

Application of the C.45 Data Mining Algorithm to Predict Reading Interest in the Rejang Lebong Regional Library

Penerapan Data Mining Algoritma C.45 Untuk Memprediksi Minat Baca Pada Perpustakaan Daerah Rejang Lebong

Wiwi Sukanti^{*1}, Murlena², Putri Milanda Binamus³, Nurfitri Andayani⁴
^{1,2,3,4} Program Studi Ilmu Komputer, Fakultas Teknik, Universitas Pat Petulai Rejang Lebong, Indonesia

*Email korespondensi: wiwiksukanti1@gmail.com

Abstract

Data Mining is a computer science method commonly used in the knowledge search process. This method is often found in the fields of machine learning and statistics. The c4.5 algorithm is a classification technique in Machine Learning which is used in the data mining process by forming a decision tree. The data used in this research is primary data originating from the source where the research was conducted. This research was created in order to determine the classification of reading interest and the level of accuracy using the C4.5 algorithm data mining method to produce a model in the form of a decision tree to predict the types of books that are of interest based on the type of book, gender and occupation of the reader at the Rejang Lebong Regional Library. From the Decision Tree that is formed, the main factors that influence interest reading, namely work and followed by gender and depending on the type of book. From the results of the Entropy and gain calculations, it was found that the most popular type of book was the Encyclopedia and Reference book type. From the results of the gain and Entropy calculations, the final decision tree and the final validation accuracy results were found. with a value of 44.48%

Keywords: Data Mining, C4.5 Algorithm, Regency Regional Library, Prediction of Reading Interest.

Abstrak

Data Mining adalah metode ilmu komputer yang biasa digunakan dalam proses pencarian Knowledge. metode ini banyak ditemukan pada bidang machine learning dan statistika. Algoritma c4.5 yaitu teknik klasifikasi pada Machine Learning yang digunakan pada proses data mining dengan membentuk sebuah pohon keputusan (Decision Tree). Data yang digunakan dalam penelitian ini yaitu data primer berasal dari sumber lokasi penelitian dilakukan. Penelitian ini dibuat agar dapat menentukan klasifikasi minat baca serta tingkat keakuratan menggunakan metode data mining algoritma C4.5 sehingga dihasilkan sebuah model berbentuk decision tree guna memprediksi jenis buku yang diminati berdasarkan jenis buku, jenis kelamin dan pekerjaan pembaca di Perpustakaan Daerah Rejang Lebong. Dari Decision Tree yang terbentuk, faktor utama yang mempengaruhi minat baca yaitu pekerjaan dan disusul jenis kelamin dan tergantung pada jenis buku. Dari hasil perhitungan Entropy dan gain maka ditemukan hasil terbanyak jenis buku yang diminati adalah jenis buku Ensiklopedia Dan Referensi, Dari hasil perhitungan gain dan Entropy maka ditemukan pohon keputusan akhir serta Hasil akhir akurasi validation dengan nilai sebesar 44,48%.

Kata Kunci: Data Mining, Algoritma C4.5, Perpustakaan, Prediksi Minat Baca.

PENDAHULUAN

Dengan perkembangan teknologi komunikasi dan informasi yang sangat pesat, semua informasi akan didapat melalui membaca, mendengarkan dan melihat. Saat ini informasi mudah dicari dimana saja dan kapan saja biasanya berbentuk cetak, digital dan terekam.

