

Terbit : 23 Desember 2024

# Penerapan Algoritma Naïve Bayes Untuk Prediksi Waktu Kelulusan Mahasiswa

Muhammad Tafsir  
STMIK Citra Mandiri Padangsidempuan  
[tansaku001@gmail.com](mailto:tansaku001@gmail.com)

## ABSTRAK

Administrator universitas memerlukan perkiraan kelulusan dalam menentukan langkah-langkah untuk mencegah keterlambatan kelulusan mahasiswa dari masa awal masa studi. Lamanya waktu yang dihabiskan seorang mahasiswa untuk belajar bergantung pada banyak faktor. Oleh karena itu, penting untuk mengetahui dan mendalami siapa mahasiswa yang tidak dapat lulus dengan tepat waktu atau terlambat. Teknik penambangan data dapat digunakan untuk mengeksplorasi dan mengaktifkan pengetahuan baru serta membuat prediksi tentang kelulusan mahasiswa. Dalam memprediksi kelulusan mahasiswa algoritma digunakan salah satunya ialah algoritma Naïve Bayes. Penelitian berikut bertujuan memprediksi atau menyelidiki dalam konteks kelulusan mahasiswa pada program studi Pendidikan Teknik Informatika & Komputer Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi menggunakan metode algoritma *Naïve Bayes*. Penelitian ini memanfaatkan data kelulusan mahasiswa pada tahun 2019-2020 sebanyak 131 data yang diambil dari Teknologi Informasi dan Pangkalan Data Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi. Data tersebut merupakan data mentah yang belum dilakukan *preprocessing*. Setelah proses *cleaning* data dilakukan, didapatkan 123 data mahasiswa yang akan diolah ke dalam tahap selanjutnya dengan atribut jenis kelamin, asal sekolah, indeks prestasi 1 sampai 8, ipk dan sks. Data kelulusan mahasiswa tersebut dibagi kedalam data *training* 103 dan *testing* 20 data. Adapun atribut yang menjadi acuan ialah, jenis kelamin, indeks prestasi 1 sampai 8, ipk, dan sks. Dari penelitian yang dilakukan menggunakan metode *Naïve Bayes* dan dibuktikan menggunakan *tool RapidMiner 9.10* menunjukkan hasil akurasi sebesar 95%. Dengan *class precision* 100%, dan *class recall* 80%. Sehingga dari nilai akurasi tersebut menunjukkan bahwa metode *Naïve Bayes* sangat baik dalam melakukan prediksi kelulusan mahasiswa. Serta hasil yang diprediksi pada mahasiswa lulusan yang terlambat diharapkan akan dievaluasi sedemikian rupa sehingga menghasilkan lulusan yang tiba tepat waktu.

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

## PENDAHULUAN

Perguruan tinggi adalah program pendidikan Pendidikan yang memiliki beban studi sekurang-kurangnya 144 sks (satuan kredit semester), yang dijadwalkan untuk 8 semester. Sehingga dapat ditempuh minimal 7 semester dan paling lama 14 semester [1]. Perguruan tinggi memiliki kewajiban untuk menghasilkan lulusan yang kompeten. Hal tersebut dapat dinilai dari tingkat kelulusan mahasiswanya. Selain itu, lulus tepat waktu juga menjadi impian para mahasiswa. Mahasiswa tidak perlu membayar biaya kuliah lagi dan bisa bekerja lebih cepat. Tetapi kondisi di lapangan, mahasiswa belum tentu dapat menuntaskan masa studi tepat pada waktunya. Banyak faktor yang menjadi pengaruh kelulusan mahasiswa terlambat, seperti status perkawinan mahasiswa, status mahasiswa (bekerja/tidak bekerja), tingkat pemahaman mahasiswa terhadap materi kuliah yang dapat dilihat dari indeks prestasi kumulatif (IPK) mahasiswa. Selain itu, tingkat kelulusan juga berpengaruh pada akreditasi suatu perguruan tinggi sehingga perguruan tinggi berusaha untuk membantu mahasiswa agar lulus tepat waktu[2]. Setiap perguruan tinggi tentunya memiliki *database* untuk menyimpan semua data mahasiswa disetiap akademik, data mahasiswa tersebut terus menerus bertambah disetiap tahun dan menumpuk seperti data yang terabaikan karena data tersebut tidak dimanfaatkan atau diolah kembali[3].

