

Implementasi Decision Tree Algoritma C4.5 Untuk Memprediksi Kesuksesan Pendidikan Karakter

M.A.Abdillah¹, Arief Setyanto², Sudarmawan^{3*}

Magister Teknik Informatika Universitas Amikom Yogyakarta

Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta 55283 INDONESIA

¹Abdillah.1989@students.amikom.ac.id, ²arief_s@amikom.ac.id, ³sudarmawan@amikom.ac.id

INTISARI

Pencanangan kurikulum Pendidikan Karakter dalam sistem pendidikan di Indonesia, adalah sesuatu hal yang baru dan belum banyak dikaji dalam penelitian Educational Data Mining (EDM). Sebagian besar penelitian dalam ranah EDM masih menggunakan faktor kognitif dalam penilaiannya, berbeda dengan pendidikan karakter yang lebih berorientasi kepada pengajaran nilai-nilai karakter, serta mempertimbangkan latar belakang peserta didik. Oleh karena itu, diperlukan cara atau metode untuk mengidentifikasi calon peserta didik, serta memprediksi kesuksesannya dalam sistem pendidikan karakter. Algoritma C4.5 dapat digunakan untuk melakukan prediksi dan klasifikasi terhadap calon siswa dengan cara membuat pohon keputusan berdasarkan data-data yang sudah ada dan melakukan prediksi terhadap calon siswa baru, dalam penelitian ini peneliti menggunakan data mahasiswa Unires Yogyakarta sebagai objek penelitian. Dengan penelitian ini, juga diharapkan dapat diketahui tingkat akurasi Decision Tree Algoritma C.45 dalam mengukur pengaruh atribut-atribut latar belakang siswa tersebut terhadap kesuksesan pendidikan karakter, sehingga akan diketahui apakah Decision Tree Algoritma C.45 memenuhi aspek reliabilitas dan validitas sebagai alat ukur kesuksesan pendidikan karakter. Dari hasil pengukuran, diketahui bahwa kombinasi atribut Bidang Bahasa dan Sosial, Latar belakang pendidikan agama dan orang tua yang menjadi seorang pendidik/guru, serta kemampuan untuk membaca Al-Qur'an berkorelasi positif terhadap kesuksesan pendidikan karakter. Nilai accuracy sebesar 60,91%, menunjukkan bahwa algoritma decision tree C4.5 layak digunakan untuk melakukan prediksi tingkat kesuksesan pada pendidikan karakter. Namun masih butuh banyak kajian mendalam terutama dalam menentukan atribut-atribut yang memang benar-benar mempengaruhi kesuksesan pendidikan karakter sehingga diharapkan akan didapat accuracy yang lebih baik, serta terjadi efisiensi dalam pengelompokan atribut.

Kata kunci— data mining, klasifikasi, Decision Tree ,Algoritma C4.5, prediksi kelulusan, pendidikan karakter.

ABSTRACT

The launching of the Character Education curriculum in the education system in Indonesia, is something new and has not been much studied in Educational Data Mining (EDM) research. Most of the research in the realm of EDM still uses cognitive factors in its assessment, in contrast to character education which is more oriented towards teaching character values, and taking into account the background of students. Therefore, we need a way or method to identify potential students, and predict their success in the character education system. C4.5 algorithm can be used to make predictions and classifications of prospective students by making a decision tree based on existing data and predicting new prospective students, in this study the researchers used the data of Yogyakarta Unires students as research objects. With this research, it is also expected to know the accuracy of the Decision Tree Algorithm C.45 in measuring the influence of the background attributes of these students on the success of character education, so it will be known whether the Decision Tree C.45 algorithm meets the aspects of reliability and validity as a measure of success character building. From the measurement results, it is known that the combination of attributes of Language and Social Affairs, the background of religious education and parents who become educators/teachers, and the ability to read Al-Qur'an are positively correlated to the success of character education. Accuracy value of 60.91%, shows that the decision tree C4.5 algorithm is feasible to uses to predict success rates in character education. But it still needs a lot of in-depth study, especially in determining the attributes that really affect the success of character education so that it is expected to get better accuracy, and efficiency in the grouping of attributes.

Keywords - data mining, classification, Decision Tree, C4.5 algorithm5, graduation prediction, character education.

I. PENDAHULUAN

Menurut [1] konsep data mining dan metodenya dapat diaplikasikan pada berbagai bidang seperti marketing, kesehatan, real estate, customer relationship management, engineering, web mining, dll. Salah satu implementasi teknik data mining yang sedang marak adalah pada data dalam bidang pendidikan, sehingga muncul konsep *Educational data mining* (EDM). Educational Data Mining digunakan untuk mempelajari data yang tersedia dalam database pendidikan dan memunculkan informasi tersembunyi di dalamnya, dimana salah satu implementasinya adalah untuk memprediksi performa siswa.

Beberapa penelitian dalam ranah Educational Data Mining ini, diantaranya adalah penelitian [2] yang dilakukan berdasar statistik yang menunjukkan bahwa tingkat pendidikan di Portugal berada di peringkat terbawah, disebabkan banyak siswa yang putus sekolah. Faktor eksternal berpengaruh pada kegagalan siswa dalam menyelesaikan bidang studi khususnya bidang studi matematika. Hasil pengujian dan analisis dalam penelitian tersebut menunjukkan bahwa Algoritma Decision Tree C.45 akurat diterapkan untuk memprediksi nilai akhir siswa sekolah menengah dengan tingkat akurasi 60%.

