

# PENERAPAN ALGORITMA C4.5 UNTUK PENENTUAN KRITERIA SISWA TELADAN BERDASARKAN PERATURAN DISIPLIN SISWA (STUDI KASUS DI SMK NEGERI 1 LINTAU BUO)

Dodi Nofri Yoliadi<sup>1)</sup>

<sup>1)</sup> Program Studi Ilmu Perpustakaan dan Informasi Islam, IAIN Batusangkar  
e-mail: dodinofriyadi@iainbatusangkar.ac.id <sup>1)</sup>

## ABSTRAK

Siswa merupakan salah satu elemen kunci dalam sistem pendidikan, maka dari itu butuh petunjuk untuk dapat dinilai dengan baik. Peraturan disiplin siswa diharapkan dapat menjadi pedoman pelaksanaan seluruh kegiatan siswa sehari-hari baik di dalam maupun di luar sekolah. Penentuan siswa teladan diukur berdasarkan kriteria yang harus dimiliki oleh setiap siswa sesuai peraturan disiplin siswa yang diterapkan di sekolah. Di SMK Negeri 1 Lintau Buo penentuan siswa teladan masih dilakukan secara manual dan hasilnya belum objektif. Sehingga untuk hasil yang diperoleh terkadang penuh pro dan kontra. *Data Mining* digunakan untuk mempelajari data yang tersedia di bidang pendidikan dan menghasilkan pengetahuan tersembunyi. Metode klasifikasi seperti Algoritma C 4.5, diterapkan pada data pendidikan dengan memanfaatkan pengetahuan tersembunyi untuk memprediksi seluruh aspek kinerja siswa. Prediksi ini juga membantu dalam penentuan kriteria siswa teladan. Hasil prediksi tersebut juga untuk perbaikan kinerja siswa secara keseluruhan terutama dalam proses pembelajaran. Penelitian dilakukan terhadap siswa SMK Negeri 1 Lintau Buo sebanyak 75 data siswa yang memiliki peringkat kelas 1, 2 dan 3 di setiap kelas. Hasil yang diperoleh dari proses klasifikasi ini adalah kriteria terbaik untuk penentuan siswa teladan dengan node yang diperoleh yaitu, Prestasi Non Akademis, Kehadiran serta Kegiatan sosial. Keakuratan proses klasifikasi pada *Data Training* diperoleh sebesar 1.00, pada *Data Testing1* diperoleh sebesar 0.90 serta pada *Data Testing2* diperoleh sebesar 1.00.

**Kata Kunci:** Siswa Teladan, *Data Mining*, Kriteria, Klasifikasi, *Algoritma C 4.5*

## 1. PENDAHULUAN

Siswa merupakan salah satu elemen kunci dalam sistem pendidikan, maka dari itu butuh petunjuk untuk dapat dinilai dengan baik. Peraturan disiplin siswa diharapkan dapat menjadi pedoman pelaksanaan seluruh kegiatan siswa sehari-hari baik di dalam maupun di luar sekolah.

Penentuan siswa teladan diukur berdasarkan spesifikasi/kriteria yang harus dimiliki oleh setiap siswa sesuai peraturan disiplin siswa yang diterapkan di sekolah. Di SMK Negeri 1 Lintau Buo penentuan siswa teladan masih dilakukan secara manual dan hasilnya belum objektif. Sehingga untuk hasil yang diperoleh terkadang penuh pro dan kontra.

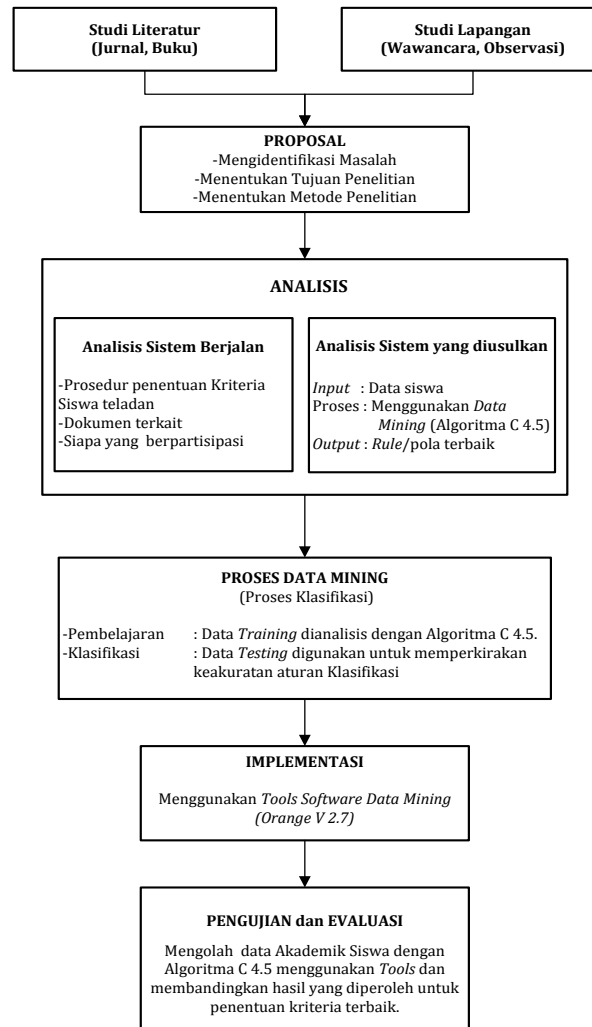
*Data Mining* digunakan untuk mempelajari data yang tersedia di bidang pendidikan dan menghasilkan pengetahuan tersembunyi. Metode klasifikasi seperti Algoritma C 4.5, diterapkan pada data pendidikan dengan memanfaatkan pengetahuan tersembunyi untuk memprediksi seluruh aspek kinerja siswa

Tentang Algoritma C 4.5 juga diteliti oleh HSSINA, dkk, yang membandingkan ID3 / C4.5, C4.5 / C5.0 dan C5.0 / Kereta. Hasil yang diperoleh dari penelitian tersebut metode yang paling kuat dan disukai dalam mesin pembelajaran adalah C4.5.

