

KLASIFIKASI PERKIRAAN KELULUSAN MAHASISWA JENJANG MAGISTER MENGGUNAKAN METODE *NAÏVE BAYES*

Galih Hery Herlambang¹⁾, Aryo Nugroho²⁾, Badrus Zaman³⁾

^{1,2)}Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Narotama, Surabaya.

³⁾Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Airlangga, Surabaya.

e-mail: galih18@fasilkom.narotama.ac.id¹⁾, aryo.nugroho@narotama.ac.id²⁾, badruszaman@fst.unair.ac.id³⁾

ABSTRAK

Pendidikan tinggi adalah jenjang pendidikan pada jalur pendidikan formal setelah pendidikan menengah yang dapat berupa program pendidikan diploma, sarjana, magister, spesialis dan doktor, yang diselenggarakan oleh perguruan tinggi. Dilihat dari banyaknya peminat dari tiap tahun akademik perguruan tinggi dapat memperhatikan berbagai faktor yang berpengaruh terhadap pengelolaan daya tampung mahasiswa, salah satunya ketepatan dari masa studi. Tujuan dari penelitian ini untuk melakukan klasifikasi perkiraan ketepatan lulus dari calon mahasiswa menggunakan metode *naive bayes*. Data yang digunakan adalah data pendaftar calon mahasiswa jenjang magister pada sebuah perguruan tinggi dengan delapan atribut yaitu usia, jenis kelamin, nilai toefl, IPK S1, sumber biaya, dan pekerjaan pendaftar. Hasil dari penelitian menunjukkan bahwa nilai akurasi terbaik sebesar 80%.

Kata Kunci: Data mining, Klasifikasi, Kelulusan Mahasiswa, *Naive Bayes*.

ABSTRACT

Higher education is a level of education in the formal education pathway after secondary education that can form diploma, bachelor, master, specialist and doctoral education programs, which are organized by universities. Judging from the number of enthusiasts from each academic year of higher education can pay attention to various factors related to the management of student capacity, one of which is the accuracy of the study period. The purpose of this study is to predict the accuracy of graduation from prospective students using the Naive Bayes method. The data used are the data of prospective master level students in tertiary institutions with these attributes, age, gender, toefl value, S1 GPA, source of cost and applicant's work. The results of the study showed that the best accuracy value was 80%.

Keywords: Data mining, Classification, Naive Bayes, Student Graduation

I. PENDAHULUAN

Pendidikan tinggi adalah jenjang pendidikan pada jalur pendidikan formal setelah pendidikan menengah yang dapat berupa program pendidikan diploma, sarjana, magister, spesialis dan doktor, yang diselenggarakan oleh perguruan tinggi[1]. Jenjang pendidikan magister (S2) adalah jenjang pendidikan untuk menghasilkan lulusan yang memiliki kemampuan mengembangkan dan menerapkan ilmu pengetahuan dan teknologi, dan/atau seni dengan cara menguasai dan memahami, pendekatan, metode, kaidah ilmiah disertai keterampilan penerapannya. Mempunyai kemampuan memecahkan permasalahan di bidang keahlian kegiatan penelitian dan pengembangan kaidah ilmiah. Serta mempunyai kemampuan pengembangan-kinerja profesionalnya yang ditunjukkan dengan merumuskan pendekatan penyelesaian masalah dengan cara penalaran ilmiah atau profesi yang serupa[2]. Dilihat dari banyaknya peminat dari tiap tahun akademik perguruan tinggi dapat memperhatikan berbagai faktor

yang berpengaruh terhadap pengelolaan daya tampung mahasiswa, salah satunya ketepatan dari masa studi-mahasiswa sesuai dengan waktu yang telah ditentukan[3]. Masa studi adalah waktu studi terjadwal yang harus ditempuh oleh mahasiswa sesuai dengan rentang waktu yang dipersyaratkan[4].

Data pendaftar calon mahasiswa magister dimungkinkan untuk menghasilkan sebuah klasifikasi perkiraan kecenderungan lama studi berdasarkan variable atau atribut data yang ada pada data pendaftar. Klasifikasi yang dihasilkan dari tiap peserta dapat dijadikan sebagai acuan dalam melakukan seleksi calon mahasiswa berdasarkan ketepatan waktu lulus.

