

ANALISA DAN PENERAPAN METODE KLASIFIKASI DALAM DATA MINING UNTUK PENERIMAAN SISWA JALUR NON-TULIS

Reni Kurniah

Universitas Prof. Dr. Hazairin SH Bengkulu Jl. A.Yani No. 1 .Kota Bengkulu.38115Indonesia

INFORMASI ARTIKEL

Sejarah Artikel

Diterima Redaksi: 11 Februari 2020

Revisi Akhir: 09 Maret 2020

Diterbitkan Online: 25 Maret 2020

KATA KUNCI

Algorithm C4.5

Decision Tree

Clasification

KORESPONDENSI

E-mail:renikurniah88@gmail.com

A B S T R A C T

Admission of new students is always held every year in senior high school (SLTA) both public and private. At the admission of new students at SMA Plus Negeri 7 Bengkulu City is needed an information that can help the school to be younger in accepting new students in accordance with the criteria school. From the admission data the student will get useful information for the school where the existing data will be an information and policy in accepting students who will go to school. Importance Information obtained from the school will help the school in determining students who will accepted. C4.5 algorithm is a classification technique in which there is a data mining process using the CRISP-DM method so that a simple but accurate data classification will be obtained. Therefore the use of the C4.5 algorithm will make it easier for schools to make policies in accepting students. new

1. PENDAHULUAN

Seiring dengan perkembangan dunia teknologi informasi, kebutuhan masyarakat akan data dan informasi semakin meningkat. Data yang tadinya hanya sebatas bentuk kumpulan arsip dan tidak di olah lebih lanjut. Namun, saat ini data umumnya telah diolah sehingga dapat menghasilkan informasi. Untuk mengolah data menjadi informasi dibutuhkan suatu sistem yang terkomputerisasi dengan menggunakan database. Saat ini sistem komputerisasi tidak hanya terbatas pada perusahaan bisnis, namun juga sampai pada lembaga pendidikan. Salah satu fungsi dari informasi hasil pengolahan data bagi lembaga yang bergerak di bidang pendidikan (Dalam hal ini Sekolah Menengah Atas atau SLTA).

Lembaga pendidikan yang dapat memanfaatkan ketersediaan informasi yang tepat sangat berpengaruh pada meningkatnya kualitas Penerimaan siswa baru merupakan salah satu proses yang ada di instansi pendidikan seperti layanan pada instansi pendidikan terutama untuk menyaring calon siswa baru yang akan diterima pada sekolah tersebut sehingga dapat membantu menghasilkan keputusan yang tepat.

Keputusan yang akan diambil seringkali harus didukung oleh tersedianya informasi yang cepat, ini sering terjadi bila keputusan harus segera diambil. Dengan menggunakan komputer maka perolehan informasi yang cepat dan tepat dapat direalisasikan. Menggunakan komputer tidak berarti semua informasi pasti dapat diperoleh dari komputer, tetapi harus ada

data-data yang disimpan di dalam komputer dan ada prosedur-prosedur yang dapat memperbaharui dan mengolah data-data tersebut menjadi sebuah informasi yang dibutuhkan sekolah menengah atas yang berguna untuk menyaring calon siswa yang terpilih sesuai kriteria yang ditentukan oleh sekolah tersebut untuk menjadi siswa didiknya. Salah satu jalur penerimaan siswa baru pada Sekolah Menengah Atas adalah melalui jalur non-tulis, pada jalur ini sekolah tidak lagi melakukan tes tertulis terhadap siswa yang mendaftar, tetapi berdasarkan nilai raport siswa dari semester satu sampai dengan semester lima, dan prestasi pendukung sewaktu di SLTP/MTS.

Masalah menentukan calon siswa ini dapat dilakukan dengan metode klasifikasi *data mining*. Adapun pohon keputusan ini adalah sebuah jawaban akan sebuah sistem yang manusia kembangkan untuk membantu mencari dan membuat keputusan untuk masalah-masalah tersebut dan dengan memperhitungkan berbagai macam faktor yang ada di dalam lingkup masalah tersebut. Dengan pohon keputusan, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah dan dapat mencari penyelesaian terbaik dengan memperhitungkan faktor-faktor tersebut.

2. TINJAUAN PUSTAKA

Data Mining

Data mining merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan secara otomatis [1]. Data mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data.[2]

Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan di masa depan. Data mining ini juga dikenal dengan istilah pattern recognition [2].KDD atau Knowledge Discovery from Data, merupakan proses terstruktur, yaitu sebagai berikut: [3]

1. Data Cleaning adalah Proses membersihkan data dari data noise dan tidak konsisten.
2. Data Integration adalah Proses untuk menggabungkan data dari beberapa sumber yang berbeda.
3. Data Selection adalah Proses untuk memilih data dari database yang sesuai dengan tujuan analisis.
4. Data Transformation adalah Proses mengubah bentuk data menjadi data yang sesuai untuk proses Mining.
5. Data Mining adalah Proses penting yang menggunakan sebuah metode tertentu untuk memperoleh sebuah pola dari data.
6. Pattern Evaluation adalah Proses mengidentifikasi pola.
7. Knowledge Presentation adalah yang dapat mempresentasikan informasi yang dibutuhkan, proses dimana informasi yang telah didapatkan kemudian digunakan oleh pemilik data.

2.1 Algoritma C4.5

Algoritma C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial.Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru. Algoritma C4.5 sendiri merupakan pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam hal, bisa mengatasi missing data, bisa mengatasi data kontinu dan pruning. [6]

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.