[3] dalam jurnalnya melakukan proses penelitian untuk memprediksi tingkat kelulusan siswa SMPN 1 Kintamani dengan menggunakan metode decision tree. Data yang diolah pada penelitian ini adalah data siswa dengan menggunakan kriteria nilai ujian harian, nilai ujian tengah semester, nilai ujian akhir semester dan nilai ujian nasional. Penerapan data mining digunakan untuk memprediksi tingkat kelulusan dengan menggunakan metode decision tree dan algoritma C45 sebagai pendukung sekaligus perbandingan untuk mengetahui informasi tingkat kelulusan siswa SMPN1 Kintamani berdasarkan kedua metode tersebut. Berdasarkan pohon keputusan yang dibuat peneliti diketahui bahwa ujian nasional sebagai root pohon merupakan kriteria yang paling menentukan hasil prediksi kelulusan siswa, yang kemudian dapat diselidiki lagi lebih dalam bahwa nilai harian sebagai node pohon juga ikut mempunyai pengaruh yang sama terhadap hasil prediksi kelulusan siswa. Setiap kategori nilai dari nilai harian akan membentuk jalur aturan kondisi yang baru dalam menentukan prediksi hasil kelulusan siswa baik itu siswa yang lulus maupun yang tidak lulus.

Selain dua penelitian di atas, telah banyak penelitian sejenis mengenai Educational Data Mining, namun kebanyakan indikator yang digunakan didasarkan hanya pada aspek kognitif semata, dimana indikator yang digunakan adalah hasil (produk) yang didapat melalui proses tes obyektif. Peneliti merasa bahwa indikator-indikator di luar aspek kognitif juga harus mulai dipertimbangkan. Penilaian dengan mempertimbangkan aspek di luar kognitif tersebut telah mulai dikembangkan dalam konsep Pendidikan Karakter. Pengembangan konsep pendidikan karakter bahkan telah menjadi unsur utama dalam kurikulum 2013 yang dicanangkan oleh Kementerian Pendidikan Indonesia, di mana dapat dilihat dalam kompetensi inti yang memuat sikap religius dan sikap sosial pada semua mata pelajaran.

Pendidikan karakter tidak hanya berorientasi pada aspek kognitif saja, akan tetapi lebih berorientasi pada proses pembinaan potensi yang ada dalam diri peserta didik, dikembangkan melalui pembiasaan sifat-sifat baik yaitu berupa pengajaran nilai-nilai karakter yang baik. Pendidikan karakter adalah konstruksi teoritik dan praktik, urgensi pendidikan progresif dan revitalisasi peran orangtua dan guru [4]. Oleh karena itu, penelitian mengenai Educational Data Mining ini akan menjadi sangat menarik jika diterapkan dalam kajian Pendidikan Karakter dikarenakan selain indikator non kognitif, pertimbangan latar belakang siswa seperti variabel orang tua juga dapat dijadikan sebagai bahan kajian serta diolah dengan data mining. Dalam konsep Data Mining dikenal banyak teknik, dimana salah satunya adalah teknik klasifikasi. Teknik klasifikasi adalah teknik pembelajaran untuk memprediksi suatu nilai dari variabel target kategori [5]. Karena dalam penelitian ini mempunyai variabel target yaitu Sangat baik, Baik, Cukup dan Gagal, maka penelitian ini akan menggunakan data mining teknik klasifikasi. Cukup banyak algoritma dalam teknik klasifikasi, namun berdasar penelitian [6] terkait uji komparasi algoritma klasifikasi yaitu *logistik regression*, *naive bayes*, *neural network* dan *decision tree* menunjukkan bahwa decision tree memiliki performansi yang paling tinggi. Oleh karena itu penulis memilih menggunakan algoritma decision tree. Algoritma C.45 akan digunakan untuk penelitian decision tree kali ini, dikarenakan keunggulannya dalam *breakdown* proses pengambilan keputusan menjadi lebih simple diharapkan dapat

membantu memberikan gambaran urutan tingkat pengaruh atribut yang tersedia.

Dengan penelitian data mining ini, diharapkan dapat diketahui tingkat akurasi Decision Tree Algoritma C.45 dalam mengukur pengaruh atribut-atribut latar belakang siswa tersebut terhadap kesuksesan pendidikan karakter. Hal ini dirasa sangat menarik dikarenakan pendidikan karakter sejatinya aspek penilaiannya bersifat abstrak. Penilaian yang bersifat abstrak ini dalam kacamata ilmu psikologi adalah penilaian yang didasarkan pada item atau aspek tertentu yang secara teoritis dimungkinkan munculnya suatu karakter tertentu. Sehingga dengan penelitian ini juga akan diketahui apakah Decision Tree Algoritma C.45 memenuhi aspek reliabilitas dan validitas sebagai alat ukur kesuksesan pendidikan karakter.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif. Penelitian kuantitatif ini diterapkan untuk menjawab pertanyaan peneliti mengenai tingkat akurasi Decision Tree Algoritma C.45. Untuk mencapai tujuan tersebut, peneliti menggunakan metode eksperimen. Secara garis besar, penelitian ini bertujuan untuk pengujian decision tree C4.5 serta untuk mendapatkan model prediksi kesuksesan pendidikan karakter.

A. Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data yang didapatkan langsung dari objek penelitian yaitu Unires Asrama mahasiswa Universitas Muhammadiyah Yogyakarta, data yang diambil adalah data yang terkait dengan siswa tahun 2014 serta 2015. Data yang didapat adalah data yang masih utuh di mana terdapat banyak field atau atribut seperti “nama, alamat rumah, asal sekolah, jenis kelamin, agama, nomer telephone, tahun masuk, program studi, dan lain sebagainya”.

Metode pengumpulan data dalam penelitian ini adalah sebagai berikut :

1. Wawancara

Wawancara dilakukan secara langsung terhadap pihak Unires (university Residence) Universitas Muhammadiyah Yogyakarta untuk mendapatkan informasi dan data-data terkait pendidikan karakter di Unires Yogyakarta.

