

Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode *Naive Bayes*

¹Yovi Apridiansyah, ²Nuri David Maria Veronika, ³Erwin Dwika Putra

^{1,2,3}Universitas Muhammadiyah Bengkulu, Indonesia

¹yoviapridiansyah@umb.ac.id; ²tmisz_nuricubv@gmail.com; ³erwindwikap@gmail.com;

Article Info

Article history:

Received, 04-06-2021

Revised, 18-06-2021

Accepted, 28-06-2021

Kata Kunci:

Prediksi
Kelulusan
Klasifikasi
Precision
Recall
Akurasi

ABSTRAK

Kelulusan mahasiswa tepat waktu merupakan salah satu point penilaian dalam proses akreditasi perguruan tinggi. Namun kelulusan mahasiswa tidak selalu dapat dideteksi secara cepat sehingga dapat mengurangi penilaian suatu perguruan tinggi dalam proses akreditasi. Permasalahan inilah yang muncul untuk mengetahui mahasiswa nantinya bisa lulus tepat waktu atau tidak. Dalam penelitian inilah digunakan metode klasifikasi untuk prediksi lulusan mahasiswa menggunakan algoritma *Naive Bayes*. Tujuan dari penelitian ini untuk mengetahui kelulusan mahasiswa tepat waktu atau tidak, yang diharapkan hasilnya dapat memberikan informasi dan masukan bagi pihak perguruan tinggi dalam membuat kebijakan kedepannya. Dari hasil pengujian ini didapatkan hasil bahwa dengan menerapkan algoritma *Naive Bayes* sistem dapat memprediksi kelulusan mahasiswa dengan tepat waktu, tingkat keberhasilan prediksi algoritma *naive bayes* ditentukan oleh nilai IPK mahasiswa dan lama Studi mahasiswa yang bersangkutan dengan nilai *precision* sebesar 90 %, *recall* 100 % dan akurasi sebesar 90 %.

Keywords:

Prediction
Graduation
Classification
Precision
Recall
Accurasy

ABSTRACT

On-time student graduation is one of the assessment points in the college accreditation process. But student graduation can not always be detected quickly so it can reduce the assessment of a college in the accreditation process. This problem arises to know whether students can graduate on time or not. In this study used classification method for prediction of graduate students using naive bayes algorithm. The purpose of this study is to know the graduation of students on time or not, which is expected that the results can provide information and input for the college in making policies in the future. From the results of this test obtained the results that by applying the algorithm naive bayes system can predict student graduation in a timely manner, the success rate of prediction algorithm naive bayes determined by the student's GPA value and the length of study of the student concerned with a procession value of 90%, Recal 100% and accuracy of 90%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Korespondensi:

Yovi Apridiansyah,
Program Studi Teknik Informatika,
Universitas Muhammadiyah Bengkulu,
Email: yoviapridiansyah@umb.ac.id

1. PENDAHULUAN

Kelulusan mahasiswa dapat dikatakan salah satu item penilaian penting dalam proses akreditasi institusi suatu perguruan tinggi. BAN-PT memiliki point nilai tersendiri untuk lulusan mahasiswa bagi setiap perguruan tinggi, sehingga jika dalam suatu perguruan tinggi memiliki lulusan tepat waktu setiap semesternya maka dapat membantu suatu perguruan tinggi tersebut dalam proses

penilaian akreditasi, pada penerepannya tingkat kelulusan mahasiswa tidak bisa kita pastikan mereka lulus dengan tepat waktu [1]. Maka untuk mengatasi permasalahan tersebut dibutuhkanlah teknik dalam memprediksi lulusan mahasiswa. Dengan adanya teknik prediksi ini, maka data mahasiswa dan data jumlah kelulusan mahasiswa per tahun dapat diprediksi tingkat kelulusan per tahunnya [2][3][4]. Profil dan hasil akademik mahasiswa selama menempuh proses kegiatan belajar mengajar di suatu perguruan tinggi juga merupakan sekumpulan data yang dapat menghasilkan informasi [2][5]. Salah satu teknik yang sering digunakan adalah dengan menggunakan data mining dengan metode klasifikasi [3][6][7].

Klasifikasi merupakan salah satu poin utama dalam pengolahan data mining [6][8]. Klasifikasi juga merupakan pengelompokan data dimana data yang digunakan mempunyai kelas atau target yang akan diklasifikasi [9][10]. Sehingga nantinya algoritma dapat digunakan untuk menyelesaikan masalah klasifikasi dikategorisasikan ke dalam pembelajaran yang diawasi [11].

Universitas Muhammadiyah Bengkulu merupakan sebuah perguruan tinggi swasta yang terletak di Kota Bengkulu yang menghasilkan banyak lulusan setiap tahunnya, khususnya Fakultas Teknik Informatika penerapan klasifikasi pada lulusan setiap tahunnya belum dapat kita lihat data-datanya. Data yang dapat dilihat pada lulusan fakultas teknik hanya data lulusan pada setiap periode wisuda, Oleh karena itu dengan penerapan klasifikasi data dengan algoritma *Naïve Bayes* diharapkan dapat mengetahui atau memprediksi lulusan mahasiswa setiap angkataannya.

Algoritma *Naive Bayes* itu sendiri merupakan salah satu algoritma yang digunakan untuk teknik klasifikasi data [5][11]. *Naive Bayes* digunakan untuk teknik pengklasifikasian data dengan metode kemungkinan dan statistik yang memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes* [6]. *Teorema* tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antara atribut yang satu dengan atribut yang lainnya itu saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak adanya ciri-ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas yang lainnya [4][6].