3. METODOLOGI

Metode Penelitian

Jenis penelitian menggunakan jenis penelitian kualitatif untuk memperoleh informasi dengan jelas tentang bagaimana sistem penerimaan siswa pada tingkat SLTA melalui jalur Non-Tulis.Instrumen penelitian menggunakan pedoman wawancara

semi terstruktur, buku catatan,serta alat perekam sebagai alat bantu melakukan pengumpulan data juga dengan menggunakan lembar observasi. Cara pengumpulan data dengan menggunakan triangulasi. Analisa data Kualitatif yaitu reduksi data, penyajian data serta penarikan dan simpulan atau verivikasi

Analisa Data Mining Dengan Algoritma C4.5

Data mining ialah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Dalam data mining ada beberapa teknik menganalisis data, salah satunya adalah dengan algoritma C4.5 Algoritma C4.5 merupakan pengembangan dari algoritma ID3, yang digunakan untuk membentuk pohon keputusan.

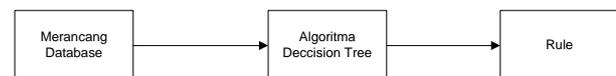
Proses untuk membentuk pohon keputusan adalah diawali dengan pengambilan data calon siswa baru dan dikelompokkan menjadi beberapa kriteria sebagai berikut :

1. Rata-rata nilai raport Bahasa Indonesia (semester 1 – semester 5)
2. Rata-rata nilai raport Bahasa Inggris (semester 1 - semester 5)
3. Rata-rata nilai rapor Matematika (semester 1 – semester 5)
4. Rata-rata nilai raport IPA (semester 1 – semester 5)
5. Kriteria Prestasi pendukung
6. Kriteria total nilai rata-rata.

Perancangan Sistem

Perencanaan ini dimulai dari menentukan alur kerja sistem berikut data yang mengalir dalam sub sistem yang ada, lanjutkan dengan merancang database, dan merancang proses-proses yang ada di setiap sub sistem.

Berdasarkan pada sistem yang ada telah diuraikan di atas, maka secara sederhana proses penemuan rule yang akan digunakan dalam proses prediksi penerimaan siswa baru tersebut dapat dilihat pada gambar 1



Gambar 1. Gambaran Sederhana Proses Penemuan Knowledge dengan Algoritma Decision Tree

Algoritma Decision Tree dipakai untuk menemukan sebuah rule untuk pengetahuan yang baru dari tumpukan basis data Nilai evaluasi yang banyak. Perancangan ini dimulai dari menentukan alur kerja sistem berikut data yang mengalir dalam sub sistem yang ada di setiap sub sistem.

Data Flow Diagram

Dalam Data Flow Diagram Ini akan terlihat dengan jelas aliran data yang terjadi pada proses penemuan rule dengan model decision tree,sehingga diharapkan orang yang tidak memahami model ini akan dapat mengerti tentang rancangan sistem.

Contex Diagram

Contex Diagram adalah merupakan langkah pertama dalam pembuatan data flow diagram,yang mana dalam contex diagram

digambarkan secara umum proses pembentukan sistem. Untuk lebih jelasnya dapat dilihat pada gambar 2



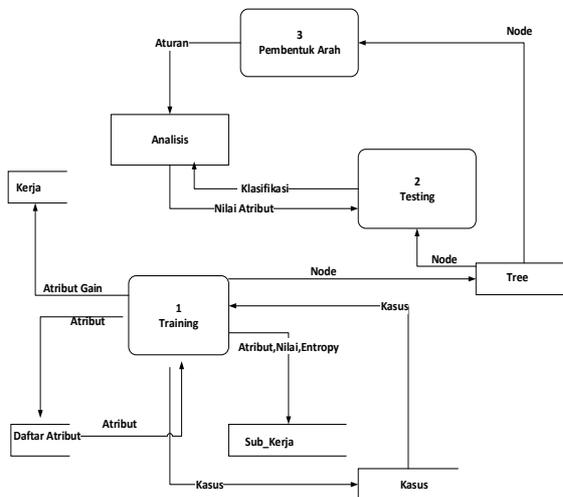
Gambar.2 Context Diagram Decision Tree

Dari *Context* diagram di atas terlihat peranan seorang analis dalam pencarian rule dalam sistem memprediksi tingkat keberhasilan penerimaan siswa baru. Gambar 2 di atas dapat dijelaskan sebagai berikut :

1. Analisis memilih atribut dari nilai siswa yang akan diproses sebagai training.
2. Kemudian mentransformasi data sehingga menjadi sama dan untuk dilakukan testing.
3. Data *training* yang akan telah normal akan diproses dengan teknik klasifikasi *decision tree*.
4. Menemukan *rule* dari *tree*.

Diagram Alir Data

Aliran data yang terjadi pada proses pembentukan sistem kelulusan calon siswa baru dalam aplikasi ini ditunjukkan dalam Diagram Alir Data level 1 seperti pada gambar 3 di bawah ini :



Gambar 3. Diagram Alir Data Level 1

Komponen-komponen yang terdapat dalam Algoritma C4.5 adalah seperti yang terdapat pada Gambar 3 yaitu :

1. Subkerja
Bagian ini digunakan untuk memasukkan pengetahuan, mengkonstruksi atau memperluas pengetahuan dalam basis pengetahuan.
2. Kasus
Bagian ini merupakan data atau sampel yang akan diolah pada algoritma C4.5
3. Atribut
Bagian ini berisi jenis-jenis data yang akan diproses. Contoh atribut pada penelitian ini adalah nilai total rata-

rata,nilai rata-rata Bahasa Indonesia,nilai rata-rata bahasa inggris,nilai rata-rata matematika, nilai rata-rata IPA, nilai total rata-rata,dan prestasi pendukung.

