

Implementasi Algoritma C4.5 dalam Menentukan Preferensi Pemilihan Jenis Buku Berdasarkan Minat Baca Pengunjung (Studi Kasus: Perpustakaan Daerah Kota Tangerang Selatan)

Istikomah¹, Ari Mulyoto²

^{1,2} Program Studi Teknik Informatika, Universitas Pamulang
e-mail: istiistikomah774@gmail.com¹, dosen00236@unpam.ac.id²

Abstrak

Preferensi pengunjung dalam memilih jenis buku merupakan salah satu informasi penting yang dapat dimanfaatkan oleh perpustakaan daerah untuk mendukung pengambilan keputusan dalam penyediaan koleksi buku yang sesuai dengan kebutuhan pengunjung. Penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam preferensi pemilihan jenis buku berdasarkan minat baca pengunjung di Perpustakaan Daerah Kota Tangerang Selatan. Data yang digunakan dalam penelitian ini meliputi usia, status, tingkat minat baca, frekuensi membaca, tujuan membaca, dan pertimbangan dalam memilih bacaan. Seluruh data tersebut digunakan untuk mengklasifikasikan jenis buku yang disukai pengunjung. Proses penelitian dilakukan melalui tahapan pengumpulan data, pembersihan data, transformasi data, pembentukan pohon keputusan menggunakan algoritma C4.5. Selanjutnya, hasil klasifikasi dievaluasi menggunakan confusion matrix, serta hasilnya diimplementasikan ke dalam sistem berbasis website. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu membentuk model klasifikasi untuk menentukan preferensi jenis buku yang diminati berdasarkan minat baca pengunjung. Selain itu, sistem yang dirancang dapat menampilkan hasil analisa dan rekomendasi jenis buku yang sesuai dengan preferensi jenis buku. Berdasarkan hasil tersebut, model yang dibuat dapat digunakan secara lebih efisien dan terarah untuk mendukung keputusan tentang pengembangan koleksi perpustakaan.

Kata Kunci: Minat Baca, Preferensi Jenis Buku, Algoritma C4.5, Data Mining, Perpustakaan Daerah.

Abstract

Visitors' preferences in selecting book types are important information that can be utilized by regional libraries to support decision-making in developing book collections that meet visitors' needs. This study aimed to implement the C4.5 algorithm to determine visitors' preferences for book types based on their reading interests at the South Tangerang City Regional Library. The data used in this study included age, status, reading interest level, reading frequency, reading purposes, and considerations in selecting reading materials. These variables were utilized to classify visitors' preferred book types. The research process was conducted through the stages

of data collection, data cleaning, data transformation, and decision tree construction using the C4.5 algorithm. Furthermore, the classification model was evaluated using a confusion matrix, and the results were implemented into a web-based system. The results show that the C4.5 algorithm is capable of generating a classification model for determining visitors' preferred book types based on their reading interests. The developed web-based system provides classification results and book recommendations that correspond to visitors' preferences. Therefore, the proposed model can be used to support library decision-making in developing book collections more effectively and according to visitors' reading preferences.

Keywords : C4.5 Algorithm, Book Preference, Reading Interest, Classification, Decision Tree.

LATAR BELAKANG

Minat baca merupakan salah satu faktor yang memengaruhi kecenderungan seseorang dalam memilih bahan bacaan. Perbedaan tingkat minat baca dapat memengaruhi preferensi terhadap jenis buku yang dipilih, baik buku fiksi maupun nonfiksi. Informasi tentang preferensi ini penting bagi perpustakaan karena dapat digunakan sebagai dasar untuk menambah koleksi yang lebih relevan untuk memenuhi kebutuhan dan karakteristik pengunjung. Dengan memahami preferensi ini, perpustakaan dapat menyediakan koleksi yang lebih relevan sehingga mampu meningkatkan pemahaman pembaca (Muaidi, 2022).

Perpustakaan umum memiliki peran strategis sebagai pusat penyedia informasi dan sumber pembelajaran bagi masyarakat. Seiring dengan kemajuan teknologi dan meningkatnya kebutuhan informasi, perpustakaan tidak hanya diharapkan menyediakan koleksi bahan bacaan yang beragam, tetapi juga memahami kebutuhan dan preferensi pengunjung agar pengembangan koleksi dapat dilakukan dengan lebih baik. Kesesuaian antara koleksi perpustakaan dengan preferensi pengunjung dapat meningkatkan pemanfaatan koleksi dan kualitas layanan yang diberikan. Oleh karena itu, informasi mengenai kecenderungan pengunjung dalam memilih jenis buku menjadi salah satu aspek penting yang dapat digunakan sebagai dasar untuk membuat keputusan tentang pengembangan koleksi perpustakaan (Apriliya & Ali, 2024).

