

TEMPERATURE CLASSIFICATION IN HIGH VOLTAGE ELECTRICAL EQUIPMENT USING K-NEAREST NEIGHBOR

KLASIFIKASI SUHU PADA PERALATAN LISTRIK TEGANGAN TINGGI MENGGUNAKAN K-NEAREST NEIGHBOR

Giovanni Dimas Prenata¹, Aswanda Harja²

¹Program Studi Teknik Elektro, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya, Indonesia
Email: gprenata@untag-sby.ac.id

²Program Studi Teknik Elektro, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya, Indonesia
Email: aswanda.harja23@gmail.com

Received: Revised: Accepted:
DOI: <https://doi.org/10.24912/tesla.v25i2.27065>

Abstract

This research develops AI for classifying safe/hazardous conditions from thermal photos. Thermal photos display the temperature of objects, with white color indicating high temperature (dangerous) and dark color showing low temperature (safe). Feature extraction methods are used to obtain the percentage of white and non-white colors as classification parameters. The research used 10 thermal images as training data and 2 images as test data. The distance between the test and training data is calculated using Euclidean Distance, then sorted based on proximity. Classification is performed with several K values (1-7). The test results show an accuracy of 71.42%. This accuracy is low due to the limited training data for the hazard category and the use of a non-optimal K value. To improve accuracy, the following are needed: proper selection of K, addition of training data, and better data preprocessing techniques. With these improvements, the system will be more reliable in classifying conditions from thermal images.

Keywords: Thermal photos, danger, safety, classification, K-Nearest Neighbor.

Abstrak

Penelitian ini mengembangkan AI untuk klasifikasi kondisi aman/berbahaya dari foto thermal. Foto thermal menampilkan suhu objek, dengan warna putih menandakan suhu tinggi (berbahaya) dan warna gelap menunjukkan suhu rendah (aman). Metode ekstraksi fitur digunakan untuk mendapatkan persentase warna putih dan non-putih sebagai parameter klasifikasi. Penelitian menggunakan 10 gambar thermal sebagai data latih dan 2 gambar sebagai data uji. Jarak antara data uji dan latih dihitung dengan Euclidean Distance, kemudian diurutkan berdasarkan kedekatan. Klasifikasi dilakukan dengan beberapa nilai K (1-7). Hasil pengujian menunjukkan akurasi 71,42%. Akurasi ini rendah karena keterbatasan data latih kategori bahaya dan penggunaan nilai K yang tidak optimal. Untuk meningkatkan akurasi diperlukan: pemilihan K yang tepat, penambahan data latih, dan teknik pra-pemrosesan data yang lebih baik. Dengan perbaikan ini, sistem akan lebih andal dalam klasifikasi kondisi dari gambar thermal.

Kata Kunci: Foto thermal, bahaya, aman, klasifikasi, K-Nearest Neighbor.

PENDAHULUAN

Penerapan kecerdasan buatan dalam sektor ketenagalistrikan merupakan sebuah inovasi yang masih tergolong baru, tetapi memiliki potensi besar dalam meningkatkan efisiensi dan keakuratan berbagai proses dalam sistem tenaga listrik. Kecerdasan buatan memungkinkan berbagai konsep dan metode dalam ketenagalistrikan diterapkan dengan lebih cepat, efisien, dan akurat dibandingkan metode konvensional. Dengan adanya teknologi ini, berbagai tantangan dalam pengelolaan dan distribusi tenaga listrik dapat diatasi dengan pendekatan berbasis data serta algoritma pembelajaran mesin yang canggih.

Salah satu penelitian yang membuktikan manfaat kecerdasan buatan dalam ketenagalistrikan dilakukan oleh Prenata, yang berfokus pada klasifikasi keandalan sistem distribusi tenaga listrik di PT PLN (Persero) UP3 Surabaya Selatan. Dalam penelitian tersebut, Prenata mengembangkan metode berbasis kecerdasan buatan untuk menganalisis dan menentukan tingkat keandalan sistem distribusi tenaga listrik dengan lebih sistematis dan akurat.

Untuk melakukan klasifikasi keandalan sistem, penelitian ini menggunakan tiga metode utama dalam kecerdasan buatan, yaitu Single Perceptron [1], Support Vector Machine (SVM) [2], dan K-Nearest Neighbor (KNN) [3]. Single Perceptron merupakan model jaringan saraf tiruan paling sederhana yang mampu mengklasifikasikan data berdasarkan pemetaan linear. SVM bekerja dengan mencari hyperplane optimal yang memisahkan data dalam ruang berdimensi tinggi, sedangkan KNN mengklasifikasikan data berdasarkan kedekatannya dengan sejumlah titik data latih tertentu.

