

# IMPLEMENTASI DATA MINING ALGORITMA C4.5 UNTUK MENENTUKAN MINAT SISWA MEMBACA PADA PERPUSTAKAAN SMP NEGERI 1 LASEM KABUPATEN REMBANG

**Maulana Rezza Alvian, Acun Kardianawati**

Universitas Dian Nuswantoro , Fasilkom Udinus, Jurusan Sistem Informasi  
Jalan 5-11 Nakula 1, Semarang, 50131, (+6224) 3520165/ (+6224) 3569684  
[Marealalvian1994@gmail.com](mailto:Marealalvian1994@gmail.com)

---

## **Abstrak**

*Di Era modern ini banyak sekali penggunaan media internet yang berdampak semakin sedikitnya minat seseorang dalam membaca buku. Pada SMP Negeri 1 Lasem memiliki masalah peminatan siswa dalam membaca buku yang masih belum diketahui seberapa banyak siswa minat meminjam buku di perpustakaan SMP Negeri 1 Lasem. Oleh karena itu untuk mengetahui seberapa besar minat siswa meminjam buku di perpustakaan SMP Negeri 1 Lasem perlu melakukan pencarian informasi dari data peminjaman buku perpustakaan dengan menggunakan teknik klasifikasi data mining dalam meningkatkan nilai ujian nasional dan ujian sekolah yang lebih baik. Dengan teknik klasifikasi data mining algoritma C4.5 dengan data setnya adalah peminjaman buku perpustakaan. Hasil yang diperoleh adalah node dan posisi terbaik, dengan posisi terbaik ditentukan pada atribut jangka pinjam karena bersifat kontinu yang memiliki gain tertinggi. Informasi yang dihasilkan adalah posisi terbaik dari atribut jangka pinjam , serta perhitungan entropy mencari node akar dengan gain tertinggi diambil dari atribut jangka pinjam, jenis kelamin, kelas, jenis buku. Sehingga proses tersebut diulang ulang sampai pada posisi terbaik jangka waktu dan node untuk mendapatkan hasil akhir minat atau tidak minat sebagai pembentuk pohon keputusan. Dengan diketahuinya siswa yang paling banyak meminjam buku di perpustakaan akan dapat memprediksikan minat baca siswa baru.*

**Kata Kunci:** Data Mining, Teknik Klasifikasi, Algoritma C4.5, Peminjaman buku, Perpustakaan, pohon keputusan

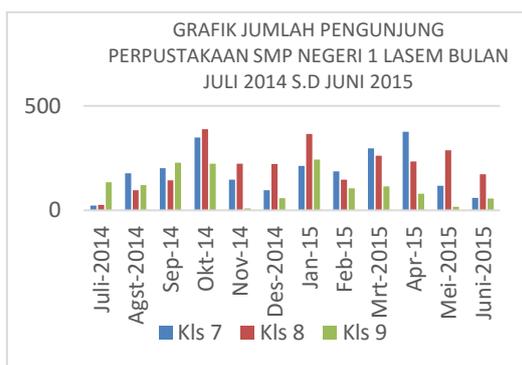
## **Abstract**

*In the modern era is a lot of use of internet media which affects fewer person's interest in reading books. At SMPN 1 Lasem specialization students have problems in reading the book is not yet known how many students interest to borrow books at the library of SMP Negeri 1 Lasem. Therefore, to find out how much interest students in the library to borrow books SMPN 1 Lasem need to search information from the data library book lending by using techniques classification of data mining to improve the national test scores and better school exam. With the technique classification of data mining algorithms C4.5 to the data disk set is a book lending library. The results obtained are the nodes and the best position, the best position is determined on the attributes and loan term due to be continuous which has the highest gain. The resulting information is the best position of attribute-term loans, as well as the calculation of entropy looking for the root node with the highest gain is taken of attribute-term loans, gender, class, type of book. So the process is repeated again until the best position and node time period to get the final result as interest or no interest in forming a decision tree. By knowing the students most borrowed books in the library will be able to predict the new students' interest in reading.*