2. Dokumen

Melakukan pengumpulan data history berupa data internal dan eksternal yang telah tersusun rapi berupa database secara

elektronik maupun data dalam bentuk *hardcopy* dalam buku besar Unires Yogyakarta.

3. Studi Pustaka

Mengetahui Informasi secara teori mengenai pokok permasalahan dan teori pendukung yang digunakan sebagai dasar pemikiran dalam membahas implementasi Decision Tree Algoritma C4.5.

B. Metode Analisis Data

Tahapan ini merupakan tahapan dimana data-data diperlukan dikumpulkan yang kemudian akan diolah dengan metode Decision Tree C4.5, dimana sebelumnya dilakukan praproses dengan seleksi fitur. Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan proses tahapan knowledge discovery in databases (KDD) yang terdiri dari Database, Data Cleaning, Data integration, Task relevan data, Data transformation, Data mining, Pattern evolution, knowledge :

1. Database

Koleksi data yang saling berhubungan untuk dipergunakan secara bersama kemudian dirancang untuk memenuhi kebutuhan informasi penelitian.

2. Data Cleaning

Data yang diperoleh dari pengumpulan data di Unires Yogyakarta ini tentu memiliki isian-isian yang tidak sempurna seperti data yang hilang atau data yang tidak valid. Data yang tidak relevan itu juga lebih baik dibuang karena keberadaanya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. Pembersihan data juga akan mempengaruhi performansi dari implementasi data mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya.

3. Data Integration

Integrasi data dilakukan pada atribut-attribut yang mengidentifikasi entitas-entitas yang unik pada siswa Unires Yogyakarta tahun 2014-2015. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil data yang menyimpang.

4. Task relevan data

Setelah semua sumber data digabung atau diintegrasikan menjadi satu database, maka tahap selanjutnya adalah melakukan task relevant data. Pada tahap ini melakukan relevansi atribut dari data yang relevant atau yang sesuai dengan target atau output yang akan dihasilkan.

5. Data transformation

Dalam tahap ini, data yang tersedia dilakukan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses ini dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam database.

6. Data mining

Dalam tahap ini akan dilakukan proses implementasi data mining Algoritma C4.5.

7. Pattern evaluation

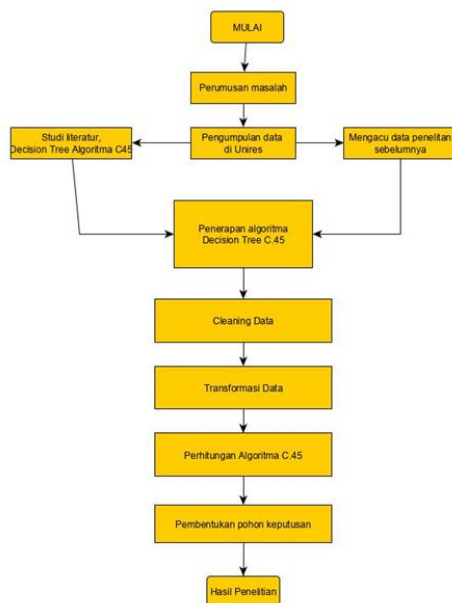
Dalam tahap ini, hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai dengan hipotesa, ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki data mining lain yang lebih sesuai, atau menerima hasilnya sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat.

8. Knowledge

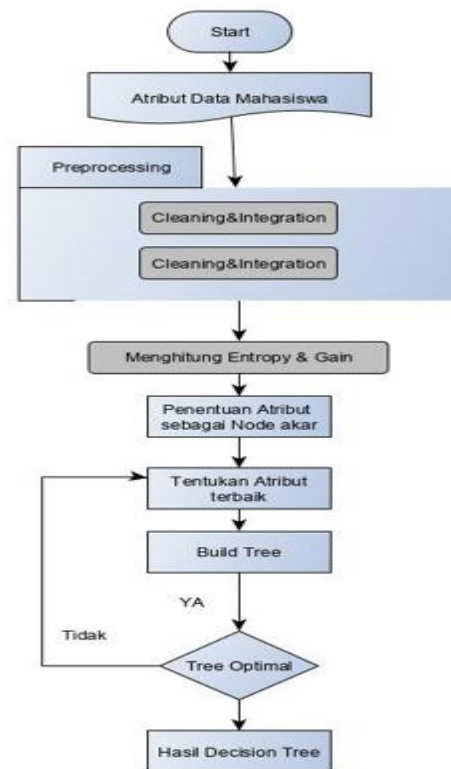
Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

C. Alur dan Desain Penelitian

Bertikut adalah alur dan desain penelitian yang merujuk pada proses KDD.



Gambar 1. Alur Penelitian



Gambar 2. Desain Penelitian

III. HASIL DAN PEMBAHASAN

A. Perumusan Masalah

Perumusan masalah dalam penelitian ini, meliputi proses studi literatur terkait penelitian sebelumnya mengenai implementasi data mining Decision Tree C4.5 serta pengumpulan data di Unires. Dimana didapat gambaran mengenai penelitian serta data mentah yang akan diproses.

B. Data Selection

Dalam proses ini dilakukan pembersihan pada data seperti memeriksa kondisi data, menghilangkan redundansi data, memperbaiki kesalahan pada data, dan menghilangkan beberapa data yang tidak lengkap ataupun data yang tidak diperlukan.