## 2. METODE PENELITIAN

Metodologi adalah kerangka teoritis yang dipergunakan oleh penulis untuk menganalisa, mengerjakan / mengatasi masalah yang dihadapi. Kerangka teoritis atau kerangka ilmiah merupakan metode-metode ilmiah yang akan diterapkan dalam pelaksanaan tugas (Keraf, 2001:310).

Dalam metodologi penelitian ada beberapa urutan kerangka kerja yang harus diikuti, urutan kerangka kerja ini merupakan gambaran dari langkah-langkah yang harus dilalui agar penelitian ini bisa berjalan dengan baik. Adapun kerangka kerja penelitian dapat dilihat pada Gambar 3.1.

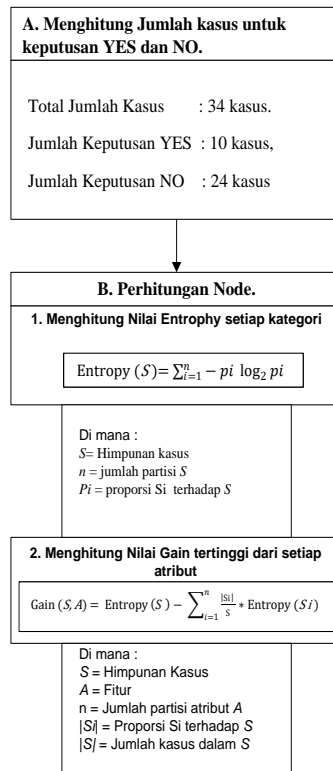


Gambar 1. Kerangka Kerja Penelitian

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Proses Klasifikasi dengan Algoritma C 4.5

Proses klasifikasi memperhatikan jumlah kasus, atribut yang dijadikan target keputusan dan perhitungan pada *Node*. Untuk Proses Klasifikasi menggunakan Algoritma C 4.5 menggunakan *Data Training* seperti terlihat pada Gambar 4.2.



Gambar 2. Proses Klasifikasi dengan Algoritma C 4.5 Menggunakan Data Training

**3.2 Pembuatan Pohon Keputusan**

Pohon keputusan untuk penentuan kriteria siswa teladan berasal dari hasil analisis data menggunakan teknik klasifikasi terlebih dahulu. Nilai Atribut Pohon keputusan yang ditetapkan menggunakan tipe kategorikal. Langkah-langkah dalam pembentukan pohon keputusan menggunakan Algoritma C 4.5 yaitu :

- a. Melakukan analisis data sesuai perhitungan menggunakan Algoritma C 4.5, dimulai dengan menghitung nilai *Entropy*.
- b. Melakukan perhitungan nilai *Gain* Tertinggi dari Atribut.
- c. Menggunakan hasil analisis untuk merancang pohon keputusan

**a. Proses Perhitungan *Entropy* berdasarkan Keputusan YES dan NO.**

Menghitung *Entropy* Target Keputusan :

**1 TELADAN**

TOTAL	34
YES	10
NO	24

**Entropy NO**

$$= \sum_{i=1}^n \left( - \frac{24}{34} \right) \log_2 \left( \frac{24}{34} \right)$$

$$= -0.7059 \quad * \quad -0.5025$$

$$= 0.35471$$

**Entropy YES**

$$= \sum_{i=1}^n \left( - \frac{10}{34} \right) \log_2 \left( \frac{10}{34} \right)$$

$$= -0.29412 \quad * \quad -1.7655$$

$$= 0.51927$$

Entropy (S) = Entropy YES + Entropy NO

= 0.51927 + 0.35471

= 0.87398

**2. TEMPAT TINGGAL**

ORTU Entropy (ORTU) =0.90593  
 TOTAL=28 YES =9  
 NO =19

Lakukan perhitungan yang sama untuk Atribut lainnya.

**b. Menghitung Nilai Gain**

KOS Entropy (KOS) =0.65002 = 0.01321

TOTAL =6 YES = 1  
 NO = 5

Hasil yang diperoleh pada *Node 1* :

Gain (S,Tempat Tinggal)

$$= \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{S} * \text{Entropy}(S_i)$$

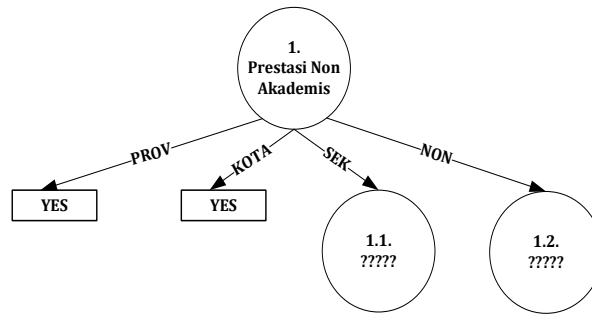
$$= 0.87398 - \left\{ \left( \frac{28}{34} \right) * 0.90593 + \left( \frac{6}{34} \right) * 0.65002 \right\}$$

