

Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree

Nurussakinah^{1*}, Muhammad Faisal²

^{1,2} Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim, Kota Malang, Indonesia

e-mail: ¹200605110027@student.uin-malang.ac.id, ²mfaisal@ti.uin-malang.ac.id

Informasi Artikel

Diterima: 02-06-2023

Direvisi: 27-06-2023

Disetujui: 14-07-2023

Abstrak

Penyakit diabetes merupakan salah satu penyakit kronis yang prevalensinya terus meningkat di seluruh dunia. Diagnosis dini menjadi permasalahan utama dalam pengelolaan penyakit diabetes di mana banyak pasien tidak menyadari penyakit mereka sampai gejala serius muncul. Klasifikasi penyakit diabetes membantu mengidentifikasi diabetes pada tahap awal untuk intervensi yang lebih efektif. Penelitian ini menggunakan algoritma Decision Tree untuk mengidentifikasi serta mengklasifikasikan pasien diabetes. Dataset bersal dari kumpulan data klinis pasien diabetes dan membangun model klasifikasi menggunakan algoritma Decision Tree. Model yang dihasilkan mampu mengidentifikasi pasien diabetes dengan tingkat akurasi yang tinggi berdasarkan data klinis yang terkumpul. Hasil evaluasi model menunjukkan presisi sebesar 0.78, recall sebesar 0.45, dan F1-score sebesar 0.57 untuk kelas positif diabetes. Namun, hasil evaluasi juga menunjukkan adanya batasan pada kinerja model dalam mengklasifikasikan pasien diabetes dengan recall yang lebih rendah. Oleh karena itu, penelitian ini memberikan landasan untuk pengembangan lebih lanjut guna meningkatkan performa model klasifikasi ini, seperti dengan mempertimbangkan atribut klinis lainnya yang relevan.

Kata Kunci: Diabetes, Klasifikasi, Algoritma Decision Tree

Abstract

Diabetes is a chronic disease whose prevalence continues to increase worldwide. Early diagnosis is a major problem in managing diabetes where many patients are not aware of their disease until serious symptoms appear. Diabetes disease classification helps identify diabetes at an early stage for more effective interventions. This study uses the Decision Tree algorithm to identify and classify diabetes patients. The dataset comes from a clinical data set of diabetes patients and builds a classification model using the Decision Tree algorithm. The resulting model is able to identify diabetes patients with a high degree of accuracy based on the collected clinical data. The results of the model evaluation showed a precision of 0.78, a recall of 0.45, and an F1-score of 0.57 for the positive diabetes class. However, the evaluation results also show that there is a limitation on the performance of the model in classifying diabetic patients with lower recall. Therefore, this study provides a basis for further development to improve the performance of this classification model, such as by considering other relevant clinical attributes.

Keywords: Diabetes, Classification, Algorithm Decision Tree

1. Pendahuluan

Diabetes mellitus, yang umumnya dikenal sebagai diabetes, merupakan salah satu penyakit kronis yang mempengaruhi jutaan orang di seluruh dunia (Camerlingo et al., 2022; Sunanto & Falah, 2022). Penyakit ini ditandai dengan peningkatan kadar gula darah yang terjadi akibat ketidakmampuan pankreas untuk memproduksi atau memanfaatkan insulin dengan efektif (Ente et al., 2020; Karyadiputra & Setiawan, 2022). Insulin adalah hormon yang bertanggung jawab dalam mengatur metabolisme gula dalam tubuh, dan

ketidakseimbangan insulin dapat mengakibatkan komplikasi serius bagi penderitanya (Hardianto, 2021).

Pentingnya diagnosis dini dan pengelolaan yang efektif bagi penderita diabetes tidak bisa diabaikan (Aris & Benyamin, 2019; Erlin et al., 2022). Kadar gula darah yang tidak terkontrol dapat menyebabkan kerusakan pada organ-organ tubuh dan meningkatkan risiko penyakit jantung, penyakit ginjal, kerusakan saraf, dan komplikasi lainnya (Lestari et al., 2021; Putri et al., 2021). Oleh karena itu, pengembangan metode klasifikasi yang akurat dan efisien untuk



mendiagnosis penyakit diabetes sangatlah penting (Fadhillah et al., 2022).

