CIPASUNG TECHNO PESANTREN
Scientific Journal
p-ISSN: 1978-0842 e-ISSN 2963-1114

Analisis *Orange* Data Mining Untuk Klasifikasi Penyakit Diabetes Menggunakan Model *Decision Tree*

Fajar Widianto[1,*]

[1]*Universitas Syeikh Nawawi Banten*

[*]*Corresponding author's email*: Fajarwidianto95@gmail.com

Diterima: DD-MM-YYYY Direvisi: DD-MM-YYYY Diterima setelah revisi: DD-MM-YYYY

Abstract. *One of the global health problems today is diabetes, the prevalence of which continues to increase and therefore requires an effective method for its classification. The purpose of this study is the implementation of Orange Data Mining in the classification of diabetes using the Decision Tree method. The selection of these specifications is due to the fact that the resulting model is easy to understand and can be interpreted. The data analyzed were taken from a public diabetes dataset that includes various health attributes. The analysis process was carried out through preprocessing, splitting, and Juvenile Decision Tree model training. The results showed that the Decision Tree model achieved an accuracy of up to 85% with adequate sensitivity and specificity. decision. Therefore, the conclusion of the study is that increasing the accuracy and quality of diabetes classification can be achieved by the Decision Tree method in Orange Data Mining.*

Keywords: Diabetes; Orange; Decision Tree

Abstrak. Penelitian ini menerapkan Salah satu masalah kesehatan global saat ini adalah penyakit diabetes, yang prevalensinya terus meningkat dan karenanya diperlukan metode yang efektif untuk klasifikasinya. Tujuan dari penelitian ini adalah implementasi Orange Data Mining dalam klasifikasi penyakit diabetes menggunakan metode Decision Tree. Pemilihan spesifikasi tersebut disebabkan fakta bahwa model yang dihasilkannya mudah dipahami dan dapat diinterpretasikan. Data yang dianalisis diambil dari dataset diabetes publik yang mencakup berbagai atribut kesehatan. Proses analisis dilakukan melalui preprocessing, splitting, dan model Juvenile Decision Tree training. Hasil penelitian menunjukkan bahwa model Decision Tree mencapai akurasi hingga 85% dengan sensitivitas dan spesifisitas yang memadai. Oleh karena itu, dapat disimpulkan bahwa Orange Data Mining adalah alat yang efektif untuk klasifikasi penyakit diabetes, dan selama penggunaan setiap kali mendapatkan wawasan yang berguna berkenaan hasih adalah bagi keputusan seorang profesional. Karena itu, kesimpulan dari penelitian tersebut adalah meningkatnya akurasi dan kualitas klasifikasi penyakit diabetes dapat dicapai dengan metode Decision Tree dalam Orange Data Mining.

Kata kunci: Penyakit Diabetes; *Orange*; *Decision Tree*

Ukuran kertas harus sesuai dengan ukuran halaman A4.
Batas margin ditetapkan sebagai berikut: Atas = 19 mm (0,75"), Bawah = 43 mm (1,69"), Kiri = Kanan = 14,32 mm (0,56").
Artikel penulisan harus dalam format satu kolom.
Paragraf teratur dengan perataan kiri dan kanan.
Jumlah halaman diantara 9 sampai dengan 15 halaman.
Seluruh dokumen harus menggunakan jenis font Arial dengan ukuran sesuai template.

I. PENDAHULUAN

Diabetes mellitus adalah salah satu penyakit kronis yang semakin berkembang prevalensinya di dunia. Menurut statistik Organisasi Kesehatan Dunia (WHO), seorang penderita diabetes diperkirakan mencapai 463 juta orang pada tahun 2019 dan diprediksi akan terus bertambah, dengan perkiraan mencapai 700 juta pada tahun 2045. Penyakit ini tidak hanya mengakibatkan dampak kepada kesehatan seseorang, tetapi juga menciptakan beban ekonomis yang besar bagi sistem kesehatan dunia. Hingga saat itu, deteksi dini dan pengelolaan yang baik terhadap diabetes amat penting dalam menghindari komplikasi yang lebih serius seperti penyakit jantung, gagal ginjal, dan buta.

