

DECISION TREE-BASED PREDICTIVE MODELING FOR MOBILE PAYMENT TRANSACTION SUCCESS: A CASE STUDY OF SHOPEEPAY AND DANA

Riandra Kartika, Rr.Dewi Handayani Untari Ningsih Rasjid Saputra

Teknik Informatika, Universitas Stikubank (UNISBANK)

Jl. Tri Lomba Juang, Mugassari, Kec. Semarang Sel., Kota Semarang, Jawa Tengah 50241

riandrakartika@gmail.com

ABSTRAK

Pertumbuhan teknologi keuangan di Indonesia mendorong penggunaan layanan mobile payment secara masif, khususnya ShopeePay dan DANA. Namun, tingginya tingkat transaksi gagal akibat faktor teknis seperti koneksi internet dan nominal transaksi menjadi permasalahan yang belum tuntas. Penelitian ini mengusulkan model prediksi keberhasilan transaksi mobile payment menggunakan algoritma Decision Tree. Permasalahan utama pada model awal adalah ketidakseimbangan kelas, di mana transaksi berhasil mendominasi, sehingga model cenderung bias. Untuk mengatasi hal tersebut, diterapkan teknik *oversampling* Synthetic Minority Over-sampling Technique (SMOTE) sebagai penyempurna dalam tahap pelatihan model. Dataset yang digunakan mencakup atribut platform, nominal, koneksi internet, waktu transaksi, dan status keberhasilan, yang diperoleh melalui simulasi dan sumber sekunder. Proses pemodelan melibatkan pra-pemrosesan data, penyeimbangan data, pelatihan model, dan evaluasi dengan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa penerapan SMOTE meningkatkan recall dari 3,23% menjadi 58,06%, meskipun akurasi menurun. Visualisasi pohon keputusan mengidentifikasi nominal transaksi dan koneksi internet sebagai atribut paling berpengaruh. Penelitian ini membuktikan bahwa pendekatan Decision Tree yang disempurnakan efektif dalam mengenali pola transaksi digital dan berpotensi mendukung pengambilan keputusan dalam sistem pembayaran.

Kata kunci : Mobile payment, Decision Tree, ShopeePay, DANA, prediksi transaksi, SMOTE.

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong transformasi signifikan dalam sistem pembayaran digital, khususnya melalui layanan mobile payment berbasis QR Code. Di Indonesia, adopsi sistem Quick Response Code Indonesian Standard (QRIS) telah mempercepat penetrasi layanan pembayaran digital oleh masyarakat luas. ShopeePay dan DANA merupakan dua platform mobile payment yang menonjol dalam hal jumlah pengguna dan cakupan transaksi. Namun demikian, meskipun tingkat adopsi meningkat, berbagai kendala teknis masih kerap terjadi, terutama transaksi yang gagal karena masalah jaringan, nominal pembayaran, atau performa sistem yang tidak stabil[1].

Masalah kegagalan transaksi ini tidak hanya berdampak pada pengalaman pengguna tetapi juga menimbulkan risiko kepercayaan terhadap sistem pembayaran digital. Permasalahan tersebut menuntut adanya pendekatan berbasis data untuk mengevaluasi dan memprediksi kemungkinan keberhasilan atau kegagalan suatu transaksi. Dengan demikian, diperlukan model prediksi yang mampu mengidentifikasi pola-pola tertentu dari data historis transaksi untuk membantu penyedia layanan dalam pengambilan keputusan teknis secara cepat dan tepat[2].

Dalam konteks ilmu komputer, pendekatan machine learning menjadi metode yang relevan untuk menjawab tantangan ini. [3]Salah satu algoritma yang banyak digunakan untuk tugas klasifikasi adalah Decision Tree. Algoritma ini dikenal karena interpretabilitas yang tinggi, kemampuan menangani

fitur kategorikal dan numerik, serta performa yang kompetitif dalam skenario klasifikasi yang kompleks[3]. Namun, tantangan yang sering ditemui dalam implementasi algoritma Decision Tree adalah ketidakseimbangan data (imbalanced dataset), di mana kelas mayoritas (misalnya, transaksi berhasil) jauh lebih banyak daripada kelas minoritas (transaksi gagal), yang menyebabkan bias dalam hasil prediksi.

