
SUPERVISED LEARNING METODE K-NEAREST NEIGHBOR UNTUK PREDIKSI DIABETES PADA WANITA

Arina Prima Silalahi✉, Harlen Gilbert Simanullang, Marlyna Infryanty Hutapea

Universitas Methodist Indonesia, Medan, Indonesia

Email: primaarinasilalahi@gmail.com

DOI: <https://doi.org/10.46880/jmika.Vol7No1.pp144-149>

ABSTRACT

Supervised learning is a technique of machine learning by doing learning that has a reference value to direct something, one of which is the K-Nearest Neighbor (KNN) method. This method is for object classification through learning data that is closest to the object (neighbor) using Euclidean distance to calculate the distance. KNN can be used for data classification that already has a reference, in this case the dataset used is the diabetes mellitus dataset in women. DM is a disease that can cause complications in parts of the body that cause death. DM in women can be seen from several parameters such as glucose levels, blood pressure, skin thickness, insulin hormone, body index mass, age, number of pregnancies, and the number of family history of diabetes. In this research, KNN will be used for the classification of diabetes in women with two classes, namely DM Positive and DM Negative, in other words, a woman can be predicted to suffer from DM disease or not. This method will be implemented into a system with PHP programming language and Codeigniter Framework. KNN testing is carried out with three test scenarios, the 1st test with 150 test data gets an 82% accuracy rate, the 2nd data test with 200 test data gets an 84% accuracy rate, and the 3rd data test with 300 test data gets an 82% accuracy rate.

Keywords: *Supervised Learning, K-Nearest Neighbor, Diabetes Mellitus, Euclidean Distance, Framework Codeigniter.*

ABSTRAK

Supervised learning adalah teknik dari machine learning dengan melakukan pembelajaran yang memiliki nilai acuan untuk mengarahkan sesuatu hal, salah satunya metode K-Nearest Neighbor (KNN). Metode ini untuk klasifikasi objek melalui data pembelajaran yang jaraknya paling dekat dengan objek tersebut (tetangga) menggunakan euclidean distance untuk menghitung jaraknya. KNN dapat digunakan untuk klasifikasi data yang sudah memiliki acuan, dalam hal ini dataset yang digunakan adalah dataset diabetes mellitus pada wanita. DM adalah salah satu penyakit yang dapat menyebabkan komplikasi pada bagian-bagian tubuh sehingga menyebabkan kematian. DM pada wanita dapat dilihat dari beberapa parameter seperti kadar glukosa, tekanan darah, ketebalan kulit, hormon insulin, massa indeks tubuh, umur, jumlah kehamilan, dan jumlah riwayat penderita diabetes dalam keluarga. Dalam penelitian ini, KNN akan digunakan untuk klasifikasi diabetes pada wanita dengan dua kelas yaitu DM Positif dan DM Negatif dengan kata lain seorang wanita dapat diprediksi menderita penyakit DM atau tidak. Metode ini akan diimplementasikan menjadi sebuah sistem dengan bahasa pemrograman PHP dan Framework Codeigniter. Pengujian KNN dilakukan dengan tiga skenario pengujian, uji ke-1 dengan 150 data uji mendapatkan tingkat akurasi 82%, uji data ke-2 dengan 200 data uji mendapatkan tingkat akurasi 84%, dan uji data ke-3 dengan 300 data uji mendapatkan tingkat akurasi 82%.

Kata Kunci: *Supervised Learning, K-Nearest Neighbor, Diabetes Mellitus, Euclidean Distance, Framework Codeigniter.*

PENDAHULUAN

Pengelolaan data komputer tidak terlepas menggunakan cara ataupun teknik yang membuat komputer mampu mengelola pengetahuan. Pengetahuan dari berbagai bidang dapat diterapkan ke dalam bidang teknologi sehingga dapat membantu manusia dalam melakukan banyak hal seperti melakukan prediksi pengenalan pola wajah,

pengenalan suara, dan banyak lagi. Dengan melakukan pendekatan ini maka manusia membuat komputer yang dapat berpikir seperti cara manusia berpikir dan memecahkan sebuah masalah sehingga komputer seolah-olah memiliki kecerdasan (Makinde, Vincent, Akinwale, Oguntuase, & Acheme, 2020; Saura, Palacios-Marqués, & Ribeiro-Soriano, 2021; Simanullang, Silalahi, & Sartika, 2022).