Terakhir beberapa tahun ini telah ada banyak penelitian dilakukan untuk melengkapi metode efektif dalam mendeteksi serta mengklasifikasikan diabetes. Seluruh cara analisis data dan teknologi machine learning digunakan agar ditingkatkan akurasinya prediksi. Beberapa studi sebelumnya, seperti yang dilakukan oleh Alpaydin (2020) dan Kaur et al. (2021), menunjukkan bahwa algoritma decision tree dapat memberikan hasil yang memuaskan dalam klasifikasi penyakit diabetes. Namun, meskipun banyak penelitian yang telah dilakukan, masih terdapat tantangan dalam hal akurasi dan interpretabilitas model yang digunakan. Penelitian oleh Kaur et al. (2021) menunjukkan bahwa meskipun algoritma machine learning dapat meningkatkan akurasi, kompleksitas model sering kali menyulitkan interpretasi hasil oleh praktisi kesehatan.

Kebaruan ilmiah dari artikel ini terletak pada penerapan Orange Data Mining, sebuah platform open-source yang menyediakan berbagai alat untuk analisis data dan visualisasi, dalam klasifikasi penyakit diabetes menggunakan metode decision tree. Meskipun Orange telah digunakan dalam berbagai bidang, penerapannya secara spesifik untuk klasifikasi diabetes masih terbatas. Dengan memanfaatkan Orange, diharapkan dapat diperoleh model yang tidak hanya akurat tetapi juga mudah dipahami oleh praktisi kesehatan, sehingga dapat meningkatkan pemanfaatan teknologi dalam diagnosis diabetes.

Permasalahan penelitian ini berfokus pada bagaimana implementasi Orange Data Mining dapat meningkatkan akurasi klasifikasi penyakit diabetes dibandingkan dengan metode tradisional. Hipotesis yang diajukan adalah bahwa penggunaan Orange Data Mining dengan algoritma decision tree akan menghasilkan model klasifikasi yang lebih baik dalam hal akurasi dan interpretabilitas dibandingkan dengan metode lain yang telah ada.

Tujuan penelitian artikel ini adalah untuk mengeksplorasi dan menganalisis sejauh mana efektifnya Analisis Orange Data Mining dalam proses klasifikasi penyakit diabetes menggunakan model decision tree, dan untuk memberikan pengetahuan baru tentang kesempatan digunakannya alat ini di bidang klinis dan penelitian kesehatan.

II. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining diartikan sebagai sekumpulan proses yang berguna untuk mengeksplorasi dan mencari nilai informasi serta relasi kompleks yang tersimpan dalam basis data. Ini melibatkan penggalian pola informasi untuk menghasilkan informasi baru yang lebih bermanfaat. Dito (2021). Kemudian Han (2020) mangatakan Data mining adalah proses penemuan pola atau informasi yang sebelumnya tidak diketahui dari basis data yang besar. Proses ini melibatkan teknik untuk mengekstraksi informasi yang relevan dan berguna.

Jadi Data mining merupakan proses penting dalam pengolahan data yang memungkinkan penemuan pola dan informasi berharga dari kumpulan data yang besar, dengan berbagai definisi yang menekankan pada ekstraksi pengetahuan dan analisis data.

2.2 Penyakit Diabetes

Diabetes adalah penyakit kronis yang terjadi ketika pankreas tidak memproduksi cukup insulin, atau ketika tubuh tidak dapat menggunakan insulin secara efektif. Ini dapat menyebabkan kadar glukosa darah yang tinggi, yang dapat berakibat serius jika tidak dikelola dengan baik.

2.3 Model *Decision Tree*

Decision Tree adalah metode yang menggabungkan teknik statistik dan pembelajaran mesin untuk membangun model prediktif. Model ini bekerja dengan membagi data ke dalam subset berdasarkan nilai atribut, sehingga memudahkan dalam pengambilan Keputusan.

