



Implementation of Decision Tree Method for Diabetes Mellitus Type 2 Prediction

Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2

**Muhammad Fahrul Aditya¹, Andri Pramuntadi^{2*},
Dhina Puspasari Wijaya³, Yanuar Wicaksono⁴**

^{1,2,3}Departemen of Informatics, Faculty of Computers and Engineering,
Universitas Alma Ata, Yogyakarta, Indonesia

⁴Departemen of Information Systems, Faculty of Computers and Engineering
Universitas Alma Ata, Yogyakarta, Indonesia

E-Mail: ¹193200058@almaata.ac.id, ²andripramuntadi@almaata.ac.id,
³dhina.puspa@almaata.ac.id, ⁴yanuar@almaata.ac.id

Received Feb 13th 2024; Revised Apr 05th 2024; Accepted Jun 12th 2024
Corresponding Author: Andri Pramuntadi

Abstract

Type 2 Diabetes Mellitus is a chronic condition wherein the pancreas either fails to produce sufficient insulin or the body is unable to effectively utilize the produced insulin. Various efforts have been made to alleviate diabetes, and one approach is early detection. The rapid advancement of technology has significantly impacted the medical field, particularly in the realm of early disease detection. The process involves the use of programmed tools intervened by artificial intelligence. Several artificial intelligence methods serve as predictive models, and one such method is the Decision Tree (DT). This research aims to implement and construct a predictive model for diabetes using a dataset sourced from the Mlati II Public Health Center in Sleman Regency. The model in this study utilizes the Decision Tree method, with parameters set as follows: criterion = 'entropy', splitter = 'best', max_depth = None, min_samples_split = 4, min_samples_leaf = 10. The model's evaluation process employs accuracy, precision, recall, and f1-score, yielding values of 92%, 0.92, 0.915, and 0.915, respectively.

Keyword: Accuracy, Decision Tree, Diabetes Melitus Type 2, K-Fold Cross Validation, Prediction

Abstrak

Diabetes Melitus Tipe 2 merupakan penyakit kronis yang apabila pankreas tidak memproduksi cukup insulin atau ketika tubuh tidak mampu menggunakan insulin yang diproduksi secara efektif. Berbagai cara telah di upayakan untuk mengurangi penderita diabetes salah satunya adalah dengan deteksi dini. Kemajuan teknologi yang sangat pesat berdampak pada dunia medis, salah satunya adalah untuk deteksi dini suatu penyakit. Proses deteksi suatu penyakit menggunakan alat yang telah diprogram dan di intervensi oleh kecerdasan buatan. Terdapat banyak metode dari kecerdasan buatan yang digunakan sebagai model prediksi, salah satunya adalah Decision Tree (DT). Penelitian ini bertujuan untuk mengimplementasikan dan membangun sebuah model prediksi penyakit Diabetes. Dataset yang digunakan adalah dataset yang berasal dari Puskesmas Mlati II Kabupaten Sleman. Model dalam penelitian ini menggunakan metode Decision Tree (Pohon Keputusan), adapun parameter yang digunakan adalah criterion = 'entropy', splitter = 'best', max_depth = None, min_samples_split = 4, min_samples_leaf = 10. Proses evaluasi model menggunakan akurasi, presisi, recall, dan f1-score yang masing-masing menghasilkan nilai 92%, 0.92, 0.915, 0.915.

Kata Kunci: Akurasi, Decision Tree, Diabetes Melitus Tipe 2, K-Fold Cross Validation, Prediksi

1. PENDAHULUAN

Diabetes merupakan penyakit kronis yang apabila pankreas tidak memproduksi cukup insulin atau ketika tubuh tidak mampu menggunakan insulin yang diproduksi secara efektif. Insulin merupakan hormon yang mengatur gula darah. Diabetes, adalah akibat umum dari gula darah yang tidak kontrol dan seiring berjalannya waktu menyebabkan kerusakan pada banyak sistem tubuh, terutama saraf dan pembuluh darah[1].

