



## KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE *K-NEAREST NEIGHBORS* (KNN)

Luluk Nuril Mukarromah <sup>a\*</sup>, Zaehol Fatah <sup>b</sup>, Irma Yunita <sup>c</sup>

<sup>a</sup> Fakultas Sains & Teknologi, [luluknuril1306@gmail.com](mailto:luluknuril1306@gmail.com), Universitas Ibrahimy Situbondo, Jawa Timur

<sup>b</sup> Fakultas Sains & Teknologi, [zaeholfatah@gmail.com](mailto:zaeholfatah@gmail.com), Universitas Ibrahimy Situbondo Jawa Timur

<sup>c</sup> Fakultas Sains & Teknologi, [irmayunitasaid@gmail.com](mailto:irmayunitasaid@gmail.com), Universitas Ibrahimy Jawa Timur

\* Korespondensi

### ABSTRACT

*Diabetes is a chronic disease caused by impaired insulin production, which causes an increase in blood sugar levels and has the potential to cause serious complications. Early detection of this disease is very important to prevent the risk of complications in patients. This research aims to implement a data mining method with the K-Nearest Neighbors (KNN) algorithm in the classification of diabetes, using attributes such as blood pressure, age, obesity and family history as variables. The KNN method is used to identify patterns in data that are relevant to potential diabetes, with stages of model learning and performance evaluation. The analysis results show that the KNN algorithm is able to classify data with a fairly good level of accuracy, showing its effectiveness in detecting possible diabetes in patients. The implementation of this algorithm shows potential as a supporting tool in the early diagnosis of diabetes.*

**Keywords:** *Diabetes, data mining, KNN*

### ABSTRAK

Diabetes adalah penyakit kronis yang disebabkan oleh gangguan produksi insulin, yang menyebabkan peningkatan kadar gula darah dan berpotensi menimbulkan komplikasi serius. Deteksi dini penyakit ini sangat penting untuk mencegah risiko komplikasi pada pasien. Penelitian ini bertujuan untuk mengimplementasikan metode *data mining* dengan algoritma *K-Nearest Neighbors* (KNN) dalam klasifikasi penyakit diabetes, menggunakan atribut-atribut seperti tekanan darah, usia, obesitas, dan riwayat keluarga sebagai variabel. Metode KNN digunakan untuk mengidentifikasi pola dalam data yang relevan terhadap potensi diabetes, dengan tahapan pembelajaran model dan evaluasi kinerja. Hasil analisis menunjukkan bahwa algoritma KNN mampu mengklasifikasikan data dengan tingkat akurasi yang cukup baik, menunjukkan efektivitasnya dalam mendeteksi kemungkinan diabetes pada pasien. Implementasi algoritma ini menunjukkan potensi sebagai alat pendukung dalam diagnosis dini terhadap penyakit diabetes.

**Kata Kunci:** *penyakit diabetes, data mining, KNN*

### 1. PENDAHULUAN

Karena pankreas tidak mampu menghasilkan insulin secara optimal, diabetes adalah penyakit yang tidak menular dan termasuk cukup serius bagi manusia. Dibandingkan dengan individu yang tidak memiliki riwayat keluarga diabetes, individu dengan tekanan darah tinggi, obesitas, umur, gaya hidup, makanan yang tidak sehat, dan makanan yang tidak sehat adalah beberapa faktor risiko yang dapat menyebabkan diabetes. Untuk menangani masalah ini, diabetes harus dideteksi sejak dini. Diharapkan bahwa diagnosis dini dapat menurunkan risiko komplikasi diabetes. Untuk mencegah diabetes, pasien diperiksa sejak dini. Diabetes adalah jenis penyakit yang diturunkan. (Prasetya & Sujatmiko, 2022)