Selain itu dalam proses ini atribut-atribut yang secara teori tidak relevan terhadap hasil pendidikan akan dihilangkan. Teori yang digunakan untuk proses ini mengacu pada teori pendidikan, dimana untuk mencapai prestasi belajar yang optimal dipengaruhi oleh faktor internal dan faktor eksternal. Menurut *M. Dalyono (2009: 55)* faktor-faktor yang mempengaruhi prestasi belajar yaitu:

1. Faktor Internal
 - a. Kesehatan
 - b. Intelegensi dan Bakat

- c. Minat dan Motivasi
 - d. Cara belajar
2. Faktor Eksternal
- a. Keluarga
 - b. Sekolah
 - c. Masyarakat
 - d. Lingkungan sekitar

Dalam proses pemilihan atribut ini, penulis juga mempertimbangkan paradigma-paradigma yang sesuai dengan teori di atas, seperti bahwa pendidikan di pulau Jawa lebih maju dibandingkan dengan di luar pulau Jawa. Selain itu pemilihan atribut juga didasarkan pada kesesuaian data dengan Algoritma di mana data yang mempunyai *range* terlalu panjang serta tidak dapat dikelompokkan tidak akan dimasukkan sebagai atribut penilaian.

Dalam proses *cleaning* atribut data ini tercipta kondisi:

- a. Data pada tabel tes nilai n masuk tidak mengalami penghilangan atribut.
- b. Data pada tabel Format penilaian Unires, hanya diambil nilai agregat yang diperoleh dari transformasi nilai rata-rata.
- c. Tabel Atribut data siswa mengalami proses penghapusan variabel-variabel yang tidak memungkinkan proses transformasi data sesuai kebutuhan Algoritma C4.5 serta secara teori di atas tidak berpengaruh terhadap pendidikan karakter.

C. Transformasi Data

Dalam proses transformasi data ini, data mengalami proses transformasi dan penggabungan kedalam satu database, adapun data hasil dari proses ini adalah sebagai mana yang terlihat dalam tabel 1 dibawah, dimana kolom Possible Value menggambarkan proses pengelompokan data kedalam range masing-masing agar sesuai dengan kebutuhan pengolahan data pada Decision Tree C4.5.

Tabel I. Hasil seleksi atribut

N	Atribut	Possible Value
0		
1	NO_RES	Char
2	JENIS_KELAMIN	Laki-laki, Perempuan
3	FAKULTAS	Agama Islam, Ekonomi dan Bisnis, Hukum, Ilmu Sosial dan Politik, Kedokteran dan Ilmu Kesehatan, Pertanian, Teknik, Pendidikan Bahasa, Politeknik UMY
4	KELAS	Reguler, Internasional

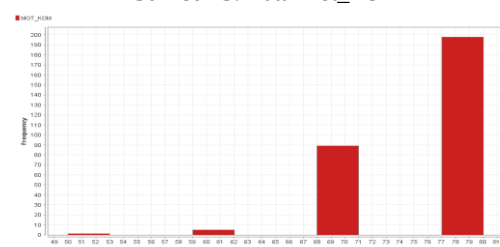
5	PROPINSI RESIDEN	Jawa, Luar Jawa
6	JURUSAN SEKOLAH	IPA, IPS, Agama, Bahasa, Jurusan Lain
7	PEKERJAAN AYAH/IBU	PNS, Wiraswasta, Swasta, Guru, Petani, Lain-lain
8	PENGHASILAN AYAH/IBU	<Rp1.000.000, Rp1.000.000-Rp2.500.000, Rp2.500.000-Rp4.000.000, >Rp4.000.000
9	MOT_LB	Numerik (>75, 66-75, 56-65, <56)
10	MOT_B&P	Numerik (>75, 66-75, 56-65, <56)
11	MOT_KOM	Numerik (>75, 66-75, 56-65, <56)
12	QUR_QIR	Numerik (>75, 66-75, 56-65, <56)
13	QUR_TAJ	Numerik (>75, 66-75, 56-65, <56)
14	PREDIKAT	A, B, C, D

Dalam proses transformasi data, dilakukan proses pengelompokan data atribut. Beberapa atribut telah mengalami pengelompokan dalam penyimpanan data di Unires, namun ada beberapa atribut yang belum mengalami pengelompokan sehingga perlu dilakukan proses transformasi data, acuan pengelompokan data didasarkan pada frekuensi kemunculan data, berikut beberapa atribut yang mengalami proses ini:

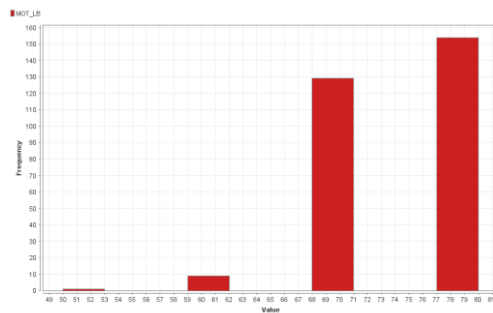
- a. Atribut jurusan sekolah: IPA, IPS, Agama, Bahasa dan Jurusan lain.
- b. Pekerjaan Ayah/Ibu: PNS, Wiraswasta, Swasta, Guru, Petani, lain-lain.
- c. Atribut numerik
Frekuensi kemunculan data numerik dapat dilihat dalam gambar di bawah,



