

## PENERAPAN ALGORITMA *K-NEAREST NEIGHBORS* DALAM MENGKLASIFIKASI PENYAKIT *MULTIPLE SCLEROSIS*

Andrew Efraim Nicholas Sitompul, Margaretha Yohanna✉, Arina Prima Silalahi

Fakultas Ilmu Komputer, Universitas Methodist Indonesia, Medan, Indonesia

Email: [yohanna.na2@gmail.com](mailto:yohanna.na2@gmail.com)

DOI: <https://doi.org/10.46880/jmika.Vol9No2.pp307-315>

### ABSTRACT

*The central nervous system is impacted by multiple sclerosis (MS), a chronic autoimmune disease that requires early identification for successful treatment. Because of its many symptoms and similarities to other neurological disorders, MS can be difficult to diagnose. Artificial intelligence techniques like the K-Nearest Neighbors (KNN) algorithm can be used to help with quicker and more precise classification in order to solve this problem. The goal of this study is to classify MS using the KNN technique and assess how well it performs in this regard. The Kaggle platform provided the dataset, which consists of 273 patient records with 18 clinical characteristics. With  $k = 3$  as the number of neighbors, the data was split into 80% for training and 20% for testing. The Python programming language was used to implement the classification procedure. According to the findings, the KNN algorithm classified MS with an accuracy of 81.82%. The precision, recall, and f1-score for class 1 were 0.83, 0.76, and 0.79, respectively, according to additional analysis utilizing a classification report, whereas the scores for class 2 were 0.81, 0.87, and 0.84. These findings suggest that the KNN method has the potential to serve as a supportive tool in the diagnosis of Multiple Sclerosis.*

**Keyword:** *Multiple Sclerosis, K-Nearest Neighbors, Classification, Machine Learning, Disease Diagnosis.*

### ABSTRAK

*Untuk penanganan MS (multiple sclerosis) yang menyerang sistem saraf pusat, diagnosis dini diperlukan. Namun, diagnosis penyakit ini sering sulit karena gejalanya sangat beragam dan sering terlihat seperti gangguan neurologis lainnya. Untuk mengatasi masalah ini, teknologi kecerdasan buatan seperti algoritma K-Nearest Neighbors (KNN) dapat digunakan untuk mempercepat dan meningkatkan akurasi proses klasifikasi penyakit ini. Penelitian ini bertujuan untuk menggunakan algoritma KNN untuk mengelompokkan penyakit Multiple Sclerosis dan mengevaluasi tingkat akurasi metode tersebut. Data yang digunakan terdiri dari 273 data pasien dengan total 18 atribut klinis yang diperoleh dari platform Kaggle. Penelitian ini membagi data menjadi dua bagian, dengan 80% data pelatihan dan 20% data pengujian, dan nilai  $k$  adalah 3. Python adalah bahasa pemrograman yang digunakan untuk menjalankan proses klasifikasi. Hasil penelitian menunjukkan bahwa metode KNN mampu melakukan klasifikasi penyakit MS dengan tingkat akurasi sebesar 81,82%. Nilai akurasi, recall, dan f1-nilai untuk kelas 1 adalah 0,83; 0,76; dan 0,79, sedangkan nilai untuk kelas 2 adalah 0,81; 0,87; dan 0,84. Hasil menunjukkan bahwa algoritma KNN dapat digunakan sebagai salah satu alternatif untuk membantu proses diagnosis penyakit Multiple Sclerosis.*

**Kata Kunci:** *Multiple Sclerosis, K-Nearest Neighbors, Klasifikasi, Machine Learning, Diagnosis Penyakit.*

### PENDAHULUAN

Kesehatan merupakan aspek penting dalam kehidupan manusia, di mana kemajuan teknologi telah berperan besar dalam meningkatkan kualitas layanan medis, termasuk dalam bidang diagnosis penyakit. Seiring dengan berkembangnya teknologi informasi, penerapan kecerdasan buatan (Artificial Intelligence) dalam dunia medis semakin meningkat, terutama dalam mendukung proses diagnosis penyakit secara lebih cepat dan akurat. Salah satu pendekatan yang banyak digunakan dalam klasifikasi penyakit adalah

pembelajaran mesin (Machine Learning), yang dapat membantu dokter dalam menganalisis data pasien dengan lebih efisien.

Diagnosis dini yang akurat diperlukan untuk mendeteksi Multiple Sclerosis (MS). Multiple Sclerosis merupakan penyakit autoimun kronis yang menyerang sistem saraf di otak dan juga bagian sumsum tulang belakang. Penyakit ini menyebabkan sistem kekebalan tubuh menyerang lapisan pelindung saraf (myelin), yang mengakibatkan gangguan komunikasi antara otak dan tubuh (Dahl et al., 2025).

Akibatnya, penderita MS dapat mengalami berbagai gejala seperti gangguan penglihatan, kelemahan otot, kehilangan keseimbangan, kesulitan berbicara, serta gangguan kognitif. Penyakit ini bersifat progresif dan dapat memburuk seiring waktu jika tidak ditangani dengan baik (Statsenko et al., 2023).

