



Prediction of Type 2 Diabetes Mellitus Using The K-Nearest Neighbor (K-NN) Algorithm

Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor (K-NN)

Agfa Oktaviana^{1*}, Dhina Puspasari Wijaya², Andri Pramuntadi³, Dadang Heksaputra⁴

^{1,2,3}Department of Informatics, Faculty of Computers and Engineering,
Universitas Alma Ata, Yogyakarta, Indonesia

⁴Department of Information Systems, Faculty of Computers and Engineering,
Universitas Alma Ata, Yogyakarta, Indonesia

E-Mail: ¹193200037@almaata.ac.id, ²dhina.puspa@almaata.ac.id,
³andripramuntadi@almaata.ac.id, ⁴dadang@almaata.ac.id

Received Feb 2nd 2024; Revised Mar 28th 2024; Accepted May 06th 2024
Corresponding Author: Dhina Puspasari Wijaya

Abstract

Diabetes Mellitus (DM) is one of the Non-Communicable Diseases (NCDs) known for high blood sugar levels. The International Diabetes Federation (IDF) predicts that in 2045, DM will increase to 629 million people. In today's modern era, lifestyle patterns are something that must be considered because they go hand in hand with the development of technology to be easy and fast. Persistent poor lifestyle patterns can potentially lead to Type 2 DM disease. Various efforts have been made to reduce the growth rate of this disease, one of which is conducting research to make a prediction of a person using various methods such as the K-NN classification method. The purpose of this research is to implement and build a model. The dataset used comes from the Mlati II Health Center, Mlati District, Sleman Regency, Yogyakarta Special Region Province. In building a prediction model, the dataset is preprocessed using MinMax Normalization, the division of the train set and test set using Stratified 5-fold CV. The parameters used from K-NN are manhattan distance and the value of n_neighbors = 13. Using the evaluation of accuracy, precision, recall, and f1-score, the results are 88%, 83%, 87%, and 85%, respectively.

Keyword: K-NN, Manhattan Distance, MinMax Normalization, Stratified 5 Fold CV, Type 2 Diabetes Mellitus

Abstrak

Diabetes Melitus (DM) merupakan salah satu Penyakit Tidak Menular (PTM) yang dikenal dengan tingginya kadar gula dalam darah. International Diabetes Federation (IDF) memprediksi di tahun 2045, penyakit DM akan mengalami peningkatan menjadi 629 juta penduduk. Pada era modern saat ini, pola gaya hidup menjadi hal yang harus diperhatikan karena beriringan dengan berkembangnya teknologi menjadi mudah dan cepat. Pola gaya hidup yang buruk terus-menerus dapat berpotensi untuk terkena penyakit DM Tipe 2. Berbagai upaya dilakukan untuk menekan angka pertumbuhan penyakit ini salah satunya melakukan penelitian untuk membuat prediksi terhadap seseorang menggunakan berbagai metode seperti metode klasifikasi K-NN. Tujuan dari penelitian ini adalah mengimplementasikan dan membangun sebuah permodelan. Dataset yang digunakan berasal dari Puskesmas Mlati II Kecamatan Mlati, Kabupaten Sleman, Provinsi Daerah Istimewa Yogyakarta. Dalam membangun sebuah model prediksi, dataset dipreprocessing menggunakan *MinMax Normalization*, pembagian train set dan test set menggunakan *Stratified 5-fold CV*. Adapun parameter yang digunakan dari K-NN adalah *manhattan distance* dan nilai *n_neighbors = 13*. Dengan menggunakan evaluasi akurasi, presisi, recall, dan *f1-score*, masing-masing memberikan hasil yaitu 88%, 83%, 87%, dan 85%.

Kata Kunci: Diabetes Melitus Tipe 2, K-NN, Manhattan Distance, MinMax Normalization, Stratified 5-fold CV

1. PENDAHULUAN

Diabetes Melitus (DM) merupakan penyakit kronis yang dikenal dengan tingginya kadar gula dalam darah dan menjadi salah satu Penyakit Tidak Menular (PTM) yang membahayakan pada tubuh seseorang. Dalam istilah bahasa Yunani, diabetes berarti curahan, sedangkan militus atau melitus yang artinya gula atau

madu. Secara bahasa dapat diartikan bahwa diabetes melitus adalah curahan cairan dari dalam tubuh yang mengandung banyak gula [1].

