

**ANALISIS *DATA MINING* UNTUK PEMETAAN MAHASISWA YANG  
MEMBUTUHKAN BIMBINGAN DAN KONSELING MENGGUNAKAN  
ALGORITMA *NAÏVE BAYES CLASSIFIER***

**Hadi Kurnia Saputra<sup>1</sup>**

**ABSTRACT**

*The students' success on educational process in the university can be measured through academic performance and study completion. However, on the students' educational process can't be separated from a variety of problems, both from the environmental education and family and also social environment. One of the efforts made by educational institutions to encourage low students' academic performance are to provide guidance and counseling services. Data Mining can be done with the classification of students who have good and bad academic performance. Naïve Bayes Classifier Algorithm is chosen to perform Classification of student slow academic performance and assessed that they need Guidance and Counseling process. A study of 507 records of student academic datasets Sign 2008 Strata-1 program of the Faculty of Engineering, State University of Padang, Naïve Bayes algorithm is able to classify students who need Guidance and Counseling process. Based on the tests performed on the students of Engineering Studies Program representing that all of them obtained an accuracy of 81%.*

**Keywords** : *Guidance and Counseling, Data Mining, Academic Performance, Classification, Naïve Bayes Classifier*

**INTISARI**

Keberhasilan mahasiswa dalam menjalani proses pendidikan di Perguruan Tinggi dapat diukur melalui kinerja akademik dan masa penyelesaian studi. Namun, dalam perjalanan mahasiswa mengikuti proses pendidikan tidak bisa lepas dari berbagai permasalahan, baik yang bersumber dari lingkungan pendidikan maupun lingkungan keluarga dan pergaulan. Salah satu upaya yang dilakukan oleh institusi pendidikan untuk mendorong peningkatan kinerja akademik mahasiswa yang rendah adalah dengan memberikan pelayanan Bimbingan dan Konseling (BK). Dengan *Data Mining* dapat dilakukan klasifikasi mahasiswa yang memiliki kinerja akademik yang baik dan buruk. Algoritma *Naïve Bayes Classifier* dipilih untuk melakukan klasifikasi mahasiswa yang berkinerja akademik rendah dan dinilai membutuhkan proses Bimbingan dan Konseling. Penelitian yang dilakukan terhadap 507 *record* set data akademik mahasiswa Tahun Masuk 2008 program Strata-1 Fakultas Teknik Universitas Negeri Padang, algoritma *Naïve Bayes* mampu mengklasifikasi mahasiswa yang membutuhkan proses Bimbingan dan Konseling. Berdasarkan pengujian yang dilakukan terhadap mahasiswa Fakultas Teknik yang mewakili keseluruhan Program Studi diperoleh akurasi sebesar 81%.

**Kata Kunci** : *Bimbingan dan Konseling, Data Mining, Kinerja Akademik, Klasifikasi, Naïve Bayes Classifier*

## PENDAHULUAN

Pemasalahan akademik yang sering dihadapi oleh perguruan tinggi adalah permasalahan prestasi akademik mahasiswa yang rendah. Dalam perjalanan mahasiswa menuntut ilmu tidak akan luput dari masalah yang dihadapi baik masalah akademik maupun non akademik. Masalah-masalah tersebut dapat mempengaruhi tidak tercapainya prestasi optimal mahasiswa, atau salah satu indikator yang dapat dilihat dialaminya masalah oleh mahasiswa adalah dengan pencapaian Indeks Prestasi (IP) yang rendah, atau dibawah rata-rata [7].

Melalui teknik *Data Mining*, data akademik dapat dimanfaatkan untuk memperoleh informasi prestasi akademik, masa studi serta tingkat kelulusan mahasiswa. Data-data tersebut dapat diolah lebih lanjut untuk memetakan mahasiswa yang berpotensi menerima sanksi akademik, selanjutnya pihak manajemen dapat mengambil kebijakan dan tindakan untuk mangantisipasi hal tersebut. Salah satu tindakan yang dapat dilakukan adalah dengan memberikan layanan bimbingan dan konseling kepada setiap mahasiswa yang berpotensi dikenakan sanksi akademik.

## PENDEKATAN PEMECAHAN MASALAH

***Knowledge Discovery in Database***  
*Knowledge Discovery in Database* (KDD) adalah penerapan metode saintifik pada *Data Mining*. Dalam konteks ini *Data Mining* merupakan satu langkah dari proses KDD [6].

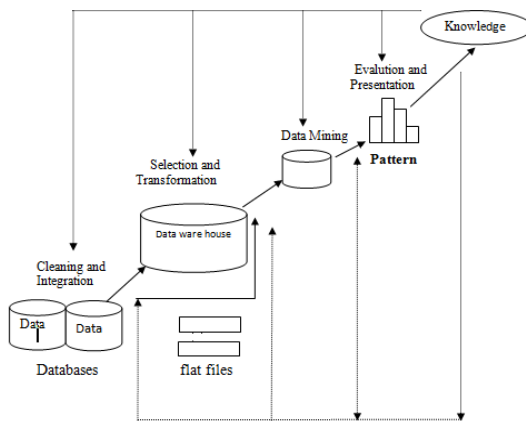
KDD adalah kegiatan atau proses untuk mendapatkan informasi yang meliputi pengumpulan,

pemakaian data historis untuk menemukan keteraturan, pola dan hubungan dalam set data berukuran besar. Fungsi-fungsi dalam *Data Mining* antara lain fungsi deksripsi, estimasi, prediksi, klasifikasi, klasterisasi dan asosiasi.

KDD merupakan keseluruhan proses konversi data mentah menjadi pengetahuan yang bermanfaat yang terdiri dari serangkaian tahap transformasi meliputi *data preprocessing* dan *postprocessing*. *Data preprocessing* merupakan langkah awal untuk mengubah data mentah menjadi format yang sesuai untuk tahap analisis selanjutnya. *Data preprocessing* kemungkinan akan membutuhkan waktu yang sangat lama, hal ini dikarenakan data yang mentah kemungkinan disimpan dengan format dan *database* yang berbeda. Sedangkan *Postprocessing* meliputi semua operasi yang harus dilakukan agar hasil *Data Mining* dapat diakses dan lebih mudah diinterpretasikan oleh para pihak analisis [10].

## ***Data Mining***

*Data Mining* sebenarnya merupakan bagian dari proses *Knowledge Discovery in Database* (KDD), bukan sebagai teknologi yang utuh dan berdiri sendiri. *Data Mining* merupakan salah satu bagian langkah yang penting dalam proses KDD terutama berkaitan dengan ekstraksi dan penghitungan pola-pola dari data yang ditelaah. Adapun tahapan atau rangkaian proses *Data Mining* dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat iterative. Terdapat 7 tahapan pada *Data Mining*, yaitu [1] :



**Gambar 1. Tahapan Data Mining**

Tahapan-tahapan *Data Mining* seperti gambar 1 dapat dijelaskan sebagai berikut :

**a. Data Cleaning**

Membersihkan data dari data yang tidak konsisten dan bersifat *noise* karena secara umum data yang tersedia pada *database* memiliki isian-isian yang tidak sempurna karena hilang ataupun tidak *valid*. Selain itu, ada juga karakteristik data yang tidak relevan dengan kebutuhan untuk *Data Mining*.