Berdasarkan Penelitian berikut ini hasil pengujian yang telah dilakukan untuk klasifikasi kelulusan mahasiswa yang menggunakan jaringan saraf tiruan *Learning Vector Quantization* ini, dapat menjadi rekomendasi bagi penasehat akademik untuk memprediksi dini kelulusan mahasiswa sehingga tindakan pencegahan *drop out* (DO) dapat diambil lebih awal. Hasil Penelitian ini menunjukkan tingkat akurasi 70% untuk data pengujian yang dikenal jaringan sesuai target dan 30 % tidak sesuai dengan target. Hal ini bisa disebabkan oleh banyak faktor lain yang mempengaruhi proses belajar mahasiswa dalam mengikuti perkuliahan baik faktor internal, maupun *external* dalam menyelesaikan studinya [2].

Penelitian berikutnya menyimpulkan Pertama, dari hasil penerapan algoritma C4.5 dalam memprediksi kelulusan mahasiswa Prodi Informatika, menghasilkan kesimpulan bahwa atribut yang paling dominan dalam klasifikasi kelulusan mahasiswa adalah IPK, TOEFL, asal daerah, dan yang terakhir adalah jenis kelamin. Kedua, Pemotongan pohon keputusan pada algoritma C4.5 dapat meningkatkan akurasi. Pohon yang tanpa pemotongan menghasilkan nilai rata-rata *precision* 63.93%, *recall* 60.73%, dan akurasi 60.52%. Sedangkan pohon keputusan yang dipotong menggunakan metode *error-based pruning* dengan menggunakan nilai *confidence* 0,4 menghasilkan *precision* 70.70%, *recall* 50.65%, dan akurasi 61.57%. Pohon yang dipotong menggunakan nilai *confidence* 0,25 menghasilkan *precision* 73.77%, *recall* 48.84%, dan akurasi 62.44%. Dari hasil tersebut menyimpulkan bahwa penggunaan nilai *confidence* 0,25 meningkatkan akurasi lebih baik daripada nilai *confidence* 0,4 [13].

Penelitian selanjutnya melakukan perancangan model prediksi kelulusan mahasiswa tepat waktu pada UIN Raden Fatah menghasilkan perhitungan data mining menggunakan algoritma *Naïve Bayes*, dengan kesimpulan bahwa kelas tahun lulus “no” / tidak lulus tepat waktu lebih kecil daripada kelas tahun lulus “yes” / lulus tepat waktu. Kedua, dari hasil observasi terhadap dataset mahasiswa Fakultas Sains dan Teknologi UIN Raden Fatah Palembang dan melalui proses perhitungan menggunakan metode klasifikasi *Naïve Bayes* dengan atribut yang telah dijelaskan di pembahasan sebelumnya, didapatkan sebuah hasil bahwa nilai akurasi terhadap klasifikasi kelulusan sebesar 82.08 %. Dimana 82.08 % bisa juga disebabkan oleh kurang kompleksitas data yang mengakibatkan model dapat memprediksi cukup akurat [14].

Dari beberapa penelitian tersebut yang menjelaskan tentang prediksi kululusan, terdapat beberapa perbedaan dengan penelitian ini diantaranya adalah penelitian ini menggunakan algoritma

Naive Bayes dengan metode klasifikasi dengan menggunakan atribut nama mahasiswa, NPM, jalur masuk, IPK mahasiswa dan Jenis kelamin. Dengan atribut dan data kriteria tersebut algoritma dapat memprediksi tingkat kelulusan mahasiswa Fakultas Teknik Universitas Muhammadiyah Bengkulu. Sehingga dengan adanya sistem prediksi kelulusan mahasiswa ini dapat membantu untuk mengetahui kelulusan mahasiswa tepat waktu atau tidak, yang diharapkan hasilnya dapat memberikan informasi dan masukan bagi pihak perguruan tinggi dalam membuat kebijakan kedepannya.

2. METODE PENELITIAN

Dalam penelitian ini model pengembangan sistem yang dipakai adalah model Incremental. Incremental model itu sendiri merupakan model pengembangan sistem yang pada merekayasa perangkat lunak berdasarkan perangkat lunak sebagai persyaratan yang dipecah menjadi beberapa bagian sehingga model pengembangannya dilakukan secara bertahap mulai dari pengumpulan data, analisa, desain, pengkodean, dan pengujian [15]. Pada tahapan pengumpulan data ini proses pengumpulan data didapat dari Prodi Teknik Informatika Universitas Muhammadiyah Bengkulu. Data yang di uji coba berupa 20 orang mahasiswa sebagai data yang lulus September 2019, 20 orang data tersebut merupakan data uji sebagai contoh untuk diterapkan ke dalam sistem serta dengan menggunakan ke-20 orang data tersebut sebagai data *testing* untuk melihat hasil prediksi terlambat atau tepat waktu kelulusan. Dengan menggunakan atribut nama mahasiswa, NPM, jalur masuk, IPK mahasiswa dan Jenis kelamin.

