



## **PEMODELAN PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAÏVE BAYES DI UNIBA**

**Nur Mahar Aji<sup>1</sup>, Vihi Atina<sup>2</sup>, Nugroho Arif Sudibyo<sup>3</sup>**

<sup>1,3</sup>Program Studi Teknik Informatika, Universitas Duta Bangsa Surakarta

<sup>2</sup>Program Studi Manajemen Informatika, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara No. 55, Tipes, Kec. Serengan, Kota Surakarta, Jawa Tengah 57154

<sup>1</sup> [190103194@fikom.udb.ac.id](mailto:190103194@fikom.udb.ac.id), <sup>2</sup> [vihi\\_atina@udb.ac.id](mailto:vihi_atina@udb.ac.id), <sup>3</sup> [nugroho\\_arif@udb.ac.id](mailto:nugroho_arif@udb.ac.id)

### **Abstract**

*The quality of tertiary institutions can be measured based on the accreditation issued by the Badan Akreditasi Nasional Perguruan Tinggi or BAN-PT. One of the accreditation assessment criteria is the number of student graduates. Universitas Islam Batik Surakarta (UNIBA), is a private tertiary institution in the Surakarta area that has four faculties, with the number of students continuing to increase or decrease every year. This research uses the naïve Bayes method to make predictions about student graduation on time or late, with the prediction results that can be used as information and data input for universities to improve the quality of higher education. The naïve Bayes method has advantages, including being able to predict the number of graduates based on concrete data, so that the results obtained can be accounted for and used for further predictions. In making predictions, 14 variables are needed, including gender, student status, age, married status, grade point index 1, grade point index 2, grade point index 3, grade point index 4, grade point index 5, grade point index 6, grade point index 7, grade point index 8, grade point average, and graduation certificate. From a data set of 302 students with 14 attributes of class 2018 students at the Universitas Islam Batik Surakarta using the python programming language and jupyter notebook software, the authors obtained accurate prediction results of 85%, with a late precision value of 0.42, precise precision of 0.95, late recall of 0.65, recall is exactly 0.88, f1-score is late 0.51, f1-score is exactly 0.91.*

**Keywords :** Predictions, Student Graduation, Data Mining, Naive Bayes

### **Abstrak**

Kualitas perguruan tinggi dapat diukur berdasarkan akreditasi yang dikeluarkan oleh Badan Akreditasi Nasional Perguruan Tinggi atau BAN-PT. Salah satu kriteria akreditasi yaitu jumlah lulusan mahasiswa. Universitas Islam Batik Surakarta (UNIBA), merupakan salah satu perguruan tinggi swasta di kawasan Surakarta yang memiliki empat fakultas, dengan jumlah mahasiswa yang terus mengalami peningkatan maupun penurunan disetiap tahunnya. Penelitian ini menggunakan metode *naïve bayes* untuk melakukan prediksi terhadap kelulusan mahasiswa tepat waktu ataupun terlambat, dengan hasil prediksi yang dapat dijadikan sebagai informasi dan masukan data bagi perguruan tinggi untuk meningkatkan kualitas perguruan tinggi. Metode *naïve bayes* memiliki kelebihan, diantaranya dapat memprediksi jumlah kelulusan berdasarkan data-data konkrit, agar hasil yang diperoleh dapat dipertanggungjawabkan dan digunakan untuk prediksi selanjutnya. Dalam melakukan prediksi, diperlukan 14 variabel, antara lain jenis kelamin, status mahasiswa, umur, status menikah, indeks prestasi semester 1, indeks prestasi semester 2, indeks prestasi semester 3, indeks prestasi semester 4, indeks prestasi semester 5, indeks prestasi semester 6, indeks prestasi semester 7, indeks prestasi semester 8, indeks prestasi kumulatif, dan keterangan lulus. Dari data set sebanyak 302 mahasiswa dengan 14 atribut mahasiswa angkatan 2018 Universitas Islam Batik Surakarta dengan menggunakan bahasa pemrograman python dan software jupyter notebook, penulis memperoleh hasil akurasi ketepatan prediksi sebesar



85%, dengan nilai *precision* terlambat 0.42, *precision* tepat 0.95, *recall* terlambat 0.65, *recall* tepat 0.88, *f1-score* terlambat 0.51, *f1-score* tepat 0.91.

**Kata kunci :** Prediksi, Kelulusan Mahasiswa, Data Mining, *Naive Bayes*

## 1. PENDAHULUAN

Dalam penyelenggaraan pendidikan, perguruan tinggi dituntut untuk menyelenggarakan pendidikan berkualitas, sehingga dapat menghasilkan mahasiswa yang kompeten, kreatif, dan memiliki daya saing. Kualitas perguruan tinggi di Indonesia dapat dilihat dari akreditasi dari Badan Akreditasi Nasional Perguruan Tinggi atau BAN-PT. Tingkat kelulusan menjadi salah satu acuan penilaian untuk memperoleh akreditasi[1], seperti tertera melalui Lampiran Peraturan Badan Akreditasi Nasional Perguruan Tinggi No. 23 Tahun 2022.

Universitas Islam Batik Surakarta (UNIBA) termasuk sebagai perguruan tinggi besar di kawasan Surakarta, yang berdiri pada tahun 1983. UNIBA mempunyai sejumlah empat Fakultas, antara lain Fakultas Ekonomi, Fakultas Pertanian, Fakultas Hukum, dan Fakultas Teknik[2]. Setiap tahun jumlah mahasiswa di UNIBA terjadi kenaikan dan penurunan. Namun tidak ada keseimbangan antara jumlah mahasiswa yang masuk yang keluar setelah menyelesaikan studinya. Hal tersebut dapat berpengaruh terhadap kualitas dan akreditasi dari Universitas Islam Batik Surakarta.

