

# Analisis Prediksi Masa Studi Mahasiswa Menggunakan Algoritma Naïve Bayes

**Fajar Edi Prabowo\*, Achmad Kodar\*\***

\*Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana  
Jl. Meruya Selatan, Kembangan, Jakarta 11650  
41515010079@student.mercubuana.ac.id

\*\*Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana  
Jl. Meruya Selatan, Kembangan, Jakarta 11650  
achmad.kodar@mercubuana.ac.id

## ABSTRACT

*Universitas Mercu Buana merupakan salah satu perguruan tinggi swasta dengan lebih dari 25.000 mahasiswa aktif pada tahun 2018. Salah satu program studi yang ada di Universitas Mercu Buana dan membutuhkan dukungan sarana dan prasarana yang baik adalah Teknik Informatika. Namun, pengadaan sarana perkuliahan yang memadai tidak mudah untuk dilakukan karena diperlukan biaya yang tidak sedikit. Berdasarkan permasalahan tersebut, penulis melakukan penelitian untuk menganalisa dan memprediksi lama masa studi mahasiswa sehingga dapat dijadikan acuan dalam pengadaan sarana dan prasarana perkuliahan. Metode yang digunakan adalah Naïve Bayes yang akan diimplementasikan pada data kelulusan mahasiswa Teknik Informatika Universitas Mercu Buana menggunakan RapidMiner. Setelah dilakukan pengujian, diperoleh nilai akurasi sebesar 82,26%. Hasil tersebut dapat dimanfaatkan sebagai strategi dalam meningkatkan kualitas pembelajaran di Universitas Mercu Buana.*

*Kata Kunci: Klasifikasi, Naïve Bayes, Prediksi, Kelulusan*

## PENDAHULUAN

Universitas adalah suatu institusi pendidikan tinggi yang dibangun untuk membentuk dan menguji mahasiswa dari berbagai bidang akademik dengan pembelajaran lebih lanjut. Dalam pelaksanaan kegiatan belajar mengajar yang baik, diperlukan sarana dan prasarana yang mumpuni untuk menunjang kegiatan perkuliahan. Namun, pengadaan sarana perkuliahan yang memadai tidak mudah untuk dilakukan karena diperlukan biaya yang tidak sedikit. Oleh karena itu, dukungan dalam pengambilan keputusan diperlukan untuk pengadaan sarana dan prasarana yang mumpuni.

Universitas Mercu Buana merupakan salah satu perguruan tinggi swasta dengan lebih dari 25.000 mahasiswa aktif pada tahun 2018. Salah satu program studi yang ada di Universitas Mercu Buana dan membutuhkan dukungan sarana dan prasarana yang baik adalah Teknik Informatika. Pengadaan

sarana dan prasarana perkuliahan yang efisien masih sulit untuk dilakukan karena belum ada acuan mengenai jumlah mahasiswa apabila ada mahasiswa yang sudah lulus maupun mahasiswa yang baru masuk. Berdasarkan permasalahan tersebut, diperlukan prediksi lama masa studi mahasiswa untuk memberikan dukungan dalam pengambilan keputusan.

Penelitian sebelumnya telah dilakukan untuk menjaga kualitas dan performa perkuliahan dengan melakukan prediksi pada data kelulusan mahasiswa (Artaye, 2015). Pada tahun 2017, Supardi Salmu dan Achmad Solichin melakukan penelitian untuk menemukan faktor-faktor yang mempengaruhi prediksi kelulusan mahasiswa melalui data kinerja akademik mahasiswa (Salmu & Solichin, 2017). Pada tahun yang sama, Deny Wiria Nugraha, dkk melakukan klasifikasi penyakit stroke menggunakan metode Naïve Bayes (Nugraha, Dodu, & Chandra, 2017). Pada penelitian ini, akan dilakukan prediksi

pada data kelulusan mahasiswa Teknik Informatika Universitas Mercu Buana menggunakan metode klasifikasi Naïve Bayes. Penelitian ini berkontribusi untuk memberikan dukungan pengambilan keputusan dalam meningkatkan kualitas pembelajaran di Universitas Mercu Buana.