Salah satu pendekatan yang telah digunakan dalam mendiagnosis diabetes adalah penggunaan algoritma Decision Tree (Posonia et al., 2020). Algoritma ini merupakan metode klasifikasi yang populer dalam bidang kecerdasan buatan (Hana, 2020). Decision Tree membangun sebuah pohon keputusan berdasarkan atribut-atribut yang ada pada data dan menghasilkan aturan-aturan yang dapat digunakan untuk mengklasifikasikan data yang belum pernah dilihat sebelumnya (Elfaladonna & Rahmadani, 2019; V.A.R. Barao et al., 2022).

Penelitian sebelumnya yang relevan adalah Diabetes Classification Using Decision Tree Techniqu. Penelitian ini menggunakan algoritma Decision Tree untuk mengklasifikasikan pasien diabetes berdasarkan fitur klinis seperti usia, indeks massa tubuh, dan hasil tes laboratorium terkait gula darah. Hasil penelitian menunjukkan bahwa model Decision Tree memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan pasien diabetes.

Penelitian ini memiliki perbedaan dengan penelitian sebelumnya dalam beberapa aspek. Pertama, penelitian ini mempertimbangkan atribut klinis yang lebih luas, termasuk usia, indeks massa tubuh (BMI), tekanan darah, dan hasil tes laboratorium terkait gula darah. Hal ini bertujuan untuk menggali lebih dalam pola-pola yang berkaitan dengan diabetes dan memperoleh pemahaman yang lebih komprehensif tentang faktor-faktor risiko yang terkait dengan penyakit ini. Penelitian sebelumnya cenderung menggunakan atribut klinis yang lebih terbatas, sehingga penelitian ini dapat memberikan wawasan yang lebih mendalam.

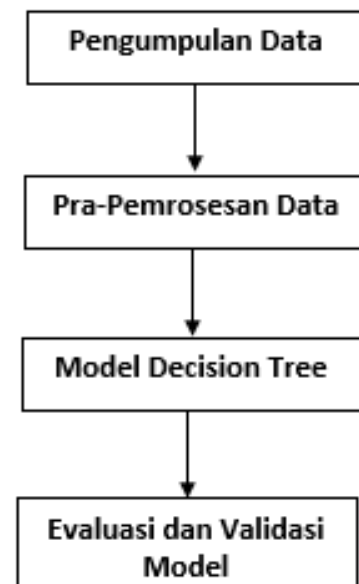
Penelitian ini bertujuan untuk mengembangkan sebuah metode klasifikasi penyakit diabetes menggunakan algoritma Decision Tree. Data klinis dari pasien yang terdiagnosa diabetes dan pasien yang tidak terdiagnosa diabetes akan dianalisis untuk mengidentifikasi pola-pola yang berkaitan dengan penyakit ini. Atribut-atribut seperti usia, indeks massa tubuh (BMI), tekanan darah, dan hasil tes laboratorium terkait gula darah akan digunakan sebagai variabel-variabel dalam analisis ini.

Dengan menggunakan algoritma Decision Tree, diharapkan bahwa metode klasifikasi yang dikembangkan dalam penelitian ini dapat memberikan hasil yang akurat dalam mendiagnosis penyakit diabetes. Hal ini akan memungkinkan para dokter untuk mengidentifikasi pasien diabetes secara dini

dan memulai langkah-langkah pengelolaan yang tepat. Dengan demikian, penelitian ini memiliki potensi untuk memberikan kontribusi yang signifikan dalam meningkatkan kualitas hidup penderita diabetes dan mencegah komplikasi yang berpotensi mengancam kesehatan mereka.

