

PREDIKSI PEMINATAN PROGRAM STUDI PADA PENERIMAAN MAHASISWA STMIK ROYAL MENGGUNAKAN NAÏVE BAYES

Wiwin Handoko¹, Muhammad Iqbal²

^{1,2}Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Royal
Email: ¹win.van.handoko@gmail.com, ²codeegoc@gmail.com

Abstract: STMIK Royal has 2 majors, namely Information Systems and Computer Systems. Almost every year at the time of admission of new students, there is a decrease in the interest of prospective students who choose the Computer Systems major, while the Information Systems major shows an increasing trend. Especially for the Computer Systems department, this decline will certainly have an impact on the development and quality of the department in the future. To overcome this, it is possible to predict the specialization of majors in new student admissions. If the prediction that appears is the Computer Systems major, the prospective student will be given a special reward so that it is hoped that interest for other prospective students will increase. Naïve Bayes is one of the classification algorithms in data mining that can be used to predict this. Based on the tests that have been carried out, the system can provide predictions with an accuracy rate of 65%.

Keywords: Prediction, Majors, Naïve Bayes, Accuracy

Abstrak: STMIK Royal memiliki 2 Program Studi yaitu Sistem Informasi dan Sistem Komputer. Hampir tiap tahun pada saat penerimaan mahasiswa baru, terjadi penurunan minat calon mahasiswa yang memilih Program Studi Sistem Komputer sedangkan untuk Program Studi Sistem Informasi menunjukkan tren yang terus meningkat. Khusus untuk Program Studi Sistem Komputer, penurunan ini tentunya akan berdampak bagi perkembangan dan kualitas Program Studi tersebut kedepannya. Untuk mengatasi hal tersebut dapat dilakukan prediksi peminatan Program Studi pada penerimaan mahasiswa baru. Apabila prediksi yang muncul adalah Program Studi Sistem Komputer, maka calon mahasiswa tersebut akan diberikan *reward* khusus sehingga diharapkan minat untuk calon mahasiswa yang lain akan semakin meningkat. Naïve Bayes merupakan salah satu Algoritma klasifikasi dalam data mining yang dapat digunakan untuk memprediksi hal tersebut. Berdasarkan pengujian yang telah dilakukan, sistem dapat memberikan prediksi dengan tingkat akurasi 65%.

Kata kunci: Prediksi, Program Studi, Naïve Bayes, akurasi

PENDAHULUAN

Setiap Perguruan Tinggi terutama kampus swasta tentunya berharap dapat menghasilkan lulusan yang berkualitas dan berdaya saing. Namun dalam menghasilkan lulusan yang berkualitas tentunya tidak terlepas dari input calon mahasiswa dalam proses penerimaan mahasiswa baru (Suryadi & Harahap, 2018). STMIK Royal merupakan salah satu kampus swasta yang berada di Kota

Kisaran Kabupaten Asahan Provinsi Sumatera Utara. Saat ini STMIK Royal memiliki 2 Prodi yaitu Sistem Informasi dan Sistem Komputer. Hampir tiap tahun pada saat penerimaan mahasiswa baru, terjadi penurunan minat calon mahasiswa yang memilih Prodi Sistem Komputer sedangkan untuk Prodi Sistem Informasi menunjukkan tren yang terus meningkat. Khusus untuk Prodi Sistem Komputer, penurunan ini tentunya akan berdampak bagi perkembangan dan kualitas prodi

tersebut kedepannya. Tidak adanya mahasiswa, bahkan penutupan prodi bisa saja terjadi. Permasalahan ini harus segera ditangani oleh pimpinan akademik maupun yayasan dalam rangka meningkatkan kembali minat dan jumlah mahasiswa prodi Sistem Komputer. data-data yang terkait harus segera dikumpulkan dan dianalisa untuk mendapatkan alternatif solusi yang baik. Salah satu alternatif solusi yang dapat dilakukan adalah pencarian *insight* pada dataset penerimaan mahasiswa baru gelombang pertama Tahun Akademik 2021/2022. Dari *insight* tersebut diharapkan dapat memberikan gambaran distribusi sebaran asal daerah mahasiswa yang memilih Program Studi Sistem Informasi dan Sistem Komputer. selanjutnya dari data sebaran tersebut akan dibuatkan sebuah model yang akan dapat memprediksi peminatan program studi pada penerimaan mahasiswa baru pada gelombang yang berikutnya. Model yang digunakan pada penelitian ini adalah *Naïve Bayes*. Apabila prediksi yang muncul adalah Prodi Sistem Komputer, maka calon mahasiswa tersebut akan diberikan *reward* khusus sehingga diharapkan minat untuk calon mahasiswa akan semakin meningkat.

