

## PREDIKSI KELULUSAN MAHASISWA PROGRAM PENDIDIKAN MULTI PROFESI 1 TAHUN DENGAN METODE NAÏVE BAYES

Suhendra Anjar Dinata<sup>1</sup>, Hijrah<sup>2</sup>, Rahmat Musfekar<sup>3</sup>

<sup>1</sup>STIKOM El Rahma, Prodi Informatika, Jl. Dr. Semeru no 21G Kel. Kebon Kalapa  
Kec. Bogor Tengah Kota Bogor , 161112

<sup>2</sup>Jurusan Dakwah dan Komunikasi Islam, Sekolah Tinggi Agama Islam Negeri  
Teungku Dirundeng Meulaboh

<sup>3</sup>Prodi Pendidikan Teknologi informasi, Fakultas Tarbiyah dan Keguruan, Universitas  
Islam Negeri Ar-Raniry Banda Aceh

<sup>3</sup>[rahmat.musfekar@ar-raniry.ac.id](mailto:rahmat.musfekar@ar-raniry.ac.id)

---

### ABSTRAK

Tingkat kelulusan mahasiswa program pendidikan multi profesi satu tahun El Rahma sejak tahun akademik 2014 sampai 2016 mengalami penurunan. Kondisi tersebut sehingga mendorong pihak El Rahma melakukan evaluasi serta menyusun langkah langkah strategis terhadap permasalahan tersebut untuk meningkatkan tingkat kelulusan mahasiswa dengan cara sasaran mutu bisa tercapai. Evaluasi terhadap permasalahan program studi tersebut perlu ditingkatkan dengan membuat sebuah sistem untuk memprediksi tingkat kelulusan pada proses perkuliahan. Dengan teknik data mining khususnya dalam klasifikasi untuk prediksi dengan algoritma klasifikasi naive bayes dapat dilakukan prediksi terhadap kelulusan studi dari mahasiswa. Data training dan testing yang digunakan diambil secara random pada data master yang digunakan adalah data kelulusan tahun 2016. Algoritma naive bayes mampu menghitung perbandingan peluang antara jumlah masing-masing kriteria nilai atribut terhadap nilai hasil prediksi kelulusan mahasiswa. Dari penelitian dengan menggunakan data mahasiswa sebanyak 229 dengan jumlah 171 mahasiswa lulus dan 58 mahasiswa tidak lulus dengan menerapkan algoritma Naïve bayes didapatkan hasil bahwa tingkat akurasi menunjukkan 90,85 % hal ini menunjukkan bahwa algoritma naive bayes baik diterapkan dalam melakukan prediksi kelulusan mahasiswa adapun hasil dari klasifikasi algoritma naive bayes dapat diketahui atribut yang memberikan pengaruh pada tingkat kelulusan mahasiswa program pendidikan multi profesi satu tahun El Rahma diantaranya jenis kelamin, daerah asal, status marital, asal sekolah, status kerja, SPP bulanan, umur dan gelombang.

*Kata kunci : Prediksi kelulusan, naive bayes, kelulusan mahasiswa, data mining*

---

### 1. PENDAHULUAN

Perkembangan teknologi informasi berkembang sangat cepat dari berbagai segmen kehidupan. Banyak data yang dihasilkan dari teknologi informasi dari berbagai bidang, seperti bidang pendidikan , ekonomi, industri, ekonomi dan teknologi serta bidang lainnya. Salah satunya pada bidang pendidikan yang menghasilkan data yang banyak tentang siswa dengan proses pembelajaran. Pada penelitian ini dilakukan pengambilan data mahasiswa dan data jumlah kelulusan mahasiswa pada STIKOM El Rahma, guna menggali informasi yang yang tidak terlihat agar dapat diketahui agar dapat berguna bagi STIKOM El Rahma . Proses Pengolahan data mahasiswa yang dilakukan untuk dapat mengetahui informasi penting merupakan pengetahuan baru (*Knowledge Discovery*),[1] sebagai contoh data yang digali

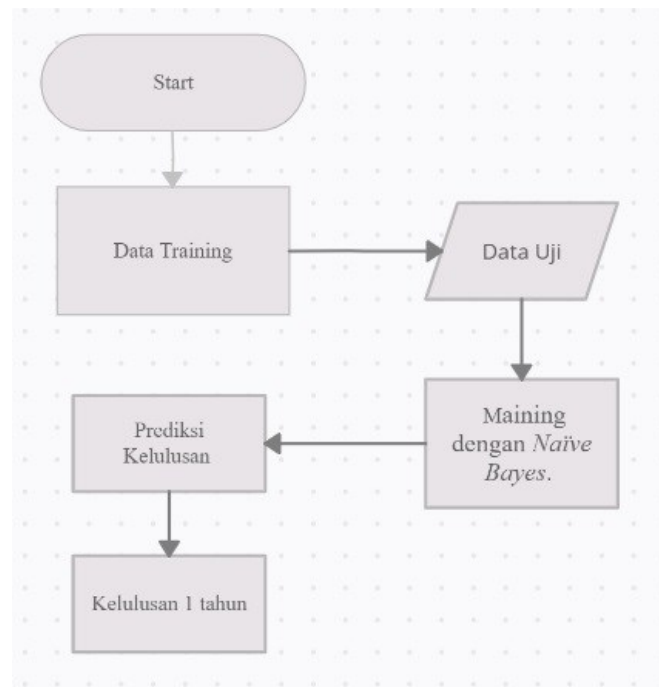
untuk mengetahui informasi terkait pengklasifikasian data mahasiswa yang ditinjau dari profil dan data akademik mahasiswa untuk diperoleh informasi baru. Dengan informasi baru tersebut dapat memberikan pengetahuan kepada pihak institusi agar dapat melakukan klasifikasi tentang tingkat kelulusan mahasiswa dalam menentukan strategi untuk meningkatkan kelulusan dan kualitas mahasiswa pada tahun - tahun berikutnya.

