

PERBANDINGAN METODE NAÏVE BAYES DAN ALGORITMA C4.5 DALAM PREDIKSI KELULUSAN SISWA BERDASARKAN UJI KOMPETENSI

Arinal Khaq^{1*}, Heri Kurniawan²

^{1,2} Matematika, Universitas Terbuka, Tangerang Selatan Indonesia

*Penulis korespondensi: 042132118@ecampus.ut.ac.id

ABSTRAK

Uji Kompetensi Profesi menjadi salah satu penentu kelulusan siswa SMK Tunas Harapan Pati dan penerbitan sertifikat kompetensi. Namun, hasil Uji Kompetensi menunjukkan adanya ketidaklulusan siswa pada beberapa klaster. Ketidaklulusan terbanyak terjadi pada klaster Mikrobiologi dimana pada klaster tersebut sangat bergantung pada keterampilan siswa dalam menjaga kesterilan lingkungan kerjanya. Penelitian ini bertujuan untuk memprediksi kelulusan siswa pada Uji Kompetensi melalui penerapan teknik data mining. Dengan menggunakan Metode Naïve Bayes dan Algoritma C4.5, penelitian ini bertujuan untuk menentukan model prediksi terbaik untuk membantu pihak sekolah mengidentifikasi siswa yang berpotensi tidak lulus dan memberikan solusi yang tepat dalam pencegahan hal yang dapat mempengaruhi kelulusan siswa. Hasil penelitian diharapkan dapat meningkatkan tingkat kelulusan siswa pada Uji Kompetensi. Hasil penelitian menunjukkan bahwa dengan metode *Naïve Bayes* memperoleh tingkat akurasi 99,64% sedangkan metode Algoritma C4.5 memperoleh tingkat akurasi sebesar 99,29%. Diperoleh metode yang tepat untuk kasus prediksi kelulusan siswa di SMK Tunas Harapan Pati dengan menggunakan metode *Naïve Bayes*.

Kata Kunci: Algoritma C4.5, Data Mining, Naïve Bayes, Prediksi Kelulusan Siswa, Uji Kompetensi

1 PENDAHULUAN

Uji kompetensi profesi merupakan serangkaian tes yang dirancang untuk menilai sejauh mana keterampilan professional siswa SMK sesuai dengan standar kompetensi di bidang pekerjaan tertentu. Uji kompetensi bertujuan untuk mengukur dan memastikan bahwa siswa telah menguasai kompetensi yang dibutuhkan dalam profesi yang telah dipelajari selama SMK. Uji ini dilakukan dengan beberapa metode seperti tes tulis, tes praktik serta wawancara.

Tujuan dilakukannya uji kompetensi profesi di SMK yaitu validasi kompetensi yaitu mengukur penguasaan siswa terhadap standar kompetensi yang dibutuhkan dalam dunia industri. Sertifikasi yaitu memberikan sertifikat kompetensi bagi siswa yang lulus uji sebagai bukti bahwa mereka telah memenuhi standar profesi yang diakui, kesiapan kerja dimana memastikan siswa siap terjun ke dunia industri dengan keterampilan yang sesuai dengan tuntutan pasar. SMK Tunas Harapan Pati melakukan uji kompetensi profesi sebagai salah satu syarat kelulusan sekolah. Uji kompetensi ini dilakukan pada siswa kelas XII sebelum diputuskannya informasi kelulusan sekolah. Standar uji kompetensi yang digunakan mengacu pada SKKNI LEVEL II dengan 5 klaster uji oleh LSP SMK Tunas Harapan Pati dengan penerbit sertifikat oleh BNSP.

Ketidakmampuan siswa untuk lulus dalam seluruh kluster pengujian dalam uji kompetensi dihadapi oleh beberapa jurusan di SMK Tunas Harapan Pati. Jurusan Kimia Analisis adalah salah satu hal yang mengalami hal tersebut. Dalam beberapa kasus, siswa mengalami ketidakkulusan dalam melakukan uji kompetensi yang diselenggarakan. Dari 284 data siswa kimia analisis tahun lulusan 2019/2020 sampai tahun 2023/2024 terdapat 241 siswa yang dinyatakan lulus untuk semua kluster uji kompetensi sedangkan 43 siswa mengalami ketidaklulusan di salah satu kluster yang diujikan. Hal tersebut mengakibatkan tidak bisa diterbitkannya sertifikat uji kompetensi oleh BNSP untuk siswa yang tidak bisa lulus dikeseluruhan kluster uji.

