



---

## KLASIFIKASI KETEPATAN WAKTU KELULUSAN MAHASISWA DENGAN METODE NAÏVE BAYES

Oleh

Tias Mugi Rahayu<sup>1)</sup>, Besse Arnawisuda Ningsi<sup>2)</sup>, Isnurani<sup>3)</sup>, Irvana Arofah<sup>4)</sup>

<sup>1,2,3,4</sup>Universitas Pamulang

Email: <sup>1</sup>[tiasmugi45@gmail.com](mailto:tiasmugi45@gmail.com), <sup>2</sup>[dosen00205@unpam.ac.id](mailto:dosen00205@unpam.ac.id), <sup>3</sup>[dosen01193@unpam.ac.id](mailto:dosen01193@unpam.ac.id) & <sup>4</sup>[dosen00351@unpam.ac.id](mailto:dosen00351@unpam.ac.id)

### Abstrak

Ketepatan waktu kelulusan mahasiswa merupakan salah satu tujuan untuk menyelesaikan pendidikan mahasiswa di perguruan tinggi dengan status tepat waktu atau tamat. Butuh waktu  $\leq 4$  tahun untuk mencapai tujuan kelulusan tepat waktu dengan gelar sarjana, namun pada kenyataannya masih terdapat kasus mahasiswa yang tidak lulus tepat waktu. Penelitian ini bertujuan untuk mengetahui klasifikasi ketepatan waktu kelulusan mahasiswa Fakultas Ekonomi Universitas Pamulang tahun ajaran 2018/2019. Algoritma yang digunakan dalam klasifikasi data adalah klasifikasi Naïve Bayes. Klasifikasi Naive Bayesian adalah teknik klasifikasi data yang menggunakan teori dan statistik probabilitas untuk memprediksi probabilitas masa depan berdasarkan data masa lalu. Atribut yang digunakan dalam metode klasifikasi Naïve Bayes adalah ketepatan waktu kelulusan siswa, daerah asal, jenis kelamin, profesi orang tua, jenis sekolah, program studi dan predikat nilai rata-rata kumulatif (IPK). Hasil penelitian menunjukkan bahwa data kelulusan mahasiswa S1 Fakultas Ekonomi Universitas Pamulang tahun ajaran 2018/2019 sebanyak 61,9% yang mampu menyelesaikan pendidikan tepat waktu dengan tingkat akurasi sebesar 69,33%.

**Kata Kunci:** Ketepatan Waktu Kelulusan, Klasifikasi & Naïve Bayes

### PENDAHULUAN

Salah satu peran yang berguna dalam mendukung kemajuan suatu negara adalah bidang pendidikan. Pendidikan yang berkualitas akan menghasilkan generasi yang berkualitas. Salah satu jenjang pendidikan yang harus mengutamakan mutu adalah pendidikan tinggi. Setiap perguruan tinggi akan berupaya meningkatkan mutu mutu yang disebut mutu lulusan. Salah satu cara untuk mencapai kualitas lulusan yang baik adalah dengan lebih cepat meluluskan mahasiswa. Sesuai dengan Keputusan Menteri Pendidikan Nomor 232/U/2000 Pasal 5 bahwa beban studi untuk program sarjana direncanakan selama 8 semester (4 tahun) dan dapat ditempuh paling lama 14 semester, yaitu ketika mahasiswa menyelesaikan program sarjana dengan pembulatan lebih 4 tahun sampai dengan  $\leq 7$

tahun, berarti lulus tetapi tidak tepat waktu karena melebihi waktu yang direncanakan.

Setiap mahasiswa tentunya berharap dapat mencapai waktu lulus yang ditentukan dalam arti memperoleh gelar sarjana tepat waktu. Namun, masih banyak kasus mahasiswa yang belum dapat menyelesaikan kuliahnya dengan tepat waktu. Masalah tidak tepat waktu dalam menyelesaikan pendidikan juga muncul di kalangan mahasiswa Universitas Pamulang. Dari data kelulusan mahasiswa S1 Fakultas Ekonomi yang diperoleh dari database akademik Universitas Pamulang tahun ajaran 2018/2019 terdapat kasus sebanyak 38,1% mahasiswa tidak memenuhi standar waktu kelulusan yang ditetapkan.

