

## Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes

Armansyah<sup>1,\*</sup>, Rakhmat Kurniawan Ramli<sup>1</sup>

<sup>1</sup> Program Studi Ilmu Komputer, UIN Sumatera Utara, Indonesia

\* Correspondence: armansyah@uinsu.ac.id

Copyright: © 2022 by the authors

Received: 20 Desember 2021 | Revised: 10 Januari 2022 | Accepted: 17 Januari 2022 | Published: 20 Juni 2022

### Abstrak

Terjadi kesenjangan jumlah mahasiswa masuk dan keluar lulus pada program studi ini. Kesenjangan itu terjadi karena rendahnya kelulusan mahasiswa tepat waktu. Oleh karena itu, penelitian ini bertujuan untuk mendisain model prediksi kelulusan mahasiswa tepat waktu dan tidak tepat waktu dalam mencari solusi dari kesenjangan itu. Model prediksi yang digunakan pada penelitian ini adalah Naïve Bayes. Data yang digunakan berupa 44 data lulusan tahun 2020 dibagi dua bagian pada tahapan analisis dengan RapidMider, yaitu 38 lulusan (training) dan 6 data untuk pengujian. Hasil temuan kami menunjukkan bahwa model prediksi penelitian yang dihasilkan adalah sangat baik, dengan 5 data dari 6 lulusan sesuai dengan prediksi di pengujian pertama, sementara satu data tidak terbaca. Namun pada pengujian yang kedua 6 data lulusan sama persis dengan pemodelan, yang akurasi menunjukkan nilai 100%.

**Kata kunci:** model prediksi; naïve bayes; kelulusan tepat waktu; klasifikasi; machine learning

### Abstract

*There is a gap in the number of students in and out of graduating from this study program. The gap occurs due to the low graduation of students on time. Therefore, this study aims to design a model of student graduation predictions on time and not on time in finding solutions to that gap. The predictive model used in this study is Naïve Bayes. The data used in the form of 44 graduate data in 2020 is divided into two parts of the analysis stage with RapidMider, namely 38 graduates (training) and six data for testing. Our findings showed that the resulting research prediction model was excellent, with five data from six graduates matching the predictions in the first test, while one data was illegible. However, in the second test, six graduate data are the same as modeling, whose accuracy shows a value of 100%.*

**Keywords:** prediction model; nave bayes; on-time graduation; classification; machine learning

## PENDAHULUAN

Kelulusan mahasiswa menjadi perhatian penting bagi hampir semua perguruan tinggi. Karena lulusan merupakan produk bimbingan yang dihasilkan melalui tahapan pendidikan. Kata “Kelulusan” merupakan nomina yang berasal dari kata dasar “lulus” Dalam *kbbi.web.id* (yang diakses pada Januari, 11, 2022), mendaftarkan 7 arti diantara: 1) dapat masuk; 2) dapat lepas; 3) terperosok masuk; 4) berhasil (dalam ujian); 5) diperkenankan; 6) hilang lenyap; dan, 7) lelap. Sementara Arti kata kelulusan dalam *kbbi.lectur.id* (yang juga diakses pada Januari, 11, 2022) mencantumkan dua pengertian yaitu keguguran (melahirkan anak sebelum masanya), dan, hal (keadaan) lulus (ujian dan sebagainya). Dalam konteks pendidikan, kelulusan sangat relevan dengan pengertian bahwa kelulusan merupakan keberhasilan seorang pelajar (mahasiswa) dalam menghadapi suatu proses pendidikan, yang diantaranya adalah evaluasi atau ujian pada setiap mata kuliah termasuk mata kuliah skripsi, yang merupakan mata kuliah



wajib bagi hampir semua perguruan tinggi program sarjana (Strata-1). Setiap pelajar, dalam hal ini mahasiswa yang dinyatakan sah lulus berijazah umumnya disebut dengan alumni suatu perguruan tinggi. Melalui alumni yang dihasilkan itu, suatu perguruan tinggi (kampus) dapat menunjukkan eksistensinya, termasuk keunggulan-keunggulan program yang ditawarkan kepada masyarakat. Faktor kepercayaan masyarakat terhadap suatu kampus, juga dapat dilihat dari rekam jejak alumni yang telah tersebar di banyak perusahaan dan lembaga-lembaga yang memberdayakan alumni suatu kampus. Namun, ketika jumlah alumni kampus belum maksimal, rekam jejak alumni sulit untuk dicapai atau diketahui. Kelulusan mahasiswa juga merupakan perhatian bagi pimpinan Program Studi Ilmu Komputer UIN Sumatera Utara Medan, mengingat tingkat kelulusan di kampus ini masih terlalu rendah.

Data kelulusan Oktober tahun 2020 jumlah lulusan masih berada diangka 44 alumni. Dengan rincian 38 lulusan pada November 2019 - Agustus 2020 dari mahasiswa angkatan pertama (2015), dan disusul oleh 6 lulusan pada Oktober 2020 dari mahasiswa angkatan kedua (2016). Hingga tahun 2020 mahasiswa kampus ini sudah 1201 mahasiswa dan belum termasuk dengan mahasiswa angkatan 2021. Data lulusan yang tertuang pada awal paragraf ini masih sangat kecil, dan kami pantau sepanjang 2021 sampai penelitian ini dilaporkan, jumlah lulusan tidak mencapai 100 alumni (tabel lulusan prodi, 2021). Sementara untuk mahasiswa angkatan 2017 yang berjumlah 253 mahasiswa, beberapa masih dalam proses bimbingan tugas akhir sejak riset ini dilakukan.