Data mining merupakan istilah untuk menjabarkan penemuan pengetahuan dalam database, melalui penggunaan matematika, Teknik statistik, *machine learning*, ataupun kecerdasan buatan guna melakukan identifikasi dan ekstraksi berbagai informasi yang dapat dimanfaatkan, serta pengetahuan yang relevan dari berbagai database besar. Data mining dalam suatu penelitian pada dasarnya bukan topik baru, sebab dapat berguna dalam meningkatkan efektivitas berbagai teknik sebelumnya dimana dapat digunakan sehingga dapat menangani beragam masalah yang biasa dijumpai (Mahar et al., 2023). Menurut Etriyanti (2020) terdapat beberapa cara untuk menyelesaikan masalah prediksi, salah satunya dengan teknik penambangan data (*data mining*). Teknik *data mining* merupakan salah satu cara yang mudah dan cepat untuk mendapatkan pengetahuan secara otomatis (*knowledge*).

Bagian penting pada *data mining* adalah teknik klasifikasi, yaitu cara yang digunakan untuk mempelajari data set untuk didapatkan hubungan antar data yang membentuk pola (*pattern*) (Setiyani et al., 2020). Terdapat beberapa metode untuk data mining yang digunakan untuk klasifikasi. Algoritma yang sering digunakan seperti *Artificial Neural Network*, *Algoritma C4.5*, *Nearest Neighbour Rule*, *Fuzzy Logic*, *Naïve Bayes*, *K-Mean*, dan lain-lain. Penelitian sebelumnya yang mengangkat tentang penerapan data mining telah banyak digunakan sebelumnya (Etriyanti 2020).

Beberapa penelitian sebelumnya yang mengangkat topik serupa dalam mengukur tingkat akurasi masing-masing metode *data mining*. Penelitian Etriyanti (2020) yang memprediksi kelulusan mahasiswa dengan *Algoritma Naïve Bayes* dan *Algoritma C4.5*. Hasil penelitian tersebut menunjukkan akurasi. *Algoritma C4.5* (79,08%) lebih baik dibandingkan dengan *Algoritma Naïve Bayes* (78,46%). Hasil yang sama diperoleh dari penelitian selanjutnya (Susanti et al., 2023) yang membandingkan penerapan *Algoritma Decision Tree C4.5* dan *Naïve Bayes* dalam analisa kelulusan siswa dimana tingkat akurasi dengan *Algoritma C4.5* (80%) lebih baik dibandingkan *Algoritma Naïve Bayes* (75%). Hal yang sama juga terjadi dalam penelitian (Zamsuri dan Muladi, 2023) dengan membandingkan akurasi *Algoritma C4.5* dan *Naïve Bayes* terhadap klasifikasi kelulusan mahasiswa dengan hasil penelitian menunjukkan *Algoritma C4.5* dengan akurasi 86,7%, sedangkan *Naïve Bayes* 76%. Pada penelitian (Rahmayanti, Rusdiana, dan Suratno 2022) dengan membandingkan metode *Algoritma C4.5* dan *Naïve Bayes* untuk memprediksi kelulusan mahasiswa didapatkan hasil bahwa dengan *Algoritma C4.5* (90%) lebih baik daripada dengan *Algoritma Naïve Bayes* (85%). Namun, hasil berbeda menjadi temuan pada penelitian Rovidatul et al. (2023) dengan memprediksi kelulusan mahasiswa dengan membandingkan *Algoritma C4.5* dan *Naïve Bayes* dengan nilai akurasi 78,95% dengan metode *Algoritma C4.5* dan 81,58% dengan metode *Algoritma Naïve Bayes*.

