

---

---

**PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU  
MENGUNAKAN METODE DECISION TREE DAN NAÏVE  
BAYES**

***PREDICTING ON-TIME GRADUATION OF STUDENTS USING  
DECISION TREE AND NAÏVE BAYES METHODS***

Azis Wahyudi<sup>1</sup>, Kusri<sup>2</sup>, Ferry Wahyu Wibowo<sup>3</sup>

1,2,3 Magister Teknik Informatika, Universitas Amikom Yogyakarta E-mail: \*1

Azis.wahyudi@students.amikom.ac.id

---

Abstrak

Kelulusan tepat waktu sendiri merupakan salah satu indikator keberhasilan hasil kinerja akademik mahasiswa. Ketentuan masa studi sendiri sudah diatur dalam ketetapan Peraturan Menteri dan Pendidikan Kebudayaan Indonesia. Untuk mengatasi hal tersebut perlu ada teknik untuk bisa melakukan prediksi terhadap kelulusan, Adapun teknik yang sering digunakan adalah dengan menggunakan data mining. Dalam penelitian ini penulis akan membandingkan dua metode data mining yaitu metode Naive Bayes Classifier dan Decision Tree. sehingga diperoleh metode dengan akurasi prediksi kelulusan mahasiswa yang terbaik. Atribut yang digunakan untuk Klasifikasi Data Mining terdiri atas 10 atribut yaitu NIM, Jenis Kelamin, Status Mahasiswa, Umur, Indeks Prestasi Semester 1, Indeks Prestasi Semester 2, Indeks Prestasi Semester 3, Indeks Prestasi Semester 4, Indeks Prestasi Kumulatif, Keterangan Sebagai atribut hasil. Dari hasil proses pengujian dengan tools RapidMiner Menggunakan dua metode yang telah dilakukan. Decision Tree(C4.5) memperoleh hasil akurasi tertinggi sebesar 88.92% dan Metode Naïve Bayes memperoleh hasil akurasi tertinggi sebesar 84.98%.

---

*Abstract*

*Timely graduation itself is one of the indicators of the success of students' academic performance. The study period regulations are already set in the provisions of the Minister of Education and Culture of Indonesia. To address this issue, there needs to be a technique to predict graduation. One of the techniques commonly used is data mining. In this study, the authors will compare two data mining methods, namely Naive Bayes Classifier and Decision Tree, to obtain the method with the best accuracy in predicting student graduation. The attributes used for Data Mining Classification consist of 10 attributes: Student ID, Gender, Student Status, Age, Semester 1 Grade Point Average, Semester 2 Grade Point Average, Semester 3 Grade Point Average, Semester 4 Grade Point Average, Cumulative Grade Point Average, and Result attribute. From the test results using RapidMiner tools with two methods that have been conducted, the Decision Tree (C4.5) obtained the highest accuracy result of 88.92%, and the Naïve Bayes method obtained the highest accuracy result of 84.98%.*

*Keywords* — *Decision Tree (C4.5), Naïve Bayes, Prediction*

## 1. PENDAHULUAN

Kelulusan merupakan salah satu komponen bagi dalam penilaian akreditasi. Mahasiswa lulus tepat waktu maka akan membantu penilaian akreditasi terhadap program studi hingga perguruan tinggi. (BANPT, 2007) Kelulusan tepat waktu sendiri merupakan salah satu indikator keberhasilan hasil kinerja akademik mahasiswa. Ketentuan masa studi sendiri sudah diatur dalam ketetapan Peraturan Menteri dan Pendidikan Kebudayaan Indonesia yang menjelaskan bahwa kompetensi lulusan bagi mahasiswa diploma tiga dapat menyelesaikan beban wajib minimal 108 SKS dengan masa studi 3 (tiga) tahun dan paling lama 5 (lima) tahun akademik untuk program diploma tiga. (Permenristekdikti No. 44 Tahun 2015).

Untuk mengatasi hal tersebut perlu ada teknik untuk bisa melakukan prediksi terhadap kelulusan, Adapun teknik yang sering digunakan adalah dengan menggunakan data mining. Dan metode yang sering digunakan untuk prediksi kelulusan mahasiswa adalah metode klasifikasi. (Rohmawan, 2018) Menurut (Suntoro, 2019) data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan. Data mining adalah proses menganalisa data dari yang berbeda dan menyimpulkannya menjadi informasi atau pengetahuan atau pola yang penting untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya (Witten, 2016).