Pada tahapan selanjutnya yaitu analisis, dalam proses analisis data merupakan upaya mengolah data menjadi informasi, sehingga karakteristik atau sifat-sifat yang ada pada data tersebut dapat dipahami dan berguna untuk menjawab masalah yang berkaitan dengan kegiatan penelitian. Persamaan 1 merupakan persamaan yang digunakan dalam menganalisa data [11] :

$$P(X|H) = P(H|X)P(X) \quad (1)$$

Dengan X sebagai Data kelas yang belum diketahui, H merupakan Kemungkinan hipotesis data X, merupakan suatu kelas yang spesifik, P(H|X) Kemungkinan hipotesis H berdasarkan kondisi X, P(H) Kemungkinan hipotesis, P(X|H) kemungkinan X berdasarkan kondisi pada hipotesis H P(X) Kemungkinan dari X. Tahapan proses dari algoritma Naïve Bayes yang pertama menghitung jumlah kelas. Kedua, menghitung jumlah kasus yang sama dengan class yang sama. Ketiga, kalikan semua hasil variable TEPAT & TERLAMBAT, Keempat, Bandingkan hasil class TEPAT & TERLAMBAT.

Tahapan selanjutnya desain, dalam proses tahapan desain ini nantinya menghasilkan rancangan yang akan memenuhi kebutuhan yang ditentukan selama tahapan analisis berlangsung. Sehingga hasil akhirnya nanti berupa spesifikasi rancangan yang terperinci sehingga nantinya diharapkan dapat dengan mudah diwujudkan pada saat pemrograman dijalankan.

Pada proses tahapan selanjutnya yaitu pengkodean, merupakan proses yang mengimplementasikan hasil desain kedalam sebuah kode atau bahasa yang dimengerti oleh mesin komputer yang menggunakan perintah atau sintak bahasa pemrograman Web. Sistem pengkodean ini menggunakan bahasa pemrograman PHP, database MySQL XAMPP.

Tahapan terakhir yaitu pengujian, dalam tahapan pengujian ini merupakan tahap pengujian dalam model pengembangan sistem yang dibuat dengan melakukan secara berurutan pada setiap bagian-bagian agar mengurangi kendala dalam proses pemakaian dari pengguna. Pada tahapan sesudah *software* dibuat perlu dilakukan pengujian perangkat lunak. Setelah program dan *database* selesai penulis memilih metode pengujian sistem menggunakan *Blackbox* untuk menguji sistem secara keseluruhan apakah sudah berjalan dengan baik apa belum selanjutnya pengujian dengan menggunakan *precision and recall* untuk mencari tingkat keakurasian dari metode *Naive Bayes*. Serta dengan pengujian *precision dan recall* untuk mencari tingkat akurasi dari algoritma *Naive Bayes*.

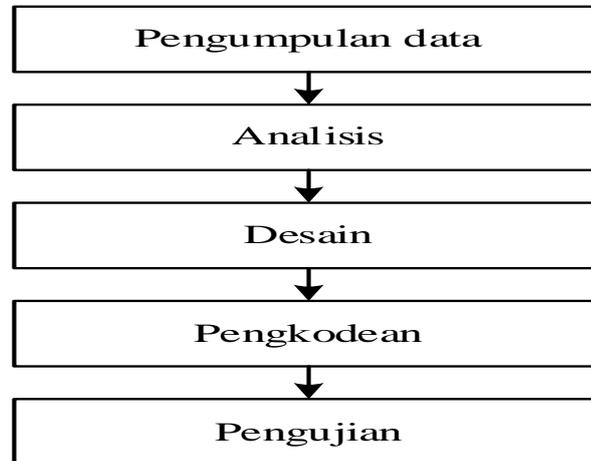
Pengujian dengan *precision and recall* serta akurasi [16] :

$$Precision = \frac{\text{Jumlah dokumen yang relevan}}{\text{Jumlah semua dokumen yang ditemukan}} \times 100\% \quad (2)$$

Sedangkan nilai *relatif Recall* dihitung dengan rumus :

$$Recall = \frac{\text{Jumlah dokumen relevan yang ditemukan}}{\text{Jumlah semua dokumen relevan didalam koleksi}} \times 100\% \quad (3)$$

$$Accuracy = \frac{\text{Jumlah dokumen relevan yang ditemukan} + \text{Dokumen relevan yang ditemukan}}{\text{Jumlah total semua dokumen relevan didalam koleksi}} \times 100\% \quad (4)$$



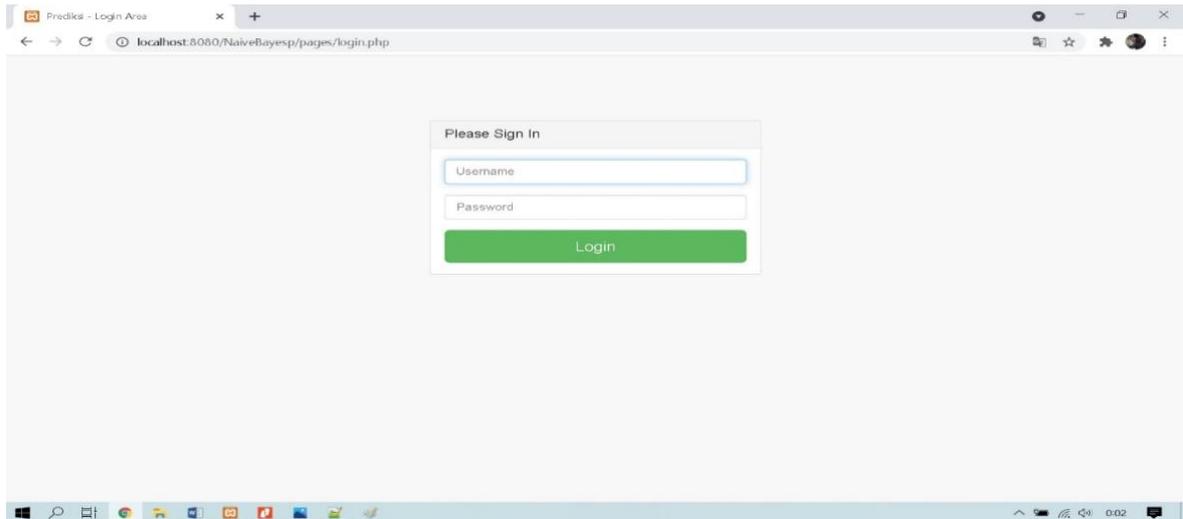
Gambar 1. Alur Penelitian [15]

