

PREDIKSI RISIKO DIABETES MENGGUNAKAN ALGORITMA *DECISION TREE* DENGAN APLIKASI *RAPID MINER*

Muhammad Gilang Febrian*¹, Riko Ferdiansyah², Erwin Ardian Nugraha³, Dzaki Satriatama⁴, Rajnaparamitha Kusumastuti⁵

¹²³⁴⁵Program Studi Informatika, STMIK Amikom Surakarta
¹²³⁴⁵Sukoharjo-Indonesia

Email: ¹muhammad.10393@mhs.amikomsolo.ac.id,
²riko.10390@mhs.amikomsolo.ac.id, ³erwin.10395@mhs.amikomsolo.ac.id,
⁴dzaki.10394@mhs.amikomsolo.ac.id, ⁵rajna@dosen.amikomsolo.ac.id

Abstract

Diabetes is a medical condition characterized by high levels of sugar (glucose) in the blood. Early detection and prediction of diabetes are very important to support effective prevention and management efforts. This research aims to predict the risk of diabetes using Decision Tree algorithms, with the hope of helping to identify individuals at risk of developing diabetes based on medical data and health parameters. The method used is the Decision Tree algorithm as a machine learning approach for classification and prediction, with individual medical data as input. The research results show that this Decision Tree model has an accuracy rate of 97.19% in predicting the risk of diabetes. In conclusion, the Decision Tree algorithm has great potential to be used in diabetes prediction. This machine learning-based predictive model can support medical practitioners in early detection and more effective decision-making related to the prevention and management of diabetes.

Keywords: *Diabetes, Algorithm, Decision Tree*

Abstraksi

Diabetes adalah kondisi medis yang ditandai oleh tingginya kadar gula (glukosa) dalam darah. Deteksi dini dan prediksi penyakit diabetes menjadi sangat penting untuk mendukung upaya pencegahan dan pengelolaan yang efektif. Penelitian ini bertujuan untuk memprediksi risiko diabetes menggunakan algoritma pohon keputusan, dengan harapan dapat membantu mengidentifikasi individu yang berisiko terkena diabetes berdasarkan data medis dan parameter kesehatan. Metode yang digunakan adalah algoritma pohon keputusan sebagai pendekatan pembelajaran mesin untuk klasifikasi dan prediksi, dengan data medis individu sebagai input. Hasil penelitian menunjukkan bahwa model pohon keputusan ini memiliki tingkat akurasi sebesar 97,19% dalam memprediksi risiko diabetes. Kesimpulannya, algoritma pohon keputusan memiliki potensi besar untuk digunakan dalam prediksi diabetes. Model prediksi berbasis pembelajaran mesin ini dapat mendukung praktisi medis dalam deteksi dini dan pengambilan keputusan yang lebih efektif terkait pencegahan serta penanganan diabetes.

Kata Kunci: *Diabetes, Algoritma, Pohon Keputusan*

1. PENDAHULUAN

Kesehatan masyarakat menjadi perhatian utama secara global karena prevalensi penyakit tidak menular yang terus meningkat, termasuk diabetes mellitus. Diabetes merupakan penyakit metabolik kronis yang ditandai dengan gangguan pengendalian kadar gula darah, yang dapat mengarah pada komplikasi serius seperti penyakit jantung, kerusakan saraf, dan gagal ginjal [1]. Menurut Federasi Diabetes Internasional (IDF), pada tahun 2019 terdapat sekitar 463 juta penderita diabetes di seluruh dunia, dan jumlah ini diproyeksikan terus meningkat hingga mencapai 700 juta orang pada tahun 2045 [2]. Salah satu tantangan utama dalam pencegahan dan pengelolaan diabetes adalah kemampuan mendeteksi risiko secara dini sehingga langkah-langkah intervensi dapat dilakukan secara tepat [3].

