

Klasifikasi Hobi Membaca pada Mahasiswa menggunakan Metode C4.5 pada Aplikasi RapidMiner

Aurelia Yolanda^{1*}, Muhammad Abdillah Nur², Eeng Sri Yuliyani³, Muhammad Rifqi Setiawan⁴, Lutfi Rizalwan⁵, Annida Purnamawati⁶

^{1,2,3,4,5}Program Studi Sistem Informasi, Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98, Jakarta Pusat, Indonesia

⁶Program Studi Sistem Informasi, Universitas Bina Sarana Informatika Kampus Kota Yogyakarta
Jl. Ringroad Barat, Ambar Ketawang, Gamping, Yogyakarta, Indonesia

email korespondensi: aureliayolada97@gmail.com

Submit: 19-12-2025 | Revisi: 23-12-2025 | Terima: 26-12-2025 | Terbit online: 28-12-2025

Abstrak - Berbagai macam hobi di dunia yang dapat menjadi hal yang positif bagi individu tersebut. Bermula dari minatnya seseorang terhadap sesuatu yang akan menjadi hobi individu itu sendiri, Salah satunya adalah hobi dalam membaca yang sangat penting bagi individu tersebut dalam hal mencari informasi ataupun mengisi waktu luang. Berbagai banyak jenis genre buku yang sangat disukai oleh individu yaitu genre fantasi, fiksi ilmiah, romansa, horor, misteri, biografi, sejarah, sains, motivasi dan sebagainya dengan berbagai tempat bacaan dalam media cetak maupun *online*. Dalam penelitian ini akan bertujuan mengetahui hobi membaca dan mengetahui genre minat yang menjadi hobi membaca bagi mahasiswa pada Universitas Tanjungpura di Pontianak dengan menggunakan Metode C4.5 serta model *Decision Tree* yang di mana metode ini akan di uji menggunakan perangkat lunak *RapidMiner Altair AI Studio Ver 2026* dan telah didapatkan hasil *accuracy* sebesar 64,42% pada kategori atribut hobi membaca dengan 3 pilihan yaitu Sangat Minat, Netral dan Kurang Minat.

Kata Kunci : Hobi Membaca, Genre, Metode C4.5

Abstract - There are various hobbies in the world that can be positive for the individual. Starting from a person's interest in something that will become the individual's hobby, one of which is a hobby in reading which is very important for the individual in terms of seeking information or filling free time. Various types of book genres that are very popular with individuals are fantasy, science fiction, romance, horror, mystery, biography, history, science, motivation and so on with various reading places in print and online media. In this study, the aim will be to find out reading hobbies and find out the genre of interest that becomes a reading hobby for students at Tanjungpura University in Pontianak using the C4.5 Method and the Decision Tree model where this method will be tested using *RapidMiner Altair AI Studio Ver 2026* software and has obtained an *accuracy* result of 64.42% in the reading hobby attribute category with 3 options, namely Very Interested, Neutral and Less Interested.

Keywords: Hobby of Reading, Genre, C4.5 Method

1. Pendahuluan

Hobi merupakan hal-hal yang kita minati secara individu maupun kelompok yang mana hobi dapat berupa hal fisik maupun lainnya, banyak macam hobi di dunia yang salah satunya membaca, membaca adalah hobi yang kita minati dan merupakan jendela dunia dengan berbagai genre yang diminati karena pada dasarnya membaca sangat penting bagi masyarakat saat ini dengan berbagai ilmu – ilmu serta informasi yang penting terdapat di dalamnya, Banyak jenis bacaan yang menarik untuk di kulik dari berbagai bacaan yang bukan hanya terdapat di buku cetak/fisik saja tetapi juga terdapat dalam bentuk *E-Book* yang diberikan oleh penyedia tempat bacaan yang mengutamakan bacaan yang lebih fleksibel karena bisa di baca melalui ponsel pintar secara gratis ataupun berbayar. Masyarakat maupun muda sampai tua saat ini pasti memiliki jenis genre atau jenis buku yang sering dibaca dikala senggang dalam bentuk novel, komik, surat kabar, majalah dan lain-lain dalam bentuk fisik maupun



elektronik dalam menambah minat membaca dari berbagai kalangan muda dan tua. Dalam kalangan muda terutama mahasiswa dituntut untuk banyak meningkatkan minat membaca dalam mendapat informasi dan meningkatkan banyak pengetahuan walau dari berbagai genre.