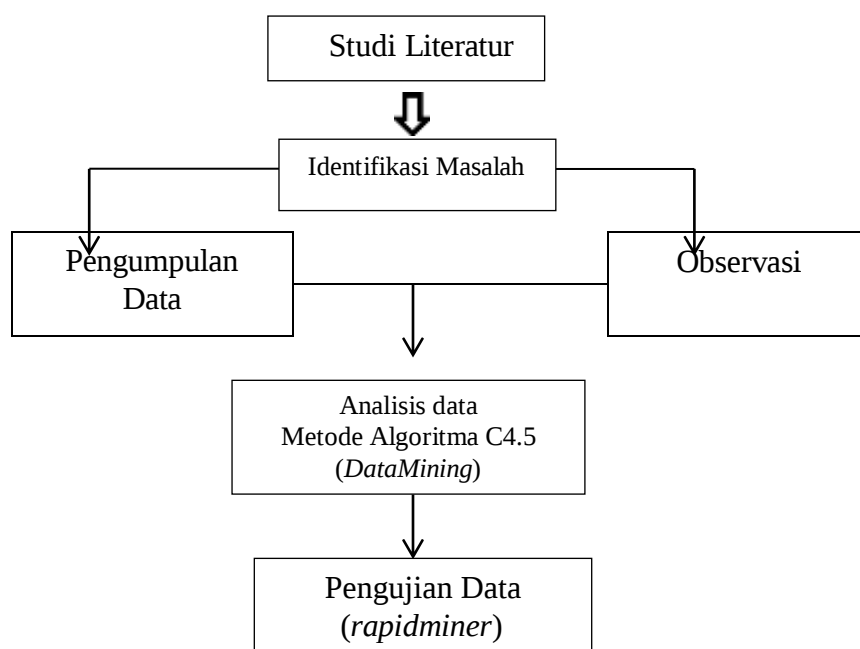
adanya perpustakaan menjadi suatu hal yang sangat penting dalam mendapatkan informasi. Salah satu jenis perpustakaan yang penting adalah perpustakaan daerah yang berkerja sama dengan pemerintah kabupaten untuk melestarikan adat dan budaya daerah. Perpustakaan memiliki peran yang begitu penting dalam penyebaran informasi hal ini dikarenakan didalam sebuah perpustakaan terdapat banayak sekali buku dan setiap bukunya itu memiliki banyak berbagai informasi yang sangat berguna bagi pembaca. Karena dianggap sebagai sumber informasi maka perpustakaan juga sangat berperan dalam menciptakan masyarakat yang literat, yaitu masyarakat yang giat akan informasi.

Data mining adalah metode dalam ilmu komputer yang biasa digunakan dalam proses pencarian knowledge. Tahapan di dalamnya berguna untuk mencari pola-pola tertentu dari data yang ada pada database. Biasanya, metode ini banyak ditemukan pada bidang machine learning dan statistika. Algoritma c4.5 merupakan salah satu teknik klasifikasi pada machine learning yang digunakan pada proses data mining dengan membentuk sebuah pohon keputusan (decision tree) yang direpresentasikan dalam bentuk aturan (rule) menghasilkan sebuah pengetahuan yang baru. Oleh karena itu penelitian ini, bertujuan untuk membantu memecahkan masalah yang dimiliki Perpustakaan Daerah Kabupaten Rejang Lebong untuk dapat memprediksi minat baca dalam data peminjaman buku perpustakaan menggunakan metode data mining algoritma C4.5 dengan cara mengidentifikasi variabel yang tersedia dalam data peminjaman buku.

Hasil dari penelitian ini nantinya dapat membantu dalam mengambil keputusan bagi karyawan petugas perpustakaan saat menentukan penambahan jumlah buku serta menyelenggarakan kegiatan yang menarik minat pengunjung. maka tujuan dari penelitian ini yaitu untuk memprediksi minat baca dengan data mining menggunakan algoritma c4.5 pada perpustakaan daerah. serta untuk mengetahui akurasi algoritma c4.5 melakukan data mining dalam memprediksi minat baca pada perpustakaan daerah rejang lebong.

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode penelitian kuantitatif, untu k bahan analisis peneliti menggunakan data daftar peminjaman buku perpustakaan daerah kabupaten rejang lebong. Adapun kerangka pikir pada penelitian ini yaitu:



Gambar 1 kerangka berikir

1. Stuidi Literaituir, aidailaih suiaitui yang saingait perlui diguinaikain dailaim melaiikuikain penelitiain mempelaijairi teori dain pengetaihuiaian mengenai semuaia yang berkaitain dengain penelitiain.
2. Identifikaisi maisailaih, yaitui dengain cairai menggaili maisailaih yang ditemuikain paidai objek yang diteliti untuik memberikain penjelaisain yang teruikuir. Guinai mencairi solusi terbaik terkait dengain permaisailaihain yang aikain diteliti.
3. Observaisi, yaitui dengain cairai mengaimaiti maisailaih laingsuing kelokaisi.
4. Setelah taihaip observaisi selesai selainjuitnyai yang dilaikuikain aidailaih proses menguimpuilkain daitai. Daitai yang dibuituihkain yaitui daitai penguinjuing dain daitai peminjaimain buikui Perpuistaikaiaian Daieraih Kaibuipaiten Rejaing Lebong.
5. AInailisis daitai dengain membuiait daitai mining yang diguinaikain untuik memilih teknik dain ailgoritmai C4.5 untuik menemuikain polai aitaui memprediksi daitai penguinjuing minait baicai paidai Perpuistaikaiaian Daieraih Kaibuipaiten Rejaing Lebong.
6. Pengujian, Ditaihaip ini kitai daipait menguiji haisil prediksi dairi daitai yang telaih diolaih daitai mining ailgoritmai C4.5 mengguinaikain Sofwaire Raipidminer untuik memvailidaisikain keaikuiraitain daitai paidai kinerjaai model yang telaih buiait.

