

PREDIKSI KELULUSAN MAHASISWA DENGAN NAÏVE BAYES DAN FEATURE SELECTION INFORMATION GAIN *STUDENT GRADUATION PREDICTION WITH NAÏVE BAYES AND FEATURE SELECTION INFORMATION GAIN*

Sri Hartati^{*1}, Nur Ariesanto Ramdhan², Haries Anom SAN³

1, Informatika, Fakultas Ilmu Komputer, Universitas Bhamada Slawi

2, Teknik Informatika, Fakultas Teknik Informatika, Universitas Muhadi Setiabudi Brebes,

3 Informatika, Fakultas Ilmu Komputer, Universitas Bhamada Slawi

*e-mail: ^{*1}sri.bhamada1305@gmail.com, ²ariesantoramdhan@gmail.com, ³anom.haries@gmail.com*

ABSTRAK

Kelulusan seorang mahasiswa dipengaruhi oleh banyak faktor, diantaranya kondisi ekonomi keluarga, nilai mahasiswa atau karena faktor lain yang berhubungan dengan tempat mahasiswa belajar. Kelulusan merupakan salah satu nilai penting pada saat proses akreditasi suatu perguruan tinggi. Oleh karena itu apabila mahasiswa banyak yang lulus tepat waktu akan mempengaruhi nilai akreditasinya. Permasalahan tersebut di atas harus segera diatasi dengan suatu metode. Data mining salah satu metode yang paling tepat untuk mengatasi masalah tersebut di atas. Suatu keilmuan yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar. Permasalahan kelulusan tepat waktu merupakan hal yang prioritas pada suatu perguruan tinggi. Oleh karena itu peneliti mengusulkan untuk melakukan mengembangkan penelitian tentang Prediksi kelulusan tepat waktu yang semula hanya menggunakan metode naive bayes, peneliti menambahkan feature selection information gain sebagai feature untuk menyeleksi atribut yang berbobot.

Kata kunci _ Naive bayes, Feature Selection, Information Gain, kelulusan

Abstract

A student's graduation is influenced by many factors, including family economic conditions, student grades or because of other factors related to where students study. Graduation is one of the important values during the accreditation process of a university. Therefore, if many students graduate on time, it will affect the accreditation value. To overcome the problems mentioned above, a technique is needed to predict student graduation. Data mining is one of the most appropriate methods to collect a database with a very large capacity. Data mining is a discipline that studies methods for extracting knowledge or finding patterns from large data. The problem of timely graduation is a priority at a university. Therefore, the researcher proposes to develop research on the prediction of on-time graduation which originally only used the Naive Bayes algorithm, the researcher added a feature selection information gain as a feature to select a weighted attribute.

Keywords _ Naive bayes, Feature Selection, Information Gain, graduation

PENDAHULUAN

Akhir studi merupakan *moment* yang sangat ditunggu oleh siswa atau mahasiswa sebagai *goal* dari perjalanan studi mereka. Sejak adanya wabah pandemi sejak bulan maret tahun 2020 yan sampai saat ini masih terasa dampaknya pada semua bidang, baik bidang ekonomi, pendidikan

Informasi Artikel:

Submitted: November 2022, **Accepted:** November 2022, **Published:** November 2022

ISSN: 2685-4902 (media online), **Website:** <http://jurnal.umus.ac.id/index.php/intech>

dan infra struktur yang lain. Pada bidang pendidikan hal yang langsung dirasakan oleh mahasiswa atau siswa adalah masalah kelulusan [1].

Kelulusan juga merupakan salah satu nilai penting pada saat proses akreditasi suatu perguruan tinggi. Oleh karena itu apabila mahasiswa banyak yang lulus tepat waktu akan mempengaruhi nilai akreditasinya [2]. Kelulusan mahasiswa juga dipengaruhi oleh banyak faktor, seperti kondisi ekonomi keluarga, nilai mahasiswa atau karena faktor lain yang berhubungan dengan kampus tempat mahasiswa belajar [3].

Untuk mengatasi permasalahan tersebut di atas, maka diperlukan suatu teknik untuk memprediksi kelulusan mahasiswa. Penulis menggunakan teknik data mining untuk memprediksi kelulusan mahasiswa, diharapkan data mahasiswa yang lulus tepat waktu jumlah kelulusan mahasiswa per tahun dapat meningkat [4].

