

## Penerapan Metode Naive Bayes Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu di Program Studi Sistem Informasi Universitas Pamulang

Risah Subariah<sup>1</sup>, Afrizal Zein<sup>2</sup>

<sup>1</sup> Sistem Informasi, Universitas Pamulang, <sup>2</sup>Sistem Informasi, Universitas Pamulang  
e-mail : [dosen02695@unpam.ac.id](mailto:dosen02695@unpam.ac.id), [dosen01495@unpam.ac.id](mailto:dosen01495@unpam.ac.id)

### Abstrak

Penelitian ini menganalisis data historis menggunakan algoritma Naive Bayes untuk memprediksi berapa banyak siswa baru yang akan menyelesaikan pendidikan tepat waktu. Metode ini dipilih karena mampu mempelajari data kasus sebelumnya yang digunakan sebagai data uji. Dari sekian banyak sifat yang ditandai dalam penelitian ini, beberapa di antaranya adalah ujian nasional, seperti UN. Atribut lainnya termasuk jenis kelamin, sekolah dan daerah asal, pendapatan keluarga dan pekerjaan orang tua. Selain itu, jumlah saudara kandung yang dimiliki seseorang diperhitungkan. Dengan menganalisis data baru, seorang dosen dapat menentukan apakah mahasiswa memiliki tepat waktu untuk menyelesaikan studi atau tidak.

Penelitian ini telah menghasilkan sebuah aplikasi sistem pendukung keputusan dengan tingkat akurasi sebesar mendekati 85%.

**Kata Kunci:** Prediksi Kelulusan Tepat Waktu, Metode Naive Bayes, Data Mining

### Abstract

*This study analyzes historical data using the Naive Bayes algorithm to predict how many new students will complete education on time. This method was chosen because it is able to study previous case data used as test data. Of the many traits marked in this study, some of them are national exams, such as the National Examination. Other attributes include gender, school and area of origin, family income and parents' occupation. Additionally, the number of siblings a person has is taken into account. By analyzing the new data, a lecturer can determine whether students have time to complete their studies or not.*

*This research has produced a decision support system application with an accuracy rate of close to 85%.*

**Keywords:** Prediction of on Time Graduation, Naive Bayes Method, Data Mining

## 1. PENDAHULUAN

Lulus tepat waktu adalah Penilaian dalam Proses Akreditasi Institusi Universitas. Elemen Penilaian Lulusan salah satu perguruan tinggi adalah perguruan tinggi tinggi memiliki arti pendidikan yang ideal (BANPT, 2007).

Jadi kalau mahasiswa lulus tepat waktu akan membantu dalam penilaian akreditasi universitas. di samping itu, Siswa akan beruntung jika mereka lulus dengan benar tepat waktu karena tidak akan membayar uang kuliah tambahan dan bisa bekerja dengan cepat setelah kuliah.

Tetapi ketika seorang siswa tidak lulus selalu bisa dideteksi sejak dini sehingga bisa mengakibatkan kelulusan tertunda. Untuk mengatasi kebutuhan ini ada tekniknya kelulusan yang dapat diprediksi mahasiswa. Adapun teknik umum digunakan adalah penggunaan datamining dan metode yang sering digunakan Prediksi kelulusan mahasiswa adalah Klasifikasi.

Klasifikasi juga merupakan proses mengidentifikasi objek Dalam suatu kelas, kelompok atau kategori menurut prosedur, karakteristik dan definisi yang telah ditentukan (Setiawan, 2018). Klasifikasi adalah sebuah komponen Dalam penambangan data, sehingga data dapat disediakan Signifikan/dapat dieksploitasi. klasifikasi adalah Salah satu tugas terpenting dalam penambangan data, Dalam klasifikasi, classifier dibuat dari satu set data pelatihan dengan kategori Dijadwalkan.

