

Using data mining with C4.5 algorithm for student department selection at MTs N Kaliangkrik

Faruq Ardana Kurniawan, Emilya Uly Artha, Ardhin Primadewi*

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Magelang, Indonesia

*email: ardhin@ummgl.ac.id

DOI: <https://doi.org/10.31603/binr.4989>

Abstract

Psychological tests can determine the characteristics of behavior, personality, attitudes, interests, motivation, attention, perceptions, thinking power, intelligence, fantasies of students. MTs N Kaliangkrik routinely conducts tests for the selection of majors on its students assisted by Pelita Harapan Bangsa Magelang. In the implementation of the test for students at MTs N Kaliangkrik, processing and calculating the score still used Ms. Excel which requires extra time to recap and know the test results and the school needs to recap the existing results. The system developed applies data mining using the C4.5 Algorithm to predict the selection of majors. The test that is used as system input is the grade IX test score of MTs N Kaliangkrik which includes verbal, non-verbal, general intelligence, language knowledge, definite knowledge, general knowledge, and qualitative power tests. The accuracy of the similarity in the system reaches 80% (good) so that the system is suitable for use as a prediction tool for selecting majors in other schools.

Keywords: Data Mining; Decision tree; Algorithm C4.5; IQ test;

Abstrak

Tes Psikologi dapat mengetahui karakteristik perilaku, kepribadian, sikap, minat, motivasi, perhatian, presepsi, daya pikir, kecerdasan, fantasi dari peserta didik. MTs N Kaliangkrik secara rutin melakukan tes untuk pemilihan jurusan pada peserta didiknya dibantu oleh Pelita Harapan Bangsa Magelang. Dalam pelaksanaan tes peserta didik di MTs N Kaliangkrik, pengolahan dan penghitungan skor nilai masih menggunakan Ms. Excel yang membutuhkan waktu ekstra dalam merekap dan mengetahui hasil tes dan pihak sekolah perlu merekap ulang hasil yang ada. Sistem yang dikembangkan menerapkan data mining menggunakan Algoritma C4.5 untuk memprediksi pemilihan jurusan. Tes yang dijadikan input sistem adalah nilai tes kelas IX MTs N Kaliangkrik meliputi tes verbal, non verbal, intelegensi umum, pengetahuan bahasa, pengetahuan pasti, pengetahuan umum, daya kualitatif. Akurasi kesamaan pada sistem mencapai 80% (baik) sehingga sistem layak digunakan sebagai alat prediksi pemilihan jurusan pada sekolah-sekolah lain.

Kata Kunci: Data Mining; Pohon keputusan; Algoritma C4.5; Tes IQ;



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

1. Pendahuluan

Psikologi diperlukan oleh dunia pendidikan untuk mengetahui karakteristik perilaku, kepribadian, sikap, minat, motivasi, perhatian, persepsi, daya pikir, kecerdasan, fantasi dari peserta didik. Setiap peserta didik memiliki karakteristik unik yang berbeda satu sama lain. Perbedaan karakteristik psikologis yang dimiliki oleh para peserta didik seyogyanya dipahami oleh guru, wali kelas dan guru Bimbingan Konseling. Karakteristik psikologis ini dapat membantu peserta didik dalam pemecahan atas masalah-masalah yang dialami peserta didik khususnya saat proses pembelajaran (Sakerebau, 2018).

Madrasah Tsanawiyah Negeri Kaliangkrik (MTs N Kaliangkrik) terletak di daerah Kabupaten Magelang. Lokasi MTs N Kaliangkrik terbagi menjadi dua tempat, yaitu: Kelas 7/VII terdiri dari 6 Kelas, yang terletak di kampung Torip, dan Kelas 8/VIII terdiri 6 kelas, dan juga kelas 9/IX juga terdiri 6 kelas, yang berlokasi di kampung Beseran, Kecamatan Kaliangkrik, Kabupaten Magelang. MTs N Kaliangkrik mulai didirikan pada tahun 1970. MTs N Kaliangkrik merupakan sekolah umum setingkat SMP yang berciri khas agama Islam. Materi pelajaran yang dipelajari sama seperti sekolah umum ditambah dengan materi pelajaran agama islam yang meliputi: Akidah Akhlaq, Al Qur'an Hadits, Fiqih, Sejarah Kebudayaan Islam (SKI) dan Bahasa Arab. Dalam mengelola pembelajaran peserta didik, MTs N Kaliangkrik kerap melakukan beberapa tes psikologi dengan bantuan sebuah lembaga resmi.

Pelita Harapan Bangsa Lembaga Pengembangan Sumber Daya Manusia (PHB LPSDM) merupakan lembaga resmi yang membantu MTs N Kaliangkrik untuk mengukur tingkat kecerdasan, dan untuk konsultasi mengenai peserta didik. PHB LPSDM memberikan pelayanan tes seperti (1) Tes Intelegensi (IQ), (2) Tes Kecerdasan Emosional (EQ), (3) Tes Kepribadian, (4) Tes Intelegensi Umum (IQ) dan (5) Tes Kecerdasan Spiritual (SQ). Dalam pelaksanaan tes peserta didik di MTs N Kaliangkrik, pengolahan dan penghitungan skor nilai pada PHB LPSDM masih menggunakan Ms. Excel yang membutuhkan waktu ekstra dalam merekap dan mengetahui hasil tes. Beberapa kali terjadi kesalahan input data tes pada peserta didik dan kerusakan *file* rekap tes. Selain itu hasil Tes dari PHB LPSDM masih perlu diolah lagi oleh pihak sekolah dalam memberikan masukan kepada peserta didik dan wali dalam pemilihan penjurusan.