**Keywords:** Data Mining, Techniques Classification, Algorithm C4.5, book lending, library, decision trees

## 1. PENDAHULUAN

Di era modern ini banyak sekali penggunaan media internet yang berdampak semakin sedikitnya seseorang dalam membaca buku. Padahal buku adalah ilmu pengetahuan yang sangat banyak dan eksklusif untuk ditelusuri daripada media internet hanya menyediakan sedikit referensi. Banyak perpustakaan-perpustakaan di Indonesia termasuk di dalam sekolah maupun jenjang yang lebih tinggi seperti perguruan tinggi. Perpustakaan sekolah didirikan untuk mengajak para siswa siswi untuk lebih mendalami dari setiap pelajaran yang diberikan oleh guru dikelas. Selain itu dengan sering membaca maupun meminjam buku di perpustakaan, kita dapat mengerjakan tugas sekolah dengan benar. Pada Perpustakaan SMP NEGERI 1 Lasem, Kabupaten Rembang memiliki tingkat kunjungan digambarkan dengan grafik sebagai berikut:



**Gambar 1.** Grafik Pengunjung Perpustakaan

Dari pernyataan pada **Gambar 1.** Grafik Pengunjung Perpustakaan dapat diketahui banyaknya siswa-siswi berkunjung ke perpustakaan dari bulan Juli 2014-Juni 2015 mengalami penurunan dari kelas 7 maupun kelas 8. Dan puncak tertinggi pada kelas 7 adalah pada bulan April 2015 dan kelas 8 adalah bulan Oktober 2014, khusus untuk kelas 9 memang jumlahnya

sedikit karena menghadapi ujian nasional. Akan tetapi banyak sedikitnya jumlah kunjungan masih belum dapat menentukan banyaknya buku yang dipinjam karena mungkin penyebab utamanya malas mencari buku yang jumlahnya banyak dan akhirnya hanya sekedar berkunjung. Padahal membaca buku dengan cara meminjam buku berpengaruh pada nilai ujian sekolah maupun ujian nasional dan berdampak pada sekolah SMP Negeri 1 Lasem sebagai sekolah terbaik sekabupaten Rembang.

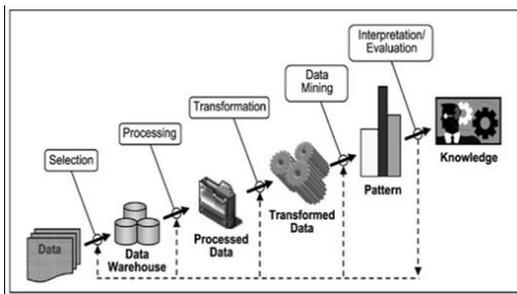
Maka dari itu diperlukannya suatu penelitian untuk mengetahui seberapa besar minat membaca jika dilihat berdasarkan seberapa banyak meminjam buku di perpustakaan bukan dari data berkunjung. Dengan data mining algoritma C4.5 dapat membangun sebuah aplikasi dengan perhitungan algoritma C4.5 berbasis web sehingga nantinya jika terdapat kelas yang memiliki data sedikit dalam meminjam buku perpustakaan akan disosialisasikan betapa pentingnya membaca buku untuk keperluan ujian sekolah maupun ujian nasional nanti dimasa mendatang dan juga untuk memprediksi peminatan membaca oleh siswa baru. Dalam membangun aplikasi data mining C4.5 tersebut akan berguna dalam mengelolah data peminjaman buku dari tahun-tahun selanjutnya

Data mining sendiri adalah proses pengolahan atau menggali data dari suatu kumpulan database yang sangat besar dan masih belum diketahui ilmu ilmu yang terkandung didalamnya dengan pola teknik statistic, matematika, kecerdasan buatan, dan machine learning [1]. Data mining mempunyai banyak metode, Salah satunya C4.5. C4.5 merupakan metode yang digunakan pada data mining yang

berguna untuk membuat pohon keputusan (Decision Tree), sedangkan pohon keputusan adalah pengklasifikasian data yang dibentuk dengan node [2].