4. *Training*
Bagian ini merupakan proses untuk menjalankan data atau atribut yang telah dimasukkan.
5. *Node*
Bagian ini merupakan hasil dari algoritma inialisasi
6. *Tree*
Bagian ini merupakan hasil dari pembentukan *node* kemudian menjadi salah satu akar dari pohon keputusan Testing Data yang sudah terbentuk pada *tree* kemudian diproses pada sistem ini.
7. Pembentukan Aturan
Pada tahapan ini sudah terjadi pembentukan aturan pada algoritma C4.5
Untuk mendapatkan model, kita harus melakukan analisis terhadap data latihan (*training set*). Sedangkan data diuji (*test set*) digunakan untuk mengetahui tingkat akurasi dari model yang telah dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai kelas dari suatu data.Tahapan – tahapan yang terjadi pada algoritma C4.5 untuk penelian ini dijelaskan sebagai berikut :
Data yang sudah terbentuk pada *tree* kemudian diproses pada sistem ini.
8. Pembentukan Aturan
Pada tahapan ini sudah terjadi pembentukan aturan pada algoritma C4.5.
Untuk mendapatkan model, kita harus analisis terhadap data latihan (*training set*). Sedangkan data diuji (*test set*) digunakan untuk mengetahui tingkat akurasi dari model yang telah dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai kelas dari suatu data. Tahapan – tahapan yang terjadi pada algoritma C4.5 untuk penelian ini dijelaskan sebagai berikut :

Pemilihan Variabel

Data penerimaan calon siswa baru pada SLTA Negeri 7 plus kota Bengkulu, memiliki format seperti berikut :

1. Nomor Induk Siswa
2. Nama
3. Sekolah Asal
4. Nilai *raport* Bahasa Indonesia (semester 1 – semester 5)
5. Nilai *raport* Bahasa Inggris (semester 1 – semester 5)
6. Nilai *raport* Matematika (semester 1 – semester 5)
7. Nilai *raport* IPA (semester 1 – semester 5)
8. Prestasi Pendukung
9. Jumlah nilai rapor dan pendukung
10. Rata-rata per 20 nilai
11. Keterangan Penerimaan

Dari data-data tersebut,yang diambil sebagai variabel keputusannya adalah nilai kelulusan “LULUS”, dan “TIDAK”. Sedangkan yang diambil sebagai variabel penentu dalam pembentukan pohon keputusan adalah nilai rata-rata Bahasa Indonesia, Bahasa Inggris, matematika, IPA, Prestasi Pendukung dan rata-rata per 20 nilai (Bahasa Indonesia, Bahasa Inggris, matematika, IPA).Pemilihan variabel-variabel tersebut engan pertimbangan bahwa jumlah nilai variabelnya tidak banyak sehingga diharapkan penerimaan siswa yang masuk dalam satu klasifikasi nilai variabel tersebut cukup banyak..

Dari 39 jumlah data siswa yang mendaftar setelah dilakukan perhitungan yang menggunakan aturan struges didapatkan maka akan mendapatkan 6 kelas,dan data tersebut akan dikelompokkan berdasarkan atribut sebagai berikut.

1. Mengelompokkan Nilai rata – rata bidang studi Bahasa

Indonesia, Bahasa Inggris, matematika, IPA, pengelompokan ini berdasarkan dari nilai raport siswa dari semester satu sampai dengan semester lima. Sehingga nilai tersebut dapat dikelompokkan seperti terlihat pada tabel 2.

Tabel.1 Nilai Rata-Rata Bahasa Indonesia

Nilai rata-rata Bahasa Indonesia	Klasifikasi
69-72	1
73-76	2
77-80	3
81-84	4
85-88	5
89-88	6

Tabel.2. Nilai Rata-Rata Bahasa Inggris

Nilai rata-rata Bahasa Inggris	Klasifikasi
71-74	1
75-78	2
79-82	3
83-86	4
87-90	5
91-94	6

Tabel.3 Nilai Rata-Rata IPA

Nilai rata-rata IPA	Klasifikasi
67-70	1
71-74	2
75-78	3
79-82	4
83-86	5
87-90	6

2. Mengelompokkan Prestasi pendukung berdasarkan prestasi yang telah didapatkan oleh siswa.

Tabel.4 Prestasi

Prestasi	Klasifikasi
Ada	Y

3. Mengelompokkan nilai total rata-rata Bahasa Indonesia, Bahasa Inggris, Matematika, dan IPA, berdasarkan dari hasil nilai raport siswa dari semester satu sampai dengan semester lima sehingga nilai tersebut dikelompokkan seperti terlihat pada tabel 4.7

Tabel 5. Nilai Total Rata-Rata

Nilai Tot Rata-Rata	Klasifikasi
7,10-7,49	1
7,50-7,89	2
7,90-8,29	3
8,30-8,69	4
8,70-9,09	5
9,10-9,49	6