El Rahma Education Centre adalah salah satu satuan pendidikan yang memiliki visi terwujudnya kultur lembaga sebagai wahana yang kondusif untuk memberdayakan peserta didik yang berkembang menjadi insan yang berakhlak mulia, cerdas, kreatif inovatif dan kompetitif dalam dunia usaha dan dunia industri. Saat ini menghadapi penurunan jumlah mahasiswa sebagaimana data jumlah kelulusan mahasiswa program pendidikan multi profesi pada 2016 menunjukkan penurunan dari 229 mahasiswa yang kuliah hanya 171 mahasiswa atau sekitar 75% yang lulus hal ini dijadikan sebagai rujukan untuk melakukan proses klasifikasi, maka penulis akan meneliti tentang klasifikasi prediksi kelulusan mahasiswa berdasarkan data mahasiswa program pendidikan multi profesi satu tahun angkatan 2016. Untuk membantu menyelesaikan permasalahan yang dihadapi El Rahma Education Centre dalam hal penurunan jumlah mahasiswa yang lulus. Penelitian ini akan melakukan pengklasifikasian menggunakan data mahasiswa Program Pendidikan Multi Profesi satu Tahun El Rahma Education Centre angkatan 2016. Atribut yang akan digunakan dalam melakukan klasifikasi prediksi kelulusan angkatan 2016 adalah jenis kelamin, daerah asal, status marital, asal sekolah, status kerja, SPP bulanan, umur dan gelombang. Oleh karena itu, penelitian ini ditunjukkan untuk menguji algoritma klasifikasi yang tepat dan membuat sebuah sistem yang dapat memprediksi kelulusan mahasiswa program pendidikan multi profesi 1 tahun El Rahma dengan mengevaluasi pada trimester ke dua, apakah berpeluang menyelesaikan kuliah atau tidak dengan menggunakan salah satu algoritma data mining yaitu *Naïve Bayes*.

## 2. METODOLOGI PENELITIAN

Metode penelitian ini dilakukan dengan maksud untuk menerapkan metode Naive Bayes classification dalam prediksi kelulusan mahasiswa program pendidikan multi profesi 1 tahun, dalam penerapan metode ini akan diklasifikasikan berbagai informasi terkait kebutuhan data atau informasi-informasi yang dapat membantu mengklasifikasi dan memprediksi kelulusan mahasiswa. pada dasarnya metode tersebut akan dapat memperoleh informasi baru berdasarkan informasi- informasi yang ada sebelumnya. Penerapan metodologi pada penelitian ini dapat dilihat pada Gambar 1.

Pada bagan alir diatas dapat dilihat tahapan peneliti dalam menggali informasi dari data yang ada dengan mengumpulkan data training yang terdiri dari data attribute yang akan di uji dengan menggunakan perangkat lunak pengujian seperti rapidminer, setelah dilakukan pengujian dengan data uji maka data tersebut akan di mining dengan algoritma naïve Bayesian Adapun langkah – langkah yang dilakukan pada tahap ini adalah dengan menghitung jumlah *class* serta menghitung jumlah kasus yang sama dengan *class* setelah itu semua nilai dengan data *x* akan dikalikan sesuai dengan *class* untuk memperoleh klasifikasi yang sesuai. Dari hasil klasifikasi akan diperoleh hasil prediksi yang dapat dikelompokkan berdasarkan peluang kelulusan mahasiswa profesi dengan durasi 1 tahun. Hasil pengujian dengan metode eksperimen yang dilakukan tersebut juga merupakan kondisi yang sesuai dengan kebutuhan serta permasalahan yang di hadapi pada penelitian ini, sehingga diharapkan hasil berupa informasi yang relevan agar dapat membantu manajemen El Rahma Education Centre dalam menganalisa serta mengambil keputusan dalam prediksi kelulusan Mahasiswa.



**Gambar 1. Bagan Alir Penelitian**

### 3. TEKNIK ANALISIS

Berdasarkan arsitektur data mining CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai berikut :[9]

#### a. **Busines Understanding**

Adanya kesulitan manajemen dalam melakukan prediksi kelulusan mahasiswa program pendidikan multi profesi satu tahun El Rahma Education Centre untuk penanganan permasalahan mahasiswa. Maka dalam penelitian ini akan dilakukan pembuatan prototype prediksi kelulusan Mahasiswa.

#### b. **Data Understanding**

Untuk pemilihan prediksi kelulusan mahasiswa, di dapat data dari LKP El Rahma Education Centre Bogor yang terdiri dari 13 Attribut dimana 21 attribut predictor dan 1 attribut hasil. Attribut-attribut yang menjadi parameter dipaparkan pada Tabel 1.