## 2. METODE PENELITIAN

Metode yang diusulkan metode C4.5 dalam mencari dan menemukan frekuensi minat siswa dalam setiap kelas ke perpustakaan tertinggi pada data peminjaman buku di SMP Negeri 1 Lasem Kabupaten Rembang.



**Gambar 2.** Process Knowledge Discovery in Database

Langkah analisa yang digunakan untuk mengolah dataset berdasarkan metode data mining menggunakan proses knowledge discovery in databases (KDD), yang terdiri dari [3]

:

### 1. Data selection

Kumpulan kumpulan data yang diseleksi adalah data peminjaman buku perpustakaan 2014-2015.

### 2. Pre-Processing

Pada proses ini melakukan pembersihan data dari data-data mentah dengan mengambil attribute jenis kelamin, kelas, lama meminjam, range kode buku.

### 3. Transformation

Dalam tahap transformasi mengubah format data yaitu data peminjaman buku yang kedalam mircosoft excel dengan

bantuan kompter agar data tersebut dapat terintegrasi dengan aplikasi yang dibuat.

### 4. Data Mining

Dengan proses pengestrakan data mining ini mendapatkan pola data peminjaman yang sudah dirubah dengan menerapkan algoritma C4.5.

### 5. Interpretation (Evaluation)

Pola yang sudah telah diidentifikasi nantinya diolah untuk diterjemahkan kedalam bahasa yang mudah dipahami agar kelak dapat membantu memperoleh keputusan strategi bisnis.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Data Training yang Digunakan

Data yang digunakan untuk perhitungan C4.5 adalah peminjaman buku perpustakaan pada SMP Negeri 1 Lasem. Pada data ini menggunakan data sampel berjumlah 50 dari data keseluruhan yaitu 1500. Berikut ini data sampel 50 dari total data yang akan dilakukan proses perhitungan c4.5 .

**Tabel 1:** Data Training

No	J.K el	Kls	J.Buku	J.Pinjam	Peminta n
1	P	VII I C	T.Map el	5	tidak minat
2	P	VII I C	T.Map el	5	tidak minat
3	P	VII I C	T.Map el	5	tidak minat
4	P	VII I C	T.Map el	10	minat
5	P	VII I C	T.Map el	10	minat
6	P	VII I E	T.Map el	8	minat
7	P	IX A	Mapel	11	tidak minat
8	P	IX A	T.Map el	11	tidak minat

9	P	IX A	Mapel	11	tidak minat
10	P	VII II	Mapel	9	minat
11	P	VII II	T.Map el	9	minat
12	P	VII IF	T.Map el	8	tidak minat
13	P	VII IA	T.Map el	10	minat
14	P	VII IA	T.Map el	2	minat
15	P	VII IH	T.Map el	3	minat
16	P	VII IH	T.Map el	6	minat
17	P	VII H	T.Map el	4	minat
18	P	VII H	T.Map el	4	minat
19	P	VII H	T.Map el	4	minat
20	P	VII A	T.Map el	4	minat
21	P	VII A	T.Map el	4	minat
22	P	VII A	T.Map el	4	tidak minat
23	P	VII A	T.Map el	4	minat
24	P	VII IH	T.Map el	3	minat
25	P	VII IG	T.Map el	6	tidak minat
26	P	VII IG	T.Map el	6	tidak minat
27	P	VII ID	T.Map el	2	tidak minat
28	P	VII ID	T.Map el	2	tidak minat
29	P	VII IE	T.Map el	2	minat
30	L	VII IE	T.Map el	2	minat
31	L	VII IE	T.Map el	6	minat
32	P	VII IE	T.Map el	6	tidak minat
33	P	VII G	T.Map el	5	tidak minat
34	P	VII ID	T.Map el	5	tidak minat
35	P	VII ID	T.Map el	5	tidak minat
36	P	VII IE	T.Map el	5	minat
37	P	VII	T.Map	5	minat