Karena masih ada mahasiswa yang tidak lulus tepat waktu, maka diperlukan upaya untuk meningkatkan ketepatan waktu kelulusan. Salah satunya dengan melakukan penelitian apa



saja yang membuat para mahasiswa belum mampu memenuhi standar waktu kelulusan. Ketepatan waktu kelulusan mahasiswa ada beberapa factor yang mempengaruhi yaitu faktor internal dan faktor eksternal. Sebagai upaya untuk memenuhi standar waktu kelulusan siswa maka perlu diperhatikan faktor - faktor yang mempengaruhi waktu kelulusan mahasiswa. Penelitian ini akan mengklasifikasikan ketepatan waktu kelulusan mahasiswa berdasarkan jenis kelamin, jenis sekolah, daerah asal, faktor pekerjaan orang tua, faktor program studi, dan faktor predikat indeks (prestasi kumulatif).

Data mining merupakan suatu proses yang menggunakan ilmu statistika, ilmu matematika, teori kecerdasan buatan, dan *machine learning method*. Kombinasi ilmu tersebut digunakan untuk proses ekstraksi dan identifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban et al., 2005). Salah satu fungsi dari data mining adalah metode klasifikasi data. Metode klasifikasi merupakan salah satu metode yang digunakan untuk melakukan migrasi keadaan apapun, klasifikasi data memerlukan suatu metode untuk menemukan model yang menjelaskan atribut kelas sebagai fungsi dari atribut masukan. Metode *naïve bayes* merupakan salah satu metode klasifikasi. *Naive Bayes* adalah metode klasifikasi menggunakan teori dan statistik probabilitas yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yang memprediksi probabilitas masa depan berdasarkan data atau pengalaman dari *veraguen & as theory* (2016).

Dalam pengklasifikasian ketepatan waktu kelulusan mahasiswa, tentu saja data yang diperoleh dari *database* sebuah universitas dalam jumlah data yang besar atau banyak, oleh karena itu membutuhkan teknik pengolahan yang tepat dan mempermudah dalam menganalisis datanya. Seperti yang dijelaskan pada buku (Turban et al., 2005) teknik data mining ini mampu melakukan proses ekstraksi pola yang menarik dari data dalam jumlah yang

sangat besar, sehingga data mining ini dapat digunakan untuk menyelesaikan permasalahan yang diambil dari penelitian ini yaitu pengklasifikasian data kelulusan mahasiswa. Berdasarkan latar belakang yang sudah dijelaskan, maka akan dilakukan penelitian dengan judul “Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Dengan Metode *Naïve Bayes*”. Dan hasilnya nanti diharapkan bisa menjadi bahan evaluasi untuk meningkatkan kualitas ketepatan waktu kelulusan dari mahasiswa.

## LANDASAN TEORI

### Data Mining

Teknik data mining berasal dari kata data dan mining. Data berarti kumpulan fakta-fakta atau entitas yang terekam yang tidak memiliki arti dan selama ini terabaikan, sedangkan mining artinya proses penambangan. Dari dua pengertian kata tersebut maka dapat disimpulkan bahwa data mining merupakan proses penambangan data yang akan menghasilkan sebuah *keluaran* (output) berupa pengetahuan. Selain itu juga definisi data mining disampaikan oleh (Turban et al., 2005) bahwa data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Hal ini juga disampaikan oleh (Larose, 2005) data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak terduga dan meringkas data dengan cara berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.

