

Klasifikasi Tingkat Kelulusan Mahasiswa Elektronika Menggunakan Algoritma *Naïve Bayes Classifier*

(Studi Kasus : Pendidikan Teknik Informatika FT-UNP)

Helga Yolanda Sardi^{1*}, Khairi Budayawan²

¹Prodi Pendidikan Teknik Informatika Fakultas Teknik Universitas Negeri Padang

²Jurusan Teknik Elektronika Universitas Negeri Padang

Jl. Prof. Hamka Kampus UNP Air Tawar Padang

*Corresponding author e-mail : helgayolandas@gmail.com

ABSTRAK

Universitas Negeri Padang merupakan perguruan tinggi yang penyimpanan datanya telah terkomputerisasi, salah satunya data kelulusan mahasiswa. Berdasarkan data kelulusan mahasiswa Program Studi Pendidikan Teknik Informatika tahun masuk 2013-2015 menunjukkan bahwa masih banyak mahasiswa yang menempuh masa studi lebih dari 9 semester. Penelitian ini bertujuan untuk mengklasifikasi tingkat kelulusan mahasiswa dan menerapkan metode *data mining* dengan menggunakan algoritma *naïve bayes classifier* yang dibantu dengan aplikasi Rapidminer. Klasifikasi ini menggunakan 20 atribut sebagai atribut kontrol yang merupakan transkrip nilai semester 1 sampai dengan semester 3. Berdasarkan pengujian yang dilakukan terhadap mahasiswa tahun masuk 2014 diperoleh akurasi sebesar 79,07% dengan 43 sampel data dan pengujian yang dilakukan terhadap mahasiswa tahun masuk 2015 akurasi yang dihasilkan sebesar 68% dengan 50 sampel data.

Kata kunci : Kelulusan Tepat Waktu, *Data mining*, Klasifikasi, *Naïve Bayes Classifier*

ABSTRACT

Padang State University is a university whose data storage has been computerized, one of which is student graduation data. Based on data on student graduation of the Informatics Engineering Education Study Program in 2013-2015, it shows that there are still many students who take study periods of more than 9 semesters. This study aims to classify student graduation rates and apply data mining methods using the *naïve Bayes classifier* algorithm assisted by the Rapidminer application. This classification uses 20 attributes as control attributes which are transcripts of grades from semester 1 to semester 3. Based on the tests carried out on students in 2014, an accuracy of 79.07% was obtained with 43 data samples and tests carried out on students in 2015 with the accuracy that resulted in 68% with 50 data samples.

Keywords: Graduation rate, *Data mining*, Classification, *Naïve Bayes Classifier*

I. PENDAHULUAN

Berdasarkan Lampiran Peraturan Badan Akreditasi Nasional Perguruan Tinggi (BAN PT) Nomor 5 tahun 2019 Tentang Instrumen Akreditasi Program Studi point 56 tentang Masa Studi disebutkan bahwa poin (nilai) tertinggi yaitu 4 diperoleh apabila masa studi mahasiswa lebih dari 3,5 tahun dan kurang atau sama dengan 4,5 tahun, sehingga dapat disimpulkan bahwa kriteria kelulusan tepat waktu mengikuti poin tertinggi dari penilaian BAN-PT tersebut. (Badan Akreditasi Nasional Perguruan Tinggi, 2019)

Masa penyelesaian studi oleh mahasiswa memiliki pengaruh penting bagi mahasiswa yang bersangkutan dan juga bagi institusi khususnya program studi dalam menentukan langkah-langkah yang diperlukan sehingga proses kegiatan studi mahasiswa menjadi lancar.

Berdasarkan data kelulusan mahasiswa Program Studi Pendidikan Teknik Informatika yang diambil dari tahun 2013-2015 terdapat sedikit mahasiswa yang lulus tepat waktu dan hal ini menunjukkan bahwa masih banyak mahasiswa Program Studi Pendidikan Teknik Informatika yang menempuh masa studi lebih dari 9 semester.

Jumlah mahasiswa lulusan tepat waktu dapat meningkat dengan cara meningkatkan kualitas pembelajaran serta layanan akademik untuk mahasiswa. Disamping itu, apabila waktu penyelesaian studi mahasiswa dapat diprediksikan lebih awal maka solusi penanganan mahasiswa akan lebih efektif dan selanjutnya bisa menjadi strategi bagi jurusan dalam melakukan suatu tindakan yang dirasa penting agar mahasiswa dapat lulus dengan tepat waktu sehingga kualitas program studi dapat meningkat.

