

## Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Naive Bayes* di Program Studi Teknik Informatika UHAMKA

Dwi Anugrah Putra<sup>1)</sup> & Mia Kamayani<sup>2)</sup>

<sup>1,2)</sup>Program Studi Teknik Informatika

Universitas Muhammadiyah Prof. Dr. Hamka

Jl. Tanah Merdeka No.6 Kp. Rambutan, Ps. Rebo, Jakarta Timur

[dwianugrah@gmail.com](mailto:dwianugrah@gmail.com)

### **Abstrak**

Berdasarkan pengamatan dan data yang ada pada Program Studi Teknik Informatika FT UHAMKA bahwa banyaknya mahasiswa yang tidak lulus tepat waktu (8 semester) pada setiap angkataannya akan menyebabkan penumpukan jumlah mahasiswa, kurangnya ruang kelas, dan kurangnya lahan parkir. Salah satu cara untuk meningkatkan kelulusan mahasiswa tepat waktu adalah memprediksi dari awal mahasiswa yang berpotensi untuk lulus tidak tepat waktu, sehingga dapat dilakukan tindakan pencegahan oleh pihak manajemen prodi atau fakultas. Prediksi dapat dilakukan menggunakan data mining dengan memanfaatkan data mahasiswa yang sudah lulus. Metode data mining yang digunakan pada penelitian ini adalah naive bayes dengan menggunakan atribut jenis kelamin, index prestasi semester satu sampai semester empat dan SKS semester satu sampai semester empat. Algoritma naive bayes ini akan dibuat beberapa model dan akan dicari nilai akurasi tertinggi dari model tersebut. Evaluasi model menggunakan K-fold Cross Validation dan hasil prediksi akan digunakan oleh dosen pembimbing akademik untuk mengevaluasi mahasiswa yang hasil prediksinya kurang memuaskan. Model dengan hasil terbaik yaitu model ke-3 dengan tingkat akurasi sebesar 80.19%, recall 80.26%, precision 92.75% dan F-Measure 86.05% yang nantinya akan digunakan untuk implementasi pada aplikasi prediksi kelulusan mahasiswa.

**Kata Kunci:** Naive Bayes, prediksi kelulusan, k-Fold Cross-Validation

### **Abstract**

Based on observations and existing data in the UHAMKA Informatics Engineering Study Program, the number of students who do not graduate on time (8 semesters) in each generation will cause an accumulation of the number of students, lack of classrooms, and lack of parking space. One of the ways to increase student graduation on time is to predict from begin which students have the potential to didn't graduate on time, so that preventive action can be taken by the study program management or faculty. Prediction can be done using data mining by utilizing data from students who have graduated. The data mining method used in this study is Naive Bayes using gender attributes, achievement index from semester one to semester four and semester one to semester four credits. The Naive Bayes algorithm will be made several models and the highest accuracy value will be sought from the model. The model evaluation uses K-fold Cross Validation and the prediction results will be used by the academic supervisor to evaluate students whose prediction results are unsatisfactory. The model with the best results is the 3rd model with an accuracy rate of 80.19%, a recall of 80.26%, precision 92.75% and F-Measure 86.05% which will be used for implementation in the student graduation prediction application.

**Keywords:** Naive Bayes, graduation prediction, k-Fold Cross-Validation

## 1 PENDAHULUAN

Di era digitalisasi sekarang ini sudah banyak instansi dan perusahaan yang menyimpan data mereka di dalam sebuah basis data yang terkomputerisasi. Dunia pendidikan pun tidak terlepas dari perkembangan teknologi ini. Universitas Prof. Dr. Hamka termasuk salah satu perguruan tinggi yang telah menyimpan datanya dalam basis data yang terkomputerisasi. Data tersebut merupakan data mahasiswa, data dosen, serta berbagai data lain yang berhubungan dengan Universitas Prof. Dr. Hamka.

