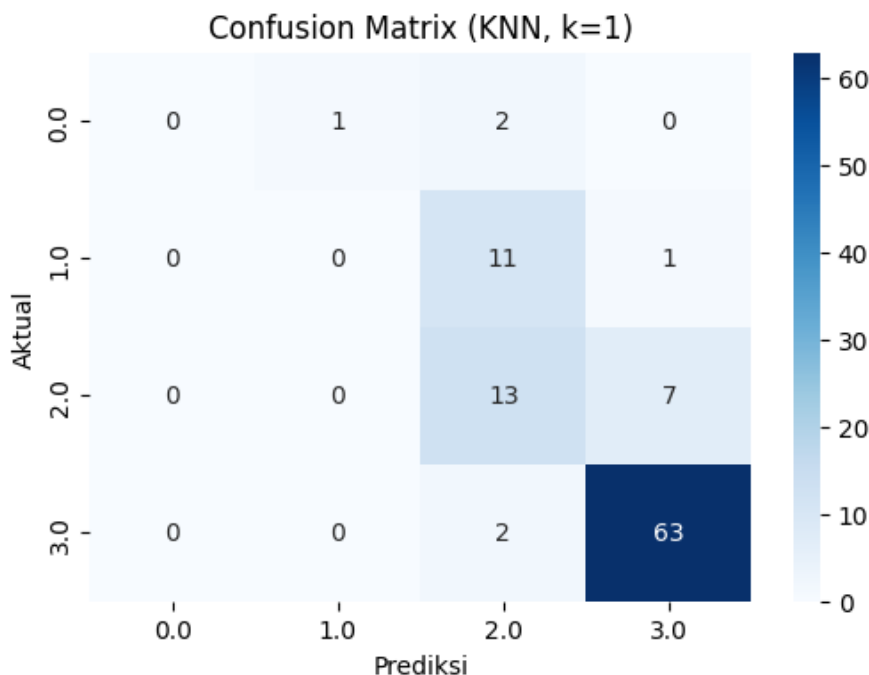
Tabel 4. Hasil pengujian model K-Nearest Neighbors (KNN) $K=1$

| | | | | |
|---------------------|------|------|------|-----|
| accuracy | | | 0,85 | 100 |
| macro avg | 0,66 | 0,64 | 0,63 | 100 |
| weighted avg | 0,85 | 0,85 | 0,85 | 100 |

Hasil pengujian model K-Nearest Neighbors (KNN) dengan $k = 1$ menunjukkan bahwa nilai akurasi keseluruhan mencapai 0,85 atau 85% dari total 100 data uji. Artinya, sebagian besar data uji dapat diprediksi dengan benar oleh model.

Dari sisi rata-rata per kelas, nilai macro average untuk precision, recall, dan f1-score masing-masing adalah 0,66; 0,64; dan 0,63. Nilai ini relatif lebih rendah dibandingkan akurasi keseluruhan, yang mengindikasikan bahwa performa model tidak merata antar kelas. Beberapa kelas, khususnya kelas minoritas (0.0 dan 1.0), memiliki nilai evaluasi rendah sehingga menurunkan skor rata-rata makro.

Sebaliknya, pada weighted average, precision, recall, dan f1-score semuanya bernilai 0,85, sama dengan akurasi keseluruhan. Hal ini terjadi karena kelas dominan (kelas 3.0) memiliki jumlah data jauh lebih banyak, sehingga berkontribusi besar terhadap rata-rata tertimbang. Dengan demikian, model terlihat sangat baik pada kelas mayoritas, namun masih kurang optimal pada kelas dengan jumlah data yang sedikit.



Gambar 3. Confusion matrix hasil klasifikasi

Gambar di atas menunjukkan confusion matrix hasil klasifikasi menggunakan algoritma K-Nearest Neighbors (KNN) dengan nilai $k = 1$. Matriks ini menggambarkan perbandingan antara label aktual (sumbu vertikal) dengan label prediksi (sumbu horizontal).