		IE	el		
38	L	VII IF	T.Map el	5	minat
39	L	VII IF	T.Map el	5	minat
40	P	VII IF	T.Map el	5	minat
41	L	IX B	T.Map el	3	minat
42	L	IX B	Mapel	3	minat
43	P	IX E	Mapel	9	tidak minat
44	P	IX B	Mapel	2	minat
45	P	IX H	Mapel	5	minat
46	P	VII IB	T.Map el	4	tidak minat
47	L	VII IE	T.Map el	8	minat
48	L	VII IE	Mapel	6	minat
49	L	VII IE	T.Map el	11	minat
50	P	VII IH	Mapel	8	minat

### 3.2 Proses Perhitungan

Pada proses perhitungan c4.5, pertama kali yang harus dilakukan adalah menghitung entropy dari semua data. Berikut ini perhitungannya:

$$\begin{aligned}
 E(\text{Semua}) &= - ((p(\text{minat/semua}) \times \log_2 p(\text{minat/semua})) + (p(\text{tidak minat/semua}) \times \log_2 p(\text{tidak lancar/semua}))) \\
 &= - (((32/50) \times \log_2 (32/50)) + ((18/50) \times \log_2 (18/50))) \\
 &= 0.9427 (1)
 \end{aligned}$$

Langkah selanjutnya menentukan posisi terbaik. Untuk menentukan posisi terbaik dibutuhkan data yang bersifat kontinu atau data yang tidak ada batas jumlahnya, dalam data perpustakaan attribute yang bersifat kontinu adalah jangka pinjam, lalu penentuan posisi terbaik didapatkan dari gain yang paling

tinggi sedangkan entropy jangka pinjam didapatkan dengan menggunakan rumus:

$$E(\text{jangka pinjam}) = - ((p(\text{minat/jangka pinjam}) \times \log_2 p(\text{minat/jangka pinjam})) + (p(\text{tidak minat/}) \times \log_2 p(\text{tidak minat/jangka pinjam})) \quad (2)$$

Berikut ini hasil perhitungan entropy dan gain untuk mencari posisi terbaik dari atribut plafond.

**Tabel 2:** Pencarian Posisi Terbaik

No	Jangka Pinjam	Minat	Tidak Minat	Entropy	Gain
1	<=2	4	2	0.9183	0.0003
	>2	28	16	0.9457	
2	<=3	8	2	0.7219	0.0215
	>3	24	16	0.971	
3	<=4	14	4	0.7642	0.0348
	>4	18	14	0.9887	
4	<=5	20	10	0.9183	0.0033
	>5	12	8	0.971	
5	<=6	23	13	0.9436	0
	>6	9	5	0.9403	
6	<=8	26	14	0.9341	0.0012
	>8	6	4	0.971	
7	<=9	28	15	0.933	0.0024
	>9	4	3	0.9852	
8	<=10	31	15	0.9109	0.0398
	>10	1	3	0.8113	
9	<=11	32	18	0.9427	0
	>11	0	0	0	

Dari **Tabel 2:** Pencarian Posisi Terbaik sehingga diperoleh posisi terbaik dalam menentukan gain terbesar pada attribute jangka pinjam yaitu <= 10 dan > 10. Selanjutnya melakukan perhitungan entropy dan gain yang nantinya untuk memperoleh node akar. Dibawah ini tabel perhitungan entropy dan gain.

**Tabel 3:** Pencarian Node

Kriteria	Jumlah	Minat	Tidak Minat	Entropy	Gain

Entropy Total					
	50	32	18	0.9427	
Jangka Pinjam					
<= 10	46	31	15	0.9109	0.0398
> 10	4	1	3	0.8113	
Jenis Kelamin					
L	9	9	0	0	0.1316
P	41	23	18	0.9892	
Kelas					
IX A	3	0	3	0	0.622
IX B	3	3	0	0	
IX E	1	0	1	0	
IX H	1	1	0	0	
VII A	4	3	1	0.8113	
VII G	1	0	1	0	
VII H	3	3	0	0	
VIII A	2	2	0	0	
VIII B	1	0	1	0	
VIII C	5	2	3	0.971	
VIII D	4	0	4	0	
VIII E	10	9	1	0.469	
VIII F	4	3	1	0.8113	
Jenis Buku					
mata pelajaran	9	6	3	0.9183	0.0005
tidak mata pelajaran	41	26	15	0.9474	

Dari **Tabel 3:** Pencarian Node didapatkan suatu gain tertinggi yang terletak pada attribute kelas, Nantinya attribute kelas sebesar 0.622 tersebut akan dijadikan node akar dimana attribute kelas tersebut memiliki 3 cabang antara lain minat, tidak minat, dan belum diketahui. Sehingga diperlukan melakukan perhitungan untuk mendapatkan nilai gain yang masih belum diketahui hasilnya.