Data tersebut tidak banyak mempunyai kegunaan dan seakan-akan menjadi kumpulan data

terabaikan yang semakin bertambah besar pada tiap tahunnya. Data tersebut dipergunakan ketika universitas memerlukan satu informasi tertentu atau ketika proses akreditasi. Ketika mahasiswa sudah lulus maka data mereka kurang bermanfaat dan jarang dipergunakan. Sebenarnya data tentang mahasiswa yang sudah lulus ialah data yang penting dan dapat dipergunakan dalam proses akreditasi. Data tentang mahasiswa yang lulus dapat menghasilkan informasi yang bermanfaat bagi universitas jika dipergunakan dengan baik.

Berdasarkan hasil pengamatan di Universitas Prof. Dr. Hamka, ditemukan sebaran yang tak seimbang antara jumlah mahasiswa baru yang

masuk dan keluar yang telah menyelesaikan masa studi. Mahasiswa baru masuk dalam jumlah besar, sedangkan jumlah mahasiswa yang lulus tepat waktu jauh lebih kecil. Berdasarkan data yang didapat bahwa tingkat kelulusan tepat waktu pada angkatan 2013 adalah 22.36% dengan jumlah mahasiswa aktif pada saat itu adalah 76 orang dan yang lulus tepat waktu adalah 17 orang. Pada angkatan tahun 2014 tingkat kelulusan tepat waktu adalah 11.23% dengan mahasiswa aktif saat itu adalah 89 orang dan mahasiswa yang lulus tepat waktu 10 orang. Lalu pada angkatan 2015 tingkat kelulusan tepat waktu hanya 7.69% dengan mahasiswa aktif pada saat itu adalah 104 orang dan yang lulus tepat waktu hanya 8 orang. Sedangkan target kelulusan mahasiswa tepat waktu pada prodi informatika yaitu sebesar 50%. Tentunya dalam jangka panjang, hal tersebut dapat menyebabkan terjadinya penumpukan jumlah mahasiswa serta kurangnya ruang kelas yang tersedia di Universitas Prof. Dr. Hamka.

Jumlah mahasiswa lulus tepat waktu dapat ditingkatkan dengan meningkatkan kualitas pembelajaran. Selain itu, cara untuk mencari informasi yang tersembunyi tentang kelulusan mahasiswa adalah menggunakan *data mining* dengan memanfaatkan data mahasiswa yang telah lulus. Penelitian terdahulu dalam memanfaatkan data mining yang dilakukan oleh Supardi Salmu dan Achmad Solihin, Algoritma yang digunakan adalah *Naive bayes*. Lalu atribut yang digunakan adalah Jenis kelamin, Jenis seleksi, Pendapatan ayah, Pendapatan ibu, IP (Index Prestasi) semester 1 sampai 4 dan SKS semester 1 dan 4. Akurasi dari pengujian pada penelitian tersebut adalah 80.72% [3].

Oleh karena itu penulis memutuskan untuk menggunakan algoritma *Naive bayes* untuk penelitian ini. Atribut yang digunakan untuk penelitian ini adalah Jenis kelamin, IP semester 1 sampai 4 dan SKS semester 1 sampai 4, hal ini mengacu pada keadaan data yang ada pada Fakultas Teknik Informatika Universitas Prof. Dr. Hamka.

Pada penelitian ini penulis merancang sebuah aplikasi *data mining* dengan menggunakan algoritma *Naive bayes* untuk memprediksi kelulusan mahasiswa yang dapat digunakan oleh prodi. Dengan mengetahui prediksi status kelulusan mahasiswa selama berjalannya perkuliahan, diharapkan prodi di bantu pembimbing akademik dapat memberi perhatian khusus terhadap mahasiswa yang di prediksi tidak lulus tepat waktu sehingga mahasiswa tersebut dapat memperbaiki

indeks prestasinya pada tiap semester agar dapat lulus tepat waktu. Penelitian ini akan dilaksanakan di program studi Informatika yang memiliki sasaran untuk menjadi program studi unggulan. Oleh karena itu diharapkan hasil prediksi kelulusan mahasiswa dapat membantu program studi Informatika dalam mengambil langkah strategis dan mengambil keputusan.