Untuk mendapatkan prediksi ketepatan waktu kuliah-mahasiswa dibutuhkan sebuah metode yang dapat-mengolah data yang telah tersedia[5]. Hasil dari pengelolaan pola dapat dianalisis dengan menggunakan-konsep data mining. Data mining, sering juga disebut *Knowledge Discovery In Database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data

historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar[6]. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Terdapat dua metode dalam pengolahan data. Metode tersebut adalah metode klasifikasi dan klasterasi[7]. Metode Klasifikasi mampu mengolah dataset dengan hasil akurasi tinggi yakni *Naïve Bayes*[8]. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan-konsep atau kelas data, untuk kepentingan tertentu. Dalam mencapai tujuan tersebut, proses klasifikasi-membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bias berupa aturan atau formula matematis[9]. Hasil yang didapatkan selanjutnya dapat diuji menggunakan *percentage split*. *Percentage split* merupakan teknik mengevaluasi dari hasil training data dimana cara ini akan menemukan akurasi terbaik dengan memanfaatkan angka dalam persentasenya[10].

Hasil klasifikasi sangat bergantung pada aturan keputusan. Aturan tersebut digunakan untuk menentukan hipotesis yang paling mungkin atau dikenal dengan *Maximum A Posteriori* (MAP). *Naive Bayes Classifier* merupakan salah satu metode dalam data mining yang menggabungkan teorema bayes dengan MAP. Algoritma *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya-sehingga algoritma ini relative lebih terukur serta memiliki akurasi tinggi dalam mengelola objek dalam-jumlah besar[10]. *Naive Bayes* merupakan metode klasifikasi yang berdasarkan pada Teorema Bayes. Dinamakan Teorema Bayes karena disesuaikan dengan nama penemunya, yaitu Reverend Thomas Bayes, walaupun sebenarnya ada beberapa penelitian yang mengatakan bahwa Teorema Bayes telah ditemukan oleh orang lain sebelum Reverend Thomas Bayes[10]. *Naive bayes* memiliki keunggulan dalam pengolah teks dibandingkan dengan metode lain[11]. Metode *Naive Bayes* tidak memerlukan proses generalisasi yang kompleks[12]. *Naive bayes* hanya perlu menghitung statistik fitur per kelas dengan satu kali melewati dokumen pelatihan sehingga menghemat waktu komputasi[13].

Naive bayes classifier (NBC) merupakan salah satu metode pada teknik klasifikasi dan termasuk dalam *classifier* statistik yang dapat memprediksi probabilitas keanggotaan *class*[14]. NBC berprinsip pada teori bayes. NBC mengasumsikan bahwa nilai atribut pada sebuah *class* adalah independen terhadap nilai pada atribut yang lain. Dalam *naive bayes* terdapat langkah-dimana cara untuk mendapatkan nilai dalam tabel *mean*, standar deviasi dan probabilitas. Hal tersebut bertujuan untuk mendapatkan kelayakan validitas dis-

kriminatif, yang dapat dilihat oleh skor AVE dari setiap atribut yang harus lebih dari varian atribut lainnya[15].

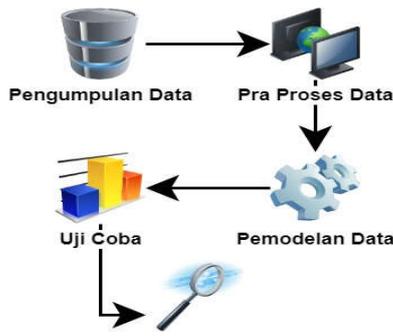
Penelitian yang pernah dilakukan oleh Agus Romadhona[5] tentang presiksi kelulusan mahasiswa berdasarkan usia, jenis kelamin dan indeks prestasi menggunakan algoritma *decision tree* dengan tujuan mengidentifikasi mahasiswa yang gagal atau lulus tepat waktu sehingga dapat diberikan peningkatan bimbingan pada mahasiswa tersebut. Hasil penelitian menyimpulkan bahwa indeks prestasi semester 4 mendapat nilai *information gain* tertinggi sehingga nilai tersebut berhak menjadi *root*. Abdul Rohman juga melakukan penelitian[9] tentang prediksi kelulusan mahasiswa menggunakan *K-Nearest Neighbor (KNN)* dengan tujuan mengetahui tingkat akurasi dari algoritma KNN. Hasil penelitian tersebut menunjukkan bahwa akurasi KNN dipengaruhi oleh jumlah *clustering* data. Fitry, Diana Laily melakukan penelitian[16] dengan tujuan mengetahui tingkat kelulusan mahasiswa dalam satu tahun ajaran menggunakan metode *naive bayes*. Hasil penelitian tersebut menunjukkan *naive bayes* mempunyai akurasi 93% dalam mengolah data dengan 8 atribut serta *naive bayes* mampu mengolah nilai probabilitas dari ketepatan lulus mahasiswa.