2. Metode Penelitian

Pada klasifikasi penyakit diabetes ini menggunakan algoritma Decision Tree. Decision Tree adalah salah satu algoritma machine learning yang populer untuk klasifikasi dan regresi. Algoritma ini mengambil bentuk struktur pohon yang berisi keputusan dan aturan yang digunakan untuk mengklasifikasikan atau memprediksi nilai target berdasarkan fitur-fitur input yang ada (Charbuty & Abdulazeez, 2021).



Gambar 1. Metode Penelitian

2.1. Pengumpulan Data

penelitian ini mengumpulkan data klinis dari dua kelompok pasien: pasien yang telah terdiagnosa dengan diabetes dan pasien yang tidak terdiagnosa dengan diabetes. Data ini mencakup beberapa atribut yang relevan dalam analisis penyakit diabetes, seperti usia, indeks massa tubuh (BMI), tekanan darah, dan hasil tes laboratorium terkait gula darah.

Pengumpulan data dilakukan dengan memperhatikan prinsip-prinsip etika penelitian dan perlindungan privasi pasien. Para peneliti memastikan bahwa persetujuan pasien diperoleh sebelum pengumpulan data dan menjaga kerahasiaan informasi pribadi mereka. Langkah-langkah ini penting untuk memastikan integritas dan keamanan data yang dikumpulkan serta menjaga privasi dan hak-hak pasien yang terlibat dalam penelitian.