Kecilnya jumlah lulusan itu merupakan masalah bagi program studi, karena terjadinya kesenjangan antara jumlah mahasiswa masuk dengan mahasiswa keluar lulus, dan akan semakin tinggi pula kesenjangan itu bila tidak dicari penyelesaiannya. Kesenjangan itu akan membawa dampak buruk bagi program studi pada aspek-aspek lain, seperti akan terjadi kesenjangan jumlah sarana dan fasilitas kampus, kesenjangan jumlah tenaga pendidik (dosen) serta berpeluang pada rendahnya penilaian akreditasi program studi. Mengingat, kelulusan juga merupakan bagian dari *asesment* akreditasi program studi seperti yang tertuang dalam (BANPT, 2019), yakni sebagai indikator penilaian kinerja kuantitatif perguruan tinggi, dan secara spesifik indikator penilaian aspek pembelajaran (Guswandi et al., 2021). Ada beberapa cara menyelesaikan kesenjangan ini. Satu diantaranya adalah melakukan prediksi kelulusan tepat waktu mahasiswa. Dengan melakukan prediksi, maka melalui pimpinan program studi, hasil prediksi nantinya dapat dijadikan sebagai informasi (referensi) dalam membuat kebijakan untuk mendorong tingkat kelulusan tepat waktu mahasiswa di masa mendatang.

Prediksi terhadap kelulusan mahasiswa seperti yang disinggung diparagraf sebelumnya, dapat dilakukan dengan pendekatan *machine learning*. *Machine learning* merupakan teknik yang dilakukan oleh mesin atau aplikasi dalam mempelajari data untuk tujuan mengelompokkan, mengklasifikasikan ataupun memprediksi masa depan (Hurwitz & Kirsch, 2018). Dengan kata lain mesin atau aplikasi dapat melakukan berbagai tugas manusia seperti mengelompokkan, mengklasifikasikan sesuatu dalam kategori tertentu, serta melakukan prediksi melalui pembelajaran terhadap data-data yang terpola. Pendekatan *machine learning* juga sering diterapkan pada *data mining*, yakni kombinasi bidang ilmu komputer dan statistik yang digunakan untuk menemukan pola dalam sekumpulan besar informasi (Han, Kamber, & Pei, 2012). Klasifikasi merupakan bentuk analisis dari berbagai model data ke dalam label-label kelas. Dimana proses klasifikasi datanya dilakukan melalui tahapan, (a) pembelajaran (*training/learning*), yang menganalisis data menggunakan teknik klasifikasi (algoritma klasifikasi), (b) klasifikasi, yakni memilah data dari kumpulan data uji. Tahapan klasifikasi umumnya dilakukan pada tahapan pengujian (*testing*) data, dimana metode klasifikasinya membentuk aturan klasifikasi.

Ada banyak metode klasifikasi yang diterapkan pada *machine learning* untuk tujuan prediksi, diantaranya adalah *Decision Tree*, *Multilaye Perceptron Backpropagation*, *K-Nearest Neighbor*, *Naïve Bayes* dan masih banyak metode klasifikasi lainnya. Klasifikasi dengan

*Decision Tree* sangat cocok untuk penelitian dengan data yang besar (stackoverflow, 2021), begitu juga dengan metode *Multilayer Perceptron Backpropagation*, dan *K-Nearest Neighbor*. Sementara *Naïve Bayes* sangat cocok untuk penelitian dengan data analisis yang berukuran kecil. *Naïve Bayes* adalah metode algoritma yang bekerja atas bagaimana menghitung frekuensi atas setiap term yang ada dalam dokumen (Prakoso et al., 2019). *Naïve Bayes* juga adalah klasifikasi berdasarkan statistik dan probabilitas sederhana yang banyak digunakan untuk metode pembelajaran identifikasi citra medis saat ini (Yunianto et al., 2021). Secara sederhana pengklasifikasian *Naïve Bayes* mengasumsikan bahwa ada atau tidaknya suatu fitur kelas tidak terkait dengan fitur lain di kelas yang sama (Jain, 2019). Prinsip kerja teori *bayes* dalam klasifikasi diasumsikan seperti sebuah ponsel yang dapat dianggap cerdas jika memiliki layar sentuh, fasilitas internet, kamera yang bagus, dan lain-lain. Meskipun semua fitur itu saling bergantung satu sama lain, fitur-fitur tersebut berkontribusi secara independen terhadap kemungkinan ponsel tersebut sebagai ponsel pintar (Joseph, 2019). Selain itu, (Pramadhana, 2021) juga melakukan penelitian untuk mengklasifikasi penyakit diabetes, dengan menggunakan metode *Correlation Feature Selection (CFS)* yang dikombinasikan dengan metode *random over sampling* dan *adaboost*, klasifikasi berhasil dilakukan dengan akurasi 92,3%.