**b. Data Integration**

Mengintegrasikan data dari berbagai sumber, dimana integrasi ini dilakukan pada data-data yang memiliki atribut unik. Pada proses integrasi perlu dilaksanakan dengan hati-hati karena kesalahan dapat mengakibatkan hasil yang menyimpang.

**c. Data Selection**

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

**d. Data Transformation**

Mentransformasikan data supaya sesuai untuk kebutuhan *Data Mining*. Aturan pemrograman untuk mengenali pola dari data untuk pelaksanaan transformasi ini sering disebut dengan

algoritma. Beberapa algoritma *Data Mining* yang populer seperti : *Association Rule*, *K-Means Clustering*, *C.45*, *Naïve Bayes Classifier*, *Apriori*, *Tree Structured*, dan lain-lain. Masing-masing algoritma ini akan menghasilkan *output* yang berbeda-beda.

**e. Proses Mining**

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

**f. Pattern Evaluation**

Evaluasi atas pola yang ditemukan berguna untuk menemukan informasi emas (*knowledge*). Pada tahapan ini diharapkan adanya hasil berupa pola yang unik yang dapat digunakan sebagai model untuk memprediksi perilaku-perilaku bisnis yang sekiranya akan mempengaruhi sebuah keputusan manajemen.

**g. Knowledge Presentation**

Mempresentasikan *knowledge* yang sudah didapatkan dari *user*. Tahapan terakhir ini merangkum dan menerjemahkan analisa yang telah didapat dengan bahasa yang semudah mungkin karena melibatkan orang lain pada manajemen yang kemungkinan besar awam dengan *Data Mining*.

**Teknik Klasifikasi Data Mining**

Klasifikasi *Data Mining* adalah suatu metode pembelajaran, untuk memprediksi nilai dari sekelompok atribut dalam menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [9].

Klasifikasi merupakan suatu fungsionalitas *Data Mining* yang akan menghasilkan model untuk

memprediksi kelas atau kategori dari objek-objek di dalam basis data. Klasifikasi merupakan proses yang terdiri dari dua tahap, yaitu tahap pembelajaran dan tahap pengklasifikasian. Pada tahap pembelajaran, sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis training data. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan  $Y=F(X)$  di mana  $Y$  adalah kelas hasil prediksi dan  $X$  adalah *tuple* yang ingin diprediksi kelasnya.

*Data Classification* memiliki dua tahap proses. Tahap pertama adalah membangun suatu model yang berdasarkan serangkaian data *class* yang disebut *learned model*. Model tersebut dibangun dengan menanalisa *database tuple*. Setiap *tuple* diasumsikan menjadi *predefined class* yang ditentukan oleh satu atribut yang disebut *class label attribute*. Akibat terdapat *class label* maka tahap ini juga dikenal dengan *supervised learning* [2].

Teknik klasifikasi merupakan pendekatan untuk menjalankan fungsi klasifikasi dalam *Data Mining* yaitu untuk menggolongkan data. Teknik klasifikasi ini dapat pula digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma *Decision Tree* C.45, *Artificial Neural Networks* (ANN), *K-Nearest Neighbor* (KNN), algoritma *Naive Bayes*, *Neural Network* serta algoritma lainnya.

#### **Algoritma Naive Bayes Classifier**

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive 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 sehingga dikenal sebagai *Teorema Bayes*. 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 [4].

*Naive Bayes Classifier* merupakan salah satu algoritma dalam teknik *Data Mining* yang menerapkan teori *Bayes* dalam klasifikasi. Teorema keputusan *Bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*) [8]. *Naive bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output*. Dengan kata lain, diberikan nilai *output*, probabilitas mengamati secara bersama adalah produk dari probabilitas individu.

Bentuk umum dari teorema *Bayes* sebagai berikut :

$$P(Y|X) = \frac{P(X|Y).P(Y)}{P(X)} \quad (1)$$

Di mana :

- $X$  : Data dengan *class* yang belum diketahui
- $Y$  : Hipotesis data  $X$  merupakan suatu *class* spesifik
- $P(Y|X)$  : Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis  $Y$  terjadi jika diberikan bukti (*evidence*)  $X$  terjadi.
- $P(Y)$  : Probabilitas awal (priori) hipotesis  $Y$  terjadi tanpa mengandung bukti apapun.

$P(X|Y)$  : Probabilitas sebuah bukti  $X$  terjadi akan mempengaruhi hipotesis  $Y$ .  
 $P(X)$  : Probabilitas awal (priori) bukti  $X$  terjadi tanpa mengandung hipotesis/bukti yang lain.

Kaitan antara *Naïve Bayes* dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah bahwa hipotesis dalam teorema *Bayes* merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi [5].

Jika  $X$  adalah vektor masukan yang berisi fitur dan  $Y$  adalah label kelas, *Naïve Bayes* dituliskan dengan  $P(Y|X)$ . Notasi tersebut berarti probabilitas label kelas  $Y$  didapatkan setelah fitur-fitur  $X$  diamati. Notasi ini disebut juga probabilitas akhir (*posterior probability*) untuk  $Y$ , sedangkan  $P(Y)$  disebut probabilitas awal (*prior probability*)  $Y$ .

Formulasi *Naïve Bayes* untuk klasifikasi adalah :

$$P(Y|X) = \frac{P(Y)\prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (2)$$

$P(X|Y)$  : Probabilitas data dengan *vector*  $X$  pada kelas  $Y$ .

$P(Y)$  : Probabilitas awal kelas  $Y$ .

$\prod_{i=1}^q P(X_i|Y)$  : Probabilitas independen kelas  $Y$  dari semua fitur dalam vektor  $X$ .

Nilai  $P(X)$  selalu tetap sehingga dalam perhitungan prediksi nantinya tinggal menghitung bagian  $P(Y)\prod_{i=1}^q P(X_i|Y)$  dengan memilih yang terbesar sebagai kelas yang

dipilih sebagai hasil prediksi. Sementara probabilitas independen  $\prod_{i=1}^q P(X_i|Y)$  tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas  $Y$ , yang dinotasikan dengan :  
 $P(X|Y = y) = \prod_{i=1}^q P(X_i|Y = y)$

3)  
 Setiap set fitur  $X = \{X_1, X_2, X_3, \dots, X_q\}$  terdiri atas  $q$  atribut ( $q$  dimensi).