Proses selanjutnya mencari node internal, caranya sama seperti tabel 2 Pencarian Posisi Terbaik dan tabel 3 Pencarian Node. Proses ini dilakukan



ini digunakan oleh karyawan Perpustakaan.

### 3.3.5 Halaman Rules



Gambar 8. Rules

Pada Gambar 8. Rules menjelaskan tentang menu berupa hasil data dari data yang sudah diolah pada menu input data training. Hasil dari pengolahan tersebut berbentuk rules. Rules atau pohon keputusan tersebut sangat panjang maka dari itu dibuat berbentuk kalimat. Halaman ini digunakan oleh karyawan Perpustakaan.

### 3.3.6 Halaman Akurasi



Gambar 8. Akurasi

Pada Gambar 8. Akurasi menjelaskan tentang berupa hasil dari sebuah proses atau pengujian rules tadi. Pengujian tersebut untuk membuat suatu akurasi. Halaman ini digunakan oleh karyawan Perpustakaan. Akurasi tersebut dapat diperoleh dari hasil rules dicocokkan

dengan data actual atau data asli dari data training tersebut.

### 3.3.7 Halaman Memprediksi Siswa Baru



Gambar 9. Memprediksi Siswa Baru

Pada Gambar 9. Memprediksi Siswa Baru menjelaskan tentang menu yang berguna untuk memprediksi peminatan siswa baru dalam meminjam buku perpustakaan dari hasil rules data training tadi yang di dalamnya mengisikan jenis kelamin, kelas, jenis buku, jangka pinjam. Halaman ini digunakan oleh karyawan perpustakaan.

### 3.3.8 Halaman Login Admin



Gambar 10. Login Admin

Pada Gambar 10. Login Admin menjelaskan tentang menu yang digunakan khusus untuk admin untuk mengolah system tersebut. Dalam halaman tersebut admin harus mengisikan username dan passwordnya.

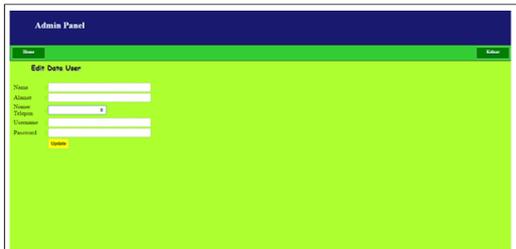
### 3.3.9 Halaman Utaman Admin



Gambar 11. Halaman Utama Admin

Pada **Gambar 11**. Halaman Utama Admin menjelaskan tentang menu yang menampilkan status dari setiap admin dan admin yang melakukan login. Dalam halaman tersebut menampilkan informasi admin yang terdiri dari nama (nama admin), alamat, nomor telepon, username, password, dan opsi hapus serta edit user admin. Untuk tombol edit user admin akan dialihkan ke halaman baru yaitu halaman edit user admin.

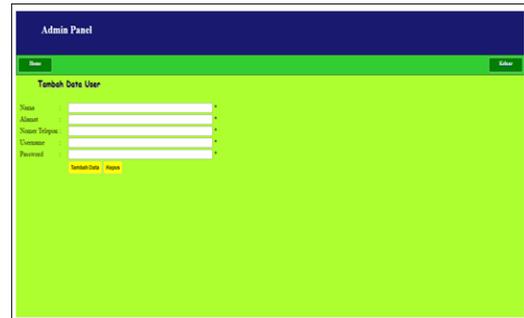
### 3.3.10 Edit User Admin



Gambar 12. Edit User Admin

Pada **Gambar 12**. Edit User Admin menjelaskan tentang halaman yang digunakan untuk mengedit dengan merubah nama, alamat, nomor telepon, username, password dari data diri admin yang sebelumnya telah dibuat. Dan jika sudah mengisikan apa saja yang ingin diedit dari 5 atribute pengisian diatas selanjutnya klik tombol update dan nantinya akan otomatis tergantikan admin lama dengan admin baru yang sudah diedit.