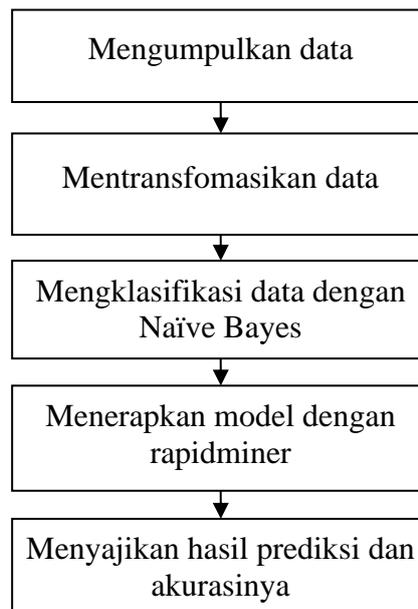
Riset dengan metode klasifikasi *naïve bayes* juga tidak kalah banyak yang menerapkannya. Seperti, Sidiq et al., (2020) yang menggunakan metode klasifikasi *naïve bayes* untuk menganalisis sintemen dari komentar *toxic* para *netizen* di ruang grup *facebook game online*. Dalam keterangan riset mereka menyimpulkan bahwa kinerja metode menghasilkan akurasi 75%. Penelitian lainnya seperti pada penelitian (Khoirunnisa et al., 2021) yang menggunakan metode klasifikasi *decision tree*, *naïve bayes*, dan *k-nearest neighbor (knn)* untuk memprediksi siswa smk yang melanjutkan ke perguruan tinggi dengan menggunakan atribut jenis kelamin, usia, penghasilan, masa asuransi. Dari penelitian mereka, metode tersebut menunjukkan akurasi masing-masing 96,60%, 92,40%, dan 94,96%. Lebih lanjut penelitian lain yang menggunakan metode klasifikasi *naïve bayes* yaitu (Zed & Muhammad, 2018) pada risetnya untuk menentukan prediksi karyawan tetap menggunakan atribut pelatihan berupa atribut absensi, kepribadian, dan prestasi. Dalam penelitiannya akurasi 94% diperoleh dari pengujian datanya. Selain itu, (Heryana, 2019) juga melakukan riset prediksi kelulusan tepat waktu, dengan menggunakan beberapa atribut pelatihan, dan kinerja metode menunjukkan akurasi 77, 2% pada pengujian 142 mahasiswa. Kemudian penelitian lainnya yang bertujuan mengklasifikasikan dan memprediksi data dengan metode klasifikasi *naïve bayes* diantaranya, (Harianto et al., 2020; Putra, 2018; Ramadhani, 2019; Roy & Hendriyawan, 2019) juga menerapkan metode *naïve bayes* juga dalam risetnya untuk mengklasifikasikan aktivitas mencurigakan dalam suatu jaringan dengan akurasi 91,62%, selain itu, Sari et al. (2020) yang menggunakan metode yang sama untuk mengklasifikasi kualitas kopi *Arabika* berdasarkan atribut aroma, *aftertaste*, *flavor*, *acidity*, *balance*, *body*, *uniformity*, *sweetness*, *clean cup*, dan *copper points* dengan akurasi 78%.

Berdasarkan permasalahan di atas serta dengan melihat penelitian terdahulu kami melakukan penelitian ini untuk menyelesaikan permasalahan yang kami hadapi. Dimana penelitian ini bertujuan untuk merancang model prediksi kelulusan mahasiswa tepat waktu dengan menggunakan metode *naïve bayes*, dengan *dataset* yang kami gunakan berupa 44 data lulusan yang terdiri dari 9 feature, yang akan diklasifikasikan ke dalam label kelulusan “tepat waktu” dan “tidak tepat waktu”.

## **METODE**

Penelitian ini merupakan penelitian eksperimen, yakni eksperimen model prediksi kelulusan mahasiswa tepat waktu dan tidak tepat waktu. Sementara tahapan pelaksanaan

penelitian yang kami lakukan ini mengacu kepada urutan tahapan yang tertuang pada gambar 1.



**Gambar 1.** Tahapan penelitian

Dimana, tahap pertama, mengumpulkan data lulusan per Oktober 2020 dari database Program Studi Ilmu Komputer UIN Sumatera Utara Medan berupa 9 atribut sebagai variabel independen yang terdiri dari atribut jenis kelamin (Zulfiqar et al., 2020), indeks prestasi semester 1, 2, 3, dan 4 (Rahayu et al., 2021), jumlah sks pada semester 4 (Reni, Widodo, & Adhi, 2020), lama tugas akhir dalam semester (Amalia, et al., 2014), lama studi mahasiswa dalam satuan tahun, serta jalur masuk kuliah (Syarli & Muin, 2016), dan 1 atribut sebagai variabel dependen yang merupakan kelas yang dicari dengan label tepat waktu dan tidak tepat waktu. Tahap kedua, mentransformasikan data ke dalam bentuk kategori sesuai pada tabel 1.