3. HASIL DAN ANALISIS

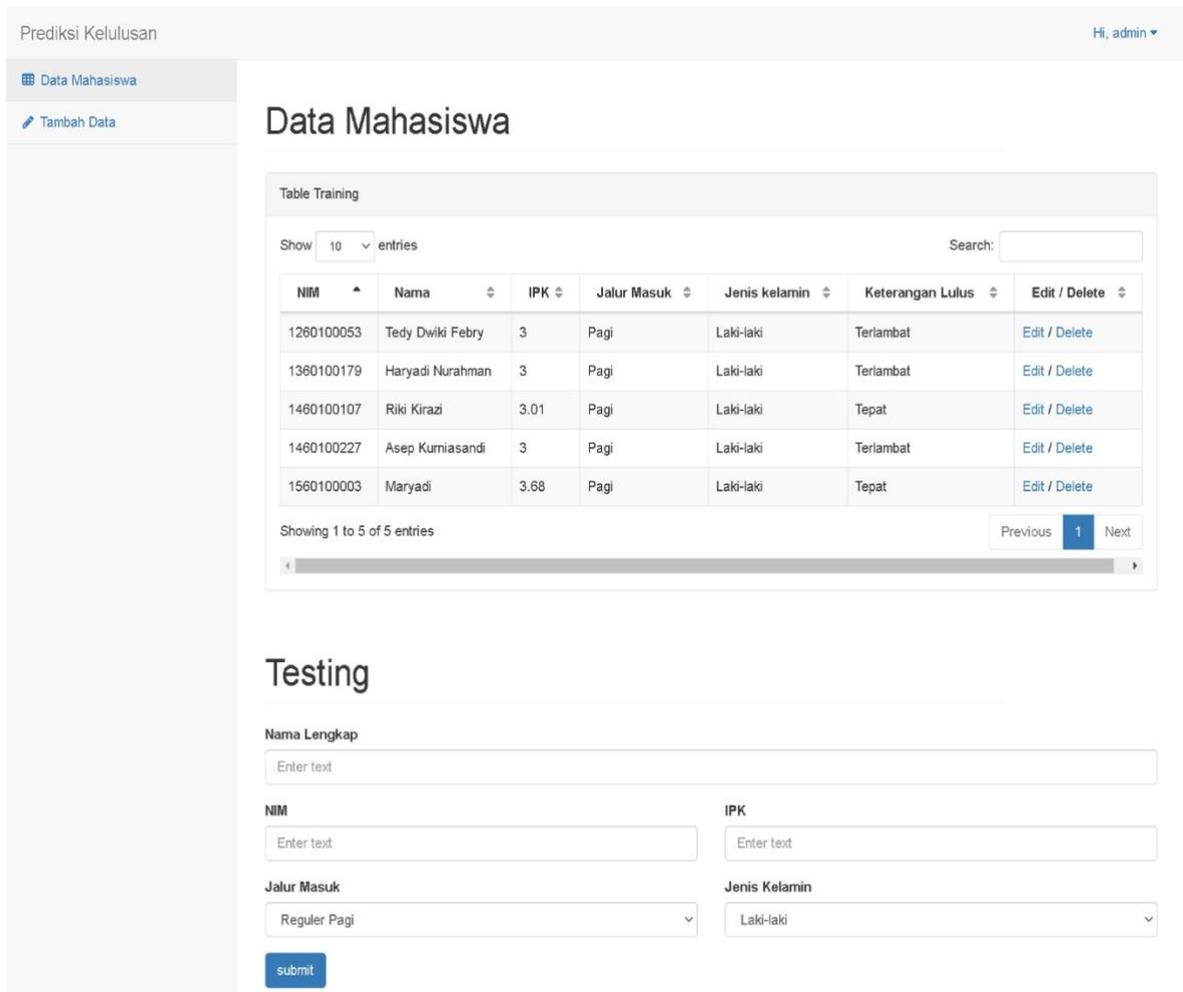
Dalam penelitian ini diharapkan dapat menghasilkan sebuah sistem untuk memprediksi kelulusan mahasiswa yang dimulai dengan proses pengumpulan data, analisis, design, kode dan pengujian agar dalam pembuatan aplikasi sistem prediksi kelulusan mahasiswa ini dapat berjalan dengan lancar.

3.1. Hasil

Tahapan pertama, adalah pengumpulan data yang merupakan tahapan proses pengumpulan data yang didapat dari Prodi Teknik Informatika Universitas Muhammadiyah Bengkulu. Data yang di uji coba berupa 20 orang mahasiswa yang lulus September 2019 dengan menggunakan atribut nama mahasiswa, NPM, jalur masuk, IPK mahasiswa dan Jenis kelamin. Setelah data terkumpul tahapan selanjutnya adalah menganalisis data yang telah diperoleh. Tahapan kedua, adalah proses analisis data merupakan upaya mengolah data menjadi informasi, sehingga karakteristik atau sifat-sifat yang ada pada data tersebut dapat dipahami dan berguna untuk menjawab masalah yang berkaitan dengan kegiatan penelitian. Tahapan ketiga, *design* yaitu tahapan dimana membuat atau merancang sistem untuk prediksi kelulusan mahasiswa.