Dengan menerapkan ketiga metode ini, penelitian bertujuan untuk menemukan pendekatan yang paling efektif dalam mengukur dan mengklasifikasikan keandalan sistem distribusi tenaga listrik. Penggunaan kecerdasan buatan dalam penelitian ini membuktikan bahwa metode berbasis AI dapat meningkatkan efisiensi dan keakuratan dalam menentukan keandalan suatu sistem, dibandingkan dengan metode manual yang membutuhkan lebih banyak waktu dan tenaga dalam proses analisisnya.

Berdasarkan Standar Perusahaan Listrik Negara (SPLN) 59-1985, suatu sistem distribusi tenaga listrik dikatakan memiliki tingkat keandalan yang baik apabila nilai System Average Interruption Duration Index (SAIDI) lebih kecil dari 12,842 dan nilai System Average Interruption Frequency Index (SAIFI) lebih kecil dari 2,145. Kedua parameter ini digunakan sebagai indikator utama dalam menilai keandalan sistem distribusi tenaga listrik, di mana SAIDI mengukur rata-rata durasi gangguan listrik yang dialami pelanggan dalam satu tahun, sedangkan SAIFI menunjukkan rata-rata frekuensi gangguan listrik yang terjadi dalam periode yang sama.

Dengan kemajuan teknologi dan penerapan kecerdasan buatan (Artificial Intelligence/AI), proses klasifikasi tingkat keandalan sistem distribusi listrik dapat dilakukan secara otomatis, cepat, dan akurat. Sistem berbasis AI yang telah dikembangkan mampu mengolah data SAIDI dan SAIFI secara langsung, tanpa perlu perhitungan manual oleh operator. Dengan memanfaatkan algoritma pembelajaran mesin, sistem dapat mengenali pola-pola dalam data historis dan melakukan prediksi atau klasifikasi keandalan sistem dengan tingkat presisi yang tinggi.

Penerapan AI dalam analisis keandalan jaringan listrik tidak hanya menghemat waktu dan tenaga, tetapi juga meningkatkan efisiensi serta mengurangi potensi kesalahan dalam pengambilan keputusan. Dengan demikian, perusahaan penyedia listrik dapat lebih proaktif dalam mengidentifikasi dan mengatasi potensi gangguan, sehingga kualitas layanan kepada pelanggan dapat ditingkatkan secara signifikan.

Hasil penelitian menunjukkan bahwa metode K-Nearest Neighbor (KNN) mampu mencapai tingkat akurasi 100% dalam proses klasifikasi keandalan sistem distribusi tenaga listrik. Keakuratan ini dicapai untuk berbagai nilai K, mulai dari $K = 1$ hingga $K = 7$, yang berarti bahwa metode ini secara konsisten dapat mengidentifikasi tingkat keandalan sistem dengan sangat baik. Keunggulan KNN dalam klasifikasi ini disebabkan oleh kemampuannya dalam mengelompokkan data berdasarkan kedekatan dengan data latih, sehingga memberikan hasil yang presisi dalam menentukan apakah suatu sistem tergolong andal atau tidak.

Selain itu, metode Single Perceptron juga menunjukkan kinerja yang sangat baik dengan mencapai akurasi 100% dalam klasifikasi keandalan. Model ini mampu memberikan hasil yang akurat dalam menentukan tingkat keandalan sistem berdasarkan data yang diberikan. Bahkan, ketika dibandingkan dengan beberapa metode lain seperti Decision Tree, Random Forest, Naïve

Bayes, dan Multilayer Perceptron (MLP), metode Single Perceptron tetap unggul. Meskipun keempat metode tersebut juga memiliki tingkat akurasi yang tinggi, yaitu di atas 85%, Single Perceptron tetap menunjukkan performa yang lebih optimal dalam konteks klasifikasi keandalan sistem distribusi tenaga listrik.

Selain kedua metode di atas, penelitian ini juga menguji kinerja metode Support Vector Machine (SVM) dalam klasifikasi keandalan sistem. Hasil yang diperoleh menunjukkan bahwa SVM juga berhasil mencapai tingkat akurasi 100% setelah ditemukan garis hyperplane yang optimal dengan persamaan $y = -1,82693x + 18,3422$. Hyperplane ini berfungsi sebagai pemisah antara kategori data yang menunjukkan sistem yang andal dan tidak andal. Dengan demikian, metode SVM terbukti mampu mengklasifikasikan keandalan sistem distribusi tenaga listrik dengan sangat akurat.

