



Jurnal Cakrawala Informasi

Journal Homepage: <http://www.itbsemarang.ac.id/sijies/index.php/jci>

e-Mail: jci@itbsemarang.ac.id



Algoritma *Naive Bayes* untuk Prediksi Kelulusan Mahasiswa

Sri Hartati ^{1*}

Haries Anom SAN ²

^{1,2} Informatika, Fakultas Ilmu Komputer, Universitas Bramada Slawi

INFO ARTIKEL

Histori artikel:

Diterima : 15 November 2022
 Revisi : 17 Desember 2022
 Disetujui : 26 Desember 2022
 Publikasi : 30 Desember 2022

Kata kunci:

Prediksi
 Kelulusan
 Mahasiswa
Naive Bayes
 Graduation
 Student

ABSTRACT

Graduation is something that students look forward to. This is very reasonable because the ultimate goal of learning is to pass. One of the assessments at the time of accreditation is student graduation. The more students who graduate on time, the better the score. To find out the number of students who graduate on time, a technique can be used to get precise and accurate results. The technique that researchers use is data mining, which is a technique to extract knowledge. There are several kinds of data mining techniques, including classification, clustering and prediction. The researcher uses naive Bayes prediction technique, to solve the problem. Naive Bayes is an algorithm that is classical, simple and independent. The timely graduation factor is influenced by several factors such as the motivation of the students themselves, the cost factor, the implementation factor of learning on campus, the status of regular students in the morning or evening.

ABSTRAK

Kelulusan merupakan sesuatu yang sangat dinantikan oleh peserta didik. Hal tersebut sangat wajar karena tujuan akhir dari pembelajaran adalah lulus. Salah satu penilaian pada saat akreditasi adalah kelulusan mahasiswa. Semakin banyak jumlah mahasiswa yang lulus tepat waktu, semakin bagus nilainya. Untuk mengetahui jumlah mahasiswa yang lulus tepat waktu dapat menggunakan suatu teknik agar mendapatkan hasil yang tepat dan akurat. Teknik yang peneliti gunakan adalah *data mining*, yaitu suatu teknik untuk mengekstrak ilmu pengetahuan. Macam teknik *data mining* ada beberapa, diantaranya klasifikasi, *clustering*, dan prediksi. Peneliti menggunakan teknik prediksi *Naive Bayes*, untuk menyelesaikan permasalahan. *Naive Bayes* merupakan algoritma yang bersifat klasikal, sederhana, dan independen. Faktor kelulusan tepat waktu

dipengaruhi oleh beberapa faktor seperti motivasi dari mahasiswa itu sendiri, faktor biaya, faktor pelaksanaan pembelajaran di kampus, status mahasiswa regular pagi atau malam. Klasifikasi, *clustering*, dan prediksi. Peneliti menggunakan teknik prediksi *Naive Bayes*, untuk menyelesaikan permasalahan. *Naive Bayes* merupakan algoritma yang bersifat klasikal, sederhana dan, independen. Faktor kelulusan tepat waktu dipengaruhi oleh beberapa faktor seperti motivasi dari mahasiswa itu sendiri, faktor biaya, faktor pelaksanaan pembelajaran di kampus, status mahasiswa regular pagi atau malam.

PENDAHULUAN

Pendidikan adalah proses belajar mengajar yang terjadi antara pendidik dan peserta didik untuk mengembangkan potensi diri [1]. Saat pandemi proses pendidikan dilaksanakan secara daring, banyak kendala yang terjadi karena menggunakan internet yang kadang mengalami susah sinyal. Setelah pandemi berakhir mahasiswa sudah melaksanakan perkuliahan secara tatap muka secara bertahap. Perjuangan panjang yang mereka tempuh selama masa pendidikan akan berganti dengan saat bahagia ketika pengumuman kelulusan. Namun saat terjadi pandemi *Covid-19* di Indonesia, kelulusan menjadi momen yang berbeda, karena tidak adanya upacara sakral secara langsung semua dilakukan secara virtual. Pandemi yang dimulai sejak bulan Maret tahun 2020 sampai dengan saat ini gaungnya masih terasa, sangat berdampak pada semua bidang baik bidang ekonomi, pendidikan, dan infrastruktur yang lain. Sektor pendidikan hal yang langsung dirasakan oleh mahasiswa adalah salah satunya adalah masalah kelulusan [2].