HASIL DAN PEMBAHASAN

Dalam penelitian ini data yang digunakan sebagai bahan penelitian adalah data pinjaman buku sebanyak 309 data yang diambil pada bulan April 2024 lalu di Perpuustakaan Daerah Kabupaten Lebong dan dipilih atributnya sesuai dengan yang dibutuhkan, seperti jenis kelamin, jenis buku, dan pekerjaan.

Setelah dilakukan beberapa proses selection dan cleaning maka total data pinjaman buku tersebut menjadi 54 data berdasarkan keterangan Minat = 28 dan Tidak = 26. Kemudian mengelompokkan dan menghitung semua total atribut dan disimpan dalam kolom "jumlah" dan total "atribut". Berdasarkan data terbaik data set pinjaman yang terdapat dalam kolom "Minat" dan Tidak ". seperti tabel data set dibawah ini:

Tabel 1 Data Set

No	JenisBuku	JenisKelamin	Pekerjaan
1	Agama	Laki-Laki	Pelajar
2	Agama	Laki-Laki	Mahasiswa
3	Agama	Perempuan	Guru
4	Agama	Perempuan	Pelajar
5	Agama	Perempuan	Mahasiswa
6	Agama	Perempuan	Umum
7	EnsiklopediaMini	Laki-Laki	Pelajar
8	EnsiklopediaMini	Perempuan	Pelajar
9	Kamus	Laki-Laki	Pelajar
10	Kamus	Perempuan	Pelajar
11	Kamus	Perempuan	Mahasiswa
12	Kamus	Perempuan	Umum
13	Motivasi	Laki-Laki	Pelajar
14	Motivasi	Laki-Laki	Umum
15	Motivasi	Perempuan	Umum
16	Motivasi	Perempuan	Guru
17	Motivasi	Perempuan	Pelajar
18	Motivasi	Perempuan	Mahasiswa
19	Ensiklopedia	Laki-Laki	Pelajar
20	Ensiklopedia	Laki-Laki	Guru
21	Ensiklopedia	Laki-Laki	Mahasiswa
22	Ensiklopedia	Laki-Laki	Umum
23	Ensiklopedia	Laki-Laki	TNI
24	Ensiklopedia	Perempuan	Guru
25	Ensiklopedia	Perempuan	Pelajar
26	Ensiklopedia	Perempuan	Umum
27	Ensiklopedia	Perempuan	Mahasiswa
28	Ensiklopedia	Perempuan	IRT
29	Cerpen	Laki-Laki	Pelajar

30	Novel	Laki-Laki	Mahasiswa
31	Novel	Laki-Laki	Umum
32	Novel	Laki-Laki	Pelajar
33	Novel	Perempuan	Guru
34	Novel	Perempuan	Mahasiswa
35	Novel	Perempuan	Umum
36	Novel	Perempuan	Pelajar
37	Refensi	Laiki-Laiki	Dosen
38	Refensi	Laiki-Laiki	Umuim
39	Refensi	Laiki-Laiki	Mahasiswa
40	Refensi	Laiki-Laiki	LSM
41	Refensi	Laiki-Laiki	Guirui
42	Refensi	Laiki-Laiki	Kepalasekolaih
43	Refensi	Laiki-Laiki	Pelajar
44	Refensi	Perempuan	Guirui
45	Refensi	Perempuan	Kepalasekolaih
46	Refensi	Perempuan	Pelajar
47	Refensi	Perempuan	Umuim
48	Sastra	Laiki-Laiki	Mahasiswa
49	Sastra	Perempuan	Pelajar
50	Sosial	Laiki-Laiki	Mahasiswa
51	Sosial	Laiki-Laiki	Pelajar
52	Sosial	Perempuan	Pelajar
53	Sosial	Perempuan	Guirui
54	Kesehatan	Perempuan	Pelajar

Selanjutnya yaitu menghitung total data yang terdiri atas 54 total data berdasarkan keterangan Minat= 28 dan Tidak= 26. Kemudian mengelompokkan dan menghitung semua total atribut dan disimpan dalam kolom “jumlah” dan total “atribut”. Berdasarkan dari banyak data set peminjaman yang terdapat dalam kolom “Minat” dan Tidak”.

