

Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree

Classification of Diabetes Disease Using Decision Tree Algorithm

Mutiara Dwi Triapanca¹, Putri Aryaningsih², Tommy Dwi Putra³

^{1,2,3} Program Studi Informatika, STMIK Amikom Surakarta

e-mail: *¹ mutiara.10317@mhs.amikomsolo.ac.id, *² putri.10316@mhs.amikomsolo.ac.id,

*³ tommy@dosen@dosen.amikom.ac.id

Abstrak - Penelitian ini bertujuan untuk membangun model klasifikasi guna memprediksi penyakit diabetes menggunakan algoritma Decision Tree dengan bantuan perangkat lunak RapidMiner. Dataset yang digunakan bersumber dari Kaggle dan terdiri atas 100.000 data pasien dengan 16 atribut kesehatan yang relevan. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan, pembangunan model klasifikasi, serta evaluasi menggunakan Confusion Matrix. Model dilatih dengan 70% data dan diuji menggunakan 30% sisanya. Hasil pengujian menunjukkan tingkat akurasi sebesar 97,16%, dengan error klasifikasi sebesar 2,84%. Model ini juga menunjukkan kemampuan yang baik dalam mendeteksi lokasi geografis dengan jumlah prediksi diabetes yang tinggi, seperti New Jersey. Dengan tingkat akurasi yang tinggi serta interpretasi model yang mudah, algoritma Decision Tree terbukti efektif dalam mendeteksi risiko diabetes berdasarkan atribut-atribut pasien yang dianalisis..

Kata kunci – algoritma; data mining; decision tree; diabetes; RapidMiner

Abstract - This research aims to build a classification model to predict diabetes using the Decision Tree algorithm with the help of RapidMiner software. The dataset used is sourced from Kaggle and consists of 100,000 patient data with 16 relevant health attributes. The research stages include data collection, pre-processing, classification model building, and evaluation using Confusion Matrix. The model was trained with 70% of the data and tested using the remaining 30%. The test results showed an accuracy rate of 97.16%, with a classification error of 2.84%. The model also showed a good ability to detect geographical locations with a high number of diabetes predictions, such as New Jersey. With a high accuracy rate and easy model interpretation, the Decision Tree algorithm proved to be effective in detecting diabetes risk based on the analyzed patient attributes.

Keywords – algorithm; data mining; decision tree; diabetes; RapidMiner

I. PENDAHULUAN

Penyakit diabetes adalah kumpulan penyakit metabolik di mana kadar gula darah tinggi dan bertahan lama. Ini terjadi karena badan tidak dapat menghasilkan cukup hormon insulin, yang menyebabkan peningkatan kadar gula dalam darah [1]. Diabetes mellitus merupakan salah satu penyakit kronis paling umum di seluruh dunia, yang memengaruhi jutaan orang. Kondisi ini muncul ketika tubuh tidak mampu memproduksi insulin dalam jumlah cukup atau menggunakannya secara efektif, yang mengakibatkan tingginya kadar gula darah. Perawatan dan manajemen diabetes sangat penting untuk mencegah komplikasi serius seperti penyakit jantung, stroke, gagal ginjal, kebutaan, dan amputasi ekstremitas bawah.

Klasifikasi adalah proses pengembangan model atau fungsi yang menggambarkan

atau mengatur konsep atau serangkaian data untuk menilai serangkaian data yang terkait dengan suatu objek yang labelnya tidak diketahui. Tujuan dari proses klasifikasi ini adalah untuk menciptakan model yang dapat secara efektif menerapkan data ke berbagai kelas berdasarkan hukum atau fungsi tertentu [2].

Penelitian ini memanfaatkan algoritma decision tree dengan menggunakan software Rapid Miner. Decision tree, juga dikenal sebagai pohon keputusan, adalah algoritma yang paling umum digunakan untuk memecahkan masalah klasifikasi. Pohon keputusan terdiri dari beberapa komponen, antara lain daun, akar pohon, dan simpul dalam. Konsep entropi digunakan untuk menentukan atribut mana yang paling mungkin berubah. Pohon keputusan diperoleh dengan menggunakan algoritma klasifikasi untuk mengekstrak data [2].