Penelitian ini bertujuan untuk mengklasifikasi ketepatan waktu lulus dari calon mahasiswa jenjang magister pada sebuah universitas. Kemudian, dari hasil klasifikasi ketepatan kelulusan mahasiswa dapat dijadikan sebagai bahan pertimbangan pada proses seleksi mahasiswa masuk perguruan tinggi jenjang magister pada suatu universitas.

II. METODE PENELITIAN

Penelitian ini adalah eksperimen untuk menentukan klasifikasi ketepatan waktu kelulusan mahasiswa berdasarkan atribut usia, jenis kelamin, nilai IPK S1, nilai toefl, sumber biaya, dan pekerjaan pendaftar. Tahap penelitian menggunakan teknik data mining dengan metode *naive bayes*. Sedangkan, untuk pengujian nilai akurasi dari metode menggunakan presentase *split* untuk mencari nilai optimal dari metode yang digunakan. Secara garis besar terdapat 5 tahap yang dilakukan untuk melaksanakan penelitian ini.

Tahap penelitian ini dapat digambarkan sebagai berikut:



Gambar 1. Tahap Penelitian

A. Pengumpulan Data

Pengumpulan data adalah prosedur yang sistematis dan standar untuk memperoleh data yang diperlukan. Pengumpulan data dilakukan dengan cara observasi-dokumen dan laporan yang ada pada sebuah perguruan tinggi negeri di kota Surabaya. Data yang dikumpulkan adalah data penerimaan mahasiswa baru jalur penerimaan jenjang magister dari tahun 2016 sampai dengan 2019. Pengumpulan data dilakukan pada awal-bulan September sampai dengan pertengahan bulan September.

B. Pra Proses Data

Tahap pra proses data adalah tahap mentransformasikan data mentah kedalam format data yang sesuai dengan kebutuhan analisis dalam penelitian. Proses ini bertujuan menyiapkan dan mengolah data dengan teknik-teknik tertentu sehingga terbentuk data yang siap diolah pada proses selanjutnya. Teknik dan metode yang dilakukan dalam proses ini diantaranya adalah:

1) Data Cleaning

Data Cleaning adalah menghilangkan nilai-nilai data yang salah, memperbaiki kecacauan data dan memeriksa konsistensi data.

2) Data Transformasi

Data Transformasi normalisasi dan pengumpulan data sehingga data yang diperoleh mempunyai nilai yang sesuai dengan kebutuhan analisa data.

3) Data Reduksi

Data Reduksi adalah menguraikan data kedalam bentuk yang lebih kecil tetapi tetap menghasilkan analisa yang sama. Dalam tahap ini juga dilakukan sebuah proses diskretisasi data yang bertujuan untuk mereduksi sekumpulan nilai yang terdapat pada atribut *continuous*, dengan membagi *range* dan atribut kedalam interval yang telah dihitung sebelumnya.

C. Pemodelan Data

Tahap pemodelan data adalah pengolahan data dengan metode *naïve bayes* sesuai dengan tahapan *naïve bayes* yang diperlukan. Dalam tahap ini dilakukan pemrosesan

san data sebanyak 2251 data. Data tersebut adalah data dari hasil pra proses data. Data yang diperoleh selanjutnya diolah sesuai dengan langkah-langkah pada metode *naïve bayes* untuk menghasilkan sebuah model.