Penjelasan tentang Kelas 0.0, Dari total data uji kelas ini, tidak ada yang terprediksi benar. Sebanyak 1 data salah diprediksi sebagai kelas 1.0, dan 2 data lainnya salah diprediksi sebagai kelas 2.0. Hal ini menunjukkan bahwa model kesulitan mengenali kelas 0.0. Kemudian pada Kelas 1.0, Dari 12 data uji, sebanyak 11 diprediksi sebagai kelas 2.0 dan hanya 1 data yang diprediksi sebagai kelas 3.0. Tidak ada data yang terklasifikasi dengan benar sebagai kelas 1.0. Artinya, model gagal membedakan pola kelas 1.0 dari kelas lainnya.

Selanjutnya Kelas 2.0: Dari 20 data uji, sebanyak 13 terprediksi sebagai kelas 2.0 (benar), sedangkan 7 data salah diprediksi sebagai kelas 3.0. Meskipun masih ada kesalahan, mayoritas data pada kelas ini berhasil dikenali dengan baik. Pada Kelas 3.0: Kinerja terbaik terlihat pada kelas 3.0. Dari 65 data uji, sebanyak 63 berhasil diprediksi dengan benar, dan hanya 2 data yang salah diklasifikasikan sebagai kelas 2.0. Hal ini menunjukkan bahwa model sangat baik dalam mengenali kelas ini.

Pembahasan

Hasil pengujian dengan algoritma K-Nearest Neighbors (KNN) menunjukkan bahwa pemilihan nilai $k = 1$ menghasilkan performa terbaik dibanding nilai k lainnya. Hal ini ditunjukkan dengan tingkat akurasi keseluruhan sebesar 85%, yang berarti sebagian besar data uji dapat diprediksi dengan benar oleh model.

Jika ditinjau dari confusion matrix dan classification report, terlihat bahwa kinerja model berbeda-beda pada tiap kelas. Kelas 3.0 memiliki performa paling unggul dengan precision, recall, dan f1-score mencapai 0,95. Mayoritas data kelas ini berhasil terklasifikasi dengan benar, hanya sebagian kecil yang salah diprediksi ke kelas 2.0. Tingginya performa pada kelas 3.0 kemungkinan disebabkan oleh jumlah data yang jauh lebih banyak dibandingkan kelas lainnya, sehingga distribusi polanya lebih mudah dipelajari oleh algoritma.

Pada kelas 2.0, hasil evaluasi cukup baik dengan recall 0,85 dan f1-score 0,77. Sebagian besar data kelas ini berhasil diprediksi dengan benar, meskipun masih ada kesalahan klasifikasi ke kelas 3.0. Hal ini menunjukkan adanya kedekatan karakteristik antara kelas 2.0 dan kelas 3.0, sehingga KNN terkadang sulit membedakan keduanya.

Sebaliknya, performa model pada kelas 0.0 dan 1.0 masih rendah. Kelas 0.0 hanya memiliki f1-score 0,29, sementara kelas 1.0 memiliki f1-score 0,53. Rendahnya performa ini diduga kuat karena jumlah data pada kedua kelas tersebut relatif sedikit, sehingga representasi pola menjadi kurang kuat. Akibatnya, model sering salah memetakan data ke kelas lain yang memiliki distribusi lebih dominan.

Dari sisi rata-rata, nilai macro average precision, recall, dan f1-score masing-masing sebesar 0,66; 0,64; dan 0,63. Nilai ini lebih rendah dari akurasi keseluruhan, menandakan adanya ketidakseimbangan performa antar kelas. Namun, pada weighted average, nilai precision, recall, dan f1-score sama dengan akurasi keseluruhan yaitu 0,85. Hal ini menunjukkan bahwa kelas mayoritas (kelas 3.0) sangat memengaruhi performa model secara keseluruhan.

