

MODEL KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE DECISION TREE

Eviyanti¹⁾, Fauziah ²⁾, Muhammad Kahfi Aulia ³⁾, Fakhru Zikri⁴⁾

^{1,3,4}Prodi Informatika Medis Fakultas Kesehatan Teknologi dan Sains Universitas Bumi Persada

*Correspondence : eviyanti@bumipersada.ac.id¹⁾

ABSTRAK:

Latar Belakang: Diabetes melitus merupakan gangguan metabolik kronis yang ditandai dengan kondisi hiperglikemia akibat defisiensi sekresi insulin, resistensi insulin, atau kombinasi keduanya. Sebagai salah satu penyakit tidak menular yang paling berbahaya secara global, diabetes memerlukan pendeteksian dini yang akurat. Klasifikasi tepat terhadap hasil tes laboratorium pasien (positif atau negatif diabetes) sangat penting untuk menentukan intervensi medis yang sesuai. Penelitian ini mengembangkan model klasifikasi berbasis Decision Tree C4.5 guna meningkatkan akurasi deteksi diabetes. Tujuan Penelitian: Studi ini bertujuan untuk membangun model klasifikasi status diabetes berdasarkan hasil laboratorium dan mengevaluasi model menggunakan berbagai metrik evaluasi. Metode Penelitian: Penelitian menggunakan dataset dari platform Kaggle yang terdiri dari 520 sampel dengan 17 variabel prediktor meliputi: faktor demografi (Age, Gender), gejala klinis (Polyuria, Polydipsia, sudden weight loss), tanda-tanda fisik (weakness, delayed healing, Alopecia), dan variabel target (class). Hasil Penelitian: Decision Tree C4.5 menunjukkan performa yang mengungguli dengan hasil akurasi: 91.35%, presisi: 93.55%, recall: 92.06%, spesifisitas: 90.24%, dan F1-score: 92.80%. Kesimpulan: Berdasarkan hasil analisis, dapat disimpulkan bahwa Decision Tree C4.5 merupakan algoritma yang efektif untuk klasifikasi diabetes, model yang dikembangkan mencapai akurasi di atas 90% pada semua metrik evaluasi, dan sistem ini berpotensi sebagai alat pendukung diagnosis diabetes di fasilitas kesehatan.

Kata kunci: Diabetes Melitus, Klasifikasi, Decision Tree C4.5, Analisis Prediktif, Metrik Evaluasi

ABSTRACT:

*Background: Diabetes mellitus is a chronic metabolic disorder characterized by hyperglycemia due to insulin secretion deficiency, insulin resistance, or a combination of both. As one of the most dangerous non-communicable diseases globally, diabetes requires early and accurate detection. Accurate classification of patient laboratory test results (positive or negative diabetes) is essential to determine appropriate medical intervention. This study developed a classification model based on Decision Tree C4.5 to improve the accuracy of diabetes detection. Research Objective: This study aims to build a diabetes status classification model based on laboratory results and evaluate the model using various evaluation metrics. Research Methods: The study used a dataset from the Kaggle platform consisting of 520 samples with 17 predictor variables including: demographic factors (Age, Gender), clinical symptoms (Polyuria, Polydipsia, sudden weight loss), physical signs (weakness, delayed healing, Alopecia), and target variables (class). Research Results: Decision Tree C4.5 showed superior performance with accuracy results: 91.35%, precision: 93.55%, recall: 92.06%, specificity: 90.24%, and F1-score: 92.80%. Conclusion: Based on the analysis results, it can be concluded that Decision Tree C4.5 is an effective algorithm for diabetes classification, the developed model achieved accuracy above 90% on all evaluation metrics, and this system has the potential as a supporting tool for diabetes diagnosis in health facilities.*Key Word: *Migraine, Neurology, Classification, Naive Bayes, Prediction*

Keyword : Diabetes Mellitus, Classification, Decision Tree C4.5, Predictive Analysis, Evaluation Metrics

PENDAHULUAN

Diabetes mellitus merupakan gangguan metabolik kronis yang ditandai dengan kondisi hiperglikemia akibat defisiensi sekresi insulin, resistensi insulin, atau kombinasi keduanya (Zhang & Tan, 2000). Data epidemiologi terkini menunjukkan peningkatan prevalensi yang mengkhawatirkan, dimana Organisasi Kesehatan Dunia melaporkan 422 juta kasus diabetes secara global - angka yang meningkat empat kali lipat dalam tiga dekade terakhir (Najib et al., 2019). Manifestasi klinis penyakit ini bervariasi, mulai dari poliuria, polidipsia, penurunan berat badan tidak terjelaskan, hingga komplikasi seperti gangguan penyembuhan luka, neuropati perifer, dan gangguan penglihatan. Deteksi dini melalui pemeriksaan laboratorium dan klasifikasi yang akurat menjadi krusial untuk menentukan intervensi terapeutik yang tepat.