Masalah utama yang menjadi fokus penelitian ini adalah prediksi risiko diabetes pada individu menggunakan algoritma *Decision Tree*. Diabetes sering kali tidak terdiagnosis pada tahap awal, yang menyebabkan penyakit ini berkembang secara bertahap hingga menyebabkan komplikasi yang parah [4]. Hal ini disebabkan oleh kurangnya akses terhadap alat prediksi yang efektif, serta keterbatasan pada pendekatan tradisional yang bergantung pada diagnosis medis manual yang memakan waktu dan biaya. Dengan semakin berkembangnya data kesehatan dalam jumlah besar, penting untuk memanfaatkan metode yang dapat mengolah data ini secara efektif guna menghasilkan model prediksi yang andal dan mudah diinterpretasikan [5].

Permasalahan ini muncul karena kurangnya pemanfaatan teknologi berbasis data dalam menganalisis faktor risiko yang kompleks dan saling terkait. Faktor-faktor seperti usia, indeks massa tubuh (BMI), tekanan darah, dan riwayat keluarga merupakan beberapa variabel yang sangat mempengaruhi risiko diabetes, namun sering kali sulit diprediksi dengan pendekatan konvensional [6]. Oleh karena itu, pendekatan berbasis data mining menggunakan algoritma seperti *Decision Tree* menjadi penting untuk menangani tantangan ini. *Decision Tree* mampu memecah dataset berdasarkan variabel penting dan memberikan representasi yang intuitif serta interpretasi hasil yang lebih jelas bagi tenaga medis [7].

Tujuan dari penelitian ini adalah mengembangkan model prediksi risiko diabetes menggunakan algoritma *Decision Tree* dan mengevaluasi performanya dalam hal akurasi, sensitivitas, dan spesifisitas. Melalui evaluasi ini, diharapkan model yang dihasilkan dapat memberikan manfaat praktis dalam membantu tenaga medis dan sistem kesehatan dalam melakukan deteksi dini risiko diabetes, serta memberikan wawasan tambahan mengenai variabel-variabel yang paling berpengaruh terhadap penyakit ini [8].

2. TINJAUAN PUSTAKA

Diabetes mellitus merupakan salah satu penyakit kronis yang kompleks dengan prevalensi yang meningkat setiap tahun. Pendeteksian dini penyakit ini sangat penting untuk meminimalkan risiko komplikasi jangka panjang. Dengan kemajuan teknologi

informasi, penggunaan algoritma pembelajaran mesin, khususnya *Decision Tree*, menjadi populer dalam memprediksi risiko diabetes karena kemampuannya dalam melakukan klasifikasi yang akurat [9].

Algoritma *Decision Tree* bekerja dengan cara membagi data ke dalam cabang-cabang berdasarkan atribut yang memiliki pengaruh terbesar pada target yang diprediksi. Model ini dianggap efektif untuk klasifikasi penyakit karena mudah diinterpretasikan dan memiliki proses komputasi yang efisien [10]. Penelitian yang dilakukan oleh Dikan Ismafillah et al. [9] menunjukkan bahwa algoritma *Decision Tree* yang dikombinasikan dengan metode oversampling SMOTE dapat meningkatkan performa dalam memprediksi penyakit diabetes. Dalam penelitian tersebut, algoritma ini berhasil mencapai akurasi yang cukup tinggi dalam mengklasifikasikan data pasien diabetes.

Selain itu, Hana [10] juga memaparkan bahwa algoritma *Decision Tree* C4.5 dapat digunakan untuk mengklasifikasikan penderita diabetes dengan mempertimbangkan beberapa variabel seperti kadar gula darah, usia, dan berat badan. Metode ini terbukti efektif dalam mengidentifikasi faktor risiko utama pada penderita diabetes. Beberapa studi lain turut memperkuat temuan ini, seperti penelitian oleh Afifuddin dan Hakim [11], yang menggunakan arsitektur *Decision Tree* C4.5 untuk mendeteksi diabetes dengan akurasi yang cukup baik.

