

PERBANDINGAN ALGORITMA C4.5 DAN NAIVE BAYES UNTUK PEMINATAN PENJURUSAN PADA SMK NEGERI 6 BATANGHARI

Wahyu Amrullah^{1*}, Kurnia budi², Joni Devitra³

¹Fakultas Ilmu Komputer, Program Studi Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: ¹wahyuambuk73@gmail.com, ²kurniabudi@unama.ac.id, ³devitrajoni@yahoo.co.id

Email Penulis Korespondensi: wahyuambuk73@gmail.com

Artikel Info :

Artikel History :

Submitted : 28-03-2025

Accepted : 04-06-2025

Published : 30-09-2025

Kata Kunci: Data Mining, Decision Tree C4.5, Naive Bayes, Penjurusan Siswa

Keyword: Data Mining, Decision Tree C4.5, Naive Bayes, Student Major

Abstrak— Penentuan jurusan di sekolah menengah kejuruan (SMK) merupakan langkah krusial dalam menentukan masa depan siswa. Secara global, proses ini mempengaruhi kualitas sumber daya manusia yang dihasilkan, karena keputusan tersebut berkaitan langsung dengan kompetensi dan karier siswa di masa depan. SMK Negeri 6 Batang Hari ialah sekolah menengah kejuruan negeri yang terletak di Desa Tapah Sari, Kec. Mersam, Kab. Batang Hari, Prov. Jambi. Penjurusan terdiri dari Multimedia, Teknik dan Bisnis Sepeda Motor, serta Agribisnis Tanaman Pangan dan Hortikultura. Pada SMK Negeri 6 Batanghari terdapat kendala dalam menentukan penjurusan siswa, dimana pada saat penentuan jurusan hanya berdasarkan penilaian subjektif atau hasil akademik tertentu tanpa mempertimbangkan data yang lebih komprehensif, seperti minat dan bakat siswa. Akibatnya, beberapa siswa merasa kurang cocok dengan jurusan yang dipilihkan, yang berdampak pada menurunnya motivasi belajar dan tingkat keberhasilan mereka. Salah satu solusi yang dapat dilakukan dengan menerapkan *data mining*. Perbandingan antara algoritma C4.5 dan *Naive Bayes* dilakukan karena efektivitas dan akurasi dalam pengambilan keputusan. Metode C4.5 sebagai algoritma pohon keputusan, memberikan interpretasi yang jelas dan mudah dipahami mengenai keputusan yang diambil, sedangkan *Naive Bayes* dengan pendekatan probabilitiknya, sering kali lebih cepat dan efisien dalam menangani dataset besar. Dari hasil analisis yang dilakukan, penelitian ini mengungkap bahwa kedua algoritma memiliki karakteristik dan performa yang berbeda dalam melakukan klasifikasi data peminatan siswa. Algoritma C4.5 yang berbasis pohon keputusan cenderung lebih mudah diinterpretasikan karena menghasilkan aturan keputusan yang jelas, sementara *Naive Bayes* yang berbasis probabilitas memiliki keunggulan dalam menangani ketidakpastian data.

Abstract— The determination of majors in vocational high schools (SMK) is a crucial step in shaping students' futures. Globally, this process influences the quality of human resources produced, as the decision is directly related to students' competencies and future careers. SMK Negeri 6 Batang Hari is a state vocational high school located in Tapah Sari Village, Mersam District, Batang Hari Regency, Jambi Province. The school offers majors in Multimedia, Motorcycle Engineering and Business, and Food Crops and Horticulture Agribusiness. At SMK Negeri 6 Batang Hari, there are challenges in determining students' majors, where the selection process is often based solely on subjective assessments or certain academic results without considering more comprehensive data, such as students' interests and talents. As a result, some students feel mismatched with the assigned major, leading to decreased learning motivation and academic success. One possible solution is the application of data mining. A comparison between the C4.5 and Naive Bayes algorithms is conducted due to their effectiveness and accuracy in decision-making. The C4.5 method, as a decision tree algorithm, provides clear and easily interpretable decision rules, while Naive Bayes, with its probabilistic approach, is often faster and more efficient when handling large datasets. The analysis reveals that both algorithms have different characteristics and performance in classifying students' interest data. The decision tree-based C4.5 algorithm tends to be more interpretable because it produces clear decision rules, while the probability-based Naive Bayes algorithm has advantages in dealing with data uncertainty.

