

Literature Review: Pemanfaatan *Decision Tree* Pada Berbagai Bidang

Chintya Mei^{1*}, Hasria², Irwan Hardi Kusuma³, Lulu Kaoniah⁴, Novariman⁵,
Zhaki Apriyan Zhodik⁶

¹⁻⁶Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46,
Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}chintyam10@gmail.com, ²hasria0511@gmail.com, ³irwanhardikusuma@gmail.com,
⁴lulukaoniah12@gmail.com, ⁵novariman@gmail.com, ⁶zhakiapriyanzhodik97@gmail.com

(* : coresponding author)

Abstrak—Algoritma Decision Tree adalah metode pembelajaran mesin yang populer dan banyak diterapkan dalam berbagai bidang, seperti kesehatan, pendidikan, keuangan, dan pengelolaan sumber daya alam. Keunggulan utama algoritma ini terletak pada kemampuannya untuk memvisualisasikan proses pengambilan keputusan dalam bentuk hierarki yang mudah dipahami. Meskipun sederhana, algoritma ini memiliki berbagai varian dan metode pengoptimalan, seperti Random Forest, Gradient Boosting, dan Ensemble Learning, yang dirancang untuk meningkatkan akurasi dan mengurangi risiko overfitting. Artikel ini bertujuan untuk meninjau literatur mengenai implementasi dan pengembangan algoritma Decision Tree dalam berbagai aplikasi praktis. Melalui pendekatan tinjauan literatur sistematis, penelitian ini menganalisis penerapan Decision Tree dalam bidang perbankan, kesehatan, dan manajemen, dengan hasil yang menunjukkan bahwa algoritma ini efektif untuk tugas klasifikasi dan prediksi. Dalam sektor perbankan, Decision Tree berhasil mengidentifikasi faktor-faktor penting dalam klasifikasi pelanggan aktif dan tidak aktif. Dalam bidang kesehatan, model ini juga terbukti dapat memprediksi faktor-faktor penyebab penyakit dengan akurasi yang tinggi. Dengan mengidentifikasi pola dalam data, penelitian ini menggarisbawahi pentingnya algoritma Decision Tree dalam meningkatkan pemahaman dan pengambilan keputusan berbasis data yang lebih efektif.

Kata Kunci: Algoritma *Decision Tree*, *Machine Learning*, *Random Forest*, *Gradient Boosting*, *Ensemble Learning*

Abstract—The Decision Tree algorithm is a popular machine learning method widely applied across various fields such as healthcare, education, finance, and natural resource management. The primary advantage of this algorithm lies in its ability to visualize the decision-making process in an easily understandable hierarchical structure. Despite its simplicity, the algorithm offers various variants and optimization methods, such as Random Forest, Gradient Boosting, and Ensemble Learning, designed to improve accuracy and reduce the risk of overfitting. This article aims to review the literature on the implementation and development of the Decision Tree algorithm in various practical applications. Through a systematic literature review approach, this study analyzes the application of Decision Tree in banking, healthcare, and management, with results indicating that the algorithm is effective for classification and prediction tasks. In the banking sector, Decision Tree successfully identifies key factors in classifying active and inactive customers. In the healthcare field, the model has also proven capable of predicting disease-causing factors with high accuracy. By identifying patterns in data, this study highlights the significance of the Decision Tree algorithm in enhancing understanding and enabling more effective data-driven decision-making.

Keywords: *Decision Tree Algorithm*, *Machine Learning*, *Random Forest*, *Gradient Boosting*, *Ensemble Learning*

1. PENDAHULUAN

Algoritma Decision Tree merupakan salah satu metode pembelajaran mesin yang telah mendapatkan popularitas luas dan sering diterapkan dalam berbagai penelitian serta aplikasi praktis. Keunggulan algoritma ini terletak pada kemampuannya untuk memvisualisasikan proses pengambilan keputusan dalam bentuk hierarki yang mudah dipahami, serta fleksibilitasnya dalam menangani data kategorikal maupun numerik. Dalam beberapa dekade terakhir, Decision Tree telah diimplementasikan secara luas dalam berbagai bidang seperti kesehatan, pendidikan, keuangan, manufaktur, hingga pengelolaan sumber daya alam.

