

**Implementasi *Data Mining* Untuk Memprediksi Kelulusan Mahasiswa Fakultas Sains Dan Teknologi Menggunakan Metode *Naive Bayes***  
(Studi Kasus : Universitas Muhammadiyah Jambi)

**Hetty Rohayani<sup>1\*</sup>, Rico<sup>2</sup>, Kevin Kurniawansyah<sup>3</sup>, Arpan Saputra Harahap<sup>4</sup>**  
Fakultas Sains dan Teknologi, Universitas Muhammadiyah Jambi, Indonesia<sup>1,3,4</sup>  
Fakultas Teknik dan Ilmu Komputer, Universitas Adiwangsa Jambi, Indonesia<sup>2</sup>  
Email : hettyrohayani@gmail.com<sup>1</sup>,  
reecho86@gmail.com<sup>2</sup>, kevin.kurniawansyah4h@gmail.com<sup>3</sup>,  
Arpannsaputra@gmail.com<sup>4</sup>

**ABSTRAK**

Fakultas Sains dan Teknologi universitas muhammadiyah Jambi memiliki cukup banyak mahasiswa. Dengan bertambah pesatnya jumlah mahasiswa baru setiap tahunnya maka secara otomatis akan terciptanya data-data yang sangat banyak dan akan bertambah melimpahnya data yang tersimpan dalam database. Agar data yang awalnya sangat minim informasi tersebut dapat diubah menjadi suatu informasi yang sangat bermanfaat dengan menggunakan teknik data mining. Salah satunya penelitian ini difokuskan untuk mengevaluasi kategori mahasiswa yang dapat lulus tepat waktu atau tidak berdasarkan atribut-atribut yang digunakan, Metode naive bayes digunakan untuk menganalisis data dalam pengenalan pola dan memprediksi kelulusan mahasiswa. Data yang dibutuhkan adalah data-data mahasiswa yang sudah lulus sebagai data training sedangkan data-data mahasiswa yang masih menempuh studi kuliah sebagai data uji. Berdasarkan hasil prediksi setelah melakukan uji coba sebanyak 6x percobaan hasil didapatkan dengan keakuratan 93.88% yaitu pred Lulus dengan true lulus 18 dan true tidak 2 class precision 90.00% dan pred Tidak dengan true lulus 1 dan true tidak 28 class precision 96.55%.

**Kata Kunci :** *Data Mining, Naive Bayes, Prediksi*

**ABSTRACT**

*The Faculty of Science and Technology, Jambi Muhammadiyah University, has quite a number of students. With the rapid increase in the number of new students each year, a lot of data will automatically be created and the abundance of data stored in the database will increase. So that the data which was originally very minimal information can be turned into a very useful information by using data mining techniques. One of these studies is focused on evaluating the categories of students who can graduate on time or not based on the attributes used. The Naive Bayes method is used to analyze data in pattern recognition and predict student graduation. The data needed is the data of students who have graduated as training data while the data of students who are still studying in college as test data. Based on the prediction results after testing 6x trials, the results were obtained with an accuracy of 93.88%, namely pred Pass with true pass 18 and true no 2 class precision 90.00% and pred No true pass 1 and true no 28 class precision 96.55%.*

*Keywords :* *Data Mining, Naive Bayes, Prediction*

## I. PENDAHULUAN

Mahasiswa merupakan salah satu aspek penting dalam evaluasi keberhasilan penyelenggaraan program studi pada suatu perguruan tinggi. Pemantauan mahasiswa yang masuk, peningkatan kemampuan mahasiswa, prestasi yang dicapai mahasiswa, rasio kelulusan terhadap jumlah total mahasiswa, dan kompetensi lulusan seyogyanya mendapatkan perhatian yang serius untuk memperoleh kepercayaan stakeholder dalam menilai dan menetapkan penggunaan lulusannya [1][2]. Berdasarkan uraian di atas, pada penelitian ini akan dibuat sebuah sistem untuk mengklasifikasikan tumpukan-tumpukan data tersebut dan dapat dijadikan suatu informasi yang sangat berharga. Untuk membantu dalam menemukan informasi-informasi berharga tersebut diperlukan adanya teknik *data mining*. *Data Mining* merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut [3]. Teknik data mining yang akan digunakan dalam penelitian ini adalah algoritma *Naive Bayes Classifier* yang merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes. Ide dasar dari Teorema Bayes adalah menangani masalah yang bersifat hipotesis yakni mendesain suatu klasifikasi untuk memisahkan objek [4]. Tujuan dari penelitian ini yaitu memprediksi masa studi mahasiswa di Fakultas Sains dan Teknologi Universitas Muhammadiyah Jambi.