Dari identifikasi permasalahan perlu adanya sebuah aplikasi penerapan *Data Mining* yang dapat memudahkan PHB LPSDM dalam mengelola data tes, mengolahnya menjadi informasi yang tepat sesuai kebutuhan guru dan wali kelas MTs N Kaliangkrik dalam memberikan informasi prediksi bagi peserta didik yang sedang memilih penjurusan. Aplikasi yang dikembangkan ini diharapkan dapat memberikan solusi seperti (1) Penentuan kriteria dalam pemilihan penjurusan pada MTs N Kaliangkrik sesuai kebutuhan sekolah, (2) Penyimpanan data hasil tes secara terpusat yang dapat diakses oleh MTs N Kaliangkrik dan PHB LPSDM, (3) Data dapat diolah secara bersamaan baik oleh pihak MTs N Kaliangkrik dan PHB LPSDM dan (4) Meminimalisir adanya kesalahan input data tes peserta didik.

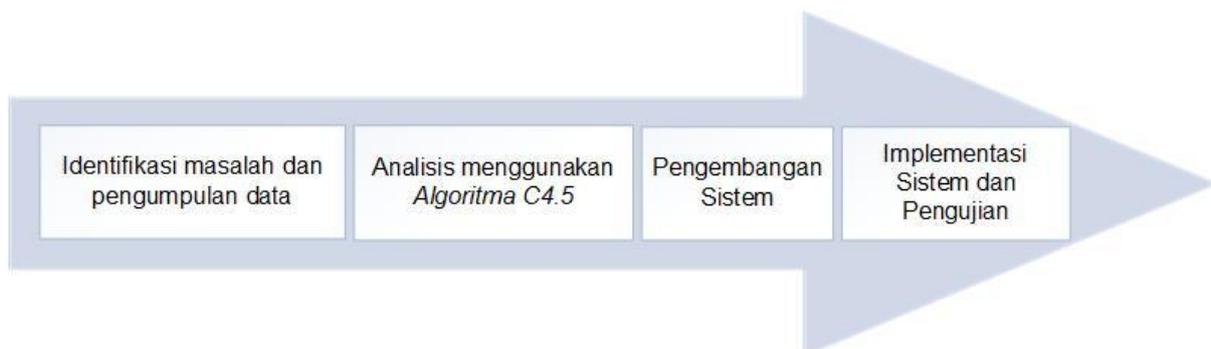
Untuk mengatasi permasalahan yang ditemukan, *Data Mining* sebagai proses menemukan hubungan baik berupa pola, dan kecenderungan sangat sesuai untuk menyelesaikan permasalahan tersebut (R & G, 2012; Yuliana & Pratomo, 2017). Penentuan pola dan kecenderungan dengan memeriksa sekumpulan data besar yang tersimpan dalam penyimpanan dengan menggunakan teknik

pengenalan pola seperti teknik statistik, matematika, kecerdasan buatan, dan *Machine Learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Harahap, 2015; Subiyanto, 2017). Dalam penentuan kriteria untuk melakukan prediksi dalam pemilihan jurusan, algoritma yang sesuai adalah Algoritma C4.5 (Nurzahputra & Muslim, 2017; Sinambela et al., 2016).

Pada penelitian sebelumnya, *Data Mining* khususnya menggunakan Algoritma C4.5 membantu penentuan strategi dalam promosi sebuah program studi pada Perguruan Tinggi (Latifah et al., 2018) dan penentuan keberhasilan mahasiswa dalam perkuliahan (Rijayana et al., 2021), pencapaian Indeks Prestasi (Rizki et al., 2020), pemilihan dalam beasiswa (Sugiyarti et al., 2018) dan kelulusan (Pradipta et al., 2019).

2. Metode

Pada Gambar 1 ditunjukkan metode penelitian yang diawali dengan identifikasi masalah dan pengumpulan data. Pada tahapan identifikasi masalah dilakukan proses wawancara dan observasi baik di MTs N Kaliangkrik dan PHB LPSPDM untuk mengetahui detail permasalahan dan mengambil data primer dan sekunder.



Gambar 1. Metode Penelitian

Pada tahapan analisis menggunakan metode Algoritma C4.5 digunakan untuk membuat pohon keputusan yang dapat membantu dalam memprediksi pemilihan jurusan. Diawali dengan pengelompokan nilai tes psikologi peserta didik dan memetakan data sampel sesuai pengelompokan nilai. Langkah-langkah dalam membuat pohon keputusan dengan metode C4.5 yaitu (1) Mencari atribut Root (Akar), (2) Mencari atribut Entropy dan Gain pada Node 1, (3) Dilanjutkan mencari atribut Entropy dan Gain pada Node 1.1 dan seterusnya. Dengan rumus entropy $entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$ (Sugiyarti et al., 2018) yang disesuaikan dengan variabel penelitian seperti pada Eq (1):

$$Gain(Total, V) = Entropy(Total) - \sum_{i=1}^n \frac{|Verbal|}{|Total|} x Entropy(Verbal) \quad (1)$$

Pada tahap pengembangan sistem, sistem dikembangkan menggunakan PHP dengan Framework Code Igniter. Pada tahap akhir sistem yang telah didesain dan di kembangkan kemudian

diujicobakan pada PHB LPSDM untuk mengolah data tes peserta didik MTs N Kaliangkrik. Pengujian dilakukan baik secara kesesuaian fungsi pada antarmuka dan perhitungan sesuai metode Algoritma C4.5 yang sebelumnya telah dihitung pada *spreadsheet*.