Lebih lanjut, penerapan teknik data mining menjadi salah satu solusi yang dapat digunakan untuk menggali informasi tersembunyi dari sejumlah besar data. Data mining memungkinkan proses pencarian pola, tren, dan hubungan antar data yang sebelumnya tidak terlihat secara langsung. Dalam penelitian ini, teknik *data mining* dimanfaatkan untuk mengolah data hasil kuesioner mengenai tingkat minat baca pengunjung sehingga dapat menghasilkan informasi mengenai preferensi pemilihan jenis buku. Informasi tersebut diharapkan dapat menjadi dasar bagi pihak perpustakaan dalam mendukung pengambilan keputusan terkait pengembangan koleksi buku yang lebih sesuai dengan kebutuhan dan preferensi pengunjung (Wijaya et al., 2024).

Salah satu metode dalam data mining yang sering digunakan untuk proses klasifikasi adalah algoritma C4.5, yang bekerja dengan membentuk pohon keputusan berdasarkan atribut tertentu. Menurut penelitian yang dilakukan oleh (Nugraha et al., 2022), algoritma C4.5 mampu menghasilkan klasifikasi data dan menghasilkan aturan keputusan yang mudah

dipahami. Karakteristik tersebut menjadikan algoritma C4.5 sesuai untuk mengolah data kuesioner yang memiliki variasi jawaban serta perbedaan tingkat minat baca responden. Data tentang minat baca pengunjung dapat digunakan untuk mengklasifikasikan preferensi jenis buku dengan menggunakan algoritma C4.5. Hasil analisis ini dapat digunakan sebagai dasar untuk merancang program literasi dan pengembangan koleksi yang lebih tepat sasaran di perpustakaan daerah.

Penelitian oleh (Nugraha et al., 2022) menunjukkan bahwa algoritma C4.5 mampu menghasilkan model klasifikasi yang mudah dipahami dalam proses pengolahan data. Penelitian lain yang dilakukan oleh (Irfiani et al., 2023) menerapkan algoritma C4.5 dalam mengklasifikasikan minat baca pada perpustakaan daerah. Di sisi lain, penelitian yang dilakukan oleh (Sibuea et al., 2025) menganalisis preferensi bacaan fiksi dan nonfiksi tanpa menerapkan teknik *data mining*. Berdasarkan penelitian-penelitian tersebut, belum banyak penelitian yang mengimplementasikan algoritma C4.5 untuk mengklasifikasikan preferensi pemilihan jenis buku berdasarkan tingkat minat baca pengunjung perpustakaan serta mengintegrasikan hasil klasifikasi tersebut ke dalam sistem berbasis web sebagai pendukung pengambilan keputusan pengembangan koleksi. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut.

Berdasarkan penjelasan ini, dapat disimpulkan bahwa minat baca dan preferensi pemilihan jenis buku merupakan informasi penting yang dapat dimanfaatkan oleh perpustakaan dalam mendukung pengambilan keputusan terkait pengembangan koleksi. Namun, minat baca dan preferensi jenis bacaan yang berbeda pada setiap pengunjung menyebabkan kebutuhan informasi yang berbeda, sehingga perpustakaan perlu memahami pola dan kecenderungan minat baca secara lebih dalam. Oleh karena itu, penelitian ini menerapkan teknik *data mining* menggunakan algoritma C4.5 untuk mengklasifikasikan preferensi pemilihan jenis buku berdasarkan tingkat minat baca pengunjung. Diharapkan hasil penelitian ini dapat memberikan informasi yang membantu Dinas Perpustakaan dan Kearsipan Kota Tangerang Selatan dalam menentukan pengembangan koleksi buku serta meningkatkan kualitas layanan perpustakaan sesuai dengan kebutuhan dan preferensi pengunjung.