Pada tahun 2014, 8,5 persen orang di atas usia 18 tahun menderita diabetes. Diabetes bertanggung jawab langsung atas 1,5 juta kematian pada tahun 2019, dan hampir setengah dari seluruh kematian akibat diabetes terjadi pada orang di bawah usia 70 tahun. Penyakit ginjal menyebabkan 460.000 kematian, beberapa di antaranya disebabkan oleh diabetes, dan sekitar 20% penyakit kardiovaskular disebabkan oleh tingginya kadar gula darah. Antara tahun 2000 dan 2019, angka kematian akibat diabetes berdasarkan usia meningkat sebesar 3 persen. Di negara-negara berkembang dengan pendapatan rata-rata lebih rendah, angka kematian akibat diabetes meningkat sebesar 13%. Sebaliknya, kemungkinan kematian akibat salah satu dari empat penyakit tidak menular antara usia 30 dan 70 tahun menurun sebesar 22% di seluruh dunia antara tahun 2000 hingga 2019[2]. Diabetes melibatkan beberapa faktor risiko yang dapat dikelompokkan menjadi faktor risiko yang tidak dapat dimodifikasi seperti ras dan etnis; usia; seksual; dan riwayat keluarga serta faktor risiko yang dapat dimodifikasi seperti obesitas; kegemukan; kurang aktivitas fisik, hipertensi; penurunan toleransi glukosa atau gula darah puasa sebelumnya, dan merokok. Dengan banyaknya data faktor resiko penderita diabetes tersebut sangat diperlukan bantuan teknologi *Artificial Intelligence* (AI) untuk membantu pakar kesehatan dalam mendiagnosis pasien terkena diabetes. AI telah diterapkan di berbagai bidang medis untuk berbagai tujuan. AI didefinisikan sebagai "bidang ilmu pengetahuan dan teknologi yang biasanya berkaitan dengan pemahaman komputasi perilaku cerdas dengan menciptakan artefak yang menunjukkan perilaku tersebut." [3].

Kedokteran modern menghadapi tantangan terkait sejumlah besar perolehan data, analisis, dan penerapan pengetahuan yang diperoleh untuk memecahkan masalah klinis yang kompleks, maka perlu menggunakan kemampuan AI untuk tujuan ini [4]. Algoritma AI memiliki potensi untuk memberikan perawatan yang lebih baik terutama dalam kombinasi dengan teknologi terkini [5]. AI sangat dibutuhkan pada era medis sekarang ini dikarenakan proses analisisnya yang sangat cepat dan mampu untuk memproses banyak sekali data, tentu hal ini akan terbalik jika AI dalam dunia medis tidak ditemukan maka proses diagnosa akan berjalan lambat lalu memiliki efek samping berupa kesalahan diagnosis hingga keterlambatan penanganan pasien. Cabang dari AI yang dapat membantu pakar kesehatan dalam mendiagnosis pasien diabetes dengan algoritma *Machine Learning* (ML). ML didefinisikan sebagai teknik mengubah data menjadi informasi yang bermakna. Dalam hal diabetes, informasi terkait diabetes dapat dihasilkan atau dikumpulkan melalui sensor canggih, mesin digital, pengurutan DNA lanjutan, tes darah, mikroskop digital resolusi super, spektrometri massa, *Magnetic Resonance Imagery* (MRI) dan peralatan medis canggih lainnya di mana a data produksi massal yang sangat besar dapat disimpan ke dalam *Electronic Health Record* (EHR) [6]. Algoritma ML dapat digunakan dalam berbagai bidang untuk mempelajari data. Beberapa penerapan ML dalam kehidupan adalah antara lain pada bidang kesehatan. Bidang kesehatan merupakan salah satu bidang yang menerapkan ML dalam proses diagnosa penyakit. ML dalam kesehatan digunakan untuk memprediksi penyakit yang diderita oleh pasien, contohnya adalah Diabetes Mellitus Tipe 2 (DMT2) [7]. Untuk menyediakan teknologi yang lebih aman melalui AI, direkomendasikan untuk memiliki desain yang aman, cadangan keselamatan, dan perlindungan prosedural, dengan semua ketidakpastian diidentifikasi untuk semua sistem teknis potensial [8].