Evaluasi tingkat kelulusan yang dilakukan selama ini hanya berpedoman terhadap data pendaftaran wisuda, tanpa memperhatikan mahasiswa yang masih mengalami permasalahan akademis maupun administratif. Sedangkan tindak lanjut universitas terhadap mahasiswa yang tidak lulus tepat waktu dimungkinkan melalui cara membujuk, mengarahkan, dan membimbing mahasiswa untuk segera menyelesaikan studinya.

Permasalahan yang terjadi di UNIBA adalah belum adanya proses prediksi kelulusan mahasiswa dengan sistem terintegrasi. Akibat dari hal tersebut antara lain : bagian Biro Administrasi Akademik dan Kemahasiswaan tidak bisa menjamin satu angkatan tersebut lulus tepat waktu, terjadinya situasi dimana mahasiswa tidak mendapatkan solusi atas keterlambatan kelulusan yang terjadi.

Berdasarkan permasalahan di atas, maka dibutuhkan adanya aplikasi prediksi terhadap kelulusan mahasiswa di Universitas Islam Batik Surakarta. Aplikasi ini mencakup data mahasiswa dalam satu angkatan semua program studi. Semoga dengan adanya aplikasi ini, kualitas mahasiswa dan akreditasi Universitas Islam Batik Surakarta dapat terus terpantau dan terus meningkat.

## 2. TINJAUAN PUSTAKA

### 2.1. Tinjauan Pustaka

Terkait kasus prediksi kelulusan mahasiswa, telah dilakukan penelitian dari berbagai universitas dengan metode yang beragam. Penelitian sebelumnya dilakukan oleh Armansyah, Rakhmat Kurniawan Ramli dengan judul “Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve bayes”[3], dengan permasalahan yang digunakan berdasarkan rendahnya jumlah lulusan yang terjadi akibat kesenjangan antara jumlah mahasiswa masuk dengan yang lulus atau keluar, yang berdampak buruk terhadap program studi dalam berbagai aspek. Metode yang digunakan yaitu dengan cara eksperimen dan *naïve bayes*, dengan hasil yang diperoleh berupa data prediksi tingkat kelulusan mahasiswa yang memperlihatkan performa metode atas model prediksi tersebut dapat bekerja dengan begitu baik, hingga bahkan akurasi mencapai 100%.

Selanjutnya, penelitian dari Lydia Yohana Lumban Gaol, M. Safii, dan Dedi Suhendro dengan judul “Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5”[4], permasalahan yang digunakan berdasarkan fakta bahwa kelulusan termasuk sebagai aspek penilaian akreditasi yang penting dalam perguruan tinggi, sehingga ketika didapati semakin banyak mahasiswa yang lulus tepat waktu, artinya juga mendorong semakin baiknya penilaian akreditasi suatu perguruan tinggi. Metode yang digunakan yaitu dengan klasifikasi C4.5 dalam mengklasifikasikan data dengan atribut numerik



ataupun kategorial. Hasil dari penelitian tersebut adalah data prediksi kelulusan mahasiswa yang memperlihatkan bahwa pengaruh yang paling tinggi dalam penentu kelulusan mahasiswa adalah atribut IPK.

Berikutnya, penelitian yang dilakukan oleh Sukarna Royan, Ansori Yulian, dan Syaechurodji dengan judul "Implementasi Data Mining Menggunakan Metode Naïve Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu"[5], permasalahan yang digunakan berdasarkan rendahnya nilai angka efisiensi edukasi yang diketahui turut menjadi indikator kualitas program pendidikan, yang berimbas terhadap penilaian akreditasi, baik pada Program Studi maupun Perguruan Tinggi. Adapun metodenya yaitu dengan *naïve bayes* guna melakukan klasifikasi tanpa memerlukan banyak data pelatihan. Hasil dari penelitian tersebut berupa data prediksi kelulusan mahasiswa menggunakan *feature selection* yang menghasilkan tingkat akurasi lebih baik.

Pada penelitian berikutnya, yakni oleh Ray Mondow Sagala dengan judul "Prediksi Kelulusan mahasiswa Menggunakan Data Mining Algoritma K-means"[6], permasalahan yang digunakan berdasarkan kelulusan yang menjadi pengharapan besar untuk mahasiswa, sebab ada keterikatan atau keterkaitan antara beberapa mata kuliah. *K-means* yang dipergunakan sebagai metode bisa memungkinkan data penelitiannya berupa angka. Hasil dari penelitian tersebut berupa data prediksi yang memperlihatkan bahwa berdasarkan  $k = 3$  dari 118 data yang diolah, ditemukan sejumlah 13 mahasiswa tidak lulus, kemudian sejumlah 36 mahasiswa lulus dalam klasifikasi nilai cukup, serta sejumlah 69 mahasiswa lulus dalam klasifikasi nilai baik.

