

Rancangan Aplikasi Prediksi Kelulusan Siswa Menggunakan Algoritma C4.5

Lastri Yuningsih^{1*}, Iwan Rizal Setiawan², Asril Adi Sunarto³

Program Studi Teknik Informatika, Universitas Muhammadiyah Sukabumi
 Jl.R Syamsudin, S.H. No.50 Cikole, Kec. Cikole, Kota Sukabumi Jawa Barat. (0266) 218 345

*Corresponding Author: lastriuningsih@gmail.com

Abstrak

SMK Al-Mizab Sukabumi merupakan lembaga pendidikan sekolah menengah kejuruan yang ada di Jampang Tengah Sukabumi. Sekolah ini memiliki banyak data yang terkait dengan kegiatan akademik, misalnya data kelulusan siswa. Data-data tersebut belum dimanfaatkan semaksimal mungkin, misalnya untuk memprediksi kelulusan siswa, sehingga dapat diambil tindakan untuk memaksimalkan persiapan pelaksanaan ujian akhir. Penelitian ini dilakukan untuk membuat rancangan sistem aplikasi menggunakan teknik klasifikasi yang dapat mengolah data dalam jumlah besar untuk menemukan pola yang terjadi pada data siswa. Pengolahan data tersebut digunakan untuk memprediksi kelulusan siswa. Teknik klasifikasi yang digunakan yaitu decision tree dengan penerepan algoritma C4.5 Inputan yang digunakan yaitu berupa atribut dari data siswa meliputi nilai persemester dari semester 1 sampai 5, dan nilai sikap. Pengujian aplikasi menggunakan data training 104 data siswa yang sudah lulus pada tahun 2016 sampai 2018. Pengetahuan yang diperoleh dari hasil training aplikasi diharapkan dapat dimanfaatkan oleh manajemen SMK Al-Mizab sebagai alat bantu pengambilan keputusan yang lebih efektif dalam rangka perencanaan dan persiapan ujian akhir siswa.

Kata kunci: Rancangan Aplikasi, Prediksi Kelulusan Siswa, Algoritma C4.5, Decision Tree

Abstract

“SMK Al-Mizab Sukabumi” is a vocational high school educational institution in Jampang Tengah Sukabumi. This school has a lot of data related to academic activities, for example student graduation data. These data have not been fully utilized, for example to predict student graduation, so that action can be taken to maximize preparation for the final exam. This research was conducted to design an application system using classification techniques that can process large amounts of data to find patterns that occur in student data. Data processing is used to predict student graduation. The classification technique used is a decision tree with a C4.5 algorithm. The input used is in the form of attributes from student data including the scores per semester from semester 1 to 5, and attitude values. Application testing uses training data of 104 student data who have passed in 2016 to 2018. The knowledge gained from the results of application training is expected to be used by the management of SMK Al-Mizab as a more effective decision-making tool in planning and preparing for student final exams.

Keywords: Application Design, Student Graduation Prediction, C4.5 Algorithm, Decision Tree

1. Pendahuluan

Sekolah menengah kejuruan (SMK) adalah salah satu bentuk satuan pendidikan formal yang menyelenggarakan pendidikan kejuruan pada jenjang pendidikan menengah sebagai lanjutan dari SMP/MTs atau bentuk lain yang sederajat atau lanjutan dari hasil belajar yang diakui sama/setara SMP/MTs. Mengacu pada isi Undang-Undang Sistem Pendidikan Nasional No.20 Tahun 2003 pasal 3 mengenai tujuan pendidikan nasional dan penjelasan pasal 15 yang menyebutkan bahwa pendidikan kejuruan merupakan pendidikan menengah yang mempersiapkan peserta didik terutama untuk bekerja di bidang tertentu.

Data tentang siswa yang lulus merupakan sebuah data yang penting baik bagi departemen atau pun sekolah karena data tersebut digunakan dalam proses akreditasi. Data tentang siswa

yang lulus terus bertambah di tiap tahunnya dan menumpuk seperti data yang terabaikan karena jarang digunakan, padahal data tentang siswa yang lulus dapat memberikan informasi yang berguna jika dimanfaatkan dengan maksimal.

Kelulusan merupakan hasil dari suatu keputusan. Siswa dinyatakan lulus tidaknya merupakan sebuah proses dalam pengambilan keputusan. Seperti halnya yang dilakukan di SMK Al-Mizab Sukabumi ini merupakan salah satu sekolah swasta yang berada di Kecamatan Jampang Tengah Sukabumi. SMK ini berdiri pada tahun 2010 berdasarkan surat ijin operasional menteri pendidikan nasional dengan nomor: 421.5/KEP.1373/DISDIK sekolah ini mempunyai 13 ruang kelas, 1 ruang perpustakaan, lab komputer. Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang pengambilan keputusan tidak cukup hanya mengandalkan data operasional saja, tetapi diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada, seperti informasi untuk mendukung pengambilan suatu keputusan yang terkait dengan prediksi kelulusan siswa. Selama ini sekolah menengah kejuruan di SMK Al-Mizab ini guru-guru dan kepala sekolah selalu mengadakan rapat terkait persiapan pelaksanaan ujian akhir siswa, agar siswa dapat mempersiapkan diri sebaik mungkin dalam mengikuti ujian akhir tersebut.