Setiap tahunnya, banyak pasien yang kehilangan nyawa karena menderita penyakit diabetes. Prediksi pada tahun 2045 juga dikatakan bahwa akan mengalami peningkatan penyakit diabetes menjadi 629 juta penduduk [2]. Pada penyakit DM Tipe 1 atau disebut dengan *Insulin-Dependent Diabetes Mellitus* (IDDM) disebabkan terjadi karena kerusakan sel beta pankreas (reaksi autoimun). Hanya sel beta pankreas yang dapat menghasilkan insulin yang berfungsi untuk mengatur kadar glukosa dalam tubuh [3].

Sedangkan penyakit DM Tipe 2 atau dikenal juga *Non-Insulin-Dependent Diabetes Mellitus* (NIDDM) penyakit yang berisiko lebih besar dialami oleh semua orang, dengan nilai prevalensi yaitu 90-95%. yang disebabkan oleh faktor risiko modifikasi seperti gaya hidup yang tidak sehat, cenderung mengonsumsi makanan cepat saji, dan mengonsumsi minuman dengan kadar gula tinggi. DM tipe 2 ini terjadi karena insulin yang dihasilkan cukup namun tidak cukup untuk menjaga kadar gula agar normal. Penderita DM tipe 2 ini biasanya memasuki usia di atas 40 tahun [4]. Berbagai upaya dilakukan untuk menekan angka pertumbuhan penyakit DM terutama pada DM tipe 2 yang dipengaruhi oleh pola gaya hidup sehari-hari. Pada bidang kesehatan, teknologi mulai berperan di dalamnya yang dikenal sebagai Kecerdasan Buatan atau *Artificial Intelligence* (AI). Profesi sebagai tenaga kesehatan memang tidak bisa digantikan oleh siapapun, namun AI diciptakan di dalamnya untuk melakukan transformasi dan meningkatkan hasil yang baik [5].

Salah satu algoritma AI yang telah banyak digunakan pada penelitian sebelumnya adalah K-Nearest Neighbor (K-NN) merupakan algoritma dari sub-bidang *machine learning* yang dapat diimplementasikan untuk memprediksi, klasifikasi, dan regresi. Algoritma K-NN merupakan metode sederhana yang dilakukan dengan cara mencari nilai *neighbor* (k) dari tetangga terdekatnya dan mencari hasil prediksi berdasarkan mayoritas dan jumlah yang ganjil [6].

Beberapa penelitian yang menggunakan *machine learning* dalam memprediksi telah dilakukan sebelumnya. Saut Dohot Siregar, dkk (2023) dengan implementasi algoritma K-NN untuk memprediksi penyakit DM menggunakan dataset *Diabetes Foot Ulcer* (DFU) mendapat akurasi 94% [7]. Jubeda Jamal Khanam dan Simon Y. Foo (2021), penelitian menggunakan *Pima Indiana Diabetes Dataset* (PIDD) untuk melakukan prediksi awal pada seorang pasien menggunakan metode klasifikasi K-NN. Hasil menunjukkan performa pada metode K-NN lebih dari 70% [8].

Fayroza Alaa Khaleel dan Abbas M. Al-Bakry (2021) penelitian yang serupa yaitu menggunakan PIDD pada metode Logistic Regression (LR), Naïve Bayes (NB), dan K-NN dengan hasil presisi, *recall*, dan *F-1 score* masing-masing pada LR yaitu 94%, 70%, 79%, pada NB 79%, 68%, 72%, dan K-NN yaitu 69%, 68% dan 68% [9]. Raharjo Putra Kurniadi, dkk (2021) melakukan prediksi menggunakan K-NN dan LR. Hasil menunjukkan metode K-NN lebih unggul dengan akurasi 85,06% [10].