Diabetes adalah penyakit yang ditandai dengan kenaikan kadar gula dalam darah. Penyakit ini dapat disebabkan oleh keturunan atau dapat disebabkan oleh infeksi dan racun makanan. Penyakit ini

mengganggu mata, ginjal, syaraf, dan pembuluh darah. Suatu model dibuat dengan algoritma klasifikasi menggunakan dataset pelatihan. Model ini kemudian digunakan untuk meramalkan label kelas untuk data baru yang belum diketahui. Dalam pengolahan data, klasifikasi adalah fungsi yang membuat model untuk memprediksi kelas atau kategori dari objek yang ada dalam basis data. Proses klasifikasi terdiri dari dua tahap: pembelajaran dan pengklasifikasian. Tahap kedua melibatkan pengelompokan data, di mana data diatur dan dikelompokkan ke dalam berbagai jenis, bentuk, atau kelas. Proses ini memungkinkan pemisahan dan kategorisasi data sesuai dengan kebutuhan untuk mengumpulkan data untuk berbagai tujuan. (Dewi Nasien et al., 2024)

Metode data mining memungkinkan pengguna mengakses jumlah data yang besar dalam waktu yang cepat. Dengan kata lain, data mining adalah alat dan aplikasi yang menggunakan analisis statistik untuk ekstraksi atau panggilan data dan informasi yang belum diketahui sebelumnya. Secara sederhana, data mining adalah proses panggilan data yang menghasilkan penemuan informasi terbaru dengan mencari pola atau aturan tertentu dalam sejumlah besar data. Oleh karena itu, cara kerja data mining adalah untuk memeriksa. (Ardilla et al., 2021)

Banyak sekali algoritma yang dapat digunakan dalam proses data mining, oleh karena itu, dalam penelitian ini penerapan data mining dilakukan dengan bantuan pemodelan pada machine learning yang menggunakan algoritma yaitu K-Nearest Neighbor. algoritma tersebut juga turut membantu untuk menganalisis serta mengekstraksi data menjadi pengetahuan baru yang berguna. diantara banyaknya algoritma yang sering digunakan dan populer digunakan oleh peneliti di bidang ilmu data mining adalah algoritma K-Nearest Neighbors. algoritma ini terbilang sederhana serta bisa terbilang efektif untuk data yang ukurannya besar, dan digunakan untuk mengklasifikasi data dengan menentukan kedekatan objek yang baru atau tertangga terdekat. tetangga terdekat biasa disimbolkan dengan (K) yang berpengaruh terhadap proses pengambilan keputusan dari metode KNN. Metode ini merupakan bagian dari algoritma supervised learning, yang berarti dataset diharuskan memiliki target, pada penelitian kali ini penentuan nilai K diukur berdasarkan perhitungan dengan rumus Euclidean Distance (Ningsih & Zakaria, 2023). Penelitian ini bertujuan untuk mengimplementasikan metode K-NN dalam Klasifikasi penyakit diabetes dan mengetahui tingkat akurasi dari metode tersebut. (Yogianto et al., 2024)

## **2. METODE PENELITIAN**

### **2.1 Metode Pengambilan Data**

Pada bagian ini merupakan suatu tata cara yang dilakukan untuk dapat mengumpulkan data terkait dengan penelitian yang sedang dilakukan, metode pengumpulan data ini merupakan suatu cara independen untuk melakukan analisis data atau bisa dijadikan alat utama dalam menganalisis data. Untuk mendapatkan data-data yang berkaitan dengan penyakit diabetes, penulis mengumpulkan data yang terkait dengan penelitian ini dengan cara Study Literature dengan jenis data sekunder yang didapatkan dari Website Kaggle dengan judul Klasifikasi Penyakit Diabetes menggunakan metode KNN. Metode yang dipakai oleh peneliti adalah metode kuantitatif karena data yang di ambil bersifat numerik atau statistic atau metode penelitian kuantitatif. Oleh karena itu metode kuantitatif diartikan sebagai suatu proses pengumpulan dan analisis data yang dilakukan secara sistematis dan logis untuk mencapai tujuan tertentu. Pengumpulan dan analisis data menggunakan metode-metode ilmiah, baik yang bersifat kuantitatif dan kualitatif, eksperimental atau non eksperimental, interaktif atau non interaktif. Metodemetode tersebut telah dikembangkan secara intensif, melalui berbagai uji coba sehingga telah memiliki prosedur yang baku. Penelitian kuantitatif merupakan penelitian ilmiah yang Sistematis terhadap bagian-bagian dan fenomena serta Kausalitas hubungan-hubungannya. Penelitian kuantitatif Didefinisikan sebagai investigasi sistematis terhadap Fenomena dengan mengumpulkan data yang dapat Diukur dengan melakukan teknik statistik, matematika Atau komputasi. Penelitian kuantitatif sebagian besar dilakukan dengan Menggunakan metode statistik yang digunakan untuk Mengumpulkan data kuantitatif dari studi penelitian. (Rustamana et al., 2024)