Berdasarkan uraian diatas, penelitian ini bertujuan untuk memprediksi kelulusan siswa jurusan Kimia Analisis SMK Tunas Harapan Pati dengan 2 metode yaitu *Algoritma C4.5* dan *Naïve Bayes*. Data yang digunakan pada penelitian ini berjumlah 284 data hasil uji kompetensi siswa Jurusan Kimia Analisis tahun ajaran 2019/2020 sampai tahun ajaran 2023/2024 yang telah lulus. Secara teoritis, penelitian ini berkontribusi dalam penerapan metode data mining untuk memprediksi kelulusan siswa. Bagi institusi, hasil penelitian ini dapat dimanfaatkan untuk merumuskan strategi yang lebih efektif, terutama dengan memberikan perhatian khusus kepada siswa yang diprediksi berpotensi mendapatkan kesulitan pada saat melaksanakan uji kompetensi.

2 METODE

Tahap pertama melakukan pengumpulan data. Diperoleh data sebanyak 284 data set siswa Jurusan Kimia Analisis tahun ajaran 2019/2020 sampai tahun ajaran 2023/2024 yang telah lulus dengan 10 atribut. Tahap kedua, dilakukan pengolahan awal data untuk memastikan kualitas data sebelum diolah menggunakan metode *Algoritma C4.5* dan *Naïve Bayes*. Setelah pengolahan awal, data set yang digunakan pada proses mining tetap berjumlah 284 data siswa dengan 10 atribut. Tahap ketiga melibatkan proses data mining menggunakan metode *Algoritma C4.5* dan *Naïve Bayes* dengan bantuan tools RapidMiner. Validasi akurasi kedua metode dilakukan menggunakan teknik *K-Fold Cross Validation* dan hasil akurasi dianalisis melalui *Confusion Matrix*. Tahap terakhir adalah membandingkan hasil pengujian antara *Algoritma C4.5* dan *Naïve Bayes* untuk menentukan metode terbaik berdasarkan tingkat akurasi tertinggi.

2.1 Pengumpulan Data

Pengumpulan data dilakukan secara langsung di lapangan yaitu data siswa jurusan Kimia Analisis tahun ajaran 2019/2020 sampai tahun ajaran 2023/2024 yang telah lulus sekolah. Data yang diperoleh sebanyak 284 data set dengan 10 atribut atau variabel. Atribut yang digunakan yaitu NIS, Nama, Jenis Kelamin, Tahun Lulus, Hasil Dasar, Hasil Proksimat, Hasil Instrumen, Hasil Sampler, Hasil Mikro dan Sertifikat yang diterbitkan. Adapun contoh data yang digunakan terlihat pada Tabel 1.

Tabel 1. Contoh Data

NIS	Nama	Jenis Kelamin	Tahun Lulus	DS	AP	AI	SAMP	Mikro	Sertifikat
14370	Ahmad Supriyadi	Laki-laki	2020	K	K	K	K	K	Garuda
14371	Alfitra Hayani	Perempuan	2020	K	K	K	K	K	Garuda
14372	Alkhirani Eka Putri Jannisya K	Perempuan	2020	K	K	K	K	K	Garuda
14373	Annis Nisa'ul Muna	Perempuan	2020	K	K	K	K	K	Garuda
14375	Ardzia Wahyu Kurniawan	Laki-laki	2020	K	K	K	K	K	Garuda
14376	Arnis	Perempuan	2020	K	K	K	K	K	Garuda
14377	Ayu Martha Adhelia	Perempuan	2020	K	K	K	K	K	Garuda

NIS	Nama	Jenis Kelamin	Tahun Lulus	DS	AP	AI	SAMP	Mikro	Sertifikat
14378	Bagus Chusen Hidayat	Laki-laki	2020	K	K	K	K	K	Garuda
14379	Bayu Nurrohman	Laki-laki	2020	K	K	K	K	K	Garuda
14381	Dea Syifa Fitriani	Perempuan	2020	K	K	K	K	K	Garuda

2.2 Pengolahan Data Awal

Pada penelitian ini melakukan prediksi berdasarkan data historis yaitu data siswa yang telah lulus sekolah. Data yang diolah mencakup variabel target berupa sertifikat yang diterima siswa. Tujuan penelitian ini yaitu mengukur akurasi prediksi melalui penerapan dua metode data mining yang digunakan. Penelitian ini memiliki kesamaan dengan penelitian yang dilakukan oleh Etriyanti (2020) dimana data kelulusan siswa digunakan sebagai data set dan diimplementasikan menggunakan metode *Algoritma C4.5* dan *Naïve Bayes* dengan bantuan *tool RapidMiner*.