Naïve Bayes merupakan metode klasifikasi yang berdasarkan pada Teorema Bayes. Dinamakan Teorema Bayes karena disesuaikan dengan nama penemunya, yaitu Reverend Thomas Bayes, walaupun sebenarnya ada beberapa penelitian yang mengatakan bahwa Teorema Bayes telah ditemukan oleh orang lain sebelum Reverend Thomas Bayes.

Naïve bayes classifier (NBC) merupakan salah satu metode pada teknik klasifikasi dan termasuk dalam *classifier* statistik yang dapat memprediksi probabilitas keanggotaan *class*. NBC berprinsip pada teori bayes. NBC mengasumsikan bahwa nilai atribut pada sebuah *class* adalah independen terhadap nilai pada atribut yang lain seperti pada persamaan 1.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

Dimana x adalah data dengan *class* yang belum diketahui. C adalah hipotesis data merupakan suatu *class* spesifik. $P(c|x)$ adalah Probabilitas hipotesis berdasar kondisi (*posteriori probability*). $P(c)$ adalah Probabilitas hipotesis (*prior probability*). $P(x|c)$ adalah Probabilitas berdasarkan kondisi pada hipotesis. $P(x)$ adalah Probabilitas dari hipotesis data merupakan suatu *class* spesifik.

Dimana Variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana seperti pada persamaan 2.

$$Posterior = \frac{Prior \times likelihood}{evidence} \quad (2)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan menggunakan aturan perkalian seperti pada persamaan 3 berikut:

$$P(C|F_1 \dots F_n) = P(C)P(F_1 \dots F_n|C) \quad (3)$$

Persamaan 3 diatas dapat dijabarkan menjadi $P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1)$ sehingga nilai $P(C|F_1 \dots F_n)$ dapat menjadi persamaan seperti pada persamaan 4 berikut:

$$P(C)P(F_1|C)P(F_2|C, F_1) \dots P(F_n|C, F_1, F_2, \dots, F_{n-1}) \quad (4)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing masing petunjuk saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu persamaan seperti pada persamaan 5 sebagai berikut:

$$P(P_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (5)$$

Untuk $i \neq j$, sehingga $P(F_i|C, F_j) = P(F_i|C)$

Dari persamaan 5 diatas dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ dapat disederhanakan seperti pada persamaan 6 berikut:

$$P(C|F_1, \dots, F_n) = P(C)P(F_1|C)P(F_2|C)P(F_3|C) \quad (6)$$

Dari persamaan diatas didapatkan $(C|F_1, \dots, F_n)$ sama dengan $P(C) \prod_{i=1}^n P(F_i|C)$

Persamaan diatas merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* seperti pada persamaan 7 berikut :

$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (7)$$

Keterangan:

P : Peluang

X_i : Atribut ke i

x_i : Nilai atribut ke i

Y : Kelas yang dicari

y_j : Sub kelas Y yang dicari

μ : rata-rata seluruh atribut

σ : varian seluruh atribut (*Deviasi standart*)

Adapun alur dari metode *Naive Bayes* yang digunakan sebagai berikut:

1. Pengkajian data training dengan menggunakan beberapa skenario komposisi (*percentage split*).
2. Penghitungan Jumlah dan probabilitas, namun apabila data numerik maka:
 - a. Cari nilai *mean* dan standar deviasi dari masing-masing parameter yang merupakan data numerik.
 - b. Cari nilai probabilistic dengan cara menghitung jumlah data yang sesuai dari-kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
 - c. Mendapatkan nilai dalam tabel *mean*, standar deviasi dan probabilitas. Hal tersebut bertujuan untuk mendapatkan kelayakan validitas diskriminatif, yang dapat dilihat oleh skor AVE dari setiap atribut yang harus lebih dari varian atribut-lainnya.

D. Uji Coba

Tahap uji coba adalah tahap dimana dilakukan uji coba dari model yang sudah diperoleh menggunakan beberapa skenario untuk mencari nilai optimal dalam penelitian. Uji coba dilakukan dengan menggunakan *percentage split*. Dengan pengujian tersebut akan dihasilkan nilai persentase ideal dalam melakukan pengujian data dengan metode *naïve bayes*.

