

Prediksi Predikat Kelulusan Mahasiswa Menggunakan Naive Bayes dan Decision Tree pada Universitas XYZ

Agung Wibowo¹, Abdul Rohman²

Jurusan Teknik Informatika, Fakultas Komputer dan Pendidikan

^{1,2}Universitas Ngudi Waluyo

Semarang, Indonesia

¹agungwibowo@unw.ac.id, ²abdulrohman15@gmail.com

ABSTRACT – *The success of universities in managing learning standards can be seen from the number of students who can graduate with high predicates and on time. The graduation predicate of a student is influenced by several factors. This study aims to find out the profile of students who graduated with a set predicate and what are the factors that influence it. The method used in this study is the Cross Industry Standard Process for Data Mining CRISP-DM by utilizing the Naïve Bayes algoritma in looking for graduate predicate patterns and Decision Tree in looking for causative factors. In the calculations using the NBC algorithm it was found that the profiles of students who passed the predicate were less satisfactory, satisfactory, very satisfactory and with praise. In the Decision Tree calculation, the highest gain value is obtained at the attributes of IPK4, IPS5 and IPK5. The factors that most influence graduation are the cumulative achievement index in semesters 4 and 5 and the achievement index in semester 5. The pattern of graduating with predicate can be known from the second year of the incoming student to the third year. This research needs to be developed again by increasing the number of attributes and data, and it is necessary to create a system for determining student graduation predicates from the patterns that have been produced in order to help universities to improve the quality of student graduation in each period.*

Keywords: *Data Mining, Decision Tree, Graduation, Naive Bayes.*

ABSTRAK -Keberhasilan perguruan tinggi dalam mengelola standar pembelajaran dapat diketahui dari banyaknya peserta didik yang bisa lulus dengan predikat tinggi dan tepat waktu. Predikat kelulusan dari mahasiswa dipengaruhi oleh beberapa faktor. Penelitian ini bertujuan untuk mengetahui profil mahasiswa yang lulus dengan predikat yang ditetapkan dan apa saja faktor-faktor yang mempengaruhinya. Metode yang digunakan pada penelitian ini adalah Cross Industry Standar Process for Data Mining CRISP-DM dengan memanfaatkan algoritma Naive Bayes dalam mencari pola predikat lulusan dan Decision Tree dalam mencari faktor-faktor penyebabnya. Dari hasil perhitungan menggunakan algoritma NBC pada penelitian ini, ditemukan profil mahasiswa yang lulus dengan predikat kurang memuaskan, predikat memuaskan, predikat sangat memuaskan dan predikat dengan pujian. Hasil perhitungan Decision Tree, nilai gain tertinggi didapatkan pada atribut IPK4, IPS5 dan IPK5. Dapat disimpulkan pada penelitian ini bahwa faktor-faktor yang paling memengaruhi kelulusan adalah indeks prestasi kumulatif pada semester 4 dan 5 serta indeks prestasi pada semester 5. Pola lulus dengan predikat yang telah ditentukan dapat diketahui dari tahun kedua mahasiswa masuk hingga tahun ketiga. Dari pola tersebut program studi dapat berupaya menyiapkan mahasiswa supaya dapat lulus dengan predikat kelulusan yang ditetapkan. Penelitian ini perlu dikembangkan lagi dengan peningkatan jumlah atribut dan data, dan perlu dibuat sebuah sistem penentu predikat kelulusan mahasiswa dari pola yang telah dihasilkan supaya dapat membantu perguruan tinggi untuk meningkatkan kualitas kelulusan mahasiswa di setiap periodenya

Kata Kunci: *Data Mining, Decision Tree, Kelulusan, Naive Bayes.*

1. PENDAHULUAN

Data akademik Mahasiswa pada perguruan tinggi yang dikelola secara strategis dapat memberikan informasi yang detail dan memiliki analisis yang baik dalam mendapatkan informasi tentang prestasi akademik mahasiswa secara khusus dan keberhasilan studi mahasiswa pada capaian tingkatan kelulusan setiap

mahasiswa. Pengelolaan data akademik setiap perguruan tinggi dilakukan oleh bagian akademik dibawah koordinasi program studi.

Universitas XYZ merupakan perguruan tinggi swasta yang sedang berkembang di daerah Semarang. Sebagai lembaga pendidikan tinggi ternama di Indonesia,



Universitas XYZ harus meningkatkan kualitas pendidikannya agar diharapkan dapat memiliki lulusan berdaya saing, berinovasi dan bisa beradaptasi dengan perkembangan di era globalisasi saat ini. Di dalam menghasilkan lulusan yang berkualitas, dipengaruhi oleh keberhasilan studi dari mahasiswa itu sendiri. Bentuk analisa dari data keberhasilan studi/kelulusan mahasiswa berupa predikat kelulusan yang diprediksikan dalam kriteria “Predikat Memuaskan”, “Predikat Sangat Memuaskan”, dan “ Predikat Dengan Pujian”. Predikat kelulusan saat ini dijadikan sebuah pedoman oleh lulusan

dalam mendapatkan rekomendasi pekerjaan. membutuhkan waktu yang lama

dan tidak mudah tersebut Dalam analisis dan pencarian informasi predikat kelulusan yang dilakukan secara konvensional akan mengingat data yang disajikan sangat banyak dan terdiri dari banyak variable. Analisis data akademik dengan menggunakan data mining dapat dilakukan dengan berbagai pemodelan.

