

Optimalisasi Metode *Naive Bayes* untuk Menentukan Program Studi bagi Calon Mahasiswa Baru dengan Pendekatan *Unsupervised Discretization*

Wildani Eko Nugroho^{1*}, Teguh Prihandoyo², Oman Somantri³

^{1,2}Program Studi Teknik Komputer, Politeknik Harapan Bersama Tegal

³Program Studi Teknik Informatika, Politeknik Negeri Cilacap

^{1,2} Jl. Mataram No.9 Pesurungan Lor, Kota Tegal, 52147, Indonesia

³Jln. Dr. Soetomo No.1 Karangcengis Sidakaya, Kabupaten Cilacap, 53212, Indonesia

E-mail: wild4n1@gmail.com¹, m_teguh_70@yahoo.co.id², oman_mantri@yahoo.com³

Info Naskah:

Naskah masuk: 30 November 2021
Direvisi: 30 Desember 2021
Diterima: 26 Januari 2022

Abstrak

Penerimaan calon mahasiswa baru harus mempertimbangkan berbagai prosedur yang bertujuan untuk mengarahkan calon mahasiswa baru dalam menentukan program studi yang ditempuh dan diminati. Penelitian ini akan membahas tentang Optimalisasi metode *Naive Bayes* untuk menentukan program studi atau jurusan bagi calon mahasiswa baru dengan pendekatan metode *Unsupervised Discretization*. Ada beberapa tahapan metode penelitian yang dilakukan dalam penelitian ini, diantaranya *Cleaning Data*, *Collection Data*, Penentuan Kriteria, Penentuan Probabilitas, serta *testing Data*. Penelitian ini sudah pernah dilakukan dengan metode yang sama yaitu metode *Naive Bayes* yang dipakai buat klasifikasi minat calon mahasiswa baru pada menentukan program studi dengan hasil nilai akurasi 96.68%. Penelitian yang sedang berjalan digunakan metode yang sama yaitu *Naive Bayes* lalu dilakukan optimalisasi dengan pendekatan metode *Unsupervised Discretization*. Untuk data *testing* ada 1671 record data siswa. Setelah dilakukan pengujian dengan metode yang sama dan dioptimalisasikan, maka nilai akurasi yang tadinya sebesar 96.68% menjadi 97.66% dengan hasil klasifikasi menunjukan pada program studi DIII Farmasi. Tujuan dari penelitian ini adalah menghasilkan klasifikasi dalam menentukan program studi atau jurusan bagi calon mahasiswa baru memakai metode *Naive Bayes* menggunakan optimalisasi pendekatan metode *Unsupervised Discretization*. Dari hasil pengujian data tersebut maka metode *Naive Bayes* setelah dilakukan optimalisasi dengan pendekatan metode *Unsupervised Discretization* sangat baik dibandingkan sebelum dilakukan optimalisasi.

Abstract

The admission of prospective new students must consider various procedures to direct prospective new students in determining the study program they are interested in. This study will discuss the optimization of the Naive Bayes method to determine the study program or major for prospective new students with the Unsupervised Discretization method approach. There are several stages of research methods carried out in this study, including Data Cleaning, Data Collection, Criteria Determination, Probability Determination, and Data Testing. This research has been carried out using the same method, namely the Naive Bayes method which is used to classify the interests of prospective new students in determining the study program with an accuracy value of 96.68%. Ongoing research uses the same method, namely Naive Bayes, then optimization is carried out with the Unsupervised Discretization method approach. For data testing, there are 1671 student data records. After testing with the same method and optimizing it, the accuracy value from 96.68% became 97.66% with the classification results showing the DIII Pharmacy study program. The purpose of this research is to produce a classification in determining the study program or major for prospective new students using the Naive Bayes method by the optimization of the Unsupervised Discretization method. From the results of testing the data, the Naive Bayes method after optimization with the Unsupervised Discretization method is very good compared to the application before optimization.

Keywords:

interests;
naive bayes;
unsupervised discretization;
study

*Penulis korespondensi:

Wildani Eko Nugroho

E-mail: wild4n1@gmail.com

1. Pendahuluan

Pemanfaatan teknologi informasi terutama pada bidang *data mining* menjadi sumber yang sangat penting bagi perguruan tinggi terutama dalam hal penelusuran dan pencarian data mahasiswa baru dalam pemilihan dan penentuan program studi. Untuk semua perguruan tinggi baik negeri atau swasta yang namanaya penentuan prodi bagi calon mahasiswa baru merupakan peran yang sangat penting. Agar sesuai dengan kompetensi diri maka calon mahasiswa untuk memilih program studi yang diinginkan. Dengan berjalannya waktu adapun sistem yang ada di Politeknik Harapan Bersama Tegal untuk mengetahui apakah program studi tersebut masih belum efektif sebagai calon mahasiswa baru dalam memilih program studi dilakukan secara pribadi di ketika registrasi [1]. Setiap memilih program studi artinya hal yang sangat penting bagi perguruan tinggi, sebab perihal tersebut berkaitan dengan pilihan dan kepentingan masyarakat. Peminat berarti kemampuan untuk meningkatkan hal-hal tertentu. Banyak universitas di Indonesia menawarkan berbagai jenis program studi yang warga minati [1].

Pada penelitian ini akan membahas topik tentang klasifikasi. Klasifikasi yang digunakan dalam penelitian ini untuk menguji data numerik dan data kategorikal. Selain untuk pengujian data numerik dan data kategorikal, dapat juga dipakai untuk mengolah data dan mengidentifikasi gambar-gambar dengan kriteria yang diinginkan, hal ini merupakan salah satu fungsi klasifikasi [2].

Model yang dipakai dalam klasifikasi penelitian ini adalah metode *Naive Bayes*. *Naive Bayes* dapat menangani penugasan label relevan. Tugas ini dilakukan untuk memperluas jangkauan peringkat pada satu tag. Secara khusus, penjabaran klasifikasi pada *Naive Bayes* dapat menilai pengujian data eksperimental dan pemodelan prediktif [3].