Pengumpulan data yang diperoleh mendapatkan *record* sebanyak 284 data set siswa yang telah lulus yakni data set siswa tahun ajaran 2019/2020 sampai tahun ajaran 2023/2024 dengan 10 atribut. Mahasiswa dapat memperoleh sertifikat garuda apabila dinyatakan lulus 5 klaster uji kompetensi yang diujikan. Hasil pengumpulan data yang diperoleh menunjukkan bahwa data *record* dan atribut seluruhnya dapat dipakai setelah dilakukan pengolahan data awal yang menandakan bahwa data tersebut baik. Adapun rincian 10 atribut yang belum dilakukan pengolahan data awal seperti pada Tabel 2 berikut:

Tabel 2. Atribut Sebelum Pengolahan Data Awal

No	Nama	Jenis Data
1	Nis	Numerik
2	Nama	Kategori
3	Jenis Kelamin	Kategori
4	Tahun Lulus	Numerik
5	Ds	Kategori
6	Ap	Kategori
7	Ai	Kategori
8	Samp	Kategori
9	Mikro	Kategori
10	Sertifikat	Kategori

Pengolahan data awal penting dilakukan untuk memastikan kualitas data set, sebagaimana yang telah diterapkan dalam beberapa penelitian sebelumnya. Penelitian yang dilakukan (Zainuddin 2019), teknik *preprocessing* digunakan untuk meningkatkan kualitas data melalui beberapa Langkah seperti validasi data untuk menghilangkan *outlier*, *noise*, data kosong dan data tidak konsisten, serta diskretisasi data untuk seleksi atribut kelulusan. Sementara itu, penelitian oleh (Etriyanti 2020) menerapkan teknik seperti pembersihan data, reduksi data serta transformasi data.

Berdasarkan beberapa penelitian diatas, pada penelitian ini pengolahan data awal akan dilakukan untuk mendapatkan data dengan kualitas baik. Pengolahan data awal yang digunakan oleh penulis yaitu transformasi data, digunakan untuk membuat nilai interval yang lebih lebar serta mempunyai kedalaman yang sama. Implementasi dilakukan dengan *tool* RapidMiner dimana pengolahan data awal dilakukan dengan operator *Discretize*.

Pembersihan data tidak dilakukan karena tidak adanya data kosong yang mengakibatkan terbentuknya *missing value* dalam data set. Setelah dilakukan pengolahan data awal, data set yang digunakan ketika proses mining sebanyak 284 data siswa dengan 10 atribut. Rincian atribut yang digunakan pada saat proses mining dapat dilihat pada Tabel 3 berikut:

Tabel 3. Atribut Sebtelah Pengolahan Data Awal

No	Nama	Jenis Data
1	Nis	Numerik
2	Nama	Kategori
3	Jenis Kelamin	Kategori
4	Tahun Lulus	Numerik
5	Ds	Kategori
6	Ap	Kategori
7	Ai	Kategori
8	Samp	Kategori
9	Mikro	Kategori
10	Sertifikat	Kategori

2.3 RapidMiner

Pada penelitian ini *tool* RapidMiner adalah *tool data mining* yang menampilkan visualisasi hasil olahan data. RapidMiner dikembangkan dengan menggunakan metode *open core*. RapidMiner menurut (Nofitri dan Irawati 2019) adalah platform perangkat lunak data yang dikembangkan oleh perusahaan yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*). *Tool* RapidMiner memiliki banyak operator data mining yang meliputi operator *input*, *output*, data *preprocessing* dan lain-lain.