Hasil keseluruhan penelitian ini mengonfirmasi bahwa penerapan metode kecerdasan buatan dalam klasifikasi keandalan sistem tenaga listrik dapat memberikan hasil yang sangat presisi. Dengan akurasi yang mencapai 100%, baik pada KNN, Single Perceptron, maupun SVM, metode berbasis AI ini dapat diandalkan untuk meningkatkan efisiensi dan keakuratan dalam menganalisis keandalan sistem distribusi tenaga listrik, sehingga membantu dalam pengambilan keputusan yang lebih tepat dalam pengelolaan jaringan listrik.

Penelitian lain yang membahas penerapan kecerdasan buatan (Artificial Intelligence/AI) dalam bidang ketenagalistrikan, yang dilakukan oleh Prenata, berfokus pada upaya validasi hasil pengukuran rele dalam sistem tenaga listrik. Rele merupakan salah satu komponen penting dalam sistem proteksi tenaga listrik yang berfungsi untuk mendeteksi gangguan dan mengisolasi bagian jaringan yang bermasalah guna mencegah dampak yang lebih luas.

Dalam penelitian ini, rele digunakan untuk mengukur nilai impedansi pada jaringan listrik Gas Insulated Switchgear (GIS) 150 kV [4]. Nilai impedansi tersebut sangat penting karena digunakan untuk menentukan lokasi gangguan hubung singkat, yaitu kondisi ketika terjadi hubungan langsung antara dua atau lebih konduktor dalam sistem tenaga listrik, yang dapat menyebabkan arus berlebih dan berpotensi merusak peralatan atau bahkan memicu pemadaman luas.

Validasi langsung hasil pengukuran rele masih menjadi tantangan dalam praktik. Prenata mengatasi kendala ini dengan menerapkan Particle Swarm Optimization (PSO) untuk memperoleh data pembandingan dan mengevaluasi hasil pengukuran impedansi rele. PSO adalah metode optimasi berbasis kecerdasan buatan yang terinspirasi perilaku kawanan. Dalam penelitian ini, PSO mencari nilai impedansi optimal sebagai referensi untuk menilai akurasi pengukuran rele.

Hasil penelitian menunjukkan penerapan PSO dalam validasi pengukuran rele memberikan akurasi lebih dari 98%. Dengan tingkat akurasi ini, PSO dapat diandalkan sebagai referensi utama untuk mengevaluasi hasil pengukuran rele. Penelitian ini berkontribusi signifikan dalam pengembangan metode berbasis kecerdasan buatan untuk meningkatkan keandalan sistem tenaga listrik. Dengan PSO sebagai alat validasi, pengukuran rele dapat dievaluasi lebih objektif dan akurat, membantu perusahaan listrik meningkatkan efisiensi dan mengurangi risiko kesalahan deteksi gangguan.

Penelitian terkait penerapan metode K-Nearest Neighbor (KNN) telah dilakukan dalam berbagai bidang, menunjukkan efektivitasnya dalam proses klasifikasi dan prediksi. Beberapa penelitian yang telah dilakukan di antaranya adalah studi yang dilakukan oleh Khairi, Ghozali, dan Hidayah [5], yang menerapkan KNN untuk klasifikasi masyarakat pra-sejahtera di Desa Sapikerep, Kecamatan Sukapura. Dalam penelitian ini, mereka menggunakan beberapa nilai K untuk melihat tingkat akurasi metode KNN dalam mengklasifikasikan kelompok masyarakat pra-sejahtera berdasarkan kriteria tertentu. Hasil yang diperoleh menunjukkan bahwa untuk $K =$

3, akurasi klasifikasi mencapai 97,36%, sedangkan untuk $K = 5$ dan $K = 7$, tingkat akurasinya meningkat menjadi 98,68%.

Selain itu, penelitian yang dilakukan oleh Nanda, Pudjiantoro, dan Sabrina juga menerapkan metode KNN, namun dalam konteks yang berbeda, yaitu prediksi curah hujan di Kota Bandung [6]. Dengan mempertimbangkan berbagai faktor meteorologi yang berpengaruh terhadap curah hujan, mereka menggunakan KNN dengan $K = 5$ dan memperoleh tingkat akurasi sebesar 86,199%. Meskipun akurasi yang diperoleh tidak setinggi penelitian sebelumnya dalam klasifikasi masyarakat pra-sejahtera, metode KNN tetap menunjukkan keandalannya dalam menangani permasalahan prediksi berbasis data cuaca.