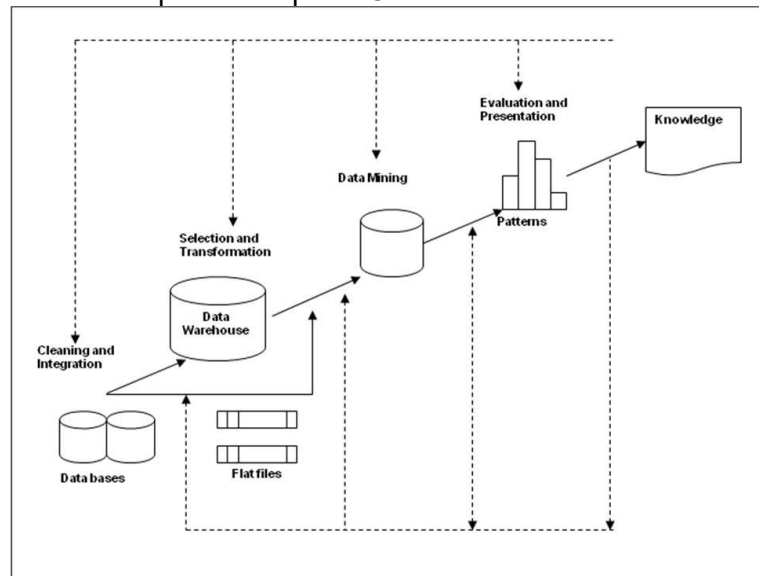
METODE

Penelitian ini menggunakan pendekatan kuantitatif karena data yang dianalisis berupa data hasil kuesioner yang diolah secara sistematis menggunakan teknik data mining. Dengan demikian, pendekatan kuantitatif digunakan untuk menghasilkan hasil analisis yang objektif berdasarkan data numerik, yang dapat membantu proses klasifikasi preferensi jenis buku (Sugiyono, 2013). Proses *data mining* pada penelitian ini mengacu pada metode *Knowledge Discovery in Databases* (KDD) yang terdiri atas tahapan *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation* (Wijaya et al., 2024).

Penelitian ini dilakukan di Dinas Perpustakaan dan Kearsipan Kota Tangerang Selatan. Objek penelitian ini berupa data hasil kuesioner pengunjung terkait minat baca yang digunakan untuk mengklasifikasikan preferensi pemilihan jenis buku. Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh melalui penyebaran kuesioner kepada pengunjung Dinas Perpustakaan dan Kearsipan Kota Tangerang Selatan. Kuesioner digunakan untuk memperoleh informasi terkait karakteristik responden, tingkat minat baca,

frekuensi membaca, tujuan membaca, dan pertimbangan memilih bacaan. Data yang diperoleh dari hasil penyebaran sebanyak 92 responden, namun setelah dilakukan proses pemeriksaan dan pembersihan data diperoleh 83 data yang digunakan sebagai dataset penelitian.

Penelitian ini menerapkan metode *Knowledge Discovery in Database (KDD)* sebagai tahapan dalam proses *data mining*. Menurut (Wijaya et al., 2024) tahapan KDD terdiri atas lima tahapan, yaitu *data selection*, *data preprocessing*, *data transformation*, *data mining*, dan *evaluation*. Tahap penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan *Knowledge Discovery in Database (KDD)*

1. *Selection*

Tahap seleksi data merupakan proses pemilihan data yang akan digunakan dalam penelitian. Data diperoleh melalui penyebaran kuesioner kepada pengunjung Perpustakaan Daerah Kota Tangerang Selatan. Atribut yang digunakan meliputi usia, status, tingkat minat baca, frekuensi membaca, tujuan membaca, pertimbangan dalam memilih buku, dan preferensi jenis buku sebagai atribut target.

2. *Preprocessing*

Pada tahap ini, dilakukan pemeriksaan terhadap data yang telah dikumpulkan untuk mengetahui adanya data yang kosong (*missing value*), data ganda, maupun data yang tidak konsisten.

3. *Transformation*

Tahap transformasi merupakan proses mengubah data ke dalam bentuk yang sesuai dengan kebutuhan algoritma C4.5. Pada penelitian ini dilakukan pengelompokan data berdasarkan kategori yang telah ditentukan.

4. *Data Mining*

Tahap *data mining* dilakukan dengan menerapkan algoritma C4.5 untuk membentuk model klasifikasi berupa *decision tree*. Model yang dihasilkan digunakan untuk mengklasifikasikan preferensi pemilihan jenis buku berdasarkan tingkat minat baca pengunjung.

5. *Evaluation*

Tahap terakhir adalah *evaluation*, yaitu mengevaluasi kinerja model menggunakan *confusion matrix*. Evaluasi dilakukan dengan menghitung nilai accuracy, precision, dan recall untuk mengetahui tingkat performa model klasifikasi yang dihasilkan.