Pada penelitian berikutnya, dilakukan oleh Nursetia Wati dengan judul "Prediksi Kelulusan mahasiswa Menggunakan *K-Nearest Neighbor* Berbasis *Particle Swarm Optimization*"[7], permasalahan yang digunakan berdasarkan persentase kelulusan tidak tepat waktu terutama pada Fakultas Teknik sangatlah tinggi, kondisi tersebut mendorong setiap program studi terus meningkatkan kelulusan agar saran mutu bisa tercapai. Guna mengelompokkan data berdasarkan jarak dari data baru ke data yang terbaru, digunakanlah metode eksperimen melalui *K-Nearest Neighbor (KNN)*. Hasil penelitian tersebut berupa pernyataan bahwa

pengujian yang dilakukan mendapatkan nilai terbaik saat menerapkan algoritma *K-Nearest Neighbor*.

## 2.2. Dasar Teori

Data mining merupakan istilah untuk menjabarkan penemuan pengetahuan dalam database, melalui penggunaan matematika, teknik statistik, *machine learning*, ataupun kecerdasan buatan guna melakukan identifikasi dan ekstraksi berbagai informasi yang dapat dimanfaatkan, serta pengetahuan yang relevan dari beragam database besar [8]. Maulana dan Fajrin mengatakan bahwa data mining dalam suatu penelitian pada dasarnya bukan topik baru, sebab dapat berguna dalam meningkatkan efektivitas berbagai teknik sebelumnya acap digunakan sehingga dapat menangani beragam masalah yang biasa dijumpai[9].

Aplikasi merupakan program yang bersifat siap guna untuk dimanfaatkan dalam melaksanakan berbagai perintah para penggunaannya, sehingga akan memperoleh hasil lebih akurat sebagaimana yang dikehendaki ketika membuat aplikasi tersebut. Selain itu, aplikasi juga diartikan sebagai upaya pemecahan masalah melalui penggunaan salah satu teknik pemrosesan data aplikasi yang diharap atau dikehendaki, dan juga sebagai pemrosesan data yang pada khususnya berfungsi secara terpadu berdasarkan kemampuannya. Kemudian juga disampaikan bahwa aplikasi adalah perangkat komputer yang siap pakai untuk para penggunaannya [10].

Algoritma pertama kali dicetuskan oleh Al Khawarizmi, seorang ilmuwan dari Persia. Pada awalnya, algoritma pertama kali digunakan untuk memecahkan masalah aritmatika, tetapi algoritma telah digunakan dari waktu ke waktu untuk memecahkan masalah matematika. Algoritma juga memiliki hubungan yang penting dengan matematika[11]. T.S. Alasi mengatakan algoritma merupakan serangkaian langkah yang bersifat logis untuk menyelesaikan masalah dengan penyusunan yang sistematis[12].

Prediksi adalah kegiatan yang dapat dijalankan guna memperkirakan hal yang dimungkinkan terjadi pada masa mendatang, melalui pemanfaatan berbagai data lama dengan berdasarkan pada suatu indikator tertentu. Berbagai permasalahan yang memerlukan



prediksi, antara lain seperti dalam hal harga, hasil produksi, prediksi tingkat kelulusan, ataupun lain sebagainya [13].

Kelulusan mahasiswa merupakan hal penting untuk dapat diperhatikan, sebab besaran persentase kelulusan akan berdampak terhadap status akreditasi program studi dan penilaian pemerintah. Serta tentunya juga penting untuk para mahasiswa itu sendiri. Terlebih faktor-faktor yang berpengaruh pada kelulusan pastinya sangat erat kaitannya dengan mahasiswa, antara lain yaitu biaya perkuliahan, penghasilan orangtua, IPS semester dari IPS 1 hingga IPS 7, IPK, dan SKS persemester [14].

*Naïve bayes classifier* dijelaskan sebagai suatu metode untuk mengklasifikasikan berdasarkan teknik probabilitas dan statistik sebagai hasil pengembangan oleh *Thomas Bayes*, ilmuwan asal Inggris. *Naïve bayes* sebagai suatu algoritma ini digunakan untuk memprediksi peluang yang akan terjadi pada masa mendatang dengan didasarkan pada pengalaman dimasa sebelumnya[15]. *Naïve bayes classifier* mendapati ciri, yakni terdapat asumsi kuat (naif) pada independensi atas setiap kejadian. Algoritma ini memiliki dua tahapan klasifikasi data, yakni dengan membuat data latih yang berlabel positif dan negatif, selanjutnya membuat data uji guna mengidentifikasi performa dari algoritma yang sebelumnya telah dilatih [16].

Klasifikasi adalah sebuah proses untuk mengelompokkan dan memprediksi suatu data baru dari sebuah atribut atau variabel berdasarkan standar dari data yang sudah ada. Klasifikasi memiliki dua jenis data, yakni data uji (*testing data*) beserta data latih (*training data*)[17].

*Jupyter Notebook* terdiri dari tiga bahasa pemrograman yang disingkat, yang terdiri dari *Julia (Ju)*, *Python (Py)*, *R*[18]. *Jupyter Notebook* adalah aplikasi web gratis yang bisa dipergunakan dalam pembuatan dan berbagi berbagai dokumen komputasi dengan pengalaman yang sederhana, efisien, dan hanya berfokus terhadap dokumennya [19].

