



## Perbandingan Teknik Data Mining untuk Penentuan Jurusan Siswa Menggunakan Metode Decision Tree C4.5 dan Naïve Bayes (Studi Kasus : SMA PGRI 109 Kota Tangerang)

Halili Maar<sup>1</sup>

<sup>1</sup>Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang

[dosen02957@unpam.ac.id](mailto:dosen02957@unpam.ac.id)

### Abstract

Since students enter high school with the aim of improving their academic abilities, attitudes, and talents in various disciplines, student interest is based on their interests. The information used in this study comes from the results of the entrance exam for new students who graduated from SMA PGRI 109 Tangerang City. To find out which algorithm produces a model or value with superior accuracy results, this study compared two data mining classification methods using the Decision Tree C4.5 Algorithm and the Naive Bayes Algorithm, from the test results the Naive Bayes method is superior to the Decision Tree C4.5 method with an accuracy value of Naive Bayes = 81.63% and AUC = 0.925, while the Decision Tree accuracy value = 75.51% and AUC = 0.725.

Keywords: Data Mining Decision Tree C4.5, Naïve Bayes, student majors

### Abstrak

Karena siswa masuk SMA dengan tujuan untuk meningkatkan kemampuan akademik, sikap, dan bakat dalam berbagai disiplin ilmu, maka peminatan siswa dilakukan berdasarkan minat. Informasi yang digunakan dalam penelitian ini bersumber dari hasil ujian masuk siswa baru yang lulus di SMA PGRI 109 Kota Tangerang. Untuk mengetahui algoritma mana yang menghasilkan model atau nilai dengan hasil akurasi yang lebih unggul, dalam penelitian ini dilakukan perbandingan dua metode klasifikasi data mining dengan menggunakan Algoritma Decision Tree C4.5 dan Algoritma Naive Bayes, dari hasil pengujian metode Naive Bayes lebih unggul dari metode Decision Tree C4.5 dengan nilai akurasi Naive Bayes = 81,63% dan AUC = 0,925, sedangkan Decision Tree nilai akurasi = 75,51% dan AUC = 0,725.

Kata Kunci: Data Mining Decision Tree C4.5, Naïve Bayes, Jurusan Siswa.

### 1. Pendahuluan

SMA merupakan jenjang pendidikan formal yang berfungsi sebagai lanjutan dari SMP atau sederajat, sesuai dengan Peraturan Pemerintah Republik Indonesia Dalam proses pembelajarannya, siswa perlu diarahkan sesuai dengan minat, bakat, dan kemampuannya agar dapat mencapai hasil belajar yang optimal dan relevan dengan dunia kerja maupun pendidikan lanjut.

Salah satu bentuk pengelompokan ini adalah pemilihan jurusan (IPA/IPS), yang idealnya didasarkan pada kemampuan akademis dan minat siswa. Siswa yang ditempatkan pada jurusan yang sesuai dengan kepribadiannya cenderung lebih fokus, berprestasi, dan menikmati proses belajar.

Berdasarkan hasil wawancara dengan pihak sekolah, SMA PGRI 109 Kota Tangerang menawarkan jurusan IPA dan IPS. Setelah ujian masuk selesai, panitia penerimaan siswa baru biasanya memberikan rekomendasi jurusan apa yang cocok untuk siswa dan hasil ujian masuk digunakan untuk menentukan jurusan untuk siswanya. Selama ini SMA PGRI 109 Kota Tangerang masih melakukan penentuan jurusan minat siswa secara manual dengan melihat hasil ujian masuk. Maka dari itu, peneliti akan membandingkan dua algoritma penentuan jurusan siswa di SMA PGRI 109 Kota Tangerang dengan menggunakan klasifikasi data mining, yaitu metode Decision Tree C4.5 dan Naïve Bayes. Hal ini dikarenakan hasil tes terkadang mengakibatkan jurusan anak tidak sesuai dengan minat dan kemampuan siswa dalam memilih jurusan yang tepat dengan bakat dan kemampuannya. Hal ini dilakukan sebagai upaya untuk membantu siswa yang



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

kesulitan dalam pemilihan jurusan dan memudahkan pihak sekolah dalam memilih jurusan siswa secara tepat dengan mempertimbangkan keterampilan, minat, dan kesukaan pribadi.