Minat adalah seseorang yang menginginkan sesuatu, dalam segi etomolgi minat disimpulkan kemauan dalam mempelajari serta mencari sesuatu. Dalam hal minat pada sesuatu akan menjadi hal yang disukai untuk dilakukan. Pengertian minat membaca adalah keinginan yang kuat dalam menafsirkan setiap kata-kata yang tertulis dalam sebuah buku atau yang kita telah baca [1]. Minat membaca memberikan peranan penting dalam mengasah kemampuan dalam hal berpikir kritis serta meningkatkan sikap menganalisis suatu bacaan [2] Minat baca merupakan satu hal faktor penting untuk meningkatkan kualitas dalam pendidikan mahasiswa [3]. Dapat disimpulkan bahwa minat baca bahwasannya merupakan sumber memberikan peranan mengasah berpikir intelektual, menganalisis sebuah bacaan serta memberikan kualitas pendidikan yang baik bagi mahasiswa. Dari hal diatas bahwa minat baca pada mahasiswa akan menjadi hobi yang sering dilakukan oleh mahasiswa itu sendiri.

Genre dapat didefinisikan sebagai pola-pola struktur karakteristik dalam sebuah cerita berdasarkan genre itu sendiri, Pada Kamus Besar Bahasa Indonesia (KBBI), genre dapat dimaksud jenis, tipe, kelompok sastra atas dasar bentuknya [4] [5]. Berdasarkan penelitian, penulis ingin mengetahui hobi membaca pada mahasiwa dengan menggunakan metode data mining yaitu algoritma C4.5 dan *decision tree* dalam mengklasifikasikan kategori hobi membaca yang menjadi pemicu hobi membaca yang dilakukan mahasiswa dengan lokasi pengambilan data di salah satu universitas negeri di pontianak yaitu Universitas Tanjungpura, Universitas Tanjungpura adalah Univeristas tua yang didirikan pada tahun 1959 hingga sekarang yang memiliki 9 fakultas yaitu Hukum, Ekonomi dan Bisnis, Pertanian, Teknik, Isip, Kip, Kehutanan, MIPA, dan Kedokteran [6]. Penulis ingin mengklasifikasikan berdasarkan 3 atribut yaitu Hobi Membaca, *Genre* dan Jenis Kelamin pada mahasiswa Universitas Tanjungpura dan penelitian ini dibuat dengan tujuan mengetahui hobi membaca buku dengan *genre* yang paling diminati oleh mahasiswa yang dimana penelitian ini dapat menjadi acuan keputusan bagi petugas perpustakaan dalam penambahan buku sebagai penarik minat dalam mengunjungi perpustakaan di Universitas Tanjungpura.

2. Metode Penelitian

2.1. Pengumpulan Data

Dataset penelitian ini diambil dari Universitas Tanjungpura, sebuah universitas negeri di Pontianak. pengambilan data dilakukan menggunakan *Google Form* yang dibagikan berupa *link* ke semua fakultas di Universitas Tanjungpura, dimulai dari 09 Oktober 2025 sampai 30 Oktober 2025. Dari pengumpulan data tersebut, diperoleh 517 data responden mahasiswa perempuan dan laki-laki. Pengumpulan data dilakukan dalam beberapa tahap seperti gambar di bawah ini



Gambar 1. Tahapan Pengumpulan Data

Gambar 1 menunjukkan beberapa tahapan pengumpulan data. Tahapan ini dimulai dengan pengumpulan data di Universitas Tanjungpura, dilanjutkan dengan pengolahan data menggunakan algoritma C4.5 dan *Decision Tree* yang diproses pada aplikasi perangkat lunak *RapidMiner*, sehingga diperoleh hasil dan pembahasan dari model klasifikasi dan *accuracy*. Hasil penelitian ini dapat disimpulkan bahwa hobi membaca pada mahasiswa dengan berbagai genre bacaan yang menjadi pemicu minat membaca. Selain itu penelitian ini dapat menjadi acuan menarik minat pengunjung, baik mahasiswa maupun umum, terhadap layanan baca buku *online* serta perpustakaan di dalam maupun di luar lingkungan universitas.

