



ISSN : 2339 - 1871

JURNAL ILMIAH BETRIK

Besemah Teknologi Informasi dan Komputer

Editor Office : LPPM Sekolah Tinggi Teknologi Pagar Alam, Jln. Masik Siagim No. 75
Simpang Mbacang, Pagar Alam, SUM-SEL, Indonesia
Phone : +62 852-7901-1390.
Email : betrik@sttpagaralam.ac.id | admin.jurnal@sttpagaralam.ac.id
Website : <https://ejournal.sttpagaralam.ac.id/index.php/betrik/index>

IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR (KNN) UNTUK PREDIKSI BENCANA GUNUNG BERAPI

Iftar Ramadhan¹, Syarifuddin Elmi², Rahmaddeni³, Lusiana Efrizoni⁴

Program Studi Teknik Informatika, STMIK AMIK RIAU

Jl. Pendidikan No.17, Pekanbaru, Indonesia

Sur-el : ¹iftarramadhan09@gmail.com, ²syarifuddin09@gmail.com, ³rahmaddeni@sar.ac.id, ⁴lusiana@stmik-amik-riau.ac.id

Abstrak: Potensi bencana alam, terutama erupsi gunung berapi, meningkat, yang membutuhkan inovasi teknologi untuk meningkatkan sistem deteksi dan respons. Studi ini berfokus pada penggunaan algoritma K-Nearest Neighbor (KNN) untuk mengidentifikasi bencana yang terjadi di gunung Merapi. Data yang digunakan mencakup berbagai parameter geofisika dan meteorologi yang terkait dengan aktivitas vulkanik. Pengujian dilakukan dengan nilai k-terdekat ($k=3$) dan menunjukkan akurasi 100%. Namun, ketika k kembali ke setting umum (k tidak spesifik), akurasi sebesar empat puluh empat persen dicapai. Hasil penelitian ini menunjukkan bahwa, dengan mempertimbangkan elemen lingkungan dan geologis, penggunaan KNN dapat meningkatkan deteksi dini bencana gunung berapi. Penelitian ini memberikan kontribusi penting untuk pengembangan sistem deteksi bencana yang lebih canggih, meskipun perlu upaya tambahan untuk meningkatkan akurasi secara keseluruhan.

Kunci Utama: *machine learning, prediksi, k-nearest neighbor, bencana gunung berapi*

Abstract: *The potential for natural disasters, particularly volcanic eruptions, is on the rise, necessitating technological innovations to enhance detection and response systems. This study focuses on the application of the K-Nearest Neighbor (KNN) algorithm to identify disasters occurring at Mount Merapi. The data utilized encompass various geophysical and meteorological parameters relevant to volcanic activity. Testing was conducted with the k-nearest value set at $k=3$, yielding an accuracy of 100%. However, when returning to the general setting (non-specific k), an accuracy of forty-four percent was achieved. The results of this research demonstrate that, by considering environmental and geological elements, the use of KNN can enhance early detection of volcanic disasters. This study makes a significant contribution to the development of more sophisticated disaster detection systems, although additional efforts are needed to improve overall accuracy.*

Keywords : *machine learning, prediction, k-nearest neighbor, volcano disaster*

1. PENDAHULUAN

Dalam struktur tektonik global, gugusan pulau Indonesia adalah wilayah di mana tiga lempeng besar berinteraksi: Lempeng Samudera Hindia-Australia (bagian selatan), Lempeng Samudera

Pasifik (bagian timur laut), dan Lempeng Eurasia (bagian barat laut). Hal ini berdampak pada morfologi, geologi, dan topografi Indonesia [1]. Gunung Merapi adalah gunung berapi strato (stratovolcano) dengan ketinggian 2980

meter dari permukaan laut. Menurut Pangestu (2010), gunung berapi strato (stratovolcano) adalah gunung berapi (gunung berapi) yang tinggi dan mengerucut yang terdiri dari lava dan abu vulkanik yang membeku [2].

Fakta menunjukkan bahwa gunung berapi di Indonesia memiliki potensi yang sangat besar untuk meletus. Ini dapat dilihat dari lokasi geografisnya dan bagaimana gunung berapi berada di jalur pertemuan lempeng tektonik dengan barisan gunung berapi aktif, juga dikenal sebagai cincin api.