Dalam penerapannya, komputer yang memiliki kecerdasan layaknya manusia dapat menerima pengetahuan untuk mengelola banyak bidang, salah satunya bidang kesehatan. Diabetes Mellitus (DM) merupakan penyakit kritis yang menurut World Health Organization (WHO), sebanyak 300 juta populasi di dunia akan terpengaruh DM pada tahun 2025. DM juga dikatakan sebagai salah satu penyakit penyebab kematian menurut International Diabetes Federation yang di rilis pada 2021 dan Indonesia menduduki posisi keenam dengan angka kematian 236.711 jiwa. DM muncul akibat glukosa pada tubuh meningkat sehingga pankreas menemukan kesulitan dalam menghasilkan insulin yang bertugas mengolah glukosa menjadi komponen sederhana agar dapat menghasilkan zat-zat yang dibutuhkan oleh tubuh (Aulia, Abdurrahman, & Putrada, 2019). Pada dasarnya, diabetes dapat diidap oleh semua kalangan tanpa memandang umur maupun gender karena biasanya pengidap diabetes disebabkan oleh pola hidup yang tidak teratur dan faktor genetik (Karmana, 2023). Namun dalam penelitian ini difokuskan untuk meneliti data pengidap diabetes mellitus (DM) pada Wanita (Adam et al., 2023).

Seorang wanita dapat dikatakan memiliki diabetes mellitus apabila hasil pengujian lab menunjukkan bahwa wanita tersebut seorang penderita diabetes mellitus (DM) dan tentu harus melakukan banyak pengujian karena penyakit tersebut menyebabkan komplikasi di tubuh penderitanya (Syafar, Baharuddin, Zainal, & Fuadi, 2021). Adapun yang menjadi acuan untuk melihat seseorang mengidap DM adalah kadar glukosa, tekanan darah, ketebalan kulit, hormon insulin, massa indeks tubuh, umur, jumlah kehamilan, dan jumlah riwayat penderita diabetes dalam keluarga.

Dalam pengembangan dunia penelitian, data pengidap DM dapat digunakan sebagai dataset untuk melakukan pengukuran seorang wanita termasuk kelas DM Positif ataupun DM Negatif. Dataset diperoleh dari Keagle terdapat 768 data dengan delapan parameter pengujian yang sudah diolah dalam bentuk angka dan sudah dilakukan tahapan preprocessing. Dengan memanfaatkan data tersebut dapat dilakukan prediksi pasien DM positif dan DM negatif pada pasien dengan gender wanita menggunakan metode K-Nearest Neighbor. Metode tersebut dapat mengklasifikasikan berdasarkan data training melalui perhitungan jarak target ke tetangganya. Penelitian terdahulu terkait Penerapan metode K-Nearest Neighbor mendapatkan hasil akurasi sebesar maksimum sebesar 84,28%, (Andriana & Hilab, 2023), dan penelitian lainnya dengan akurasi 92,3% (Maliha, Santoso, & Furqon,

2019). Melihat beberapa kajian tersebut, dapat disimpulkan bahwa kelebihan metode ini memiliki perhitungan yang akurat dan konsisten menggunakan berbagai platform untuk pendukung klasifikasi maupun prediksi.

Melalui deskripsi permasalahan tersebut, perlu dilakukan penelitian dengan memanfaatkan salah satu bagian dari supervised learning yaitu metode K-Nearest Neighbor untuk prediksi diabetes pada wanita. Output dari prediksi ini adalah DM Positif dan DM Negatif. Dalam penelitian ini metode tersebut akan diuji tingkat akurasi serta diimplementasikan menggunakan bahasa pemrograman PHP dan Framework Codeigniter 4 untuk merancang sistemnya.