Diagnosis Multiple Sclerosis sering kali menjadi tantangan bagi tenaga medis karena gejalanya yang bervariasi dan dapat menyerupai penyakit neurologis lainnya. Proses diagnosis biasanya melibatkan berbagai pemeriksaan seperti pencitraan resonansi magnetik (Magnetic Resonance Imaging), analisis cairan serebrospinal, dan pemeriksaan neurologis lainnya (Helme et al., 2025).

Dalam bidang kecerdasan buatan, algoritma K-Nearest Neighbors (KNN) merupakan salah satu metode klasifikasi yang banyak digunakan dalam pengolahan data medis (Mansouri et al., 2024). Algoritma ini bekerja dengan cara mengklasifikasikan data berdasarkan kedekatan dengan data lain yang sudah memiliki label (Kiram et al., 2025). KNN memiliki keunggulan dalam kesederhanaan implementasi dan kemampuannya untuk bekerja dengan data dalam berbagai skala (Sabita & Yahfizham, 2024).

Penelitian yang dilakukan oleh Ameer Ali et al (Ali et al., 2020) memanfaatkan berbagai varian algoritma K-Nearest Neighbors (KNN) seperti Fine, Weighted, Medium, dan Cubic untuk klasifikasi penyakit diabetes menggunakan perangkat lunak MATLAB. Hasil pengujian menunjukkan bahwa varian Fine KNN menghasilkan akurasi tertinggi sehingga direkomendasikan sebagai metode yang efektif untuk mengklasifikasikan data pasien berdasarkan kriteria yang ditetapkan oleh American Diabetes Association. Studi lain oleh Haviluddin et al. (Haviluddin et al., 2022) melakukan perbandingan kinerja antara algoritma Naïve Bayes (NB) dan KNN dalam diagnosis dini Diabetes Mellitus dengan menggunakan dataset dari Rumah Sakit Abdul Moeis di Samarinda. Hasil penelitian tersebut memperlihatkan bahwa KNN memiliki kinerja lebih unggul dibanding NB, dengan selisih akurasi sekitar 10 % lebih tinggi serta nilai presisi, spesifisitas, dan Area Under the Curve (AUC) yang juga lebih baik. Sementara itu, Jamil, Warsito, dan Wibowo (Jamil et al., 2023) mensimulasikan deteksi dini diabetes mellitus menggunakan algoritma KNN pada lingkungan cloud-based runtime COLAB. Penelitian tersebut juga memanfaatkan dataset Pima Indians Diabetes, dan hasilnya memperlihatkan kinerja model yang efektif untuk pengolahan data medis secara daring. Meskipun

kedua penelitian tersebut berhasil menunjukkan efektivitas KNN, fokusnya masih pada penyakit yang gejalanya relatif terukur dan dataset yang sudah umum digunakan. Berbeda dengan itu, penelitian ini menerapkan KNN untuk mengklasifikasikan penyakit Multiple Sclerosis (MS) yang memiliki gejala kompleks dan mirip dengan gangguan neurologis lain. Dengan memanfaatkan dataset klinis dari platform Kaggle yang memuat 18 atribut medis, penelitian ini diharapkan dapat mengisi kekosongan penelitian sebelumnya dan menawarkan kontribusi baru dalam penerapan pembelajaran mesin untuk diagnosis penyakit neurologis yang sulit diidentifikasi.

Kebaruan dalam penelitian ini adalah fokus penerapan algoritma K-Nearest Neighbors (KNN) khusus untuk klasifikasi penyakit Multiple Sclerosis (MS), yang hingga kini masih jarang dibahas dalam penelitian terdahulu. Sebagian besar studi sebelumnya memanfaatkan KNN untuk deteksi penyakit umum seperti diabetes, penyakit jantung, atau obesitas, sementara penerapannya pada MS dengan kompleksitas gejala yang beragam masih terbatas. Selain itu, penelitian ini menggunakan dataset klinis spesifik dari platform Kaggle yang terdiri dari 18 atribut medis, yang diolah melalui pembagian data latih dan uji serta evaluasi menyeluruh menggunakan akurasi, precision, recall, dan F1-score pada kedua kelas. Hasil penelitian ini diharapkan tidak hanya memberikan model klasifikasi yang akurat, tetapi juga menjadi langkah awal pengembangan sistem pendukung keputusan medis berbasis kecerdasan buatan untuk membantu diagnosis dini MS secara lebih cepat dan efisien.

## **KAJIAN LITERATUR/TINJAUAN PUSTAKA**

### **Multiple Sclerosis**

Multiple Sclerosis (MS) adalah suatu penyakit neurologis kronis yang memengaruhi sistem saraf pusat, yaitu otak dan sumsum tulang belakang. Penyakit ini termasuk dalam kategori autoimun, di mana sistem kekebalan tubuh yang seharusnya melindungi tubuh dari infeksi justru menyerang jaringan tubuh sendiri, khususnya lapisan pelindung yang disebut mielin. Mielin berfungsi sebagai pelapis saraf dan memungkinkan transmisi impuls listrik yang cepat dan efisien antara otak dan seluruh tubuh. Ketika mielin rusak atau hilang akibat serangan sistem imun, sinyal antara otak dan tubuh menjadi terganggu atau bahkan terputus sama sekali. Kerusakan ini dapat menyebabkan berbagai gejala fisik dan kognitif yang bervariasi antar individu.