Data mining salah satu metode yang paling tepat untuk menghimpun sebuah basis data yang berkapasitas sangat besar. Data mining adalah disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar [5]. Pekerjaan yang berkaitan dengan data mining terdiri atas empat kelompok yaitu model prediksi, analisis kelompok, analisis asosiasi, dan deteksi anomali [6].

Algoritma *naive bayes* merupakan algoritma klasifikasi untuk prediksi berbasis *probabilistik* sederhana, atribut bersifat *independen* kondisional (yaitu, tidak ada hubungan ketergantungan antar atribut). Dengan kata lain *naive bayes* merupakan model fitur yang *independen* [6]. Algoritma *naive bayes* sudah banyak digunakan pada beberapa bidang untuk mendukung suatu keputusan, seperti pada bidang keuangan, asuransi, kesehatan, pendidikan dan lainnya. Tahapan bekerja dengan algoritma *naive bayes* ada empat yaitu baca data training, hitung jumlah class, hitung jumlah kasus yang sama dengan class yang sama, kalikan semua nilai hasil sesuai dengan data X yang dicari class-nya [6].

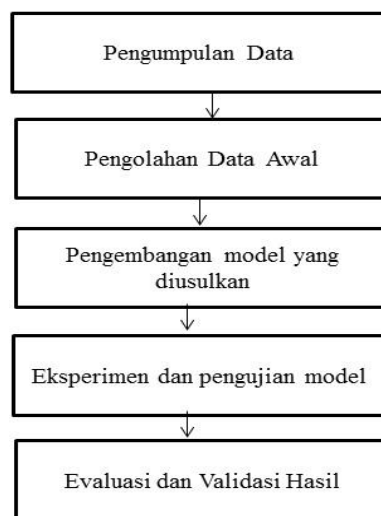
Penelitian ini mengacu pada penelitian sebelumnya yang membahas tentang hal serupa, yaitu penelitian tentang prediksi tingkat kelulusan mahasiswa STEKOM Semarang menggunakan algoritma *naive bayes* [4], penelitian tentang kelulusan mahasiswa Stikom Tunas Bangsa menggunakan algoritma C45 [7] dan penelitian tentang kelulusan tepat waktu menggunakan neural network dan visualisasi data [8] serta penelitian tentang prediksi stroke menggunakan metode KNN [9].

Berdasarkan uraian tersebut di atas, permasalahan kelulusan tepat waktu merupakan hal yang prioritas pada suatu perguruan tinggi. Oleh karena itu peneliti mengusulkan untuk melakukan mengembangkan penelitian tentang Prediksi kelulusan tepat waktu yang semula hanya menggunakan algoritma *naive bayes*, peneliti menambahkan *feature selection information gain* sebagai feature untuk menyeleksi atribut yang berbobot.

METODE PENELITIAN

Jenis penelitian yang dilakukan adalah *experimen*, melakukan pengujian menggunakan suatu algoritma dalam hal ini peneliti menggunakan *naive bayes* dan *filtering feature selection information gain*. Supaya hasilnya lebih maksimal dalam pembuatan penelitian ini perlu ada tahapan dari pengumpulan data, pengolahan data awal, pengembangan model, eksperimen dan evaluasi [10].

Tahapan Metodologi Penelitian



Gambar 1 Tahapan Metodologi Penelitian

1. Pengumpulan Data

Sebelum melaksanakan penelitian, proses pengumpulan data diperlukan agar mendapatkan hasil yang lebih baik. Adapun teknik pengumpulan data menurut [10] ada 3 cara yaitu : wawancara, dokumentasi dan tinjauan pustaka.

2. Pengolahan Data Awal

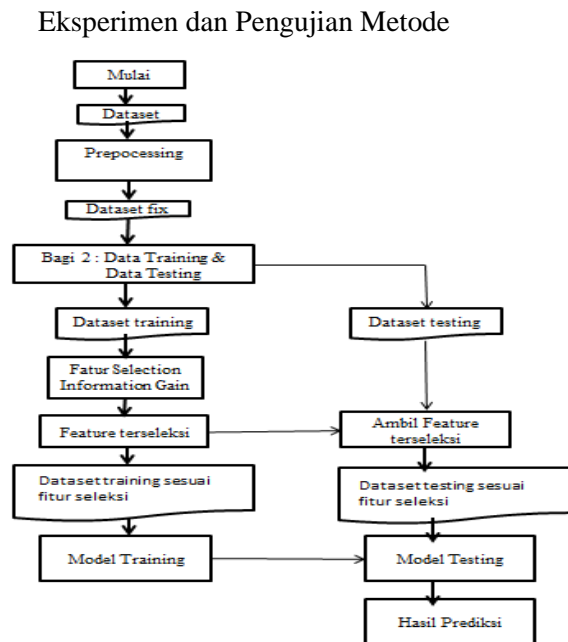
Setelah data diperoleh kemudian dilakukan *preprocessing* [11] yaitu melakukan penghapusan data yang kurang penting dan hanya mengambil yang berguna sebagai parameter prediksi kelulusan mahasiswa. Selanjutnya menyiapkan data training untuk proses klasifikasi menentukan kelas yang tepat waktu dan tidak tepat waktu.