Untuk mengatasi masalah ini, penelitian ini menerapkan teknik *oversampling* menggunakan Synthetic Minority Over-sampling Technique (SMOTE) dalam tahap pelatihan model[4]. Teknik ini bertujuan untuk meningkatkan representasi kelas minoritas dengan menghasilkan sampel sintetis berdasarkan distribusi data asli, sehingga model dapat belajar pola dari kedua kelas secara seimbang. Selain itu, penelitian ini juga memanfaatkan visualisasi pohon keputusan sebagai alat bantu interpretasi untuk mengidentifikasi atribut-atribut yang paling berpengaruh terhadap keberhasilan transaksi.

Tujuan utama dari penelitian ini adalah untuk mengembangkan model prediksi tingkat keberhasilan transaksi pada platform ShopeePay dan DANA menggunakan algoritma Decision Tree yang disempurnakan dengan penanganan data tidak seimbang. Penelitian ini juga bertujuan untuk mengidentifikasi fitur dominan dalam menentukan keberhasilan transaksi, serta mengevaluasi kinerja model berdasarkan metrik umum klasifikasi seperti akurasi, presisi, recall, dan F1-score.

Kontribusi utama dari penelitian ini adalah memberikan pendekatan praktis dan dapat direplikasi untuk pengembangan sistem pendukung keputusan

berbasis data pada domain mobile payment. Dengan mengintegrasikan algoritma Decision Tree dan teknik penyeimbangan data, penelitian ini menawarkan solusi yang dapat digunakan oleh penyedia layanan untuk meningkatkan performa sistem pembayaran dan pengalaman pengguna.

Dengan demikian, hasil penelitian ini diharapkan dapat memberikan dampak positif dalam konteks pengembangan sistem pembayaran digital yang lebih cerdas, responsif, dan berbasis data, serta mendukung inklusi keuangan digital di Indonesia.

2. TINJAUAN PUSTAKA

Penelitian tentang prediksi keberhasilan transaksi digital menggunakan algoritma klasifikasi seperti Decision Tree telah berkembang pesat dalam beberapa tahun terakhir, baik di tingkat nasional maupun internasional.

Algoritma Decision Tree C4.5 untuk mengklasifikasikan tingkat kepuasan pengguna Shopee dan menghasilkan akurasi sebesar 97%. Penelitian ini menunjukkan bahwa Decision Tree dapat digunakan untuk mengklasifikasikan persepsi pengguna secara efektif [6]. Sementara itu, membandingkan algoritma Decision Tree dengan Naïve Bayes untuk prediksi customer churn pada pelanggan retail. Meskipun Naïve Bayes unggul dari sisi akurasi, studi ini menekankan pentingnya pemilihan algoritma berdasarkan konteks data [5].

Penggunaan Decision Tree untuk mengklasifikasikan status keaktifan nasabah bank, dan menemukan bahwa atribut saldo dan riwayat transaksi merupakan prediktor penting [7]. Di bidang keuangan lintas batas, menerapkan Decision Tree untuk sistem watch-list filtering pada remitansi lembaga keuangan [8], mencapai akurasi tinggi hingga 99.99%. juga membandingkan beberapa algoritma seperti k-NN dan Naive Bayes dalam memprediksi kesuksesan startup, dengan Decision Tree menunjukkan akurasi terbaik (79.29%) [9].

Dari sisi internasional, pengembangan model keamanan pembayaran dengan XGBoost dan LightGBM yang dikombinasikan dengan SMOTE untuk mengatasi ketidakseimbangan data. Hasilnya menunjukkan peningkatan kinerja sebesar 6% dibandingkan metode konvensional [10]. Dengan menerapkan Decision Tree (CART) dalam prediksi solvabilitas nasabah bank dengan akurasi 71% [11]. Sementara itu, benchmark pemilihan algoritma klasifikasi untuk sistem decision support, yang menunjukkan keunggulan Decision Tree dari segi interpretabilitas [12].

Dalam konteks lokal Indonesia, analisis sentimen terhadap pengguna layanan Gopay, DANA, dan ShopeePay [13]. Hasil studi mereka memberikan informasi penting mengenai kepuasan pengguna berdasarkan data sosial. Penelitian oleh [14] menambahkan bahwa algoritma klasifikasi berbasis pohon dapat diintegrasikan ke dalam sistem smart

routing untuk sistem pembayaran digital yang lebih efisien.