Beberapa penelitian sebelumnya yang melakukan penelitian terhadap predikat kelulusan mahasiswa yang ditunjukkan pada table 1.

Tabel 1. Penelitian Sebelumnya

No.	Judul	Algoritma	Hasil
1	Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan Algoritma C45 di STIMIK Rosma di Karawang	C.45	nilai accuracy sebesar 85,83%, nilai recall sebesar 92,52%, nilai precision sebesar 90,83%, dan memiliki nilai AUC sebesar 0,79
2	Sistem Prediksi Masa Studi Dan Predikat Kelulusan Mahasiswa Menggunakan model Decision Tree Dan NBC	C.45 dan NBC	perpaduan dari algoritma pohon keputusan dan NBC menghasilkan akurasi sebesar 93.02% untuk lama studi dan total akurasi sebesar 76.74% untuk predikat akhir.
3	Algoritma J48 Dan Logistic Model Tree Untuk Memprediksi Predikat Kelulusan Mahasiswa: Studi Kasus STT XYZ	J48 dan Logistic Model Tree	Hasil pengukuran pada penelitian menggunakan algoritma J48 menghasilkan akurasi sebesar 75,5%, dan pada saat menggunakan LMT akurasi didapatkan adalah 73,5%.

Pencapaian predikat kelulusan mahasiswa dapat ditentukan dengan cara berikut, yaitu membuat kelompok data dari Jumlah total sks yang diambil oleh mahasiswa, indeks kumulatif dari mahasiswa, dan dapat berupa keterangan kelulusan tugas akhir mahasiswa [1]. Predikat saat mahasiswa lulus dapat disebabkan banyak faktor, beberapa diantaranya yang mempengaruhi adalah faktor dari data akademik mahasiswa tersebut, yang berisikan biodata mahasiswa dan capaian akademik yang didapatkan oleh mahasiswa. Berdasarkan penelitian sebelumnya dan beberapa faktor yang dapat berpengaruh pada predikat kelulusan, maka pada penelitian ini dilakukan analisis prediksi predikat indeks prestasi kelulusan mahasiswa dengan penggunaan model decision tree dan naive bayes yang pada akhirnya dapat dimanfaatkan oleh perguruan tinggi dalam mengatasi persoalan yang ada, sehingga perguruan tinggi dapat berkompetisi dengan perguruan tinggi lainnya dan dapat meningkatkan kualitas akademik dari mahasiswa pada perguruan tinggi tersebut.

2. DASAR TEORI

A. Prediksi

Peramalan atau sering disebut juga prediksi merupakan proses dalam peramalan secara sistematis yang kemungkinan dapat terjadi di waktu masa yang akan datang berdasarkan informasi di saat masa lalu dan pada waktu sekarang yang dimiliki dan meminimalisirkan ketidaksesuaian (perbedaan antara apa terjadi sebenarnya

dan hasil yang diramalkan). Prediksi belum bisa dipastikan dapat memberikan jawaban pasti atas peristiwa yang terjadi, tetapi tetap mengusahakan mencari jawaban yang paling dekat dengan fenomena yang saat ini terjadi [2]. Berikut adalah penjelasan dari atribut yang digunakan untuk perhitungan :

1) Atribut Gender

Jenis kelamin laki-laki (L) dan perempuan (F) digunakan untuk menentukan persentase penyelesaian. Menurut [3], karakteristik gender merupakan variabel yang digunakan dalam penentuan tingkat kelulusan.

2) Atribut tempat dan tanggal lahir

Tempat lahir mahasiswa dipakai dalam menentukan predikat lulus pada mahasiswa. Tempat lahir merupakan asal daerah mahasiswa yang berasal dari daerah jawa, ntb, kalimantan, bali dan provinsi lain di Indonesia.