Indonesia memiliki 13% dari seluruh gunung api di dunia, dengan 500 gunung api yang tidak aktif dan 129 yang berstatus aktif. Selain itu, 60% dari seluruh gunung api di seluruh dunia memiliki potensi letusan yang signifikan [3].

Sebaliknya, letusan gunung api dapat menyebabkan banyak kerusakan. Wesnawa dan Christiawan menjelaskan bahwa bencana letusan gunung api menyebabkan banyak kerusakan dan kerugian, termasuk kehancuran tata ruang wilayah, penurunan kualitas lingkungan, kerusakan sarana lalu lintas, kerusakan bangunan yang berfungsi sebagai pusat aktivitas masyarakat, dan lainnya.

Kerugian yang ditimbulkan juga dapat mengganggu aktivitas kehidupan masyarakat, terhentinya sektor perekonomian, dan bahkan dapat terhentinya proses pembangunan nasional [4]. Penelitian ini bertujuan untuk mengklarifikasi deteksi bencana gunung Merapi dengan menggunakan metode KNN. Karena mudah dan sederhana, algoritma K-nearest neighbour (k-nn) adalah salah satu algoritma mesin pembelajaran yang paling populer. Selain itu, k-nn adalah salah satu algoritma pembelajaran yang memiliki proses belajar berdasarkan nilai variabel target yang terasosiasi dengan nilai variabel prediktor [5]. Hasil classification matrix dari model prediksi bencana gunung merapi di sajikan pada table 1.

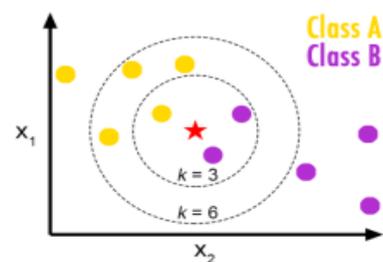
Tabel 1. Hasil Prediksi

	prec	rec	f1-	sup
anthropology	0.00	0.00	0.00	1
Ar/ar	0.00	0.00	0.00	1
Denbrochrology	0.00	0.00	0.00	0
Fumarolic	0.00	0.00	0.00	7
Historical	0.62	0.10	0.18	49
Holocene	0.43	0.96	0.60	48
Ice core	0.00	0.00	0.00	1
Magnetism	0.00	0.00	0.00	2
Radiocarbon	0.00	0.00	0.00	8
Tephrochronology	0.00	0.00	0.00	3
Uncertain	0.44	1.00	0.62	4
Accuracy			0.44	124
Macrp avg	0,14	0.19	0.13	124
Wighted avg	0.43	0.44	0.32	124

Penelitian ini diharapkan dapat membantu mengurangi bencana dan melindungi masyarakat melalui pemahaman mendalam tentang karakteristik Gunung Merapi dan teknik kecerdasan buatan. Oleh karena itu, penggunaan algoritma K-Nearest Neighbor ini merupakan kemajuan dalam informasi teknologi serta solusi yang dapat meningkatkan ketangguhan komunitas terhadap ancaman bencana alam yang kompleks.

1.1. K-Nearest Neighbor (KNN)

KNN adalah metode klasifikasi yang menentukan kategori berdasarkan mayoritas kategori pada kNearest Neighbor. KNN dilakukan dengan mencari kelompok k objek dalam data pelatihan yang paling dekat (atau sebanding) dengan objek dalam data baru atau data pengujian [6].



Gambar 1. Ilustrasi Penggunaan Nilai K pada metode KNN

Nilai k terbaik untuk KNN bergantung pada data; Umumnya, nilai k yang tinggi akan mengurangi efek noise pada klasifikasi tetapi membuat perbedaan antara klasifikasi menjadi lebih jelas. Ilustrasi penggunaan nilai k pada metode disajikan pada Gambar 1, Nilai k yang ideal diperoleh dengan melakukan optimasi parameter, seperti menggunakan cross-validation. Algoritma paling dekat tetangga digunakan dalam situasi tertentu di mana prediksi klasifikasi didasarkan pada data pembelajaran yang paling dekat (dengan kata lain, k = 1) [7].