2.2. Pembersihan Data

Data yang didapatkan tidak semua digunakan. Oleh karena itu, dibutuhkan pemilihan data sesuai dengan kebutuhan penelitian. Beberapa data yang diambil terdiri atas 3 atribut yaitu:

- Hobi Membaca sebagai *label* dengan 3 pilihan: Sangat Minat, Netral, Tidak Minat.
- Genre dengan 11 Pilihan yaitu Biografi, Fantasi, Fiksi Ilmiah, Horror, Misteri, Pengembangan Diri/Motivasi, Romansa, Sains, Sejarah, Thriller, dan Lainnya.
- Jenis Kelamin terdiri 2 pilihan: Laki-laki dan Perempuan.

2.3. Proses Data Mining

Data Mining adalah proses dalam menganalisa sebuah data yang besar menggunakan sebuah berbagai teknik salah satunya *machine learning* dimana akan mendapat informasi mengenai dataset yang di uji [7]. Secara makna bahwa data mining berasal dari dua kata, yaitu data berarti kumpulan fakta yang terekam dan mining berarti proses penggalian yang dapat disimpulkan bahwa data mining adalah sesuatu kumpulan fakta informasi yang akan kita gali dari sebuah data yang belum memiliki makna [8]. Data mining sendiri adalah tahap dalam suatu proses *Knowledge Discovery in Databases* (KDD) yang dapat melakukan klasifikasi, prediksi, estimasi dan mendapatkan banyak informasi berharga dalam jumlah data yang besar [9]. Metode C4.5 adalah metode yang menghasilkan pohon keputusan yang bertujuan melihat keakuratan dan juga metode ini banyak digunakan karena algoritma yang mudah dipahami [12]. Metode C4.5 adalah algoritma dengan kegunaannya juga menggunakan informasi *entropy*, *gain* dan *gain ratio* dalam memilih sebuah atribut [13].

Klasifikasi adalah sebuah algoritma yang dibutuhkan dalam mengklasifikasi sebuah data dengan kategori tertentu contohnya *decision tree*, *random forest*, dan *support vector machines (SVM)* [10]. Dalam pengertian lain klasifikasi juga adalah proses menemukan model dengan membedakan kelas yang bertujuan memprediksi kelas dari suatu objek yang belum di ketahui kelasnya [11].

RapidMiner adalah *software* atau perangkat lunak dengan menyediakan analisis data mentah yang akan diproses dalam sebuah data yang bersih serta siap untuk digunakan [14]. Dalam pengertian lain *RapidMiner* merupakan *open source* yang telah di rancang untuk menganalisis dengan dilengkapi berbagai operator seperti input data, output hasil, visualisasi dan preprocessing data [15]. Pada tahapan ini akan dilakukan beberapa prosedur dalam uji di *RapidMiner* yaitu:

- Read Excel* (karena data terletak di *Excel*) dengan Pemilihan data semuanya *Polynomial*
- Set Role* dengan tiga kategori: *label* yaitu “Hobi Membaca”, dan *reguler* yaitu “Genre” dan “Jenis Kelamin”.
- Cross Validation* tetap *default* (tanpa di ubah)
- Spilt Data* dengan data *training* 70% dan data *testing* 30%
- Decision Tree* dengan pilihan *default* (tanpa di ubah)
- Apply model*
- Performance* dengan pilihan *default* (tanpa di ubah)

2.4. Evaluasi

Tahapan ini merupakan tahap akhir penelitian, yang akan menghasilkan nilai *accuracy* yang berfungsi sebagai gambaran dari atribut yang berperan besar dalam hobi membaca pada mahasiswa di Universitas Tanjungpura.