Predikat kelulusan merupakan capaian yang dihasilkan oleh mahasiswa dalam menyelesaikan studi/perkuliahannya, Capaian tersebut dikelompokkan menjadi tiga kelompok yaitu lulus dengan hasil yang memuaskan, lulus dengan hasil sangat memuaskan dan lulus dengan pujian (cumlaude). Menurut [1] penentuan predikat mahasiswa dilakukan dengan mengelompokkan data yang diterima berdasarkan Jumlah beban studi yang diselesaikan, indeks prestasi keseluruhan mahasiswa, dan data skripsi. Mahasiswa akan memperoleh predikat kelulusan jika memenuhi persyaratan sebagai berikut:

- a. Nilai rata-rata kumulatif yang mewakili setidaknya skor minimum yang telah dibuat ketetapanannya oleh universitas..
- b. Total dari beban studi yang telah ditentukan.
- c. Dinyatakan lulus pada ujian sidang skripsi

Persyaratan tersebut dapat berbeda-beda untuk setiap perguruan tinggi. Mahasiswa dikatakan lulus jika mahasiswa telah memenuhi segala syarat dari perguruan tinggi asal.

B. Data Mining

Data mining merupakan bidang multidisiplin ilmu yang menggabungkan pola, statistik, *basis data* dan visualisasi dalam memecahkan masalah untuk mengekstraksi informasi dari basis data yang besar.[4]. Menurut [5] Data Mining adalah proses penggunaan teknik statistik, matematika, pembelajaran mesin, dan kecerdasan buatan untuk mengekstraksi dan proses identifikasi data dan informasi yang berguna terkait dengan basis data yang besar. Data mining identik dengan penemuan pengetahuan dalam basis data (KDD). [6] KDD bertujuan sebagai fungsi dari pengumpulan dan penggunaan data, historis dalam penemuan sebuah keadaan yang teratur, pola atau hubungan didalam satu paket data yang ukurannya besar [7]. Arti lain tentang data mining oleh [8] adalah proses dalam pencarian relasi dan proses penemuan sebuah data yang penggunaanya tidak mengetahui darimana data itu, yang kemudian dibuat dengan cara yang mudah dimengerti, sehingga relasi data dapat menjadi pedoman dalam sebuah sistem pendukung pengambilan keputusan.

C. Klasifikasi

Pengelompokan data merupakan model dalam klasifikasi yang mempelajari data latih menggunakan algoritma klasifikasi [9]. Algoritma Pengelompokan data diantaranya Decision Tree, NBC, SVM dan KNN [10] .

D. Decision Tree

Algoritma Decision Tree adalah sebuah model klasifikasi yang menggunakan model pohon keputusan [11].. Untuk menentukan atribut yang menjadi akar pada model pohon keputusan ini dengan melihat nilai gain yang paling besar dari atribut lain yang ada. Perhitungan entropy dan gain diperoleh dengan persamaan(1) :

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i \log_2 p_i \dots\dots(1)$$

Dimana pi adalah jumlah data class dibagi total data.

E. Naive Bayes (NBC)

Naive Bayes adalah teknik peramalan dengan dasar probabilitas sederhana yang didasarkan dari penerapan teorema bayesian dengan perkiraan independensi kuat. Manfaat NBC yang lain, dapat menganalisis variabel yang paling mendukung dalam hal peluang [9].

Rumus teori bayes adalah pada persamaan (2)

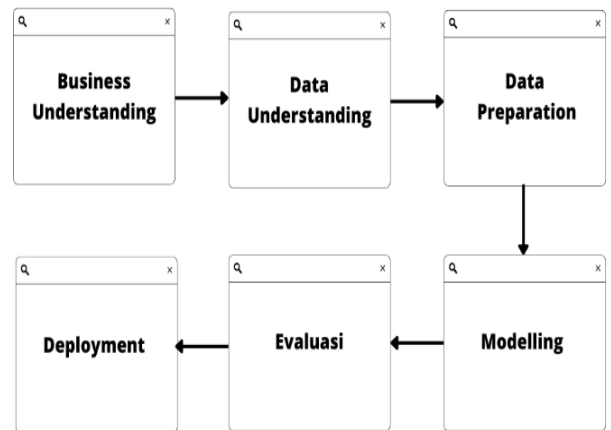
$$: P(H|X) = \frac{P(X|H)P(H)}{P(X)} \dots\dots\dots(2)$$

X merupakan data yang *class*-nya tidak diketahui, H adalah hipotesa, data X adalah class yang spesifik, P(H | X)

adalah probabilitas hipotesis H berdasar kondisi X (posteriori probability), P(H) yaitu hipotesa probabilitas H (prior probability), P(X|H) adalah probabilitas X berdasar kondisi pada hipotesa H, P(X) adalah probabilitas X.

3. METODOLOGI

Dalam penelitian ini, kami memecahkan masalah menggunakan pendekatan kuantitatif dengan pendekatan eksperimental. Prosedur pengujian ini disesuaikan dengan langkah-langkah dari metode CRISP-DM. Metode tersebut memanfaatkan model proses pengembangan data seringkali digunakan oleh para ahli dalam membantu memecahkan masalah. Proses penelitian ini berkaitan dengan enam tahapan dari metode CRISP-DM yang digambarkan pada gambar 1. Tahapan dalam metode tersebut dimulai dari fase pemahaman bisnis, dilanjutkan fase pemahaman data, berikutnya dalam fase persiapan data, kemudian fase pemodelan, fase evaluasi, dan fase ke enam adalah fase penyebaran [11].