**Tabel 1**  
**Attribut dan Nilai Kategori**

No	Attribut	Nilai	Keterangan
1	Gelombang	1	1,2,3
2	Jurusan	1	ITK,KBI, ADM, KDG, AKT, KTO
3	Jenis kelamin	1	Laki-laki
		2	Perempuan
4	Umur	1	Remaja
		2	Dewasa
5	Asal daerah	1	Jawa Barat
		2	Jawa Tengah
		3	Jawa Timur
		4	Luar Jawa
		5	Banten
		6	Bogor
6	Asal Sekolah	1	Negeri
		2	Swasta
7	Status	1	Menikah
		2	Single
8	Satus Kerja	1	Kerja
		2	Belum Kerja
9	SPP Bulanan	1	Ringan
		2	Sedang
		3	Berat
10	Biaya Kuliah	1	Orang Tua
		2	Mandiri
11	Nilai trimester 1	Nilai trimester 1	Cumlaude
			Sangat Memuaskan
			Memuaskan
12	Nilai trimester 2	Nilai trimester 2	Tidak Lulus
			Cumlaude
			Sangat Memuaskan
13	Status Lulus	TRUE	Memuaskan
			FALSE
			Tidak Lulus

c.  
d.

**e. Klasifikasi Kategori Atribut**

Dalam menentukan atribut agar dapat diolah sehingga menghasilkan informasi maka dari data primer yang tersedia penulis melakukan pengelompokan kategori seperti Umur, SPP Bulanan, dan Nilai trimester, seperti yang ditampilkan pada Tabel 2, Tabel 3, dan Tabel 4.

**Tabel 2**  
**Pengelompokan Umur**

No	Umur	Keterangan
1	<25	Remaja
2	>25	Dewasa

**Tabel 3**  
**Pengelompokan SPP Bulanan**

No	Jumlah SPP	Keterangan
1	< 400.000	Ringan
2	450.000	Sedang
3	>500.000	Berat

**Tabel 4**  
**Pengelompokan Nilai Trimester**

No	Nilai	Keterangan
1	>3.5	Cumlaude
2	2.75-3.45	Sangat Memuaskan
3	2.0-2.70	Memuaskan
4	< 2.0	Tidak Lulus

#### **f. Data Preparation**

Dalam penelitian ini, diperlukan proses validasi agar menemukan serta mengkonversi data tersebut agar dapat digunakan pada algoritma data mining. Hal ini dikarenakan ada beberapa variable yang kurang lengkap, tidak konsisten serta tidak rapi yang tidak bisa ditangani secara sistematis oleh algoritma data mining. Pada penelitian ini digunakan data yang mempunyai nilai pada setiap variable. Data pada setiap dataset yang tidak mempunyai nilai maka data tersebut akan dihapus dan tidak akan digunakan. Hal ini dilakukan supaya data yang digunakan menjadi valid dan akurasi mempunyai nilai yang benar.

#### **g. Modelling Fase**

Tahap ini merupakan tahap Learning dikarenakan pada tahap ini dilakukan data training klasifikasi oleh model yang kemudian dapat menghasilkan beberapa aturan. Pada penelitian ini pembuatan model dengan menggunakan algoritma *Naïve Bayes*.

#### **h. Evaluation Fase**

Pada tahap evaluasi di lakukan pengujian terhadap model-model agar mendapatkan informasi model yang akurat. Evaluasi dan validasi pada tahap ini menggunakan *precision, recall, f-measure* dan akurasi.[10]

### i. Deployment

Setelah pembentukan model serta melakukan analisa dan penguatan pada tahap sebelumnya, kemudian pada tahapan ini menerapkan model yang sangat akurat untuk penentuan klasifikasi prediksi kelulusan Mahasiswa.

## 4. HASIL DAN PEMBAHASAN

### a. Pengumpulan Data

Pada penelitian ini dilakukan proses pengumpulan data dan penyeleksian data dimana data yang sudah dikumpulkan dari bagian akademik Program Pendidikan Multi Profesi satu Tahun El Rahma akan melakukan proses validasi agar mendapatkan dan mengkonversi data untuk dapat dipergunakan pada algoritma data mining serta untuk memperoleh hasil akurasi serta performa yang baik adapun proses validasi yang digunakan pada tahap ini yakni menghapus data yang tidak lengkap atau yang tidak memiliki nilai (null).

Setelah proses penghapusan data yang tidak memiliki nilai (null) maka langkah selanjutnya adalah penyeleksian atribut, langkah tersebut merupakan proses pemilihan atribut yang dibutuhkan dalam kalasifikasi prediksi kelulusan mahasiswa, setelah dilakukan proses pemilihan attribute maka attribute yang akan digunakan yaitu: *Gelombang Masuk, Jenis Kelamin, Jurusan, Nilai IP 1, Nilai IP 2, Kuantitas Umur, Asal Sekolah, Status Kerja, SPP Bulanan, Daerah Asal, Status Marital, Sumber Biaya Kuliah, Status Kelulusan*