Secara umum, tujuan dari prediksi kelulusan mahasiswa adalah untuk menganalisa data dan merancang model menggunakan algoritma *Naïve Bayes* dapat memprediksi waktu kelulusan mahasiswa. Penelitian-penelitian tentang prediksi kelulusan mahasiswa menggunakan algoritma *Naïve Bayes* telah banyak dilakukan. Beberapa di antaranya penelitian Perez dan Perez (2021), memiliki dampak yang signifikan terhadap kelulusan khususnya untuk administrator universitas karena dapat menjadi alat untuk mengidentifikasi mahasiswa yang akan menyelesaikan kuliah berdasarkan variabel yang termasuk dalam model [4]. Noronha, *et al.* (2023) menggunakan aplikasi *Rapid Miner* yang menghasilkan data klasifikasi untuk evaluasi kelulusan setiap mahasiswa [5]. Matar, *et al.* (2022) menggunakan pendekatan *Naïve Bayes classification* untuk membantu mengurangi ketidakpastian penerimaan dan meningkatkan kemampuan pengambilan keputusan pada universitas[6]. Nurul Khasanah dkk (2022) menggunakan metode *Naïve Bayes* dengan jumlah data 379, sehingga hasil yang diperoleh adalah kategori *good classification*[7]. Mohammed Ali, *et al.* (2020) *Naïve Bayes* digunakan untuk membantu mengidentifikasi putus sekolah dan memberikan nasihat atau konseling bagi manajemen pendidikan untuk mempertimbangkan dan merestrukturisasi kurikulum Pendidikan[8]. Robi Sepriansyah, *et al.* (2023) menggunakan *Naïve Bayes* dengan data training 151 *record* dan data *testing* 120 *record*. Maka dapat mempengaruhi hasil pengujian yang dilakukan dalam memprediksi kelulusan mahasiswa[9]. Sri Wahyuni dan Adinda (2021) dengan metode algoritma *Naïve Bayes* dengan 129 sampel data latih dan 10 data uji. Hasil penelitian tersebut digunakan sebagai langkah untuk menghindari penurunan kelulusan mahasiswa setiap tahunnya[10]. Royan Habibie Sukarna dan Yulian Ansori (2021) dengan metode *Naïve Bayes* ditemukan pola dalam menganalisis dan memprediksi kelulusan mahasiswa tepat waktu dengan atribut *gain ration* dan tool *Weka*[11]. Neni Purwati dan Agnes Dwi Januanti (2021), *Naïve Bayes* diterapkan untuk prediksi kelulusan mahasiswa sehingga mahasiswa dapat memprediksi kelulusan tepat/terlambat sejak dini[12]. Selanjutnya Muryan dkk (2022) prediksi kelulusan mahasiswa dengan mengoptimalkan parameter *Naïve Bayes* dengan data trainig 41 dan testing 25 data [13]. MT Sembiring dan RH Tambunan (2021) dengan metode *Naïve Bayes* prediksi kelulusan mahasiswa dengan pengujian 173 sampel data dan memiliki akurasi kecocokan[14]. Hartatik dkk (2021) *Naïve Bayes* untuk prediksi kelulusan mahasiswa dengan parameter indeks prestasi semester 1 sampai dengan 4[15]. Godinez dan Lomibao (2022), algoritma *Naïve Bayes* memprediksi kecenderungan prokatinasi akademik dimana berhasil memprediksi perilaku penundaan siswa dikemudian hari, apakah mereka non penunda, rendah, sedang atau penunda tinggi [16]. Olalekan, *et al.* (2020) metode *Naïve Bayes* dapat membantu penyelenggara Pendidikan untuk mengidentifikasi siswa yang berisiko tinggi lulus lebih awal[17]. Armansyah dan Rahkmat Kurniawan Rahmli (2022) metode *Naïve Bayes* digunakan untuk prediksi kelulusan mahasiswa dengan 44 data lulusan tahun 2020 dengan *Rapid Miner* dan 6 data untuk pengujian. Hasil penelitian adalah sangat baik[18]. Rovidatul Hikmah Tanjung dkk (2023) algoritma yang digunakan *C4.5* dan *Naïve Bayes* serta atribut yang digunakan yaitu umur masuk kuliah, jenis kelamin, indeks prestasi semester 1-4[18].

## TINJAUAN PUSTAKA

### Prediksi

Prediksi/*forecasting* Merupakan salah satu metode yang digunakan untuk memprediksi atau meramalkan suatu hal yang belum pernah terjadi. Prediksi merupakan suatu kegiatan meramalkan secara terstruktur mengenai sesuatu yang mungkin terjadi dimasa akan datang yang bersumber pada informasi dari masa lalu dan sekarang yang dimiliki, agar kesalahannya dapat diperkecil (Lusi Suryadi dkk, 2022).

### *Naïve Nayes*

Algoritma *Naïve Bayes* merupakan satu di antara algoritma penggalian data dan pengklasifikasi statistik sebuah klasifikasi berpeluang mudah yang menerapkan teorema *Bayes* dengan melalui asumsi antar variabel-variabel ketidak ketergantungan yang luhur. Algoritma *Gaussian Naïve Bayes* adalah varian dari *Naive Bayes*, ini adalah algoritma pembelajaran mesin probabilistik yang digunakan untuk banyak fungsi klasifikasi dan didasarkan pada teorema *Bayes* dan memiliki asumsi kuat bahwa prediktor harus *independent* satu sama lain (Elsharkawy, *et al.*

2022).

### Data Mining

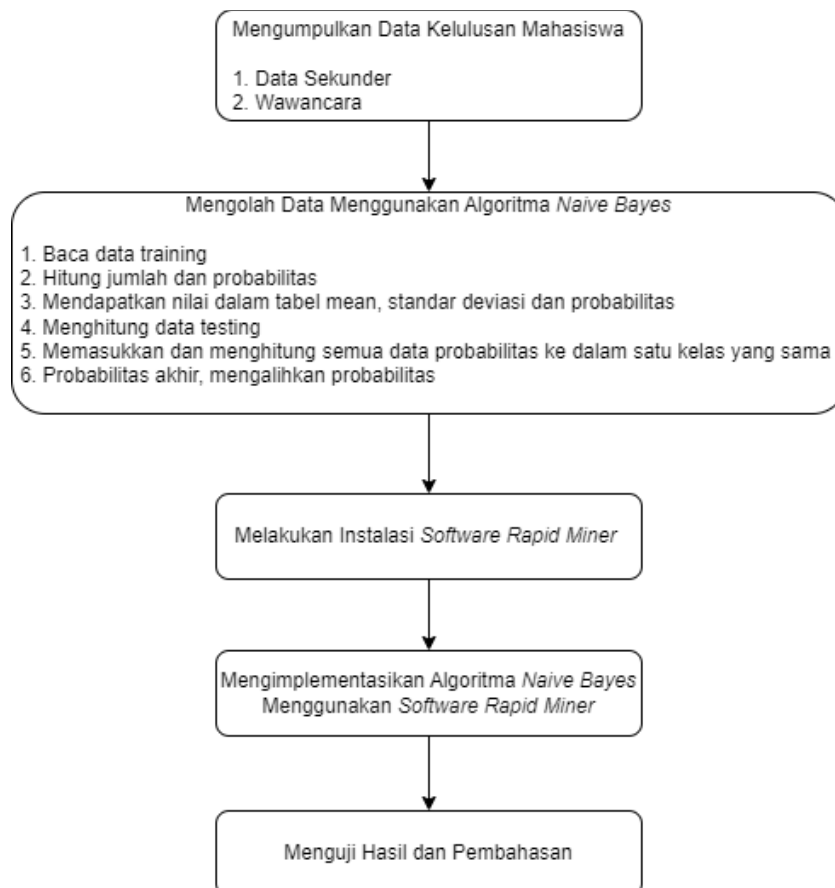
*Data Mining* merupakan sebuah proses dalam mengekstraksi sebuah data yang sebelumnya bersifat implisit, tidak terstruktur menjadi sebuah informasi dan pengetahuan atau pola data terstruktur yang menjadi lebih berguna (Witen, Ian H. Frank). *Data Mining* sendiri merupakan bagian dari *Knowledge Discovery in Databases* (KDD). *Knowledge Discovery* merupakan pencarian informasi atau pola tertentu yang tersembunyi dalam suatu data (Hilda Nuraliza dkk, 2022).