KESIMPULAN

Algoritma K-Nearest Neighbors (KNN) dengan nilai $k = 1$ menghasilkan performa terbaik pada dataset ini dengan tingkat akurasi Prediksi aktifitas gunung karangetan keseluruhan sebesar 85%, Kelas 3.0 memiliki kinerja paling tinggi dengan precision, recall, dan f1-score masing-masing 0,95, menunjukkan model mampu mengenali kelas mayoritas dengan sangat baik, Kelas 2.0 menunjukkan performa cukup baik (recall 0,85, f1-score 0,77), meskipun masih terdapat kesalahan prediksi ke kelas 3.0, Performa model masih rendah pada kelas 0.0 dan 1.0 (f1-score 0,29 dan 0,53). Hal ini disebabkan oleh jumlah data yang sedikit, sehingga pola distribusi tidak dapat dipelajari secara optimal, Perbedaan antara nilai macro average (precision 0,66; recall 0,64; f1-score 0,63) dengan weighted average (precision, recall, dan f1-score 0,85) menunjukkan adanya ketidakseimbangan data, di mana model lebih baik mengenali kelas dominan dibanding kelas minoritas.

REFERENSI

- Adi, A. C. (2023). *Status Vulkanik G. Karangetang Turun ke Level Waspada*. Kepala Biro Komunikasi, Layanan Informasi Publik, Dan Kerja Sama. <https://www.esdm.go.id/id/media-center/arsip-berita/status-vulkanik-g-karangetang-turun-ke-level-waspada>
- Anisa, F., Wijaya, M. A. Z., Erpindo, R., Dini, P., & Bencana, M. (2025). *PENERAPAN TEKNOLOGI INTERNET OF THINGS DALAM SISTEM PERINGATAN DINI TSUNAMI UNTUK PENGUATAN STRATEGI MITIGASI BENCANA*. 19, 1196–1202.
- Budiyanto, M. N., Santosa, P. I., & Sumaryono, S. (2012). Purwarupa Sistem Peringatan Dini Awan Panas Gunungapi Berbasis Sistem Informasi Geografis (Kasus Gunung Merapi di Perbatasan Jawa Tengah dan Daerah Istimewa Yogyakarta). *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 1(1), 24–30.
- Diandika, D. (2018). *Rancangan Interface Sistem Mitigasi Gunung Merapi Menggunakan Metode XP Untuk Penyampaian Informasi Dari BPPTKG Kepada Masyarakat*. <https://repository.its.ac.id/58344/>
- Gentur Dwi Teguh Santoso, S. K. (2025). *Penurunan Tingkat Aktivitas G. Karangetang, Sulawesi Utara, dari Level III (Siaga) menjadi Level II (Waspada)*. PVBMG.
- Hermawan, H. (2019). Sistem Pemantauan Gunung Berapi Berbasis Iot Menggunakan Nodemcu Dan Lora. *Jurnal Teknik Elektro*.
- Mandita, F., & Lintang Ragadanu Arizona. (2023). ``Klasifikasi Status Erupsi Gunung Berapi dengan menggunakan Analisa Template Matching dan Algoritma K Nearest Neighbors. *Journal of Computer Science and Visual Communication Design*, 8(2), 275–289. <https://doi.org/10.55732/jikdiskomvis.v8i2.1061>
- Pangestu, I. A., & Rosyani, P. (2023). Perancangan Aplikasi Mitigasi Bencana Alam Berbasis Android Menggunakan Metode Extreme Programming (XP). *LOGIC: Jurnal Ilmu Komputer Dan Pendidikan*, 1(6), 1458–1464.
- Samsugi, S., Rusliyanti, & Jupriadi. (2021). Rancang Bangun Sistem Monitoring Aktivitas Gunung Anak Krakatau Berbasis IoT. *Jurnal Teknologi Elektro Dan Kejuruan*, 31(1), 14–22.
- Siswanto, S., Ngatono, & Febri Saputra, S. (2022). Prototype Sistem Peringatan Dini Bencana Gempa Bumi Dan Tsunami Berbasis Internet of Things. *PROSISKO: Jurnal Pengembangan Riset Dan Observasi Sistem Komputer*, 9(1), 60–66.

<https://doi.org/10.30656/prosisko.v9i1.4743>

- Vincensio Simon, M. A., Setiawan, W., & Sastra, N. P. (2020). Rancang Bangun Sistem Peringatan Dini Bahaya Aktivitas Gunung Berapi Berbasis Mikrokontroler Arduino. *Jurnal SPEKTRUM*, 7(3), 42. <https://doi.org/10.24843/spektrum.2020.v07.i03.p6>
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). Wiley.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.