1. PENDAHULUAN

Penentuan jurusan di sekolah menengah kejuruan (SMK) merupakan langkah krusial dalam menentukan masa depan siswa. Secara global, proses ini mempengaruhi kualitas sumber daya manusia yang dihasilkan, karena keputusan tersebut berkaitan langsung dengan kompetensi dan karier siswa di masa depan. Pentingnya penentuan jurusan yang tepat semakin ditekankan oleh kebutuhan industri yang terus berkembang, menuntut siswa memiliki keahlian spesifik sesuai dengan minat dan kemampuan mereka[1] Kenyataannya, masih banyak sekolah yang mengandalkan metode manual atau tidak terstruktur dalam menentukan jurusan siswa, sehingga

mengurangi akurasi dalam mengidentifikasi minat dan bakat mereka. Dengan perbandingan algoritma C4.5 dan Naïve Bayes dapat memprediksi pemintanan siswa sehingga siswa tidak kesulitan dalam menentukan jurusan. Tujuan dari penelitian ini adalah untuk menguji kemampuan algoritma C4.5 dan *Naïve Bayes* dalam memprediksi peminatan penjurusan siswa di SMK Negeri 6 Batang Hari.

Beberapa penelitian yang menggunakan metode *data mining* klasifikasi untuk penjurusan, diantaranya Penelitian dengan menggunakan metode C4.5 dengan hasil penelitian menghasilkan nilai akurasi sebesar 92% [3]. Penelitian yang dilakukan dengan tujuan membuat klasifikasi program studi untuk calon mahasiswa baru dengan metode *naïve bayes*. Dengan hasil penelitian yaitu hasil nilai akurasi sebesar 96,68% [4]. Penelitian menggunakan metode KNN untuk menentukan jurusan SMA, dengan hasil penelitian akurasi sebesar 79 % [5].

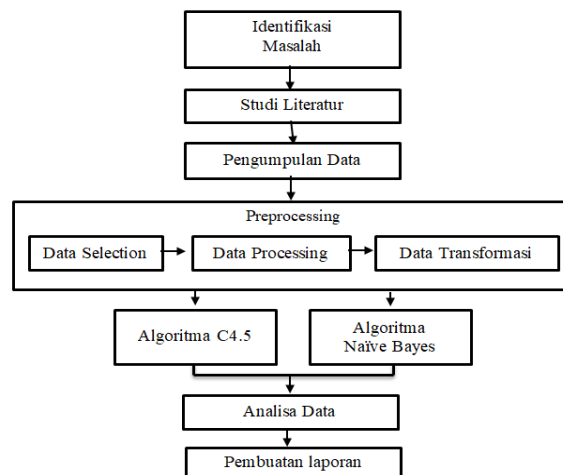
Selanjutnya penelitian yang menggunakan perbandingan metode klasifikasi, diantaranya: penelitian dengan tujuan untuk melakukan perbandingan antara algoritma klasifikasi *naïve bayes* dan Support Vector Machine dengan hasil akurasi 75% [6]. Perbandingan tujuan utama penelitian ini adalah untuk menilai tingkat akurasi perbandingan klasifikasi dengan hasil akurasi 71,25% dan 66,67% [7]. Perbandingan metode C4.5 dan *naïve bayes* Pemilihan penggunaan Algoritma C4.5 dan *Naïve Bayes* didasarkan pada beberapa alasan, yaitu kedua algoritma memiliki kesamaan mudah diimplementasikan dan dapat memberikan hasil yang baik dalam kasus klasifikasi [8]. Pemilihan metode C4.5 dan *naïve bayes* mudah dalam pengolahan data dengan hasil akurasi metode C4.5 91,25% dan *naïve bayes* hasil akurasi sebesar 87,11% [9]. penelitian dengan perbandingan metode C4.5 dan *naïve bayes* didasarkan karena algoritma C4.5 memberikan aturan rule dan pohon keputusan dalam meningkatkan keakuratan sebuah prediksi yang diteliti, sedangkan *naïve bayes* klasifikasi statistik yang mampu diterapkan dalam memprediksi probabilitas pengelompokan suatu class [10].

Berdasarkan penelitian sebelumnya yang membandingkan beberapa metode klasifikasi dapat disimpulkan bahwa perbandingan algoritma C4.5 dan *Naïve Bayes* dipilih karena metode C4.5 dan *naïve bayes* menghasilkan nilai akurasi yang tinggi dan merupakan metode yang mudah diimplementasikan. Oleh sebab itu dalam penelitian ini peneliti menggunakan perbandingan algoritma C4.5 dan *Naïve Bayes*. Perbandingan antara algoritma C4.5 dan *Naïve Bayes* dilakukan karena efektivitas dan akurasi dalam pengambilan keputusan. Metode C4.5 sebagai algoritma pohon keputusan, memberikan interpretasi yang jelas dan mudah dipahami mengenai keputusan yang diambil, sedangkan *Naïve Bayes* dengan pendekatan probabilitiknya, sering kali lebih cepat dan efisien dalam menangani dataset besar [11]. Perbandingan ini diharapkan dapat memberikan wawasan tentang metode mana yang menghasilkan nilai *accuracy*, *precision* yang tinggi untuk penentuan peminatan penjurusan.