Metode lain untuk klasifikasi pada penelitian ini metode *Unsupervised Discretization*. Alasan pemilihan metode ini karena metode ini dapat mentransformasikan kriteria numerikal atau kontinyu menjadi kriteria kategorikal dengan menghilangkan atau mengeliminasi satu kriteria yang dianggap tidak mempunyai pengaruh terhadap hasil keakuratan pengujian [4]. Metode ini adalah salah satu model aproksimasi yang sederhana dan digunakan untuk melakukan optimalisasi pada metode *Naive Bayes* [2].

Penggunaan klasifikasi *Naive Bayes* bisa diimplementasikan dan digunakan pada berbagai macam aplikasi, seperti pengenalan wajah, dan diagnosis kanker. Hal tersebut merupakan salah satu contoh implementasi dari metode *Naive Bayes* pada penelitian tentang aplikasi. Peran klasifikasi dalam penelitian ini adalah pada layanan server. Layanan server fokus pada proses pembentukan pengklasifikasi privasi *Naive Bayes* yang digunakan untuk menahan serangan yang mudah dijalankan di server akan tetapi sulit untuk mendeteksi serangan tersebut [3].

Penelitian terkait tentang penentuan program studi menggunakan metode *Naive Bayes* dilakukan pada MA Al Mahrusiyah Lirboyo. Penelitian ini dilatarbelakangi hasil pengamatan yang dilakukan pada siswa yang naik dari kelas X naik ke kelas XI, dimana masalah yang dialami oleh para siswa dalam hal menentukan bidang jurusan yang diminati. Masalah tersebut adanya kesalahan dari para siswa dalam menentukan penjurusan, yaitu; rendahnya prestasi belajar siswa atau

terjadinya ketidakcocokan dalam memilih jurusan sehingga mempengaruhi ketika akan masuk ke jenjang perguruan tinggi. Dengan adanya metode *Naive Bayes* yang di bantu dengan Bahasa pemrograman PHP dengan teknik klasifikasi pada data mining, maka para guru dapat dengan mudah mengklasifikasikan dalam menentukan jurusan berdasarkan nilai prestasi dan minat sebagai acuan [5].

Penelitian menggunakan metode *Naive Bayes* dilakukan oleh Tzu Tsung dimana dalam penelitian ini diusulkan ukuran non parametrik untuk mengevaluasi tingkat ketergantungan antara atribut kontinu dan kelas. Ukuran non parametrik digunakan untuk mengembangkan hibrida dan mendiskritisasi atribut kontinu sehingga menghasilkan akurasi klasifikasi *Naive Bayes* dapat meningkat. Metode *hybrid* dipakai untuk menguji 20 *dataset* [6].

Penerapan metode *Naive Bayes* yang lain dipergunakan buat penjabaran hostpot pada modul fotovoltaik. Kategori host yang terdeteksi digunakan untuk pelatihan metode *Naive Bayes* dengan data uji pada system PV 42,24 kWp, ini menggambarkan akibat taraf pengenalan rata homogen lebih kurang 94,1% dari gugus contoh data sebanyak 375 [7]. Rujukan yang lain pada penelitian ini yaitu tentang klasifikasi artikel berita dengan *Naive Bayes*, dimana ini yang telah mengalami modifikasi pada pembobotan digunakan buat mengklasifikasikan isu berdasarkan posisinya. Terdapat 100 dokumen dari setiap 9 kategori yang merupakan hasil pembagian dari 900 dokumen pengujian. Buat dokumen pelatihan dan dokumen pengujian apa pun adalah hasil dari kombinasi dari sejumlah data latih pada klasifikasi metode *Naive Bayes* yang sudah memiliki efektifitas [8].

Penelitian yang sama tentang menentukan jurusan dengan metode *Naive Bayes* dimana dalam penelitian ini, metode *Naive Bayes* digunakan untuk mencari informasi dalam menentukan jurusan siswa. Ada 100 data siswa yang digunakan untuk data uji dimana data tadi ada beberapa kriteria yang mengandung nilai homogen rata-rata mata pelajaran sains dan sosial dan yang akan terjadi rekomendasi guru serta nilai angket calon siswa. Nilai akurasi yang dihasilkan dalam menentukan jurusan sebesar 90% dan untuk *error*-nya 10% [9].

Aplikasi sistem pendukung keputusan mengenai penggunaan minat untuk menentukan prodi merupakan kajian terkait, yaitu penentuan jurusan SMA Islam di kecamatan Bumiayu Kabupaten Brebes. Pihak sekolah memberikan dukungan dalam membangun sistem dalam bentuk aplikasi pendukung keputusan yang digunakan untuk menentukan jurusan siswa. Sistem ini nantinya diperuntukkan sebagai proses akhir dalam penentuan dan pemilihan jurusan untuk para siswa [10].

Penelitian tentang penentuan atau pemilihan jurusan yang dipimpin oleh Alfa Saleh dan Fina Nasari bertajuk Penggunaan Diskritisasi Tanpa Pengawasan dalam metode *Naive Bayes* untuk menentukan jurusan siswa Madrasah Aliyah pada tahun 2017. Di penelitian ini melakukan optimalisasi metode *Naive Bayes* menggunakan dan menerapkan teknik diskritisasi tanpa pengawasan yang mengubah kriteria numerik atau kontinu menjadi kriteria kategoris dan menerapkan kriteria yang diklaim tidak mempunyai efek terhadap keakuratan terhadap yang akan terjadi pengujian, maka dengan begitu keakuratan hasil pembagian terstruktur mengenai dapat semakin tinggi. Berasal hasil pengujian 120 data peserta didik memperoleh

bukti bahwa penerapan teknik *Unsupervised Discretization* pada metode *Naive Bayes* meningkat dari 90% menjadi 92,8% [4].

