

# Penggunaan Algoritma K-Nearest Neighbor Untuk Klasifikasi Data Diabetes Pada Wanita

**Diterima:** 10 Juni 2024  
**Revisi:** 10 Juli 2024  
**Terbit:** 1 Agustus 2024

<sup>1\*</sup>Maha Shelin Sahira, <sup>2</sup>Fitria Dessela Putri, <sup>3</sup>Aidina Ristyawan, <sup>4</sup>Erna Daniati  
<sup>1-4</sup>Universitas Nusantara PGRI Kediri Fakultas Teknik dan Ilmu Komputer  
<sup>1</sup>[shellinsahira@gmail.com](mailto:shellinsahira@gmail.com), <sup>2</sup>[desselaputri@gmail.com](mailto:desselaputri@gmail.com),  
<sup>3</sup>[aidinaristi@unpkediri.ac.id](mailto:aidinaristi@unpkediri.ac.id), <sup>4</sup>[ernadaniati@unpkediri.ac.id](mailto:ernadaniati@unpkediri.ac.id)

**Abstrak**— Diabetes adalah kadar gula darah yang tinggi atau di atas nilai normal, dalam menguji berbagai metode pada kumpulan data yang relevan adalah salah satu cara untuk menentukan metode klasifikasi yang tepat untuk mengelola diabetes. Dalam penelitian ini, masalah yang diangkat adalah bagaimana mengukur kinerja metode klasifikasi dalam mengelola diabetes yang tidak dikontrol dengan baik. Selama proses klasifikasi, algoritma *K-Nearest Neighbor* dan tools yang digunakan *RapidMiner* untuk menguji nilai *accuracy*, *class precision*, dan *class recall* dari data yang digunakan dalam penelitian ini. Data yang digunakan dalam penelitian ini berasal dari dataset diabetes *Kaggle*. Oleh karena itu, algoritma *K-Nearest Neighbor* digunakan untuk mengklasifikasikan penyakit diabetes pada wanita dapat digunakan karena proses seleksinya cepat, metodenya mudah dipahami, dan memiliki nilai akurasi yang baik (79.93%), ketepatan kelas (78.19%), dan recall kelas 96.45%.

**Kata Kunci**— *Diabetes, K-Nearest Neighbor, RapidMiner, Kaggle, Klasifikasi, KDD*

**Abstract-** *Diabetes is a blood sugar level that is high or above normal values, in testing various methods on relevant data sets is one way to determine the right classification method for managing diabetes. In this study, the problem raised is how to measure the performance of classification methods in managing poorly controlled diabetes. During the classification process, the K-Nearest Neighbor algorithm and RapidMiner tools are used to test the accuracy, class precision, and class recall values of the data used in this study. The data used in this study comes from the Kaggle diabetes dataset. Therefore, the K-Nearest Neighbor algorithm used to classify diabetes in women can be used because the selection process is fast, the method is easy to understand, and has a good accuracy value (79.93%), class precision (78.19%), and class recall 96.45%.*

**Keywords-** *Diabetes, K-Nearest Neighbor, RapidMiner, Kaggle, Classification. Knowledge Discovery on Database*

This is an open access article under the CC BY-SA License.



---

## Penulis Korespondensi:

Nama Penulis, *Aidina Ristyawan*  
Departemen Penulis, *Prodi Sistem Informasi*  
Institusi Penulis, *Universitas Nusantara PGRI Kediri*  
Email: [aidinaristi@unpkediri.ac.id](mailto:aidinaristi@unpkediri.ac.id)  
ID Orcid: [<https://orcid.org/0009-0003-2712-1507>]  
Handphone: 081232624460

---

## I. PENDAHULUAN

Diabetes adalah penyakit yang disebabkan oleh banyak komplikasi dan penyakit lain. Ini bertentangan dengan pendapat umum bahwa diabetes hanyalah penyakit yang ditandai dengan kadar gula darah tinggi dari normal. Diabetes biasanya didefinisikan sebagai penyakit kronis yang memiliki kadar glukosa darah meningkat secara tidak teratur sebagai hasil dari sekresi insulin yang tidak normal. Risiko yang termasuk yaitu: stroke, penyakit jantung, kebutaan, gagal ginjal, dan bahkan kematian. Perawatan membutuhkan pemantauan medis yang tepat.[1]

Meningkatnya penderita diabetes tidak sama dengan tindakan yang memadai. Pada saat ini, transformasi digital mengubah cara orang menggunakan teknologi informasi secara efektif dan efisien. Dengan demikian, penggunaan teknik data mining untuk memprediksi dan pengklasifikasian penyakit berdasarkan dataset *kaggle*. [2]