Berdasarkan penelitian terdahulu disimpulkan bahwa algoritma C4.5 dan *Naïve Bayes* dipilih karena metode C4.5 dan *naïve bayes* menghasilkan nilai akurasi yang tinggi dan merupakan metode yang mudah diimplementasikan. Oleh sebab itu dalam penelitian ini peneliti menggunakan perbandingan algoritma C4.5 dan *Naïve Bayes*. Perbandingan antara algoritma C4.5 dan *Naïve Bayes* dilakukan karena efektivitas dan akurasi dalam pengambilan keputusan. Metode C4.5 sebagai algoritma pohon keputusan, memberikan interpretasi yang jelas dan mudah dipahami mengenai keputusan yang diambil, sedangkan *Naïve Bayes* dengan pendekatan probabilitiknya, sering kali lebih cepat dan efisien dalam menangani dataset besar .

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



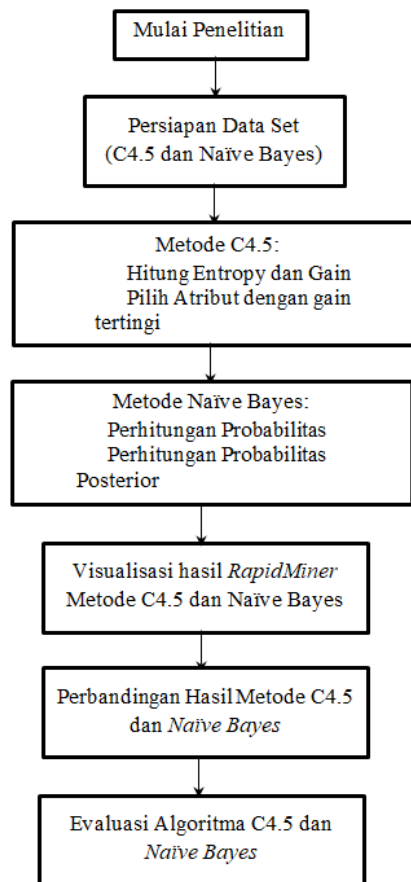
Gambar 1. Tahapan Penelitian

Berdasarkan alur penelitian pada gambar diatas, maka dapat diuraikan menjadi pembahasan pada masing-masing tahapan. alur penelitian yang digunakan sebagai berikut:

1. Identifikasi Masalah
Pada tahap ini, melakukan identifikasi dan merumuskan permasalahan penelitian, dengan melakukan observasi langsung ke objek penelitian serta melakukan wawancara kepada pihak terkait, untuk kemudian merumuskan masalah yang terjadi pada objek penelitian tersebut.
2. Studi Literatur
Tahap ini dilakukan peincarian landasan-landasan teori yang reileivan meilalui beirbagai artikeil, buku, jurnal dan juga internet untuk yang bertujuan meileingkapi konsep dan teori dan keiilmuan yang baik seirta meinambah wawasan dan peingeitahuan peineiliti.
3. Pengumpulan Data
Tahap ini adalah pengumpulan data penelitian, Data memiliki kedudukan yang sangat penting dalam penelitian. Kualitas data menentukan kualitas penelitian, beberapa metode pengumpulan data diterapkan, antara lain:
 1. Pengamatan (Observasi)
Pada tahap ini melakukan pengamatan secara langsung pada objek yang akan diteliti. Dalam hal ini penulis mengamati mengenai pemilihan jurusan yang dilakukan oleh siswa di SMK Negeri 6 Batang Hari.
 2. Wawancara
Pada metode ini penulis melakukan wawancara secara langsung dengan Kepala Sekolah SMK Negeri 6 Batanghari yang bertujuan untuk memperoleh data dan informasi akurat mengenai penjurusan siswa.
4. Preprocessing
Pada proses Preprocessing dilakukan melalui beberapa tahapan, yang akan dijelaskan di bawah ini
 1. Pemilihan Data
Sebelum melakukan penggalian informasi dalam data mining, perlu dilakukan seleksi pada kumpulan data operasional.Pada tahap ini, peneliti memilih data untuk memastikan kesesuaian dengan kebutuhan analisis.
 2. Pengolahan Data
Di tahap ini, peneliti menyiapkan data yang diperoleh dari riwayat data nilai siswa Setelah itu, peneliti melakukan seleksi terhadap atribut yang berpengaruh.Pada tahap ini juga dilakukan penghapusan data duplikat dan pemeriksaan terhadap data yang tidak konsisten.
 3. Transformasi Data
Pada tahap ini, peneliti melakukan proses transformasi terhadap data yang telah dipilih, agar data tersebut siap untuk proses data mining. Peneliti mengubah atau menggabungkan data ke dalam format yang sesuai untuk analisis lebih lanjut.
 4. Klasifikasi dengan Algoritma C4.5 dan Naïve Bayes
Di tahap ini, penulis menganalisis hasil perhitungan menggunakan metode C4.5 dan Naïve Bayes serta membandingkannya melalui perangkat lunak *Rapidminer Studio*
5. Analisis Data
Pada tahap ini penulis menganalisis hasil perbandingan dari perhitungan algoritma C4.5 dan Naïve Bayes menggunakan dengan tools Rapidminer Studio.
6. Penulisan Laporan Penelitian
Pada tahap ini peimbuatan laporan akhir peineilitian, dilakukan beirdasarkan keirangka keirja yang telah di rancang yang teirdiri dari Peindahuluan, Landasan Teiori, Meitodologi Peineilitian, Analisis, Hasil dan Reikomeindasi dan Peinutup seirta lampiran bukti hasil peineilitian.

