

OPTIMASI ALGORITMA C4.5 DAN NAIVE BAYES MENGUNAKAN K-MEANS UNTUK PREDIKSI KELULUSAN MAHASISWA

Arif Budiman

Fakultas Ekonomi dan Bisnis, Universitas Muhammadiyah Berau

Jl. DR. Murjani II, Berau, Kalimantan Timur

¹arif_budiman@umberau.ac.id

Abstract

Today's education system demands quality-oriented education. The quality of Indonesian higher education is measured based on accreditation issued by the Badan Akreditasi Nasional Perguruan Tinggi. One indicator of success in the process of managing education on higher education is the period of student graduation. Undergraduate students have a study load of 144 credits which can be taken in 8 semesters. but in fact, there are still many students who cannot complete their studies for 8 semesters due to various factors such as lack of motivation, intelligence factors, and economic factors. There is a need for continuous monitoring and evaluation of periods in student graduation using the C4.5 and Naive Bayes algorithms. Optimization is needed to increase the accuracy value of the C4.5 and Naive Bayes algorithms by using K-means for the data discretization process. The experimental result show C4.5 algorithm with K-means produces an accuracy value of 89.74%, a precision value of 90.60%, and a recall value of 98.00% while Naive Bayes with K-means produces an accuracy value of 80.73%, a precision value of 89.60%, a value recall of 87.20%. The comparison of two classification algorithms combined with K-means shows that the C4.5 algorithm has a better performance than Naive Bayes.

Keywords: C4.5 Alogrithm, C4.5 Discretization, C4.5 Optimazation, Naive Bayes, Student Graduation

Abstrak

Sistem pendidikan masa kini menuntut pendidikan yang berorientasi pada kualitas. Kualitas perguruan tinggi indonesia diukur berdasarkan akreditasi yang dikeluarkan oleh Badan Akreditasi Nasional Perguruan Tinggi. Salah satu indikator keberhasilan dalam proses pengelolaan pendidikan di perguruan tinggi adalah ketepatan kelulusan mahasiswa. Mahasiswa jenjang sarjana memiliki beban studi sebanyak 144 sks yang dapat ditempuh dalam waktu 8 semester. namun kenyatannya masih banyak mahasiswa yang tidak dapat meyelesaikan studinya selama 8 semester karena berbagai faktor seperti kurangnya motivasi, faktor kecerdasan dan faktor ekonomi. Diperlukan adanya pemantauan dan evaluasi yang berkelanjutan terhadap ketepatan dalam kelulusan mahasiswa menggunakan algoritma C4.5 dan Naive bayes. Optimasi diperlukan untuk meningkatkan nilai akurasi algoritma algoritma C4.5 dan Naive bayes dengan menggunakan K-means untuk proses dikritisasi data. Hasil pengujian menunjukan algoritma C4.5 dengan K-means menghasilkan nilai akurasi 89.74%, nilai presisi 90.60%, nilai recall 98.00% sedangkan Naive bayes dengan K-means menghasilkan nilai akurasi 80.73%, nilai presisi 89.60%, nilai recall 87.20%. Hasil komparasi kedua algoritma klasifikasi yang dikominasikan dengan K-means menunjukan algoritma C4.5 memiliki nilai kinerja yang lebih baik dibandingkan Naive bayes.

Kata kunci: Algoritma C4.5, Diskritisasi C4.5, Kelulusan Mahasiswa, Naive Bayes, Optimasi C4.5