### **2.2 Klasifikasi**

Klasifikasi merupakan teknik untuk mencari model atau fungsi yang membedakan atau menjelaskan konsep atau kelas data tertentu. Tujuannya adalah untuk memprediksikan kelas dari objek yang tidak diketahui berdasarkan pola atau karakteristik yang telah dianalisis dalam data tersebut. Proses ini melibatkan pembuatan pengklasifikasi menggunakan data latih yang sudah diberi label kelas, dan mengevaluasi kinerjanya dengan menggunakan metrik akurasi untuk menilai kemampuan pengklasifikasi dalam mengidentifikasi kelas objek yang tidak diketahui. Proses klasifikasi melibatkan pelatihan pada fungsi target yang menghubungkan setiap kelompok atribut dengan label kelas yang tersedia. Model yang

dihasilkan digunakan untuk memperkirakan label kelas pada data baru yang tidak diketahui. (Marwah et al., 2024) (Sholikhul Fiqri & Dwi Bhakti, 2024)

Klasifikasi juga bisa disebut proses menilai objek data untuk memasukkannya ke dalam salah satu kelas yang tersedia. Klasifikasi membangun model berdasarkan data latih, yang kemudian digunakan untuk mengklasifikasikan data baru. Klasifikasi dapat didefinisikan sebagai proses pembelajaran atau pelatihan terhadap fungsi target yang memetakan setiap kumpulan atribut (fitur) ke salah satu label kelas yang ada. Sistem klasifikasi diharapkan dapat mengklasifikasikan semua data dengan akurat, namun, kinerjanya tidak selalu 100% tepat. Oleh karena itu, performa sistem klasifikasi perlu diukur. Biasanya, pengukuran kinerja klasifikasi dilakukan menggunakan matriks konfusi. (Arrohman & Fatah, 2024)

### 2.3 Diabetes

Diabetes adalah kondisi kekurangan hormon insulin yang menyebabkan kadar gula darah (glukosa) tinggi secara kronis. Insulin adalah hormon yang membantu penyerapan glukosa dalam sel-sel tubuh untuk diubah menjadi energi. Insulin juga membantu menyimpan sebagian glukosa sebagai cadangan energi. Kondisi ini juga dikenal sebagai penyakit gula atau kencing manis. Gula dalam darah seharusnya diserap oleh sel-sel tubuh untuk diubah menjadi energi. Diabetes adalah kondisi kekurangan hormon insulin yang menyebabkan kadar gula darah (glukosa) tinggi secara kronis. Insulin adalah hormon yang membantu penyerapan glukosa dalam sel-sel tubuh untuk diubah menjadi energi. Insulin juga membantu menyimpan sebagian glukosa sebagai cadangan energi. Kondisi ini juga dikenal sebagai penyakit gula atau kencing manis. Gula dalam darah seharusnya diserap oleh sel-sel tubuh untuk diubah menjadi energi. (Aswin Ardiansyah et al., 2023)

### 2.4 RapidMiner

Rapid Miner adalah suatu Analisis teks yang bekerja untuk memproses data mining, dengan melibatkan penggalian pola dari kumpulan data besar dan menggabungkannya dengan Teknik statistik, kecerdasan buatan, dan basis data. Tujuan dari analisis teks ini adalah untuk menghasilkan informasi yang akurat dan bermakna dari teks yang sangat besar atau kompleks. (Wasik et al., 2024)