Seiring dengan perkembangan teknologi, penggalian pola dapat dilakukan dengan teknik klasifikasi data yang didapat dari proses data mining. Data mining biasanya dilakukan karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan pola dan knowledge yang berguna untuk proses bisnis suatu perusahaan. Salah satu teknik yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak ada dalam basis data yang tersimpan. Dalam kasus ini model klasifikasi dibuat untuk mengidentifikasi pola data untuk kelas status "lulus" atau "tidak lulus", dari hasil penentuan pola training data. Pola atau model dari training data tersebut selanjutnya diuji dengan menggunakan test set data. Singkatnya Model klasifikasi dibuat dengan cara menganalisis training data (terdiri dari variable variabel yang kelasnya sudah diketahui). Model yang dihasilkan kemudian akan digunakan untuk memprediksi kelas dari *unknown* data (variable-variabel yang kelasnya belum diketahui), test set data digunakan untuk pengujian dari model yang telah didapatkan pada training data. Model klasifikasi yang digunakan dalam kasus ini adalah algoritma C4.5. Algoritma C4.5 yaitu untuk memprediksi atau mengklasifikasi suatu kejadian dengan pembentukan pohon keputusan [1].

Algoritma *Decision Trees* adalah salah satu penerapan *data mining* model klasifikasi. Algoritma ini mengonstruksi pohon keputusan dari sebuah data training yang berupa *record-record* dalam basis data. Algoritma C4.5 merupakan salah satu algoritma klasifikasi data dengan teknik pohon keputusan yang dapat mengolah data numerik (*kontinyu*) dan diskrit, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah untuk diinterpretasikan. Algoritma C4.5 telah diujicoba dalam banyak kasus klasifikasi, misalnya pada bidang medis [2][3], bidang perdagangan [4][5], bidang kepegawaian [6][7], dan berbagai bidang lainnya.

Paper ini menyajikan model Aplikasi untuk memprediksi kelulusan siswa menggunakan Algoritma C4.5 Berbasis

2. Tinjauan Pustaka

Kelulusan merupakan salah satu hal yang menjadi perhatian bagi pengelola Perguruan Tinggi, baik Perguruan Tinggi Negeri maupun di Perguruan Tinggi swasta. Kelulusan merupakan salah satu item penilaian dalam proses akreditasi institusi suatu perguruan tinggi. Jika mahasiswa lulusnya tepat waktu, akan membantu penilaian akreditasi suatu perguruan tinggi. Di sisi lain, mahasiswa akan beruntung jika lulus tepat waktu karena akan efisien dalam hal membayar biaya kuliah. Namun waktu kelulusan mahasiswa tidak selalu dapat dideteksi secara dini, sehingga bisa mengakibatkan keterlambatan lulusan. Untuk mengatasi hal tersebut perlu ada teknik untuk bisa melakukan prediksi terhadap kelulusan mahasiswa dengan tepat.

Penelitian mengenai prediksi kelulusan telah banyak dilakukan. Prediksi kelulusan dengan menggunakan metode *Neural Network* dan *Particle Swarm Optimization* telah diteliti oleh [8] dan [9]. Sebuah *Neural Network* dikonfigurasi untuk aplikasi tertentu, seperti pengenalan pola atau klasifikasi data, dan kemudian disempurnakan melalui proses pembelajaran. Hakim [10] dan Rahmatullah [11] telah menguji algoritma *K-Nearest Neighbor* (K-NN) dalam memprediksi kelulusan siswa. Prinsip kerja algoritma K-NN adalah melakukan pencarian jarak paling dekat antara data baru dengan sejumlah k tetangga terdekat dari data latih. Selanjutnya, Sumpena [12] menggunakan algoritma *Naïve Bayes* untuk memprediksi Kelulusan Mahasiswa PKBM Paket C. *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa

sebelumnya. Naïve Bayes Classifier ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

Paper ini menyajikan Algoritma C4.5 untuk memprediksi kelulusan siswa. Algoritma C4.5 merupakan algoritma klasifikasi data dengan teknik pohon keputusan yang dapat mengolah data numerik (kontinyu) dan diskrit, dapat menangani nilai atribut yang hilang, dan menghasilkan aturan-aturan yang mudah untuk diinterpretasikan

3. Metodologi

Model Algoritma C4.5 ditunjukkan sebagai berikut:

- 1) Menyiapkan data training. Data ini diambil dari data yang sudah pernah ada sebelumnya dan sudah dikelompokkan kedalam kelas tertentu.
- 2) Setelah itu tentukan akar dari pohon. Pilih akar dari atribut, cara adalah dengan menghitung nilai gain dari semua atribut, yang menjadi akar pertama adalah nilai gain yang paling. Sebelum menentukan nilai gain, terlebih dahulu hitung nilai entropy. Untuk menentukan nilai entropy gunakan rumus:

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

Keterangan: S = himpunan kasus n = jumlah partisi S pi = proporsi Si terhadap S

- 3) Tentukan nilai gain menggunakan rumus:

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

Keterangan :

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A

| Si | = Jumlah Kasus pada partisi ke-i

| S | = Jumlah Kasus dalam S

- 4) Setelah itu ulangilah langkah ke-2 sampai semua record terpartisi secara sempurna.
- 5) Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua record dalam simpul N mendapat kelas yang sama
 - b. Tidak ada atribut di dalam record yang dipartisi lagi
 - c. Tidak ada record di dalam cabang yang kosong.

Metode penelitian yang dilakukan oleh penulis yaitu menggunakan algoritma C4.5. Data training yang digunakan yaitu 104 data siswa tahun 2018-2019. Parameter yang digunakan nantinya yaitu nilai semester 1-5 dan nilai sikap untuk menentukan kelulusan siswa. Algoritma C4.5 merupakan algoritma klasifikasi data dengan teknik pohon keputusan yang dapat mengolah data numerik (*kontinyu*) dan diskrit, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah untuk diinterpretasikan. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Adapun tahapan penelitian yang dilakukan penulis yaitu menggunakan siklus SDLC (*Software Development Life Cycle*), dengan susunan berikut:

1. Perencanaan Sistem
 Dalam perencanaan sistem peneliti menentukan apa saja yang akan dikerjakan, algoritma apa yang digunakan dalam penelitian serta penentuan judul yang akan diajukan.
2. Pengumpulan Data
 Pengumpulan data yang dilakukan oleh peneliti yaitu dengan cara observasi ke SMK Al-Mizab, melakukan wawancara dan studi pustaka.
3. Analisis Sistem
 Menganalisis sistem yang sedang berjalan di SMK Al-Mizab dan analisis perancangan.
4. Perancangan Sistem
 Perancangan sistem yang akan dibuat, perancangan basis data.
5. Implementasi Sistem
 Implementasi yang digunakan yaitu menggunakan bahasa pemrograman PHP dan MySQL sebagai basis data.
6. Pengujian Sistem
 Pengujian sistem yang akan dilakukan nantinya menggunakan *black-box testing*.