1. Pendahuluan

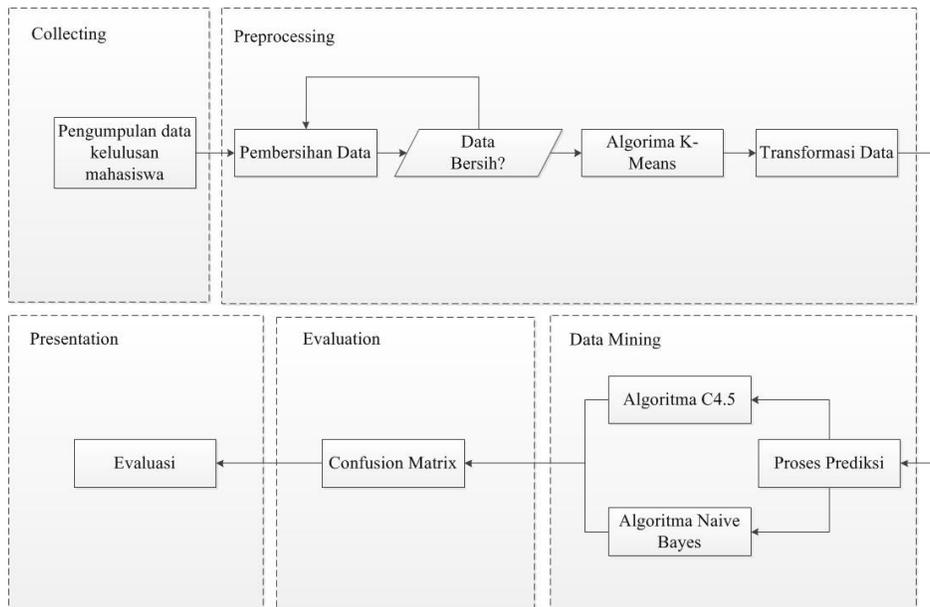
Sistem pendidikan masa kini menuntut pendidikan yang berorientasi pada kualitas [1]. Kualitas perguruan tinggi Indonesia diukur berdasarkan akreditasi yang dikeluarkan oleh Badan Akreditasi Nasional Perguruan Tinggi [2]. Salah satu indikator keberhasilan dalam proses pengelolaan pendidikan di perguruan tinggi adalah ketepatan kelulusan mahasiswa [3]. Mahasiswa jenjang sarjana memiliki beban studi sebanyak 144 sks yang dapat ditempuh dalam waktu 8 semester. Namun kenyataannya masih banyak mahasiswa yang tidak dapat menyelesaikan studinya selama 8 semester karena berbagai faktor seperti kurangnya motivasi, faktor kecerdasan dan faktor ekonomi [4]. Perguruan tinggi saat ini menggunakan database untuk menyimpan data-data yang berkaitan dengan akademik [2], [5]. Semestinya data tersebut dapat dimanfaatkan oleh pengelola perguruan tinggi dalam pembuatan suatu kebijakan, salah satu cara yang dapat diterapkan adalah mengolah data tersebut menggunakan teknik data mining.

Data Mining adalah proses pencarian pengetahuan tersembunyi didalam data yang jumlahnya besar yang tersedia didalam database [3]. Terdapat banyak metode dalam data mining diantaranya seperti Decision Trees, Bayesian, Artificial Neural Networks, Nearest Neighbor, Support Vector Machines dan lainnya [6]. Beberapa algoritma klasifikasi dapat digunakan untuk memprediksi kelulusan mahasiswa antara lain naive bayes [3], [7] dan algoritma C4.5 [8]. Algoritma naive bayes menghasilkan nilai akurasi yang baik [3]. Komparasi algoritma C4.5 dan naive bayes untuk memprediksi hasil seleksi masuk perguruan tinggi menunjukkan naive bayes memiliki akurasi yang lebih baik dibandingkan C4.5 [9]. Untuk meningkatkan nilai akurasi algoritma klasifikasi dapat menerapkan diskritisasi pada tahapan pemrosesan data awal. Teknik diskritisasi dapat digunakan antara lain k-means, binning, dan entropy [10]–[12]. Kombinasi k-means dengan naive bayes dapat meningkatkan nilai akurasi [12].

Berdasarkan permasalahan tersebut diperlukan adanya pemantauan dan evaluasi yang berkelanjutan terhadap ketepatan dalam kelulusan mahasiswa dengan algoritma menggunakan algoritma C4.5 dan Naive Bayes. Optimasi dilakukan menggunakan K-means untuk meningkatkan akurasi pada algoritma C4.5 dan Naive Bayes.

2. Metodologi Penelitian

Adapun tahapan penelitian ketepatan kelulusan mahasiswa dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Pengumpulan data dilakukan untuk mendapatkan data yang berkaitan dengan kelulusan mahasiswa. Data yang diperoleh berasal dari Innovation Center Amikom adalah data mahasiswa jurusan teknik informatika tahun ajaran 2011-2014. Selanjutnya dataset tersebut akan diolah menggunakan Teknik data mining untuk memperoleh informasi yang dapat digunakan untuk memprediksi ketepatan kelulusan mahasiswa.