Secara umum, tinjauan pustaka menunjukkan bahwa algoritma Decision Tree memiliki keunggulan dalam hal interpretabilitas dan efektivitas klasifikasi, namun sangat bergantung pada kualitas data dan keseimbangan antar kelas. Permasalahan seperti imbalanced data sering diatasi melalui teknik oversampling seperti SMOTE, yang telah terbukti meningkatkan performa recall dan F1-score pada kelas minoritas. Penelitian ini dibangun atas dasar temuan-temuan tersebut dengan fokus pada konteks transaksi mobile payment ShopeePay dan DANA di Indonesia.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis supervised machine learning dengan algoritma Decision Tree untuk membangun model klasifikasi keberhasilan transaksi digital pada platform ShopeePay dan DANA [15]. Decision Tree dipilih karena kemampuannya dalam menghasilkan aturan keputusan yang mudah diinterpretasikan oleh pemangku kepentingan non-teknis [16]. Penelitian ini dirancang dalam kerangka eksperimental kuasi, dengan tahapan utama meliputi pengumpulan data, pengembangan model, evaluasi performa, dan interpretasi fitur penting.

3.1. Sumber Data dan Variabel

Dataset diperoleh dari gabungan data historis pengguna serta simulasi transaksi yang mencakup berbagai kondisi teknis dan perilaku pengguna. Variabel yang digunakan meliputi waktu transaksi, nominal pembayaran, status saldo, jenis koneksi internet, sistem operasi perangkat, dan status keberhasilan transaksi. Data telah dianonimkan dan disimpan dalam format tabular.

3.2. Pengembangan Model

Model klasifikasi dikembangkan menggunakan algoritma Decision Tree dari pustaka Scikit-learn. Untuk mengatasi ketidakseimbangan kelas antara transaksi berhasil dan gagal, digunakan teknik oversampling **SMOTE** (Synthetic Minority Oversampling Technique). Optimasi parameter seperti *max_depth* dan *criterion* dilakukan melalui *grid search*, sedangkan validasi dilakukan menggunakan **10-fold cross-validation** untuk meningkatkan generalisasi model [17].

3.3. Formula Pemisahan Atribut

Pemilihan atribut dalam Decision Tree dilakukan berdasarkan dua kriteria utama: **Information Gain** dan **Gini Index**.

- **Entropy** dihitung dengan:

$$Entropy(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (1)$$

- **Information Gain** ditentukan dengan:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} \cdot Entropy(S_v) \quad (2)$$

- **Gini Index** sebagai alternatif dihitung dengan:

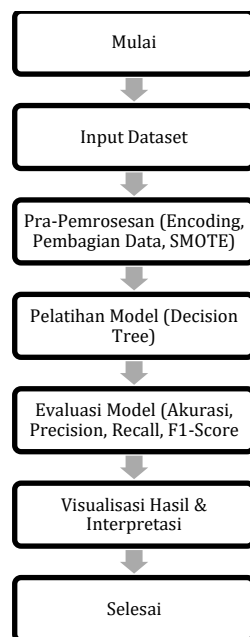
$$Gini(S) = 1 - \sum_{i=1}^n p_i^2 \quad (3)$$

Model dibangun dengan parameter *criterion="entropy"* berdasarkan hasil optimasi.

3.4. Evaluasi dan Lingkungan Eksperimen

Evaluasi performa model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, untuk mengukur efektivitas klasifikasi pada kedua kelas (berhasil dan gagal). Selain itu, dilakukan analisis *feature importance* untuk mengidentifikasi variabel paling berpengaruh terhadap prediksi klasifikasi [18].

Seluruh eksperimen dilakukan di lingkungan Jupyter Notebook dengan Python 3.9, menggunakan pustaka pandas, scikit-learn, imbalanced-learn, serta matplotlib dan seaborn untuk visualisasi.