Tabel 2 total data, atribut dan kelas

Kategori	Atribut	Jumlah	Minat	Tidak
Total		54	28	26
JenisBuku				
	Agama	6	3	3
	EnsiklopediaMini	2	2	0
	Kamus	4	0	4

	Motivasi	6	3	3
	Ensiklopedia	10	6	4
	Cerpen	1	0	1
	Novel	7	6	1
	Referensi	11	5	6
	Sastra	2	1	1
	Sosial	4	1	3
	Kesehatan	1	0	1
JenisKelamin				
	Laki-Laki	25	10	15
	Perempuan	29	17	12
Pekerjaian				
	Pelajar	19	13	6
	Wiru	11	3	6
	Umuim	13	4	9
	Mahasiswa	11	7	4

A. Proses Pembentukan Decision Tree Menggunakan Algoritma C4.5

Menghitung nilai Entropy total, nilai Entropy dari nilai setiap atribut kemudian dihitung nilai gain dari setiap atribut. Nilai Entropy total dan Entropy gain dari nilai setiap atribut pada tabel data 4.5 diketahui jumlah kaisusnyai adalah 54 dengan kelas minoritas sebanyak 28 dan tidak sebanyak 26 dihitung menggunakan persamaan Rumus sebagai berikut;

$$\text{Entropy (s)} = -p * \log_2 p$$

Nilai Entropy (Total)

$$= (-28/54 * \log_2 (28/54)) + (-26/54 * \log_2 (26/54))$$

$$= 0,99901$$

kemudian yaitu mencari Entropy kategori (Jenis Buku, Jenis Kelamin, Dan Pekerjaan). dari masing-masing atribut yang ada pada tabel 4.5 dihitung berdasarkan jumlah kelas (keterangan Minoritas Dan Tidak Minoritas).

Entropy Jenis Buku

$$\text{Aigaimai} = (-3/6 * \log_2 (3/6)) + (-3/6 * \log_2 (3/6)) = 0,21541$$

$$\text{Ensiklopedia} = (-2/2 * \log_2 (2/2)) + (0/2 * \log_2 (0/2)) = 0$$

$$\text{Kamus} = (-0/4 * \log_2 (0/4)) + (-4/4 * \log_2 (4/4)) = 0$$

$$\text{Motivasi} = (-3/6 * \log_2 (3/6)) + (-3/6 * \log_2 (3/6)) = 1$$

$$\text{Ensiklopedia} = (-6/10 * \log_2 (6/10)) + (-4/10 * \log_2 (4/10)) = 0,97095$$

$$\text{Cerpen} = (-0/1 * \log_2 (0/1)) + (-1/1 * \log_2 (1/1)) = 0$$

$$\text{Novel} = (-6/7 * \log_2 (6/7)) + (-1/7 * \log_2 (1/7)) = 0,59167$$

$$\text{Referensi} = (-5/11 * \log_2 (5/11)) + (-6/11 * \log_2 (6/11)) = 0,24662$$

$$\text{Sastra} = (-1/2 * \log_2 (1/2)) + (-1/2 * \log_2 (1/2)) = 1$$

$$\text{Sosial} = (-1/4 * \log_2 (1/4)) + (-1/4 * \log_2 (1/4)) = 0,81128$$

$$\text{Kesehatan} = (-0/1 * \log_2 (0/1)) + (-1/1 * \log_2 (1/1)) = 0$$

Entropy Jenis Kelamin

$$\text{Laki-laki} = (-10/25 * \log_2 (10/25)) + (-15/25 * \log_2 (15/25))$$

$$= 0,97095$$

Perempuan = (-17/29Log2 (17/29)) + (-12/29*Log2 (12/29))
= 0,97845

Entropy Pekerjaan

Pelajar = (-13/19*Log2 (13/19)) + (-6/19*Log2 (6/29)) = 0,89974

Guru = (-3/11*Log2 (3/11)) + (-6/19*Log2 (6/19)) = 0,9882

Umum = (-3/11*Log2 (3/11)) + (-6/11)*Log2 (6/11)) = 0, 89049

Maihasiswa=(-7/11*Log2 (7/11))+(-4/11)*Log2 (4/11))=0,94566

Setelah mendapatkain nilai hasil Entropy dilanjutkan untuk perhitungan nilai Gan dengan persamaan sebagai berikut:

Gainratio(S,A)=Entropy(S)–

Perhitungan Gain:

Jenis Buku:

=0,99901-(6/54*1)+6/54*1)+(10/54*0,2922)+(7/54*0,1781)
+(11/54*0,0120)+(2/54*0,1505)+(4/54*0,5681) = 1,49006

Jenis Kelamin: = 0,99901-(25/54*0,8797) + (29/54*0,0215) = 1,07496

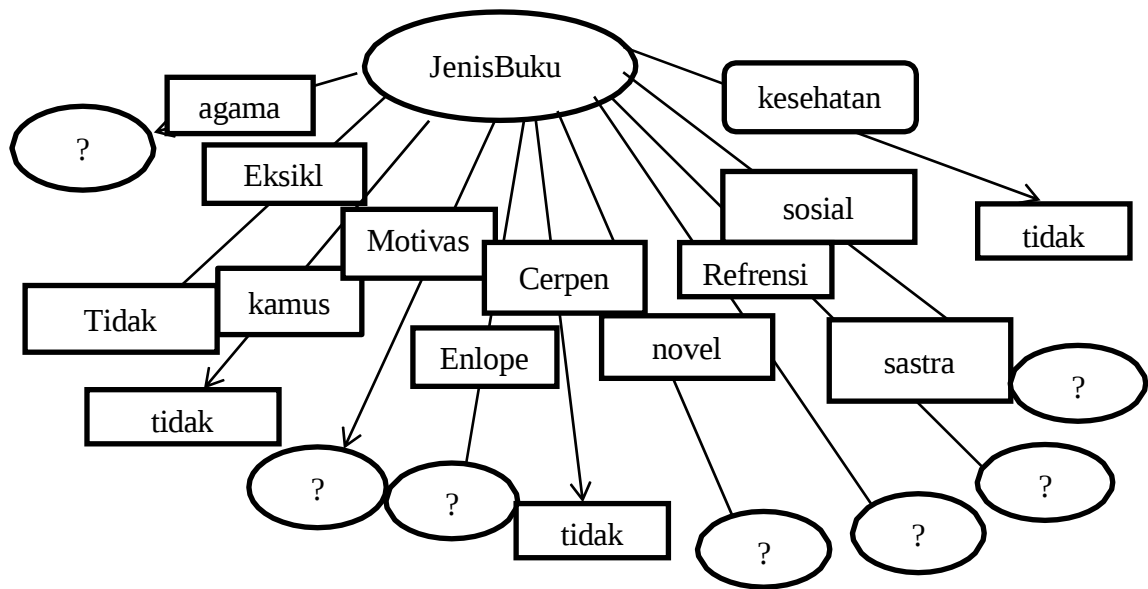
Pekerjaan: =

=0,77193-(19/54*0,0750)+(11/54*0,9882)+(13/54*0,2097)+(11/54*0,1157)
= 1,29075

Tabel 3 Nilai Hasil Perhitungan Entropy dan gain

Kategori	Atribut	Jumlah	Minat	Tidak	Entrop y	Gai n
Total		54	28	26	0,99901	
JenisBukui						1,49 006
	Agama	6	3	3	0,21541	
	EnsiklopediaMini	2	2	0	0	
	Kamus	4	0	4	0	
	Motivasi	6	3	3	1	
	Ensiklopedia	10	6	4	0,97095	
	Cerpen	1	0	1	0	
	Novel	7	6	1	0,59167	
	Referensi	11	5	6	0,24662	
	Sastra	2	1	1	1	
	Sosial	4	1	3	0,81128	
	Kesehatan	1	0	1	0	
Jenis Kelamin						1,07 496

	Laki-Laki	25	10	15	0,97095	
	Perempuan	29	17	12	0,97845	
Pekerjaan						1,29075
	Pelajar	19	13	6	0,89974	
	Guru	11	3	6	0,9882	
	Umum	13	4	9	0,89049	
	Mahasiswa	11	7	4	0,94566	



Gambar 2 Pohon Keputusan Node 1

Selanjut menghitung Entropy dari jenis kelamin.

$$\text{Entropy Total} = (-3/4 * \log_2 (3/4) + (-1/4 * \log_2 (1/4)) = 0,11278124$$