## II. LANDASAN TEORI

1. **Data Mining**  
Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *datawarehouse*, atau penyimpanan informasi lainnya [5][6]. Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk menginteraksi mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terikat dari berbagai database besar [7].
2. **Prediksi**  
Prediksi adalah memperkirakan sesuatu yang akan terjadi pada masa yang mendatang. Prediksi didapatkan melalui metode ilmiah maupun hanya subjektif belaka. Prediksi juga dapat digunakan dalam pengklasifikasian, tidak hanya untuk memprediksi *time series*, karena sifatnya yang bisa menghasilkan *class* berdasarkan berbagai atribut yang kita sediakan [8][9].
3. **Kelulusan Mahasiswa**  
Mahasiswa yang memenuhi persyaratan kelulusan ditetapkan dalam yudisium kelulusan Fakultas/ Program Studidan ditetapkan dengan keputusan Rektor. Tanggal kelulusan ditetapkan berdasarkan tanggal yudisium kelulusan dan merupakan tanggal penetapan IPK akhir (transkrip nilai) [4][10].
4. **Naive Bayes**  
NBC merupakan salah satu algoritma dalam teknik data mining yang menerapkan teori Bayes dalam klasifikasi. Teorema keputusan Bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern recognition) [1].

Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan Teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat.[11].

### III. METODOLOGI PENELITIAN

#### a. Data Mining

- Pengelompokan Data Mining Menurut Larose, terdapat enam kelompok data mining berdasarkan tugas yang dapat dilakukan, yaitu[4]:

##### a. Deskripsi

Analisis mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data secara sederhana.

##### b. Estimasi

Memperkirakan sebuah nilai yang belum diketahui.

##### c. Prediksi

Menerka / memperkirakan nilai di masa yang akan datang.

##### d. Klasifikasi

Membedakan - bedakan data berdasarkan kelas dari suatu objek yang labelnya tidak diketahui.

##### e. Pengklusteran

Mengelompokkan data yang memiliki karakteristik tertentu.

##### f. Asosiasi

Hubungan antara kejadian yang terjadi pada satuan waktu.

- Metode Data Mining

Menurut Larose, terdapat enam fase data mining CRISP-DM (Cross Industry Standard Process for Data Mining)[4].

a. Fase Pemahaman Bisnis (Business Understanding Phase)

b. Fase Pemahaman Data (Data Understanding Phase)

c. Fase Pengolahan Data (Data Preparation Phase)

d. Fase Pemodelan (Modelling Phase)

e. Fase Evaluasi (Evaluation Phase)

f. Fase Penyebaran (Deployment Phase)

#### b. Penentuan Atribut

Dalam memprediksi kelulusan mahasiswa maka diperlukan data-data seluruh mahasiswa baik yang sudah lulus maupun yang belum lulus. Semua data-data diperoleh dari Biro Administrasi Akademik (BAA) dan Laboratorium Fakultas Sains dan Teknologi. Data yang digunakan sebagai data training yaitu mahasiswa yang sudah lulus, sedangkan data yang digunakan sebagai data uji yaitu mahasiswa yang masih aktif mengikuti perkuliahan diambil sampel secara acak. Atribut-atribut yang akan digunakan yaitu:

**Tabel 1.** Daftar Atribut

Atribut	Isi dalam Atribut-Atribut
Jurusan sekolah	IPA, IPS dan Lain
Gender	Pria dan Wanita
Daerah asal mahasiswa	Kota Jambi dan Luar Kota Jambi
Asal sekolah	Kota Jambi dan Luar Kota Jambi
Rata SKS (satuan kredit semester)	SKS $\leq$ 16 dan SKS $>$ 16
Rata MK (Mata kuliah)	MK $\leq$ 8 dan MK $>$ 8
Lama studi	Tepat (lulus $\leq$ 4 tahun) dan Terlambat (lulus $>$ 4 tahun)