### Klasifikasi

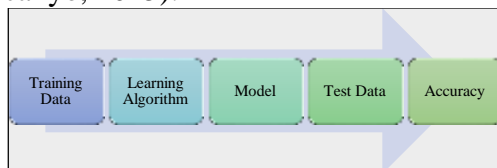
Klasifikasi merupakan sebuah proses training (pembelajaran) suatu fungsi tujuan (target) yang digunakan untuk memetakan tiap himpunan atribut suatu objek ke satu dari label kelas tertentu. Proses klasifikasi didasarkan pada empat komponen mendasar :



1. Label, variabel terikat dari model yang merupakan variabel kategori mewakili objek klasifikasinya.
2. Prediktor, variabel bebas dari model diwakili oleh karakteristik (atribut) dari data yang harus diklasifikasikan.
3. Data latih (training data), merupakan bagian dataset yang dilatih untuk membuat prediksi atau menjalankan fungsi dari sebuah algoritma.
4. Data uji (testing data), merupakan bagian dataset yang ditest untuk melihat keakuratannya.

### Naïve Bayes

Naïve bayes merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Naïve Bayes merupakan salah satu algoritma pemecahan masalah yang termasuk kedalam metode klasifikasi pada data mining. Naïve bayes mengadopsi ilmu statistika yaitu menggunakan teori kemungkinan (probabilitas) untuk menyelesaikan sebuah kasus *supervised learning*, artinya dalam himpunan data terdapat label, *class* atau target sebagai acuan. Naïve Bayes dalam konsepnya ada beberapa fase penyelesaian yaitu dimulai dari training dan diakhiri dengan proses testing sehingga dihasilkan sebuah keputusan yang akurat. Berikut ini adalah gambar alur pemecahan metode klasifikasi (Nofriansyah & Widi Nurcahyo, 2015):



**Gambar** Fase Penyelesaian Metode Klasifikasi

Probabilitas atas peluang bersyarat dinyatakan dalam penjelasan sebagai berikut (Iskandar, 2013)

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

Dimana  $X$  adalah bukti,  $H$  adalah hipotesis,  $P(H|X)$  adalah probabilitas bahwa hipotesis  $H$  benar untuk bukti  $X$  atau dengan kata lain  $P(H|X)$  merupakan probabilitas posterior  $H$  dengan syarat  $X$ ,  $P(X|H)$  adalah probabilitas bahwa bukti  $X$  benar hipotesis  $H$  atau probabilitas posterior  $X$  dengan syarat  $H$ ,  $P(H)$  adalah probabilitas prior hipotesis  $H$ , dan  $P(X)$  adalah probabilitas prior bukti  $X$ .

Dalam data mining,  $X$  adalah sebuah tuple atau objek data,  $H$  adalah hipotesis atau dugaan bahwa tuple  $X$  adalah kelas  $C$ . Secara spesifik dalam klasifikasi anda dapat menghitung  $P(H|X)$  sebagai probabilitas bahwa hipotesis  $H$  benar untuk tuple  $X$  atau dengan kata lain  $P(H|X)$  adalah probabilitas bahwa tuple  $X$  berada dalam kelas  $C$ . Sementara itu  $P(H)$  adalah probabilitas prior bahwa hipotesis  $H$  benar untuk setiap tuple tidak peduli nilai-nilai atributnya sedangkan  $P(X)$  adalah probabilitas prior dari tuple  $X$ .

Langkah-langkah metode *naïve bayes* (Suyanto, 2019) :

1. Misalkan  $A$  adalah himpunan data latih (*training set*) yang berisi sejumlah tuple beserta label kelasnya. Setiap tuple adalah berdimension  $n$  yang dinyatakan sebagai  $X = (x_1, x_2, \dots, x_n)$  yang didapat dari  $n$  atribut  $A_1, A_2, \dots, A_n$ .
2. Misalkan terdapat  $m$  kelas, yaitu  $C_1, C_2, \dots, C_m$ , untuk sebuah tuple masukan  $X$ , *naïve bayes classifier* memprediksi bahwa tuple  $X$  termasuk kedalam kelas  $C_i$  jika dan hanya jika :

$$P(C_i|X) > P(C_j|X) \text{ untuk } \leq j \leq m, j \neq i$$

Dengan kata lain *naïve bayes classifier* bekerja dengan cara memaksimalkan (mencari nilai maksimal)  $P(C_i|X)$ . Kelas  $C_i$  yang membuat  $P(C_i|X)$  bernilai maksimal disebut *maximum posteriori hypothesis*. Dengan teorema Bayes,  $P(C_i|X)$  diestimasi dengan menggunakan formula :