Oleh karena itu, Rapid Miner juga bisa disebut alat penambangan data yang dikembangkan menggunakan java sebagai bahasa pemrograman. Bahasa java dianggap sebagai salah satu bahasa yang memiliki berbagai keunggulan, seperti kesederhanaan, keamanan, kekuatan, dampak, kemampuan berorientasi objek tingkat tinggi dan banyak keunggulan lainnya. Rapid Miner dikembangkan untuk tugas penambangan data umum. (Adab, n.d.)

### 2.5 K-Nearest Neighbors (KNN)

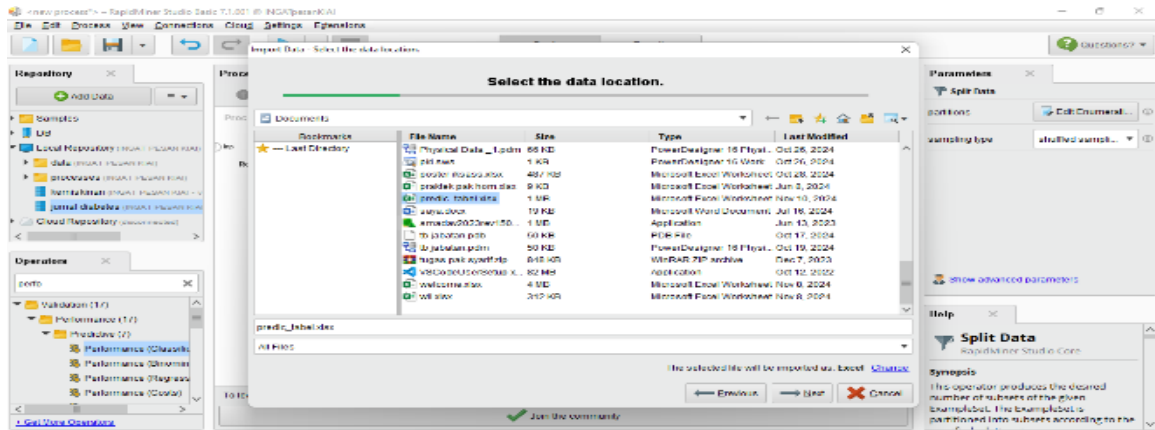
K-Nearest Neighbor (KNN) termasuk kelompok instance-based learning. Algoritma ini juga merupakan salah satu Teknik lazy learning. KNN dilakukan dengan mencari kelompok ke objek dalam data mining training yang paling dekat dengan objek pada data baru atau data testing. Diperlukan suatu sistem klasifikasi sebagai sebuah sistem yang mampu mencari informasi. Contoh kasus, misal diinginkan untuk mencari Solusi terhadap masalah seorang pasien baru dengan menggunakan Solusi dari pasien lama.

Algoritma K-Nearest Neighbor (K-NN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-NN. (Admojo, 2020)