Untuk itu, peneliti berinisiatif membandingkan dua algoritma klasifikasi data mining, yaitu Decision Tree C4.5 dan Naïve Bayes, guna mengidentifikasi jurusan siswa secara lebih tepat dan objektif. Pemanfaatan teknik ini diharapkan dapat membantu siswa dalam menentukan jurusan yang sesuai, sekaligus memudahkan pihak sekolah dalam proses penjurusan berdasarkan data akademik dan minat siswa.

Judul penelitian yang diangkat adalah "Perbandingan Teknik Data Mining untuk Penentuan Jurusan Siswa Menggunakan Metode Decision Tree C4.5 dan Naïve Bayes (Studi Kasus: SMA PGRI 109 Kota Tangerang)".

## 2. Metode Penelitian

### Data Nilai Siswa

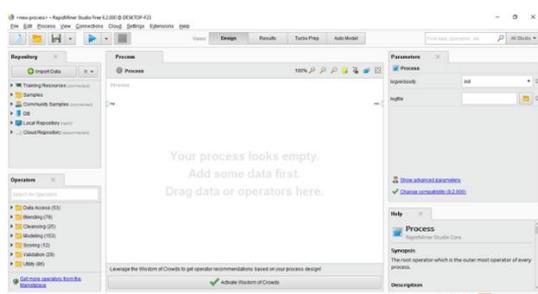
Penelitian menggunakan data nilai siswa baru SMA PGRI 109 Tangerang pada tahun 2019/2020 sebanyak 149 Siswa.

Tabel 1. Data Nilai Siswa

No	No Pendaftaran	In do	Ingg ris	P AI	PK N	MT K	Fisi ka	Biol ogi	Ki mia
1	192010001	21	20	17	17	20	21	20	21
2	192010002	20	10	20	20	22	23	18	15
3	192010003	16	20	20	22	20	13	17	10
4	192010004	23	7	19	20	21	23	18	16
5	192010005	22	17	16	8	22	13	21	23
6	192010006	17	18	16	9	21	22	17	18
7	192010007	16	12	12	11	18	11	17	11
8	192010008	21	22	21	20	17	16	18	22
9	192010009	17	16	17	21	16	10	18	10
10	192010010	15	5	12	10	7	6	7	6

### Tools/Aplikasi

Aplikasi RapidMiner, yang menggunakan pendekatan algoritma C4.5 dan Naïve Bayes, adalah alat/aplikasi yang mendukung analisis.



Gambar 1. Tampilan Aplikasi RapidMiner

### Persiapan Data (Preprocessing Data)

Kumpulan data pertama, yang terdiri dari 149 item, merupakan hasil dari prosedur pengumpulan data.

pengujian dilakukan dengan menggunakan 100 data training dan 49 data testing, dan untuk meningkatkan pengembangan model, tidak semua informasi yang telah dikumpulkan dapat dimanfaatkan. Oleh karena itu, kumpulan data akan melalui sejumlah fase persiapan data awal, termasuk transformasi data menggunakan nilai numerik yang ditetapkan untuk setiap subjek (khususnya B.Indo, B.Bahasa Inggris, PAI, Pkn, Mtk, Fisika, Biologi, dan Kimia) dan transformasi data melalui transformasi data. Dengan 25 pertanyaan di setiap subjek, skor tes berkisar dari 1 hingga 25, yang kemudian dikalikan dengan 4 untuk menghasilkan skala penilaian 1–100, yang dibagi menjadi kategori berikut:

Skor yang  $\leq 60$  diklasifikasikan sebagai Rendah, Skor mulai dari  $\geq 61$  hingga  $\leq 80$  diklasifikasikan sebagai Sedang, Skor yang  $\geq 81$  diklasifikasikan sebagai Tinggi.