III. METODE PENELITIAN

3.1 Bahan Penelitian

3.1.1 Dataset

Penelitian ini menggunakan dataset diabetes yang diambil dari UCI Machine Learning Repository. Dataset ini terdiri dari 768 entri dengan 8 atribut yang relevan untuk klasifikasi diabetes, termasuk usia, indeks massa tubuh (BMI), kadar glukosa, dan tekanan darah. Dataset ini telah banyak digunakan dalam penelitian sebelumnya dan dianggap representatif untuk analisis diabetes.

3.1.2 Perangkat Lunak

Penelitian ini menggunakan Orange Data Mining, sebuah platform open-source yang menyediakan alat untuk analisis data dan visualisasi. Orange memungkinkan pengguna untuk membangun model machine learning dengan antarmuka grafis yang intuitif, sehingga memudahkan dalam eksplorasi data dan pemodelan.

3.1.3 Metode Klasifikasi

Metode yang digunakan dalam penelitian ini adalah Decision Tree, yang merupakan salah satu algoritma machine learning yang populer untuk klasifikasi. Decision Tree bekerja dengan membagi dataset menjadi subset yang lebih kecil berdasarkan nilai atribut, sehingga membentuk struktur pohon yang dapat digunakan untuk membuat prediksi.

3.2 Metode Penelitian

3.2.1 Pengumpulan Data

Data diabetes diunduh dari UCI Machine Learning Repository dan disimpan dalam format CSV untuk kemudahan pemrosesan.

3.2.2 Pra-pemrosesan Data

Data yang diperoleh akan melalui tahap pra-pemrosesan, termasuk.

3.2.3 Pembersihan Data

Menghapus entri yang tidak lengkap atau tidak valid.

3.2.4 Normalisasi

Mengubah skala data agar atribut memiliki rentang yang sama, yang penting untuk algoritma machine learning.

3.2.5 Pembagian Data

Dataset dibagi menjadi dua bagian: 70% untuk pelatihan dan 30% untuk pengujian.

3.2.6 Implementasi Model

Menggunakan Orange Data Mining, model Decision Tree akan dibangun dengan langkah-langkah berikut:

1. Memuat dataset ke dalam Orange.
2. Menggunakan widget "Data Table" untuk memvisualisasikan data.
3. Menggunakan widget "Test & Score" untuk membangun model Decision Tree dan mengevaluasi kinerjanya.