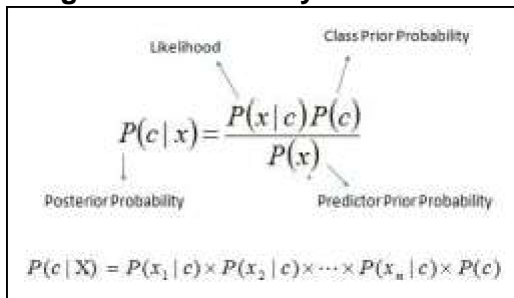
**c. Penentuan sampel**

Untuk menentukan jumlah sampel yang akan digunakan dalam menggambar dari jumlah populasi menggunakan bantuan metode slovin dengan nilai maksimal  $e = 5\%$  (Umar, 2004)[4].

$$n = \frac{N}{1 + N e^2} \quad (1)$$

Bahwa :  $n$  = jumlah sampel  
 $N$  = jumlah populasi  
 $e$  = nilai toleransi

**d. Algoritma Naive Bayes**



Naive bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara berkondisi saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Dengan memasukkan Persamaan 1 ke Persamaan 2 akan diperoleh pendekatan yang digunakan dalam NBC[4].

$$P(a_1, a_2, a_3, \dots, a_n | V_j) = \prod_i P(a_i | v_j) \quad (1)$$

$$V_{nb} = \arg \max_{v \in V} P(V_j) \prod_i P(a_i | v_j) \quad (2)$$

dengan  
 VNB : nilai *ouput* hasil klasifikasi Naive Bayes  
 $P(a_i|v_j)$  : rasio antara  $n_c / n$ , di mana  $n_c$  adalah jumlah data training untuk

$v=v_j$  dan  $a=a_i$ ; dan  $n$  adalah total kemungkinan output

**e. RapidMiner**

*RapidMiner* adalah sebuah lingkungan *machine learning data mining, text mining dan predictive analytics* yang akan digunakan pada kasus ini.

**IV. HASIL DAN PEMBAHASAN**

The screenshot shows a table with 20 rows of student data. Columns include NIM, Nama, Jurusan, Daerah, Jenis Kelamin, SKS, and Status Kelulusan. The data is used for training a Naive Bayes classifier.

**Gambar 1.** Hasil Dataset

Penjelasan dataset:

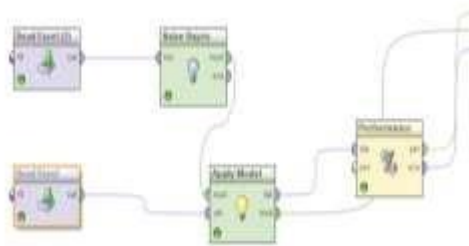
Gambar 1 di atas merupakan dataset yang digunakan. Dengan variabel NIM, nama, jurusan, daerah asal, jenis kelamin, SKS yang ditempuh, IPK sebelum kelulusan, dan status kelulusan. Dimana variabel tersebut akan digunakan jenjang, jurusan, daerah asal, jenis kelamin, SKS yang ditempuh, dan IPK

sebelum kelulusan sebagai pengelompokan variabel klasifikasi status kelulusan mahasiswa

Penelitian ini menggunakan metode Naive Bayesian Classification (NBC). NBC merupakan algoritma klasifikasi yang sangat efektif (mendapatkan hasil yang tepat) dan efisien (proses penalaran dilakukan memanfaatkan input yang ada dengan cara yang relatif cepat).

Implementasi *Data Mining* Untuk Memprediksi Kelulusan Mahasiswa Fakultas Sains Dan Teknologi Menggunakan Metode *Naive Bayes* (Studi Kasus : Universitas Muhammadiyah Jambi)

Algoritma NBC bertujuan untuk melakukan klasifikasi data pada kelas tertentu. Model statistik merupakan salah satu model yang terpercaya sangat andal sebagai pendukung pengambilan keputusan. Konsep probabilitas merupakan salah satu bentuk model statistik. Salah satu metode yang menggunakan konsep probabilitas adalah *Naive Bayesian Classification (NBC)*. Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satusama lain[4]. Penggunaan *Rapid Miner* untuk mendapatkan kalkulasi klasifikasi *naive bayes*