Klasifikasi data adalah proses dua langkah. Langkah pertama adalah membangun model menggambarkan sekumpulan kelas data atau konsep populasi data ditentukan sebelumnya (seperti data akun surat kabar/pendapatan). bangunan model Analisis data pelatihan yang dijelaskan oleh Berdasarkan properti. Asumsikan bahwa setiap tupel termasuk dalam kategori tertentu, ditentukan oleh salah satu sifat, yaitu Disebut atribut label kelas. tahap kedua Hal ini untuk menguji model yang telah dibangun untuk menguji

data untuk mengukur akurasi atau Performa model saat mengklasifikasikan data uji. Setelah pengukuran kinerja selesai, maka pengambil keputusan dapat memutuskan gunakan model atau ulangi Buat model menggunakan data atau metode pelatihan model produksi yang berbeda Klasifikasi yang Lebih Baik (Setiawan, 2018).

Metode klasifikasi adalah pendekatan Mengkategorikan fungsi data. Data mining adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Teknologi Klasifikasi ini juga dapat digunakan untuk Membuat prediksi berdasarkan informasi yang tidak lengkap tahu sebelumnya. beberapa algoritma Salah satu hal yang dapat digunakan adalah algoritma Pohon Keputusan C.45, Jaringan Syaraf Tiruan (ANN), K-Nearest Neighbors (KNN), Algoritma Naive Bayes, Algoritma Genetika, Rough Set, pendekatan berbasis aturan, berbasis memori Mesin Vektor Inferensi dan Dukungan (Widodo et al, 2013).

## 2. TINJAUAN PUSTAKA

### 2.1 DATAMINING

Datamining adalah proses penggalian pola dan informasi berguna lainnya dari kumpulan data besar. Kadang-kadang dikenal sebagai penemuan pengetahuan dalam data atau KDD. Berkat munculnya data besar dan kemajuan teknologi pergudangan data, penggunaan teknik penambangan data telah berkembang dalam beberapa dekade terakhir, mengubah data mentah menjadi pengetahuan berharga yang dapat digunakan perusahaan. Meskipun teknologi telah maju untuk menangani kumpulan data yang substansial, para eksekutif masih menghadapi tantangan otomatisasi dan skalabilitas.

Penambangan data telah meningkatkan pengambilan keputusan perusahaan melalui analitik data yang cerdas. Teknik data mining dapat secara luas diklasifikasikan menjadi dua kategori:

- Menentukan kumpulan data target

- Peramalan hasil menggunakan metode pembelajaran mesin

Taktik ini digunakan untuk mengatur dan memfilter data – memberikan informasi paling penting, mulai dari deteksi penipuan hingga perilaku pengguna, hambatan, dan bahkan pelanggaran keamanan.

Masuk ke ranah penambangan data tidak pernah semudah ini, dan mengumpulkan wawasan yang bermakna tidak pernah secepat ini – terutama saat menggabungkan penambangan data dengan analitik data dan alat visualisasi seperti Apache Spark. Kemajuan kecerdasan buatan mempercepat adopsi teknik penambangan data di seluruh industri.

## 2.2 KLASIFIKASI

Klasifikasi dapat didefinisikan sebagai proses memprediksi kelas atau kategori dari nilai yang diamati atau titik data yang diberikan. Keluaran yang dikategorikan dapat memiliki bentuk seperti "Hitam" atau "Putih" atau "spam" atau "tanpa spam".

Secara matematis, klasifikasi adalah tugas untuk mengaproksimasi fungsi pemetaan ( $f$ ) dari variabel input ( $X$ ) ke variabel output ( $Y$ ). Ini pada dasarnya milik pembelajaran mesin yang diawasi di mana target juga disediakan bersama dengan set data input.

Contoh masalah klasifikasi dapat berupa deteksi spam di email. Hanya ada dua kategori keluaran, "spam" dan "tanpa spam"; karenanya ini adalah klasifikasi tipe biner.

Untuk mengimplementasikan klasifikasi ini, pertama-tama kita perlu melatih classifier. Untuk contoh ini, email "spam" dan "tanpa spam" akan digunakan sebagai data pelatihan. Setelah berhasil melatih classifier, itu dapat digunakan untuk mendeteksi email yang tidak dikenal.