Langkah – Langkah Klarifikasi data menggunakan metode K- Neighbor [8]:

- a. Tentukan nilai K.
- b. Hitung setiap jarak dari data baru ke setiap label data.
- c. Tentukan k labeled data yang jaraknya paling minimal.
- d. Klarifikasi data baru ke dalam label data yang minoritas K-NN berdasarkan metrik jarak

Untuk tujuan penelitian ini, teks data yang dikumpulkan akan diubah menjadi vektor. Cosine kesamaan adalah metode yang paling umum digunakan untuk menghitung jarak antar vektor. Metode ini mudah dijelaskan pada data yang jarang, seperti data teks. Persamaan cosine, yang sebanding dengannya adalah:

$$S_{\text{Cosine}}(x, y) = \frac{x^T y}{\|x\| \|y\|} \quad (1)$$

Di mana $\|x\| = \sqrt{\sum_{i=1}^l x_i^2}$

Dan $\|y\| = \sqrt{\sum_{i=1}^l y_i^2}$ Adalah

Panjang dari Vector [9].

1.2. Evaluasi

Untuk menentukan model yang paling cocok untuk hasil klasifikasi, evaluasi klasifikasi melakukan pengujian pada objek yang benar dan yang salah [10].

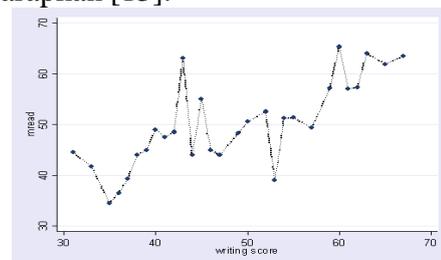
Tujuan dari pengujian adalah untuk menunjukkan hasil dari pengujian data untuk mengukur akurasi. Pada tahap ini, model evaluasi dilakukan untuk mengukur tingkat akurasi dari setiap teknik yang

mengacu pada dataset yang digunakan [11].

Dalam klasifikasi model evaluasi, urutan tes digunakan untuk menentukan apakah objek benar atau salah. Urutan tes ditabulasikan dalam Confusion Matrix, dengan kelas yang diharapkan ditampilkan di atas matriks dan kelas yang sebenarnya ditampilkan di sisi kiri. Setiap sel berisi angka yang menunjukkan jumlah kasus sebenarnya dari kelas yang diukur untuk diprediksi [12].

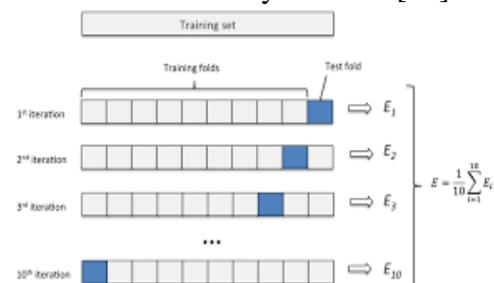
1.3. Validasi

Salah satu elemen penting dalam masalah klasifikasi dan regresi adalah model fit, yang menunjukkan kemampuan model untuk memprediksi data baru yang belum dibor.. Overfitting terjadi ketika data prediksi sangat berbeda dari yang diharapkan [13].



Gambar 2. Ilustrasi Overfitting

K-Fold Cross Validation membagi data ke dalam k bagian dan kemudian melakukan proses klasifikasi pada masing-masing bagian. Dengan metode ini, percobaan akan dilakukan sebanyak k kali [14].



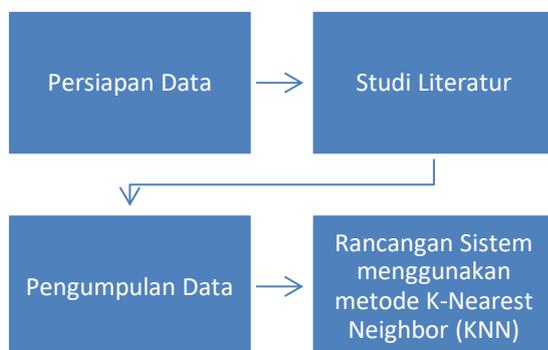
Gambar 3. Ilustrasi K-Fold Cross Validation

Untuk validasi cross-fold k, model yang telah dibuat dibagikan menjadi k bagian yang sama atau hampir sama ukurannya. Uji data akan digunakan untuk menguji akurasi model pada setiap lipatan dan terus sampai

selesai. Akurasi akan ditotalkan dan dibagi dengan banyaknya k [15].