Metode DT merupakan *supervised learning* digunakan sebagai klasifikasi dan regresi. Metode DT memiliki karakteristik seperti pohon atau hierarki yang terdiri dari simpul akar, cabang, simpul internal, dan simpul daun. Penelitian terdahulu yang dilakukan oleh Dewi Anggraeni dan Ramdhani (2018) dalam penelitiannya yang berjudul "Analisa Decision Tree Untuk Prediksi Diagnosa Diabetes Melitus". Jenis penelitian ini merupakan deskriptif dengan metode pendekatan kuantitatif. Persamaan penelitian sebelumnya dengan penelitian ini adalah metode yang digunakan yaitu Decision Tree (DT) dan juga objek penelitian adalah Diabetes Melitus [9].

Indonesia memiliki fasilitas kesehatan di tingkat kecamatan yang dinamakan dengan Pusat Kesehatan Masyarakat (Puskesmas). Puskesmas merupakan fasilitas layanan kesehatan yang menyelenggarakan upaya pelayanan kesehatan untuk tingkat pertama [10]. Puskesmas tersebar di seluruh kecamatan seindonesia salah satunya di Kecamatan Mlati, Kabupaten Sleman, Yogyakarta. Kecamatan Mlati, Kabupaten Sleman memiliki dua fasilitas kesehatan tingkat pertama yaitu Puskesmas Mlati 1 dan Puskesmas Mlati 2. Pada penelitian ini, eksperimen dilakukan dengan menggunakan dataset yang berasal dari Puskesmas Mlati 2. Berdasarkan uraian diatas, permasalahan DMT 2 merupakan kasus penyakit dengan kasus terbesar secara global sehingga sampai saat ini masih dilakukan penelitian mengenai *early warning system* (deteksi dini) terhadap penyakit DMT 2. ML merupakan salah satu cara untuk melakukan deteksi dini dengan memprediksi sebuah dataset yang berisi faktor diabetes terhadap seorang pasien. Penelitian ini dimaksudkan untuk menguji metode Decision Tree menggunakan dataset dari Puskesmas Mlati II Sleman dalam memprediksi penyakit Diabetes Melitus Tipe 2.

Adapun penelitian yang telah dilakukan oleh Mariawan Ahmed Hama Saeed (2023) dalam jurnalnya yang berjudul "*Diabetes Type 2 Classification using Machine Learning Algorithms with Up-Sampling Technique*", penelitian ini bertujuan membandingkan akurasi dengan menggunakan dua dataset yang berbeda yaitu PIDD dan BRFS. Hasil yang didapatkan akurasi tertinggi diperoleh pada dataset BRFS, hal ini dapat terjadi karena jumlah variabel yang berbeda antara dua dataset tersebut [11].

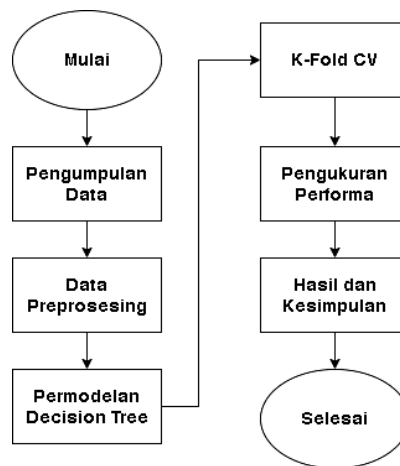
Penelitian kedua dilakukan oleh Dewi Rahma Ente et. al (2020), pada penelitian yang tersebut menggunakan dataset yang berasal dari rumah sakit dengan metode DT dan teknik validasi K-Fold Cross

Validation dengan 10 sebagai nilai K. Penelitian ini menghasilkan akurasi sebesar 97,69% pada fold ke 7 hingga fold ke 10 [12].

Pada penelitian yang dilakukan menggunakan metode DT dan teknik K-Fold Cross Validation. Penelitian ini bertujuan untuk menguji akurasi dari klasifikasi oleh metode Decision Tree dengan menggunakan dataset yang didapatkan dari Puskesmas Mlati II Sleman. Proses validasi menggunakan K-Fold Cross Validation dengan 5 sebagai nilai K. Penelitian ini juga melakukan uji performa dari metode DT dengan *Confusion Matrix* dan menghitung nilai akurasi, presisi, *recall*, dan *f1-score*.