Tabel 1. Hasil Perhitungan *Node 1*

NO	ATRIBUT / KATEGORI	JUMLAH KASUS			ENTROPY	GAIN
		YES	NO	TOTAL		
1	TELADAN	10	24	34	0.87398	
2	TEMPAT TINGGAL					
	ORTU	9	19	28	0.90593	0.01321
	KOS	1	5	6	0.65002	
3	KEHADIRAN					
	BAIK	10	19	29	0.92936	0.08129
	KURANG	0	5	5	0.00000	
4	PRESTASI AKADEMIC					
	JU	0	0	0	0.00000	0.04474
	JK1	5	6	11	0.99403	
	JK2	3	9	12	0.81128	
	JK3	2	9	11	0.68404	
5	PRESTASI NON AKADEMIC					
	PROV	3	0	3	0.00000	0.61841
	KOTA	5	0	5	0.00000	
	SEK	1	2	3	0.91830	
	NON	1	22	23	0.25802	
6	KEGIATAN SOSIAL					
	PANITIA INTI	0	0	0	0.00000	0.27158
	ANGGOTA	6	1	7	0.59167	
	NON	4	23	27	0.60519	
7	KEGIATAN KEAGAMAAN					
	PANITIA INTI	0	0	0	0.00000	0.09350
	ANGGOTA	8	10	18	0.99108	
	NON	2	14	16	0.54356	

**3.3 Merancang pohon keputusan**

Berdasarkan hasil perhitungan di Tabel 4.10, Nilai *Gain* tertinggi adalah Atribut Prestasi Non Akademis sebesar 0.61841. Dengan begitu Atribut Prestasi Non Akademis dapat dijadikan *Node* Akar yaitu *Node 1*. Pohon keputusan yang didapat seperti Gambar 4.3 yaitu :



Gambar 3. Pohon Keputusan dari Akar Prestasi Non Akademis (Node 1)

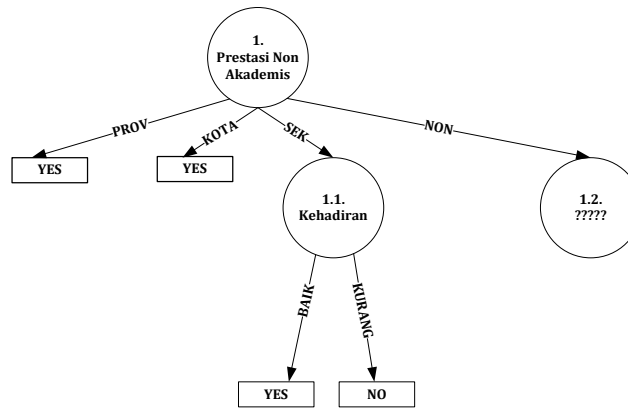
Nilai yang terdapat pada Atribut Prestasi Non Akademis adalah NAS, PROV, KOTA, SEK, NON. Berikut penjelasan di setiap cabangnya :

1. Nilai pada *Node 1* Prestasi Non Akademis-PROV sebesar 0.00000 sehingga tidak perlu dilakukan perhitungan lagi. Untuk jumlah kasus pada Prestasi Non Akademis-PROV tersebut pada keputusan YES terdapat data sebanyak 3 dan pada NO sebanyak 0. Untuk kasus ini keputusan yang didapat adalah YES.
2. Nilai pada *Node 1* Prestasi Non Akademis-KOTA sebesar 0.00000 sehingga tidak perlu dilakukan perhitungan lagi. Untuk jumlah kasus pada Prestasi Non Akademis-KOTA tersebut pada keputusan YES terdapat data sebanyak 5 dan pada NO sebanyak 0. Untuk kasus ini keputusan yang didapat adalah YES.
3. Nilai pada *Node 1* Prestasi Non Akademis-SEK sebesar 0.59167, sehingga perlu dilakukan perhitungan lagi sebagai *Node 1.1*.
4. Nilai pada *Node 1* Prestasi Non Akademis-NON sebesar 0.60519, sehingga perlu dilakukan perhitungan lagi sebagai *Node 1.2*.

Tabel 4.11 Hasil Perhitungan Node 1.1

NO	ATRIBUT / KATEGORI	JUMLAH KASUS			ENTROPY	GAIN
		YES	NO	TOTAL		
1	PRESTASI NON AKADEMIS--SEK	1	2	3	0.91830	
2	TEMPAT TINGGAL					
	ORTU	1	2	3	0.91830	0.00000
	KOS	0	0	0	0.00000	
3	KEHADIRAN					
	BAIK	1	0	1	0.00000	0.91830
	KURANG	0	2	2	0.00000	
4	PRESTASI AKADEMIS					
	JU	0	0	0	0.00000	0.91830
	JK1	0	0	0	0.00000	
	JK2	1	0	1	0.00000	
	JK3	0	2	2	0.00000	
6	KEGIATAN SOSIAL					
	PANITIA INTI	0	0	0	0.00000	0.25163
	ANGGOTA	1	1	2	1.00000	
	NON	1	0	1	0.00000	
7	KEGIATAN KEAGAMAAN					
	PANITIA INTI	0	1	1	0.00000	0.91830
	ANGGOTA	1	0	1	0.00000	
	NON	0	1	1	0.00000	

Berdasarkan hasil perhitungan di Tabel 4.11, Nilai *Gain* tertinggi yang pertama kali ditemukan adalah Atribut (Prestasi Non Akademis-SEK)---(Kehadiran) sebesar 0.91830. Dengan begitu Atribut (Prestasi Non Akademis-SEK)---(Kehadiran) dapat dijadikan sebagai *Node 1.1*. Pohon keputusan yang didapat seperti Gambar 4.4 yaitu :



Gambar 4. Pohon Keputusan dari Node1 (Prestasi Non Akademis-SEK) dan Node 1.1 (Kehadiran)