## STUDI LITERATUR

### A. Klasifikasi

Klasifikasi adalah pemrosesan untuk menemukan sebuah model (atau fungsi) yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu yang bisa menggunakan pemodelan untuk memprediksi kelas objek yang label nya tidak diketahui (Amelia, Lumenta, & Jacobus, 2017).

### B. Naïve Bayes

Algoritma Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Naïve Bayes bertujuan untuk memprediksi kelas dari contoh kasus berdasarkan data training yang disediakan (Sharma, Sharma, & Mansotra, 2016).

Teorema Bayes memiliki bentuk umum sebagai berikut (Jananto, 2013):

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan:

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu kelas spesifik

$P(H|X)$  = Probabilitas hipotesis H berdasarkan kondisi X

$P(H)$  = Probabilitas hipotesis H

$P(X|H)$  = Probabilitas X berdasarkan kondisi tersebut

$P(X)$  = Probabilitas dari X

## METODOLOGI

Metodologi yang digunakan secara keseluruhan dapat dilihat pada Gambar 1. Terdapat beberapa tahap yaitu Pengumpulan Data, *Pre-processing* Data, Implementasi Algoritma dan yang terakhir adalah Evaluasi.



Gambar 1 Metodologi Penelitian

### A. Pengumpulan Data

Dataset yang digunakan pada penelitian ini adalah data kelulusan mahasiswa Teknik Informatika Universitas Mercu Buana selama 4 tahun (2011-2014) yang berjumlah 306 data.

Contoh dataset yang digunakan dapat dilihat pada tabel 1.

Tabel 1 Contoh Dataset

No	Atribut	Nilai
1.	Tahun Masuk	Int
2.	Jenis Kelamin	L P
3.	SKS Semester 1	Int
4.	SKS Semester 2	Int
5.	SKS Semester 3	Int
6.	SKS Semester 4	Int
7.	IPK Semester 4	Tinggi Rendah Sedang
8.	Pendidikan Asal	SMU SMEA
9.	Status Kelulusan	Cepat Tepat Terlambat

### B. *Pre-processing* Data

Jenis *Pre-processing* data yang digunakan adalah transformasi data. Nilai IPS dari semester 1 sampai 4 dihitung rata-rata untuk mendapatkan IPK semester 4. Kemudian data dirubah nilainya menjadi “Tinggi”, “Sedang”, atau “Rendah” berdasarkan hasilnya. Transformasi data dapat dilihat pada tabel 2.

$3 \leq \text{IPK} \leq 4$	Tinggi
$2 \leq \text{IPK} < 3$	Sedang
$\text{IPK} < 2$	Rendah

Dataset yang akan digunakan setelah dilakukan transformasi data dapat dilihat pada tabel 3.