Gambar 2. Login



Gambar 3. Menu Admin

Skripsi Hi, admin

Data Mahasiswa

Tambah Data

Tambah Data Mahasiswa

Nama Lengkap

NIM IPK

Jalur Masuk Jenis Kelamin

Gambar 4. Tambah Data

Aplikasi

Screen Capture Result

Screen Capture Result

localhost:8080/NaiveBayesp/pages/index.php#focusin

Testing

Nama Lengkap

NIM IPK

Jalur Masuk Jenis Kelamin

Table Testing

NIM	Nama	IPK	Jalur Masuk	Jenis kelamin	Keterangan Lulus
1260100053	Tedy Dwiki Febry	3.00	Pagi	Laki-laki	TERLAMBAT

Gambar 5. Test Uji Data mahasiswa

3.2 Analisis Implementasi Sistem

Dalam tahapan ini merupakan tahapan implementasi sistem untuk menguji data mahasiswa dengan mengimplementasikan algoritma *Naïve Bayes* untuk prediksi ketepatan lulusan mahasiswa Fakultas Teknik Universitas Muhammadiyah Bengkulu. Sebagai contoh sistem, mencoba data mahasiswa yang bernama Anggi fratama dikategorikan lulus tepat waktu dengan pola *Naïve Bayes* terlihat pada Gambar 6. hasil prediksi dengan *Naïve Bayes*

Naive Bayes

- Tahap 1 menghitung jumlah class/label

$P(Y=TEPAT) = 3/6$ "jumlah data "TEPAT" pada kolom "STATUS KELULUSAN" dibagi jumlah data"
 $P(Y=TERLAMBAT) = 3/6$ "jumlah data "TERLAMBAT" pada kolom "STATUS KELULUSAN" dibagi jumlah data"

- Tahap 2 menghitung jumlah kasus yang sama dengan class yang sama

$P(IPK = 3.48 | Y = TEPAT) = 1/3$ "jumlah data IPK "3.48" dengan keterangan "TEPAT" dibagi jumlah data TEPAT"
 $P(IPK = 3.48 | Y = TERLAMBAT) = 0/3$ "jumlah data IPK "3.48" dengan keterangan "TERLAMBAT" dibagi jumlah data TERLAMBAT"

$P(JALUR MASUK = Pagi | Y = TEPAT) = 3/3$ "jumlah data jalur masuk "Pagi" dengan keterangan "TEPAT" dibagi jumlah data TEPAT"
 $P(JALUR MASUK = Pagi | Y = TERLAMBAT) = 3/3$ "jumlah data jalur masuk "Pagi" dengan keterangan "TERLAMBAT" dibagi jumlah data TERLAMBAT"

$P(JENIS KELAMIN = Laki-laki | Y = TEPAT) = 3/3$ "jumlah data jenis kelamin "Laki-Laki" dengan keterangan "TEPAT" dibagi jumlah data TEPAT"
 $P(JENIS KELAMIN = Laki-laki | Y = TERLAMBAT) = 3/3$ "jumlah data jenis kelamin "Laki-laki" dengan keterangan "TERLAMBAT" dibagi jumlah data TERLAMBAT"

- Tahap 3 kalikan semua hasil variable TEPAT & TERLAMBAT

$P(IPK = 3.48) \cdot (JALUR MASUK = Pagi) \cdot (KELAMIN = Laki-laki) | TEPAT$
 $= (P(IPK = 3.48 | Y=TEPAT) \cdot P(JALUR MASUK = Pagi | Y= TEPAT) \cdot P(JENIS KELAMIN = Laki-laki | Y=TEPAT))$

- Tahap 2 menghitung jumlah kasus yang sama dengan class yang sama

$P(IPK = 3.48 | Y = TEPAT) = 1/3$ "jumlah data IPK "3.48" dengan keterangan "TEPAT" dibagi jumlah data TEPAT"
 $P(IPK = 3.48 | Y = TERLAMBAT) = 0/3$ "jumlah data IPK "3.48" dengan keterangan "TERLAMBAT" dibagi jumlah data TERLAMBAT"

$P(JALUR MASUK = Pagi | Y = TEPAT) = 3/3$ "jumlah data jalur masuk "Pagi" dengan keterangan "TEPAT" dibagi jumlah data TEPAT"
 $P(JALUR MASUK = Pagi | Y = TERLAMBAT) = 3/3$ "jumlah data jalur masuk "Pagi" dengan keterangan "TERLAMBAT" dibagi jumlah data TERLAMBAT"

- Tahap 2 menghitung jumlah kasus yang sama dengan class yang sama

$P(IPK = 3.48 | Y = TEPAT) = 1/3$ "jumlah data IPK "3.48" dengan keterangan "TEPAT" dibagi jumlah data TEPAT"
 $P(IPK = 3.48 | Y = TERLAMBAT) = 0/3$ "jumlah data IPK "3.48" dengan keterangan "TERLAMBAT" dibagi jumlah data TERLAMBAT"

$P(JALUR MASUK = Pagi | Y = TEPAT) = 3/3$ "jumlah data jalur masuk "Pagi" dengan keterangan "TEPAT" dibagi jumlah data TEPAT"
 $P(JALUR MASUK = Pagi | Y = TERLAMBAT) = 3/3$ "jumlah data jalur masuk "Pagi" dengan keterangan "TERLAMBAT" dibagi jumlah data TERLAMBAT"