2. METODE PENELITIAN

Proses awal termasuk memilih topik penelitian, menemukan pertanyaan penelitian, dan menetapkan hipotesis atau tujuan penelitian. Setelah menentukan tujuan penelitian, langkah selanjutnya adalah mempersiapkan data yang dibutuhkan. Ini melibatkan memilih sumber data, mengumpulkan data awal, dan membersihkan data untuk memastikan bahwa data tersebut aman dan berkualitas tinggi. Penelitian melibatkan membaca literatur sebelum memulai eksperimen atau analisis untuk menemukan penelitian sebelumnya yang relevan. Ini membantu menemukan perbedaan pengetahuan, membangun landasan teori, dan memahami teknik yang digunakan dalam penelitian serupa. Pengumpulan data ini bersifat analisis skunder artinya dari data yang sudah ada. Studi pengembangan sistem atau teknologi melibatkan perancangan sistem secara rinci. Ini mencakup pemilihan teknologi, arsitektur sistem, desain antarmuka pengguna, dan spesifikasi teknis lainnya. Studi pengembangan sistem atau teknologi melibatkan perancangan sistem secara rinci. menggunakan teknik analisis knn atau metode lainnya yang sesuai dengan pertanyaan penelitian. Hasil analisis digunakan untuk membuat kesimpulan dan menarik hasil penelitian. Tahapan Algoritma KNN disajikan pada Gambar 4.



Gambar 4. Tahapan Algoritma K-Nearest Neighbor (KNN)

2.1 Persiapan Data

Tahap awal yang sangat penting dalam implementasi algoritma K-Nearest Neighbor (KNN) untuk mendeteksi bencana Gunung Merapi adalah persiapan data. Data persiapan proses termasuk :

a. Pemahaman Terhadap Data

Memahami struktur dan karakteristik data yang akan digunakan, termasuk atribut yang relevan seperti data aktivitas gunung, kondisi cuaca, dan variabel lain yang dapat mempengaruhi deteksi bencana

b. Pembersih Data

Menemukan dan menangani nilai-nilai yang hilang atau outlier. Pembersihan data membuat data lebih konsisten dan dapat diandalkan dengan menghapus atau mengganti nilai-nilai yang tidak valid.

c. Normalisasi Data

Jika perlu, normalisasikan data untuk memastikan bahwa semua atribut memiliki skala yang sama

d. Pembagian Data

Dataset terdiri dari dua bagian: data pelatihan (training set) dan data pengujian (test set). Pembagian ini penting untuk melatih model dan menguji keakuratannya, dan ini dapat meningkatkan kinerja algoritma KNN.

2.2 Studi Literatur

Studi literatur adalah langkah penting menuju pemahaman lebih baik tentang kerangka kerja penelitian sebelumnya dan konsep-konsep yang terkait dengan penerapan algoritma KNN untuk deteksi bencana Gunung Merapi. Studi sastra akan mencakup:

a. Review jurnal dan artikel terkait

Membaca jurnal dan artikel sebelumnya yang membahas tentang deteksi bencana alam menggunakan teknik KNN atau kecerdasan buatan lainnya.

b. Pemahaman Algoritma KNN

Pelajari konsep dan prinsip dasar dari algoritma K-Nearest Neighbor secara menyeluruh. Pahami cara KNN bekerja dan bagaimana itu dapat digunakan untuk mendeteksi bencana.

2.3 Pengumpulan Data

Pengumpulan data meliputi pengumpulan informasi tentang aktivitas Gunung Merapi, cuaca, dan komponen lain yang berhubungan dengan deteksi bencana. Berikut ini adalah prosedur yang digunakan untuk mengumpulkan data:

a. Sumber Data

Menemukan sumber data yang dapat diandalkan, seperti observasi lapangan, stasiun meteorologi, atau data sejarah dari lembaga terkait.

b. Pengumpulan Data

Pastikan data memiliki kualitas yang baik dengan mengumpulkannya dengan hati-hati sesuai dengan fitur yang ditetapkan dalam rencana sistem.

c. Integrasi Data

Menggabungkan data dari berbagai sumber untuk membuat dataset yang lengkap yang memenuhi kebutuhan penelitian.