Rapid Miner merupakan perangkat lunak data yang dikenal sebagai penambangan teks, pembelajaran mesin, pembelajaran dalam, dan analisis prediktif dalam lingkungan terintegrasi. Untuk dapat diproses lebih lanjut dan menghasilkan dataset yang lebih bersih, Rapid Miner harus mendapatkan data mentah dari web Kaggle. Sangat sulit untuk mengevaluasi kinerja secara menyeluruh karena volume data diabetes yang sangat besar. Oleh karena itu, untuk membantu proses bisnis, penulis memerlukan bantuan dari Rapid Miner [3].

Beberapa penelitian sebelumnya telah menunjukkan efektivitas algoritma Decision Tree dalam klasifikasi penyakit diabetes. Sebagai contoh, dalam studi yang dilakukan oleh Sari & Putri (2023), Decision Tree regression menjadi metode terbaik dalam memprediksi kemungkinan diabetes dibanding metode regresi lainnya, dengan akurasi di atas 90% [4]. Sementara itu, penelitian oleh Apriliah et al. (2021) membandingkan berbagai algoritma klasifikasi seperti Support Vector Machine, Naïve Bayes, dan Random Forest, dan menyimpulkan bahwa Random Forest memiliki akurasi tertinggi dalam memprediksi diabetes, namun Decision Tree tetap menunjukkan performa kompetitif dan memberikan hasil yang sangat mudah dipahami oleh pengguna non-teknis [5]. Penelitian lain yang memanfaatkan Decision Tree dalam bidang medis juga menunjukkan keandalan metode ini. Dalam studi oleh Pradana et al. (2018), kombinasi algoritma C4.5 dengan Binary Particle Swarm Optimization (BPSO) digunakan untuk klasifikasi data kanker, yang menunjukkan akurasi tinggi dan peningkatan performa signifikan saat fitur yang kurang relevan dihilangkan [6]. Hal ini mengindikasikan bahwa penggunaan Decision Tree tidak hanya terbatas pada satu domain penyakit, tetapi juga dapat diadaptasikan untuk berbagai kondisi medis lainnya, termasuk diabetes. Meskipun demikian, algoritma ini bukan tanpa kelemahan. Salah satu tantangan utama dari Decision Tree adalah kecenderungannya terhadap overfitting, terutama ketika jumlah data latih terbatas atau terdapat atribut-atribut yang tidak relevan. Untuk mengatasi masalah ini, berbagai teknik dapat digunakan seperti pruning, pemilihan atribut berbasis information gain atau Gini index, serta validasi silang untuk mengevaluasi kinerja model secara objektif.

Tujuan dari penelitian ini adalah untuk mengembangkan model klasifikasi menggunakan metode decision tree untuk memprediksi terjadinya diabetes berdasarkan berbagai faktor risiko yang ditemukan dalam dataset yang disediakan. Model ini diharapkan dapat membantu mendeteksi diabetes dengan lebih efektif dan mengobatinya.

II. TINJAUAN TEORI

1) Penyakit Diabetes

Diabetes melitus (DM) atau penyakit kencing manis merupakan suatu kelompok dengan karakteristik hiperglikemia yang terjadi karena kelainan sekresi insulin, kerja insulin atau keduanya [7]. Menurut penelitian Safitri dan Praba, Diabetes Mellitus adalah penyakit kronis yang terjadi akibat gangguan produksi atau fungsi insulin, yang menyebabkan peningkatan kadar gula darah dalam jangka panjang. Kondisi ini berisiko menimbulkan komplikasi serius seperti penyakit jantung, stroke, hingga gagal ginjal apabila tidak ditangani dengan baik [1].