3. Hasil dan pembahasan

3.1. Identifikasi masalah dan pengumpulan data

Pada tahap identifikasi masalah diketahui bahwa dalam pelaksanaan tes peserta didik di MTs N Kaliangkrik, pengolahan dan penghitungan skor nilai pada PHB LPSDM masih menggunakan Ms. Excel yang membutuhkan waktu ekstra dalam merekap dan mengetahui hasil tes. Beberapa kali terjadi kesalahan input data tes pada peserta didik dan kerusakan *file* rekap tes. Selain itu hasil Tes dari PHB LPSDM masih perlu diolah lagi oleh pihak sekolah dalam memberikan masukan kepada peserta didik dan wali dalam pemilihan penjurusan. Pada tahap pengumpulan data, diambil data sampel untuk selanjutnya diolah menggunakan algoritma C4.5. Data sampel pada [Tabel 1](#) diambil dari 1 kelas di MTs N Kaliangkrik.

Tabel 1. Data Sampel tes psikologi MTs N Kaliangkrik

NO	Nilai Tes							JURUSAN
	V	NV	I	PB	PP	PU	DK	
1	110	101	106	109	121	103	112	IPA
2	101	96	99	105	99	99	99	IPS
3	107	100	104	107	113	101	107	IPS
4	107	100	104	107	113	101	107	IPS
5	107	105	106	109	110	103	107	IPS
6	111	120	116	115	115	103	109	IPA
7	105	100	103	101	113	101	107	IPA
8	111	113	112	119	113	101	107	IPA
9	116	120	118	115	131	103	117	IPA
10	111	113	112	107	119	107	113	IPA
11	111	115	113	109	115	109	112	IPA
12	109	113	111	107	119	101	110	IPA
13	113	120	117	115	121	103	112	IPA
14	110	111	110	105	127	99	113	IPA
15	113	104	109	113	119	107	113	IPS

3.2. Analisis menggunakan Algoritma C4.5

Pada tahapan analisis menggunakan metode Algoritma C4.5 diawali dengan pengelompokan nilai tes psikologi peserta didik dan memetakan data sampel sesuai pengelompokan nilai seperti yang ditampilkan pada Tabel 2. Dari data pada Tabel 1 kemudian dibuat pengelompokan sesuai data sampel pada Tabel 1 menjadi seperti pada Tabel 3.

Tabel 2. Pengelompokan Nilai Test

No.	Range Nilai	Klasifikasi
1	70-79	Sangat Rendah
2	80-90	Rendah
3	91-110	Normal
4	111-120	Tinggi
5	>120	Superior

Pada Tabel 3, V merupakan Kemampuan Verbal, NV merupakan Kemampuan Non Verbal, I merupakan Intelegensi Umum, PB merupakan Pengetahuan Bahasa, PP merupakan Pengetahuan Pasti, PU merupakan Pengetahuan Umum dan DK merupakan Daya Kualitatif. Dilanjutkan langkah pembuatan pohon keputusan dengan metode C4.5.

Tabel 3. Hasil Pengelompokan Nilai

No	Nilai Test							Jurusan
	V	NV	I	PB	PP	PU	DK	
1	Normal	Normal	Normal	Normal	Superior	Normal	Tinggi	IPA
2	Normal	Normal	Normal	Normal	Normal	Normal	Normal	IPS
3	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
4	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
5	Normal	Normal	Normal	Normal	Normal	Normal	Normal	IPS
6	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Normal	Normal	IPA
7	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPA
8	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Normal	Normal	IPA
9	Tinggi	Tinggi	Tinggi	Tinggi	Superior	Normal	Tinggi	IPA
10	Tinggi	Tinggi	Tinggi	Normal	Tinggi	Tinggi	Tinggi	IPA
11	Tinggi	Tinggi	Tinggi	Normal	Tinggi	Normal	Tinggi	IPA
12	Normal	Tinggi	Tinggi	Normal	Tinggi	Normal	Normal	IPA
13	Tinggi	Tinggi	Tinggi	Tinggi	Superior	Normal	Tinggi	IPA
14	Normal	Tinggi	Normal	Normal	Superior	Normal	Tinggi	IPA
15	Tinggi	Normal	Normal	Tinggi	Tinggi	Normal	Tinggi	IPS

a. Mencari atribut *root* (akar)

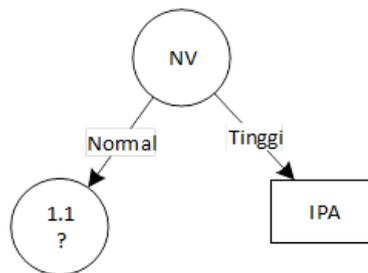
Dari data hasil pengelompokan nilai pada Tabel 3 diketahui bahwa jumlah kasus (S) sebanyak 15 kasus dengan Jumlah Kasus – IPA (S1) = 10 dan Jumlah Kasus – IPS (S2) = 5. Dengan menggunakan rumus penghitungan entropy dan gain pada Eq (1) diketahui entropy total :

$$Gain(Total, V) = 0.918296 - \left(\left(\frac{8}{15} \times 1 \right) + \left(\frac{7}{15} \times 0.591673 \right) \right)$$

$$Gain(Total, V) = 0.108849$$

b. Mencari *entropy* dan *gain* pada *node* 1

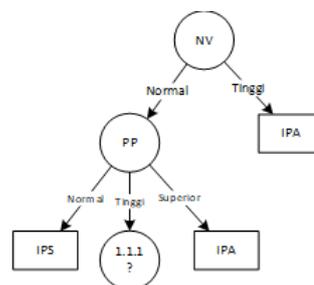
Nilai Entropy dan Gain yang ada untuk tes Kemampuan Verbal, tes Kemampuan *Non Verbal*, tes Intelegensi Umum, tes Pengetahuan Bahasa, tes Pengetahuan Pasti, tes Pengetahuan Umum dan tes Daya Kualitatif dapat dilihat pada Tabel 4. Atribut yang dapat digunakan sebagai akar adalah atribut yang memiliki Gain tertinggi, pada data Tabel 4 atribut yang memiliki *Gain* tertinggi adalah NV (tes Kemampuan *Non Verbal*), sehingga NV dapat digunakan sebagai akar. Sehingga dapat digambarkan pohon pada node pertama seperti yang terlihat pada Gambar 2.