2.4 Perancangan Sistem Menggunakan Metode KNN

Sistem perencanaan menggunakan metode KNN yang mencakup tahap perencanaan dan persyaratan implementasi algoritma KNN untuk deteksi bencana Gunung Merapi. Berikut ini adalah langkah-langkah perencanaan sistem:

a. Pemilihan Atribut

Pilih atribut yang paling penting untuk digunakan untuk mendeteksi bencana. Ini termasuk data seismik, gas vulkanik, suhu, dan faktor lain yang dapat mempengaruhi aktivitas Gunung Merapi.

b. Penentuan Parameter KNN

Menemukan nilai terbaik untuk parameter K (jumlah tetangga terdekat) dan fungsi jarak yang akan digunakan berdasarkan analisis data dan eksperimen.

c. Membangun Model

Menggunakan data pelatihan yang telah disiapkan sebelumnya, membangun

model KNN. Komputasi jarak, memilih tetangga terdekat, dan membuat model prediktif adalah semua bagian dari proses ini.

d. Membangun Model

Menggunakan pengujian data untuk kinerja model KNN. Ini termasuk akurasi, presisi, recall, dan skor F1 untuk mengukur keberhasilan sistem.

e. Optimasi dan Peningkatan

Jika diperlukan, optimalkan model KNN dengan menguji berbagai nilai parameter untuk meningkatkan kinerja sistem.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Perancangan Sistem

Data yang akan digunakan termasuk atribut yang relevan seperti data aktivitas gunung, kondisi cuaca, dan variable lain yang dapat mempengaruhi deteksi bencana. Data ini di ambil dari situs web kegggle, sebanyak 617 data dan data ini bersifat publik.

3.2 Pembersihan Data

Region	Number	Volcano name	Country	Location	Latitude	Longitude	Elevation (m)	Type	Status	Last known eruption	#
West Indies	1000-21	Pico de Teide	Spain	Tenerife	28.2667	-16.8667	3715	Stratovolcano	Historical	D1	NaN
West Indies	1000-30	Parícutin	Mexico	Michoacán	19.1667	-101.8667	1486	Stratovolcano	Historical	U	NaN
West Indies	1000-33	Parícutin	Mexico	Michoacán	19.1667	-101.8667	1486	Stratovolcano	Historical	D1	NaN
West Indies	1000-18	Merapi	Indonesia	Sumatra	-6.9167	101.0167	2964	Stratovolcano	Historical	D2	NaN
West Indies	1000-19	Merapi	Indonesia	Sumatra	-6.9167	101.0167	2964	Stratovolcano	Historical	D1	NaN

Gambar 1. Data Gunung Berapi

```

[4] df = df.drop('Region', axis=1)
df = df.drop('d', axis=1)
df = df.drop(["Volcano Name"], axis=1)
df = df.drop(["Country"], axis=1)

[5] df = df.drop(["Type"], axis=1)

[6] df = df.drop(["Location"], axis=1)

[7] df = df.drop(["Number"], axis=1)

[8] df = df.drop(["Last Known Eruption"], axis=1)
    
```

Gambar 2. Script Menghapus Kolom

```
[9] print(df)

   Latitude Longitude Elevation (m) Status
0      10.000    -61.00         140 Historical
1      12.150    -61.67          840 Holocene
2     16044.000  -61664.00        1467 Historical
3     14809.000  -61165.00        1394 Historical
4       16.720    -62.18          915 Historical
..      ...      ...      ...      ...
612     0.079     99.98        2919 Holocene
613    -0.979    100681.00        2575 Historical
614    -0.390    100331.00        2854 Historical
615    4769.000    96821.00        2617 Historical
616     2.580     98.83         2157 Holocene

[617 rows x 4 columns]
```