2) Klasifikasi Dan Data Mining

Klasifikasi dalam data mining adalah proses mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan atribut-atribut yang ada [8]. Menurut penelitian Faid dan kawan-kawannya, Klasifikasi merupakan salah satu teknik dalam data mining yang bertujuan untuk memetakan data ke dalam kelompok-kelompok tertentu berdasarkan pola yang ditemukan dalam data historis. Model klasifikasi sangat penting untuk prediksi penyakit karena dapat mengelompokkan pasien berdasarkan risiko atau status kesehatannya secara otomatis [2].

3) Algoritma Decision Tree

Decision Tree adalah salah satu metode klasifikasi yang paling populer, karena mudah untuk diinterpretasi oleh manusia. Decision Tree adalah model prediksi menggunakan struktur pohon atau struktur berhieraki [9]. Menurut penelitian Pradana dan kawan-kawan, Decision Tree adalah salah satu metode klasifikasi yang sangat populer karena mudah dipahami dan diimplementasikan. Algoritma ini bekerja dengan membangun pohon keputusan berdasarkan atribut-atribut penting yang mempengaruhi hasil klasifikasi. Algoritma C4.5 adalah salah satu varian paling terkenal yang menggunakan nilai gain informasi untuk menentukan pembagian terbaik pada setiap simpul [6]

4) RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi [10]. Menurut penelitian Faid dan kawan-kawan, RapidMiner adalah sebuah platform perangkat lunak open-source yang digunakan untuk analisis data, pembelajaran mesin (machine learning), dan data mining. Platform ini menyediakan antarmuka visual berbasis drag-and-drop, sehingga memudahkan pengguna dalam merancang alur kerja (workflow) analisis data tanpa perlu menulis kode secara manual. RapidMiner mendukung berbagai tahap proses data mining seperti ekstraksi data, pra-pemrosesan, pemodelan, evaluasi, dan visualisasi hasil [2].

III. METODE

Adapun metode penelitian yang digunakan sebagai berikut:

1) Pengumpulan Dataset

Pada penelitian ini menggunakan dataset bersumber dari website kaggle.com, yang dapat di temukan dengan alamat website <https://www.kaggle.com/datasets/priyamchoksi/100000-diabetes-clinical-dataset>.

Atribut yang dipakai pada penelitian ini sebanyak 16 atribut dengan jumlah data sebanyak 100.000. Meliputi tahun, jenis kelamin, usia, lokasi, ras, hipertensi, penyakit

jantung, kebiasaan merokok, indeks massa tubuh (BMI), kadar HbA1c, dan kadar glukosa darah. Atribut terakhir dalam kumpulan data adalah indikator diabetes (0 atau 1).

2) Pre-processing Data

Pada pre-processing data dilakukan pembagian data atau split data. Pada tahap ini, 100.000 data diabetes dibagi menjadi data training dan data testing. Dimana presentase 70% untuk data training dan 30% untuk data testing. Data training bertindak sebagai pembentuk model dan data testing sebagai data pengujian.

3) Metode Klasifikasi

Penelitian ini menggunakan metode klasifikasi Pohon Keputusan atau Decision Tree untuk memprediksi indikasi diabetes. Decision Tree merupakan metode populer karena kemampuannya menangani data yang kompleks dan menyediakan model yang mudah ditafsirkan. Langkah-langkah utama metode klasifikasi decision tree ini adalah sebagai berikut:

a. Mengembangkan Model

Model Pohon Keputusan akan dikembangkan menggunakan data training (pelatihan). Model ini akan belajar dari berbagai atribut yang disertakan dalam dataset untuk menghasilkan prediksi status diabetes.

b. Pengujian Model

Model yang dikembangkan akan dievaluasi menggunakan data testing (pengujian). Data pengujian digunakan untuk menilai kinerja model dalam memprediksi status diabetes.