$P(JENIS KELAMIN = Laki-laki | Y = TEPAT) = 3/3$ "jumlah data jenis kelamin "Laki-Laki" dengan keterangan "TEPAT" dibagi jumlah data TEPAT"
 $P(JENIS KELAMIN = Laki-laki | Y = TERLAMBAT) = 3/3$ "jumlah data jenis kelamin "Laki-Laki" dengan keterangan "TERLAMBAT" dibagi jumlah data TERLAMBAT"

- Tahap 3 kalikan semua hasil variable TEPAT & TERLAMBAT

$P(IPK = 3.48) \cdot (JALUR MASUK = Pagi) \cdot (KELAMIN = Laki-laki) | TEPAT$
 $= (P(IPK = 3.48 | Y=TEPAT) \cdot P(JALUR MASUK = Pagi | Y= TEPAT) \cdot P(JENIS KELAMIN = Laki-laki | Y=TEPAT))$

Gambar 6. Hasil Prediksi Dengan Naive Bayes

Hasil Analisa

Dari hasil percobaan diatas Implementasi Algoritma *Naïve Bayes* untuk Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu. Dijelaskan bahwa Klasifikasi dengan *Naïve Bayes* bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *Naïve Bayes* sebagai berikut:

Contoh pengujian algoritma *Naïve Bayes* pada aplikasi dengan nama mahasiswa Anggi Fratama :

Tahap 1 menghitung jumlah class/label

- $P(Y = TEPAT) = \frac{19}{20}$ = Jumlah data “Tepat” pada kolom ‘STATUS KELULUSAN’ dibagi jumlah data.
- $P(Y = Terlambat) = \frac{1}{20}$ = Jumlah data “Terlambat” pada kolom ‘STATUS KELULUSAN’ dibagi jumlah data.

Tahap 2 menghitung jumlah kasus yang sama dengan class yang sama

- $P(IPK = 3.48|Y = Tepat) Y \frac{1}{19}$ "jumlah data IPK “3.48” dengan keterangan “TEPAT” dibagi jumlah data TEPAT"
- $P(IPK = 3.48|Y = Terlambat) Y \frac{0}{19}$ "jumlah data IPK “3.48” dengan keterangan “Terlambat” dibagi jumlah data Terlambat "
- $P(Jalur Masuk = Pagi|Y = Tepat) Y \frac{19}{19}$ "jumlah data jalur masuk “Pagi” dengan keterangan “Tepat” dibagi jumlah data Tepat"
- $P(Jalur Masuk = Pagi|Y = Terlambat) Y \frac{1}{1}$ "jumlah data jalur masuk “Pagi” dengan keterangan “Terlambat” dibagi jumlah data Terlambat"
- $P(Jenis kelamin = Laki - laki|Y = Tepat) Y \frac{1}{1}$ "jumlah data jenis kelamin “Laki-lakii” dengan keterangan “Tepat” dibagi jumlah data Tepat"
- $P(Jenis kelamin = Laki - laki|Y = Terlambat) Y \frac{1}{1}$ "jumlah data jenis kelamin “Laki-lakii” dengan keterangan “Terlambat” dibagi jumlah data Terlambat"

Tahap 3 kalikan semua hasil variable TEPAT & TERLAMBAT

- $P(IPK = 3.48). (Jalur masuk = Pagi). (Kelamin = Laki - laki)|Tepat) = \{P(P(IPK = 3.48|Y = Tepat). P(Jalur masuk = Pagi|Y = Tepat). P(Jenis Kelamin = Laki - laki|Y = Tepat) \} = \frac{1}{19} \cdot \frac{19}{19} \cdot \frac{18}{19} \cdot \frac{19}{20} = 0.047368421052632$
- $P(IPK = 3.48). (Jalur masuk = Pagi). (Kelamin = Laki - laki)|Terlambat) = \{P(P(IPK = 3.48|Y = Terlambat). P(Jalur masuk = Pagi|Y = Terlambat). P(Jenis Kelamin = Laki - laki|Y = Terlambat) = \frac{0}{1} \cdot \frac{1}{1} \cdot \frac{1}{1} \cdot \frac{1}{20} = 0$

Tahap 4 Bandingkan hasil class TEPAT & TERLAMBAT

- Karena hasil $(P|Tepat)$ Lebih besar dari $(P|Terlambat)$ maka keputusannya adalah " Tepat"
- Jadi kesimpulannya mahasiswa yang bernama Anggi Fratama lulus dengan TEPAT WAKTU.

Contoh satu kasus lainnya yaitu mahasiswa yang bernama Rejon Saputra Diono.

Tahap 1 menghitung jumlah class/label

- $P(Y = TEPAT) = \frac{19}{20}$ = Jumlah data “Tepat” pada kolom ‘Status Kelulusan’ dibagi jumlah data

- $P(Y = Terlambat) = \frac{1}{20}$ Jumlah data “Terlambat” pada kolom ‘Status Kelulusan’ dibagi jumlah data