Tabel 2. Kategori Nilai Data Siswa

NILAI ANGKA	KATEGORI
4 – 60	Rendah
61 – 80	Sedang
81 – 100	Tinggi

Tabel di atas berdasarkan data aktual tentang hasil ujian dari SMA PGRI 109 Kota Tangerang. Banyak bidang akademik yang tercakup dalam dataset yang digunakan dalam penelitian ini, dengan penekanan khusus pada Bahasa Inggris, Bahasa Indonesia, Pendidikan Islam, Kewarganegaraan, Matematika, Fisika, Biologi, dan Kimia. Prosedur penambangan data dilakukan setelah pengumpulan data selesai, khususnya:

1. Pemurnian data  
Data yang tidak lengkap akan ditemukan setelah data dikumpulkan. Oleh karena itu, diperlukan langkah pembersihan data, yang mencakup penghapusan data yang hilang.
2. Integrasi Data (data integration).  
Integrasi data, yang melibatkan penggabungan data dari Bahasa Indonesia, Bahasa Inggris, PAI, Pkn, Mtk, Fisika, Biologi, dan Kimia, dilakukan setelah data dibersihkan.
3. Pemilihan Data  
Pemilihan data, atau memilih data mana yang akan digunakan, adalah langkah berikutnya. Data dari Fisika, Biologi, Kimia, PAI, Pkn, MTK, Bahasa Inggris, dan Bahasa Indonesia digunakan.

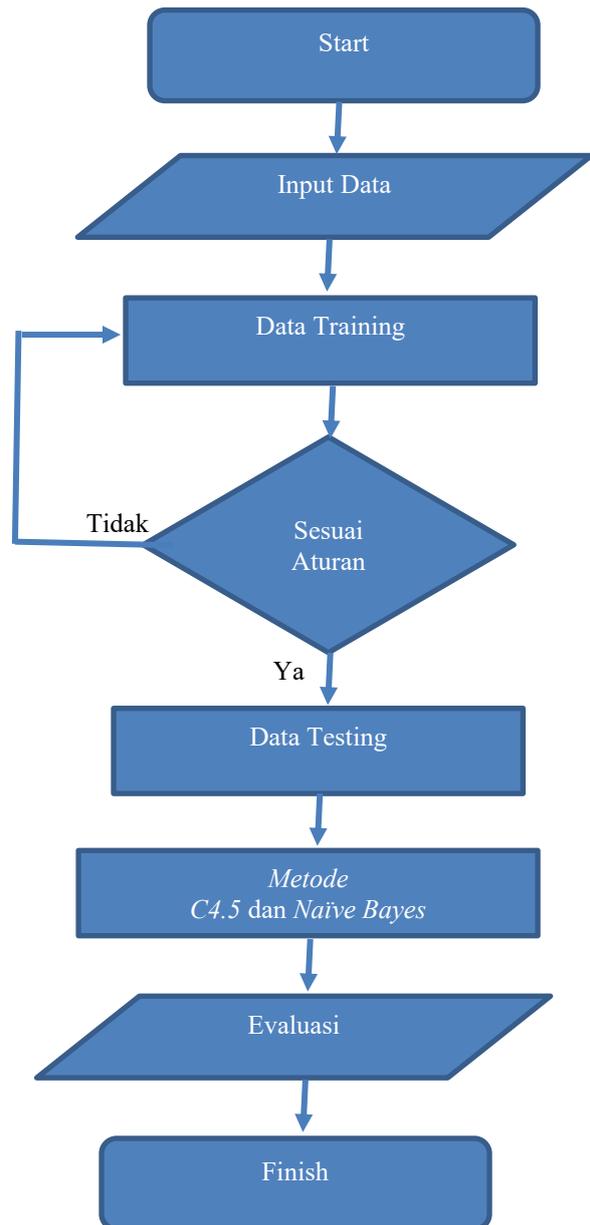
Atribut yang dimanfaatkan peneliti untuk menentukan salah satu dari kedua Algoritma Decision Tree C4.5 dan Naive Bayes yang memiliki nilai akurasi lebih tinggi adalah Bahasa Indonesia, Bahasa Inggris, Pendidikan Agama Islam, PKn, Matematika, Fisika, Biologi, dan Kimia beserta label jurusan atau kelasnya.

Tabel 3. Atribut Data Pelajaran

NO	ATRIBUT	NILAI
1	B.Indo	Rendah
		Sedang
		Tinggi
2	B.Ingggris	Rendah
		Sedang
		Tinggi
3	PAI	Rendah
		Sedang
		Tinggi
4	Pkn	Rendah
		Sedang
		Tinggi
5	Mtk	Rendah
		Sedang
		Tinggi
6	Fisika	Rendah
		Sedang
		Tinggi
7	Biologi	Rendah
		Sedang
		Tinggi
8	Kimia	Rendah
		Sedang
		Tinggi

### Perancangan Penelitian

Tahapan diagram alur desain penelitian adalah sebagai berikut. Berikut ini adalah diagram alurnya:



Gambar 2 Alur/Flowcart Perancangan Penelitian.