*Naive Bayes Classifier* mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberikan dengan label kelas  $y$ . Asumsi independen bersyarat dapat dinyatakan dalam bentuk berikut :

$$P(X|Y = y) = \prod_{i=1}^d P(X_i|Y = y) \quad (4)$$

Dengan tiap set atribut  $X = \{X_1, X_2, \dots, X_d\}$  terdiri dari  $d$  atribut. Atribut kategorikal, peluang bersyarat  $P(X_i = x_i|Y = y)$  diestimasi menurut pecahan *training instances* pada kelas  $y$  yang membuat nilai atribut khusus  $x_i$ . (3)

Distribusi *Gaussian* sering dipilih untuk merepresentasikan peluang kelas bersyarat untuk atribut kontinyu. Distribusi dikarakterisasi dengan dua parameter yaitu *mean*,  $\mu$ , dan varian,  $\sigma^2$ . Untuk tiap kelas  $y_j$ , peluang kelas bersyarat untuk atribut  $X_i$  adalah :

$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp \left( -\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right) \quad (5)$$

Parameter  $\mu_{ij}$  dapat diestimasi berdasarkan sampel *mean*  $X_i$  ( $\bar{x}$ ) untuk seluruh *training record* yang

dimiliki kelas  $y_j$ . Dengan cara sama,  $\sigma^2_{ij}$  dapat diestimasi dari sampel varian ( $s^2$ ) *training record* tersebut.

### Bimbingan Konseling

Bimbingan adalah membantu pemecahan masalah seseorang sehingga dapat membuat keputusan yang tepat atau dengan kata dengan bimbingan diharapkan memperoleh sebuah solusi dan perencanaan yang tepat [12]. Konseling adalah suatu proses wawancara yang bertujuan untuk memberikan bantuan kepada seseorang sehingga orang yang dilayani dapat lebih berkembang dalam kehidupannya [3].

Bimbingan Konseling merupakan proses pemberian bantuan yang dilakukan melalui wawancara konseling (*face to face*) oleh seorang ahli (disebut konselor) kepada individu yang sedang mengalami sesuatu masalah (disebut konseli) yang bermuara pada teratasinya masalah yang dihadapi konseli serta dapat memanfaatkan berbagai potensi yang dimiliki dan sarana yang ada, sehingga individu atau kelompok individu itu dapat memahami dirinya sendiri untuk mencapai perkembangan yang optimal, mandiri serta dapat merencanakan masa depan yang lebih baik untuk mencapai kesejahteraan hidup.

## ANALISA DAN PERANCANGAN

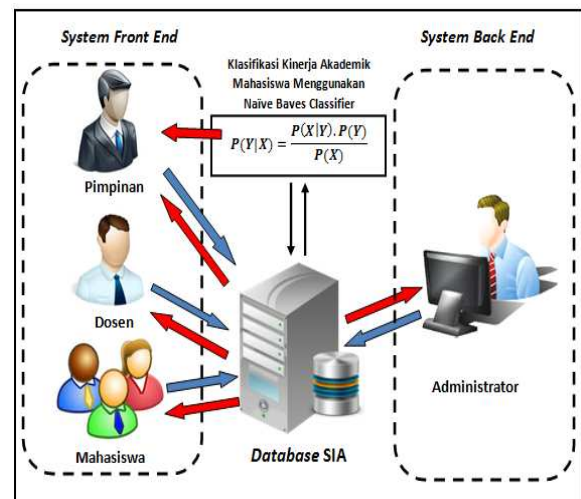
### Analisa Sistem

Analisis sistem dilakukan untuk mengetahui faktor-faktor apa saja yang mempengaruhi sistem. Faktor-faktor tersebut yang akan dijadikan sebagai acuan dalam proses pengembangan sistem selanjutnya.

Berdasarkan hasil observasi dan wawancara yang dilakukan dengan pihak-pihak terkait dalam

penentuan kriteria mahasiswa yang membutuhkan proses bimbingan dan konseling yang menjadi *domain* utama dari penelitian ini, belum ada ketentuan baku dan aturan khusus yang mengharuskan seorang mahasiswa harus mengikuti proses bimbingan dan konseling.

Setelah melakukan analisis dan identifikasi permasalahan pada sistem yang sedang berjalan, diperlukan adanya sebuah pola yang terotomasi oleh sistem untuk penentuan kriteria mahasiswa yang membutuhkan proses bimbingan konseling dengan memaksimalkan pemberdayaan set data akademik mahasiswa yang tersedia pada *database* akademik.



Gambar 2. Arsitektur Sistem yang Diusulkan

### Analisa Data Mining

Teknik *Data Mining* yang digunakan dalam penelitian ini adalah teknik *Classification*. Sedangkan algoritma yang akan digunakan adalah *Naive Bayes Classifier*, sehingga analisa yang dilakukan mengacu kepada kebutuhan analisa algoritma *Naive Bayes Classifier*.

#### 1. Pengumpulan Data Awal

Penelitian dilakukan menggunakan set data akademik mahasiswa program studi sarjana (S1) Fakultas Teknik



Universitas Negeri Padang. Sampel data yang digunakan adalah data akademik mahasiswa tahun masuk 2008 semua program studi yang ada di Fakultas Teknik. Pemilihan sampel data ini didasarkan pada aspek kelengkapan data serta relevansinya dengan topik penelitian.

2. Mendeskripsikan Data

Set data akademik mahasiswa yang digunakan dalam penelitian ini terdiri dari data profil mahasiswa, orang tua, sekolah asal (SMTA), indeks prestasi dan kelulusan. Data-data tersebut dinilai bisa digunakan untuk penelitian yang berkaitan dengan kinerja akademik mahasiswa. Data-data tersebut terekam dalam bentuk tabel-tabel pada *database* akademik UNP.

3. Evaluasi Kualitas Data

Hasil evaluasi terhadap kualitas data yaitu menemukan banyak nilai kosong/*null* yang disebut dengan *missing value* pada atribut dalam set data akademik mahasiswa. Dari 507 *record* data terdapat 12 *record* yang memiliki atribut bernilai kosong/*null*, sedangkan 495 *record* memiliki atribut yang sudah memiliki data lengkap. Data yang memiliki atribut bernilai kosong/*null*.

*Record* yang memiliki nilai atribut kosong/*null* atau *record missing value* adalah :

$$\frac{12}{507} \times 100\% = 2.36\%.$$

data yang memiliki nilai atribut yang lengkap adalah :

$$\frac{495}{507} \times 100\% = 97.63\%.$$

4. Pemilihan Atribut

Set data akademik mahasiswa yang digunakan dalam penelitian ini terdiri dari beberapa atribut antara lain : Angkatan/Tahun Masuk, Nama Mahasiswa, Jenis Kelamin, Status (aktif, cuti, lulus),

Program Studi, Status Masuk, Jenis SMTA, Pekerjaan Orang Tua (Ayah), IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, Masa Studi.

Pemilihan atribut ini didasarkan pada faktor kebutuhan data yang relevan dengan topik penelitian dengan mempertimbangkan kelengkapan dan ketersediaan data pada *database* akademik.