2.2 Alur Experiment

Penelitian ini menggunakan dua metode klasifikasi, yaitu C4.5 dan Naïve Bayes. Pada kedua metode tersebut memiliki cara kerja yang berbeda dalam mengklasifikasikan data dan masing-masing memberikan hasil klasifikasi yang spesifik sesuai dengan karakteristik data. Adapun diagram alur experiment terdapat pada gambar



Gambar 2. Diagram Alur Experiment

Berdasarkan tabel 3.2 dapat diuraikan penjelasan sebagai berikut:

1. Persiapan Data Set
Pada tahap ini peneliti mempersiapkan dataset yang akan digunakan untuk metode C4.5 dan *Naïve Bayes*. Pada penelitian ini data yang digunakan adalah data nilai siswa.
2. Metode C4.5
Pada tahap ini menghitung entropy dan gain untuk setiap atribut. Dan memilih atribut dengan nilai gain tertinggi untuk membangun pohon keputusan. Kemudian dihitung kembali untuk entropy dan gain hingga sampai memenuhi kriteria tertentu.
3. Metode *Naïve Bayes*
Pada tahap ini menghitung probabilitas prior dan likelihood. Dan Menghitung probabilitas posterior untuk prediksi.
4. Visualisasi Hasil Rapidminer
Pada tahap ini peneliti memvisualisasikan hasil perhitungan metode C4.5 dan *Naïve Bayes* dengan menggunakan tools *Rapidminer* dengan menghitung nilai *accuracy*, *5 fold cross validation* dan *10 fold cross validation*
5. Evaluasi Algoritma C4.5 dan *Naïve Bayes*
Evaluasi dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa pengujian benar. Pengujian dilakukan mengukur keakuratan hasil, tujuan evaluasi untuk mengetahui mana yang terbaik dari hasil uji. Akurasi didefinisikan tingkat kedekatan antara prediksi dan nilai aktual, pengukuran akurasi terhadap model menggunakan *Confusion Matrix*[12]. Berikut rumus yang digunakan untuk mencari TPR (*True Positive Rate*), FPR (*False Positive Rate*), Akurasi, dan *F-Measure* pada *confusion matrix*[13]:

$$TPR = \frac{TP}{(TP+FN)}$$

$$FPR = \frac{FP}{(FP+TN)}$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

$$F - Measure = \frac{2*Precistion*Recall}{Precistion+Recall}$$

(1)

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Algoritma C4.5 dan Naïve Bayes Untuk Peminatan Penjurusan Siswa Pada SMK Negeri 6 Batanghari

3.1.1 Data Set

Tabel 1. Variabel Data

NO	NAMA	NO PENDAFTARAN	MTK	B.INNGRIS	B.INDO	IPA	IPS	PSIKOTES	JURUSAN
1	Ahlan Najla	20240611	sedang	Tinggi	sedang	sedang	sedang	Rendah	ATPH
2	Ahmad Nuril Anwar	20240629	sedang	sedang	Tinggi	sedang	sedang	Rendah	ATPH
3	Alhaji Diof Saputra	20240635	rendah	sedang	sedang	sedang	Tinggi	Sedang	ATPH
.....
138	Romidi	20240613	sedang	sedang	rendah	sedang	sedang	Rendah	TBSM
139	Solman	20240619	sedang	sedang	rendah	sedang	sedang	Rendah	TBSM
140	Walyasir	20240614	sedang	sedang	rendah	sedang	sedang	Rendah	TBSM