### 3. Hasil dan Pembahasan

#### Hasil Penelitian dan Pengujian Model

Dengan menghitung dan membandingkan dua algoritma yang disarankan, lalu menguji data ke dalam model untuk mendapatkan hasil yang lebih baik, temuan studi dan pengujian menggambarkan bagaimana model dan hasil diperoleh.

#### Algoritma Decision Tree

Salah satu strategi untuk membuat pohon keputusan adalah algoritma C4.5, metode pohon keputusan mengubah data menjadi pohon keputusan yang menampilkan aturan.

**Data Training**

Penyiapan data awal sebanyak 149 record, meliputi informasi tentang Bahasa Indonesia, Bahasa Inggris, Pendidikan Agama Islam, PKn, Matematika, Fisika, Biologi, dan Kimia, beserta label atau mata kuliah utama, akan dikonversi menjadi data nominal. Setiap atribut memiliki 25 pertanyaan, dan hasilnya dikalikan 4 untuk membuat skala nilai mulai dari 1 hingga 100, yang digunakan untuk menghasilkan data siswa.

Tabel 4. Nilai Siswa

No	Indo	Inggris	PAI	Pkn	Mtk	Fisika	Biologi	Kimia	Jurusan
1	84	80	68	84	80	84	80	84	IPA
2	80	40	80	80	88	92	72	60	IPA
3	64	80	80	88	80	52	68	40	IPS
4	92	28	76	80	84	92	72	64	IPA
5	88	68	64	32	88	52	84	92	IPA
6	68	72	64	36	84	88	68	72	IPA
7	64	48	48	44	72	44	68	44	IPS
8	84	88	84	80	68	64	72	88	IPA
9	68	64	68	84	64	40	72	40	IPA
10	60	20	48	40	28	24	28	24	IPS

Data awal hasil ujian sebelum nilai numerik dibagi menjadi tiga kelompok Rendah, Sedang, dan Tinggi.

Untuk memudahkan pembuatan model, nilai-nilai yang berbeda kemudian dikelompokkan. Hasil pengujian dikelompokkan sebagai berikut:

1. Nilai yang jumlahnya kurang dari 60 tergolong rendah.
2. Nilai total > 61 hingga < 80 dianggap Sedang.
3. Nilai 81 atau lebih dianggap tinggi.

Setelah nilai-nilai yang disebutkan di atas dikonversi sesuai dengan tabel kategori, 100 record data merupakan data akhir dari konversi data training.

Tabel 5. Data Training

No	Indo	Inggris	PAI	Pkn	Mtk	Fisika	Biologi	Kimia	Jurusan
1	Tinggi	Sedang	Sedang	Tinggi	Sedang	Tinggi	Sedang	Tinggi	IPA
2	Sedang	Rendah	Sedang	Sedang	Tinggi	Tinggi	Sedang	Rendah	IPS
3	Sedang	Sedang	Sedang	Tinggi	Sedang	Rendah	Sedang	Rendah	IPS
4	Tinggi	Rendah	Sedang	Sedang	Tinggi	Tinggi	Sedang	Sedang	IPA

5	Tinggi	Sedang	Sedang	Rendah	Tinggi	Rendah	Tinggi	Tinggi	IPA
6	Sedang	Sedang	Sedang	Rendah	Tinggi	Tinggi	Sedang	Sedang	IPS
7	Sedang	Rendah	Rendah	Rendah	Sedang	Rendah	Sedang	Rendah	IPS
8	Tinggi	Tinggi	Tinggi	Sedang	Sedang	Sedang	Sedang	Tinggi	IPA
9	Sedang	Sedang	Sedang	Tinggi	Sedang	Rendah	Sedang	Rendah	IPS
10	Rendah	IPS							

**Data Testing**

Data pengujian selanjutnya harus disiapkan setelah data pengujian yang berjumlah 49 data telah disiapkan. Lampiran berisi Tabel Data Pengujian. 49 data pengujian yang telah diubah dibagi menjadi tiga kategori Rendah, Sedang, dan Tinggi.