Salah satu metode klasifikasi dalam teknik *data mining* yang digunakan pada penelitian ini adalah algoritma C4.5. Algoritma ini bekerja dengan membentuk pohon keputusan (*decision tree*) berdasarkan atribut yang memiliki nilai *information gain* tertinggi sehingga mampu menghasilkan aturan klasifikasi yang mudah dipahami dan diinterpretasikan (Quinlan, 1993). Proses pembentukan *decision tree* diawali dengan menghitung nilai *entropy*. Nilai *entropy* dihitung menggunakan Persamaan (1).

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

Selanjutnya dilanjutkan perhitungan *information gain* untuk menentukan atribut terbaik yang akan dijadikan sebagai node pada *decision tree*. Nilai *information gain* dihitung dengan Persamaan (2).

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

Berdasarkan hasil perhitungan *entropy* dan *information gain*, algoritma C4.5 membentuk *decision tree* yang digunakan sebagai model klasifikasi untuk menentukan preferensi pemilihan jenis buku.

HASIL DAN PEMBAHASAN

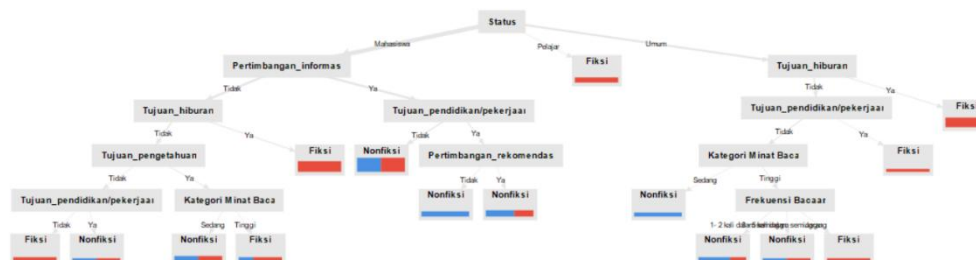
Penelitian ini menggunakan data hasil kusioner pengunjung Dinas Perpustakaan dan Kearsipan Kota Tangerang Selatan. Setelah melalui proses *preprocessing*, diperoleh sebanyak 83 data yang digunakan sebagai dataset penelitian. Data tersebut terdiri atas beberapa atribut, yaitu usia, status, tingkat minat baca, frekuensi membaca, tujuan membaca, pertimbangan dalam memilih buku, dan preferensi jenis buku sebagai atribut target.

Tabel 1. Atribut Dataset Penelitian

Atribut	Keterangan
Usia	Remaja, Dewasa Awal, Dewasa
Status	Pelajar, Mahasiswa, Umum
Tingkat Minat Baca	Rendah, Sedang, Tinggi
Frekuensi Membaca	Jarang, 1-2 kali/minggu, 3-5 kali/minggu, Setiap hari

Tujuan Membaca	Hiburan, Pengetahuan, Pendidikan/Pekerjaan, Waktu Luang
Pertimbangan Memilih Bacaan	Isi Cerita, Informasi, Kebutuhan, Rekomendasi
Target	Preferensi Jenis Buku

Selanjutnya, algoritma C4.5 digunakan untuk memproses dataset untuk membentuk model klasifikasi berupa pohon keputusan. Proses pembentukan pohon keputusan didasarkan pada perhitungan nilai entropy dan information gain pada setiap atribut. Aturan klasifikasi yang dihasilkan digunakan untuk menentukan preferensi pemilihan jenis buku berdasarkan tingkat minat baca pengunjung. Hasil pembentukan *decision tree* dapat dilihat pada Gambar 2.



Gambar 2. Pohon Keputusan

Berdasarkan *decision tree* yang dihasilkan, diperoleh beberapa aturan klasifikasi yang menunjukkan hubungan antara atribut-atribut yang digunakan dengan preferensi jenis buku pengunjung. Aturan-aturan tersebut menjadi dasar dalam proses klasifikasi sehingga sistem dapat menentukan apakah pengunjung memiliki preferensi terhadap buku fiksi atau nonfiksi berdasarkan karakteristik yang dimiliki.

Untuk mengetahui tingkat kinerja algoritma C4.5, model klasifikasi diuji menggunakan *confusion matrix*. Hasilnya menunjukkan bahwa model memperoleh nilai *accuracy* sebesar 64,71%, nilai *precision* sebesar 81,82%, dan nilai *recall* sebesar 69,23% pada kelas fiksi. Sementara itu, pada kelas Nonfiksi, diperoleh *precision* sebesar 33,33% dan *recall* sebesar 50,00%. Nilai tersebut menunjukkan bahwa kemampuan model dalam mengklasifikasikan preferensi buku nonfiksi masih lebih rendah dibandingkan kelas fiksi.