Penelitian lain yang juga menggunakan metode KNN dilakukan oleh Choli, Handayani, dan Ardianita dalam proses klasifikasi seleksi penerima beasiswa [7]. Dalam penelitian ini, mereka menerapkan metode KNN untuk membantu menentukan kandidat yang berhak menerima beasiswa berdasarkan sejumlah kriteria tertentu, seperti prestasi akademik, kondisi ekonomi, dan faktor pendukung lainnya. Hasil yang diperoleh menunjukkan bahwa metode KNN mampu mengklasifikasikan calon penerima beasiswa dengan tingkat akurasi sebesar 90,5%, yang membuktikan bahwa metode ini dapat diandalkan dalam proses seleksi berbasis data.

Dari berbagai penelitian di atas, dapat disimpulkan bahwa metode K-Nearest Neighbor (KNN) memiliki fleksibilitas dan efektivitas yang tinggi dalam berbagai bidang aplikasi, baik dalam klasifikasi sosial, prediksi cuaca, maupun seleksi penerima beasiswa. Dengan tingkat akurasi yang relatif tinggi pada berbagai skenario penggunaan, KNN dapat menjadi salah satu metode pilihan dalam berbagai penerapan kecerdasan buatan dan analisis data untuk menghasilkan keputusan yang lebih tepat dan akurat.

KAJIAN PUSTAKA

Metode K-Nearest Neighbor (KNN) merupakan salah satu teknik klasifikasi yang bekerja dengan membandingkan data uji terhadap sejumlah data latih berdasarkan perhitungan jarak di antara keduanya. Prinsip utama dari metode ini adalah bahwa kondisi data uji sangat dipengaruhi oleh data latih yang paling dekat dengannya, di mana kedekatan ini diukur menggunakan suatu metrik jarak tertentu. Semakin dekat suatu data uji dengan sekelompok data latih yang memiliki label tertentu, semakin besar kemungkinan bahwa data uji tersebut akan diklasifikasikan ke dalam kategori yang sama dengan data latih tersebut.

Dalam metode KNN, jumlah data latih yang dipertimbangkan dalam menentukan klasifikasi data uji disebut sebagai nilai K . Nilai K ini menunjukkan berapa banyak tetangga terdekat yang akan diperhitungkan dalam menentukan kelas atau kategori dari data uji yang sedang dianalisis. Pemilihan nilai K menjadi faktor krusial dalam metode ini, karena jika nilai K terlalu kecil, maka klasifikasi bisa menjadi terlalu sensitif terhadap outlier atau data yang tidak representatif. Sebaliknya, jika K terlalu besar, klasifikasi bisa menjadi terlalu umum dan kehilangan detail spesifik dari pola yang ada.

Nilai K dapat berupa bilangan genap maupun ganjil, namun dalam banyak kasus, penggunaan nilai ganjil lebih disarankan. Hal ini karena nilai K yang ganjil dapat membantu menghindari kemungkinan terjadinya jumlah suara (vote) yang sama dalam proses klasifikasi. Sebagai contoh, jika $K = 3$, maka klasifikasi dilakukan berdasarkan mayoritas dari tiga tetangga terdekat, yang memastikan bahwa selalu ada keputusan akhir yang jelas. Namun, jika $K = 4$, terdapat kemungkinan bahwa dua tetangga memiliki satu label dan dua lainnya memiliki label yang berbeda, sehingga keputusan

klasifikasi menjadi kurang pasti. Oleh karena itu, pemilihan nilai K yang optimal perlu dilakukan dengan mempertimbangkan karakteristik data serta kebutuhan spesifik dari sistem yang sedang dikembangkan.

Untuk menentukan jarak antara data uji dan data latih, penelitian ini menerapkan metode perhitungan menggunakan rumus Euclidean Distance, yang merupakan salah satu metrik jarak yang paling umum digunakan dalam algoritma KNN. Euclidean Distance mengukur jarak lurus antara dua titik dalam ruang dimensi tertentu dan dapat dihitung menggunakan rumus sebagai berikut:

$$d = \sqrt{\sum_{i=1}^k (X_i - Y_i)^2} \quad (1)$$

Dalam rumus di atas, p dan q merepresentasikan dua titik dalam ruang multidimensi, sedangkan koordinat (x, y, z, ...) adalah nilai-nilai fitur dari data yang dibandingkan. Dengan menggunakan metode ini, sistem dapat mengidentifikasi seberapa dekat suatu data uji dengan sejumlah data latih yang sudah diketahui labelnya, sehingga dapat menentukan klasifikasi yang paling sesuai berdasarkan pola yang ada.


Dengan demikian, metode KNN menawarkan pendekatan yang intuitif dan efektif dalam klasifikasi berbasis jarak, serta memiliki fleksibilitas dalam menangani berbagai jenis data dengan beragam karakteristik. Meskipun sederhana dalam konsepnya, pemilihan parameter seperti nilai K dan metode perhitungan jarak tetap menjadi aspek penting yang perlu diperhatikan untuk mendapatkan hasil klasifikasi yang optimal dan akurat.