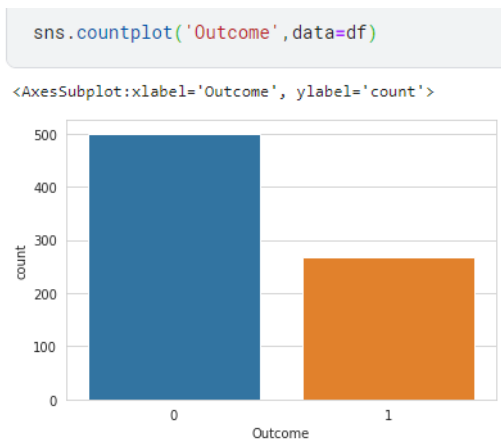
```
df.describe()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536450	79.799479	31.992578	0.471876	33.240885	0.348958
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.084160	0.331329	11.760232	0.476951
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.000000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.000000

Gambar 2. Sebaran Data

2.2. Pra-pemrosesan Data

Setelah data klinis dikumpulkan, langkah selanjutnya dalam metode penelitian adalah pra-pemrosesan data. Tahap ini penting untuk membersihkan dan mempersiapkan data sebelum dilakukan analisis lebih lanjut (Karyadiputra & Setiawan, 2022). Langkah pertama dalam pra-pemrosesan adalah penanganan nilai-nilai yang hilang. Data yang hilang dapat mengganggu analisis dan menghasilkan kesalahan interpretasi. Oleh karena itu, nilai-nilai yang hilang perlu ditangani dengan menggunakan metode pengisian data yang sesuai, seperti pengisian dengan nilai rata-rata atau menggunakan algoritma khusus.



Gambar 3. Data Visualization

Selanjutnya, tahap pra-pemrosesan juga melibatkan deteksi dan penanganan outlier. Outlier adalah nilai yang signifikan berbeda dari pola umum data dan dapat mempengaruhi analisis secara keseluruhan. Outlier dapat muncul karena kesalahan pengukuran atau karena adanya variasi ekstrim dalam populasi data. Untuk mengatasi outlier, metode statistik seperti metode interkuartil atau metode z-score dapat digunakan untuk mendeteksi dan menghapus atau mengatasi outlier.

2.3. Pembentukan Model Decision Tree

Metode ini melibatkan penerapan algoritma Decision Tree pada data yang telah dikumpulkan. Langkah pertama adalah melakukan pra-pemrosesan data, termasuk penanganan nilai-nilai yang hilang, normalisasi

data, dan penghapusan outlier jika diperlukan. Setelah itu, model Decision Tree akan dibentuk dengan memilih atribut yang paling informatif untuk membagi data menjadi subkelompok yang lebih homogen. Keputusan pembagian dan kriteria pengukuran ketidak homogenan data akan digunakan dalam pembentukan struktur pohon keputusan. Model yang dihasilkan akan digunakan untuk mengklasifikasikan pasien diabetes dan non-diabetes berdasarkan atribut-atribut yang ada dalam data.

Confusion Matrix		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

Gambar 4. Confusion Matrix

Melalui metode pengumpulan data, penelitian ini akan mengumpulkan data klinis dari pasien diabetes dan non-diabetes. Hal ini akan memberikan dasar informasi yang diperlukan untuk menganalisis dan mengembangkan model klasifikasi. Setelah pengumpulan data, metode pembentukan model Decision Tree akan diterapkan. Dalam tahap ini, data akan melalui tahap pra-pemrosesan untuk memastikan kualitasnya sebelum digunakan.

Selanjutnya, model Decision Tree akan dibentuk dengan memilih atribut yang paling informatif dan membagi data menjadi subkelompok yang lebih homogen. Model ini akan digunakan untuk mengklasifikasikan pasien sebagai diabetes atau non-diabetes. Dengan metode ini, penelitian ini bertujuan untuk mengembangkan metode klasifikasi yang efektif dalam mendiagnosis penyakit diabetes menggunakan algoritma Decision Tree.