Gambar 3. Menampilkan Kolom Yang Belum Di Hapus

```
[10] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 617 entries, 0 to 616
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  ---          -
0   Latitude         617 non-null   float64
1   Longitude        617 non-null   float64
2   Elevation (m)    617 non-null   int64
3   Status           617 non-null   object
dtypes: float64(2), int64(1), object(1)
memory usage: 19.4+ KB
```

Gambar 4. Melihat Info Dataset

```
[14] from sklearn.model_selection import train_test_split

[15] x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaler.fit(x_train)
x_train = scaler.transform(x_train)
x_test = scaler.transform(x_test)
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=2)
y_train = y_train.astype(str)
label_encoder = LabelEncoder()
y_train = y_train.astype(str)
label_encoder = LabelEncoder()
y_train_encoded = label_encoder.fit_transform(y_train)

knn.fit(x_train, y_train_encoded)
knn.fit(x_train, y_train)

KNeighborsClassifier
KNeighborsClassifier(n_neighbors=2)
```

Gambar 5. Menampilkan KNeighborsClassifier Dengan Model Prediksi 80.20

```
[29] print(classification_report(y_test,y_pred))

              precision    recall  f1-score   support

 Anthropology      0.00      0.00      0.00         2
    Ar/Ar          0.00      0.00      0.00         1
 Dendrochronology  0.00      0.00      0.00         1
    Fumarolic      0.17      0.33      0.22         6
    Historical     0.38      0.69      0.49        32
    Holocene       0.60      0.47      0.53        55
    Holocene?     0.00      0.00      0.00         3
    Magnetism      0.00      0.00      0.00         2
    Radiocarbon    0.40      0.20      0.27        10
 Tephrochronology 0.00      0.00      0.00         6
    Uncertain      0.00      0.00      0.00         6

 accuracy          0.42         124
 macro avg         0.14         124
 weighted avg      0.41         124
```

Gambar 6. Menampilkan Hasil

3.3 Normalisasi Data

Data awal terletak I gambar.7

Region	Number	Volcano Name	Country	Location	Latitude	Longitude	Elevation (m)	Type	Status	Last Known Eruption	d
West Indies	1600-21	Piparo	Tinidad	Tinidad	10.000	-61.00	140	Mud volcano	Historical		D1 R/HR
West Indies	1600-30	poira	Grenada	W Indies	12.150	-61.67	840	Stratovolcano	Holocene		D1 R/HR
West Indies	1600-33	zila	Guadeloupe	W Indies	16044.000	-61664.00	1467	Stratovolcano	Historical		D1 R/HR
West Indies	1600-16	alctra	Marthique	W Indies	14809.000	-61165.00	1394	Stratovolcano	Historical		D2 R/HR
West Indies	1600-19	Piparo	Montserrat	W Indies	16.720	-62.18	915	Stratovolcano	Historical		D1 R/HR
...
Indonesia	0601-11	Sibualbuli	Indonesia	Sumatra	0.079	99.98	2919	Complex volcano	Holocene		U R/HR
Indonesia	0601-08	Sinabung	Indonesia	Sumatra	-0.979	100681.00	2575	Stratovolcano	Historical		D1 R/HR
Indonesia	0601-07	Singkat	Indonesia	Sumatra	-0.390	100331.00	2854	Stratovolcano	Historical		D2 R/HR
Indonesia	0601-12	Sonkarapu	Indonesia	Sumatra	4769.000	96821.00	2617	Stratovolcano	Historical		D2 R/HR
Indonesia	0601-18	Sumbing	Indonesia	Sumatra	2.580	98.83	2157	Caldera	Holocene		U R/HR

Gambar 7. Data Awal

Hasil dari pembersihan data terletak di gambar.8

	Latitude	Longitude	Elevation (m)	Status
0	10.000	-61.00	140	Historical
1	12.150	-61.67	840	Holocene
2	16044.000	-61664.00	1467	Historical
3	14809.000	-61165.00	1394	Historical
4	16.720	-62.18	915	Historical
..
612	0.079	99.98	2919	Holocene
613	-0.979	100681.00	2575	Historical
614	-0.390	100331.00	2854	Historical
615	4769.000	96821.00	2617	Historical
616	2.580	98.83	2157	Holocene