No	Nama	GEL	JKL	JUR	IP.1	IP.2	Umur	Sekolah	Status Kerja	SPP	Daerah ASAL
1	ANGGA KUSUMAH	1	Laki-laki	ADM	3.43	3.55	18 Th	NEGERI	BELUM KERJA	450000	KAB. SUKABUMI
2	DEDE MUYANA	1	Laki-laki	ADM	3.06	3.24	22 Th	NEGERI	BELUM KERJA	440000	KAB. SUKABUMI
3	FARIDI	1	Laki-laki	ADM	3.67	3.66	19 Th	NEGERI	BELUM KERJA	455000	KAB. LEBAK
4	M. FAHRUL MARDIANA	1	Laki-laki	ADM	3.27	3.38	19 Th	SWASTA	BELUM KERJA	410000	KAB. BOGOR
5	NANDAR MUHAMMAD	1	Laki-laki	ADM	3.05	3.18	19 Th	NEGERI	BELUM KERJA	420000	KAB. SUKABUMI
6	RIFAL M ISWANDI	1	Laki-laki	ADM	2.95	3.10	19 Th	NEGERI	BELUM KERJA	455000	KAB. LEBAK
7	ADELIA PRATIWI	1	Perempuan	ADM	3.58	3.56	17 Th	SWASTA	BELUM KERJA	460000	KAB. PANDEGLANG
8	DEDE EKA PERTINI	1	Perempuan	ADM	3.12	3.07	20 Th	SWASTA	BELUM KERJA	490000	KAB. BOGOR
9	ENENG LESTARI	1	Perempuan	ADM	3.23	3.27	19 Th	SWASTA	BELUM KERJA	465000	KAB. BOGOR
10	ENENG YULI Nuraeni	1	Perempuan	ADM	3.42	3.46	19 Th	SWASTA	BELUM KERJA	475000	KOTA SUKABUMI
11	ERNAWATI	1	Perempuan	ADM	3.32	3.42	20 Th	SWASTA	BELUM KERJA	465000	KAB. LEBAK
12	ESY SAWITRI	1	Perempuan	ADM	3.65	3.70	26 Th	NEGERI	BELUM KERJA	455000	KAB. CIANJUR
13	GUSTUR LUTFIAH	1	Perempuan	ADM	3.20	3.29	18 Th	NEGERI	BELUM KERJA	417500	KAB. BOGOR
14	IIS ISMAYANTI	1	Perempuan	ADM	3.89	3.90	19 Th	SWASTA	BELUM KERJA	425000	KAB. BOGOR
15	MIA MEIDINA	1	Perempuan	ADM	3.37	3.26	18 Th	SWASTA	BELUM KERJA	460000	KAB. PANDEGLANG
16	MILA RAHAYU	1	Perempuan	ADM	3.26	3.32	19 Th	NEGERI	BELUM KERJA	470000	KOTA BOGOR
17	NANDA RIZKIA	1	Perempuan	ADM	3.27	3.22	18 Th	SWASTA	BELUM KERJA	475000	KAB. BOGOR
18	NONI ASIH	1	Perempuan	ADM	3.45	3.33	19 Th	SWASTA	BELUM KERJA	465000	KAB. BOGOR

**Gambar 2**  
**Potongan Sampel Dataset (18 record)**

Sebelum digunakan dan diproses maka data-data tersebut akan disiapkan (preprocessing) terlebih dahulu sesuai dengan kebutuhan proses klasifikasi, sehingga nantinya akan memiliki kelas binominal atau polynominal sesuai aturan yang telah dibuat dan berdasarkan nilai datanya. Pada Tabel 5 ditunjukkan pembagian variable dan kelas data yang akan digunakan pada proses analisis data mining klasifikasi.

**Tabel 5**  
**Pembagian Variable dan Kelas Data**

Variable	Nama Field	Jenis Kelas Data	Kelas Data yang Digunakan
Y	Status Kelulusan	Binominal	Lulus, Tidak Lulus
X1	Gelombang Masuk	Polynomial	Satu, dua , tiga
X2	Jenis Kelamin	Binominal	Laki-laki, Perempuan
X3	Jurusan	Polynomial	ADM, ITK, KDG, KTO, AKT, KBI
X4	Nilai IP 1	Polynomial	Cumlaude, Sangat Memuaskan, Memuaskan, Tidak Lulus
X5	Nilai IPK 2	Polynomial	Cumlaude, Sangat Memuaskan, Memuaskan, Tidak Lulus
X6	Umur	Binominal	Remaja , Dewasa
X7	Asal Sekolah	Binominal	Negeri, Swasta
X8	Status Kerja	Binominal	Kerja, Belum Kerja
X9	Biaya SPP	Polynomial	Ringan , Sedang, Berat
X10	Daerah Asal	Polynomial	Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Luar Jawa
X11	Marital Status	Binominal	Menikah, Belum Menikah
X12	Biaya Kuliah	Binominal	Mandiri, Orang tua

Adapun data hasil dari *preprocessing* sesuai dengan jenis kelas data atribut dapat dilihat pada Gambar 2.