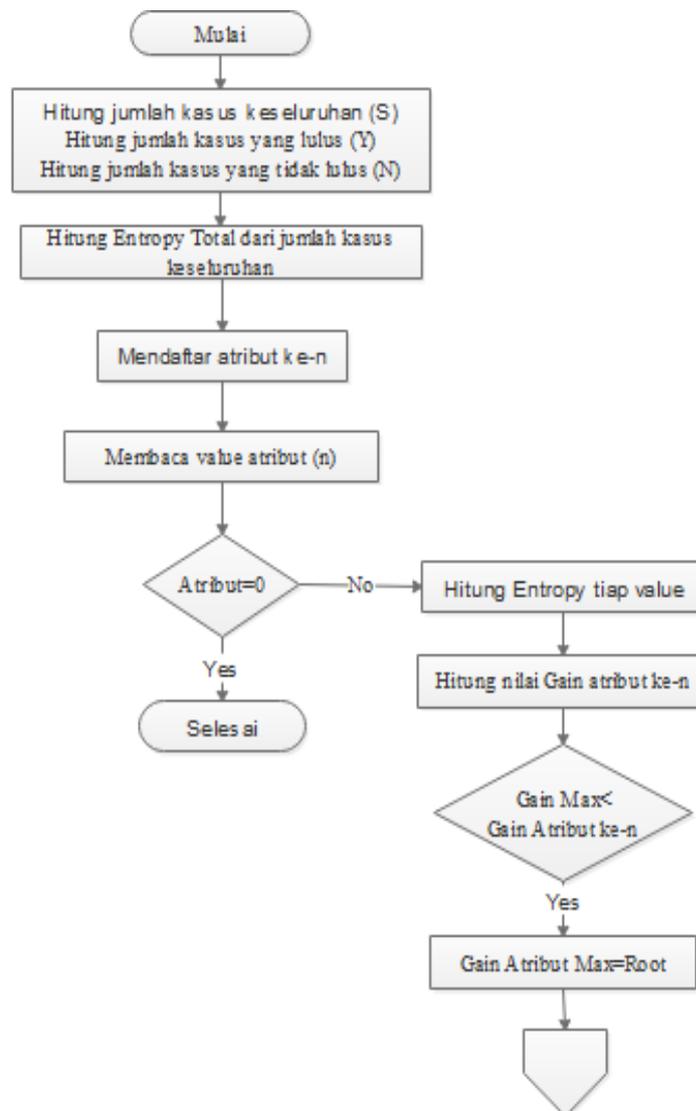
4. Hasil dan Pembahasan

Data yang diinput dari sistem yang akan dibangun oleh penulis yaitu berupa data siswa lama dalam bentuk tabel berisi beberapa atribut meliputi nilai semester 1-5 dan nilai sikap. Data siswa tersebut akan diklasifikasikan berdasarkan target yang akan ditentukan dan dihitung menggunakan metode *Decision Tree* yaitu algoritma C4.5 untuk mencari nilai *entropy*, nilai *gain*, nilai *splitinfo*, nilai *gainratio*, dan proses pembentukan node akar. Setelah proses perhitungan selesai maka akan menghasilkan rule yang digunakan dalam penentuan keputusan pada proses prediksi.

Sistem yang akan dibangun memiliki kemampuan sebagai berikut:

1. Melakukan perhitungan terhadap data yang telah diperoleh dari data sampel training.
2. Melakukan perhitungan data untuk menentukan nilai *entropy*, *gain*, *splitinfo*, dan *gainratio*.
3. Memberikan informasi dalam menentukan target yang akan diprediksi, yaitu lulus atau tidaknya setiap siswa.

Di bawah ini merupakan flowchart pembentukan pohon keputusan yang berjalan pada sistem:

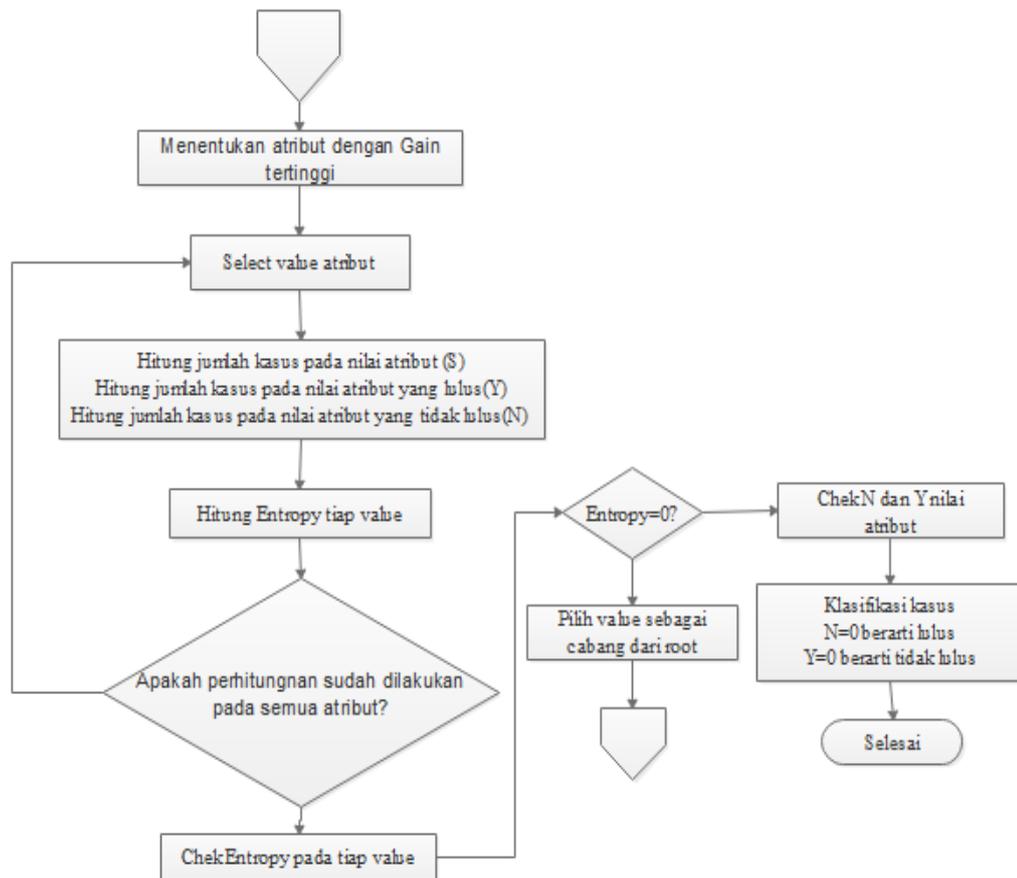


Gambar 1. Flowchart Penentuan Root

Berikut sistem yang berjalan pada saat penentuan akar(root):

- a. Sistem akan menghitung total nilai informasi dari data trainingnya
- b. Sistem akan mendaftar atribut A
- c. Tiap-tiap atribut dihitung nilai *entropy* dan *gain*

- d. Sistem akan membandingkan nilai *gain* terbesar dari tiap atribut
- e. Setelah nilai *gain* terbesar didapat, maka sistem akan memilih atribut dengan *gain* terbesar sebagai atribut terbaik untuk dijadikan akar
- f. Sistem akan melakukan proses ini sampai semua atribut dalam daftar habis hitung