### 3.3.11 Tambah User Admin



Gambar 13. Tambah User Admin

Pada **Gambar 13**. Edit User Admin menjelaskan tentang halaman yang digunakan untuk mengedit dengan merubah nama, alamat, nomor telepon, username, password dari data diri admin yang sebelumnya telah dibuat.

## 4. KESIMPULAN DAN SARAN

### 4.1 Kesimpulan

Dari hasil analisis data dengan memakai teknik data mining yang diterapkan pada sistem yang sudah dijelaskan maka dapat diambil kesimpulan yaitu

1. Dengan menggunakan algoritma C4.5 dengan penentuan entropy, gain, node akar dan internal dapat membentuk suatu pohon keputusan atau decision tree yang hasilnya dapat mengetahui seberapa besar minat siswa dalam membaca buku dilihat dari data peminjaman buku dan juga untuk memprediksi siswa baru yang meminjam buku dipertustakaan yang jika diketahui siswa tidak minat membaca buku akan disosialisasikan demi meningkatkan nilai ujian nasional dan ujian sekolah dan penambahan buku-buku baru baik buku matapelajaran maupun buku tidak matapelajaran bagi siswa yang sudah tinggi minat membacanya supaya mereka tidak bosan.

2. Dan pada algoritma C4.5 telah dibuktikan tingkat akurasi dari prediksi dengan menggunakan confusion matrik sebanyak 71,5% dari semua data berjumlah 1520 dengan diambil data training sebesar 70% dan data testing sebesar 30%. Attribute yang mempengaruhi peminatan siswa dalam membaca buku dalam algoritma C4.5 tersebut adalah Jangka pinjam, Jenis Kelamin, Kelas, Jenis Buku..

#### 4.2 Saran

1. Dalam penelitian ini memiliki keterbatasan dalam pengetahuan sehingga perlu dilakukan perbaikan antara lain:

a. Sangat diperlukan lagi penelitian lebih lanjut agar penelitian saat ini dapat dibandingkan penelitian-penelitian selanjutnya untuk mendapatkan hasil yang lebih baik.

b. Diperlukan objek yang beda dengan data sama yaitu data peminjaman buku perpustakaan untuk dijadikan perbandingan objek lama dengan objek yang baru.

2. Untuk pihak-pihak yang berhubungan dengan sekolah dalam meningkatkan minat siswa membaca buku hendaknya:

a. Bagi pihak guru seharusnya siswa yang kurang atau tidak minat membaca buku memberlakukan siswa untuk diberikan suatu tugas tambahan seperti pekerjaan rumah atau PR yang nantinya jawaban tugas tersebut diperoleh dari buku buku perpustakaan sehingga secara tidak langsung siswa mau meminjam dan membaca buku.

b. Bagi pihak Perpustakaan seharusnya untuk siswa yang sudah minat membaca buku diberikan tambahan buku buku

baru supaya lebih menambah wawasan pengetahuan dan bagi yang belum minat membaca buku memberikan tambahan jangka pinjam dan buku-buku selain matapelajaran seperti komik bergambar tentang kepahlawanan Indonesia dan lain-lain.

#### DAFTAR PUSTAKA

[1] Budanis Dwi Meilani Achmad dan Fauzi Slamet, Mei 2012. *Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree*, vol. 16, No.1.

[2] Ferry Kristian Sukendro Wimmie Handiwidjojo, Agustus 2015. *Sistem Informasi Penjualan Sepeda Motor Bekas Menggunakan Algoritma C4.5*, vol. 01, No. 02.

[3] Dicky Nofriansyah, S.Kom, M.Kom., 2015. *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Yogyakarta: Deepublish.