Data mining sudah ada sejak lama dan teori-teorinya pun sudah banyak dibahas dalam literatur. Teori-teori tersebut antara

lain: Naive-Bayes dan Nearest Nighbour, Pohon Keputusan, Aturan Asosiasi, k-Means Clustering, dan Text Mining (Bramer, 2007). Metode klasifikasi merupakan pendekatan untuk menjalankan fungsi klasifikasi dalam data mining yaitu menggolongkan data. Teknik klasifikasi ini dapat pula digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma Decision Tree C.45, Artificial Neural Networks (ANN), K-Nearest Neighbour (KNN), Algoritma Naive Bayes, Algoritma Genetik, Rough Set, Metode Berbasis Aturan, Memory Based Reasoning, dan Support Vector Machine (Widodo dkk., 2013).

Penelitian yang dilakukan (Hendrawan dkk, 2021) “Klasifikasi Lama Studi dan Predikat Kelulusan Mahasiswa menggunakan Metode Naïve Bayes” pada penelitian ini diklasifikasikan lama studi dan predikat kelulusan mahasiswa dengan tujuan untuk membantu pihak program studi dan fakultas dalam menganalisis luaran pembelajaran. Metode klasifikasi yang diterapkan pada penelitian ini adalah Naïve Bayes. Data yang digunakan adalah data mahasiswa Institut Teknologi dan Bisnis STIKOM Bali tahun 2008 sampai dengan tahun 2016 dengan total jumlah data sebanyak 5.081. Atribut dataset yang digunakan untuk mengklasifikasikan Lama Studi dan Predikat Kelulusan adalah Jenis Kelamin, Prodi, Konsentrasi, Tahun Masuk, dan Tahun Lulus. Hasil eksperimen menunjukkan bahwa akurasi tes classifier untuk klasifikasi lama studi sebesar 0,74 dan untuk akurasi tes klasifikasi predikat kelulusan sebesar 0,61 pada kelompok data Program Studi Sistem Komputer. Kemudian

untuk kelompok data Program Studi Sistem Informasi akurasi tes klasifikasi lama studi

Dalam penelitian ini penulis akan membandingkan dua metode data mining yaitu metode Naive Bayes Classifier dan

sebesar 0,73 dan untuk akurasi klasifikasi predikat kelulusan sebesar 0,67.]

Decision Tree. sehingga diperoleh metode dengan akurasi prediksi kelulusan mahasiswa yang terbaik.

## 2. METODE

Dalam Penelitian ini menggunakan jenis penelitian eksperimen, yaitu melakukan pengujian performance dan memprediksi dengan menggunakan algoritma Naïve bayes dan C.4.5 dalam prediksi kelulusan mahasiswa tepat waktu menggunakan data kelulusan dan data mahasiswa di prodi Rekam Medis & Infokes. Dalam pengumpulan data penelitian ini mengambil data kelulusan mahasiswa dari bagian kemahasiswaan dan alumni. Adapun sumber data yang digunakan dalam penelitian ini yaitu data primer dan data sekunder. Data primer merupakan data yang di dapat langsung dari bagian akademik yaitu data mahasiswa jurusan Rmik tahun akademik 2017//2018, 2018/2019 dan 2019/2020 berjumlah 380 data kemudian 280 di gunakan untuk data tranning dan 100 data digunakan untuk data testing, Atribut yang menjadi parameter sebanyak 10 atribut yaitu NIM, Jenis Kelamin, Status Mahasiswa, Umur, Indeks Prestasi Semester 1, Indeks Prestasi Semester 2, Indeks Prestasi Semester 3, Indeks Prestasi Semester 4,

Indek Prestasi Kumulatif, Keterangan Sebagai atribut hasil. sedangkan data sekunder yaitu data yang di peroleh berupa buku dan jurnal. Untuk data sekunder peneliti melakukan pencarian di internet dan kunjungan ke perpustakaan.