Tabel 2 Transformasi Data

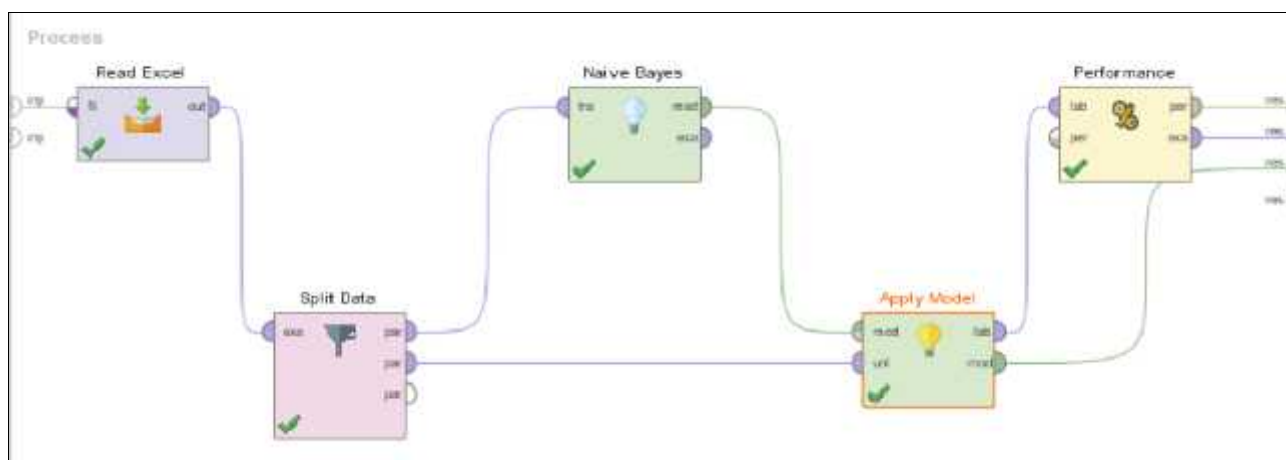
IPK Semester 4	Konversi
----------------	----------

Tabel 3 Dataset

Tahun Masuk	Jenis Kelamin	SKS 1	SKS 2	SKS 3	SKS 4	IPK	Asal Sekolah	Status Kelulusan
2011	P	18	24	24	23	Tinggi	SMU	Cepat
2011	P	18	24	20	23	Rendah	SMEA	Terlambat
2011	L	18	21	18	0	Rendah	SMU	Terlambat
2011	L	18	24	24	23	Sedang	SMU	Tepat
2011	L	18	21	20	22	Tinggi	SMEA	Tepat

### C. Implementasi Algoritma

Pada tahap ini, akan dilakukan implementasi algoritma Naïve Bayes pada data kelulusan mahasiswa tahun 2011 sampai 2014 menggunakan RapidMiner. Proses implementasi algoritma Naïve Bayes dapat dilihat pada gambar 2.



Gambar 2 Operator RapidMiner

- **Read Excel**  
Operator ini digunakan untuk membaca data dari Microsoft Excel di RapidMiner. Pada operator ini user dapat memilih data dan menentukan atribut masing-masing data yang akan diuji.
- **Split Data**  
Split Data digunakan untuk membagi data berdasarkan perbandingan yang diperlukan user. Pada penelitian ini data dibagi menjadi 2 dengan perbandingan 8:2. Dari 306 data, 244 data akan digunakan sebagai data training dan 62 data akan digunakan sebagai data testing.
- **Naïve Bayes**  
Operator ini digunakan untuk mengaplikasikan algoritma Naïve Bayes pada data training yang diambil dari Split Data. Operator ini mencari

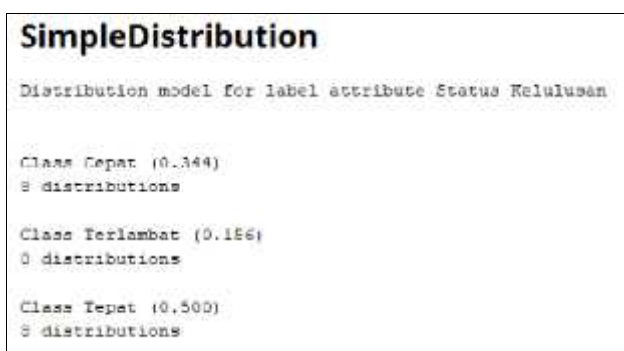
pola pada data training untuk selanjutnya digunakan pada data testing.

- **Apply Model**  
 Operator ini digunakan untuk mengimplementasikan model yang sudah diolah pada data training. Data yang diimplementasikan pada model memiliki atribut yang sama dengan model yang dihasilkan.
- **Performance**  
 Performance Operator adalah operator evaluasi yang dapat digunakan untuk mengukur performa model. Operator ini akan menentukan jenis pemodelan secara otomatis dan menghitung performa dari model tersebut.

