



Analisis Perbandingan Algoritma *Data Mining* Metode *Decision Tree C4.5* Dengan *Naïve Bayes* dalam Penjurusan Siswa (Studi Kasus MAN 1 Kota Tangerang Selatan)

Nanang¹

¹ Universitas Pamulang

² Universitas Pamulang

Dosen02599@unpam.ac.id¹, dosen02607@unpam.ac.id²

Kata kunci:

Data Mining Decision Tree C4.5, Naïve bayes, Penjurusan Siswa.

Abstrak

Penjurusan siswa pada MAN 1 Kota Tangerang Selatan ditentukan berdasarkan hasil psikotes, oleh karena itu sangat penting dilakukan penelitian terkait penjurusan siswa yang sesuai dengan kemampuan nilai akademik. Penjurusan siswa dilakukan sejak awal masuk kelas X dengan tujuan untuk memberikan kesempatan kepada siswa dalam mengembangkan kompetensi sikap, kompetensi keterampilan siswa sesuai dengan minat, bakat, dan kemampuan siswa. Dalam penelitian ini, data yang digunakan adalah data hasil ujian masuk dan data nilai dari hasil psikotes yang diselenggarakan di sekolah untuk siswa baru yang telah dinyatakan lulus seleksi masuk MAN 1 Kota Tangerang Selatan. Penelitian ini menggunakan dua algoritma klasifikasi data mining yaitu *Naïve Bayes* dan *Decision Tree*, dalam penelitian ini dapat dibuktikan *Naïve Bayes* menghasilkan akurasi lebih tinggi dibandingkan *Decision Tree*, dengan nilai akurasi 95,92 % untuk algoritma *Naïve Bayes* dan 91,84 % untuk algoritma *Decision Tree*.

Pendahuluan

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.[1]

Pada kurikulum 2013 penjurusan dilakukan pada saat awal masuk, yaitu pada kelas X. Perubahan kurikulum dimaksudkan agar memungkinkan penyesuaian program pendidikan pada satuan pendidikan dengan kondisi dan kekhasan potensi yang ada di daerah peserta didik (Peraturan Kemdikbud, 2013). Kemungkinan yang akan terjadi jika siswa mengalami kesalahan dalam penjurusan adalah rendahnya prestasi belajar siswa atau dapat menyebabkan terjadinya ketidakcocokan dengan jurusan yang sudah dipilih oleh siswa atau siswi sebelumnya.[2]

Penjurusan siswa yang dilakukan pada Sekolah Menengah Atas sesuai minat dan dilakukan sejak peserta didik mendaftar ke SMA/MA memiliki tujuan untuk memberikan kesempatan kepada peserta didik mengembangkan kompetensi sikap dan kompetensi keterampilan peserta didik sesuai dengan minat, bakat, dan kemampuan akademik dalam sekelompok mata pelajaran keilmuan.[3]

Penempatan siswa sesuai dengan kapasitas kemampuannya atau sering disebut dengan penjurusan siswa di sekolah menengah ditentukan oleh kemampuan akademik yang didukung oleh faktor minat karena karakteristik suatu ilmu menuntut karakteristik yang

sama dari yang mempelajarinya. Dengan demikian, siswa yang mempelajari suatu ilmu yang sesuai dengan karakteristik kepribadiannya akan merasa senang ketika mempelajari ilmu tersebut. Minat dapat mempengaruhi kualitas pencapaian hasil belajar siswa dalam bidang studi tertentu. Seorang siswa yang berminat pada Matematika misalnya, akan memusatkan perhatiannya lebih banyak ke bidang Matematika dari pada siswa lain. Karena pemusatan perhatian intensif terhadap materi, siswa akan belajar lebih giat dan mencapai prestasi yang diinginkan.[4]

Berdasarkan dari hasil wawancara dengan pihak Madrasah jurusan yang ada di MAN 1 Kota Tangerang Selatan meliputi jurusan IPA dan IPS. Pada penentuan jurusan yang biasa dilakukan oleh panitia penerimaan siswa baru, di lakukan setelah selesai melaksanakan tes masuk. Penentuan jurusan ini ditentukan dari hasil nilai psikotes.

Penentuan jurusan peminatan siswa yang dilakukan oleh MAN 1 Kota Tangerang Selatan selama ini masih dilakukan secara manual yaitu dengan cara melihat dari hasil psikotes yang dilakukan pada siswa yang lulus pada ujian masuk sekolah di awal saja, dan proses tersebut dilakukan oleh pihak ketiga dari luar sekolah. Oleh karena itu, terkadang apa yang dihasilkan dari tes tersebut mengakibatkan hasil penjurusan anak tidak sesuai dengan minat dan kemampuan siswa dalam memilih jurusan yang tepat dengan bakat dan kemampuannya.

Penelitian terdahulu yang dilakukan oleh [5] dengan judul Penerapan Data Mining Dalam Menentukan Jurusan Siswa untuk melihat keakuratan metode *Naive Bayes* dalam mengklasifikasikan jurusan siswa. Dari hasil penelitian tersebut mengukur tingkat akurasi dengan nilai sebesar 90%.

Penelitian yang dilakukan oleh [6] dengan judul Sistem Pendukung Keputusan Untuk Menentukan Penjurusan SLTA Dengan Metode ID3 dan C4.5 yang isinya adalah untuk mengetahui algoritma mana yang paling bagus kinerjanya. Dengan hasil pengukuran akurasinya yaitu 96% untuk nilai akurasi dari algoritma C4.5 dan 87% untuk nilai akurasi ID3.

Penelitian selanjutnya yang dilakukan oleh [7] dengan judul Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: SMA Negeri 1 Pontianak) yang menghasilkan tingkat kinerja hasil klasifikasi penjurusan siswa tersebut dengan nilai akurasi sebesar 89,74%, nilai *precision* sebesar 96,55% dan *recall* sebesar 90,32%.

Penelitian selanjutnya yang dilakukan oleh [8] dengan judul Klasifikasi dan Klustering Penjurusan Siswa SMA Negeri 3 Boyolali. Mereka yang melakukan teknik klasifikasi dan klustering dalam data mining dengan menggunakan algoritma C4.5 yang menghasilkan tingkat akurasi sebesar 79,14% nilai *precision* sebesar 75,51% dan nilai *recall* sebesar 90,80%.