2.2. Data Mining

Data mining merupakan proses untuk menemukan sebuah pengetahuan dengan mencari pola tersembunyi dalam database [11], [14]. proses tersebut merupakan gabungan dari teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat. Data mining merupakan proses untuk menganalisa data dari berbagai sudut pandang dan merangkumnya menjadi informasi yang berguna[3], [14]. Berikut ini merupakan rangkaian proses dalam data mining untuk menghasilkan sebuah pengetahuan yaitu : [7]

a. Data Collecting

Pengumpulan data yang berkaitan dengan kelulusan mahasiswa

b. Preprocessing Data

Proses pembersihan, penggabungan dan transformasi data ke dalam bentuk yang lebih sesuai sebelum dilakukan proses data mining

c. Data Mining

Menerapkan suatu metode untuk menghasilkan pengetahuan berdasarkan suatu pola-pola tertentu.

d. Pattern Evaluation

Mengevaluasi pola yang ditentukan dapat mewakili pengetahuan yang dihasilkan berdasarkan perhitungan tertentu.

e. Knowledge Presentation

Mempresentasikan pengetahuan yang sudah didapatkan dari pengguna.

2.3. Optimasi Menggunakan K-Means

Optimasi yang dilakukan adalah menggunakan k-means untuk diskritisasi data yang memiliki nilai kontinyu dibuat kedalam bentuk diskrit. k-means merupakan salah satu algoritma klusterisasi non-hirarki yang membagi sejumlah data menjadi beberapa kluster[13]. Metode ini menggolongkan data yang memiliki nilai kedekatan dengan titik pusat kluster (centroid). Tujuan K-Means adalah melakukan pembagian data dengan memaksimalkan kemiripan data dalam satu kluster yang diukur berdasarkan fungsi jarak[15]. proses perhitungan akan dilakukan secara terus-menerus sampai tidak ada perubahan data dalam setiap kluster. Berikut merupakan tahapan proses pengelompokan data menggunakan k-means yaitu :

- Tentukan nilai k sebagai jumlah cluster yang ingin dibentuk.
- Bangkitkan k centroid (titik pusat cluster) awal.
- Hitung jarak setiap data ke masing-masing centroid menggunakan rumus Euclidean Distance seperti pada persamaan 1

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2} \quad (1)$$

- Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
- Tentukan centroid baru dengan cara menghitung nilai rata-rata dari setiap nilai yang ada pada masing-masing centroid menggunakan persamaan 2

$$C_i = \left(\frac{\sum x}{n} \right) \quad (2)$$

Apabila centroid baru dengan centroid lama tidak sama, proses ini akan terus diulang kembali dari langkah 3 hingga menghasilkan centroid konvergen

2.4. Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang digunakan untuk membentuk pohon keputusan[1], [5]. Pada algoritma C4.5 menggunakan Gain Ratio dalam pemilihan atribut. Gain Ratio digunakan untuk mengatasi atribut yang memiliki nilai yang sangat bervariasi dan dihitung berdasarkan Split Information[6]. Kelebihan algoritma C4.5 adalah dapat menangani nilai yang hilang, menangani atribut yang memiliki nilai kontinu serta melakukan pemangkasan pohon keputusan [8]. Berikut merupakan tahapan proses didalam algoritma C4.5 yaitu : [1]

- a. Hitung nilai awal entropi sebagai untuk distribusi sampel menggunakan persamaan 3

$$I(T) = - \sum_{i=1}^k P_i \log_2(P_i) \tag{3}$$

- b. Hitung nilai entropi pada masing-masing atribut menggunakan persamaan 4

$$I(A, T) = \sum_i^m \frac{|T_i|}{|T|} I(T_i) \tag{4}$$

- c. Hitung information gain pada atribut A menggunakan persamaan 5

$$G(A, T) = I(T) - I(A, T) \tag{5}$$

- d. Hitung nilai split information dan rasio information gain untuk atribut A menggunakan persamaan 6

$$Split(A, T) = \sum_i^m \frac{|T|}{|T_i|} \log_2 \frac{|T|}{|T_i|} \tag{6}$$

- e. Hitung gain ratio dengan membagi information gain dengan split information menggunakan persamaan 7

$$G_R(A, T) = \frac{G(A, T)}{Split(A, T)} \tag{7}$$

Nilai gain ratio tertinggi akan digunakan sebagai simpul akar untuk membentuk pohon keputusan