Tabel 6. Format Data Akhir Calon Siswa Baru

NO	NIS	Nilai rata-rata Bahasa Indonesia	Nilai rata-rata Bahasa Inggris	Nilai rata-rata Matematika	Nilai rata-rata IPA	Prestasi Pendukung	Nilai Total Rata-rata	Keterangan Penerimaan
1	20146	6	4	7	7	Y	6	LULUS
2	73340	7	4	5	7	T	5	LULUS
3	5825	8	4	5	7	Y	5	LULUS
4	24145	6	3	5	6	T	5	LULUS
5	41801	5	3	3	5	T	4	TIDAK
6	36338	4	4	3	3	T	3	TIDAK
7	73196	6	5	5	6	Y	5	LULUS
8	61510	7	4	5	5	T	5	LULUS
9	80450	6	5	5	6	T	5	LULUS
10	3455	6	4	5	6	Y	5	LULUS
11	73208	6	4	6	6	T	5	LULUS
12	51209	6	4	4	7	T	5	LULUS
13	20533	6	4	3	6	T	5	LULUS
14	21470	6	5	5	5	T	5	LULUS
15	27461	6	4	5	5	T	5	LULUS
16	73290	5	5	4	6	Y	5	LULUS
17	93253	6	5	5	5	T	5	LULUS
18	23332	7	3	5	5	T	5	LULUS
19	6553	6	3	5	5	T	4	LULUS
20	25001	6	3	5	5	T	4	LULUS
21	38055	6	4	4	5	T	4	LULUS
22	11282	5	3	5	6	T	4	LULUS
23	34818	4	3	5	6	T	4	LULUS
24	07337	5	3	5	5	T	4	TIDAK
25	23214	5	4	3	5	T	4	TIDAK
26	88970	5	3	3	5	T	4	TIDAK
27	93272	4	3	5	4	T	4	TIDAK
28	24151	5	2	3	5	T	4	TIDAK
29	11986	5	4	2	4	T	3	TIDAK
30	24135	5	2	2	5	T	3	TIDAK
31	61609	5	2	3	3	T	3	TIDAK
32	73270	5	2	2	4	T	3	TIDAK
33	73141	4	2	4	4	T	3	TIDAK
34	59874	4	2	2	4	T	3	TIDAK
35	17693	4	2	2	3	T	2	TIDAK
36	07492	4	1	2	3	T	1	TIDAK
37	93882	5	2	1	2	T	1	TIDAK
38	94553	5	2	1	2	T	1	TIDAK
39	07665	2	1	2	4	T	1	TIDAK

Format data akhir pada tabel 4 diatas diperoleh berdasarkan dari atribut yang telah dikelompokkan atau diklasifikasi, misalkan data pada tabel 4.1 Nilai rata-rata Bahasa Indonesia "84,5" setelah diklasifikasi menjadi "6", Nilai rata-rata bahasa Inggris "83,5" setelah diklasifikasi menjadi "4", Nilai rata-rata Matematika "91,2" setelah diklasifikasi menjadi "7", Nilai IPA "89,5" setelah diklasifikasi menjadi "7" prestasi pendukung "ada", maka setelah diklasifikasi menjadi "Y", Nilai total rata-rata "87,65" menjadi "6" dan seterusnya.

Pohon Keputusan

Dari format data akhir klasifikasi penerimaan calon siswa baru maka akan dilakukan klasifikasi data algoritma C4.5 dengan membuat pohon keputusan. Seperti yang telah dijelaskan sebelumnya, algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai
3. Buat kasus dalam cabang Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki nilai yang sama.

Dalam kasus yang tertera pada tabel 8 di atas, akan dibuat pohon keputusan untuk menentukan klasifikasi penerimaan calon siswa baru (Lulus dan Tidak) dengan melihat nilai rata-rata bahasa Indonesia, nilai rata-rata bahasa Inggris, nilai rata-rata matematika, nilai rata-rata IPA, prestasi pendukung, dan Nilai total rata-rata.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus (2.1), sedangkan untuk menghitung nilai entropy dapat digunakan rumus (2.2).

Dengan menggunakan dua persamaan di atas maka akan didapatkan entropy dan gain yang digunakan sebagai akar dalam membuat pohon keputusan.

Menghitung jumlah kasus, jumlah kasus untuk keputusan “Lulus”, jumlah kasus untuk keputusan “Tidak”, dan kasus yang dibagi berdasarkan atribut Nilai rata-rata bahasa Indonesia, bahasa Inggris, matematika, IPA, prestasi pendukung dan nilai total rata-rata. Setelah itu lakukan perhitungan gain untuk setiap atribut.

Menghitung Nilai Entropy tiap-tiap atribut :

Entropy (Total)

$$\text{Entropy}(\text{total}) = \left(-\frac{21}{39} \times \log_2\left(\frac{21}{39}\right)\right) + \left(-\frac{18}{39} \times \log_2\left(\frac{18}{39}\right)\right) = 0.99573$$

Entropy (total) adalah menghitung nilai total keputusan ya (21) dan tidak (18) sedangkan 39 adalah jumlah keseluruhan kasus.

Atribut Nilai Total Rata-Rata

$$\text{Entropy}(1) = \left(-\frac{0}{7} \times \log_2\left(\frac{0}{7}\right)\right) + \left(-\frac{7}{7} \times \log_2\left(\frac{7}{7}\right)\right) = 0$$

$$\text{Entropy}(2) = \left(-\frac{0}{9} \times \log_2\left(\frac{0}{9}\right)\right) + \left(-\frac{9}{9} \times \log_2\left(\frac{9}{9}\right)\right) = 0$$

$$\text{Entropy}(3) = \left(-\frac{11}{13} \times \log_2\left(\frac{11}{13}\right)\right) + \left(-\frac{2}{13} \times \log_2\left(\frac{2}{13}\right)\right) = 0.61938$$

$$\text{Entropy}(4) = \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0$$

$$\text{Entropy}(5) = \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = 0$$

$$\text{Entropy}(6) = \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) = 0$$

Menghitung nilai entropy atribut Nilai Total Rata-Rata berdasarkan dari tiap-tiap kelas (1,2,3,4,5,6), pada atribut Nilai Total Rata-Rata .