Penelitian ini melanjutkan penelitian sebelumnya, termasuk pengujian statistik hingga 1.671 data. Dari jumlah data ini dibagi sebagai dua bagian, pertama merupakan pedagogi keterangan menggunakan jumlah 1158 data menggunakan persentase 70% sedangkan yang ke 2 merupakan pembuktian data menggunakan jumlah 513 menggunakan persentase 30%. Dari *output* pengujian menggunakan metode yang sama tetapi menggunakan jumlah warta yang tidak sama mengenai dokumen tadi akan suatu taraf presesi akurasi. Penelitian terdahulu menggunakan metode *Naive Bayes* dengan nilai akurasi 96,68% [11].

Pokok pertarungan yang asal kajian ini adalah bagaimana nilai presisi akurasi pada metode *Naive Bayes* dan sesudah dilakukan optimalisasi menggunakan pendekatan metode *Unsupervised Discretization* dapat berperan dalam menentukan kurikulum bagi calon mahasiswa baru untuk menjadi bagian dari Unit Penerimaan Mahasiswa Baru Politeknik Harapan Bersama Tegal. Selain permasalahan itu ada juga pemasalahan lain yang dibahas dalam penelitian ini yaitu, apakah metode *Naive Bayes* lebih baik dengan metode *Naive Bayes* yang sudah dilakukan optimalisasikan dengan metode *Unsupervised Discretization* dalam melakukan klasifikasi penentuan program studi bagi calon mahasiswa baru.

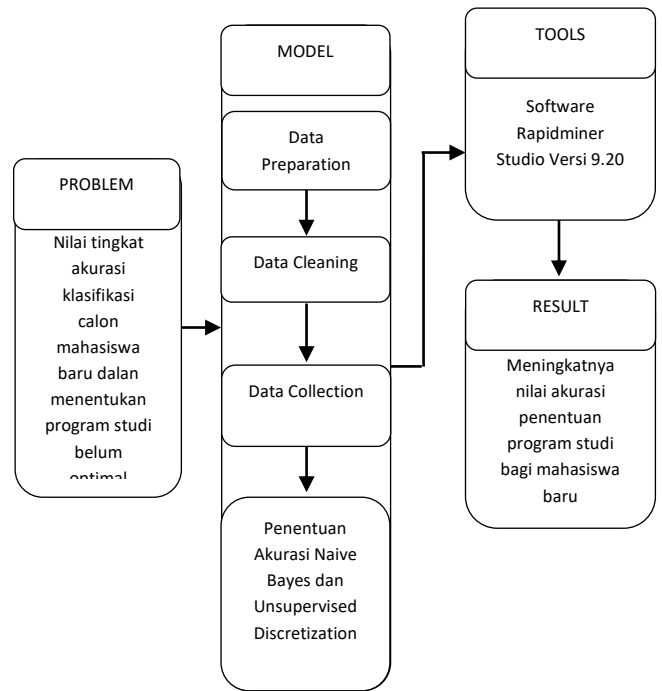
Pada penelitian ini dalam penentuan prodi untuk calon mahasiswa baru dengan menggunakan diusulkan sebuah metode *Naive Bayes* dan *Naive Bayes* yang dioptimalisasi dengan *Unsupervised Discretization*. Penelitian ini bertujuan buat menciptakan penjabaran pada penentuan program studi calon mahasiswa baru memakai metode *Naive Bayes* dan pendekatan metode *Unsupervised Discretization*, dimana *output* yang dibutuhkan akan dipakai sebagai acuan peningkatan mutu dari setiap program studi.

2. Metode

2.1. Kerangka Penelitian

Langkah-langkah penelitian yang dilakukan, bagian pertama merupakan pencarian data pada unit penerimaan mahasiswa baru. Riset data bisa berupa melihat pribadi proses penerimaan mahasiswa baru dan wawancara melalui petugas penerimaan mahasiswa baru. Setelah menyelidiki data, langkah selanjutnya merupakan mengumpulkan data dan menganalisa lebih lanjut variabel dan megeksekusi desain. Langkah terakhir merupakan menguji variabel dan menganalisa konsekuensi berdasarkan *testing* data.

Terdapat beberapa data calon mahasiswa tahun 2019 yang telah dikonversi menjadi kriteria dengan data string dan data numerik, dan data tersebut diagregasi menggunakan *Naive Bayes*. Langkah-langkah pengolahan data yang dilakukan dalam metode penelitian ini adalah *data Cleaning*, *data Collection*, penentuan kriteria, menentukan peluang atau probabilitas dari setiap kriteria, pengujian akhir. Langkah-langkah dapat ditunjukkan dalam kerangka penelitian menggunakan metode yang diusulkan dengan bagan pada Gambar 1.



Gambar 1. Kerangka Penelitian

Gambar 1 Merupakan kerangka atau tahapan melakukan penelitian. Langkah-langkah penelitian yang akan dilakukan terdiri dari pengolahan data mentah dari data penerimaan mahasiswa baru. Dalam metode *Naive Bayes* dibedakan antara data kategorikal menjadi dua macam, yaitu data numerik yang bersifat kontinu, sebagai akibatnya disparitas yang dihasilkan akan terlihat pada saat menentukan nilai probabilitas dari setiap kriteria, baik kriteria yang menggunakan nilai data string maupun kriteria yang menggunakan data numerik. Adapun tahapannya yang dilakukan artinya menjadi berikut:

- 1) *Problem*, yaitu mencari masalah yang akan dijadikan acuan penelitian dimana pada hal ini adalah nilai taraf akurasi pembagian terstruktur mengenai calon mahasiswa baru pada memilih program studi belum optimal.
- 2) Dalam tahapan penelitian ini terdapat beberapa tahapan antara lain:
 - *Data preparation*, yaitu proses dimana dalam tahapan ini mempersiapkan kebutuhan data yang akan diolah.
 - *Data Collection*, yaitu data yang dijadikan sebagai data pembinaan, pada hal ini adalah data calon mahasiswa baru.
 - *Data Cleaning*, yaitu di termin dalam hal ini masih ada kriteria yang dihilangkan lantaran kriteria tadi tidak berpengaruh terhadap *output* akurasi penjabaran metode *Naive Bayes*.
 - Memilih kriteria, Ketika memilih kriteria ini, kriteria data yang sah menurut data yang dikumpulkan yang akan digunakan.
 - Memilih probabilitas, dalam hal ini menentukan kriteria yang dijadikan acuan dalam penilaian prodi calon mahasiswa baru.
- 3) Pengujian, di tahap ini merupakan langkah implementasi dari *Naive Bayes* dengan *optimalisasi Unsupervised Discretization* pada dengan data yang siap untuk diuji.