Kelulusan menjadi salah satu nilai penting saat akreditasi pada perguruan tinggi. Oleh sebab itu institusi selalu memberi perhatian lebih pada

mahasiswa tingkat akhir. Jika banyak mahasiswa lulus tepat waktu akan berpengaruh dalam nilai akreditasinya [3]. Hal yang dapat berpengaruh pada kelulusan mahasiswa ada beberapa faktor, seperti faktor status mahasiswa, nilai semester akhir, kondisi ekonomi, dan faktor kampus tempat mahasiswa belajar [4].

Perlu adanya suatu teknik untuk mengatasi permasalahan tersebut agar mahasiswa banyak yang lulus tepat waktu. Peneliti menggunakan teknik *data mining* untuk memprediksi permasalahan tersebut di atas, diharapkan agar data mahasiswa yang lulus tepat waktu jumlah kelulusan mahasiswa per tahun dapat meningkat [5]. *Data mining* suatu metode yang digunakan untuk menghimpun sebuah basis data yang volumenya sangat besar. Teknik ini merupakan untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar [6]. *Data mining* mempunyai empat model yaitu model prediksi, analisis kelompok, analisis asosiasi, dan deteksi anomali [7].

Salah satu algoritma yang digunakan untuk klasifikasi prediksi data adalah *Naive Bayes*. *Naive Bayes* digunakan untuk data yang berbasis probabilitas sederhana, atribut bersifat independen kondisional [7]. Algoritma *Naive Bayes* banyak digunakan pada beberapa bidang seperti pada bidang keuangan, asuransi, kesehatan, pendidikan, pertanian, dan ekonomi. Penelitian dibidang ekonomi oleh Annur (2018) menggunakan algoritma *Naive Bayes* sebagai salah satu teknik klasifikasi untuk mengklasifikasi tingkat kemiskinan di Gorontalo [8]. Hasil menunjukkan angka 73% untuk akurasinya. Penelitian oleh Mukti (2020) tentang tugas akhir menggunakan algoritma *Support Vector Machine* (SVM) untuk prediksi mahasiswa yang tepat waktu atau

terlambat dalam kelulusannya [9]. Teknik tersebut sangat membantu dalam proses administrasi mahasiswa tugas akhir. Penelitian tentang kelulusan mahasiswa sebelumnya sudah dibahas oleh Murtopo (2015), hasil dari penelitiannya bahwa IPK semester 4 sangat berpengaruh dalam menentukan kelulusan tepat waktu [10]. Menurut Nasution (2020) ada beberapa faktor yang berpengaruh dalam menentukan mahasiswa lulus tepat waktu diantaranya pemilihan kelas [11]. Kelas reguler pagi diprediksi cepat lulus daripada reguler malam. Penelitian serupa menggunakan metode Naive Bayes memberikan keuntungan bagi pihak manajemen dapat menaikkan tingkat kelulusan mahasiswa tepat waktu [12]. Berdasarkan uraian tersebut di atas permasalahan kelulusan tepat waktu merupakan hal yang sangat menarik untuk diadakan penelitian, hal tersebut karena berpengaruh terhadap akreditasi suatu perguruan tinggi. Oleh karena itu peneliti mengusulkan untuk melakukan penelitian tentang prediksi kelulusan tepat waktu menggunakan algoritma *Naive Bayes*.

TINJAUAN PUSTAKA

Data Mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu *database* [13], dimana serangkaian proses tersebut melalui beberapa tahapan antara lain:

1. Pembersihan data, yaitu membuang data yang tidak konsisten.
2. Integrasi data, menggabungkan data dari berbagai sumber.
3. Transformasi data, mengubah data sesuai dengan data yang akan diekstrak.
4. Proses ekstrak data dari pola yang akan dipakai.

5. Evaluasi pola yang ditemukan.
6. Presentasi pengetahuan.

Ada beberapa teknik dalam *data mining* [6], yaitu:

1. Klasifikasi

Klasifikasi merupakan pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia.

2. Klastering

Klastering adalah kumpulan objek data yang serupa (atau terkait) satu sama lain dalam grup yang sama atau data berbeda (atau tidak terkait) dengan objek dalam kelompok lain. Dalam klastering terdapat kesamaan antara data sesuai dengan karakteristik yang ditemukan dalam data dan mengelompokkan objek data yang serupa ke dalam kelompok.

