

PREDIKSI PENGUNDURAN DIRI MAHASISWA UNIVERSITAS AMIKOM YOGYAKARTA MENGGUNAKAN METODE NAIVE BAYES

Andika Mahanggara

Fakultas Ilmu Komputer, Program Studi Informatika
Universitas Amikom Yogyakarta
Email: andika.m@students.amikom.ac.id

Arif Dwi Laksito

Fakultas Ilmu Komputer, Program Studi Informatika
Universitas Amikom Yogyakarta
Email: arif.laksito@amikom.ac.id

ABSTRAK

Universitas AMIKOM Yogyakarta merupakan Universitas yang telah berdiri sejak tahun 1994 dan telah banyak menghasilkan mahasiswa yang berbakat dalam bidang komputer. Setiap tahunnya Universitas AMIKOM Yogyakarta memiliki mahasiswa baru yang jumlahnya terus bertambah hingga saat ini, namun sering kali sejumlah peserta didik pada Universitas AMIKOM Yogyakarta mengundurkan diri. Dengan menggunakan teknik klasifikasi probabilistik sederhana yaitu Algoritma *Naive Bayes* dapat dilakukan prediksi terhadap pengunduran diri mahasiswa dengan menghitung sekumpulan probabilitas dari jumlah frekuensi dan kombinasi nilai dari dataset yang didapatkan. Implementasi *Naive Bayes* diharapkan mampu untuk memprediksi pengunduran diri agar pihak lembaga dapat melakukan pencegahan terhadap pengunduran diri mahasiswa. Hasil uji coba dari 120 dataset yang dibagi menjadi 70% data training dan 30% data testing diperoleh nilai error sebesar 22,22%. Sedangkan tingkat akurasi yang diperoleh sebesar 77,78% dengan hasil prediksi 22 mahasiswa diprediksi bertahan dan 14 mahasiswa diprediksi mengundurkan diri.

Kata kunci: *naive bayes*; prediksi; pengunduran diri.

ABSTRACT

Universitas Amikom Yogyakarta has been established in 1994. It has produced many talented students in the computer field. The number of students at this university is increasing every year. But, there were also several students who resigned or did not continue their studies. Using probabilistic classification technique, named Naive Bayes Algorithm, it can be predicted on student resignation by calculating a set of probabilities from the number of frequencies and combinations of values from the dataset. The Naive Bayes implementation is expected to be able to predict the resignation so the institution can prevent the resignation of students. The trial results of 120 datasets, which is divided into 70% training data and 30% testing data, resulted an error value of 22.22%. The accuracy level was 77.78%, it contained 22 students predicted to survive and 14 students predicted to resign.

Keywords: *naive bayes*; prediction; resignation.

1. PENDAHULUAN

Universitas AMIKOM Yogyakarta merupakan Universitas yang telah berdiri sejak tahun 1994 dan telah banyak menghasilkan mahasiswa yang berbakat dalam bidang komputer. Setiap tahunnya Universitas AMIKOM Yogyakarta memiliki mahasiswa baru yang jumlahnya terus bertambah hingga saat ini, namun sering kali sejumlah peserta didik pada Universitas AMIKOM Yogyakarta mengundurkan diri. Pihak lembaga pun tidak mampu mempertahankan mahasiswanya karena kurangnya informasi penyebab pengunduran diri mahasiswa. Apabila kemungkinan pengunduran diri mahasiswa dapat diketahui sejak dini, maka pihak lembaga dapat melakukan tindakan-tindakan yang diperlukan untuk mempertahankan mahasiswa tersebut. Keterbatasan kemampuan pihak lembaga untuk mempertahankan dan memprediksikan pengunduran diri mahasiswa dalam jumlah besar tidak mungkin dilakukan secara manual karena membutuhkan waktu yang cukup lama. Serta belum adanya suatu sistem yang membantu dalam memprediksi pengunduran diri mahasiswa, merupakan beberapa faktor penyebab lembaga ini kehilangan mahasiswanya. Oleh karena itu diperlukan suatu metode yang dapat memprediksi pengunduran diri.