Gambar 1. Tahap CRISP-DM [11]

A. Business Understanding (Fase Pemahaman Bisnis)

Selama ini data yang disimpan dalam bentuk data predikat kelulusan belum dimanfaatkan untuk mengetahui tingkat keberhasilan mahasiswa. Data yang digunakan dalam penelitian ini berasal dari data akademik dari universitas, yang diproses sehingga menghasilkan informasi nilai keberhasilan studi mahasiswa yang dikelompokkan dan dilakukan prediksi untuk mengetahui predikat kelulusannya.

B. Data Understanding (Pemahaman Data)

Penggunaan data dalam penelitian ini dibagi 2 kategori yang terdiri dari *training* data dan *testing* data.. Training data yang digunakan dari data lulusan angkatan 2017 dan angkatan 2018 yang lulus di tahun 2021 dan 2022 Kedua data ini mempunyai jumlah dan jenis atribut yang sama dengan rincian yang tertera pada tabel 2.

Tabel 2. Jenis Atribut Training dan Testing Data

Jenis	Keterangan
Nomor Induk Mahasiswa	Nomor Registrasi Mahasiswa terdapat informasi tahun masuk yang berguna untuk menghitung lama studi dari awal masuk
Nama	Nama Lengkap Mahasiswa
Jenis Kelamin	Laki-laki/Perempuan untuk mengidentifikasi bahwa lulusan yang paling banyak lulus tepat waktu itu L atau P
Tempat Tanggal Lahir	Tempat dan tanggal mahasiswa lahir sesuai Akte kelahiran
Pekerjaan Orangtua/wali	Pekerjaan orangtua/wali mahasiswa
IPS 4, 5, 6	Indeks Prestasi Semester ke 4, 5, 6
IPK 4, 5, 6	Indeks Prestasi Kumulatif semester 4, 5, 6

C. Data Preparation (Pengolahan Data)

Pada tahap pengolahan data, data pelatihan berasal dari pengolahan data alumni dan data pengujian berasal dari pengolahan data mahasiswa semester 4 atau level 2. Di bawah ini adalah tahap preprocessing data mining.

- 1) Cleaning data : Penghapusan data tidak lengkap dan tidak sesuai yang nantinya tidak mempengaruhi akurasi perkiraan.
- 2) Integrasi Data : Penggabungan data yang dibutuhkan. Atribut target lulus merupakan penggabungan dari
- 3) ahun masuk mahasiswa hingga tahun lulus mahasiswa

- 4) Transformasi Data: proses normalisasi data yang dibutuhkan
- 5) Reduksi dan Diskretisasi Data: proses penguraian data menjadi bagian yang lebih spesifik supaya data yang dipilih merupakan data yang informatif.

Atribut yang telah melalui fase pengolahan data ditunjukkan pada tabel 3, dan pada tabel 4 merupakan data training yang digunakan setelah melalui fase pengolahan data.

Tabel 3. Atribut setelah melalui fase pengolahan data.

Jenis Atribut	Keterangan
Jenis Kelamin	Laki-laki/Perempuan untuk mengidentifikasi bahwa lulusan yang paling banyak lulus tepat waktu itu L atau P
Tempat	Asal daerah mahasiswa
Usia	Usia mahasiswa ketika mendaftar
Pekerjaan Orangtua/wali	Pekerjaan orangtua/wali mahasiswa
IPS 4, 5, 6	Indeks Prestasi Semester ke 4, 5, 6
IPK 4, 5, 6	Indeks Prestasi Kumulatif semester 4, 5, 6
Predikat Keulusan	Lulus A = lulus dengan indeks kurang memuaskan, Lulus B = Lulus dengan indeks memuaskan, Lulus C = Lulus dengan indeks prestasi sangat memuaskan, berikutnya Lulus D = Lulus dengan indeks tertinggi dengan pujian.