Teknik yang digunakan dalam mengoptimalkan data supaya memiliki nilai manfaat sehingga menjadi suatu informasi dan pengetahuan adalah dengan menggunakan *data mining*. *Data mining* dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi kelulusan mahasiswa berdasarkan transkrip nilai akademik mahasiswa.

Metode dalam *data mining* diantaranya yaitu prediksi, klasifikasi, pengklusteran, dan asosiasi. Algoritma *data mining* yang akan digunakan dalam penelitian ini adalah Algoritma *Naïve Bayes Classifier (NBC)* yaitu merupakan suatu pengklasifikasi probabilitas yang sederhana dalam mengaplikasikan Teorema Bayes. Ide yang diperoleh dari Teorema Bayes adalah menangani suatu masalah yang bersifat hipotesis yakni mendesain suatu klasifikasi untuk memisahkan objek.

Dalam menggunakan Algoritma *Naïve Bayes Classifier* untuk menghitung nilai probabilitas tiap kelas pengklasifikasi kelulusan mahasiswa menggunakan beberapa atribut pendukung, sehingga menghasilkan suatu rekomendasi berdasarkan data pembelajaran atau *data training*.

II. METODE

Pada penelitian ini alur penelitian yang akan dilakukan dengan melakukan studi pustaka, pengumpulan data, praproses data, klasifikasi dengan *Naïve Bayes Classifier*, serta membuat Kesimpulan Penelitian. Penelitian ini dilakukan untuk mengklasifikasikan tingkat kelulusan Mahasiswa dengan beberapa tahap yaitu (1) proses analisis data (2) perhitungan *Naïve Bayes Classifier* (3) pengujian data menggunakan *tools data mining RapidMiner*.

Proses Analisis Data

Dalam proses analisis data menggunakan data latih (*data training*) berupa data mahasiswa tahun masuk 2013 yang diperoleh dari *database SIE* (Sistem Informasi Eksekutif) dan *SIA* (Sistem Informasi Akademik) Universitas Negeri Padang, data tersebut diklasifikasikan menggunakan algoritma *naïve bayes classifier*, sehingga dari proses klasifikasi menghasilkan suatu model yang akan dilakukan penerapannya pada data uji (*data testing*)

data mahasiswa tahun masuk 2014 dan 2015 yang nantinya akan menghasilkan suatu keluaran data uji tersebut.

Perhitungan *Naïve Bayes Classifier*

Adapun tahapan dalam perhitungan *naïve bayes classifier* yaitu (a) menyiapkan data set yang digunakan sebagai bahan pelatihan. (b) menghitung probabilitas setiap kelas dan fitur yang digunakan. (c) menghitung probabilitas akhir. (d) membandingkan probabilitas akhir.

III. HASIL DAN PEMBAHASAN

Analisis Data Mahasiswa dengan Algoritma *Naïve Bayes Classifier*

Atribut yang digunakan dalam penelitian ini terangkum pada tabel berikut:

Tabel 1. Atribut yang digunakan dalam penelitian

Atribut	Penjelasan
X1	Nilai matakuliah Algoritma Pemrograman
X2	Nilai matakuliah Desain Grafis
X3	Nilai matakuliah Elektronika Komputer
X4	Nilai matakuliah Fisika Teknik
X5	Nilai matakuliah Kalkulus dan Aljabar
X6	Nilai matakuliah Konsep Teknologi Informasi
X7	Nilai matakuliah Sistem Informasi
X8	Nilai matakuliah Matematika Diskrit
X9	Nilai matakuliah Organisasi dan Arsitektur Komputer
X10	Nilai matakuliah Praktikum Fisika
X11	Nilai matakuliah Praktikum Algoritma dan Struktur Data
X12	Nilai matakuliah Struktur Data
X13	Nilai matakuliah Jaringan Komputer dan Komunikasi Data
X14	Nilai matakuliah Pemeliharaan Perangkat Komputer
X15	Nilai matakuliah Pemrograman Berorientasi Objek
X16	Nilai matakuliah Pemrograman Berbasis Web
X17	Nilai matakuliah Praktikum Pemrograman Berorientasi Objek
X18	Nilai matakuliah Sistem Basis Data
X19	Nilai matakuliah Sistem Operasi
X20	Nilai matakuliah Teknik Komputasi
Status Kelulusan	Terdiri dari lulus tepat waktu dan lulus tidak tepat waktu