Atribut Nilai Rata-Rata Bahasa Indonesia

$$\text{Entropy}(1) = \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0$$

$$\text{Entropy}(2) = \left(-\frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \times \log_2\left(\frac{3}{4}\right)\right) = 0.41504$$

$$\text{Entropy}(3) = \left(-\frac{1}{14} \times \log_2\left(\frac{1}{14}\right)\right) + \left(-\frac{3}{14} \times \log_2\left(\frac{3}{14}\right)\right) = 0.10962$$

$$\text{Entropy}(2) = \left(-\frac{4}{4} \times \log_2\left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4} \times \log_2\left(\frac{0}{4}\right)\right) = 0$$

$$\text{Entropy}(4) = \left(-\frac{14}{15} \times \log_2\left(\frac{14}{15}\right)\right) + \left(-\frac{1}{15} \times \log_2\left(\frac{1}{15}\right)\right) = 3.90689$$

$$\text{Entropy}(5) = \left(-\frac{5}{5} \times \log_2\left(\frac{5}{5}\right)\right) + \left(-\frac{0}{5} \times \log_2\left(\frac{0}{5}\right)\right) = 0$$

$$= 0$$

$$\text{Entropy}(6) = \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) = 0$$

Menghitung nilai entropy atribut Nilai Rata-Rata Bahasa Indonesia berdasarkan dari tiap kelas (1,2,3,4,5,6)

Atribut Nilai Rata-Rata Bahasa Inggris

$$\text{Entropy}(1) = \left(-\frac{0}{8} \times \log_2\left(\frac{0}{8}\right)\right) + \left(-\frac{8}{8} \times \log_2\left(\frac{8}{8}\right)\right) = 0$$

$$\text{Entropy}(2) = \left(-\frac{0}{30} \times \log_2\left(\frac{0}{30}\right)\right) + \left(-\frac{30}{30} \times \log_2\left(\frac{30}{30}\right)\right) = 0.94556$$

$$\text{Entropy}(3) = \left(-\frac{4}{11} \times \log_2\left(\frac{4}{11}\right)\right) + \left(-\frac{7}{11} \times \log_2\left(\frac{7}{11}\right)\right) = 0.81128$$

$$\text{Entropy}(4) = \left(-\frac{11}{12} \times \log_2\left(\frac{11}{12}\right)\right) + \left(-\frac{1}{12} \times \log_2\left(\frac{1}{12}\right)\right) = 0.41382$$

$$\text{Entropy}(5) = \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$\text{Entropy}(6) = \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) = 0$$

Menghitung nilai entropy atribut Nilai Rata-Rata Bahasa Inggris berdasarkan tiap-tiap kelas (1,2,3,4,5,6)

Atribut Nilai Rata-Rata Matematika

$$\text{Entropy}(1) = \left(-\frac{0}{6} \times \log_2\left(\frac{0}{6}\right)\right) + \left(-\frac{0}{6} \times \log_2\left(\frac{0}{6}\right)\right) = 0$$

$$\text{Entropy}(2) = \left(-\frac{0}{4} \times \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{0}{4} \times \log_2\left(\frac{0}{4}\right)\right) = 0$$

$$\text{Entropy}(3) = \left(-\frac{3}{9} \times \log_2\left(\frac{3}{9}\right)\right) + \left(-\frac{6}{9} \times \log_2\left(\frac{6}{9}\right)\right) = 0.9183$$

$$\text{Entropy}(4) = \left(-\frac{10}{13} \times \log_2\left(\frac{10}{13}\right)\right) + \left(-\frac{3}{13} \times \log_2\left(\frac{3}{13}\right)\right) = 0.77935$$

$$\text{Entropy}(5) = \left(-\frac{6}{6} \times \log_2\left(\frac{6}{6}\right)\right) + \left(-\frac{0}{6} \times \log_2\left(\frac{0}{6}\right)\right) = 0$$

$$\text{Entropy}(6) = \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = 0$$

Menghitung nilai entropy atribut nilai rata-rata matematika berdasarkan tiap kelas (1,2,3,4,5,6)

Nilai Rata-Rata IPA

$$\text{Entropy}(1) = \left(-\frac{0}{2} \times \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right)\right) = 0$$

$$\text{Entropy (3)} = \left(-\frac{6}{6} \times \log_2\left(\frac{6}{6}\right)\right) + \left(-\frac{0}{6} \times \log_2\left(\frac{0}{6}\right)\right) = 0$$

$$\text{Entropy (4)} = \left(-\frac{7}{12} \times \log_2\left(\frac{7}{12}\right)\right) + \left(-\frac{5}{12} \times \log_2\left(\frac{5}{12}\right)\right) = 0,97987$$

$$\text{Entropy (5)} = \left(-\frac{10}{11} \times \log_2\left(\frac{11}{11}\right)\right) + \left(-\frac{1}{11} \times \log_2\left(\frac{1}{11}\right)\right) = 0,4395$$

Atribut Nilai Rata-Rata IPA

$$\text{Entropy (1)} = \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$\text{Entropy (2)} = \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$\text{Entropy (3)} = \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$\text{Entropy (4)} = \left(-\frac{9}{9} \times \log_2\left(\frac{9}{9}\right)\right) + \left(-\frac{0}{9} \times \log_2\left(\frac{0}{9}\right)\right) = 0$$

$$\text{Entropy (5)} = \left(-\frac{4}{4} \times \log_2\left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4} \times \log_2\left(\frac{0}{4}\right)\right) = 0$$

$$\text{Entropy (6)} = \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) = 0$$

Menghitung nilai entropy atribut Nilai Rata-Rata IPA berdasarkan dari tiap kelas (1,2,3,4,5,6)

4. HASIL DAN PEMBAHASAN

Hasil

Untuk pengujian kebenaran dari hasil pengolahan data yang dikerjakan secara manual kita dapat menggunakan salah satu software aplikasi WEKA Knowledge Explorer dengan langkah-langkah sebagai berikut :

1. Seluruh variabel-variabel (terdiri dari atribut kondisi dan atribut keputusan) yang digunakan untuk menentukan klasifikasi penerimaan calon siswa baru disimpan pada microsoft excel dengan nama file data penerimaan siswa.xls (yang berisi kasus atau kriteria dalam menghasilkan rule) seperti yang terlihat pada gambar 5.1.
2. File data penerimaan siswa.xls kemudian disimpan sebagai csv.Selanjutnya,buka file tersebut dari microsoft word,notepad,atau editor teks lainnya dan data sudah berubah dalam format comma-separated.Lalu sesuaikan data tersebut dengan menambahkan informasi awal hasilnya,data tersebut sudah dapat digunakan sebagai input dalam WEKA.