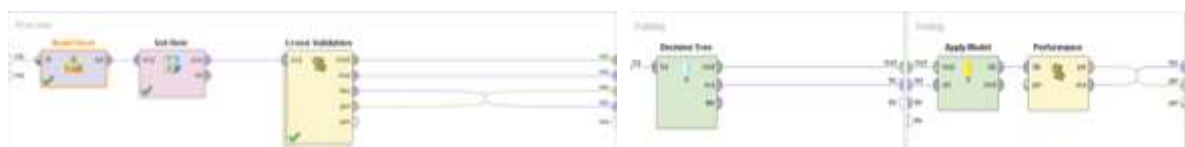
3. Hasil dan Pembahasan

3.1 Pengumpulan data

Pada tahap ini, terdapat dataset dari mahasiswa sebanyak 517 responden yang terdiri dari nama, jenis kelamin, jurusan, umur, hobi membaca, dan genre buku. Dari data tersebut, dilakukan penghapusan atau pembersihan data di *Microsoft Excel* sehingga tersisa 3 atribut pilihan, yaitu jenis kelamin, hobi membaca dan *genre*.

3.2 Proses Data Mining

Proses ini merupakan tahap awal dalam membangun klasifikasi hobi membaca dengan algoritma C4.5 dan *Decision Tree*.



Gambar 2. Pre Process

Gambar 2 menunjukkan pre process yang menggunakan perangkat lunak RapidMiner dalam implementasi algoritma C4.5. Proses dimulai dengan operator Excel untuk memanggil data dan menentukan tipe data. Selanjutnya, data diarahkan ke operator Set Role yang menentukan atribut target (label) yang wajib dalam

klasifikasi dengan algoritma C4.5. Kemudian, data diarahkan ke Cross Validation untuk membagi data training dan testing secara bergantian. Proses dilanjutkan dengan membangun model menggunakan operator Decision Tree. Model yang dihasilkan akan dilanjutkan ke operator Apply Model untuk menghasilkan hasil prediksi, yang kemudian dikirim ke operator Performance untuk menampilkan hasil perhitungan matriks seperti akurasi, presisi dan recal.

3.3 Evaluasi *Performance* dan *Decision Tree*

Berikut adalah hasil dari gambaran *Decision Tree* dapat dilihat dari Gambar 3 di bawah ini.



Gambar 3. Hasil *Decision Tree*

Berdasarkan hasil *Decision Tree* pada Gambar 3, atribut paling atas dan menepati akar pohon adalah *Genre*. Untuk *node* Misteri, Romansa dan Sains pada atribut *genre*, klasifikasi dapat didasarkan jenis kelamin.



Gambar 4. *Description Decision Tree*

Dari gambar 4, penjelasan lengkap pada *Decision Tree* sebagai berikut:

- Jika Genre Biografi tertinggi adalah kategori **Netral** maka hasil hobi membacanya yaitu Kurang Minat=2, **Netral=14**, Sangat Minat=0
- Jika Genre Fantasi tertinggi adalah kategori **Netral** maka hasil hobi membacanya yaitu Kurang Minat=4, **Netral=151**, Sangat Minat=72
- Jika Genre Fiksi Ilmiah tertinggi adalah kategori **Netral** maka hasil hobi membacanya yaitu Kurang Minat=1, **Netral=32**, Sangat Minat=15
- Jika Genre Horor tertinggi adalah kategori **Netral** maka hasil hobi membacanya yaitu Kurang Minat=3, **Netral=14**, Sangat Minat=3
- Jika Genre Lainnya tertinggi adalah kategori **Netral** maka hasil hobi membacanya yaitu Kurang Minat=3, **Netral=11**, Sangat Minat=3
- Jika Genre Misteri:

Pada Jenis Kelamin Laki-Laki tertinggi adalah kategori **Kurang Minat** maka hasilnya yaitu **Kurang Minat=2**, Netral=1, Sangat Minat=1

