

## PREDIKSI JURUSAN PADA SELEKSI NASIONAL MASUK PERGURUAN TINGGI NEGERI (SNMPTN) MENGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES

Herliyani Hasanah<sup>1</sup>, Nugroho Arif Sudibyo<sup>2</sup>, Edy Kurniawan<sup>3</sup>

<sup>1,2,3</sup>Fakultas Ilmu Komputer, Universitas Duta Bangsa

Email: <sup>1</sup>herliyani\_hasanah@udb.co.id

### Abstrak

SNMPTN merupakan salah satu jalur masuk perguruan tinggi negeri yang banyak diminati siswa karena hanya menggunakan parameter nilai raport, prestasi siswa, dan prestasi sekolah. Setiap jurusan mempunyai nilai diterima minimal yang berbeda-beda, besaran kuota yang ditetapkan LTMP 2019 minimal hanya 20% dari daya tampung program studi di setiap perguruan tinggi negeri. Besarnya minat siswa dan kecilnya jumlah kuota tidak sebanding sehingga menyebabkan persaingan diterima pada jalur ini semakin ketat. Namun, masih banyak siswa yang belum mempertimbangkan parameter tersebut saat mendaftar sehingga kemungkinan diterima pada jalur SNMPTN semakin kecil. Oleh karena itu diperlukan sebuah sistem yang dapat memprediksi kemungkinan diterimanya siswa pada jurusan SNMPTN berdasarkan atribut yang sudah ditentukan. *Naïve Bayes* diterapkan untuk mencari nilai probabilitas terbesar dalam setiap variabel yang ada. Variabel yang digunakan meliputi nilai rata-rata matematika, bahasa indonesia, dan bahasa inggris semester 1 sampai 5 serta prestasi siswa yang dilampirkan saat mendaftar dan prestasi sekolah. Hasilnya dengan *naïve bayes* mampu menghasilkan akurasi sebesar 83,3%.

**Kata kunci:** SNMPTN, klasifikasi, *naïve bayes*, probabilitas

### Abstract

*SNMPTN is one of the entrance paths of state tertiary institutions which is of great interest to students because it only uses the parameters of report card, student achievement, and school achievement. Each department has a different acceptable minimum value, the quota amount specified in LTMP 2019 is at least only 20% of the capacity of study programs in each state university. The amount of student interest and the small amount of the quota are not comparable, causing competition to be accepted on this path more stringently. However, there are still many students who have not considered these parameters when registering so the possibility of being accepted on the SNMPTN path is getting smaller. Therefore we need a system that can predict the possibility of acceptance of students in the SNMPTN majors based on predetermined attributes. Naïve Bayes is applied to find the greatest probability value in every variable. The variables used include the average value of mathematics, Indonesian, and English semester 1 to 5 as well as student achievement that is attached when registering and school achievement. The result with Naïve Bayes is able to produce an accuracy of 83,3 %.*

**Keywords:** SNMPTN, classification, *naïve bayes*, probability

### PENDAHULUAN

Penerimaan mahasiswa baru telah mengalami perubahan dan penyempurnaan dari waktu ke waktu. Perubahan tersebut dilakukan dengan tujuan meningkatkan kualitas dan

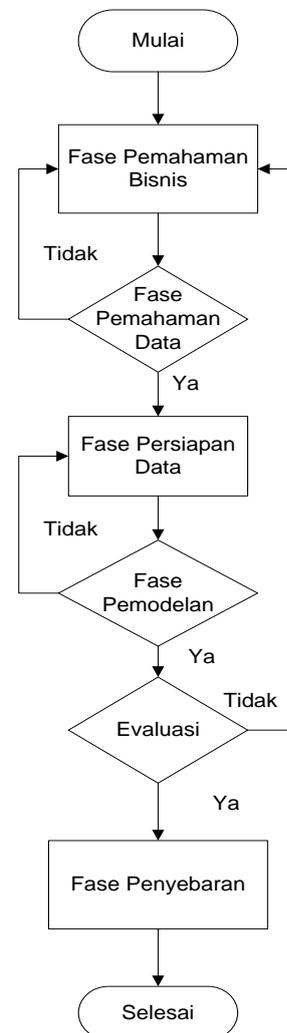
keefektifan sistem pendidikan di Indonesia (Lestari, 2016). Salah satu jalur masuk ke Perguruan Tinggi Negeri adalah melalui SNMPTN dimana besaran kuota yang ditetapkan oleh Lembaga Tes Masuk Perguruan Tinggi (LTMP) pada 2019