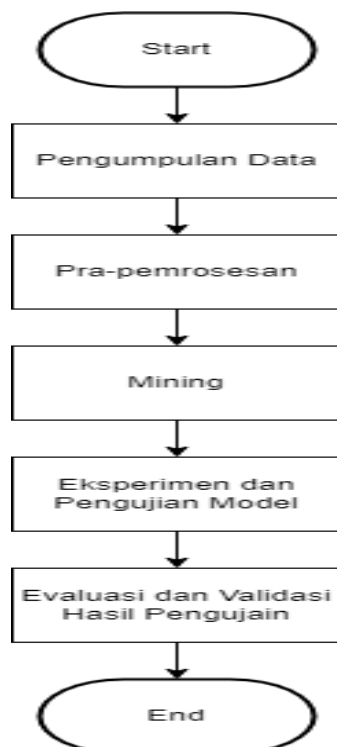
*Python* diketahui termasuk sebagai bahasa pemrograman tingkat tinggi yang sifatnya interaktif, interpreter, *object-oriented*, serta bisa berjalan pada nyaris setiap platform, antara lain *Linux*, *Mac*, *Windows*, ataupun yang lain sebagainya. Selain itu, *Python* juga termasuk sebagai bahasa pemrograman tingkat tinggi yang

dinilai tidak sulit untuk dipelajari, sebab sintaksnya jelas dan elegan, melalui kombinasi terhadap pemanfaatan berbagai modul yang memiliki struktur data bertingkat tinggi. Pada umumnya, *source code* aplikasi *python* akan dikompilasi sebagai format perantara yang disebut *byte code* sebelum dieksekusi[20].

### 3. METODOLOGI PENELITIAN

#### 3.1. Tahapan Penelitian

Penelitian ini dimulai dengan perumusan masalah dan studi literatur. Diperlukan sejumlah data yang diperoleh melalui metode observasi dan dokumentasi untuk dapat memenuhi permasalahannya [21]. Berikutnya pengumpulan data mahasiswa yang dibutuhkan, dengan perolehan sejumlah 395 data set mahasiswa angkatan 2018 yang sudah selesai studinya dalam 16 atribut. Kemudian tahapan kedua dilaksanakan pra-pemrosesan data untuk mendapatkan data yang baik sebelum data diolah menggunakan *naïve bayes*. Jika pra-pemrosesan data sudah selesai dilakukan, artinya akan digunakan sejumlah 302 data mahasiswa dalam 14 atribut untuk melaksanakan proses mining. Ketiga, proses mining dengan menerapkan algoritma *naïve bayes* dengan menggunakan *python*. Tahap keempat eksperimen dan uji model. Tahap kelima evaluasi dan validasi hasil pengujian.



**Gambar 1.** Alur penelitian

### 3.2. Pengumpulan Data

Adapun untuk teknik yang digunakan untuk mengumpulkan data penelitiannya terdiri dari :

1. Metode observasi

Upaya dalam memperoleh atau mengumpulkan data secara langsung dengan mengamati di lapangan, untuk kemudian akan berguna dalam perhitungan ketika memprediksikan kinerja sistem yang berjalana pada perusahaan[22]. Penulis melakukan pengamatan secara langsung di Universitas Islam Batik Surakarta.

2. Metode wawancara

Melaksanakan wawancara secara langsung dalam beberapa pertanyaan yang disampaikan pada Biro Administrasi Akademik dan Kemahasiswaan untuk menentukan data yang diperlukan.

3. Metode kepustakaan

Mengumpulkan data dengan cara mempelajari buku, jurnal ilmiah, serta informasi dari internet yang mendukung penelitian.

### 3.3. Analisa Data

Data mahasiswa pada penelitian ini adalah data dari mahasiswa di Universitas Islam Batik Surakarta angkatan 2018, berupa jenis kelamin, status mahasiswa, status menikah, umur, Indeks Prestasi Semester (IPS) dari semester 1 sampai dengan semester 8, dan Indeks Prestasi Kumulatif (IPK). Dengan kelas target klasifikasinya yaitu status kelulusan yang mencakup apakah tepat atau terlambat.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Pengumpulan Data

Proses pengumpulan datanya dilangsungkan secara langsung di lapangan dan dari website *siakad* UNIBA dan *neofeeder* dikti. Data yang digunakan diambil secara manual dengan cara menulis satu persatu data pada file *microsoft excel* yang berformat *xlsx*. Data yang diperoleh dari angkatan masuk tahun 2018 yang telah menyelesaikan studinya. Data tersebut sebanyak 395 dengan 16 atribut. Adapun contoh data yang digunakan terdapat dalam gambar 2 dan 3.

	A	B	C	D	E	F
1	NPM	Nama	Jenis Kelamin	Status Mahasiswa	Umur	Status Menikah
2	'2018020001	FITRIA NUR JANAH	Wanita	Mahasiswa	23	Belum
3	'2018020002	DWI YUDIYANTO	Pria	Bekerja	35	Sudah
4	'2018020003	RINDIANI PUTRI WIDAWATI	Wanita	Mahasiswa	23	Belum
5	'2018020004	RIZAL BAYU PRASETYA	Pria	Bekerja	24	Belum
6	'2018020005	MUHAMMAD FAHRIZAL AMRU	Pria	Mahasiswa	23	Belum
7	'2018020006	ALMIRA DIAH ASMOROWATI	Wanita	Mahasiswa	22	Belum
8	'2018020007	DHITA APRIANI	Wanita	Mahasiswa	22	Belum
9	'2018020008	WAHYU EDY SAPUTRO	Pria	Bekerja	24	Belum
10	'2018020009	NADA ANISA KAMILIA	Wanita	Mahasiswa	22	Belum
11	'2018020010	ERVINA SRIMA'RIFAA	Wanita	Mahasiswa	22	Belum
12	'2018020011	JUWITA EKA MURSANI	Wanita	Mahasiswa	22	Belum
13	'2018020012	MELINDA PUTRI WARDANI	Wanita	Mahasiswa	23	Belum
14	'2018020013	AMALIA NOVEL AGSA	Wanita	Mahasiswa	22	Belum
15	'2018020014	ZAINI AZHARI ZULFAN	Pria	Bekerja	24	Belum
16	'2018020015	TUTUT SHOLEH BUDI SANTOSO	Pria	Mahasiswa	23	Belum
17	'2018020016	ASYA KURNIA SYAPUTRA	Pria	Bekerja	24	Belum
18	'2018020017	NOLA DWIJA AMELIA	Wanita	Mahasiswa	22	Belum
19	'2018020018	AWAHITA HERMAWANTI	Wanita	Mahasiswa	23	Belum
20	'2018020019	FADHLURROHMAN ARKAAN	Pria	Bekerja	24	Belum
21	'2018020020	SELYSTIANA YUNIAR	Wanita	Mahasiswa	22	Belum