## 3. HASIL DAN PEMBAHASAN

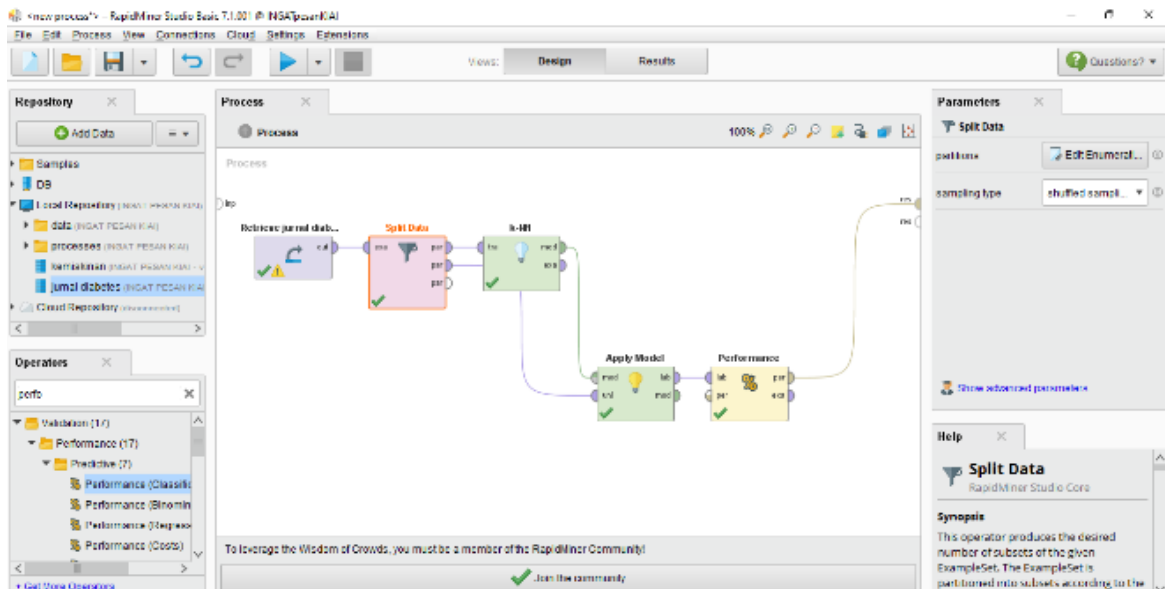
### 3.1 Penerapan Data Mining Menggunakan Rapidminer

Tahapan ini diawali dengan pemilihan data lalu penentuan label kelas seperti dibawah ini:



Gambar 1. Penentuan Kelas Label

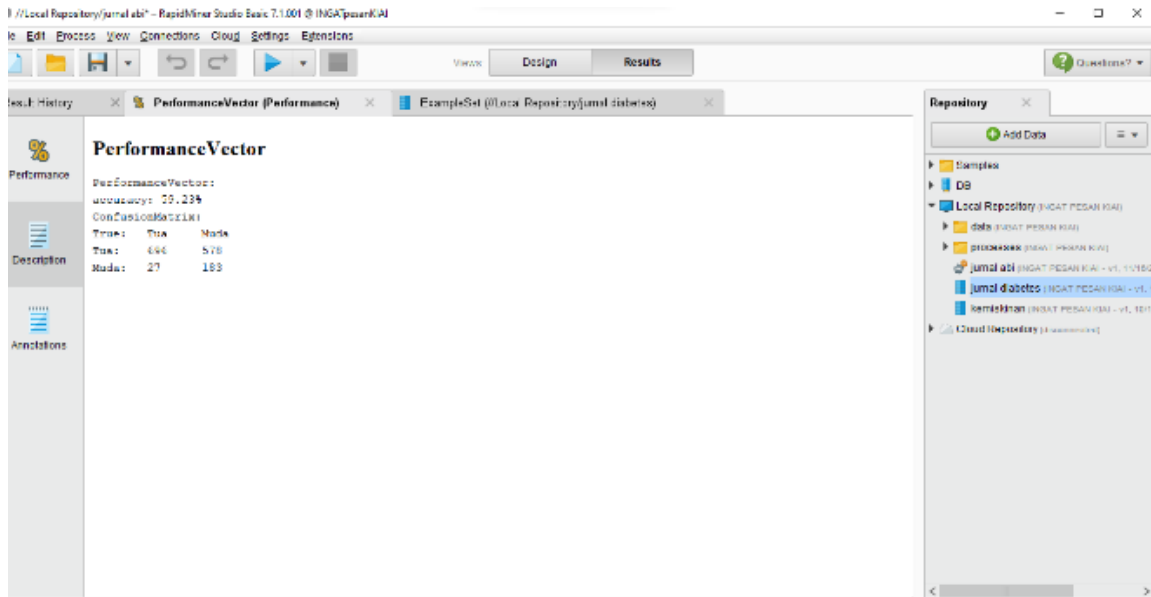
Model yang digunakan sesuai dengan proses algoritma K-NN yang terdiri dari data operator retrieve data set, operator split data sebagai tahapan transformasi. Operator k-NN sebagai bentuk dari operator itu sendiri guna untuk menggabungkan antara data split dengan k-NN agar menjadi kesatuan model dan operator performance untuk mengukur akurasi kinerja dari model k-NN yang telah terbentuk digunakan operator apply model. Setelah model k-NN diterapkan pada data testing.