Tahap ketiga, mengklasifikasikan data dengan metode *naïve bayes*, yang dilakukan dengan mentabulasi data terlebih dahulu dalam bentuk tabel frekuensi, kemudian menentukan probabilitas kelas/label dan probabilitas kejadian atribut terhadap kelas menggunakan persamaan 1.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

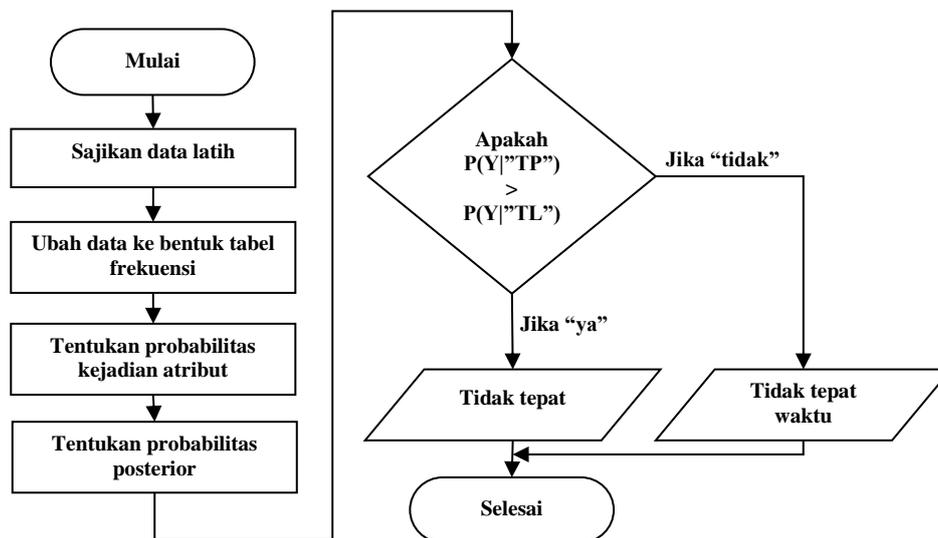
Dimana,  $P(A/B)$  adalah Probabilitas posterior, yakni probabilitas hipotesis A pada kejadian yang diamati B. Sementara  $P(B/A)$  adalah Probabilitas kemungkinan, yaitu probabilitas bukti yang diberikan bahwa probabilitas hipotesis adalah benar. Dan  $P(A)$  adalah probabilitas sebelumnya, yaitu probabilitas hipotesis sebelum mengamati bukti. Sedangkan  $P(B)$  adalah probabilitas marginal yang merupakan probabilitas bukti (Tuhenay, 2021). Tahapan klasifikasi metode *bayes* pada penelitian ini mengacu kepada gambar 2.

Pada gambar 2, secara *natural language* berarti (1) menyajikan data awal yang berisi 9 atribut dan 1 label/kelas, (2) mengubah data kedalam tabel frekuensi, (3) menentukan probabilitas kelas dan probabilitas atribut kejadian dari masing-masing kelas, (4) menentukan probabilitas kelas (*posterior*) dengan teori *bayes*, dan (5) membandingkan nilai probabilitas posterior dengan skema: jika probabilitas *posterior* “tepat waktu” > probabilitas *posterior* “tidak tepat waktu”, maka prediksi yang dipilih adalah kelas “tepat waktu”, dan berlaku sebaliknya (javatpoint.com, 2021). Tahap keempat, menerapkan model atau analisis data, berupa eksperimen model dengan *RapidMiner* yang menggunakan 44 data lulusan tahun 2020

yang dibagi dalam dua bagian, yaitu 38 data untuk pelatihan model, dan 6 data untuk pengujian model. Tahap selanjutnya adalah menyajikan hasil prediksi dengan menampilkan akurasi dari model yang digunakan, yang akan disajikan pada sub materi hasil penelitian.

**Tabel 1.** Kriteria atribut

No	Atribut	Keterangan
1	Jenis Kelamin (JK)	L (laki-laki), P(perempuan)
2	Indeks prestasi semester 1(IPS1)	Kriteria $\geq 3,00$ dan $< 3,00$
3	Indeks prestasi semester 2(IPS2)	Kriteria $\geq 3,00$ dan $< 3,00$
4	Indeks prestasi semester 3(IPS3)	Kriteria $\geq 3,00$ dan $< 3,00$
5	Indeks prestasi semester 4(IPS4)	Kriteria $\geq 3,00$ dan $< 3,00$
6	Jumlah SKS semester 4(SKS4)	Kriteria $\geq 80$ dan $< 80$ sks
7	Lama tugas akhir (TA)	Kriteria $\leq 1$ dan $> 1$ semester
8	Lama studi (STUDI)	Kriteria $\leq 4$ tahun dan $> 4$ tahun
9	Jalur masuk kuliah (MASUK)	Jalur SBMPT dan MANDIRI
10	Kelulusan (KL)	Tepat waktu (TP), dan tidak tepat waktu(TL)