**Gambar 2.** Contoh Data



G	H	I	J	K	L	M	N	O	P
IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPS8	IPK	Lulus
3,45	3,5	3,65	3,63	3,63	3,81	3,65	3,32	3,58	Terlambat
3,73	3,5	2,96	3,63	3,84	3,76	3,82	3,24	3,56	Terlambat
3,14	3,48	3,65	3,75	3,53	3,86	3,82	3,41	3,58	Tepat
2,86	3,48	2,91	3,38	3,58	3,9	3,82	2,79	3,34	Terlambat
3	3,38	3,61	3,38	3,68	4	3,82	2,57	3,43	Tepat
3,36	3,76	3,87	3,63	4	3,86	4	3	3,78	Tepat
3,45	3,33	3,39	3,5	3,42	3,43	3,82	3,18	3,44	Tepat
3,23	3,68	3,65	4	3,84	4	3,82	4	3,73	Tepat
3,14	3,41	3,65	3,88	4	3,89	3,65	3,34	3,62	Tepat
3,55	4	3,87	4	3,84	3,81	3,65	4	3,84	Tepat
3,5	4	3,87	4	3,68	4	3,65	4	3,84	Tepat
3,36	3,67	3,65	3,75	4	3,9	4	4	3,73	Tepat
3,64	3,86	3,87	4	3,84	4	3,65	3	3,86	Tepat
3,73	3,71	3,91	4	4	4	3,69	4	3,88	Tepat
3,36	3,62	4	3,88	3,74	4	3,65	3	3,77	Tepat
3,27	3,48	3,87	3,75	3,58	3,76	3,65	3,76	3,64	Terlambat
3,59	3,76	3,91	3,63	3,84	4	4	4	3,8	Tepat
2,91	3,48	3,17	3,63	3,26	3,57	3,82	2,56	3,3	Terlambat
3,5	3,48	3,57	3,75	3,58	3,62	3	3,74	3,53	Tepat
3,5	3,71	3,57	3,63	3,84	4	4	4	3,69	Tepat

Gambar 3. Contoh Data Lanjut

#### 4.2. Pra-pemrosesan

Berdasarkan hasil pengumpulan datanya, diperoleh *record* sejumlah 395 data set mahasiswa yang sudah selesai masa studinya. Untuk kategori yang digunakan pada mahasiswa adalah yang masa studinya selesai tetap waktu, atau selesai dalam 7 semester (3.5 tahun) dan 8 semester (4 tahun). Kemudian yagn termasuk kategori terlambat, yakni yang menempuh atau selesai masa studinya di atas 8 semester. Kendati demikian, tidak semua data *record* dan atribut tidak dapat dipakai, sehingga diperlukan adanya pra-pemrosesan data atau upaya mengolah data awal guna memperoleh data lebih baik. Kemudian berikut ini merupakan rincian dari sejumlah 16 atribut data yang belum melalui pra-pemrosesan :

Tabel 1. Data Sebelum Pra Pemrosesan

No	Nama	Jenis Data
1	NPM	Karakter
2	Nama	Karakter
3	Jenis Kelamin	Kategorial
4	Status Mahasiswa	Kategorial
5	Umur	Numerik
6	Status Menikah	Kategorial
7	IPS 1	Numerik
8	IPS 2	Numerik
9	IPS 3	Numerik
10	IPS 4	Numerik
11	IPS 5	Numerik
12	IPS 6	Numerik

13	IPS 7	Numerik
14	IPS 8	Numerik
15	IPK	Numerik
16	Lulus	Kategorial

Adapun untuk ra-pemrosesan yang digunakan oleh penulis, terdiri dari:

- Pembersihan Data, dengan cara menghapus atau menghilangkan data kosong dan kruang lengkap. Seperti, *record* data mahasiswa yang berstatus tidak aktif dan berhenti, data tersebut dihapus karena sebab tidak lengkap muatan data nilai mata kuliahnya. Hingga kemudian hanya diperoleh sejumlah 302 data yang dapat digunakan dari sebelumnya, atau data awal sejumlah 395, dengan kata lain dilakukan pembersihan data sebesar 23,54% dalam tahap pra-pemrosesan data guna mencegah adanya *missing value*.
- Reduksi data, berguna untuk memperoleh data set dalam *record* berikut jumlah atribut yang sesuai pada proses mining. Hal ini dicontohkan dengan tidak digunakannya atribut NPM dan nama, sebab dinilai tidak atau kurang relevan. Artinya, dalam proses miningnya hanya menggunakan beberapa atrubut, yaitu jenis kelamin, umur, status mahasiswa, status menikah, Indeks Prestasi Semester (IPS) dari semester 1 sampai 8, serta Indeks Prestasi Kumulatif (IPK), dan Lulus.
- Mengubah data dengan jenis data kategorial kedalam jenis data numerik. Atribut jenis kelamin dengan data pria dan wanita, diubah menjadi 0 untuk pria, dan 1 untuk wanita. Atribut status mahasiswa dengan data bekerja dan mahasiswa diubah menjadi 0 untuk bekerja, dan 1 untuk mahasiswa. Atribut status menikah dengan data sudah dan belum diubah menjadi 0 untuk sudah, dan 1 untuk belum. Atribut lulus dengan data tepat dan terlambat, diubah menjadi 0 untuk terlambat, dan 1 untuk tepat.