3. Pengembangan Model Yang Diusulkan

Pada penelitian ini, peneliti menambahkan *feature* untuk mendapatkan hasil yang lebih maksimal. Peneliti membahas tentang prediksi kelulusan mahasiswa menggunakan metode algoritma *naïve bayes*, seperti [12] diharapkan dengan menambah suatu *feature* akan mendaptkan hasil yang lebih baik. *Feature* yang akan peneliti tambahkan adalah *Feature selection Inforation Gain*.

4. Eksperimen dan Pengujian Metode

Pada tahap ini menjelaskan teknik pengujian dalam penelitian dengan modeling dari dataset *privat* yaitu dataset kelulusan mahasiswa STMIK YMI Tegal yang disimpan pada *microsoft excel*. *Feature selection* yang diusulkan adalah metode jenis filter, yaitu *information gain* yang digunakan secara integrasi agar akurasi algoritma *naive bayes* dapat meningkat [13], seperti terlihat pada gambar di bawah ini :



Gambar 2 Alur eksperimen dan pengujian metode

Langkah awal dari penelitian ini adalah pengambilan dataset yaitu dataset kelulusan mahasiswa STMIK YMI sejumlah 321 . Setelah data diperoleh kemudian dilakukan *preprocessing* [11] yaitu melakukan penghapusan data yang kurang penting dan hanya mengambil yang berguna sebagai parameter prediksi kelulusan mahasiswa. Sebelum menghitung algoritma *Information Gain* terlebih dahulu harus mengetahui nilai *entropy* masing-masing.

Selanjutnya menyiapkan data training untuk proses klasifikasi menentukan kelas yang tepat waktu dan tidak tepat waktu berdasarkan nilai hitungan probabilitas dari rumus bayes yang lebih besar. Cara penghitungan algoritma *Naive bayes* sebagai berikut :

- a. Menghitung prior *probability* dari setiap kelas/label

Pada tahap ini digunakan untuk menghitung prior probabilitas setiap klas (1) :

$$P(H|X) = \frac{P(H|X).P(H)}{P(X)} \quad (1)$$

- b. Menghitung probabilitas *hipotesis* (*posteriori probability*)

Pada tahap ini perlu dicari *standar deviasi* dari masing-masing *class variabel* yang bernilai *numeric* menggunakan rumus persamaan (2) :

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

Akan tetapi apabila atributnya bernilai diskret, maka $P(x_i|C)$ diestimasi sebagai frekuensi relatif dari sampel yang memiliki nilai x_i sebagai atribut i dalam *class* C . Jika atribut ke i bersifat *continue* maka $P(x_i|C)$ diestimasi dengan fungsi *densitas gauss* dengan rumus (3) :

$$f(x|y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

Keterangan :

σ : standar deviasi

e : 2,7183

π : Mean, rata-rata dari seluruh atribut

f : peluang

x : nilai atribut i

y : kelas yang dicari

c. Membandingkan hasil klas

Tahap ini membandingkan hasil penentuan tepat waktu maupun tidak tepat waktu sehingga diperoleh satu prediksi kelulusan menggunakan *likelihood* dan *Hob*.

5. Evaluasi dan Validasi Hasil

Evaluasi terhadap dataset kelulusan mahasiswa dilakukan dengan menggunakan *Confusion Matrix* yang berisi AUC [13]. Bentuk tabel *Confusion Matrix* seperti dibawah ini :

Tabel 1. *Confusion matrix*

Correct Classification	Classification	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan pada tabel di atas

1. TP adalah True Positif, jumlah data positif yang terklasifikasi dengan benar oleh sistem
2. TN adalah True Negatif, jumlah data negatif yang terklasifikasi dengan benar oleh sistem
3. FN adalah False Negatif, jumlah data negatif namun terklasifikasi salah oleh sistem
4. FP adalah False Positif, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Untuk meningkatkan *performance* pada metode *Naive bayes*, maka dilakukan dengan menggabungkan metode *feature selection information gain*. Diharap hasil akan menjadi lebih baik daripada yang hanya menggunakan *naive bayes* saja.