Pemanfaatan aplikasi seperti *RapidMiner* dalam mengimplementasikan algoritma *Decision Tree* juga diangkat oleh Elfaladonna dan Rahmadani [12]. Mereka menganalisis data kesehatan menggunakan metode C4.5 di *RapidMiner* dan berhasil memprediksi pasien dengan risiko diabetes. Studi ini menunjukkan bahwa algoritma *Decision Tree* tidak hanya efektif tetapi juga fleksibel untuk diintegrasikan dengan perangkat lunak data mining yang umum digunakan.

Penelitian yang dilakukan Setio et al. [13] dan Habibah et al. [14] mengindikasikan bahwa *Decision Tree* C4.5 adalah salah satu algoritma yang sering diadopsi dalam klasifikasi data kesehatan, termasuk prediksi penyakit diabetes. Kedua penelitian ini menunjukkan bahwa faktor-faktor seperti usia, tekanan darah, dan indeks massa tubuh sangat berpengaruh dalam klasifikasi risiko diabetes. Meskipun algoritma ini sangat bermanfaat, ada tantangan dalam menangani *overfitting* yang sering terjadi ketika model pohon keputusan terlalu dalam.

Untuk mengatasi kelemahan ini, beberapa penelitian menggunakan metode lain sebagai pembanding atau pelengkap. Misalnya, Apriliah et al. [15] menggunakan algoritma Random Forest sebagai alternatif karena algoritma ini menghasilkan pohon keputusan yang lebih stabil dan akurat, sedangkan Maulidah et al. [16] membandingkan hasil prediksi dari metode *Support Vector Machine* dan *Naive Bayes* dengan *Decision Tree* untuk melihat algoritma yang paling optimal dalam mendeteksi risiko diabetes.

Dengan pertimbangan efektivitasnya, algoritma *Decision Tree* terbukti menjadi pilihan yang layak untuk diterapkan dalam deteksi dini dan prediksi risiko diabetes. Namun, optimalisasi model diperlukan untuk mengurangi risiko *overfitting*, serta

adaptasi lebih lanjut dengan metode atau algoritma lain dapat membantu meningkatkan akurasi dan kestabilan model.

3. METODE PENELITIAN

Pendekatan penelitian yang dilakukan dalam studi ini adalah penelitian eksperimen. Penelitian eksperimen bertujuan untuk mengembangkan dan menguji model prediksi risiko penyakit diabetes menggunakan algoritma *Decision Tree*. Dalam penelitian ini, pendekatan kuantitatif diterapkan dengan memanfaatkan data medis pasien, yang mencakup berbagai variabel seperti usia, indeks massa tubuh (BMI), tekanan darah, dan riwayat keluarga. Pada bab ini akan dijelaskan tahapan-tahapan penelitian, mulai dari tahap tinjauan pustaka hingga tahap evaluasi hasil prediksi Gambar 1.



Gambar 1. Tahap Penelitian

3.1. Literatur Review

Tahap pertama dalam penelitian ini adalah melakukan literatur review. Proses ini dilakukan untuk memahami teori dan penelitian sebelumnya yang relevan dengan topik prediksi penyakit diabetes menggunakan algoritma *Decision Tree*. Literatur yang dikaji meliputi jurnal, buku, dan artikel ilmiah yang membahas algoritma *Decision Tree*, metode klasifikasi pada data medis, serta teknik pengolahan data yang tepat untuk memprediksi diabetes. Tinjauan pustaka ini membantu dalam menentukan metodologi yang paling sesuai untuk mencapai tujuan penelitian.