Renata Rizki Raffi Athallah, dkk (2022) menggunakan beberapa kernel pada K-NN. Hasil yang didapatkan adalah pada K-NN 85%, kernel K-NN linier 96%, dan kernel K-NN polynomial 1 yaitu 96% [11]. A. Muhammad Syafar (2021) melakukan penelitian dengan PIDD pada metode K-NN dengan nilai *neighbor* (k) 2, 3, dan 5 mendapat hasil akurasi berturut-turut 74,09%, 71,09%, dan 70,70% [12]. Adetunji Olosogo Julius, dkk (2021) melakukan penelitian menggunakan dataset repositori Irvine (UCI) dengan perbandingan antara K-NN, Random Forest (RF), Support Vector Machine (SVM), dan Functional Tree (FT). Hasil menunjukkan K-NN unggul dengan hasil akurasi, presisi, dan *recall* berturut-turut 98,08%, 99,36%, dan 99% [13]. Umikulsum Indah Lestari, dkk (2021) yang melakukan prediksi menggunakan metode K-NN dengan PIDD mendapat hasil akurasi 96% [14].

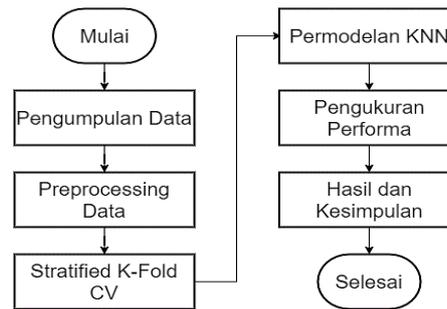
Berdasarkan latar belakang dan beberapa penelitian yang telah dilakukan sebelumnya, metode K-NN dapat digunakan sebagai klasifikasi untuk memprediksi pada sebuah data. Tahap klasifikasi merupakan tahap pengelompokan data yang berlandaskan relasi atau hubungan data terhadap data sampel [15]. Pada penelitian ini bertujuan untuk melakukan prediksi penyakit DM tipe 2 dengan membangun permodelan klasifikasi dari dataset yang bersumber dari Puskesmas Mlati II yang terletak di Kecamatan Mlati menggunakan K-NN dengan teknik *Manhattan Distance* sebagai eksperimental. Penelitian ini melakukan pembagian data menggunakan metode *Stratified K-fold CV* dengan tujuan prediksi dapat memberikan hasil yang baik.

2. METODOLOGI PENELITIAN

Pada tahap ini merupakan alur dari penelitian yang akan dilakukan. Pada gambar 1 menjelaskan detail tahap pelaksanaan. Dimulai dengan pengumpulan data yang bersumber dari Puskesmas Mlati II Sleman, kemudian dataset dilakukan *preprocessing* agar skala dataset atribut satu dengan lainnya seimbang. Selanjutnya implementasi metode K-NN dengan pengujian nilai k dimana ditentukan dengan data training, tahap akhir yaitu melakukan evaluasi terhadap permodelan.

Pada tahap ini merupakan alur dari penelitian yang akan dilakukan. Pada gambar 1 menjelaskan detail tahap pelaksanaan. Dimulai dengan pengumpulan data yang bersumber dari Puskesmas Mlati II Sleman, kemudian dataset dilakukan *preprocessing* agar skala dataset atribut satu dengan lainnya seimbang. Selanjutnya dilakukan pembagian data menggunakan *stratified k-fold cv*, kemudian implementasi metode K-

NN dengan pengujian nilai k dimana ditentukan dengan data training, tahap akhir yaitu melakukan evaluasi terhadap permodelan untuk mendapatkan hasil dan kesimpulan.



Gambar 1. Alur penelitian

2.1 Diabetes Melitus Tipe 2

DM tipe 2 merupakan penyakit yang berisiko lebih besar dialami oleh semua orang, dengan nilai prevalensi yaitu 90-95% yang disebabkan oleh faktor risiko modifikasi seperti gaya hidup yang tidak sehat dan cenderung mengonsumsi makanan cepat saji. DM tipe 2 ini terjadi karena insulin yang dihasilkan cukup namun tidak cukup untuk menjaga kadar gula agar normal. Penderita DM tipe 2 ini biasanya memasuki usia di atas 40 tahun [16]. Pada penelitian ini, cara untuk mengukur seseorang apakah terkena dan tidak terkena diabetes adalah dengan melakukan prediksi pada dataset yang mendukung faktor utama dari DM tipe 2.