Entropy pekerjaan :

$$\text{Pelajar} = (-1/1 * \log_2 (1/1) + (-0/1 * \log_2 (0/1)) = 0$$

$$\text{Mahasiswa} = (-1/1 * \log_2 (1/1) + (-0/1 * \log_2 (0/1)) = 0$$

$$\text{Umum} = (-1/1 * \log_2 (1/1) + (-0/1 * \log_2 (0/1)) = 0$$

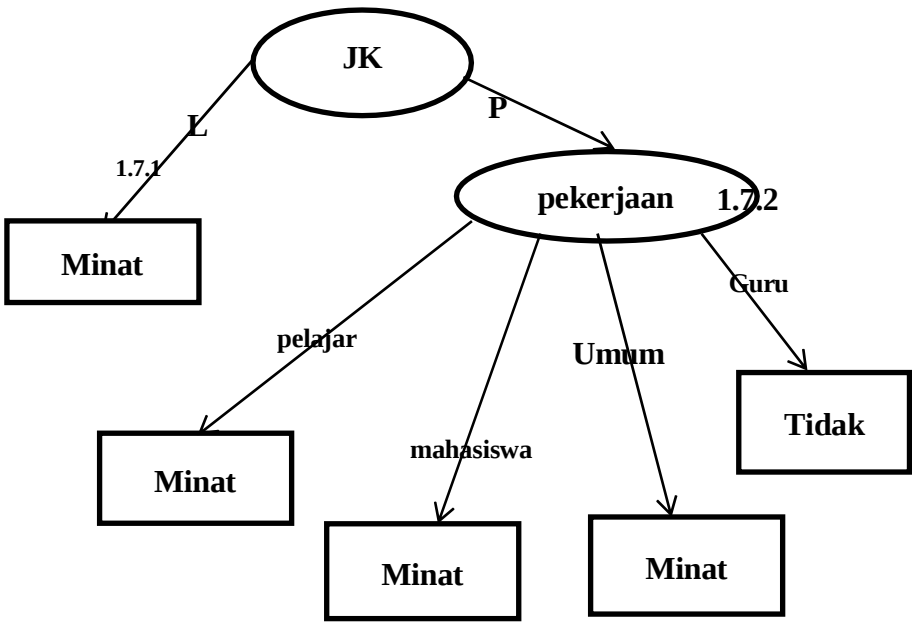
$$\text{Guru} = (-0/1 * \log_2 (0/1) + (-1/1 * \log_2 (1/1)) = 0$$

$$\text{Gain pekerjaan} = 0,11278124 - (1/48 * 0) + (1/4 * 0) + (1/4 * 0) + (1/4 * 0) = 0$$

Hasil perhitungan Entropy dan gain Node 1.7.2

Tabel 4. Hasil Perhitungan Entropy Dan Gain

kategori	atribut	jumlah	minat	tidak	Entropy	Gain
total		4	3	1	0,11278124	
pekerjaan						0
	pelajar	1	1	0	0	
	mahasiswa	1	1	0	0	
	guru	1	0	1	0	
	umum	1	1	0	0	



Gambar 3 Pohon Keputusan Node

Tabel 5 Daitai Jk Perempuan

JenisBukui	JenisKelamin	Pekerjaan	Minat/Tidak
Referensi	Perempuan	Guru	Minat
Referensi	Perempuan	Guru	Tidak
Referensi	Perempuan	Pelajar	Minat
Referensi	Perempuan	Umum	Minat

$$\text{Entropy Total} = (-2/4 * \text{Log}_2 (2/4)) + (-2/4 * \text{Log}_2 (2/4)) = 0,811278$$

Entropy Pekerjaan :

$$\text{Pelajar} = (-1/1 * \text{Log}_2 (1/1)) + (-0/1 * \text{Log}_2 (0/1)) = 0$$