3.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data yang berasal dari *Kaggle Diabetes prediction dataset* oleh Mohammed Mustafa[17]. Data ini Memiliki 9 atribut yaitu kumpulan indikator atau faktor risiko diabetes. Contoh data bisa dilihat pada Gambar 2.

gender	age	hypertension	heart_disea...	smoking_hi...	bmi	HbA1c_level	blood_gluco...	diabetes
Female	80	0	1	never	25.190	6.600	140	0
Female	54	0	0	No Info	27.320	6.600	80	0
Male	28	0	0	never	27.320	5.700	158	0
Female	36	0	0	current	23.450	5	155	0
Male	76	1	1	current	20.140	4.800	155	0
Female	20	0	0	never	27.320	6.600	85	0
Female	44	0	0	never	19.310	6.500	200	1
Female	79	0	0	No Info	23.860	5.700	85	0
Male	42	0	0	never	33.640	4.800	145	0
Female	32	0	0	never	27.320	5	100	0
Female	53	0	0	never	27.320	6.100	85	0
Female	54	0	0	former	54.700	6	100	0
Female	78	0	0	former	36.050	5	130	0

Gambar 2. Dataset

3.3. Penerapan Algoritma Decision Tree

Tahap ini merupakan proses inti dalam penelitian, yaitu penerapan algoritma *Decision Tree*. Algoritma *Decision Tree* adalah metode yang digunakan untuk membuat model prediksi dengan membangun serangkaian aturan keputusan dalam bentuk struktur pohon. Algoritma ini bekerja dengan memulai dari simpul akar (*root node*) yang berisi atribut atau fitur paling penting dalam data, yang ditentukan berdasarkan kriteria seperti *Information Gain* atau *Gini Index*. Data kemudian dibagi menjadi beberapa subset berdasarkan nilai-nilai dari atribut tersebut, dengan setiap subset diwakili oleh cabang yang menghubungkan simpul akar ke simpul anak (*child nodes*).

Proses pembagian ini berlangsung secara rekursif, di mana setiap subset kembali dianalisis untuk menentukan atribut terbaik berikutnya. Pembagian ini berlanjut hingga kondisi tertentu tercapai, misalnya ketika semua data dalam simpul memiliki kelas yang sama atau mencapai kedalaman maksimum yang telah ditentukan [5].

Daun pohon (*leaf nodes*) adalah simpul yang tidak bisa dibagi lagi dan mewakili hasil akhir atau kelas yang diprediksi. Untuk memprediksi data baru, algoritma mengarahkan data tersebut melalui pohon, mengikuti aturan-aturan dari simpul akar hingga mencapai simpul daun yang menghasilkan prediksi [7]. Keunggulan algoritma ini terletak pada interpretabilitas yang tinggi, sehingga sangat bermanfaat untuk aplikasi medis seperti prediksi penyakit, di mana hasil dari model dapat dijelaskan dan digunakan dalam pengambilan keputusan klinis [8].

3.3.1. Membangun Model

Kami menggunakan perangkat lunak *RapidMiner* untuk pembangunan model *Decision Tree*. *RapidMiner* dipilih karena kemampuannya dalam melakukan analisis data secara cepat dan intuitif.

a. Pembagian Data

Pembagian data dilakukan menggunakan metode stratified sampling. *Stratified sampling* berarti membagi data ke dalam beberapa subset (*fold*) sambil memastikan

bahwa proporsi kelas dalam setiap subset mencerminkan distribusi kelas pada dataset asli. *Stratified* sampling dipilih karena dataset memiliki distribusi kelas yang sangat tidak merata.

b. Decision Tree Model

Pada penelitian ini model *Decision Tree* dibangun dengan parameter sebagai berikut :

- **Criterion: Gain Ratio**
Gain ratio Digunakan untuk memilih atribut terbaik di setiap node berdasarkan rasio keuntungan informasi.
- **Maximal Depth: 20**
Hal ini mencegah pohon menjadi terlalu kompleks (*overfitting*) dengan membatasi jumlah level yang dapat dibangun.
- **Apply Pruning: Checked**
Mengurangi kompleksitas pohon dan risiko *overfitting*.
- **Confidence: 0.25**
Nilai ini menentukan batas kepercayaan (*confidence threshold*) dalam memutuskan apakah cabang tertentu akan dipangkas.
- **Apply Prepruning: Checked**
Membantu mengurangi risiko *overfitting* dengan menghentikan split yang tidak signifikan.
- **Minimal Gain: 0.1**
Mengurangi kemungkinan split yang tidak relevan atau terlalu kecil kontribusinya.
- **Minimal Leaf Size: 2**
Menghindari *overfitting* dengan mencegah node terbentuk dari jumlah data yang terlalu kecil.