Sebelumnya telah dilakukan penelitian oleh Mila Listiana dari Universitas Muhammadiyah Surakarta (2015), dengan judul penelitian “Perbandingan Algoritma *Decision Tree* (C4.5) dan *Naive Bayes* Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita (Studi Kasus Puskesmas Kartasura)”. Hasil dari penelitian ini didapatkan *Naive Bayes* dengan nilai total 258.03% dengan tingkat akurasi 76.97%, nilai *recall* 96.89%, dan tingkat *precision* 84.17% sedangkan *Decision Tree* (C4.5) memiliki nilai total 250.67% dengan tingkat akurasi 75.66%, nilai *recall* 89.78%, dan tingkat *precision* 85.23% [1]. Dengan hasil penelitian tersebut didapatkan bahwa *Naive Bayes* lebih baik dibandingkan *Decision Tree*. Sedangkan penelitian yang lain oleh Hera Wasati dan Dwi Wijayanti (2014) dari STMIK AKAKOM Yogyakarta, dengan judul penelitian “Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode *Naive Bayes* (Studi Kasus: P.T. Karyatama Mitra Sejati Yogyakarta)”, menggunakan beberapa parameter berupa usia, pendidikan, tinggi badan, berat badan, nilai tes. Dari hasil pengujian menggunakan 180 data testing hasil akurasi yang diperoleh sebesar 73.89% [2]. Pada penelitian Arief Jananto dari Universitas Stikubank (2013), dengan judul “Algoritma *Naive Bayes* untuk Mencari Perkiraan Waktu Studi Mahasiswa” menggunakan atribut IPK, jenis kelamin, kota lahir, tipe sekolah, kota sekolah, dan lama studi. Sebelum diolah peneliti melakukan konversi data pada setiap atribut yang memiliki banyak nilai. Hasil dari penelitian ini berdasarkan 60 data testing didapat tingkat error 20% hingga 34% [3].

Dengan hasil penelitian tersebut peneliti memilih menggunakan metode *Naive Bayes*. *Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasarkan pada penerapan Teorema *Bayes* dengan asumsi independensi (ketidaktergantungan) yang kuat (naif) [4]. Metode *Naive Bayes* pada penelitian ini digunakan untuk memprediksi pengunduran diri mahasiswa khususnya jurusan informatika. Berdasarkan penelitian yang sudah dilakukan dan salah satu kelebihan *Naive Bayes* yaitu memerlukan sejumlah kecil data training maka data yang akan digunakan adalah 120 data acak mahasiswa sebagai data training kemudian untuk perhitungan akurasi dengan confusion matrix menggunakan 30% dari total data training tersebut, dengan atribut yang meliputi jenis kelamin, IPK, dan masa studi. Data mahasiswa yang masih aktif akan digunakan sebagai data testing. Mahasiswa akan diklasifikasikan dalam kategori bertahan atau mengundurkan diri.

1.1 Landasan Teori

1.1.1 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas yang tidak diketahui dari suatu objek. Terdapat dua proses dalam klasifikasi data yaitu:

- Proses pelatihan (*training*), pada proses ini akan digunakan *training set* dimana sudah diketahui label-labelnya untuk membangun model atau fungsi.
- Proses pengujian (*testing*), untuk mengetahui akurasi model atau fungsi yang akan dibangun pada proses *training*, oleh karena itu perlu digunakan data *testing set* untuk melakukan memprediksi pada label-labelnya.

Klasifikasi dokumen adalah pemberian kategori yang telah didefinisikan kepada dokumen yang belum memiliki kategori. Mengklasifikasikan dokumen merupakan salah satu cara untuk membuat dokumen terorganisir. Dokumen-dokumen yang memiliki isi yang sama akan dikelompokkan ke dalam kategori yang sama [5].