## 2 DASAR TEORI

### 2.1 Data Mining

*Data mining* yaitu proses pencarian informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang ilmu statistik, *machine learning*, dan sistem manajemen basis data [2]. *Data mining* digunakan untuk ekstraksi informasi penting yang tersembunyi dari *dataset* yang besar. Dengan adanya *data mining* maka akan didapatkan suatu informasi berupa pengetahuan di dalam kumpulan data – data yang banyak jumlahnya.

### 2.2 Naive Bayes

*Naive bayes* yaitu salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk *machine learning* dan *data mining*. Performa naive bayes yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keindependenan atribut (tidak ada kaitan antar atribut). Asumsi keindependenan atribut ini pada data sebenarnya jarang terjadi, namun walaupun asumsi keindependenan atribut tersebut dilanggar performa pengklasifikasian naive bayes cukup tinggi, hal ini dibuktikan pada berbagai penelitian empiris [4]. Performansi klasifikasi *naive bayes* akan lebih baik ketika atribut numerik didiskritisasi daripada diasumsikan dengan pendekatan distribusi. Nilai-nilai numerik akan dipetakan ke nilai nominal dalam bentuk interval yang tetap memperhatikan kelas dari tiap-tiap nilai numerik yang dipetakan, penggambaran perhitungan Naive Bayesnya seperti berikut [4]:

$$p(I = ij|C = ci) = \frac{p(I=ij)p(C = ci|I = ij)}{p(C=ci)} \quad (1)$$

Apabila data bersifat kuantitatif (numerik) maka menggunakan pendekatan distribusi normal (gaussian) seperti berikut :

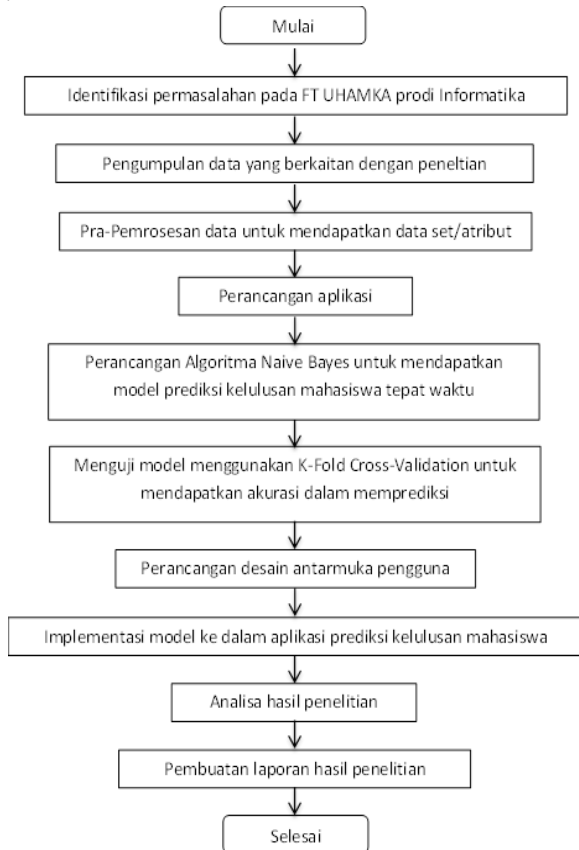
$$f = n(X_i; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1 - \mu\sigma)^2}{2\sigma^2}} \quad (2)$$

### 2.3 K-Fold Cross Validation

*Cross-validation* atau dapat disebut sebagai estimasi rotasi yaitu suatu teknik validasi model untuk mengetahui bagaimana hasil statistik analisis akan menggeneralisasi sekumpulan data independen. Teknik ini utamanya digunakan untuk menjalankan prediksi model dan memperhitungkan seberapa akurat sebuah model prediktif ketika dijalankan dalam praktiknya[7].

### 3 METODOLOGI PENELITIAN

Langkah penelitian dapat dilihat pada Gambar 1.