Gejala Multiple Sclerosis sangat beragam, tergantung pada lokasi dan tingkat keparahan kerusakan saraf. Beberapa gejala umum termasuk kelemahan otot, kesemutan atau mati rasa, gangguan penglihatan seperti penglihatan kabur atau ganda, gangguan keseimbangan dan koordinasi, serta kelelahan yang ekstrem. Dalam kasus yang lebih parah, MS dapat menyebabkan kelumpuhan atau kehilangan kemampuan untuk berbicara. Selain gejala fisik, pasien MS juga dapat mengalami gangguan emosional dan kognitif, seperti depresi, gangguan memori, dan kesulitan berkonsentrasi (Helme et al., 2025).

Penyebab pasti dari MS hingga kini masih belum sepenuhnya dipahami. Namun, para ilmuwan meyakini bahwa MS disebabkan oleh kombinasi faktor genetik dan lingkungan. Faktor risiko yang dapat meningkatkan kemungkinan seseorang terkena MS meliputi riwayat keluarga dengan MS, infeksi virus tertentu seperti Epstein-Barr Virus (EBV), kekurangan vitamin D, dan gaya hidup yang tidak sehat. MS lebih sering terjadi pada wanita dibandingkan pria, dengan perbandingan sekitar 2:1, dan biasanya mulai muncul pada usia antara 20 hingga 40 tahun (Dahl et al., 2025).

### K-Nearest Neighbor (KNN)

K-nearest neighbor (KNN) adalah metode yang mudah untuk digunakan dalam melakukan klasifikasi teks. KNN adalah salah satu metode supervised learning yang proses klasifikasinya berdasarkan jarak ketetanggaan terdekat dari dataset training. K-NN merupakan salah satu metode non parametrik yang prinsip kerjanya adalah mengklasifikasikan objek pada set pengujian berdasarkan kelas terbaik dari k tetangga terdekat pada data pelatihan. Nilai k merupakan bilangan bulat positif yang ditentukan secara acak oleh pengguna. Untuk mendapatkan nilai k yang baik dapat dipilih dengan optimasi parameter (Sipayung et al., 2024).

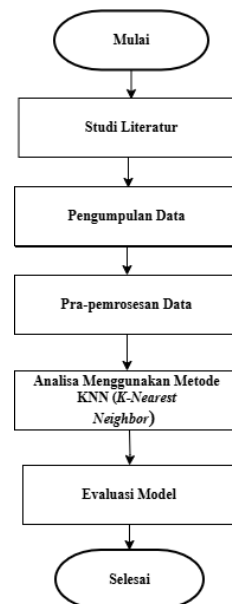
Algoritma k-Nearest Neighbor merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. k-NN termasuk algoritma supervised learning dimana hasil dari query instance yang baru diklasifikasi berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul itu yang akan menjadi kelas hasil klasifikasi. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut dan training sample (Yasin & Widodo, 2025).

K-Nearest Neighbor (KNN) adalah salah satu algoritma Machine Learning yang digunakan untuk klasifikasi dan regresi. Algoritma ini didasarkan pada

gagasan sederhana bahwa objek yang serupa dalam ruang fitur juga akan memiliki label atau nilai yang sama. Dengan kata lain, KNN memprediksi berdasarkan data baru yang didekati dengan tetangga terdekatnya. Metode Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru (testing data) dengan kasus lama (training data).

### METODE PENELITIAN

Framework penelitian dalam penelitian ini dirancang untuk menggambarkan secara sistematis seluruh tahapan yang dilakukan agar penelitian berjalan sesuai dengan tujuan yang telah ditetapkan. Kerangka kerja ini mencakup proses mulai dari studi literatur, pengumpulan data, penerapan metode, hingga evaluasi hasil. Kerangka kerja pada penelitian ini dapat dilihat pada Gambar 1.



**Gambar 1.** Framework Penelitian

Gambar 1. menampilkan bagan alur penelitian yang menggambarkan tahapan kegiatan dari awal hingga akhir secara sistematis. Proses dimulai pada tahap Mulai, yang berfungsi sebagai penanda dimulainya seluruh rangkaian penelitian. Pada bagian ini, peneliti menetapkan perumusan masalah, tujuan, serta ruang lingkup kajian agar proses yang dijalankan memiliki arah yang jelas. Tahap ini menjadi landasan penting karena akan menentukan keberhasilan langkah-langkah selanjutnya.

Tahap kedua adalah Studi Literatur, yang berfokus pada penelusuran sumber-sumber ilmiah seperti jurnal, buku, dan publikasi penelitian sebelumnya. Aktivitas ini dilakukan untuk memperoleh

pemahaman yang komprehensif mengenai teori dan metode yang relevan dengan topik penelitian. Selain itu, studi literatur membantu peneliti mengidentifikasi kekosongan penelitian (research gap) yang akan menjadi dasar kebaruan dari karya ilmiah yang disusun.