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$



3. Mengingat  $P(X)$  bernilai sama untuk semua kelas (artinya, tuple  $X$  memiliki probabilitas yang sama untuk masuk ke dalam kelas manapun), maka hanya  $P(X | C_i)P(C_i)$  yang perlu dimaksimalkan. Jika probabilitas prior untuk setiap kelas tidak diketahui, maka probabilitas setiap kelas biasanya diasumsikan sama  $P(C_1)=P(C_2)=\dots=P(C_n)$ . Dengan demikian, *naïve bayes classifier* hanya memaksimalkan  $P(X | C_i)$ . Namun jika probabilitas prior untuk setiap kelas bisa berbeda-beda, maka *naïve bayes classifier* harus memaksimalkan  $P(X | C_i)P(C_i)$ .
4. Jika anda berhadapan dengan himpunan data yang memiliki sangat banyak atribut, anda dapat mereduksi kompleksitas penghitungan  $P(X | C_i)$ . Dengan demikian *naïve bayes classifier* memaksimalkan :
- $$P(C_i | X) = \prod_{k=1}^n P(X_k | C_i)$$
- $$= P(X_1 | C_i) \times P(X_2 | C_i) \times \dots \times P(X_n | C_i)$$
- Berdasarkan tuple-tuple pada himpunan data latih, anda dapat mengestimasi  $P(X_1 | C_i), P(X_2 | C_i), \dots, P(X_n | C_i)$  berdasarkan atributnya.
5. Untuk memprediksi label kelas dari tuple  $X$  anda harus menghitung probabilitas  $P(X | C_i)P(C_i)$  untuk setiap kelas  $C_i$ . Selanjutnya anda hanya perlu memaksimalkan probabilitas tersebut yaitu mencari kelas  $C_i$  yang menghasilkan probabilitas  $P(X | C_i)P(C_i)$  maksimum sebagai kelas keputusan. Secara matematis tuple  $X$  diberi label kelas  $C_i$  jika dan hanya jika :
- $$P(X | C_i)P(C_i) > P(X | C_j)P(C_j)$$
- untuk  $1 \leq j \leq m, j \neq i$

### Faktor Kelulusan Mahasiswa

Faktor-faktor yang mempengaruhi kelulusan mahasiswa yang ditunjukkan pada

beberapa penelitian, diantaranya pada penelitian (ARIANI et al., 2013) faktor yang mempengaruhi adalah jalur masuk Perguruan Tinggi, jenis kelamin, jurusan, jalur masuk, daerah asal, asal SMA dan lama pengerjaan tugas akhir. Ada juga faktor yang ditunjukkan pada penelitian (Rahmi & Yozza, 2020) yaitu asal SMA, nilai UN, prestasi akademik SMA, mata pelajaran SMA, keikutsertaan organisasi SMA, keikutsertaan lomba akademik/non akademik, jalur masuk kuliah, pilihan program studi ketika memasuki Perguruan Tinggi, jenis kelamin, dan IPK. Hal serupa juga ditunjukkan pada penelitian (Darwin & Zurimi, 2019) bahwa yang mempengaruhi kelulusan mahasiswa diantaranya jalur masuk Perguruan Tinggi, jenis kelamin, IP semester awal, daerah asal, dan kondisi ekonomi keluarga.

Dari beberapa penelitian-penelitian sebelumnya maka pada penelitian ini akan dilakukan klasifikasi ketepatan waktu kelulusan mahasiswa berdasarkan data yang diperoleh dari *database* Universitas Pamulang diantaranya berupa jenis kelamin, jenis Sekolah (SMA, SMK, MA), asal daerah, pekerjaan orang tua, program studi, dan IPK.

### METODE PENELITIAN

#### Jenis Dan Sumber Data

Dalam penelitian ini jenis data yang digunakan adalah data kuantitatif karena angka-angka atau data menunjukkan nilai variabel yang menentukannya.