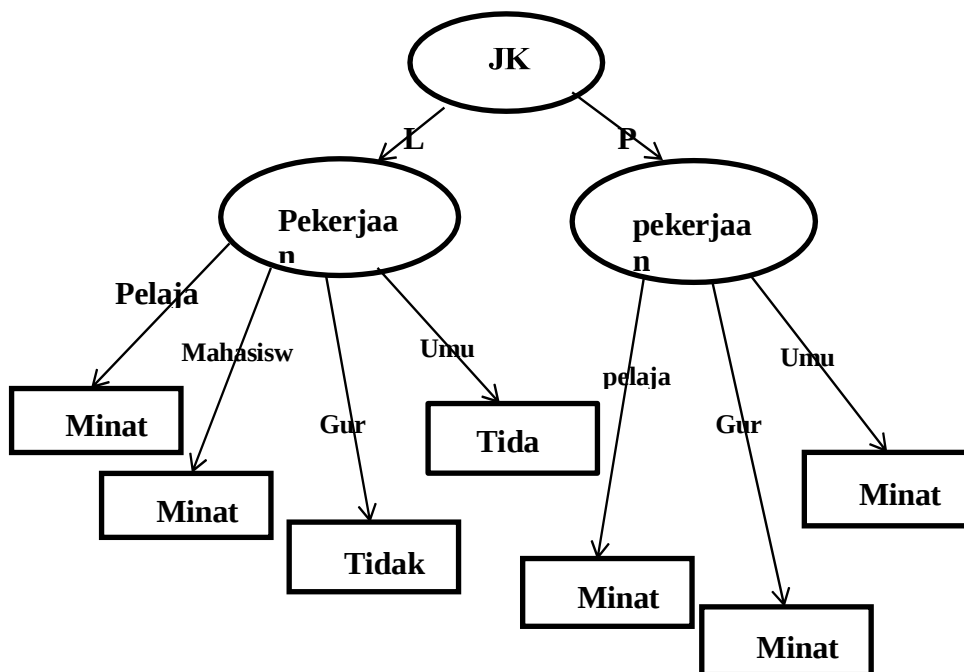
$$\text{Guru} = (-1/2 * \text{Log}_2 (1/2)) + (-1/2 * \text{Log}_2 (1/2)) = 0$$

$$\text{Umum} = (-1/1 * \text{Log}_2 (1/1)) + (-0/1 * \text{Log}_2 (0/1)) = 0$$

Tabel 6 Hasil Perhitungan Entropy Dan Gain

Kategori	Atribut	Jumlah	Minat	Tidak	Entropy	Gain
Total		4	2	2	0,811278	
Pekerjaan						0
	Pelajar	1	1	0	0	
	Guru	2	1	1	1	
	Umum	1	1	0	0	

Berikut adalah pohon keputusan dari Node cabang



Gambar 4 Pohon Keputusan node 3

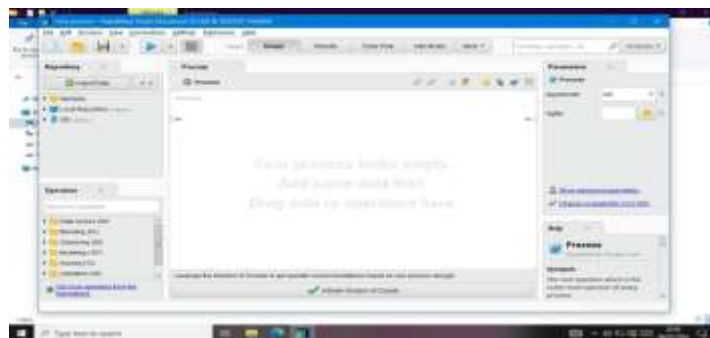
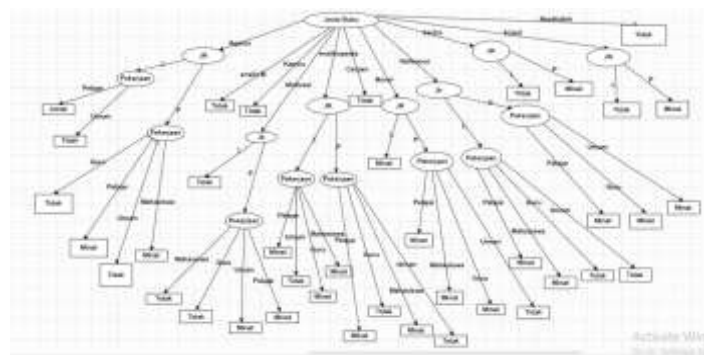
Setelah diperoleh Node cabang pada kasus jenis buku referensi terdapat kaitegori JK laki-laki memiliki kategori 4 pekerjaan yaitu pelajar, mahasiswa dan guru pada kategori pekerjaan pelajar dan mahasiswa memiliki keputusan “**Minat**” sedaingkain untuk guru dan umum menentukan keputusan “**Tidak**”.