2.4 Naïve Bayes

Dalam penelitian ini, metode Naïve Bayes digunakan untuk algoritma perhitungan untuk masalah prediksi. Teorema Bayes digunakan sebagai metode yang bekerja berdasarkan probabilitas sederhana dengan persamaan sebagai berikut:

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)} \quad (1)$$

Keterangan persamaan:

E = Bukti

H = Hipotesis

P(H|E) = Hipotesis H benar untuk bukti E

P(E|H) = Kemungkinan sebuah bukti E terjadi akan mempengaruhi hipotesis H atau dengan kata lain kemungkinan bahwa bukti E benar untuk hipotesis H

P(H) = Probabilitas awal (priori) hipotesis H terjadi tanpa memamndang bukti apapun

$P(E)$ = Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain
 Aturan dari Bayes yakni berdasarkan bukti-bukti (E) yang diamati, hasil hipotesis (H) bisa diprediksi.

2.5 Algoritma C4.5

Metode *Algoritma C4.5* digunakan penulis sebagai metode untuk menyelesaikan masalah prediksi kelulusan siswa. Menurut (Anam dan Santoso 2018) *Algoritma C4.5* berfungsi untuk membangun sebuah pohon keputusan yang mempresentasikan aturan klasifikasi.

Algoritma ini memiliki dua elemen penting yaitu Entropy dan Gain. Proses membangun pohon keputusan diantaranya adalah:

1. Menentukan atribut yang dijadikan node atau akar
2. Membuat cabang pada setiap nilai
3. Membagi kasus sesuai cabang yang dihasilkan
4. Ulangi proses pada setiap cabang hingga semua kasus dalam cabang berada di kelas yang sama

Pemilihan atribut untuk node atau akar dilakukan berdasarkan nilai Grain tertinggi dari keseluruhan atribut. Rumus untuk menghitung Grain yakni:

$$Garint(S,A) = Entropy(S) \sum_{i=1}^n \frac{|S_i|}{|S|} \quad (2)$$

Keterangan :

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi S

$|S_i|$ = Jumlah Kasus Pada Partisi Ke-i

$|S|$ = Jumlah Kasus Dalam S

Nilai *entropy* dapat dihitung menggunakan persamaan berikut:

$$Entropy(S) = \sum_{i=1}^n P_i * P_i \quad (3)$$

Keterangan :

S = Himpunan Kasus

A = Fitur

n = Jumlah Partisi S

P_i = Proporsi Dari S_i Terhadap S

2.6 Evaluasi Metode Klasifikasi

Evaluasi metode klasifikasi bertujuan untuk menganalisis dan membandingkan kinerja metode klasifikasi yang digunakan. Menurut (Anam dan Santoso 2018) evaluasi kinerja model klasifikasi dilakukan untuk memahami performa model klasifikasi berdasarkan hasil pengujian yang diterapkan. Dalam penelitian ini, kinerja metode *Naïve Bayes* akan dibandingkan dengan *Algoritma C4.5*. Validasi akurasi model dilakukan dengan metode *K-Fold Cross Validation*, sedangkan hasil akurasi dilihat berdasarkan *Confussion Matrix*.

a. K-Fold Cross Validation

Menurut (Anam dan Santoso, 2018) *K-Fold Cross Validation* adalah teknik validasi untuk mengukur akurasi metode berdasarkan data set. Umumnya, ini menggunakan 10 kali pengulangan ($k=10$) yang bertujuan memperoleh nilai akurasi dengan bias dan variansi yang sangat rendah.

b. Confussion matrix

Confussion matrix digunakan untuk menganalisis kualitas kinerja model klasifikasi dalam mengenali variabel disetiap kelas. (Anam dan Santoso 2018) menjelaskan bahwa *Confussion Matrix* memuat informasi kelas sebenarnya dan kelas prediksi dalam proses klasifikasi. Tabel matriks ini digunakan untuk mempresentasikan hasil evaluasi model klasifikasi. Berikut adalah contoh *Confussion Matrix*:

		Kelas Hasil Prediksi		Jumlah
		Ya	Tidak	
Kelas Aktual	Ya	TP	FN	P
	Tidak	FP	TN	N
Jumlah		P	N	P + N

Gambar 1. Confussion Matrix

Perhitungan nilai akurasi, precission dan reccal dinyatakan dalam persamaan sebagai berikut:

$$Accuracy = \frac{TP+TN}{P+N} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Reccal = \frac{TP}{P} \quad (6)$$

Keterangan :

TP (*true positive*) : Jumlah variabel positif yang dilabeli dengan benar oleh *classsifier*, sebagai contoh variabel dengan label sertifikat = garuda

TN (*True Negative*) : Jumlah variabel negatif yang dilabeli dengan benar oleh *classifier*

FP (*False Positive*) : Jumlah variabel positif yang salah dilabeli oleh *classifier*

P : Jumlah sampel positif

N : Jumlah sampel negatif

3 HASIL DAN PEMBAHASAN

3.1 Analisis Data

Data yang digunakan dalam proses *training* dan *testing* telah melalui tahap pengolahan data awal sehingga siap diolah. Dataset yang terdiri dari 284 data siswa dengan atribut seperti NIS, Nama, Jenis Kelamin, Tahun Lulus, Dasar, Proksimat, Instrument, Sampling, Mikro, Sertifikat. Proses *training* dan *testing* dilakukan dengan tool RapidMiner dengan metode *Naïve Bayes* dan *Algoritma C4.5*. Pengujian menggunakan operator tambahan *Cross Validation* dengan 10 iterasi. Hasil akurasi dari pengujian dianalisis dengan *Confussion Matrix* seperti pada Gambar 2 dan Gambar 3

3.2 Hasil Implementasi Metode Naïve Bayes dan Algoritma C4.5

Dari hasil implementasi metode *Naïve Bayes* menunjukkan nilai akurasi sebesar 99,64% yang dievaluasi oleh *Confussion Matrix*. Hasil detail yang diperoleh dengan proses data mining menggunakan metode *Naïve Bayes* dapat dilihat pada Gambar 2.

Table View
 Plot View

accuracy: 99.64% +/- 1.13% (micro average: 99.65%)

	true GARUDA	true LSP	class precision
pred. GARUDA	239	0	100.00%
pred. LSP	1	44	97.78%
class recall	99.58%	100.00%	

Gambar 2. Nilai Akurasi Metode *Naïve Bayes*

Gambar 2 menampilkan hasil perhitungan data set dengan metode *Naïve Bayes* berdasarkan *Confussion Matrix*, dimana terlihat bahwa *Accuracy* rata-rata sebesar 99,64%, nilai *Recall* sertifikat Garuda sebesar 99,58%, nilai *Recall* sertifikat LSP 100%, nilai *Class Precicion* sertifikat Garuda 100% dan nilai *Class Precicion* sertifikat LSP 97,78%. Dari 284 data set terdapat 239 data sesuai prediksi yaitu “GARUDA” dan 0 yang prediksi “GARUDA” ternyata “LSP”. Dan terdapat 1 data yang prediksi “LSP” ternyata termasuk klasifikasi “GARUDA” dan 44 data yang sesuai prediksi yakni “LSP”

Selanjutnya hasil implementasi metode *Algoritma C4.5* dengan RapidMiner diperoleh nilai *Accuracy* sebesar 99,29% dengan evaluasi *Confussion Matrix*. Gambar 3 memperlihatkan detail hasil yang diperoleh dari proses mining yang dilakukan dengan metode *Algoritma C4.5*

Table View
 Plot View

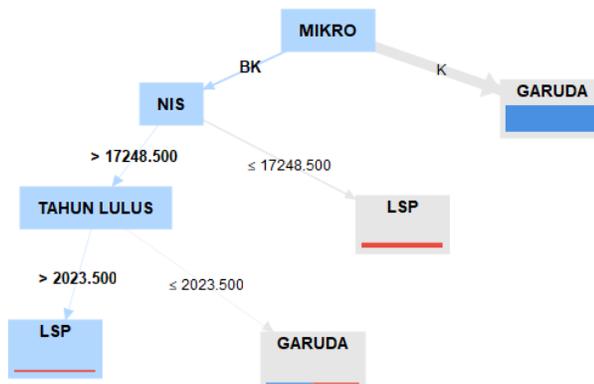
accuracy: 99.29% +/- 1.51% (micro average: 99.30%)

	true GARUDA	true LSP	class precision
pred. GARUDA	239	1	99.58%
pred. LSP	1	43	97.73%
class recall	99.58%	97.73%	