minimal 20% dari kuota daya tampung setiap prodi di PTN. Tujuan pelaksanaan SNMPTN meliputi : menjaring calon mahasiswa yang diprediksi mampu menyelesaikan studi di perguruan tinggi dengan baik serta memberi peluang bagi calon mahasiswa yang memilih lebih dari satu PTN lintas wilayah (Ratno, 2016). SNMPTN merupakan seleksi nasional masuk perguruan tinggi negeri berdasarkan nilai rapor dan prestasi-prestasi lain. Kemudahan ini menyebabkan SNMPTN menjadi jalur yang banyak diminati oleh siswa, tetapi daya tampung yang terbatas menyebabkan siswa harus menentukan pilihan jurusan dan universitas dengan tepat agar dapat diterima melalui jalur SNMPTN (Prabowo, 2015). Tetapi masih banyak siswa yang tidak mempertimbangkan hal tersebut dan memilih jurusan tanpa berdasarkan nilai dan prestasi mereka, sehingga kemungkinan mereka diterima disalah satu jurusan di SNMPTN semakin kecil. Dari permasalahan tersebut maka dibutuhkan sistem pendukung keputusan yang baik sehingga dapat menghasilkan suatu keputusan yang objektif. Sistem pendukung keputusan dapat menyelesaikan beragam persoalan yang berhubungan dengan banyaknya kriteria maupun alternatif. Pada penelitian sebelumnya naïve bayes juga digunakan dalam memprediksi besaran penggunaan listrik rumah tangga berdasarkan beberapa variabel, dan hasilnya diperoleh hasil persentase 78,3333% untuk keakuratan prediksi (Saleh, 2015). Pada penelitian lain naïve bayes digunakan untuk memprediksi ranting penjualan buku, penelitian ini bertujuan untuk mengetahui faktor yang mempengaruhi ranting penjualan buku untuk menghasilkan sebuah buku yang memenuhi kriteria pembaca, penelitian tersebut diperoleh tingkat akurasi sebesar 74,60%. Penerapan gabungan antara metode algoritma Information Gain dengan Naïve Bayes juga mampu menghasilkan nilai akurasi 90% dalam memprediksi keterlambatan pembayaran SPP sekolah (Muqorobin, 2019). Tujuan dari penelitian ini adalah membuat sebuah

sistem yang dapat memprediksi kemungkinan diterimanya siswa pada jurusan SNMPTN berdasarkan atribut yang sudah ditentukan dengan menggunakan metode Naïve Bayes. Dengan sistem ini diharapkan calon mahasiswa kemungkinan mereka diterima disalah satu jurusan di SNMPTN semakin besar.

**METODE PENELITIAN**

Alur penelitian ini menggunakan metode *Cross Industry Standard Process Data Mining* (CRISP-DM). Berikut adalah tahapan penelitian dengan metode CRISP-DM :



Gambar 1. Flowchart tahapan penelitian metode CRISP-DM

Pada gambar 1 menunjukkan enam tahapan metode CRISP-DM yang meliputi :

- a. Fase pemahaman bisnis (*business understanding phase*). Pada fase ini ada beberapa langkah yang dilakukan meliputi menentukan tujuan bisnis, menganalisis situasi serta menentukan tujuan menggunakan data mining sebagai solusi.
- b. Fase pemahaman data (*data understanding phase*): Pada fase ini langkah - langkah yang dilakukan meliputi pengumpulan data awal, deskripsi data, eksplorasi data dan verifikasi kualitas data. Data yang digunakan pada penelitian ini didapatkan dari data publik hasil survei SNMPTN UI 2017 oleh halo kampus. Data yang didapatkan akan digunakan untuk bahan penganalisisan data terhadap metode Naïve Bayes. Setelah data dikumpulkan dilakukan analisa data untuk menyesuaikan proses data yang akan diolah pada metode Naïve Bayes.
- c. Fase persiapan data (*data preparation phase*). Pada tahap ini dilakukan beberapa langkah yang meliputi pembersihan data, seleksi data dan transformasi data.
- d. Fase pemodelan (*modelling phase*) Pada tahap ini dilakukan pemodelan klasifikasi menggunakan Algoritma C45.
- e. Fase evaluasi (*evaluation*), Pada tahap evaluasi ini dinilai besarnya kekuatan model yang terbentuk. Untuk mengukur keakuratan dari model yang dibuat menggunakan Software Rapid Miner
- f. Penyebaran (*deployment*). Pada tahap ini