**Persiapan Data (*Data Preparation*)**

Persiapan data mencakup semua kegiatan untuk membangun set data akademik mahasiswa yang akan diterapkan ke dalam sistem, dari data mentah awal berupa set data akademik mahasiswa selanjutnya akan melakukan proses *data mining* dan akan diteruskan dengan proses klasifikasi menggunakan algoritma *Naive Bayes Classifier*.

1. Seleksi Data (*Data Selection*)

Atribut yang digunakan adalah ID MHS, Nama, Jenis Kelamin, Status Wisuda, Program Studi, Jenjang Program, Status Masuk, Jenis SMTA, Pekerjaan Ayah, IP Sem 1, IP Sem 2, IP Sem 3, IP Sem 4 dan Masa Studi.

Dari seluruh program studi Strata 1 Fakultas Teknik UNP yaitu Program Studi Pendidikan Teknik Bangunan, Pendidikan Teknik Elektro, Pendidikan Teknik Elektronika, Pendidikan Teknik Informatika dan Komputer, Pendidikan Teknik Mesin, Pendidikan Teknik Otomotif serta Pendidikan Kesejahteraan Keluarga, diperoleh data sebanyak 507 *record*. Data tersebut akan dibagi menjadi dua bagian untuk keperluan *data training* dan *data testing*. Pengambilan data yang akan digunakan sebagai *data training* pada proses *Data Mining* didasarkan pada teknik

pengambilan sampel penelitian. Jika populasi > 100 objek maka dapat diambil 10 % - 15 % atau 20 % - 25 % dari semua jumlah populasi. Jika populasi < 100 objek maka diambil seluruh dari populasi sebagai sampel penelitian [11]. Jika sampel untuk *data training* diambil sebanyak 10 % dari 507 *record* data, maka data yang akan digunakan sebagai *data training* adalah sebanyak 51 *record*. Sedang sisanya sebanyak 456 *record* akan digunakan sebagai *data testing*.

2. Pengolahan Data Mentah (*Data Preprocessing/ Data Cleaning*)  
Tahap ini merupakan tahap untuk memastikan data akademik mahasiswa yang dipilih telah layak untuk dilakukan proses pengolahan. *Data Cleaning* dilakukan ketika terlihat data yang tidak konsisten atau data tersebut ganda. Proses *Cleaning* data dapat dilakukan dengan cara :
  - 1) Menambahkan isi atribut yang kosong secara manual
  - 2) Menghapus data yang ganda secara manual
 Kedua cara tersebut tidak efektif kalau digunakan pada data yang banyak.
3. Transformasi Data (*Data Transformation*)  
Transformasi data diberlakukan untuk atribut/parameter yang harus diubah dulu ke format sesuai kategorinya. Transformasi data dilakukan secara manual pada data yang akan diolah. Data yang berjenis numerikal seperti Masa Studi harus dilakukan proses inialisasi data terlebih dahulu ke dalam bentuk kategorikal.

**Tabel 1. Transformasi Data**

No	Kondisi Awal		Transformasi		Keterangan
	Atribut	Kategori	Atribut	Kategori	
1	Status Wisuda	A = Aktif C = Cuti L = Lulus	Status Wisuda	B = Belum Lulus L = Lulus	Satus Aktif dan Cuti dikategorikan menjadi Belum Lulus
2	Masa Studi	Range 1 s.d. 14 (Dalam satuan Semester)	Butuh BK (Bimbingan Konseling)	Tidak = Masa studi ≤ 10 Semester Ya = masa studi ≥ 11 Semester.	Masa studi yang telah dilalui mahasiswa baik yang belum lulus maupun sudah lulus Atribut ini yang akan dijadikan sebagai dasar dalam menentukan apakah mahasiswa membutuhkan layanan bimbingan dan konseling.
3	IP Semester 1 sampai 4	Range 0.0 sampai dengan 4.0	IP Semester		Indeks prestasi mahasiswa yang bernilai <i>numeric</i>

### Klasifikasi dengan *Naive Bayes Classifier*

#### 1. Pemilihan Data Latih (*Data Training*)

*Data training* diambil secara acak (*random*) dari set data akademik mahasiswa yang sudah ditransformasi. Pengambilan data secara acak dapat dilakukan dengan membangkitkan bilangan *random*. *Data training* diambil sebanyak 51 *record* dan sampel dipilih secara acak dari semua data mahasiswa Fakultas Teknik UNP semua jenjang program. Data yang diambil hanya data akademik masiswa tahun masuk 2008.

##### a. Menghitung Probabilitas Setiap Fitur dan Kelas

Langkah awal yang dilakukan adalah menghitung nilai setiap fitur pada setiap kelasnya menggunakan teorema *Bayes*. Data yang digunakan adalah *Data training*. Untuk menghitung probabilitas setiap fitur, digunakan rumus  $P(X/C_i)$  untuk  $i = 1,2$ . Hasil perhitungan



probabilitas setiap fitur dan kelas sebagai berikut :

1) Kelas Butuh BK :

Probabilitas kelas Butuh BK "YA" adalah  $P(\text{Butuh BK "Ya"}) = 21/51 = 0.4118$ .

Probabilitas kelas Butuh BK "Tidak" adalah  $P(\text{Butuh BK "Tidak"}) = 21/51 = 0.5882$ .

2) Fitur Jenis Kelamin :

$P(\text{Jenis Kelamin} = L | \text{Butuh BK} = Ya) = 16/21 = 0.7619$

$P(\text{Jenis Kelamin} = P | \text{Butuh BK} = Ya) = 5/21 = 0.2381$

$P(\text{Jenis Kelamin} = L | \text{Butuh BK} = Tidak) = 19/30 = 0.6333$

$P(\text{Jenis Kelamin} = P | \text{Butuh BK} = Tidak) = 11/30 = 0.3667$

3) Fitur Status Wisuda :

$P(\text{Status Wisuda} = L | \text{Butuh BK} = Ya) = 6/21 = 0.2857$

$P(\text{Status Wisuda} = P | \text{Butuh BK} = Ya) = 15/21 = 0.7143$

$P(\text{Status Wisuda} = L | \text{Butuh BK} = Tidak) = 28/30 = 0.9333$

$P(\text{Status Wisuda} = P | \text{Butuh BK} = Tidak) = 2/30 = 0.667$

4) Fitur Program Studi :