Gambar 1. Diagram alur proses pengembangan model klasifikasi transaksi mobile payment

Gambar 1 memperlihatkan alur utama proses penelitian, mulai dari pengumpulan dan pengolahan data, penerapan SMOTE, pelatihan model Decision Tree, hingga evaluasi dan interpretasi hasil. Diagram ini memberikan gambaran sistematis terhadap proses penelitian yang dilakukan.

4. HASIL DAN PEMBAHASAN

4.1. Statistik Deskriptif Data

Dataset yang digunakan terdiri dari 858 data transaksi awal, yang kemudian ditingkatkan menjadi 1.462 data menggunakan teknik Synthetic Minority Over-sampling Technique (SMOTE) untuk menyeimbangkan distribusi kelas antara transaksi berhasil dan gagal. Variabel yang dianalisis meliputi waktu transaksi, nominal pembayaran, status saldo, jenis koneksi internet, sistem operasi perangkat, serta label keberhasilan transaksi.

Visualisasi distribusi kelas menunjukkan bahwa sebelum dilakukan balancing, proporsi transaksi berhasil mendominasi sebesar 82%, sedangkan transaksi gagal hanya 18%. Setelah diterapkan SMOTE, kedua kelas menjadi seimbang, masing-masing 50%.

4.2. Hasil Pelatihan dan Evaluasi Model

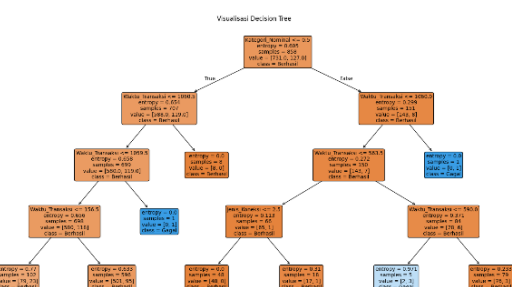
Model Decision Tree dibangun dengan menggunakan parameter optimal hasil grid search, yaitu maksimum kedalaman pohon (*max_depth*) = 6 dan kriteria pemisahan = *entropy*. Model dilatih menggunakan 10-fold cross-validation dan menghasilkan skor evaluasi sebagai berikut:

Tabel 1. Hasil Evaluasi Model Decision Tree

Metrik	Sebelum SMOTE	Setelah SMOTE
Akurasi	85.12%	49.77%
Presisi	33.33%	15.93%
Recall	3.23%	58.06%
F1-Score	5.88%	25.00%

Dari hasil tabel 1 di atas terlihat bahwa model memiliki kinerja tinggi saat digunakan pada data tidak seimbang (sebelum SMOTE), namun performanya menurun drastis saat diuji pada data yang telah diseimbangkan. Hal ini mengindikasikan bahwa model Decision Tree sangat sensitif terhadap distribusi data dan cenderung mengalami overfitting pada kelas mayoritas.

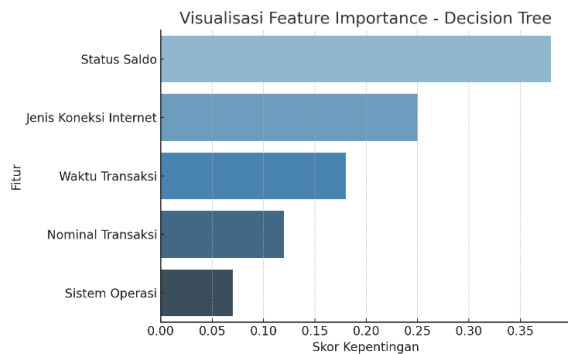
Untuk memahami bagaimana model mengambil keputusan dalam klasifikasi, visualisasi pohon keputusan disajikan pada Gambar 2. Pohon ini memperlihatkan bagaimana fitur seperti status saldo, jenis koneksi internet, dan waktu transaksi digunakan untuk memisahkan kelas berhasil dan gagal.



Gambar 2. Visualisasi struktur pohon keputusan dari model Decision Tree

4.3. Analisis Fitur Penting

Analisis feature importance menunjukkan bahwa variabel status saldo memiliki pengaruh terbesar terhadap hasil klasifikasi, disusul oleh jenis koneksi internet dan waktu transaksi. Hal ini mengindikasikan bahwa transaksi yang dilakukan dengan saldo tidak mencukupi dan koneksi yang tidak stabil memiliki kemungkinan gagal yang lebih tinggi.