3. Asosiasi

Asosiasi merupakan pola sekumpulan item, kejadian atau substruktur yang sering terjadi dalam kumpulan. Dalam asosiasi menemukan keteraturan yang melekat dalam data. Asosiasi digunakan pada data pasar dan transaksi.

4. Algoritma Estimasi

Estimasi adalah metode untuk memperkirakan nilai suatu populasi dengan menggunakan contoh.

Naive Bayes

Teori *Bayes* merupakan pendekatan statistika yang fundamental dalam *data mining*. Pendekatan yang dimaksud adalah pada kuantifikasi *trade off* antara berbagai keputusan klasifikasi yang menggunakan probabilitas [13]. Adapun rumus yang dipergunakan sebagai berikut:

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

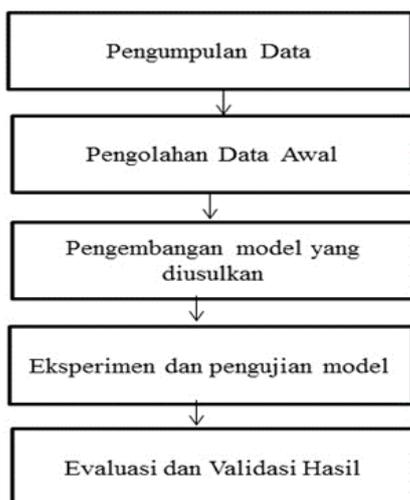
Keterangan:

- X : Data dengan *class* yang belum diketahui
 C_i : Hipotesis dari data X
 $P(C_i|X)$: Probabilitas hipotesis C_i berdasarkan kondisi X
 PC_i : Probabilitas hipotesis C_i
 $P(X|C_i)$: Probabilitas X berdasarkan kondisi hipotesis C_i
 $P(X)$: Probabilitas X

METODE PENELITIAN

Jenis penelitian ini adalah penelitian kuantitatif, yaitu penelitian yang dinyatakan dalam bentuk angka serta dianalisis menggunakan model statistika [14]. Penelitian yang dilakukan saat ini adalah penelitian eksperimen, yaitu penelitian untuk mengetahui pengaruh suatu variabel terhadap variabel lainnya, dengan melakukan pengujian tingkat akurasi suatu algoritma yang kemudian dilakukan pengujian.

Tahap penelitian dimulai dari pengumpulan data, pengolahan data, pengembangan model yang diusulkan, mulai eksperimen dan diakhiri dengan pengujian [15]. Terlihat seperti gambar di bawah:



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Penelitian ini diawali dengan pengambilan *dataset* yang dipakai yaitu data kelulusan mahasiswa STMIK YMI sejumlahnya 321 *record*. Atribut yang dipakai jenis kelamin, status perkawinan, status pekerjaan, IPK1, IPK2, IPK3, IPK4, pekerjaan orang tua, alamat, dan sumber dana.

Pengolahan Data Awal

Preprocessing [16] menghapus data yang kurang penting, hanya diambil yang berguna sebagai parameter prediksi kelulusan mahasiswa. Selanjutnya menyiapkan *data training* untuk proses klasifikasi menentukan kelas yang tepat waktu dan tidak tepat waktu. Apabila hasil probabilitas *class* tepat waktu lebih besar daripada *class* tidak tepat waktu maka termasuk kategori *class* positif. Jika hasil probabilitas *class* tepat waktu lebih kecil daripada *class* tidak tepat waktu maka termasuk kategori *class* negatif.

Pengembangan Metode yang Diusulkan

Untuk memprediksi kelulusan mahasiswa tepat waktu, metode yang diusulkan dalam penelitian ini adalah algoritma *Naive Bayes*. Langkah-langkah penghitungan algoritma *Naive Bayes* adalah sebagai berikut:

1. Menghitung *prior probability* dari setiap *class/label*
 Pada tahap ini digunakan untuk menghitung *prior* probabilitas setiap *class* atau label.
2. Menghitung probabilitas hipotesis (*posteriori probability*)
 Pada tahap ini perlu dicari standar deviasi dari masing-masing *class* variabel yang bernilai numerik menggunakan rumus persamaan sebagai berikut:

$$\text{Standar Deviasi} = \sqrt{\frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n-1}}$$