Berikutnya adalah tahap Pengumpulan Data diikuti dengan Pra-pemrosesan Data. Data yang dikumpulkan disesuaikan dengan kebutuhan penelitian, dalam hal ini meliputi data klinis terkait penyakit Multiple Sclerosis dari sumber terpercaya. Setelah diperoleh, data diproses melalui tahapan pembersihan (cleaning), standarisasi, dan pembagian ke dalam data latih serta data uji. Langkah ini memastikan bahwa data yang digunakan siap dianalisis dan dapat memberikan hasil yang akurat saat model dibangun.

Tahap inti terletak pada Analisa Menggunakan Metode KNN (K-Nearest Neighbor), di mana algoritma ini memproses data dengan mengukur kedekatan antar titik data menggunakan metode jarak tertentu seperti Euclidean Distance (Sipayung et al., 2024). Proses ini menghasilkan model klasifikasi yang kemudian diuji pada tahap Evaluasi Model untuk mengetahui performa

berdasarkan metrik seperti akurasi, presisi, recall, dan F1-score. Rangkaian penelitian diakhiri pada tahap Selesai, yang menandakan berakhirnya proses setelah seluruh evaluasi dan interpretasi hasil dilakukan secara menyeluruh.

## HASIL DAN PEMBAHASAN

Pada tahap implementasi K- Nearest Nighbor (KNN) proses perhitungan dilakukan menggunakan dengan K=3 dilakukan menggunakan perbandingan data yaitu 80:20, yaitu data 1-218 merupakan data latih (total data latih 218 data) dan data uji yaitu data 219-273 (total data uji 55 data).

Perhitungan jarak dari setiap data uji terhadap masing- masing data latih digunakan rumus jarak Euclidean Distance hingga ditemukan nilai minimum terhadap setiap data latih. Berikut merupakan tahapan dari perhitungan menggunakan K-Nearest Neighbor (KNN) dengan mencari nilai Euclidean Distance (Farkhatun Zaidah & Supatman, 2025). Data yang digunakan dalam penelitian ini yaitu data penyakit multiple sclerosis dapat dilihat pada Tabel 1.

**Tabel 1.** Data Penyakit Multiple Sclerosis

No	Gender	Age	Breastfeeding	Varicella	Initial_Symptom	Mono_or_Polysymptomatic	Oligoclonal_Bands	LLSSEP	...	group
1	1	34	1	1	2	1	0	1		1
2	1	61	3	2	10	2	1	1		1
3	1	22	3	1	3	1	1	0		1
4	2	41	1	1	7	2	1	0		1
5	2	34	2	1	6	2	0	1		1
6	1	29	1	1	6	2	0	1		1
7	2	53	1	1	14	2	0	1		1
8	2	24	1	1	14	2	0	1		1
9	1	36	1	1	8	2	0	1		1
10	2	28	1	1	8	2	0	0		1
11	2	60	3	2	15	2	0	1		1
12	2	25	1	1	5	2	0	1		1
13	1	34	1	1	11	2	0	1		1
14	1	36	1	1	13	2	1	0		1
15	2	29	1	1	1	1	0	1		1
.										
.										
.										
273	2	77	3	3	2	1	0	0		2

Berikut adalah deskripsi Variabel gejala dalam dataset penyakit Multiple Sclerosis (MS) dapat dilihat pada Tabel 2

**Tabel 2.** Deskripsi Variabel

No	Variabel	Keterangan
1	Gender	Jenis kelamin: 1 = Laki-laki, 2 = Perempuan.
2	Age	Usia pasien dalam tahun.
3	Breastfeeding	Riwayat Menyusui : 1 = Ya, 2 = Tidak, 3 = Tidak diketahui.
4	Varicella	Riwayat Cacar Air : 1 = Positif, 2 = Negatif, 3 = Tidak diketahui.
5	Initial_Symptom	Gejala Awal: 1 = Visual (penglihatan), 2 = Sensorik, 3 = Motorik, 4 = Lainnya, 5-15 =

		Kombinasi berbagai gejala awal Mono atau Polisimptomatik :
6	Mono_or_Polysymptomatic	1 = Monosimptomatik (satu gejala), 2 = Polisimptomatik (lebih dari satu gejala), 3 = Tidak diketahui.
7	Oligoclonal_Bands	Pita Oligoklonal : 0 = Negatif, 1 = Positif, 2 = Tidak diketahui.
8	LLSSEP	Respon Sensorik Somatosensori Bawah: 0 = Negatif, 1 = Positif.
9	ULSSEP	Respon Sensorik Somatosensori Atas : 0 = Negatif, 1 = Positif.
10	VEP	Potensial Evoked Visual: 0 = Negatif, 1 = Positif.
11	BAEP	Potensial Evoked Auditorik Batang Otak: 0 = Negatif, 1 = Positif.
12	Periventricular_MRI	Hasil MRI Periventrikular: 0 = Negatif, 1 = Positif.
13	Cortical_MRI	Hasil MRI Kortikal: 0 = Negatif, 1 = Positif.
14	Infratentorial_MRI	Hasil MRI Infratentorial: 0 = Negatif, 1 = Positif.
15	Spinal_Cord_MRI	Hasil MRI Sumsum Tulang Belakang : 0 = Negatif, 1 = Positif
16	Initial_EDSS	Skor awal Expanded Disability Status Scale (EDSS) untuk mengukur tingkat keparahan MS.
17	Final_EDSS	Skor EDSS setelah periode pengamatan.
18	Group	Kelompok : 1 = CDMS (Clinically Definite Multiple Sclerosis), 2 = non-CDMS.