E. Evaluasi

Tahapan terakhir dalam penelitian ini adalah tahap evaluasi. Evaluasi dilakukan dengan menganalisa hasil dari pengujian menggunakan *percentage split* dengan mencari nilai optimal sehingga dapat disimpulkan nilai pengujian terbaik dalam proses uji coba yang dilakukan.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang digunakan untuk penelitian ini adalah data pendaftar sebanyak 2962 data. Data yang digunakan memiliki atribut jenis kelamin, nilai toefl ketika mendaftar, Indeks Prestasi Kumulatif (IPK S1) ketika mendaftar, usia pendaftar ketika mendaftar, pekerjaan-ketika mendaftar, dan sumber biaya ketika mendaftar. Atribut tersebut digunakan karena atribut tersebut merupakan syarat masuk yang telah ditentukan oleh perguruan tinggi sebagai persyaratan pendaftaran jenjang magister.

B. Pra Proses Data

Pra proses data dilakukan terhadap 2962 data yang telah didapatkan. Setelah dilakukan 3 tahapan pada pra proses data maka dihasilkan data siap olah sebanyak 2251 data. Data tersebut berupa data table dengan atri-

but yang telah ditentukan sebelumnya, yaitu jenis kelamin mahasiswa, nilai toefl, Indeks Prestasi Kumulatif (IPK S1), usia, pekerjaan, dan sumber biaya. Sebagai contoh data yang telah siap diolah dapat dilihat pada Tabel 1.

TABEL 1
 DATASET PENERIMAAN MAHASISWA BARU JENJANG MAGISTER

Data	Jenis Kelamin	Nilai Toefl	IPK S1	Usia	Pekerjaan	Sumber Biaya
Data 1	2	1	2	1	1	1
Data 2	1	1	1	2	2	2
Data 3	2	3	4	2	1	4
Data 4	2	4	2	4	1	3
Data 5	1	2	1	3	1	2
....
Data 2251	1	2	2	1	2	3

C. Pemodelan Data

Pada tahap ini dilakukan pemodelan dengan pengujian data kelulusan mahasiswa jenjang pendidikan magister menggunakan metode *naïve bayes*. Metode tersebut digunakan untuk memprediksi ketepatan waktu kelulusan mahasiswa. Langkah yang dilakukan adalah dengan menghitung nilai probabilitas nilai “tepat waktu” dan “tidak tepat waktu” dari masing-masing atribut pada keseluruhan kasus “tepat waktu” dan “tidak tepat waktu” dari keseluruhan data. Hasil perhitungan dapat dilihat pada Tabel 2.

TABEL 2
 PERHITUNGAN PROBABILITAS PRIOR KESELURUHAN

Atribut	Sub Atribut	Jumlah	Tepat	Tidak Tepat	P(X c1)	
					Tepat	Tidak Tepat
Total		2251	455	1796	0,202	0,798
jeniskelamin	1	813	187	626	0,230	0,770
	2	1438	268	1170	0,186	0,814
Pekerjaan	1	1247	278	969	0,223	0,777
	2	1004	177	827	0,176	0,824
Usia	1	689	172	517	0,250	0,750
	2	672	137	535	0,204	0,796
	3	355	47	308	0,132	0,868
	4	535	99	436	0,185	0,815
Toefl	1	670	145	525	0,216	0,784
	2	477	98	379	0,205	0,795
	3	558	103	455	0,185	0,815
	4	546	109	437	0,200	0,800
Biaya	1	481	77	404	0,160	0,840
	2	175	70	105	0,4	0,600
	3	1318	265	1053	0,201	0,799
	4	7	1	6	0,143	0,857
	5	167	22	145	0,132	0,868
	6	35	7	28	0,2	0,800

	7	4	0	4	0	1,000
	8	64	13	51	0,203	0,797
IPK	1	556	102	454	0,183	0,817
	2	566	104	462	0,184	0,816
	3	548	119	429	0,217	0,783
	4	581	130	451	0,224	0,776

Selain atribut di atas juga terdapat atribut program studi dengan jumlah 46 program studi. Program studi tersebut dihitung juga menggunakan pencarian nilai probabilitas keseluruhan dengan hasil akhir nilai prior pada masing-masing prodi.