$P(\text{Prodi} = \text{Bangunan} | \text{Butuh BK} = Ya) = 5/21 = 0.2381$

$P(\text{Prodi} = \text{Elektro} | \text{Butuh BK} = Ya) = 6/21 = 0.2857$

$P(\text{Prodi} = \text{Elektronika} | \text{Butuh BK} = Ya) = 4/21 = 0.1905$

$P(\text{Prodi} = \text{Informatika} | \text{Butuh BK} = Ya) = 1/21 = 0.0476$

$P(\text{Prodi} = \text{Mesin} | \text{Butuh BK} = Ya) = 0/21 = 0.0000$

$P(\text{Prodi} = \text{Otomotif} | \text{Butuh BK} = Ya) = 2/21 = 0.0952$

$P(\text{Prodi} = \text{PKK} | \text{Butuh BK} = Ya) = 3/21 = 0.1429$

$P(\text{Prodi} = \text{Bangunan} | \text{Butuh BK} = Tidak) = 4/30 = 0.1333$

$P(\text{Prodi} = \text{Elektro} | \text{Butuh BK} = Tidak) = 3/30 = 0.1000$

$P(\text{Prodi} = \text{Elektronika} | \text{Butuh BK} = Tidak) = 6/30 = 0.2000$

$P(\text{Prodi} = \text{Informatika} | \text{Butuh BK} = Tidak) = 6/30 = 0.2000$

$P(\text{Prodi} = \text{Mesin} | \text{Butuh BK} = Tidak) = 5/30 = 0.1667$

$P(\text{Prodi} = \text{Otomotif} | \text{Butuh BK} = Tidak) = 4/30 = 0.1333$

$P(\text{Prodi} = \text{PKK} | \text{Butuh BK} = Tidak) = 2/30 = 0.0667$

5) Fitur Status Masuk :

$P(\text{Status Masuk} = \text{PMDK/Undangan} | \text{Butuh BK} = Ya) = 6/21 = 0.2857$

$P(\text{Status Masuk} = \text{SPMB/SNMPTN} | \text{Butuh BK} = Ya) = 2/21 = 0.0952$

$P(\text{Status Masuk} = \text{Non Reguler} | \text{Butuh BK} = Ya) = 13/21 = 0.6190$

$P(\text{Status Masuk} = \text{PMDK/Undangan} | \text{Butuh BK} = Tidak) = 8/30 = 0.2667$

$P(\text{Status Masuk} = \text{SPMB/SNMPTN} | \text{Butuh BK} = Tidak) = 8/30 = 0.2667$

$P(\text{Status Masuk} = \text{Non Reguler} | \text{Butuh BK} = Tidak) = 14/30 = 0.4667$

6) Fitur Jenis SMTA :

$P(\text{Jenis SMTA} = \text{SMU} | \text{Butuh BK} = Ya) = 7/21 = 0.3333$

$P(\text{Jenis SMTA} = \text{SMK} | \text{Butuh BK} = Ya) = 10/21 = 0.4762$

$P(\text{Jenis SMTA} = \text{MA} | \text{Butuh BK} = Ya) = 1/21 = 0.0476$

$P(\text{Jenis SMTA} = \text{SMTA lain-lain} | \text{Butuh BK} = Ya) = 3/21 = 0.1429$

$P(\text{Jenis SMTA} = \text{SMU} | \text{Butuh BK} = Tidak) = 9/30 = 0.3000$

$P(\text{Jenis SMTA} = \text{SMK} | \text{Butuh BK} = Tidak) = 6/30 = 0.2000$

$P(\text{Jenis SMTA} = \text{MA} | \text{Butuh BK} = Tidak) = 0/30 = 0.0000$

$P(\text{Jenis SMTA} = \text{SMTA lain-lain} | \text{Butuh BK} = Tidak) = 15/30 = 0.5000$

7) Fitur Pekerjaan Ayah :

$P(\text{Pekerjaan Ayah} = \text{Dosen/Guru} | \text{Butuh BK} = Ya) = 1/21 = 0.0476$

$P(\text{Pekerjaan Ayah} = \text{Karyawan Swasta} | \text{Butuh BK} = Ya) = 1/21 = 0.0476$

$P(\text{Pekerjaan Ayah} = \text{Pegawai Negeri} | \text{Butuh BK} = Ya) = 8/21 = 0.3810$

$P(\text{Pekerjaan Ayah} = \text{Pensiunan} | \text{Butuh BK} = Ya) = 2/21 = 0.0952$

$P(\text{Pekerjaan Ayah} = \text{Petani/Nelayan} | \text{Butuh BK} = Ya) = 1/21 = 0.0476$

$P(\text{Pekerjaan Ayah} = \text{TNI/Polri} | \text{Butuh BK} = Ya) = 0/21 = 0.0000$

$P(\text{Pekerjaan Ayah} = \text{Wiraswasta} | \text{Butuh BK} = Ya) = 6/21 = 0.2857$

$P(\text{Pekerjaan Ayah} = \text{Lain-lain} | \text{Butuh BK} = Ya) = 2/21 = 0.0952$

$P(\text{Pekerjaan Ayah} = \text{Dosen/Guru} | \text{Butuh BK} = Tidak) = 1/30 = 0.0333$

$P(\text{Pekerjaan Ayah} = \text{Karyawan Swasta} | \text{Butuh BK} = Tidak) = 2/30 = 0.0667$

$P(\text{Pekerjaan Ayah} = \text{Pegawai Negeri} | \text{Butuh BK} = Tidak) = 3/30 = 0.1000$

$P(\text{Pekerjaan Ayah} = \text{Pensiunan} | \text{Butuh BK} = Tidak) = 1/30 = 0.0333$

$P(\text{Pekerjaan Ayah} = \text{Petani/Nelayan} | \text{Butuh BK} = Tidak) = 6/30 = 0.2000$

$P(\text{Pekerjaan Ayah} = \text{TNI/Polri} | \text{Butuh BK} = Tidak) = 3/30 = 0.1000$

$P(\text{Pekerjaan Ayah} = \text{Wiraswasta} | \text{Butuh BK} = Tidak) = 10/30 = 0.3333$

$P(\text{Pekerjaan Ayah} = \text{Lain-lain} | \text{Butuh BK} = Tidak) = 4/30 = 0.1333$

8) Fitur IP Semester 1 :

$$\bar{X}_{\text{Butuh BK "Ya"}} = 2.2862$$

$$S^2_{\text{Butuh BK "Ya"}} = 0.5383$$

$$S_{\text{Butuh BK "Ya"}} = 0.7337$$

$$\bar{X}_{\text{Butuh BK "Tidak"}} = 3.0623$$

$$S^2_{\text{Butuh BK "Tidak"}} = 0.2184$$

$$S_{\text{Butuh BK "Tidak"}} = 0.4674$$

9) Fitur IP Semester 2 :

$$\bar{X}_{\text{Butuh BK "Ya"}} = 2.1933$$

$$S^2_{\text{Butuh BK "Ya"}} = 0.6014$$

$$S_{\text{Butuh BK "Ya"}} = 0.7755$$

$$\bar{X}_{\text{Butuh BK "Tidak"}} = 2.8590$$

$$S^2_{\text{Butuh BK "Tidak"}} = 0.3038$$

$$S_{\text{Butuh BK "Tidak"}} = 0.5512$$

10) Fitur IP Semester 3 :