Gambar 2. Proses Perhitungan Menggunakan Tools Rapidminer Metode K-NN

### 3.2 Evaluasi

Tahapan selanjutnya ini adalah tahapan evaluasi untuk mengevaluasi hasil dari algoritma K-Nearst untuk mengetahui kesimpulan dari proses analisis data mining.

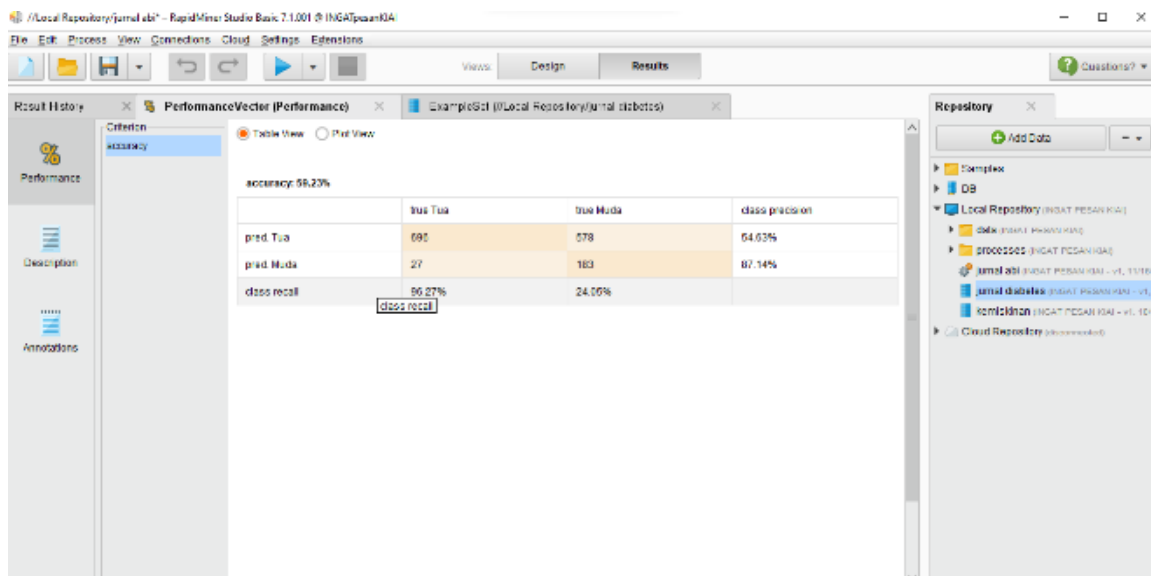


**Gambar 3. Performance Vektor**

Pada gambar di atas adalah hasil evaluasi dari Performance Vektor memberikan Gambaran tentang kinerja model K-Nearest Neighbor yang telah ditetapkan dalam proses analisis data.

### 3.3 Klasifikasi

Tahapan selanjutnya ini adalah evaluasi untuk mengevaluasi hasil dari algoritma K-Nearest untuk mengetahui kesimpulan dari proses analisis data mining.



**Gambar 4. Hasil Klasifikasi**

Kode pemrograman ditempatkan segera setelah itu disebut dalam tubuh teks. Judul untuk konten program kode ditulis di bawahnya dan ada satu baris spasi tunggal kosong antara itu dan program kode. Keterangannya ditulis dalam ukuran font 8, dan ditempatkan seperti pada contoh. Antara program kode dengan badan teks ada satu baris spasi tunggal kosong.