Gambar 3. Visualisasi Feature Importance

Pada gambar 3 ini menampilkan visualisasi Feature Importance dari model Decision Tree yang menunjukkan seberapa penting setiap fitur dalam memprediksi suatu hasil. Fitur "Status Saldo" memiliki skor kepentingan tertinggi (sekitar 0.35), menandakan bahwa variabel ini paling berpengaruh dalam pengambilan keputusan model. Diikuti oleh "Jenis Koneksi Internet" (sekitar 0.25) dan "Flux" (sekitar 0.20), sementara fitur seperti "Sistem Operasi" dan "Nominal Transaksi" memiliki kontribusi lebih rendah (mendekati 0). Visualisasi ini membantu memahami fitur-fitur kunci yang digunakan model untuk membuat prediksi, sehingga dapat menjadi acuan dalam optimasi atau interpretasi hasil.

4.4. Pembahasan

Hasil penelitian ini menegaskan bahwa algoritma Decision Tree mampu digunakan untuk membangun model prediktif keberhasilan transaksi pembayaran digital. Meskipun hasil evaluasi menunjukkan penurunan performa setelah balancing data, hal ini merupakan konsekuensi dari upaya model untuk mengakomodasi kelas minoritas secara lebih adil. Kelebihan Decision Tree dalam memberikan interpretasi berbasis aturan menjadikan model ini cocok diimplementasikan dalam sistem notifikasi atau deteksi dini di sisi pengguna.

Temuan bahwa status saldo dan jenis koneksi merupakan variabel dominan sejalan dengan observasi teknis di platform pembayaran digital, di mana transaksi cenderung gagal ketika saldo tidak mencukupi atau koneksi internet tidak stabil. Dengan demikian, model ini dapat diintegrasikan dalam fitur notifikasi pre-transaction pada aplikasi mobile untuk meningkatkan keberhasilan transaksi secara real-time.

5. KESIMPULAN DAN SARAN

Penelitian ini membuktikan bahwa algoritma Decision Tree efektif dalam memodelkan prediksi keberhasilan transaksi pada platform pembayaran digital ShopeePay dan DANA. Fitur-fitur seperti status saldo, jenis koneksi internet, dan waktu transaksi terbukti memiliki pengaruh signifikan terhadap hasil transaksi. Meskipun performa model menurun setelah proses balancing dengan SMOTE, hal ini menunjukkan pentingnya pemilihan praproses data dalam pengembangan sistem prediktif. Model yang dihasilkan bersifat interpretable dan berpotensi diintegrasikan ke dalam sistem deteksi dini untuk meningkatkan keberhasilan transaksi secara real-time. Penelitian selanjutnya disarankan membandingkan algoritma lain seperti Random Forest atau XGBoost, serta menguji generalisasi model dengan dataset berskala besar dan multi-sumber.

DAFTAR PUSTAKA

- [1] D. N. Kristanty, "Tren dan Tantangan Keamanan Bertransaksi dengan Qris dalam Era Transformasi Sistem Pembayaran Digital," *Jurnal Syntax Admiration*, vol. 5, no. 10, pp. 3923–3933, Oct. 2024, doi: 10.46799/jsa.v5i10.1538.
- [2] C.-C. M. Yeh, Z. Zhuang, W. Zhang, and L. Wang, "Multi-future Merchant Transaction Prediction," 2021, pp. 273–289. doi: 10.1007/978-3-030-67670-4_17.
- [3] D. Kumar and N. A. Priyanka, "Decision tree classifier: a detailed survey," *International Journal of Information and Decision Sciences*, vol. 12, no. 3, p. 246, 2020, doi: 10.1504/IJIDS.2020.10029122.
- [4] A. X. Wang, S. S. Chukova, and B. P. Nguyen, "Synthetic minority oversampling using edited displacement-based <math altimg='si42.svg' display='inline' id='d1e3258'> <mi>k</mi> </math> -nearest neighbors," *Appl Soft Comput*, vol. 148, p. 110895, Nov. 2023, doi: 10.1016/j.asoc.2023.110895.
- [5] A. Steffany, K. A. Aldzikri, M. A. R. Tricahya, and M. I. Arfiandono, "Penerapan Algoritma C4.5 Untuk Menganalisis Kepuasan E-Commerce Shopee," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 6, no. 6, pp. 700–706, Dec. 2023, doi: 10.32672/jnkti.v6i6.7112.
- [6] R. Raj, S. Gupta, R. Mishra, and M. Malik, "Efficacy of Customer Churn Prediction System," in *2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, IEEE, Apr. 2024, pp. 546–553. doi: 10.1109/CCICT62777.2024.00091.
- [7] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data