2.2. Metode Naive Bayes

Metode *Naive Bayes* merupakan model penjabaran probabilistic sederhana buat menghitung kombinasi probabilitas dan menjumlahkan frekuensi dan kombinasi berdari perpaduan data [12].

Naive Bayes memiliki beberapa keunggulan, termasuk fakta bahwa ia membutuhkan sedikit pemrosesan data. Orientasi data dipakai buat memilih perkiraan parameter pada proses klasifikasi [6]. Persamaan metode *Naive Bayes* seperti pada persamaan (1).

$$P(H|X) = \frac{p(H|X).P(H)}{P(X)} \tag{1}$$

dimana:

- X : tidak diketahui kelas data.
- H : hipotesis data dari kelas spesifik
- P(H|X) : a posteriori probabilitas (probabilitas hipotesis H berdasarkan kondisi X)
- P(H) : mengandung nilai probabilitas hipotesis H
- P(X|H) : memuat nilai probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : memuat nilai probabilitas hipotesis X

Diketahui bahwa prosess konstruksi atau klasifikasi menggunakan metode *Naive Bayes* memerlukan sejumlah instruksi untuk menentukan kelas yang tepat untuk data sampel dalam analisis [13]. Misalnya, metode *Naive Bayes* diatas telah diadaptasi untuk:

$$P(C|F1..Fn) = \frac{P(C)P(F1...Fn|C)}{P(F1...Fn)} \tag{2}$$

dimana:

- Kelas nilai variabel C
- Gambaran karakteristik klasifikasi (Nilai variable F1 – Fn)

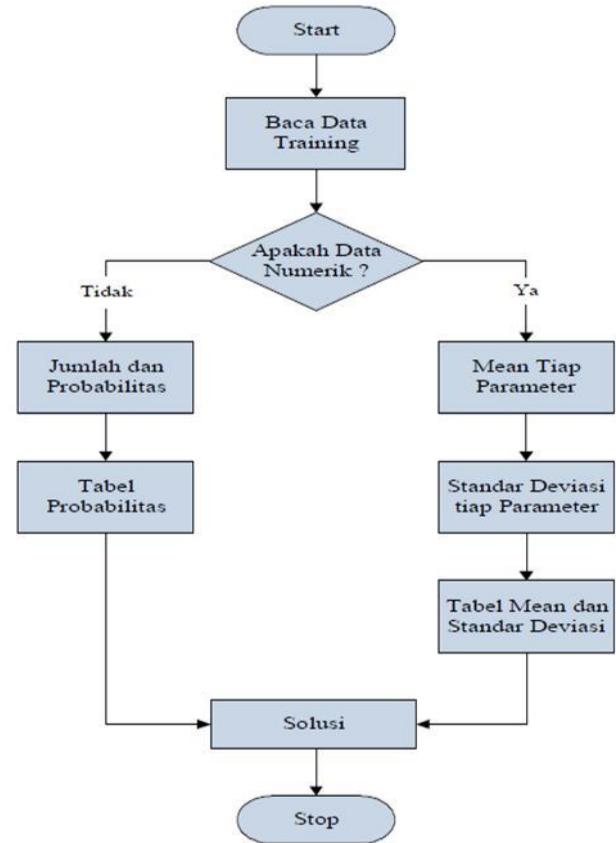
Dari penerangan rumus tadi dijelaskan bahwa peluang masuknya sampel ciri yang tertentu dalam kelas derajat C dikalikan menggunakan peluang keluarnya contoh ciri kelas C dibagi menggunakan peluang keluarnya berdasarkan model ciri kelas C dalam skala global. Kemudian rumus ini disederhanakan menjadi persamaan (3)

$$Posterior = \frac{prior \times likelihood}{evidence} \tag{3}$$

Berdasarkan apa yang akan terjadi menggunakan penyederhanaan rumus diatas, nilai bukti dipakai buat setiap kelas dalam setiap sampel dan selalu permanen, membangun nilai posterior buat dibandingkan menggunakan nilai posterior lainnya. Sedangkan buat menentukan suatu nilai dimana kelas suatu sampel data bisa diklasifikasikan, bisa digambarkan menggunakan rumus *Naive Bayes* menjadi: (C|F1,...,Fn) [14], sedangkan perkalian dari aturan rumus ini adalah:

$$\begin{aligned} P(C|F1, \dots, Fn) &= P(C)P(F1, \dots, Fn|C) \\ &= P(C)P(F1|C)P(F2, \dots, Fn|C, F1) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3, \dots, Fn|C, F1, F2) \\ &= \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2)P(F4, \dots, Fn|C, F1, F2, F3) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2) \dots P(Fn|C, F1, F2, F3, \dots, Fn-1) \end{aligned}$$

Dari uraian tersebut, bisa disimpulkan bahwa rumus ini bisa mengakibatkan factor syarat yang semakin kompleks yang mensugesti nilai probabilitas, sebagai akibatnya hampir mungkin buat melakukan perhitungan dan analisa [15]. Dari rumus dan berbagai penjelasan diatas, Gambar 2 adalah diagram alir penggunaan metode *Naive Bayes*.