Gambar 1 Flowchart Alur Metodologi Penelitian

#### 3.1 Identifikasi Permasalahan

Pada tahap ini penulis mengidentifikasi hal-hal yang harus diperhatikan dalam pemilihan atribut yang signifikan dalam dataset, pengumpulan dataset, analisis outlier dari dataset dan analisis kebutuhan untuk implementasi algoritma.

#### 3.2 Pengumpulan Data

Pengumpulan data ini bertujuan untuk memudahkan dan menguatkan selama pelaksanaan penelitian berlangsung. Terdapat 2 tahap dalam pengumpulan data, yaitu :

a. Observasi

Pada tahapan observasi penulis melakukan pengamatan di fakultas teknik UHAMKA prodi informatika serta terlibat secara langsung dengan objek penelitian serta wawancara kepada pihak yang berkaitan dengan penelitian ini.

b. Studi Pustaka

Studi pustaka berguna untuk pengumpulan data berupa penelitian yang terdahulu, pembelajaran dari berbagai macam sumber dan dokumen seperti *e-book*, *journal* dan teori-teori yang mendukung penelitian seperti *data mining*, *naive bayes*, *k-Fold Cross-Validation*, *tools* yang akan digunakan seperti *Rapidminer* dan data pendukung lainnya yang berhubungan dengan penelitian ini.

c. Pengumpulan data mahasiswa

Pada tahap ini peneliti mengumpulkan data daftar indeks prestasi mahasiswa prodi informatika UHAMKA mulai dari semester 1 sampai dengan semester 8 dan mulai dari angkatan 2009 sampai dengan angkatan 2015.

Peneliti juga mengumpulkan data daftar kelulusan mahasiswa dan masa studi mulai dari tahun 2012 sampai tahun 2019. Data daftar indeks prestasi mahasiswa dan data daftar kelulusan mahasiswa ini didapat dari sekretariat untuk nantinya dipelajari agar dapat dilakukan pra-pemrosesan data.

#### 3.3 Pra-Pemrosesan Data

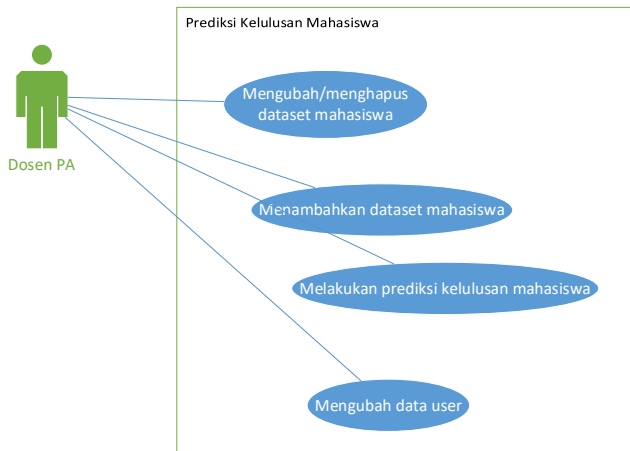
Pada tahap ini penulis melakukan pra-pemrosesan data, yaitu merekap data mahasiswa prodi informatika dengan memastikan bahwa data mahasiswa yang dijadikan dataset adalah data mahasiswa yang mengikuti perkuliahan sampai selesai/lulus (tidak mengambil cuti dan keluar/pindah). Dalam hal ini cara yang digunakan penulis yaitu melihat satu persatu apakah mahasiswa tersebut selalu termasuk dalam daftar indeks prestasi mahasiswa pada tiap semester (semester 1 s/d semester 8). Dataset mahasiswa prodi informatika ini berbentuk file *excel* yang berisikan NIM, nama, jenis kelamin, IP semester 1 sampai semester 4, SKS semester 1 sampai semester 4, dan keterangan (lulus tepat waktu/tidak).

#### 3.4 Perancangan Aplikasi

Proses desain akan mengartikan syarat kebutuhan sebuah perancangan perangkat lunak yang dapat diperhitungkan sebelum lanjut pada tahap implementasi.

Perancangan mencakup perancangan fungsional, non-fungsional dan perancangan

database. Perancangan fungsional menggunakan diagram UML melingkupi use case diagram. (lihat Gambar 2).