Penelitian berikutnya yang dilakukan oleh [9] dengan judul Aplikasi Sistem Pendukung Keputusan Menggunakan Algoritma C4.5 Untuk Penjurusan Siswa. Dalam penelitiannya menjelaskan tentang penggunaan metode Algoritma C4.5 terhadap data penjurusan siswa yang menghasilkan nilai akurasi sebesar 96% sedangkan dengan *pruning* sebesar 87%.

Berdasarkan permasalahan di atas maka dalam penelitian ini akan dilakukan perbandingan terhadap dua algoritma untuk penentuan jurusan siswa MAN 1 Kota Tangerang Selatan dengan menggunakan klasifikasi data mining, yaitu metode *decision tree* C4.5 dan *Naive Bayes*. Dengan harapan dapat memberikan solusi bagi siswa-siswi yang mengalami kebingungan dalam memilih jurusan, serta dapat memberikan kemudahan kepada pihak sekolah dalam melakukan penentuan jurusan kepada siswa dengan baik dan sesuai dengan kemampuan, minat, dan bakat siswa.

Metode

2.1 Data Siswa

Penelitian ini menggunakan data sekunder, yaitu data siswa baru MAN 1 Kota Tangerang Selatan, pada tahun pelajaran 2020/2021 sejumlah 149 Siswa. Pengumpulan data dilakukan dengan metode observasi dan teknik sampling dengan metode sensus yaitu

suatu metode pemilihan sampel dimana seluruh populasi di jadikan sebagai sampel. Data ini diperoleh dari panitia penerimaan siswa baru MAN 1 Kota Tangerang Selatan.

Tabel 2. 1 Data Siswa

Data Siswa								
NO URUT	NO PENDAFTARAN	MTK	INGG	INDO	FISIKA	PAI	BIO	PSIKOTES
1	20200200010	22	19	22	22	22	18	139
2	20200200021	22	14	23	19	17	20	100
3	20200200030	19	13	23	19	17	16	137
4	20200200043	21	15	24	19	20	20	136
5	20200200054	22	17	22	19	20	22	136
6	20200200060	18	13	22	19	10	15	133
7	20200200076	19	16	21	19	15	22	132
8	20200200084	19	21	24	19	21	19	132
9	20200200098	22	16	21	19	19	21	129
10	20200200109	19	17	22	20	13	21	128
11	20200200122	18	12	21	21	13	16	127
12	20200200131	19	21	13	19	22	16	127
13	20200200147	20	21	21	22	16	20	125
14	20200200150	21	19	21	21	17	18	125
15	20200200164	10	10	19	21	15	17	122
16	20200200165	21	19	21	19	17	22	121
17	20200200186	18	10	21	19	19	18	120

2.2 Persiapan Data

Jumlah data awal diperoleh dari pengumpulan data yaitu sebanyak 149 *record*. Persiapan juga dapat berupa konversi nilai-nilai relevan dan atau nilai yang terlalu beragam kedalam kelompok yang lebih kecil untuk mempermudah pembentukan model. Dari semua data yang sudah terkumpul, tidak semua data dapat digunakan. Oleh karena itu data tersebut akan melewati beberapa tahap persiapan awal data (*preparation data*) antara lain menggunakan tranformasi data dengan cara nilai dari setiap mata pelajaran berupa numerical, untuk *field* matematika, fisika, biologi, agama, bahasa inggris dan bahasa indonesia akan ditransformasikan menjadi nominal. Jumlah soal masing - masing mata pelajaran adalah 25 soal, jadi nilai hasil ujian terdiri dari 1-25, nilai tersebut di kali 4 jadi skala nilai 1-100 dapat ditegorikan sebagai berikut.

- A. Nilai dengan jumlah ≤ 60 dikategorikan Rendah,
- B. Nilai dengan jumlah ≥ 61 sampai dengan ≤ 80 dikategorikan Sedang
- C. Nilai dengan jumlah ≥ 81 dikategorikan Tinggi.

Tabel 2. 2 Kategori Nilai Atribut Data Siswa

NILAI ANGKA	KATEGORI
4 – 60	Rendah
61 – 80	Sedang

81 – 100	Tinggi
----------	--------

Tabel 2.2 pengkategorian atribut di atas berdasarkan dari data asli hasil nilai tes ujian masuk pada sekolah MAN 1 Kota Tangerang Selatan yang isinya berupa nilai-nilai mata pelajaran yang di ujian diantaranya adalah: MTK, B.Inggris, B.Indonesia, Fisika, PAI, Biologi dan Psikotes.

Data yang digunakan dalam penelitian ini terdiri dari beberapa sumber data, yaitu data MTK, B.Inggris, B.Indonesia, Fisika, PAI, Biologi dan Psikotes, Setelah seluruh data terkumpul, selanjutnya dilakukan proses tahapan data mining yaitu :

1. Pembersihan data (cleaning).
Setelah data di peroleh, maka akan ditemukan data-data yang tidak lengkap. Untuk itu diperlukan tahapan pembersihan data yaitu membuang data-data yang tidak lengkap
2. Integrasi Data (data integration).
Setelah data di cleaning, maka proses selanjutnya adalah integrasi data yaitu penggabungan data MTK, B.Inggris, B.Indonesia, Fisika, PAI, Biologi dan Psikotes.
3. Seleksi Data.
Proses selanjutnya adalah seleksi data, yaitu menyeleksi data-data yang akan digunakan, data yang digunakan adalah, data MTK, B.Inggris, B.Indonesia, Fisika, PAI, Biologi dan Psikotes.

Proses selanjutnya adalah seleksi data, yaitu menyeleksi data-data yang akan digunakan. Data yang digunakan adalah, data MTK, B.Inggris, B.Indonesia, Fisika, PAI, Biologi dan Psikotes.

Peneliti dalam menentukan nilai akurasi yang lebih baik dari kedua Algoritma *Decesion Tree C4.5* dan *Naive Bayes* maka artibut yang digunakan adalah MTK, Ingg, Indo, Fisika, PAI, Bio dan Psikotes dengan label atau kelas penjurusan.