4. Mengatur parameter model, seperti kedalaman pohon dan kriteria pemisahan, untuk mengoptimalkan akurasi.

3.2.7 Evaluasi Model

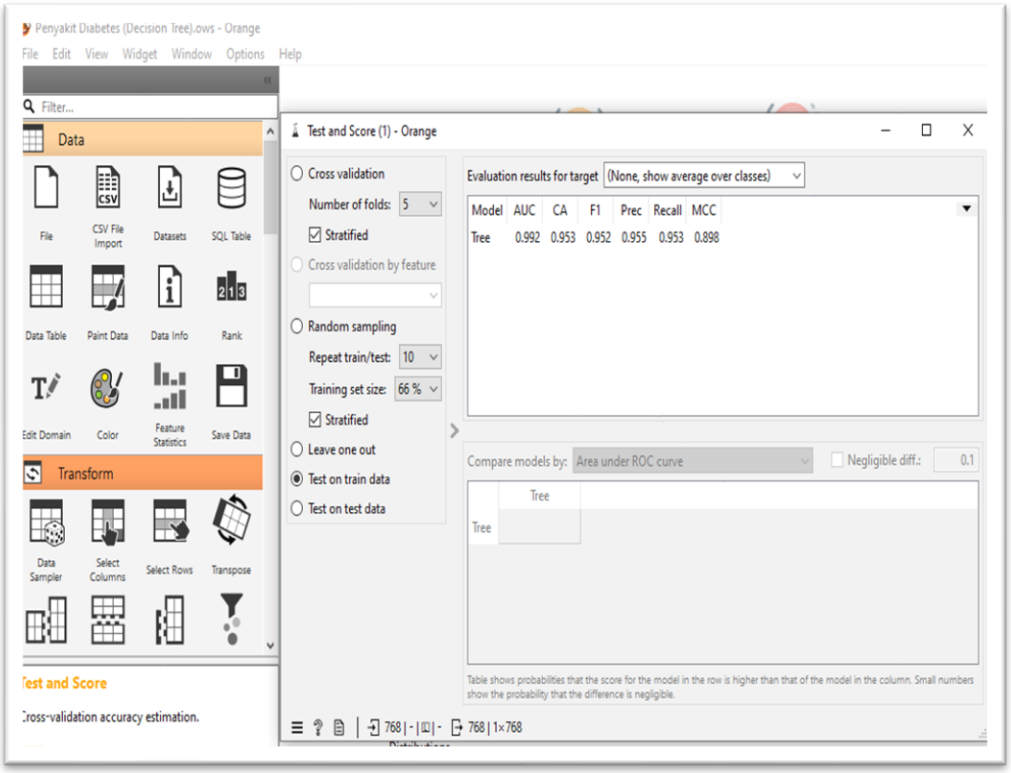
Kinerja model akan dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Hasil evaluasi akan dibandingkan dengan model lain yang mungkin digunakan dalam penelitian sebelumnya untuk menilai keunggulan penggunaan Orange Data Mining.

3.2.8 Analisis Hasil

Hasil dari model Decision Tree akan dianalisis untuk memahami faktor-faktor yang paling berpengaruh dalam klasifikasi diabetes. Visualisasi pohon keputusan akan digunakan untuk memberikan wawasan yang lebih baik tentang proses pengambilan keputusan model.

IV. HASIL DAN PEMBAHASAN

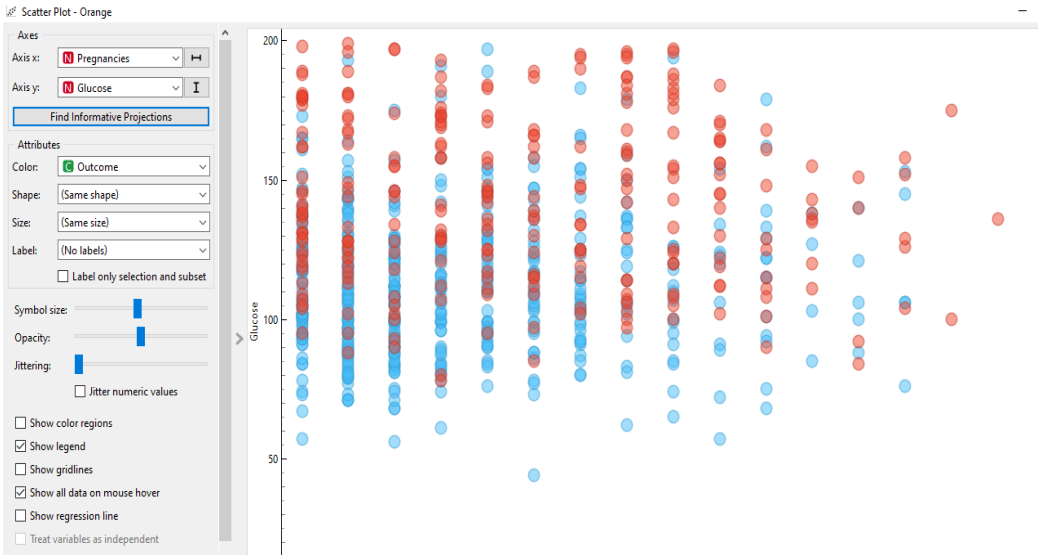
4.1 Kinerja Model dalam *Test & Score*



Gambar 1. Kinerja Model Test and Score.