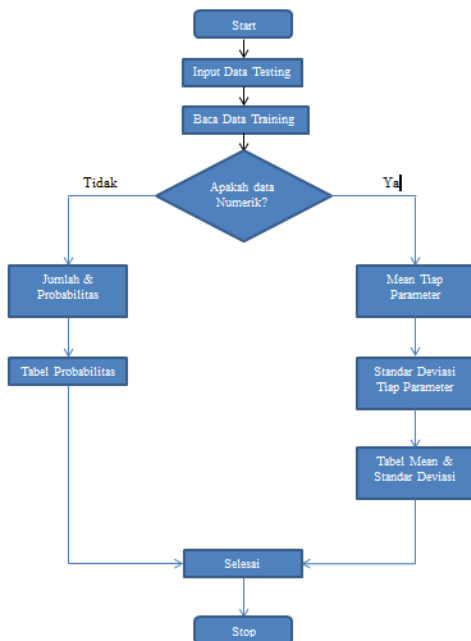


Gambar 2 Use case diagram

### 3.5 Implementasi Algoritma Naive Bayes

Setelah dataset yang diperlukan sudah lengkap, langkah selanjutnya yaitu pengujian model algoritma *naive bayes* menggunakan *tools Rapidminer* dan nanti hasilnya dijadikan rujukan oleh penulis pada penghitungan manual yang dimasukkan ke dalam coding program. Coding program berbasis website dengan menggabungkan PHP, JavaScript, Css, dll.

Adapun flowchart algoritma naive bayes pada gambar 3 berikut.



Gambar 3 Flowchart algoritma naive bayes

### 3.6 Pengujian Model Menggunakan *k-Fold Cross Validation*

Sebelum pembuatan aplikasi, pengujian pada *naive bayes* diperlukan untuk mencari model yang memiliki akurasi terbaik untuk nantinya diimplementasikan ke dalam aplikasi. Pengujian pada *naive bayes* menggunakan *tools rapidminer* dan di evaluasi menggunakan *k-fold Cross Validation*. Pada tahapan ini, algoritma *naive bayes* yang telah dihitung dan didapatkan hasilnya akan dilihat seberapa besar akurasi yang didapat dari model tersebut menggunakan *Rapidminer*.

### 3.7 Perancangan Desain Antarmuka Pengguna

Pada tahap ini desain antarmuka akan dibuat simple agar nantinya user mudah dalam pengoperasiannya. Pada aplikasi ini terdapat beberapa menu antara lain, yaitu : menu mengubah data user, menu mengubah dataset mahasiswa, menu menambahkan dataset mahasiswa, dan menu memprediksi kelulusan mahasiswa.

### 3.8 Implementasi Model ke Dalam Aplikasi

Setelah dilakukan pengujian pada model, pada tahap ini akan dilakukan implementasi dengan menggabungkan model dan coding sehingga terbentuk suatu aplikasi yang disebut sebagai aplikasi prediksi kelulusan mahasiswa, diharapkan aplikasi tersebut dapat digunakan oleh perguruan tinggi untuk memprediksi kelulusan mahasiswa.

### 3.9 Analisa Hasil Penelitian

Pada tahap ini akan berisi uraian tentang uji fungsionalitas dan uji efektivitas pada aplikasi prediksi kelulusan mahasiswa dari segi kelebihan maupun kekurangan yang dapat dilihat pada bab pembahasan, dan nantinya dapat digunakan sebagai saran untuk penelitian selanjutnya.