$$\bar{X}_{\text{Butuh BK "Ya"}} = 2.0586$$

$$S_{Butuh\ BK\ "Ya"}^2 = 0.6042$$

$$S_{Butuh\ BK\ "Ya"} = 0.7773$$

$$\bar{X}_{Butuh\ BK\ "Tidak"} = 2.7867$$

$$S_{Butuh\ BK\ "Tidak"}^2 = 0.4326$$

$$S_{Butuh\ BK\ "Tidak"} = 0.6577$$

11) Fitur IP Semester 4 :

$$\bar{X}_{Butuh\ BK\ "Ya"} = 1.8976$$

$$S_{Butuh\ BK\ "Ya"}^2 = 0.3966$$

$$S_{Butuh\ BK\ "Ya"} = 0.6298$$

$$\bar{X}_{Butuh\ BK\ "Tidak"} = 2.8233$$

$$S_{Butuh\ BK\ "Tidak"}^2 = 0.2923$$

$$S_{Butuh\ BK\ "Tidak"} = 0.5406$$

b. Menghitung Probabilitas Akhir Setiap Kelas

Misalkan akan dihitung probabilitas akhir menggunakan sebuah contoh data training ID MHS = 200807060 dengan nilai fitur Jenis Kelamin = P, Status Wisuda=B, Program Studi = Pendidikan Kesejahteraan Keluarga, Status Masuk = SPMB/SNMPTN, Jenis SMTA = SMK, Pekerjaan Ayah = Wiraswasta, IP Sem 1= 2.24, IP Sem 2 = 2.0, IP Sem 3 = 1.86, IP Sem 4 = 2.2.

Berikut adalah tahapan penyelesaiannya :

1) Harus diketahui nilai probabilitas setiap fitur pada setiap kelasnya, menggunakan rumus  $P(X_i|Y_j)$ .

Nilai probabilitas setiap fitur mengacu kepada nilai yang telah diperoleh rincian sebagai berikut :

$$P(\text{Jenis Kelamin} = P | \text{Butuh BK} = Ya) = 0.2381$$

$$P(\text{Jenis Kelamin} = P | \text{Butuh BK} = Tidak) = 0.3667$$

$$P(\text{Status Wisuda} = B | \text{Butuh BK} = Ya) = 0.7143$$

$$P(\text{Status Wisuda} = B | \text{Butuh BK} = Tidak) = 0.0667$$

$$P(\text{Prodi} = PKK | \text{Butuh BK} = Ya) = 0.1429$$

$$P(\text{Prodi} = PKK | \text{Butuh BK} = Tidak) = 0.0667$$

$$P(\text{Status Masuk} = SPMB/SNMPTN | \text{Butuh BK} = Ya) = 0.0952$$

$$P(\text{Status Masuk} = SPMB/SNMPTN | \text{Butuh BK} = Tidak) = 0.2667$$

$$P(\text{Jenis SMTA} = SMK | \text{Butuh BK} = Ya) = 0.4762$$

$$P(\text{Jenis SMTA} = SMK | \text{Butuh BK} = Tidak) = 0.2000$$

$$P(\text{Pekerjaan Ayah} = Wiraswasta | \text{Butuh BK} = Ya) = 0.2857$$

$$P(\text{Pekerjaan Ayah} = Wiraswasta | \text{Butuh BK} = Tidak) = 0.3333$$

Nilai probabilitas untuk fitur dengan tipe numerik yakni fitur IP Sem 1, IP Sem 2, IP Sem 3 dan IP Sem 4 dihitung menggunakan rumus :

Probabilitas IP Sem 1 :

Diketahui :

$$\bar{X}_{Butuh\ BK\ "Ya"} = 2.2862$$

$$S_{Butuh\ BK\ "Ya"}^2 = 0.5383$$

$$S_{Butuh\ BK\ "Ya"} = 0.7337$$

$$\bar{X}_{Butuh\ BK\ "Tidak"} = 3.0623$$

$$S_{Butuh\ BK\ "Tidak"}^2 = 0.2184$$

$$S_{Butuh\ BK\ "Tidak"} = 0.4674$$

$$P(\text{IP Sem 1} = 2.24 | \text{Butuh BK} = "Ya") = \frac{1}{\sqrt{2\pi \cdot 0.7337}} \exp\left[-\frac{(2.24 - 2.2862)^2}{2 \cdot 0.5383}\right] = 0.464956656$$

$$P(\text{IP Sem 1} = 2.24 | \text{Butuh BK} = "Tidak") = \frac{1}{\sqrt{2\pi \cdot 0.4674}} \exp\left[-\frac{(2.24 - 3.0623)^2}{2 \cdot 0.2184}\right] = 0.124147275$$

Probabilitas IP Sem 2 :

Diketahui :

$$\bar{X}_{Butuh\ BK\ "Ya"} = 2.1933$$

$$S_{Butuh\ BK\ "Ya"}^2 = 0.6014$$

$$S_{Butuh\ BK\ "Ya"} = 0.7755$$

$$\bar{X}_{Butuh\ BK\ "Tidak"} = 2.8590$$

$$S_{Butuh\ BK\ "Tidak"}^2 = 0.3038$$

$$S_{Butuh\ BK\ "Tidak"} = 0.5512$$

$$P(\text{IP Sem 2} = 2 | \text{Butuh BK} = "Ya") = \frac{1}{\sqrt{2\pi \cdot 0.7755}} \exp\left[-\frac{(2 - 2.1933)^2}{2 \cdot 0.6014}\right] = 0.439264892$$

$$P(\text{IP Sem 2} = 2 | \text{Butuh BK} = "Tidak") = \frac{1}{\sqrt{2\pi \cdot 0.5512}} \exp\left[-\frac{(2 - 2.859)^2}{2 \cdot 0.3038}\right] = 0.159598066$$

Probabilitas IP Sem 3 :

Diketahui :

$$\bar{X}_{Butuh\ BK\ "Ya"} = 2.0586$$

$$S_{Butuh\ BK\ "Ya"}^2 = 0.6042$$

$$S_{Butuh\ BK\ "Ya"} = 0.7773$$

$$\bar{X}_{Butuh\ BK\ "Tidak"} = 2.7867$$

$$S_{Butuh\ BK\ "Tidak"}^2 = 0.4326$$

$$S_{Butuh\ BK\ "Tidak"} = 0.6577$$

$$P(\text{IP Sem 4} = 1.86 | \text{Butuh BK} = "Ya") = \frac{1}{\sqrt{2\pi \cdot 0.7773}} \exp\left[-\frac{(1.86 - 2.0586)^2}{2 \cdot 0.6042}\right] = 0.438079908$$

$$P(IP \text{ Sem } 4 = 1.86 \mid \text{Butuh BK "Tidak"}) = \frac{1}{\sqrt{2\pi \cdot 0.6577}} \exp^{-\frac{(1.86 - 2.7867)^2}{2 \cdot 0.4326}} = 0.182397747$$