2.2 Data Mining

Data mining atau disebut sebagai *Knowledge Discovery Databases (KDD)* karena pada umumnya digunakan untuk mencari pengetahuan yang terdapat dalam basis data dengan jumlah yang besar. *Data mining* merupakan proses pengolahan dari suatu basis data untuk mendapat pengetahuan berupa informasi yang belum diketahui sebelumnya secara manual.

2.3 Metode K-Nearest Neighbor (K-NN)

Algoritma K-NN merupakan salah satu dari bentuk *machine learning* populer sederhana yang termasuk bagian dari *supervised learning* yang digunakan untuk menentukan keputusan menggunakan perhitungan algoritma agar mendapat pola baru yang didapat dari pola data lama. Tujuan dari supervised learning adalah mengevaluasi hubungan antara variabel independen dan variabel dependen pada metodologi, seperti klasifikasi dan regresi [17]. Algoritma K-NN atau dikenal dengan *lazy learner* karena hanya mempelajari dan mengklasifikasi data tanpa membangun sebuah model, algoritma ini hanya perlu mengingat semua data training dalam memori [18]. Pada penelitian ini, parameter yang digunakan adalah nilai k (*neighbor*) dan perhitungan *Manhattan Distance*. Berikut proses tahapan dalam klasifikasi K-NN:

1. Menentukan parameter k, sebagai nilai tetangga (*neighbor*) terdekat (nilai k bisa berupa ganjil, sesuai dengan n data training yang digunakan). Rumus dalam menentukan nilai k sebagai berikut [19]:

$$k = \sqrt{n} \quad (1)$$

2. Menghitung jarak antara data training dan data testing menggunakan perhitungan *Manhattan Distance*. Rumus dalam menentukan jarak sebagai berikut:

$$d_m(x, y) = \sum |x_i - x_j| \quad (2)$$

Dengan keterangan:

- d : jarak antara x dan y
- x : data pusat kluster
- y : data pada variabel
- x_i : data pada pusat kluster ke i
- x_j : data pada setiap data ke j [20].

3. Urutkan hasil perhitungan jarak dari yang paling kecil hingga terbesar
4. Melakukan klasifikasi tetangga terdekat menggunakan nilai k
5. Menentukan hasil klasifikasi berdasarkan kelas yang paling banyak teridentifikasi [21].

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Dataset yang digunakan pada penelitian ini yaitu dataset klinis Puskesmas Mlati II Kabupaten Sleman, Daerah Istimewa Yogyakarta. Pada pengumpulan dataset didapatkan melalui laboratorium puskesmas yang sebelumnya pasien sudah dilakukan pemeriksaan. Dalam dataset ini berisi 200 baris yang melibatkan pasien pria dan wanita. Dataset ini terdapat 5 atribut yaitu usia, tekanan darah, BMI, GDP atau Gula Darah Puasa, dan sebagai klasifikasi yaitu atribut catatan terdiri dari kelas terkena dan tidak terkena.

Tabel 1. Data Awal

| No. | Usia | Tekanan Darah | BMI | GDP | Catatan Diabetes |
|-----|------|---------------|------|-----|------------------|
| 1 | 43 | 126 | 22,3 | 121 | Tidak Terkena |
| 2 | 37 | 127 | 21 | 112 | Tidak Terkena |
| 3 | 27 | 136 | 28,4 | 178 | Terkena |
| 4 | 28 | 114 | 34,6 | 104 | Terkena |
| 5 | 40 | 161 | 27,8 | 125 | Terkena |
| ... | ... | ... | ... | ... | ... |
| 199 | 53 | 135 | 26,2 | 93 | Tidak Terkena |

3.2. Preprocessing Data

Pada tahap *preprocessing* ini melakukan pemeriksaan dataset dari nilai null atau *missing values* dan normalisasi. Pada dataset penelitian ini terdapat jenis tipe data objek yaitu pada variabel catatan diabetes. Jenis data objek tersebut tidak bisa dilakukan proses ke tahap selanjutnya. Maka dari itu, dilakukan perubahan ke dalam jenis tipe numerik menggunakan fungsi *map()* pada *library* seperti Pandas. Pemetaan yang dilakukan adalah mengubah Terkena menjadi nilai 1 dan Tidak Terkena menjadi nilai 0. Berikut hasil dari penerapan *map()* pada tabel 2.