3.4. Evaluasi Hasil

Tahap akhir dari proses pembangunan model prediksi adalah melakukan evaluasi terhadap performa model. Evaluasi ini mencakup pengukuran akurasi, presisi, recall, dan nilai F1-score pada data uji. Hasil evaluasi tersebut bertujuan untuk menilai seberapa efektif algoritma *Decision Tree* dalam memprediksi penyakit diabetes pada dataset yang digunakan.

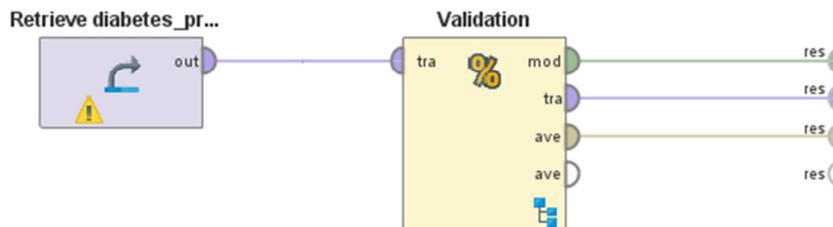
4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan aplikasi *RapidMiner* untuk melakukan perhitungan data dengan algoritma *Decision Tree*. Data yang dipakai memiliki 9 atribut dengan *class target* berupa diabetes. Pada atribut diabetes 1 berarti memiliki penyakit diabetes dan 0 berarti tidak memiliki penyakit diabetes.

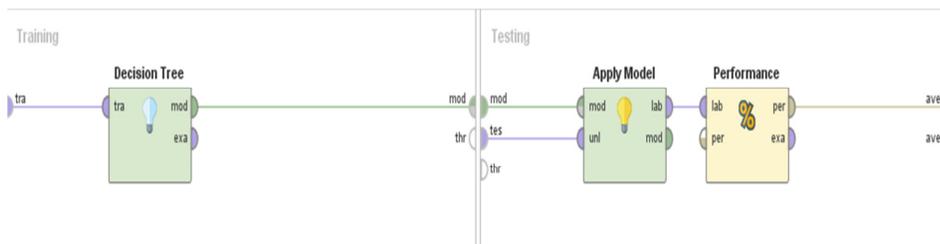
Pada aplikasi *RapidMiner* langkah pertama adalah memasukan dataset. Setelah dataset berhasil dimasukan set role pada atribut. Untuk atribut diabetes diberi role label. Gambar 3 adalah dataset setelah dimasukan dan diolah pada aplikasi *RapidMiner*. Data yang sudah diolah akan digunakan untuk melakukan perhitungan algoritma *Decision Tree*.

Row No.	diabetes	gender	age	hypertension	heart_disea...	smoking_hi...	bmi	HbA1c_level	blood_gluco...
1	0	Female	80	0	1	never	25.190	6.600	140
2	0	Female	54	0	0	No Info	27.320	6.600	80
3	0	Male	28	0	0	never	27.320	5.700	158
4	0	Female	36	0	0	current	23.450	5	155
5	0	Male	76	1	1	current	20.140	4.800	155
6	0	Female	20	0	0	never	27.320	6.600	85
7	1	Female	44	0	0	never	19.310	6.500	200
8	0	Female	79	0	0	No Info	23.860	5.700	85
9	0	Male	42	0	0	never	33.640	4.800	145
10	0	Female	32	0	0	never	27.320	5	100

Gambar 3. Dataset setelah diolah



Gambar 4. Model



Gambar 5. Model algoritma dan pengujian

Gambar 4 adalah model untuk melakukan perhitungan *Decision Tree*. Dalam menyusun model terdapat operator yang diperlukan yaitu Retrieve dan X-validation. Gambar 5 merupakan isi dari X-validation yang terdapat operator *Decision Tree*, Apply Model, dan Performance.