Probabilitas IP Sem 4 :  
Diketahui :

$$\bar{X}_{\text{Butuh BK "Ya"}} = 1.8976$$

$$S^2_{\text{Butuh BK "Ya"}} = 0.3966$$

$$S_{\text{Butuh BK "Ya"}} = 0.6298$$

$$\bar{X}_{\text{Butuh BK "Tidak"}} = 2.8233$$

$$S^2_{\text{Butuh BK "Tidak"}} = 0.2923$$

$$S_{\text{Butuh BK "Tidak"}} = 0.5406$$

$$P(IP \text{ Sem } 4 = 2.2 \mid \text{Butuh BK "Ya"}) = \frac{1}{\sqrt{2\pi \cdot 0.6298}} \exp^{-\frac{(2.2 - 1.8976)^2}{2 \cdot 0.3966}} = 0.448087164$$

$$P(IP \text{ Sem } 4 = 2.2 \mid \text{Butuh BK "Tidak"}) = \frac{1}{\sqrt{2\pi \cdot 0.5406}} \exp^{-\frac{(2.2 - 2.8233)^2}{2 \cdot 0.2923}} = 0.277548728$$

2) Langkah berikutnya adalah menghitung probabilitas akhir untuk setiap kelas :

$$\begin{aligned} P(X|\text{Butuh BK} = \text{Ya}) &= P(\text{Jenis Kelamin} = P \mid \text{Butuh BK} = \text{Ya}) \times P(\text{Status Wisuda} = B \mid \text{Butuh BK} = \text{Ya}) \times P(\text{Prodi} = \text{PKK} \mid \text{Butuh BK} = \text{Ya}) \times P(\text{Status Masuk} = \text{SPMB/SNMPTN} \mid \text{Butuh BK} = \text{Ya}) \times P(\text{Jenis SMTA} = \text{SMK} \mid \text{Butuh BK} = \text{Ya}) \times P(\text{Pekerjaan Ayah} = \text{Wiraswasta} \mid \text{Butuh BK} = \text{Ya}) \times P(\text{IP Sem 1} = 2.24 \mid \text{Butuh BK} = \text{Ya}) \times P(\text{IP Sem 2} = 2.0 \mid \text{Butuh BK} = \text{Ya}) \times P(\text{IP Sem 3} = 1.86 \mid \text{Butuh BK} = \text{Ya}) \times P(\text{IP Sem 4} = 2.2 \mid \text{Butuh BK} = \text{Ya}) \\ &= 0.2381 \times 0.7143 \times 0.1429 \times 0.0952 \times 0.4762 \times 0.2857 \times 0.464956656 \times 0.439264892 \times 0.438079908 \times 0.448087164 \\ &= 0.0000126213 \end{aligned}$$

$$\begin{aligned} P(X|\text{Butuh BK} = \text{Tidak}) &= P(\text{Jenis Kelamin} = P \mid \text{Butuh BK} = \text{Tidak}) \times P(\text{Status Wisuda} = B \mid \text{Butuh BK} = \text{Tidak}) \times P(\text{Prodi} = \text{PKK} \mid \text{Butuh BK} = \text{Tidak}) \times P(\text{Status Masuk} = \text{SPMB/SNMPTN} \mid \text{Butuh BK} = \text{Tidak}) \times P(\text{Jenis SMTA} = \text{SMK} \mid \text{Butuh BK} = \text{Tidak}) \times P(\text{Pekerjaan Ayah} = \text{Wiraswasta} \mid \text{Butuh BK} = \text{Tidak}) \times P(\text{IP Sem 1} = 2.24 \mid \text{Butuh BK} = \text{Tidak}) \times P(\text{IP Sem 2} = 2.0 \mid \text{Butuh BK} = \text{Tidak}) \times P(\text{IP Sem 3} = 1.86 \mid \text{Butuh BK} = \text{Tidak}) \times P(\text{IP Sem 4} = 2.2 \mid \text{Butuh BK} = \text{Tidak}) \\ &= 0.3667 \times 0.0667 \times 0.0667 \times 0.2667 \times 0.2000 \times 0.0333 \times 0.124147275 \times \end{aligned}$$

$$\begin{aligned} &0.159598066 \times 0.182397747 \times 0.277548728 \\ &= 0.0000000291 \end{aligned}$$

a. Selanjutnya nilai tersebut dimasukkan untuk mendapatkan probabilitas akhir :

$$P(\text{Butuh BK} = \text{Ya} \mid X) = \alpha \times 0.4118 \times 0.0000126213$$

$$= 0.0000051970$$

$$P(\text{Butuh BK} = \text{Tidak} \mid X) = \alpha \times 0.5882 \times 0.0000000291$$

$$= 0.0000000171$$

$\alpha = 1/P(X)$  nilainya konstan sehingga tidak perlu diketahui karena yang terbesar dari dua kelas tersebut tidak dapat dipengaruhi oleh  $P(X)$ . Karena nilai probabilitas akhir terbesar ada di kelas Butuh BK "Ya", data uji mahasiswa NIM 200807060 diprediksi sebagai kelas Butuh BK "Ya". Hasil pengujian atas nama Sesri Noviarlis NIM 200807060 dapat dilihat pada Tabel 2 berikut :

**Tabel 2. Hasil Klasifikasi**

ID MHS	Nama	Kelas:	Prediksi	Hasil
		Butuh BK	<i>Naïve Bayes</i>	
200807060	Sesri Noviarlis	Ya	Ya	AKURAT

Data mahasiswa tersebut diambil dari *data testing* dan masuk ke kategori yang membutuhkan proses bimbingan dan konseling. Setelah dilakukan perhitungan menggunakan *Naïve Bayes Classifier* hasilnya adalah masuk kelas Butuh BK "Ya". Hasil pengujian yang diperoleh adalah AKURAT.

## IMPLEMENTASI DAN PENGUJIAN

Berdasarkan pengujian yang telah dilakukan dengan berbagai variasi, diperoleh hasil seperti terangkum pada Tabel 3 berikut :

**Tabel 3 Rangkuman Hasil Pengujian**

No	Sampel <i>Testing</i>	Jumlah Data	Confusion Matrix	Accuracy	Plot Area Under ROC		
1	Fakultas : Mewakili mahasiswa dari seluruh Program Studi	200	Class	Ya	Tidak	81%	0.9363 <i>Excellent Classification</i>
				61	33		
2	Mahasiswa Program Studi Pendidikan Teknik Bangunan	56	Class	Ya	Tidak	82,1429 %	0.9505 <i>Excellent Classification</i>
				24	10		
3	Mahasiswa Program Studi Pendidikan Teknik Elektro	53	Class	Ya	Tidak	90,566 %	0.9856 <i>Excellent Classification</i>
				27	5		
4	Mahasiswa Program Studi Pendidikan Teknik Elektronika	65	Class	Ya	Tidak	84,6154 %	0.9233 <i>Excellent Classification</i>
				25	8		
5	Mahasiswa Program Studi Pendidikan Teknik Informatika dan Komputer	82	Class	Ya	Tidak	82,9268 %	0.9469 <i>Excellent Classification</i>
				11	14		
6	Mahasiswa Program Studi Pendidikan Teknik Mesin	67	Class	Ya	Tidak	79,1045 %	0.9944 <i>Good Classification</i>
				12	9		
7	Mahasiswa Program Studi Pendidikan Teknik Otomotif	66	Class	Ya	Tidak	81,8182 %	0.9407 <i>Excellent Classification</i>
				20	9		
8	Mahasiswa Program Studi Pendidikan Kesejahteraan	67	Class	Ya	Tidak	73,1343 %	0.9002 <i>Excellent Classification</i>
				13	15		