No	Nama	GEL	JKL	JUR	IPK1	IPK2	Umur
1	ANGGA KUSUMAH	satu	Laki-laki	ADM	Sangat Memuaskan	Cumlaude	Remaja
2	DEDE MULYANA	satu	Laki-laki	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
3	FARIDI	satu	Laki-laki	ADM	Cumlaude	Cumlaude	Remaja
4	M. FAHRUL MARDIANA	satu	Laki-laki	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
5	NANDAR MUHARAM	satu	Laki-laki	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
6	RIFAL M ISWANDI	satu	Laki-laki	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
7	ADELIA PRATIWI	satu	Perempuan	ADM	Cumlaude	Cumlaude	Remaja
8	DEDE EKA PERTIWI	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
9	ENENG LESTARI	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
10	ENENG YULI Nurani	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
11	ERNAWATI	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
12	ESY SAWFITRI	satu	Perempuan	ADM	Cumlaude	Cumlaude	Remaja
13	GUSTUR LUTFIAH	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
14	IIS ISMAYANTI	satu	Perempuan	ADM	Cumlaude	Cumlaude	Remaja
15	MIA MEIDINA	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
16	MILA RAHAYU	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
17	NANDA RIZKIA	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
18	NONI ASIH	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
19	SALMA RAANIYAH	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
20	TRIA SUKMA WAHYUNINGSIH	satu	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
21	MOH FUAD BAWAZIR	Dua	Laki-laki	ADM	Cumlaude	Cumlaude	Remaja
22	RIZKI ANUGRAH	Dua	Laki-laki	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
23	SANDRA DWI LESMANA	Dua	Laki-laki	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
24	SOFIAN SYIDIK	Dua	Laki-laki	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja
25	ADHA DIATUL HARTINA SUHANDA	Dua	Perempuan	ADM	Cumlaude	Cumlaude	Remaja
26	HERA NURUL RAHMAWATI	Dua	Perempuan	ADM	Sangat Memuaskan	Sangat Memuaskan	Remaja

**Gambar 2**  
**Potongan Hasil Preprocessing (26record)**

**b. Klasifikasi Data dengan Naïve Bayes**

Pada tahap penelitian ini dilakukan perhitungan manual menggunakan algoritma Naïve Bayes dengan 229 sampel data, adapun langkah – langkah klasifikasi data tersebut dengan menggunakan algoritma naïve bayes sebagai berikut :

1. Data training dibutuhkan serta data yang digunakan pada penelitian ini adalah 229 sampel data training dan data testing.
2. Setelah data disiapkan, kemudian melakukan proses penghitungan jumlah data Mahasiswa yang Lulus dan Tidak Lulus berdasarkan acuan nilai atribut.
3. Setelah melakukan proses penghitungan tersebut, kemudian melakukan penghitungan nilai prior total dimana yang diketahui pada proses penghitungan sebelumnya jumlah mahasiswa yang lulus berjumlah 171 dan mahasiswa yang tidak lulus berjumlah 58
4. Setelah dilakukan penghitungan nilai priorpropabilitas , maka dilakukan juga penghitungan nilai gain untuk masing – masing atribut sehingga diperoleh nilai prior probabilitas pada masing-masing atribut sebagaimana pada Tabel 6.

Untuk memperoleh nilai tertinggi maka nilai gain yang diperoleh akan dijadikan sebagai akar maka harus ditentukan oleh hasil perhitungan nilai prior probabilitas dari semua atribut.



**Tabel 6**  
**Hasil Perhitungan Nilai Prior Probabilitas**

Atribut		Jml Kasus	Lulus	Tidak Lulus	Prior Lulus	Prior Tidak Lulus
Total		229	171	58	0.747	0.253
Gel	1	91	73	18	0.319	0.079
	2	53	45	8	0.197	0.035
	3	85	53	32	0.231	0.140
Jur	ITK	104	73	31	0.319	0.135
	ADM	71	55	16	0.240	0.070
	KDG	25	17	8	0.074	0.035
	KBI	9	8	1	0.035	0.004
	AKT	17	16	1	0.070	0.004
	KTO	3	2	1	0.009	0.004
Jkl	Male	152	48	104	0.210	0.454
	Female	77	22	55	0.096	0.240
N-IPK-1	Cum	48	107	7	0.467	0.031
	SM	116	46	2	0.201	0.009
	M	39	16	23	0.070	0.100
	TL	26	26	0	0.114	0.000
N-IPK-2	Cum	43	41	1	0.179	0.004
	SM	120	114	5	0.498	0.022
	M	32	14	19	0.061	0.083
	TL	34	34	0	0.148	0.000
Umur	Remaja	221	121	33	0.528	0.144
	Dewasa	8	7	1	0.031	0.004
Asal Sekolah	Swasta	146	102	44	0.445	0.192
	Negeri	83	69	14	0.301	0.061
Status Kerja	Bekerja	91	53	38	0.231	0.166
	Belum Kerja	138	118	20	0.515	0.087
SPP Bulanan	Berat	36	20	16	0.087	0.070
	Sedang	125	91	34	0.397	0.148
	Ringan	65	57	8	0.249	0.035
	Sangat Ringan	3	3	0	0.013	0.000
Daerah Asal	Jabar	166	120	46	0.524	0.201
	Jateng	6	4	2	0.017	0.009
	Jatim	1	1	0	0.004	
	Banten	36	32	4	0.140	0.017
	L. Jawa	18	14	4	0.061	0.017
Status Marital	Meried	22	10	12	0.044	0.052
	Single	207	161	46	0.703	0.201
Biaya Kuliah	Mandiri	77	51	26	0.223	0.114
	Ortu	152	120	32	0.524	0.140