Tabel 4. Data Training setelah melalui fase pengolahan data

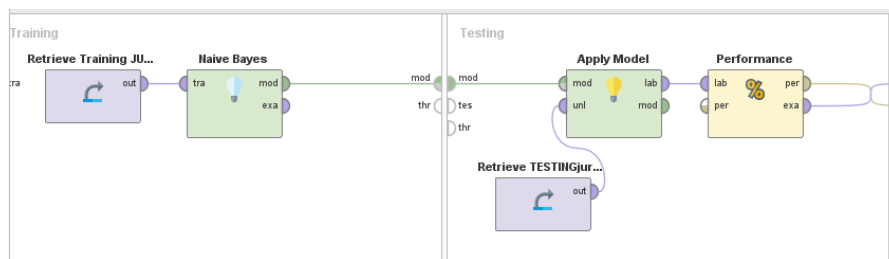
Gender	Ips4	Ips5	Ips6	Ipk4	Ipk5	Ipk6	Asal	Usia Masuk	Pekerjaan Ortu	Predikat Lulus
L	2,89	2,21	2,50	2,55	2,53	1,83	Ntb	19	Wiraswasta	Lulus B
L	2,74	2,84	3,00	2,97	2,86	2,40	Jawa	18	Wiraswasta	Lulus C
P	3,42	3,05	2,70	3,29	3,05	2,26	Jawa	18	Swasta	Lulus C
P	3,05	3,42	3,40	3,29	3,21	2,51	Ntb	18	Pns	Lulus C
P	2,11	1,89	1,25	2,00	1,74	1,89	Ntb	18	Wiraswasta	Lulus C
P	3,11	2,74	2,45	2,92	2,76	2,15	Ntb	18	Pns	Lulus C
P	3,16	3,05	2,85	3,11	2,97	2,23	Jawa	19	Wiraswasta	Lulus C
P	2,89	2,95	3,05	2,92	3,00	2,26	Jawa	18	Wiraswasta	Lulus C
P	2,89	3,16	3,15	3,24	3,07	2,33	Jawa	18	Wiraswasta	Lulus C
P	2,95	3,00	2,85	3,03	2,93	2,15	Jawa	19	Swasta	Lulus C
P	3,05	3,21	3,03	3,13	3,03	2,41	Jawa	18	Swasta	Lulus C

P	3,63	3,74	3,40	3,68	3,55	2,60	Ntb	17	Tani	Lulus D
P	3,11	2,79	2,60	2,95	2,83	2,22	Kalimantan	18	Swasta	Lulus C
P	2,63	2,42	2,25	2,53	2,43	1,83	Jawa	18	Swasta	Lulus C
P	2,74	3,53	2,85	3,39	2,84	2,28	Jawa	18	Wiraswasta	Lulus C
P	1,53	1,84	1,15	1,68	1,50	1,19	Ntb	18	Wiraswasta	Lulus B
P	1,89	2,16	1,40	2,03	1,81	1,36	Ntb	18	Wiraswasta	Lulus B
P	2,84	3,21	3,25	3,24	3,02	2,55	Jawa	18	Pns	Lulus C
L	2,74	3,11	3,00	3,08	2,91	2,33	Bali	18	Pns	Lulus C
L	3,11	2,68	2,65	2,89	2,81	2,12	Bali	19	Wiraswasta	Lulus C

D. Model Building (Pembangunan Model)

Pemodelan ini dilakukan sesuai dengan metode pemodelan pilihan dan kemudian diterapkan pada data besar yang telah disiapkan guna pemenuhan kebutuhan bisnis yang sesuai. Pemodelan yang dipakai disini

menggunakan teknik klasifikasi, dengan decision tree dan NBC. Model perhitungan menggunakan tools Rapidminer versi 9.10.011 dengan desain NBC pada gambar 2.



Gambar 2. Design NBC Menggunakan Rapidminer

1) Naive Bayes Classifier (NBC)

Perhitungan *Naive Bayes* menggunakan data testing pada tabel 5.

Tabel 5. Data Testing

Gender	Ips4	Ips5	Ips6	Ipk4	Ipk5	Ipk6	Asal	Usia Masuk	Pekerjaan Ortu	Predikat Lulus
Wanita	2,89	2,74	2,70	2,53	2,83	2,60	NTB	19	Wiraswasta	???

Terdapat data yang belum diketahui classnya pada tabel 5 diatas. Perhitungan dengan menggunakan model NBC :

- Hasil perhitungan jumlah class/label pada tabel 6

Tabel 6. Perhitungan Jumlah Class

Kelulusan	Lulus A	Lulus B	Lulus C	Lulus D
Nilai	0,02	0,18	0,66	0,14

- Menghitung kasus serupa didalam kelas yang serupa pada tabel 7

Tabel 7 Jumlah kasus yang dihitung pada setiap class

Class	Nilai	Lulus A	Lulus B	Lulus C	Lulus D
Gender	Wanita	1	0,4	0,84	0,90
IPS4	2,89	-	0,33	0,17	-
IPS5	2,74	-	-	0,17	-
IPS6	2,70	-	-	0,08	-
IPK4	2,53	-	-	0,08	-
IPK5	2,83	-	-	0,33	-
IPK6	2,60	-	-	-	1

Asal	NTB	0,16	-	-	-
Usia	19	0,16	-	0,50	0,33
Pekerjaan	Wiraswasta	0,42	1	1	0,33

c) Hasil perkalian variable ditunjukkan pada tabel 8

Tabel 8 Perhitungan Variabel

Kelulusan	Lulus A	Lulus B	Lulus C	Lulus D
Nilai	0	0	0,72	0

d) Membandingkan Hasil class

Karena hasil P(Lulus C) lebih besar maka keputusannya adalah **Lulus dengan predikat Sangat Memuaskan**.