### Knowledge Discovery in Database

*Knowledge Discovery In Database* (KDD) adalah sekumpulan proses untuk memilih pengetahuan yang bermanfaat dari data. *Knowledge Discovery In Database* (KDD) yang berasal dari serangkaian langkah perubahan, termasuk data *preprocessing* serta *post processing*. Data *preprocessing* merupakan dari selangkah perubahan untuk megubah data mentah menjadi format yang *sinkron* untuk tahap analisa berikutnya. Selain itu data *preprocessing* juga digunakan untuk membantu dalam pengenalan atribut dan data yang relevan dengan *task Data Mining*. *Data Mining* dan *Knowledge Discovery in Database* seringkali dipergunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Gurita Wijaya dan Khaerul Anam, 2023).

### METODE PENELITIAN

Kerangka kerja penelitian merupakan gambaran urutan langkah-langkah yang dilakukan oleh peneliti dalam menganalisa dan menyelesaikan masalah yang akan diteliti. Penelitian ini menggunakan proses metode *Naïve Bayes* dengan tujuan untuk mendapatkan hasil yang optimal dan akurasi yang tinggi. Adapun tahapan penelitian dapat disajikan pada gambar 1 dibawah.



Gambar 1. Tahapan Proses Penelitian

1. Mengumpulkan Data  
Adapun tahapan mengumpulkan data, peneliti melakukan penghimpunan data lulusan mahasiswa tahun 2019 sampai dengan 2020 kepada pihak Teknologi Informasi dan Pangkalan Data (TIPD) Universitas Sjech M. Djamil Djambek Bukittinggi serta wawancara guna mendapatkan informasi mengenai data kelulusan mahasiswa.
2. Mengolah Data Menggunakan Algoritma *Naïve Bayes*  
Adapun langkah-langkah dalam pengolahan data sebagai berikut :
  - a. Baca data *training*.
  - b. Hitung jumlah dan probabilitas.
    1. Jika terdapat data *numerik*, maka temukan nilai *mean* dan standar deviasi dari masing-masing parameter yang menggambarkan data angka.
    2. Jika tidak, hitung nilai probabilitas tiap kategori yang sama.
  - c. Mendapatkan nilai dalam tabel *mean*, standar deviasi dan probabilitas.
  - d. Menghitung data *testing*.
  - e. Probabilitas akhir tiap kelas, memasukkan dan menghitung semua data probabilitas ke dalam satu kelas yang sama.
  - f. Probabilitas akhir, mengalihkan probabilitas.
3. Melakukan Instalasi *Software Rapid Miner*  
*Rapid Miner* adalah sebuah aplikasi atau *software* yang berfungsi dalam pengolahan *Data Mining*. Versi *software* yang akan digunakan adalah 9.10.
4. Mengimplementasikan *Software Rapid Miner* Menggunakan Algoritma *Naïve Bayes*  
Setelah proses perhitungan manual selesai, langkah selanjutnya ialah melakukan tahapan implementasi menggunakan *software Rapid Miner* 9.10 dengan algoritma *Naïve Bayes* dengan spesifikasi *software* dan *hardware* sebagai berikut :
  - a. *Hardware*
    1. *Lenovo* Laptop V110.
    2. *Processor* AMD A9-9420 RADEON R5, 5 COMPUTE CORES 2C+3G 3.00 GHz.
    3. *Hard Disk* 500 GB.
    4. *RAM* 4 GB
  - b. *Software*
    1. *Windows* 10 Pro.
    2. *Software Rapid Miner* 5.3.
    3. *Microsoft Excel*.
    4. *Microsoft Office* 2010.
5. Menguji Hasil dan Pembahasan  
Dalam tahapan ini, selanjutnya dilakukan proses menguji hasil baik dari perhitungan secara manual serta menggunakan *software Rapid Miner* 9.10 dengan menerapkan formula algoritma *Naïve Bayes*. Berikut tahapan-tahapan dalam proses pengujian :
  - a. Melakukan proses pengolahan data secara manual menggunakan *software Microsoft Excel* dengan algoritma *Naïve Bayes*.
  - b. Melakukan pengujian pada *software Rapid Miner* menggunakan algoritma *Naïve Bayes* dengan data yang sama.
  - c. Melakukan perbandingan hasil pengujian manual dengan aplikasi untuk validasi hasil.
  - d. Langkah terakhir adalah mengukur akurasi hasil.