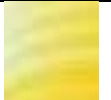
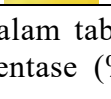
METODE PENELITIAN

Langkah pertama dalam proses ini adalah mengumpulkan data yang akan digunakan sebagai data latih dan data uji. Data yang dikumpulkan berupa gambar thermal yang berisi informasi suhu dari objek yang diamati. Berbeda dengan gambar konvensional yang menampilkan warna asli suatu objek, gambar thermal menggunakan warna untuk merepresentasikan suhu di setiap bagian objek. Warna gelap menunjukkan bahwa suhu pada bagian tersebut rendah atau dingin, yaitu di bawah 50°C, sedangkan warna terang menunjukkan suhu yang lebih tinggi atau panas, mendekati 160°C.

Data latih dan data uji terdiri dari gambar thermal berbagai objek, termasuk contact panel, housing lightning arrester, serta clamp circuit breaker pada bagian atas dan bawah. Setiap gambar thermal kemudian dipecah menjadi beberapa bagian kecil untuk memperbanyak jumlah sampel yang digunakan dalam analisis.

Tabel 1 Foto thermal

No	Foto thermal	Nama objek
1		Contact panel

No	Foto thermal	Nama objek
2		Bagian dari contact panel (data latih 1)
3		Bagian dari contact panel (data latih 2)
4		Bagian dari contact panel (data uji 1)
5		Housing Lightning arrester
6		Bagian dari housing Lightning arrester (data latih 3)
7		Bagian dari housing Lightning arrester (data latih 4)
8		Bagian dari housing Lightning arrester (data latih 5)
9		Clamp Circuit Breaker bagian atas (data latih 6)
10		Clamp Circuit Breaker bagian bawah (data latih 7)
11		Clamp Circuit Breaker bagian bawah (data latih 8)
12		Bagian Clamp Circuit Breaker bagian bawah (data latih 9)
13		Bagian Clamp Circuit Breaker bagian bawah (data uji 2)
14		Bagian Clamp Circuit Breaker bagian bawah (data latih 10)

Gambar dalam tabel sebelumnya akan melalui proses ekstraksi fitur untuk menghitung persentase (%) warna putih dan persentase (%) warna selain putih.

Data hasil ekstraksi fitur tersebut kemudian digunakan untuk melatih model K-Nearest Neighbor. Dalam pelatihan ini, jika persentase warna putih $\geq 30\%$, kondisi dianggap berbahaya, sedangkan jika persentasenya $< 30\%$, kondisi dikategorikan aman. Tahap berikutnya adalah menguji model K-Nearest Neighbor yang telah menghitung jarak antara data uji dengan masing-masing data latih menggunakan rumus 1. Hasil perhitungan akan disimpan lalu diurutkan.

Data uji ini sebelumnya juga telah melalui proses ekstraksi fitur untuk memperoleh persentase warna putih dan warna selain putih. Dengan menentukan nilai K (tingkat ketentangan) maka hasil klasifikasi K-Nearest Neighbor bisa mengklasifikasikan data uji termaksud kategori aman/berbahaya dengan menghitung jumlah data yang sudah diurutkan sebelumnya.

HASIL DAN PEMBAHASAN

Sebelum data latih dan data uji dipergunakan, proses ekstraksi fitur dilakukan untuk mendapat persentase warna putih dan persentase warna selain putih.

Tabel 2 Data ekstraksi fitur data latih dan data uji

Jenis data	% warna putih	% warna selain putih	Kategori
Data latih ke- 1	18.5556	81.4444	aman
Data latih ke- 2	36.5556	70.1111	bahaya
Data latih ke- 3	56.5476	43.4524	bahaya
Data latih ke- 4	0	101.695	aman
Data latih ke- 5	0	101.149	aman
Data latih ke- 6	1.12989	98.9585	aman
Data latih ke- 7	1.40157	98.5984	aman
Data latih ke- 8	1.63177	98.6321	aman
Data latih ke- 9	0.093985	99.906	aman
Data latih ke- 10	0.127714	103.576	aman
Data uji ke- 1	83.2647	20.439	??
Data uji ke- 2	0.086	99.91	??