Gambar 2. Pohon Keputusan Node 1

c. Mencari *Entropy* dan *gain* pada *node* 1.1

Langkah selanjutnya adalah mencari atribut untuk *node* 1.1 yang saat ini masih belum diketahui. Langkah yang digunakan sama dengan cara pada mencari attribut akar tetapi data yang digunakan hanya data dengan atribut NV Normal dengan data dapat dilihat pada Tabel 5. Data tersebut kemudian dihitung *entropy* dan *gain* pada *node* 1.1 sehingga didapatkan hasil sesuai pada Tabel 6. Dari nilai *gain* pada Tabel 6 diketahui kalau Node 1.1 adalah atribut PP (tes Kemampuan Pasti) dengan nilai gain 0.399534, sehingga pohon keputusan yang didapat seperti Gambar 3.



Gambar 3. Pohon Keputusan Node 1.1

Tabel 4. Hasil perhitungan *entropy* dan *gain* pada *node* 1

Node	Atribut	Value	Jml Kasus (S)	IPA (S1)	IPS (S2)	Entropy	Gain
1	Total		15	18	18	1,3547556	
	V						0,1917102
		Rata-rata	17	6	11	1,265200	
		Diatas Rata-rata	19	11	8	1,071644	
	NV						-0,304549
		Rata-rata	22	6	16	1,728479	
		Diatas Rata-rata	11	9	2	2,900052	
		Cerdas	3	3	0	0,000000	
	I						-1,613447
		Rata-rata	21	5	16	2,057062	
		Diatas Rata-rata	15	13	2	4,243800	
	PB						0,3547556
		Rata-rata	22	11	11	1,000000	
		Diatas Rata-rata	14	7	7	1,000000	
	PP						-0,206173
		Rata-rata	2	1	1	1,000000	
		Diatas Rata-rata	24	9	15	1,186742	
		Cerdas	10	8	2	2,571162	
	PU						0,3547556
		Rata-rata	36	18	18	1,000000	
	DK						0,1718525
		Rata-rata	15	5	10	1,349984	
		Diatas Rata-rata	20	12	8	1,116738	
		Cerdas	1	1	0	0	

Tabel 5. Data dengan *filter* atribut NV bernilai normal

No	Nilai Test							Jurusan
	V	NV	I	PB	PP	PU	DK	
1	Normal	Normal	Normal	Normal	Superior	Normal	Tinggi	IPA
2	Normal	Normal	Normal	Normal	Normal	Normal	Normal	IPS
3	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
4	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
5	Normal	Normal	Normal	Normal	Normal	Normal	Normal	IPS
6	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPA
7	Tinggi	Normal	Normal	Tinggi	Tinggi	Normal	Tinggi	IPS

Tabel 6. Hasil perhitungan *entropy* dan *gain* pada *node* 1.1

Node	Atribut	Value	Jml Kasus (S)	IPA (S1)	IPS (S2)	Entropy	Gain
1	Total		7	2	5	0.863121	
	V						0.076010
		Normal	6	2	4	0.918296	
		Tinggi	1	0	1	0	
	I						0
		Normal	7	2	5	0.863121	
	PB						0.076010
		Normal	6	2	4	0.918296	
		Tinggi	1	0	1	0	
	PP						0.399534
		Normal	2	0	2	0	
		Tinggi	4	1	3	0.811278	
		Superior	1	1	0	0	
	PU						0
		Normal	7	2	5	0.863121	
	DK						0.061743
		Normal	5	1	4	0.721928	
		Tinggi	2	1	1	1	