**HASIL DAN PEMBAHASAN**

**Fase pemahaman bisnis (*business understanding phase*).**

Pada fase ini ada beberapa langkah yang dilakukan meliputi menentukan tujuan bisnis, menganalisis situasi serta menentukan tujuan menggunakan data mining sebagai solusi.

**Fase pemahaman data (*data understanding phase*).**

Pada fase ini langkah - langkah yang dilakukan meliputi pengumpulan data awal, deskripsi data, eksplorasi data dan verifikasi kualitas data.

**Fase persiapan data (*data preparation phase*).**

Pada tahap ini dilakukan beberapa langkah yang meliputi pembersihan data, seleksi data dan transformasi data.

Penelitian ini menggunakan dua jenis data dari hasil pengumpulan yaitu, data pelatihan dan data uji. Tabel 1 merupakan hasil dari data yang sudah dilakukan cleaningdata, yaitu pembersihan data yang tidak terpakai pada proses perhitungan data mining.

Tabel 1. Data Latih

Jurusa n	Mtk	Bhs. Ing	Bhs. Indo	P. sis wa	P. sklh wa
Matem a-tika	92	86	88	Ya	Cu- kup
Matem a-tika	90	86	85	Ya	Baik
Sastra Inggris	82	87	92	Tid ak	Cu- kup
Matem a-tika	93	82	85	Ya	Cu- kup
Sastra Inggris	86	85	88	Ya	Cu- kup
Matem a-tika	91	83	84	Tid ak	Baik
Sastra Indone sia	88	80	90	Tid ak	Cu- kup
Sastra Indone sia	84	82	88	Ya	Cu- kup
Sastra Inggris	79	89	85	Ya	Cu- kup
Sastra Inggris	83	88	81	Tid ak	Baik
Sastra Inggris	82	85	84	Tid ak	Cu- kup

Keterangan :

- Mtk : nilai matematika
- Bhs.Ing : nilai bahasa inggris
- Bhs.Indo : nilai bahasa indonesia
- P.siswa : prestasi siswa

P.skhl : prestasi sekolah

Tabel 2 merupakan data uji yang akan digunakan untuk menguji hasil dari pengolahan data latih dengan metode naïve bayes. Isi dari data pelatihan di kategorikan berdasarkan hasil dari penentuan variabel dan konversi data.

Tabel 2. Data Uji

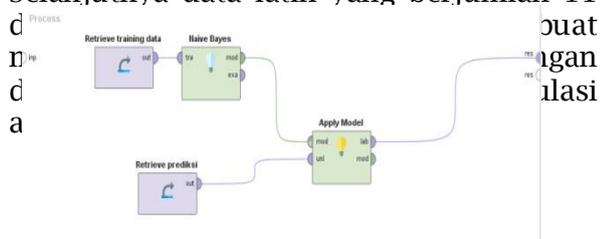
Jurusan	Mtk	Bhs. Ing	Bhs. Indo	P. siswa	P. sklh
Matematika	92	82	85	Ya	Cukup
Matematika	91	83	85	Tidak	Baik
Sastra Inggris	82	80	88	Ya	Cukup
Sastra Indonesia	88	80	90	Tidak	Cukup
Sastra Inggris	83	85	84	Tidak	Cukup
Sastra Indonesia	88	82	88	Ya	Cukup

Keterangan :

- Mtk : nilai matematika
- Bhs.Ing : nilai bahasa inggris
- Bhs.Indo : nilai bahasa indonesia
- P.siswa : prestasi siswa
- P.skhl : prestasi sekolah

**Fase pemodelan (modelling phase).**

Pada tahap ini dilakukan pemodelan klasifikasi menggunakan Algoritma C45. Selanjutnya data latih yang berjumlah 11