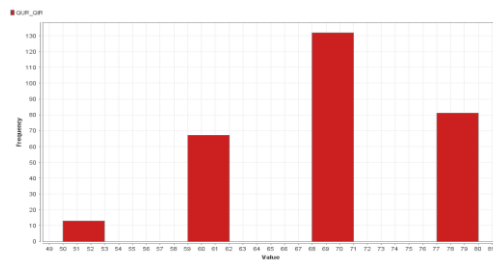
Gambar 3. Data Mot_B&P



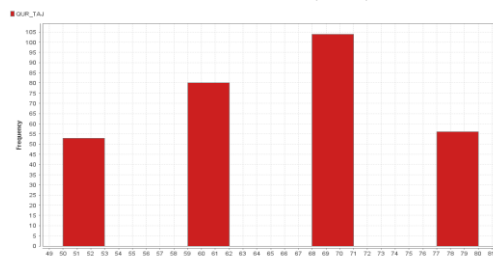
Gambar 4. Data Mot_Kom



Gambar 5. Data Mot_LB



Gambar 6. Data Qur_Qir



Gambar 7. Data Qur_Taj

D. Data Mining Algoritma C4.5

Secara umum node dalam proses Decision Tree adalah atribut yang mempunyai nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai gain suatu atribut haruslah terlebih dahulu diketahui nilai entropinya. Sebagai parameter untuk mengukur tingkat keberagaman (heterogenitas) dari kumpulan data. Untuk menghitung entropy digunakan rumus seperti yang tertera dalam persamaan berikut:

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

Keterangan :

- S adalah himpunan (dataset) kasus
- n adalah banyaknya partisi S
- p_i adalah probabilitas yang di dapat dari $\text{Sum}(Y_i)$ dibagi Total Kasus

Proses selanjutnya adalah menghitung gain, gain adalah Ukuran efektifitas suatu variabel dalam mengklasifikasikan data. Gain dapat dirumuskan dengan:

$$Gain(A) = Entropy(S) - Entropy_A(S)$$

Algoritma C4.5 adalah pengembangan dari Algoritma ID3, namun berbeda dengan algoritma ID3 yang menggunakan information gain, proses pembetukan tree pada algoritma C4.5 menggunakan Gain Ratio. Sehingga dalam proses selanjutnya perlu dicari Split Info yang akan digunakan sebagai pembagi dari $Gain(A)$, yang akan menghasilkan Gain Ratio. Di bawah adalah rumus Split Info serta Gain Ratio:

$$SplitInfo_A(D) = \sum_{i=1}^v \frac{|D_j|}{|D|} \cdot \log_2 \left(\frac{|D_j|}{|D|} \right)$$

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(D)}$$

E. Pembentukan Decision Tree

Proses pembentukan Decision Tree adalah sebagai berikut:

1. Hitung jumlah data, jumlah data berdasarkan anggota atribut hasil dengan syarat tertentu. Untuk proses pertama syaratnya masih kosong.
2. Pilih atribut sebagai Node.
3. Buat cabang untuk tiap-tiap anggota dari Node.
4. Periksa apakah nilai entropy dari anggota Node ada yang bernilai nol. Jika ada, tentukan daun yang terbentuk. Jika seluruh nilai entropy anggota Node adalah nol, maka proses pun berhenti.
5. Jika ada anggota Node yang memiliki nilai entropy lebih besar dari nol, ulangi lagi proses dari awal dengan Node sebagai syarat sampai semua anggota dari Node bernilai nol.

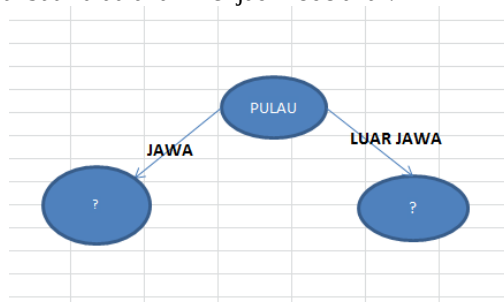
Pada penelitian ini, menggunakan perhitungan algoritma C4.5 dalam membangun pohon keputusan. Tabel II di bawah menunjukkan hasil dari perhitungan Algoritma C4.5 untuk mencari node awal Decision Tree.

TABEL II.
PERHITUNGAN ALGORITMA C4.5

	S	ENTROPY	INFORMATIO N GAIN	SPLIT LEVEL	GAIN RATIO
Total	307	1,465540426			
Jenis Kelamin			0,032635203	0,999624939	0,032647448
Laki – Laki	150	1,600376978			
Perempuan	157	1,272900361			
Jurusan			0,075070828	2,952750677	0,025424
Ilmu sosial dan ilmu politik	63	1,615250265			
Ekonomi dan bisnis	51	1,294414371			
Agama islam	49	1,183345557			
Kedokteran dan ilmu kesehatan	44	1,364331215			
Teknik	35	1,680422032			
Pertanian	21	0,856995718			
Hukum	20	1,765957321			
Pendidikan bahasa	18	1,122113677			
Politeknik umy	6	1,459147917			
Kelas			0,015101575	0,155269505	0,0972604
Reguler	290	1,433496572			
International	17	1,739454196			
Pulau			0,207443468	0,97839285	0,2120247
Jawa	180	1,420173612			
Luar jawa	127	1,520613253			
Jurusan Sekolah			0,017263742	1,555934521	0,0110954
Ipa	173	1,397682599			
Ips	91	1,545566516			
Bahasa	6	0,650022422			
Agama	8	1,548794941			
Jurusan Lain	29	1,582234674			
Pekerjaan Ayah/Ibu			-0,042458423	2,27525268	-0,018660
PNS	112	1,605930693			
WIRASWASTA	66	1,427482563			
SWASTA	52	1,40900614			
GURU	16	1,324460293			
PETANI	13	1,334679141			
LAIN-LAIN	48	1,605563479			
Mot_LB			0,144848083	1,748762423	0,0828289
>Rp 4.000.000,00	110	1,486732377			
Rp 2.500.000,00 – Rp 4.000.000,00	119	1,505247235			
Rp 1.000.000,00 - Rp 2.500.000,00	62	1,135739597			
<Rp 1.000.000,00	16	-0,476768			
Mot_B&P			0,04687044	1,194610046	0,0392349
> 75	164	1,25215911			
66 -75	132	1,645628585			
56 – 65	10	1,295461844			
< 56	1	0			
Mot_Kom			0,025903108	1,375184465	0,018836
> 75	140	1,303063005			
66 -75	143	1,533975792			
56 – 65	19	1,714297751			
< 56	5	1,521928095			
Qur_Qir			0,028062378	1,051826904	0,0266796