1.1.2 *Naive Bayes*

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan dalam suatu class. Teorema *Bayes* menjadi dasar dari *Bayesian classification* dimana memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [6]. Dalam hal ini, diasumsikan bahwa kehadiran atau ketiadaan dari suatu kejadian tertentu dari suatu kelompok tidak berhubungan dengan kehadiran atau ketiadaan dari kejadian lainnya. *Naive Bayes* dapat digunakan untuk berbagai macam keperluan antara lain untuk klasifikasi dokumen, deteksi atau filtering spam, dan masalah klasifikasi lainnya. Rumus bayes dinyatakan pada rumus nomor 1 berikut:

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

X = Data dengan class yang belum diketahui

- H = Hipotesis data X merupakan suatu class spesifik
 $P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (posteriori prob)
 $P(H)$ = Probabilitas hipotesis H (prior prob)
 $P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut
 $P(X)$ = Probabilitas dari X

1.1.3 Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining atau Sistem Pendukung Keputusan. Di penelitian Siti Masripah [7] dan Aida Indriani [8] menggunakan confusion matrix untuk melakukan pengukuran akurasi klasifikasi. Pada pengukuran kinerja menggunakan *confusion matrix* terdapat 4 istilah sebagai representasi dari hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* [8].

Tabel 1. Confusion matrix [8]

		<i>Kelas prediksi</i>	
		1	0
<i>Kelas sebenarnya</i>	1	TP	FN
	0	FP	TN

Keterangan confusion matrix untuk tabel 1 diatas adalah sebagai berikut:

- TP (*True Positive*) merupakan banyaknya data yang kelas aktualnya yaitu kelas positif dengan kelas prediksinya merupakan kelas positif.
- FN (*False Negative*) merupakan banyaknya data yang kelas aktualnya yaitu kelas positif dengan kelas prediksinya merupakan kelas negatif.
- FP (*False Positive*) merupakan banyaknya data yang kelas aktualnya yaitu kelas negatif dengan kelas prediksinya merupakan kelas positif.
- TN (*True Negative*) merupakan banyaknya data yang kelas aktualnya yaitu kelas negatif dengan kelas prediksinya merupakan kelas negatif.

Akurasi merupakan suatu metode pengujian berdasar tingkat kedekatan antara nilai prediksi dengan nilai aktual. Dengan mengetahui jumlah data klasifikasi secara benar maka dapat diketahui akurasi hasil prediksi. Nilai akurasi dapat dicari dengan menggunakan persamaan nomor 2 berikut:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

Nilai presisi merupakan proporsi jumlah kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Nilai recall merupakan proporsi jumlah kasus positif yang sebenarnya yang diprediksi positif secara benar pula. Sedangkan nilai *accuracy* merupakan prosentase jumlah data yang diklasifikasi secara benar secara keseluruhan [9]. Nilai presisi bisa diperoleh dengan persamaan nomor 3 berikut:

$$\text{Presisi} = \frac{TP}{FP + TP} \times 100\% \quad (3)$$

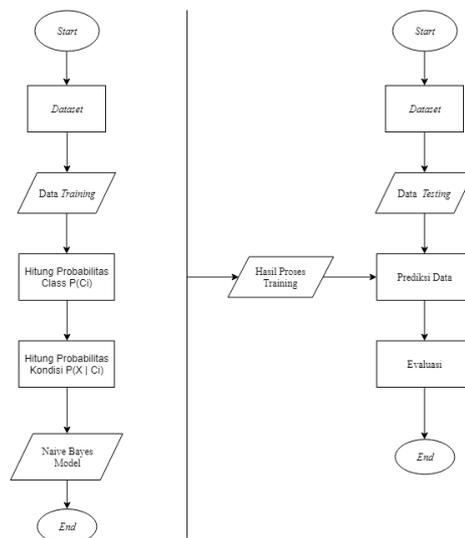
Dan nilai *recall* bisa diperoleh dengan menggunakan persamaan nomor 4 berikut:

$$\text{Recall} = \frac{TP}{FN + TP} \times 100\% \quad (4)$$

2. METODOLOGI PENELITIAN

Pada bagian ini dijelaskan tahap-tahap dalam penelitian ini dari awal hingga akhir. Adapun tahap-tahap tersebut adalah sebagai berikut ini:

- a) *Data Understanding*: yaitu melakukan pengumpulan data awal, deskripsi data, eksplorasi data, dan verifikasi kualitas data.
- b) *Data Preparation*: pada tahap ini dilakukan identifikasi dan konversi data, yaitu penentuan parameter-parameter yang akan digunakan dalam proses pembelajaran.
- c) *Data Conversion*: untuk mempermudah dalam proses perhitungan maka data yang ada dilakukan konversi kedalam bentuk yang dibutuhkan dan dapat diimplementasikan kedalam bentuk program.
- d) Membangun Model *Naive Bayes*. Proses dalam membangun model Algoritma *Naive Bayes* dilakukan seperti pada gambar 1 berikut ini:



Gambar 1. Flowchart Membangun Algoritma *Naive Bayes*

- e) Pengujian, dilakukan untuk mengetahui bagaimana kinerja algoritma *Naive Bayes* dalam melakukan prediksi pengunduran diri mahasiswa Universitas AMIKOM Yogyakarta. Disini pengujian menggunakan *Confusion Matrix*.

3. HASIL DAN PEMBAHASAN

3.1 *Data Understanding dan Data Preparation*

Penelitian yang dilakukan ini menggunakan data sekunder yang sudah ada dari pihak Universitas AMIKOM Yogyakarta. Adapun parameter-parameter yang akan digunakan dalam proses pembelajaran adalah sebagai berikut:

- a) Jumlah data yang digunakan sebanyak 120 data acak mahasiswa Universitas AMIKOM Yogyakarta, kemudian 70% dari jumlah data akan digunakan sebagai data training dan 30% akan digunakan sebagai data testing.
- b) Atribut yang digunakan berjumlah 3 yaitu jenis kelamin, masa studi, dan IPK.

3.2 *Data Conversion*

Tahap selanjutnya yaitu dilakukan konversi kedalam bentuk yang dibutuhkan dan dapat diimplementasikan kedalam bentuk program. Adapun proses konversi data yang telah diperoleh dari tahap persiapan data adalah sebagai berikut:

- a) Jenis Kelamin, untuk jenis kelamin karena hanya memiliki 2 nilai maka tidak dilakukan konversi.
- b) Masa Studi, untuk mendapatkan nilai masa studi dilakukan konversi dari atribut NIM yang ada didata mahasiswa dengan rumus tahun sekarang – 2 digit diawal NIM seperti pada tabel 2 berikut ini:

Tabel 2. Nilai Masa studi

<i>Nim</i>	<i>Proses</i>	<i>Masa studi</i>
14.11.8464	18 - 14	4
16.11.0532	18 - 16	2

- c) IPK, untuk nilai IPK tidak akan mengubah nilai sebenarnya dari data namun untuk tetap mempermudah perhitungan maka akan dibuat range seperti pada tabel 3:

Tabel 3. Range IPK

<i>Range IPK</i>
1 <= IPK < 2
2 <= IPK < 3
3 <= IPK <= 4

- d) Keterangan atau Status, untuk keterangan memiliki 3 nilai yaitu SK, NA, dan MD maka dilakukan konversi agar memiliki 2 nilai saja, proses konversi dilakukan dengan memberi nilai keterangan pada tabel 4:

Tabel 4. Nilai Keterangan

<i>Status asli</i>	<i>Status baru</i>
SK	Bertahan
MD	Mengundurkan diri
NA	Mengundurkan diri

Keterangan:

SK = sudah krs

MD = mengundurkan diri

NA = tanpa status

Karena kurangnya data mahasiswa yang mengundurkan diri maka mahasiswa dengan status NA (tidak memiliki status apapun) akan diasumsikan sebagai mahasiswa yang mengundurkan diri. Hasil akhir dari konversi data akan diperoleh susunan data seperti pada tabel 4 dengan jumlah 120 *record* data yang sudah siap untuk dilakukan proses perhitungan. Hasil akhir dari konversi data akan diperoleh susunan data seperti pada tabel 5 dengan jumlah 120 *record* data yang sudah siap untuk dilakukan proses perhitungan.