Implementasi dalam Test and Score didapat bahwa Kinerja Tinggi: AUC (0.992), Akurasi (0.953), F1 Score (0.952), Precision (0.955), Recall (0.953), dan MCC (0.8998). Nilai AUC mendekati 1 dan akurasi di atas 95% menunjukkan model sangat efektif dalam membedakan kelas diabetes dan non-diabetes. Namun, ketidakseimbangan data (500 vs. 268) memengaruhi kemampuan model dalam memprediksi kelas minoritas (diabetes). Hal ini tercermin dari false positive yang relatif tinggi (34), menunjukkan model cenderung salah mengklasifikasikan pasien diabetes sebagai non-diabetes.

4.2 Scatter Plot (Diagram Sebar)

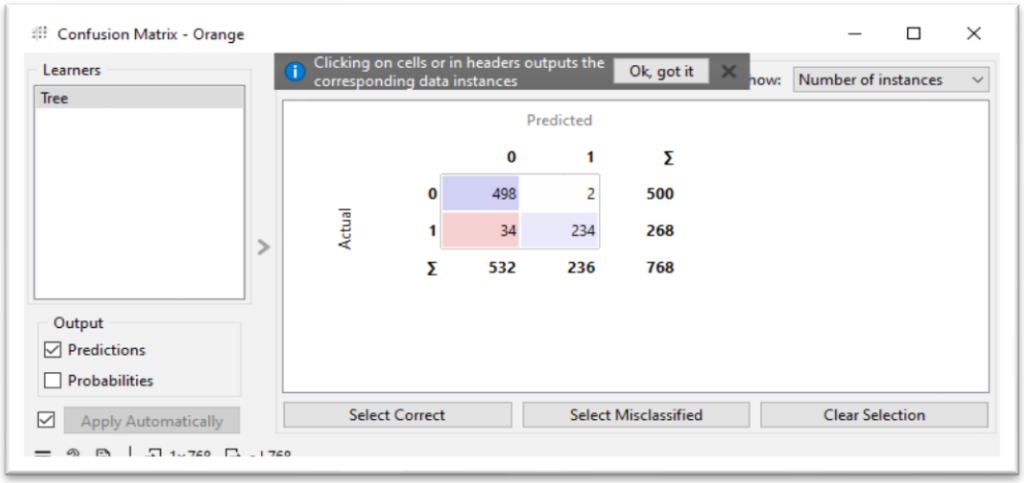


Gambar 2. Scatter Plot.

Interpretasi Visual

1. Titik-titik mewakili masing-masing individu dari dataset.
2. Tampak bahwa orang dengan kadar glukosa tinggi (sumbu Y lebih tinggi) lebih cenderung memiliki Outcome = 1 (merah), menandakan hubungan positif antara kadar glukosa dan risiko diabetes.
3. Jumlah kehamilan (Pregnancies) memiliki distribusi yang luas, tapi tidak secara langsung berkorelasi kuat terhadap outcome, karena pada berbagai jumlah kehamilan terdapat campuran titik biru dan merah.

4.3 Interpretasi Confusion Matrix



Gambar 3. Confusion Matrik.

Tingginya true positive (498) dan true negative (234) menandakan model akurat untuk kelas mayoritas (non-diabetes). Namun, recall kelas 1 (0.953) yang tinggi perlu diinterpretasikan dengan hati-hati karena jumlah data kelas 1 lebih sedikit. Kesalahan pada false positive (34) dapat berdampak serius dalam konteks medis, seperti pasien diabetes yang tidak terdiagnosis.