HASIL DAN PEMBAHASAN

1. Metode Feature Selection Information Gain

Langkah awal dari penelitian ini adalah menentukan *feature* yang terbaik. Jika sudah mendapatkan *feature* yang terbaik maka digunakan untuk penghitungan pada metode *Naive Bayes*. Sebelum menghitung *Information Gain* terlebih dahulu harus mengetahui nilai *entropy* total dan *entropy* masing-masing atribut.

a. Menghitung Total Entropy

Pada data ini atribut label ada 2 yaitu tepat waktu dan tidak tepat waktu. Dimana Jumlah data kelas tepat waktu ada 197, data kelas tidak tepat waktu 94. Jumlah data keseluruhan 291. Untuk mengklasifikasikan kelas tersebut caranya (3)

$$Entropy\ Total = \left(-\frac{197}{291}\right) \log_2 \left(\frac{197}{291}\right) + \left(-\frac{94}{291}\right) \log_2 \left(\frac{94}{291}\right) = 0,90764$$

b. Menghitung Masing-masing Entropy Atribut

Langkah selanjutnya adalah menghitung *entropy* masing-masing atribut, yaitu Jenis kelamin, Status perkawinan, Status pekerjaan, IPK SM1, IPK SM2, IPK SM3, IPK SM4, Pekerjaan Orang Tua, Alamat, Total SKS SM 4 dan Sumber Dana.

Dibawah ini contoh penghitungan untuk atribut jenis kelamin :

<p>a. Laki-laki</p> <p>Diketahui :</p> <p>Jumlah data kelas tepat waktu 69</p> <p>Jumlah data kelas tidak tepat waktu 91</p> <p>Jumlah data keseluruhan Laki-laki : 160</p>	<p>b. Perempuan</p> <p>Diketahui :</p> <p>Jumlah data kelas tepat waktu 128</p> <p>Jumlah data kelas tidak tepat waktu 3</p> <p>Jumlah data keseluruhan perempuan 131</p>
<p>Jumlah entropy =</p> $\left(-\frac{69}{160}\right) \log_2 \left(\frac{69}{160}\right) + \left(-\frac{91}{160}\right) \log_2 \left(\frac{91}{160}\right) = 0,98632$	<p>Jumlah entropy =</p> $\left(-\frac{128}{131}\right) \log_2 \left(\frac{128}{131}\right) + \left(-\frac{3}{131}\right) \log_2 \left(\frac{3}{131}\right) = 0,15743$

Jadi Gain Jenis Kelamin =

Gain Jenis Kelamin

$$= 0,90763974 - \left(\left(\frac{160}{291} \times 0,98632 \right) + \left(\frac{131}{291} \times 0,15743 \right) \right) = 0,294$$

Dengan cara yang sama untuk menghitung atribut yang lain, seperti Status pekerjaan, IPK SM1, IPK SM2, IPK SM3, IPK SM4, Pekerjaan Orang Tua, Alamat, Sumber Dana.

c. Perangkingan Bobot Nilai Gain Dari Tiap Atribut

Berdasarkan penghitungan *gain* atribut di atas, maka diperoleh hasil pembobotan sebagai berikut.

Tabel 2. Hasil pembobotan *Gain* atribut

NO	ATRIBUT PARAMETER	NILAI GAIN	RANKING
1	Sumber Dana	1	1
2	Alamat	0,868	2
3	IPK SM 4	0,789	3
4	Status Perkawinan	0,638	4
5	IPK SM 1	0,595	5
6	IPK SM 3	0,415	6
7	IPK SM 2	0,415	7
8	Jenis Kelamin	0,294	8
9	Status Pekerjaan	0,049	9
10	Pekerjaan Orang Tua	0	11

Dari tabel tersebut di atas, diambil 8 atribut terbaik diurutkan dari yang bernilai tertinggi ke yang terendah, digunakan pada implementasi *naive bayes*.