Tabel 6. Data Testing

No	Indo	Inggris	PAI	Pkn	Mtk	Fisika	Biologi	Kimia	Jurusan
1	Sedang	Sedang	Sedang	Sedang	Sedang	Rendah	Sedang	Sedang	IPS
2	Tinggi	Rendah	Sedang	Sedang	Sedang	Rendah	Sedang	Sedang	IPS
3	Tinggi	Rendah	Tinggi	Tinggi	Sedang	Rendah	Sedang	Sedang	IPA
4	Tinggi	Sedang	Tinggi	Sedang	Tinggi	Tinggi	Sedang	Tinggi	IPA
5	Tinggi	Rendah	Sedang	Sedang	Tinggi	Sedang	Sedang	Rendah	IPA
6	Tinggi	Rendah	Sedang	Sedang	Sedang	Sedang	Sedang	Tinggi	IPS
7	Sedang	Rendah	Sedang	Sedang	Sedang	Rendah	Sedang	Rendah	IPS

**Entropi dan Gain**

Menurut tabel di bawah, hasil perhitungan Entropi dan Gain adalah sebagai berikut.:

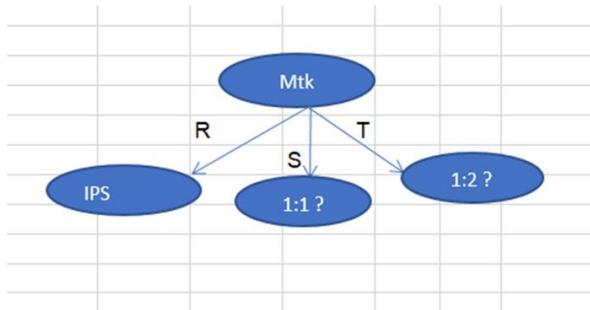
Tabel 7. Hitung Entropi dan Gain

Tentukan nilai gain dan entropi untuk setiap atribut.								
No	Atribut	Value	Jumlah Kasus	IP A	IP S	Entropi	Gain	
1	Total		149	59	90	0.9685		%
	Indo						0.163	16%
	Tinggi		80	48	32	0.971		

Tentukan nilai gain dan entropi untuk setiap atribut.								
No de	Atri but	Valu e	Juml ah Kas us	IP A	IP S	Entr opi	Gai n	
		Seda ng	64	9	55	0.5859		
		Ren dah	5	2	3	0.971		
	Ingg ris						0.1728	17%
		Ting gi	18	17	1	0		
		Seda ng	62	20	42	0.9072		
		Ren dah	69	22	47	0.9031		
	PAI						0.122	12%
		Ting gi	13	12	1	0.3912		
		Seda ng	93	39	54	0.9812		
		Ren dah	43	8	35	0.6931		
	Pkn						0.0035	0%
		Ting gi	18	8	10	0.9911		
		Seda ng	101	41	60	0.9743		
		Ren dah	30	10	20	0.9183		
	Mtk						0.2122	21%
		Ting gi	61	43	18	0.8752		
		Seda ng	85	16	69	0.6978		
		Ren dah	3	0	3	0		
	Fisik a						0.1204	12%
		Ting gi	26	20	6	0.7793		
		Seda ng	43	20	23	0.9965		
		Ren dah	80	19	61	0.7909		
	Biol ogi						0.1755	18%
		Ting gi	23	20	20	0.3507		
		Seda ng	115	38	77	0.9154		
		Ren dah	11	1	10	0.4395		
	Kimi a						0.1393	14%
		Ting gi	47	32	15	0.9035		

Tentukan nilai gain dan entropi untuk setiap atribut.								
No de	Atri but	Valu e	Juml ah Kas us	IP A	IP S	Entr opi	Gai n	
		Seda ng	41	16	25	0.965		
		Ren dah	61	11	50	0.6808		

Seperti yang dapat dilihat dari tabel di atas, karakteristik dengan nilai manfaat tertinggi atribut Mtk sebesar 21%, Mtk menjadi akar atau simpul pertama.