> 75	208	1,369681804		
66 -75	91	1,612752319		
56 – 65	74	1,390817185		
< 56	13	1,419556299		
Qur_Taj		0,025177921	1,943502873	0,0129549
> 75	61	1,272737014		
66 -75	110	1,353390396		
56 – 65	82	1,513334299		
< 56	54	1,696073612		

Dari tabel iterasi tersebut dapat dilihat gain ratio tertinggi adalah atribut Pulau, sehingga atribut Pulau akan menjadi node akar.



Gambar 8. Node Akar

Dalam atribut pulau, terdapat dua simpul, yaitu Jawa dan Luar Jawa, dari kedua simpul tersebut tidak ada satupun yang entropy bernilai nol (0), seperti yang terlihat dalam tabel 2 di bawah, oleh karena itu perlu dilakukan proses iterasi berikutnya sampai semua simpul bernilai nol(0). Hal ini juga menunjukkan bahwa atribut pulau berpengaruh terhadap proses pendidikan karakter di Unires, namun belum secara langsung menghasilkan nilai akhir A, B, C ataupun D.

TABEL III.
ENTROPY PULAU

	S	A	B	C	D	Entropy
Jawa	180	117	30	7	26	1,420173612
Luar Jawa	127	76	28	6	17	1,520613253

Dalam proses selanjutnya, dilakukan pengulangan penghitungan entropy, information gain serta gain ratio, namun kali ini Jawa dan Luar Jawa masing-masing menjadi node.

Dari proses iterasi kedua ini didapatkan hasil sebagai berikut

1. Proses Iterasi “Pulau Jawa”, didapatkan gain ratio tertinggi adalah atribut “Jurusan Sekolah” dengan nilai 0,067845601, namun dari 5 sub di atribut sekolah yaitu IPA, IPS, Bahasa, Agama serta Jurusan Lain, belum satupun yang menghasilkan entropi 0(nol) sehingga proses iterasi cabang ini akan

dilanjutkan dengan masing-masing sub atribut menjadi node berikutnya.

2. Proses Iterasi “Pulau Luar Jawa”, didapatkan hasil atribut Kelas sebagai node berikutnya dikarenakan mempunyai Gain Ratio Tertinggi(0,655007555). Namun dari dua sub atribut yang ada yaitu Kelas International maupun kelas Reguler, belum ditemukan entropy 0(nol), sehingga proses iterasi cabang ini pun akan dilanjutkan dengan masing-masing sub atribut menjadi node berikutnya.

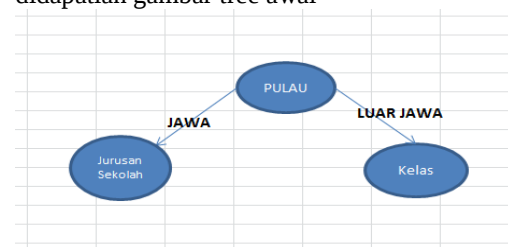
TABEL IV.
ENTROPY JURUSAN SEKOLAH(PULAU JAWA)

	S	A	B	C	D	Entropy
IPA	100	67	16	3	14	1,358997868
IPS	56	34	9	3	10	1,530975814
Bahasa	3	2	1	0	0	0,918295834
Agama	5	2	1	1	1	1,921928095
Jurusa n Lain	16	12	3	0	1	1,014097656

TABEL V.
ENTROPY KELAS (PULAU LUAR JAWA)

	S	A	B	C	D	Entropy
Reguler	119	75	23	6	15	1,471990362
Internas ional	8	1	5	0	2	1,298794941

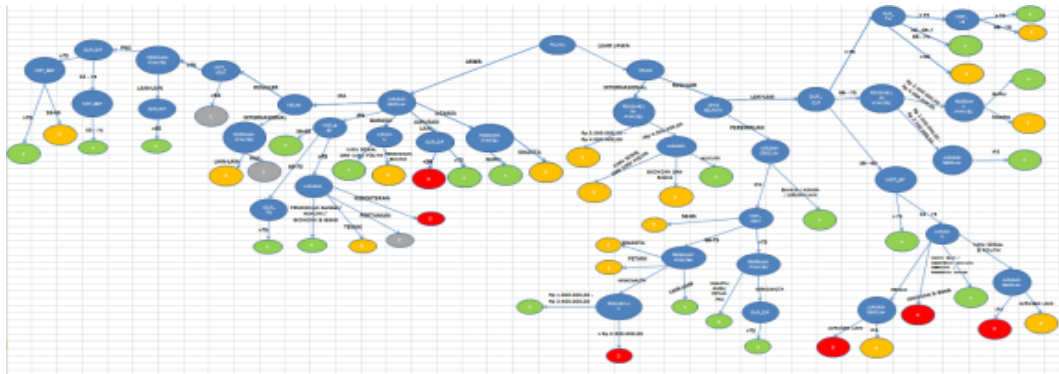
Dari hasil proses iterasi kedua ini didapatlah gambar tree awal



Gambar 9. Tree Awal

Dikarenakan masing-masing anggota Node masih memiliki nilai entropy lebih besar dari nol, proses ini akan diulangi lagi dari awal

dengan Node sebagai syarat sampai semua anggota dari Node bernilai nol. Berikut adalah gambar Decision Tree yang terbentuk