### c. Perhitungan Algoritma Naïve Bayes

Proses untuk melakukan perhitungan kelulusan mahasiswa lulus menggunakan data yang telah dipilih. Dari data training jumlah mahasiswa lulus sebanyak 171 dan jumlah tidak lulus 58 sehingga jumlah data yang dipakai adalah 229 data. Maka Langkah dalam melakukan perhitungan algoritma naive bayes sebagai berikut :

1. Menghitung jumlah kelas
2. Menghitung jumlah kasus yang sama dengan class yang sama
3. Kalikan semua hasil variable LULUS & TIDAK LULUS
4. Bandingkan hasil class LULUS & TIDAK LULUS

Kemudian dapat dihitung dengan menggunakan formula Naive Bayes Clasification, adapun cara kerjanya sebagai berikut :

Untuk masalah klasifikasi, yang dihitung adalah  $P(H|X)$ , yaitu peluang bahwa hipotesa benar (valid) untuk data sample X yang diamati :

Dimana :

X adalah data sample dengan kelas (label) yang tidak diketahui

H merupakan hipotesa bahwa X adalah data dengan (label) C.

$P(H)$  adalah peluang dari hipotesa H

$P(X)$  adalah peluang data sampel yang diamati

$P(X|H)$  adalah peluang data sampel X, bila diasumsikan bahwa hipotesa benar (valid)

Jadi Rumusnya adalah :

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

Jumlah kelas :

$$P(\text{Lulus}) = 171/229 = 0,746$$

$$P(\text{Tidak Lulus}) = 58/229 = 0,253$$

$$P(\text{Gel}=1|\text{Lulus}) = 73/91 = 0,318$$

$$P(\text{Gel}=1|\text{Tidak Lulus}) = 18/91 = 0,078$$

$$P(\text{Gel}=2|\text{Lulus}) = 45/53 = 0,196$$

$$P(\text{Gel}=2|\text{Tidak Lulus}) = 8/53 = 0,034$$

$$P(\text{Gel}=3|\text{Lulus}) = 53/85 = 0,231$$

$$P(\text{Gel}=3|\text{Tidak Lulus}) = 32/85 = 0,139$$

$$P(\text{Jur} = \text{ADM} | \text{Lulus}) = 55/71 = 0,240$$

$$P(\text{Jur} = \text{ADM} | \text{Tidak Lulus}) = 16/71 = 0,069$$

$$P(\text{Jur} = \text{ITK} | \text{Lulus}) = 73/104 = 0,318$$

$$P(\text{Jur} = \text{ITK} | \text{Tidak Lulus}) = 31/104 = 0,135$$

$$P(\text{Jur} = \text{KDG} | \text{Lulus}) = 17/25 = 0,074$$

$$P(\text{Jur} = \text{KDG} | \text{Tidak Lulus}) = 8/25 = 0,034$$

$$P(\text{Jur} = \text{KBI} | \text{Lulus}) = 8/9 = 0,034$$

$$P(\text{Jur} = \text{KBI} | \text{Tidak Lulus}) = 1/9 = 0,004$$

$$P(\text{Jur} = \text{AKT} | \text{Lulus}) = 16/17 = 0,069$$

$$P(\text{Jur} = \text{AKT} | \text{Tidak Lulus}) = 1/17 = 0,004$$

$$P(\text{Jur} = \text{KTO} | \text{Lulus}) = 2/3 = 0,008$$

$$P(\text{Jur} = \text{KTO} | \text{Tidak Lulus}) = 1/3 = 0,004$$

$$P(\text{Jkl} = \text{Laki-Laki} | \text{Lulus}) = 48/152 = 0,209$$

$$P(\text{Jkl} = \text{Laki-Laki} | \text{Tidak Lulus}) = 104/152 = 0,454$$

$$P(\text{Jkl} = \text{Perempuan} | \text{Lulus}) = 22/77 = 0,096$$

$$P(\text{Jkl} = \text{Perempuan} | \text{Tidak Lulus}) = 55/77 = 0,204$$

Kemudian :

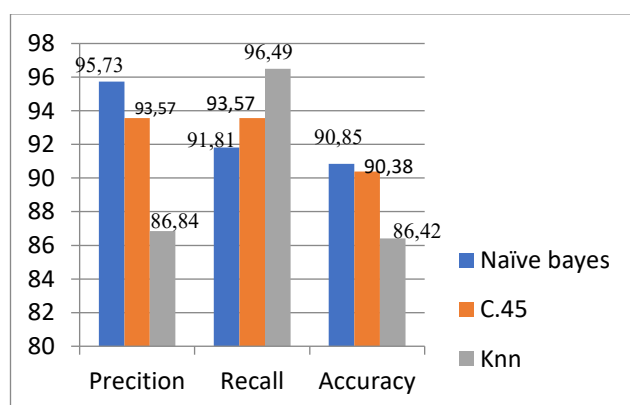
$$\begin{aligned}
 &P(\text{Lulus}) P(\text{Gel}=1|\text{Lulus}) P(\text{Gel}=3|\text{Lulus})P(\text{Jur} = \text{ADM} | \text{Lulus})P(\text{Jur}= \text{ITK} | \\
 &\text{Lulus})P(\text{Jur}=\text{KDG} | \text{Lulus})P(\text{Jur}=\text{KBI} | \text{Lulus})P(\text{Jur}=\text{AKT} | \text{Lulus})P(\text{Jur}=\text{KTO} | \\
 &\text{Lulus})P(\text{Jkl} = \text{Laki-Laki} | \text{Lulus})P(\text{Jkl} = \text{Perempuan} | \text{Lulus}) \\
 &=0,746 \times 0,318 \times 0,196 \times 0,231 \times 0,240 \times 0,318 \times 0.074 \times 0.034 \times 0.069 \times 0.008 \times 0.209 \\
 &\times 0.096 \\
 &=2.28423
 \end{aligned}$$

