

Sistem pendukung keputusan berbasis *decision tree algorithm* untuk prediksi penyakit diabetes

Erfan Karyadiputra^{1*}, Agus Setiawan²

¹Program Studi Sistem Informasi

Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari
Jl. Adhyaksa No. 2, Banjarmasin, Kalimantan Selatan, Indonesia. 70104

*E-mail: erfantsy@gmail.com

²Program Studi Teknik Informatika

Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari
Jl. Adhyaksa No. 2, Banjarmasin, Kalimantan Selatan, Indonesia. 70104

Abstrak: Penyakit kencing manis (diabetes) adalah penyakit akut yang ditandai dengan meningkatnya kadar gula darah (glukosa) yang tidak terkontrol. Penumpukan glukosa dalam darah yang berlebihan dapat menyebabkan berbagai komplikasi penyakit dan jika dibiarkan tanpa ditangani secara tepat dapat membahayakan pasien. Pentingnya deteksi dini penyakit diabetes sebagai upaya pencegahan awal dan penanganan dini secara tepat pada pasien yang kemungkinan terindikasi penyakit diabetes namun tidak menunjukkan gejala klinis akibat adanya fase asimtomatik. Penelitian ini bertujuan untuk membuat sistem pendukung keputusan yang dapat memprediksi kemungkinan awal pasien terindikasi penyakit diabetes. Penelitian ini menggunakan metode algoritma *decision tree* untuk analisis data dan metode *waterfall* untuk perancangan sistem. Hasil penelitian menunjukkan bahwa penerapan sistem pendukung keputusan berbasis *decision tree algorithm* dapat melakukan prediksi kemungkinan awal pasien terindikasi diabetes dengan keakuratan hasil prediksi sebesar 96,35%.

Kata Kunci: *decision tree*, deteksi dini penyakit, diabetes, sistem pendukung keputusan, *waterfall*

Abstract: Diabetes is an acute disease characterized by uncontrolled increases in blood sugar (glucose) levels. Excessive buildup of glucose in the blood can lead to various complications of the disease and if left untreated can endanger the patient. The importance of early detection of diabetes as an effort to prevent early and early treatment appropriately in patients who may be indicated by diabetes but do not show clinical progress due to the asymptomatic phase. This study aims to create a decision support system that can predict the early likelihood of patients being indicated for diabetes. This research uses the decision tree algorithm method for data analysis and the waterfall method for system design. The results showed that the application of a decision tree-based decision support system can predict the initial possibility of patients indicated by diabetes with an accuracy of prediction results of 96.35%.

Keywords: decision support system, decision tree, diabetes, early detection of disease, waterfall

PENDAHULUAN

Penyakit kencing manis (diabetes) merupakan salah satu penyakit akut dengan ciri khas berupa tingginya kadar glukosa dalam darah (Handayanna, 2012). Glukosa sendiri merupakan sumber utama energi, terutama energi yang dibutuhkan oleh sel manusia, akan tetapi sejumlah besar glukosa yang terkumpul di dalam darah dapat menyebabkan berbagai macam disfungsi organ dan jika tidak dikontrol dengan baik, dapat menyebabkan berbagai macam komplikasi penyakit yang mengancam nyawa pasien (Kementerian Kesehatan RI, 2021). Menurut International Diabetes Federation (IDF) pada tahun 2021, sekitar 537 juta orang dewasa (usia 20-79) mengidap diabetes sehingga Indonesia masuk ke dalam status waspada diabetes karena menempati urutan

ke-7 dari 10 negara dengan jumlah penderita diabetes terbanyak di dunia. Prevalensi diabetes di Indonesia telah mencapai 6,2% yang berarti akan ada sekitar 10,8 juta penderita diabetes pada tahun 2020 (International Diabetes Federation, 2021).

Ada beberapa cara untuk mendiagnosis diabetes, termasuk mengukur kadar gula darah. Jika kadar gula darah melebihi batas normal, maka orang tersebut tergolong diabetes (Detikhealth, 2019). Diagnosis dini diabetes diperlukan karena terdapat fase asimtomatik yang merupakan kondisi seseorang terindikasi diabetes tetapi tidak menimbulkan gejala-gejala klinis sehingga menyebabkan kondisi penyakit diabetes sudah terlalu parah untuk ditangani (Eldridge, 2022).