Berikut contoh perhitungan nilai jarak antara data uji yaitu data uji D219 dengan setiap data latih menggunakan Euclidean Distance.

$$D1 = \sqrt{(D1_1 - D219_1)^2 + (D1_{\dots} - D219_{\dots})^2 + (D1_{17} - D219_{17})^2}$$

$$D1 = \sqrt{(2 - 2)^2 + (\dots - \dots)^2 + (1 - 2)^2}$$

$$D1 = 24,23$$

$$D2 = \sqrt{(D2_1 - D219_1)^2 + (D2_{\dots} - D219_{\dots})^2 + (D2_{17} - D219_{17})^2}$$

$$D2 = \sqrt{(2 - 2)^2 + (\dots - \dots)^2 + (2 - 2)^2}$$

$$D2 = 3,74$$

$$D3 = \sqrt{(D3_1 - D219_1)^2 + (D3_{\dots} - D219_{\dots})^2 + (D3_{17} - D219_{17})^2}$$

$$D3 = \sqrt{(2 - 2)^2 + (\dots - \dots)^2 + (1 - 2)^2}$$

$$D3 = 6,16$$

·  
·  
·

$$D218 = \sqrt{(D218_1 - D219_1)^2 + (D218_{\dots} - D219_{\dots})^2 + (D218_{17} - D219_{17})^2}$$

$$D218 = \sqrt{(2 - 2)^2 + (\dots - \dots)^2 + (0 - 2)^2}$$

$$D218 = 14,56$$

Berdasarkan hasil perhitungan kemudian diurutkan jaraknya (Rank) antara data latih dengan data uji, dari yang terdekat ke terjauh. Setelah dilakukan pengurutan jarak hasil perhitungan nilai Euclidean Distance dilakukan penentuan Label berdasarkan dari tetangga terdekat (Yasin & Widodo, 2025).

Perhitungan jarak dilakukan terhadap semua data uji dengan setiap data latih menggunakan Euclidean Distance. Selanjutnya mengurutkan jarak dari hasil perhitungan nilai Euclidean Distance dari setiap data latih terhadap data uji untuk menentukan tetangga terdekat berdasarkan jarak minimum K (Silalahi & Simanullang, 2023). Hasil lengkap

perhitungan Euclidean Distance terhadap semua data uji yaitu D219 -D273 dapat dilihat pada Tabel 3.

**Tabel 3.** Hasil Perhitungan Euclidean Distance

Data Latih	Data Uji D219		Data Uji D220		Data Uji D221		...	Data Uji D273	
	Nilai	Rank	Nilai	Rank	Nilai	Rank		Nilai	Rank
	<b>Distance</b>		<b>Distance</b>		<b>Distance</b>			<b>Distance</b>	
D1	24,23	202	20,37	172	37,39	204		15,00	139
D2	3,74	2	10,86	53	13,30	60		12,08	102
D3	6,16	29	13,45	86	9,64	32		15,36	146
D4	14,25	155	17,44	138	23,11	139		12,69	115
D5	10,44	99	14,83	111	5,83	9		20,27	192
D6	7,68	52	6,00	3	19,34	105		7,81	46
D7	9,11	77	5,48	2	19,70	109		9,00	64
D8	6,71	36	8,89	26	19,24	104		6,71	39
D9	12,88	138	10,25	42	21,75	133		12,25	105
D10	9,75	88	15,30	116	23,28	141		4,36	20
D11	10,82	108	19,75	168	4,90	6		20,47	195
D12	13,67	144	20,88	177	12,33	52		20,37	194
D13	15,65	171	17,72	142	12,00	48		22,38	206
D14	10,72	106	10,00	40	16,79	92		13,89	131
D15	3,32	1	10,68	51	13,49	62		11,36	86
D16	13,67	144	16,31	126	26,38	167		6,86	40
D17	14,32	156	15,07	114	14,00	69		19,97	190
D18	8,60	69	7,68	16	16,16	87		12,49	109
D19	7,14	45	6,40	7	20,25	112		7,00	41
.									
.									
.									
D218	14,56	161	34,73	213	26,59	168		8,00	49

Nilai rank diperoleh dari pengurutan hasil perhitungan jarak menggunakan Euclidean Distance, hasil jarak terendah akan menjadi rank pertama (Firmansyah, 2020), dengan k=3 maka dipilih 3 jarak terdekat terhadap data latih. Berdasarkan hasil pengurutan nilai jarak (rank) untuk data uji D219 yang telah diurutkan berdasarkan rangking dapat yang dilihat pada tabel 4, diperoleh bahwa tiga data latih terdekat adalah sebagai berikut:

1. Urutan pertama (terdekat) adalah D15 dengan label 1,
2. Urutan kedua adalah D2 dengan label 1,
3. Urutan ketiga adalah D106 dengan label 2.