Klasifikasi dengan *Naïve Bayes Classifier* Menghitung Probabilitas Kelas

Berdasarkan *data training* yang digunakan, terdapat 2 kelas klasifikasi yang dibentuk dari probabilitas kelas tepat waktu, yang dapat dilihat pada tabel di bawah:

Tabel 2. Probabilitas kelas lulus tepat waktu

Kelas Lulus Tepat Waktu	
Lulus Tepat Waktu "Ya" = 13	Lulus Tepat Waktu "Tidak" = 77
P(Lulus Tepat Waktu "Ya") = 13/90 = 0.1444	P(Lulus Tepat Waktu "Tidak") = 77/90 = 0.8556

Menghitung Nilai Mean dan Standar Deviasi

Berdasarkan atribut yang digunakan dalam penelitian ini bernilai kontinu. Langkah pertama yang perlu dilakukan yaitu mencari suatu nilai rata-rata (*mean*) dan standar deviasi setiap atribut pada kelas Ya dan Tidak.

Notasi yang digunakan dalam menghitung suatu nilai *mean* dapat menggunakan persamaan sebagai berikut :

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \tag{1}$$

atau,

$$\mu = \frac{x_1+x_2+ x_3+\dots+ x_n}{n} \tag{2}$$

Dimana :

μ : Rata-rata hitung (*mean*)

x_i : Nilai sample ke-i

n : Jumlah sample

Adapun dalam menghitung suatu nilai simpangan baku (standar deviasi) dapat dilihat pada persamaan berikut :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \tag{3}$$

Dimana :

σ : Standar deviasi

x_i : Nilai x ke-i

μ : Rata-rata hitung

n : Jumlah sampel

Nilai *mean* dan standar deviasi pada tiap kelas YA dan TIDAK dari ke-20 atribut terangkum dalam tabel di bawah :

Tabel 2. Nilai *mean* dan standar deviasi

Nama Atribut	Mean		Standar Deviasi	
	Tepat Waktu	Terlambat	Tepat Waktu	Terlambat
(X1)	9	8.7584	0.9721	1.1811
(X2)	7.0615	7.0701	0.2631	0.2934
(X3)	9.2308	8.6844	0.8321	1.1597
(X4)	6.0615	4.7636	1.5565	1.8037
(X5)	6.4769	5.6468	1.4912	1.2985
(X6)	6.8769	6.7195	0.3113	0.5696
(X7)	6.9692	6.7247	0.3038	0.3129
(X8)	11.2615	10.5857	1.1266	1.7253
(X9)	9.8769	8.8286	1.5990	1.71745
(X10)	2.6538	2.4260	0.5825	0.7049
(X11)	7.3385	7.4571	0.7136	0.7880
(X12)	5.1231	4.3221	0.8584	1.3223
(X13)	7.6846	7.1532	1.8270	1.5016
(X14)	6.7231	6.6390	0.7049	0.4963

(X15)	10.5923	10.4104	0.3947	0.2891
(X16)	6.8308	6.6727	0.8199	0.7055
(X17)	7.2308	6.5792	0.7111	1.2106
(X18)	9.9231	8.3545	1.3989	2.4254
(X19)	10.8462	10.5	1.3685	1.4705
(X20)	6.8769	6.2519	0.9400	1.1227

Menghitung Probabilitas Akhir setiap Kelas

Dalam menghitung suatu nilai probabilitas pada fitur mengacu berdasarkan nilai yang telah didapatkan pada tabel 2. Notasi yang dipakai adalah

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp \frac{-(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \tag{4}$$

Tabel 4. Perhitungan probabilitas akhir setiap atribut

Nilai Probabilitas	Lulus Tepat Waktu "Ya"	Lulus Tepat Waktu "Tidak"
X1 = 9.9	0.2637	0.2302
X2 = 6.6	0.1670	0.2040
X3 = 9	0.4210	0.3571
X4 = 4.6	0.2058	0.2959
X5 = 5.2	0.2265	0.3301
X6 = 6.6	0.4815	0.5172
X7 = 6.6	0.3459	0.6589
X8 = 10.8	0.3457	0.3015
X9 = 7.8	0.1358	0.2545
X10 = 2	0.2785	0.3960
X11 = 8	0.3074	0.3564
X12 = 2	0.0006	0.0734
X13 = 6.9	0.2692	0.3210
X14 = 6.6	0.4681	0.5647
X15 = 9.9	0.1364	0.3468
X16 = 6	0.2638	0.3015
X17 = 6.6	0.3193	0.3626
X18 = 9	0.2714	0.2473
X19 = 9.9	0.2686	0.3028
X20 = 6	0.2664	0.3672