Akan tetapi apabila atributnya bernilai diskret, maka $P(X_i|C)$ diestimasi sebagai frekuensi relatif dari sampel yang memiliki nilai X_1 sebagai atribut i dalam class C . Jika atribut ke i bersifat *continue* maka $P(X_i|C)$ diestimasi dengan fungsi *densitas gauss* sebagai berikut:

$$f(x|y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Keterangan:

- σ : Standar deviasi
- π : Mean, rata-rata dari seluruh atribut
- X : Nilai atribut i
- e : 2,7183
- f : Peluang
- y : Class yang dicari

Eksperimen dan Pengujian Model

Tahap selanjutnya adalah melakukan pengujian, dari *dataset* yang dipakai yang disimpan pada *Microsoft Excel*. Pada metode *Naive Bayes*, setelah dilakukan eksperimen, selanjutnya membandingkan hasil. Jumlah yang tepat waktu atau tidak tepat waktu menggunakan *likelihood* dan *hob*. Dilanjutkan dengan melakukan evaluasi data dengan menggunakan *confusion matrix* [17] seperti tabel di bawah ini:

Tabel 1. *Confusion Matrix*

Correct Classification	Classification	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Evaluasi dari pengujian dengan tabel di atas, menghasilkan akurasi, presisi, dan *recall*. Rumusnya yang digunakan adalah sebagai berikut:

$$\text{Akurasi} = \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} * 100\%$$

$$\text{Presisi} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (FP_i + TP_i)} * 100\%$$

$$\text{Recall} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FN_i)} * 100\%$$

PEMBAHASAN DAN HASIL

Dari *dataset* mahasiswa lulus sejumlah 321 yang sudah dilakukan *preprocessing* [16] yaitu menghapus data yang tidak diperlukan, pada metode *Naive Bayes* data dibagi menjadi 2 yaitu *data training* dan *data testing*, *data training* sebanyak 291 dan 30 sebagai *data testing*. Kemudian diproses menjadi data tipe numerik agar dapat diproses lebih lanjut menggunakan *Naive Bayes*. *Data training* sebagai berikut:

Tabel 2. *Data Training*

NO	Jenis Kelamin	Status Perkawinan	Status Pekerjaan	IPKSM1	IPKSM2	IPKSM3	IPKSM4	Pekerjaan Orang Tua	Alamat	asal sekolah	Sumber dana
1	Laki-laki	Menikah	Kerja	2.45	2.45	2.45	Kurang	PNS	Kab. Tanah	80	Mandiri
2	Perempuan	Belum Menikah	Belum	3.00	3.00	3.00	Sangat Bagus	PNS	Luar	90	Orang Tua
3	Perempuan	Belum Menikah	Kerja	3.24	3.24	3.24	Sangat Bagus	PNS	Luar	75	Mandiri
4	Perempuan	Belum Menikah	Belum	3.50	2.48	2.48	Sangat Bagus	PNS	Luar	80	Orang Tua
-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-
288	Laki-laki	Menikah	Kerja	Kurang	Kurang	Kurang	Kurang	Swastraku	Tengah	90	Mandiri
289	Laki-laki	Menikah	Belum	Kurang	Kurang	Kurang	Kurang	Swastraku	Tengah	90	Mandiri
290	Laki-laki	Menikah	Kerja	Kurang	Kurang	Kurang	Kurang	Swastraku	Luar	90	Mandiri
291	Perempuan	Belum Menikah	Kerja	Kurang	Kurang	Kurang	Kurang	Wiraswastaku	Luar	80	Mandiri

Algoritma Naive Bayes

Langkah selanjutnya mencari nilai setiap *class* dan nilai setiap atribut dengan langkah-langkahnya adalah sebagai berikut:

1. Menghitung *prior* probabilitas dari setiap *class* yang ada

Jumlah kasus untuk jenis kelamin, status perkawinan, status pekerjaan, pekerjaan orang tua, alamat, dan sumber dana, IPK SM1,

IPK2, IPK3, IPK4. Ada dua *class* yaitu tepat waktu dan tidak tepat waktu. Untuk mendapatkan probabilitas dari atribut yang ada, adalah sebagai berikut:

Diketahui:

Data *class* yang lulus tepat waktu : 197
 Data *class* yang lulus tidak tepat waktu : 94
 Jumlah keseluruhan : 291