**Tabel 4.** Hasil Ranking Data Uji D219

DataLatih	Label	Distance	Rank
	Data Latih		
D15	1	3,32	1
D2	1	3,74	2
D106	2	3,74	3
...			
D194	2	48,24	218

Dari ketiga data latih terdekat tersebut, terdapat dua data yang memiliki label 1 dan satu data yang memiliki label 2. Oleh karena itu, berdasarkan prinsip mayoritas dalam metode K-Nearest Neighbors (KNN),

data uji D219 diklasifikasikan ke dalam label 1, karena label tersebut lebih dominan (mayoritas) di antara K=3 tetangga terdekatnya. Perbandingan label aktual dan label prediksi KNN dapat dilihat pada Tabel 5

**Tabel 5.** Perbandingan Label Aktual dan Label Prediksi KNN

No Data	Label Aktual	Label Prediksi KNN
Data Uji 219	1	1
Data Uji 220	1	1
Data Uji 221	1	1
Data Uji 222	1	2
Data Uji 223	1	1
Data Uji 224	1	1
Data Uji 225	1	1
Data Uji 226	1	1
Data Uji 227	1	2
Data Uji 228	1	1
Data Uji 229	1	1
Data Uji 230	1	1
Data Uji 231	1	1
Data Uji 232	1	1
Data Uji 233	1	1
Data Uji 234	1	1
Data Uji 235	1	1
Data Uji 236	1	1
Data Uji 237	1	1
Data Uji 238	1	2
Data Uji 239	1	2
Data Uji 240	1	2
Data Uji 241	1	2
Data Uji 242	1	1

Data Uji 243	1	1
Data Uji 244	2	2
Data Uji 245	2	1
Data Uji 246	2	2
Data Uji 247	2	1
Data Uji 248	2	2
Data Uji 249	2	2
Data Uji 250	2	2
Data Uji 251	2	2
Data Uji 252	2	2
Data Uji 253	2	2
Data Uji 254	2	2
Data Uji 255	2	2
Data Uji 256	2	2
Data Uji 257	2	2
Data Uji 258	2	2
Data Uji 259	2	2
Data Uji 260	2	2
Data Uji 261	2	2
Data Uji 262	2	2
Data Uji 263	2	2
Data Uji 264	2	2
Data Uji 265	2	2
Data Uji 266	2	2
Data Uji 267	2	2
Data Uji 268	2	2
Data Uji 269	2	2
Data Uji 270	2	2
Data Uji 271	2	2
Data Uji 272	2	2

Label 1 merupakan pasien yang terdeteksi penyakit MS (Multiple Sclerosis) dan Label 2 merupakan pasien non-MS.

Untuk mengevaluasi kinerja model klasifikasi, salah satu metode yang umum digunakan adalah confusion matrix. Confusion matrix merupakan representasi tabel yang menyajikan perbandingan antara hasil prediksi model dengan label sebenarnya dari data uji (Firyal Laila Ramadhina et al., 2024), yang dapat dilihat pada tabel 3. Melalui confusion matrix, dapat diketahui jumlah prediksi yang benar dan salah dari masing-masing kelas, sehingga memberikan gambaran yang lebih detail mengenai kinerja model dibandingkan hanya menggunakan nilai akurasi saja. Evaluasi ini sangat penting dalam memahami bagaimana model mengklasifikasikan data, terutama dalam kasus data tidak seimbang antar kelas. Dari confusion matrix, dapat dihitung metrik evaluasi penting seperti akurasi, presisi, recall, dan F1-score yang digunakan untuk menilai efektivitas model klasifikasi secara menyeluruh (Pratama et al., 2024).

Untuk menghitung confusion matrix dari tabel 3 di cari True Positive (TP) True Negative (TN), False Positive (FP) dan False Negative (FN). Terdapat empat nilai utama yang tercantum pada matriks ini:

1. True Positive (TP)=26: Data yang berasal dari kelas 2 dan diprediksi dengan benar sebagai kelas 2.

2. True Negative (TN) = 19: Data yang berasal dari kelas 1 dan diprediksi dengan benar sebagai kelas 1.
3. False Positive (FP)= 6: Data yang berasal dari kelas 1 tetapi salah diprediksi sebagai kelas 2.
4. False Negative (FN) = 4: Data yang berasal dari kelas 2 tetapi salah diprediksi sebagai kelas 1.