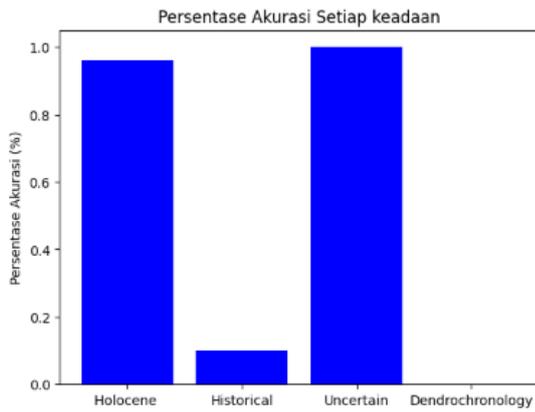
Gambar 8. Pembersihan Data

3.4 Model Prediksi

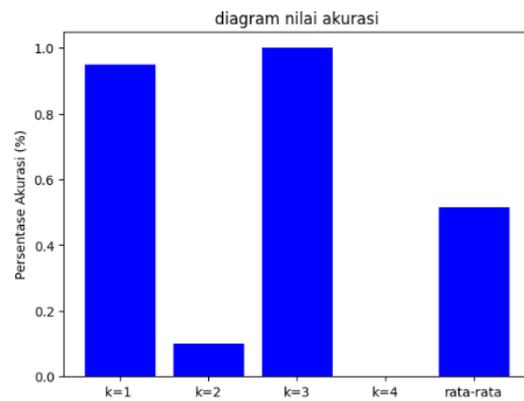
Model prediksi ini menggunakan algoritma K-NN yang mana data akan di bagi menjadi data latih dan di uji dengan rasio 80:20. Algoritma yang telah ditentukan akan digunakan untuk membuat data latih untuk model classifier, sedangkan data uji akan diprediksi menggunakan model yang telah dibangun. Hasil prediksi model akan menentukan akurasi dan waktu proses

3.5 Evaluasi Prediksi

Penentuan K terbaik pada K-NN Untuk menentukan nilai K terbaik pada K-NN, evaluasi dilakukan, dengan nilai K dari K=1 hingga K=4. Nilai K terbaik adalah nilai K dengan akurasi tertinggi, dan hasilnya menunjukkan bahwa nilai K tertinggi ditemukan pada dua nilai K, yaitu K=3 dengan presentase 1,00. Diagram batang perbandingan presentase K dengan K=1 hingga K=4 ditunjukkan di diagram 1.



Gambar 9. Diagram 1



Gambar 10. Diagram 2

Tabel 1. Perhitungan Precision, Recall Dan F1-Score

Model	precision	recall	Fi-score	support
K=1	0,34	0,96	0,60	48
2	0,62	0,10	0,18	49
3	1,00	1,44	0,62	4
4	0,00	0,00	0,00	0

Nilai K pada K-NN yang memiliki akurasi terbaik, yaitu K=1 digunakan untuk validasi pada model classifier K-NN, sedangkan nilai k pada metode validasi cross-fold adalah k=4 Hasil validasi menunjukkan bahwa akurasi optimal K-NN adalah 100% pada iterasi kedua, dengan rata-rata 51,5%. Dengan K=4, akurasi optimal K-NN adalah 95% pada iterasi kedua, dengan rata-rata 51,5% dan akurasi tertinggi 100%. Presentase akurasi K dengan K=1 hingga K=4 ditunjukkan di diagram 2

Tabel 2. Hasil Validasi

literasi	Nilai akurasi
1	0,95
2	0,10
3	1,00
4	0,00
Rata_rata	0,515

4. SIMPULAN

Hasil penelitian menunjukkan bahwa data secara keseluruhan diklasifikasikan dengan baik, dengan rata-rata 0,515 untuk k=3 dan 100% untuk k=4. Penggunaan crossvalidation pada metode knn cukup untuk mengidentifikasi perubahan nilai k pada setiap uji coba.