d. Mencari *entropy* dan *gain* pada *node* 1.1.1

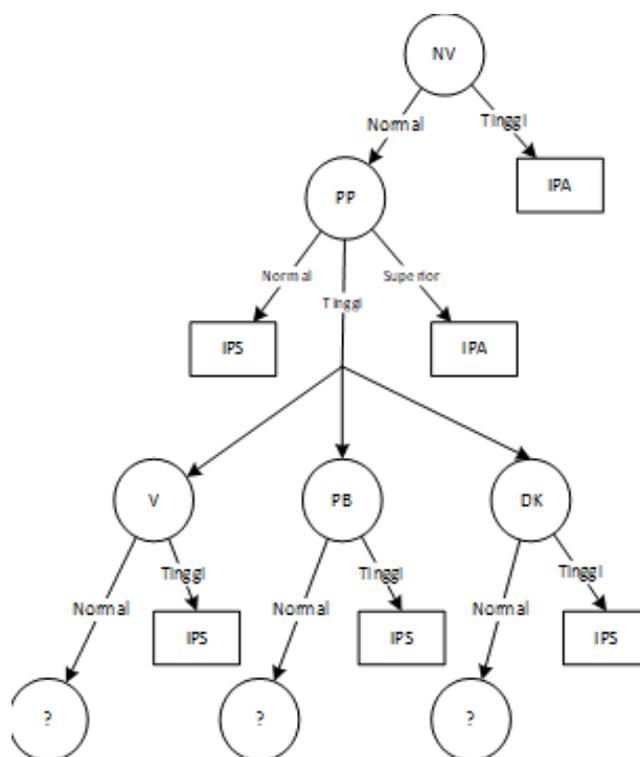
Langkah selanjutnya adalah mencari atribut untuk *node* 1.1.1 yang saat ini masih belum diketahui. Langkah yang digunakan sama dengan cara pada mencari atribut 1.1, yaitu dengan mencari nilai *entropy* dan *gain* dengan data yang digunakan hanya data dengan atribut PP (tes Kemampuan Pasti) seperti pada Tabel 7 dengan hasil *entropy* dan *gain* 1.1.1 pada Tabel 8. Dari nilai *gain* pada Tabel 7 diketahui kalau *node* 1.1.1 adalah atribut V (tes Verbal), PB (tes Pengetahuan Bahasa) atau DK (tes Daya Kualitatif) dikarenakan memiliki nilai *gain* yang sama. Sehingga pohon keputusan yang didapat seperti Gambar 4.

Tabel 7. Data dengan Filter Atribut PP Bernilai Tinggi

No	Nilai Test							Jurusan
	V	NV	I	PB	PP	PU	DK	
1	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
2	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
3	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPA
4	Tinggi	Normal	Normal	Tinggi	Tinggi	Normal	Tinggi	IPS

Tabel 8. Hasil perhitungan *entropy* dan *gain* pada *node* 1.1.1

Node	Atribut	Value	Jml Kasus (S)	IPA (S1)	IPS (S2)	Entropy	Gain
1.1.1	Total		4	1	3	0.811278	
	V						0.122556
		Normal	3	1	2	0.918296	
		Tinggi	1	0	1	0	
	I						0
		Normal	4	1	3	0.811278	
	PB						0.122556
		Normal	3	1	2	0.918296	
		Tinggi	1	0	1	0	
	PU						0
		Normal	4	1	3	0.811278	
	DK						0.122556
		Normal	3	1	2	0.918296	
		Tinggi	1	0	1	0	



Gambar 4. Pohon keputusan *node* 1.1.1

e. Mencari *entropy* dan *gain* pada *node* Selanjutnya

Langkah selanjutnya adalah mencari atribut untuk *node* selanjutnya yang saat ini masih belum diketahui, langkah yang digunakan sama dengan cara pada mencari atribut 1.1.1, yaitu dengan mencari nilai *entropy* dan *gain*, akan tetapi data yang digunakan hanya data dengan atribut V Normal yang hasilnya ada pada Tabel 9 dan Tabel 10.

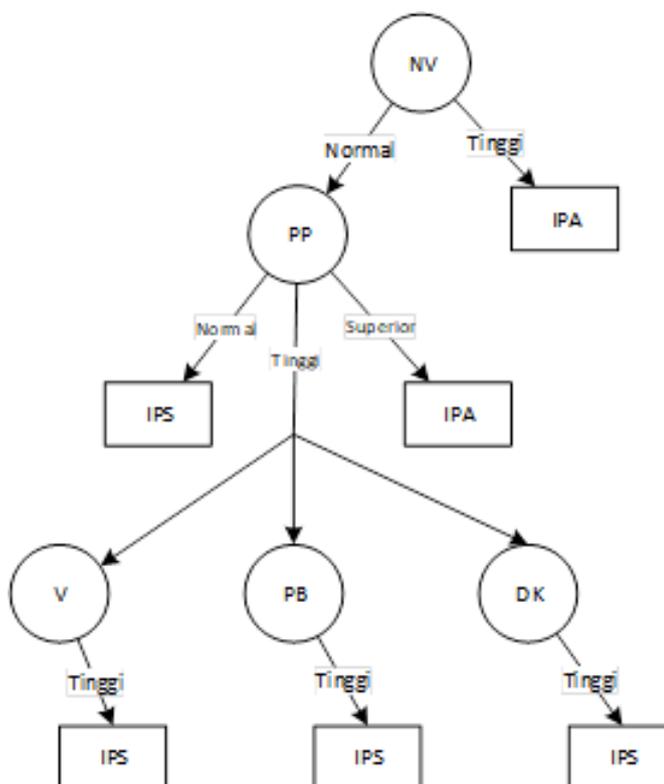
Tabel 9. Data dengan Filter Atribut V Bernilai Normal

No	V	NV	I	Nilai Test				Jurusan
				PB	PP	PU	DK	
1	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
2	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
3	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPA

Tabel 10. Hasil Perhitungan Entropy dan Gain pada Node Selanjutnya

Node	Atribut	Value	Jml Kasus (S)	IPA (S1)	IPS (S2)	Entropy	Gain
1	Total		3	1	2	0.918296	
	I						0
		Normal	3	1	2	0.918296	
	PB						0
		Normal	3	1	2	0.918296	
	PU						0
		Normal	3	1	2	0.918296	
	DK						0
		Normal	3	1	2	0.918296	

Dari Tabel 10 tidak ditemukan gain yang memiliki nilai tertinggi (gain bernilai 0) sehingga atribut selanjutnya tidak ditemukan. Dengan demikian pembuatan pohon keputusan dihentikan, sehingga hasil akhir pohon keputusan seperti yang terlihat pada Gambar 5.