## 4 TEMUAN DAN PEMBAHASAN

### 4.1 Dataset

Data yang digunakan adalah data mahasiswa prodi informatika jenjang strata 1 yang bersumber dari sekretariat FT UHAMKA, mulai dari angkatan 2009 sampai dengan angkatan 2015 yang berjumlah 546 data mahasiswa. Atribut yang digunakan adalah jenis kelamin, IP semester 1 sampai semester 4, SKS semester 1 sampai semester 4, dan keterangan (tepat waktu/tidak). (lihat Tabel 1)

Tabel 1 Contoh dataset mahasiswa

JE	IPS	nSKS	T
----	-----	------	---

N K E L									E P A T?
	1	2	3	4	1	2	3	4	
L	3.21	3.00	2.91	2.62	19	24	23	21	Y
P	2.68	2.60	2.75	2.30	19	20	20	20	N
L	3.58	3.08	3.38	2.88	19	24	24	24	Y
L	3.84	3.83	3.67	3.38	19	24	24	24	Y
L	2.68	2.60	2.24	2.44	19	20	21	18	N
L	2.47	2.44	2.65	3.00	19	18	17	21	N
L	2.53	2.62	1.81	2.53	19	21	21	15	N
L	3.42	3.17	2.92	2.33	19	24	24	18	N
L	2.37	2.83	2.48	3.06	19	18	21	18	N
L	3.68	3.25	2.96	2.43	19	24	23	21	N
L	2.63	3.24	3.13	2.46	19	21	24	24	N
P	3.53	3.46	3.50	3.33	19	24	22	24	N
L	3.21	3.00	3.04	2.67	19	24	24	24	Y
L	2.74	1.95	1.80	2.21	19	21	15	14	N
L	2.11	1.83	1.14	1.00	19	18	14	12	N
L	3.11	2.77	2.75	3.00	19	22	20	20	Y
L	2.42	1.72	2.33	2.41	19	18	15	17	N
L	2.58	3.00	3.38	2.88	19	21	24	24	Y
L	2.42	2.44	2.29	2.28	19	18	17	18	N
L	2.63	3.00	2.13	1.35	19	20	23	17	N

### 4.2 Pengujian Model

Terdapat beberapa model yang diuji pada penelitian ini, antara lain :

1. Model pertama, menggunakan keseluruhan atribut yang tersedia yaitu jenis kelamin, index prestasi(IP) semester dan satuan kredit semester(SKS).
2. Model kedua, menggunakan atribut IP semester dan SKS serta menghilangkan atribut jenis kelamin.
3. Model ketiga, menggunakan atribut jenis kelamin dan IP semester serta menghilangkan atribut SKS

### 4.3 Evaluasi Menggunakan Cross Validation

Pada tahap evaluasi *cross validation* ini jumlah *fold* yang digunakan adalah  $k=10$ . Akurasi yang didapat dari ketiga model yang diuji dapat dilihat pada Tabel 2.

Tabel 2 Hasil Pengujian Ketiga Model Menggunakan Cross Validation

Model	Hasil Uji Coba			
	Accuracy	Precision	Recall	F-

				Measure
1	79.84% (±5.45%)	94.73% (±3.11%)	77.86% (±6.05%)	85.47%
2	78.74% (±5.57%)	93.78% (±2.28%)	77.14% (±6.36%)	84.65%
3	80.19% (±5.83%)	92.75% (±3.33%)	80.26% (±6.09%)	86.05%

Berdasarkan Tabel 2 dapat diketahui bahwa model yang optimal adalah model ketiga dengan tingkat akurasi 80.19% yang terdiri dari atribut jenis kelamin dan index prestasi semester 1 sampai semester 4.

### 4.4 Contoh Penyelesaian Kasus

Menyelesaikan masalah dengan contoh kasus salah satu mahasiswa dengan jenis kelamin laki-laki (JENKEL = L), nilai IP semester 1  $IPS_1 = 3.26$ , IP semester 2,  $IPS_2 = 3.04$ , IP semester 3,  $IPS_3 = 3.05$ , dan IP semester 4,  $IPS_4 = 3.32$ , ingin diketahui apakah dia diprediksi lulus tepat waktu (Y=TEPAT) atau terlambat (Y=TERLAMBAT).