D. Uji Coba

Untuk menentukan kelas yang digunakan maka diperlukan sebuah data testing atau sampel yang selanjutnya data itu disebut data probabilitas posterior karena data tersebut bersumber dari data prior yang telah dihitung sebelumnya.

TABEL 3
 PERHITUNGAN PROBABILITAS PRIOR

Atribut	Nilai	Tepat	Tidak Tepat
total kasus	2251	0,202	0,798
Jenis Kelamin	1438	0,186	0,814
pekerjaan	672	0,204	0,796
usia	535	0,185	0,815
toefl	670	0,216	0,784
biaya	481	0,160	0,840
IPK	581	0,224	0,776
prodi	118	0,017	0,983

Berdasarkan Tabel 3, dapat dihitung probabilitas dari tiap atribut yang ada. Perhitungan probabilitas lulus “tepat waktu” untuk tiap atribut:

$$= 0,202 * 0,186 * 0,204 * 0,185 * 0,216 * 0,160 * 0,224 * 0,017 = 1,866108465414144e-7$$

Perhitungan probabilitas lulus “tidak tepat waktu” untuk tiap atribut:

$$= 0,798 * 0,814 * 0,796 * 0,815 * 0,784 * 0,840 * 0,776 * 0,983 = 0,2116940055961304$$

Perbandingan probabilitas antara lulus “tepat waktu” dan “tidak tepat waktu” adalah

Probabilitas tepat waktu: 1,86610846e-7

Probabilitas tidak tepat waktu: 0,2116940055

Hasil dari perhitungan probabilitas tersebut menunjukkan nilai probabilitas tidak tepat lebih kecil dari pada nilai tepat waktu. Sehingga dapat disimpulkan bahwa data testing tersebut memiliki hasil kualifikasi “lulus tepat waktu”.

E. Evaluasi

Evaluasi dilakukan dengan menganalisa hasil pengujian dengan *presentage split*. Hal tersebut dilakukan untuk mencari nilai optimal dalam pengujian. Berdasarkan data sejumlah 2251 diperoleh hasil training dengan klasifikasi *naïve bayes* dengan tingkat akurasi 79,78%. Selanjutnya dilakukan evaluasi menggunakan *percentage split* dari 55% hingga 95% dengan menggunakan 9 skenario pengujian. Skenario dalam proses evaluasi dapat dilihat pada Tabel 4.

Skenario 1 dilakukan dengan menggunakan data training 55 % dan data testing 45 % Hasil evaluasi dapat dilihat pada Tabel 5.

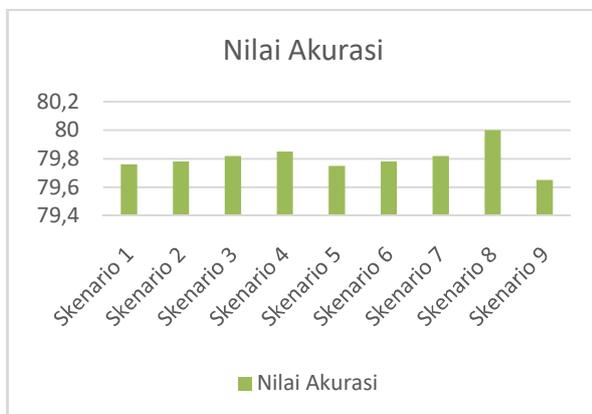
TABEL 4
TABEL SKENARIO EVALUASI

Skenario	Data Training (%)	Data Testing (%)
Skenario 1	55	45
Skenario 2	60	40
Skenario 3	65	35
Skenario 4	70	30
Skenario 5	75	25
Skenario 6	80	20
Skenario 7	85	15
Skenario 8	90	10
Skenario 9	95	5

TABEL 5
TABELHASILEVALUISKENARIO

Skenario	Data Training (%)	Data Testing (%)	Nilai Akurasi (%)
Skenario 1	55	45	79,76
Skenario 2	60	40	79,78
Skenario 3	65	35	79,82
Skenario 4	70	30	79,85
Skenario 5	75	25	79,75
Skenario 6	80	20	79,78
Skenario 7	85	15	79,82
Skenario 8	90	10	80
Skenario 9	95	5	79,65

Perbandingan hasil evaluasi juga disajikan dalam bentuk grafik pada Gambar 2.