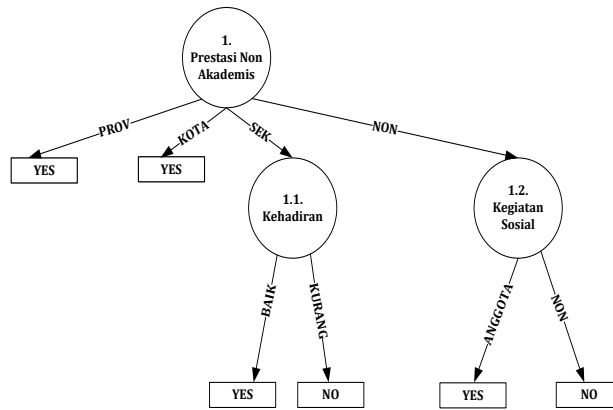
Nilai yang terdapat pada Atribut (Prestasi Non Akademis-SEK)---(Kehadiran) adalah BAIK dan KURANG. Berikut penjelasannya :

1. Nilai pada Node 1.1 (Prestasi Non Akademis-SEK)---(Kehadiran-BAIK) sebesar 0.00000 sehingga tidak perlu dilakukan perhitungan lagi. Untuk jumlah kasus pada Node 1.1 (Prestasi Non Akademis-SEK)---(Kehadiran-BAIK) tersebut pada keputusan YES terdapat data sebanyak 1 dan pada NO sebanyak 0. Untuk kasus ini keputusan yang didapat adalah YES.
2. Nilai pada Node 1.1 (Prestasi Non Akademis-SEK)---(Kehadiran- KURANG) sebesar 0.00000 sehingga tidak perlu dilakukan perhitungan lagi. Untuk jumlah kasus pada Node 1.1 (Prestasi Non Akademis-SEK)---(Kehadiran- KURANG) tersebut pada keputusan YES terdapat data sebanyak 0 dan pada NO sebanyak 2. Untuk kasus ini keputusan yang didapat adalah NO.

Tabel 4.13 Hasil Perhitungan Node 1.2

NO	ATRIBUT / KATEGORI	JUMLAH KASUS			ENTROPY	GAIN
		YES	NO	TOTAL		
1	<b>PRESTASI NON AKADEMIS--NON</b>	1	22	23	0.25802	
2	<b>TEMPAT TINGGAL</b>					
	ORTU	0	17	17	0.00000	0.08845
	KOS	1	5	6	0.65002	
3	<b>KEHADIRAN</b>					
	BAIK	1	19	20	0.28640	0.00898
	KURANG	0	3	3	0.00000	
4	<b>PRESTASI AKADEMIS</b>					
	JU	0	0	0	0.00000	0.07794
	JK1	1	6	7	0.59167	
	JK2	0	9	9	0.00000	
	JK3	0	7	7	0.00000	
6	<b>KEGIATAN SOSIAL</b>					
	PANITIA INTI	0	0	0	0.00000	0.25802
	ANGGOTA	1	0	1	0.00000	
	NON	0	22	22	0.00000	
7	<b>KEGIATAN KEAGAMAAN</b>					
	PANITIA INTI	0	1	1	0.00000	0.25802
	ANGGOTA	1	0	1	0.00000	
	NON	0	13	13	0.00000	

Berdasarkan hasil perhitungan di Tabel 4.13, Nilai Gain tertinggi yang pertama kali ditemukan adalah Atribut (Prestasi Non Akademis-NON)---(Kegiatan Sosial) sebesar 0.25802. Dengan begitu Atribut (Prestasi Non Akademis-NON)---(Kegiatan Sosial) dapat dijadikan sebagai Node cabang 1.2. Pohon keputusan yang didapat seperti Gambar 4.6 yaitu :



Gambar 5. Pohon Keputusan dari Node 1 (Prestasi Non Akademis-NON) dan Node 1.2 (Kegiatan Sosial)

Nilai yang terdapat pada Atribut (Prestasi Non Akademis-NON)---(Kegiatan Sosial) adalah PANITIA INTI, ANGGOTA dan NON. Berikut penjelasannya :

1. Nilai pada Node 1.2 (Prestasi Non Akademis-NON)---(Kegiatan Sosial—PANITIA INTI) sebesar 0.00000 sehingga tidak perlu dilakukan perhitungan lagi. Untuk jumlah kasus pada Node 1.2 (Prestasi Non Akademis-NON)---(Kegiatan Sosial—PANITIA INTI) tersebut pada keputusan YES terdapat data sebanyak 0 dan pada NO sebanyak 0. Untuk kasus ini tidak dibuatkan cabang karena tidak ada keputusan.
2. Nilai pada Node 1.2 (Prestasi Non Akademis-NON)---(Kegiatan Sosial—ANGGOTA) sebesar 0.00000 sehingga tidak perlu dilakukan perhitungan lagi. Untuk jumlah kasus pada Node 1.2 (Prestasi Non Akademis-NON)---(Kegiatan Sosial—ANGGOTA) tersebut pada keputusan YES terdapat data sebanyak 1 dan pada NO sebanyak 0. Untuk kasus ini keputusan yang didapat adalah YES.
3. Nilai pada Node 1.2 (Prestasi Non Akademis-NON)---(Kegiatan Sosial—NON) sebesar 0.00000 sehingga tidak perlu dilakukan perhitungan lagi. Untuk jumlah kasus pada Node 1.2 (Prestasi Non Akademis-NON)---(Kegiatan Sosial—NON) tersebut pada keputusan YES terdapat data sebanyak 0 dan pada NO sebanyak 22. Untuk kasus ini keputusan yang didapat adalah NO.