Gambar 2. Masukkan setiap dataset

ID	nama	jenis_kelamin	status_kelulusan
1	14000001	laki	lulus
2	14000002	laki	lulus
3	14000003	laki	lulus
4	14000004	laki	lulus
5	14000005	laki	lulus
6	14000006	laki	lulus
7	14000007	laki	lulus
8	14000008	laki	lulus
9	14000009	laki	lulus
10	14000010	laki	lulus
11	14000011	laki	lulus
12	14000012	laki	lulus
13	14000013	laki	lulus
14	14000014	laki	lulus
15	14000015	laki	lulus
16	14000016	laki	lulus
17	14000017	laki	lulus
18	14000018	laki	lulus
19	14000019	laki	lulus
20	14000020	laki	lulus
21	14000021	laki	lulus
22	14000022	laki	lulus
23	14000023	laki	lulus
24	14000024	laki	lulus
25	14000025	laki	lulus
26	14000026	laki	lulus
27	14000027	laki	lulus
28	14000028	laki	lulus
29	14000029	laki	lulus
30	14000030	laki	lulus
31	14000031	laki	lulus
32	14000032	laki	lulus
33	14000033	laki	lulus
34	14000034	laki	lulus
35	14000035	laki	lulus
36	14000036	laki	lulus
37	14000037	laki	lulus
38	14000038	laki	lulus
39	14000039	laki	lulus
40	14000040	laki	lulus
41	14000041	laki	lulus
42	14000042	laki	lulus
43	14000043	laki	lulus
44	14000044	laki	lulus
45	14000045	laki	lulus
46	14000046	laki	lulus
47	14000047	laki	lulus
48	14000048	laki	lulus
49	14000049	laki	lulus
50	14000050	laki	lulus
51	14000051	laki	lulus
52	14000052	laki	lulus
53	14000053	laki	lulus
54	14000054	laki	lulus
55	14000055	laki	lulus
56	14000056	laki	lulus
57	14000057	laki	lulus
58	14000058	laki	lulus
59	14000059	laki	lulus
60	14000060	laki	lulus
61	14000061	laki	lulus
62	14000062	laki	lulus
63	14000063	laki	lulus
64	14000064	laki	lulus
65	14000065	laki	lulus
66	14000066	laki	lulus
67	14000067	laki	lulus
68	14000068	laki	lulus
69	14000069	laki	lulus
70	14000070	laki	lulus
71	14000071	laki	lulus
72	14000072	laki	lulus
73	14000073	laki	lulus
74	14000074	laki	lulus
75	14000075	laki	lulus
76	14000076	laki	lulus
77	14000077	laki	lulus
78	14000078	laki	lulus
79	14000079	laki	lulus
80	14000080	laki	lulus
81	14000081	laki	lulus
82	14000082	laki	lulus
83	14000083	laki	lulus
84	14000084	laki	lulus
85	14000085	laki	lulus
86	14000086	laki	lulus
87	14000087	laki	lulus
88	14000088	laki	lulus
89	14000089	laki	lulus
90	14000090	laki	lulus
91	14000091	laki	lulus
92	14000092	laki	lulus
93	14000093	laki	lulus
94	14000094	laki	lulus
95	14000095	laki	lulus
96	14000096	laki	lulus
97	14000097	laki	lulus
98	14000098	laki	lulus
99	14000099	laki	lulus
100	14000100	laki	lulus

Gambar 3. Hasil Dataset

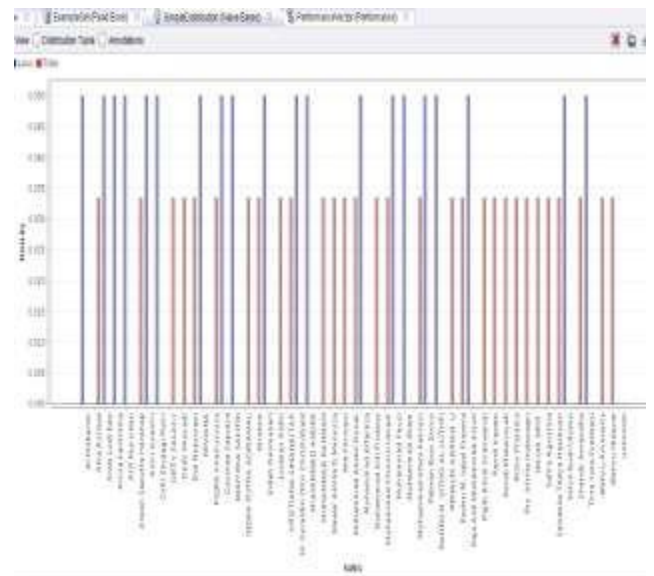
Import dataset ke rapid miner dengan NIM sebagai id polynomial , Program Studi polynomial, Provinsi polynomial, Gender binomial karena hanya ada Laki-laki dan perempuan, sks, dan IPK yang telah dikelompokkan menjadi kelompok ipk.