Oleh karena itu, terdapat teknologi alternatif yang menggunakan pembelajaran mesin, seperti algoritma *K-Nearest Neighbors (KNN)*. KNN mengklasifikasikan objek baru berdasarkan jumlah K tetangga terdekat. Karena sederhana dan mudah dipahami, algoritma KNN banyak digunakan. Dengan algoritma ini, gambar dikategorikan berdasarkan jarak terdekatnya dengan gambar tetangganya. Penelitian ini mengangkat masalah mengenai pengukuran kinerja algoritma *K-Nearest Neighbors (KNN)* dalam mendiagnosis penyakit diabetes pada wanita. Pengujian algoritma *K-Nearest Neighbors (KNN)* digunakan untuk mempelajari data diabetes pada wanita. Tujuan dari pengujian ini adalah untuk meningkatkan pemahaman kita bagaimana cara data mining digunakan untuk klasifikasi diabetes pada wanita. Dataset diabetes yang penulis ini adalah "*Diabetes Data-set*" yang diperoleh dari *Kaggle* dan berisi 9 atribut. [3]

## II. METODE

Peneliti menggunakan metodologi yaitu *Knowledge Discovery on Database (KDD)*. metode untuk menciptakan pola atau aturan dalam data, [4] penggunaan proses komputasi untuk perhitungan matematis dan mengekstrak informasi yang melakukan perhitungan probabilitas yang dapat diperoleh dari probabilitas tindakan di masa depan. Metode penemuan pengetahuan pada database meliputi tahapan pemilihan data, preprocessing data, transformasi data, penambahan data, dan evaluasi. [5]

## 2.1 Metode KDD



Gambar 1. Metode KDD

### 1) Selection

Pemilihan data berdasarkan atribut dan diintegrasikan ke dalam kumpulan data selama fase pemilihan data. Proses mempelajari data mining dan menemukan pola baru bergantung pada dataset yang dibuat, sehingga proses pembuatan dataset sangatlah penting. [6]

### 2) Preprocessing

Pada tahap preprocessing data, ada proses pembersihan, yaitu memperbaiki kesalahan, menghapus duplikat, pemeriksaan data tidak konsisten. Selain itu, pada tahap ini, data akan diperkaya dengan informasi atau data luar yang diperlukan. [7]

#### a. Normalization Data

Preprocessing termasuk fase transformasi data. Fase ini dibagi lagi menjadi beberapa fase lagi, salah satunya adalah normalisasi. Ini adalah teknik yang digunakan untuk memetakan data ke skala tertentu selama proses data mining. *Min-Max* dan *Z-score* normalisasi adalah beberapa metode umum untuk normalisasi data. [3]

#### b. Split Data

Pemisahan data melibatkan kumpulan data menjadi 2 bagian yaitu data pelatihan dan data pengujian. Tujuannya adalah menguji data menggunakan titik data yang tidak diketahui, bukan menguji pada titik yang sama dengan tempat model dilatih. Data uji digunakan untuk menguji model klasifikasi menghasilkan dan mengetahui kinerja model klasifikasi memiliki cara perbandingan hasil klasifikasi model dengan label sebenarnya untuk seluruh data pada data uji.[3]

### 3) Transformation

Proses perubahan bentuk data yang memengandung entitas yang tidak jelas menjadi bentuk data yang dapat diproses lebih lanjut (data mining) dikenal sebagai transformasi data.[7] transformasi tipe data non-numerik menjadi data numerik melalui

tahapan transformasi data, yang digunakan dalam penelitian ini untuk membuat data tersedia untuk proses data mining.[8]

#### 4) Data Mining

Data Mining adalah proses pengelolaan data menjadi informasi yang akurat dan mempermudah memecahkan masalah dengan bermacam-macam metode atau algoritma. perangkat lunak pengolahan data yang memungkinkan analisis data mining, text mining.