	A	B	C	D	E	F	G
	Bk.Ind	Bk.ing	MM	IPA	Pendukung	Total Rata-rata	Keterangan penerimaan
2	6	4	7	7	Y	6	LULUS
3	7	4	5	7	T	5	LULUS
4	8	4	3	7	Y	5	LULUS
5	6	3	5	6	T	5	LULUS
6	5	3	3	5	T	4	TIDAK
7	4	4	3	3	T	3	TIDAK
8	6	5	5	6	Y	5	LULUS
9	7	4	5	5	T	5	LULUS
10	6	5	5	6	T	5	LULUS
11	6	4	5	6	Y	5	LULUS
12	6	4	6	6	T	5	LULUS
13	6	4	4	7	T	5	LULUS
14	6	4	3	6	T	5	LULUS
15	6	5	5	5	T	5	LULUS
16	6	4	5	5	T	5	LULUS
17	5	5	4	6	Y	5	LULUS

Gambar 4. Data siswa Pada Excel

Agar data tersebut dapat digunakan pada aplikasi WEKA decision tree,maka data tersebut harus disimpan dalam format “ARFF” (Attribute-Relation File Format) adalah sebuah file teks ASCH yang berisi daftar instance dalam sekumpulan atribut.File ARFF dikembangkan oleh Machine Learning Project di Departement of computers science of The University of wakaito untuk digunakan dalam perangkat lunak WEKA.Data dalam format.arff tersebut dapat dipenuhi dengan cara :

1. Data dipisahkan dengan koma,dengan kelas sebagai atribut terakhir.
2. Bagian header diawali dengan @Relation, Weka. data generation.classifier.calssification.
3. Tiap atribut ditandai dengan @ATTRIBUTE. Tipe-tipe data dalam WEKA: numerik (REAL atau INTEGER),nominal,String,dan Date.
4. Bagian data diawali dengan @DATA.

```

@relation weka.datagenerators.classifiers.classification
@attribute BkInd numeric {1,2,3,4,5,6,7,8}
@attribute BkIng numeric {1,2,3,4,5,6,7,8}
@attribute MM numeric {1,2,3,4,5,6,7,8}
@attribute IPA numeric {1,2,3,4,5,6,7,8}
@attribute Pendukung {Y,T}
@attribute Total numeric {1,2,3,4,5,6,7,8}
@attribute Keterangan {LULUS,TIDAK}

@data
6,4,7,7,5,LULUS
7,4,5,7,5,LULUS
8,4,3,7,5,LULUS
6,3,6,7,5,LULUS
5,3,3,7,4,TIDAK
4,3,3,3,3,TIDAK
6,5,3,6,5,LULUS
7,4,3,6,5,LULUS
6,5,6,5,5,LULUS
6,4,5,6,5,LULUS
6,4,6,6,5,LULUS
6,4,4,5,5,LULUS
6,5,5,5,5,LULUS
7,5,5,5,5,LULUS
6,5,5,7,4,LULUS
6,3,5,7,4,LULUS
5,3,5,7,4,LULUS
5,3,6,7,4,LULUS
4,3,5,6,4,LULUS
5,4,5,5,4,TIDAK
5,3,5,4,4,TIDAK
4,3,5,4,7,4,TIDAK
    
```

Gambar 5. Data Telah Disimpan Pada Format ARFF

Pengujian Program

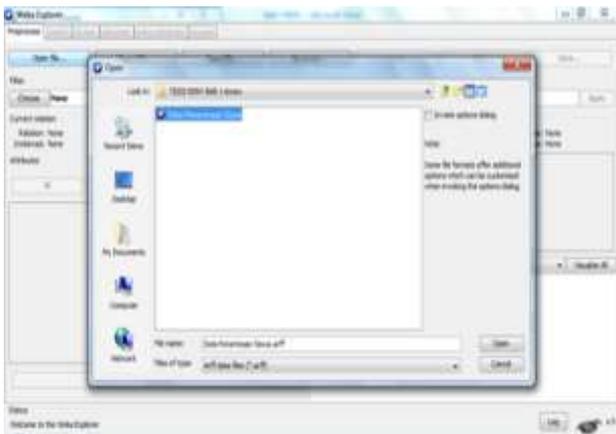
Tampilan Utama WEKA



Gambar 6. Tampilan Utama WEKA

Tampilan Weka Explorer

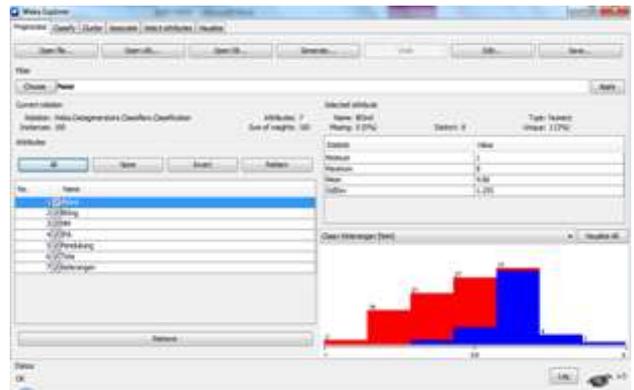
Kemudian klik explorer button dan pilih open file, pilih data test kita “Data penerimaan siswa.arff” sehingga muncul tampilan seperti :