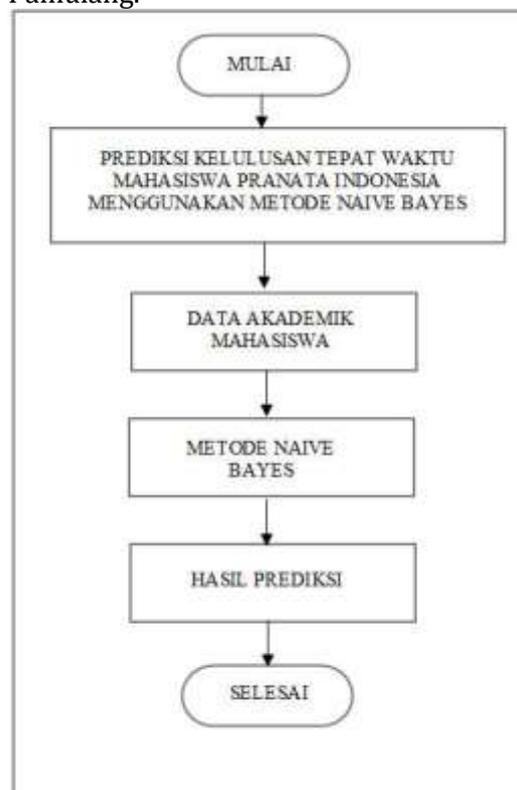
## 2.3 NAÏVE BAYES

Naive Bayes adalah pengklasifikasi pembelajaran mesin yang sederhana, namun efektif dan umum digunakan. Ini adalah pengklasifikasi probabilistik

yang membuat klasifikasi menggunakan aturan keputusan Maximum A Posteriori dalam pengaturan Bayesian. Itu juga dapat direpresentasikan menggunakan jaringan Bayesian yang sangat sederhana. Pengklasifikasi Naive Bayes sangat populer untuk klasifikasi teks, dan merupakan solusi tradisional untuk masalah seperti deteksi spam.

## 3. METODOLOGI PENELITIAN

Dalam tahap ini, penulis menggunakan metode klasifikasi yaitu Naive Bayes. Penulis mencoba menggunakan metode ini dalam memprediksi kelulusan mahasiswa Sistem Informasi Universitas Pamulang.



Gambar 1. Desain Penelitian

Berdasarkan desain penelitian Gambar 1. Prediksi Kelulusan Siswa Dilakukan dengan mengumpulkan data akademik Mahasiswa, lalu Data Akademik Siswa Preprocess dengan menggunakan metode naive bayes. **Kelebihan metode naive bayes :**

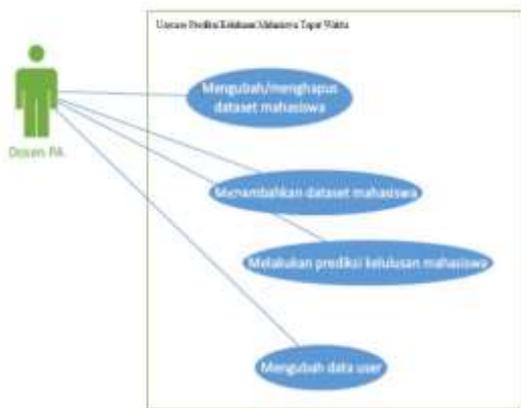
- Klasifikasi yang dilakukan dengan 40 data uji yang relative lebih cepat
- Tidak sensitive terhadap fitur yang tidak relevan
- Lebih mudah dipahami
- Pengkodean yang diperlukan relative sederhana

**Kekurangan metode naïve bayes :**

- Asumsi data yang diberikan adalah independent
- Metode bayes memerlukan pengetahuan awal dalam mengambil keputusan, tingkat keberhasilan metode ini tergantung seberapa baik pengetahuan awal yang diberikan.

**Perancangan Aplikasi**

Proses dalam perancangan aplikasi menjelaskan istilah-istilah tersebut Persyaratan Desain Perangkat Lunak dapat dihitung sebelum melanjutkan Fase implementasi. Didalam desain meliputi desain Fungsionalitas, non-fungsionalitas dan desain basis data. penggunaan desain fungsional Diagram UML termasuk diagram use case.