Perhitungan probabilitas akhir dalam tiap kelas menggunakan notasi :

$$P(X|Y = y) = \prod_{i=1}^q P(X_i | Y = y) \tag{5}$$

Hasil $P(X|Lulus\ Tepat\ Waktu = Ya) = 9.3528 \cdot 10^{-15}$ dan $P(X|Lulus\ Tepat\ Waktu = Tidak) = 8.4393 \cdot 10^{-11}$.

Berdasarkan hasil perhitungan diatas dapat diketahui hasil akhir probabilitas terbesar ada di kelas Lulus Tepat Waktu "TIDAK". Hasil pengujian atas nama Nur Fadillah ID 73 dapat dilihat pada tabel di bawah:

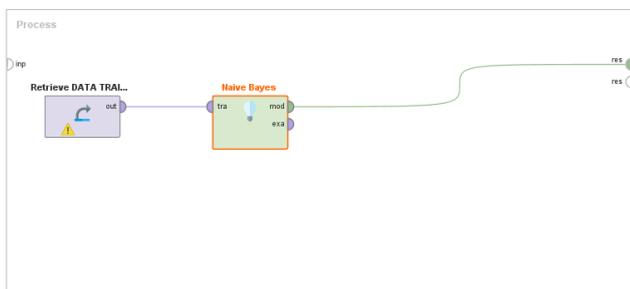
Tabel 5. Hasil akhir klasifikasi

ID MHS	Nama MHS	Kelas :		Hasil
		Lulus Tepat Waktu Actual	Lulus Tepat Waktu Prediksi	
73	Nur Fadillah	Tidak	Tidak	AKURAT

Data mahasiswa di atas merupakan salah satu *data testing* dengan kategori mahasiswa yang tidak lulus tepat waktu. Setelah melakukan perhitungan menggunakan *naïve bayes classifier* hasil yang diperoleh masuk kelas Lulus Tepat Waktu “Tidak”. Dengan predikat AKURAT.

Pengujian dengan Tools RapidMiner

Pengujian ini bertujuan untuk menguji keakuratan klasifikasi tingkat kelulusan mahasiswa menggunakan algoritma *naïve bayes classifie*. Data yang di analisis adalah data mahasiswa prodi Pendidikan Teknik Informatika Universitas Negeri Padang tahun 2013 sampai dengan 2015 dan data didapat dari SIA dan SIE UNP.



Gambar 1. Pelatihan *data training*

Gambar di atas menunjukkan proses pelatihan data training menggunakan algoritma Naïve Bayes yang menghasilkan data keluaran seperti gambar dibawah:

SimpleDistribution

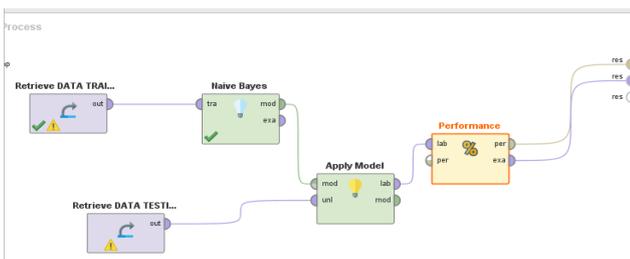
```
Distribution model for label attribute Lulus Tepat Waktu

Class Tidak (0.856)
20 distributions

Class Ya (0.144)
20 distributions
```

Gambar 2. Hasil pelatihan *data training*

Gambar di atas merupakan model distribusi pengujian dengan algoritma *naïve bayes*. Hasil perincian kelas ya = 0.144 dengan 20 distribusi, dan hasil perincian kelas tidak = 0.865 dengan 20 distribusi.



Gambar 3. Implementasi *naïve bayes* untuk uji *performance* untuk data tahun masuk 2015

Gambar di atas merupakan suatu pengaplikasian model dengan rapidminer pada kasus data mahasiswa tahun masuk 2014.