Gambar 2. Flowchart Penentuan Node

Berikut sistem yang berjalan saat penentuan node pada pembentukan *Tree*:

- a. Sistem akan menghitung jumlah kasus pada atribut dengan *gain* tertinggi
- b. Menghitung nilai *entropy* total dari atribut dengan *gain* tertinggi
- c. Mendaftar atribut A selain atribut dari *gain* tertinggi
- d. Menghitung *entropy* dari masing-masing nilai atribut ke-n hingga habis semua
- e. Menghitung nilai *gain* dari masing-masing atribut
- f. Menentukan *gain* maksimal sebagai penentusan node selanjutnya
- g. Setelah node selanjutnya terpilih maka proses perhitungan akan berulang lagi mulai dari penentuan cabang dan penentuan node hingga semua atribut habis dieksekusi dan mencapai end of *Tree*

Pada langkah awal pada usulan sistem akan ditentukan beberapa atribut yang digunakan sebagai parameter dalam pengklasifikasian data sampel. Atribut menyatakan suatu parameter yang dibuat sehingga kriteria dalam pembentukan pohon keputusan.

Berikut langkah-langkah perhitungannya:

1. Data dikelompokkan berdasarkan atribut beserta nilai didalamnya
2. Menghitung jumlah data pada tiap nilai atribut yang ada
3. Mengklasifikasi data yang sudah dihitung menjadi dua kelompok berdasarkan target tujuan yaitu Lulus (Y) dan Tidak Lulus (T)
4. Menghitung nilai *entropy* total 104 data siswa

5. Menghitung nilai *entropy* dari masing-masing nilai atribut
6. Menghitung nilai *gain* dari tiap atribut
7. Mencari atribut dengan *gain* tertinggi untuk dijadikan root
8. Menentukan nilai atribut yang akan dijadikan cabang
9. Menentukan nilai node cabang selanjutnya dari atribut yang terpilih berdasarkan nilai *gain* tertinggi

Berikut ini adalah mekanisme algoritma C4.5 pada pembentukan *decision tree*:

Uji coba menggunakan menggunakan 104 data siswa yang telah lulus di tahun 2018. Contoh data yang digunakan disajikan pada tabel 1.

Tabel 1. Himpunan data Siswa dengan 6 atribut kategorial (nilai semester 1-5, nilai sikap).

No	NIS	DATA POKOK						DATA TAMBAHAN		
		SEMESTER						tahun masuk	tahun lulus	Ket
		1	2	3	4	5	Nilai Sikap			
1	161701001	74	76	77	77	79	B	2015	2018	Lulus
2	161701002	79	78	78	78	81	B	2015	2018	Lulus
3	161701003	74	76	77	76	77	B	2015	2018	Tidak Lulus
4	161701004	70	76	77	76	78	B	2015	2018	Tidak Lulus
5	161701005	74	75	77	76	77	B	2015	2018	Tidak Lulus
104	161701110	78	78	80	80	82	B	2015	2018	Lulus

1) Rumus mencari nilai *entropy*

Data dikelompokkan berdasarkan atribut dan nilai atributnya lalu dihitung jumlah keseluruhan, jumlah siswa yang lulus dan yang tidak, lalu dihitung nilai *entropy* dan *gain* dimasing-masing atributnya.

Baris total kolom *entropy* pada tabel di atas dihitung dengan rumus adalah sebagai berikut:

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

Keterangan : S = himpunan kasus

n = jumlah partisi

Pi = proporsi dari Si terhadap S

$$\text{Entropy (Total)} = \left(-\frac{99}{104} * \log_2 \left(\frac{99}{104} \right) + \left(-\frac{19}{104} * \log_2 \left(\frac{19}{104} \right) \right) = 0,5673$$

Jadi, dari data yang telah diambil memiliki nilai *entropy* sebesar 0,5673.

Perhitungan total kolom *entropy* pada tiap nilai atribut dihitung dengan cara yang sama dengan *entropy* total:

$$\text{Entropy}(\text{Smstr1,A}) = \left(-\frac{4}{4} * \log_2 \left(\frac{4}{4} \right) + \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4} \right) \right) = \text{galat matematika}/0$$

$$\text{Entropy}(\text{Smstr1,B}) = \left(-\frac{62}{64} * \log_2 \left(\frac{62}{64} \right) + \left(-\frac{2}{64} * \log_2 \left(\frac{2}{64} \right) \right) = 0,2006$$

$$\text{Entropy}(\text{Smstr1,C}) = \left(-\frac{32}{32} * \log_2 \left(\frac{32}{32} \right) + \left(-\frac{0}{32} * \log_2 \left(\frac{0}{32} \right) \right) = 0,9544$$

$$\text{Entropy}(\text{Smstr1,D}) = \left(-\frac{0}{4} * \log_2 \left(\frac{0}{4} \right) + \left(-\frac{4}{4} * \log_2 \left(\frac{4}{4} \right) \right) = \text{galat matematika} \Rightarrow 0$$

$$\left(\frac{4}{4} \right) = \text{galat matematika} \Rightarrow 0$$

Jadi dari data diatas dapat diambil kesimpulan bahwa nilai *entropy* dari semester 1 yang bernilai A menghasilkan galat matematika, nilai atribut B menghasilkan 0,2006, nilai atribut C menghasilkan 0,9544, nilai atribut D bernilai galat matematika.