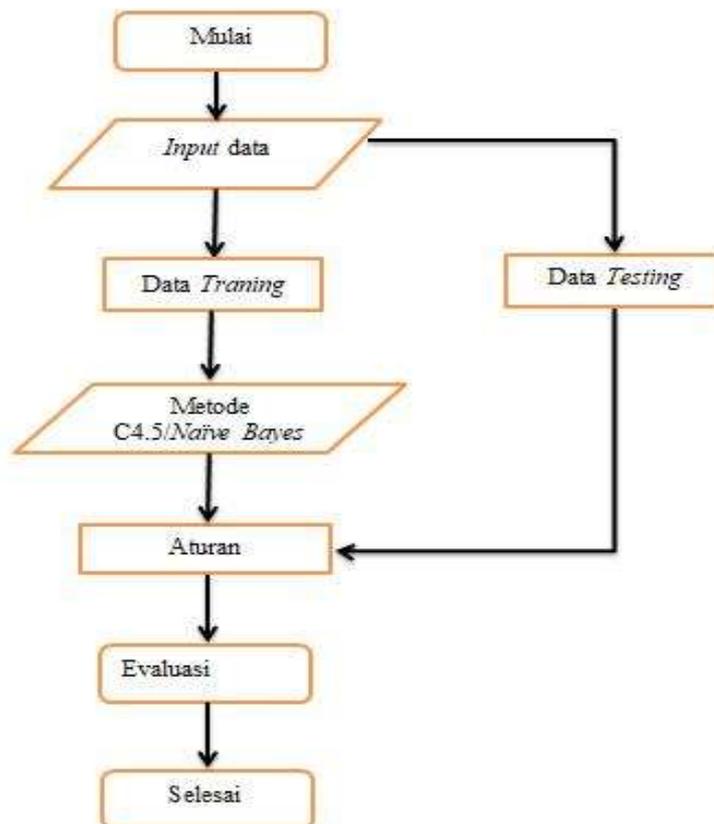
Tabel 2. 3 Atribut Data Siswa

NO	ATRIBUT	NILAI
1	MTK	Rendah
		Sedang
		Tinggi
2	B.Ingg	Rendah
		Sedang
		Tinggi
3	Indo	Rendah
		Sedang
		Tinggi
4	Fisika	Rendah
		Sedang
		Tinggi
5	PAI	Rendah

		Sedang
		Tinggi
6	Bio	Rendah
		Sedang
		Tinggi
7	Psikotes	Rendah
		Sedang
		Tinggi

2.3 Perancangan Penelitian

Tahapan ini adalah tahap dimana dibuatkan suatu rancangan untuk memproses data mulai dari input data sampai mendapatkan nilai akurasi yang terbaik dari algoritma *Decision tree c4.5* dan *Naïve bayes* dengan aplikasi *Rapidminer*.



Gambar 2. 1 Flowchart Perangkat Lunak

2.4 Teknik Analisis

Setelah ini hasil akan di analisa dan data yang di dapat di olah untuk membuat akurasi penjurusan siswa. Algoritma yang akan digunakan adalah:

A. Algoritma C4.5

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 yaitu:

1. Menyiapkan data *training*
2. Menentukan akar dari pohon.
3. Hitung nilai *Gain*: (1)
4. Ulangi langkah ke-2 hingga semua tupel terpartisi (2)

5. Proses partisi pohon keputusan akan berhenti saat semua tupel dalam node N mendapat kelas yang sama dan atau tidak ada atribut di dalam tupel yang dipartisi lagi dan atau tidak ada tupel di dalam cabang yang kosong.

B. Algoritma Naive Bayes

Alur dari metode *Naive Bayes* sebagai berikut;

1. Baca data *training*
2. Hitung Jumlah dan *probabilitas*, namun apabila data numerik maka:
 - A. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik.
 - B. Cari nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Mendapatkan nilai dalam tabel mean, standard deviasi dan *probabilitas*.
4. Solusi kemudian dihasilkan.

2.5 Evaluasi dan Validasi

Pada tahap evaluasi dilakukan pengujian terhadap model-model untuk mendapatkan informasi model yang akurat. Evaluasi dan validasi menggunakan AUC, ROC dan akurasi. Kemudian menganalisa akurasi dengan membandingkan kedua metode tersebut.

Hasil dan Pembahasan

4.1 Hasil Penelitian dan Pengujian Model

Pada bab ini di jelaskan hasil penelitian dan pengujian untuk mendapatkan model dan hasil dari menghitung dan membandingkan kedua algoritma yang diusulkan, selanjutnya dengan pengujian data ke dalam model untuk mendapatkan hasil yang lebih baik. Berikut data awal hasil ujian sebelum nilai angkanya dikonversikan tabel 4.1 dibawah ini.

4.1.1 Algoritma Decision tree C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk pohon keputusan, metode pohon keputusan mengubah fakta menjadi pohon keputusan yang mempresentasikan sebuah aturan.

A. Data Training

Persiapan awal data (preparation data) Jumlah data diperoleh dari pengumpulan data yaitu sebanyak 149 record, dengan atribut MTK, Ingg, Indo, Fisika, PAI, Bio dan Psikotes dengan label atau kelas penjurusan, akan ditransformasikan menjadi nominal. Jumlah soal masing - masing atribut berjumlah 25 soal, nilai tersebut di kali 4 di jadi skala nilai 1-100 sehingga menghasilkan data siswa seperti terlihat pada Tabel 4.1.

Tabel 4. 1 Data Nilai Siswa

NO.	MTK	INGG	INDO	FISIKA	PAI	BIO	PSIKOTES	JURUSAN
1	68	77	89	89	87	72	139	IPA
2	88	57	80	77	69	80	100	IPA
3	72	76	72	57	64	66	100	IPA
4	82	60	96	77	80	80	136	IPA
5	88	66	89	75	80	87	136	IPA
6	82	77	80	63	72	72	115	IPA
7	72	60	72	57	56	66	100	IPA

Langkah selanjutnya adalah dengan melakukan konversi nilai-nilai yang beragam ke dalam kelompok untuk mempermudah pembentukan model, pengelompokan nilai hasil ujian sebagai berikut.

- C. Nilai dengan jumlah ≤ 60 dikategorikan Rendah.
- D. Nilai dengan jumlah ≥ 61 sampai dengan ≤ 80 dikategorikan Sedang.
- E. Nilai dengan jumlah ≥ 81 dikategorikan Tinggi.