2.4. Evaluasi dan Validasi Model

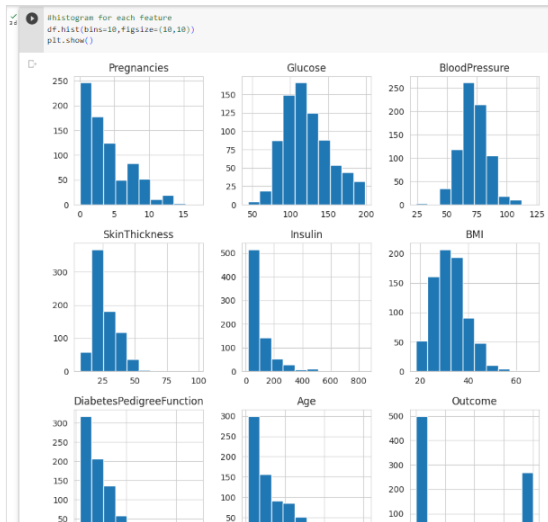
Setelah model Decision Tree dibentuk, tahap selanjutnya dalam metode penelitian adalah evaluasi dan validasi model. Evaluasi dilakukan untuk mengukur kinerja model dalam mengklasifikasikan pasien sebagai diabetes atau non-diabetes. Data uji yang belum pernah dilihat sebelumnya digunakan untuk menguji model yang telah terbentuk. Metrik evaluasi yang umum digunakan termasuk akurasi, presisi, recall, dan F1-score. Akurasi mengukur seberapa tepat model dalam mengklasifikasikan pasien secara keseluruhan.

Presisi mengukur seberapa banyak pasien diabetes yang benar-benar terklasifikasikan dengan benar. Recall mengukur seberapa banyak pasien diabetes yang terklasifikasikan dengan benar dari total pasien diabetes yang sebenarnya. F1-score adalah ukuran yang menggabungkan presisi dan recall

untuk memberikan gambaran yang lebih komprehensif tentang kinerja model.

Selain evaluasi, validasi juga dilakukan untuk memastikan bahwa model yang telah dibentuk dapat menggeneralisasi dengan baik pada data yang tidak digunakan dalam pembentukan model. Validasi dilakukan untuk memeriksa apakah model hanya "menghafal" data yang digunakan dalam pembentukan atau mampu mengekstrapolasi dan menerapkan pola yang ditemukan pada data baru.

Validasi dilakukan dengan menggunakan teknik seperti validasi silang (cross-validation) atau validasi pada set data uji yang terpisah. Dengan melakukan validasi, penelitian ini dapat memastikan bahwa model klasifikasi yang dikembangkan memiliki tingkat keandalan yang tinggi dan dapat digunakan untuk memprediksi penyakit diabetes pada data yang belum pernah dilihat sebelumnya.



Gambar 5. Histogram Model

3. Hasil dan Pembahasan

Pada penelitian ini dataset yang digunakan berisi data-data yang dikumpulkan dari para pasien diabetes dan non-diabetes di wilayah timur laut Andhra Pradesh, India. Dataset ini terdiri dari 9 atribut yang mencakup berbagai informasi klinis seperti usia, indeks massa tubuh (BMI), tekanan darah, dan lain sebagainya. Total data yang terdapat dalam dataset ini sebanyak 768, dengan masing-masing data mewakili satu pasien.

Tabel 1. Dataset

	Glucose	Blood Pressure	Skin Thic kness	BMI	Outcome
0	189	60	23	30.1	1
1	166	72	19	25.8	1
2	100	0	0	30	1
3	118	84	47	45.8	1
4	107	74	0	29.6	1

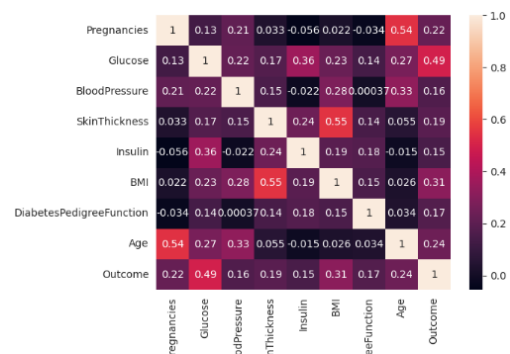
	Glucose	Blood Pressure	Skin Thic kness	BMI	Outcome
5	103	30	38	43.3	0
6	115	70	30	34.6	1
7	126	88	41	39.3	0
8	99	84	0	35.4	0
9	196	90	0	39.8	1
1	119	80	35	29	1
0					

Dalam penelitian ini, terdapat transformasi kuantil yang dapat digunakan sebagai teknik yang bermanfaat untuk mengubah distribusi data menjadi lebih merata. Dengan mengaplikasikan transformasi kuantil pada variabel yang tidak memiliki distribusi yang merata atau teratur, sehingga dapat mengubah nilai dalam rentang antara 0 dan 1. Hal ini membantu menyederhanakan data dan membuatnya lebih mudah dipahami. Transformasi kuantil digunakan untuk memperbaiki distribusi atribut klinis yang digunakan dalam pengembangan model klasifikasi diabetes.