2) *Algoritma Decision Tree*

Perhitungan nilai entropy dengan data pada tabel 4 adalah sebagai berikut :

$$\text{Nilai Entropy total : } \sum_{i=1}^n \left(-\frac{12}{20} * \log_2 \left(\frac{12}{20} \right) \right) + \left(-\frac{1}{20} * \log_2 \left(\frac{1}{20} \right) \right) + \left(-\frac{1}{20} * \log_2 \left(\frac{1}{20} \right) \right) + \left(-\frac{3}{20} * \log_2 \left(\frac{3}{20} \right) \right) + \left(-\frac{3}{20} * \log_2 \left(\frac{3}{20} \right) \right) = 1,70$$

Hasil perhitungan nilai entropy ditunjukkan pada tabel 9

Tabel 9 Hasil Perhitungan Entropy dan Gain

Rangking Atribut	Atribut	Entropy	Nilai	Nilai Gain
1	IPS4	IPS	0	1,70
		IPK	0	
2	IPK4	IPK	0	1,69
		IPK5	0	
3	IPS4	IPK	0	1,59
		IPS	1	
		IPK	0	
		IPS	1,83	
4	IPS6	IPK	0	1,59
		IPK	0	
		IPS	1	
		IPS	3,00	
5	Asal	IPS6	1	1,49
		IPS	0	
		Lombok	1,84	
		Jawa	0,47	
		Bali	0	
		Kalimantan	0	
6	Pekerjaan Ortu	Wiraswasta	1,68	0,82
		Swasta	0,97	
		PNS	0	
		Tani	0	
		MAN	1	
		90	1,5	
7	Gender	Laki-laki	1,5	0,06
		Wanita	1,67	
		A	0	
8	Predikat	B	0,47	0,19
		C	1,6	
		D	0	
		D	0	

Dari hasil perhitungan didapatkan atribut IPS4 dengan nilai gain 1,70 , IPK4 dan IPK 5 dengan nilai gain 1,69 memiliki nilai gain terbesar, dengan demikian ketiga atribut tersebut yang akan menjadi node akar(root node).

E. *Evaluasi*

Dalam tahap evaluasi dilakukan proses pengukuran hasil yang dicapai. Nilai akurasi pada masing-masing dapat dihitung dengan rumus berikut (3):

$$\frac{TP+TN}{TP+FN+FP+TN} \times 100\% \dots\dots\dots(3)$$



TP adalah True Positif adalah keseluruhan data yang sesuai dari class 1 yang digolongkan dalam class 1, TN adalah True Negatif adalah Jumlah data yang sesuai dari class 2 yang tergolong dalam class 2, FP adalah False Positif adalah Jumlah data yang tidak sesuai dari class 2 yang tergolong sebagai class 1, FN adalah False Negatif adalah jumlah data yang tidak sesuai dari class 1 yang digolongkan sebagai class 2

F. Deployment (Penyebaran)

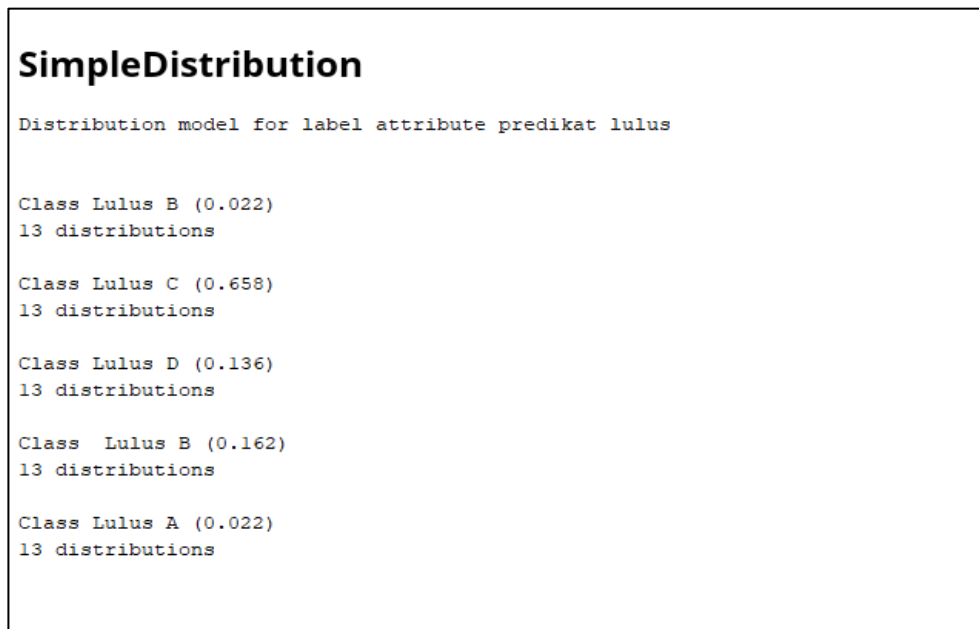
Setelah menyelesaikan proses pengujian model yang akurat dapat diperoleh untuk memprediksi predikat kelulusan pelajar menggunakan dua algoritma, NBC dan C.45. Model yang dihasilkan dapat digunakan untuk memprediksi data uji baru.