Untuk transformasi nilai tersebut dapat dilihat pada Tabel 4.2 dibawah ini:

Tabel 4. 2 Kategori Nilai Data Siswa

NILAI ANGKA	KATEGORI
40 – 60	Rendah
61 – 80	Sedang
81 – 100	Tinggi

Berikut adalah data hasil psikotes sebelum nilai angkanya dikonversikan terdapat pada tabel 4.3 dibawah ini.

Tabel 4. 3 Data Hasil Psikotes

NO	IQ	JURUSAN	GRADE
01	129	IPA	Cerdas
02	123	IPA	Cerdas
06	120	IPA	Cerdas
07	119	IPA	Rata - rata Cerdas
10	115	IPA	Rata - rata Cerdas
11	110	IPA	Rata - rata Cerdas
12	110	IPS	Rata - rata Cerdas
13	108	IPS	Sedang
14	108	IPS	Sedang

Pada tabel 4.3 data hasil Psikotes di atas, peneliti mengolah kembali data tersebut dengan melakukan konversi nilai-nilai yang beragam ke dalam kelompok untuk mempermudah pembentukan model pengelompokan nilai hasil Psikotes sebagai berikut.

- A. Nilai dengan jumlah ≤ 109 dikategorikan sedang / Rendah,
- B. Nilai dengan jumlah ≥ 110 sampai dengan ≤ 119 dikategorikan Rata-Rata cerdas / Sedang
- C. Nilai dengan jumlah ≥ 120 dikategorikan cerdas / Tinggi.

Untuk transformasi nilai nilai psikotes, pada Tabel 4.4 dibawah ini:

Tabel 4. 4 Kategori Nilai Psikotes

NILAI ANGKA	GRADE	KATEGORI
90 – 109	Sedang	Rendah
110 – 119	Rata-Rata Cerdas	Sedang
120 – 139	Cerdas	Tinggi

Setelah nilai di atas dikonversikan dengan berpedoman pada tabel kategori, hasil konversi data training berjumlah 100 data record seperti pada Tabel 4.5 Data Training

Tabel 4. 5 Data Training

NO	MTK	INGG	INDO	FISIKA	PAI	BIO	PSIKOTES	JURUSAN
1	Sedang	Sedang	Tinggi	Tinggi	Tinggi	Sedang	Tinggi	IPA
2	Tinggi	Rendah	Sedang	Sedang	Sedang	Sedang	Rendah	IPA
3	Sedang	Sedang	Sedang	Rendah	Sedang	Sedang	Rendah	IPS
4	Tinggi	Rendah	Tinggi	Sedang	Sedang	Sedang	Tinggi	IPA
5	Tinggi	Sedang	Tinggi	Sedang	Sedang	Tinggi	Tinggi	IPA
6	Tinggi	Sedang	Sedang	Sedang	Sedang	Sedang	Sedang	IPA
7	Sedang	Rendah	Sedang	Rendah	Rendah	Sedang	Rendah	IPS
8	Sedang	Sedang	Tinggi	Sedang	Rendah	Tinggi	Tinggi	IPA
9	Sedang	Tinggi	Tinggi	Sedang	Tinggi	Sedang	Tinggi	IPA
10	Sedang	Sedang	Sedang	Rendah	Sedang	Sedang	Rendah	IPS
11	Tinggi	Rendah	Tinggi	Sedang	Sedang	Sedang	Sedang	IPA
12	Sedang	Sedang	Sedang	Sedang	Rendah	Tinggi	Tinggi	IPA

B. D
ata
Testin
g

S
etelah
melakukan persiapan data testing berjumlah 49 data record, langkah selanjutnya adalah melakukan persiapan data testing. Tabel 4.6 data Testing.

Tabel 4. 5 Data Testing

NO.	MTK	INGG	INDO	FISIKA	PAI	BIO	SIKOTES	JURUSAN
1	Sedang	Sedang	Sedang	Rendah	Sedang	Sedang	Sedang	IPS
2	Sedang	Rendah	Tinggi	Rendah	Sedang	Sedang	Sedang	IPS
3	Sedang	Rendah	Tinggi	Rendah	Tinggi	Sedang	Sedang	IPA
4	Tinggi	Sedang	Tinggi	Tinggi	Tinggi	Sedang	Tinggi	IPA
5	Tinggi	Rendah	Tinggi	Sedang	Sedang	Sedang	Rendah	IPA

Peneliti dalam menentukan nilai akurasi dari Decision Tree dan Naive Bayes dari atribut MTK, Ingg, Indo, Fisika, PAI, Bio dan Psikotes serta Penjurusan Sebagai labelnya.

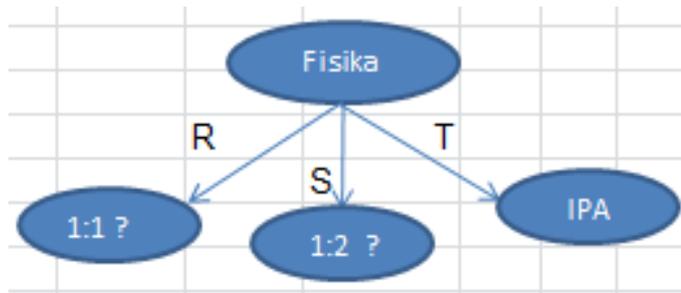
C. Entropy dan gain

Tabel 4. 6 Hitung Entropy dan Gain

HITUNG NILAI ENTROPY DAN GAIN SELURUH ATRIBUT								
Node	Atribut	Value	Jumlah Kasus	IPA	IPS	Entropy	Gain	
1	Total		100	61	39	0,9648		%
	MTK						0,23833	24%

HITUNG NILAI ENTROPY DAN GAIN SELURUH ATRIBUT							
		Tinggi	43	39	4	0,4465	
		Sedang	56	21	35	0,9544	
		Rendah	1	1	0	0	
	INGG					0,10549	11%
		Tinggi	12	12	0	0	
		Sedang	43	27	16	0,9523	
		Rendah	45	22	23	0,9996	
	INDO					0,10672	11%
		Tinggi	58	44	14	0,7973	
		Sedang	41	16	25	0,965	
		Rendah	1	1	0	0	
	FISIKA					0,31578	32%
		Tinggi	17	17	0	0	
		Sedang	28	25	3	0,4912	
		Rendah	55	19	36	0,9299	
	PAI					0,03949	4%
		Tinggi	10	9	1	0,469	
		Sedang	66	40	26	0,9673	
		Rendah	24	12	12	1	
	BIO					0,06863	7%
		Tinggi	18	16	2	0,5033	
		Sedang	76	43	33	0,9875	
		Rendah	6	2	4	0,9183	
	PSIKOTES					0,07979	8%
		Tinggi	30	25	5	0,65	
		Sedang	21	13	8	0,9587	
		Rendah	49	23	26	0,9973	

Dari tabel 4.7 di atas di dapat atribut Fisika adalah mendapat nilai gain terbesar yang selanjutnya menjadi akar / node pertama.