Akurasi adalah metrik evaluasi yang menunjukkan proporsi dari seluruh prediksi yang benar. Ini mencerminkan seberapa sering model membuat prediksi yang tepat, baik untuk kelas positif maupun negatif (Valerian et al., 2025). Untuk menghitung nilai akurasi:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{26 + 19}{26 + 19 + 6 + 4} = \frac{45}{55} = 0.818 \text{ atau } 81.8\%$$

Presisi mengukur seberapa banyak dari semua prediksi kelas positif (kelas 2) yang benar-benar merupakan kelas positif. Ini penting ketika kesalahan positif (false positive) perlu diminimalkan (Valerian et al., 2025). Presisi untuk setiap kelas:

$$\text{Precision}_1 = \frac{TP}{TP + FP} = \frac{19}{19 + 4} = \frac{19}{23} = 0.8261$$

$$\text{Precision}_2 = \frac{TP}{TP + FP} = \frac{26}{26 + 6} = \frac{26}{32} = 0.8125$$

Recall mengukur seberapa banyak dari seluruh data kelas positif yang berhasil dikenali oleh model. Ini penting ketika kita ingin meminimalkan kesalahan negatif (false negative) (Valerian et al., 2025). Recall setiap kelas:

$$\text{Recall}_1 = \frac{TP}{TP + FN} = \frac{19}{19 + 6} = \frac{19}{25} = 0.76$$

$$\text{Recall}_2 = \frac{TP}{TP + FN} = \frac{26}{26 + 4} = \frac{26}{30} = 0.8667$$

Terakhir dilakukan perhitungan F1 Score. F1-score adalah metrik yang menggabungkan precision dan recall secara harmonis. Nilai ini sangat berguna ketika terdapat ketidakseimbangan antara data positif dan negatif, serta ketika kita ingin mempertimbangkan keseimbangan antara precision dan recall (Valerian et al., 2025). Recall setiap kelas:

$$F1_1 = 2 \cdot \frac{\text{Precision}_1 \cdot \text{Recall}_1}{\text{Precision}_1 + \text{Recall}_1} = 2 \cdot \frac{0.8261 \cdot 0.76}{0.8261 + 0.76} = 2 \cdot \frac{0.6278}{1.5861} = 0.7913$$

$$F1_2 = 2 \cdot \frac{0.8125 \cdot 0.8667}{0.8125 + 0.8667} = 2 \cdot \frac{0.7042}{1.6792} \approx 0.8385$$

Berdasarkan hasil evaluasi performa model K-Nearest Neighbor (KNN) yang ditunjukkan melalui confusion matrix dan classification report, dapat disimpulkan bahwa model memiliki tingkat akurasi yang cukup baik, yaitu sebesar 81,82%. Hal ini menunjukkan bahwa dari seluruh data yang diuji, sekitar 82% di antaranya berhasil diklasifikasikan dengan benar oleh model.

Lebih lanjut, dilihat dari nilai precision, recall, dan f1-score, performa model juga cukup seimbang antara kedua kelas yang diklasifikasikan. Precision untuk kelas 1 mencapai 82,61%, artinya dari seluruh prediksi yang diklasifikasikan sebagai kelas 1, sekitar 82,61% merupakan prediksi yang benar. Sementara itu, precision untuk kelas 2 adalah 81,25%, yang juga menunjukkan tingkat ketepatan prediksi yang tinggi.

Recall untuk kelas 1 berada pada angka 76%, yang berarti model mampu mengenali 76% dari seluruh data aktual kelas 1. Sedangkan recall untuk kelas 2 berada pada angka 86,67%, menandakan bahwa model lebih baik dalam mengenali data dari kelas 2. Hal ini juga terlihat dari nilai f1-score, di mana kelas 1 memperoleh skor sebesar 79,13% dan kelas 2 memperoleh skor sebesar 83,85%, yang mencerminkan keseimbangan antara precision dan recall.

## KESIMPULAN

Dari hasil penelitian yang telah dilakukan dengan metode K-Nearest Neighbor (KNN) untuk klasifikasi penyakit Multiple Sclerosis, maka dapat disimpulkan bahwa algoritma KNN dapat diterapkan secara efektif dalam proses klasifikasi data pasien yang mengidap penyakit Multiple Sclerosis. Penggunaan metode ini terbukti mampu mengklasifikasikan data dengan akurasi sebesar 81,82%, yang menunjukkan bahwa sebagian besar prediksi yang dilakukan oleh model telah sesuai dengan label sebenarnya. Evaluasi lebih lanjut melalui classification report juga menunjukkan performa yang seimbang pada masing-masing kelas, di mana untuk kelas 1 (pasien MS), nilai precision, recall, dan f1-score masing-masing adalah 0,83; 0,76; dan 0,79, sedangkan untuk kelas 2 (non-MS) masing-masing mencapai 0,81; 0,87; dan 0,84. Hasil ini menunjukkan bahwa metode KNN memiliki potensi yang baik untuk digunakan sebagai alat bantu dalam proses diagnosis penyakit Multiple Sclerosis, khususnya dalam memberikan hasil prediksi awal yang cepat dan akurat berdasarkan data klinis pasien. Temuan ini diharapkan dapat memberikan kontribusi

positif dalam pengembangan sistem pendukung keputusan medis berbasis kecerdasan buatan di masa yang akan datang.