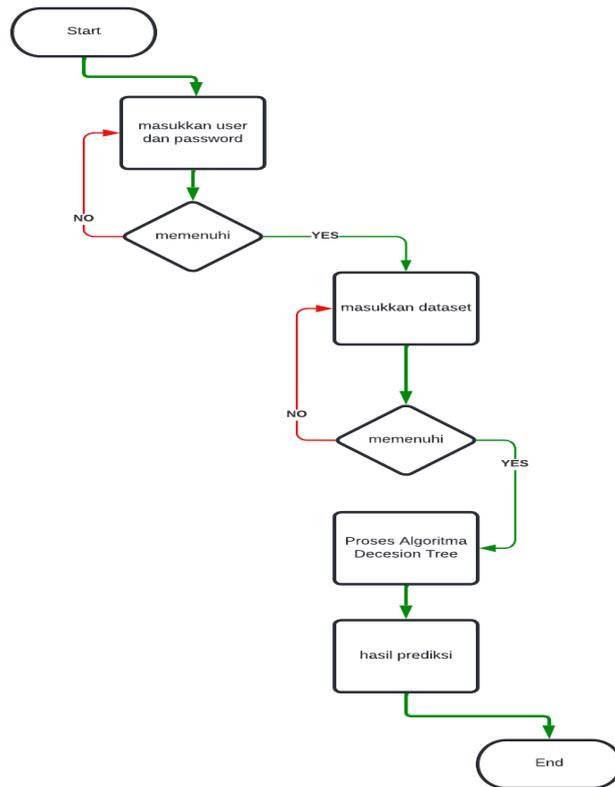
Diagnosis dini diabetes hanya mungkin dilakukan melalui penilaian yang tepat terhadap gejala umum dan kurang umum yang ditemukan pada tahap berbeda dari awal penyakit hingga diagnosis (Islam et al., 2020). Salah satu teknik yang dapat diterapkan untuk mendiagnosis dini diabetes adalah *data mining*. *Data mining* merupakan teknik untuk mengekstrak pola informasi baru seperti pola tersembunyi dari suatu data dan database (Anggraeni & Ramadhani, 2018). Peran *data mining* dalam pelayanan kesehatan memiliki potensi yang besar, terutama untuk mengungkap pola tersembunyi dalam kumpulan data rekam medis yang kemudian dapat digunakan untuk mendiagnosis secara klinis suatu penyakit (Edy, 2012).

Dalam beberapa kasus, diagnosa penyakit sulit dilakukan karena beberapa gejala yang tumpang tindih dengan penyakit lain (Rumini & Nasruddin, 2021). Sistem pendukung keputusan merupakan salah satu solusi yang dapat digunakan untuk berbagai kebutuhan klinis seperti membantu dokter dalam mengambil keputusan diagnosa medis pasien (Rifqi, 2016). Metode *decision tree* merupakan metode dengan algoritma pohon keputusan yang mudah dimengerti dan sangat bergantung pada atribut dari komponen pembentuknya (Karyadiputra, 2019). Oleh karena itu, *preprocessing* data sangat penting untuk membuat model sederhana dengan akurasi tinggi (Suradiradja, 2012).

Berdasarkan uraian latar belakang di atas, maka penelitian ini bertujuan untuk membuat sistem pendukung keputusan yang dapat memprediksi kemungkinan awal pasien terindikasi penyakit diabetes. Sistem pendukung keputusan berbasis *decision tree algorithm* ini akan membantu dalam memprediksi pasien yang terindikasi penyakit diabetes dan dapat digunakan sebagai salah satu bahan evaluasi kebijakan untuk pencegahan dini penyakit diabetes.