2) Rumus mencari nilai *Gain*

Sementara itu nilai *gain* pada baris nilai semester 1 dihitung dengan menggunakan rumus *gain*, sebagai berikut:

$$Gain(S,A)=Entropy(s)\sum_{i=1}^n \frac{S_i}{s} * entropy(S_i)$$

Dengan:

- S : Himpunan Kasus
- A : Atribut
- N : Jumlah partisi A
- |S_i| : Jumlah kasus pada partisi ke 1
- |S| : Jumlah kasus dalam S

$$Gain(Total,Smstr1)=Entropy(Total)-\sum_{i=1}^n \frac{smstr1}{total} * entropy(smstr1)$$

$$Gain(Total,Smstr1) = 0,567-(\frac{99}{104}*0) + (\frac{88}{104} * 0,267)+ (\frac{13}{104} * 0,391) =0,985$$

Jadi, nilai *gain* dari data diatas berjumlah 0,985.

3) Rumus mencari nilai SplitInfo

$$SplitInfo(S,A) = -\sum_{i=1}^n \frac{S_i}{s} * log_2 \frac{|s_i|}{|s|}$$

$$NilaiSplitInfo=(-\frac{3}{104}*log_2(\frac{3}{104})+(-\frac{88}{104}*log_2(\frac{88}{104}) + (\frac{13}{104}*log_2(\frac{13}{104})) =0,727$$

Jadi, nilai *SplitInfo* dari data diatas berjumlah 0,727.

4) Rumus mencari nilai GainRatio

$$GainRatio = \frac{Gain(S,A)}{SplitInfo(S,A)}$$

$$Nilai GainRatio(Total,smstr1) = \frac{0,985}{1,316} = 0,76$$

Jadi, nilai *gainratio* dari data diatas bernilai 0,76

Lakukan diskritisasi data bernilai numerik pada tabel 2 ke dalam dua nilai (*binary, split*). Pertama,nilai-nilai numerik didalam suatu atribut diambil yang unik(duplikasi dihilangkan) lalu diurutkan dari yang terkecil ke besar (*ascending*).

Selanjutnya memilih nilai batas yang menghasilkan partisi terbaik berdasarkan ukuran ke tidakmurnian *gainratio*. Untuk atribut semester 1, semester 2, semester 3, semester 4, semester 5, dan Nilai sikap, kita mendapatkan nilai batas terbaik secara berturut.turut.

Tabel 2
Tabel Hasil *Entropy, Gain, SplitInfo, dan GainRatio*

Atribut	Nilai Atribut	S	Y	N	Entropy	Gain	Split Info	Gain Ratio
Total	Total	104	95	19	0,5673			
SMSTR 1								
	A	4	4	0	0	0,985	1,316	0,758
	B	64	62	2	0,2006	0,985	1,316	0,748
	C	32	20	12	0,954	0,985	1,316	0,748
	D	4	0	4	0	0,985	1,316	0,748
SMSTR 2								
	A	8	8	0	0	1,057	0,891	1,185
	B	87	76	11	0,547	1,057	0,891	1,185
	C	4	1	3	0,811	1,057	0,891	1,185
	D	5	0	5	0	1,057	0,891	1,185
SMSTR 3								
	A	29	29	0	0	1,117	0	0

Atribut	Nilai Atribut	S	Y	N	Entropy	Gain	Split Info	Gain Ratio
Total	Total	104	95	19	0,5673			
	B	73	56	17	0,783	1,117	0	0
	C	0	0	0	0	1,117	0	0
	D	2	0	2	0	1,117	0	0
SMSTR 4								
	A	23	23	0	0	1,12	0	0
	B	79	63	16	0,727	1,12	0	0
	C	0	0	0	0	1,12	0	0
	D	2	0	2	0	1,12	0	0
SMSTR 5								
	A	57	57	0	0	1,0012	0	0
	B	47	29	18	0,96	1,0012	0	0
	C	0	0	0	0	1,0012	0	0
	D	0	0	0	0	1,0012	0	0
NILAI SIKAP								
	A	2	2	0	0	1,225	0	0
	B	89	74	15	0,654	1,225	0	0
	C	13	10	3	0,779	1,225	0	0
	D	0	0	0	0	1,225	0	0

Dari hasil tabel 2 menunjukkan bahwa atribut smstr 2 dengan *gainratio* paling besar diantara 6 atribut yang ada, menjadi atribut pemilah terbaik (*thebest split attribute*) untuk diletakan sebagai *root*.