Tahapan pengujian menggunakan 2 data uji. Untuk data uji pertama jarak antara data uji dan data latih sebagai berikut :

Tabel 3 Jarak antara data uji pertama dengan masing-masing data latih

Terhadap	Nilai X1	Nilai Y1	Nilai X2	Nilai Y2	Jarak	Kategori
Data latih ke- 1	18.5556	81.4444	83.2647	20.439	88.9323	Aman (0)
Data latih ke- 2	36.5556	70.1111	83.2647	20.439	68.1841	Bahaya (1)
Data latih ke- 3	56.5476	43.4524	83.2647	20.439	35.2622	Bahaya (1)
Data latih ke- 4	0	101.695	83.2647	20.439	116.342	Aman (0)
Data latih ke- 5	0	101.149	83.2647	20.439	115.962	Aman (0)
Data latih ke- 6	1.12989	98.9585	83.2647	20.439	113.629	Aman (0)
Data latih ke- 7	1.40157	98.5984	83.2647	20.439	113.183	Aman (0)
Data latih ke- 8	1.63177	98.6321	83.2647	20.439	113.04	Aman (0)
Data latih ke- 9	0.093985	99.906	83.2647	20.439	115.032	Aman (0)
Data latih ke- 10	0.127714	103.576	83.2647	20.439	117.574	Aman (0)

Pada tabel diatas, nilai X1 dan Y1 adalah nilai % warna putih dan % warna

selain putih pada data latih. Sedangkan nilai X2 dan Y2 adalah nilai % warna putih dan % warna selain putih pada data uji. Sedangkan kategori merupakan kondisi klasifikasi gambar thermal. Lalu data tabel 2 diurutkan dari yang bernilai kecil menuju bernilai besar.

Tabel 4 Pengurutan Jarak antara data uji pertama dengan masing-masing data latih

Urutan	Jarak	Kategori
1	35.2622	Bahaya (1)
2	68.1841	Bahaya (1)
3	88.9323	Aman (0)
4	113.04	Aman (0)
5	113.183	Aman (0)
6	113.629	Aman (0)
7	115.032	Aman (0)
8	115.962	Aman (0)
9	116.342	Aman (0)
10	117.574	Aman (0)

Berdasarkan tabel diatas yang telah diurutkan berdasarkan kedekatan jarak, tahap pengujian dapat dilakukan dengan menggunakan berbagai nilai K untuk menentukan kategori data uji. Nilai K ini menunjukkan jumlah data tetangga terdekat yang digunakan dalam proses klasifikasi.

- Jika $K = 1$, maka hanya data dengan urutan pertama yang digunakan. Jika data ini termasuk dalam kategori bahaya (1), maka hasil klasifikasi juga bahaya.
- Jika $K = 2$, maka dua data pertama digunakan. Jika kedua data ini termasuk dalam kategori bahaya (1,1), maka hasil klasifikasi tetap bahaya.
- Jika $K = 3$, tiga data pertama diperhitungkan. Jika terdapat dua data kategori bahaya (1,1) dan satu data kategori aman (0), maka hasil akhir mengikuti kategori terbanyak, yaitu bahaya.
- Jika $K = 4$, empat data pertama digunakan. Jika terdapat dua data kategori bahaya (1,1) dan dua data kategori aman (0,0), maka terjadiimbang, sehingga sistem gagal mengidentifikasi kategori dengan pasti.
- Jika $K = 5$, lima data pertama diperhitungkan. Jika terdapat dua data kategori bahaya (1,1) dan tiga data kategori aman (0,0,0), maka sistem mengklasifikasikan data sebagai aman, karena kategori ini lebih dominan.
- Jika $K = 6$, enam data pertama dipertimbangkan. Dengan dua data kategori bahaya (1,1) dan empat data kategori aman (0,0,0,0), hasil klasifikasi tetap aman, karena jumlah data kategori aman lebih banyak.
- Jika $K = 7$, tujuh data pertama digunakan. Jika terdapat dua data kategori bahaya (1,1) dan lima data kategori aman (0,0,0,0,0), maka sistem menetapkan bahwa data termasuk dalam kategori aman.

Dari hasil klasifikasi ini, dapat disimpulkan bahwa semakin besar nilai K, semakin besar kemungkinan sistem memilih kategori dengan jumlah data terbanyak. Hal ini menunjukkan bahwa pemilihan nilai K yang optimal sangat penting agar sistem dapat memberikan klasifikasi yang lebih akurat, terutama ketika data dalam kategori tertentu lebih sedikit dibandingkan kategori lainnya.