4) Evaluasi

Performa model klasifikasi akan dievaluasi menggunakan Confusion Matrix. Confusion Matrix adalah hasil evaluasi dari klasifikasi penambahan data yang ditampilkan dalam bentuk tabel. (Kudus, Ganesha, and Kudus 2020), yang akan menunjukkan akurasi, kesalahan klasifikasi, dan kesalahan relatif. Evaluasi ini meliputi:

- Accuracy (akurasi): Proporsi prediksi yang benar dibandingkan dengan jumlah total prediksi yang dibuat.
- Classification error (kesalahan klasifikasi): Proporsi prediksi yang salah dibandingkan dengan keseluruhan prediksi yang dibuat.
- Relative error (kesalahan relatif): Perbedaan antara kesalahan aktual dan kesalahan dasar, memberikan wawasan tentang seberapa baik model dibandingkan dengan prediksi yang acak.

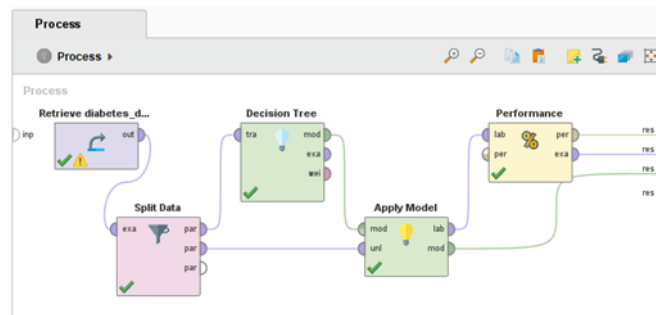
Evaluasi ini penting untuk memastikan bahwa model yang dikembangkan memiliki kinerja yang baik dan dapat digunakan untuk memprediksi timbulnya diabetes.

IV. HASIL DAN PEMBAHASAN

Pada penelitian ini dari dataset yang berjumlah 100.000, data dibagi menjadi dua dengan presentase 70% data training (pelatihan) didapatkan 70.000 data dan 30% data testing (pengujian) didapatkan 30.000 data untuk menguji kinerja pada model, memastikan bahwa model dapat membuat prediksi yang akurat. Atribut yang digunakan ada tahun, jenis

kelamin, usia, lokasi, ras, hipertensi, penyakit jantung, kebiasaan merokok, indeks masa tubuh (BMI), kadar HbA1c, kadar glukosa darah, dan indikator diabetes yang merupakan label untuk menentukan berapa banyak data yang mengalami penyakit diabetes.

Untuk menghasilkan aturan dan membangun pohon keputusan yang dapat menghitung jumlah orang yang menderita diabetes berdasarkan data, digunakan klasifikasi C4.5. Atribut dipilih sebagai akar dengan menghitung gain informasi. Untuk membangun pohon keputusan, atribut dengan nilai gain informasi tertinggi dipilih sebagai akar. Proses penerapan dan pengujian dataset dilakukan dengan Rapid Miner. Berikut ini adalah proses penelitian klasifikasi diabetes menggunakan RapidMiner 2024.0.1:



Gambar 1. Proses Klasifikasi Menggunakan RapidMiner 2024.0.1

Dalam proses ini, Rapid Miner digunakan untuk memprediksi diabetes. Pertama, data tentang diabetes diimpor ke dalamnya. Data kemudian dibagi menjadi dua bagian: data training (pelatihan) dan data testing (pengujian). Data pelatihan digunakan untuk melatih model, sedangkan data testing (pengujian) digunakan untuk mengukur kinerja model. Kemudian dibangun model prediksi dengan menggunakan algoritma Decision Tree yang menggunakan atribut-atribut data. Setelah model dilatih, model tersebut diterapkan pada data pengujian untuk menghasilkan prediksi. Menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan skor F1, kinerja model dinilai dengan membandingkan hasil prediksi dengan label sebenarnya pada data pengujian. Setelah diuji maka dapat dilihat hasilnya:

1) Evaluasi dengan confusion matrix

Hasil dari evaluasi dengan confusion matrix dilakukan dua kali yaitu evaluasi hasil klasifikasi terhadap data pelatihan dan data pengujian. Dari data tersebut menunjukkan hasil akurasi sebesar 97.16% dan Dari accuracy evaluasi tersebut terdapat juga klasifikasi error sebesar 2.84%. berikut tabel evaluasi hasil nilai akurasi/accuracy dengan Confusion Matrix:

accuracy: 97.16%

	true 0	true 1	class precision
pred. 0	27360	762	97.29%
pred. 1	90	1788	95.21%
class recall	99.67%	70.12%	