- Pada Jenis Kelamin Perempuan yang tertinggi adalah kategori **Netral** maka hasilnya yaitu Kurang Minat=0, **Netral=6**, Sangat Minat=2
- g. Jika Genre Pengembangan diri / Motivasi yang tertinggi adalah Kategori **Netral** maka hasil hobi membacanya yaitu Kurang Minat=8, **Netral=29**, Sangat Minat=19
 - h. Jika Genre Romansa,
Pada Jenis Kelamin Laki-Laki nilai tertinggi adalah kategori **Sangat Minat** maka hasilnya yaitu Kurang Minat=1, Netral=2, **Sangat Minat=3**
Pada Jenis Kelamin Perempuan nilai tertinggi adalah kategori **Netral** maka hasilnya yaitu Kurang Minat=5, **Netral=41**, Sangat Minat=11
 - i. Jika Genre Sains,
Pada Jenis Kelamin Laki-Laki nilai tertinggi adalah kategori **Sangat Minat** maka hasilnya yaitu Kurang Minat=1, Netral=2, **Sangat Minat=3**
Pada Jenis Kelamin Perempuan nilai tertinggi adalah kategori **Netral** maka hasilnya yaitu Kurang Minat=0, **Netral=4**, Sangat Minat=1
 - j. Jika Genre Sejarah nilai tertinggi adalah Kategori **Netral** maka hasil hobi membacanya yaitu Kurang Minat=5, **Netral=21**, Sangat Minat=8
 - k. Jika Genre Thriller nilai tertinggi adalah kategori **Netral** maka hasil hobi membacanya yaitu Kurang Minat=0, **Netral=11**, Sangat Minat=2

3.4 Performance

Pada tahapan ini hasil uji *accuracy* dalam penggunaan perangkat lunak *RapidMiner* adalah sebagai berikut:

accuracy: 64.42% ± 1.63% (micro average: 64.41%)

	true Kurang Minat	true Netral	true Sangat Minat	class precision
pred. Kurang Minat	2	6	2	20.00%
pred. Netral	31	329	139	65.93%
pred. Sangat Minat	2	4	2	25.00%
class recall	5.71%	97.05%	1.40%	

Gambar 5. Hasil *Accuracy*

Pada gambar 5, hasil *accuracy* dapat dijelaskan sebagai berikut

- a. Hasil *accuracy* sebesar 64,42% ± 1,63% dengan micro average 64,41. Hal ini dapat dipahami bahwa tingkat akurasi tergolong sedang tetapi belum optimal.
- b. Kurang Minat hanya memperkirakan 2 data yang benar dengan *class precision* sebesar 20% (rendah).
- c. Prediksi Netral memperkirakan 329 data benar (netral – netral) dengan *class precision* sebesar 65,3% (cukup baik)
- d. Prediksi Sangat Minat hanya memperkirakan 2 data yang benar dengan *class precision* sebesar 25% (rendah).
- e. Pada *class recall*, kelas Netral menunjukkan angka 97,05% sedangkan kategori kelas Kurang Minat dan Sangat Minat masing-masing (5,71%) dan (1,40%) artinya kelas Netral sangat baik dikenali, terlihat dari angka yang tinggi. Berbeda dengan Kurang Minat dan Sangat Minat yang hampir tidak mampu dideteksi, sehingga angka yang dihasilkan kecil.

4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan menggunakan algoritma C4.5 pada dataset dari 517 responden mahasiswa, dapat disimpulkan bahwa tingkat akurasi klasifikasi hobi membaca yang diperoleh adalah 64,42%. Hasil penelitian menunjukkan bahwa kategori Netral memiliki jumlah responden tertinggi (329 orang) dibandingkan kategori lainnya. Secara keseluruhan, penelitian ini dapat dikembangkan lebih lanjut dengan menggunakan metode atau algoritma lain, menambahkan atribut, memperbanyak dataset, dan menggunakan aplikasi selain *RapidMiner* untuk mendapatkan hasil perbandingan yang lebih komprehensif.

Referensi

- [1] I. P. R. Bangsawan, *Mengembangkan Minat Baca*. PT Pustaka Adhikara Mediatama, 2023. [Online]. Available: <https://books.google.co.id/books?id=hyWyEAAAQBAJ>
- [2] Mahmudi Mahmudi and Zaehol Fatah, "Analisis Pengaruh Jenis Buku Terhadap Minat Baca Mahasiswa di Perpustakaan Ibrahimiy dengan Algoritma K-Means Clustering," *JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI*, vol. 3, no. 1, pp. 53–65, Dec. 2024, doi: 10.59024/jiti.v3i1.1013.