2. Menentukan nilai setiap atribut

a. Mencari atribut dari data yang berjenis kategori yaitu jenis kelamin, status perkawinan, status perkawinan, status pekerjaan, pekerjaan orang tua, dan sumber dana. Adapun cara penghitungannya sebagai berikut:

Atribut jenis kelamin:

- Atribut *class* tepat waktu
 - Laki-laki = $69/197 = 0,35$
 - Perempuan = $128/197 = 0,65$
 - Jumlah data keseluruhan = 197
- Atribut *class* tidak tepat waktu
 - Laki-laki = $91/94 = 0,97$
 - Perempuan = $3/94 = 0,03$
 - Jumlah data keseluruhan = 94

Dengan cara yang sama seperti di atas, digunakan untuk mencari nilai atribut yang lainnya.

b. Mencari atribut dari data yang berjenis numerik yaitu IPK1, IPK2, IPK3, dan IPK4. Pada tahap ini dicari *mean* dan standar deviasi dari masing-masing *class* variabel untuk mencari kelas tepat waktu dan tidak tepat waktu menggunakan persamaan sebagai berikut:

$$\text{Standar Deviasi} = \sqrt{\frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n - 1}}$$

Dengan menggunakan rumus tersebut di atas, berikut perhitungan standar deviasi IPK SM1 kategori *class* tepat waktu:

$$S = \sqrt{\frac{409446,41 - 2078,41}{197 - 1}} = 0,31$$

Dengan menggunakan rumus tersebut di atas, berikut perhitungan standar deviasi IPK SM1 kategori *class* tidak tepat waktu:

$$S = \sqrt{\frac{50789,34 - 970,34}{94 - 1}} = 0,23$$

Untuk perhitungan standar deviasi IPK2, IPK3, IPK4 kategori *class* tepat waktu dan tidak tepat waktu caranya sama dengan di atas. Untuk lebih jelasnya berikut tabel yang menjelaskan *mean* dan standar deviasi IPK SM1, IPK SM2, IPK SM3, IPK4.

Tabel 3. *Mean* dan Standar Deviasi untuk IPK, SM1, SM2, SM3, SM4

	IPK1		IPK2		IPK3		IPK4	
	Tepat	Tidak tepat						
Mean	3.24	1.75	3.18	1.78	3.19	1.78	3.45	1.74
Standar Deviasi	0.31	0.23	0.41	0.28	0.41	0.28	0.17	0.22

c. Menghitung nilai probabilitas dari data variabel yang bersifat *continue* yaitu IPK1, IPK2, IPK3, IPK4. Apabila atribut ke *i* bernilai diskret, maka $P(X_i|C)$ diestimasi sebagai frekuensi relatif dari sampel yang ada nilai *X* sebagai nilai *i* dalam *class C* dan apabila atribut ke *i* bersifat *continue* maka $P(X_i|C)$, diestimasi dengan fungsi densitas *gauss*. Rumusnya sebagai berikut:

$$f(x|y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Kasus berikut berupa data yang belum diketahui keterangan lulusnya.

Tabel 4. Data yang Belum Diketahui Keterangan Lulusnya

Jenis	Status	Status	IPK1	IPK2	IPK3	IPK4	Ket
Kelamin	Perkawinan	Pekerjaan					lulus
Perempuan	Belum Menikah	Belum	3.0	3.12	3.14	3.43	?

- Menghitung probabilitas IPK1 tepat waktu

Diketahui:

Standar deviasi $\sigma = 0,31$

Mean $\pi = 3,24$

$$f(\text{ips1} = 3,00 | \text{tepat waktu})$$

$$= \frac{1 \times 2,7183 \frac{(2,76-3,24)^2}{2(0,31)^2}}{\sqrt{2 \times 3,24(0,31)}} = 0,832$$

- Menghitung probabilitas IPK1 tidak tepat waktu

Diketahui:

Standar deviasi $\sigma = 1,75$

Mean $\pi = 0,23$

$$f(\text{ips1} = 3,00 | \text{tidak tepat waktu})$$

$$= \frac{1 \times 2,7183 \frac{(2,67-0,23)^2}{2(0,31)^2}}{\sqrt{2 \times 0,23(1,75)}} = 0,092$$

- d. Membandingkan hasil klas tepat waktu dan *class* tidak tepat waktu. Untuk membandingkan hasil yang tepat waktu dan tidak tepat waktu menggunakan rumus *likelihood* dan *hob*. Untuk caranya seperti di bawah ini:

- *Likelihood* tepat waktu = (jenis kelamin × status perkawinan × status perkawinan × status pekerjaan × IPK SM1 × IPK SM2 × IPK SM3 × IPK SM4 × pekerjaan orang tua × sumber dana) = $(0,30 \times 0,89 \times 0,47 \times 0,83 \times 0,52 \times 0,46 \times 0,003 \times 0,16 \times 0,50 \times 0,08) = 0,000002318$.