### 3.4 Pembentukan Rule

Rule merupakan aliran data sehingga untuk menelusurinya dimulai dari Akar sampai pada cabang dan daun. Rule biasanya berdasarkan pohon keputusan yang diperoleh dari hasil analisis data dengan metode klasifikasi. Rule menjadi kriteria dalam penentu keputusan. Berikut Rule yang diperoleh berdasarkan Pohon keputusan pada Gambar 4.6, seperti dapat dilihat pada Tabel 4.14.

Tabel 4.14 Rule yang Dihasilkan dari Penerapan Algoritma C 4.5.

No	Rule
1	Jika Prestasi Non Akademis = PROV, maka Teladan = YES
2	Jika Prestasi Non Akademis = KOTA, maka Teladan = YES
3	Jika Prestasi Non Akademis = SEK, maka Teladan = NO
4	Jika Prestasi Non Akademis = SEK, dan Kehadiran = BAIK, maka Teladan = YES
5	Jika Prestasi Non Akademis = SEK, dan Kehadiran = KURANG, maka Teladan = NO

- 
- 6      Jika Prestasi Non Akademis = NON, maka Teladan = NO
- 7      Jika Prestasi Non Akademis = NON, dan Kegiatan Sosial = NON maka Teladan = NO
- 8      Jika Prestasi Non Akademis = NON, dan Kegiatan Sosial = ANGGOTA, maka Teladan = YES
- 

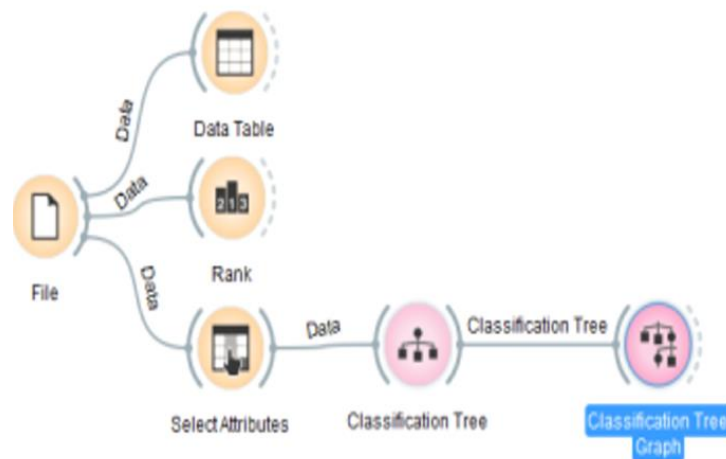
Atribut yang menjadi akar (*root*) dalam penelitian ini adalah Prestasi Non Akademis. Berarti bahwa atribut Prestasi Non Akademis merupakan atribut yang paling menentukan dalam penentuan kriteria siswa teladan. Atribut yang menjadi cabang adalah Kehadiran dan Kegiatan Sosial, yang berarti bahwa setelah kriteria Prestasi Non Akademis, atribut Kehadiran dan Kegiatan Sosial merupakan kriteria selanjutnya yang menjadi faktor penentu.

*Rule* yang dihasilkan ini bermanfaat untuk penentuan prioritas kriteria yang dapat menghasilkan keputusan terbaik. Dengan memanfaatkan *rule* ini, kegiatan untuk menentukan siswa teladan dapat terprogram dengan baik dan sesuai aturan.

### 3.5 Implementasi Menggunakan *Data Training*

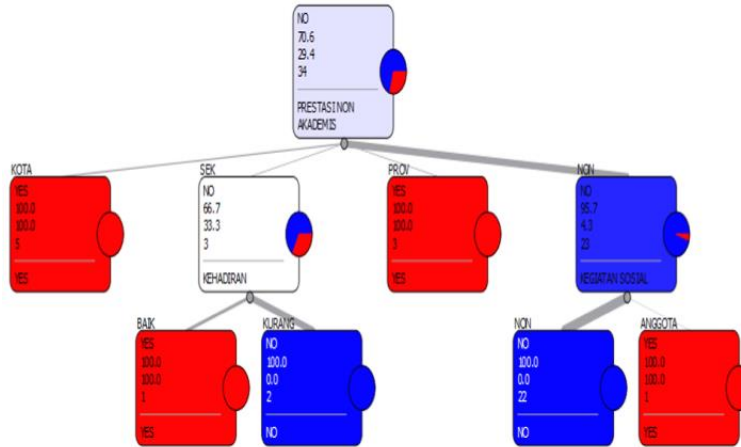
*Data Training* yang dipersiapkan dapat dilihat pada Tabel 5.1. Seperti yang sudah dijelaskan pada Bab IV, *Data Training* merupakan *Data Cleaning* yang sudah ditransformasi sesuai dengan Atribut dan kategorinya. *Data Training* yang digunakan untuk implementasi ini sebanyak 37 *record*. *File Data Training* maupun *file Data Testing* sebaiknya disesuaikan dengan tipe *file* yang dapat diterapkan pada Aplikasi Orange V 2.7 yaitu *file* yang berekstensi berbentuk seperti salah satu di bawah ini :