**Gambar 2.** Alur klasifikasi kelas dengan *naïve bayes* (Wahyudi et al., 2019)

## HASIL DAN PEMBAHASAN

### Hasil

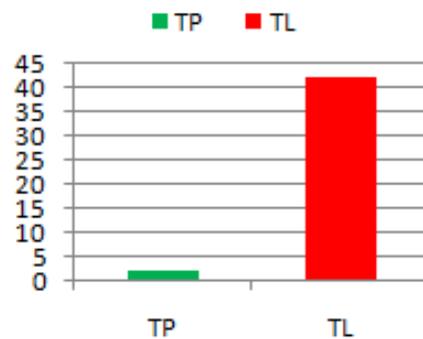
Setelah tahapan pengumpulan data, diperoleh data kelulusan mahasiswa per Oktober 2020. Data tersebut kemudian di transformasikan berdasarkan kriteria atribut dengan menggunakan tabel 1. Selanjutnya, data ditabulasi untuk melihat frekuensi kejadian atribut terhadap masing-masing kelas seperti pada tabel 2. Dari tabel frekuensi kejadian atribut, diperoleh probabilitas kelas seperti yang ditampilkan pada tabel 3, dan probabilitas atribut terhadap masing-masing kelas seperti yang tampilkan pada tabel 4. Dari kedua tabel tersebut maka diperoleh probabilitas kelas lulus tepat waktu sebesar 0,0526 atau 5,30% atau sama dengan 2 lulusan, sementara probabilitas lulus tidak tepat waktu sebesar 0,9473 atau 94,70% atau sama dengan 36 lulusan, yang secara visual ditampilkan pada grafik kelulusan di gambar 3.

**Tabel 2.** Variabel dan frekuensi kejadian atribut

Variabel	Atribut	Sub Atribut	Kelas (Y)		Total	Variabel	Atribut	Sub Atribut	Kelas (Y)		Total
			TP	TL					TP	TL	
X <sub>1</sub>	JK	L	1	15	38	X <sub>6</sub>	SKS4	>=80	2	36	38
		P	1	21				<80	0	0	
	Total	2	36	Total				2	36		
X <sub>2</sub>	IPS	>=3,00	2	36	38	X <sub>7</sub>	TA	<=1	2	0	38
		<3,00	0	0				>1	0	36	
	Total	2	36	Total				2	36		
X <sub>3</sub>	IPS	>=3,00	2	35	38	X <sub>8</sub>	Studi	<=4	2	0	38
		<3,00	0	1				>4	0	36	
	Total	2	36	Total				2	36		
X <sub>4</sub>	IPS	>=3,00	2	35	38	X <sub>9</sub>	Masuk	SBMPTN	1	1	38
		<3,00	0	1				MANDIRI	1	35	
	Total	2	36	Total				2	36		
X <sub>5</sub>	IPS	>=3,00	2	36	38	Y	Kelulusan	Tepat Waktu	2		38
		<3,00	0	0				Tidak tepat Waktu		36	
	Total	2	36								

**Tabel 3.** Probabilitas kelas

Variabel	Kelas	Jumlah	Probabilitas Kelas
Y(KL)	TP	2	0,053
	TL	36	0,947



**Gambar 3.** Grafik kelulusan tahun 2020

Berdasarkan nilai probabilitas pada tabel 3 dan 4 diperoleh probabilitas *posterior* kelas TP sebesar 0,01325, dan untuk probabilitas *posterior* kelas TL sebesar 0,50701. Maka perbandingannya ditentukan dengan skema bahwa  $P(Y | \text{“Tepat Waktu”}) < P(Y | \text{“Tidak Tepat Waktu”})$ . Karena probabilitas kelas TL lebih besar dari probabilitas kelas TP, maka penelitian dari model prediksi kami menyimpulkan bahwa mahasiswa memiliki keterangan lulus tidak tepat waktu. Rancangan model prediksi itu kemudian diterapkan untuk dilakukan eksperimen dengan *RapidMiner7.6.0*. menggunakan 38 data latih. Hasil pelatihan model kemudian digunakan sebagai pengujian model. Tahap pengujian model dilakukan 2 kali uji coba terhadap 6 data pengujian. Dari pengujian tersebut diperoleh hasil prediksi yang sama persis dengan model. Yakni keterangan bahwa ke-6 lulusan pada data uji merupakan kelulusan tidak tepat

waktu seperti yang ditampilkan pada gambar 4. Adapun dari 2 kali pengujian model itu diperoleh akurasi 100% seperti yang ditunjukkan pada gambar 5.

**Tabel 4.** Probabilitas atribut terhadap kelas

Variabel	Peluang/kejadian		Probabilitas		Variabel	Peluang/kejadian		Probabilitas	
	TP	TL	TP	TL		TP	TL	TP	TL
X <sub>1</sub>	1/2	15/36	0,500	0,417	X <sub>6</sub>	2/2	36/36	1,000	1,000
	1/2	21/36	0,500	0,583		0/2	0/36	0,000	0,000
Total	2/38	36/38	0,053	0,947	Total	2/38	36/38	0,053	0,947
X <sub>2</sub>	2/2	36/36	1,000	1,000	X <sub>7</sub>	2/2	0/36	1,000	0,000
	0/2	0/36	0,000	0,000		0/2	36/36	0,000	1,000
Total	2/38	36/38	0,053	0,947	Total	2/38	36/38	0,053	0,947
X <sub>3</sub>	2/2	35/36	1,000	0,972	X <sub>8</sub>	2/2	0/38	1,000	0,000
	0/2	1/36	0,000	0,028		0/2	36/36	0,000	1,000
Total	2/38	36/38	0,053	0,947	Total	2/38	36/38	0,053	0,947
X <sub>4</sub>	2/2	35/36	1,000	0,972	X <sub>9</sub>	1/2	1/36	0,500	0,028
	0/2	1/36	0,000	0,028		1/2	35/36	0,500	0,972
Total	2/38	36/38	0,053	0,947	Total	2/38	36/38	0,053	0,947
X <sub>5</sub>	2/2	36/36	1,000	1,000					
	0/2	0/36	0,000	0,000					
Total	2/38	36/38	0,053	0,947					