Gambar 5. Pohon keputusan hasil metode C4.5

f. Penerapan pola pada kasus baru

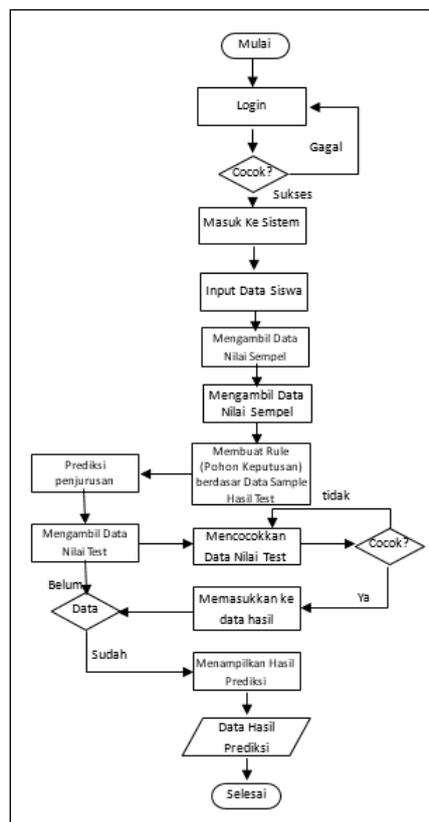
Setelah pohon keputusan selesai dibuat maka dapat menerapkan kasus baru terhadap pohon keputusan tersebut. Dengan kasus pertama IF NV = Tinggi THEN Jurusan = IPA dan kasus kedua IF NV = Normal AND PP = Tinggi AND V = Tinggi THEN Jurusan = IPS. Dari 2 kasus tersebut dibuat prediksinya seperti terlihat pada [Tabel 11](#).

Tabel 11. Hasil Prediksi Pola Kasus Baru

No	Nilai Test							Jurusan
	V	NV	I	PB	PP	PU	DK	
1	Normal	Tinggi	Tinggi	Normal	Tinggi	Normal	Normal	IPA
2	Tinggi	Normal	Normal	Tinggi	Tinggi	Normal	Normal	IPS

3.3. Pengembangan sistem

Setelah data selesai dianalisis menggunakan Algoritma C4.5 pada *spreadsheet*, maka dilanjutkan pengembangan sistem. Sistem yang dikembangkan seperti yang terlihat pada [Gambar 6](#) diawali dengan memasukkan data tes oleh petugas ke dalam sistem. Petugas mengisi/unggah data sampel terdahulu yang sudah ada hasil prediksinya. Secara otomatis, sistem akan membuat rule (pohon keputusan) menggunakan metode C4.5. Selanjutnya petugas mengisi/unggah data hasil tes IQ dan sistem akan secara otomatis memprediksi dan menampilkan hasil prediksi berdasar hasil pohon keputusan yang sudah dibuat.



Gambar 6. Flowchart Sistem yang Diajukan

3.4. Implementasi sistem dan pengujian

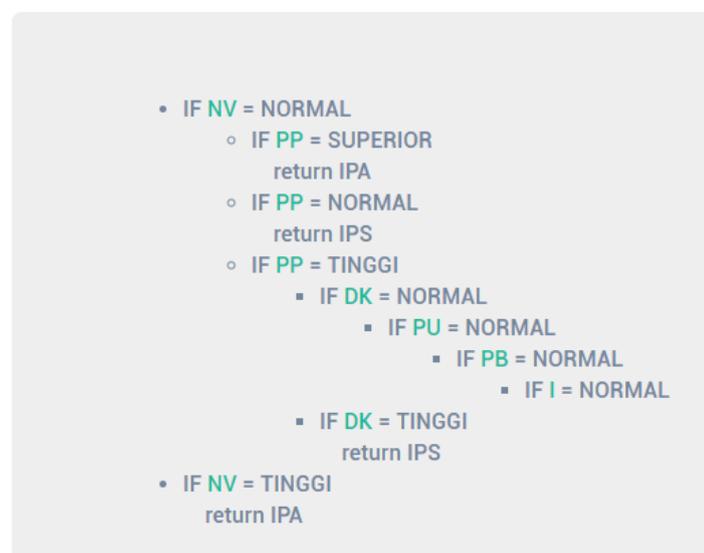
Sistem yang telah dikembangkan dilakukan pengujian sistem. Semua fungsi dari input data siswa, input data nilai, input data tes, pembuatan pohon keputusan sampai dengan pemilihan jurusan sudah sesuai dengan desain sistem yang ada. Pada aplikasi data yang akan digunakan untuk membuat pohon keputusan dapat dilihat pada [Gambar 7](#).

No.	V	NV	I	PB	PP	PU	DK	Jurusan
1	Normal	Normal	Normal	Normal	Superior	Normal	Tinggi	IPA
2	Normal	Normal	Normal	Normal	Normal	Normal	Normal	IPS
3	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
4	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPS
5	Normal	Normal	Normal	Normal	Normal	Normal	Normal	IPS
6	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Normal	Normal	IPA
7	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	IPA
8	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Normal	Normal	IPA
9	Tinggi	Tinggi	Tinggi	Tinggi	Superior	Normal	Tinggi	IPA
10	Tinggi	Tinggi	Tinggi	Normal	Tinggi	Tinggi	Tinggi	IPA
11	Tinggi	Tinggi	Tinggi	Normal	Tinggi	Normal	Tinggi	IPA
12	Normal	Tinggi	Tinggi	Normal	Tinggi	Normal	Normal	IPA
13	Tinggi	Tinggi	Tinggi	Tinggi	Superior	Normal	Tinggi	IPA
14	Normal	Tinggi	Normal	Normal	Superior	Normal	Tinggi	IPA
15	Tinggi	Normal	Normal	Tinggi	Tinggi	Normal	Tinggi	IPS