Data mining adalah proses ekstraksi untuk menemukan sebuah arti, pola dan kebiasaan baru dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan pengenalan pola seperti teknik statistik dan matematika (Mafakhir & Solichin, 2020). Data mining terbagi beberapa kelompok pembahasan salah satunya adalah klasifikasi (Classifier). Dalam klasifikasi terdapat sebuah metode yaitu *Naïve Bayes*. Algoritme *Naïve Bayes* merupakan suatu metode pengklasifikasian data dengan model statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes dan dapat digunakan untuk memprediksi probabilitas keanggotaan pada suatu kelas yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya (Lianda & Atmaja, 2021).

Berikut bentuk umum dari teorema *Naïve Bayes*:

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

Keterangan:

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

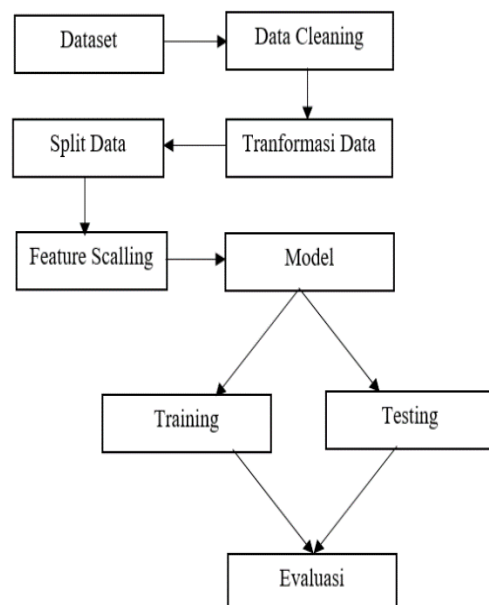
P(H) = Probabilitas hipotesis H (prior prob.)

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

P(X) = Probabilitas dari X

METODE

Metodologi Penelitian merupakan tahapan sistematis yang harus dilalui dalam penelitian untuk membantu penelitian menjadi terarah dengan baik (Ristianto & Yoraeni, 2021). Pengolahan data pada penelitian ini menggunakan python. Berikut tahapan yang dilakukan untuk melakukan prediksi dengan menggunakan *naïve bayes* pada penelitian ini:



Gambar 1. Tahapan Pengolahan Data

Dataset

Dataset yang digunakan pada penelitian ini adalah data penerimaan mahasiswa baru gelombang pertama Tahun Akademik 2021/2022 dalam bentuk file excel (.xlsx)

Data Cleaning

Data Cleaning merupakan tahapan yang dilakukan untuk menangani data apabila terjadi missing value, membuang kolom yang tidak terlalu digunakan maupun imbalance pada data.

Transformasi data

Tranformasi data merupakan tahapan preprocessing yang dilakukan untuk mengubah data agar dapat dibaca oleh mesin.

Split Data

Split data merupakan tahapan yang dilakukan untuk membagi dataset menjadi dua bagian yaitu data training dan data testing

Feature scalling

Feature scalling merupakan tahapan yang digunakan untuk melakukan normalisasi atau standarisasi nilai yang ada variabel fitur.

Model

Model merupakan tahapan yang digunakan untuk membuat sebuah model yang akan digunakan untuk mengolah data.

Evaluasi

Evaluasi merupakan tahapan yang digunakan untuk mengukur sejauh mana kualitas hasil dari model yang dipakai.