Gambar 2. Alur Metode *Naive Bayes*

2.3. Metode Unsupervised Discretization

Metode *Unsupervised Discretization* adalah metode yang sederhana. Kesalahan metode sederhana lainnya serupa (diskritisasi interval lebar yang sama dan diskritisasi interval frekuensi yang sama) dan kesalahan lain yang lebih canggih, berdasarkan analisis kluster, mirip dengan diskritisasi Kmeans. Rentang konstan dibagi menjadi lebar yang ditentukan pengguna atau sub-rentang frekuensi. Namun dalam penelitian ini peneliti memakai teknik equal interval discretization, yaitu metode diskritisasi paling sederhana yang membagi rang nilai observasi kepada setiap fitur atau atribut, dimana k adalah parameter yang disediakan pengguna. Prosesnya terdiri menurut pengurutan nilai yang diamati menurut fitur atau atribut konstan dan menemukan nilai minimum (*Vmin*) dan maksimum (*Vmax*). Interval bisa dihitung menggunakan membagi rentang nilai yang diamati buat variabel sebagai k yang dipengaruhi sang pengguna [16].

$$Interval = \frac{Vmax - Vmin}{k}$$

$$Boundaries = Vmin + (i \times Interval)$$

(4)

Kemudian batas-batas $i=1..k1$ dapat dibangun menggunakan persamaan diatas. Jenis diskritisasi ini tidak tergantung pada struktur data multi-relasional. Namun, metode diskritisasi ini sensitif terhadap outlier yang secara signifikan dapat mengurangi jangkauan. Keterbatasan metode ini diberikan oleh distribusi titik data yang tidak merata: beberapa interval dapat berisi lebih banyak titik data daripada yang lain.

3. Hasil dan Pembahasan

Tahapan penelitian yang akan dilakukan data preposisi yaitu pengolahan data mentah dari penerimaan mahasiswa baru tahun akademik 2019 / 2020 dengan jumlah *dataset* 2256. Metode *Naive Bayes* membedakan antara string atau data kategorikal yang konstan sebagai 2 macam yaitu data numerik yang bersifat kontinyu, sehingga disparitas yang dihasilkan, ini akan ditampilkan di waktu menentukan nilai probabilitas setiap kriteria dengan data string atau kriteria menggunakan data numerik. Adapun tahapannya yang dilakukan adalah sebagai berikut:

- 1) *Problem* adalah mencari permasalahan yang akan dilakukan dalam penelitian ini. Adapun masalah yang diambil dalam penelitian ini adalah mencari nilai akurasi dalam klasifikasi penentuan program studi atau jurusan untuk calon mahasiswa baru.
- 2) *Data Preparation* yaitu sebuah tahapan yang didalamnya adalah mempersiapkan kebutuhan data untuk menunjang penelitian. Adapun data yang dipersiapkan adalah data penerimaan calon mahasiswa baru yang berjumlah 2256.
- 3) Pengumpulan data (*Data Collection*) terdiri dari data yang dijadikan sebagai data latih dan *testing*, pada hal ini adalah data calon mahasiswa baru. Kriteria buat memilih data pembinaan serta data *testing* merupakan untuk data pelatihan 70% dan data *testing* 30%.
- 4) *Data Cleaning* yaitu di tahap dalam hal ini ada kriteria yang di-drop karena kriteria tersebut tidak mempengaruhi akurasi klasifikasi *Naive Bayes*. Jumlah *dataset* 2256 *record*, setelah dilakukan eliminasi terhadap *dataset* tadi, maka jumlah *dataset*nya sebanyak 1671, sedangkan atribut yang digunakan sebesar 23.
- 5) Dalam menentukan kriteria, dalam hal pemilihan kriteria ini telah dilakukan penggunaan data yang konsisten dengan data yang dikumpulkan.
- 6) Menentukan peluang atau probabilitas dari setiap kriteria, pada tahap ini memilih kriteria yang dijadikan acuan dalam penilaian prodi untuk calon mahasiswa baru.
- 7) Pengujian pada tahap ini merupakan Langkah implementasi metode *Naive Bayes* dengan Optimalisasi metode *Unsupervised Discretization* dengan beberapa data yang siap untuk diuji.

Adapun pada Gambar 3 terdapat 20 contoh data yang siap dijalankan dengan langkah-langkah sebagai berikut:

	A	B	C	D	E	F	G	H	I	J
1	No Pend...	Nama	Kelas	Gelomb...	Alamat	Jenis K...	Tempat ...	Tanggal...	Usia	Agan
2	1901175...	PUJI AS...	Reguler...	Gelomb...	KALIPU...	P	JAKARTA	Oct 21, 1...	27.879	ISLAI
3	1901175...	SAFITRI ...	Reguler...	Gelomb...	JL-RAK...	P	Tegal	Jun 23, 1...	23.205	ISLAI
4	1901275...	RAMANG...	Reguler...	Gelomb...	PERUM...	L	tegal	Jul 21, 2...	19.126	ISLAI
5	1901275...	ALDI AF...	Reguler...	Gelomb...	PESAN...	L	BREBES	Jun 3, 20...	20.258	ISLAI
6	1901275...	NANDA ...	Reguler...	Gelomb...	DESA-U...	P	TEGAL	Jul 10, 2...	20.156	ISLAI
7	1901275...	DWI AMA...	Reguler...	Gelomb...	GUMAY...	P	TEGAL	Aug 27, ...	22.027	ISLAI
8	1901375...	SUSMO...	Reguler...	Gelomb...	JLPROF...	L	TEGAL	Nov 23, ...	20.786	ISLAI
9	1902075...	WAHYU ...	Reguler...	Gelomb...	DUSUN...	L	BREBES	Sep 29, ...	23.937	ISLAI
10	1902075...	DANU H...	Reguler...	Gelomb...	DESA-K...	L	Brebes	Jun 17, 1...	23.222	ISLAI
11	1902075...	NUR MIL...	Reguler...	Gelomb...	JALAN...	P	TEGAL	Aug 26, ...	22.030	ISLAI
12	1902075...	MOHAMA...	Reguler...	Gelomb...	DSKER...	L	TEGAL	Oct 3, 19...	21.926	ISLAI
13	1902075...	DANW Y...	Reguler...	Gelomb...	JLSETIA...	L	BREBES	Dec 3, 1...	20.759	ISLAI