Selanjutnya dalam penelitian ini sumber data yang digunakan adalah data sekunder dari *database* akademik Universitas Pamulang berupa data pascasarjana untuk mahasiswa S1 Fakultas Ekonomi yang terdiri dari studi Akuntansi dan program studi Manajemen tahun ajaran 2018/2019. Data yang diperoleh dengan jumlah 1.126 data siswa, terdiri dari 697 siswa yang lulus tepat waktu dan 429 siswa yang tidak lulus tepat waktu. Data yang digunakan sebagai atribut jenis kelamin, jenis sekolah, asal daerah, pekerjaan orang tua, program studi, dan



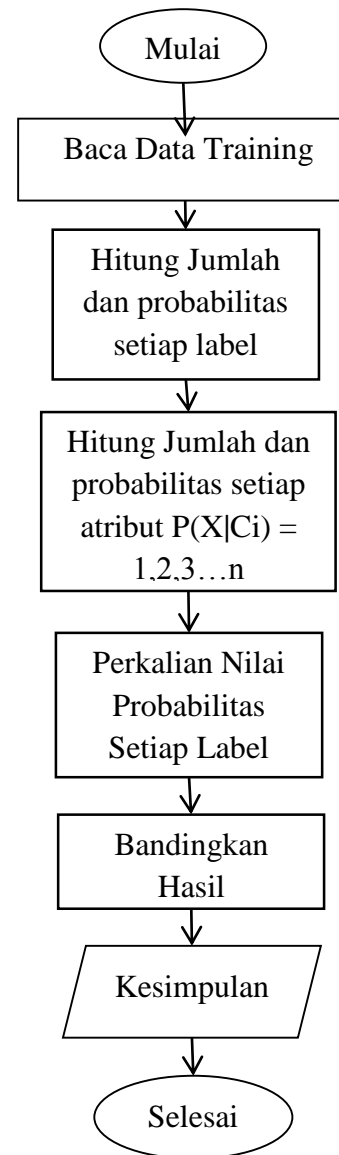
predikat IPK, sedangkan sebagai label adalah ketepatan waktu kelulusan mahasiswa.

### Variabel Penelitian

**Tabel 1. Penjelasan Operasional Atribut dan Label**

No	Jenis Variabel	Penjelasan Operasional	
1	Atribut	Jenis Kelamin ( $X_1$ )	laki-laki dan perempuan
2	Atribut	Jenis Sekolah ( $X_2$ )	SMK, SMA, dan MA
3	Atribut	Asal Daerah ( $X_3$ )	Jabodetabek dan Luar Jabodetabek
4	Atribut	Pekerjaan Orang Tua ( $X_4$ )	PNS, Wiraswasta, Karyawan Swasta, Petani, Buruh, Lainnya (selain PNS, Wiraswasta, Karyawan Swasta, Petani, Buruh).
5	Atribut	Program Studi ( $X_5$ )	Akuntansi dan Manajemen
6	Atribut	Predikat IPK ( $X_6$ )	Nilai akhir kelulusan mahasiswa S-1 dari 2,00 – 4,00 dengan predikat sebagai berikut : 3,51 – 4,00 = Dengan Pujian 3,01 – 3,50 = Sangat Memuaskan 2,76 – 3,00 = Memuaskan
7	Label	Waktu Kelulusan Mahasiswa	Lama studi mahasiswa : $\leq 4$ tahun = Tepat Waktu $> 4$ tahun = Tidak Tepat Waktu

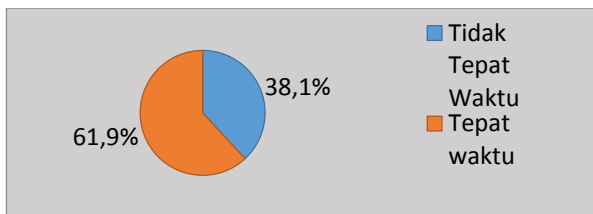
**Gambar 1. Alur Penelitian Naïve Bayes**