Gambar 10. Decision Tree

F. Patern Evaluation

Dari hasil Tree tersebut, dapat diketahui model untuk prediksi nilai yaitu

a. Prediksi nilai A

1. If(Pulau=Jawa)^(Jurusan Sekolah=IPA)
 $\wedge$ (Kelas=Reguler) $\wedge$ (Mot_Kom>75) $\wedge$
(Pekerjaan Ayah/Ibu=PNS) $\wedge$
(Qur_Qir>75) $\wedge$ (Mot_B&P>75)
2. If(Pulau=Jawa)^(Jurusan Sekolah=IPA)
 $\wedge$ (Kelas=Reguler)^(Mot_Kom>75) $\wedge$
(Pekerjaan Ayah/Ibu=PNS)^(Qur_Qir
66-75)^(Mot_BP 66-75)
3. If(Pulau=Jawa)^(Jurusan Sekolah=IPA)
 $\wedge$ (Kelas=Reguler)^(Mot_Kom>75) $\wedge$
(Pekerjaan Ayah/Ibu=Lain-lain) $\wedge$
(Qur_Qir > 65)
4. If(Pulau=Jawa)^(Jurusan Sekolah=IPS)
 $\wedge$ (Mot_BP 56-65)
5. If(Pulau=Jawa)^(Jurusan Sekolah=IPS)
 $\wedge$ (Mot_BP 66-75)^(Qur_Taj > 75)
6. If(Pulau=Jawa)^(Jurusan Sekolah=IPS)
 $\wedge$ (Mot_BP >75)^(Jurusan=Bahasa)
7. If(Pulau=Jawa)^(Jurusan Sekolah=IPS)
 $\wedge$ (Mot_BP >75)^(Jurusan=Hukum)
- 8.If(Pulau=Jawa)^(Jurusan Sekolah=IPS) $\wedge$
(Mot_BP>75)^(Jurusan=Ekonomi)
9. If(Pulau=Jawa) $\wedge$ (Jurusan Sekolah =
Bahasa) $\wedge$ (Jurusan=Ilmu Sosial dan
Ilmu Politik)
10. If(Pulau=Jawa)^(Jurusan Sekolah =
Jurusan Lain) $\wedge$ (Qur_Qir >75)
11. If(Pulau=Jawa)^(Jurusan Sekolah =
Agama) $\wedge$ (Pekerjaan Ayah/Ibu=Guru)

12. If(Pulau=Luar Jawa) ^ (Kelas = Internasional) ^ (Penghasilan Ayah/Ibu >4.000.000) ^ (Jurusan=Hukum)

b. Prediksi nilai D

1. If(Pulau=Jawa)^(Jurusan Sekolah=IPS)
 $\wedge$ (Mot_B&P>75) $\wedge$ (Jurusan =
 Kedokteran)
2. If(Pulau=Jawa)^(Jurusan Sekolah=
 Jurusan Lain)^(Qur_Qir<56)
3. If(Pulau=Luar Jawa) $\wedge$ (Kelas= Reguler)
 $\wedge$ (Jenis Kelamin=Perempuan)^(Jurusan
 Sekolah=IPA) $\wedge$ (Mot_Kom 66-75) $\wedge$
 (Pekerjaan Ayah/Ibu Wiraswasta) $\wedge$
 (Penghasilan >Rp.2.500.000)
4. If(Pulau=Luar Jawa)^(Kelas=Reguler) $\wedge$
 (Jenis Kelamin=Laki-laki) $\wedge$ (Mot_BP
 66-75)^(Jurusan=Teknik)^(Jurusan
 Sekolah=Jurusan Lain)
5. If(Pulau=Luar Jawa)^(Kelas=Reguler) $\wedge$
 (Jenis Kelamin=Laki-laki)^(Mot_BP
 66-75)^(Jurusan=Ekonomi dan Bisnis)
6. If(Pulau=Luar Jawa)^(Kelas=Reguler) $\wedge$
 (Jenis Kelamin=Laki-laki)^(Mot_BP
 66-75)^(Jurusan=Ilmu Sosial dan Ilmu
 Politik)^(Jurusan Sekolah=IPA)

Pada proses selanjutnya dilakukan proses pengujian dengan menggunakan confuse matrix

accuracy: 60.90% +/- 3.66% (mikro 60.91%)					
	true A	true B	true D	true C	class precision
pred. A	185	56	42	12	62.71%
pred. B	6	1	0	1	12.50%
pred. D	2	1	0	0	25.00%
pred. C	0	0	0	0	0.00%
class recall	95.95%	1.72%	2.33%	0.00%	

Gambar 11. Confuse Matrix

Testing model dilakukan dengan menggunakan toll RapidMiner, dengan menggunakan Cross Validation, dan didapat nilai accuracy sebesar 60,91%. Berdasarkan nilai True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP) dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan Persamaan 1. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan Persamaan 2. Sementara itu, recall menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai recall diperoleh dengan Persamaan 3.

Persamaan 1

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

Persamaan 2

$$\text{Presisi} = \frac{TP}{TP + FP} * 100\%$$

Persamaan 3

$$\text{Recall} = \frac{TP}{TP + FN} * 100\%$$

IV. KESIMPULAN

Hasil Accuracy sebesar 60,91%, menunjukkan bahwa algoritma decision tree C4.5 layak digunakan untuk melakukan prediksi tingkat kesuksesan pada pendidikan karakter. Namun masih butuh banyak kajian mendalam terutama dalam menentukan atribut-atribut yang memang benar-benar mempengaruhi kesuksesan pendidikan karakter sehingga diharapkan akan didapat accuracy yang lebih baik, serta terjadi efisiensi dalam pengelompokan atribut agar tidak terlalu banyak atribut-atribut, dimana dalam penelitian ini dari 12 reguler atribut didapat atribut turunan hingga 64 atribut.