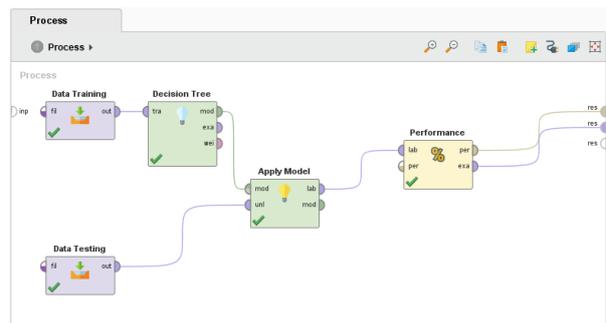
Gambar 3. akar atau simpul pertama

Mirip dengan perhitungan pertama, simpul 1.1 diperoleh dengan memfilter data menggunakan nilai Rendah dalam properti Mtk.

Atribut Mtk menjadi akar / node pertama, Bila Nilai Mtk R (Rendah) maka akan langsung masuk IPS, dan S (sedang), T (Tinggi) akan dilakukan perhitungan lebih lanjut dengan menghitung nilai entropy dan gain untuk mendapatkan hasil kategori Sedang dan Tinggi.

**Pengujian Model Decision Tree**

Pengujian data dilakukan dengan menggunakan aplikasi RapidMiner Studio, berikut gambar alur pengujian algoritma Decision Tree C4.5



Gambar 4. Alur Pengujian Decision Tree C4.5

Pada Gambar alur Pengujian Decision Tree C4.5 yaitu dengan menginputkan Data Training dan Data Testing kedalam aplikasi RapidMiner kemudian inputkan metode Decision Tree, Apply Model dan Performance

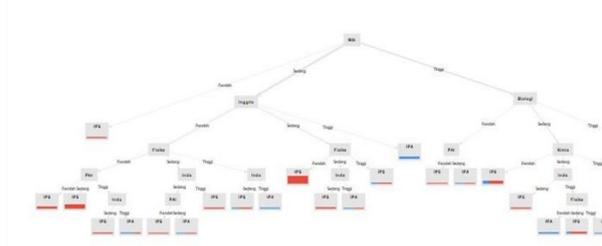
pada menu operator untuk mendapatkan hasil Akurasi dari metode Decision Tree.

	true IPS	true IPA	class precision
pred. IPS	19	11	63.33%
pred. IPA	1	18	94.74%
class recall	95.00%	82.07%	

Gambar 5. Hasil Akurasi Decision Tree

Akurasi sebesar 75,51% dicapai dalam pengujian yang dilakukan dengan model Decision Tree C4.5.

**Pohon Keputusan Decision Tree C4.5**

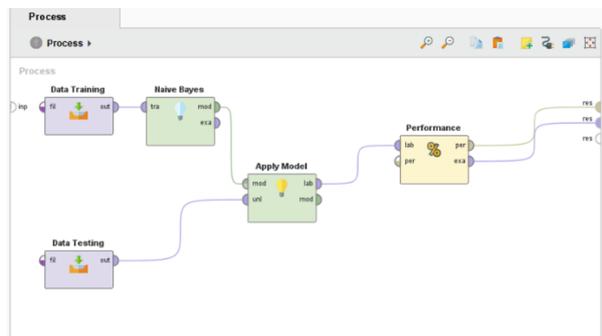


Gambar 7. Pohon Keputusan Decision Tree C4.5

Hal ini terbukti dari hasil prosedur klasifikasi jurusan siswa yang menggunakan Algoritma Decision Tree C4.5 pada gambar diatas bahwa karakteristik matematika memiliki dampak terbesar pada cara klasifikasi jurusan siswa. Variabel yang berada di simpul akar menunjukkan hal ini.

**Pengujian Model Naive Bayes**

Aplikasi RapidMiner Studio digunakan untuk menguji data, alur pengujian metode Naive Bayes ditunjukkan pada gambar berikut.



Gambar 8. Alur Pengujian Metode Naive Bayes

Tahapan pertama diawali dengan menginputkan Data Training melalui read excel, dan juga menginputkan Data Testing melalui read excel, kemudian hubungkan

Data Training menggunakan model Naive Bayes dan Insert Apply model, hubungkan Data Testing ke Apply model kemudian hubungkan dengan performance untuk mendapatkan hasil yang diinginkan.