TINJAUAN PUSTAKA

Kecerdasan Buatan

Kecerdasan buatan dapat dikatakan sebagai pengetahuan yang diadaptasikan berdasarkan kecerdasan manusia dengan cara mencangkok pengetahuan kedalam sebuah sistem. Data-data yang sudah tersedia kemudian dipelajari agar dapat dijadikan basis pengetahuan untuk memperoleh pemodelan ataupun simulasi seperti layaknya manusia berfikir. Kecerdasan buatan merupakan sebuah teknologi baru yang dikembangkan manusia dengan melakukan proses pembelajaran berbasis data-data yang banyak untuk menghasilkan pengetahuan yang baru dan berharga (Kutyauripo, Rushambwa, & Chiwazi, 2023).

Kecerdasan diperoleh dan disisipkan ke dalam komputer agar dapat berpikir dan bertindak layaknya seorang manusia. Kecerdasan buatan terdiri dari beberapa bagian seperti sistem pakar, permainan komputer, logika samar, jaringan saraf tiruan dan robotika. Secara teknis, dapat dikatakan sebagai model statistik dalam membuat keputusan dengan menggeneralisasi karakteristik suatu objek dari data yang selanjutnya diterapkan dalam alat elektronik yang berbeda (Dwilaga, 2022).

Machine Learning

Dalam pengaplikasiannya, kecerdasan buatan dibagi menjadi beberapa bagian, salah satunya *Machine Learning*. *Machine Learning* ialah cabang dari Kecerdasan Buatan dengan disiplin ilmu yang didalamnya membahas penerapan dan pengembangan algoritma dengan tujuan komputer mengadaptasi perilaku sesuai dengan data empiris (Kavakiotis et al., 2017).

Machine Learning dapat diimplementasikan pada beberapa teknologi seperti:

1. Deteksi Wajah: Sebuah sistem dapat mengenali wajah seseorang sehingga dapat merefleksikan tentang identitas diri
2. Penyaringan Spam: sistem mampu melakukan penyaringan pesan spam.
3. Segmentasi Pelanggan: sistem dapat melakukan peramalan pola perilaku konsumen agar produsen dapat mengatur biaya tertentu misalnya bidang promosi.

Untuk mempelajari Machine Learning ini ada beberapa teknik yang harus diketahui, secara umum ada 2 teknik dasar, yaitu; Supervised dan unsupervised. Pembelajaran terawasi/terarah artinya pada pembelajaran ini ada acuan yang digunakan sebagai pengarah untuk mengarahkan sesuatu hal misalnya melakukan klasifikasi (classification). Jika terdapat dua kelas dikatakan sebagai binary classification, tetapi jika lebih dari dua kelas, dikatakan multi-class classification. Jenis klasifikasi lainnya yaitu multi-label classification, untuk mengklasifikasi sampel kedalam suatu himpunan kelas. Pada multi-class classification, suatu sampel hanya bisa berkorespondensi dengan satu kelas. Sedangkan multi-label classification, satu sampel dapat berkorespondensi dengan lebih dari satu kelas. Salah satu metode Machine Learning dengan pembelajaran terawasi adalah K-Nearest Neighbour (Chen, Zhou, Li, Zhang, & Huo, 2020).

Sedangkan pembelajaran tak terarah tidak memerlukan target keluaran. Pada metode ini, hasil tidak bisa ditentukan selama prosesnya. Nilai bobot disusun pada rentang tertentu tergantung nilai inputan yang diberikan. Pembelajaran ini berguna untuk mengelompokkan unit yang mirip dalam suatu area tertentu seperti pengelompokan pola.

K-Nearest Neighbor

K-Nearest Neighbor (KNN) biasanya digunakan untuk klasifikasi objek melalui data pembelajaran yang memiliki jarak terdekat dengan tetangganya. Dekat atau jauhnya jarak dengan tetangganya dapat dihitung berdasarkan jarak Euclidean (Deng, Wang, Guan, Li, & Wang, 2023). Tahapan metode terdapat dua fase, yaitu pembelajaran (training) dan klasifikasi atau pengujian (testing). Pada fase pertama, algoritma ini menyimpan vektor fitur dan klasifikasi dari data training. Di fase kedua, fitur-fitur yang sama dihitung untuk menguji data yang klasifikasinya belum diketahui. Jarak dari vektor yang baru ini ke seluruh vektor data pembelajaran dihitung, dan sejumlah k neighbor yang paling dekat diambil (Ma, Han, Qin, & Wang, 2023). Perhitungan jarak