Gambar 3. Contoh *Dataset* Training dan *Testing*

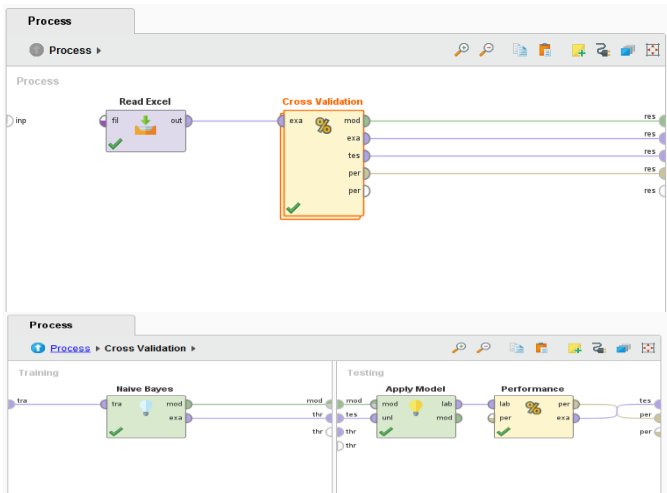
Data pada gambar 3 kemudian siap untuk pelatihan dan pengujian dengan metode optimasi *Naive Bayes* dan pendekatan *Unsupervised Discretization*. Tetapi sebelum memasuki tahap Optimalisasi *Naive Bayes* menggunakan Rapidminer, terlebih dahulu harus dilakukan metode *Naive Bayes* dengan menggunakan rapidminer. Ada beberapa tahap pengujian yang dilakukan yaitu dari mulai merubah nilai *Cross Validation* dan *Sampling Type* yang dapat dilakukan dengan rapidminer.

Untuk melakukan pengujian *dataset* ini dilakukan beberapa skenario penelitian seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Hasil Eksperimen *Naive Bayes*

Pengujian	Number Of <i>Cross Validation</i>	Sampling Type	Accuracy
1	5	Automatic	18.07%
	5	Linear Sampling	15.80%
	5	Saffled Sampling	19.27%
	5	Stratified Sampling	18.07%
2	10	Automatic	19.33%
	10	Linear Sampling	15.44%
	10	Saffled Sampling	18.67%
	10	Stratified Sampling	19.33%
3	15	Automatic	18.79%
	15	Linear Sampling	15.15%
	15	Saffled Sampling	18.20%
	15	Stratified Sampling	18.79%
4	20	Automatic	18.74%
	20	Linear Sampling	14.54%
	20	Saffled Sampling	18.56%
	20	Stratified Sampling	18.74%

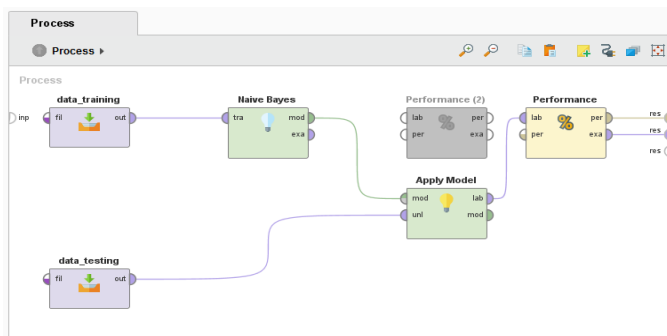
Dari beberapa pengujian pada tabel 1 memperoleh nilai akurasi paling tinggi 19.33%, nilai akurasi tersebut merupakan nilai tidak sesuai dengan yang diharapkan, karena memakai nilai dari *Number of Cross Validation* 10 dan *Sampling Type* menggunakan *Stratified Sampling*. Sedangkan desain yang digunakan di rapidminer seperti desain Gambar 4.



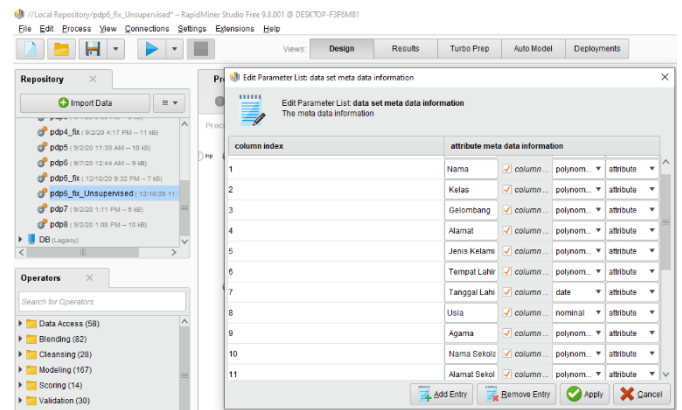
Gambar 4. Hasil Desain Eksperimen Dengan *Cross Validation* Pada *Naive Bayes*

30% data dengan jumlah catatan 513 data yang berasal jumlah total record asal *dataset* 1671. Serta hasil akurasi diperoleh menurut desain pengujian diatas dengan memakai metode *Naive Bayes* adalah nilai akurasi 96.68%. Nilai akurasi tersebut masih belum menggunakan optimalisasi pendekatan *Unsupervised Discretization*. Nilai akurasi tersebut dapat dilihat melalui grafik penentuan program studi pilihan mahasiswa yang terdapat pada Gambar 6.