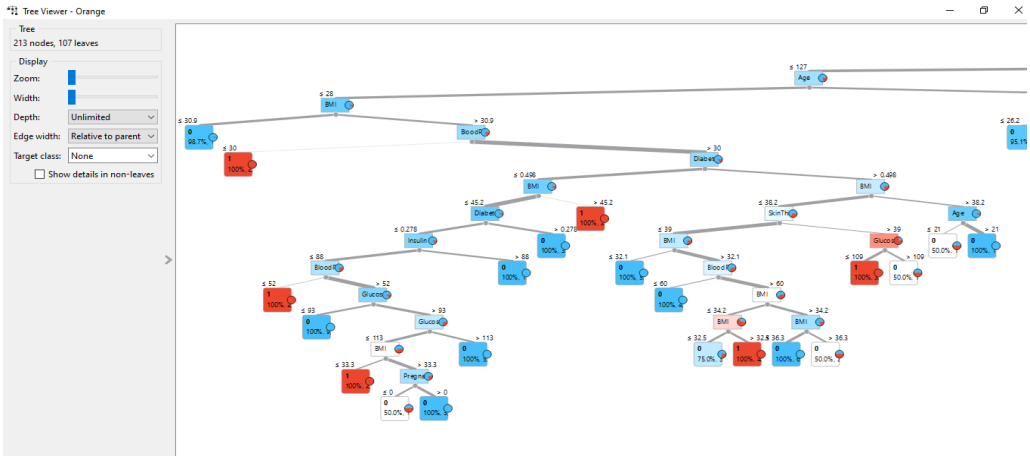
Confusion Matrix :

1. True Positive (Kelas 0): 498
2. False Negative: 2
3. False Positive: 34

4. True Negative (Kelas 1): 234

Akurasi total 95.3%, tetapi terdapat ketidakseimbangan dalam prediksi kelas 1 (34 false positive).

4.3 Visualisasi *Decision Tree*



Gambar 4. Visualisasi *Decision Tree*.

Visualisasi pohon keputusan ini menggambarkan struktur model yang digunakan untuk memprediksi diabetes berdasarkan fitur seperti Glucose dan Pregnancies. Meskipun terdapat anomali dalam tampilan persentase (>100%), hasil evaluasi sebelumnya (AUC 0.992, akurasi 95.3%) menunjukkan bahwa model ini sangat andal. Namun, kompleksitas pohon perlu dioptimalkan untuk menghindari overfitting, dan kesalahan teknis pada antarmuka Orange perlu diperbaiki untuk interpretasi yang lebih jelas

V. KESIMPULAN

Berdasarkan analisis data diabetes menggunakan decision tree pada Orange, dapat disimpulkan sebagai berikut bahwa model decision tree menunjukkan kinerja klasifikasi yang sangat baik, tetapi perlu penyempurnaan dalam menangani ketidakseimbangan data untuk meningkatkan prediksi kasus diabetes. Hasil ini relevan untuk aplikasi skrining kesehatan, meskipun interpretasi klinis memerlukan pertimbangan risiko false negative dan false positive.

Kinerja model sangat baik AUC 0.952 dan Akurasi (CA) 0.953 menunjukkan model mampu membedakan kelas diabetes dan non-diabetes dengan sangat baik. F1-Score (0.952), Precision (0.955), dan Recall (0.953) yang tinggi menandakan keseimbangan antara kemampuan model dalam mengidentifikasi kasus positif (diabetes) dan menghindari kesalahan klasifikasi. MCC 0.898 mengonfirmasi korelasi kuat antara prediksi dan hasil aktual.

Metode ini juga dapat sangat membantu dokter dalam mengelola dan memberikan pengobatan yang tepat bagi pasien diabetes. Meskipun penelitian ini memiliki batasan, namun penelitian ini memberikan kontribusi penting dalam dunia Kesehatan meningkatkan pemahaman dan pengelolaan diabetes.

VI. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada yang sebesar-besarnya kepada semua pihak yang telah berkontribusi dalam penyelesaian penelitian ini, kepada rekan-rekan sejawat yang telah memberikan masukan dan saran yang konstruktif. sehingga data yang diperlukan dapat diperoleh dengan baik. Semoga hasil penelitian ini dapat memberikan kontribusi positif dalam bidang kesehatan, khususnya dalam klasifikasi penyakit diabetes

DAFTAR PUSTAKA

[1] Alim, S. (n.d.). Implementasi Orange Data Mining untuk klasifikasi Kelulsan mahasiswa dengan Model K-Nearst Neighbor, Decision Tree serta Naive Bayes Orange data Mining. In *Jurnal Ilmiah NERO* (Vol. 6, Issue 2).