Tabel 1. Contoh Sample dataset

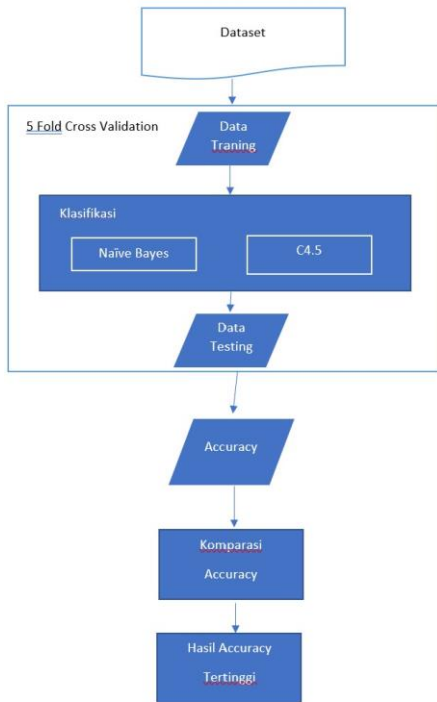
NIM	JK	Umur	Status
2019133001	Perempuan	20	Belum Menikah
2019133002	Perempuan	22	Belum Menikah
2019133003	Laki-laki	21	Belum Menikah
2019133004	Perempuan	22	Belum Menikah
2019133005	Perempuan	21	Belum Menikah
2019133006	Laki-laki	24	Belum Menikah
2019133007	Perempuan	23	Belum Menikah

Tabel 2. Atribut Penelitian

No	Nama Atribut	Keterangan
1	NIM	- NIM Mahasiswa
2	JK	- L dan P
3	Umur	- 17-25 Remaja - 26-35 Dewasa
4	Status Mahasiswa	- Menikah dan Belum Menikah
5	IPS 1	- 0.00-1.99 Kurang Memuaskan - 2.00-2.75 Memuaskan - 2.76-3.50 Sangat Memuaskan - 3.51-4.00 Dengan Pujian
6	IPS 2	- 0.00-1.99 Kurang Memuaskan - 2.00-2.75 Memuaskan - 2.76-3.50 Sangat Memuaskan - 3.51-4.00 Dengan Pujian
7	IPS 3	- 0.00-1.99 Kurang Memuaskan - 2.00-2.75 Memuaskan - 2.76-3.50 Sangat Memuaskan - 3.51-4.00 Dengan Pujian
8	IPS 4	- 0.00-1.99 Kurang Memuaskan - 2.00-2.75 Memuaskan - 2.76-3.50 Sangat Memuaskan - 3.51-4.00 Dengan Pujian
9	IPK	- 0.00-1.99 Kurang Memuaskan - 2.00-2.75 Memuaskan - 2.76-3.50 Sangat Memuaskan - 3.51-4.00 Dengan Pujian
10	KET	- Terlambat - Tepat

Jenis data yang digunakan dalam penelitian ini adalah data Primer. Data

Primer adalah sumber data yang berlangsung memberikan data kepada pengumpul data (Sugiyono, 2016).



Gambar 1. Metode komparasi Decision Tree, Naïve Bayes

Pada penelitian ini dilakukan dengan mengkomparasi 3 metode yaitu Decision Tree, Naive Bayes ada Gambar 1. Proses yang dilakukan adalah training dan testing dataset dengan menggunakan metode Decision Tree dan Naïve Bayes untuk menghasilkan akurasi. Akurasi yang dihasilkan kedua metode tersebut kemudian dikomparasi untuk mendapatkan akurasi tertinggi (Sani, 2018).

### 2.1.1. Naïve Bayes

Klasifikasi Bayes juga dikenal dengan Naïve Bayes, memiliki kemampuan sebanding dengan dengan pohon keputusan dan neural network (Han & Kamber, 2007). Klasifikasi Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas (Kusrini, 2009).

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah dengan cara menyebarkan

$$P(H|X) = \frac{P(X|H)}{P(X)} \cdot P(H)$$

(1)

Keterangan

- X : Data dengan class yang belum diketahui
- H : Hipotesis data merupakan suatu class spesifik
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- P(H) : Probabilitas hipotesis H (prior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X [7]

### 2.1.2. Decision Tree

Algoritma decision tree yang digunakan adalah C4.5. secara umum untuk membangun pohon keputusan adalah sebagai berikut:

- a. Pilih atribut sebagai node akar
- b. Buat cabang untuk tiap-tiap nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang

sama.