Tabel 2. Pemetaan Dataset

| No. | Usia | Tekanan Darah | BMI | GDP | Catatan Diabetes |
|-----|------|---------------|------|-----|------------------|
| 1 | 43 | 126 | 22,3 | 121 | 0 |
| 2 | 37 | 127 | 21 | 112 | 0 |
| 3 | 27 | 136 | 28,4 | 178 | 1 |
| 4 | 28 | 114 | 34,6 | 104 | 1 |
| 5 | 40 | 161 | 27,8 | 125 | 1 |
| ... | ... | ... | ... | ... | ... |
| 199 | 53 | 135 | 26,2 | 93 | 0 |

Tahap berikutnya adalah melakukan normalisasi dataset menggunakan metode *min-max normalization*. *Min-max normalization* adalah metode normalisasi untuk mengukur performa transformasi linear pada data asli sehingga menghasilkan nilai perbandingan yang seimbang antara data sebelum dan sesudah proses. Berikut rumus dari *min-max normalization* [22]:

$$X_{new} = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{3}$$

Dengan keterangan:

- X_{new} : Nilai baru dari hasil normalisasi
- X : Nilai lama
- Max (X) : Nilai maksimum dalam dataset
- Min (X) : Nilai minimum dalam dataset

Untuk menemukan hasil X_{new} dari setiap variabel, tahap awal adalah menentukan nilai minimum dan nilai maksimum. Berikut nilai minimum dan maksimum pada setiap variabel pada tabel 3.

Tabel 3. Nilai *Min-Max* Atribut

| Atribut | Nilai Minimum | Nilai Maksimum |
|---------------|---------------|----------------|
| Usia | 19 | 100 |
| Tekanan Darah | 95 | 215 |
| BMI | 16,4 | 38 |
| GDP | 74 | 364 |

Tabel 4 merupakan hasil setelah kalkulasi *min-max normalization*.

Tabel 4. Hasil Kalkulasi Min-Max

| No. | Usia | Tekanan Darah | BMI | GDP | Catatan Diabetes |
|-----|----------|---------------|----------|----------|------------------|
| 1 | 0,296296 | 0,258333 | 0,273148 | 0,162069 | 0 |
| 2 | 0,222222 | 0,266667 | 0,212963 | 0,131034 | 0 |
| 3 | 0,098765 | 0,341667 | 0,555556 | 0,358621 | 1 |
| 4 | 0,111111 | 0,158333 | 0,842593 | 0,103448 | 1 |
| 5 | 0,259259 | 0,550000 | 0,527778 | 0,175862 | 1 |
| ... | ... | ... | ... | ... | ... |
| 199 | 0,419753 | 0,333333 | 0,453704 | 0,065517 | 0 |

3.3. Stratified K-Fold Cross Validation

Pada tahapan berikutnya dataset dibagi menjadi data training dan data testing menggunakan *stratified K-fold cross validation*. Nilai K yang digunakan pada penelitian adalah 5, dimana pembagian menjadi 80% pada data training dan 20% pada data testing, atau dalam jumlah 200 data menjadi 160 pada data training dan 40 pada data testing.

3.4. Permodelan K-Nearest Neighbor (K-NN)

Pada tahap permodelan metode K-NN yaitu menentukan nilai k dengan menggunakan data training sebanyak 160 data. Berikut kalkulasi dalam menentukan nilai k:

$$k = \sqrt{160} = 12,6491106407$$

Hasil yang didapat adalah 12,6491106407 atau dalam penelitian ini jika dibulatkan maka hasilnya k = 13.

3.5. Pengukuran Performa

Setelah tahap membangun permodelan menggunakan metode K-NN, tahap berikutnya adalah melakukan pengukuran performa menggunakan *confusion matrix*, dilanjutkan dengan evaluasi menggunakan akurasi, presisi, *recall*, dan *f1-score*. Berikut hasil *confusion matrix* pada tabel 5.