Row No.	Jurusan	prediction	confidence	confidence	Indo	Inggris	PAI	IPA	IPA
1	IPS	IPS	0.022	0.078	Sedang	Sedang	Sedang	Sedang	Sed
2	IPS	IPS	0.156	0.844	Tinggi	Rendah	Sedang	Sedang	Sed
3	IPA	IPA	0.671	0.329	Tinggi	Rendah	Tinggi	Tinggi	Sed
4	IPA	IPA	0.998	0.002	Tinggi	Sedang	Tinggi	Sedang	Ting
5	IPA	IPA	0.508	0.492	Tinggi	Rendah	Sedang	Sedang	Ting
6	IPA	IPS	0.499	0.501	Tinggi	Rendah	Sedang	Sedang	Sed
7	IPS	IPS	0.009	0.991	Sedang	Rendah	Sedang	Sedang	Sed
8	IPA	IPA	0.992	0.008	Tinggi	Sedang	Sedang	Sedang	Ting
9	IPA	IPS	0.109	0.891	Tinggi	Rendah	Rendah	Rendah	Sed
10	IPS	IPS	0.009	0.991	Sedang	Sedang	Sedang	Sedang	Sed
11	IPA	IPA	0.713	0.287	Tinggi	Rendah	Sedang	Sedang	Sed
12	IPA	IPA	0.855	0.145	Sedang	Sedang	Rendah	Rendah	Ting
13	IPA	IPA	0.929	0.071	Tinggi	Rendah	Rendah	Sedang	Ting
14	IPA	IPA	0.994	0.006	Rendah	Tinggi	Tinggi	Sedang	Sed

Gambar 9. Hasil Apply Model Naive Bayes

Hasil prediksi penjurusan IPA dan IPS dengan menggunakan Metode Naive Bayes yang berjumlah 49 record dengan 8 atribut.

	true IPS	true IPA	class precision
pred. IPS	19	8	70.37%
pred. IPA	1	21	95.45%
class recall	95.00%	72.41%	

Gambar 10. Hasil pengujian Metode Naive Bayes

Hasil pengujian yang telah dilakukan dengan menggunakan model Naive Bayes memperoleh akurasi sebesar 81.63%.

**Analisa Hasil Perbandingan**

Model dengan menggunakan Algoritma C4.5 dan Naive Bayes yang diuji tingkat akurasinya menghasilkan Nilai Accuracy, Precision dan Recall dari kedua model.

Tabel 8. Nilai Accuracy, Precision, dan Recall

Metode →	Decision Tree	Naive Bayes
Hasil ↓		
Accuracy	75,51%	81,63%
Precision	63,33%	70,37%
Recall	95,00%	95,00%

Berdasarkan tabel di atas, yang membandingkan kinerja kedua algoritma, hasil pengujian Naive Bayes

menunjukkan nilai akurasi yang lebih besar daripada Decision Tree, dengan nilai akurasi Decision Tree sebesar 75,51% dan nilai akurasi algoritma Naïve Bayes sebesar 81,63%. Metode Naïve Bayes memiliki nilai presisi sebesar 70,37%, sedangkan algoritma Decision Tree memiliki nilai presisi sebesar 63,33%. Kedua algoritma memiliki nilai recall yang identik sebesar 95,00%. **Algoritma Naïve Bayes mengungguli** algoritma Decision Tree dalam hasil pengujian ini. Selanjutnya, bandingkan nilai AUC (Area Under Curve) kedua model.

Tabel 9. Perbandingan Performa Metode

Nilai ↓	Decision Tree	Naive Bayes
AUC	0,725	0,925

Berdasarkan tabel di atas, metode Naïve Bayes menghasilkan nilai tertinggi dengan nilai AUC sebesar 0,925. Dapat disimpulkan dari hasil perbandingan bahwa baik metode Decision Tree maupun Naïve Bayes memiliki prediksi yang sangat baik dengan nilai AUC masing-masing sebesar 0,725 dan 0,925 untuk metode Naïve Bayes dan Decision Tree. Jika dibandingkan dengan pendekatan decision tree, **Naïve Bayes menghasilkan nilai tertinggi**.