Gambar 7. Data Sample Untuk Perhitungan

Hasil pohon keputusan dari data yang ada dapat dilihat pada [Gambar 8](#). Pohon keputusan pada [Gambar 8](#) menunjukkan bahwa (1) Jika NV Tinggi maka Jurusan IPA, (2) Jika NV Normal maka PP, (3) Jika PP Superior maka Jurusan IPA, (4) Jika PP Normal maka Jurusan IPS, (5) Jika PP Tinggi maka DK, (6) Jika DK Tinggi maka Jurusan IPS, (7) Jika DK Normal maka PU dan (8) Untuk selanjutnya tidak ditemukan hasil.

Hasil Pohon Keputusan



Gambar 8. Data Hasil Pohon Keputusan pada Sistem

Sistem yang dikembangkan telah berhasil melakukan prediksi jurusan pada MTs N Kaliangkrik dengan data hasil perhitungan pada spreadsheet seperti pada Gambar 9 selanjutnya ditampilkan pada sistem sesuai Gambar 10.

1	NO	NAMA	KELAS	SEKOLAH	Tanggal Lahir	Tanggal Tes	SC
2							
3	001	Adi Asmo Bagus	MTs Negeri Kaliangkrik	IX A	06 November 2003	12 Desember 2017	IPA
4	002	Adiah	MTs Negeri Kaliangkrik	IX A	14 Februari 2002	12 Desember 2017	IPS
5	003	Akadiyah Mafudoh	MTs Negeri Kaliangkrik	IX A	05 September 2002	12 Desember 2017	IPS
6	004	Alvien Mafaza	MTs Negeri Kaliangkrik	IX A	21 April 2003	12 Desember 2017	IPS
7	005	Alwi Muhammad Kafi	MTs Negeri Kaliangkrik	IX A	28 Juli 2003	12 Desember 2017	IPS
8	006	Alya Nur Aini Marifah	MTs Negeri Kaliangkrik	IX A	04 Juni 2003	12 Desember 2017	IPA
9	007	Andira Pramudita	MTs Negeri Kaliangkrik	IX A	25 Desember 2002	12 Desember 2017	IPA
10	008	anis Muflikhah	MTs Negeri Kaliangkrik	IX A	30 Agustus 2002	12 Desember 2017	IPA
11	009	Bintan Zubdatun Nisa'	MTs Negeri Kaliangkrik	IX A	24 November 2003	12 Desember 2017	IPA
12	010	Catur Fajar Nugroho	MTs Negeri Kaliangkrik	IX A	01 Januari 2003	12 Desember 2017	IPA
13	011	Daniyatun	MTs Negeri Kaliangkrik	IX A	27 Desember 2003	12 Desember 2017	IPA
14	012	Dewi Utami	MTs Negeri Kaliangkrik	IX A	04 November 2002	12 Desember 2017	IPA
15	013	Dini Renati	MTs Negeri Kaliangkrik	IX A	08 September 2003	12 Desember 2017	IPA
16	014	Divi Kamilasari	MTs Negeri Kaliangkrik	IX A	10 April 2002	12 Desember 2017	IPA
17	015	Endah Ayu Mawadah	MTs Negeri Kaliangkrik	IX A	13 Juni 2002	12 Desember 2017	IPS

Gambar 9. Data Hasil Prediksi Jurusan Pada Sistem Lama

Pencarian

No.	NISN	Nama Lengkap	V	NV	I	PB	PP	PU	DK	Prediksi Jurusan
1	000000001	Adi Asmo Bagus	Normal	Normal	Normal	Normal	Superior	Normal	Tinggi	IPA
2	000000002	Adiah	Normal	Normal	Normal	Normal	Normal	Normal	Normal	IPS
3	000000003	Akadiyah Mafudoh	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	-
4	000000004	Alvien Mafaza	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	-
5	000000005	Alwi Muhammad Kafi	Normal	Normal	Normal	Normal	Normal	Normal	Normal	IPS
6	000000006	Alya Nur Aini Marifah	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Normal	Normal	IPA
7	000000007	Andira Pramudita	Normal	Normal	Normal	Normal	Tinggi	Normal	Normal	-
8	000000008	anis Muflikhah	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Normal	Normal	IPA
9	000000009	Bintan Zubdatun Nisa'	Tinggi	Tinggi	Tinggi	Tinggi	Superior	Normal	Tinggi	IPA
10	000000010	Catur Fajar Nugroho	Tinggi	Tinggi	Tinggi	Normal	Tinggi	Normal	Tinggi	IPA
11	000000011	Daniyatun	Tinggi	Tinggi	Tinggi	Normal	Tinggi	Normal	Tinggi	IPA
12	000000012	Dewi Utami	Normal	Tinggi	Tinggi	Normal	Tinggi	Normal	Normal	IPA
13	000000013	Dini Renati	Tinggi	Tinggi	Tinggi	Tinggi	Superior	Normal	Tinggi	IPA
14	000000014	Divi Kamilasari	Normal	Tinggi	Normal	Normal	Superior	Normal	Tinggi	IPA
15	000000015	Endah Ayu Mawadah	Tinggi	Normal	Normal	Tinggi	Tinggi	Normal	Tinggi	IPS

Gambar 10. Data Hasil Prediksi Jurusan Pada Aplikasi

Selain pengujian fungsi, fitur dan antarmuka pada sistem, pengujian validasi data *output* juga dilakukan. Berikut ringkasan data prediksi antara penghitungan *spreadsheet* dengan sistem yang dikembangkan seperti pada Tabel 12. Dapat diketahui bahwa akurasi data prediksi $Kesamaan = \frac{12}{15} \times 100\% = 80\%$. Dimana 3 buah ketidaksamaan dihasilkan dari data sistem yang tidak menghasilkan prediksi. Dengan data ini diketahui bahwa sistem yang dikembangkan berbasis metode C4.5 dapat digunakan untuk memprediksi jurusan berdasarkan atribut dan nilai yang dimasukkan.