Pemanfaatan algoritma dalam prediksi kelulusan siswa ini dilakukan terlebih dahulu menghitung data training untuk dijadikan pola prediksi kedepannya. Dalam penelitian ini menggunakan data training 104 sampel data siswa lama ditahun 2018-2019, seperti yang tertera pada gambar dibawah ini:

Dari dataset yang telah di *inputkan* maka sistem akan menemukan pola parameter mana yang menjadi atribut pemilah terbaik. Halaman ini merupakan form inputan data siswa atau data sampel training untuk prediksi yang meliputi beberapa atribut yaitu id, nis, nama siswa, semester 1, semester 2, semester 3, semester 4, semester 5, sikap, tahun masuk, keterangan dan aksi. Pada form ini atribut dari masing-masing siswa telah dikelompokan dalam bentuk tabel, lengkap dengan keterangan jumlah siswa yang lulus dan tidak lulus.

Dataset Nilai Siswa
Berikut ini adalah contoh dataset yang digunakan sebagai bahan pembelajaran

Buat baru

Copy Print Excel PDF

ID	NIS	Nama Siswa	SM1	SM2	SM3	SM4	SM5	Sikap	Tahun Masuk	Keterangan	Aksi
1	161701110	Yanti	A	A	A	B	A	A	2015	L	
2	161701109	Suzi Sunilawati	B	B	B	B	B	A	2015	L	
3	161701108	Supriatna Nurkafi	C	C	B	B	B	A	2015	L	
4	161701107	Septian Dwi Putra	B	B	B	B	B	A	2015	L	
5	161701106	Sahnul Gunawan	C	C	B	B	B	B	2015	T	
6	161701105	Riza Ruswandi	B	B	B	B	A	A	2015	L	
7	161701104	Reza Jakaria	C	B	B	B	B	A	2015	L	
8	161701103	Priatna	B	D	B	B	B	B	2015	T	
9	161701101	Nurjaman Abdumohman	B	B	B	B	A	A	2015	L	

Gambar 3. Dataset 104 Siswa lulusan 2018-2019

Untuk Proses perhitungannya C4.5 bisa dilihat pada gambar 4.

Perhitungan

Perhitungan algoritma C 45
Berikut ini merupakan detail perhitungan algoritma C 45

Lakukan Perhitungan

#Step 1
Atribut (SM1)

Kriteria	Hasil			Entropi
	Total Kasus	L	T	
C	39	26	13	0.91829583405449
B	59	58	1	0.12394996771101
D	5	0	5	0
A	1	1	0	0

Gain : 0.27125287150448 Split info : 1.2895140746564 Gain Ratio : 0.21388680089505

Gambar 4. Perhitungan C4.5

Perhitungan C4.5 di dapat dari dataset atribut yang digunakan yaitu nilai semester 1-5 dan nilai sikap. Halaman perhitungan ini merupakan halaman untuk memprediksi kelulusan siswa. Proses prediksi dilakukan secara otomatis di ambil dari dataset yang telah diinput. Halaman ini hanya menampilkan hasil perhitungan C4.5 yaitu nilai *entropy*, *gain*, *splitinfo* dan *gainratio*. keterangan lulus dan tidaknya ada di halaman dataset dan data siswa. Dalam kasus ini atribut indeks prestasi *gain* tertinggi yaitu di semester 2.

The screenshot shows the C4.5 web application interface. On the left is a navigation menu with options: Dashboard, Data Siswa, Dataset, Perhitungan, Rule, and Logout. The main content area displays a table of student data with columns: ID, NIS, Nama Siswa, SM1, SM2, SM3, SM4, SM5, Sikap, Tahun Masuk, Keterangan, and Aksi. Below the table is a summary table with the following data:

KETERANGAN	Jumlah
LULUS	19
TIDAK LULUS	85
Jumlah Total Siswa	104

At the bottom of the interface, it shows 'Showing 101 to 104 of 104 entries' and a pagination control with 'Previous', '1', '...', '7', '8', '9', '10', '11', and 'Next'.

Gambar 5. Hasil Lulus dan Tidak Lulus

Pada gambar 5 diatas terdapat keterangan jumlah siswa yang lulus dan tidak lulus dengan jumlah total 104 siswa, terdapat 11 halaman dan untuk setiap halaman hanya berjumlah 10 siswa. Contoh gambar diatas adalah halaman ke 11.