## HASIL DAN PEMBAHASAN

### Gambaran Umum

Pada penelitian ini jumlah data yang diperoleh sebanyak 1126 data mahasiswa. Berikut gambaran umum ketepatan waktu kelulusan mahasiswa S1 Fakultas Ekonomi Universitas Pamulang tahun ajaran 2018/2019:



Berdasarkan gambar di atas diketahui persentase waktu kelulusan mahasiswa tepat waktu sebesar 61.9% sedangkan untuk mahasiswa tidak tepat waktu sebesar 38,1%. Dari gambaran tersebut menunjukkan bahwa kelulusan mahasiswa yang lulus tepat waktu lebih banyak, namun masih banyak juga kasus mahasiswa yang belum mampu menyelesaikan pendidikannya sesuai waktu yang dijadwalkan.

**Hasil Analisis Data**

**1. Membaca Data Training**

**Tabel 2. Jumlah Data Training**

Atribut	Sub Atribut	Label		Total
		Tepat Waktu	Tidak Tepat Waktu	
X <sub>1</sub>	L	173	204	
	P	385	139	
<b>Total</b>		<b>558</b>	<b>343</b>	<b>901</b>
X <sub>2</sub>	SMK	333	225	
	SMA	193	100	
	MA	32	18	
<b>Total</b>		<b>558</b>	<b>343</b>	<b>901</b>
X <sub>3</sub>	J	365	223	
	LJ	193	120	
<b>Total</b>		<b>558</b>	<b>343</b>	<b>901</b>
X <sub>4</sub>	W	253	151	
	KW	102	59	
	P	34	35	
	B	65	36	
	PNS	32	24	
	L	72	38	
<b>Total</b>		<b>558</b>	<b>343</b>	<b>901</b>
X <sub>5</sub>	A	149	181	
	M	409	162	
<b>Total</b>		<b>558</b>	<b>343</b>	<b>901</b>
X <sub>6</sub>	DP	234	63	
	SM	323	256	
	M	1	24	
<b>Total</b>		<b>558</b>	<b>343</b>	<b>901</b>

**2. Menghitung Jumlah dan Probabilitas Setiap Label**

**Tabel Probabilitas Ketepatan Waktu kelulusan Mahasiswa**

Label	Jumlah	Probabilitas
Tepat Waktu	558	0,619
Tidak Tepat Waktu	343	0,381
Total	901	1

**3. Menghitung Probabilitas Jumlah Setiap Atribut P(X|Ci) = 1,2,3...n**

**Tabel 3. Probabilitas Setiap Kriteria Ketepatan Waktu kelulusan Mahasiswa**

Atribut		Probabilitas	
		Tepat Waktu	Tidak Tepat Waktu
X <sub>1</sub>	Laki-laki	0,310	0,595
	Perempuan	0,690	0,405
X <sub>2</sub>	SMK	0,597	0,656
	SMA	0,346	0,292
	MA	0,057	0,052
	Jabodetabek	0,654	0,650
X <sub>3</sub>	L.Jabodetabek	0,346	0,350
	Wiraswasta	0,453	0,440
X <sub>4</sub>	K.Swasta	0,102	0,172
	Petani	0,061	0,102
	Buruh	0,116	0,105
	PNS	0,057	0,070
	Lainya	0,129	0,111
	Akuntansi	0,267	0,528
X <sub>5</sub>	Manajemen	0,733	0,472
	DP	0,419	0,184
	SM	0,579	0,746
X <sub>6</sub>	M	0,002	0,070

**4. Perkalian Nilai Probabilitas Atribut Untuk Setiap Label**

Dengan menggunakan satu contoh data testing dengan kategori atribut seperti pada tabel contoh data testing, akan disimulasikan perkalian dalam menentukan klasifikasi label waktu kelulusan apakah tepat waktu atau tidak tepat waktu.

**Tabel 4. Contoh Kasus Data Testing**

X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	Label
L	SMK	J	PE	AK	M	?