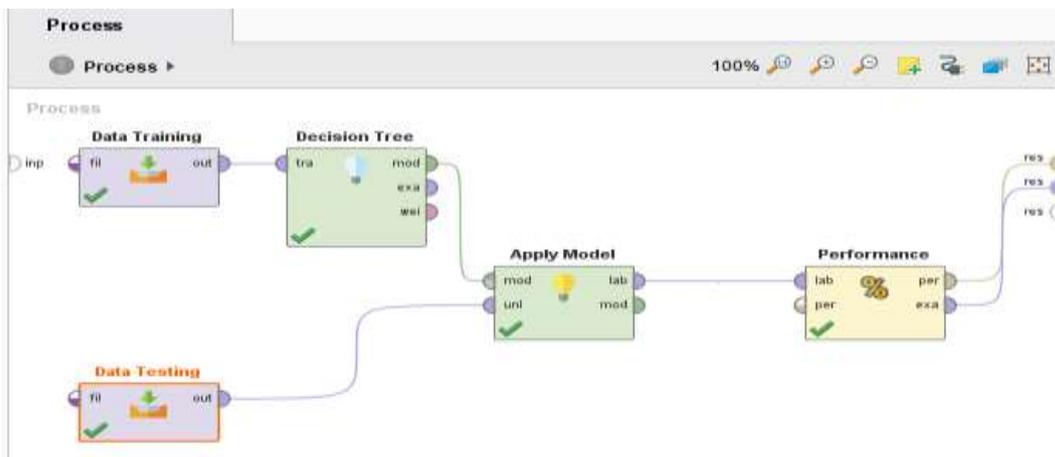


Gambar 4. 1 Node/Akar Psikotes

Untuk mendapatkan node 1.1 dilakukan perhitungan seperti yang pertama tetapi datanya sudah di filter oleh value Tinggi yang ada dalam atribut Fisika.

D. Pengujian Model *Decision Tree*

Pengujian data dilakukan dengan menggunakan aplikasi RapidMiner 9.0 berikut gambar 4.2 alur pengujian algoritma C4.5.



Gambar 4. 1 Alur Pengujian *Decision Tree*

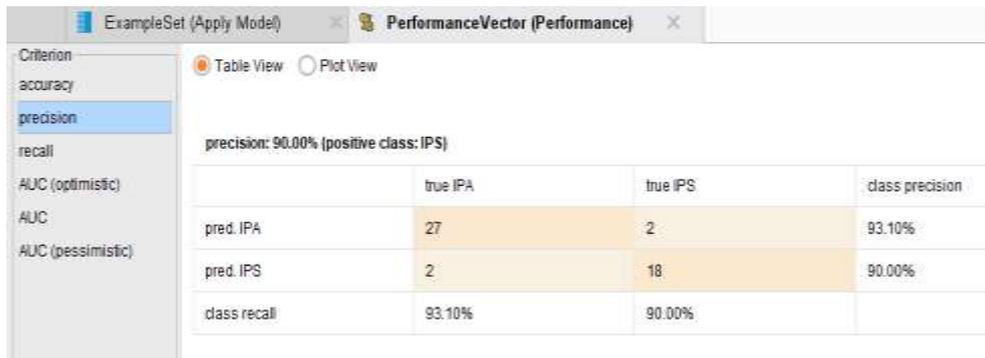
ExampleSet (Apply Model) PerformanceVector (Performance)

Table View Plot View

accuracy: 91.84%

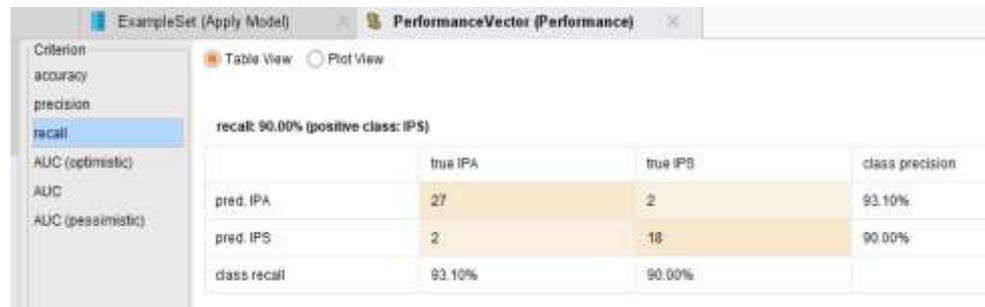
	true IPS	true IPA	class precision
pred. IPS	18	2	90.00%
pred. IPA	2	27	93.10%
class recall	90.00%	93.10%	

Gambar 4. 2 Hasil Akurasi Decision Tree



	true IPA	true IPS	class precision
pred. IPA	27	2	93.10%
pred. IPS	2	18	90.00%
class recall	93.10%	90.00%	

Gambar 4. 3 Hasil Precision Decision Tree



	true IPA	true IPS	class precision
pred. IPA	27	2	93.10%
pred. IPS	2	18	90.00%
class recall	93.10%	90.00%	

Gambar 4. 4 Hasil Recall Decision Tree

Hasil pengujian pada gambar 4.3 menggunakan metode *Decision Tree* di atas di dapatkan hasil dengan tingkat akurasi 91,84% yang artinya tingkat akurasi data sudah baik.