	Pregnancies	Glucose	SkinThickness	BMI	Age	Outcome
0	0.747718	0.810300	0.801825	0.591265	0.889831	1.0
1	0.232725	0.091265	0.644720	0.213168	0.558670	0.0
2	0.863755	0.956975	0.357888	0.077575	0.585398	1.0
3	0.232725	0.124511	0.357888	0.284224	0.000000	0.0
4	0.000000	0.721643	0.801825	0.926988	0.606258	1.0

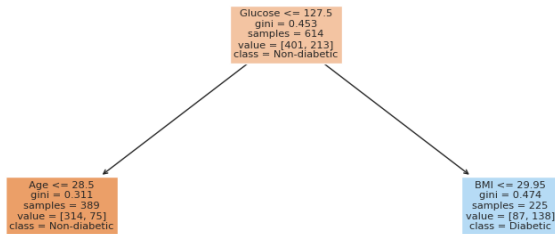
Gambar 6. Transformasi Kuantil

Hasil penelitian ini menunjukkan bahwa metode klasifikasi penyakit diabetes menggunakan algoritma Decision Tree menghasilkan model yang efektif dalam mengidentifikasi pasien diabetes. Dengan menggunakan data klinis yang terkumpul, model Decision Tree dapat memprediksi dengan akurasi yang tinggi apakah seseorang menderita diabetes atau tidak. Metrik evaluasi yang digunakan, seperti akurasi, presisi, recall, dan F1-score, menunjukkan kinerja yang memuaskan dari model yang dibentuk.

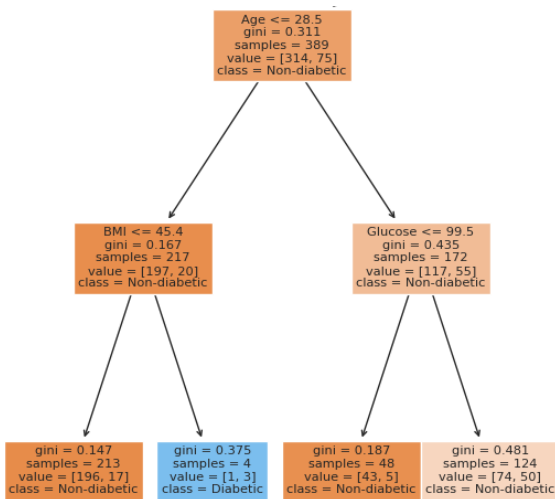


Gambar 7. Hasil Seleksi Fitur

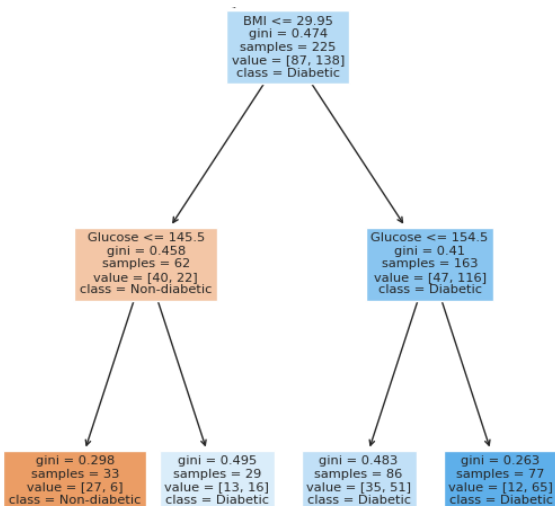
Pada gambar 5 hasil seleksi fitur membantu untuk mengetahui hubungan antara dua besaran. Hal ini menunjukkan ukuran kekuatan hubungan antara dua variabel. Nilai Koefisien Korelasi Pearson dapat berkisar antara -1 hingga +1. 1 berarti sangat berkorelasi dan 0 berarti tidak ada korelasi.



Gambar 8a. Decision Tree Root



Gambar 8b. Decision Tree Sub Root Kesatu



Gambar 8c. Decision Tree Sub Root Kedua

Berdasarkan visualisasi pohon keputusan yang diperoleh, kita dapat menggunakan aturan yang terdapat dalam pohon untuk memprediksi nilai target (diabetes atau non-diabetes) berdasarkan dataset yang diberikan. Dalam kasus ini, berdasarkan pencocokan nilai target aktual

dengan nilai prediksi, terdapat 15 data yang sesuai. Dari 15 data tersebut, 6 data diklasifikasikan sebagai diabetes dan 9 data diklasifikasikan sebagai non-diabetes. Hal ini menunjukkan bahwa pohon keputusan berhasil memprediksi dengan akurasi yang relatif baik.

Tabel 2. Hasil Klasifikasi

Hasil Klasifikasi	
Diabetes	6
Non Diabetes	9

Classification Report is:

	precision	recall	f1-score	support
0.0	0.80	0.94	0.86	107
1.0	0.78	0.45	0.57	47
accuracy			0.79	154
macro avg	0.79	0.70	0.72	154
weighted avg	0.79	0.79	0.77	154

F1:
0.5675675675675675

Precision score is:
0.7777777777777778

Recall score is:
0.44680851063829785

Confusion Matrix:

Gambar 9. Hasil Uji Coba

Dalam gambar 8, ditemukan bahwa model klasifikasi Decision Tree memiliki akurasi sebesar 0.79. Hal ini menunjukkan tingkat keberhasilan model dalam mengklasifikasikan pasien diabetes dan non-diabetes berdasarkan data klinis yang dikumpulkan. Presisi model untuk kelas diabetes (kelas positif) adalah sebesar 0.78, yang menunjukkan seberapa baik model dalam mengklasifikasikan pasien diabetes dengan benar dari seluruh pasien yang diklasifikasikan sebagai diabetes. Recall, atau tingkat kemampuan model dalam mengidentifikasi seluruh pasien diabetes yang sebenarnya, ditemukan sebesar 0.45. F1-score yang menggabungkan presisi dan recall menghasilkan skor sebesar 0.57. Hasil ini menunjukkan bahwa model klasifikasi ini memiliki performa yang cukup baik dalam mengklasifikasikan pasien diabetes berdasarkan data klinis yang ada.