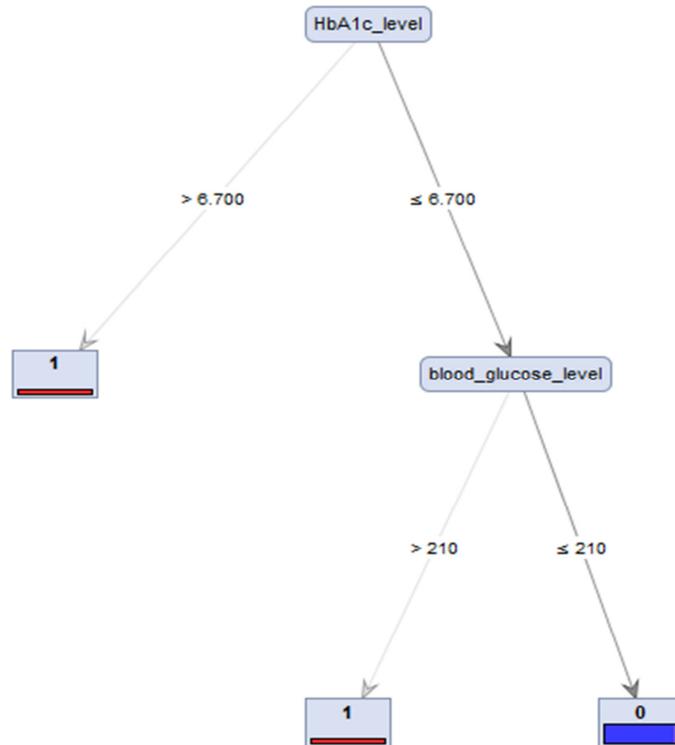
accuracy: 97.19% +/- 0.13% (mikro: 97.19%)

	true 0	true 1	class precision
pred. 0	91500	2813	97.02%
pred. 1	0	5687	100.00%
class recall	100.00%	66.91%	

Gambar 6. Akurasi

Model diatas menghasilkan tingkat akurasi pengukuran seperti Gambar 6. Model memiliki akurasi sebesar 97.19% dengan margin error $\pm 0.13\%$. Akurasi ini menunjukkan proporsi total prediksi yang benar dari seluruh data, baik untuk kelas 0 (negatif) maupun kelas 1 (positif). Untuk *Class Precision* kelas 0: 97.02%, artinya dari seluruh prediksi negatif, 97.02% benar-benar negatif. Untuk kelas 1: 100.00%, artinya dari seluruh

prediksi positif, 100% benar-benar positif. *Class Recall* untuk kelas 0: 100.00%, yang berarti semua sampel negatif diprediksi dengan benar dan untuk kelas 1: 66.91%, yang berarti hanya 66.91% dari sampel positif yang diprediksi dengan benar.



Gambar 7. Pohon Keputusan

Berdasarkan pohon keputusan Gambar 7 :

- jika $HbA1c_level > 8.7$, model mengklasifikasikan individu tersebut ke kelas 1 (positif diabetes).
- Jika $HbA1c_level \leq 8.7$, keputusan selanjutnya didasarkan pada nilai $blood_glucose_level$.
- Jika $blood_glucose_level > 210$, maka individu diklasifikasikan ke kelas 1 (positif diabetes).
- Jika $blood_glucose_level \leq 210$, individu diklasifikasikan ke kelas 0 (negatif diabetes).