## DAFTAR PUSTAKA

- Ali, A., Alrubei, M. A. T., Hassan, L. F. M., Al-Ja'afari, M. A. M., & Abdulwahed, S. H. (2020). Diabetes Diagnosis Based on KNN. *IJUM Engineering Journal*, 21(1), 175–181. <https://doi.org/10.31436/iijum.v21i1.1206>
- Dahl, J. R., Weier, A., Winter, C., Hintze, M., Rothhammer, V., Tsaktanis, T., Proebstel, A. K., Neziraj, T., Poessnecker, E., Oechtering, J., Kuhle, J., Kallmann, B. A., Lubner, G., Heider, T., Klotz, L., Chunder, R., & Kuerten, S. (2025). Modulator of VRAC Current 1 Is a Potential Target Antigen in Multiple Sclerosis. *Neurology(R) Neuroimmunology & Neuroinflammation*, 12(2), e200374. <https://doi.org/10.1212/NXI.0000000000200374>
- Farkhatun Zaidah, & Supatman. (2025). Implementasi Metode K-Nearest Neighbor Dalam Menentukan Klasifikasi Strata Posyandu Di Kabupaten Brebes. *JEKIN - Jurnal Teknik Informatika*, 5(1), 181–192. <https://doi.org/10.58794/jekin.v5i1.1124>
- Firmansyah, A. A. (2020). *Pengembangan Pencarian Produk Terkait Menggunakan Euclidean Distance dan Cosine Similarity Pada Aplikasi Halal Nutrition Food*. 1–79.
- Firyal Laila Ramadhina, A., Sofian, E., Esq, S., & Jakarta Selatan, C. (2024). *Perbandingan Metode K-Nearest Neighbor dan Pohon Keputusan dalam Analisis Sentimen Data Ulasan Aplikasi Pinjaman Online Berizin OJK Di Google Play*. VII, 115–124.
- Haviluddin, Puspitasari, N., Burhandeny, A. E., Nurulita, A. D. A., & Trahutomo, D. (2022). Naïve Bayes and K-Nearest Neighbor Algorithms Performance Comparison in Diabetes Mellitus Early Diagnosis. *International Journal of Online and Biomedical Engineering (IJOE)*, 18(15), 202–215. <https://doi.org/10.3991/ijoe.v18i15.34143>
- Helme, A., Kalra, D., Brichetto, G., Peryer, G., Vermersch, P., Weiland, H., White, A., & Zaratin, P. (2025). Artificial intelligence and science of patient input: a perspective from people with multiple sclerosis. *Frontiers in Immunology*, 16(February), 1–6. <https://doi.org/10.3389/fimmu.2025.1487709>
- Jamil, M., Warsito, B., & Wibowo, A. (2023). Diabetes Mellitus Early Detection Simulation using The K-Nearest Neighbors Algorithm with Cloud-Based Runtime (COLAB). *ILKOM Jurnal Ilmiah*, 15(2), 215–221. <https://doi.org/10.33096/ilkom.v15i2.1510.215-221>



- Kiram, M. A., Darnila, E., & Sahputra, I. (2025). *Machine Learning Klasifikasi Penyakit Jiwa Menggunakan Metode K-Nearest Neighbor Berbasis Web*. 9, 2445–2456.
- Mansouri, S., Boulares, S., & Chabchoub, S. (2024). Machine Learning for Early Diabetes Detection and Diagnosis. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 15(1), 216–230. <https://doi.org/10.58346/JOWUA.2024.II.015>
- Pratama, F., Hadryan Nst, Z., Khairi, Z., & Efrizoni, L. (2024). Perbandingan Algoritma Random Forest Dan K-Nearest Neighbor Dalam Klasifikasi Kesehatan Mental Mahasiswa. *Jurnal Ilmiah Betrik*, 15(1), 31–37.
- Sabita, S. A., & Yahfizham, Y. (2024). Penerapan algoritma klasifikasi nearest neighbor dalam mendeteksi penyakit diabetes. *Jurnal Bintang Pendidikan Dan Bahasa*, 2(1), 149–158.
- Silalahi, A. P., & Simanullang, H. G. (2023). Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita. *METHOMIKA: Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 7(1), 144–149. <https://doi.org/10.46880/jmika.Vol7No1.pp144-149>
- Sipayung, S. M., Yohanna, M., Manullang, H. G., & Mandala, R. (2024). Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap Film Ngeri-Ngeri Sedap. *Jurnal Ilmiah Teknik Informatika*, 4(2).
- Statsenko, Y., Smetanina, D., Arora, T., Östlundh, L., Habuza, T., Simiyu, G. L., Meribout, S., Talako, T., King, F. C., Makhnevych, I., Gelovani, J. G., Das, K. M., Gorkom, K. N. Van, Almansoori, T. M., Al Zahmi, F., Szólics, M., Ismail, F., & Ljubisavljevic, M. (2023). Multimodal diagnostics in multiple sclerosis: predicting disability and conversion from relapsing-remitting to secondary progressive disease course - protocol for systematic review and meta-analysis. *BMJ Open*, 13(7), 1–9. <https://doi.org/10.1136/bmjopen-2022-068608>
- Valerian, F. R., Syarief, M., Fatah, D. A., Informasi, S., Madura, U. T., Kamal, K., & Timur, J. (2025). *Klasifikasi tingkat obesitas menggunakan metode gbm dan confusion matrix*. 9(2), 2242–2249.
- Yasin, S. Q. F., & Widodo, A. W. (2025). Klasifikasi Tingkat Obesitas Berdasarkan Pola Hidup dan Kebiasaan Konsumsi Makanan menggunakan metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 9(3).