$$\begin{aligned}
 &P(\text{Tidak Lulus}) P(\text{Gel}=1|\text{Tidak Lulus}) P(\text{Gel}=3|\text{Tidak Lulus})P(\text{Jur} = \text{ADM} | \text{Tidak} \\
 &\text{Lulus})P(\text{Jur}= \text{ITK} | \text{Tidak Lulus})P(\text{Jur}=\text{KDG} | \text{Tidak Lulus})P(\text{Jur}=\text{KBI} | \text{Tidak} \\
 &\text{Lulus})P(\text{Jur}=\text{AKT} | \text{Tidak Lulus})P(\text{Jur}=\text{KTO} | \text{Tidak Lulus})P(\text{Jkl} = \text{Laki-Laki} | \text{Tidak} \\
 &\text{Lulus})P(\text{Jkl} = \text{Perempuan} | \text{Tidak Lulus}) \\
 &=0,253 \times 0,078 \times 0,034 \times 0,139 \times 0,069 \times 0,135 \times 0.034 \times 0.004 \times 0.004 \times 0.454 \times 0.204 \\
 &= 4.37700074
 \end{aligned}$$

#### d. Pengujian dengan Tools

Pengujian pada penelitian ini dilakukan dengan tools rapid miner dan tiga metode data mining, proses pengujian dilakukan untuk menganalisa dan membuktikan bahwa metode data mining yang digunakan yakni algoritma Naïve Bayes adalah metode yang paling baik jika diterapkan dalam pemrosesan data untuk prediksi kelulusan mahasiswa. Berikut adalah proses pengujian yang dilakukan dengan *tools rapid miner*.

Pada penelitian ini dilakukan pencarian tingkat akurasi yang paling baik dengan membandingkan tingkat akurasi yang dihasilkan oleh beberapa metode data mining, metode yang digunakan sebagai pembanding yaitu Algoritma *Naïve Bayes Classifier*, C4.5 dan K-NN. Pengujian dilakukan dengan menggunakan *Tools Rapid Miner* yang berbeda dengan tujuan untuk lebih meningkatkan kepastian akan tingkat akurasi yang diperoleh. Adapun *tools* yang digunakan adalah Rapid Miner 5.3. Dari hasil pengujian dengan menggunakan *tools* tersebut maka diperoleh hasil tingkat akurasi yang paling baik yakni algoritma Naïve Bayes. Grafik pada Gambar 3 adalah rangkuman dari hasil yang telah dilakukan.



**Gambar 3**  
Hasil Pengujian dengan *Rapid Miner*

#### e. Penerapan Algoritma Naïve Bayes dalam Sistem

Setelah dilakukan berbagai pengujian dan analisa serta dari hasil evaluasi dan validasi maka metode data mining yang memiliki akurasi dan performa lebih tinggi akan diterapkan pada aplikasi prediksi kelulusan mahasiswa program pendidikan multi

profesi 1 tahun dengan menggunakan bahasa pemograman PHP dan MySQL sebagai databasenya. Dari pengujian dan analisa yang dilakukan algoritma Naïve Bayes memiliki nilai persentase yang lebih baik dibandingkan metode data mining lainnya yakni C45 dan Knn.

Pengujian dilakukan dengan confusion matrix yang terdiri dari precision, recall, f-measure dan akurasi dilakukan pada dataset yang diolah dengan menggunakan algoritma Naïve Bayes. Pengujian confusion matrix untuk dataset yang diolah menggunakan algoritma Naïve Bayes untuk nilai akurasi dapat dilihat pada Gambar 4.

accuracy: 90.85% +/- 4.52% (mikro: 90.83%)			
	true LULUS	true TIDAK LULUS	class precision
pred. LULUS	157	7	95.73%
pred. TIDAK LULUS	14	51	78.46%
class recall	91.81%	87.93%	

**Gambar 4**  
**Hasil Pengujian Algoritma Naïve Bayes**

Pada lamiran 1 berisi data training sebanyak 229 data, dimana jumlah yang lulus sebanyak 171 dan jumlah tidak lulus sebanyak 58. Pada lampiran 2 berisi data tsting sebanyak 50 data. Mengapa hasil *precision*, *recall* dan *f-measure* berbeda jauh dengan akurasi, karena terlalu banyak data baik di data training dan data testing yang lulus tidak lulus.

Precision menggambarkan proporsi kasus dari hasil yang lulus dianggap benar. Recall menggambarkan proporsi kasus dari hasil yang tepat waktu diidentifikasi dengan benar. *F-measure* menggambarkan evaluasi algoritma yang mengkombinasikan recall dan precision. Akurasi menggambarkan perbandingan kasus yang diidentifikasi secara benar (Lulus dan Tidak Lulus) dengan semua kasus.