DAFTAR RUJUKAN

- [1] S. Haeriah, A. Laila Nugraha, And B. Sudarsono, "Analisis Kerentanan Pada Wilayah Permukiman Akibat Bencana Erupsi Gunung Merapi (Studi Kasus : Kabupaten Sleman)," 2018.
- [2] Ariyadi Nugroho Susilo Dan Iwan Rudiarto, "Dampak Erupsi Gunung Merapi," Vol. 3, No. 1, Pp. 34–49, 2014.
- [3] L. Christie Gosal, R. Ch Tarore, And H. H. Karongkong, "Analisis Spasial Tingkat Kerentanan Bencana Gunung Api Lokon Di Kota Tomohon," *Jurnal Spasial*, Vol. 5, No. 2, Pp. 229–237, 2018.
- [4] A. S. Ardi And D. R. S. Sumunar, "Analisis Risiko Bencana Erupsi Gunung Merapi Di Kecamatan Dukun Kabupaten Magelang," *Geomedia: Majalah Ilmiah Dan Informasi Kegeografian*, Vol. 15, No. 1, Pp. 99–110, 2017, Doi: 10.21831/Gm.V15i1.16243.
- [5] F. Tempola, M. Muhammad, And A. Khairan, "Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada

- Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation,” *Jurnal Teknologi Informasi Dan Ilmu Komputer*, Vol. 5, No. 5, Pp. 577–584, 2018, Doi: 10.25126/Jtiik.201855983.
- [6] M. Reza Noviansyah, T. Rismawan, D. Marisa Midyanti, J. Sistem Komputer, And F. H. Mipa Universitas Tanjungpura Jl Hadari Nawawi, “Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya),” *Jurnal Coding, Sistem Komputer Untan*, Vol. 06, No. 2, Pp. 48–56, 2018.
- [7] I. A. Angreni, S. A. Adisasmita, M. I. Ramli, And S. Hamid, “Pengaruh Nilai K Pada Metode K-Nearest Neighbor (Knn) Terhadap Tingkat Akurasi Identifikasi Kerusakan Jalan,” *Rekayasa Sipil*, Vol. 7, No. 2, P. 63, 2019, Doi: 10.22441/Jrs.2018.V07.I2.01.
- [8] N. Ibrahim, T. F. Bacheramsyah, B. Hidayat, And S. Darana, “Pengklasifikasian Grade Telur Ayam Negeri Menggunakan Klasifikasi K-Nearest Neighbor Berbasis Android,” *Elkomika: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, Vol. 6, No. 2, P. 288, 2018, Doi: 10.26760/Elkomika.V6i2.288.
- [9] F. S. Al-Anzi And D. Abuzeina, “Toward An Enhanced Arabic Text Classification Using Cosine Similarity And Latent Semantic Indexing,” *Journal Of King Saud University - Computer And Information Sciences*, Vol. 29, No. 2, Pp. 189–195, 2017, Doi: 10.1016/J.Jksuci.2016.04.001.
- [10] A. N. Kasanah, M. Muladi, And U. Pujiyanto, “Penerapan Teknik Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma Knn,” *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, Vol. 3, No. 2, Pp. 196–201, 2019, Doi: 10.29207/Resti.V3i2.945.
- [11] W. Yunus, “Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Ginjal Kronik,” *Jurnal Teknik Elektro Cosphi*, Vol. 2, No. 2, Pp. 51–55, 2018.
- [12] Ardiyansyah, P. A. Rahayuningsih, And R. Maulana, “Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner,” *Jurnal Khatulistiwa Informatika*, Vol. Vi, No. 1, Pp. 20–28, 2018.
- [13] J. Lever, M. Krzywinski, And N. Altman, “Points Of Significance: Model Selection And Overfitting,” *Nature Methods*, Vol. 13, No. 9, Pp. 703–704, 2016, Doi: 10.1038/Nmeth.3968.
- [14] Z. Sidik, “Klasifikasi Kelancaran Kredit Furniture Menggunakan Algoritma K-Nearest Neighbor Berbasis Forward Selection,” 2019.
- [15] M. R. A. Nasution And M. Hayaty, “Perbandingan Akurasi Dan Waktu Proses Algoritma K-Nn Dan Svm Dalam Analisis Sentimen Twitter,” *Jurnal Informatika*, Vol. 6, No. 2, Pp. 226–235, 2019, Doi: 10.31311/Ji.V6i2.5129.