## a. Probabilitas Label Tepat Waktu

Dari data training diperoleh nilai probabilitas setiap atribut sebagai berikut:

$$P(\text{Tepat Waktu}) = 0,619$$

$$P(X_1(L)|C(\text{Tepat Waktu})) = 0,310$$

$$P(X_2(\text{SMK})|C(\text{Tepat Waktu})) = 0,597$$

$$P(X_3(J)|C(\text{Tepat Waktu})) = 0,654$$

$$P(X_4(\text{PE})|C(\text{Tepat Waktu})) = 0,061$$

$$P(X_5(\text{AK})|C(\text{Tepat Waktu})) = 0,267$$

$$P(X_6(M)|C(\text{Tepat Waktu})) = 0,002$$

- Probabilitas Label Tepat Waktu:  

$$= P(X_1|\text{Tepat Waktu}) \times P(X_2|\text{Tepat Waktu}) \times P(X_3|\text{Tepat Waktu}) \times P(X_4|\text{Tepat Waktu}) \times P(X_5|\text{Tepat Waktu}) \times P(X_6|\text{Tepat Waktu}) \times P(\text{Tepat Waktu})$$

$$= 0,310 \times 0,597 \times 0,654 \times 0,061 \times 0,267 \times 0,002 \times 0,619$$

$$= 0,00000244$$

## b. Probabilitas Label Tidak Tepat Waktu

Dari data training diperoleh nilai probabilitas setiap atribut sebagai berikut:

$$P(\text{Tidak Tepat Waktu}) = 0,381$$

$$P(X_1(L)|C(\text{Tidak Tepat Waktu})) = 0,595$$

$$P(X_2(\text{SMK})|C(\text{Tidak Tepat Waktu})) = 0,656$$

$$P(X_3(J)|C(\text{Tidak Tepat Waktu})) = 0,650$$

$$P(X_4(\text{PE})|C(\text{Tidak Tepat Waktu})) = 0,102$$

$$P(X_5(\text{AK})|C(\text{Tidak Tepat Waktu})) = 0,528$$

$$P(X_6(M)|C(\text{Tidak Tepat Waktu})) = 0,070$$

- Probabilitas Label Tidak Tepat Waktu:  

$$= P(X_1|\text{Tidak Tepat Waktu}) \times P(X_2|\text{Tidak Tepat Waktu}) \times P(X_3|\text{Tidak Tepat Waktu}) \times P(X_4|\text{Tidak Tepat Waktu}) \times P(X_5|\text{Tidak Tepat Waktu}) \times P(X_6|\text{Tidak Tepat Waktu}) \times P(\text{Tidak Tepat Waktu})$$

$$= 0,595 \times 0,656 \times 0,650 \times 0,102 \times 0,528 \times 0,070 \times 0,381$$

$$= 0,00013884$$

## 5. Membandingkan Nilai Probabilitas Setiap label

Berdasarkan hasil perhitungan, diketahui bahwa contoh kasus data testing di atas memiliki nilai probabilitas pada label tepat waktu sebesar 0,00000244 dan memiliki nilai probabilitas pada label tidak tepat waktu sebesar 0,00013884. Itu berarti contoh data testing tersebut memiliki probabilitas label waktu kelulusan tidak tepat waktu lebih besar dibandingkan dengan probabilitas label waktu kelulusan tepat waktu.

## 6. Membuat kesimpulan

Setelah diketahui hasil perhitungan dan membandingkannya, dapat disimpulkan bahwa contoh kasus data testing dengan atribut jenis kelamin laki-laki, jenis sekolah SMK, asal daerah Jabodetabek, pekerjaan orang tua petani, program studi Akuntansi, predikat IPK “Memuaskan” masuk dalam klasifikasi label waktu kelulusan tidak tepat waktu.