2. Metode *Naive Bayes*

Dataset sejumlah 321 dibagi menjadi 2 yaitu data training dan data testing. Data training sebanyak 291 dan data testing 30. Data training untuk melatih algoritma klasifikasi dalam hal ini *naive bayes* sedangkan data testing untuk menguji algoritma klasifikasi (*naive bayes*) yang telah dilatih [14].

Setelah mendapatkan bobot dari tiap atribut pada *seleksi information gain*, langkah selanjutnya mencari nilai setiap kelas dan nilai setiap atribut dengan metode *naive bayes* adapun rumus sebagai berikut :

- a. Menghitung Prior Probabilitas dari setiap kelas yang ada.

Data kelas yang lulus tepat waktu 197	Data kelas yang lulus tidak tepat waktu 94
Jumlah keseluruhan 291	

- b. Menentukan nilai setiap atribut

- b.1. Mencari atribut dari data yang berjenis kategori yaitu jenis kelamin, status perkawinan, status perkawinan, status pekerjaan, pekerjaan orang tua dan sumber dana. Adapun cara penghitungan Atribut Jenis Kelamin sebagai berikut :

1. Atribut Kelas Tepat Waktu	2. Atribut Kelas Tidak Tepat Waktu
a. Laki-laki : $69/197 = 0,35$	a. Laki-laki : $91/94 = 0,97$
b. Perempuan : $128/197 = 0,65$	b. Perempuan : $3/94 = 0,03$
Jumlah data keseluruhan : 197	Jumlah data keseluruhan : 94

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

- b.2. Mencari atribut dari data yang berjenis numerik yaitu IPK SM1, IPK SM2, IPK SM3 dan IPK SM4. Pada tahap ini dicari mean dan standar deviasi dari masing-masing klas variabel untuk mencari kelas tepat waktu dan tidak tepat waktu menggunakan persamaan sebagai berikut :

Mencari standar deviasi IPK SM1 dengan kategori kelas tepat waktu :

$$S = \sqrt{\frac{409446,41 - 2078,41}{197-1}} = 0.31$$

Mencari standar deviasi IPK SM1 dengan kategori kelas tidak tepat waktu

$$S = \sqrt{\frac{50789,34 - 970,34}{94-1}} = 0.23$$

Untuk perhitungan standar deviasi IPKSM2, IPKSM3, IPKSM4 kategori kelas tepat waktu dan tidak tepat waktu caranya sama seperti di atas. Untuk lebih jelasnya berikut tabel yang menjelaskan mean dan standar deviasi IPK SM1, IPK SM2, IPK SM3, IPK SM4.

Tabel 3. Mean dan standar deviasi

	IPK1		IPK2		IPK3		IPK4	
	Tepat	Tidak tepat	Tepat	Tidak tepat	Tepat	Tidak tepat	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

- b.3. Menghitung nilai probabilitas dari data variabel yang bersifat kontinue yaitu IPK SM1, IPK SM2, IPK SM3, IPK SM4.

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 klas C dan apabila atribut ke i bersifat kontinue maka $P(x_i | C)$, diestimasi dengan fungsi densitas gauss. Seperti contoh kasus di bawah ini.

Tabel 4. Kasus yang belum diketahui keterangan lulusnya

Jenis Kelamin	Status Perkawinan	Status Pekerjaan	IPK Sm1	IPK Sm2	IPK Sm3	IPK Sm4	Keterangan Lulus
Perempuan	Belum Menikah	Belum	3.0	3.12	3.14	3.43	?

1. Menghitung probabilitas IPK SM1 Tepat waktu

Diketahui

$$\begin{aligned}
 &\text{Standar deviasi } \sigma = 0.31 \\
 &\text{Mean } \pi = 3.24 \\
 &f(\text{ips1} = 3.00 \mid \text{tepat waktu}) = \frac{1 * 2.7183}{\sqrt{2 * 3.24(0.31)}} \frac{((2.76 - 3.24)^2)}{2(0.31)^2} \\
 &= 0.832
 \end{aligned}$$

2. Menghitung probabilitas IPK SM1 Tidak Tepat waktu

Diketahui

Standar deviasi $\sigma = 1.75$

Mean $\pi = 0.23$

$$\begin{aligned}
 f(\text{ips1} = 3.00 \mid \text{tepat waktu}) &= \frac{1 * 2.7183}{\sqrt{2 * 0.23(1.75)}} \frac{((2.67 - 0.23)^2)}{2(0.31)^2} \\
 &= 0.092
 \end{aligned}$$

Dengan cara penghitungan yang sama untuk menghitung IPK SM2, IPK SM3 dan IPK SM4 .

c. Membandingkan hasil klas tepat waktu dan klas 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** dan 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 penghitungan Likelihood & Hob.

d. 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, dan **recall** merupakan kasus positif yang sebenarnya yang diprediksi positif secara benar [4].