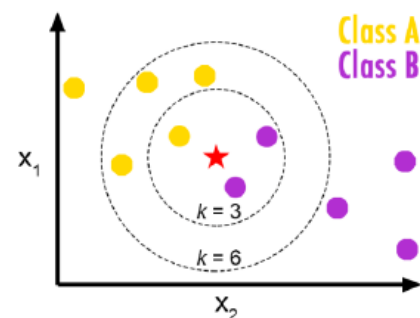
ketetangga menggunakan algoritma eucliden seperti yang ditunjukkan pada formula berikut:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Dimana:

- p, q = dua titik di ruang-n Euclidean,
- qi, pi = vektor Euclidean, dimulai dari titik awal
- n = ruang-n, untuk atribut dengan nilai kategori.

Sebuah titik akan diketahui kelasnya melalui klasifikasi terbanyak dari neighbor di sekitarnya, ilustrasinya dapat dilihat pada Gambar 1.



Gambar 1. Ilustrasi Nilai k pada KNN

Nilai k yang terbaik untuk KNN tergantung pada data. Secara general, nilai k tinggi dapat mengurangi efek noise di klasifikasi, namun membentuk batasan dengan setiap klasifikasi menjadi lebih kabur. Untuk mendapatkan nilai k yang baik dapat dipilih dengan optimasi parameter (Ajijah & Kurniawan, 2023).

Framework Codeigniter

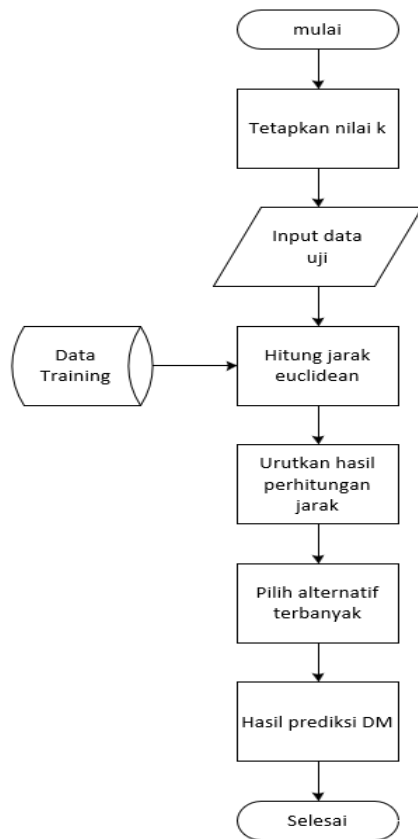
Codeigniter adalah framework PHP yang digunakan dalam merancang sistem berbasis web agar lebih dinamis (Simanullang & Silalahi, 2022). Manfaat penggunaan Framework Codeigniter versi 4 adalah:

1. Lebih Cepat Dan Efisien,
2. Memudahkan kolaborasi,
3. Ringan dan hemat sumber daya,
4. Keamanan baik,
5. Terstruktur.

METODE PENELITIAN

Dalam penelitian ini, tahap penelitian merupakan prosedur dalam pencapaian penelitian melalui empat tahapan yaitu pengumpulan data perancangan sistem, prediksi dengan metode KNN dan

pengujian akurasi. Tahapan metode KNN untuk prediksi dapat dilihat pada Gambar 2.



Gambar 2. Tahapan metode KNN

Pengumpulan data yang dijadikan sebagai objek penelitian ini diperoleh dari dataset Keagle. Objek penelitian didapatkan sebanyak 768 data dengan delapan parameter yang sudah lengkap dengan nilai parameternya. Parameter prediksi DM dilihat pada Tabel 1.

Tabel 1. Parameter prediksi DM

No	Parameter	Satuan
1	Kadar Glukosa	Mg/dl
2	Tekanan Darah	mmhg
3	Ketebalan kulit	mm
4	Hormon Insulin	mu U/ml
5	Massa Indeks Tubuh (BMI)	Kg/m ²
6	Umur	Tahun
7	Angka kehamilan	Kali
8	Jumlah riwayat penderita diabetes dalam keluarga	-

Tahapan pengujian dilakukan dengan metode confusion matrix dengan menganalisa tingkat precision, recall dan accuracy dapat dilihat pada persamaan berikut.