- *Likelihood* tidak tepat waktu = (jenis kelamin × status perkawinan × status perkawinan × status pekerjaan × IPK SM1 × IPK SM2 × IPK SM3 × IPK SM4 × pekerjaan orang tua × sumber dana = $(0,03 \times 0,06 \times 0,41 \times 0,09 \times 0,30 \times 0,74 \times 0,07 \times 0,16 \times 0,53 \times 0,01) = 0,000000001$.

- *Hob* tepat waktu = $0,000002318 / (0,000002318 + 0,000000001) = 0,999622571$.

- *Hob* tidak tepat waktu = $0,000000001 / (0,000000001 + 0,000002318) = 0,000377429$.

Apabila *hob* tepat waktu lebih besar daripada *hob* tidak tepat waktu, maka diprediksi tepat waktu. Namun sebaliknya apabila *hob* tepat waktu lebih kecil daripada *hob* tidak tepat waktu, maka diprediksi tidak tepat waktu. Dari contoh kasus di atas diprediksi bahwa mahasiswa tersebut hasilnya adalah tepat waktu. *Data testing* tersebut di atas digunakan untuk menguji akurasi metode *Naive Bayes* dengan menggunakan perhitungan *likelihood* dan *hob*. Penghitungan *likelihood* dan *hob* digunakan untuk mendapatkan hasil prediksi dan

sebagai dasar untuk pengujian pada tabel *confusion matrix*, sehingga terlihat jumlah yang *True Positif* (TP), *True Negatif* (TN), *False Negatif* (FN), dan *True Negatif* (TN).

Pengujian *Confusion Matrix*

Confusion matrix merupakan suatu perhitungan yang menghasilkan akurasi, presisi, dan *recall*. Dimana akurasi adalah persentasi ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian, presisi adalah kasus yang diprediksi positif yang juga positif benar pada data sebenarnya, sedangkan *recall* merupakan kasus positif yang sebenarnya yang diprediksi positif secara benar [4]. *Confusion matrix* rumus penghitungannya sebagai berikut:

$$\text{Akurasi} = \frac{(\text{TN} + \text{TP})}{(\text{TN} + \text{FN} + \text{TP} + \text{FP})} \times 100\%$$

$$\text{Presisi} = \frac{\text{TP}}{(\text{TP} + \text{TN})} \times 100\%$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \times 100\%$$

CONFUSION MATRIX

CORRECT CLASSIFICATION	CLASSIFICATION	
	POSITIF	NEGATIF
POSITIF	22	3
NEGATIF	3	2

Hasil Pengujian

$$\text{Akurasi} = 80,00\%$$

$$\text{Presisi} = 88,00\%$$

$$\text{Recall} = 88,00\%$$

Dari tabel tersebut di atas, terlihat hasilnya akurasi 80%. Berarti ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian nilainya 80,00%, presisi 88,00%, berarti kasus yang diprediksi positif yang juga positif benar pada data sebenarnya nilainya 88,00%. *Recall* 88,00%, berarti kasus positif yang

sebenarnya yang diprediksi positif secara benar nilainya 88,00%.

KESIMPULAN

Dari *dataset* kelulusan mahasiswa sejumlah 321 terlihat akurasi 80%, nilai tersebut perlu ditingkatkan agar mencapai hasil yang lebih maksimal. Pada penelitian kali ini terlihat bahwa atribut yang berpengaruh adalah sumber dana. Sumber dana ada beberapa yaitu mandiri dan beasiswa. Dengan demikian metode *Naive Bayes* sangat bagus digunakan untuk memprediksi suatu permasalahan.