accuracy: 64.71%			
	true Fiksi	true Nonfiksi	class precision
pred. Fiksi	9	2	81.82%
pred. Nonfiksi	4	2	33.33%
class recall	69.23%	50.00%	

Gambar 3. Hasil Confusion Matrix

Hasil klasifikasi kemudian diterapkan ke dalam sistem berbasis web yang dibuat menggunakan PHP dan MySQL. Sistem menyediakan beberapa fitur utama, antara lain pengelolaan data responden, visualisasi decision tree, hasil evaluasi model, serta rekomendasi jenis buku berdasarkan preferensi pengunjung. Implementasi sistem ini diharapkan dapat membantu Dinas Perpustakaan dan Kearsipan Kota Tangerang Selatan dalam menentukan pengembangan koleksi buku secara lebih tepat sasaran sesuai dengan kebutuhan pengunjung.

Hasil penelitian ini sejalan dengan penelitian (Nugraha et al., 2022) yang menunjukkan bahwa algoritma C4.5 mampu menghasilkan model klasifikasi yang mudah dipahami melalui pembentukan *decision tree*. Perbedaan penelitian ini terletak pada objek penelitian, yaitu pengunjung Dinas Perpustakaan dan Kearsipan Kota Tangerang Selatan, serta implementasi hasil klasifikasi ke dalam sistem berbasis web yang digunakan sebagai pendukung pengambilan keputusan dalam pengembangan koleksi perpustakaan.

SIMPULAN

Penelitian ini menunjukkan bahwa algoritma C4.5 dapat diterapkan untuk mengklasifikasikan preferensi pemilihan jenis buku berdasarkan tingkat minat baca pengunjung Dinas Perpustakaan dan Kearsipan Kota Tangerang Selatan dengan tingkat accuracy sebesar 64,71%. Hasil klasifikasi berhasil diimplementasikan pada sistem berbasis web yang membantu perpustakaan membuat keputusan tentang pengembangan koleksi buku. Selanjutnya, pihak perpustakaan disarankan untuk menambah jumlah data dan atribut penelitian agar model klasifikasi menjadi lebih akurat serta memanfaatkan hasil klasifikasi sebagai dasar dalam pengembangan koleksi yang sesuai dengan kebutuhan pengunjung.

DAFTAR PUSTAKA

- Apriliya, D., & Ali, M. N. (2024). Upaya Perpustakaan Umum Kota Tebing Tinggi Dalam Peningkatan Literasi Informasi Melalui Program Ngopi Literasi. *Jurnal Ilmiah Dinamika Sosial*, 8(2), 160–173. <https://doi.org/10.38043/Jids.V8i2.5409>
- Irfiani, E., Kusnadi, Y., & Handayanna, F. (2023). Implementasi Data Mining Dalam Klasifikasi Minat Baca Pada Perpustakaan Daerah Menggunakan Algoritma C4 . 5. *Journal Of Information System*, 8(2), 106–114. <https://doi.org/10.33633/Joins.V8i2.8004>
- Muaidi, M. (2022). Minat Baca Praja Terhadap Buku Novel Di Perpustakaan IPDN Kampus Nusa Tenggara Barat. *THE LIGHT : Journal Of Librarianship And Information Science*, 2(1), 23–34.
- Nugraha, F. F., Sunandar, I., & Juliane, C. (2022). Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Teknik Informatika Dan Sistem Informasi*, 9(4), 2862–2869.
- Quinlan, J. R. (1993). *C4.5: Programs For Machine Learning* (C. Leyba, D. Sery, & J. Hammett (Eds.)). Morgan Kaufmann Publishers.
- Sibuea, P., Angin, B. S. P., Ramadani, S. O., Dumila, Z. A., & Farel, M. (2025). Analisis Preferensi Minat Baca Mahasiswa Terhadap Bacaan Fiksi Dan Nonfiksi Di Universitas Islam Negeri Sumatera Utara. *Didaktik : Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 11(4), 250–257. <https://doi.org/10.36989/Didaktik.V11i04.10349>

- Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif Dan R & D*. Alfabeta.
- Sugiyono. (2014). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sugiyono. (2015). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sugiyono. (2016). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sugiyono. (2018). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sugiyono. (2020). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D (Edisi 2)*. Bandung: Alfabeta.
- Wijaya, A., Pinastawa, I. W. R., Gusti, K. W., Christo, P., Judijanto, L., Kurniawan, M., Risqiwahid, M., Budianingsih, Sabilirasyad, I., Izzati, A. N., Pratama, J. D., Zuriati, & Sriyanto. (2024). *Data Mining*. PT Penamuda Media.