Sesudah pra-pemrosesan data selesai dilakukan, kemudian berlanjut pada proses mining terhadap sejumlah 302 data mahasiswa dalam 14 atribut yang sudah melalui penormalisasian dan penghindaran atas kemungkinan *missing value*. Berikut rinciannya yang tercantum dalam tabel berikut :

**Tabel 2.** Data Setelah Pra Pemrosesan

No	Nama	Jenis Data
1	Jenis Kelamin	Numerik
2	Status Mahasiswa	Numerik
3	Umur	Numerik
4	Status Menikah	Numerik
5	IPS 1	Numerik
6	IPS 2	Numerik
7	IPS 3	Numerik
8	IPS 4	Numerik
9	IPS 5	Numerik
10	IPS 6	Numerik
11	IPS 7	Numerik
12	IPS 8	Numerik
13	IPK	Numerik
14	Lulus	Numerik

#### 4.3. Mining

Mengubah data berjenis kategorial menjadi numerik, contoh data tercantum dalam gambar berikut :

	A	B	C	D
1	JenisKelamin	StatusMahasiswa	Umur	StatusMenikah
2	1	1	23	1
3	0	0	35	0
4	1	1	23	1
5	0	0	24	1
6	0	1	23	1
7	1	1	22	1
8	1	1	22	1
9	0	0	24	1
10	1	1	22	1
11	1	1	22	1
12	1	1	22	1
13	1	1	23	1
14	1	1	22	1
15	0	0	24	1
16	0	1	23	1
17	0	0	24	1
18	1	1	22	1
19	1	1	23	1
20	0	0	24	1
21	1	1	22	1

**Gambar 4.** Data Setelah Diubah

E	F	G	H	I	J	K	L	M	N
IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPS8	IPK	Lulus
3,45	3,5	3,65	3,63	3,63	3,81	3,65	3,32	3,58	0
3,73	3,5	2,96	3,63	3,84	3,76	3,82	3,24	3,56	0
3,14	3,48	3,65	3,75	3,53	3,86	3,82	3,41	3,58	1
2,86	3,48	2,91	3,38	3,58	3,9	3,82	2,79	3,34	0
3	3,38	3,61	3,38	3,68	4	3,82	2,57	3,43	1
3,36	3,76	3,87	3,63	4	3,86	4	3	3,78	1
3,45	3,33	3,39	3,5	3,42	3,43	3,82	3,18	3,44	1
3,23	3,68	3,65	4	3,84	4	3,82	4	3,73	1
3,14	3,41	3,65	3,88	4	3,89	3,65	3,34	3,62	1
3,55	4	3,87	4	3,84	3,81	3,65	4	3,84	1
3,5	4	3,87	4	3,68	4	3,65	4	3,84	1
3,36	3,67	3,65	3,75	4	3,9	4	4	3,73	1
3,64	3,86	3,87	4	3,84	4	3,65	3	3,86	1
3,73	3,71	3,91	4	4	4	3,69	4	3,88	1
3,36	3,62	4	3,88	3,74	4	3,65	3	3,77	1
3,27	3,48	3,87	3,75	3,58	3,76	3,65	3,76	3,64	0
3,59	3,76	3,91	3,63	3,84	4	4	4	3,8	1
2,91	3,48	3,17	3,63	3,26	3,57	3,82	2,56	3,3	0
3,5	3,48	3,57	3,75	3,58	3,62	3	3,74	3,53	1
3,5	3,71	3,57	3,63	3,84	4	4	4	3,69	1

**Gambar 5.** Data Setelah Diubah Lanjutan

#### 4.4. Eksperimen dan Pengujian Model

Penulis menggunakan software *jupyter* untuk melakukan eksperimen data kelulusan mahasiswa dengan metode *naïve bayes*.

1. Memanggil *library* yang diperlukan

```
import pandas as pan
import numpy as npy
```

**Gambar 6.** Perintah Memanggil *Library*

2. Membaca data kelulusan mahasiswa pada file *excel*

```
mahasiswa = pan.read_excel("Data Kelulusan Mahasiswa Ok.xlsx")
mahasiswa.head(302)
```

**Gambar 7.** Perintah Membaca Data *Excel*

3. Hasil data yang ditampilkan sebagai berikut

Out[2]:	JenisKelamin	StatusMahasiswa	Umur	StatusMenikah
0	1	1	23	1
1	0	0	35	0
2	1	1	23	1
3	0	0	24	1
4	0	1	23	1
...	...	...	...	...
297	0	0	25	0
298	0	0	37	0
299	0	1	23	1
300	0	1	22	1
301	0	1	23	1