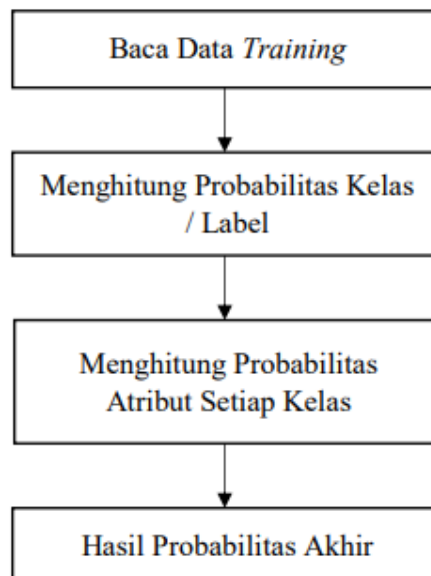
### HASIL DAN PEMBAHASAN

Data merupakan informasi yang diperoleh melalui pengamatan atau pencarian dari beberapa sumber. Sumber data penelitian kelulusan mahasiswa ini diperoleh dari Teknologi Unit dan Pangkalan Data (TIPD) Universitas Negeri Sjech M. Djamil Djambek Bukittinggi 2019-2020 sebanyak 131 data. Data tersebut merupakan data mentah yang belum dilakukan preprocessing dengan atribut nim, nama mahasiswa, jenis kelamin, tempat lahir, tanggal lahir, asal sekolah, program studi, ips 1, ips 2, ips 3, ips 4, ips 5, ips 6, ips 7, ips 8, ipk, sks dan lama studi. Dalam memperoleh data tersebut peneliti melakukan observasi langsung pada Teknologi Unit dan

Pangkalan Data (TIPD). Dalam penelitian ini, data training yang digunakan sebanyak 103 data, dan 20 data testing.

Data mentah yang dikumpulkan selanjutnya diolah menggunakan *Microsoft Excel* untuk dilakukan *cleaning* terhadap data yang akan diperlukan dalam penelitian. Hasil dari pembersihan data tersebut, selanjutnya dilakukan transformasi pada data yang bersifat numerik sesuai dengan ketentuan parameter yang digunakan. Setelah semua proses data tersebut dilakukan, maka data siap untuk diolah menggunakan algoritma *Naïve Bayes*. Data yang dikumpulkan masih banyak yang berisikan *null* sehingga pada tahap selanjutnya dilakukan pre-processing. Pada tahap pre-processing meliputi proses *cleaning* data dan transformasi data. Proses *cleaning* data dilakukan untuk atribut yang tidak digunakan seperti : nim, tempat lahir, program studi dan masa studi. Sedangkan untuk transformasi data dilakukan untuk merubah numerik menjadi kategorikal sehingga memudahkan dalam proses penambangan dan pengolahan data menggunakan algoritma *Naïve Bayes*. Setelah proses *cleaning* data selesai dilakukan, didapatkan 123 data mahasiswa yang akan diolah ke dalam tahap selanjutnya dengan atribut jenis kelamin, asal sekolah, indeks prestasi 1 sampai 8, ipk serta sks. Setelah proses *cleaning* data selesai, maka selanjutnya yaitu melakukan transformasi data.

Transformasi data dilakukan bertujuan untuk mengubah jenis data yang sifatnya numerik menjadi kategorikal agar dapat memudahkan dalam penambangan data menggunakan algoritma *Naïve Bayes*. Sebelum data diolah menggunakan algoritma *Naïve Bayes*, data terlebih dahulu ditransformasikan. Proses transformasi dilakukan pada beberapa atribut data yang bersifat numerik yaitu ips 1 sampai dengan ips 8, ipk dan sks. Sebelum data diolah menggunakan algoritma *Naïve Bayes*, data dikelompokkan terlebih dahulu menjadi dua bagian yaitu data training dan testing. Pembagian data tersebut berdasarkan data yang telah melalui tahap proses *cleaning* dan transformasi data. Tujuan dari pembagian data adalah untuk menentukan data yang akan digunakan sebagai data *training* dan data testing dalam proses menentukan hasil prediksi waktu kelulusan mahasiswa menggunakan algoritma *Naïve Bayes*. Adapun data training yang digunakan pada penelitian ini sebanyak 103 data dan data testing sebanyak 20 data. Data *training* dan testing meliputi atribut jenis kelamin, asal sekolah, ips1 sampai dengan 8, ipk dan sks. Data training tersebut akan digunakan untuk melatih algoritma dalam mencari model yang diinginkan. Pada data testing tersebut juga telah selesai melalui tahapan proses *cleaning* dan transformasi data. Sehingga memudahkan dalam proses pengolahan data menggunakan algoritma *Naïve Bayes*. Data testing ini akan dilakukan pengujian menggunakan algoritma *Naïve Bayes* untuk melihat model dari data tersebut berfungsi dengan benar dengan data sebelumnya yang tidak terlihat. Algoritma *Naïve Bayes* merupakan salah satu metode klasifikasi perhitungan probabilitas dan statistik. Berikut langkah-langkah penyelesaian yang dilakukan dalam menentukan prediksi waktu kelulusan mahasiswa menggunakan algoritma *Naïve Bayes*.



Gambar 2. Tahapan Alur Proses Algoritma *Naïve Bayes*

A. Baca data training

Terdapat 103 data training. Sedangkan untuk data testing sebanyak 20 data.