3.1.2 Pengolahan Data Algoritma C4.5

Setelah data nilai siswa melalui tahap seleksi dan pembersihan, langkah berikutnya adalah pengolahan data menggunakan algoritma C4.5. Proses pengelompokan ini dilakukan dengan mengikuti beberapa tahapan berikut:

1. Pilihlah atribut sebagai akar
2. Buatlah cabang untuk setiap nilai
3. Beri kasus dalam cabang
4. Ulangi setiap proses untuk setiap cabang sehingga seluruh kasus pada kasus memiliki kelas yang sama.

Untuk memilih atribut menjadi node akar, berdasarkan pada nilai Gain paling tinggi dari atribut-atribut yang telah ada. Untuk menghitung Gain dipergunakan rumus yang tertera dalam persamaan berikut:

$$Entropy(s) = -\sum_{i=1}^n p_i \cdot \log_2 p_i \tag{2}$$

$$Gain(S,A) = Entropy(s) - \sum_{i=1}^n \left(\frac{|S_i|}{|S|} \cdot Entropy(S_i) \right) \tag{3}$$

Untuk perhitungannya disajikan sebagai berikut:

$$Entropy = ((-89/136) \cdot \log_2(89/136)) + ((-36/136) \cdot \log_2(36/136)) + ((-89/136) \cdot \log_2(89/136)) + ((-59/136) \cdot \log_2(59/136)) = 0.927973$$

Dari perhitungan diatas disajikan dalam tabel 2.

Tabel 2. Hasil Perhitungan Jarak Dengan Centroid (Iterasi 1)

NODE	ATRIBUT	JUMLAH KASUS (S)	ATPH	MULTIMEDI A	TBSM	ENTROPY	GAIN
1	TOTAL	136	84	70	59	1,445214	
	MTK						1,289
	Seda	83	30	3	50	0	
	ng						
	Ren	17	7	1	9	1,253298	
	dah						
	Ting	36	5	31	0	0	
	gi						
	B.INGGRIS						1,262
	Seda	97	37	13	47	0	
ng							
Ren	19	3	4	12	1,312431		
dah							
Ting	20	2	18	0	0		
gi							
B.INDONESIA							1,126

	Seda	72	26	0	46	0	
	ng						
	Ren	28	7	9	12	1,550199	
	dah						
	Ting	34	9	25	0	0	
	gi						
IPA							0,894
	Seda	59	23	2	34	0	
	ng						
	Ren	50	12	13	25	1,499423	
	dah						
	Ting	27	7	20	0	0	
	gi						
IPS							0,929
	Seda	59	24	1	34	0	
	ng						
	Ren	48	14	9	25	1,461448	
	dah						
	Ting	29	4	25	0	0	
	gi						
PSIKOTES							0,833
	Ting	78	23	19	36	0	
	gi						
	Seda	28	10	6	12	1,530619	
	ng						
	Ren	28	9	9	10	0	
	dah						

Setelah dilakukan perhitungan nilai *entropy* dan *gain*, nilai *gain* yang tertinggi akan dijadikan *node* (simpul) pada pohon keputusan. Dilakukan perhitungan hingga nilai pada *gain* tidak memenuhi kriteria minimum.

3.1.3 Pengolahan Data Naïve Bayes

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

1. Perhitungan Probabilitas

Probabilitas prior adalah peluang awal setiap kelas sebelum memperhatikan fitur data. Hasil perhitungan probabilitas disajikan pada tabel 3

Tabel 3. Probabilitas Kelas

Probabilitas Kelas	
Rekomendasi	NILAI
ATPH	31%
MULTIMEDIA	25%
TBSM	44%
Jumlah	100%

Dari tabel 3 dapat disimpulkan bahwa untuk jumlah keseluruhan kelas mencapai 100%.Selanjutnya dilakukan perhitungan untuk mencari nilai probabilitas dari setiap atribut.

a. Atribut MTK

Pada atribut MTK diketahui bahwa siswa pada jurusan TSM dengan kriteria nilai Baik memiliki tingkat probabilitas tinggi yaitu sebesar 100%. Perhitungan dapat dilihat pada tabel 4