Meskipun terlihat sederhana, algoritma Decision Tree memiliki beragam varian dan metode pengoptimalan, seperti Random Forest, Gradient Boosting, dan Ensemble Learning, yang dirancang untuk meningkatkan akurasi dan mengurangi risiko overfitting. Selain itu, penelitian terus

berkembang untuk memahami bagaimana algoritma ini dapat disesuaikan dengan kebutuhan spesifik dalam berbagai konteks, termasuk data besar (big data), data tidak seimbang, dan data dengan dimensi tinggi.

Artikel ini bertujuan untuk menyajikan tinjauan literatur tentang implementasi dan pengembangan algoritma Decision Tree dalam berbagai konteks aplikasi. Dengan memahami perkembangan terkini dan berbagai pendekatan yang telah diterapkan, diharapkan artikel ini dapat menjadi acuan untuk peneliti dan praktisi dalam memilih dan mengimplementasikan algoritma Decision Tree secara lebih efektif.

2. METODE PENELITIAN

Penelitian ini dilakukan dengan pendekatan tinjauan literatur sistematis untuk mengidentifikasi, menganalisis, dan merangkum studi-studi terkait algoritma Decision Tree pada berbagai konteks aplikasi. Proses pengumpulan data dilakukan dengan menelusuri artikel-artikel ilmiah dan sumber dari Google Scholar. Pencarian difokuskan pada kata kunci seperti "Decision Tree," "Machine Learning," "Random Forest," "Gradient Boosting," dan "Decision Tree Applications". Seleksi artikel dilakukan dalam beberapa tahap, dimulai dari peninjauan judul dan abstrak untuk menentukan relevansi, diikuti oleh analisis isi penuh pada artikel yang memenuhi kriteria inklusi.

Kriteria inklusi meliputi studi yang membahas implementasi algoritma Decision Tree dalam konteks tertentu, seperti kesehatan, pendidikan, keuangan, dan data besar, serta publikasi yang diterbitkan dalam 5 tahun terakhir untuk memastikan pembahasan terkini. Kriteria eksklusi diterapkan pada artikel yang tidak memiliki metode yang jelas atau yang hanya membahas teori tanpa aplikasi. Data yang dikumpulkan kemudian diolah melalui analisis kualitatif untuk menemukan pola-pola, mengidentifikasi tantangan, dan mengeksplorasi peluang dalam pengembangan algoritma Decision Tree.

3. ANALISA DAN PEMBAHASAN

3.1 Pembahasan

Penelitian ini mengkaji secara mendalam penggunaan pohon keputusan sebagai alat analisis untuk menentukan klasifikasi pelanggan, terutama dalam membedakan pelanggan aktif dan tidak aktif pada sektor perbankan. Melalui pendekatan ini, pohon keputusan mampu memberikan visualisasi yang jelas mengenai hubungan kompleks antara variabel yang mempengaruhi status pelanggan. Model yang dihasilkan juga memungkinkan pemahaman yang lebih dalam terhadap pola data, sehingga dapat diinterpretasikan dengan mudah oleh pengambil keputusan.

Dalam penerapannya, algoritma pohon keputusan memproses dataset dengan berbagai atribut, termasuk informasi lokasi, identitas produk, nama pelanggan, saldo akun, dan status pelanggan. Hasil analisis menunjukkan bahwa atribut tertentu memiliki pengaruh lebih signifikan dibandingkan yang lain dalam menentukan klasifikasi pelanggan. Misalnya, saldo pelanggan dan ID produk sering kali menjadi variabel utama dalam membedakan pelanggan aktif dan tidak aktif.

Proses klasifikasi ini melibatkan pembangunan model yang mampu mengelompokkan data ke dalam dua kategori utama: pelanggan aktif dan pelanggan tidak aktif. Melalui evaluasi performa model, diperoleh hasil akurasi yang memadai, sehingga membuktikan bahwa algoritma pohon keputusan efektif untuk menangani masalah klasifikasi dalam konteks ini. Selain itu, model juga memberikan wawasan penting terkait atribut-atribut yang dapat dioptimalkan dalam strategi pemasaran layanan perbankan.