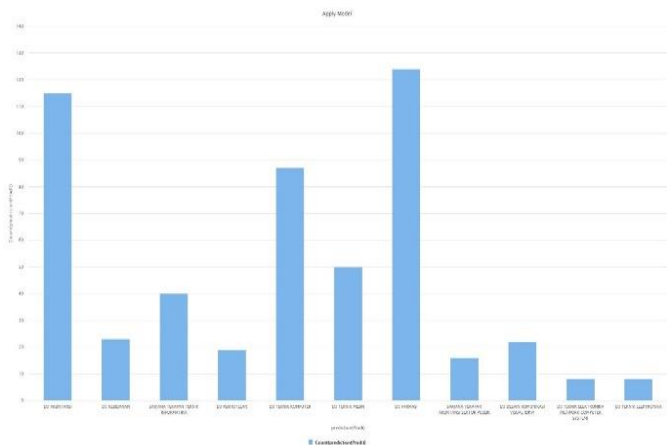
Dari hasil grafik eksperimen bisa menyimpulkan bahwa pemilihan program studi bagi calon mahasiswa baru mengarah pada program studi DIII Farmasi. Sedangkan untuk optimalisasi dengan pendekatan *Unsupervised Discretization* masih menggunakan desain eksperimen seperti tampilan diatas tetapi disini menitikberatkan pada transformasi kriteria numerik atau kontinyu menjadi kriteria kategorikal dan menghilangkan satu kriteria yang diklaim berpengaruh dalam keakuratan yang akan terjadi pengujian, memakai keakuratan output pembagian terstruktur mengenai meningkat. Adapun gambar 7. merupakan tampilan perubahan kriteria numerik/kontinyu menjadi kriteria kategori.



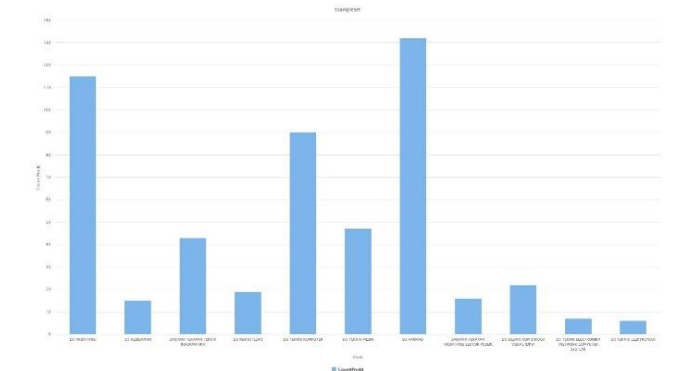
Gambar 5. Hasil Desain Pengujian Data Training dan *Testing* Pada *Naive Bayes*



Gambar 7. Hasil Perubahan Kriteria Numerik Menjadi Kriteria Kategorikal



Gambar 6. Grafik Hasil Eksperimen Pada *Naive Bayes*



Gambar 8 Grafik Hasil Eksperimen Pada *Naive Bayes* Dan *Unsupervised Discretization*

Pada eksperimen berikutnya dengan desain eksperimen yang berbeda dengan yang terdahulu mendapatkan nilai akurasi 96.68%. Adapun desain eksperimen yang digunakan pada gambar 5. Desain eksperimen yang terdapat pada gambar 5 mendeskripsikan bahwa *dataset* yang digunakan dipecah sebagai 2 bagian menggunakan rasio 70% dipergunakan buat pelatihan data menggunakan jumlah catatan 1158 data serta tes

Pada Gambar 7 merupakan hasil perubahan data training dan *testing* yang tadinya mempunyai kriteria numerik menjadi kriteria kategorikal. Hasil akurasi yang diperoleh dari rancangan percobaan di atas dengan metode *Naive Bayes* dengan optimalisasi pendekatan *Unsupervised Discretization* adalah nilai akurasi 97.66%. Nilai akurasi klasifikasi penerapan optimalisasi *Unsupervised Discretization* pada metode *Naive Bayes* dari 96.68% menjadi 97.66%. Nilai

akurasi tersebut dapat dilihat melalui grafik gambar 8. Dari hasil grafik eksperimen disimpulkan bahwa pemilihan program studi bagi calon mahasiswa baru menunjuk dalam program studi Diploma III Farmasi.

Apabila dibandingkan antara penelitian sebelumnya yaitu dengan metode *Naive Bayes* dan penelitian yang sedang dikerjakan dengan metode *Naive Bayes* dan dioptimalisasikan dengan pendekatan metode *Unsupervised Discretization* maka hasilnya lebih baik berdasarkan output penelitian terkini. Pengujian dengan metode lain seperti halnya t-test dengan metode Anova tidak dilakukan karena metode tersebut digunakan untuk analisis statistik, sedangkan metode pengujian pada penelitian ini hanya dibatasi untuk mencari nilai akurasi klasifikasi. Perbandingan hasil penelitian terdahulu dan terkini bisa dilihat pada tabel 2 berikut ini:

Tabel 2. Perbandingan Hasil Penelitian

No	Metode	Hasil Akurasi
1.	<i>Naive Bayes</i>	96.68%
2.	<i>Naive Bayes + Unsupervised Discretization</i>	97.66%

Dari tabel 2, membagikan bahwa antara penelitian sebelumnya dengan penelitian saat ini lebih baik karena dari segi apa yang akan terjadi, akurasi yang diperoleh antara penelitian sebelumnya mendapatkan nilai akurasi sebesar 96,68% sedangkan penelitian saat ini mengalami kenaikan nilai akurasi sejumlah 97.66%.