Menurut hasil klasifikasi menggunakan algoritma *Naive Bayes Classifier* seperti terangkum pada Tabel 3. diterapkan pada Fakultas Teknik Universitas Negeri Padang untuk memetakan atau mengelompokkan mahasiswa yang memiliki kinerja akademik rendah dan membutuhkan proses bimbingan dan konseling, hasilnya diperkirakan akan mendekati seperti Tabel 3.

Tingkat *Excellent Classification* (sangat baik) tercapai ketika model yang dibuat digunakan untuk mengklasifikasi mahasiswa tingkat Fakultas yang mewakili keseluruhan Prgram Studi, serta Program Studi Pendidikan Teknik Bangunan, Pendidikan Teknik Elektro, Pendidikan Teknik Elektronika, Pendidikan Teknik Informatika dan Komputer, Program Studi Pendidikan Teknik Otomotif dan Pendidikan Kesejahteraan Keluarga.

Program Studi Pendidikan Teknik Elektro. edangkan ketika digunakan untuk mengklasifikasi mahasiswa Program Studi Teknik Mesin tingkat klasifikasi yang didapatkan adalah *Good Classification* (Baik).

## KESIMPULAN

Berdasarkan analisis dan pengujian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut :

1. Teknik klasifikasi menggunakan Algoritma *Naive Bayes Classifier* sangat baik dalam melakukan pemetaan dan mengidentifikasi mahasiswa yang memiliki kinerja akademik rendah dan membutuhkan Bimbingan dan Konseling.
2. Tingkat akurasi yang dihasilkan proses klasifikasi menggunakan algoritma *Naive Bayes* sangat ditentukan oleh penentuan dan pemilihan *data training*.
3. Proses transformasi data dan pemilihan atribut *input* sangat berpengaruh terhadap *output* / hasil prediksi.
4. Atribut yang menjadi *output* pada penelitian ini adalah Butuh BK “Ya” dan “Tidak”. Mahasiswa yang masuk *Class* Butuh BK “Ya” diidentifikasi sebagai mahasiswa yang memiliki kinerja akademik rendah dan membutuhkan Bimbingan dan Konseling, sedangkan mahasiswa yang masuk *Class* Butuh BK “Tidak” adalah mahasiswa yang memiliki kinerja akademik normal dan tidak membutuhkan Bimbingan dan Konseling.

## Saran

Saran yang dapat diberikan untuk pengembangan lebih lanjut adalah sebagai berikut :

1. Pada penelitian ini, penulis hanya mencoba salah satu teknik klasifikasi untuk

memetakan mahasiswa yang membutuhkan proses Bimbingan dan Konseling yaitu teknik klasifikasi menggunakan Algoritma *Naïve Bayes Classifier*. Untuk mendapatkan hasil klasifikasi yang lebih baik dapat digunakan beberapa atau penggabungan beberapa teknik klasifikasi seperti penerapan Algoritma CART, ID3, C4.5, *Linear Discriminant Analysis* serta algoritma klasifikasi yang lain.

2. Untuk mengimplementasikan pola klasifikasi yang dihasilkan dari teknik klasifikasi menggunakan Algoritma *Naïve Bayes Classifier* tersebut, perlu dilakukan konsultasi lebih lanjut dengan pihak pengambil keputusan, sehingga dapat dipilih pola klasifikasi yang terbaik dalam pemetaan mahasiswa memiliki kinerja akademik rendah dan membutuhkan Bimbingan dan Konseling.

#### DAFTAR PUSTAKA

- [1] Afen Prana Utama Sembiring, et al, "Analisis Faktor-Faktor yang Mempengaruhi Pengunduran Mahasiswa dengan Aplikasi Data Mining ADD-INS – Studi Kasus Pada STMIK Mikroskil". *JSM STMIK Mikroskil*. Vol. 14. 139 -146, 2013.
- [2] Arif Junanto, "Algoritma *Naïve Bayes* untuk Mencari Perkiraan Waktu Studi Mahasiswa". *Jurnal Teknologi Informasi Dinamik*. Vol. 18. 9-16, 2013.
- [3] Arintoko, "Wawancara Konseling di Sekolah". Ed. I. Yogyakarta : ANDI, 2011.
- [4] Bustami, "Penerapan Algoritma *Naïve Bayes* untuk Mengklasifikasi Data Nasabah Asuransi". *Jurnal Informatika*. Vol. 8, 2014.
- [5] Eko Prasetyo, "Data Mining, Konsep dan Aplikasi Menggunakan Matlab". Ed. I. Yogyakarta : ANDI, 2012
- [6] Fajar Astuti Hermawati, "Data Mining". Ed. I. Yogyakarta : ANDI, 2013.
- [7] Kadek Suranata, "Pengembangan Model Tutor Bimbingan Konseling Sebaya (peer counseling) untuk Mengatasi Masalah Mahasiswa Fakultas Ilmu Pendidikan Undiksha". *Jurnal Pendidikan Indonesia*. Vol. 2. No. 2. 255-263, 2013.
- [8] Mujib Ridwan, Hadi Suyono, dan M. Sarosa, "Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naïve Bayes Classifier*". *Jurnal EECCIS*. Vol. 7. 59-64, 2013.
- [9] Rizal Amegia Saputra, "Komparasi Algoritma Data Mining untuk Memprediksi Penyakit Tuberculosis (TB) : Studi Kasus Pada Puskesmas Karawang Sukabumi". *Seminar Nasional Inovasi dan Tren (SNIT) 2014*. 1-8, 2014.
- [10] Sandro Romario. S, "Penerapan Data Mining Pada RSUP DR. Moh Hoesin Sumatera Selatan Untuk Pengelompokan Hasil Diagnosa Pasien Pengguna Asuransi Kesehatan Miskin". Universitas Binadarma 2013, 2013.
- [11] Suharsimi Arikunto, "Prosedur Penelitian : Suatu Pendekatan Praktik". Ed. Revisi VI, Jakarta : PT Rineka Cipta, 2006.
- [12] Sutirna, "Bimbingan dan Konseling. Pendidikan Formal, Nonformal dan Informal". Ed. I. Yogyakarta : ANDI, 2013