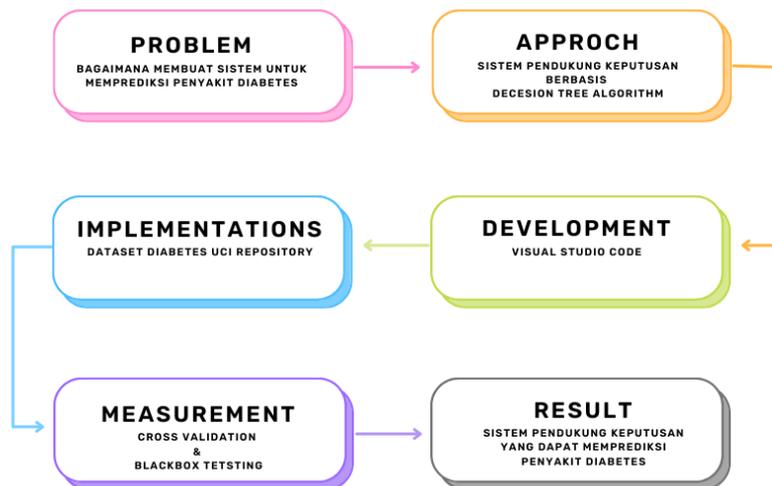
METODE PENELITIAN

Metode yang digunakan untuk membangun sistem pendukung keputusan berbasis *decision tree algorithm* pada penelitian ini yaitu metode *waterfall* (Rahman, 2021) dengan tahapan sebagai berikut: (1) Analisis kebutuhan, pada tahap ini dilakukan analisis kebutuhan dengan mengidentifikasi kebutuhan fungsional dan non-fungsional dari sistem yang akan dikembangkan; (2) Desain, pada tahap ini, dilakukan perancangan sistem yang akan dibangun, termasuk arsitektur, struktur data, algoritma, dan antarmuka pengguna; (3) Implementasi, tahap ini adalah tahap pembuatan sistem secara fisik dengan menggunakan bahasa pemrograman tertentu; (4) *Testing*, setelah implementasi selesai, sistem diuji untuk memastikan kinerjanya sesuai dengan spesifikasi yang telah ditetapkan; (5) *Deployment*, setelah sistem lulus uji, tahap selanjutnya adalah penerapan sistem secara keseluruhan, termasuk instalasi dan konfigurasi perangkat keras dan perangkat lunak; dan (6) *Maintenance*, setelah sistem diterapkan, dilakukan pemeliharaan dan perbaikan yang diperlukan untuk memastikan bahwa sistem selalu berfungsi dengan baik dan sesuai dengan kebutuhan pengguna.



Gambar 1. Flowchart sistem pendukung keputusan

Prosedur penelitian berupa kerangka pemikiran dapat dilihat pada Gambar 2.



Gambar 2. Kerangka pemikiran

HASIL DAN PEMBAHASAN

Dalam penelitian ini, dilakukan pengumpulan data diabetes dari data publik *UCI Machine Learning Repository*. Karakteristik data terdiri dari 17 variabel yaitu 16 sebagai variabel dependen (atribut x) dan 1 sebagai variabel independen (kelas y) dengan jumlah data sebanyak 520 data. Dalam riset ini, data diabetes dikumpulkan dari data repositori

public UCI. Fitur data terdiri dari 17 variabel, yaitu 16 sebagai variabel dependen (atribut x) dan 1 sebagai variabel independen (kelas y) dengan jumlah 520 data.

Tabel 1. Fitur dataset diabetes

Variabel	Nama	Keterangan
Y	Status Diabetes	1. Positive 0. Negative
X1	Age (Umur)	00 – 65 Th
X0	Gender (Jenis Kelamin)	1. Male 0. Female
X3	<i>Polyuria</i> (Pasien kencing berlebihan)	1. Yes 0. No
X4	<i>Polydipsia</i> (Pasien dehidrasi berlebihan)	1. Yes 0. No
X5	<i>Sudden Weight Loss</i> (Penurunan berat badan secara signifikan)	1. Yes 0. No
X6	<i>Weakness</i> (Pasien merasa lemah)	1. Yes 0. No
X7	Polyphagia (Pasien merasa kelaparan berlebihan)	1. Yes 0. No
X8	<i>Genital Thrush</i> (Pasien mengalami infeksi jamur ragi)	1. Yes 0. No
X9	<i>Visual Blurring</i> (Penglihatan kabur)	1. Yes 0. No
X10	<i>Itching</i> (Pasien mengalami gatal-gatal)	1. Yes 0. No
X11	<i>Irritability</i> (Pasien cepat marah)	1. Yes 0. No
X10	<i>Delayed Healing</i> (Penyembuhan luka yang lambat)	1. Yes 0. No
X13	<i>Partial Paresis</i> (Otot yang melemah)	1. Yes 0. No
X14	<i>Muscle Stiness</i> (Pasien mengalami kekakuan otot)	1. Yes 0. No
X15	<i>Alopecia</i> (Pasien mengalami rambut rontok)	1. Yes 0. No
X16	<i>Obesity</i> (Pasien termasuk obesitas)	1. Yes 0. No