2.5. Naive Bayes

Naive bayes merupakan metode klasifikasi dan statistik yang dikemukakan oleh ilmuwan inggris thomas bayes. Naive bayes dikenal lebih baik dari beberapa algoritma klasifikasi lainnya. kelebihan naive bayes adalah memiliki asumsi bahwa Semua atribut independen tidak saling bergantung yang berdampak pada nilai variabel kelas[10], sederhana dan mudah dibuat dan dapat diterapkan pada data yang berjumlah besar[16]. Metode ini menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari suatu kumpulan data[3], [7]. Algoritma naive bayes menggunakan teorema bayes untuk memprediksi kemungkinan keanggotaan dalam suatu kelas. perhitungan teorema bayes menggunakan persamaan 8

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \tag{8}$$

$P(H|E)$ yang dicari merupakan probabilitas hipotesis H berdasarkan kondisi E (posteriori probabilitas). Dimana $P(E|H)$ merupakan probabilitas E berdasarkan kondisi pada hipotesis H merupakan data dengan class yang belum diketahui; H merupakan data hipotesis suatu class tertentu; $P(H)$ merupakan probabilitas hipotesis H (prior probabilitas) dan $P(E)$ merupakan Probabilitas E [10]

2.6. Evaluasi

Confusion matrix merupakan salah satu metode evaluasi yang digunakan untuk mengukur kinerja algoritma klasifikasi, perhitungan kinerja berdasarkan kesesuaian ketepatan prediksi terhadap keadaan aktual. Berikut adalah tabel confusion matrix ditunjukkan pada tabel 1.

Tabel 1. Confusion Matrix

Nilai Prediksi	Nilai Aktual	
	TP	TN
FP	FN	

Dimana TP (True Positive) adalah data positif yang diprediksi benar, TN (True Negative) adalah data negatif yang diprediksi benar, FP (False Positive) adalah data negatif yang diprediksi sebagai positif, dan FN (False Negative) adalah data negatif yang diprediksi sebagai negatif. Dengan menggunakan tabel confusion matrix dapat dihitung nilai akurasi, presisi dan recall. Akurasi adalah perhitungan nilai berdasarkan kedekatan nilai prediksi dengan nilai sebenarnya menggunakan persamaan 9

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Presisi adalah perhitungan dalam mengukur tingkat keakuratan perhitungan prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif dengan Persamaan 10

$$Presisi = \frac{TP}{TP + FP} \tag{10}$$

Recall digunakan dalam mengukur tingkat keberhasilan sistem dalam mengambil suatu informasi menggunakan persamaan 11

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Dataset mahasiswa yang dikumpulkan berjumlah 1941 record yang disimpan didalam basis data Universitas Amikom Yogyakarta, adapun data yang dikumpulkan terdiri dari faktor internal antara lain Nomor Induk, Jenis Kelamin, Asal Daerah, Angkatan, Tahun Lulus, Jumlah SKS. IPK Semester 1-8, Presensi Semester 1-4, Matakuliah Konsentrasi, Nilai Ujian Nasional, Nilai TOEFL, Asal SLTA, Jumlah Mata kuliah Mengulang dan Jumlah Mata kuliah Remedi dan faktor eksternal mahasiswa antara lain Pekerjaan Ayah, Pekerjaan Ibu, Penghasilan Ayah, Penghasilan Ibu, Beasiswa.