Perhitungan yang pertama adalah perhitungan Gain pada Persamaan 2 dan perhitungan Entropy pada Persamaan 3

Perhitungan Gain

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

(2)

Keterangan:

- S : himpunan
- A : atribut
- n : jumlah partisi atribut A
- |S<sub>i</sub>| : jumlah kasus pada partikel ke-i
- |S| : jumlah kasus dalam S

Menghitung Nilai Entropy

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

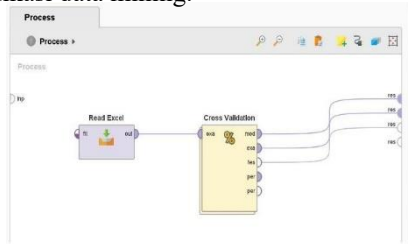
(3)

Keterangan :

- S : himpunan kasus
- A : fitur
- n : jumlah partisi S
- p<sub>i</sub> : Proporsi dari S<sub>i</sub> terhadap S [5]

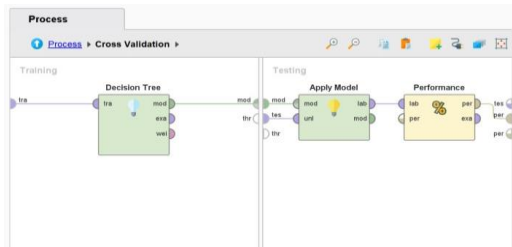
#### 4. PEMBAHASAN

Tujuan dari penelitian adalah untuk menganalisa prediksi mahasiswa lulus tepat waktu dengan menerapkan teknik klasifikasi data mining.

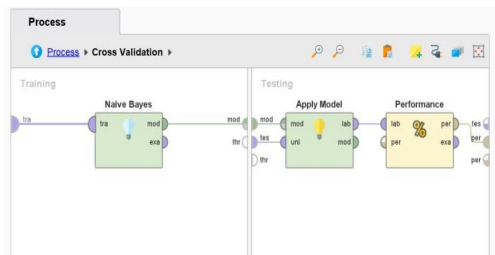


Gambar 2. Model Proses Desain Import Data

1. Read Excel: Operator ini dapat digunakan untuk memuat data dari spreadsheet Microsoft Excel.
2. Cross Validation: Operator yang bersarang. Ini memiliki dua subproses: subproses pelatihan dan subproses pengujian. Subproses pelatihan digunakan untuk melatih model. Model yang terlatih kemudian diterapkan dalam subproses pengujian. Kinerja model diukur selama fase Pengujian



Gambar 3. Model Validasi C4.5



Gambar 4. Model Validasi Naïve Bayes

3. Model Validasi: Metode klasifikasi yang digunakan dalam penelitian ini yaitu Decision Tree dapat dilihat pada gambar 3, Naïve Bayes Dapat dilihat pada gambar 4.
4. Apply Model: Operator yang digunakan untuk penghubung metode Decision Tree, dan Naïve Bayes ke performance. Dapat dilihat pada gambar 3 dan 4
5. Performance: Operator yang digunakan untuk mengukur performance akurasi dari model. Pada gambar 5.

	true TERLAMBAT	true TEPAT	class precision
pred TERLAMBAT	140	19	88.05%
pred TEPAT	23	197	89.55%
class recall	85.89%	91.20%	

accuracy: 88.92% +/- 6.64% (micro average: 88.92%)

Gambar 5. Hasil Akurasi Decision Tree

Dari proses evaluasi model Decision Tree pada gambar 5 dan proses validasi terbentuk hasil matrix akurasi sebesar 88.92 % pada pengujian ke 2 .Dibawah ini merupakan perhitungan akurasi menggunakan Confusion Matrix dari gambar 5 diatas.

	true TERLAMBAT	true TEPAT	class precision
pred TERLAMBAT	140	34	80.46%
pred TEPAT	23	182	88.78%
class recall	85.89%	84.26%	

accuracy: 84.97% +/- 6.42% (micro average: 84.96%)