Interpretasi hasil penelitian ini menunjukkan bahwa algoritma *Decision Tree* mampu memprediksi risiko diabetes dengan tingkat akurasi tinggi (97,19%), yang sesuai dengan tujuan penelitian, yaitu untuk mengembangkan model prediksi yang dapat membantu mendeteksi individu dengan risiko diabetes secara efektif. Hasil ini menunjukkan bahwa algoritma *Decision Tree* berhasil menangkap pola yang relevan dalam data medis, terutama melalui variabel utama seperti $HbA1c$ level dan $blood$ glucose level, yang keduanya merupakan indikator signifikan dalam diagnosis diabetes.

Namun, meskipun tujuan utama penelitian tercapai, terdapat aspek yang perlu diperhatikan, yaitu recall pada kelas positif yang rendah (66,91%). Hal ini menunjukkan bahwa model memiliki keterbatasan dalam mendeteksi semua individu yang benar-benar memiliki diabetes. Rendahnya recall ini kemungkinan muncul karena distribusi data yang tidak seimbang (*class imbalance*) atau karena kompleksitas model yang kurang mampu menangkap semua variasi pada kelas positif. Di sisi lain, precision yang tinggi pada kelas positif (100%) mengindikasikan bahwa model sangat baik dalam memastikan individu yang diprediksi positif memang benar-benar memiliki diabetes, sehingga dapat meminimalkan risiko kesalahan diagnosis positif.

Secara keseluruhan, hasil yang diperoleh sesuai dengan tujuan penelitian dalam hal akurasi dan kemudahan interpretasi, tetapi menunjukkan perlunya optimalisasi untuk mengatasi kelemahan pada recall kelas positif. Optimalisasi ini bisa dilakukan dengan teknik seperti *oversampling*, penggunaan metrik evaluasi tambahan, atau penggabungan dengan algoritma lain yang lebih sensitif terhadap data minoritas. Hasil yang muncul ini sejalan dengan pemahaman bahwa algoritma *Decision Tree*, meskipun efektif dan mudah diinterpretasikan, sering menghadapi tantangan dalam mengelola ketidakseimbangan data.

Perbandingan hasil eksperimen dengan penelitian sebelumnya Penelitian ini menghasilkan model prediksi diabetes menggunakan algoritma *Decision Tree* dengan akurasi 97,19%, yang sedikit lebih tinggi dibandingkan penelitian sebelumnya oleh Fida Maisa Hana[10] yang menggunakan algoritma *Decision Tree* C4.5 dengan akurasi 97,12%. Perbedaan kecil ini menunjukkan bahwa kedua model memiliki performa yang hampir setara dalam memprediksi diabetes, meskipun perbedaan atribut dataset memengaruhi hasil akhir. Hal ini mengonfirmasi keandalan algoritma *Decision Tree* dalam aplikasi klasifikasi medis, termasuk deteksi dini risiko diabetes.

5. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa algoritma *Decision Tree* dapat digunakan secara efektif untuk memprediksi risiko diabetes berdasarkan data medis. Model yang dibangun dengan menggunakan aplikasi *RapidMiner* ini mencapai akurasi sebesar 97,19% dengan margin error $\pm 0,13\%$, yang menunjukkan tingkat prediksi yang sangat baik untuk mendeteksi diabetes. Namun, model ini memiliki kelemahan dalam recall pada kelas positif (66,91%), yang berarti tidak semua kasus positif terdeteksi dengan akurat. Variabel *HbA1c level* dan *blood glucose level* ditemukan sebagai faktor utama dalam klasifikasi risiko diabetes, sehingga keduanya menjadi penentu dalam hasil prediksi. Kelebihan utama dari algoritma *Decision Tree* adalah kemampuannya untuk menghasilkan model yang mudah diinterpretasikan, yang membantu tenaga medis dalam pengambilan keputusan klinis. Meski demikian, masih terdapat risiko *overfitting* yang perlu diatasi agar model lebih dapat diandalkan dalam berbagai kondisi data. Pengembangan lebih lanjut dapat dilakukan dengan mengoptimalkan model atau

menggabungkannya dengan algoritma lain untuk meningkatkan recall dan akurasi, serta menguji model ini pada dataset yang lebih besar untuk meningkatkan generalisasi.