Tabel 4. Atribut MTK

P(MTK= 1	ATPH	MULTIMEDIA	TBSM
Sedang	71,43%	8,57%	84,75%
Rendah	16,67%	2,86%	15,25%
Tinggi	12,20%	85,71%	0,00%

b. Atribut Bahasa Inggris

Pada atribut Bahasa Inggris diketahui bahwa siswa pada jurusan TSM dengan kriteria nilai Baik memiliki tingkat probabilitas tinggi yaitu sebesar 100%. Perhitungan dapat dilihat pada tabel 5

Tabel 5. Atribut Bahasa Inggris

P(B.INGGRIS= 1	ATPH	MULTIMEDIA	TBSM
Sedang	88%	37%	80%
Rendah	7%	11%	20%
Tinggi	5%	51%	0%

c. Atribut Bahasa Indonesia

Pada atribut Bahasa Indonesia diketahui bahwa siswa pada jurusan TSM dengan kriteria nilai Baik memiliki tingkat probabilitas tinggi yaitu sebesar 100%. Perhitungan dapat dilihat pada tabel 6.

Tabel 6. Atribut Bahasa Indonesia

P(B.INDONESIA= 1	ATPH	MULTIMEDIA	TBSM
Sedang	62%	0%	78%
rendah	17%	26%	20%
tinggi	21%	71%	0%

d. Atribut IPA

Pada atribut IPA diketahui bahwa siswa pada jurusan TSM dengan kriteria nilai Baik memiliki tingkat probabilitas tinggi yaitu sebesar 100%. Perhitungan dapat dilihat pada tabel 7.

Tabel 7. Atribut IPA

P(IPA= 1	ATPH	MULTIMEDIA	TBSM
Sedang	55%	6%	58%
Rendah	29%	37%	42%
Tinggi	17%	57%	0%

e. Atribut IPS

Pada atribut IPS diketahui bahwa siswa pada jurusan TSM dengan kriteria nilai Baik memiliki tingkat probabilitas tinggi yaitu sebesar 100%. Perhitungan dapat dilihat pada tabel 8.

Tabel 8. Atribut IPS

P(IPS= 1	ATPH	MULTIMEDIA	TBSM
Sedang	57%	3%	58%
Rendah	33%	26%	42%
Tinggi	10%	71%	0%

f. Atribut Psikotes

Pada atribut Psikotes diketahui bahwa siswa pada jurusan ATPH dengan kriteria nilai Baik memiliki tingkat probabilitas tinggi yaitu sebesar 86%. Perhitungan dapat dilihat pada tabel 9.

Tabel 9. Atribut Psikotes

P(PSIKOTES= 1	ATPH	MULTIMEDIA	TBSM
Sedang	24%	17%	20%
Rendah	21%	26%	17%
Tinggi	55%	54%	61%

3.1.4 Hasil Algoritma C4.5 dan Naïve Bayes Dengan Tools RapidMiner

Pada tahap ini dilakukan perhitungan algoritma C4.5 dan Naïve Bayes dengan *Tools Rapidminer* sebagai perbandingan dengan perhitungan manual yang telah dilakukan dengan menggunakan *Microsoft excel*.

1. Algoritma C4.5

Klasifikasi algoritma C4.5 dengan menggunakan data set sebanyak 136 data yang telah dilakukan *cleaning data* dan transformasi data. Hasil klasifikasi algoritma C4.5 menggunakan *tools Rapidminer*.

a. Hasil Akurasi Algoritma C4.5

Tabel 10. Hasil Akurasi Algoritma C4.5

accuracy : 62.50% +/-5.20%				
	true ATPH	true MULTIMEDIA	true TBSM	class precision
pred. APTH	9	3	14	34.62%
pred. MULTIMEDIA	6	31	0	83.78%
pred. TBSM	27	1	45	61.64%
class recall	21.43%	88.57%	76.27%	

b. Hasil Validation 5 Fold Cross Validation

Tabel 11. Hasil Validation 5 Fold Cross Validation

accuracy : 66.93% +/-8.96% (micro average : 66.91%)				
	true ATPH	true MULTIMEDIA	true TBSM	class precision
pred. APTH	15	2	11	53.57%
pred. MULTIMEDIA	2	31	3	86.11%
pred. TBSM	25	2	45	62.50%
class recall	35.71%	88.57%	76.27%	

c. Hasil Validation 10 Cross Validation

Tabel 12. Hasil Validation 10 Fold Cross Validation

accuracy : 69.34% +/-8.96% (micro average : 69.12%)				
	true ATPH	true MULTIMEDIA	true TBSM	class precision
pred. APTH	19	4	14	51.35%
pred. MULTIMEDIA	2	31	1	91.18%
pred. TBSM	21	0	44	67.69%
class recall	45.24%	88.57%	74.58%	