## Implementasi Aplikasi Rapidminer

Karena data penelitian berjumlah besar, maka peneliti menggunakan bantuan *software* untuk memudahkan dalam melakukan perhitungan. Pada penelitian ini akan diklasifikasikan data sejumlah 225 data mahasiswa. Berikut hasil output *naïve bayes* melalui aplikasi *rapidminer*:

	Yes	No
prob tepat waktu	0,619	0,381
prob tidak tepat waktu	0,381	0,619
hasil akhir	0,00000244	0,00013884

## Pembahasan

Berdasarkan hasil perhitungan di atas diketahui bahwa data kelulusan mahasiswa program studi Akuntansi dan program studi Manajemen Universitas Pamulang tahun ajaran 2018/2019 berjumlah 2478 mahasiswa, dari



total data diketahui 61,9% yang lulus tepat waktu yaitu berjumlah 697 mahasiswa dan 38,1% yang lulus tidak tepat waktu yaitu berjumlah 429 mahasiswa.

Pada perhitungan menggunakan *naïve bayes* diperoleh hasil dari penelitian bahwa dari 225 data testing yang diuji menunjukkan 156 data yang terklasifikasi dengan benar. Hasil prediksi klasifikasi waktu kelulusan tersebut diperoleh tingkat akurasi sebesar 69,33%.

## PENUTUP

### Kesimpulan

1. Data kelulusan mahasiswa S1 Fakultas Ekonomi Universitas Pamulang tahun ajaran 2018/2019 menunjukkan 61,9% yang lulus tepat waktu dan 38,1% yang lulus tidak tepat waktu.
2. Pada 225 data testing yang diuji menggunakan perhitungan *naïve bayes* menunjukkan hasil bahwa data testing tersebut terklasifikasi dengan benar sebanyak 156 data.
3. Hasil perhitungan *naïve bayes* pada 156 data yang terklasifikasi dengan benar menunjukkan tingkat akurasi sebesar 69,33%.

### Saran

1. Untuk peneliti selanjutnya bisa ditambahkan jumlah atribut yang digunakan dalam klasifikasi waktu kelulusan, dengan harapan bisa mendapatkan informasi lebih mengenai faktor yang berpengaruh pada lama studi mahasiswa.
2. Pada penelitian klasifikasi data juga bisa digunakan metode lain selain *naïve bayes* seperti jaringan saraf tiruan, *decision tree*, *K-nearest neighbor*, dan lainnya.

## DAFTAR PUSTAKA

- [1] Ariani, N. K. D., Sumarjaya, I. W., & Bagus Oka, T. (2013). Analisis Faktor-Faktor Yang Memengaruhi Waktu Kelulusan Mahasiswa Dengan Menggunakan Metode GOMPIT (Studi Kasus: Mahasiswa Fakultas MIPA Universitas Udayana). *E-Jurnal Matematika*, 2(3), 40. <https://doi.org/10.24843/mtk.2013.v02.i03.p047>
- [2] Darwin, D., & Zurimi, S. (2019). Analisis model aplikatif Multivariate Adaptive Regression Spline (MARS) terhadap klasifikasi faktor yang mempengaruhi masa studi mahasiswa FKIP Universitas Darussalam Ambon. *Jurnal Simetrik*, 9(2), 250–255. <http://ejournal-polnam.ac.id/index.php/JurnalSimetrik/article/viewFile/426/278>
- [3] Iskandar, D. (2013). Perbandingan akurasi tingkat kemiskinan antara algoritma C4.5 dan Naive Bayes Classifier. *JAVA Journal of Electrical and Electronic Engineering*.
- [4] Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. In *Discovering Knowledge in Data: An Introduction to Data Mining*. <https://doi.org/10.1002/0471687545>
- [5] Nofriansyah, D., & Widi Nurcahyo, G. (2015). *Algoritma Data Mining dan Pengujian*. Yogyakarta: Deepublish.
- [6] Rahmi, I., & Yozza, H. (2020). Analisis Kausal Masa Studi Mahasiswa Program Studi Matematika Universitas Andalas Dengan Menggunakan Metode CART.
- [7] Suyanto. (2019). *Data mining: untuk klasifikasi dan klasterisasi data*. Bandung: Informatika.
- [8] Syarli, S., & Muin, A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*, 2(1), 22–26.
- [9] Turban, E., Aronson, J. E., & Peng Liang, T. (2005). *Decision-Support-and-Intelligence.Pdf*. Andi: Yogyakarta.