4. Hasil dan Pembahasan

Data yang dipakai disini diambil dari mahasiswa yang lulus dengan predikat indeks prestasi kurang sejumlah 5, data dari predikat indeks prestasi memuaskan sejumlah 42, data dari predikat indeks prestasi sangat memuaskan sejumlah 150, dan data dari predikat dengan pujian sejumlah 42. *Training* data yang digunakan sebanyak 228 wisudawan dan 5 wisudawan merupakan *testing* data yang digunakan untuk mengetahui tingkat akurasi algoritma NBC dan C.45. Atribut yang digunakan sebagai parameter adalah 11 atribut, dimana 10 adalah prediktor dan 1 adalah hasil.

a) Hasil Percobaan dan pengujian NBC

Pada gambar 3 didapatkan hasil pengujian NBC dengan *tools* rapidminer yang berisikan distribusi model untuk label predikat lulus. Pada tabel 10 ditunjukkan nilai class recall masing-masing atribut predikat lulus.



Gambar 3. Hasil Pengujian menggunakan NBC

Tabel 10. Hasil Pengujian NBC

true	true Lulus B	true Lulus C	true Lulus A	true Lulus D
pred. B	10	0	5	0
pred. C	0	30	0	0
pred. A	0	0	5	0
pred. D	0	0	0	0
class recall	80.00%	100.00%	50.00%	0.00%

Hasil percobaan dengan menggunakan NBC pada tabel 10, diketahui label Lulus C yaitu lulus dengan predikat sangat memuaskan lebih tinggi dari label lainnya. Akurasi dan nilai AUC dan waktu perhitungan dari NBC ditunjukkan pada tabel 11

Tabel 11. Akurasi dan AUC yang dihasilkan NBC

Naive Bayes Classifier	
Accuracy Confusion Matrix (%)	86.00
Kappa	0,757
Time	1

Penggunaan NBC dalam penelitian ini dihasilkan perhitungan dengan pola seperti pada tabel 12.

Tabel 12. Profil Mahasiswa yang dihasilkan dari perhitungan NBC

Gender	Ips1	Ips2	Ips3	Ips4	Ipk2	Ipk3	Ipk4	Asal	Usia Masuk	Pekerjaan Ortu	Pola
Wanita	2,27	2,20	1,96	2,14	2,16	2,08	2,10	Jawa	17	Wiraswasta	LULUS B
Wanita	2,80	2,98	2,98	3,02	2,98	2,90	2,78	Jawa	18	PNS	LULUS C
Wanita	3,47	3,54	3,61	3,35	3,54	3,53	3,28	Jawa	18	Wiraswasta	LULUS D
Laki-Laki	0,70	0,78	0,50	0,83	0,45	0,62	0,43	Kalimantan	18	TANI	LULUS A
Wanita	2,26	2,42	2,25	2,43	2,31	2,25	2,26	Jawa	18	Wiraswasta	LULUS B
Wanita	2,65	2,64	2,64	2,55	2,60	2,62	2,37	Jawa	18	Wiraswasta	LULUS C
Wanita	3,68	3,66	3,76	3,40	3,66	3,69	3,08	NTB	19	PNS	LULUS D

b) Hasil Percobaan Decision tree

Pada tabel 13 ditunjukkan nilai class recall masing-masing atribut predikat lulus pada pengujian dengan decision tree

Tabel 13 Hasil Pengujian Decision Tree

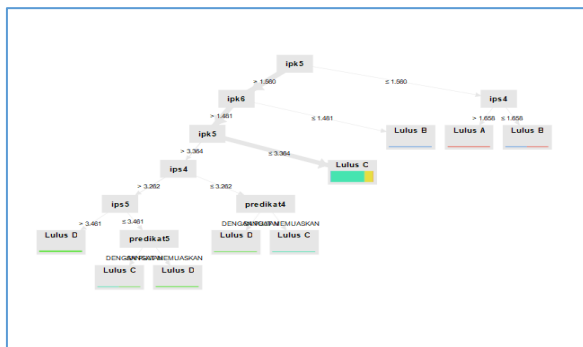
	true Lulus B	true Lulus C	true Lulus A	true Lulus D
pred. Lulus B	0	0	0	0
pred. Lulus C	10	30	9	0
pred. Lulus A	0	0	0	0
pred. Lulus D	0	0	0	0
class recall	0.00%	100.00%	0.00%	0.00%

Hasil pengujian dengan decision tree yang ditunjukkan pada tabel 13, diketahui bahwa nilai label C yaitu lulus dengan predikat sangat memuaskan lebih tinggi dari label lainnya. Akurasi, nilai AUC dan waktu perhitungan dari decision tree ditunjukkan pada tabel 14

Tabel 14 Hasil akurasi AUC Decision Tree

Decision Tree	
Accuracy Confusion Matrix (%)	60,00
Kappa	0,023
Time	2

Result dari decision tree yang dihasilkan dalam bentuk tree ditunjukkan pada gambar 4.