Class precision dapat di artikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut, jadi kecocokan antara permintaan dan prediksi kelas IPS adalah 90,00% dan kecocokan dengan prediksi kelas IPA adalah 93,10% pada gambar 4.4

Recall di definisikan sebagai rasio dari item relevan yang di pilih terhadap total jumlah item relevan yang tersedia. Jadi dari data di atas dapat di simpulkan bahwa tingkat relevan true IPS sebesar 90,00% dan True IPA sebesar 93,10 %. gambar 4.5

Open in Turbo Prep Auto Model Filter (45 / 49 examples): correct_predictions

...	JURUSAN	predicti...	confide...	confide...	MTK	INGG	INDO	FISIKA	PAI	BIO	PSIKOT...
1	IPS	IPS	0.059	0.941	Sedang	Sedang	Sedang	Rendah	Sedang	Sedang	Sedang
2	IPS	IPS	0.053	0.947	Sedang	Rendah	Tinggi	Rendah	Sedang	Sedang	Sedang
3	IPA	IPA	1	0	Tinggi	Sedang	Tinggi	Tinggi	Tinggi	Sedang	Tinggi
4	IPA	IPA	0.893	0.107	Tinggi	Rendah	Tinggi	Sedang	Sedang	Sedang	Rendah
5	IPA	IPA	0.893	0.107	Sedang	Rendah	Tinggi	Sedang	Sedang	Sedang	Tinggi
6	IPS	IPS	0.053	0.947	Sedang	Rendah	Sedang	Rendah	Sedang	Sedang	Rendah
7	IPA	IPA	0.893	0.107	Tinggi	Sedang	Tinggi	Sedang	Sedang	Tinggi	Tinggi
8	IPA	IPA	0.893	0.107	Sedang	Rendah	Tinggi	Sedang	Rendah	Rendah	Tinggi
9	IPS	IPS	0.059	0.941	Sedang	Sedang	Sedang	Rendah	Sedang	Sedang	Rendah
10	IPA	IPA	0.893	0.107	Tinggi	Rendah	Tinggi	Sedang	Sedang	Sedang	Sedang

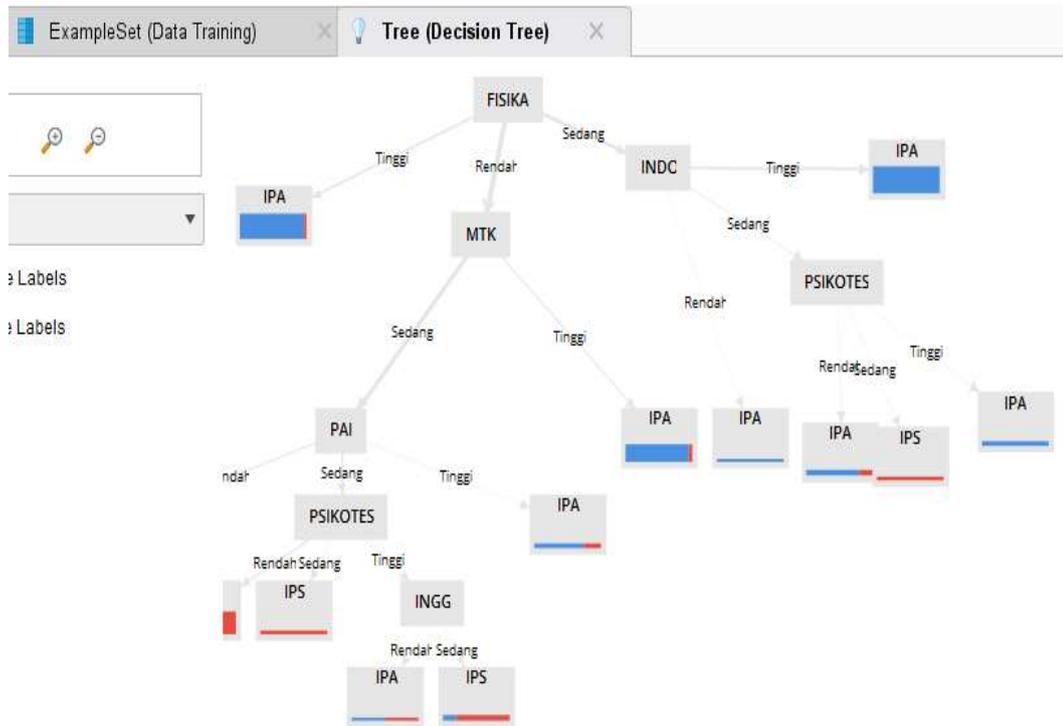
Gambar 4. 6 Data Set Decision Tree

Dari hasil analisis di atas, kita dapat melihat bahwa terdapat 12 kolom atribut yang terdiri 7 kolom atribut biasa dan 4 kolom spesial atribut dengan nama Jurusan, Prediction Jurusan, Confidence IPA dan Confidence IPS, hal ini dikarenakan kolom tersebut menjadi penentu dari hasil analisis data siswa. Dari data siswa 49 ada sebagian data yang dinyatakan tidak cocok dengan prediksi decision tree seperti pada gambar 4.5

Open in Turbo Prep Auto Model Filter (4 / 49 examples): wrong_predictions

Ro...	JURUSAN	predicti...	confide...	confide...	MTK	INGG	INDO	FISIKA	PAI	BIO	PSIKOT...
1	IPA	IPS	0.053	0.947	Sedang	Rendah	Tinggi	Rendah	Tinggi	Sedang	Sedang
2	IPS	IPA	1	0	Sedang	Sedang	Tinggi	Tinggi	Sedang	Sedang	Sedang
3	IPA	IPS	0.059	0.941	Sedang	Sedang	Tinggi	Rendah	Sedang	Sedang	Tinggi
4	IPS	IPA	0.893	0.107	Sedang	Rendah	Sedang	Sedang	Rendah	Sedang	Sedang