**Pembahasan**

Hasil rancangan model diperoleh total probabilitas untuk kelas lulus ‘tidak tepat waktu’ lebih besar daripada kelas lulus ‘tepat waktu’. Kemudian, rancangan model tersebut diterapkan menggunakan *RapidMiner* yang dikonfigurasi berdasarkan kriteria masing-masing atribut kelas mengacu kepada tabel variabel 1 untuk dilakukan pelatihan. Pelatihan model menggunakan 38 data latih, dan 6 data uji.

ExampleSet (6 examples, 4 special attributes, 9 regular attributes)

Row...	KL	prediction(KL)	confidence(TL)	confidence(TP)	JK	IPS1
1	TL	TL	1.000	0.000	P	IPS1>=3,00
2	TL	TL	1.000	0.000	L	IPS1>=3,00
3	TL	TL	1.000	0.000	P	IPS1>=3,00
4	TL	TL	1.000	0.000	L	IPS1>=3,00
5	TL	TL	1.000	0.000	L	IPS1>=3,00
6	TL	TL	1.000	0.000	P	IPS1>=3,00

IPS2	IPS3	IPS4	SKS4	TA	STUDI	JALUR
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4<3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	S
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M

**Gambar 4.** Hasil pengujian model

accuracy: 100.00%

	true TL	true TP	class precision
pred. TL	6	0	100.00%
pred. TP	0	0	0.00%
class recall	100.00%	0.00%	

**Gambar 5.** Performance vector

Pada tahap pelatihan model, distribusi perbandingan atribut kelulusan diperoleh sebesar 0,053 untuk kelas TP dan 0,947 untuk kelas TL. Nilai itu sesuai dengan probabilitas kelas TP dan TL yang tertuang pada tabel 3 pada hasil rancangan model. Adapun atribut jenis kelamin diperoleh perbandingan sebesar 0,579 perempuan, dan 0,421 laki-laki, pada IPS1 dan IPS4 nilai lebih menonjol atau seluruh data memiliki kriteria  $IPS \geq 3,00$ , dan atribut jumlah SKS semester 4 (SKS4) juga dipenuhi oleh kriteria  $SKS \geq 80$ . Adapun atribut IPS2 dan IPS3 diperoleh 0,974 pada nilai  $\geq 3,00$  dan 0,026 pada nilai  $< 3,00$ , sementara atribut lama masa TA 0,947 untuk kriteria  $TA > 1$  semester, dan 0,053  $\leq 1$  semester, atribut lama masa STUDI 0,947 untuk  $STUDI > 4$  tahun, dan 0,053 untuk  $STUDI \leq 4$  tahun, serta atribut jalur MASUK perguruan tinggi 0,947 untuk kriteria MANDIRI dan 0,053 untuk kriteria SBMPTN. Karena distribusi perbandingan kelas TP dan TL sebesar 0,053 dan 0,947, berarti kelas TP sama dengan 5,30% dan kelas TL sama dengan 94,70% dari 38 data latih. Dengan demikian 5,30% dibagi 100% dikali 38 sama dengan 2,014 atau 2 dan 94,70% dibagi 100% dikali 38 data sama dengan 35,986 atau 36. Sehingga dari 38 data pelatihan diperoleh data kelulusan seperti yang ditampilkan pada grafik kelulusan pada gambar 3.

Kemudian, dari pelatihan model itu dilakukan dua kali pengujian model menggunakan data uji yang terdiri dari 6 data lulusan. Dari pengujian pertama diperoleh hasil bahwa 5 dari 6 data lulusan yang digunakan merupakan lulusan tidak tepat waktu, sementara 1 data tidak terbaca. Namun pada pengujian yang kedua, diperoleh 6 data lulusan tersebut merupakan lulus tidak tepat waktu atau sama persis seperti yang ditunjukkan pada gambar 4, dengan akurasi model menunjukkan 100% seperti yang ditampilkan pada gambar 5.