Gambar 3. Nilai Akurasi Metode *Algoritma C4.5*

Gambar 3 menampilkan hasil perhitungan data set dengan metode *Algoritma C4.5* berdasarkan *Confussion Matrix*, dimana terlihat bahwa *Accuracy* rata-rata sebesar 99,29%, nilai *Recall* sertifikat garuda sebesar 99,58%, nilai *Recall* Sertifikat LSP 97,73%, nilai *Class Precicion* Sertifikat Garuda 99,58% dan nilai *Class Precicion* sertifikat LSP 97,73%. Dari 284 data set terdapat 239 data sesuai prediksi yaitu “GARUDA” dan 1 yang prediksi “GARUDA” ternyata “LSP”. Dan terdapat 1 data yang prediksi “LSP” ternyata termasuk klasifikasi “GARUDA” dan 43 data yang sesuai prediksi yakni “LSP”

Dari hasil implementasi *Algoritma C4.5* dengan *tool* RpidMiner terbentuk pohon keputusan berdasarkan nilai *gain* tertinggi sebagai berikut:



Gambar 4. Pohon Keputusan Berdasarkan Informasi Grain

Pada gambar 4 terlihat kriteria yang berpengaruh dalam prediksi sertifikat yang diperoleh siswa kimia analisis adalah ujian klaster MIKRO, NIS, dan tahun lulus. Dari keseluruhan kriteria yang digunakan, kriteria mikro menjadi simpul akar karena nilai *gain* tertinggi terdapat pada kriteria MIKRO. Kriteria MIKRO merupakan klaster ujian terakhir yang dilaksanakan oleh siswa, sehingga dapat menggambarkan bagaimana kemampuan siswa dalam menjalankan keseluruhan proses ujian yang dilaksanakan. Data tersebut merupakan data yang paling dekat untuk menggambarkan data prediksi sertifikat yang diperoleh siswa dibandingkan dengan kriteria yang lain.

Selain MIKRO yang menarik dari hasil pohon keputusan *Algoritma C4.5* adalah NIS yang menjadi salah satu kriteria dominan dalam penelitian ini. Pohon keputusan memperlihatkan bahwa NIS lebih dari 17248.500 diprediksi mendapatkan sertifikat GARUDA jika tahun lulusnya kurang dari 2023.500. sehingga dapat disimpulkan bahwa NIS dan tahun lulus mempengaruhi hasil sertifikat yang diperoleh siswa.

3.3 Perbandingan Hasil Akurasi Metode Naïve Bayes dan Algoritma C4.5

Implementasi yang telah dilakukan menghasilkan perbandingan tingkat akurasi metode *Naïve Bayes* dan *Algoritma C4.5*.

Tabel 4. Perbandingan Nilai Akurasi Metode Naïve Bayes dan Algoritma C4.5

No	Metode	Nilai Akurasi
1	Naïve Bayes	99,64%
2	Algoritma C4.5	99,29%

Berdasarkan table diatas, prediksi kelulusan siswa yang memperoleh sertifikat garuda dengan metode *Naïve Bayes* memiliki nilai akurasi yang lebih tinggi dibandingkan dengan metode *Algoritma C4.5* yaitu 99,64% dengan selisih nilai akurasi dari kedua metode sebesar 0,35%. Hal ini sejalan dengan penelitian yang telah dilakukan oleh (Rovidatul et al., 2023) dimana nilai akurasi metode *Naïve Bayes* lebih besar dari metode *Algoritma C4.5*.