B. Menghitung Probabilitas Kelas

Berdasarkan data dengan label rincian data Tepat sebanyak 60 dan Terlambat sebanyak 43. Pada tahap ini dilakukan untuk menghitung jumlah probabilitas keseluruhan. Adapun perhitungan menggunakan rumus

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

Jumlah data label “Tepat”  $60 / 103 = 0,582524$

Jumlah data label “Terlambat”  $43 / 103 = 0,417476$

C. Menghitung Probabilitas atribut setiap kelas

Pada tahap ini dilakukan proses perhitungan pada atribut setiap kelas yang digunakan dalam penelitian ini. Adapun pada masing-masing atribut kelas tersebut dilakukan proses perhitungan terhadap jumlah label Tepat dan jumlah label Terlambat, sebagai berikut :

a. Jenis Kelamin Kelamin

Atribut kelas “Tepat”

1) Laki-laki =  $22 / 60 = 0,366667$  2)

2) Perempuan =  $40 / 60 = 0,666667$

Atribut kelas “Terlambat”

1) Laki-laki =  $27 / 43 = 0,627907$  2)

2) Perempuan =  $14 / 43 = 0,325581$

b. Atribut Asal sekolah

Atribut kelas SMK

1) SMK Tepat =  $18 / 60 = 0,3$  2)

2) SMK Terlambat =  $13 / 43 = 0,302326$

Atribut kelas SMA

1) SMA Tepat =  $31 / 60 = 0,516667$  2)

2) SMA Terlambat =  $19 / 43 = 0,44186$

Atribut kelas MA

1) MA Tepat =  $12 / 60 = 0,2$  2)

2) MA Terlambat =  $10 / 43 = 0,232558$

c. Atribut IPS 1

Atribut kelas IPS 1

1) Tinggi Tepat =  $12 / 60 = 0,2$  2)

2) Tinggi Terlambat =  $3 / 43 = 0,069767$

Atribut kelas IPS 1 Cukup

1) Cukup Tepat =  $49 / 60 = 0,816667$

2) Cukup Terlambat =  $39 / 43 = 0,906977$

Atribut kelas IPS 1

1) Kecil Tepat =  $0 / 60 = 0$

2) Kecil Terlambat =  $0 / 43 = 0$

d. Atribut IPS 2

Atribut kelas IPS 2

1) Tinggi Tepat =  $0 / 60 = 0$

2) Tinggi Terlambat =  $0 / 43 = 0$

Atribut kelas IPS 2

1) Cukup Tepat =  $61 / 60 = 1,016667$

$$2) \text{ Cukup Terlambat} = 34 / 43 = 0,837209$$

Atribut kelas IPS 2

$$1) \text{ Kecil Tepat} = 0 / 60 = 0$$

$$2) \text{ Kecil Terlambat} = 6 / 43 = 0,139535$$

e. Atribut IPS 3

Atribut kelas IPS 3

$$1) \text{ Tinggi Tepat} = 19 / 60 = 0,316667$$

$$2) \text{ Tinggi Terlambat} = 5 / 43 = 0,116279$$

Atribut kelas IPS 3

$$1) \text{ Cukup Tepat} = 42 / 60 = 0,7$$

$$2) \text{ Cukup Terlambat} = 32 / 43 = 0,744186$$

Atribut kelas IPS 3

$$1) \text{ Kecil Tepat} = 0 / 60 = 0$$

$$2) \text{ Kecil Terlambat} = 5 / 43 = 0,116279$$

#### D. Menghitung Probabilitas Kelas

Perhitungan data testing yang akan dilakukan yaitu dengan mengambil 20 data testing dari 123 dataset. Data testing merupakan data asli dari dataset, maka untuk menguji akurasi algoritma *Naïve Bayes* dari data tersebut dilakukan perhitungan menggunakan metode algoritma *Naïve Bayes*. Kemudian hasil dari proses perhitungan tersebut nantinya dibandingkan dengan data asli sebenarnya. Sehingga akan dapat dihitung tingkat akurasi dari metode algoritma *Naïve Bayes*. Proses perhitungan metode *Naïve Bayes* ini, menggunakan tool *Microsoft Excel* yaitu menguji 20 data testing. Berikut ini adalah contoh beberapa perhitungan dari data testing ke 1 sampai dengan data testing ke 3 :

Data testing ke 1

Adapun data testing yang pertama mempunyai atribut : Jenis Kelamin = Laki-laki, Asal Sekolah = SMK, IPS 1 = Cukup, IPS 2 = Cukup, IPS 3 = Cukup, IPS 4 = Tinggi, IPS 5 = Cukup, IPS 6 = Cukup, IPS 7 = Tinggi, IPS 8 = Tinggi, IPK = Sangat Memuaskan, SKS = Tuntas. Selanjutnya dilakukan pengujian pada semua atribut tersebut. Pengujian hasil "Tepat" =  $(0,366 * 0,3 * 0,816 * 1,016 * 0,7 * 0,583 * 0,5 * 0,183 * 1,016 * 0,666 * 0,4 * 1,016) = 0,00093$  Pengujian hasil "Terlambat" =  $(0,627 * 0,302 * 0,906 * 0,837 * 0,744 * 0,139 * 0,581 * 0,627 * 0,86 * 0,325 * 0,813 * 0,976) = 0,001199725$

Data testing ke 2

Adapun data testing yang kedua mempunyai atribut : Jenis Kelamin = Laki-laki, Asal Sekolah = SMK, IPS 1 = Cukup, IPS 2 = Cukup, IPS 3 = Cukup, IPS 4 = Tinggi, IPS 5 = Tinggi, IPS 6 = Tinggi, IPS 7 = Tinggi, IPS 8 = Tinggi, IPK = Sangat Memuaskan, SKS = Tuntas. Selanjutnya dilakukan pengujian pada semua atribut tersebut. Pengujian hasil "Tepat" =  $(0,366 * 0,3 * 0,816 * 1,016 * 0,7 * 0,583 * 0,583 * 0,833 * 1,016 * 0,666 * 0,4 * 1,016) = 0,00496$  Pengujian hasil "Terlambat" =  $(0,627 * 0,302 * 0,906 * 0,837 * 0,744 * 0,139 * 0,139 * 0,302 * 0,86 * 0,325 * 0,813 * 0,976) = 0,000138248$