$$\text{Precision} = \frac{TP}{TP+FP} \dots\dots\dots (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \dots\dots\dots (2)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (3)$$

Keterangan:

TP = True Positive (Jika hasil prediksi positif dan data sebenarnya positif)

TN = True Negative (Jika hasil prediksi negatif dan data sebenarnya negatif)

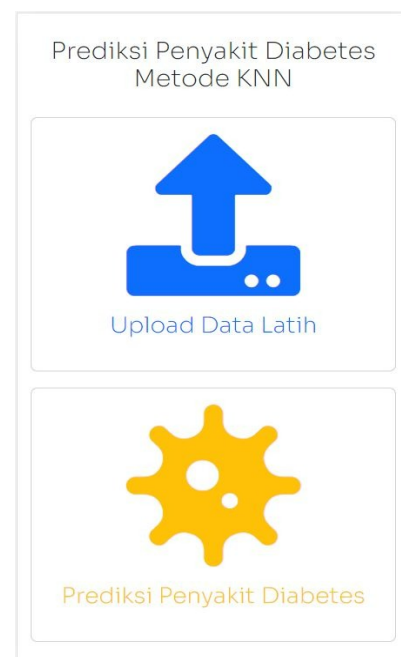
FN = False Negative (Jika hasil prediksi negatif dan data sebenarnya positif)

FP = False Positive (Jika hasil prediksi positif dan data sebenarnya negatif)

Precision ialah tingkat ketelitian metode, recall ialah tingkat ketepatan metode dan accuracy menyatakan tingkat kesesuaian atau dekatnya suatu hasil pengukuran terhadap kondisi yang sebenarnya.

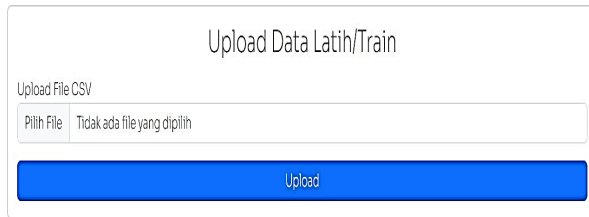
HASIL DAN PEMBAHASAN

Penelitian ini diimplementasikan dalam sebuah sistem dengan bahasa pemrograman PHP dan framework Codeigniter 4. Halaman awal sistem ditampilkan pada Gambar 3.



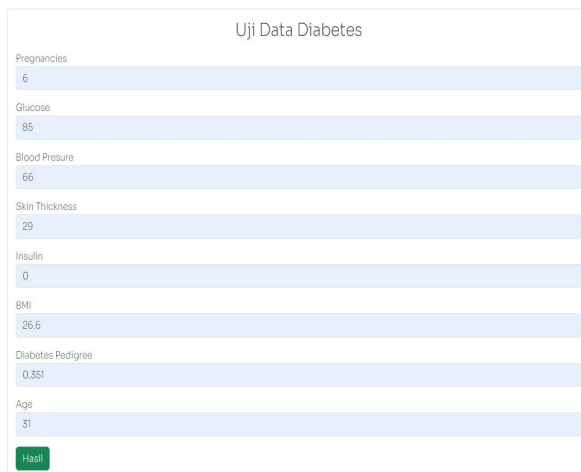
Gambar 3. Halaman awal sistem prediksi

Pada tampilan diatas disajikan menu untuk upload data latih dan prediksi DM. Menu unggah Data Latih ditampilkan pada Gambar 4.



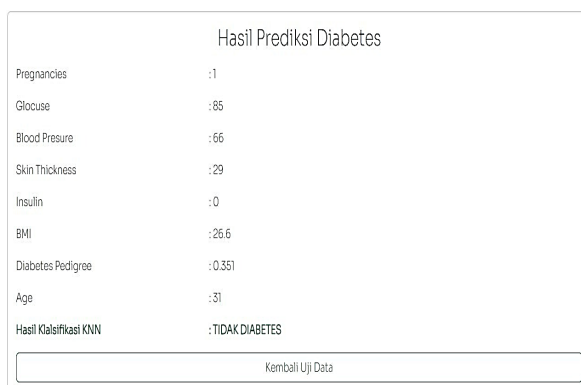
Gambar 4. Menu Unggah Data Latih/Train

Pada tampilan gambar tersebut disajikan halaman untuk upload file data training dengan format csv yang sudah berisi dataset DM. Setelah tahapan tersebut lalu dilakukan tahapan prediksi seperti pada Gambar 5.