4 KESIMPULAN

Berdasarkan penelitian ini terlihat bahwa prediksi kelulusan siswa Jurusan Kimia Analisis SMK Tunas Harapan Pati dengan data set yang diimplementasikan dengan metode *Naïve Bayes* diperoleh nilai *Accuracy* 99,64% dan *Algoritma C4.5* diperoleh nilai *Accuracy* 99,29%. Karena metode *Naïve Bayes* memiliki nilai akurasi yang lebih tinggi dibandingkan dengan *Algoritma C4.5*. Metode *Naïve Bayes* direkomendasikan untuk digunakan sebagai metode penyelesaian masalah prediksi kelulusan siswa Jurusan Kimia Analisis SMK Tunas Harapan Pati.

UCAPAN TERIMAKASIH

Peneliti mengucapkan terima kasih kepada SMK Tunas Harapan Pati terkhusus Jurusan Kimia Analisis yang telah memberikan dukungan terhadap penelitian ini yang memperbolehkan peneliti menggunakan data internal sekolah, sehingga penelitian ini dapat terlaksana.

DAFTAR PUSTAKA

- Anam, C., & Santoso, H. B. (2018). Perbandingan kinerja algoritma C4.5 dan Naive Bayes untuk klasifikasi penerima beasiswa. *Jurnal Energy*, 8(1).
- Etriyanti, E., Syamsuar, D., & Kunang, Y. N. (2020). Implementasi data mining menggunakan algoritme Naive Bayes Classifier dan C4.5 untuk memprediksi kelulusan mahasiswa. *Telematika*, 13(1). <https://doi.org/10.35671/telematika.v13i1.881>
- Mahar, N. M. A., Atina, V., & Sudibyoy, N. A. (2023). Pemodelan prediksi kelulusan mahasiswa dengan metode Naïve Bayes di UNIBA. *Jurnal Manajemen Informatika dan Sistem Informasi*, 6(2). <https://doi.org/10.36595/misi.v6i2.875>
- Nofitri, R., & Irawati, N. (2019). Analisis data hasil keuntungan menggunakan software RapidMiner. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 5(2). <https://doi.org/10.33330/jurteks.v5i2.365>
- Rahmayanti, A., Rusdiana, L., & Suratno, S. (2022). Perbandingan metode algoritma C4.5 dan Naïve Bayes untuk memprediksi kelulusan mahasiswa. *Walisongo Journal of Information Technology*, 4(1), 11–22. <https://doi.org/10.21580/wjit.2022.4.1.9654>
- Rovidatul, Y., Yunus, Y., & Nurcahyo, G. W. (2023). Perbandingan algoritma C4.5 dan Naïve Bayes dalam prediksi kelulusan mahasiswa. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1). <https://doi.org/10.37859/coscitech.v4i1.4755>
- Setiyani, L., Wahidin, M., Awaludin, D., & Purwani, S. (2020). Analisis prediksi kelulusan mahasiswa tepat waktu menggunakan metode data mining Naïve Bayes: Systematic review. *Faktor Exacta*, 13(1), 35. <https://doi.org/10.30998/faktorexacta.v13i1.5548>
- Susanti, Y., Choyyin, M. G., Priyatna, A., & Lestari, S. (2023). Perbandingan penerapan algoritma Decision Tree C4.5 dan Naïve Bayes dalam analisa kelulusan siswa pada SMK Swadhipa 2 Natar Kabupaten Lampung Selatan. *Jurnal SIMADA (Sistem Informasi dan Manajemen Basis Data)*, 6(2). <https://doi.org/10.30873/simada.v6i2.3772>
- Zainuddin, M. (2019). Perbandingan 4 algoritma berbasis Particle Swarm Optimization (PSO) untuk prediksi kelulusan tepat waktu mahasiswa. *Jurnal Ilmiah Teknologi Informasi Asia*, 13(1).
- Zamsuri, A., & Muladi, D. (2023). Perbandingan akurasi algoritma C4.5 dan Naïve Bayes terhadap klasifikasi kelulusan berdasarkan data mahasiswa Universitas Lancang Kuning. *Prosiding SEMASTER: Seminar Nasional Teknologi Informasi & Ilmu Komputer*, 2(1)