Gambar 2. Hasil Akurasi

classification_error: 2.84%

	true 0	true 1	class precision
pred. 0	27360	762	97.29%
pred. 1	90	1788	95.21%
class recall	99.67%	70.12%	

Gambar 3. Hasil Klasifikasi Error

Confusion Matrix:

- a. True Positive (TP) = 1788 pasien yang benar-benar menderita diabetes
- b. True Negative (TN) = 27360 pasien yang tidak menderita diabetes
- c. False Positive (FP) = 762 pasien sehat sebagai pasien diabetes
- d. False Negative (FN) = 90 pasien diabetes sebagai pasien sehat

Presisi:

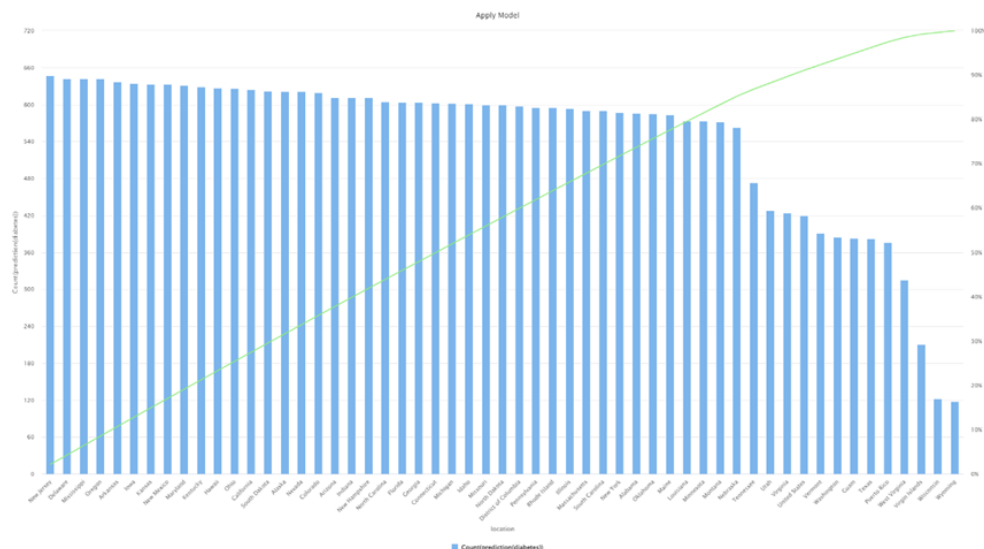
- a. Presisi Kelas 0 (Tidak Diabetes) = 97.29%
- b. Presisi Kelas 1 (Diabetes) = 95.21%

Recall:

- a. Recall Kelas 0 (Tidak Diabetes) = 99.67% (hampir semua pasien yang sebenarnya tidak menderita diabetes)
- b. Recall Kelas 1 (Diabetes) = 70.12% (pasien yang sebenarnya menderita diabetes)

2) Prediksi Diabetes Berdasarkan Lokasi

Evalusi pada data diabetes berdasar lokasi. Berikut diagram yang menunjukkan lokasi berapa yang mengalami diabetes:



Gambar 4. Grafik Lokasi

Grafik pada batang menunjukkan jumlah prediksi kasus diabetes (Positif) di berbagai lokasi, garis hijau yang melintang pada grafik menunjukkan tren umum dari jumlah kasus prediksi diabetes, sumbu x menampilkan nama-nama lokasi, sedangkan sumbu y menunjukan jumlah kasus prediksi diabetes di setiap lokasi. Berdasarkan diagram pada gambar, menunjukkan lokasi dengan jumlah prediksi diabetes tertinggi adalah New Jersey, sedangkan lokasi dengan jumlah prediksi diabetes terendah adalah Wyoming. Ini menunjukkan bahwa fokus pada beberapa lokasi utama dapat secara signifikan mempengaruhi total jumlah prediksi diabetes, mungkin akan berguna untuk intervensi kesehatan masyarakat yang lebih efektif