**D. Evaluasi**  
 Proses evaluasi pada penelitian ini akan dilakukan menggunakan Confusion Matrix. Evaluasi diperlukan untuk mengukur tingkat akurasi penerapan algoritma terhadap dataset.

## HASIL DAN DISKUSI

Metode Naïve Bayes mencari pola pada data training untuk diterapkan pada data testing. Prediksi kelulusan mahasiswa ditentukan dengan atribut Status Kelulusan dengan nilai Cepat, Tepat, dan Terlambat. Probabilitas posterior data testing yang dihasilkan pada metode Naïve Bayes menggunakan RapidMiner dapat dilihat pada Gambar 3.



Gambar 3 Probabilitas Posterior

Dari gambar di atas dapat diketahui bahwa untuk class Cepat (Mahasiswa lulus dibawah 4 tahun) memiliki probabilitas posterior 0,344, untuk class Terlambat (Mahasiswa lulus diatas 4 tahun) memiliki probabilitas posterior 0,156, sedangkan

untuk class Tepat (Mahasiswa lulus 4 tahun) memiliki probabilitas posterior 0,5.

Berdasarkan data yang digunakan, dapat dilakukan perhitungan akurasi menggunakan Confusion Matrix. Confusion Matrix adalah tool yang berguna untuk menganalisis seberapa baik klasifikasi yang dilakukan dari kelas-kelas yang berbeda (Han, Kamber, & Pei, 2012).

Tabel 4 Confusion Matrix

	true Cepat	True Terlambat	True Tepat	class precision
pred. Cepat	21	1	7	72,41%
pred. Terlambat	0	7	1	87,50%
pred. Tepat	0	2	23	92%
class recall	100%	70%	74,19%	

$$Accuracy = \left( \frac{(21 + 7 + 23)}{62} \right) * 100\% = 82,26\%$$

Berdasarkan perhitungan didapatkan nilai akurasi sebesar 82,26%. Hal ini menunjukkan bahwa algoritma Naïve Bayes dapat digunakan untuk memprediksi lama masa studi mahasiswa.

## KESIMPULAN

Berdasarkan hasil pengujian dapat disimpulkan bahwa algoritma Naïve Bayes bisa digunakan untuk memprediksi lama masa studi mahasiswa. Data yang digunakan adalah data kelulusan mahasiswa tahun 2011 hingga 2014 dengan 244 data sebagai data training dan 62 data sebagai testing. Atribut yang digunakan antara lain tahun masuk, jenis kelamin, sks semester, IPK dan asal sekolah mahasiswa. Perhitungan akurasi yang dilakukan menghasilkan nilai sebesar 82,26%. Hasil penelitian yang diperoleh dapat dijadikan sebagai pendukung pengambilan keputusan dalam peningkatan sarana dan prasarana untuk meningkatkan efektifitas kegiatan belajar mengajar di Universitas Mercu Buana.

## DAFTAR PUSTAKA

- Amelia, M. W., Lumenta, A. S. M., & Jacobus, A. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes, *11*(1).
- Artaye, K. (2015). Implementation of Naïve Bayes Classification Method to Predict Graduation Time of IBI Darmajaya Scholar, (August), 284–290.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, *itd: Morgan Kaufmann*.
- Jananto, A. (2013). Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa, *18*(1), 9–16.
- Nugraha, D. W., Dodu, A. Y. E., & Chandra, N. (2017). Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes Classifier (Studi Kasus Pada Rumah Sakit Umum Daerah Undata Palu), *3*(2), 13–22.
- Salmu, S., & Solichin, A. (2017). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta, (April), 701–709.
- Sharma, T., Sharma, A., & Mansotra, P. V. (2016). Performance Analysis of Data Mining Classification Techniques on Public Health Care Data, 11381–11386.