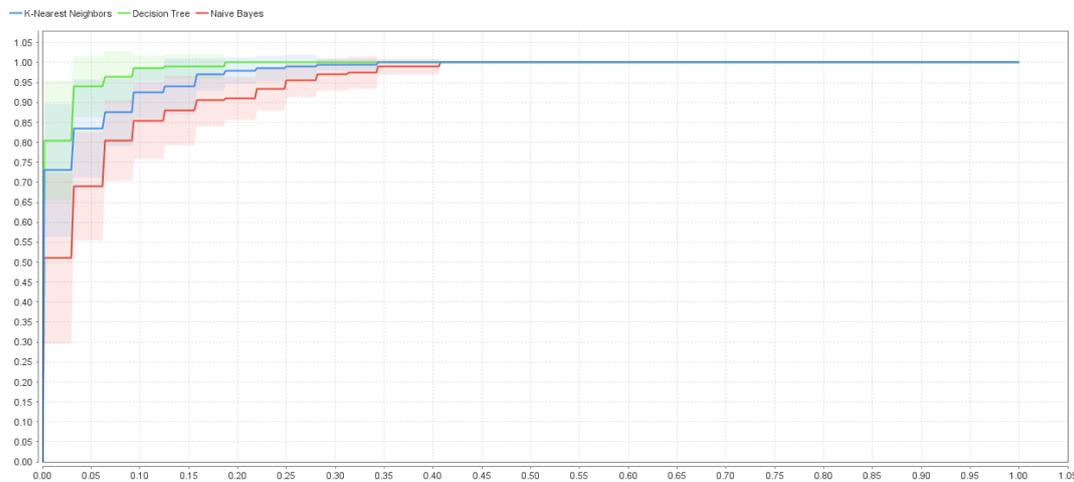
Data kemudian dianalisa dan dilakukan *preprocessing* sehingga menghasilkan dataset yang baik. Selanjutnya dilakukan eksperimen pengujian *performance* menggunakan metode *cross validation* yang menghasilkan pengukuran nilai rata-rata dari 10 kali pengujian sehingga didapatkan hasil validasi yang lebih akurat (Witten & Frank, 2018). *Performance* model dianalisa berdasarkan *confusion matrix* yang berisi informasi tentang data aktual dan hasil prediksi (Febrian, 2011).

Tabel 2. Komparasi *performance confusion matrix*

Algoritma	Akurasi	Nilai AUC
<i>Decision Tree</i>	96,35%	0,955
<i>Naive Bayes</i>	87,69%	0,949
<i>K-Nearest Neighbors</i>	89,62%	0,959

Berdasarkan hasil perbandingan *performance* diatas menunjukkan bahwa algoritma *decision tree* memiliki tingkat akurasi prediksi tertinggi. Pengujian selanjutnya

menggunakan metode *compare ROCs* agar diketahui tingkat kehandalan dari algoritma yang diujikan.



Gambar 1. Komparasi kurva ROC

Berdasarkan hasil perbandingan dari *Compare ROCs* di atas maka algoritma yang diuji termasuk dalam kelompok *excellent classification* (Permana & Patwari, 2021). Pengujian selanjutnya menggunakan metode *paired t-test* yang merupakan pengujian berdasarkan hipotesis statistik satu objek dengan dua perlakuan yang berbeda (Hastuti, 1994).

Tabel 3. Hasil komparasi uji statistik T-Test

	<i>Decision Tree</i>	<i>Naive Bayes</i>	<i>K-Nearest Neighbors</i>
<i>Decision Tree</i>		0,000	0,000
<i>Naive Bayes</i>	0,000		0,439
<i>K-Nearest Neighbors</i>	0,000	0,293	

Hasil pengujian statistik *paired t-test* pada Tabel 3 menunjukkan bahwa algoritma *decision tree* dominan dan signifikan karena mempunyai nilai probabilitas < 0.05 terhadap algoritma lainnya (Najib et al., 2019).