Hasil ini memberikan implikasi praktis bagi sektor perbankan, terutama dalam mengembangkan pendekatan pemasaran yang lebih tepat sasaran. Dengan memahami faktor utama yang memengaruhi status pelanggan, bank mampu menyusun strategi yang lebih fokus untuk mempertahankan pelanggan aktif sekaligus mengaktifkan kembali pelanggan pasif. Secara keseluruhan, penelitian ini menunjukkan potensi besar pohon keputusan sebagai alat pendukung pengambilan keputusan yang efektif dalam konteks pemasaran dan manajemen pelanggan.

3.2 Hasil Rangkuman

Tabel 1. Hasil Rangkuman

No	Penulis	Judul	Thn	Metode	Hasil
1	Rafael Nuansa Ramadhon, Aldino Ogi, Ari Permana Agung, Ryandra Putra, Siti Sarah Febrihartina, Uus Firdaus	Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank	2024	Decision Tree Metode (Pohon Keputusan)	Algoritma pohon keputusan memungkinkan bank untuk menganalisis pola perilaku nasabah, mengidentifikasi variabel kunci seperti saldo akun dan frekuensi transaksi, serta menyempurnakan strategi pemasaran guna meningkatkan tingkat retensi, kepuasan nasabah, dan profitabilitas perusahaan.
2	Agung Wibowo, Indarti, Dewi Laraswati	Komparasi Algoritma Decision Tree, Random Forest dan SVM untuk Prognosis COVID-19	2024	Decision Tree, Random Forest, SVM (Support Vector Machine)	Algoritma decision tree menunjukkan akurasi, recall, f1-score, dan macro average terbaik dengan nilai false positif terendah dibandingkan algoritma random forest dan SVM.
3	M. Fahrul Rizki Aditya, Nuril Lutvi Azizah dan Uce Indahyanti	Prediksi Penyakit Hipertensi Menggunakan metode Decision Tree dan Random Forest	2024	Decision Tree dan Random Forest	Metode pohon keputusan menghasilkan nilai akurasi, presisi, recall, dan F1-score mencapai 100%. Demikian pula, metode random forest memberikan hasil serupa dengan nilai akurasi, presisi, recall, dan F1-score yang juga mencapai 100%.
4	Retnani Latifah, Emi Setia Wulandari dan Priadhana Edi Kreshna	Model Decision Tree untuk Prediksi Jadwal Kerja menggunakan Scikit-Learn	2019	Mengembangkan model Decision Tree menggunakan Scikit-Learn	Penelitian ini menemukan bahwa model Decision Tree yang dibangun menggunakan Scikit-Learn mampu memprediksi jadwal kerja dari 54 data yang terdiri atas variabel numerik dan kategorikal, dengan akurasi melebihi 0,8 dan presisi lebih dari 0,9. Pemisahan data terbaik dilakukan dengan pembagian 70% untuk data latih dan 30% untuk data uji. Namun, penggunaan LabelEncoder untuk mengonversi data kategorikal menjadi numerik kurang ideal, karena dapat

					menyebabkan interpretasi data kategorikal sebagai urutan angka, sehingga berpotensi memengaruhi keakuratan prediksi.
5	Asep Budiyanto, Dodi Solihudin, Ryan Hamonangan, Cep Lukman Rohmat, Ade Rizki Rinaldi	Analisis Keadaan Stunting pada Kelompok Balita di Kecamatan Tukdana dengan Pendekatan Decision Trees	2024	Decision Tree Metode (Pohon Keputusan, Algoritma C4.5)	Algoritma C4.5 mengidentifikasi bahwa STATUS BB/U dan STATUS PB/U merupakan atribut utama dalam analisis stunting di Kecamatan Tukdana. Model ini menunjukkan hasil yang akurat dengan prevalensi stunting yang relatif rendah, meskipun tetap diperlukan pemantauan lebih lanjut untuk memastikan keakuratan dan keberlanjutan analisis.
6	Baiq Andriska Candra Permana, Muhammad Sadali, Ramli Ahmad	Penerapan Model Decision Tree Menggunakan Python Untuk Prediksi Faktor Dominan Penyebab Penyakit Stroke	2024	Decision Tree Metode (Pohon Keputusan)	Hasil penelitian menunjukkan bahwa model decision tree memiliki kinerja yang cukup baik dalam memprediksi faktor-faktor utama penyebab stroke, dengan tingkat akurasi mencapai 91% dan nilai AUC sebesar 0,5. Penelitian ini juga menemukan bahwa usia merupakan faktor dominan penyebab stroke, yang turut dipengaruhi oleh faktor lain seperti kadar gula darah, berat badan, dan hipertensi.