#### 5) Evaluation

*Evaluation* adalah proses yang menggambarkan hasil dari model dan menguji kesesuaian dan akurasi data.[9] Membahas hasil data mining menjadi bentuk yang mudah dipahami disebut evaluasi. Tahap ini termasuk pemeriksaan pola informasi yang telah dihasilkan. [10]

### 2.2 Klasifikasi

Klasifikasi adalah Pembelajaran mempunyai fungsi untuk menentukan setiap atribut suatu objek, dan klasifikasi dikatakan sebagai proses pengelompokan data. Klasifikasi adalah istilah yang digunakan untuk menggambarkan kumpulan data di mana setiap jenis data memiliki format nominal atau *biner*. Ada beberapa algoritma yang digunakan untuk klasifikasi data, seperti : *algoritma KNN*, *algoritma Naïve Bayes* *algotirma C4.5*, *algoritma C5* dll. [7]

### 2.3 Algoritma K-Nearest Neighbors (K-NN)

Metode pengklasifikasian objek berdasarkan data latih yang paling dekat dengannya adalah algoritma *K-Nearest Neighbor (K-NN)*. Teknik ini mudah digunakan dan mudah dipelajari. Mengikuti pendekatan *clustering*. Dengan kata lain, mengelompokkan data baru berdasarkan seberapa dekat mereka dengan beberapa data atau tetangga. [11]

### 2.4 Rapid Miner

*Rapid Miner* adalah *software* pengolahan data untuk data mining yang bersifat open source yang menawarkan lingkungan yang terintegrasi untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif.[12] Penelitian penulis menggunakan Metode *K-Nearest Neighbor (KNN)* untuk mengklasifikasikan dataset diabetes pada wanita yang dari *kaggle*. *Rapid Miner* diterbitkan sebagai perangkat lunak sumber terbuka, dan pengguna

bebas menggunakan dan memodifikasinya.[13] Aplikasi ini menawarkan operator-operator yang memungkinkan pengguna melakukan perhitungan besar data dan melakukan modifikasi. Untuk penelitian kali ini menggunakan algoritma *K-Nearest Neighbors (K-NN)*. [14]

### III. HASIL DAN PEMBAHASAN

Penelitian ini berfokus pada proses *klasifikasi* dengan menggunakan program *RapidMiner* dan algoritma KNN. Tujuan penelitian adalah untuk menghasilkan nilai akurasi pada algoritma yang digunakan untuk mengklasifikasikan penyakit diabetes pada wanita. Dataset yang digunakan untuk penelitian ini dideskripsikan dalam tabel 1.

Tabel 1. *Atribut Dataset*

| No. | Atribut                          | Keterangan  |
|-----|----------------------------------|---|
| 1.  | <i>Pregnancies</i>               | Jumlah kehamilan                                    |
| 2.  | <i>Glucoses</i>                  | Kadar glukosa dalam darah                           |
| 3.  | <i>Blood Pressure</i>            | Tekanan darah                                       |
| 4.  | <i>Insulin</i>                   | Tingkat insulin dalam darah                         |
| 5.  | <i>BMI</i>                       | Menyatakan indeks massa tubuh                       |
| 6.  | <i>Diabetes Pedigree Fuction</i> | Presentase diabetes                                 |
| 7.  | <i>Age</i>                       | Menyatakan umur                                     |
| 8.  | <i>Outcome</i>                   | Menyatakan hasil akhir <i>True</i> dan <i>False</i> |
| 9.  | <i>Skin Thickness</i>            | Menyatakan ketebalan tubuh                          |

#### 4.1 Selection

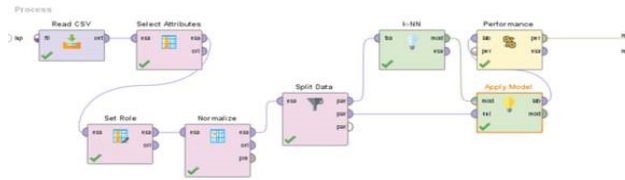
Tahap pemilihan data meliputi pemilihan kolom-kolom yang akan digunakan dalam penelitian. Penelitian tersebut menggunakan tujuh karakteristik *Pregnancies*, *Glucose*, *Blood Pressure*, *Insulin*, *BMI*, *Diabetes Pedigree Fuction*, *Age*, *Outcome*. Atribut yang digunakan penulis 'outcome' sebagai *label*. Tabel 2 menunjukkan atribut yang dipilih untuk digunakan.