HASIL DAN PEMBAHASAN

Pada bagian ini, akan dijelaskan hasil penelitian yang sudah di uji coba dengan model prediksi sesuai dengan pembahasan yang ada di bab sebelumnya

mulai dari import dataset sampai evaluasi dengan menggunakan python.

```
Import Library
In [1]: import pandas as pd
import numpy as np

Import Dataset
In [2]: df = pd.read_excel('mahasiswa')
df
Out[2]:
```

NO	NAMA MAHASISWA	TTL	JENIS KELAMIN	ALAMAT LENGKAP	KABUPATEN/KOTA	PROVINSI	ASAL SEKOLAH	JENJANG	TAHUN LULUS	JURUSAN
0	1	Irfan Asidq Masya P	Male	Desa III Kal Mengkal Lama Kec. Ims Pulau Kab...	BATUBARA	SUMATERA UTARA	SMA Negeri 1 Luhk Putih Evulbas	SMA	2019	SI
1	2	Sul Ranjaya	Ar Genteng	Ar Genteng Kec. Ar Batu Asahan	ASAHAN	SUMATERA UTARA	SMA Negeri 1 Sibang Empat Asahan	SMA	2020	SI
420	421	YAZID BUSTON NASUTION	Male	Hala II Teluk Lapatr Kec. Teluk Lapatr Kec.	SIMALUNGUN	SUMATERA UTARA	SMK Swasta Al Mubtahir Kualan	SMK	2021	SI
421	422	RISKA SYDIA AL NUDA	Sumud	Dusun I Upeh Sukan Kec. Jo Jomb. Kab. Paseran	ASAHAN	SUMATERA UTARA	SMA Negeri 1 Kibaran Asahan	SMA	2021	SI

422 rows x 11 columns

Gambar 2. Import Dataset

Pada gambar 2 dataset yang diimport kedalam jupyter notebook adalah data penerimaan mahasiswa baru gelombang pertama dimana terdapat 422 baris dan 11 kolom. Tahap selanjutnya adalah data cleaning

Data Cleaning

```
In [3]: #Cara mengecek Missing Value pada Dataset
df.isna().sum()
Out[3]:
```

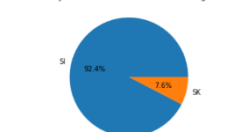
NO	0
NAMA MAHASISWA	0
TTL	0
JENIS KELAMIN	0
ALAMAT LENGKAP	0
KABUPATEN/KOTA	0
PROVINSI	0
ASAL SEKOLAH	0
JENJANG	0
TAHUN LULUS	0
JURUSAN	0
dtype:	int64

Gambar 3. Mengecek Missing Value

Pada gambar 3, setelah dilakukan data cleaning, missing value tidak ditemukan pada dataset. selanjutnya kita dapat melakukan Visualisasi untuk melihat insight pada dataset. Kolom Jurusan pada Dataset ini merupakan variabel target sedangkan kolom yang lainnya variabel fitur.

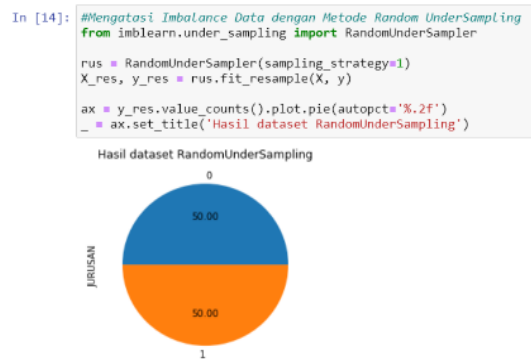
```
In [5]: #Import library untuk visualisasi
import matplotlib.pyplot as plt
import seaborn as sns

#Visualisasi data untuk Menampilkan Persentase Jumlah Mahasiswa berdasarkan Program Studi
plt.pie(df['JURUSAN'].value_counts(), labels=df['JURUSAN'].value_counts().index, autopct = '%1.1f%%')
plt.title('Persentase Jumlah Mahasiswa berdasarkan Program Studi')
Out[5]: Text(0.5, 1.0, 'Persentase Jumlah Mahasiswa berdasarkan Program Studi')
```



Gambar 4. Visualisasi Dataset

berdasarkan hasil visualisasi diatas, kita dapat menemukan Imbalance data, dimana Jumlah mahasiswa untuk yang memilih Program Studi Sistem Informasi sebanyak 92% (390 Orang) dan Sistem Komputer 7.6% (32 Orang). untuk mengatasi hal tersebut kita dapat menggunakan berbagai metode. pada kasus ini akan menggunakan metode Random Undersampling