Tahap 2 menghitung jumlah kasus yang sama dengan class yang sama

- $P(IPK = 2.75|Y = Tepat) Y \frac{1}{19}$ "jumlah data IPK “3.48” dengan keterangan “TEPAT” dibagi jumlah data TEPAT”
- $P(IPK = 2.75|Y = Terlambat) Y \frac{0}{19}$ "jumlah data IPK “3.48” dengan keterangan “Terlambat” dibagi jumlah data Terlambat ”
- $P(Jalur Masuk = Pagi|Y = Tepat) Y \frac{19}{19}$ "jumlah data jalur masuk “Pagi” dengan keterangan “Tepat” dibagi jumlah data Tepat”
- $P(Jalur Masuk = Pagi|Y = Terlambat) Y \frac{1}{1}$ "jumlah data jalur masuk “Pagi” dengan keterangan “Terlambat” dibagi jumlah data Terlambat”
- $P(Jenis kelamin = Laki - laki|Y = Tepat) Y \frac{1}{1}$ "jumlah data jenis kelamin “Laki-lakii” dengan keterangan “Tepat” dibagi jumlah data Tepat”
- $P(Jenis kelamin = Laki - laki|Y = Terlambat) Y \frac{1}{1}$ "jumlah data jenis kelamin “Laki-lakii” dengan keterangan “Terlambat” dibagi jumlah data Terlambat”

Tahap 3 kalikan semua hasil variable TEPAT & TERLAMBAT

- $P(IPK = 2.75). (Jalur masuk = Pagi). (Kelamin = Laki - laki)|Tepat) = \{P(P(IPK = 2.75|Y = Tepat). P(Jalur masuk = Pagi|Y = Tepat). P(Jenis Kelamin = Laki - laki|Y = Tepat) \} = \frac{0}{19} \cdot \frac{19}{19} \cdot \frac{18}{19} \cdot \frac{19}{20} = 0$
- $P(IPK = 2.75). (Jalur masuk = Pagi). (Kelamin = Laki - laki)|Terlambat) = \{P(P(IPK = 2.75|Y = Terlambat). P(Jalur masuk = Pagi|Y = Terlambat). P(Jenis Kelamin = Laki - laki|Y = Terlambat) = \frac{1}{1} \cdot \frac{1}{1} \cdot \frac{1}{1} \cdot \frac{1}{20} = 0.05$

Tahap 4 Bandingkan hasil class TEPAT & TERLAMBAT

- Karena hasil $(P|Tepat)$ Lebih kecil dari $(P|Terlambat)$ maka keputusannya adalah " Terlambat"

Kesimpulanya mahasiswa yang bernama Rejon Saputra Diono LULUS DENGAN TERLAMBAT.

Dikarenakan *Naïve Bayes* mengklasifikasikan bahwa mahasiswa yang bernama Rejon Saputra Diono lulus terlambat karena faktor NPM lebih tua dan Nilai IPK yang cukup kecil. Dari contoh *implementasi* sistem pengujian tersebut dapat dilihat bahwa dengan menggunakan algoritma *naïve bayes* prediksi kelulusan banyak mahasiswa dapat dengan mudah diketahui oleh system yang menggunakan algoritma *Naïve Bayes* dengan menggunakan atribut NPM, lama studi, dan IPK.

3.3. Pengujian

Sebelum sistem ini digunakan atau diimplementasikan maka perlu dilakukan pengujian. Pengujian ini bermaksud untuk meminimalkan jumlah kesalahan atau kegagalan sistem pada saat implementasi sistem. Metode pengujian sistem yang digunakan adalah metode *blackbox testing*. *Blackbox testing* merupakan metode pengujian perangkat lunak yang menguji fungsionalitas atau *nonfungsionalitas* sistem tanpa mengetahui struktur internal modul atau program yang akan diuji. Dalam pengujian sistem ini data uji yang dipakai sebanyak 20 orang nama mahasiswa lulusan September 2019 pada Prodi Teknik Informatika Universitas Muhammadiyah Bengkulu. 20 orang data tersebut merupakan data uji sebagai contoh untuk diterapkan ke dalam sistem serta dengan menggunakan ke-20 orang data tersebut sebagai data *testing* untuk melihat hasil prediksi terlambat atau tepat waktu kelulusan. Berikut Pengujian sistem dengan *Blackbox testing* :

Tabel 1. Pengujian Black box

Skenario Pengujian	Test	Hasil Pengujian	Kesimpulan
Login	Input password	Bisa diakses	valid
Tambah Data	Proses tambah data	Bisa diakses	valid
Submit	Submit	Bisa diakses	valid
Keluar	keluar	Bisa diakses	Valid
Hasil naive Bayes	Perhitungan Naive Bayes	Bisa diakses	Valid
Entri Data	click	Bisa diakses	Valid
Previous	click	Bisa diakses	Valid
Next	click	Bisa diakses	Valid

Pengujian data dengan *Precision*, *Recall*, dan akurasi

Dalam pengujian ini data uji yang diinput sebanyak 20 orang nama mahasiswa lulusan September 2019. Dari ke-20 data uji tersebut dilakukan dengan 20 data testing dan yang mendapatkan 18 orang termasuk kategori True Positif (TP), dan 2 orang termasuk dalam False Positif (FP) yang dapat dilihat pada tabel berikut :

Tabel 2. Tabel Pengujian Precision, Recall, dan akurasi

Nilai Prediksi	Nilai Sebenarnya	
	TRUE	FALSE
TRUE	TP (True Positif) Hasil Yang Benar/Relevan	FP (False Positif) Hasil Yang Tidak Diharapkan/Tidak Relevan
FALSE	FN (False Negatif) Hasil yang Hilang/Negatif Ditemukan	TN (True Negatif) Tidak Adanya Hasil Yang Benar/Tidak Ditemukan