302 rows × 4 columns

**Gambar 8.** Hasil Data *Python*

IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPS8	IPK	Lulus
3.45	3.50	3.65	3.63	3.63	3.81	3.65	3.32	3.58	0
3.73	3.50	2.96	3.63	3.84	3.76	3.82	3.24	3.56	0
3.14	3.48	3.65	3.75	3.53	3.86	3.82	3.41	3.58	1
2.86	3.48	2.91	3.38	3.58	3.90	3.82	2.79	3.34	0
3.00	3.38	3.61	3.38	3.68	4.00	3.82	2.57	3.43	1
...	...	...	...	...	...	...	...	...	...
3.78	3.78	3.40	3.70	3.73	3.76	3.71	4.00	3.69	1
3.78	3.91	3.65	3.70	3.86	3.79	3.71	4.00	3.79	1
3.11	3.05	2.92	3.09	3.63	3.38	2.89	3.00	3.14	1
3.58	2.80	2.83	3.09	4.00	3.50	3.06	3.00	3.20	1
3.68	3.70	3.83	3.82	3.94	4.00	3.44	3.00	3.75	1

Gambar 9. Tampilan Data Python Lanjutan

- Memisahkan atribut ke dalam 2 kelompok, yaitu x dan y. Variabel x terdiri dari atribut jenis kelamin, status mahasiswa, umur, status menikah, IPS dari semester 1 sampai 8, dan IPK. Sedangkan variabel y terdiri dari atribut lulus

```
x = mahasiswa.drop(["Lulus"], axis=1)
x.head(302)
```

Gambar 10. Perintah Membuang Atribut Lulus

Out[3]:

	JenisKelamin	StatusMahasiswa	Umur	StatusMenikah
0	1	1	23	1
1	0	0	35	0
2	1	1	23	1
3	0	0	24	1
4	0	1	23	1
...	...	...	...	...
297	0	0	25	0
298	0	0	37	0
299	0	1	23	1
300	0	1	22	1
301	0	1	23	1

302 rows × 5 columns

Gambar 11. Data Variabel X

IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPS8	IPK
3.45	3.50	3.65	3.63	3.63	3.81	3.65	3.32	3.58
3.73	3.50	2.96	3.63	3.84	3.76	3.82	3.24	3.56
3.14	3.48	3.65	3.75	3.53	3.86	3.82	3.41	3.58
2.86	3.48	2.91	3.38	3.58	3.90	3.82	2.79	3.34
3.00	3.38	3.61	3.38	3.68	4.00	3.82	2.57	3.43
...	...	...	...	...	...	...	...	...
3.78	3.78	3.40	3.70	3.73	3.76	3.71	4.00	3.69
3.78	3.91	3.65	3.70	3.86	3.79	3.71	4.00	3.79
3.11	3.05	2.92	3.09	3.63	3.38	2.89	3.00	3.14
3.58	2.80	2.83	3.09	4.00	3.50	3.06	3.00	3.20
3.68	3.70	3.83	3.82	3.94	4.00	3.44	3.00	3.75

Gambar 12. Data Variabel X Lanjutan

```
y = mahasiswa["Lulus"]
y.head(302)
```

Gambar 13. Perintah Mengelompokkan Atribut Lulus ke Variabel Y

```
Out[4]: 0      0
        1      0
        2      1
        3      0
        4      1
        ..
       297     1
       298     1
       299     1
       300     1
       301     1
       Name: Lulus, Length: 302, dtype: int64
```

Gambar 14. Data Variabel Y

- Import Gaussian Naïve Bayes Model

```
from sklearn.naive_bayes import GaussianNB
```

Gambar 15. Perintah Import GaussianNB Model

- Memanggil Fungsi GaussianNB

```
nbc = GaussianNB()
```

Gambar 16. Perintah Memanggil GaussianNB

- Membuat Data Training

```
data_training = nbc.fit(x,y)
```

Gambar 17. Perintah Membuat Data Training

- Melakukan Prediksi Pada Data Training

```
y_predict = data_training.predict(x)
print(y_predict)
```

Gambar 18. Perintah Prediksi Data Training







Terutama kepada Universitas Islam Batik Surakarta yang telah memberikan izin kepada penulis untuk melaksanakan penelitian ini.