Data testing ke 3

Adapun data testing yang ketiga mempunyai atribut : Jenis Kelamin = Laki-laki, Asal Sekolah = SMK, IPS 1 = Cukup, IPS 2 = Cukup, IPS 3 = Cukup, IPS 4 = Cukup, IPS 5 = Tinggi, IPS 6 = Tinggi, IPS 7 = Tinggi, IPS 8 = Tinggi, IPK = Pujian, SKS = Tuntas. Selanjutnya dilakukan pengujian pada semua atribut tersebut. Pengujian hasil "Tepat" =  $(0,366 * 0,3 * 0,816 * 1,016 * 0,7 * 0,433 * 0,583 * 0,833 * 1,016 * 0,666 * 0,4 * 1,016) = 0,00093$

$0,616 * 1,016) = 0,005675$  Pengujian hasil “Terlambat” =  $(0,627 * 0,302 * 0,906 * 0,837 * 0,744 * 0,813 * 0,139 * 0,302 * 0,86 * 0,325 * 0,093 * 0,976) = 0,00009249$

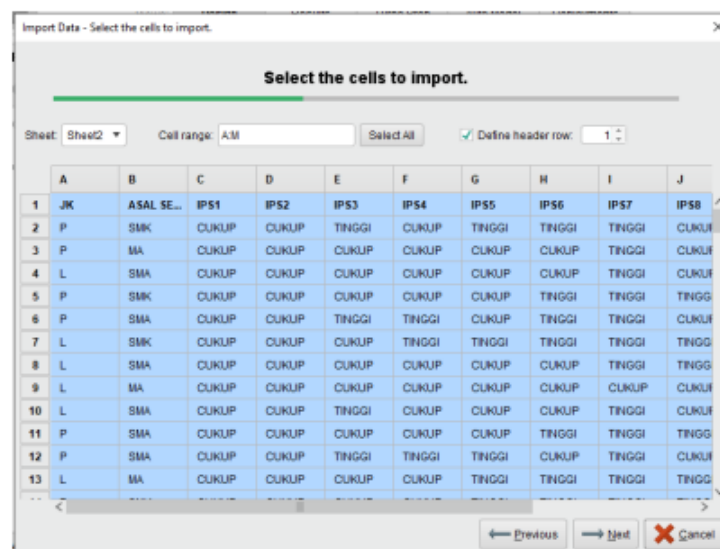
E. Pengujian dengan aplikasi

Pada tahapan awal ketika aplikasi *tool RapiMiner* dijalankan, maka akan terlihat tampilan awal aplikasi ketika sedang berjalan. Berikut adalah tampilan awal *tool RapidMiner*.



Gambar 3. Tampilan *Software Rapid Miner 9.10*

Setelah data kita *import*, maka proses selanjutnya memilih *sheet* dan blok data yang akan digunakan lalu klik tombol *next* untuk melanjutkan kepada tahap selanjutnya.



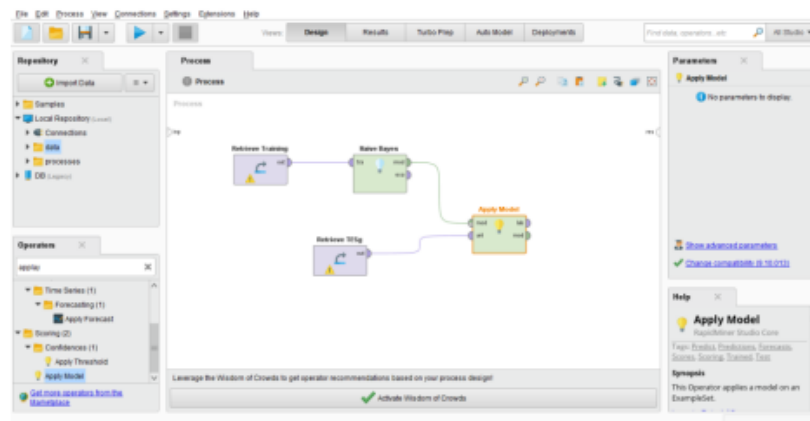
Gambar 4. Tahapan *Import Data*

etelah data *trainig* dan testing kita *drag* ke dalam main process selanjutnya yaitu mencari metode *Naïve Bayes* pada bagian *panel operators* dan menghubungkan antara data training dengan metode *Naïve Bayes* yang bertujuan untuk mempelajari data.



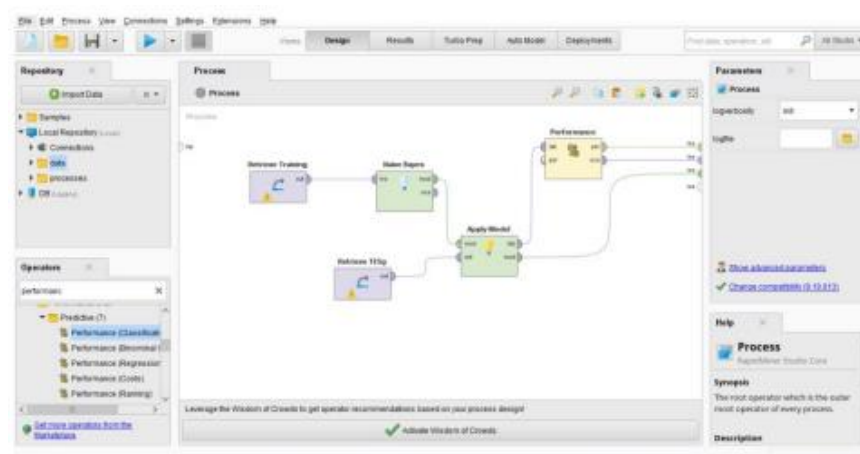
Gambar 5. Drag & Drop data

Selanjutnya yaitu mencari *apply* model pada bagian *panel operators* dan menghubungkannya dengan metode *Naive Bayes* serta data testing.