Penelitian sebelumnya yang relevan dengan penelitian kami menggunakan metode yang sama adalah Nurdiawan & Salim (2018) dengan objek penelitiannya berupa penjualan barang memperoleh akurasi metode sebesar 97,22%, penelitian Khoirunnisa et al. (2021) dengan objek penelitian siswa SMK yang masuk Perguruan Tinggi memperoleh akurasi sebesar 92,40%, penelitian Zed & Muhammad (2018) dengan objek penelitian karyawan untuk prediksi karyawan tetap memperoleh akurasi 94%, dan Syarli & Muin (2016) dengan objek kelulusan calon mahasiswa baru memperoleh akurasi 94%. Sementara penelitian dengan objek dan metode yang sama dengan penelitian ini, seperti Reni, Widodo, & Adhi, (2020), memperoleh 63% akurasi pada kelulusan tepat waktu dan 37% akurasi prediksi kelulusan terlambat, sementara pada penelitian Heryana (2019) memperoleh akurasi 75,67%. Dengan melihat akurasi metode dari model prediksi dalam penelitian ini, serta melihat laporan-laporan penelitian sebelumnya, dapat disimpulkan bahwa menggunakan metode klasifikasi *naive bayes* sangat efektif untuk aplikasi prediksi, dan dapat diterapkan pada kasus-kasus prediksi lainnya.

## SIMPULAN

Hasil penelitian ini menunjukkan performa metode dari model prediksi bekerja sangat baik bahkan menunjukkan akurasi hingga 100%. Perolehan itu dikarenakan konfigurasi atribut data latih yang sesuai pada rancangan model. Dengan kata lain bahwa pendekatan model menjadi lebih utama daripada penerapan model. Sementara, faktor yang memengaruhi kelulusan mahasiswa dalam penelitian ini dipicu oleh atribut indeks prestasi (IP) semester 1, 2, 3, dan 4 dengan kriteria indeks prestasi  $\geq 3,00$ , jumlah SKS Semester 4 dengan kriteria  $\geq 80$  sks, lama masa TA  $\leq 1$  semester, dan lama masa STUDI  $\leq 4$  tahun. Sementara untuk atribut jenis kelamin dan jalur masuk perguruan tinggi tidak terlalu memengaruhi kelulusan mahasiswa.

## REFERENSI

Amalia, N., Shaufiah, S., & Sa'adah, S. (2015). Penerapan Teknik Data Mining untuk Klasifikasi Ketepatan Waktu Lulus Mahasiswa Teknik Informatika Universitas Telkom

- Menggunakan Algoritma Naive Bayes Classifier. *EProceedings of Engineering*, 2(3). <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/1314>
- Andri, R. (2019). *Implementasi Metode Naive Bayes Classifier Untuk Perkiraan Cuaca*. [Skripsi, Universitas Teknologi Yogyakarta] Eprint. <http://eprints.uty.ac.id/2916/>
- BAN-PT. (2019). *Naskah Akademik Instrumen Akreditasi Perguruan Tinggi 3.0*. Badan Akreditasi Nasional Perguruan Tinggi. [https://www.banpt.or.id/wp-content/uploads/2019/09/Lampiran-01-PerBAN-PT-3-2019-Naskah-Akademik-IAPT-3\\_0.pdf](https://www.banpt.or.id/wp-content/uploads/2019/09/Lampiran-01-PerBAN-PT-3-2019-Naskah-Akademik-IAPT-3_0.pdf)
- Guswandi, D., Yanto, M., Hafizh, M., & Mayola, L. (2021). Analisis Hybrid Decision Support System dalam Penentuan Status Kelulusan Mahasiswa. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(6), 1127–1136. <https://doi.org/10.29207/resti.v5i6.3587>
- Han, J., Kamber, M., Pei, J. (2012). *Data mining: Data mining concepts and techniques 3rd Edition*. USA: Morgan Kaufmann Publisher
- Hariato, H., Sunyoto, A., & Sudarmawan, S. (2020). Optimasi Algoritma Naive Bayes Classifier untuk Mendeteksi Anomali dengan Univariate Fitur Selection. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 40–49. <https://doi.org/10.29408/edumatic.v4i2.2433>
- Heryana, D. (2019). *Data Mining Untuk Memprediksi Kelulusan Mahasiswa Pendidikan Matematika UIN Raden Intan Lampung Menggunakan Naive Bayes*. [Skripsi. UIN Raden Intan Lampung]. Repository. <http://repository.radenintan.ac.id/6430/1/Skripsi%20Full.pdf>
- Hurwitz, J., Kirsch, D., (2018). *Machine Learning for Dummies*, IBM Limited Edition. New Jersey: Jhon Wiley & Sons, Inc.
- Jain, K. (2019). *Machine Learning Simolified*. India: Analytics Vidhya Educon Private Limited.
- Javatpoint. (2021). Naive Bayes Classifier in Machine Learning. accessed on November, 28, 2021 from website: <https://www.javatpoint.com/machine-learning-naive-bayes-classifier>
- Joseph, S. K. (2019). *Machine Learning with Python*. India. Tutorials Point (I) Pvt. Ltd., 1–13. <https://store.tutorialspoint.com>
- Kbbi.web.id. (2022). *Lulus*. Diakses pada 11 Januari 2022 dari website: <https://kbbi.web.id/lulus>
- Khoirunnisa, K., Susanti, L., Rokhmah, I. T., & Stianingsih, L. (2021). Prediksi Siswa Smk Al-Hidayah Yang Masuk Perguruan Tinggi Dengan Metode Klasifikasi. *Jurnal Informatika*, 8(1), 26–33. <https://doi.org/10.31294/ji.v8i1.9163>
- Nurdiawan, O., & Salim, N. (2018). Penerapan Data Mining Pada Penjualan Barang Menggunakan Metode Metode Naive Bayes Classifier Untuk Optimasi Strategi Pemasaran. *Jurnal Teknologi Informasi dan Komunikasi*, 13(1), 84-95.
- Pramadhana, D. (2021). Klasifikasi Penyakit Diabetes Menggunakan Metode CFS dan ROS dengan Algoritma J48 Berbasis Adaboost. *Edumatic: Jurnal Pendidikan Informatika*, 5(1), 89–98. <https://doi.org/10.29408/edumatic.v5i1.3336>
- Pratama, R. R. (2020). Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(2), 302–311. <https://doi.org/10.30812/matrik.v19i2.688>
- Prakoso, B. S., Rosiyadi, D., Utama, H. S., & Aridarma, D. (2019). Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(2), 227-232. <https://doi.org/10.29207/resti.v3i2.1042>
- Putra, A. S. (2018). *Klasifikasi Status Gizi Balita Menggunakan Naive Bayes Classification*