Gambar 7. Tampilan WEKA Explorer

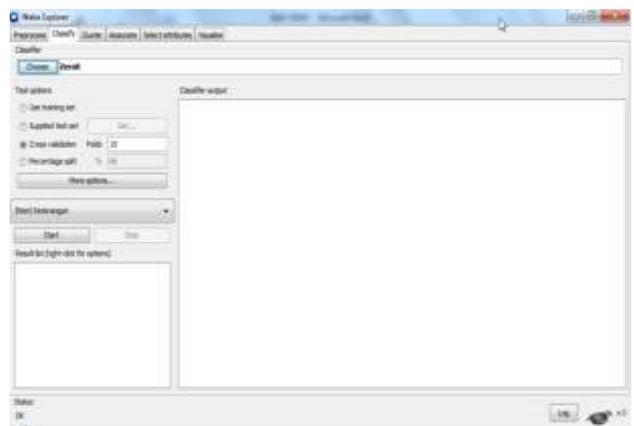
Tampilan Weka Expoler Dengan Visualisasi Data

Visualisasi tiap atribut dapat dilihat dengan meng-klik tombol visualize. Visualisasi ini menggunakan diagram batang, yang mengilustrasikan jumlah dari masing-masing tipe pada atribut yang ada



Gambar 8. Visualisasi Atribut

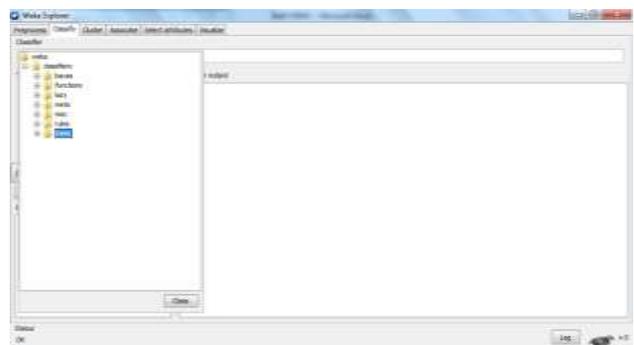
Kemudian pilih Tab Classifiers, klik Tombol choose pada pilihan Classifiers



Gambar 9. Tampilan Menu Tab Classifiers

Tampilan Pilihan Model Classifiers

Pada menu classifiers pilih Root Tree dan klasifikasi dengan pohon keputusan, yaitu metode klasifikasi dengan struktur pohon yang merepresentasikan kriteria pemabagian kelas dan kelas-kelas dengan menggunakan metode algoritma J48



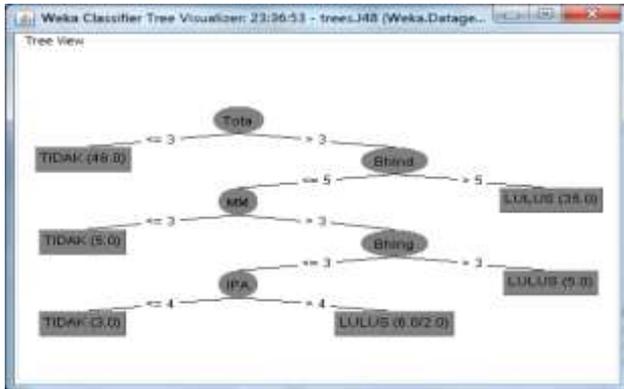
Gambar 7. Tampilan Menu Classifiers

Tersedia pula pilihan pengujian bagi user dibawah bagian pemilihan classifiers yaitu :

1. User Training set .Classifiers dievaluasi pada kemampuannya meprediksi kelas dari instances yang diujikan.

- Supplied test set. Pengujian kemampuan classifiers dilakukan terhadap himpunan instances terpisah yang di-load dari sebuah file.
- Cross-validation. Classifier dievaluasi dengan cross-validation, menurut jumlah folds yang dimasukkan pada kolom folds.
- Percentage split. Evaluasi classifiers dilakukan pada sejumlah persentase tertentu dari data yang digunakan untuk pengujian.

Hasil Visualisasi Tree, adalah hasil akhir dari pohon keputusan



Gambar 8. Tampilan Hasil Visualisasi Tree

Dimana rule yang dihasilkan pada gambar 5.10 adalah :

- IF Nilai Total Rata-Rata ≤ 3 , THEN Keterangan Penerimaan = TIDAK.
- IF Nilai Total Rata-Rata > 3 THEN Keterangan Penerimaan = LULUS.
- IF Nilai Total Rata-Rata > 3 AND Nilai Rata-Rata Bahasa Indonesia > 5 THEN Keterangan Penerimaan = LULUS.
- IF Nilai Total Rata-Rata > 3 AND Nilai Rata-Rata Bahasa Indonesia ≤ 5 THEN Keterangan Penerimaan = TIDAK.
- IF Nilai Total Rata-Rata > 3 AND Nilai Rata-Rata Bahasa Indonesia ≤ 5 AND Nilai Rata-Rata Matematika ≤ 3 THEN Keterangan Penerimaan = TIDAK.
- IF Nilai Total Rata-Rata > 3 AND Nilai Rata-Rata Bahasa Indonesia ≤ 5 AND Nilai Rata-Rata Matematika > 3 THEN Keterangan Penerimaan = LULUS.
- IF Nilai Total Rata-Rata > 3 AND Nilai Rata-Rata Bahasa Indonesia ≤ 5 AND nilai Rata-Rata Matematika > 3 AND Nilai Rata-Rata Bahasa Inggris ≤ 3 THEN Keterangan Penerimaan = TIDAK.
- IF Nilai Total Rata-Rata > 3 AND Nilai Rata-Rata Bahasa Indonesia ≤ 5 AND Nilai Rata-Rata Matematika > 3 AND Nilai Rata-Rata Bahasa Inggris > 3 THEN Keterangan Penerimaan = LULUS.
- IF Nilai Total Rata-Rata > 3 AND Nilai Rata-Rata Bahasa Indonesia ≤ 5 AND Nilai Rata-Rata Matematika > 3 AND Nilai Rata-Rata Bahasa Inggris ≤ 3 AND Nilai Rata-Rata IPA ≤ 4 THEN Keterangan Penerimaan = TIDAK.
- IF Nilai Total Rata-Rata > 3 AND Nilai Rata-Rata Bahasa Indonesia ≤ 5 AND Nilai Rata-Rata Matematika > 3 AND Nilai Rata-Rata Bahasa Inggris ≤ 3 AND Nilai Rata-Rata IPA > 4 , THEN keterangan penerimaan = LULUS.