Attribute	Parameter	Matematika	Sastra Inggris	Sastra Indonesia	Sastra Inggris
Matematika	mean	91.500	82.250	86	83
	standard deviation	1.291	2.872	2.828	0.001
Bahasa Inggris	mean	84.250	86.500	81	88
	standard deviation	2.982	1.915	1.414	0.001
Bahasa Indonesia	mean	85.500	87.250	89	81
	standard deviation	1.732	3.584	1.414	0.001
Prestasi Siswa	value=Ya	0.750	0.500	0.500	0
	value=Tidak	0.250	0.500	0.500	1
	value=unknown	0	0	0	0
Prestasi Sekolah	value=Cukup	0.500	1	1	0
	value=Baik	0.500	0	0	1
	value=unknown	0	0	0	0

Gambar 2. Hasil Naïve bayes

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
#memasukan data Latih
datalatih = pd.read_excel("data testing.xlsx")
datalatih.head(11)
#Prestasi Siswa = 1 == Ya
#Prestasi Siswa = 2 == Tidak
#Prestasi Sekolah = 1 == Cukup
#Prestasi Sekolah = 2 == Baik

Out[1]:
```

	Jurusan	Matematika	Bahasa Inggris	Bahasa Indonesia	Prestasi Siswa	Prestasi Sekolah
0	Matematika	92	86	88	1	1
1	Matematika	90	86	85	1	2
2	Sastra Inggris	82	87	92	2	1
3	Matematika	93	82	85	1	1
4	Sastra Inggris	86	85	88	1	1
5	Matematika	91	83	84	2	2
6	Sastra Indonesia	88	80	90	2	1
7	Sastra Indonesia	84	82	88	1	1
8	Sastra Inggris	79	89	85	1	1
9	Sastra Inggris	83	88	81	2	2
10	Sastra Inggris	82	85	84	2	1

Gambar 3. Hasil Prediksi Naïve bayes

Row No.	Jurusan	prediction(Jurusan)	confidence...	confidence...	confidence...	confidence...	Matematika	Bahasa Ing...
1	Matematika	Matematika	0.998	0.000	0.002	0	92	82
2	Matematika	Matematika	1	0	0	0	91	83
3	Sastra Inggris	Sastra Indonesia	0.000	0.008	0.992	0	82	80
4	Sastra Indon...	Sastra Indonesia	0.000	0.000	0.999	0	88	80
5	Sastra Inggris	Sastra Inggris	0.000	1.000	0.000	0	83	85
6	Sastra Indon...	Sastra Indonesia	0.019	0.010	0.972	0	88	82

```
In [17]: Y_predict = nbtrain.predict(x_test)
print("Prediksi Naive Bayes : ",Y_predict)

Prediksi Naive Bayes : ['Matematika' 'Matematika' 'Sastra Indonesia' 'Sastra Indonesia'
'Sastra Inggris' 'Sastra Indonesia']

In [16]: from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_uji, Y_predict)
print("Akurasi Naive Bayes : ",accuracy)

Akurasi Naive Bayes : 0.8333333333333334

In [12]: # Menghitung nilai akurasi dari klasifikasi naive bayes
from sklearn.metrics import classification_report
print(classification_report(y_uji, Y_predict))

              precision    recall  f1-score   support

Matematika      1.00      1.00      1.00         2
Sastra Indonesia  0.67      1.00      0.80         2
Sastra Inggris   1.00      0.50      0.67         2

accuracy
macro avg      0.89      0.83      0.82         6
weighted avg   0.89      0.83      0.82         6
```

Gambar 5. Data Uji

Berdasarkan proses perhitungan dengan naïve bayes di phyton didapatkan hasil prediksi yang ditampilkan dalam bentuk array dan nilai akurasi sebesar 0.8333 dalam skala 1 serta nilai *recall* dan *precision* seperti dalam gambar 6.

accuracy: 83.33%

	true Matematika	true Sastra Inggris	true Sastra Indonesia	true Sastra Inggris	class precision
pred. Matematika	2	0	0	0	100.00%
pred. Sastra Inggris	0	1	0	0	100.00%
pred. Sastra Indone..	0	1	2	0	66.67%
pred. Sastra Inggris	0	0	0	0	0.00%
class recall	100.00%	50.00%	100.00%	0.00%	

Gambar 6. Hasil Perhitungan

**Fase evaluasi (evaluation).**

Pada tahap evaluasi ini dinilai besarnya kekuatan model yang terbentuk. Untuk mengukur keakuratan dari model yang dibuat menggunakan Software Rapid Miner. Dari 6 data uji yang dilakukan pengujian dengan naïve bayes terdapat 5 data yang prediksinya sesuai dengan data real dan 1 data yang prediksinya berbeda dengan data real seperti pada gambar 3. Dan berdasarkan hasil prediksi tersebut didapatkan nilai akurasi sebesar 83.33%. Hasil evaluasi simulasi model di Rapid Miner sama dengan hasil evaluasi dengan implementasi di phyton.