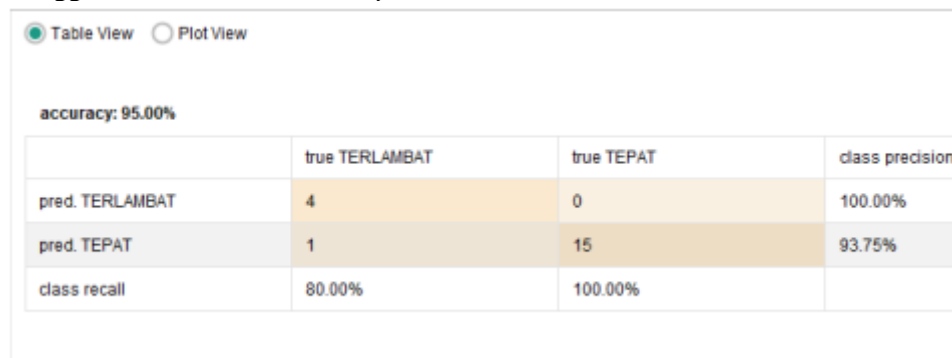
Gambar 5. Apply Models

*Apply* model bertujuan untuk mempelajari pola data dan menghasilkan sebuah model data probabilitas. *Performance* berfungsi untuk melihat tingkat akurasi metode *Naive Bayes* dan diolah oleh tool *RapidMiner* dengan data *training* dan testing. Selanjutnya yaitu menghubungkan *apply* model dengan *performance*. Menghubungkan *performance* dengan *result*.



Gambar 6. Proses menghubungkan *Apply Models* dan *Performance*

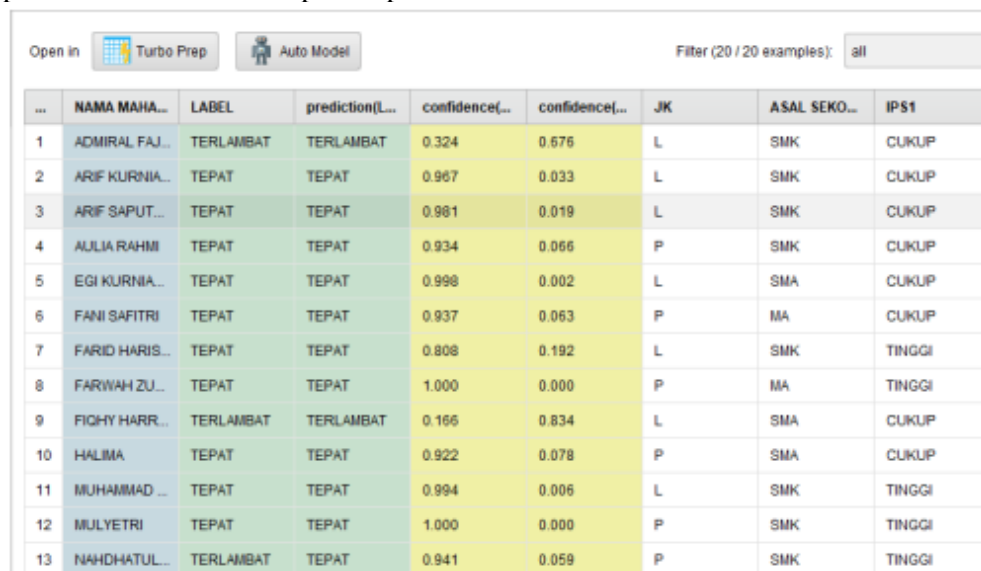
Setelah dihubungkan, maka menghasilkan keluaran dari proses yang dihasilkan adalah tingkat *accuracy* dari data *training* dan testing kelulusan mahasiswa menggunakan metode *Naïve Bayes*.



	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	4	0	100.00%
pred. TEPAT	1	15	93.75%
class recall	80.00%	100.00%	

Gambar 7. Hasil *Accuracy*

Gambar .7 diatas memperlihatkan dengan menggunakan metode *Naïve Bayes* didapatkan hasil akurasi 95,00 %. Setelah mengetahui jumlah mahasiswa Tepat dan Terlambat, selanjutnya adalah pembuktian *prediction* dengan data testing sebanyak 20 menggunakan *tool RapidMiner 9.10*. Sehingga diperoleh hasil pengujian nilai probabilitas untuk mendapatkan prediksi.



...	NAMA MAHA...	LABEL	prediction(L...	confidence(...	confidence(...	JK	ASAL SEKO...	IPST
1	ADMIRAL FAJ...	TERLAMBAT	TERLAMBAT	0.324	0.676	L	SMK	CUKUP
2	ARIF KURNIA...	TEPAT	TEPAT	0.967	0.033	L	SMK	CUKUP
3	ARIF SAPUT...	TEPAT	TEPAT	0.981	0.019	L	SMK	CUKUP
4	AULIA RAHMI	TEPAT	TEPAT	0.934	0.066	P	SMK	CUKUP
5	EGI KURNIA...	TEPAT	TEPAT	0.998	0.002	L	SMA	CUKUP
6	FANI SAFITRI	TEPAT	TEPAT	0.937	0.063	P	MA	CUKUP
7	FARID HARIS...	TEPAT	TEPAT	0.808	0.192	L	SMK	TINGGI
8	FARWAH ZU...	TEPAT	TEPAT	1.000	0.000	P	MA	TINGGI
9	FIQHY HARR...	TERLAMBAT	TERLAMBAT	0.166	0.834	L	SMA	CUKUP
10	HALIMA	TEPAT	TEPAT	0.922	0.078	P	SMA	CUKUP
11	MUHAMMAD ...	TEPAT	TEPAT	0.994	0.006	L	SMK	TINGGI
12	MULYETRI	TEPAT	TEPAT	1.000	0.000	P	SMK	TINGGI
13	NAHDHATUL...	TERLAMBAT	TEPAT	0.941	0.059	P	SMK	TINGGI