Data pada tabel diatas akan diurutkan menjadi :

Tabel 5 Jarak antara data uji kedua dengan masing-masing data latih

Terhadap	Nilai X1	Nilai Y1	Nilai X2	Nilai Y2	Jarak	Kategori
Data latih ke- 1	18.5556	81.4444	0.0886525	102.039	27.6616	Aman (0)
Data latih ke- 2	36.5556	70.1111	0.0886525	102.039	48.4688	Bahaya (1)
Data latih ke- 3	56.5476	43.4524	0.0886525	102.039	81.3634	Bahaya (1)
Data latih ke- 4	0	101.695	0.0886525	102.039	0.35533	Aman (0)
Data latih ke- 5	0	101.149	0.0886525	102.039	0.893994	Aman (0)
Data latih ke- 6	1.12989	98.9585	0.0886525	102.039	3.2517	Aman (0)
Data latih ke- 7	1.40157	98.5984	0.0886525	102.039	3.68258	Aman (0)
Data latih ke- 8	1.63177	98.6321	0.0886525	102.039	3.74011	Aman (0)
Data latih ke- 9	0.093985	99.906	0.0886525	102.039	2.133	Aman (0)
Data latih ke- 10	0.127714	103.576	0.0886525	102.039	1.53748	Aman (0)

Data pada tabel diatas akan diurutkan menjadi :

Tabel 6 Pengurutan Jarak antara data uji kedua dengan masing-masing data latih

Urutan	Jarak	Kategori
1	0.35533	Aman (0)
2	0.893994	Aman (0)
3	1.53748	Aman (0)
4	2.133	Aman (0)
5	3.2517	Aman (0)
6	3.68258	Aman (0)
7	3.74011	Aman (0)
8	27.6616	Aman (0)
9	48.4688	Bahaya (1)
10	81.3634	Bahaya (1)

Berdasarkan tabel diatas yang telah diurutkan berdasarkan kedekatan jarak, pengujian dapat dilanjutkan dengan menggunakan data uji kedua. Proses ini bertujuan untuk menentukan kategori data berdasarkan jumlah tetangga terdekat yang dianalisis, dengan berbagai nilai K yang menunjukkan jumlah data latih yang digunakan dalam klasifikasi.

- Jika $K = 1$, hanya data dengan urutan pertama yang digunakan. Data ini termasuk dalam kategori aman (0), sehingga hasil klasifikasi untuk data uji juga aman.
- Jika $K = 2$, dua data pertama dipertimbangkan. Jika kedua data ini termasuk dalam kategori aman (0,0), maka hasil klasifikasi tetap aman.
- Jika $K = 3$, tiga data pertama diperhitungkan. Jika ketiga data ini termasuk dalam kategori aman (0,0,0), maka sistem mengklasifikasikan data uji sebagai aman, karena seluruh data tetangga termasuk dalam kategori tersebut.
- Jika $K = 4$, empat data pertama digunakan. Dengan seluruh data tetangga termasuk dalam kategori aman (0,0,0,0), sistem tetap mengidentifikasi data sebagai aman.
- Jika $K = 5$, lima data pertama digunakan dalam perhitungan. Karena seluruhnya masih dalam kategori aman (0,0,0,0,0), hasil klasifikasi tidak berubah, yaitu aman.
- Jika $K = 6$, enam data pertama diperhitungkan. Dengan seluruh data tetangga masih tergolong dalam kategori aman (0,0,0,0,0,0), hasil klasifikasi tetap aman.
- Jika $K = 7$, tujuh data pertama digunakan. Karena semua data ini termasuk dalam kategori aman (0,0,0,0,0,0,0), maka sistem tetap mengidentifikasi data sebagai aman.

Dari hasil pengujian ini, terlihat bahwa semakin banyak data latih yang dipertimbangkan (nilai K lebih besar), semakin stabil hasil klasifikasi yang diperoleh. Jika seluruh data latih yang menjadi tetangga terdekat memiliki kategori yang sama, maka sistem akan lebih mudah dan konsisten dalam mengklasifikasikan data uji ke dalam kategori yang benar. Hal ini menunjukkan pentingnya distribusi data latih yang seimbang agar proses klasifikasi dapat berjalan dengan lebih optimal.

Aplikasi K-Nearest Neighbor (KNN) dalam penelitian ini dikembangkan menggunakan bahasa pemrograman C++, yang dikenal sebagai bahasa pemrograman berorientasi objek dengan performa tinggi. Untuk proses pengembangan dan implementasi, digunakan Code::Blocks versi 12.11 sebagai lingkungan pengembangannya. Code::Blocks 12.11 merupakan Integrated Development Environment (IDE) berbasis open source yang menyediakan berbagai fitur untuk mempermudah pengkodean, debugging, serta pengelolaan proyek pemrograman C++. IDE ini mendukung berbagai compiler, salah satunya adalah GNU GCC (GNU Compiler Collection), yang digunakan dalam pembuatan aplikasi KNN ini. Keunggulan utama dari penggunaan Code::Blocks 12.11 adalah fleksibilitasnya dalam menangani proyek-proyek berbasis C++ serta kemampuannya untuk berjalan di berbagai sistem operasi, termasuk Windows, Linux, dan macOS. Selain itu, karena bersifat gratis dan open source, pengguna dapat dengan mudah mengunduh dan menggunakan aplikasi ini tanpa perlu biaya lisensi. Dengan memanfaatkan Code::Blocks 12.11 serta compiler GNU GCC, aplikasi KNN yang dikembangkan dapat berjalan dengan efisien dan optimal, baik dalam melakukan perhitungan jarak antar data maupun dalam proses klasifikasi menggunakan algoritma KNN.