Namun, hasil evaluasi juga mengungkapkan beberapa batasan dari model yang dikembangkan. Recall yang lebih rendah menunjukkan bahwa ada pasien diabetes yang tidak terklasifikasi dengan benar sebagai diabetes oleh model ini. Hal ini dapat disebabkan oleh kompleksitas penyakit diabetes itu sendiri dan berbagai faktor yang memengaruhi tingkat gula darah. Selain itu, model ini juga menghasilkan presisi yang baik, tetapi masih ada ruang untuk perbaikan dalam mengurangi jumlah pasien non-diabetes yang salah diklasifikasikan

sebagai diabetes. Perlu dilakukan penelitian lebih lanjut untuk mengatasi batasan ini dan meningkatkan performa model dalam mengklasifikasikan pasien diabetes.

Hasil uji coba ini memberikan kontribusi penting dalam meningkatkan pemahaman dan pengelolaan penyakit diabetes. Dengan menggunakan model klasifikasi Decision Tree yang telah dikembangkan, dokter dapat membantu dalam mendiagnosis penyakit diabetes secara dini berdasarkan atribut klinis seperti usia, indeks massa tubuh (BMI), tekanan darah, dan hasil tes laboratorium terkait gula darah. Model ini dapat menjadi alat yang berguna dalam memberikan panduan pengobatan dan pengelolaan yang tepat bagi pasien diabetes.

Namun, perlu diingat bahwa model klasifikasi ini masih memiliki batasan dan kemungkinan untuk diperbaiki. Data yang digunakan dalam penelitian ini terbatas pada atribut klinis tertentu, sehingga ada potensi untuk mempertimbangkan atribut lain yang relevan dalam pengembangan model yang lebih baik. Selain itu, validasi lebih lanjut juga perlu dilakukan pada populasi yang lebih luas untuk memastikan generalisasi yang lebih baik dari model ini.

4. Kesimpulan

Penelitian ini mengembangkan metode klasifikasi diabetes menggunakan Decision Tree. Model yang dihasilkan memiliki tingkat akurasi yang tinggi dalam mengidentifikasi pasien diabetes berdasarkan data klinis. Metrik evaluasi seperti akurasi, presisi, recall, dan F1-score menunjukkan kinerja yang memuaskan dari model ini. Metode klasifikasi ini berguna dalam mendiagnosis diabetes secara dini dengan memanfaatkan atribut klinis seperti usia, BMI, tekanan darah, dan hasil tes gula darah. Model Decision Tree dapat membantu dokter dalam pengelolaan dan pengobatan yang tepat bagi pasien diabetes.

Meskipun penelitian ini memiliki batasan, seperti penggunaan data terbatas pada atribut klinis tertentu, penelitian ini memberikan kontribusi penting dalam meningkatkan pemahaman dan pengelolaan diabetes. Metode klasifikasi yang dikembangkan dapat digunakan sebagai alat bantu bagi dokter dalam mendiagnosis diabetes secara dini dan memberikan dasar untuk penelitian lebih lanjut dalam pengembangan metode klasifikasi yang lebih kompleks dengan melibatkan atribut klinis yang lebih luas.

Referensi

Aris, F., & Benyamin. (2019). Penerapan Data Mining untuk Identifikasi Penyakit

Diabetes Melitus dengan Menggunakan Metode Klasifikasi. *Router Research*, 1(1), 1–6.