- a. TAB/TXT
- b. XML
- c. XLS
- d. CSV



Gambar 6. Menampilkan *Classification Tree Graph*

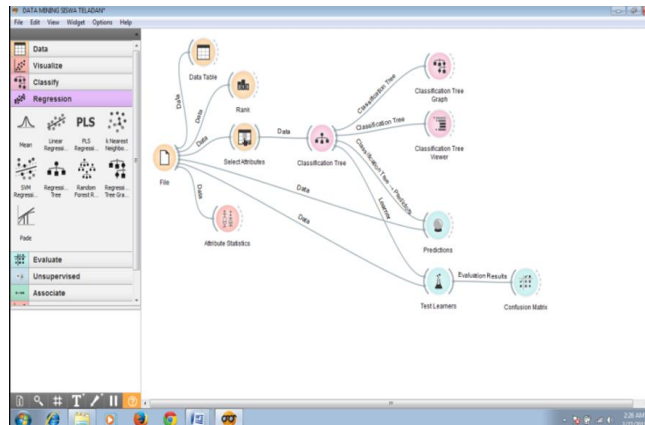




Gambar 7. Tampilan Pohon Keputusan

Classification Tree	Class	P(Class)	P(Target)	# Inst	Distribution (rel)	Distribution (abs)
└─ <root>	NO	0.706	0.706	34	0.706:0.294	24:10
├─ PRESTASI NON AKADEMIS = KOTA	YES	1.000	0.000	5	0.000:1.000	0:5
└─ PRESTASI NON AKADEMIS = NON	NO	0.957	0.957	23	0.957:0.043	22:1
├─ KEGIATAN SOSIAL = ANGGOTA	YES	1.000	0.000	1	0.000:1.000	0:1
└─ KEGIATAN SOSIAL = NON	NO	1.000	1.000	22	1.000:0.000	22:0
├─ PRESTASI NON AKADEMIS = PROV	YES	1.000	0.000	3	0.000:1.000	0:3
└─ PRESTASI NON AKADEMIS = SEK	NO	0.667	0.667	3	0.667:0.333	2:1
├─ KEHADIRAN = BAIK	YES	1.000	0.000	1	0.000:1.000	0:1
└─ KEHADIRAN = KURANG	NO	1.000	1.000	2	1.000:0.000	2:0

Gambar 8. Tampilan Rule Hasil Klasifikasi Menggunakan Data Training



Gambar 9. Tampilan Skema Data Mining untuk Data Training

	NO	YES	
NO	100.0 %	0.0 %	24
YES	20.0 %	80.0 %	10
	26	8	34

Gambar 10. Keakuratan *Data Training* Dilihat dari *Proportions of True*

	NO	YES	
NO	24	0	24
YES	2	8	10
	26	8	34

Gambar 11. Keakuratan *Data Training* Dilihat dari Jumlah Data

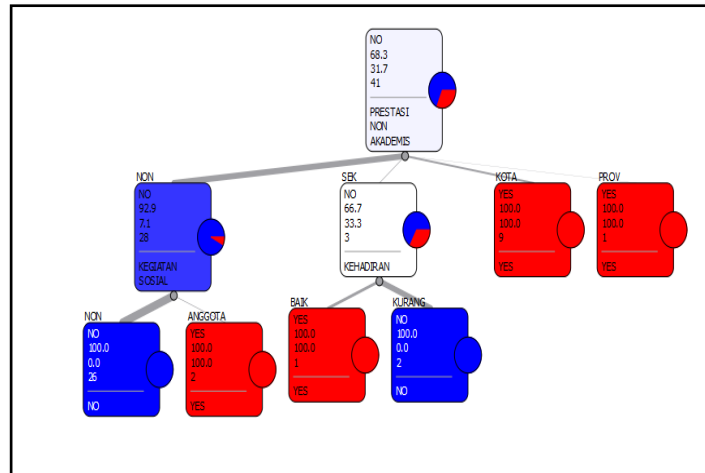
Matrik keakuratan dengan jumlah total populasi 34 data yang terlihat pada Gambar 5.24 dapat dijelaskan sebagai berikut :

1. Pada *Correct Class=NO*, pada *Data Prediction=NO* diperoleh data akurat sebesar 24 data.
2. Pada *Correct Class=NO*, pada *Data Prediction=YES* diperoleh data sebesar 0 data.
3. Pada *Correct Class=YES*, pada *Data Prediction=NO* diperoleh data sebesar 2 data.
4. Pada *Correct Class=YES*, pada *Data Prediction=YES* diperoleh data akurat YES sebesar 8 data.

Berdasarkan hasil *Confusion Matrix* dan tingkat nilai diagnosa keakuratan menurut Gorunescu (2011) yaitu,

- a. Akurasi bernilai 0.90 – 1.00 = *excellent classification*
- b. Akurasi bernilai 0.80 – 0.90 = *good classification*
- c. Akurasi bernilai 0.70 – 0.80 = *fair classification*
- d. Akurasi bernilai 0.60 – 0.70 = *poor classification*
- e. Akurasi bernilai 0.50 – 0.60 = *failure*

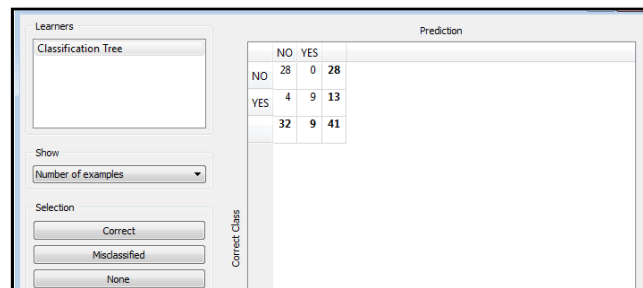
dapat terlihat bahwa jumlah keakuratan *Data Training* yang bernilai *True* diperoleh dengan cara menghitung jumlah data akurat *True* pada *Correct Class=NO* dan *Data Prediction=NO* sebesar 24 data ditambahkan dengan jumlah data akurat *True Correct Class=YES* dan *Data Prediction=YES* sebesar 8 data, sehingga total data yang bernilai akurat *True* sebanyak 32 data. Selanjutnya menghitung tingkat nilai akurasi diperoleh berdasarkan jumlah total data akurat *True* sebanyak 28 data dari total populasi sebanyak 32 data diperoleh hasil perhitungan menjadi 1.00. Melihat aturan untuk menentukan tingkat keakuratan, keakuratan pada *Data Training* termasuk kategori *excellent classification*.