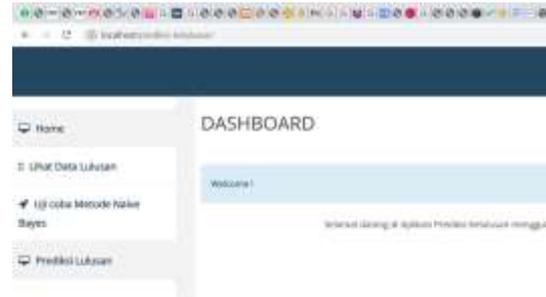


**Gambar 2.** UseCase Diagram Aplikasi Prediksi Mahasiswa Tepat Waktu

**4. HASIL DAN PEMBAHASAN**

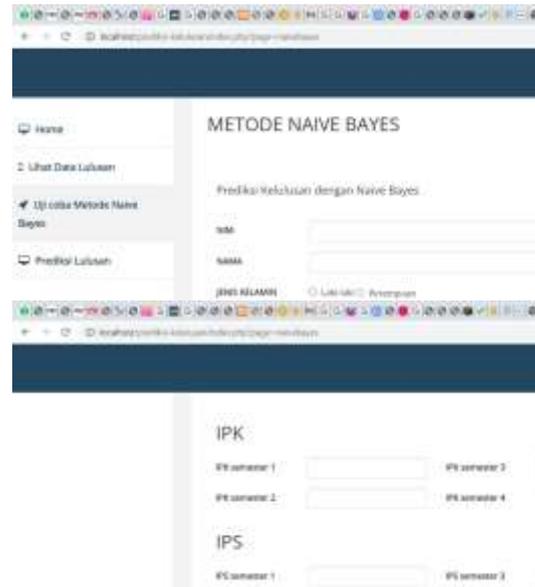
**Data Preparation**

Dalam penelitian ini digunakan 400 data mahasiswa Sistem Informasi Universitas Pamulang yang selanjutnya sebagai data data uji. Persentase distribusi data yang digunakan adalah 90:10 sehingga menghasilkan 153 data yang dijadikan data latih dan 40 data dipakai untuk data uji.



**Gambar 3.** Dashboard Aplikasi

Untuk input data dapat dilihat sebagai berikut:



**Gambar 4.** Input Data

Untuk hasil prediksi data dapat dilihat sebagai berikut:

**HASIL PREDIKSI**

NO	NIM	NAMA	JENIS KELAMIN	Status kelulusan
1.	20180940101	Nurhasni Nurhidayah	Wanita	Lulus Tepat Waktu
2.	18016081012	Nurhasni Nurhidayah	Wanita	Lulus Tepat Waktu
3.	20180940101	Dani Yulka Kania	Wanita	Lulus Tidak Tepat Waktu
4.	19160840101	Shafiqul Hafidz	Wanita	Lulus Tepat Waktu
5.	19160840104	Haniwa Azzah	Wanita	Lulus Tidak Tepat Waktu

**5. SIMPULAN**

Berdasarkan hasil analisis data dan menggunakan aplikasi naïve bayes dapat diambil kesimpulan dengan ketentuan sebagai berikut:

1. Penggunaan data mining dengan metode Naive Bayes dapat diterapkan

dalam melakukan prediksi kelulusan mahasiswa tepat waktu

2. Berdasarkan hasil Prediksi didapat akurasi untuk Kelulusan Tepat Waktu adalah mencapai 85%

#### DAFTAR PUSTAKA

**C, Bramer, Max.2017.** *Principles of Data mining.* Springer-Verlag. London..

**Huda, Nuqson M. 2018.** *Aplikasi Data mining untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa.* FMIPA Universitas Diponegoro. Semarang.

**Syahrin, Alvi. 2018.** *Implementasi Algoritma K,Means untuk Klasterisasi Mahasiswa Berdasarkan Prediksi Waktu Kelulusan.* UPN Veteran. Surabaya..

A. Indriani and D. Nbc, “Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier,” pp. 5–10, 2014.

**Nofriansyah, D (2016).** *Konsep Data Mining Vs Sistem Pendukung Keputusan.* Yogyakarta. Deepublish.

**Prasetyo, Eko. 2017.** *Data Mining Konsep dan Aplikasi Menggunakan MATLAB.* Yogyakarta: Andi.

**Widodo, P.P, dkk. 2016.** *Penerapan Data Mining Dengan Matlab.* Bandung. Rekayasa Sains.