Tabel 5. Pengujian dengan *Naive Bayes* dengan 11 atribut

<i>Classification</i>	Positif	Negatif
Positif	22	5
Negatif	1	2

Akurasi : $(22+2) / (22+2+1+5) \times 100\% = 80,00 \%$

Presisi : $22 / (22 + 5) \times 100\% = 81,48 \%$

Recall : $22 / (22 + 1) \times 100\% = 95,65 \%$

3. Metode *Feature Selection Information Gain* dan *Naive Bayes*

Implementasi *Feature Selection Information Gain* dan *Naive Bayes* hanya menggunakan 8 parameter yang paling berpengaruh yaitu jenis kelamin, status perkawinan, IPK SM1, IPK SM2, IPK SM 3, IPK SM 4, alamat dan sumber dana. Selanjutnya mencari prediksi kelulusan menggunakan Likelihood dan Hob Tepat Waktu dan Tidak Tepat Waktu, seperti berikut :

Likelihood Tepat Waktu = $(0,30 \times 0,89 \times 0,43 \times 0,25 \times 0,25 \times 0,14 \times 0,50 \times 0,08) = 0,000040184$

Likelihood Tidak Tepat Waktu = $(0,03 \times 0,06 \times 0,04 \times 0,04 \times 0,04 \times 0,10 \times 0,53 \times 0,01) = 0,000000000$

Hob Tepat Waktu

= $0,000040184 / (0,000040184 + 0,000000000) = 0,999998481$

Hob Tidak Tepat Waktu = $0,000040184 / (0,000040184 + 0,000000000) = 0,000001519$

Pengujian *Confusion Matriks*

Hasil klasifikasi akan diuji dalam tabel *Confusion Matriks*, sebagai berikut :

Akurasi = $(TN+TP) / (TN+FN+TP+FP) \times 100 \%$

Presisi = $TP / (TP + TN) \times 100 \%$

Recall = $TP / (TP + FP) \times 100 \%$

Tabel 6. . Pengujian dengan *Naïve Bayes* dengan 8 atribut

<i>Correct Classification</i>	<i>CLASSIFICATION</i>	
	Positif	Negatif
Positif	28	1
Negatif	1	0

$$\text{Akurasi} = (0+28)/(28+1+1+0) \times 100\% = 93,33\%$$

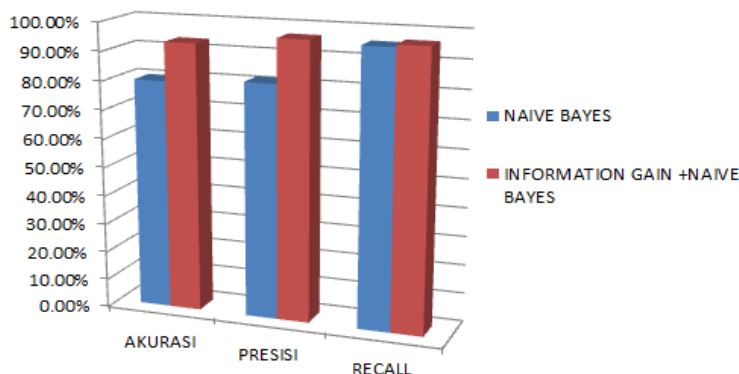
$$\text{Presisi} = (28) / (28 + 1) \times 100\% = 96,55\%$$

$$\text{Recall} = (28) / (26 + 1) \times 100\% = 96,55\%$$

Tabel 7. Hasil Komparasi dua Algoritma

No	Pengujian	<i>Naive Bayes</i>	<i>Information Gain dan Naive Bayes</i>
1	Akurasi	80,00%	93,33%
2	Presisi	81,48%	96,55%
3	Recall	96,55%	96,55%

**Hasil Komparasi Metode Naive Bayes dan
Feature Selection Information Gain + Naive Bayes**



Gambar 3 : Grafik hasil komparasi

Hasil komparasi menunjukkan bahwa menggunakan metode *Feature Selection Information Gain* dan *Naive Bayes* menghasilkan akurasi lebih baik yaitu 93,33%.