- (*Studi Kasus Posyandu Ngudi Luhur*). [Skripsi, Universitas Sanata Dharma Yogyakarta] Repository. [https://repository.usd.ac.id/18332/2/135314092\\_full.pdf](https://repository.usd.ac.id/18332/2/135314092_full.pdf)
- Rahayu, T. M., Ningsi, B. A., Isnurani, I., & Arofah, I. (2021). Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa dengan Metode Naïve Bayes. *MEDIA BINA ILMIAH*, 15(8), 4993–5000.
- Ramadhani, N. D. *Analisis kinerja algoritma naïve bayes classifier dengan seleksi fitur chi-square dan confix stripping stemmer pada sistem klasifikasi terjemahan hadits*. [thesis, Universitas Islam Negeri Syarif Hidayatullah Jakarta]. Repository. <https://repository.uinjkt.ac.id/dspace/handle/123456789/48360>
- Reni, Widodo, & Adhi, B. P. (2020). Sistem Pendukung Keputusan untuk Memprediksi Kelulusan Mahasiswa pada Program Studi Pendidikan Teknik Informatika dan Komputer dengan Menggunakan Algoritma Naïve Bayes. *PINTER: Jurnal Pendidikan Teknik Informatika Dan Komputer*, 4(1), 26–29. <https://doi.org/10.21009/pinter.4.1.6>
- Sari, V., Firdausi, F., & Azhar, Y. (2020). Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 1–9. <https://doi.org/10.29408/edumatic.v4i2.2202>
- Sidiq, R. P., Dermawan, B. A., & Umaidah, Y. (2020). Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes. *Jurnal Informatika Universitas Pamulang*, 5(3), 356–363. <https://doi.org/10.32493/informatika.v5i3.6571>
- Stackoverflow. (2021). *Decision tree vs. Naive Bayes classifier* accessed on December, 9, 2021 from website: <https://stackoverflow.com/questions/10317885/decision-tree-vs-naive-bayes-classifier>
- 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. <https://fikom-unasman.ac.id/ejournal/index.php/jikom/article/view/76>
- Tuhenay, D. (2021). Perbandingan Klasifikasi Bahasa Menggunakan Metode Naïve Bayes Classifier (NBC) Dan Support Vector Machine (SVM). *JIKO (Jurnal Informatika dan Komputer)*, 4(2), 105-111. <https://doi.org/10.33387/jiko.v4i2.2958>
- Wahyudi, I., Bahri, S., & Handayani, P. (2019). Aplikasi Pembelajaran Pengenalan Budaya Indonesia. *Jurnal Teknik Komputer*, 5(1), 71-76. <https://ejournal.bsi.ac.id/ejournal/index.php/jtk/article/view/4470>
- Yunianto, M., Soeparmi, S., Cari, C., Anwar, F., Septianingsih, D. N., Ardyanto, T. D., & Pradana, R. F. Klasifikasi Kanker Paru Paru menggunakan Naïve Bayes dengan Variasi Filter dan Ekstraksi Ciri GLCM. *INDONESIAN JOURNAL OF APPLIED PHYSICS*, 11(2), 256-268. <https://doi.org/10.13057/ijap.v11i2.53213>
- Zed, A., & Muhammad, A. L. I. (2018). Penerapan Metode Naive Bayes untuk Prediksi Menentukan Karyawan Tetap Pada PT. YSP Industries Indonesia. [Skripsi, Sekolah Tinggi Teknologi Pelita Bangsa Bekasi]. <https://ecampus.pelitabangsa.ac.id/pb/AmbilLampiran?ref=22968&jurusan=&jenis=Item&usingId=false&download=false&clazz=ais.database.model.file.LampiranLain>
- Zulfiqar, L. O. M., Renaningtias, N., & Fathoni, M. Y. (2020). Educational Data Mining in Graduation Rate and Grade Predictions Utilizing Hybrid Decision Tree and Naïve Bayes Classifier. *Proceedings of the International Conferences on Information System and Technology 2019*, 151–157. Yogyakarta: SCITEPRESS. <https://doi.org/10.5220/0009907101510157>