Gambar 8. Hasil *Probabilitas Tool RapidMiner*

## KESIMPULAN

Berdasarkan penelitian yang dilakukan mengenai prediksi waktu kelulusan mahasiswa menggunakan algoritma *Naïve Bayes* maka didapatkan kesimpulan sebagai berikut :

1. Dengan kriteria atribut seperti jenis kelamin, asal sekolah, ips 1 sampai 8, ipk dan sks dalam menentukan waktu kelulusan mahasiswa menggunakan algoritma *Naïve Bayes* dihasilkan akurasi 95%. Dengan *class precision* 100%, sedangkan *class recall* 80 %. Dengan jumlah data *training* 103 dan testing 20 data.
2. Dengan tingkat akurasi yang didapatkan 95% menunjukkan bahwa algoritma *Naïve Bayes* dapat diterapkan untuk prediksi waktu kelulusan mahasiswa.
3. Berdasarkan implementasi *tool RapidMiner* dengan metode *Naïve Bayes* menghasilkan tingkat akurasi 95% untuk prediksi waktu kelulusan mahasiswa.

## REFERENSI

- Purwati, N. and Januanti, A.D. (2021) 'Prediksi Tingkat Kelulusan Mahasiswa dengan Algoritma Naive Bayes', 2(1), pp. 126–137.
- Salim, A., Afni, N., Komarudin, R., & Maulana, Y. I. (2022). Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes. *Technologia: Jurnal Ilmiah*, 13(3), 207-214.
- Sepriansyah, R., Purnamasari, S. D., Wardani, K. R. N., & Halim, N. (2023). Prediksi Kelulusan Mahasiswa Fakultas Teknik Universitas Bina Darma Menggunakan Algoritma Naive Bayes. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8 (1), 313-322. DOI : <https://doi.org/10.1145/3369555.3369570>
- Perez, J. G., & Perez, E. S. (2021). Predicting student program completion using Naive Bayes classification algorithm. *International Journal of Modern Education and Computer Science*, 13 (3), 57-67. DOI : <https://doi.org/10.5815/ijmecs.2021.03.05>
- Noronha, Marcelino. (2022). Data Mining Informatics Engineering Student Graduation with the Naive Bayes Classifier Algorithm.
- Matar, N. A., Matar, D. W., & AlMalahmeh, D. T. (2022). A Predictive Model for Students Admission Uncertainty Using Naive Bayes Classifier and Kernel Density Estimation (KDE). *International Journal of Emerging Technologies in Learning*, 17(8). DOI : <https://doi.org/10.3991/ijet.v17i08.29827>
- Salim, A., Afni, N., Komarudin, R., & Maulana, Y. I. (2022). Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes. *Technologia: Jurnal Ilmiah*, 13(3), 207-214.
- Ali, Z. M., Hassoon, N. H., Ahmed, W. S., & Abed, H. N. (2020). The application of data mining for predicting academic performance using k-means clustering and naive bayes classification. *International Journal of Psychosocial Rehabilitation*, 24(03), 2143-2151. DOI : <https://doi.org/10.37200/IJPR/V24I3/PR200962>
- Sepriansyah, R., Purnamasari, S. D., Wardani, K. R. N., & Halim, N. (2023). Prediksi Kelulusan Mahasiswa Fakultas Teknik Universitas Bina Darma Menggunakan Algoritma Naive Bayes. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8 (1), 313-322. DOI : <https://doi.org/10.1145/3369555.3369570>
- Wahyuni, S., & Adinda. (2021). Penerapan Algoritma Naive Bayes Pada Prediksi Kelulusan Mahasiswa (Studi Kasus: Institut Medika Drg. Suherman). *Jurnal Informatika SIMANTIK*, 6(2), 29-34
- Sukarna, R. H., & Ansori, Y. (2022). Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Ilmiah Sains dan Teknologi*, 6(1), 50-61. DOI : <https://doi.org/10.47080/saintek.v6i1.1467>
- Purwati, N. and Januanti, A.D. (2021) 'Prediksi Tingkat Kelulusan Mahasiswa dengan Algoritma Naive Bayes', 2(1), pp. 126–137.
- Olalekan, A. M., Egwuche, O. S., & Olatunji, S. O. (2020, March). Performance evaluation of machine learning techniques for prediction of graduating students in tertiary institution. In *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)* (pp. 1-7). IEEE. DOI : <https://doi.org/10.1109/ICMCECS47690.2020.240888>
- Godinez, C. D., & Lomibao, L. S. (2022). A Gaussian-Bernoulli Mixed Naive Bayes Approach to Predict Students' Academic Procrastination Tendencies in Online Mathematics Learning. *American Journal of Educational Research*, 10(4), 223-232. DOI : <https://doi.org/10.3991/ijet.v17i08.29827>
- Hartatik, H., & Yuliani, E. (2022). Implementation Of The Naive Bayes Algorithm In The Design Of Information Systems For The Development Of Student Academic

Achievement. *International Journal of Multidisciplinary Research and Literature*, 1(4), 478-488.

Sembiring, M. T., & Tambunan, R. H. (2021, March). Analysis of graduation prediction on time based on student academic performance using the Naïve Bayes Algorithm with data mining implementation (Case study: Department of Industrial Engineering USU). In *IOP Conference Series: Materials Science and Engineering* (Vol. 1122, No. 1, p. 012069). IOP Publishing.

Armansyah, A., & Ramli, R. K. (2022). Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 1-10.

Yunus, Y., & Nurcahyo, G. W. (2023). Perbandingan algoritma c4. 5 dan naive bayes dalam prediksi kelulusan mahasiswa. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), 193-199. DOI : <https://doi.org/10.37859/coscitech.v4i1.4755>