Gambar 12. Tampilan Pohon Keputusan pada *Data Testing 1*

Classification Tree	Class	P(Class)	P(Target)	# Inst	Distribution (rel)	Distribution (abs)
<root>	NO	0.683	0.683	41	0.683:0.317	28:13
PRESTASI NON AKADEMIS = KOTA	YES	1.000	0.000	9	0.000:1.000	0:9
PRESTASI NON AKADEMIS = NON	NO	0.929	0.929	28	0.929:0.071	26:2
KEGIATAN SOSIAL = ANGGOTA	YES	1.000	0.000	2	0.000:1.000	0:2
KEGIATAN SOSIAL = NON	NO	1.000	1.000	26	1.000:0.000	26:0
PRESTASI NON AKADEMIS = PROV	YES	1.000	0.000	1	0.000:1.000	0:1
PRESTASI NON AKADEMIS = SEK	NO	0.667	0.667	3	0.667:0.333	2:1
KEHADIRAN = BAIK	YES	1.000	0.000	1	0.000:1.000	0:1
KEHADIRAN = KURANG	NO	1.000	1.000	2	1.000:0.000	2:0

Gambar 13. Tampilan Rule untuk *Data Testing 1*



Gambar 14. Keakuratan *Data Testing 1* Dilihat dari Jumlah Data

Matrik keakuratan dengan jumlah total populasi 41 data, pada Gambar 5.10, maka terlihat :

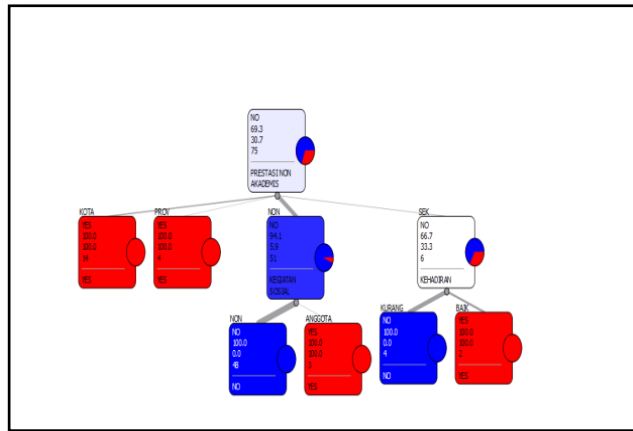
1. Pada *Correct Class=NO*, pada *Data Prediction=NO* diperoleh data akurat sebesar 28 data.
2. Pada *Correct Class=NO*, pada *Data Prediction=YES* diperoleh data sebesar 0 data.
3. Pada *Correct Class=YES*, pada *Data Prediction=NO* diperoleh data sebesar 4 data.
4. Pada *Correct Class=YES*, pada *Data Prediction=YES* diperoleh data akurat YES sebesar 9 data.

Berdasarkan hasil *Confusion Matrix* dan tingkat nilai diagnosa keakuratan menurut Gorunescu (2011) yaitu,

- a. Akurasi bernilai 0.90 – 1.00 = *excellent classification*
- b. Akurasi bernilai 0.80 – 0.90 = *good classification*
- c. Akurasi bernilai 0.70 – 0.80 = *fair classification*
- d. Akurasi bernilai 0.60 – 0.70 = *poor classification*
- e. Akurasi bernilai 0.50 – 0.60 = *failure*

dapat terlihat bahwa jumlah keakuratan *Data Training* yang bernilai *True* diperoleh dengan cara menghitung jumlah data akurat *True* pada *Correct Class=NO* dan *Data Prediction=NO* sebesar 28 data ditambahkan dengan jumlah data akurat *True Correct Class=YES* dan *Data Prediction=YES*

sebesar 9 data, sehingga total data yang bernilai akurat *True* sebanyak 37 data. Selanjutnya menghitung tingkat nilai akurasi diperoleh berdasarkan jumlah total data akurat *True* sebanyak 37 data dari total populasi sebanyak 41 data diperoleh hasil perhitungan menjadi 0.90. Melihat aturan untuk menentukan tingkat keakuratan, keakuratan pada *Data Training* termasuk kategori *excellent classification*.



Gambar 5.11 Tampilan Pohon Keputusan pada *Data Testing2*

Classification Tree	Class	P(Class)	P(Target)	# Inst	Distribution (rel)	Distribution (abs)
▲ <root>	NO	0.693	0.693	75	0.693:0.307	52:23
PRESTASI NON AKADEMIS = KOTA	YES	1.000	0.000	14	0.000:1.000	0:14
▲ PRESTASI NON AKADEMIS = NON	NO	0.941	0.941	51	0.941:0.059	48:3
KEGIATAN SOSIAL = ANGGOTA	YES	1.000	0.000	3	0.000:1.000	0:3
KEGIATAN SOSIAL = NON	NO	1.000	1.000	48	1.000:0.000	48:0
PRESTASI NON AKADEMIS = PROV	YES	1.000	0.000	4	0.000:1.000	0:4
▲ PRESTASI NON AKADEMIS = SEK	NO	0.667	0.667	6	0.667:0.333	4:2
KEHADIRAN = BAIK	YES	1.000	0.000	2	0.000:1.000	0:2
KEHADIRAN = KURANG	NO	1.000	1.000	4	1.000:0.000	4:0

Gambar 5.12 Tampilan *Rule* untuk *Data Testing2*

Matrik keakuratan dengan jumlah total populasi 75 data, pada Gambar 5.12, maka terlihat :

1. Pada *Correct Class=NO*, pada *Data Prediction=NO* diperoleh data akurat sebesar 52 data.
2. Pada *Correct Class=NO*, pada *Data Prediction=YES* diperoleh data sebesar 0 data.
3. Pada *Correct Class=YES*, pada *Data Prediction=NO* diperoleh data sebesar 0 data.
4. Pada *Correct Class=YES*, pada *Data Prediction=YES* diperoleh data akurat YES sebesar 23 data.