Gambar 5. Halaman Uji Data Diabetes

Setelah mengisi nilai setiap parameter, lalu dilakukan pengujian data baru tersebut data training yang sudah dijadikan pengetahuan di sistem. Sehingga dihasilkan prediksi DM seperti Gambar 6.



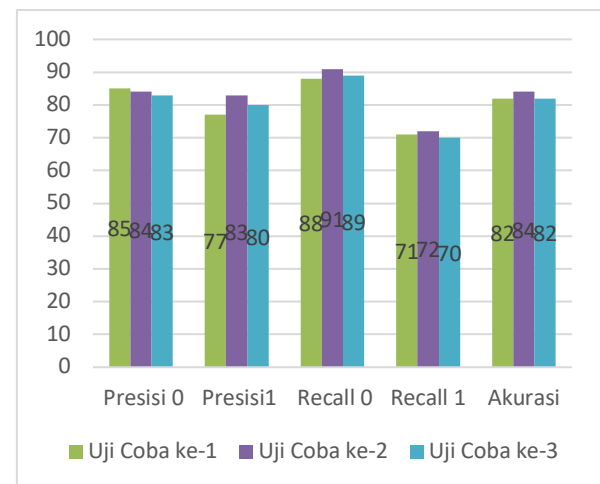
Gambar 6. Hasil Prediksi Diabetes

Uji coba metode KNN menggunakan metode Confusion matrix untuk mendapatkan nilai precision, recall dan accuracy dengan persamaan (1), (2), (3). Skenario pengujian dilakukan dengan tiga pengujian, Uji coba ke-1 data uji = 150, uji coba ke-2 data uji = 200, uji data ke-3 = 300. Nilai yang digunakan = 2 karena pada penelitian ini output yang akan dihasilkan DM Positif diberi kode =1 dan DM Negatif diberi kode = 0. Dengan menggunakan dataset sebanyak 768 dan nilai k=2, maka diuji nilai precision DM Positif (Precision 1), precision DM Negatif (Precision 0), Recall DM Positif (Recall 1), Recall DM Negatif (Recall 0) dan Akurasi. Hasil pengujian disajikan dalam satuan persentase pada Tabel 2.

Tabel 2. Hasil Pengujian Data

Training	Nilai K	Data uji	Precision		Recall		Akurasi
			0	1	0	1	
768	2	150	85	77	88	71	82
768	2	200	84	83	91	72	84
768	2	300	83	80	89	70	82

Hasil dari Uji coba metode KNN disajikan dalam gambar 7.



Gambar 7. Hasil Pengujian KNN

KESIMPULAN

Dari pengujian metode yang diteliti menggunakan dataset DM disimpulkan bahwa metode tersebut dapat di implementasikan kedalam sebuah sistem dengan bahasa pemrograman PHP dan framework CodeIgniter 4. Selain itu metode tersebut berjalan dengan baik karena berdasarkan pengujian pertama dengan 150 data uji didapatkan nilai presisi DM Positif sebanyak 77%, nilai presisi DM Negatif sebanyak 85%, nilai recall DM Positif 71%, nilai recall DM Negatif 88%, Akurasi 82%. Pada pengujian kedua dengan 200 data didapatkan nilai presisi DM Positif

sebanyak 83%, nilai presisi DM Negatif sebanyak 84%, nilai recall DM Positif 72%, nilai recall DM Negatif 91%, Akurasi 84%. Pada pengujian ketiga dengan 300 data didapatkan nilai presisi DM Positif sebanyak 80%, nilai presisi DM Negatif sebanyak 83%, nilai recall DM Positif 70%, nilai recall DM Negatif 89%, Akurasi 82%.