3) Decision Tree

Tree

```

hbA1c_level > 6.700: 1 {0=0, 1=2681}
hbA1c_level ≤ 6.700
|   blood_glucose_level > 210: 1 {0=0, 1=1258}
|   blood_glucose_level ≤ 210
|   |   hbA1c_level > 5.350
|   |   |   age > 42.500
|   |   |   |   blood_glucose_level > 113
|   |   |   |   |   bmi > 30.655
|   |   |   |   |   |   age > 54.500
|   |   |   |   |   |   |   location = Alabama
|   |   |   |   |   |   |   |   age > 78: 1 {0=0, 1=2}
|   |   |   |   |   |   |   |   age ≤ 78: 0 {0=35, 1=10}
|   |   |   |   |   |   |   |   location = Alaska
|   |   |   |   |   |   |   |   smoking_history = No Info: 0 {0=12, 1=0}
|   |   |   |   |   |   |   |   smoking_history = current: 0 {0=3, 1=0}
|   |   |   |   |   |   |   |   smoking_history = ever: 1 {0=1, 1=2}
|   |   |   |   |   |   |   |   smoking_history = former: 0 {0=9, 1=3}
|   |   |   |   |   |   |   |   smoking_history = never: 0 {0=13, 1=3}
|   |   |   |   |   |   |   |   smoking_history = not current: 0 {0=2, 1=1}
|   |   |   |   |   |   |   |   location = Arizona: 0 {0=26, 1=10}
|   |   |   |   |   |   |   |   location = Arkansas: 0 {0=39, 1=12}
|   |   |   |   |   |   |   |   location = California
|   |   |   |   |   |   |   |   |   smoking_history = No Info: 1 {0=4, 1=5}
|   |   |   |   |   |   |   |   |   smoking_history = current: 0 {0=4, 1=2}
|   |   |   |   |   |   |   |   |   smoking_history = former: 0 {0=7, 1=1}
|   |   |   |   |   |   |   |   |   smoking_history = never: 0 {0=15, 1=3}
|   |   |   |   |   |   |   |   |   smoking_history = not current: 0 {0=4, 1=2}
|   |   |   |   |   |   |   |   |   location = Colorado: 0 {0=32, 1=9}
|   |   |   |   |   |   |   |   |   location = Connecticut: 0 {0=42, 1=12}
|   |   |   |   |   |   |   |   |   location = Delaware

```

Gambar 5. Deskripsi Tree

Pohon keputusan ini menganalisis data kesehatan individu berdasarkan berbagai faktor, termasuk level hbA1c, level glukosa darah, usia, BMI, lokasi, riwayat merokok, jenis kelamin, penyakit jantung, hipertensi, ras, dan tahun. Pertama, analisis memeriksa level hbA1c, dengan nilai di atas 6.7 yang menunjukkan klasifikasi positif diabetes. Selanjutnya, analisis dilanjutkan dengan memeriksa level glukosa darah. Jika level hbA1c di bawah Beberapa percabangan menggunakan riwayat merokok dan jenis kelamin untuk memberikan klasifikasi tambahan. Faktor BMI juga berperan, terutama bagi mereka dengan nilai BMI di atas 30.655, dengan pembagian lebih lanjut berdasarkan lokasi geografis. Riwayat merokok dan jenis kelamin juga digunakan dalam beberapa percabangan untuk memperinci klasifikasi lebih lanjut. Setiap node dalam pohon ini memberikan keputusan berdasarkan kombinasi berbagai faktor, dengan klasifikasi akhir menunjukkan kemungkinan kondisi diabetes atau kesehatan lainnya berdasarkan data input.