Tabel 2. *Atribut Dipilih*

| No | Atribut                          | Keterangan  |
|----|----------------------------------|---|
| 1. | <i>Pregnancies</i>               | Jumlah Kehamilan                                    |
| 2. | <i>Glucoses</i>                  | Kadar glukosa dalam darah                           |
| 3. | <i>Blood Pressure</i>            | Tekanan darah                                       |
| 4. | <i>Insulin</i>                   | Tekanan insulin dalam darah                         |
| 5. | <i>BMI</i>                       | Menyatakan indeks masa tubuh                        |
| 6. | <i>Diabetes Pegigree Fuction</i> | Presentase diabetes                                 |
| 7. | <i>Age</i>                       | Menyatakan umur                                     |
| 8. | <i>Outcome</i>                   | Menyatakan hasil akhir <i>True</i> dan <i>False</i> |

Tabel 2 menunjukkan proses pengolahan data menggunakan algoritma KNN dan alat *RapidMiner*. Atribut yang dipilih sesuai dengan gejala wanita yang menderita diabetes diolah

menggunakan algoritma KNN dan alat *RapidMiner*, yang menghasilkan nilai seperti *accuracy*, *precision*, *recall*, *confusion matrix*, dan *kurva ROC*.



Gambar 2. *RapidMiner*

Atribut dataset yang telah dipilih dimasukkan dalam tahap *preprocessing* yang dapat dipastikan bahwa tidak ada *Missing Value*. Pada penelitian penulis, atribut dataset diabetes tidak ditemukan nilai yang hilang dan sudah lengkap. Setelah transformasi data, *RapidMiner* dibagi menjadi data latihan dan data uji, dengan persentase data latihan sebesar 60% dan data uji sebesar 40%.

Selanjutnya, tahap berikutnya adalah Data Mining dapat dilakukan klasifikasi menggunakan algoritma K-NN memiliki nilai  $K=23$  untuk menghasilkan nilai yang tepat untuk klasifikasi penyakit diabet pada wanita menggunakan *RapidMiner* untuk membantu mendeteksi diabetes lebih awal. Perhitungan evaluasi dilakukan dengan cara *confussion matrix* penghasilan performa klasifikasi. Pada tabel 3 menunjukkan nilai *confussion matrix* yang dihasilkan dengan algoritma KNN.

Tabel 3. *Confussion Matrix*

|              | true TRUE | True FALSE | Class precision |
|--------------|-----------|------------|-----------------|
| Pred TRUE    | 49        | 7          | 87.50%          |
| Pred FALSE   | 53        | 190        | 78.19%          |
| Class recall | 48.04%    | 96.45%     |                 |

*Confusion matrix* digunakan untuk memeriksa label kelas mana yang memiliki nilai benar dan teridentifikasi dengan benar. Tabel *Confusion matrix* digunakan untuk mencari hasil aktual dan prediksi yang dihasilkan oleh klasifikasi data.[15] Pengujian dilakukan dengan parameter optimal pada umumnya *confussion matrix* memiliki 3 metode untuk pengujian :

1) *Accuracy*

*Accuracy* adalah Proporsi prediksi yang benar terhadap keseluruhan prediksi. Rumus untuk perhitungan *accuracy* yaitu :

a. Nilai *accuracy* :

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN}$$

Keterangan :

- TP (*True Positive*) = 49
- TN (*True Negative*) = 190
- FP (*False Positive*) = 7
- FN (*False Negative*) = 53

b. Perhitungan accuracy :

$$Accuracy = \frac{48 + 190}{48 + 190 + 7 + 53} = \frac{239}{299} = 79.93\%$$

Keterangan :

- TP: *True Positive* (Prediksi *TRUE*, Aktual *TRUE*)
- TN: *True Negative* (Prediksi *FALSE*, Aktual *FALSE*)
- FP: *False Positive* (Prediksi *TRUE*, Aktual *FALSE*)
- FN: *False Negative* (Prediksi *FALSE*, Aktual *TRUE*)

## 2) Class Precision

*Class Precision* adalah proporsi prediksi positif yang benar terhadap keseluruhan prediksi positif. berikut rumus perhitungan pada class precision :

a. Nilai *class precision* :

$$Class\ precision = \frac{TN}{TN + FN}$$

- TN (*True Negative*) = 190
- FN (*False Negative*) = 53

b. Perhitungan Precision =  $\frac{190}{190+53} = \frac{190}{243} = 78.19\%$

## 3) Class Recall

*Class Recall* Proporsi prediksi positif yang benar terhadap keseluruhan kasus positif sebenarnya. berikut rumus perhitungan dari class recall

a. Nilai *Class Recall*

$$Class\ recall = \frac{TN}{TN + FP}$$

- TN (*True Negative*) = 190
- FP (*False Positive*) = 7

b. Perhitungan Recall =  $\frac{190}{190+7} = \frac{190}{197} = 96.45\%$

Penjelasan perhitungan :

- **Akurasi 79.93%:** Dari total 299 prediksi, 239 prediksi benar (baik TRUE maupun FALSE). Ini mencakup semua prediksi yang benar dibandingkan dengan total prediksi.
- **Class Precision 78.19%:** Dari semua kasus yang diprediksi sebagai FALSE, 78.19% dari mereka benar-benar FALSE. Ini menunjukkan seberapa baik model dalam menghindari kesalahan positif palsu untuk kelas FALSE.
- **Class Recall 96.45%:** Dari semua kasus yang benar-benar FALSE, model berhasil mengidentifikasi 96.45% dari mereka sebagai FALSE. Ini menunjukkan seberapa baik model dalam menangkap semua kasus yang benar-benar negatif.