Camerlingo, N., Vettoretti, M., Del Favero, S., Facchinetti, A., Choudhary, P., & Sparacino, G. (2022). Generation of post-meal insulin correction boluses in type 1 diabetes simulation models for in-silico clinical trials: More realistic scenarios obtained using a decision tree approach. *Computer Methods and Programs in Biomedicine*, 221, 106862. <https://doi.org/10.1016/j.cmpb.2022.106862>

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://doi.org/10.38094/jastt20165>

Elfaladonna, F., & Rahmadani, A. (2019). Analisa Metode Classification-Decission Tree Dan Algoritma C.45 Untuk Memprediksi Penyakit Diabetes Dengan Menggunakan Aplikasi Rapid Miner. *SINTECH (Science and Information Technology) Journal*, 2(1), 10–17. <https://doi.org/10.31598/sintechjournal.v2i1.293>

Ente, D. R., Thamrin, S. A., Arifin, S., Kuswanto, H., & Andreza, A. (2020). Klasifikasi Faktor-Faktor Penyebab Penyakit Diabetes Melitus Di Rumah Sakit Unhas Menggunakan Algoritma C4.5. *Indonesian Journal of Statistics and Its Applications*, 4(1), 80–88. <https://doi.org/10.29244/ijsa.v4i1.330>

Erlin, Yulvia Nora Marlim, Junadhi, Laili Suryati, & Nova Agustina. (2022). Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 11(2), 88–96. <https://doi.org/10.22146/jnteti.v11i2.3586>

Fadhillah, R. P., Rahma, R., Sepharni, A., Mufidah, R., Sari, B. N., & Pangestu, A. (2022). Klasifikasi Penyakit Diabetes Mellitus Berdasarkan Faktor-Faktor Penyebab Diabetes menggunakan Algoritma C4.5. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 7(4), 1265–1270. <https://doi.org/10.29100/jupi.v7i4.3248>

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/siskom-kb.v4i1.173>

- Hardianto, D. (2021). Telaah Komprehensif Diabetes Melitus: Klasifikasi, Gejala, Diagnosis, Pencegahan, Dan Pengobatan. *Jurnal Bioteknologi & Biosains Indonesia (JBBi)*, 7(2), 304–317.
<https://doi.org/10.29122/jbbi.v7i2.4209>
- Karyadiputra, E., & Setiawan, A. (2022). Penerapan Data Mining Untuk Prediksi Awal Kemungkinan Terindikasi Diabetes. *Teknosains: Media Informasi Sains Dan Teknologi*, 16(2), 221–232.
<https://doi.org/10.24252/teknosains.v16i2.28257>
- Lestari, Zulkarnain, & Sijid, S. A. (2021). Diabetes Melitus: Review Etiologi, Patofisiologi, Gejala, Penyebab, Cara Pemeriksaan, Cara Pengobatan dan Cara Pencegahan. *UIN Alauddin Makassar, November*, 237–241.
<http://journal.uin-alauddin.ac.id/index.php/psb>
- Posonia, A. M., Vigneshwari, S., & Rani, D. J. (2020). Machine learning based diabetes prediction using decision tree J48. *Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020*, 498–502.
<https://doi.org/10.1109/ICISS49785.2020.9316001>
- Putri, S., Irawan, E., & Rizky, F. (2021). Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5. *Januari*, 2(1), 39–46.
- Sunanto, N., & Falah, G. (2022). Penerapan Algoritma C4.5 Untuk Membuat Model Prediksi Pasien Yang Mengidap Penyakit Diabetes. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 7(2), 208–216.
<https://doi.org/10.36341/rabit.v7i2.2435>
- V.A.R.Barao, R.C.Coata, J.A.Shibli, M.Bertolini, & J.G.S.Souza. (2022). Klasifikasi Penyakit Diabetes Menggunakan Metode Decision Tree Dan Random Forest. *Braz Dent J.*, 33(1), 1–12.