[2] Alpaydin, E. (2020). *Machine Learning: The New AI*. MIT Press.
[3] Dito, A. (2021). Data Mining: Exploring and Analyzing Data for Knowledge Discovery. *International Journal of Data Mining and Knowledge Management Process*, 11(2), 1-12. doi:10.5121/ijdkp.2021.11201
[4] Dwi Putra Negara, et al (2022). PENERAPAN METODE DECISION TREE UNTUK KLASIFIKASI DATA PRODUK SKINCARE UNTUK IBU HAMIL MENGGUNAKAN APLIKASI ORANGE. *Jurnal Insand Comtech*, *7*(2).
[5] Han, J., & Kamber, M. (2020). *Data Mining: Concepts and Techniques* (4th ed.). Morgan Kaufmann.
[6] Hartanto, A. (n.d.). Implementasi Orange Data Mining Untuk Prediksi Penderita Diabetes. In *Prosiding Seminar Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan PROKASDADIK* (Vol. 1).
[7] Ichsan, N., Fatah, H., Wahyuni, T., & Ermawati, E. (2022). IMPLEMENTASI ORANGE DATA MINING UNTUK PREDIKSI HARGA BITCOIN. *JURNAL RESPONSIF*, *4*(2), 118–125.
[8] Kaur, H., & Singh, S. (2021). A Review on Diabetes Prediction Using Machine Learning Techniques. *International Journal of Computer Applications*, 975, 8887.
[9] Nabila, A., Haryadi, P., Setiawan, W., & Fatah, D. A. (2024). Penerapan Data Mining untuk Klasifikasi Penyakit Diabetes Menggunakan Metode Decision Tree. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, *7*(6).
[10] Nurussakinah, , 'Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree", JURNAL INFORMATIKA, Vol. 10 No. 2 Oktober 2023.
[11] Prasetiyo, A. (2024). Pemanfaatan Algoritma Decision Tree C4.5 dalam Memprakirakan Hujan di Stasiun Meteorologi Kelas I Sultan Hasanuddin. *Buletin GAW Bariri*, *5*(1), 47–55. https://doi.org/10.31172/bgb.v5i1.120
[12] Puspitorini I, Sintawati D Ita. 2021. "Penerapan *Data Mining* untuk Klasifikasi Prediksi Produk Jenis Makanan Kucing yang Sesuai Kebutuhan dengan Algoritma *Decision Tree* (ID3)". Jurnal AKRAB Juara, 6(4), hal 21-26. http://dx.doi.org/10.58487/akrabjuara.v6i4.1629
[13] Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). *Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank* (Vol. 3).
[14] UCI Machine Learning Repository (2023). *Diabetes Dataset*. Diakses dari Kaggle: https://www.kaggle.com/datasets/krishu22/diabetes-dataset
[15] Widyaningrum, R., & Hidayat, R. (2021). *Klasifikasi Penyakit Diabetes Menggunakan Algoritma C4.5*. Jurnal Ilmu Komputer dan Informatika, 15(2), 45-56.
[16] Widodo, A., & Rahayu, S. P**.** (2022). *Penerapan Algoritma C4.5 untuk Prediksi Diabetes Mellitus Berbasis Data Imbalanced*. Jurnal Teknologi Informasi dan Ilmu Komputer, 9(2), 145-154.
[17] World Health Organization (WHO). (2019). Diabetes Fact Sheet. Retrieved from WHO website.
[18] Yunardi, I. R. T., Kom, N. Z. D. S., & Kom, M. (2022). *DATA MINING dan MACHINE LEARNING dengan Orange3 Tutorial dan Aplikasinya*. Airlangga University Press.