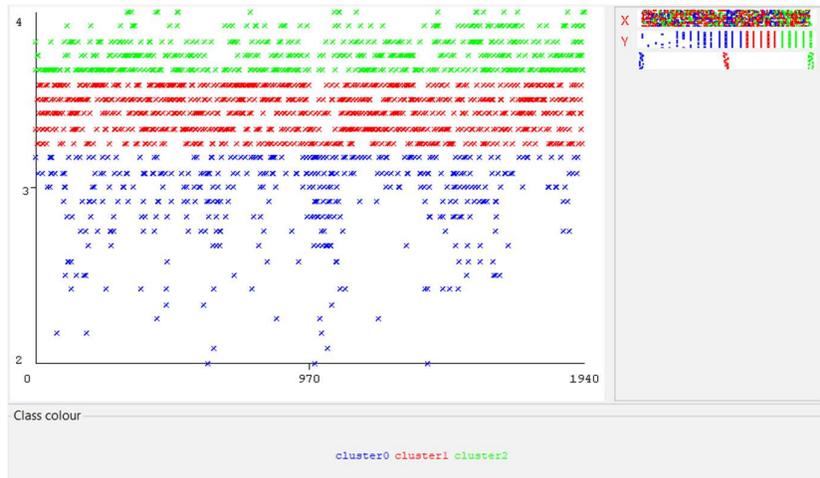
3.2. Preprocessing Data

Data yang telah dikumpulkan belum bisa digunakan untuk prediksi karena terdapat atribut yang memiliki nilai kosong dan berganda, oleh karena itu perlu dilakukan pemrosesan data awal untuk menjaga konsistensi data. Pembersihan data dilakukan dengan cara mengisi data pada atribut yang memiliki nilai kosong dengan rata-rata jumlah data dari keseluruhan data untuk data bertipe numerikal sedangkan data bertipe string diisi dengan nilai yang paling sering ada pada masing-masing atribut. Dataset mahasiswa yang telah melalui tahap pembersihan data ditunjukkan pada tabel 1.

Tabel 1. Dataset Mahasiswa

IPS_1	IPS_2	IPS_3	IPS_4	Lulus
3.17	3.08	3.17	3.15	2017
3.67	3	2.23	2.23	2016
3.83	3.83	3.86	3.83	2017
3.33	3.38	3.32	3.26	2015
3.5	3.5	3.58	3.67	2015
3.5	3.42	3.44	3.35	2015
3.25	3.29	3.28	3.17	2016
3.75	3.67	3.58	3.63	2015

Setelah diperoleh data yang konsisten kemudian dilakukan diskritisasi data. Diskritisasi bertujuan untuk merubah data yang memiliki nilai kontinu menjadi diskrit dengan menggunakan algoritma K-Means. proses ini dilakukan pada atribut Ipk Sem 1, Ipk Sem 2, Ipk Sem 3, Ipk Sem 4, total sks, presensi sem 1, presensi sem 2, presensi sem 3, presensi sem 4, makul mengulang dan makul remedi. Jumlah k yang digunakan untuk proses pengelompokan sebanyak k=3. Berikut adalah hasil pengelompokan data menggunakan k-means ditunjukkan pada gambar 2.



Gambar 2. Distribusi Kmeans atribut IPS_1

Setelah dilakukan pengelompokan data kemudian dilakukan transformasi bentuk dataset mahasiswa yang nantinya akan digunakan untuk proses prediksi. Pengelompokan data ditunjukkan pada tabel 2.

Tabel 2. Pengelompokan data

Atribut	Nilai Atribut	Kategori
IPS_1	2,00-3,17	Klaster0
	3,25-3,58	Klaster1
	3,67-4,00	Klaster2
IPS_2	1,63-3,08	Klaster0
	3,09-3,50	Klaster1
	3,52-4,00	Klaster2
IPS_3	1,53-2,95	Klaster0
	2,97-3,44	Klaster1
	3,45-4,00	Klaster2
IPS_4	1,27-2,87	Klaster0
	2,88-3,42	Klaster1
	3,43-4,00	Klaster2
....
Kelas	1-8	Tw
	>8	Te

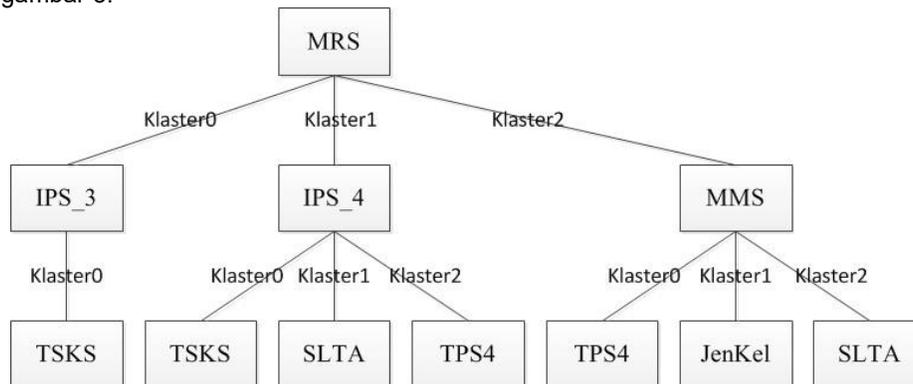
Data mahasiswa dibagi menjadi 2 kelas yaitu tepat waktu dan terlambat. Kelas tepat waktu merupakan mahasiswa yang mampu menyelesaikan studi selama 8 semester sedangkan kelas terlambat merupakan mahasiswa yang menyelesaikan studi lebih dari 8 semester. Dataset mahasiswa setelah proses transformasi ditunjukkan pada tabel 3.