## 4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa algoritma K-Nearest Neighbor efektif dalam mengklasifikasi penyakit diabetes dengan menggunakan metode ini menghasilkan hasil hasil klasifikasi yang akurat, mencapai Tingkat akurasi sebesar 59,23% dari data yang telah ada. Hasil klasifikasi ini memiliki dampak positif dalam mengklasifikasi penyakit diabetes sebagai saran, penelitian ini dapat

dikembangkan lagi dengan eksplorasi algoritma lain atau penggabungan metode data mining untuk hasil yang lebih baik lagi.

#### DAFTAR PUSTAKA

- [1] Adab, R. S. S. M. F. S. M. M. S. I. P. (n.d.). *IMPLEMENTASI DATA MINING (Clustering, Association, Prediction, Estimation, Classification)*. Penerbit Adab. <https://books.google.co.id/books?id=LsOqEAAAQBAJ>
- [2] Admojo, F. T. (2020). *Klasifikasi Aroma Alkohol Menggunakan Metode KNN*. 1(2), 34–38.
- [3] Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Manuhutu, M. A., Ridwan, M., & Wardhani, A. K. (2021). *DATA MINING DAN APLIKASINYA*. Penerbit Widina. <https://books.google.co.id/books?id=53FXEAAAQBAJ>
- [4] Arrohman, S., & Fatah, Z. (2024). *Gudang Jurnal Multidisiplin Ilmu Prediksi Diabetes Menggunakan Algoritma Klasifikasi K-Nearest Neighbors ( K-NN ) pada Perempuan Indian Pima*. 2, 220–226.
- [5] Aswin Ardiansyah, Enos C.O.Telaumbanua, Aron S. Gultom, & Angelita A. S. M. Limbong. (2023). *Klasifikasi Penyakit Diabetes Menggunakan Metode SVM Dan KNN*. *Jurnal Penelitian Rumpun Ilmu Teknik*, 3(1), 77–83. <https://doi.org/10.55606/juprit.v3i1.13151>
- [6] Dewi Nasien, Darwin, R., Cia, A., Leo Winata, A., Go, J., M.C, R., Charles Wijaya, R., & Charles Lo, K. (2024). *Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes*. *JEKIN - Jurnal Teknik Informatika*, 4(1), 10–17. <https://doi.org/10.58794/jekin.v4i1.640>
- [7] Marwah, S., Astuti, R., & M. Basysyar, F. (2024). *Implementasi Data Mining Menggunakan Algoritma Naive Bayes Untuk Diagnosis Penyakit Kulit Scabies Pada Hewan*. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3892–3897. <https://doi.org/10.36040/jati.v7i6.8276>
- [8] Prasetya, W. D., & Sujatmiko, B. (2022). *Rancang Bangun Aplikasi dengan Perbandingan Metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam Klasifikasi Penderita Penyakit Diabetes*. *Journal of Informatics and Computer Science (JINACS)*, 3(04), 515–525. <https://doi.org/10.26740/jinacs.v3n04.p515-525>
- [9] Rustamana, A., Wahyuningsih, P., Azka, M. F., & Wahyu, P. (2024). *Penelitian Metode Kuantitatif*. *Sindoro Cendikia Pendidikan*, 5(6), 1–10.
- [10] Sholikhul Fiqri, M., & Dwi Bhakti, H. (2024). *Klasifikasi Potensi Penyakit Diabetes Mellitus Tipe Ii Pada Pasien Menggunakan Algoritme Knn (K-Nearest Neighbor)*. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(4), 7305–7313. <https://doi.org/10.36040/jati.v8i4.10133>
- [11] Wasik, A., Fatah, Z., Munazilin, A., Studi, P., Informasi, S., Situbondo, U. I., Studi, P., Komputer, I., & Situbondo, U. I. (2024). *Implementasi data mining untuk memprediksi penjualan accessories handphone dan handphone terlaris menggunakan metode k-nearest neighbor (k-nn) I*. 1(2), 469–479.
- [12] Yogiarto, A., Homaidi, A., & Fatah, Z. (2024). *Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung*. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1720–1728. <https://doi.org/10.33379/gtech.v8i3.4495>