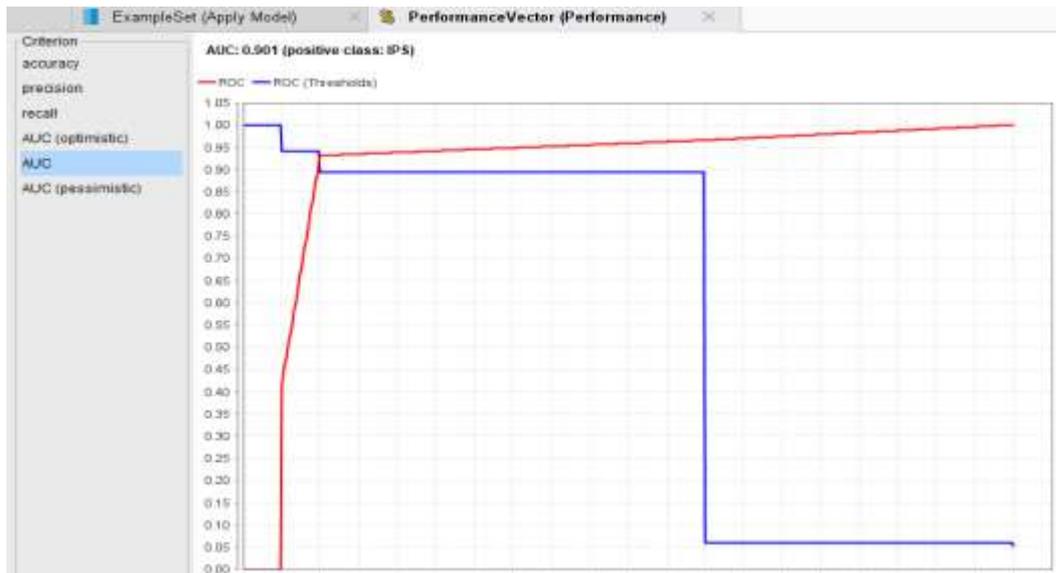
Gambar 4. 5 Data Wrong Prediction Decision Tree



Gambar 4. 6 Pohon Keputusan

Dari hasil analisis diatas dapat ditarik kesimpulan, bahwa :

- Jika nilai fisika tinggi maka masuk jurusan IPA.
- Jika nilai fisika sedang maka cek nilai indo tinggi = masuk jurusan IPA.
- Jika nilai fisika rendah maka cek nilai MTK = tinggi maka masuk jurusan IPA.
- Jika nilai fisika sedang maka cek nilai indo sedang cek nilai psikotes rendah masuk jurusan IPA.
- Jika nilai fisika sedang maka cek nilai indo sedang cek nilai psikotes sedang masuk jurusan IPS.
- Jika nilai fisika sedang maka cek nilai indo sedang cek nilai psikotes tinggi masuk jurusan IPA.
- Jika nilai fisika rendah maka cek nilai indo rendah maka masuk jurusan IPA.
- Jika nilai fisika rendah maka cek nilai MTK sedang maka cek nilai PAI rendah masuk jurusan IPS.
- Jika nilai fisika rendah maka cek nilai MTK sedang maka cek nilai PAI sedang cek nilai psikotes sedang masuk jurusan IPS.
- Jika nilai fisika rendah maka cek nilai MTK sedang maka cek nilai PAI sedang cek nilai psikotes tinggi cek nilai ingg rendah masuk jurusan IPA.
- Jika nilai fisika rendah maka cek nilai MTK sedang maka cek nilai PAI sedang cek nilai psikotes tinggi cek nilai ingg sedang masuk jurusan IPS.
- Jika nilai fisika rendah maka cek nilai MTK sedang maka cek nilai PAI tinggi masuk jurusan IPA.



Gambar 4. 7 AUC Decision Tree

AUC (*the area under curve*) dihitung untuk mengukur perbedaan performansi. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual dengan *false positive* sebagai garis horizontal dan *true positive* sebagai garis vertikal. Dari data diatas analisis menggunakan aplikasi *Rapid Miner* dengan pengukuran *Decision Tree* didapatkan hasil AUC 0,901 yang termasuk dalam kategori sangat baik (*Excellent Classification*).

4.1.2 Algoritma Naive Bayes

Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes, alur Naive Bayes sebagai berikut:

A. Baca Data Training

Baca *data training* yang ada pada tabel 4.5

B. Atribut dan Probabilitas

Probabilitas setiap atribut didapatkan dari data training pada tabel 4.5, jumlah kasus adalah 149 dengan kasus IPA sejumlah 91 dan IPS sebanyak 58 kasus maka nilai probabilitas dari setiap atribut adalah sebagai berikut.

1) Probabilitas Atribut MTK

Pada atribut MTK dapat diketahui jumlah siswa yang ada di class IPS adalah dengan nilai MTK sedang memiliki nilai probabilitas tertinggi yaitu sebesar 95 %. Hasil perhitungan dapat dilihat pada tabel 4.8 sebagai berikut.

Tabel 4. 7 Probabilitas MTK

HITUNG JUMLAH CLASS/LABEL		
MTK	JML SISWA	PROBABILITAS

	IPA	IPS	P(IPA)	P(IPS)
Tinggi	58	3	0,637	0,052
Sedang	31	55	0,341	0,948
Rendah	2	0	0,022	0,000
TOTAL	91	58		

2) Probabilitas Atribut Bio

Pada atribut Bio dapat diketahui jumlah siswa yang ada di class IPS adalah dengan nilai Bio rendah yang memiliki nilai probabilitas sebesar 53%. Hasil perhitungan dapat dilihat pada tabel 4.9 sebagai berikut.

Tabel 4. 8 Probabilitas Bio

HITUNG JUMLAH CLASS/LABEL				
Bio	JML SISWA		PROBABILITAS	
	IPA	IPS	P(IPA)	P(IPS)
Tinggi	18	0	0,198	0,000
Sedang	37	27	0,407	0,466
Rendah	36	31	0,396	0,534
TOTAL	91	58		

3) Probabilitas Atribut Indo

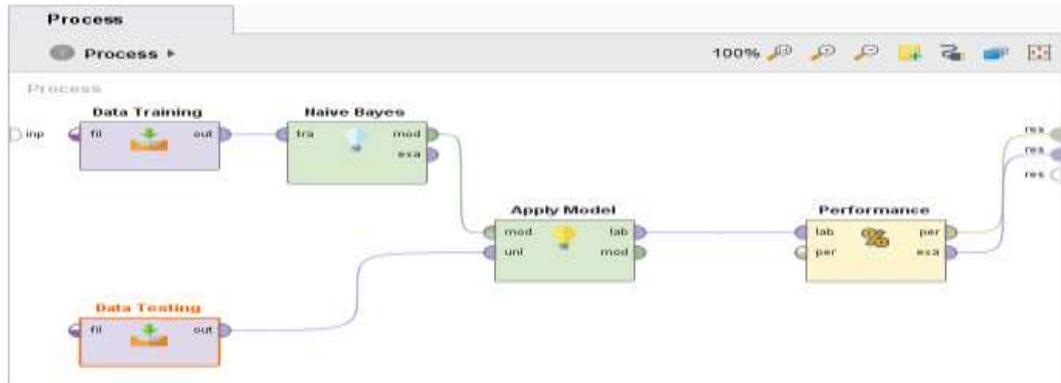
Pada atribut Indo dapat diketahui jumlah siswa yang ada di class IPA adalah dengan nilai Indo tinggi memiliki nilai probabilitas tertinggi yaitu sebesar 73%. Hasil perhitungan dapat dilihat pada tabel 4.10 sebagai berikut.