V. KESIMPULAN DAN SARAN**a. Kesimpulan**

Penggunaan Rapid Miner dan Algoritma Decision Tree dapat membantu dalam memprediksi diabetes dengan tingkat akurasi yang tinggi, serta memungkinkan identifikasi lokasi-lokasi dengan prevalensi diabetes yang tinggi untuk intervensi kesehatan yang lebih efektif. Model Klasifikasi untuk mendeteksi diabetes menunjukkan kinerja yang sangat baik dengan akurasi sebesar 97,16% ini berarti semua prediksi yang dibuat oleh model adalah benar, namun terdapat juga eror

klasifikasi sebesar 2.84% yang berarti prediksi dari model adalah salah. Berdasarkan hasil klasifikasi, lokasi dengan jumlah prediksi diabetes tertinggi adalah New Jersey, sedangkan lokasi dengan jumlah prediksi diabetes terendah adalah Wyoming.

b. Saran

Menjelaskan model lanjutan menggunakan algoritma lain seperti Random Forest atau Neural Networks untuk meningkatkan akurasi dapat memperluas penelitian ini. Dataset harus diperkaya dengan lebih banyak data dan fitur baru. Untuk mencegah overfitting, diperlukan metode pra-pemrosesan yang lebih canggih dan validasi silang yang ketat.

VI. DAFTAR PUSTAKA

- [1] M. Safitri dan A. D. Praba, "PREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN ALGORITMA C4.5," *JIKA J. Inform.*, vol. 8, no. 1, hlm. 74, Jan 2024, doi: 10.31000/jika.v8i1.9998.
- [2] M. Faid, M. Jasri, dan T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, hlm. 11–16, Jun 2019, doi: 10.34148/teknika.v8i1.95.
- [3] L.N. Salsabila, M. R. Dwi Pangga, S. M. Yasser, N. A. Riyani, S. Aminah, dan W. Wahyunengsih, "Application of Naïve Bayes Algorithm for Diabetes Prediction," *Unisda J. Math. Comput. Sci. UJMC*, vol. 10, no. 1, hlm. 56–68, Jun 2024, doi: 10.52166/ujmc.v10i1.6886.
- [4] A. Sari dan D. U. Putri, "PENERAPAN REGRESI LINEAR BERGANDA UNTUK MEMPREDIKSI DIABETES SECARA DINI," *J. Tek.*, vol. 3, no. 1, hlm. 32, Feb 2023, doi: 10.54314/teknisi.v3i1.1255.
- [5] W. Apriliah, I. Kurniawan, M. Baydhowi, dan T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *SISTEMASI*, vol. 10, no. 1, hlm. 163, Jan 2021, doi: 10.32520/stmsi.v10i1.1129.
- [6] A. C. Pradana, A. Adiwijaya, dan A. Aditsania, "Implementasi Algoritma Binary Particle Swarm Optimization (BPSO) dan C4.5 Decision Tree untuk Deteksi Kanker Berdasarkan Klasifikasi Microarray Data," 2018. [Daring]. Tersedia pada: <https://api.semanticscholar.org/CorpusID:198320631>
- [7] Y. L. Tinungki, J. S. H. Hinonaung, dan E. Efitra, *DETEKSI DINI PENYAKIT DIABETES MELLITUS (DM) DAN OBAT TRADISIONAL DM PADA LANSIA DI KABUPATEN KEPULAUAN SANGIHE*. PT. Sonpedia Publishing Indonesia, 2023. [Daring]. Tersedia pada: <https://books.google.co.id/books?id=rHPkEAAQBAJ>
- [8] R. M. Sari dan Y. A. M, *Klasifikasi Data Mining*. Serasi Media Teknologi, 2024. [Daring]. Tersedia pada: <https://books.google.co.id/books?id=xTwXEQAQBAJ>
- [9] D. Jollyta, A. Hajjah, E. Haerani, dan M. Siddik, *Algoritma Klasifikasi untuk Pemula Solusi Python dan RapidMiner*. Deepublish, 2023. [Daring]. Tersedia pada: <https://books.google.co.id/books?id=y84TEQAQBAJ>
- [10] S. T. M. K. Yahya, *Data Mining*. CV Jejak (Jejak Publisher), 2022. [Daring]. Tersedia pada: <https://books.google.co.id/books?id=0J2mEAAQBAJ>