- [3] Yeremias Bardi, Rikarda Nivani Dholo Bala Maria Ermelinda Bunga, Fransiska Putri Andriany, Anastasia Doa Nestiana Pare, Risky Afrizal Donatus Datoq, and Wahban Az Zuhaili Razak Misbah Nurrahim, "Kuranganya Minat Baca di Kalangan Mahasiswa : Studi Kasus di Universitas Muhammadiyah Maumere," *Morfologi: Jurnal Ilmu Pendidikan, Bahasa, Sastra dan Budaya*, vol. 3, no. 2, pp. 106–119, Feb. 2025, doi: 10.61132/morfologi.v3i2.1484.
- [4] M. S. Dr. Redi Panuju, *Ide Kreatif Dalam Produksi Film*. Kencana Prenada Media, 2022. [Online]. Available: <https://books.google.co.id/books?id=hSSeEAAAQBAJ>
- [5] Kamus Besar Bahasa Indonesia (KBBI), "genre." Accessed: Nov. 03, 2025. [Online]. Available: <https://kbbi.web.id/genre>
- [6] Universitas Tanjungpura, "Universitas Tanjungpura." Accessed: Nov. 05, 2025. [Online]. Available: <https://www.untan.ac.id/>
- [7] C. Carudin *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: <https://books.google.co.id/books?id=m-QGEQAAQBAJ>
- [8] R. Rahmadden, D. Wulandari, S. Aziz, S. Adrianto, R. Hidayatullah, and R. M. Sari, *Data Mining: Algoritma dan Implementasi dengan Python dan Rapidminer*. Serasi Media Teknologi, 2025. [Online]. Available: <https://books.google.co.id/books?id=udmEEQAAQBAJ>
- [9] S. Hartawan *et al.*, "Penerapan Data Mining Menggunakan Algoritma C4. 5 Untuk Memprediksi Persediaan Stok Beras," vol. 4, no. 2, 2024, doi: 10.46880/tamika.Vol 4 No 2 (Semnastik).pp114-121.
- [10] S. Syam *et al.*, *Data Mining : Teori dan Penerapannya dalam Berbagai Bidang*. PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: <https://books.google.co.id/books?id=hTAXEQAAQBAJ>
- [11] Aulia Sugarda, Saifullah, Jalaluddin, Agus Perdana Windarto, and Wendi Robiansyah, "Penerapan Metode Data Mining C4.5 dalam Penentuan Kelayakan Rehabilitas Rumah Warga," *Journal of Computing and Informatics Research*, vol. 1, no. 3, pp. 56–64, Jul. 2022, doi: 10.47065/comforch.v1i3.321.
- [12] N. Khasanah, D. Uki, E. Saputri, T. Hidayat, F. Aziz, and U. N. Mandiri, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 dengan RapidMiner: Studi Kasus Data Akademik Perguruan Tinggi XYZ," *Journal Computer Science*, vol. 4, no. 2, 2025.
- [13] N. I. S. Futri, R. N. S. Fatonah, and M. Y. H. Setyawan, *Implementasi algoritma C4.5 untuk meperediksi kapasitas penumpang pesawat menggunakan Python*. Penerbit Buku Pedia, 2023. [Online]. Available: <https://books.google.co.id/books?id=9b6-EAAAQBAJ>
- [14] R. M. Ubaidilah, Z. Anugerahwati, I. Ikko, M. Rizky, and S. Lestari, "Prediksi Kelulusan Mahasiswa Berdasarkan Data Kunjung dan Peminjaman Buku menggunakan Rapid Miner dengan Metode C.45 dan Random Forest," vol. 7, no. 2, 2023, [Online]. Available: <https://www.e-jurnal.dharmawacana.ac.id/index.php/ir>
- [15] R. M. Sari and Y. A. M, *Klasifikasi Data Mining*. Serasi Media Teknologi, 2024. [Online]. Available: <https://books.google.co.id/books?id=xTwXEQAAQBAJ>