Tabel 3. Dataset Mahasiswa setelah Tranformasi

IPS_1	IPS_2	IPS_3	IPS_4	Kelas
klaster0	klaster0	klaster1	klaster1	Tepat Waktu
klaster2	klaster0	klaster0	klaster0	Terlambat
klaster2	klaster2	klaster2	klaster2	Tepat Waktu
klaster1	klaster1	klaster1	klaster1	Terlambat
klaster1	klaster1	klaster2	klaster2	Terlambat
klaster1	klaster1	klaster1	klaster1	Terlambat
klaster1	klaster1	klaster1	klaster1	Tepat Waktu
klaster3	klaster2	klaster2	klaster2	Terlambat

3.3. Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma klasifikasi[6], namun permodelan data menggunakan algoritma C4.5 dapat digunakan untuk proses prediksi[5], [11]. Pada algoritma C4.5 akan membentuk sebuah pohon keputusan, berdasarkan pohon keputusan yang terbentuk berisi set aturan yang dapat digunakan untuk memprediksi ketepatan kelulusan mahasiswa berdasarkan dua kelas yaitu tepat waktu dan terlambat. Berdasarkan hasil perhitungan nilai gain ratio diketahui bahwa atribut MRS memiliki nilai gain ratio paling tinggi sehingga digunakan sebagai akar dalam membentuk pohon keputusan. Pohon keputusan yang terbentuk ditampilkan pada gambar 3.



Gambar 3. Pohon Keputusan

Berdasarkan pohon keputusan yang terbentuk kemudian diperoleh aturan untuk memprediksi ketepatan kelulusan mahasiswa.

3.4. Naive Bayes

Set aturan yang terbentuk pada naive bayes diperoleh berdasarkan perhitungan nilai probabilitas pada masing-masing atribut pada keanggotan kelas. Berikut adalah tabel perhitungan Naive bayes ditunjukkan pada tabel 4.

Tabel 4. Probabilitas Naive Bayes

Atribut	Nilai	Tepat Waktu (0.84)	Terlambat (0.16)
IPS_1	1	0.142	0.307
	2	0.426	0.430
	3	0.431	0.262
IPS_2	1	0.071	0.275
	2	0.433	0.506
	3	0.495	0.223
IPS_3	1	0.070	0.278
	2	0.420	0.504
	3	0.508	0.216
IPS_4	1	0.058	0.294
	2	0.433	0.508
	3	0.507	0.193
....

3.5. Evaluasi Model

Evaluasi dilakukan untuk mengukur kinerja algoritma C4.5 dan Naive Bayes menggunakan confusion matrix. Berikut adalah tabel confusion matrix algoritma C4.5 terbentuk ditunjukkan pada tabel 5.

Tabel 5. Confusion Martix C4.5

Nilai Prediksi	Nilai Aktual	
		1594
	38	144

Dengan menggunakan algoritma C4.5 diperoleh perbandingan kesesuaian nilai prediksi dan nilai aktul dalam memprediksi ketepatan kelulusan mahasiswa. Berikut adalah tabel confusion matrix naive bayes ditunjukkan pada tabel 6.

Tabel 6. Confusion Matrix Naive Bayes

Nilai Prediksi	Nilai Aktual	
		1420
	212	150

Berdasarkan tabel confusion matrix kemudia dilakukan perhitungan kinerja algoritma klasifikasi berdasarkan akurasi, presisi dan recall. tabel kinerja antara algoritma C4.5 dan naive bayes ditunjukkan pada tabel 7.