Setelah dilakukan pengujian pada software WEKA 3.7 untuk memprediksi klasifikasi penerimaan calon siswa seperti yang sudah dijelaskan di atas ternyata hasil rule yang dihasilkan pada WEKA pada gambar 5.10 hampir sama dengan rule yang

dihasilkan pada proses manual knowledge menggunakan pohon keputusan.

Pembahasan Hasil Pengujian

Dari hasil pengujian yang telah dilakukan yaitu proses manual, menggunakan software WEKA 3.7, menggunakan microsoft excel dapat kita ambil sebuah kesimpulan bahwa hasil pengujian sangat baik dan yang dihasilkan hampir sama. Berdasarkan pengujian yang dihasilkan secara umum dapat dilihat bahwa calon siswa baru yang mempunyai nilai yang bagus pada nilai rapor dari semester satu sampai dengan semester 5 pada bidang studi Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, prestasi pendukung, dan nilai total rata-rata akan diterima dalam seleksi penerimaan calon siswa baru, begitu juga sebaliknya. Algoritma C4.5 dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasi dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan maupun aturan IF-then, sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan. Tapi hasil pengujian ini belum dapat dikatakan valid karena pada penelitian ini hanya menggunakan salah satu teknik yaitu teori Decision Tree. Oleh sebab itu perlunya dilakukan penelitian lanjutan untuk membandingkan hasil prediksi dan menentukan teori mana yang menghasilkan prediksi yang baik.

5. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil Penelitian yang penulis lakukan pada SLTA Plus Negeri 7 Bengkulu, maka dapat ditarik kesimpulan bahwa sistem klasifikasi penerimaan siswa baru dengan metoda Data Mining khususnya Algoritma C4.5 akan bermanfaat sekali dalam proses pengambilan keputusan pengolahan data penerimaan siswa baru di lingkungan SMA Plus Negeri 7 kota Bengkulu.

Dari hasil analisa dan pembahasan yang dilakukan pada sistem klasifikasi penerimaan siswa baru dengan metoda Data Mining khususnya Algoritma C4.5 maka dapat ditarik kesimpulan sebagai berikut :

- Dengan adanya sistem klasifikasi penerimaan siswa baru dengan metoda Data Mining Khususnya Algoritma C4.5 maka proses pengolahan data siswa baru akan lebih optimal.
- Dengan menggunakan Metoda Data Mining khususnya Algoritma C4.5 maka kesalahan dan waktu proses pengolahan data penerimaan siswa baru dapat diminimalkan
- Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi *eksekutif* dan sistem digunakan untuk menggambarkan proses yang terkait dengan pengklasifikasian penerimaan calon siswa baru.

Saran

Berdasarkan hasil penelitian yang dilakukan penulis, maka dapat diusulkan beberapa saran untuk penelitian lanjutan sebagai berikut :

- Diharapkan Metoda Data Mining khusus Algoritma C4.5 dapat di terapkan dalam proses pengolahan data penerimaan siswa baru dan proses lainnya yang berhubungan dengan pengambilan keputusan khususnya pada SLTA Plus Negeri 7 Kota Bengkulu.
- Selanjutnya penulis menyarankan agar dapat membandingkan metoda pengambilan keputusan dengan metoda Data Mining dengan metoda yang lainnya.

DAFTAR PUSTAKA

- [1] Santosa Budi, "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis", Yogyakarta : Graha Ilmu, 2007.
- [2] Han, J. Kamber, M & Jian, Pei. Data Mining : Concepts and techniques, Third Edition. America: Morgan Kauffman, San Francisco, 2011.
- [3] Gorunescu, F. (2011). Data Mining : Concepts, Models and Techniques. New York: Springer-Verlag.
- [4] Efraim Turban, dkk (2005). "Decision Support System And Intelephant System", Yogyakarta : Andi Yogyakarta.
- [5] Feri Sulianta. Dan Dominikus Juju (2010). "Data Mining Meramalkan Bisnis Perusahaan". Jakarta Gramedia.
- [6] Kusriani. Dan Emha Taufiq Luthfi (2009). "Algoritma Data Mining". Yogyakarta : Andi Yogyakarta.
- [7] Sujana,(2010). Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi *Decission Tree*. Universitas Widyatama.

BIODATA PENULIS



Penulis Pertama

Reni Kurniah, Dosen Universitas Profesor Dr.Hazairin SH Bengkulu,Program Studi Ilmu Administrasi Publik berlatar belakang S1.Teknik Informatika Universitas Muhammadiyah Bengkulu dan S2.Teknologi Informasi UPI.YPTK Padang.