Gambar. 2. Hasil Evaluasi

Dari hasil uji coba didapatkan hasil optimal pada skenario 8 dengan nilai *percentage split* sebesar 90%.

Hasil ini menunjukkan bahwa klasifikasi perkiraan kelulusan mahasiswa jenjang magister menggunakan metode *naïve bayes* dapat menghasilkan akurasi terbaik pada *percentage split* sebesar 90% dengan nilai akurasi sebesar 80%.

IV. KESIMPULAN

Hasil penelitian menunjukkan bahwa metode Naïve Bayes dapat dilakukan untuk klasifikasi perkiraan kelulusan dengan komposisi data yang digunakan 90% data training dan 10% data testing. Komposisi ini menghasilkan nilai akurasi sebesar 80%

DAFTAR PUSTAKA

- [1] "PERATURAN PEMERINTAH REPUBLIK INDONESIA NOMOR 66 TAHUN 2010.pdf." .
- [2] "Peraturan Rektor Unair No 27 Tahun 2018.pdf." .
- [3] A. C. Lagman *et al.*, "Embedding naïve Bayes algorithm data model in predicting student graduation," in *Proceedings of the 3rd International Conference on Telecommunications and Communication Engineering*, Tokyo Japan, Nov. 2019, pp. 51–56, doi: 10.1145/3369555.3369570.
- [4] "Peraturan Rektor Unair No 29 Tahun 2018.pdf." .
- [5] A. Romadhona and H. Himawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree," vol. 13, p. 15, 2017.
- [6] D. Novianti, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. - J. Komput. Dan Inform.*, vol. 21, no. 1, pp. 49–54, Mar. 2019, doi: 10.31294/p.v21i1.4979.
- [7] N. Priandini, B. Zaman, and E. Purwanti, "Categorizing document by fuzzy C-Means and K-nearest neighbors approach," Surabaya, Indonesia, 2017, p. 020012, doi: 10.1063/1.4994415.
- [8] S. Fitri, "Perbandingan Kinerja Algoritma Klasifikasi Naïve Bayesian, LAZY-LBK, Zero-R, Dan Decision Tree- J48," vol. 15, no. 1, p. 5, 2014.
- [9] A. Rohman, "Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa," *Neo Tek.*, vol. 1, no. 1, Mar. 2015, doi: 10.37760/neoteknika.v1i1.350.
- [10] A. Nugroho, R. Hidayatillah, S. Sumpeno, and M. H. Purnomo, "Klasifikasi Interaksi Kampanye di Media Sosial Menggunakan Naïve Bayes Kernel Estimator," *J. Nas. Tek. Elektro Dan Teknol. Inf. JNTETI*, vol. 8, no. 2, p. 107, May 2019, doi: 10.22146/jnteti.v8i2.499.
- [11] B. Zaman, A. Justitia, Kretawewika Nuraga Sani, and Endah Purwanti, "An Indonesian Hoax News Detection System Using Reader Feedback and Naïve Bayes Algorithm," *Cybern. Inf. Technol.*,

- vol. 20, no. 1, pp. 82–94, Mar. 2020, doi: 10.2478/cait-2020-0006.
- [12]H. Kim, J. Kim, J. Kim, and P. Lim, “Towards perfect text classification with Wikipedia-based semantic Naïve Bayes learning,” *Neurocomputing*, vol. 315, pp. 128–134, Nov. 2018, doi: 10.1016/j.neucom.2018.07.002.
- [13]Z. E. Rasjid and R. Setiawan, “Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques,” *Procedia Comput. Sci.*, vol. 116, pp. 107–112, 2017, doi: 10.1016/j.procs.2017.10.017.
- [14]M. Ridwan and H. Suyono, “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier,” vol. 7, no. 1, p. 6, 2013.
- [15]F. I. Puspitasari, S. Soebandhi, D. H. Subagyo, and A. Nugroho, “Altruism and Egoism in e-WOM: The moderating effect of Facebook perceived ease of use,” in *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia, Sep. 2019, pp. 536–541, doi: 10.1109/ISEMANTIC.2019.8884276.
- [16]D. L. Fithri and E. Darmanto, “Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes,” p. 6, 2014.