2. METODOLOGI PENELITIAN

Pada tahap ini merupakan alur dari penelitian yang akan dilakukan. Gambar 1 menjelaskan detail tahap pelaksanaan. Dimulai dengan pengumpulan data yang bersumber dari Puskesmas Mlati II Sleman, kemudian dataset dilakukan *preprocessing* agar skala dataset atribut satu dengan lainnya seimbang. Selanjutnya implementasi metode Decision Tree dengan menggunakan 5 parameter diantaranya adalah *Criterion*, *Splitter*, *max_depth*, *min_sample_split*, *min_sample_leaf*. Proses penerapan metode dapat dilihat pada gambar 1.



Gambar 1. Alur penelitian

Proses penerapan metode diawali dengan pengumpulan data berupa dataset diabetes melitus tipe 2 yang didapatkan dari Puskesmas Mlati 2 Kabupaten Sleman. Setelah dilakukan pengumpulan data selanjutnya akan dilakukan tahap preproses yaitu dengan mencari nilai null pada atribut dataset dan memisahkan data kategori dengan numerik. Tahap berikutnya adalah melakukan implementasi metode Decision Tree untuk mengetahui akurasi dari model. Setelah didapatkan akurasi model Decision Tree, selanjutnya adalah proses validasi menggunakan teknik K-Fold Cross Validation dengan nilai K yaitu 5. Setelah didapatkan hasil akurasi yang sudah divalidasi dengan menggunakan K-Fold CV, maka selanjutnya adalah melakukan pengukuran performa dengan menghitung nilai akurasi, presisi, *recall*, dan *f1-measure*.

2.1 Diabetes Melitus Tipe 2

Diabetes melitus (DM) atau kencing manis merupakan penyakit yang umum terjadi. Diabetes adalah penyakit yang berhubungan dengan produksi insulin tidak normal, ketidakmampuan menggunakan insulin, maupun keduanya. Gejala umum penderita diabetes antara lain sering buang air kecil, rasa haus ringan, cepat merasa lapar, penurunan berat badan cepat, kulit kering dan gatal, luka sulit sembuh, penglihatan buruk, dan lemas adalah salah satu gejalanya. Hal ini menyebabkan sakit kepala, obesitas, infeksi jamur dan bakteri [13].

2.2 Data Mining

Data Mining (Penambangan data) juga disebut sebagai *Knowledge Discovery in Databases* (KDD), adalah ekstraksi otomatis dari pola yang mewakili pengetahuan yang secara tidak langsung disimpan dalam basis data besar, dan penyimpanan informasi besar lainnya dan menerapkan bidang multidisiplin termasuk teknologi basis data, kecerdasan buatan, *machine learning*, jaringan saraf, statistik, pengenalan pola, sistem berbasis pengetahuan, perolehan pengetahuan, pengambilan informasi, komputasi kinerja tinggi, dan visualisasi data [14]. *Data mining* sudah diterapkan dalam berbagai macam disiplin ilmu, salah satunya adalah kesehatan. Diantara penerapan konsep dan teknik *data mining* dalam kesehatan adalah sebagai alat untuk memprediksi suatu permasalahan kesehatan, diantaranya adalah penyakit diabetes [15]. Dalam *data mining* ada beberapa pendekatan seperti klusterisasi, klasifikasi, asosiasi. Pada penelitian ini dilakukan teknik klasifikasi yaitu pengelompokan data, data yang dikelompokkan berlandaskan relasi atau hubungan data terhadap data sampel [16].

2.3 Metode Decision Tree

Dalam penelitian ini, model machine learning yang digunakan adalah Decision Tree (DT). Model DT merupakan salah satu metode klasifikasi non parametrik. Disebut sebagai Decision Tree dikarenakan metode ini memiliki gambaran model hierarki atau struktur seperti pohon yang terdiri dari simpul akar, simpul internal dan simpul daun [17]. Menurut Feng-Jen Yang (2019), DT merupakan metode dimana setiap jalur mulai dari akar mewakili urutan pemisahan data hingga hasil Boolean tercapai di simpul daun. Setiap jalur pada metode DT merupakan aturan keputusan yang dapat dengan mudah diterjemahkan dalam bahasa manusia [18].