Kesimpulan

Pada kasus di atas komparasi metode *Feature Selection Information Gain* dan *Naive Bayes* menunjukkan hasil yang lebih baik dibandingkan hanya menggunakan Metode *Naive Bayes* saja. Terpilih 8 atribut yang digunakan pada implementasi menggunakan *naive bayes* menunjukkan bahwa atribut sumber dana mempunyai bobot yang paling tinggi yaitu 1,00 dan atribut pekerjaan orang tua memperoleh bobot yang paling kecil yaitu 0,00.

Saran

Secara keseluruhan, model yang diusulkan sudah menunjukkan hasil yang lebih baik dibanding dengan metode sebelumnya yang hanya menggunakan metode *naive bayes* saja. Namun masih

bisa dilanjutkan untuk penelitian berikutnya dan menggunakan metode klasifikasi yang lain sehingga mendapatkan hasil yang lebih baik.

Daftar Pustaka

- [1] R. H. Syah, “Dampak Covid-19 pada Pendidikan di Indonesia: Sekolah, Keterampilan, dan Proses Pembelajaran,” *SALAM J. Sos. dan Budaya Syar-i*, vol. 7, no. 5, 2020, doi: 10.15408/sjsbs.v7i5.15314.
- [2] Y. Apridiansyah, N. D. M. Veronika, and E. D. Putra, “Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naïve Bayes,” *JSAI (Journal Sci. Appl. Informatics)*, vol. 4, no. 2, pp. 236–247, 2021, doi: 10.36085/jsai.v4i2.1701.
- [3] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, “Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review,” *Fakt. Exacta*, vol. 13, no. 1, p. 35, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [4] E. Siswanto, “Optimasi Metode Naïve Bayes dalam Memprediksi Tingkat Kelulusan Mahasiswa Stekom Semarang,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 6, no. 1, pp. 1–6, 2019, [Online]. Available: <https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom/article/view/1038>
- [5] J. Han, M. Kamber, and J. Pei, “Introduction,” in *Data Mining*, Elsevier, 2012, pp. 1–38. doi: 10.1016/b978-0-12-381479-1.00001-0.
- [6] E. Prasetyo, *Data Mining : Konsep Dan Aplikasi Menggunakan Matlab*, I. Yogyakarta: Andi Yogyakarta, 2013.
- [7] L. Y. Lumban Gaol, M. Safii, and D. Suhendro, “Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5,” *Brahmana J. Penerapan Kecerdasan Buatan*, vol. 2, no. 2, pp. 97–106, 2021, doi: 10.30645/brahmana.v2i2.71.
- [8] N. Purwati, R. Nurlistiani, and O. Devinsen, “Data Mining Dengan Algoritma Neural Network Dan Visualisasi Data Untuk Prediksi Kelulusan Mahasiswa,” *J. Inform.*, vol. 20, no. 2, pp. 156–163, 2020, doi: 10.30873/ji.v20i2.2273.
- [9] M. N. Maskuri, K. Sukerti, and R. M. Herdian Bhakti, “Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke Stroke Desease Predict Using KNN Algorithm,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 1, pp. 130–140, 2022.
- [10] N. Komang Ratih Kumala and A. Savitri Puspaningrum, “E-Delivery Makanan Berbasis Mobile (Studi Kasus : Okonomix Kedaton Bandar Lampung),” *J. Teknol. dan Sist. Inf.*, vol. 1, no. 2, pp. 105–110, 2020, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [11] S. Zhang, C. Zhang, and Q. Yang, “Data preparation for data mining,” *Appl. Artif. Intell.*, vol. 17, p. 2003, 2010, doi: 10.1080/08839510390219264.
- [12] A. A. Murtopo, “Prediksi Kelulusan Tepat Waktu Mahasiswa STMIK YMI Tegal Menggunakan Algoritma Naïve Bayes Time Graduation Prediction by Using Naïve Bayes Algorithm at STMIK YMI Tegal,” pp. 145–154.
- [13] L. Dini, U. Sekolah, T. M. Informatika, D. Komputer, N. Mandiri, and R. S. Wahono, “Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naïve Bayes,” *J. Intell. Syst.*, vol. 1, no. 2, pp. 120–126, 2015.

- [14] A. Saifudin, “Metode Data Mining Untuk Seleksi Calon Mahasiswa,” vol. 10, no. 1, pp. 25–36, 2018.