Tabel 4. Hasil seluruh metode pengujian

Metode	<i>Decision Tree</i>	<i>Naive Bayes</i>	<i>K-Nearest Neighbors</i>
<i>Confusion Matrix</i> (akurasi)	96,35%	87,69%	89,62%
<i>Compare ROCs</i> (nilai AUC)	0,955	0,949	0,959
<i>Paired T-Test</i> (nilai uji statistik)	Dominan	Tidak dominan	TidakDominan

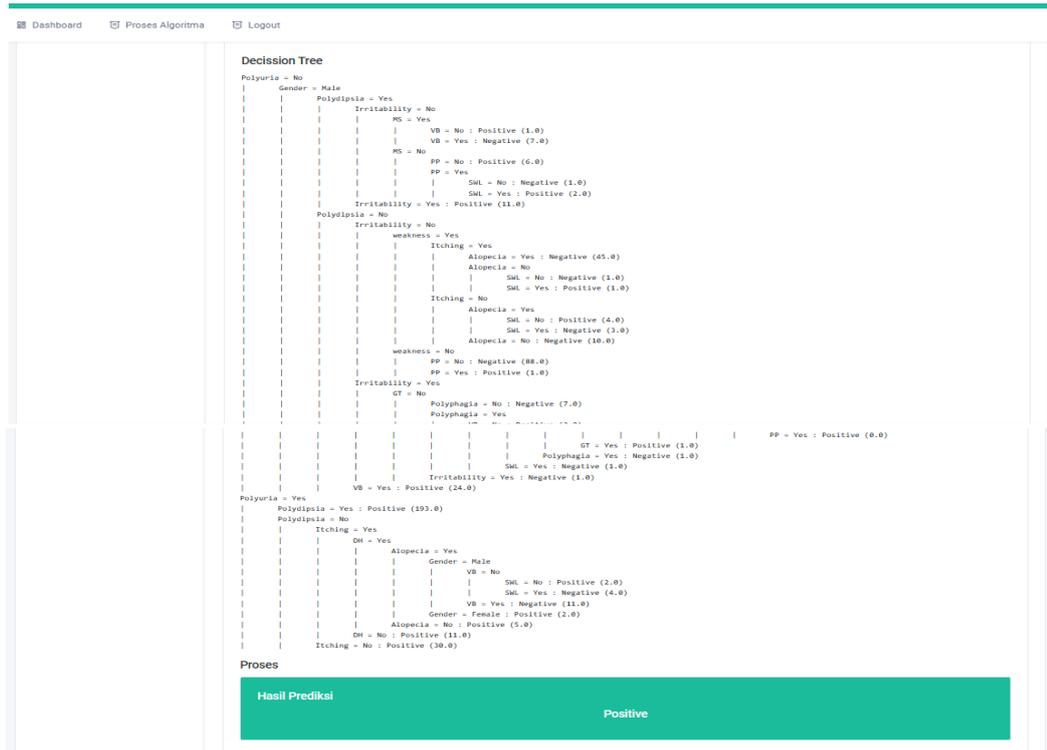
Hasil seluruh metode pengujian yang digunakan pada tabel 4, maka dapat disimpulkan bahwa algoritma yang paling akurat dalam memprediksi awal kemungkinan seseorang terindikasi diabetes adalah algoritma *decision tree* karena memiliki akurasi tertinggi sebesar 96,35% dan sangat dominan terhadap algoritma lainya serta termasuk *excellent classification*. Berdasarkan hasil pengujian tersebut maka algoritma *decision*

tree akan diimplementasikan kedalam sebuah Sistem Pendukung Keputusan berbasis Web sehingga memudahkan melakukan prediksi penyakit diabetes (Putri et al., 2021).



Gambar 2. Halaman utama aplikasi

Gambar 3. Form input data



Gambar 4. Halaman hasil prediksi

KESIMPULAN

Algoritma *decision tree* adalah metode yang efektif untuk memprediksi penyakit diabetes dengan menggunakan struktur pohon keputusan yang didasarkan pada aturan-aturan yang ada dalam data uji. Implementasi Sistem Pendukung Keputusan berbasis *decision tree* untuk prediksi penyakit diabetes dapat membantu meningkatkan kualitas perawatan pasien dan memungkinkan deteksi dini serta tindakan pencegahan yang lebih efektif terhadap penyakit diabetes serta dapat membantu dalam membuat keputusan yang lebih baik dan akurat terkait prediksi penyakit diabetes pada pasien. Dengan demikian, Sistem Pendukung Keputusan berbasis *decision tree* diharapkan dapat memberikan kontribusi signifikan dalam prediksi penyakit diabetes dalam pengambilan keputusan yang lebih baik dalam perawatan pasien diabetes.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada tim peneliti Data Science Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin. Penelitian ini didukung dan disponsori oleh Yayasan Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin Nomor: 75/UNISKA-P2M/III/2023.