2. Naïve Bayes

Klasifikasi menggunakan algoritma *naïve bayes* dengan menggunakan data set sebanyak 136 data yang telah dilakukan *cleaning data* dan transformasi data. Hasil klasifikasi *naïve bayes* menggunakan *tools Rapidminer*.

a. Hasil Akurasi Algoritma Naïve Bayes

Tabel 13. Hasil Akurasi Naïve Bayes

accuracy : 79.41% +/-8.96% (micro average : 79.41%)				
	true ATPH	true MULTIMEDIA	true TBSM	class precision
pred. APTH	24	4	6	70.59%
pred. MULTIMEDIA	1	31	0	96.88%
pred. TBSM	17	0	53	75.71
class recall	57.14%	88.57%	89.83%	

b. Hasil Validation 5 Fold Cross Validation

Tabel 14. Hasil Validation 5 Fold Cross Validation

accuracy : 83.10% +/-8.96% (micro average : 83.09%)				
	true ATPH	true MULTIMEDIA	true TBSM	class precision
pred. APTH	24	4	1	82.76%
pred. MULTIMEDIA	1	31	0	96.88%
pred. TBSM	17	0	58	77.33%

class recall	57.14%	88.57%	98.31%
--------------	--------	--------	--------

c. Hasil Validation 10 Fold Cross Validation

Tabel 15. Hasil Validation 10 Fold Cross Validation

accuracy : 83.90% +/-7.38% (micro average : 83.82%)

	true ATPH	true MULTIMEDIA	true TBSM	class precision
pred. APTH	24	4	0	85.71%
pred. MULTIMEDIA	1	31	0	96.88%
pred. TBSM	17	0	59	77.63%
class recall	57.14%	88.57%	100.00%	

3.1.5 Evaluasi Algoritma C4.5 dan Naïve Bayes

Evaluasi dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa pengujian benar. Pengujian dilakukan mengukur keakuratan hasil, tujuan evaluasi untuk mengetahui mana yang terbaik dari hasil uji. Akurasi didefinisikan tingkat kedekatan antara prediksi dan nilai aktual, pengukuran akurasi terhadap model menggunakan *Confusion Matrix*

Tabel 16. Hasil Confusion Matrix

Algoritma	Kelas	TPR	FPR	Accuracy	F-Measure
C4.5	APTH	58.54%	17.65%	79.41%	63.12%
	MULTIMEDIA	88.57%	4.60%	79.41%	92.54%
	TBSM	89.83%	9.23%	79.41%	82.16%
Naïve Bayes	APTH	58.54%	17.65%	79.41%	63.12%
	MULTIMEDIA	88.57%	4.60%	79.41%	92.54%
	TBSM	89.83%	9.23%	79.41%	82.16%

Pada tabel 16 menampilkan hasil perhitungan dari confusion matrix untuk algoritma C4.5 dan Naïve Bayes.

4. KESIMPULAN

Dari hasil analisis yang dilakukan, penelitian ini mengungkap bahwa kedua algoritma memiliki karakteristik dan performa yang berbeda dalam melakukan klasifikasi data peminatan siswa. Dimana algoritma C4.5 cenderung untuk lebih mudah diinterpretasikan karena menghasilkan aturan keputusan yang jelas, sementara Naïve Bayes yang berbasis probabilitas memiliki keunggulan dalam menangani ketidakpastian data. Dapat disimpulkan bahwa hasil penelitian adalah metode algoritma C4.5 merupakan metode yang terbaik dalam penentuan jurusan karena algoritma C4.5 memiliki nilai akurasi yang tinggi dibandingkan dengan algoritma Naïve Bayes. Penelitian ini memberikan kontribusi dalam pengembangan model rekomendasi berbasis data mining yang dapat digunakan oleh SMK Negeri 6 Batang Hari untuk meningkatkan akurasi dan objektivitas dalam menentukan peminatan jurusan siswa. Dengan adanya model rekomendasi ini, diharapkan proses penjurusan siswa tidak hanya didasarkan pada intuisi atau penilaian subjektif, tetapi juga pada analisis data yang lebih sistematis dan terukur.

Penelitian ini dapat menjadi dasar bagi pengembangan sistem pendukung keputusan dalam proses penjurusan di tingkat SMK, khususnya bagi sekolah-sekolah yang menghadapi tantangan dalam menentukan jurusan yang sesuai dengan minat dan potensi siswa. Dengan implementasi sistem yang berbasis kecerdasan buatan dan analisis data, sekolah dapat meningkatkan efisiensi dan efektivitas dalam menyusun strategi pendidikan vokasional yang lebih adaptif terhadap kebutuhan siswa dan dunia industri.