DAFTAR PUSTAKA

- Adam, S., McIntyre, H. D., Tsoi, K. Y., Kapur, A., Ma, R. C., Dias, S., ... McAuliffe, F. M. (2023). Pregnancy as an opportunity to prevent type 2 diabetes mellitus: FIGO Best Practice Advice. *International Journal of Gynecology and Obstetrics*, 160(S1), 56–67. <https://doi.org/10.1002/ijgo.14537>
- Ajjjah, N., & Kurniawan, A. (2023). Klasifikasi Teks Mining Terhadap Analisa Isu Kegiatan Tenaga Lapangan Menggunakan Algoritma K-Nearest Neighbor (KNN). *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 7(1), 254–262.
- Andriana, H., & Hilab, S. S. (2023). Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter terhadap KTT G20 di Indonesia. *JURIKOM (Jurnal Riset Komputer)*, 10(1), 60–67. <https://doi.org/10.30865/jurikom.v10i1.5427>
- Aulia, M. S., Abdurrahman, M., & Putrada, A. G. (2019). Pendeteksian Kadar Glukosa dalam Darah pada Gejala Diabetes Tipe 1 Menggunakan Algoritma K-Nearest Neighbor dengan Metode Nafas. *SMARTICS Journal*, 5(1), 14–21. <https://doi.org/10.21067/smartics.v5i1.3287>
- Chen, Z., Zhou, L. J., Li, X. Da, Zhang, J. N., & Huo, W. J. (2020). The Lao text classification method based on KNN. *Procedia Computer Science*, 166, 523–528. <https://doi.org/10.1016/j.procs.2020.02.053>
- Deng, S., Wang, L., Guan, S., Li, M., & Wang, L. (2023). Non-parametric Nearest Neighbor Classification Based on Global Variance Difference. *International Journal of Computational Intelligence Systems*, 16(1). <https://doi.org/10.1007/s44196-023-00200-1>
- Dwilaga, A. T. (2022). Implementasi Model Artificial Intelligence Dalam Warehouse: Systematic Literature Review. *JUSTI (Jurnal Sistem Dan Teknik Industri)*, 3(2), 253–261.
- Karmana, W. (2023). Penerapan Pola Hidup Sehat Untuk Mencegah Diabetes Mellitus. *Panthera : Jurnal Ilmiah Pendidikan Sains Dan Terapan*, 3(1), 1–6.
- Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology Journal*, 15, 104–116. <https://doi.org/10.1016/j.csbj.2016.12.005>
- Kutyaauripo, I., Rushambwa, M., & Chiwazi, L. (2023). Artificial intelligence applications in the agrifood sectors. *Journal of Agriculture and Food Research*, 11(January), 100502. <https://doi.org/10.1016/j.jafr.2023.100502>
- Ma, X., Han, Y., Qin, H., & Wang, P. (2023). KNN Data Filling Algorithm for Incomplete Interval-Valued Fuzzy Soft Sets. *International Journal of Computational Intelligence Systems*, 16(1). <https://doi.org/10.1007/s44196-023-00190-0>
- Makinde, A. S., Vincent, O. R., Akinwale, A. T., Oguntuase, A., & Acheme, I. D. (2020). An Improved Customer Relationship Management Model for Business-to-Business E-commerce Using Genetic-Based Data Mining Process. *2020 International Conference in Mathematics, Computer Engineering and Computer Science, ICMCECS 2020*. <https://doi.org/10.1109/ICMCECS47690.2020.240875>
- Maliha, D. Z., Santoso, E., & Furqon, M. T. (2019). Penerapan Metode Neighbor Weighted K-Nearest Neighbor Dalam Klasifikasi Diabetes Mellitus. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2910–2915.
- Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2021). Using data mining techniques to explore security issues in smart living environments in Twitter. *Computer Communications*, 179(January), 285–295. <https://doi.org/10.1016/j.comcom.2021.08.021>
- Simanullang, H. G., & Silalahi, A. P. (2022). *Pemrograman Web Menggunakan Framework Codeigniter 4* (1st ed.). Malang: Madza Media.
- Simanullang, H. G., Silalahi, A. P., & Sartika, D. (2022). Prediksi Jumlah Pasien Covid-19 Di Indonesia Menggunakan Least Square Method Berbasis Android. *INFORMATIKA*, 14(1), 86–93.
- Syafar, A. M., Baharuddin, M. S. I. F., Zainal, M., & Fuadi, N. (2021). Peramalan Penderita Diabetes Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Sintaks Logika*, 1(2), 113–115.