Gambar 4. List kegagalan dalam berproses

	Actual Lulus	Actual Tidak Lulus	Total	Presisi
Pred Lulus	75	2	77	98.7%
Pred Tidak Lulus	1	23	24	95.8%
Total	76	25	101	

Gambar 5. Hasil performance



Gambar 6. Hasil Proses dengan Naive Bayes

### Simple Distribution

Distribution model for label attribute Ket

Class Lulus (0.400)

10 distributions

Class Tidak (0.600)

10 distributions

Dilihat dari gambar 4,5,6 Setelah melakukan uji coba sebanyak 6x percobaan hasil didapatkan dengan keakuratan 93.88% yaitu pred Lulus dengan true lulus 18 dan true tidak 2 class precision 90.00% dan pred Tidak dengan true lulus 1 dan true tidak 28 class precision 96.55%.

## V. KESIMPULAN DAN SARAN

### Kesimpulan

Penerapan Data Mining Untuk Evaluasi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier berhasil dilakukan dengan aplikasi RapidMiner yang menghasilkan data pengklasifikasian evaluasi kelulusan untuk masing-masing mahasiswa. Dengan hasil akhir akurasi prediksi kelulusan mahasiswa : 93.88% dari precision 100%.

### Saran

Diharapkan dalam penelitian selanjutnya dapat dilakukan perbandingan dengan memanfaatkan metode klasifikasi lainnya seperti metode nearest neighbor atau metode C.4.5. Dengan begitu dapat diketahui kelebihan masing-masing metode dan metode mana yang lebih baik dari keakuratan nilai yang dihasilkan[4].

## DAFTAR PUSTAKA

- [1] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," vol. 7, no. 1, pp. 59–64, 2013.
- [2] H. Rohayani, "Analisis Sistem Pendukung Keputusan Dalam Memilih Program Studi Menggunakan Metode Logika Fuzzy," *J. Sist. Inf.*, vol. 5, no. Analisis Sistem Pendukung Keputusan, pp. 530–539, 2013.
- [3] S. Kasus, U. Dehasen, S. Haryati, A. Sudarsono, and E. Suryana, "IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4.5," pp. 130–138, 2015.
- [4] A. Trimanto, F. Faqih, I. M. Irfani, and S. Timur, "Penerapan Data Mining Untuk Evaluasi Status Kelulusan Mahasiswa Fakultas Teknologi Pertanian Tahun 2015 Menggunakan Algoritma Naive Bayes Classifier," 2015.
- [5] M. R. Effendi, "Akurasi Data Mining Untuk Menghasilkan Pola Kelulusan Mahasiswa dengan Metode NAIVE BAYES," pp. 101–106.
- [6] H. Rohayani, T. Mauritsius, S. Warnars, H. Leslie, and E. Abdurrachman, "Evaluation Performance Neural Network Genetic Algorithm."
- [7] A. K-means, "Penerapan data mining dalam pengelompokan kelas mahasiswa berdasarkan hasil ujian saringan masuk dengan algoritma k-means," vol. V, no. 2, 2019.
- [8] H. Rohayani, H. L. H. S. Warnars, T. Mauritsius, and E. Abdurrachman, "Wind speed forecasting in big data and machine learning: from presents, opportunities and future trends," *Commonications Math. Biol. Neurosci. J.*, vol. 3, no. 1, pp. 1–16, 2021.
- [9] Rico and H. Rohayani, "Penerapan Metode Analytical Hierarchy Process (AHP) Pada Sistem Pendukung Keputusan

- Penentuan Supplier (Studi Kasus: Toko Herifal Bags),* V-Tech, vol. 4, no. 1, pp. 1–6, 2021.
- [10] H. Rohayani and M. C. Umam, “Prediksi Penentuan Program Studi Berdasarkan Nilai Siswa dengan Algoritma Backpropagation,” *J. Inf. Syst. Res.*, vol. 3, no. 4, pp. 651–657, 2022.