Gambar 5. Visualisasi Setelah Resample

Pada gambar 5, dataset sudah menjadi balance dengan komposisi 50 : 50 untuk masing-masing program studi (32 orang untuk Sistem Informasi dan 32 orang untuk Sistem Komputer). Tahap berikutnya adalah tranformasi data



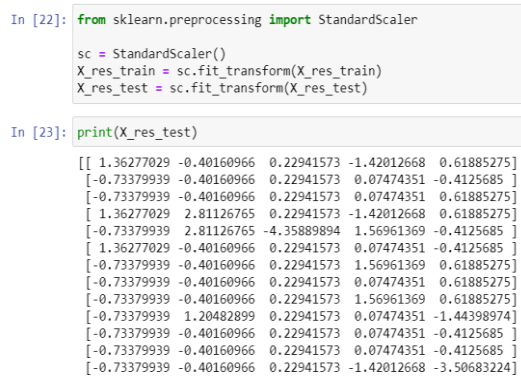
Pada gambar 6, dilakukan tranformasi data dengan menggunakan

metode Label Encoding untuk kolom yang akan digunakan dalam pengolahan data.



Gambar 7. Split Data

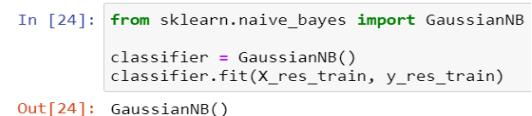
Pada gambar 7, Selanjutnya kita melakukan split data untuk membagi dataset menjadi dua bagian yaitu data training dan data test. pada kasus ini dataset dibagi 70% untuk data training dan 30% untuk data test.



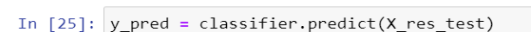
Gambar 8. Feature Scalling

Pada gambar 8 dilakukan feature scalling. Feature Scalling yang digunakan pada kasus ini adalah Standarization.

Training



Testing (Predict Result)



Gambar 9. Pembuatan Model

Pada gambar 9, dibuat sebuah model naïve bayes untuk melakukan training dan testing pada dataset.

Evaluasi

```
In [26]: from sklearn.metrics import confusion_matrix, accuracy_score
         ac = accuracy_score(y_res_test, y_pred)
```

```
In [27]: print("Tingkat Akurasi dengan menggunakan Naive Bayes sebesar ", ac)
```

Tingkat Akurasi dengan menggunakan Naive Bayes sebesar 0.65

Gambar 10. Evaluasi

Pada gambar 10, dilakukan evaluasi untuk mengukur kualitas dari hasil model yang telah dibuat. Berdasarkan pengujian untuk kasus ini, tingkat akurasi dengan menggunakan naïve bayes adalah 65%

SIMPULAN

Berdasarkan penelitian dan pengujian yang telah dilakukan dapat disimpulkan bahwa algoritma naïve bayes dapat melakukan prediksi terhadap peminatan program studi pada penerimaan mahasiswa baru STMIK Royal dengan tingkat akurasi 65%.

DAFTAR PUSTAKA

- Lianda, D., & Atmaja, N. S. (2021). Prediksi Data Buku Favorit Menggunakan Metode Naïve Bayes (Studi Kasus: Universitas Dehasen Bengkulu). *Pseudocode*, 8(1), 27–37. <https://doi.org/10.33369/pseudocode.8.1.27-37>
- Mafakhir, A. Z., & Solichin, A. (2020). Penerapan Metode Naïve Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta. *Fountain of Informatics Journal*, 5(1), 21. <https://doi.org/10.21111/fij.v5i1.4007>
- Ristiano, F., & Yoraeni, A. (2021). Implementasi Metode Naive Bayes Untuk Prediksi Harga Emas. *Jurnal CO-SCIENCE (Computer Science)*, 1(1), 62–71. <http://jurnal.bsi.ac.id/index.php/co-science/article/view/201>
- Suryadi, A., & Harahap, E. (2018). Sistem Rekomendasi Penerimaan Mahasiswa Baru Menggunakan Naive Bayes Classifier Di Institut Pendidikan Indonesia. *Joutica*, 3(2), 171. <https://doi.org/10.30736/jti.v3i2.231>