Tabel 5. Sample data untuk proses

<i>Jenis kelamin</i>	<i>Masa studi</i>	<i>IPK</i>	<i>Status</i>
L	4	3,37	Bertahan
L	4	2,88	Bertahan
L	4	2,28	Mengundurkan diri
L	4	1,71	Mengundurkan diri
L	2	3,43	Bertahan
L	2	3,20	Bertahan
L	2	3,28	Mengundurkan diri
L	2	2,14	Mengundurkan diri

3.3 Membangun Model Naive Bayes

Tahap membangun Model *Naive Bayes* seperti pada gambar 2 diatas yaitu sebagai berikut:

- Start, memulai proses untuk mendapatkan model dan prediksi *Naive Bayes*.
- Dataset, mengumpulkan seluruh dataset sebanyak 120 data yang telah melalui proses konversi.
- Data Training, mengambil 70% data dari total dataset untuk dijadikan data training.
- Hitung Probabilitas Class $P(C_i)$, pada penelitian ini kelas dibagi 2 yaitu kelas Bertahan dan Mengundurkan Diri(MD). Cara mendapatkan nilai probabilitas masing-masing kelas tersebut sebagai berikut:

$$P(\text{Class} = \text{Bertahan}) = \frac{45}{84} = 0,535714$$

$$P(\text{Class} = \text{MD}) = \frac{39}{84} = 0,464286$$

- e) Hitung Probabilitas Kondisi $P(X | Ci)$, pada langkah ini menghitung jumlah kondisi yang sama pada setiap nilai dari 2 kelas yang sudah didapatkan yaitu Bertahan dan Mengundurkan Diri (MD). Cara mencari nilai probabilitas masing-masing kondisi terhadap nilai kelas sebagai berikut:

$$P\left(\begin{array}{l} \text{Jenis Kelamin} = L \\ | \text{Class} = \text{Bertahan} \end{array}\right) = \frac{39}{45} = 0,866667$$

$$P\left(\begin{array}{l} \text{Jenis Kelamin} = P \\ | \text{Class} = \text{Bertahan} \end{array}\right) = \frac{6}{45} = 0,133333$$

$$P\left(\begin{array}{l} \text{Jenis Kelamin} = L \\ | \text{Class} = \text{MD} \end{array}\right) = \frac{36}{39} = 0,923077$$

$$P\left(\begin{array}{l} \text{Jenis Kelamin} = P \\ | \text{Class} = \text{MD} \end{array}\right) = \frac{3}{39} = 0,076923$$

$$P\left(\begin{array}{l} \text{Masa Studi} = 2 \\ | \text{Class} = \text{Bertahan} \end{array}\right) = \frac{25}{45} = 0,555556$$

$$P\left(\begin{array}{l} \text{Masa Studi} = 4 \\ | \text{Class} = \text{Bertahan} \end{array}\right) = \frac{20}{45} = 0,444444$$

$$P\left(\begin{array}{l} \text{Masa Studi} = 2 \\ | \text{Class} = \text{MD} \end{array}\right) = \frac{21}{39} = 0,538462$$

$$P\left(\begin{array}{l} \text{Masa Studi} = 4 \\ | \text{Class} = \text{MD} \end{array}\right) = \frac{18}{39} = 0,461538$$

$$P\left(\begin{array}{l} 1 \leq \text{IPK} < 2 \\ | \text{Class} = \text{Bertahan} \end{array}\right) = \frac{2}{45} = 0,044444$$

$$P\left(\begin{array}{l} 2 \leq \text{IPK} < 3 \\ | \text{Class} = \text{Bertahan} \end{array}\right) = \frac{14}{45} = 0,311111$$