Saran

Secara keseluruhan metode *Naive Bayes* untuk prediksi kelulusan mahasiswa yang diusulkan sudah menunjukkan hasil yang sangat baik. Namun agar hasil tersebut lebih maksimal dapat dilakukan pada penelitian selanjutnya dengan menambahkan *feature filtering data*. Banyak *feature filtering data* yang dapat dipakai untuk mengoptimalkan hasil seperti *information gain*. Pada penelitian selanjutnya akan dilakukan penambahan dengan *feature filtering information gain*.

DAFTAR PUSTAKA

- [1] H. Habe and A. Ahiruddin, "Sistem Pendidikan Nasional," *Ekombis Sains J. Ekon. Keuangan, dan Bisnis*, vol. 2, no. 1, pp. 39–45, 2017, doi: 10.24967/ekombis.v2i1.48.
- [2] R. H. S. Aji, "Dampak Covid-19 pada Pendidikan di Indonesia: Sekolah, Keterampilan, dan Proses Pembelajaran," *Salam J. Sos. dan Budaya Syar'i*, vol. 7, no. 2, 2020, doi: 10.15408/sjsbs.v7i5.15314.
- [3] Y. Apridiansyah, N. D. M. Veronika, and E. D. Putra, "Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes," *JSAI (Journal Sci.*

- Appl. Informatics*), vol. 4, no. 2, pp. 236–247, 2021, doi: 10.36085/jsai.v4i2.1701.
- [4] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, “Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naive Bayes: Systematic Review,” *Fakt. Exacta*, vol. 13, no. 1, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [5] E. Siswanto, “Optimasi Metode Naive Bayes dalam Memprediksi Tingkat Kelulusan Mahasiswa STEKOM Semarang,” *Jurikom (Jurnal Ris. Komputer)*, vol. 6, no. 1, pp. 1–6, 2019, doi: 10.30865/jurikom.v6i1.1038.
- [6] J. Han, M. Kamber, and J. Pei, *Data Mining, Concepts and Techniques*, Third Edit. Massachusetts: Morgan Kaufmann Publishers, 2012.
- [7] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Yogyakarta, 2012.
- [8] H. Annur, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.
- [9] Y. I. Mukti, “Sistem Prediksi Lulus Tepat Waktu Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM),” *JUTIM (Jurnal Tek. Inform. Musirawas)*, vol. 5, no. 2, pp. 110–115, 2020, doi: 10.32767/jutim.v5i2.1050.
- [10] A. A. Murtopo, “Prediksi Kelulusan Tepat Waktu Mahasiswa STMIK YMI Tegal Menggunakan Algoritma Naive Bayes Time Graduation Prediction by Using Naive Bayes Algorithm at STMIK YMI Tegal,” vol. 7, pp. 145–154, 2015.
- [11] T. Nasution, “Implementasi Algoritma k-Nearest Neighbor untuk Penentuan Kelulusan Mahasiswa Tepat Waktu,” *J. Perangkat Lunak*, vol. 2, no. 1, pp. 1–14, 2020, doi: 10.32520/jupel.v2i1.944.
- [12] S. Salmu and A. Solichin, “Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta,” *Pros. Semin. Nas. Multidisiplin*, pp. 701–709, 2017.
- [13] R. T. Vlandari, *Data Mining: Teori dan Aplikasi Rapidminer*. Yogyakarta: Gava Media, 2017.
- [14] M. B. U. B. Arifin and Nurdyansyah, *Buku Ajar Metodologi Penelitian Pendidikan*. Sidoarjo: UMSIDA Press, 2018.
- [15] M. F. Abdi, S. Y. Qodarbaskoro, A. Alfani, K. Kusriani, and D. Maulina, “Klasifikasi Pembagian Arus Lalu Lintas Menggunakan Algoritma Naive Bayes dan Model Linear,” *Tecnol. J. Ilm.*, vol. 12, no. 4, 2021, doi: 10.31602/tji.v12i4.5626.
- [16] S. Zhang, C. Zhang, and Q. Yang, “Data Preparation for Data Mining,” *Appl. Artif. Intell.*, vol. 17, pp. 375–381, 2003, doi: 10.1080/08839510390219264.
- [17] L. D. Utami and R. S. Wahono, “Integrasi Metode Information Gain untuk Seleksi Fitur dan Adaboost untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naive Bayes,” *J. Intell. Syst.*, vol. 1, no. 2, pp. 120–126, 2015.