Tabel 5. Confusion Matrix

| Prediksi | Jumlah |
|-----------------------|--------|
| <i>True Positive</i> | 20 |
| <i>False Positive</i> | 4 |
| <i>False Negative</i> | 4 |
| <i>True Negative</i> | 12 |

Evaluasi berikutnya dari hasil *confusion matrix* yaitu berupa nilai rata-rata akurasi, presisi, *recall*, dan *f1-score* pada tabel 6 di bawah ini:

Tabel 6. Evaluasi

| Parameter | Nilai |
|-----------------|--------|
| Akurasi | 0,8800 |
| Presisi | 0,8354 |
| <i>Recall</i> | 0,8750 |
| <i>F1-score</i> | 0,8536 |

4. KESIMPULAN

Pada penelitian ini, melakukan eksperimen menggunakan dataset Puskesmas Mlati II Sleman yang terdiri dari 5 variabel yaitu usia, tekanan darah, BMI, gula darah puasa, dan catatan diabetes dengan mengimplementasikan permodelan klasifikasi K-NN. Parameter yang digunakan pada metode K-NN adalah nilai k yaitu 13 dan perhitungan *manhattan distance* dalam menentukan jarak antara data testing dengan data training. Pembagian data training dan data testing pada penelitian ini menggunakan *stratified 5-fold cv*. Hasil evaluasi pada permodelan ini menggunakan akurasi, presisi, *recall*, dan *f1-score* berturut-turut adalah 88%, 83,54%, 87,5%, dan 85,36%. Hal ini menunjukkan bahwa implementasi metode klasifikasi K-NN dapat diimplementasikan dan memberikan hasil evaluasi yang baik.

Kekurangan pada penelitian ini adalah dataset yang digunakan memiliki variabel yang sedikit. Sebagai penelitian selanjutnya, dapat dikembangkan dengan menambahkan variabel seperti insulin, riwayat keluarga, maupun riwayat kehamilan. Eksperimen lainnya adalah dengan melakukan implementasi jenis metode

klasifikasi yang berbeda, seperti metode Naïve Bayes, SVM (*Support Vector Machine*), Random Forest, DNN (*Deep Neural Network*) dan metode lainnya dengan tujuan untuk mendapat hasil akurasi yang lebih baik.

UCAPAN TERIMA KASIH

Ucapan terima kasih pihak Puskesmas Mlati II Sleman yang telah memberikan kesempatan untuk melaksanakan penelitian.