Dalam bidang machine learning, decision tree merupakan algoritma klasifikasi yang merepresentasikan aturan prediktif dalam struktur pohon hierarkis. Menurut Eska (2016), struktur ini terdiri dari tiga komponen fundamental: (1) root node sebagai atribut pembeda utama, (2) internal node sebagai percabangan aturan decision, dan (3) leaf node sebagai output klasifikasi. Algoritma C4.5 yang dikembangkan oleh Quinlan menjadi salah satu implementasi decision tree yang efektif melalui proses iteratif: seleksi atribut optimal sebagai root node, pembentukan cabang berdasarkan nilai atribut, dan pembagian kasus hingga mencapai terminal node yang merepresentasikan kelas prediksi (Parung, 2018). Pendekatan ini memungkinkan interpretasi model yang transparan dengan kemampuan handling baik pada data numerik maupun kategorikal

terapan (applied research) berbasis komputasional dengan pendekatan kuantitatif-eksperimental, yang bertujuan untuk membangun dan menguji sistem deteksi penyakit diabetes menggunakan algoritma Decision Tree. Penelitian bersifat interdisipliner, menggabungkan bidang ilmu komputer dengan ilmu medis.

Data yang digunakan dalam penelitian ini diperoleh dari situs resmi UCI Machine Learning Repository. Dataset tersebut memiliki 520 entri dan terdiri atas 16 variabel independen serta satu variabel dependen, yaitu status penyakit diabetes (positif atau negatif). Adapun keenam belas variabel independen tersebut meliputi:

- a. Age: usia partisipan dalam survei.
- b. Gender: jenis kelamin.
- c. Polyuria: frekuensi buang air kecil berlebih (>20 kali per hari).
- d. Polydipsia: rasa haus ekstrem (>3 liter per hari).
- e. Sudden weight loss: penurunan berat badan secara signifikan.
- f. Weakness: penurunan daya tahan tubuh.
- g. Polyphagia: peningkatan nafsu makan secara berlebihan.
- h. Genital thrush: infeksi jamur atau bakteri pada area genital.
- i. Visual blurring: gangguan penglihatan seperti pandangan kabur.
- j. Itching: rasa gatal yang terus-menerus.
- k. Irritability: gejala seperti mudah marah, kemerahan pada kulit, atau pembengkakan di tangan dan kaki.
- l. Delayed healing: luka yang sulit sembuh.
- m. Partial paresis: kelemahan sebagian pada tubuh.
- n. Muscle stiffness: kekakuan otot atau penurunan fungsi gerak otot.

METODE

Penelitian ini merupakan penelitian

- o. Alopecia: kerontokan rambut pada area tertentu yang menyebabkan kebotakan.
- p. Obesity: kondisi kelebihan berat badan atau obesitas.
- q. Class: label target berupa status diabetes (positif atau negatif).

Penelitian ini dilakukan melalui beberapa tahap sebagai berikut antara lain dengan Deskripsi dan Pra Pemrosesan Data, Klasifikasi menggunakan metode Decision tree, Klasifikasi menggunakan metode random forest

HASIL DAN PEMBAHASAN

A. HASIL

Hasil klasifikasi status penyakit diabetes menggunakan metode Decision Tree disajikan melalui confusion matrix pada Tabel 4.6. Dari 104 data uji yang dianalisis, sebanyak 58 individu berhasil diklasifikasikan secara benar sebagai positif diabetes, sedangkan 37 individu diklasifikasikan secara tepat sebagai negatif diabetes. Sementara itu, terdapat kesalahan klasifikasi pada 5 individu yang sebenarnya negatif tetapi diprediksi sebagai positif, dan 4 individu yang sebenarnya positif namun diprediksi sebagai negatif.

Berdasarkan perhitungan metrik evaluasi, diperoleh akurasi sebesar 91,35%, presisi sebesar 93,55%, recall sebesar 92,06%, dan specificity sebesar 90,24%. Selain itu, nilai F1-score yang dihasilkan adalah 0,9280. Hasil ini menunjukkan bahwa metode Decision Tree memberikan performa yang baik dalam mengklasifikasikan status diabetes berdasarkan atribut-atribut gejala yang tersedia dalam dataset.