Tabel 5. Tabel Hasil Pengujian

Data Testing	TP	FP	FN	TN
20	18	2	0	0

$$Precision = \frac{18}{18+2} \times 100\%$$

$$= \frac{18}{20} \times 100\% = \frac{9}{10} \times 100\% = 0,9 = 90\%$$

$$Recall = \frac{18}{18+0} \times 100\%$$

$$= \frac{18}{18} \times 100\% = \frac{1}{1} \times 100\% = 1 = 100\%$$

$$Accuracy = \frac{18+0}{18+0+2+0} \times 100\%$$

$$= \frac{18}{20} \times 100\% = \frac{9}{10} \times 100\% = 0,9 = 90\%$$

Dari data pengujian diatas menggunakan 20 data mahasiswa sebagai data uji didapatkan nilai persentase keberhasilan dari *precision* sebesar 90 %, *recall* 100 % dan akurasi sebesar 90 %. Dari nilai persentase tersebut dapat disimpulkan bahwa dengan menerapkan metode *Naïve Bayes* dalam prediksi kelulusan mahasiswa dikategorikan berhasil dengan nilai persentase akurasi mencapai 90 %.

4. KESIMPULAN

Dengan menerapkan algoritma *Naïve Bayes* dapat memprediksi tingkat kelulusan mahasiswa tepat waktu atau tidak yang dapat berguna dalam memberikan informasi dan masukan bagi pihak perguruan tinggi dalam membuat kebijakan kedepannya. Serta Tingkat keberhasilan prediksi algoritma *Naïve Bayes* yang ditentukan oleh atribut nama, NPM, dan nilai IPK menghasilkan nilai persentase *precession* sebesar 90 %, *Recal* 100 % dan akurasi sebesar 90 %.

UCAPAN TERIMA KASIH

Pada kesempatan kali ini ,penulis menyampaikan ucapan terima kasih kepada seluruh rekan-rekan yang telah membantu penulis dalam menyelesaikan penelitian ini. Terima kasih kepada Ka. Prodi yang telah memberikan izin dalam menyelesaikan penelitian ini. Ka. Lab. Fakultas Teknik yang telah memberikan izin dalam menggunakan komputer di laboratorium serta rekan rekan sejawat terima kasih atas saran-sarannya.

REFERENSI

- [1] Y. Kusnia, "Pengaruh karakteristik gender dan motivasi belajar terhadap prestasi belajar matematika siswa kelas X IPA di MAN 2 Semarang," *Semin. Nas. Pendidikan, Sains dan Teknol. Fak. Mat. dan Ilmu Pengetah. Alam Univ. Muhammadiyah Semarang*, pp. 398–405, 2017.
- [2] D. Kartini, "Penerapan Data Mining dengan Algoritma Neural Network (Backpropagation) Untuk Prediksi Lama Studi Mahasiswa," in *PROSIDING seminar nasional sisfotek*, 2017, vol. 3584, pp. 235–241.
- [3] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no. no. juni, pp. 10–20, 2013.
- [4] M. A. Nurrohmat, "Aplikasi Pemrediksi Masa Studi dan Predikat Kelulusan Mahasiswa Informatika Universitas Muhammadiyah Surakarta Menggunakan Metode Naive Bayes," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 1, no. 1, p. 29, 2015, doi: 10.23917/khif.v1i1.1179.
- [5] M. F. Rifai, H. Jatnika, and B. Valentino, "Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *Petir*, vol. 12, no. 2, pp. 131–144, 2019, doi: 10.33322/petir.v12i2.471.
- [6] S. Syarli and A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2016.
- [7] Y. S. Nugroho and Setyawan, "Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi dan Informatika," *KomuniTi*, vol. VI, no. 1, pp. 84–91, 2014.
- [8] S. S. Nikam, "A Comparative Study of Classification Techniques in Data Mining Algorithms," *Int. J. Mod. Trends Eng. Res.*, vol. 4, no. 7, pp. 58–63, 2017, doi: 10.21884/ijmter.2017.4211.vxayk.
- [9] H. Bhavsar and A. Ganatra, "A Comparative Study of Training Algorithms for Supervised Machine Learning," *Int. J. Soft Comput. Eng.*, vol. 2, no. 4, pp. 74–81, 2012, doi: 10.1.1.492.6088.
- [10] S. D. Jadhav and H. P. Channe, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," *Int. J. Sci. Res.*, vol. 5, no. 1, pp. 1842–1845, 2016, doi: 10.21275/v5i1.nov153131.
- [11] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *J. Inform.*, vol. 8, no. 1, pp. 884–898, 2014, doi: 10.26555/jifo.v8i1.a2086.
- [12] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 699–714, 2005, doi: 10.1109/TPAMI.2005.93.
- [13] R. P. S. Putri and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, p. 1, 2018, doi: 10.23917/khif.v4i1.5975.
- [14] G. Testiana, "Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah," *JUSIFO (Jurnal Sist. Informasi)*, vol. 4, no. 1, pp. 49–62, 2018.
- [15] M. Utami and Y. Apridiansyah, "Implementasi Algoritma Sequential Searching Pada Sistem Pelayanan Puskesmas Menggunakan Bootstrap (Studi Kasus Puskesmas Kampung Bali Bengkulu),"

- [16] *JSAI (Journal Sci. Appl. Informatics)*, vol. 2, no. 1, pp. 81–86, 2019, doi: 10.36085/jsai.v2i1.166.
M. Martin and L. Nilawati, “Recall dan Precision Pada Sistem Temu Kembali Informasi Online Public Access Catalogue (OPAC) di Perpustakaan,” *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 77–84, 2019, doi: 10.31294/p.v21i1.5064.