Tabel 6. Hasil pengujian *system* untuk data tahun masuk 2014

	True Tidak	True Ya	class precision
pred. Tidak	30	9	76.92%
pred. Ya	0	4	100 %
class recall	100%	30.77%	

Menggunakan rapidminer pada hasil pengujian sistem yang telah dilakukan didapatkan hasil suatu nilai *accuracy* dengan nilai persentase 79,07%, *class precision* pred Ya sebesar 100 % dan *class recall* Ya 30.77%.

Tabel 8. Hasil pengujian *system* untuk data tahun masuk 2015

	True Tidak	True Ya	class precision
pred. Tidak	8	6	57.14%
pred. Ya	10	26	72.22 %
class recall	44.44%	81.25%	

Pada percobaan menggunakan data mahasiswa tahun masuk 2015 diperoleh nilai *accuracy* sebesar 68%, *class precision* pred Ya sebesar 72.22% dan *class recall* Ya 81.25%.

Tabel 9. Rangkuman hasil eksperimen *naïve bayes*

Data Set	Jumlah Record	Accuracy	Precision	Recall
2014	43	79.09%	100 %	30.77%.
2015	50	68 %	72.22%	81.25%.

IV. KESIMPULAN

Berdasarkan analisis dan pengujian data mahasiswa prodi Pendidikan Teknik Informatika tahun masuk 2013-2015 dapat diambil kesimpulan sebagai berikut :

1. Metode *data mining* dengan algoritma *naïve bayes classifier* telah dapat diterapkan kedalam klasifikasi kelulusan mahasiswa Pendidikan Teknik Informatika FT-UNP.
2. Tingkat akurasi yang dihasilkan dari suatu proses klasifikasi dengan menggunakan algoritma *naïve bayes classifier* sangat ditentukan dari pemilihan *data training* dan pemilihan atribut inputpun juga berpengaruh terhadap hasil output (hasil prediksi).
3. Data keluaran atau prediksi kelulusan tepat waktu mahasiswa yang dihasilkan oleh program

adalah suatu keputusan apakah mahasiswa tersebut termasuk dalam klasifikasi lulusan tepat waktu “Ya” atau “Tidak”, sehingga data tersebut menjadi tolak ukur dalam menunjang strategi kelulusan di program studi Pendidikan Teknik Informatika.

4. Berdasarkan pengujian yang dilakukan terhadap mahasiswa tahun masuk 2014 diperoleh akurasi sebesar 79.07% dengan 43 record data dan terhadap mahasiswa tahun masuk 2015 diperoleh akurasi sebesar 68% dengan 50 record data.

V. SARAN

Adapun hal-hal yang dapat dijadikan saran adalah sebagai berikut: 1) Mencoba menggunakan aplikasi tools *data mining* selain rapidminer dalam klasifikasi kelulusan Mahasiswa menggunakan algoritma *naïve bayes classifier*; 2) Dapat dilakukan uji coba menggunakan algoritma pengklasifikasi *data mining* seperti algoritma C4.5, ID3, CART dan hasilnya dapat dilakukan perbandingan dan analisa untuk pengembangan penelitian kedepan.

DAFTAR PUSTAKA

- [1] Saputra, H. K. (2018). Analisis Data Mining Untuk Pemetaan Mahasiswa Yang Membutuhkan Bimbingan dan Konseling Menggunakan Algoritma Naïve Bayes Classifier. *JTIP: Jurnal Teknologi Informasi dan Pendidikan*, 11(1), 14-26.
- [2] Badan Akreditasi Nasional Perguruan Tinggi. (2019). *Matriks Penilaian Laporan Evaluasi Diri dan Laporan Kinerja Program Studi*. Jakarta.
- [3] Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *jurnal EECCIS*, 7(1), 59-64.
- [4] Handayani, L., & Maulida, E. L. (2015, November). Perkiraan Waktu Studi Mahasiswa Menggunakan Metode Klasifikasi Dengan Algoritma Naive Bayes. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri*.
- [5] Salmu, S., & Solichin, A. (2017, April). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta. In *Seminar Nasional Multidisiplin Ilmu (SENMI) 2017* (pp. 701-709).
- [6] Umam, M. H. (2017). *Analisis Perbandingan Algoritma C4. 5 dan Algoritma Naïve Bayes untuk Prediksi Kelulusan Mahasiswa (Studi Kasus: Prodi Teknik Informatika Universitas*

Muhammadiyah Jember) (Doctoral dissertation, UNIVERSITAS MUHAMMADIYAH JEMBER).