Pada Node JK perempuan memiliki kategori 3 pekerjaan yaitu pelajar, guru dan umum serta menemukan keputusan dari pelajar, mahasiswa dain umum yaitu “**Minat**”

Gambar 5 pohon keputusan akhir

B. Tahapan Implementasi *Decision tree*

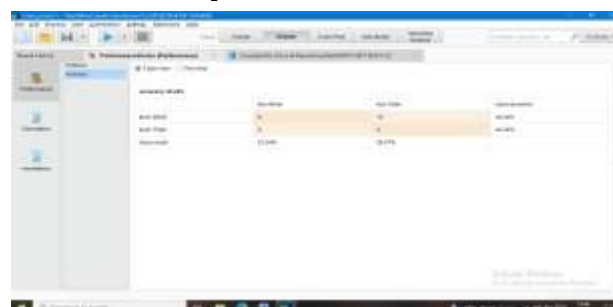
Berikut merupakan langkah-langkah menentukan pohon keputusan pada aplikasi rapidminer. Tampilan AIawal AIplikasi *Rapidminer*



Setelah beberapa tahapan sudah dilakukan pada rapidminer maka mendapatkan hasil pohoh keputusan seperti gambar dibawah ini:



Serta Menampilkan hasil akurasi sebesar 44.48%



Model	Accuracy	Area Under the Curve	Confusion Matrix
Model 1	0.4448	0.5	Confusion Matrix

KESIMPULAN

Berdasarkan rumusan masalah beserta pembahasan diatas dapat disimpulkan bahwa

1. Untuk memprediksi minat baca Perpustakaan Daerah Kabupaten Rejang Lebong langkah pertama yaitu dengan cara Pengumpulan Data, Melakukan Data Seleksi, Transformasi Data, setelah itu pisah kan data berbentuk data set untuk pengujian dengan proporsi yang sesuai dengan ketentuan jika peminjam >3 maka Tidak minat, lalu melakukann perhitungan entropy dan Gain disetiap atribut, kemudian tentukan gain tertinggi untuk sebagai akar dan entropy untuk menentukan node yang menjadi pemecah atau cabang sampai semua entropy memiliki nilai yang sama (0).
2. Dari hasil perhitungan Entropy dan Gain maka ditemukan hasil terbanyak jenis buku yang diminaiti adalah jenis buku Ensiklopedia Dan Referensi serta ditemukan pohon keputusan akhir seperti pada gambar 4.14 dan menunjukan hasil akurasi sebesar 44,48%.

DAFTAR PUSTAKA

- Anggriani, S. D., Muhammad Syahril, S. E., Kom, M., &ItaMariami, S. E. (2022). Data Mining Algoritma C4. 5 UntukMenganalisaPendudukPenerima Program KeluargaHarapan (PKH)(StudiKasus: KelurahanTualang V). *Jurnal Cyber Tech*, 1(6).
- Badriah, S., Nugroho, M. F. E., Sanjaya, N., Rismawati, I., Sari, B. N., &Rozikin, C. (2021).KlasifikasiAlgoritma C4. 5 dalamMenentukanPenerimaBantuan Covid-19:(StudiKasus: Desa di Karawang). *JurnalInformatikaPolinema*, 7(3), 23-28.
- Fitriani, E. (2020). PerbandinganAlgoritma C4. 5 Dan Naïve Bayes UntukMenentukanKelayakanPenerimaBantuan Program KeluargaHarapan. *SISTEMASI: JurnalSistemInformasi*, 9(1), 103-115.
- Murlena, DiwiApriana(2022).Penerapan Data Mining UntukMemprediksiKetersediaanStokProduk HNI HPAI MenggunakanAlgoritma C4.5.Arcitech: Journal of Computer Science and Artificial Intellegence, Volume 2 No. 1 (January - June 2022), pp. 19 – 32
- Saragih, I. A., Firmansyah, A. U., &Rohminatin, R. (2021).PenerapanMetode C4. 5 UntukPenentuanKelayakanPenerima Program KeluargaHarapan. *JUTSI (JurnalTeknologidanSistemInformasi)*, 1(3), 275-280.
- Sari, R. A. PREDIKSI CALON PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN (PKH) MENGGUNAKAN ALGORITMA C4. 5 (STUDI KASUS: KECAMATAN BANGKO KABUPATEN ROKAN HILIR).
- Sari, R. A. PREDIKSI CALON PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN (PKH) MENGGUNAKAN ALGORITMA C4. 5 (STUDI KASUS: KECAMATAN BANGKO KABUPATEN ROKAN HILIR).
- Tanjung, M. A., Poningsih, P., &Qurniawan, H. (2021).AnalisaKelayakanPenerima Program KeluargaHarapan (PKH) MenggunakanAlgoritma C4. 5. *Jurasik (JurnalRisetSistemInformasidanTeknikInformatika)*, 6(1), 217-225.