REFERENSI

- [1] Gusti Restuning Ramadhan Sabela, "Penerapan Edukasi Tentang DM Terhadap Pengetahuan Kesehatan di Dusun Mloso Desa Kedungbanjar," 2020, Diakses: 12 Februari 2023. [Daring]. Tersedia pada: <http://repository.unair.ac.id/id/eprint/110286>
- [2] N. R. Dzakiyullah, M. A. Burhanuddin, R. R. R. Ikram, K. A. Ghani, dan W. Setyonugroho, "Machine learning methods for diabetes prediction," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, hlm. 2199–2205, Okt 2019, doi: 10.35940/ijitee.L2973.1081219.
- [3] "Terapi Pada DM Tipe 1," Jl. Prof. DR. Ir. Sumatri Brojonegoro No.1, Gedong Meneng, Kec. Rajabasa, Kota Bandar Lampung, Lampung, Indonesia 35145. [Daring]. Tersedia pada: <http://jurnal.globalhealthsciencegroup.com/index.php/JPPP>
- [4] W. Dwi Prasetya dan B. Sujatmiko, "Rancang Bangun Aplikasi dengan Perbandingan Metode K-Nearest Neighbor (K-NN) dan Naive Bayes dalam Klasifikasi Penderita Penyakit Diabetes," *Journal of Informatics and Computer Science*, vol. 03, 2022.
- [5] A. Panesar, *Machine Learning and AI for Healthcare*. Apress, 2021. doi: 10.1007/978-1-4842-6537-6.
- [6] J. Xin dan S. Chen, "Bus Dwell Time Prediction Based on K-NN," dalam *Procedia Engineering*, Elsevier Ltd, 2016, hlm. 283–288. doi: 10.1016/j.proeng.2016.01.260.
- [7] J. Mantik *dkk.*, "Implementation of K-NN algorithm in classifying diabetic ulcers in patients with diabetes mellitus," Online, 2023.
- [8] J. J. Khanam dan S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, hlm. 432–439, Des 2021, doi: 10.1016/j.ict.2021.02.004.
- [9] F. Alaa Khaleel dan A. M. Al-Bakry, "Diagnosis of diabetes using machine learning algorithms," *Mater Today Proc*, Jul 2021, doi: 10.1016/j.matpr.2021.07.196.
- [10] R. Putra Kurniadi, R. Rohmat Saedudin, dan V. P. Widartha, "Perbandingan Akurasi Algoritma K-Nearest Neighbor dan Logistic Regression untuk Klasifikasi Penyakit Diabetes."
- [11] R. Rizki, R. Athallah, I. Cholissodin, dan P. P. Adikara, "Prediksi Potensi Pengidap Penyakit Diabetes berdasarkan Faktor Risiko Menggunakan Algoritme Kernel K-Nearest Neighbor," 2022. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>.
- [12] A. M. Syafar, M. Syawal, I. Fitrah Baharuddin, M. Zainal, N. Fuadi, dan I. Artikel, "Peramalan Penderita Diabetes Menggunakan Algoritma K-Nearest Neighbor," vol. 1, no. 2, 2021, doi: 10.31850/jsilog.v1i3.
- [13] A. Olusogo Julius, A. Olusola Ayokunle, dan F. Olawale Ibrahim, "Early Diabetic Risk Prediction using Machine Learning Classification Techniques," 2021. [Daring]. Tersedia pada: www.ijisrt.com.
- [14] U. I. Lestari, A. Y. Nadhiroh, dan C. Novia, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus," vol. 8, no. 4, 2021, [Daring]. Tersedia pada: <http://jurnal.mdp.ac.id>.
- [15] M. G. Pradana, P. H. Saputro, dan D. P. Wijaya, "Komparasi Metode Support Vector Machine dan Naive Bayes dalam Klasifikasi Peluang Penyakit Serangan Jantung," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 5, no. 2, hlm. 87, Des 2022, doi: 10.21927/ijubi.v5i2.2659.
- [16] "Terapi Pada DM Tipe 1," Jl. Prof. DR. Ir. Sumatri Brojonegoro No.1, Gedong Meneng, Kec. Rajabasa, Kota Bandar Lampung, Lampung, Indonesia 35145. [Daring]. Tersedia pada: <http://jurnal.globalhealthsciencegroup.com/index.php/JPPP>
- [17] N. R. Dzakiyullah, A. Pramuntadi, dan A. K. Fauziyyah, "Semi-Supervised Classification on Credit Card Fraud Detection using AutoEncoders," *Journal of Applied Data Sciences*, vol. 2, no. 1, hlm. 1–07, 2021.
- [18] M. S. Fajri, N. Septian, dan E. Sanjaya, "Evaluasi Implementasi Algoritma Machine Learning K-Nearest Neighbors (K-NN) pada Data Spektroskopi Gamma Resolusi Rendah," *Al-Fiziya: Journal of Materials Science, Geophysics, Instrumentation and Theoretical Physics*, vol. 3, no. 1, hlm. 9–14, Agu 2020, doi: 10.15408/fiziya.v3i1.16180.
- [19] S. Margaretta, I. Arwani, dan D. E. Ratnawati, "Implementasi Algoritma K-Nearest Neighbor Pada Database Menggunakan Bahasa SQL," 2020. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>.
- [20] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 4, no. 1, hlm. 20–24, Jan 2019, doi: 10.30591/jpit.v4i1.1253.

- [21] M. E. Febrian, F. X. Ferdinan, G. P. Sendani, K. M. Suryanigrum, dan R. Yunanda, "Diabetes prediction using supervised machine learning," dalam *Procedia Computer Science*, Elsevier B.V., 2022, hlm. 21–30. doi: 10.1016/j.procs.2022.12.107.
- [22] H, T. Wahyuningsih, dan E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (K-NN) Algorithm to Test the Accuracy of Types of Breast Cancer." [Daring]. Tersedia pada: <http://archive.ics.uci.edu/ml>.