DT pertama kali dikenalkan pada tahun 1950-an dan 1960-an. Pada saat itu, peneliti mulai mengeksplorasi gagasan menggunakan model berbasis pohon untuk mengambil sebuah keputusan. Salah satu yang paling awal adalah algoritma yang dikembangkan oleh C. A. R. Hoare pada tahun 1961 [19]. Menurut Alsaman et al., (2019) pohon keputusan merupakan teknik yang umum digunakan untuk membangun model klasifikasi berdasarkan kumpulan data [20]. Pada penelitian ini, parameter yang digunakan adalah *Criterion*, *Splitter*, *max_depth*, *min_sample_split*, *min_sample_leaf*. Berikut proses tahapan dalam klasifikasi Decision Tree:

1. Menentukan parameter *Criterion*='entropy', sebagai nilai awal menentukan node akar. Secara matematis ditulis sebagai berikut:

$$Entropy(S) = -p \times \log_2(n) - q \times \log_2(n) \quad (1)$$

2. Menghitung *Gain* untuk menentukan node akar yang dipakai. Rumus dalam menentukan *Gain* sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \times Entropy(S_i) \quad (2)$$

Keterangan:

S : Kasus
p : Total input positif
q : Total input negatif
n : Jumlah kasus
A : Atribut
n : Jumlah partisi S
pi : Proporsi Si terhadap S

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Dataset yang digunakan pada penelitian ini yaitu dataset klinis Puskesmas Mlati II Kabupaten Sleman, Daerah Istimewa Yogyakarta. Pada pengumpulan dataset didapatkan melalui laboratorium puskesmas yang sebelumnya pasien sudah dilakukan pemeriksaan. Dalam dataset ini berisi 200 baris yang melibatkan pasien pria dan wanita. Dataset ini terdapat 5 atribut yaitu usia, tekanan darah, BMI, GDP atau Gula Darah Puasa, dan sebagai klasifikasi yaitu atribut catatan terdiri dari kelas terkena dan tidak terkena.

3.2. Preprocessing Data

Pada tahap *preprocessing* ini melakukan pemeriksaan dataset dari nilai null atau *missing values* dan normalisasi. Pada dataset penelitian ini terdapat jenis tipe data objek yaitu pada variabel catatan diabetes. Jenis data objek tersebut tidak bisa dilakukan proses ke tahap selanjutnya, oleh karena itu dilakukan perubahan ke dalam jenis tipe numerik menggunakan fungsi *map()* pada *library* seperti Pandas. Pemetaan yang dilakukan adalah mengubah Terkena menjadi nilai 1 dan Tidak Terkena menjadi nilai 0.

Tahap berikutnya adalah melakukan perubahan data sehingga terletak pada rentang 0 hingga 1 menggunakan min-max normalization, hal ini bertujuan untuk menghindari permasalahan skala yang tidak seimbang antar atribut. Tahap *min-max normalization* setiap baris akan dicari baris minimum dan maximumnya. Berikut merupakan formula dari *min-max normalization*:

$$X_{new} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (2)$$

Keterangan:

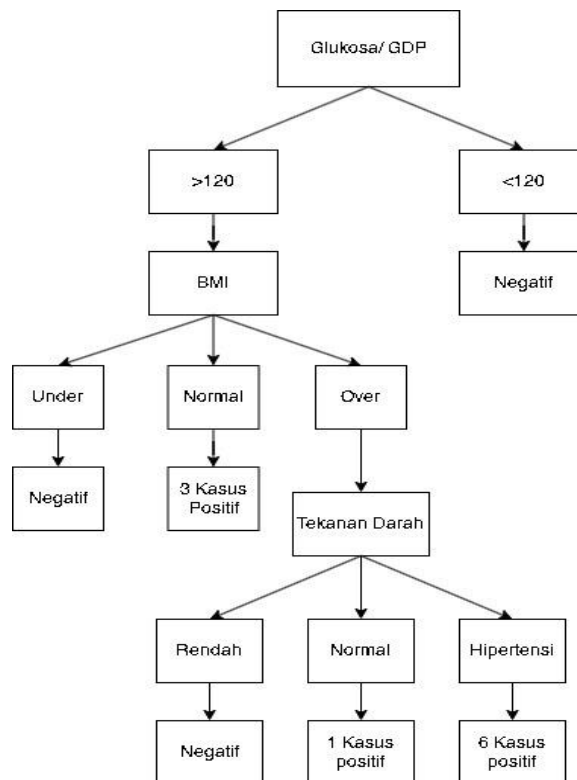
X_new : Nilai baru dari hasil normalisasi
X : Nilai lama
Max (X) : Nilai maksimum dalam dataset

Min (X) : Nilai minimum dalam dataset

Untuk menemukan hasil X_{new} dari setiap variabel, tahap awal adalah menentukan nilai minimum dan nilai maksimum.