#### 4. Kesimpulan

Beberapa kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut :

Berdasarkan hasil penelitian yang dilakukan dengan menggunakan metode Decision Tree dan Naïve Bayes pada aplikasi RapidMiner, dapat dikatakan bahwa pengujian yang dilakukan dengan menggunakan 100 data training dan 49 data testing menghasilkan nilai akurasi untuk identifikasi jurusan siswa SMA PGRI 109 Kota Tangerang dengan menggunakan dua perbandingan algoritma klasifikasi data mining. Hasil dari masing-masing algoritma menunjukkan bahwa akurasi untuk Decision Tree sebesar 75,51% dan AUC sebesar 0,725. Namun, AUC = 0,925 dan akurasi Naïve Bayes = 81,63%, pada hasil pengujian tersebut Naïve Bayes lebih unggul daripada Decision Tree, dan Rekomendasi penggunaan algoritma terbaik dalam konteks sekolah ialah Naive Bayes.

Aplikasi/perangkat lunak RapidMiner digunakan untuk mengimplementasikan metode Decision Tree dan Naïve Bayes untuk memilih prediksi untuk penentuan utama. Hal ini menunjukkan bahwa, dibandingkan dengan pendekatan Decision Tree, metode Naïve Bayes merupakan cara yang sangat efektif untuk memecahkan masalah prediksi jurusan siswa.

Pengembangan studi selanjutnya diharapkan dapat mengintegrasikan algoritma dalam sistem informasi sekolah.

#### Daftar Rujukan

- [1] A. Arif, B. "Pengolahan Data mining menggunakan Tool Orange dan Python," 2013.
- [2] Bahar, "Penentuan Jurusan Sekolah Menengah Atas dengan Algoritma Fuzzy C-Means," 2011 Tesis Program Pasca Sarjana Udinus Semarang.
- [3] E. Bonfadini, "Introduction To Orange Data Mining. PYCON SETTE," 2016
- [4] H. Naparin, "Klasifikasi Peminatan Siswa SMA Menggunakan Metode Naive Bayes," 2016 SYSTEMIC.
- [5] Andriani, A. (2012). Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Drop Out. Seminar Nasional Matematika.
- [6] Andriani, A. (2013). Aplikasi Data Mining Market Basket Analisis Penjualan Suku Cadang Sepeda Motor Menggunakan Metode Association Rules Pada PT. Sejahtera Motor Gemilang. Jurnal Ilmiah Teknik Informatika Universitas Nusantara PGRI Kediri.
- [7] Ardiansyah. (2018). Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner. Jurnal Khatulistiwa Informatika.
- [8] B, A. A. (2013). Pengolahan Data mining menggunakan Tool Orange dan Python.
- [9] Bahar. (2011). Penentuan Jurusan Sekolah Menengah Atas dengan Algoritma Fuzzy C-Means. Tesis Program Pasca Sarjana Udinus Semarang.
- [10] Berry, A, M. J., & S, G. L. (2005). Introduction to Data Mining and Knowledge Discovery. Third Edition. Potomac: Two Crows.
- [11] Bonfadini, E. (2016). Introduction To Orange Data Mining. PYCON SETTE.
- [12] Bustami. (2014). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. Jurnal Informatika.
- [13] C, D. A. (2013). Belajar Data Mining dengan RapidMiner.
- [14] Meilani, B. D. (2015). Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naive Bayes. Jurnal Ilmiah NERO.
- [15] Musthofa, M. (2015). Pengembangan Sistem Pendukung Keputusan Penjurusan Bagi Siswa Baru Menggunakan Metode Naive Bayes.
- [16] Naparin, H. (2016). Klasifikasi Peminatan Siswa SMA Menggunakan Metode Naive Bayes. SYSTEMIC.
- [17] Ridwan, M., & dkk. (2013). Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. Jurnal EECIS Vol. 7, No. 1.
- [18] Saleh, A. (2014). Klasifikasi Metode Naive Bayes Dalam Data Mining Untuk Menentukan Konsentrasi Siswa (Studi Kasus di MAS PAB 2 Medan). KeTIK.
- [19] Saleh, A. (2015). Penerapan Data Mining Dalam Menentukan Jurusan Siswa. Seminar Nasional Informatika.
- [20] Sujai, I., & dkk. (2016). Prediksi Hasil Penjurusan Siswa Sekolah Menengah Atas Dengan Menggunakan Algoritma Decision Tree C4.5. Jurnal Teknologi Informasi.
- [21] Swastina, L. (2013). Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. Jurnal GEMA AKTUALITA.
- [22] Vlandari, R. T. (2017). Data Mining Teori dan Aplikasi RAPidminer. Yogyakarta: Gava Media