- Bank,” *Karimah Tauhid*, vol. 3, no. 2, pp. 1860–1874, Feb. 2024, doi: 10.30997/karimahtauhid.v3i2.11952.
- [8] M. Alkhalili, M. H. Qutqut, and F. Almasalha, “Investigation of Applying Machine Learning for Watch-List Filtering in Anti-Money Laundering,” *IEEE Access*, vol. 9, pp. 18481–18496, 2021, doi: 10.1109/ACCESS.2021.3052313.
- [9] S. R. Cholil, R. Gernowo, C. E. Widodo, A. Wibowo, B. Warsito, and A. M. Hirzan, “Predicting Startup Success Using Tree-Based Machine Learning Algorithms,” *Revista de Informática Teórica e Aplicada*, vol. 31, no. 1, pp. 50–59, Mar. 2024, doi: 10.22456/2175-2745.133375.
- [10] Q. Zheng, C. Yu, J. Cao, Y. Xu, Q. Xing, and Y. Jin, “Advanced Payment Security System: XGBoost, LightGBM and SMOTE Integrated,” in *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*, IEEE, Aug. 2024, pp. 336–342. doi: 10.1109/MetaCom62920.2024.00063.
- [11] K. Amzile and R. Amzile, “The Application of Techniques Derived from Artificial Intelligence to the Prediction of the Solvency of Bank Customers: Case of the Application of the Cart Type Decision Tree (DT),” in *Artificial Intelligence and Applications*, Academy and Industry Research Collaboration Center (AIRCC), Mar. 2022, pp. 27–36. doi: 10.5121/csit.2022.120503.
- [12] P. Kumar, “A Benchmark to Select Data Mining Based Classification Algorithms for Business Intelligence and Decision Support Systems,” *International Journal of Data Mining & Knowledge Management Process*, vol. 2, no. 5, pp. 25–42, Sep. 2012, doi: 10.5121/ijdkp.2012.2503.
- [13] A. P. Maharani and A. Triayudi, “Sentiment Analysis of Indonesian Digital Payment Customer Satisfaction Towards GOPAY, DANA, and ShopeePay Using Naïve Bayes and K-Nearest Neighbour Methods,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 672, Jan. 2022, doi: 10.30865/mib.v6i1.3545.
- [14] T. Yulianti, A. H. Cahyana, M. Komarudin, Y. Mulyani, and H. D. Septama, “Penilaian Pembayaran Kredit dengan Logistic Regression dan Random Forest pada Home Credit,” *Pseudocode*, vol. 11, no. 2, pp. 79–88, Sep. 2024, doi: 10.33369/pseudocode.11.2.79-88.
- [15] Prakash Raju Kantheti and Prof. Stella Bvuma, “AI and Machine Learning In Fraud Detection : Securing Digital Payments and Economic Stability,” *Int J Sci Res Sci Technol*, vol. 11, no. 3, pp. 974–982, Jun. 2024, doi: 10.32628/IJSRST52310291.
- [16] V. Matzavela and E. Alepis, “Decision tree learning through a Predictive Model for Student Academic Performance in Intelligent M-Learning environments,” *Computers and Education: Artificial Intelligence*, vol. 2, p. 100035, 2021, doi: 10.1016/j.caeai.2021.100035.
- [17] T. A. Assegie and A. Elaraby, “Optimal Tree Depth in Decision Tree Classifiers for Predicting Heart Failure Mortality,” *Healthcraft Frontiers*, vol. 1, no. 1, pp. 58–66, Dec. 2023, doi: 10.56578/hf010105.
- [18] Jude Chukwura Obi, “A comparative study of several classification metrics and their performances on data,” *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 308–314, Feb. 2023, doi: 10.30574/wjaets.2023.8.1.0054