Gambar 6. Hasil Akurasi Naïve Bayes

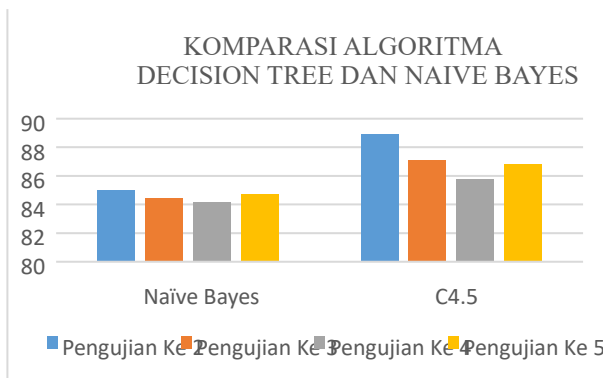
Dari proses evaluasi model Naïve Bayes dan proses validasi terbentuk hasil matrix Accuracy sebesar 84.96 % pada pengujian ke 4. Dibawah ini merupakan perhitungan akurasi menggunakan Confusion Matrix dari gambar 6 diatas.

#### Hasil Pengujian Menggunakan K-5 Cross Validation

Tabel 3. Hasil Pengujian akurasi

Metode	2	3	4	5
Naïve Bayes	84,98	84,43	84,17	84,70
C4.5	88,92	87,07	85,75	86,81

Pada tabel hasil pengujian menggunakan teknik K-5 fold cross validation terhadap data mahasiswa dengan pengujian data mulai dari 2,3,4,dan 5.



Gambar 7. Diagram Chart Hasil Komparasi

Dari hasil komparasi tersebut menunjukkan Decision Tree memiliki tingkat akurasi yang paling tinggi. Hal tersebut menunjukkan bahwa kinerja Decision Tree lebih baik dibanding dengan Naive Bayes pada gambar 7.

## 5. KESIMPULAN

Atribut yang digunakan untuk Klasifikasi Data Mining terdiri atas 10 atribut yaitu NIM, Jenis Kelamin, Status Mahasiswa, Umur, Indeks Prestasi Semester 1, Indeks Prestasi Semester 2, Indeks Prestasi Semester 3, Indeks Prestasi Semester 4, Indeks Prestasi Kumulatif, Keterangan Sebagai atribut hasil. Dari hasil Atribut yang digunakan untuk Klasifikasi Data Mining terdiri atas 10 atribut NIM, Jenis Kelamin, Status Mahasiswa, Umur, Indeks Prestasi Semester 1, Indeks Prestasi Semester 2, Indeks Prestasi Semester 3, Indeks Prestasi Semester 4, Indeks Prestasi

Kumulatif, Keterangan Sebagai atribut hasil. Dari hasil proses pengujian dengan tools RapidMiner Menggunakan dua metode yang telah dilakukan. Decision Tree(C4.5) memperoleh hasil akurasi tertinggi sebesar 88.92% dan Metode Naïve Bayes memperoleh hasil akurasi tertinggi sebesar 84.98%

## DAFTAR PUSTAKA

- BANPT. (2007). *Akreditasi Institusi Perguruan*.
- Bramer, M. (2007). *Springer Science & Business Media*.
- Hendra. (n.d.). *Klasifikasi Lama Studi dan Predikat Kelulusan Mahasiswa menggunakan Metode Naïve Bayes*.
- Ian H. Witten, E. F. (2016). *Data Mining Practical Machine Learning Tools and Techniques*. Elsevier.
- Isnain Mulia, M. (2021). *Model Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree C4.5 dan Software Weka*. *Jurnal Analisis Sistem Pendidikan Tinggi VOL. 5 NO. 1 2021, 57-64*.
- Kusrini, d. E. (2009). *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- Rohmawan, E. P. (2018). *Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Artificial Neural Network*. *Jurnal Ilmiah Matrik*, pp. 21-30.

- S. Salmu and A. Solichin. ( 2017).  
“Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta,” . *Pros. Semin. Nas. Multidisiplin Ilmu Univ. Budi Luhur*; no. April, , 701–709.
- Sani, A. (2018). Penerapan Metode K-Means Clustering Pada Perusahaan. *Jurnal Ilmiah Teknologi Informasi*, (353), 1–7.
- Suntoro, J. (2019). *Data Mining Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: Elex Media Komputindo.
- Widodo, d. (2013). *Penerapan Data mining dengan MATLAB*. Bandung: Rekayasa Sains