The screenshot shows the 'Pohon Keputusan / Tree Decision' page in the C4.5 web application. The page title is 'Pohon Keputusan / Tree Decision'. Below the title, it says 'Pohon keputusan yang terbuat dari implementasi algoritma C.45 pada data learning sebagai berikut :'. The rules are listed as follows:

- Periksa atribut SM2, Jika Nilai SM2 bernilai D Maka dinyatakan TIDAK LULUS ,Nilai SM2 bernilai A Maka dinyatakan LULUS
- Jika tidak Periksa atribut SM1, Jika Nilai SM1 bernilai B Maka dinyatakan LULUS ,Nilai SM1 bernilai D Maka dinyatakan TIDAK LULUS ,Nilai SM1 bernilai A Maka dinyatakan LULUS
- Jika tidak Periksa atribut nilaiSikap, Jika Nilai nilaiSikap bernilai A Maka dinyatakan LULUS
- Jika tidak Periksa atribut SM3, Jika Nilai SM3 bernilai D Maka dinyatakan TIDAK LULUS ,Nilai SM3 bernilai A Maka dinyatakan LULUS ,Nilai SM3 bernilai C Maka dinyatakan TIDAK LULUS

Gambar 6. Rule yang dihasilkan

Halaman rule ini merupakan halaman yang berisi kondisi dari data atribut yang dijadikan aturan dalam penentuan keputusan dalam prediksi kelulusan. Aturan tersebut diperoleh berdasarkan perhitungan *entropy*, *gain*, *splitinfo* dan *gainratio*. A

5. Kesimpulan

Dari hasil uji coba menggunakan 104 data sampel dari siswa lulusan SMKAI-Mizab tahun 2018/2019, indeks prestasi *gain* tertinggi yaitu di semester 2 dengan nilai gain 0.31667. Aplikasi sistem pendukung keputusan yang menerapkan Algoritma C4.5 untuk memudahkan dalam memprediksi kelulusan siswa, yaitu dengan menggunakan *Rule* yang dihasilkan dari aplikasi. Hasil Sistem ini hanya menjadi alat bantu bagi pengambil keputusan, keputusan akhir tetap berada ditangan pengambil keputusan.

Walaupun telah menghasilkan rule yang dapat menjadi pedoman dalam mengambil suatu keputusan, rule yang dihasilkan tersebut masih perlu pengujian faktual untuk mengetahui tingkat akurasi.

6. Ucapan Terimakasih

Ucapan terimakasih penulis sampaikan kepada:

1. Kedua orang tua atas doa dan nasihat.
2. Bapak Iwan Rizal S selaku pembimbing I yang selalu memberikan bimbingan serta masukan dalam proses penelitian ini.
3. Bapak Asril Adi S selaku pembimbing II yang senantiasa meluangkan waktunya dalam mengarahkan dan memberi semangat.
4. Semua pihak yang turut membantu dan memberi dukungan baik berupa nasihat, masukan, tenaga, usaha dan waktunya.

DAFTAR REFERENSI

- [1] Suntoro, J. *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: Elex Media Komputindo, 2019
- [2] Mujahidin, A., & Pribadi, D. (2017). Penerapan Algoritma C4. 5 Untuk Diagnosa Penyakit Pneumonia Pada Anak Balita Berbasis Mobile. *Jurnal Swabumi*, 5(2), 155-161.
- [3] Fiandra, Y. A., Defit, S., & Yuhandri, Y. Penerapan Algoritma C4. 5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10). *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2017; 1(2): 82-89
- [4] Pritalia, G. L. Penerapan Algoritma C4. 5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems*, 2018; 1(1): 47-56.
- [5] Arifin, M. F., & Fitriana, D. Penerapan Algoritma Klasifikasi C4. 5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada. *InComTech*, 2018; 8(2): 87-102.
- [6] Sunarti, S. Prediksi Promosi Jabatan Karyawan Dengan Algoritma C4. 5 (Studi Kasus: Apartemen Senayan Jakarta). *Techno. Com*, 2019; 18(4): 288-298.
- [7] Taufiq, T., & Yudihartanti, Y. PENERAPAN ALGORITMA C4. 5 KLASIFIKASI TINGKAT KELULUSAN MAHASISWA. *Soliter*, 2, 153-162, 2020
- [8] Kusumawati, D., Winarno, W. W., & Arief, M. R. Prediksi Kelulusan Mahasiswa Menggunakan Metode Neural Network dan Particle Swarm Optimization. *Semnasteknomedia Online*, 2015; 3(1): 3-8
- [9] Rohmawan, E. P. Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network. *Jurnal Ilmiah Matrik*, 2018; 20(1): 21-30.
- [10] Hakim, L. A. R., Rizal, A. A., & Ratnasari, D. (2019). Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN). *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 2019; 1(1): 30-36.
- [11] Rahmatullah, S. Prediksi Tingkat Kelulusan Tepat Waktu dengan Metode Naive Bayes dan K-Nearest Neighbor. *Jurnal Informasi Dan Komputer*, 2019; 7(1): 7-16.
- [12] Sumpena, J., & Kurnia, N. (2019). Analisis Prediksi Kelulusan Siswa PKBM Paket C Dengan Metoda Algoritma Naive Bayes. *Jurnal TEDC*, 2019; 13(2): 127-133.