Berdasarkan hasil *Confusion Matrix* dan tingkat nilai diagnosa keakuratan menurut Gorunescu (2011) yaitu,

- a. Akurasi bernilai 0.90 – 1.00 = *excellent classification*
- b. Akurasi bernilai 0.80 – 0.90 = *good classification*
- c. Akurasi bernilai 0.70 – 0.80 = *fair classification*
- d. Akurasi bernilai 0.60 – 0.70 = *poor classification*
- e. Akurasi bernilai 0.50 – 0.60 = *failure*

dapat terlihat bahwa jumlah keakuratan *Data Training* yang bernilai *True* diperoleh dengan cara menghitung jumlah data akurat *True* pada *Correct Class=NO* dan *Data Prediction=NO* sebesar 52 data ditambahkan dengan jumlah data akurat *True Correct Class=YES* dan *Data Prediction=YES* sebesar 4 data, sehingga total data yang bernilai akurat *True* sebanyak 23 data. Selanjutnya menghitung tingkat nilai akurasi diperoleh berdasarkan jumlah total data akurat *True* sebanyak 75 data dari total populasi sebanyak 75 data diperoleh hasil perhitungan menjadi 1.00. Melihat aturan untuk

menentukan tingkat keakuratan, keakuratan pada *Data Training* termasuk kategori *excellent classification*.

### 3.6 Evaluasi

Melihat hasil analisa pada pembelajaran menggunakan analisis manual dengan Algoritma C 4.5 dan pengujian menggunakan *Tools Software Data Mining Orange V 2.7* menghasilkan :

- a. *Rule* keputusan yang tepat sama,
  - b. Keakuratan data termasuk kategori *Excellent Classifications*,
- sehingga dapat ditentukan bahwa keputusan yang dihasilkan dapat diterapkan untuk penentuan kriteria siswa teladan, juga menjelaskan bahwa Algoritma C 4.5 dapat memberikan solusi yang tepat untuk menentukan kriteria siswa teladan.

### 3.7 Kesimpulan

Berdasarkan analisis dan pengujian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut :

1. Teknik klasifikasi menggunakan Algoritma C 4.5 memberikan solusi yang baik dalam penentuan kriteria siswa teladan.
2. Proses transformasi data dengan pemilihan atribut *input*, jumlah atribut dan kategori setiap atribut sangat berpengaruh pada *output* / hasil klasifikasi. Atribut yang menjadi target dalam penelitian ini adalah Teladan. Nilai yang ditetapkan untuk kategori teladan adalah Nilai "Yes" untuk siswa yang memenuhi semua kriteria siswa teladan dan Nilai "No" untuk siswa yang kurang atau tidak memenuhi kriteria siswa teladan yang telah ditentukan.
3. Hasil proses klasifikasi melalui perhitungan menurut aturan klasifikasi dengan Algoritma C 4.5 berbentuk pohon keputusan yang dijabarkan dalam *Rule*.
4. *Rule* sebagai kriteria terbaik dapat diterapkan sebagai penentu kriteria siswa teladan.
5. Tingkat Keakuratan penentuan kriteria menggunakan *Data Training*, *Data Testing1* dan *Data testing2* pada proses klasifikasi tersebut termasuk kategori *excellent classification*.

## 4. Referensi

- [1] Ian H. Witten, E. F., Mark A. Hall (2011). "Data Mining - Practical Machine Learning Tools and Techniques " MorganKaufmann Publishers is an imprint of Elsevier 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA Third Edition.
- [2] Jiawei Han, M. K. (2006). "DM Concepts & Techniques " Second Edition.
- [3] Kumar, A. and Vijayalakshmi (2011). "Implication Of Classification Techniques In Predicting Student's Recital." International Journal of Data Mining & Knowledge Management Process 1(5): 41-51.
- [4] Kusriani, E. T. L. (2009). "Definisi Data Mining." Algoritma Data Mining.
- [5] LAROSE, D. T. (2005). "DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining."
- [6] LAROSE, D.T.(2006). "Data Mining Methods and Models ".
- [7] Oded Maimon, L. R. "INTRODUCTION TO KNOWLEDGE DISCOVERY IN DATABASES." Department of Industrial Engineering Tel-Aviv University.
- [8] ShravanKumar Uppin, A. M. A. (2014). "Expert System Design to Predict Heart and Diabetes Diseases." International Journal of Scientific Engineering and Technology Volume No.3(No.8): 1054-1059.
- [9] Tan, H. C. K. a. G. "Data Mining Applications in Healthcare." Journal of Healthcare Information Management Vol. 19, No.2.
- [10] David, Hansun, S. 2014. Implementasi Data Mining Dengan Algoritma C4.5 Untuk Memprediksi Tingkat Kelulusan Mahasiswa, Tangerang
- [11] Sunjana. 2010. Klasifikasi Data Nasabah Sebuah Asuransi Menggunakan Algoritma C4.5, Seminar Nasional Aplikasi Teknologi Informasi, Vol 7.