Gambar 5 menunjukkan presentasi *precision*, *recal*, *f-measure* dan Akurasi dari hasil pengujia algoritma Naïve bayes dari aplikasi prototype.

Tabel Penilaian Naive Bayes		
	Didentifikasi Tidak oleh Naive Bayes	Didentifikasi Ya oleh Naive Bayes
Keputusan Asli: Tidak = 58	50	8
Keputusan Asli: Ya = 171	20	161

1. Precision =  $161 / (8 + 161) * 100\% = 95.27\%$   
 2. Recall =  $161 / (10 + 161) * 100\% = 94.15\%$   
 3. Accuracy =  $(50 + 161) / (171 + 58) * 100\% = 92.14\%$   
 4. F-Measure =  $((95.27 * 94.15) / ((95.27 + 94.15) * 2)) = 94.71\%$

**Gambar 5**  
**Hasil Pengujian dari Aplikasi Prototipe**

#### f. Evaluasi Model dengan *Confusion Matrix*

Penelitian ini melakukan proses evaluasi model dengan menggunakan *confusion matrix*, dengan proses tersebut akan terbentuk matrix yang terdiri dari *accuracy*, *precision* dan *recall*. Dengan data training yang sudah disiapkan ke dalam confusion matrix maka

pengujian pertama sekali dilakukan dengan perolehan nilai confusion matrix yang diperoleh dari *tools rapid miner* seperti pada tabel 4 sebelumnya.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{51+157}{51+157+14+7} = \frac{208}{229} = 0,90829 = 90,83 \%$$

$$Precision = \frac{TP}{(TP+FP)} = \frac{51}{51+14} = \frac{51}{65} = 0,7846 = 78,46 \%$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{51}{51+7} = \frac{51}{58} = 0,8793 = 87,93 \%$$

Setelah dilakukan pencarian terhadap nilai *confusion matrix* dengan *tools rapid miner* maka diperoleh hasil *Accuracy* sebesar 90,85 %, *Precision* 78,46 % dan *Recall* 87,93 %.

## 5. KESIMPULAN

Dari hasil proses penelitian yang dilakukan dapat disimpulkan bahwa dengan menggunakan metode data mining dapat diperoleh informasi prediksi kelulusan mahasiswa dengan persentase ketepatan record data yang diklasifikasi secara benar atau nilai *accuracy* sebesar 90,85% dengan nilai data yang memiliki kesalahan atau nilai rendah sebesar 9,15%

Simpulan lainnya yang diperoleh dari penelitian ini adalah prediksi kelulusan mahasiswa dari metode data mining yang telah ditentukan, bahwa algoritma *Naïve Bayes* memiliki kinerja (*precision*, *recall*, dan *accuracy*) yang lebih baik dibandingkan dengan *C4.5*, dan *k-nearest neighbors* sehingga prediksi kelulusan mahasiswa dengan studi kasus pada penelitian ini dapat dianggap baik.

Dari penelitian data mining dengan menggunakan algoritma *Naïve Bayes* diperoleh informasi atribut yang memberikan pengaruh kelulusan mahasiswa dengan detail data yang digunakan 229 data 171 mahasiswa lulus dan 58 mahasiswa tidak lulus maka atribut yang berpengaruh dari data tersebut adalah *jdj* : jenis kelamin, daerah asal, status marital, asal sekolah, status kerja, SPP bulanan, umur dan gelombang.

## DAFTAR PUSTAKA

- [1] Yuda S, (2013) Data Mining Menggunakan Algoritma *Naïve Bayes* Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro Jurnal sematik 2013
- [2] Fajar, Ardi S., 2104, 'Kajian Penerapan Model *Naïve Bayes* Dan *C4.5* Dalam Prediksi Penawaran Produk pada Bank XYZ Tbk', Tesis M.Kom, Universitas Budi Luh
- [3] Larose, Daniel T, 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, John Willey & Sons. Inc
- [4] Jefri, 2013, "Implementasi Algoritma *C4.5* Dalam Aplikasi Untuk Memprediksi Jumlah Mahasiswa Yang Mengulang Mata Kuliah Di STMIK AMIKOM Yogyakarta", Yogyakarta
- [5] Kusriani, Luthfi, E.T. (2009). "Algoritma Data Mining", Andi Offset. Surabaya.
- [6] Ogor, E. N. (2007). Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques. Fourth Congress of Electronics, Robotics and Automotive Mechanics
- [7] Han, J., & Kamber, M. (2006). *Data mining concepts and techniques*. Boston: Elsevier

- [8] Fatmawati, (2016) Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes Jurnal Techno Nusa Mandiri Vol. XIII, No. 1 Maret 2016
- [9] MacLennan, Jamie. Tang, ZhaoHui. CrivatBogdan. Data Mining with Microsoft SQL Server 2008, Wiley Publishing Inc, Indiana Polis, Indiana. 2009
- [10] Arief Jananto, Januari 2013, Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa, Jurnal Teknologi Informasi DINAMIK Volume 18, No.1
- [11] Borglet, C, 2003, Finding Association Rules with Apriori Algorithm, <http://www.fuzzy.cs.uniagdeburg.de/~borglet/apriori.pdf>, diakses tgl 23 Februari 2007.