Gambar 4. Result Decision Tree

Dalam pohon keputusan pada gambar 4 atribut yang diseleksi adalah IPK5, IPK4, IPS4 dan IPS5, untuk atribut lainnya mengalami eliminasi dari pohon keputusan. Dari percobaan tersebut diketahui bahwa dengan jumlah dan jenis data yang sudah disiapkan, tidak semua datanya dipakai, algoritma tersebut melakukan pemrosesan data dengan beberapa atribut saja untuk mendapatkan kelas output dari dataset yang sudah disiapkan.

Nilai akurasi dari NBC dan Decision tree ditunjukkan pada tabel 15.

Tabel 15 Nilai Akurasi NBC dan Decision tree

NBC		Decision Tree	
Akurasi	Kappa	Akurasi	Kappa
86,00%	0,757	60,00%	0,023
	Time		Time
	1		2

Dalam proses pemecahan masalah pada penelitian disini, pemakaian model yang terdiri dari dua algoritma ini memiliki hasil yang lebih lengkap dibandingkan dengan penelitian-penelitian sebelumnya. Naive bayes lebih unggul dalam akurasi dibandingkan decision tree, dan menghasilkan pola yang nantinya dapat dipakai dalam pendukung keputusan, sedangkan model decision tree dapat mengetahui atribut yang unggul sebagai factor penyebabnya.

5. Kesimpulan

Predikat kelulusan mahasiswa dapat diprediksi sesuai dengan kriteria indeks prestasinya dan dapat diketahui faktor-faktor apa saja yang dapat mempengaruhinya. Dengan pemanfaatan algoritma naïve bayes dan algoritma decision tree yang dilakukan pada penelitian ini dipatkan actor-faktor yang mempengaruhi predikat kelulusan adalah indeks prestasi kumulatif tahun ketiga. Pola kelulusan tersebut juga dapat dilihat sejak mahasiswa masuk di tahun kedua. Dengan pola tersebut program studi dapat berupaya menyiapkan mahasiswa dapat lulus dengan predikat kelulusan yang ditetapkan. Hasil penelitian ini perlu dikembangkan lagi dengan peningkatan jumlah atribut dan data, dan perlu dibuat sebuah sistem penentu predikat kelulusan mahasiswa dari pola yang telah dihasilkan supaya dapat membantu perguruan tinggi untuk meningkatkan predikat kelulusan mahasiswa di setiap periodenya.

Daftar Pustaka

- [1] L. R. Rosmiati, "Aplikasi Berbasis Fuzzy C-Means Dalam Penentuan Predikat Kelulusan Mahasiswa," *Jurnal Ilmiah Ilmu Komputer Fakultas*, Vol. 2, P. 1–9, 2017.
- [2] V. W. T. A. C. M. L. A. M. M. Hairul Umam, "Analisis Perbandingan Algoritma C4.5 Dan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Mahasiswa (Studi Kasus : Prodi Teknik Informatika Universitas Muhammadiyah Jember)," Repository Um Jember, Jember, 2019.
- [3] A. Y. N. Kunang And S. Murniati, "Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang," In *Seminar Nasional Informatika* , Yogyakarta, 2013.
- [4] K. D. E. T. Luthfi, *Algoritma Data Mining*, CV Andi Offset, 2009.
- [5] E. E. A. A. L. Turban, *Decision Support System And Intelligent System*, Yogyakarta: Penerbit Andi, 2005.
- [6] H. S. D. M. S. Mujib Ridwan, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Jurnal EECCIS*, Pp. 59-64, 2013.
- [7] B. Santosa, *Data Mining, Teknik Pemanfaatan Data Untuk Keperluan Bisnis : Teori Dan Aplikasi*, Yogyakarta: Graha Ilmu, 2007.
- [8] J. D. G. S. Mcleod, *Management Information System*, Vol. 10th Ed., 10th Ed Pearsoneducation, Inc, 2007.
- [9] A. S. N. A. R. K. Y. I. M. Nurul Khasanah, "Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes," *Technologia*, Vol. 13, P. 207, 2022.
- [10] Y. P. Arie Yandi Saputra, "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour," *Techno.COM*, Vol. 17, Pp. 395-403J, November 2018.
- [11] A. R. I. A. F. Dwi Astuti, "Penentuan Strategi Promosi Usaha Mikrokecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM Dengan algoritmak-Means Clustering," *Journal Of Informatics, Information System, Software Engineering And Applications (INISTA)*, Vol. 1, Pp. 60-72, 2019.