$$P\left(\begin{array}{l} 3 \leq \text{IPK} \leq 4 \\ | \text{Class} = \text{Bertahan} \end{array}\right) = \frac{29}{45} = 0,644444$$

$$P\left(\begin{array}{l} 1 \leq \text{IPK} < 2 \\ | \text{Class} = \text{MD} \end{array}\right) = \frac{9}{39} = 0,230769$$

$$P\left(\begin{array}{l} 2 \leq \text{IPK} < 3 \\ | \text{Class} = \text{MD} \end{array}\right) = \frac{17}{39} = 0,435897$$

$$P\left(\begin{array}{l} 3 \leq \text{IPK} \leq 4 \\ | \text{Class} = \text{MD} \end{array}\right) = \frac{13}{39} = 0,333333$$

- f) Naive Bayes Model, hasil klasifikasi dan perhitungan probabilitas dari proses *Naive Bayes*.
g) Data Testing, data testing diambil dari sisa data yang tidak termasuk kedalam data training dengan jumlah data 30% dari *dataset*.
h) Prediksi Data, hasil prediksi dari proses *Naive Bayes*. Ambil data dari dataset untuk memprediksikan termasuk kelas mana data tersebut. Contoh: Jenis Kelamin = L, Masa Studi = 4, IPK = 3,26086956521739
- 1) Bertahan = $P(\text{Jenis Kelamin} = L | \text{bertahan}) * P(\text{Masa studi} = 4 | \text{bertahan}) * P(3 \leq \text{IPK} \leq 4 | \text{bertahan}) * P(\text{bertahan})$
 $= 0,866667 * 0,444444 * 0,644444 * 0,535714$
 $= 0,13298$
 - 2) Mengundurkan Diri = $P(\text{JK} = L | \text{MD}) * P(\text{Masa studi} = 4 | \text{MD}) * P(3 \leq \text{IPK} \leq 4 | \text{MD}) * P(\text{MD})$
 $= 0,923077 * 0,461538 * 0,333333 * 0,464286$
 $= 0,065934$

- 3) Setelah itu kita hitung seluruh data tiap kelasnya, dengan menghitung jumlah data per nilai sebagai berikut:
 $P(X) = P(\text{Jenis Kelamin} = L = 100 \text{ data}) * P(\text{Masa Studi} = 4 = 38 \text{ data}) * P(3 \leq \text{IPK} \leq 4 = 42 \text{ data})$

$$\frac{100}{120} \times \frac{38}{120} \times \frac{42}{120} =$$

$$= 0,092361$$
- 4) Kemudian dibagi hasil tiap kelas
Bertahan = $\frac{0,13298}{0,092361} = 1,439785$
Mengundurkan Diri = $\frac{0,065934}{0,092361} = 0,713873$
- 5) Setelah mendapatkan hasil baginya kita bandingkan kedua hasil tersebut, probabilitas mana yang lebih besar. Hasil perhitungan prediksi *Naive Bayes* yaitu Bertahan dengan nilai $1,439785 > 0,713873$
- i) Pengujian dilakukan dengan menggunakan 30% total data sebagai data testing. Pada tabel 6 dibawah ini merupakan hasil uji dari penelitian.