```
In [8]: from sklearn.model_selection import train_test_split
# Import Gaussian Naive Bayes model
from sklearn.naive_bayes import GaussianNB
# Mengaktifkan/memanggil/membuat fungsi klasifikasi Naive bayes
modelnb = GaussianNB()
# Memasukkan data training pada fungsi klasifikasi naive bayes
nbtrain = modelnb.fit(x, y)
datauji = pd.read_excel("data uji akurasi.xlsx")
datauji.head(11)
```

Out[8]:

	Jurusan	Matematika	Bahasa Inggris	Bahasa Indonesia	Prestasi Siswa	Prestasi Sekolah
0	Matematika	92	82	85	1	1
1	Matematika	91	83	85	2	2
2	Sastra Inggris	82	80	88	1	1
3	Sastra Indonesia	88	80	90	2	1
4	Sastra Inggris	83	85	84	2	1
5	Sastra Indonesia	88	82	88	1	1

Gambar 7 Hasil Evaluasi dengan Rapid Miner

**Penyebaran (deployment).**

Pada tahap ini penyebaran berupa penggunaan aplikasi yang memanfaatkan phyton. Dimana data yang dimasukkan berupa *file Microsoft Excel*.

**Kesimpulan dan Saran**  
**Kesimpulan**

1. Sistem dapat memprediksi kemungkinan diterimanya siswa pada jurusan SNMPTN berdasarkan atribut nilai rata-rata matematika, bahasa indonesia, dan bahasa inggris

semester 1 sampai 5 serta prestasi siswa dan prestasi sekolah dengan menggunakan metode Naïve Bayes

2. Hasil pengujian menggunakan simulasi Rapid Miner serta pembuatan aplikasi melalui *python* mendapatkan hasil yang sama yaitu 83.33% nilai akurasi dan 5 data uji yang sesuai prediksi.

#### Saran

1. Kombinasi dengan metode lain untuk meningkatkan hasil akurasi dalam prediksi kemungkinan diterimanya dalam jurusan SNMPTN berdasarkan atribut yang ditentukan.
2. Penambahan fitur hak akses untuk pengguna.

Penjualan Buku Menggunakan Metode Naive Bayes Sulastri. *Duta.Com*. Vol. 12, No. 2. Hlm. 57-72.

Muqorobin, E. T. L dan Kusrini. 2019. Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah. *J. Ilm. SINUS*. Vol. 17, No. 1.

#### DAFTAR PUSTAKA

- Lestari, D., Anis, K., Khaeruddin., K. 2018. Analisis Hasil Belajar Mahasiswa Jurusan Fisika Berdasarkan Jalur Masuk (SNMPTN, SBMPTN dan Mandiri) Universitas Negeri Makassar. *Jurnal Sains dan Pendidikan Fisika (JSPF)*. Jilid 14, Nomor 3. ISSN Online: 2548-6373. Hlm. 15-25.
- Ratno, P. 2016. Perbedaan Kebugaran Jasmani Mahasiswa Baru Melalui Jalur SNMPTN Undangan Dengan Jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri SBMPTN FIK UNIMED Tahun 2015. *Jurnal Ilmu Keolahragaan*. Vol. 15, Nomor 1. ISSN 2549-9777. Hlm. 38 - 47.
- Prabowo, Y.S., Kusrini., dan Sunyoto, A. 2015. Sistem Pendukung Keputusan Pemilihan Jurusan SNMPTN Bagi Siswa SMAN 7 Purworejo. Konferensi Nasional Sistem & Informatika 2015. STMIK STIKOM Bali. Hlm. 180 - 185.
- Saleh, A. 2015. Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creat. Inf. Technol. J*. Vol. 2, No. 3. Hlm. 207-217.
- Nugroho, Y., S. 2017. Penerapan Data Mining Untuk Prediksi Rating