1. Menghitung jumlah class/label:

$$P(Y = TEPAT) = \frac{n_{Y=TEPAT}}{n} = \frac{130}{546} = 0.238$$

$$P(Y = TERLAMBAT) = \frac{n_{Y=TERLAMBAT}}{n} = \frac{416}{546} = 0.762$$

2. Menghitung jumlah kasus yang sama dengan class sama:

$$P(\text{JENKEL} = L | Y = \text{TEPAT}) = \frac{n_{\text{JENKEL=L} \cap n_{Y=TEPAT}}}{n_{Y=TEPAT}} = \frac{77}{130} = 0.592$$

$$P(\text{JENKEL} = L | Y = \text{TERLAMBAT}) = \frac{n_{\text{JENKEL=L} \cap n_{Y=TERLAMBAT}}}{n_{Y=TERLAMBAT}} = \frac{357}{416} = 0.858$$

3. Menghitung data kontinyu  $IPS_1$ :

$$P(IPS_1 = 3.26 | Y = \text{TEPAT})$$

untuk  $X = IPS_1 = 3.26$  dan  $Y = \text{TEPAT}$

$$\bar{X} = 3.131$$

$$\sigma = 0.388$$

$$P(X) = 0.606$$

$$P(IPS_1 = 3.26 | Y = \text{TERLAMBAT})$$

untuk  $X = IPS_1 = 3.26$  dan  $Y = \text{TERLAMBAT}$

$$\bar{X} = 2.569$$

$$\sigma = 0.491$$

$$P(X) = 0.212$$

4. Menghitung data kontinyu  $IPS_2$ :

$$P(IPS_2 = 3.04 | Y = TEPAT)$$

untuk  $X = IPS_2 = 3.04$  dan  $Y = TEPAT$

$$\bar{X} = 3.103$$

$$\sigma = 0.402$$

$$P(X) = 0.622$$

$$P(IPS_2 = 3.04 | Y = TERLAMBAT)$$

untuk  $X = IPS_2 = 3.04$  dan  $Y =$   
 $TERLAMBAT$

$$\bar{X} = 2.456$$

$$\sigma = 0.602$$

$$P(X) = 0.321$$

5. Menghitung data kontinyu  $IPS_3$ :

$$P(IPS_3 = 3.05 | Y = TEPAT)$$

untuk  $X = IPS_3 = 3.05$  dan  $Y = TEPAT$

$$\bar{X} = 3.214$$

$$\sigma = 0.376$$

$$P(X) = 0.592$$

$$P(IPS_3 = 3.05 | Y = TERLAMBAT)$$

untuk  $X = IPS_3 = 3.05$  dan  $Y =$   
 $TERLAMBAT$

$$\bar{X} = 2.433$$

$$\sigma = 0.571$$

$$P(X) = 0.294$$

6. Menghitung data kontinyu  $IPS_4$ :

$$P(IPS_4 = 3.32 | Y = TEPAT)$$

untuk  $X = IPS_4 = 3.32$  dan  $Y = TEPAT$

$$\bar{X} = 3.077$$

$$\sigma = 0.392$$

$$P(X) = 0.526$$

$$P(IPS_4 = 3.32 | Y = TERLAMBAT)$$

untuk  $X = IPS_4 = 3.32$  dan  $Y =$   
 $TERLAMBAT$

$$\bar{X} = 2.339$$

$$\sigma = 0.660$$

$$P(X) = 0.163$$

7. Membandingkan nilai P

$$P_{TEPAT} = 0.0165$$

$$P_{TERLAMBAT} = 0.002$$

Karena  $P_{TEPAT} > P_{TERLAMBAT}$ , maka keputusannya adalah "TEPAT WAKTU".

#### 4.5 Tampilan Uji Coba Menggunakan Aplikasi

Halaman antarmuka aplikasi prediksi terdapat pada Gambar 4 dan Gambar 5. Pada Gambar 4 terdapat menu data *test* untuk memprediksi

kelulusan mahasiswa dengan memasukkan form pengisian yang meliputi nama, nim, jenis kelamin, dan IP semester 1 sampai dengan semester 5. Pada Gambar 5 terdapat hasil prediksi.