3.3 Pembahasan Sub

Studi yang dilakukan oleh Rafael Nuansa Ramadhon dkk menunjukkan bahwa algoritma pohon keputusan dapat digunakan untuk menganalisis perilaku nasabah, mengidentifikasi faktor utama seperti saldo dan frekuensi transaksi, serta mengembangkan strategi pemasaran yang lebih efektif guna meningkatkan retensi, kepuasan, dan profitabilitas.

Agung Wibowo dkk, melakukan penelitian Komparasi Algoritma Decision Tree, Random Forest dan SVM untuk Prognosis COVID-19 memperoleh hasil algoritma decision tree menunjukkan akurasi, recall, f1-score, dan macro average terbaik dengan nilai false positif terendah dibandingkan algoritma random forest dan SVM.

M. Fahrul Rizki Aditya dkk membandingkan metode decision tree dengan random forest. Hasilnya, kedua metode tersebut menunjukkan performa yang sama, dengan nilai akurasi, presisi, recall, dan F1-score masing-masing mencapai 100%.

Retnani Latifah dkk menggunakan algoritma Decision Tree untuk memprediksi jadwal kerja berdasarkan data kategorikal dan numerikal. Dengan memanfaatkan library scikit-learn di Python, model ini mencapai akurasi di atas 0,7 dan presisi lebih dari 0,9 dari 54 data yang terdiri atas 3

variabel kategorikal dan 1 variabel numerikal. Hasil penelitian menunjukkan bahwa Decision Tree merupakan metode yang efektif untuk prediksi jadwal kerja dan memiliki performa yang lebih baik dibandingkan algoritma KNN.

Asep Budiyanto dkk dalam penelitiannya mengidentifikasi faktor-faktor demografis dan kesehatan yang memengaruhi stunting pada bayi di Kecamatan Tukdana, Kabupaten Indramayu. Hasil analisis menunjukkan bahwa usia ibu, akses terhadap air bersih, fasilitas sanitasi, serta status berat dan panjang badan bayi merupakan faktor signifikan yang berkontribusi pada stunting. Dengan menggunakan metode Decision Trees, penelitian ini mencapai akurasi sebesar 95,43%. Meskipun sebagian besar bayi di Kecamatan Tukdana memiliki status gizi yang baik, temuan ini menyoroti pentingnya pemantauan berkelanjutan dan langkah-langkah pencegahan untuk meningkatkan kesejahteraan gizi bayi serta merancang intervensi yang lebih efektif di wilayah tersebut dan daerah lain di Indonesia.

Baiq Andriksa Candra Permana dkk melakukan penelitian penerapan model decision tree menggunakan Python untuk memprediksi faktor dominan penyebab penyakit stroke. Hasil penelitian menunjukkan bahwa model decision tree memiliki performa yang cukup baik dalam mengidentifikasi faktor utama penyebab stroke, dengan tingkat akurasi mencapai 91% dan nilai AUC sebesar 0,5. Penelitian ini juga menemukan bahwa usia merupakan faktor utama yang memicu stroke, didukung oleh faktor lain seperti kadar gula darah, berat badan, dan hipertensi.