REFERENCES

- [1] E. A. Juliawati, "Pemilihan Jurusan SMA Menggunakan Sistem Pendukung Keputusan pada Siswa SMPN 210 Jakarta Menggunakan Metode Profile Matching," vol. 3, no. 3, pp. 397–403, 2023.
- [2] F. Huda Aminuddin, T. Djauhari, and A. Ardiansyah, "Penentuan Jurusan Pada Smkn 1 Muaro Jambi Dengan Metode K-Means Clustering," *J. Akad.*, vol. 15, no. 1, pp. 76–82, 2022, doi: 10.53564/akademika.v15i1.846.
- [3] D. Yulia, A. Pandu Kusuma, and D. Fanny Hebrasianto Permadi, "Penerapan Algoritma C4.5 Untuk

- Prediksi Minat Penjurusan Siswa Di Smkn 1 Kademangan,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 6, no. 2, pp. 893–900, 2022, doi: 10.36040/jati.v6i2.5781.
- [4] W. E. Nugroho, A. Sofyan, and O. Somantri, “Metode Naive Bayes Dalam Menentukan Program Studi Bagi Calon Mahasiswa Baru,” *Infotek Mesin,* vol. 12, no. 1, pp. 59–64, 2021, doi: 10.35970/infotek mesin.v12i1.491.
- [5] S. Nuraeni, S. P. A. Syam, M. F. Wajdi, B. Firmansyah, and M. Malkan, “Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa di SMAN 02 Manokwari,” *G-Tech J. Teknol. Terap.,* vol. 7, no. 1, pp. 89–95, 2023, doi: 10.33379/gtech.v7i1.1905.
- [6] D. D. Dewi *et al.*, “Perbandingan Naïve Bayes Classifier Dan Support Vector Machine Dalam Mengklasifikasikan Tingkat Pengangguran Terbuka Di Indonesia,” vol. 9, pp. 788–796, 2024.
- [7] M. Jannah, “Perbandingan Metode Naïve Bayes Dan K-Nearest Neighbor Dalam Mengklasifikasi Status Pertumbuhan Anak Stunting (Studi Kasus: Posyandu Cemara),” vol. 14, no. 1, pp. 250–255, 2022, [Online]. Available: https://repository.unilak.ac.id/id/eprint/2807%0Ahttps://repository.unila.ac.id/2807/1/1855201057_BA-B-I_VI_DAFTAR-PUSTAKA.pdf
- [8] Rian Pratama, B. Huda, E. Novalia, and H. Kabir, “Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Menentukan Persediaan Stok,” *Matic J.,* vol. 6, no. 2, pp. 115–122, 2022, doi: 10.47002/metik.v6i2.379.
- [9] E. Fitriani, “Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan,” *Sistemasi,* vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.
- [10] D. S. Afianti and A. Jananto, “Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Data Pesawat Delay,” *semanTIK,* vol. 9, no. 1, p. 17, 2023, doi: 10.55679/semantik.v9i1.29755.
- [11] Y. I. Kurniawan, “Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining,” *J. Teknol. Inf. dan Ilmu Komputer.,* vol. 5, no. 4, pp. 455–464, 2019, doi: 10.25126/jtiik.201854803.
- [12] D. Y. Kardono, Y. M. Pranoto, and E. Setyati, “Prediksi Kecocokan Jurusan Siswa SMK Dengan Support Vector Machine dan Random Forest,” *Teknika,* vol. 12, no. 1, pp. 11–17, 2023, doi: 10.34148/teknika.v12i1.567.
- [13] M. F. D. Ikhbal and D. Irfan, “Menentukan Penjurusan Siswa dengan Menggunakan Metode Decision Tree Algoritma C4.5,” *Jav. J. Vokasi Inform.,* pp. 2–8, 2021, doi: 10.24036/javit.v1i3.40.
- [14] V. Anestiviya, A. Ferico, and O. Pasaribu, “Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar),” *J. Teknol. dan Sist. Inf.,* vol. 2, no. 1, pp. 80–85, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [15] M. Roghib, N. Rahaningsih, and R. D. Dana, “Penerapan Algoritma C4.5 Untuk Seleksi Penjurusan Siswa Baru Pada Sekolah Menengah Kejuruan (Studi Kasus: Smk Plus Al-Hilal Arjawinangun),” *J. Mhs. Tek. Inform.,* vol. 8, no. 1, p. Vol. 8,-No. 1, 2024.