Data Test Per Individu

Gambar 4 Tampilan Form Input Data Test

NIM	Nama	Jenis Kelamin	IPS 1	IPS 2	IPS 3	IPS 4	Keterangan Lulus
1503015043	Deli Anugrah Putra	L	3.26	3.04	3.05	3.32	TEPAT WAKTU

Gambar 5 Tampilan Hasil Prediksi

## 5 SIMPULAN

Penelitian ini terdiri dari 546 data mahasiswa yang terdiri dari jenis kelamin dan index prestasi semester 1 sampai semester 4. Pada penelitian ini terdapat tiga model dan satu model yang terbaik adalah model 3 yang terdiri dari atribut jenis kelamin dan index prestasi semester 1 sampai semester 4, karena memiliki akurasi 80.19%, *precision* 92.75%, *recall* 80.26%, dan *F-Measure* 86.05%.

Perbanyak dataset mahasiswa yang dibuat agar hasil prediksi lebih akurat dan menambahkan atribut yang berkaitan dengan prediksi kelulusan mahasiswa, seperti alamat, lulusan sekolah, dan sebagainya. Mengubah target kelulusan mahasiswa tepat waktu menjadi 4.5 tahun (9 semester), karena mengacu pada standar kelulusan yang ada pada prodi informatika FT UHAMKA.

## KEPUSTAKAAN

- [1] E. P. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network," 2013.
- [2] R. Yanto and R. Khoiriah, "Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat," *Creat. Inf. Technol. J.*, vol. 2, no. 2, pp. 102–113, 2015.
- [3] S. Salmu and A. Solichin, "Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta Prediction

- of Timeliness Graduation of Students Using Naïve Bayes : A Case Study at Islamic State University Syarif Hidayatullah Jakarta,” no. April, pp. 701–709, 2017.
- [4] S. Syarli and A. A. Muin, “Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi),” *J. Ilmu Komput.*, vol. 2, no. 1, Sep. 2018.
- [5] H. Widayu, S. D. Nasution, N. Silalahi, and M. Mesran, “DATA MINING UNTUK MEMREDIKSI JENIS TRANSAKSI NASABAH PADA KOPERASI SIMPAN PINJAM DENGAN ALGORITMA C4.5,” *J. MEDIA Inform. BUDIDARMA*, vol. 1, no. 2, Jun. 2017.
- [6] A. Y. Saputra and Y. Primadasa, “Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor,” *Techno.Com*, vol. 17, no. 4, pp. 395–403, 2019.
- [7] F. Tempola, M. Muhammad, and A. Khairan, “Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation Comparison of Classification Between Knn and Naive Bayes At the Determination of the Volcanic Status With K-Fold Cross,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 577–584, 2018.
- [8] J. Junaedy and A. Munir, “Rancang Bangun Sistem Pengelolaan Data Kuliah Kerja Lapang Plus Memanfaatkan Framework Codeigniter Dengan Menggunakan Metode Waterfall,” *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 203–210, 2017.
- [9] D. Dahri, F. Agus, and D. M. Khairina, “Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman,” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 11, no. 2, p. 29, 2016.
- [10] A. R. C and Y. Lukito, “Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes,” *J. Ultim.*, vol. 9, no. 1, pp. 50–58, Jun. 2017.
- [11] G. A. Buntoro, “Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine,” *J. Din. Inform.*, vol. 3, no. 1, p. 56, 2016.
- [12] A. Muzakir and R. A. Wulandari, “Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree,” *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, Jun. 2016.
- [13] J. Pardede, M. G. Husada, A. N. Hermana, and S. A. Rumapea, “Fruit Ripeness Based on RGB, HSV, HSL, L\*a\*b\* Color Feature Using SVM,” *2019 Int. Conf. Comput. Sci. Inf. Technol. ICoSNIKOM 2019*, pp. 2–6, 2019.
- [14] S. Mawarni and P. N. Bengkalis, “SistemPrediksi Pengunduran Diri Calon Mahasiswa Baru Menggunakan Algoritma C45,” pp. 227–236, 2018.
- [15] A. Rohman *et al.*, “Implementasi Data Mining Dengan Algoritma Decision Tree C4 . 5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas,” pp. 134–139, 2019.