3.3. Permodelan Decision Tree

Dalam perhitungan menentukan suatu pohon keputusan, dilakukan perhitungan parameter entropi dan gain terlebih dahulu, hal ini digunakan untuk mengukur keberagaman dari kumpulan sampel data. Dari hasil pengukuran entropi dan gain, maka didapatkan atribut sebagai node akar yaitu Glukosa (GDP). Berdasarkan node yang didapatkan, maka *rules* dari Decision Tree ditunjukkan pada gambar 2.



Gambar 2. Pohon Keputusan Hasil Node Akhir

Dari perhitungan diatas, diadaptkan *rules* yang dijadikan sebagai referensi dalam proses prediksi diabetes. Terdapat enam *rules* keputusan diantaranya adalah tiga *rules* positif dan tiga *rules* negatif dengan penjabaran sebagai berikut:

1. Jika Glukosa/ GDP (Gula Darah Puasa) >120 dan BMI *Under*, maka hasil = Negatif.
2. Jika Glukosa/ GDP >120 dan BMI *Normal*, maka hasil = 3 kasus positif, 0 kasus negatif.
3. Jika Glukosa/ GDP >120 dan BMI *Over* dan Tekanan Darah Rendah, maka hasil = Negatif.
4. Jika Glukosa/ GDP >120 dan BMI *Over* dan Tekanan Darah Normal, maka hasil = 1 kasus positif, 0 kasus negatif.
5. Jika Glukosa/ GDP >120 dan BMI *Over* dan Tekanan Darah Hipertensi, maka hasil = 6 kasus positif, 0 kasus negatif.
6. Jika Glukosa/ GDP <120, maka hasil = Negatif

3.4. K-Fold Cross Validation

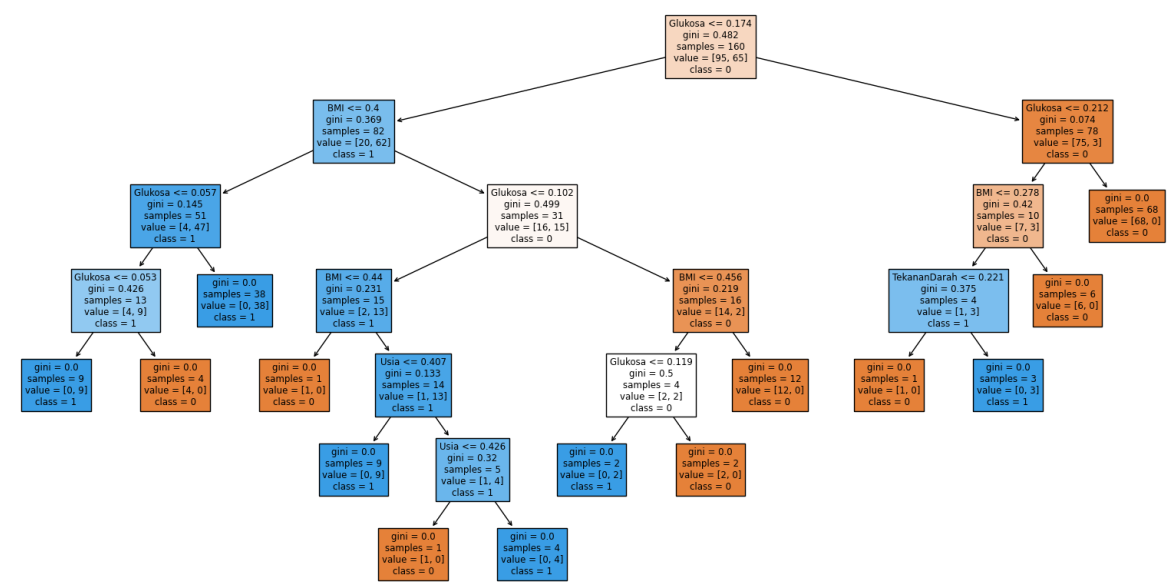
Setelah dilakukan permodelan menggunakan metode DT, langkah selanjutnya adalah memvalidasi hasil dari permodelan tersebut menggunakan K-Fold Cross Validation. Pada tahap ini data akan dibagi kedalam subset sebanyak nilai K, penelitian ini menggunakan 5 sebagai nilai K. Akurasi tertinggi pada proses validasi ini ditunjukkan oleh iterasi ke 2 atau fold ke 2 dengan nilai 92%. Tabel 1 merupakan hasil dari Cross Validation.