**DAFTAR PUSTAKA:**

- [1] C. P. Demasari, N. Ariyanti, P. G. Muhanan, and B. Iman, "Implementasi Analisis Big Data untuk Memprediksi Kelulusan Mahasiswa Fakultas Teknik Universitas Langlangbuana," vol. 6, no. 1, pp. 48–52, 2023.
- [2] A. Dharma and I. Putri, "Pengaruh Strategi Bauran Pemasaran Terhadap Keputusan Mahasiswa Memilih Fakultas Ekonomi Universitas Islam Batik Surakarta," *J. Akunt.*, vol. 6, no. 1, pp. 24–35, 2019.
- [3] A. Armansyah and R. K. Ramli, "Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes," *Edumatic J. Pendidik. Inform.*, vol. 6, no. 1, pp. 1–10, 2022, doi: 10.29408/edumatic.v6i1.4789.
- [4] L. Y. L. M. S. D. S. Gaol, "Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5," *Brahmana J. Penerapan Kecerdasan Buatan*, vol. 2, no. 2, pp. 97–106, 2021, doi: 10.30645/brahmana.v2i2.71.
- [5] S. Royan, A. Yulian, and Syaechurodji, "Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. Ilm. Sains dan Teknol.*, vol. 6, no. 1, pp. 50–61, 2022, doi: 10.47080/saintek.v6i1.1467.
- [6] R. M. Sagala, "Prediksi Kelulusan Mahasiswa Menggunakan Data mining Algoritma K-means," *J. TelKa*, vol. 11, no. 2, pp. 131–142, 2021.
- [7] N. Wati, "PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN K-NEAREST NEIGHBOR BERBASIS PARTICLE SWARM OPTIMIZATION Nursetia Wati," *Jtii*, vol. 6, no. 2, pp. 118–127, 2021.
- [8] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [9] N. A. Sudibyo, Ardymulya Iswardani, Kartika Sari, and Siti Suprihatiningsih, "Penerapan Data Mining Pada Jumlah Penduduk Miskin Di Indonesia," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 1, no. 3, pp. 199–207, 2020, doi: 10.46306/lb.v1i3.42.
- [10] I. P. Sari, A. Syahputra, N. Zaky, R. U. Sibuea, and Z. Zakhir, "Perancangan Sistem Aplikasi Penjualan dan Layanan Jasa Laundry Sepatu Berbasis Website," *Blend Sains J. Tek.*, vol. 1, no. 1, pp. 31–37, 2022, doi: 10.56211/blendsains.v1i1.67.
- [11] N. Khesya, "Mengenal Flowchart dan Pseudocode Dalam Algoritma dan Pemrograman," *Preprints*, vol. 1, pp. 1–15, 2021, [Online]. Available: <https://osf.io/dq45ef>
- [12] T. S. Alasi, A. T. Al, and A. Siahaan, "Algoritma Vigenere Cipher Untuk Penyandian Record Informasi Pada Database," *J. Inf. Komput. Log.*, vol. 1, no. 4, 2020, [Online]. Available: <http://ojs.logika.ac.id/index.php/jikl>
- [13] R. Harun, K. C. Pelangi, and Y. Lasena, "Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naive Bayes," *J. Manaj. Inform. dan Sist. Inf.*, vol. 3, no. 1, pp. 8–15, 2020, [Online]. Available: <http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/19417.pdf>
- [14] R. Thaniket, Kusriani, and E. T. Luthf, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma Support Vector Machine," *J. FATEKSA J. Teknol. dan rekayasa*, vol. 13, no. 2, pp. 69–83, 2019.
- [15] D. Ghassa, Aji; Wahyudi, Adi; Tampubolon, Silvia Ovella, Putri, Nurul Afrilia; Rasywir, Errisya; Kisbianty, "Penerapan Data Mining Algoritma Naive Bayes Clasifier Untuk Mengetahui Minat Beli Pelanggan Terhadap INDIHOME," *J. Inform. ...*, vol. 2, no. 2, pp. 240–247, 2022, [Online]. Available: <https://ejournal.unama.ac.id/index.php/jakakom/article/view/33%0Ahttps://ejournal.unama.ac.id/index.php/jakakom/article/download/33/56>
- [16] M. Idris, "Implementasi Data Mining



- Dengan Algoritma Naïve Bayes Untuk Memprediksi Angka Kelahiran,” *J. Pelita Inform.*, vol. 7, no. 3, pp. 1–33, 2019, [Online]. Available: [https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/MT\\_Globalization\\_Report\\_2018.pdf%0Ahttp://eprints.lse.ac.uk/43447/1/India\\_globalisation%2C\\_society\\_and\\_inequalities%28%29.pdf%0Ahttps://www.quora.com/What-is-the](https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/MT_Globalization_Report_2018.pdf%0Ahttp://eprints.lse.ac.uk/43447/1/India_globalisation%2C_society_and_inequalities%28%29.pdf%0Ahttps://www.quora.com/What-is-the)
- [17] M. T. Nitamia and H. Februariyanti, “Analisis Sentimen Ulasan Ekpedisi J&T Ekspres Menggunakan Algoritma Naive Bayes,” *J. Manaj. Inform. Sist. Inf.*, vol. 5, no. 1, pp. 20–29, 2022.
- [18] R. P. Asyrofi, Reva Riani Putri; Affandy, Ali Nuke; Hermoyo, “Analisis Plagiarism Artikel Ilmiah Dalam Jurnal Terakreditasi Sinta,” *J. Pendidik. dan Konseling*, vol. 4, pp. 1349–1358, 2022.
- [19] C. D. Mait, J. A. Watuseke, P. D. G. Saerang, and S. R. Joshua, “Sistem Pendukung Keputusan Menggunakan Fuzzy Logic Tahani Untuk Penentuan Golongan Obat Sesuai Dengan Penyakit Diabetes,” *J. Media Infotama*, vol. 18, no. 2, p. 344, 2022.
- [20] S. Ratna, “Pengolahan Citra Digital Dan Histogram Dengan Phyton Dan Text Editor Phycharm,” *Technol. J. Ilm.*, vol. 11, no. 3, p. 181, 2020, doi: 10.31602/tji.v11i3.3294.
- [21] V. Atina and D. Hartanti, “Knowledge Based Recommendation Modeling for Clothing Product Selection Recommendation System,” *J. Tek. Inform.*, vol. 3, no. 5, pp. 1407–1413, 2022, doi: 10.20884/1.jutif.2022.3.5.584.
- [22] T. P. Utomo, B. P. Chandra, F. A. Pratama, and I. D. Fibrian, “Implementasi Metode Exponential Smoothing Pada Sistem Informasi Peramalan Stok Di PT Atlantic Biruraya Jombang,” *J. Manaj. Inform. Sist. Inf.*, vol. 5, no. 1, 2022.