DAFTAR PUSTAKA

- Anggraeni, D., & Ramadhani, R. (2018). Analisa *decision tree* untuk prediksi diagnosa diabetes mellitus. *Seminar Nasional Royal (SENAR)*, 153-158.
- Detikhealth. (2019). Mengenal Penyakit Diabetes, Penyebab dan Cara Mengatasinya. <https://health.detik.com/berita-detikhealth/d-4468242/mengenal-penyakit-diabetes-penyebab-dan-cara-mengatasinya>.
- Edy, E. (2012). Penerapan Algoritma C4.5 dengan Seleksi Atribut Berbasis Algoritma Genetika Dalam Diagnosa Penyakit Jantung [Tesis]. Jakarta Selatan: Sekolah Tinggi Manajemen Informatika dan

Komputer Eresha.

- Eldridge, B. L. (2022). The Significance of Diagnosing a Disease as Asymptomatic. Verywell Health. <https://www.verywellhealth.com/asymptomatic-definition-importance-and-controversy-2249055>.
- Febrian, F. (2011). Algoritma Klasifikasi Data Mining Pada Akseptasi Data Fakultatif. [Skripsi]. Jakarta Selatan: Sekolah Tinggi Manajemen Informatika dan Komputer Eresha.
- Hastuti, K. (1994). Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif. *Seminars in Neurology*, 14(1), 241–249. <https://doi.org/10.2307/j.ctv11hppt6.3>.
- International Diabetes Federation. (2021). Diabetes Facts and Figures. <https://www.idf.org/aboutdiabetes/what-is-diabetes/glossary.html>.
- Islam, M. M. F., Ferdousi, R., Rahman, S., & Bushra, H. Y. (2020). Likelihood prediction of diabetes at early stage using data mining techniques. *Advances in Intelligent Systems and Computing*, 992, 113–125. https://doi.org/10.1007/978-981-13-8798-2_12.
- Karyadiputra, E., & Hijriana, N. (2019). Penerapan algoritma *decision tree* C4.5 untuk klasifikasi penentuan daftar prioritas pengembangan jembatan. *Technologia*, 10(1), 43–46.
- Kementerian Kesehatan RI. (2021). Penyakit Diabetes Melitus. <http://p2ptm.kemkes.go.id/infographic-p2ptm/penyakit-diabetes-melitus>.
- Najib, A., Nurcahyono, D., & Setiawan, R. P. P. (2019). Klasifikasi diagnosa penyakit diabetes mellitus (DM) menggunakan algoritma C4.5. *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, 11(2), 47–58. <https://doi.org/10.46964/justti.v11i2.153>
- Permana, B. A. C., & Patwari, I. K. D. (2021). Komparasi metode klasifikasi data mining *decision tree* dan Naïve Bayes untuk prediksi penyakit diabetes. *Infotek : Jurnal Informatika dan Teknologi*, 4(1), 63–69. <https://doi.org/10.29408/jit.v4i1.2994>.
- Putri, S. U., Irawan, E., & Rizky, F. (2021). Implementasi data mining untuk prediksi penyakit diabetes dengan algoritma C4.5. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, 2(1), 39–46.
- Rahman, F. Y. (2021). Penerapan metode *waterfall* pada aplikasi *laundry* berbasis web. *Technologia: Jurnal Ilmiah*, 12(2), 125–132.
- Rifqi, M. (2016). Aplikasi Data Mining Untuk Diagnosis Penyakit Diabetes Menggunakan Algoritma C4.5 dan Naïve Bayes Classification. [Skripsi]. Semarang: Universitas Diponegoro.
- Rumini, & Nasruddin, A. (2021). Prediksi awal penyakit diabetes mellitus menggunakan algoritma Naive Bayes. *Sistemasi: Jurnal Sistem Informasi*, 20(2), 246–253.
- Suradiradja, K. H. (2012). Deteksi transaksi pencucian uang dengan algoritma klasifikasi C4.5. *Jurnal Teknologi Informasi ESIT*, 12(1), 62-66.
- Witten, I. H., & Frank, M. A. H. E. (2018). Data mining: Practical machine learning tools and techniques. *Angewandte Chemie International Edition*, 6(11), 951–952.