Dari model dalam memprediksi nilai A sebagai tolak ukur kesuksesan pendidikan karakter di Unires tersebut, terdapat beberapa informasi menarik, yang dapat membuktikan beberapa asumsi terkait pendidikan khususnya pendidikan karakter, diantaranya:

1. Bidang Bahasa dan Sosial berperan berkorelasi positif dengan pendidikan karakter.

2. Latar belakang pendidikan agama serta mempunyai orang tua yang menjadi seorang pendidik/guru berkorelasi positif dengan pendidikan karakter.
3. Kemampuan untuk membaca Al-Qur'an berkorelasi positif dengan pendidikan karakter.

UCAPAN TERIMA KASIH

Terimakasih kepada ALLAH SWT yang telah memberikan limpahan rahmat dan karuniaNya kepada penulis, sehingga penulis dapat menyelesaikan penelitian ini. Terimakasih kepada orang tua yang telah memberikan doa, motivasi dan dukungan financial sehingga saya terus semangat dan termotivasi. Terimakasih kepada teman-teman, terkhusus komunitas MTI Amikom Yogyakarta 15A serta komunitas "Santri Kalong" yang sudah memberikan semangat dan membantu dalam penyusunan naskah ini. Terimakasih kepada staff MTI Universitas Amikom Yogyakarta yang sudah memberikan kemudahan dalam segala urusan akademik yang berkaitan dengan penyelesaian penelitian ini. Dan terimakasih juga kepada dosen-dosen yang telah membimbing saya, sehingga saya dapat menyelesaikan penelitian ini dengan baik.

REFERENSI

- [1] Yadav SK, Pal S. 2012. Data Mining: A Prediction for performance improvement of engineering students using classification. World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol 2, No.2, 51-56.
- [2] Pambudi R.H, Setiawan B.D, Indriati. 2018. Penerapan Algoritma C4.5 Dalam Program Untuk Memprediksi Kinerja Siswa Sekolah Menengah. Jurnal Pengembangan Teknologi Informasi dan Ilmu Vol. 2, No. 7, hlm. 2637-2643 Komputer e-ISSN: 2548-964X.
- [3] Nugraha, dkk. 2016. Penerapan metode decision tree (data mining) untuk memprediksi tingkat kelulusan siswa SMP N 1 Kintamani. Seminar Nasional Vokasi dan Teknologi (SEMNASVOKTEK). ISSN Cetak : 2541-2361 | ISSN Online : 2541-3058. Denpasar-Bali.
- [4] Mu'in, Fatchul. (2011). Pendidikan Karakter, Konstruksi Teoretik dan Praktik. Ar-Ruzz Media, Yogyakarta.
- [5] Kusriani & Luthfi, E. 2009, Algoritma Data Mining. Andi Offset, Yogyakarta.
- [6] Hastuti, K. (2012). Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif, 2012(Semantik), 241-249.
- [7] Arshad., M, P., Buniyamin, N., & AB Manan, J.,L Prediction of Engineering Students'

- Academic Performance Using Artificial Neural Network and liner regression . IEEE 5th Confrence on Engineneering Education (ICEED)
- [8] Lee, A.S., 2007, Action Is an Artifact: What Action Research and Design Science Offer to Each Other, Springer, Laredo, Texas, USA.
 - [9] Prasetyo, T. F., Susandi, D., & Widianingrum, I. S. (2016). Prediksi Kelulusan Mahasiswa pada Perguruan Tinggi Kabupaten Majalengka Berbasis Knowledge Based System. Seminar Nasional Telekomunikasi dan Informatika (SELISIK 2016) , ISSN : 2503-2844, Hal. 32-37
 - [10] Al-Barrak M.A, Al-Razgan M. 2016. Predicting Student Final GPA Using Decission Trees: A Case Study. International Journal of Information and Educational Technology, Vol. 6, No. 7. DOI: 10.7763/IJJET.2016.V6.745.
 - [11] Nichat AA, Raut AB. 2017. Predicting and Analysis of Student Performance Using Decision Tree Technique. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 5, Issue 4. ISSN(Online): 2320-9801 | ISSN(Print): 2320-9798.
 - [12] Kabra RR, Bichar RS. 2011. Performance Prediction of Engineering Students using Decision Trees. *International Journal of Computer Applications (0975-8887)* Volume 36 – No.11.
 - [13] Shaleena K.P, Paul S. 2015. Data Mining Techniques for Predicting Students Performance. IEEE International Conference on Engineering and Technology (ICETECH). Coimbatore, TN, India.
 - [14] Vandy Ferdian, Seng Hansung. 2017. Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE. Jatisi, Vol. 3 No. 2 Maret 2017.
 - [15] HSSINA, B., dkk., 2014, A Comparative Study of Decision Tree ID3 and C4.5, Sultan Moulay Slimane University, Morocco.
 - [16] Shah, N. S. (2012). Predicting Factors That Affect Students' Academic Performance by Using Data Mining Techniques. *Pakistan Business Review*, 13(4), 631-638.
 - [17] Cheewaparakobkit, P. (2015). Predicting student academic achievement by using the decision tree and neural network techniques. *Catalyst*, 12(2), 34-43
 - [18] Osmanbegović, E. and Suljić, M. (2012). Data mining approach for predicting student performance. *Economic Review: Journal of Economics & Business / Ekonomska Revija: Casopis za Ekonomiju i Biznis*, 10(1), 3-12.
 - [19] Josip Mesaric, Dario Šebalj (2016). Decision trees for predicting the academic success of students. *Croatian Operational Research Review CRORR* 7(2016), 367–388