Pada Tabel 3 diketahui bahwa nilai dari class precision sebesar 78.19% dan nilai dari class recall 96.45%. Dari Klasifikasi menggunakan algoritma KNN menghasilkan nilai akurasi yang bisa dilihat pada gambar 3.

```
PerformanceVector  
PerformanceVector:  
accuracy: 79.93%  
ConfusionMatrix:  
True:  TRUE  FALSE  
TRUE:  49    7  
FALSE:  53   190  
precision: 78.19% (positive class: FALSE)  
ConfusionMatrix:  
True:  TRUE  FALSE  
TRUE:  49    7  
FALSE:  53   190  
recall: 96.45% (positive class: FALSE)  
ConfusionMatrix:  
True:  TRUE  FALSE  
TRUE:  49    7  
FALSE:  53   190  
AUC (optimistic): 0.809 (positive class: FALSE)  
AUC: 0.809 (positive class: FALSE)  
AUC (pessimistic): 0.809 (positive class: FALSE)
```

Gambar 3. Performance Vector

Kurva ROC adalah alat dua dimensi untuk mengevaluasi kinerja klasifikasi menggunakan dua kelas keputusan. Setiap objek dikaitkan dengan elemen himpunan pasangan positif atau negatif. Teknik ini digunakan untuk memvisualisasikan, mengatur, dan memilih klasifikasi.[16]

Gambar 4 menunjukkan kurva ROC sebagai hasil dari pengujian data yang digunakan, yang menghasilkan nilai akurasi sebesar 78,50% untuk mengklasifikasikan kumpulan data diabetes pada perempuan.



Gambar 4. Kurva ROC

Kurva ROC atau *Receiver Operating Characteristic* pada gambar tersebut menunjukkan hasil klasifikasi algoritma KNN menghasilkan nilai AUC 0.809 yang berarti baik.



accuracy: 76.33% +/- 8.05% (micro average: 76.35%)

|              | true true | true false | class precision |
|--------------|-----------|------------|-----------------|
| pred. true   | 236       | 80         | 74.68%          |
| pred. false  | 60        | 216        | 78.26%          |
| class recall | 79.73%    | 72.97%     |                 |

Gambar 5. Accuracy Cross Validation

Pada gambar 5 perbandingan menggunakan SMOTE untuk upsampling dan cross validation yang menghasilkan hasil *accuracy* 76.33%, *class precision* 78.97%, dan *class recall* 78.26%

Pada penelitian sebelumnya penulis Amellia feronica agustin dan Apriade voutama dengan judul “Implementasi data mining klasifikasi penyakit diabetes pada perempuan menggunakan *Naive Bayes*”. Perbandingan penelitian sebelumnya menggunakan *naive bayes*, dan penelitian penulis dengan algoritma K-NN pada table 4.

Tabel 4. Perbandingan *accuracy*

|   | <i>Accuracy</i> | <i>Class Precision</i> | <i>Class Recall</i> |
|---|-----------------|------------------------|---------------------|
| <i>Naive Bayes</i>                                    | 78.50%          | 85.24%                 | 83.64%              |
| <i>K-NN</i>   | 79.93%          | 78.19%                 | 96.45%              |
| <i>K-NN (Smote Upsumpling &amp; Cross Validation)</i> | 76.33%          | 78.26%                 | 72.97%              |

#### IV. KESIMPULAN

Salah satu cara untuk menentukan metode klasifikasi yang tepat untuk mengelola diabetes adalah dengan menguji berbagai metode pada kumpulan data yang relevan. Dalam penelitian ini, masalah yang diangkat adalah bagaimana pengukur kinerja metode klasifikasi dalam pengelola diabetes tidak dikontrol dengan baik, Algoritma *K-Nearest Neighbor* dan alat *RapidMiner* untuk penguji nilai *accuracy*, *class precision*, dan *class recall* dari data yang digunakan dalam penelitian selama proses klasifikasi. Dalam pengujian kali ini menggunakan algoritma K-NN dengan K-23 percobaan pertama naik hasil *accuracy* 79.93%, *class precesion* 78.19%, dan *class recall* 96.45% . percobaan kedua menggunakan *tools smote* dan *cross validasi* K-23 dengan hasil *accuracy* 76.33%, *class precision* 78.26%, *class recall* 72.97%.

#### UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Bapak/Ibu Dosen Pembimbing Universitas Nusantara PGRI Kediri yang telah memberikan dukungan dan bimbingan yang sangat berarti dalam pelaksanaan penelitian ini, Semoga hasil penelitian ini bermanfaat dan dapat memberikan kontribusi yang berarti bagi pengembangan ilmu pengetahuan.

#### DAFTAR PUSTAKA

- [1] A. M. Argina, “Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, Jul. 2020.
- [2] P. SHETA, “diabetes-dataset.” [Online]. Available: <https://www.kaggle.com/datasets/pritsheta/diabetes-dataset>.
- [3] N. Azizah, M. Riyad Firdaus, R. Suyaningsih, and F. Indrayatna, “Penerapan Algoritma Klasifikasi K-Nearest Neighbor pada Penyakit Diabetes,” *Pros. Semin. Nas. Stat. Aktuaria*, vol. 2, no. 1, pp. 119–126, 2023.
- [4] J. Nasir, “Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 690–703, 2021.
- [5] A. Veronica Agustin and A. Voutama, “Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1002–1007, 2023.
- [6] F. M. Almufqi and A. Voutama, “Perbandingan Metode Data Mining Untuk Memprediksi Prestasi Akademik Siswa,” *J. Tek.*, vol. 15, no. 1, pp. 61–66, 2023.
- [7] N. M. Putry, “Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus,” *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, 2022.
- [8] J. I. Marzuki, K. Mataram, and N. T. Bar, “KOMPARASI AKURASI METODE CORRELATED NAIVE BAYES CLASSIFIER DAN NAIVE BAYES CLASSIFIER UNTUK DIAGNOSIS PENYAKIT DIABETES Hairani , Gibran Satya Nugraha , Mokhammad Nurkholis Abdillah , Muhammad Innuddin InfoTekJar ( Jurnal Nasional Informatika dan Teknolog,” *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 3, no. 1, pp. 6–11, 2018.
- [9] F. Alghifari and D. Juardi, “Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes,” *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021.
- [10] D. M. A. M. Sanjaya, A. A. I. I. Paramitha, and N. W. Utami, “Penerapan Data Mining untuk Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma C4.5: Studi Kasus STMIK Primakara,” *J. Ilm. Ilmu Terap. Univ. Jambi*, vol. 6, no. 1, pp. 84–97, 2022.
- [11] R. A. Manullang, F. A. Sianturi, [ Penerapan, A. K.-N. Neighbor, U. Memprediksi, and K. Mahasiswa, “JIKOMSI [Jurnal Ilmu Komputer dan Sistem Informasi] Penerapan Algoritma K-Nearest Neighbour Untuk Memprediksi Kelulusan Mahasiswa,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 4, no. 2, pp. 42–50, 2021.
- [12] R. Nofitri and N. Irawati, “Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer,” *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 199–204, 2019.
- [13] F. Yunita, “Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor ( K-Nn ),” *Bappeda*, vol. 2, no. 1, pp. 223–230, 2016.
- [14] C. H. P. Panjaitan, L. J. Pangaribuan, and C. I. Cahyadi, “Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner untuk Sistem Rekomendasi Tempat Wisata Labuan Bajo,” *Remik*, vol. 6, no. 3, pp. 534–541, 2022.
- [15] S. I. Fernanda, D. E. Ratnawati, and P. P. Adikara, “Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode Modified K- Nearest Neighbor ( MKNN ),” *J. Pengemb. Teknol. Inf. dan Ilmu Komputer* Fernanda, S. I., Ratnawati, D. E., Adikara, P. P. (2017). *Identifikasi Penyakit Diabetes Mellit. Menggunakan Metod. Modif. K- Nearest Neighbor ( MKNN ). J. Pengemb. Teknol. Inf. ,* vol. 1, no. 6, pp. 507–513, 2017.
- [16] F. Sistem Informasi STMIK Nusa Mandiri Jakarta Jl Damai No, W. Jati Barat, and J. Selatan, “Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes,” *J. Techno Nusa Mandiri*, vol. XIII, no. 1, p. 50, 2016.