Tabel 1. Confusion Matrix

Fold ke-	Akurasi	Training Time
1	0.875	0.0009977817535400390
2	0.950	0.0019941329956054688

Fold ke-	Akurasi	Training Time
3	0.925	0.0009958744049072266
4	0.925	0.0019943714141845703

Dari hasil pengujian K-Fold didapatkan data bahwa fold kedua memiliki nilai akurasi yang lebih tinggi. Hal ini dipengaruhi oleh faktor berupa pembagian data dan randomness. Berikut merupakan gambar pohon dari fold kedua, ditunjukkan pada gambar 3.



Gambar 2. Hasil Fold ke 2

3.5. Pengukuran Performa

Setelah tahap membangun permodelan menggunakan metode Decision Tree, tahap berikutnya adalah melakukan pengukuran performa menggunakan *confusion matrix*, dilanjutkan dengan evaluasi menggunakan akurasi, presisi, *recall*, dan *f1-score*. Berikut hasil *confusion matrix* pada tabel 2.

Tabel 2. Confusion Matrix

Prediksi	Jumlah
True Positive	70
False Positive	6
False Negative	10
True Negative	114

Tabel 2 menunjukkan bahwa terdapat 114 data yang tidak terkena diabetes melitus tipe 2, 70 data terkena diabetes melitus tipe 2, 10 data yang mengalami false negative atau kondisi prediksi tidak terkena diabetes melitus tipe 2 namun pada dataset terkena, dan 6 data mengalami false positive yang berarti kondisi prediksi terkena diabetes melitus tipe 2 namun dalam kasus nyatanya tidak terkena. Evaluasi berikutnya dari hasil *confusion matrix* yaitu berupa nilai rata-rata akurasi, presisi, *recall*, dan *f1-score* pada tabel 6.

Tabel 3. Evaluasi

Parameter	Nilai
Akurasi	0.920
Presisi	0.920
Recall	0.915
F1-score	0.915

4. KESIMPULAN

Kesimpulan dari penelitian ini adalah algoritma Decision Tree dapat digunakan sebagai metode untuk melakukan klasifikasi Diabetes Melitus Tipe 2 dengan baik. Dengan jumlah data 200 dan memiliki 5 atribut, metode DT melakukan pembagian data sebesar 80:20 atau 160 data training dan 40 data testing menggunakan metode train test split. Parameter awal yang digunakan adalah splitter, max_depth, min_samples_split, min_samples_leaf. Disamping menggunakan parameter tersebut, penelitian ini juga menggunakan metode K-

Fold Cross Validation sebagai proses validasi dengan 5 sebagai nilai K. Akurasi yang didapatkan setelah K-Fold Cross Validation adalah $0,919 = 0,92$ atau 92%. Pada tahap akhir penelitian ini adalah dilakukannya proses evaluasi dan validasi dengan menggunakan *Confusion Matrix* untuk mengidentifikasi jumlah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Terakhir pada penelitian ini juga menggunakan F1-Score untuk mengetahui keseimbangan dari model dengan menggabungkan perhitungan presisi dengan recall dengan hasil 0,915.

UCAPAN TERIMA KASIH

Ucapan terima kasih pihak Puskesmas Mlati II Sleman yang telah memberikan kesempatan untuk melaksanakan penelitian.