Tabel 4. 9 Probabilitas Indo

HITUNG JUMLAH CLASS/LABEL				
Indo	JML SISWA		PROBABILITAS	
	IPA	IPS	P(IPA)	P(IPS)
Tinggi	66	17	0,725	0,293
Sedang	22	41	0,242	0,707
Rendah	3	0	0,033	0,000
TOTAL	91	58		

C. Pengujian Model Naive Bayes

Pengujian data siswa dilakukan dengan menggunakan aplikasi *RapidMiner*, gambar 4.8 alur pengujian *Naive Bayes*.



Gambar 4. 8 Alur Pengujian *Naive Bayes*

Pertama pengolahan diawali dengan menginputkan data training dan data testing yang akan digunakan pada operator read excel kemudian di hubungkan dengan metode Naive Bayes dalam pengujian data untuk menghasilkan performa yang terbaik.

Hasil prediksi penjurusan IPA dan IPS, data yang digunakan pada metode Naive Bayes sejumlah 49 record dengan 7 atribut.

		true IPS	true IPA	class precision
pred. IPS		19	1	95.00%
pred. IPA		1	28	96.55%
class recall		95.00%	96.55%	

Gambar 4. 9 Hasil Akurasi *Naive Bayes*

Hasil pengujian menggunakan metode Naive Bayes di atas di dapatkan hasil dengan tingkat akurasi 95,92% yang artinya tingkat akurasi data sudah baik.

Class precision di dapat dengan hasil prediksi IPS sebesar 95,00% dan prediksi IPA sebesar 96,55%. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut, jadi kecocokan antara permintaan dan prediksi IPS adalah 95,00% dan kecocokan dengan prediksi IPA adalah 96,55%.

Recall didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. Jadi dari data diatas dapat disimpulkan bahwa tingkat relevan true IPS sebesar 95,00% dan true IPA sebesar 96,55%. gambar 4.9



Gambar 4. 12 Hasil AUC Naive Bayes

AUC (the area under curve) di hitung untuk mengukur perbedaan performansi. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual dengan false positive sebagai garis horizontal dan true positive sebagai garis vertikal. Dari data diatas analisis menggunakan aplikasi Rapid Miner dengan pengukuran Decision Tree didapatkan hasil AUC 0,978 yang termasuk dalam kategori sangat baik (Excellent Classification).

Pembahasan

Dari hasil perbandingan di atas maka dapat disimpulkan bahwa metode *Naïve Bayes* dan *Decision Tree* termasuk prediksi sangat baik dengan nilai AUC 0,978 dan 0,901 namun metode *Naïve Bayes* lebih unggul dalam hal akurasinya, yaitu 95,92% dan 91,84%

Setelah melakukan penelitian ini menghasilkan akurasi tertinggi oleh algoritma *Naïve Bayes* karena perhitungan *probabilitas* algoritma *Naïve Bayes* mampu memprediksi *true positif* lebih besar nilainya di dibandingkan *Decision Tree*.

Kesimpulan

Berdasarkan hasil penelitian dalam penentuan jurusan siswa pada MAN 1 Kota Tangerang Selatan yang menggunakan 100 data *training* dan 49 data *testing*. Kriteria yang digunakan adalah nilai 40-60 = rendah, 61-80 = sedang dan 81-100 = tinggi serta menggunakan metode *Naive Bayes* dan *Decision Tree* maka dapat disimpulkan bahwa *Naïve Bayes* menghasilkan akurasi sebesar 95,92% dengan $AUC=0,987$ dan *Decesion Tree* menghasilkan 91,84% dengan $AUC = 0,901$. Kedua algoritma termasuk kategori sangat baik namun algoritma *Naïve Bayes* memiliki tingkat akurasi lebih tinggi.

Dalam penelitian ini algoritma *Naive Bayes* menghasilkan model terbaik dengan akurasi tertinggi sebagai metode prediksi untuk penentuan penjurusan siswa dibandingkan dengan metode *Decision Tree*.

Daftar Pustaka

- [1] Asroni, Hidayatul, Fitri, & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Mean pada Pengelompokan data calon mahasiswa baru di Universitas Muhammadiyah Yogyakarta. *Semesta Teknika*.
- [2] Nuh, M. (2013). Pedoman Penyusunan dan Pengelolaan Kurikulum. *Peraturan Menteri Pendidikan dan Kebudayaan*.
- [3] Nuh, M. (2014). Peminatan pada Pendidikan Menengah. *Pendidikan, Kebudayaan, & Indonesia*, .
- [4] Bahar. (2011). Penentuan Jurusan Sekolah Atas Dengan Algoritma Fuzzy C-Means. *Universitas Dian Nuswantoro Semarang*.
- [5] Saleh, A. (2015). Penerapan Data Mining dalam Menentukan Jurusan Siswa. *Seminar Nasional Informatika*.
- [6] Rahmanita, E., & Kustiyahningsih, Y. (2016). Sistem Pendukung Keputusan untuk Menentukan Penjurusan SLTA dengan Metode ID3 dan C4.5. *Prosiding Seminar Nasional Multi Disiplin Ilmu*.
- [7] Novianti, B., Rismawan, T., & Bahri, S. (2016). Implementasi Data Mining dengan Algoritma C4.5 untuk Penjurusan Siswa. *Jurnal Coding Sistem Komputer Untan*.
- [8] Nugroho, Y. S., & Haryati, S. N. (2015). Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali. *Jurnal Ilmu KOMputer dan Informatika*.
- [9] Kustiyahningsih, Y., & Rahmanita, E. (2016). Aplikasi Sistem Pendukung Keputusan Menggunakan Algoritma C4.5 untuk Penjurusan SMA. *Jurnal Simantec*.