KESIMPULAN

Berikut hasil pengujian menggunakan 2 data uji dan menggunakan nilai K = 1, K = 2, K = 3, K = 4, K = 5, K = 6 dan K = 7.

Tabel 7 Hasil pengujian

Data uji ke-	Nilai K	Kategori	Kesesuaian
1	1	Bahaya	Sesuai
	2	Bahaya	Sesuai
	3	Bahaya	Sesuai
	4	Gagal mengidentifikasi	Tidak sesuai
	5	Aman	Tidak sesuai
	6	Aman	Tidak sesuai
	7	Aman	Tidak sesuai
2	1	Aman	Sesuai
	2	Aman	Sesuai
	3	Aman	Sesuai
	4	Aman	Sesuai
	5	Aman	Sesuai
	6	Aman	Sesuai
	7	Aman	Sesuai

Berdasarkan hasil perhitungan dan analisis pada tabel di atas, metode K-Nearest Neighbor (KNN) menghasilkan tingkat akurasi klasifikasi sebesar 71,42%. Persentase akurasi ini menunjukkan bahwa model KNN yang digunakan dalam penelitian ini memiliki keterbatasan dalam mengenali dan mengklasifikasikan data dengan tepat. Salah satu faktor utama yang menyebabkan tingkat akurasi tersebut adalah ketidakseimbangan jumlah data latih, khususnya pada kategori bahaya. Jumlah data

latih yang lebih sedikit dalam kategori ini mengakibatkan model kesulitan dalam mengenali pola dan karakteristik dari data berisiko tinggi. Akibatnya, ketika parameter K (jumlah tetangga terdekat) disetel ke nilai yang lebih besar dari 3, hasil klasifikasi cenderung menjadi kurang akurat dan tidak sesuai dengan kondisi sebenarnya. Kekurangan data latih dalam kategori bahaya membuat model cenderung lebih bias terhadap kategori aman, terutama ketika jumlah tetangga yang dipertimbangkan dalam klasifikasi meningkat. Oleh karena itu, untuk meningkatkan akurasi, diperlukan penambahan jumlah data latih pada kategori bahaya agar distribusi data lebih seimbang dan model dapat melakukan klasifikasi dengan lebih optimal.

DAFTAR PUSTAKA

- [1] G. D. Prenata, "Klasifikasi Keandalan Sistem Distribusi Tenaga Listrik di PT. PLN (PERSERO) UP3 Surabaya Selatan Menggunakan Metode Single Perceptron," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, pp. 20-25, 2024.
- [2] G. D. Prenata, "Klasifikasi Keandalan Sistem Distribusi Tenaga Listrik Di Pt. Pln (Persero) Up3 Surabaya Selatan Menggunakan Support Vector Machine (Svm)," *J. Tek. Elektro*, vol. 16, no. 2, pp. 62–70, 2023, doi: 10.9744/jte.16.2.62-70.
- [3] G. D. Prenata, "Klasifikasi Keandalan Sistem Distribusi Tenaga Listrik Di Pt. Pln (Persero) Up3 Surabaya Selatan Menggunakan Metode K-Nearest Neighbor (Knn)," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, 2023, doi: 10.23960/jitet.v11i3s1.3397.
- [4] G. D. Prenata and N. A. Basyarach, "Prediksi Impedansi Rele menggunakan metode PSO sebagai validasi pengukuran rele," pp. 181–186, 1945.
- [5] A. Khairi, A. F. Ghozali, and A. D. N. Hidayah, "Implementasi K-Nearest Neighbor (KNN) untuk Mengklasifikasi Masyarakat Pra-Sejahtera Desa Sapikerep Kecamatan Sukapura," *TRILOGI J. Ilmu Teknol. Kesehatan, dan Hum.*, vol. 2, no. 3, pp. 319–323, 2021, doi: 10.33650/trilogi.v2i3.2878.
- [6] D. M. Nanda, T. H. Pudjiantoro, and P. N. Sabrina, "Metode K-Nearest Neighbor (KNN) dalam Memprediksi Curah Hujan di Kota Bandung," *Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform.*, pp. 387–393, 2022, [Online]. Available: <https://doi.org/10.31284/p.snestik.2022.2750>
- [7] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.