Tabel 12. Perbandingan hasil prediksi perhitungan *spreadsheet* dengan sistem

No	Nama Siswa	Prediksi		Hasil Perbandingan
		Sistem Lama	Sistem Baru	
1	Adi Asmo Bagus	IPA	IPA	Sama
2	Adiah	IPS	IPS	Sama
3	Akadiyah Mafudoh	IPS	-	-
4	Alvien Mafaza	IPS	-	-
5	Alwi Muhammad Kafi	IPS	IPS	Sama
6	Alya Nur Aini Marifah	IPA	IPA	Sama
7	Andira Pramudita	IPA	-	-
8	anis Muflikhah	IPA	IPA	Sama
9	Bintan Zubdatun Nisa'	IPA	IPA	Sama
10	Catur Fajar Nugroho	IPA	IPA	Sama
11	Daniyatun	IPA	IPA	Sama
12	Dewi Utami	IPA	IPA	Sama
13	Dini Renati	IPA	IPA	Sama
14	Diva Kamilasari	IPA	IPA	Sama
15	Endah Ayu Mawadah	IPS	IPS	Sama

4. Kesimpulan

Berdasarkan hasil dan pembahasan maka dapat diambil kesimpulan sebagai berikut:

1. Sistem yang dikembangkan untuk memprediksi jurusan pada MTs N Kaliangkrik memudahkan penyimpanan data siswa dan data tes sehingga data lebih aman tersimpan. Sistem juga memberi kemudahan adanya fitur import data siswa, data nilai tes, sehingga hasil prediksi ng dapat diketahui lebih cepat. Sistem memiliki akurasi 80% (baik) dalam memprediksi jurusan.
2. Sistem yang dikembangkan masih perlu perbaikan khususnya jika data nilai yang dianalisa diluar jangkauan dari pohon keputusan maka sistem tidak dapat memberikan prediksinya.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih atas dukungan MTs N Kaliangkrik dan Pelita Harapan Bangsa Lembaga Pengembangan Sumber Daya Manusia atas kerjasamanya dalam kemudahan penulis mendapatkan akses data untuk penelitian.

Referensi

Harahap, F. (2015). Penerapan Data Mining dalam Memprediksi Pembelian Cat. *Konferensi Nasional Sistem & Informatika*.

- Latifah, K., Wibowo, S., & Nada, N. Q. (2018). Analisis dan Penerapan Algoritma C4.5 dalam Data Mining untuk Menunjang Strategi Promosi Prodi Informatika UPGRIS. *Jurnal Teknik Informatika*, 11(2), 109–120. <http://dx.doi.org/10.15408/jti.v11i2.6706>
- Nurzahputra, A., & Muslim, M. A. (2017). Peningkatan Akurasi Pada Algoritma C4.5 Menggunakan Adaboost Untuk Meminimalkan Resiko Kredit. *Pros Iding SNATIF Ke -4*.
- Pradipta, A., Hartama, D., Wanto, A., Saifullah, S., & Jalaluddin, J. (2019). The Application of Data Mining in Determining Timely Graduation Using the C45 Algorithm. *IJISTECH (International Journal of Information System & Technology)*, 3(1), 31–36. <https://doi.org/10.30645/ijistech.v3i1.30>
- R, L., & G, M. A. (2012). Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit. *Jurnal Komputer Dan Informatika (KOMPUTA)*.
- Rijayana, I., Fikri, M. I., Razaq, I. F., & Achlafass, R. S. (2021). Using Data Mining with C45 Algorithm for Student Data Classification. *PSYCHOLOGY AND EDUCATION*, 58(3), 670–673.
- Rizki, M. Y., Sartika, E., Pratama, Y., Rasika, S., & Windarto, A. P. (2020). Implementasi C45 Dalam Memprediksi Index Prestasi Mahasiswa / i Menurut Kebiasaan Belajar. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 297–303.
- Sakerebau, J. (2018). Memahami Peran Psikologi Pendidikan Bagi Pembelajaran. *Jurnal Teologi Dan Pendidikan*.
- Sinambela, M. S., P, M., & E, R. (2016). Sistem Pendukung Keputusan Peminatan Jurusan Siswa Di Sma Menggunakan Algoritma Klasifikasi Data Mining C4.5. *E-Proceeding of Applied Science*, 858.
- Subiyanto, S. (2017). Recommendation System For Vocational Major Streaming By C4.5 Algorithm. *Jurnal Kependidikan*, 1(1), 139–149.
- Sugiyarti, E., Jasmi, K. A., Basiron, B., Huda, M., Shankar, K., & Maselena, A. (2018). Decision Support System of Scholarship Grantee Selection using Data Mining. *International Journal of Pure and Applied Mathematics*, 119(15), 2239–2250. <https://doi.org/10.5772/47788>
- Yuliana, A., & Pratomo, B. (2017). Algoritma Decision Tree (C4.5) Untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik Tedc Bandung. *Seminar Nasional Inovasi Teknologi*.
-