4. Kesimpulan

Berdasarkan hasil klasifikasi dalam memilih program studi calon mahasiswa baru menurut metode *Naive Bayes* dan optimalisasi pendekatan metode *Unsupervised Discretization* diatas dapat disimpulkan bahwa kualitas penerimaan mahasiswa baru semakin meningkat, sehingga terpenuhi kriteria yang diinginkan untuk setiap program studi. Masih banyak atribut yang belum sinkron, sehingga membutuhkan proses pengembangan dan pengujian data yang harus diulang beberapa kali. Perlu adanya pengujian yang lebih mendalam guna menemukan metode klasifikasi yang lebih baik atau misalkan dibandingkan dengan metode t-test Anova. Banyak sekali eksperimen dengan model eksperimen pada metode *Naive Bayes* sehingga mendapatkan hasil dengan nilai akurasi 96,68%. Dari nilai hasil akurasi klasifikasi menggunakan *Naive Bayes* 96.68% dan di optimalisasi menggunakan *Unsupervised Discretization* menghasilkan nilai akurasi 97.66%.

Ucapan Terimakasih

Ucapan terimakasih tidak terhingga pada seluruh pihak yang telah membantu dan mendukung proses penelitian ini, ucapan terimakasih khususnya pada pihak institusi yang telah mendukung aktivitas penelitian ini.

Daftar Pustaka

- [1] M. S. Sungkar, "Analisis Minat Mahasiswa Memasuki Program Studi Teknik Dengan Menggunakan Metode K-Means Clustering Di Politeknik Harapan Bersama," *J. Chem. Inf. Model.*, vol. 21, no. 1, pp. 1–9, 2020, [Online]. Available: <https://doi.org/10.1016/j.tmaid.2020.101607%0Ahttps://doi.org/10.1016/j.ijisu.2020.02.034%0Ahttps://onlinelibrary.wiley.com/doi/abs/10.1111/cjag.12228%0Ahttps://doi.org/10.1016/j.ssci.2020.104773%0Ahttps://doi.org/10.1016/j.jinf.2020.04.011%0Ahttps://doi.o>
- [2] C. Gonzalez-Arias, C. C. Viafara, J. J. Coronado, and F. Martinez, "Automatic classification of severe and mild wear in worn surface images using histograms of oriented gradients as descriptor," *Wear*, vol. 426–427, no. November 2018, pp. 1702–1711, 2019, doi: 10.1016/j.wear.2018.11.028.
- [3] M. Andrejiova and A. Grincova, "Classification of impact damage on a rubber-textile conveyor belt using *Naive-Bayes* methodology," *Wear*, vol. 414–415, pp. 59–67, 2018, doi: 10.1016/j.wear.2018.08.001.
- [4] A. Saleh, F. Nasari, U. P. Utama, and K. J. Siswa, "Penggunaan Teknik *Unsupervised Discretization* Pada Metode *Naive Bayes* Dalam Menentukan Jurusan Siswa", *MAJOR*, vol. 5, no. 3, pp. 353–360, 2018, doi: 10.25126/jtiik.201853705.
- [5] M. A. Al and M. Lirboyo, "Program studi sistem informasi fakultas teknik universitas nusantara pgri kediri 2017," 2017.
- [6] T. Wong, "ve Bayesian classifiers A hybrid discretization method for nai," *Pattern Recognit.*, vol. 45, no. 6, pp. 2321–2325, 2012, doi: 10.1016/j.patcog.2011.12.014.
- [7] K. A. K. Niazi, W. Akhtar, H. A. Khan, Y. Yang, and S. Athar, "Hotspot diagnosis for solar photovoltaic modules using a *Naive Bayes* classifier," *Sol. Energy*, vol. 190, no. July, pp. 34–43, 2019, doi: 10.1016/j.solener.2019.07.063.
- [8] J. Sains *et al.*, "Klasifikasi Berita Indonesia Menggunakan Metode *Naive Bayesian* Classification dan Support Vector Machine dengan Confix Stripping Stemmer," vol. 4, no. 2, 2015.
- [9] S. U. Khan, M. Niazi, and R. Ahmad, "Factors influencing clients in the selection of offshore software outsourcing vendors: An exploratory study using a systematic literature review," *J. Syst. Softw.*, vol. 84, no. 4, pp. 686–699, Apr. 2011, doi: 10.1016/j.jss.2010.12.010.
- [10] T. Prawira and D. Kusuma, "Sistem Pendukung Keputusan Berbasis Web untuk Menentukan Penjurusan (IPA / IPS / Bahasa) pada SMA Islam Bumiayu," *Juita*, vol. I, pp. 177–189, 2011.
- [11] W. E. Nugroho, A. Sofyan, and O. Somantri, "Metode *Naive Bayes* Dalam Menentukan Program Studi Bagi Calon Mahasiswa Baru," vol. 12, no. 01, pp. 59–64, 2021, doi: 10.35970/infotekmesin.v12i1.491.
- [12] A. Saleh, "Klasifikasi Metode *Naive Bayes* Dalam Data Mining Untuk Menentukan Konsentrasi Siswa," *KeTIK*, pp. 200–208, 2015.
- [13] J. Wu, S. Pan, Z. Cai, X. Zhu, and C. Zhang, "Dual instance and attribute weighting for *Naive Bayes* classification," *Proc. Int. Jt. Conf. Neural Networks*, no. 1994, pp. 1675–1679, 2014, doi: 10.1109/IJCNN.2014.6889572.
- [14] D. Mondal, D. K. Kole, and K. Roy, "Gradation of yellow mosaic virus disease of okra and bitter melon based on entropy based binning and *Naive Bayes* classifier after identification of leaves," *Comput. Electron. Agric.*, vol. 142, no. October, pp. 485–493, 2017, doi: 10.1016/j.compag.2017.11.024.
- [15] A. H. Mirza, "Application of *Naive Bayes* Classifier Algorithm in Determining New Student Admission Promotion Strategies," *J. Inf. Syst. Informatics*, vol. 1, no. 1, pp. 14–28, 2019, doi: 10.33557/journalisi.v1i1.2.
- [16] J. J. Christopher, "*Unsupervised Discretization* : An Analysis of Classification Approaches for Clinical Datasets *Unsupervised Discretization* : An Analysis of Classification Approaches for Clinical Datasets," no. February, 2017, doi: 10.19026/rjaset.14.3991.