DAFTAR PUSTAKA

- [1] World Health Organization, "Diabetes," 2021. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- [2] International Diabetes Federation, "IDF Diabetes Atlas, 9th ed.," 2019. [Online]. Available: <https://www.diabetesatlas.org/>.
- [3] American Diabetes Association, "Diagnosis and Classification of Diabetes Mellitus," *Diabetes Care*, vol. 43, no. 1, pp. 25-33, 2020.
- [4] C. H. Lee and M. W. Yoon, "Trends in the Epidemiology of Diabetes in Asia-Pacific Regions," *Current Diabetes Reports*, vol. 15, no. 6, pp. 465-473, 2019.
- [5] F. Johnson and L. White, "Application of Data Mining in Health Sector: A Review of Literature," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 175-183, 2020.
- [6] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham: Morgan Kaufmann, 2018.
- [7] M. Ahmed, S. Rajendran, and S. Saleh, "Predicting Diabetes Using Decision Tree Algorithm with High Accuracy," *IEEE Access*, vol. 8, pp. 71221-71230, 2020.
- [8] S. L. Choi, B. B. Lee, and C. J. Kim, "Comparative Study on Decision Tree and Random Forest Algorithms in Diabetes Prediction," *Journal of Healthcare Informatics Research*, vol. 4, no. 2, pp. 123-136, 2021.
- [9] D. Ismafillah, T. Rohana, dan Y. Cahyana, "Analisis algoritma pohon keputusan untuk memprediksi penyakit diabetes menggunakan oversampling smote," *INFOTECH J. Inform. Teknol.*, vol. 4, no. 1, pp. 27-36, Jun. 2023, doi: 10.37373/infotech.v4i1.452.
- [10] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32-39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [11] A. Afifuddin dan L. Hakim, "Deteksi Penyakit Diabetes Mellitus Menggunakan Algoritma Decision Tree Model Arsitektur C4.5," *J. Krisnadana*, vol. 3, p. null, 2023, doi: 10.58982/krisnadana.v3i1.470.
- [12] F. Elfaladonna dan A. Rahmadani, "Analisa Metode Classification-Decision Tree Dan Algoritma C.45 Untuk Memprediksi Penyakit Diabetes Dengan Menggunakan Aplikasi Rapid Miner," *SINTECH (Science Inf. Technol. J.)*, vol. 2, no. 1, pp. 10-17, 2019, doi: 10.31598/sintechjournal.v2i1.293.
- [13] P. B. N. Setio, D. R. S. Saputro, dan B. Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prisma Pros. Semin. Nas. Mat.*, vol. 3, pp. 64-71, 2020.
- [14] N. N. Habibah, A. Nazir, I. Iskandar, F. Syafria, L. Oktavia, dan I. Syurfi, "Pemodelan Klasifikasi Untuk Menentukan Penyakit Diabetes dengan Faktor Penyebab Menggunakan Decision Tree C4.5 Pada Wanita," *J. Sist. Komput. dan Inform.*, vol. 4, no. 4, p. 654, 2023, doi: 10.30865/json.v4i4.6202.

- [15] W. Aprilia, I. Kurniawan, M. Baydhowi, dan T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sistemasi*, vol. 10, no. 1, p. 163, 2021, doi: 10.32520/stmsi.v10i1.1129.
- [16] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, dan A. Christian, "Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes," *Indones. J. Softw. Eng.*, vol. 7, no. 1, pp. 63–68, 2021, doi: 10.31294/ijse.v7i1.10279.
- [17] M. Tz, "Diabetes Prediction Dataset," Kaggle, Sep. 2022. Accessed: Nov. 7, 2024. [Online]. Available: <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset/data>.