B. PEMBAHASAN

Hasil klasifikasi menunjukkan bahwa metode Decision Tree mampu mengidentifikasi status diabetes dengan tingkat akurasi yang tinggi, yaitu mencapai 91,35%. Nilai presisi sebesar 93,55% menunjukkan bahwa dari seluruh individu yang diprediksi sebagai positif diabetes, sebagian besar memang benar-benar positif. Hal ini mencerminkan rendahnya tingkat false positive yang terjadi dalam model. Nilai recall sebesar 92,06% juga cukup tinggi, yang berarti sebagian besar individu yang benar-benar positif berhasil dikenali oleh model, sehingga menunjukkan efektivitas model dalam mendeteksi kasus positif.

Selain itu, specificity sebesar 90,24% menandakan kemampuan model dalam mengidentifikasi individu yang tidak memiliki diabetes secara akurat. Dengan demikian, keseimbangan antara nilai recall dan specificity menunjukkan bahwa model tidak hanya baik dalam mendeteksi kasus positif, tetapi juga cukup andal dalam menghindari kesalahan prediksi terhadap kasus negatif.

Nilai F1-score yang tinggi (0,9280) semakin menguatkan bahwa performa model tidak berat sebelah, melainkan mampu mempertahankan keseimbangan antara presisi dan recall. Hal ini sangat penting dalam konteks medis, mengingat kesalahan klasifikasi dapat berdampak langsung pada tindakan diagnosis dan penanganan terhadap pasien. Oleh karena itu, metode Decision Tree dapat dipertimbangkan sebagai pendekatan yang efektif dalam sistem pendukung keputusan untuk mendeteksi diabetes berdasarkan data gejala.

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode *Decision Tree* menunjukkan performa klasifikasi yang cukup baik dalam mendeteksi status penyakit diabetes. Dengan akurasi sebesar 91,35%, presisi 93,55%, recall 92,06%, specificity 90,24%, dan *F1-score* sebesar 0,9280, model ini mampu memberikan hasil yang seimbang dan akurat. Temuan ini mengindikasikan bahwa metode *Decision Tree* layak digunakan sebagai metode klasifikasi dalam analisis data kesehatan, khususnya dalam prediksi dini penyakit diabetes.

DAFTAR PUSTAKA

- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Media Informatika Budidarma*, 5(2), 640–651.
<https://doi.org/10.30865/mib.v5i2.2937>
- Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A. W., & Malanda, B. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Research and Clinical Practice*, 138, 271–281.
<https://doi.org/10.1016/j.diabres.2018.02.023>
- Eska, J. (2016). Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 2, 9 – 13.
https://doi.org/10.31227/osf.io/x6sv_c
- Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 32–39.
<https://doi.org/10.47970/siskomb.v4i1.173>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219.
<https://doi.org/10.22202/ei.2016.v2i2.1465>
- Najib, A., Nurcahyono, D., & Setiawan, R. P. P. (2019). Klasifikasi Diagnosa Penyakit Diabetes Mellitus (Dm) Menggunakan Algoritma C4.4. *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, 11(2), 47.
<https://doi.org/10.46964/justti.v11i2.153>
- Nasution, M. Z. F., Sitompul, O. S., & Ramli, M. (2018). PCA based feature reduction to improve the accuracy of decision tree c4.5 classification. *Journal of Physics: Conference Series*, 978(1).
<https://doi.org/10.1088/1742-6596/978/1/012058>
- Novianti, B., Rismawan, T., & Bahri, S. (2016). Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak). *Jurnal Coding, Sistem Komputer Untan*, 04(3), 75–84.
- Parung, F. (2018). Penerapan algoritma decision tree c4.5 dalam penerimaan guru pada smk sirajul falah parung. *CKI On SPOT*, 11(2), 192–198
- Punthakee, Z., Goldenberg, R., & Katz, P. (2018). Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. *Canadian Journal of Diabetes*, 42, S10–S15.
<https://doi.org/10.1016/j.icjd.2017.10.003>
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Siallagan, R. A., & Fitriyani. (2021). Prediksi Penyakit Diabetes

- Mellitus. Jurnal Responsif, 3(1), 45–46.
- Widiyati, D. K., Wati, M., & Pakpahan, H. S. (2018). Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara. Jurnal Rekayasa Teknologi Informasi (JURTI), 2(2), 126. <https://doi.org/10.30872/jurti.v2i2.1864>
- Yezli, S., Yassin, Y., Mushi, A., Balkhi, B., & Khan, A. (2021). Insulin Knowledge, Handling, and Storage among Diabetic Pilgrims during the Hajj Mass Gathering. Journal of Diabetes Research, 2021.
- Zhang, X. F., & Tan, B. K. H. (2000). Effects of an ethanolic extract of *Gynuracprocumbens* on serum glucose, cholesterol and triglyceride levels in normal and streptozotocin-induced diabetic rats. Singapore Medical Journal, 41(1), c9–13