REFERENSI

- [1] World Health Organization, "Diabetes." Accessed: Mar. 19, 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- [2] Kementerian Kesehatan RI, "Infodatin 2020 Diabetes Melitus," 2020.
- [3] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," *Metabolism*, vol. 69, pp. S36–S40, Apr. 2017, doi: 10.1016/j.metabol.2017.01.011.
- [4] M. Fallah and S. R. N. Kalhori, "Systematic review of data mining applications in patient-centered mobile-based information systems," *Healthc Inform Res*, vol. 23, no. 4, pp. 262–270, Oct. 2017, doi: 10.4258/hir.2017.23.4.262.
- [5] I. Contreras and J. Vehi, "Artificial intelligence for diabetes management and decision support: Literature review," *Journal of Medical Internet Research*, vol. 20, no. 5. JMIR Publications Inc., 2018, doi: 10.2196/10775.
- [6] N. R. Dzakiyullah, M. A. Burhanuddin, R. R. R. Ikram, K. A. Ghani, and W. Setyonugroho, "Machine learning methods for diabetes prediction," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 2199–2205, Oct. 2019, doi: 10.35940/ijitee.L2973.1081219.
- [7] A. Setiadi, "PENERAPAN ALGORITMA MULTILAYER PERCEPTRON UNTUK DETEKSI DINI PENYAKIT DIABETES." [Online]. Available: <http://archive.ics.uci.edu/ml/>.
- [8] S. Ellahham, "Artificial Intelligence: The Future for Diabetes Care," *American Journal of Medicine*, vol. 133, no. 8. Elsevier Inc., pp. 895–900, Aug. 01, 2020. doi: 10.1016/j.amjmed.2020.03.033.
- [9] D. Anggraeni, "Seminar Nasional Royal (SENAR) 2018 ISSN 2622-9986 (cetak) STMIK Royal-AMIK Royal," 2018.
- [10] Puskesmas Sidoarjo, "Tugas dan Fungsi Puskesmas." Accessed: Feb. 01, 2024. [Online]. Available: [https://puskesmasidoarjo.sidoarjokab.go.id/?page=tupoksi#:~:text=Menurut%20Kementerian%20Kesehatan%20RI%2C%20\(2014,tingkat%20pertama%2C%20dengan%20lebih%20mengutamakan](https://puskesmasidoarjo.sidoarjokab.go.id/?page=tupoksi#:~:text=Menurut%20Kementerian%20Kesehatan%20RI%2C%20(2014,tingkat%20pertama%2C%20dengan%20lebih%20mengutamakan)
- [11] M. A. Hama Saeed, "Diabetes type 2 classification using machine learning algorithms with up-sampling technique," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, p. 8, Feb. 2023, doi: 10.1186/s43067-023-00074-5.
- [12] D. Rahma Ente, S. Astuti Thamrin, H. Kuswanto, and S. Arifin, "Klasifikasi Faktor-Faktor Penyebab Penyakit Diabetes Melitus di Rumah Sakit UNHAS Menggunakan Algoritma C4.5," 2020.
- [13] N. Dzakiyullah, "Machine Learning Methods for Diabetes Prediction."
- [14] A. Verma, "The data-mining industry coming of age," *IEEE Intelligent Systems and Their Applications*, vol. 1, no. 1, Nov. 2022, doi: 10.1109/5254.809566.
- [15] Sriyanto and A. Ria Supriyatna, "Prediksi Penyakit Diabetes Menggunakan Algoritma Random Forest," *IJCCS*, vol. x, No.x, pp. 1–5.
- [16] M. G. Pradana, P. H. Saputro, and D. P. Wijaya, "KOMPARASI METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES DALAM KLASIFIKASI PELUANG PENYAKIT SERANGAN JANTUNG," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 5, no. 2, p. 87, Dec. 2022, doi: 10.21927/ijubi.v5i2.2659.
- [17] J. R. Quinlan, "Induction of Decision Trees," 1986.
- [18] F. J. Yang, "An extended idea about decision trees," in *Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019*, Institute of Electrical and Electronics Engineers Inc., Dec. 2019, pp. 349–354. doi: 10.1109/CSCI49370.2019.00068.
- [19] C. A. R. Hoare, "Algorithm 64: Quicksort," *Commun ACM*, vol. 4, no. 7, p. 321, Jul. 1961, doi: 10.1145/366622.366644.
- [20] Y. Shafer Alsalman, N. Khamees Abu Halemah, E. Saleh AlNagi, and W. Salameh, *Using Decision Tree and Artificial Neural Network to Predict Students Academic Performance*. 2019. doi: 10.1109/IACS.2019.8809106.