Tabel 7. Komparasi C4.5 dan Naive bayes

Algoritma	Akurasi	Presisi	Recall
C4.5 + K-Means	89.54%	90.60%	97.70%
Naive Bayes + K-Means	80.88%	89.90%	87.00%

4. Kesimpulan

Optimasi algoritma C4.5 dan Naive bayes menggunakan K-Means dapat digunakan untuk memprediksi ketepatan kelulusan mahasiswa. Pengujian menunjukan algoritma C4.5 menghasilkan nilai akurasi sebesar 89.74%, nilai presisi sebesar 90.60%, nilai recall sebesar 98.00% dan Naive bayes menghasilkan nilai akurasi sebesar 80.73%, nilai presisi sebesar 89.60%, nilai recall sebesar 87.20%. Hasil komparasi kedua algoritma klasifikasi tersebut menunjukan algoritma C4.5 yang dikombinasikan dengan K-means memiliki nilai kinerja yang lebih baik dibandingkan Naive bayes yang dikombinasikan menggunakan K-means.

DAFTAR PUSTAKA

- [1] X. Wang, C. Zhou, and X. Xu, "Application of C4.5 decision tree for scholarship evaluations," *Procedia Comput. Sci.*, vol. 151, pp. 179–184, 2019, doi: 10.1016/j.procs.2019.04.027.
- [2] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, p. 35, Jun. 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [3] I. B. A. Peling, I. N. Arnawan, I. P. A. Arthawan, and I. G. N. Janardana, "Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm," *Int. J. Eng. Emerg. Technol.*, vol. 2, no. 1, p. 53, Sep. 2017, doi: 10.24843/IJEET.2017.v02.i01.p11.
- [4] N. N. Sari, M. Salam, and H. Usmanto, "Analisis Faktor-Faktor Penyebab Keterlambatan Mahasiswa Dalam Menyelesaikan Tugas Akhir (Skripsi) Pada Mahasiswa," vol. 1, no. 1, p. 13, 2021.
- [5] R. P. S. Putri and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Inform. J. Ilmu Komput. Dan Inform.*, vol. 4, no. 1, p. 1, Jul. 2018, doi: 10.23917/khif.v4i1.5975.
- [6] Suyanto, *Data mining : untuk klasifikasi dan klasterisasi data*. Bandung: Informatika, 2017.
- [7] M. R. Qisthiano, T. B. Kurniawan, E. S. Negara, and M. Akbar, "Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 3, p. 987, Jul. 2021, doi: 10.30865/mib.v5i3.3030.
- [8] L. Y. Lumban Gaol, M. Safii, and D. Suhendro, "Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5," *Brahmana J. Penerapan Kecerdasan Buatan*, vol. 2, no. 2, pp. 97–106, Jun. 2021, doi: 10.30645/brahmana.v2i2.71.
- [9] I. Sulihati, "Penerapan Komparasi Algoritma C4.5 dan Naïve Bayes untuk Menentukan Hasil Seleksi Masuk Perguruan Tinggi," *J. TECNOSCENZA*, vol. 6, no. 2, pp. 311–320, Apr. 2022, doi: 10.51158/tecnoscienza.v6i2.694.
- [10] "Penerapan Naive Bayes Pada Detection Malware dengan Diskritisasi Variabel," *Telematika*, vol. 13, no. 1, pp. 11–21, Feb. 2020, doi: 10.35671/telematika.v13i1.886.
- [11] A. Budiman, "Cronic Kidney Disease Prediction Using C4.5 Algorithm and K-Means," vol. 1, no. 1, p. 7, 2020.
- [12] H. Rahmawan, "Penentuan Rekomendasi Pelatihan Pengembangan Diri Bagi Pegawai Negeri Sipil Menggunakan Algoritma C4.5 Dengan Principal Component Analysis Dan Diskritisasi," *J. Tekno Kompak*, vol. 14, no. 1, p. 5, Feb. 2020, doi: 10.33365/jtk.v14i1.531.
- [13] D. A. Effendy, "Algoritma K-Means untuk Diskretisasi Numerik Kontinyu Pada Klasifikasi Intrusion Detection System Menggunakan Naive Bayes," p. 6.
- [14] D. Supriyadi and S. T. Safitri, "The Application of C4.5 Algorithm to Classify the User Satisfaction of Online Learning System," vol. 3, no. 36, p. 9, 2020.
- [15] S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma JURSI TGD*, vol. 1, no. 1, p. 31, Jan. 2022, doi: 10.53513/jursi.v1i1.4784.
- [16] N. Salmi and Z. Rustam, "Naïve Bayes Classifier Models for Predicting the Colon Cancer," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, p. 052068, Jun. 2019, doi: 10.1088/1757-899X/546/5/052068.