Tabel 6. Hasil uji

<i>Jenis kelamin</i>	<i>Masa studi</i>	<i>IPK</i>	<i>Status</i>	<i>Prediksi</i>
L	2	3,5	MD	Bertahan
L	2	2,5	MD	MD
L	2	1,81	MD	MD
L	2	3,47	Bertahan	Bertahan
L	4	3,60	Bertahan	Bertahan
L	2	3,79	Bertahan	Bertahan
L	4	3	Bertahan	Bertahan
L	4	3,38	MD	Bertahan
L	2	2,3	MD	MD
L	4	2,53	MD	MD
L	4	3,21	Bertahan	Bertahan
L	4	2,68	Bertahan	MD
L	4	1,93	Bertahan	MD
L	2	3,52	MD	Bertahan
L	4	3,51	Bertahan	Bertahan
L	2	2,59	MD	MD
L	2	2,05	MD	MD
P	2	3,37	Bertahan	Bertahan
L	2	2,53	MD	MD
P	2	3,60	Bertahan	Bertahan
L	4	2,78	MD	MD
L	2	3,75	MD	Bertahan
L	2	3,43	Bertahan	Bertahan
L	2	1,71	MD	MD
L	2	3,67	Bertahan	Bertahan
L	4	2,51	MD	MD
P	4	3,26	Bertahan	Bertahan
L	4	3,26	Bertahan	Bertahan
P	4	3,31	Bertahan	Bertahan
L	2	3,39	Bertahan	Bertahan
P	2	2,93	Bertahan	Bertahan
L	2	2,11	MD	MD
P	2	3,18	Bertahan	Bertahan
L	4	3,07	MD	Bertahan
L	4	2,92	Bertahan	MD
L	2	3,68	Bertahan	Bertahan

Hasil dari tabel 5 diatas didapat data positif dengan prediksi positif(TP) sebanyak 17, data negatif dengan prediksi negatif (TN) sebanyak 11, data negatif dengan prediksi positif(FP) sebanyak 5, dan data

positif dengan prediksi negatif (FP) sebanyak 3. Setelah mendapatkan nilai-nilai tersebut maka untuk mendapatkan nilai akurasi menggunakan confusion matrix bisa menggunakan rumus dibawah ini:

$$\begin{aligned} \text{Akurasi} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \\ &= \frac{17 + 11}{17 + 11 + 5 + 3} \times 100\% \\ &= \frac{28}{36} \times 100\% = 77,7777777\% \end{aligned}$$

4. KESIMPULAN

Dari hasil perhitungan confusion matrix dapat disimpulkan bahwa tingkat akurasi Algoritma *Naive Bayes* dalam prediksi pengunduran diri mahasiswa sebesar 77,78% dengan tingkat error sebesar 22,22%. Kurangnya data aktual mahasiswa yang mengundurkan diri memengaruhi kinerja *Naive Bayes*. Sehingga perlu dikembangkan lagi penelitian menggunakan atribut yang berbeda atau menambah atribut dan menggunakan dataset yang lebih banyak terutama data aktual pada mahasiswa yang mengundurkan diri.

DAFTAR PUSTAKA

- [1] M. Listiana, "Perbandingan Algoritma Decision Tree (C4.5) Dan NaïVe Bayes Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita (Studi Kasus Puskesmas Kartasura)," Universitas Muhammadiyah Surakarta, 2015.
- [2] H. Wasiati and D. Wijayanti, "Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes (Studi Kasus: Di P.T. Karyatama Mitra Sejati Yogyakarta)," *IJNS - Indones. J. Netw. Secur.*, vol. 3, no. 2, pp. 45–51, 2014.
- [3] A. Jananto, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa," *Teknol. Inf.*, vol. 18, no. 1, pp. 9–16, 2013.
- [4] E. Prasetyo, *Data Mining: Konsep Dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset, 2012.
- [5] C. Goller, J. Löning, T. Will, and W. Wolff, "Automatic Document Classification: A thorough Evaluation of various Methods," *Int. Symp. für Informationswiss.*, vol. 8, no. 10, pp. 145–162, 2000.
- [6] Kusriani and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [7] S. Masripah, "Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit," *Bina Insa. ICT J.*, vol. 3, no. 1, pp. 187–193, 2016.
- [8] A. Indriani, "Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier," *Aust. Prescr.*, vol. 20, no. SUPPL. 3, pp. 28–33, 1997.
- [9] A. Andriani, "Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus : Amik ' Bsi Yogyakarta ," " *Semin. Nas. Teknol. Inf. dan Komun. 2013 (SENTIKA 2013)*, vol. 2013, no. Sentika, pp. 163–168, 2013.