4. KESIMPULAN

Berdasarkan hasil tinjauan literatur, algoritma Decision Tree telah menunjukkan efektivitasnya sebagai metode analisis yang dapat diterapkan di berbagai bidang. Penelitian-penelitian yang telah dilakukan menunjukkan bahwa Decision Tree mampu mengidentifikasi faktor penting, seperti dalam sektor perbankan untuk memahami perilaku pelanggan, serta di sektor kesehatan untuk memprediksi faktor dominan yang mempengaruhi penyakit, seperti hipertensi, COVID-19, dan stunting. Beberapa penelitian juga menunjukkan tingkat akurasi yang sangat tinggi, seperti pada prediksi penyakit hipertensi dan analisis faktor penyebab stroke, yang membuktikan kemampuan algoritma ini dalam menghasilkan model prediksi yang andal. Selain itu, fleksibilitas algoritma Decision Tree memungkinkan penggunaannya pada data kategorikal maupun numerik, serta efektif dalam menangani dataset dengan atribut yang kompleks. Variasi algoritma, seperti C4.5 dan metode ensemble termasuk Random Forest, memberikan hasil yang lebih optimal dengan mengurangi risiko overfitting.

Meskipun demikian, penelitian juga mengidentifikasi beberapa tantangan, seperti pengolahan data kategorikal yang memerlukan perhatian khusus dan kebutuhan untuk mengoptimalkan parameter model agar hasil prediksi lebih akurat. Oleh karena itu, Decision Tree dan variannya tetap menjadi pilihan yang menjanjikan bagi peneliti dan praktisi dalam menyelesaikan berbagai masalah analisis data. Penelitian mendatang diharapkan dapat mengeksplorasi lebih jauh penggunaan algoritma ini dengan dataset yang lebih besar dan kompleks, serta mengintegrasikannya dengan teknik machine learning lainnya untuk meningkatkan performa dan aplikasinya di berbagai sektor.

REFERENCES

- Aditya, M. F. R., Azizah, N. L., & Indahyanti, U. (2024). Prediksi penyakit hipertensi menggunakan metode decision tree dan random forest. Diakses 18 Desember 2024, dari <https://ejournal.jakstik.ac.id/index.php/komputasi/article/view/3503>
- Budiyanto, A., Solihudin, D., Hamonangan, R., Rohmat, C. L., & Rinaldi, A. R. (2024). Analisis keadaan stunting pada kelompok balita di Kecamatan Tukdana dengan pendekatan decision trees. Diakses 18 Desember 2024, dari <https://publikasiilmiah.unwas.ac.id/JINRPL/article/view/10230>
- Latifah, R., Wulandari, E. S., & Kreshna, P. E. (2024). Model decision tree untuk prediksi jadwal kerja menggunakan scikit-learn. Diakses 18 Desember 2024, dari <https://jurnal.umj.ac.id/index.php/semnastek/article/view/5239>

- Pandia, M., Sihombing, P., Simamora, P., & Kaban, R. (2024). Kajian literatur multimedia retrieval: Machine learning untuk pengenalan wajah. Diakses 19 Desember 2024, dari <https://ejournal.sisfokomtek.org/index.php/jikom/article/view/2758>
- Permana, B. A. C., Sadali, M., & Ahmad, R. (2024). Penerapan model decision tree menggunakan python untuk prediksi faktor dominan penyebab penyakit stroke. Diakses 18 Desember 2024, dari <https://e-journal.hamzanwadi.ac.id/index.php/infotek/article/view/23232>
- 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. Diakses 18 Desember 2024, dari <https://ojs.unida.ac.id/karimahtauhid/article/view/11952>
- Rahadyan, R. V., & Arri W, R. (2024). Literature review: Implementation of machine learning in e-learning systems. Diakses 18 Desember 2024, dari http://dosen.unimma.ac.id/public/document/seminarilmiah/2376-Article_Text-4667-1-10-20230116_4.pdf
- Wibowo, A., Indarti, & Laraswati, D. (2024). Komparasi algoritma decision tree, random forest dan SVM untuk prognosis COVID-19. Diakses 18 Desember 2024, dari <https://jurnal.bsi.ac.id/index.php/imtechno/article/download/2868/1713/18347>