



## OPTIMALISASI METODE NAIVE BAYES DAN DECISION TREE UNTUK MENENTUKAN PROGRAM STUDI BAGI CALON MAHASISWA BARU DENGAN PENDEKATAN UNSUPERVISED DISCRETIZATION

Wildani Eko Nugroho<sup>1</sup>, Heru Saputro<sup>2</sup>

Politeknik Harapan Bersama Tegal  
Universitas Islam Nahdlatul Ulama Jember  
wild4n1@gmail.com, herusaputro@unisnu.ac.id

### ABSTRACT

*Higher Education is a place for providing education that aims to produce quality human resources and is able to face increasingly fierce job competition. Therefore, from the recruitment process or the admission process, prospective new students must consider various procedures that aim to be able to direct prospective new students in determining the study program that will be taken by prospective new students. The things that have been broken in the admission process for new students include the scores of national exam results, report cards, school test scores and the admission test for new students, as well as the admission process of the achievement path and aiming for missions. From these things, performance must be improved is a supporting factor so that the process of transforming educational science to students can be carried out properly. The purpose of this study is to obtain classification in determining the study program of prospective new students by optimizing the Naïve Bayes and Decision Tree methods with an Unsupervised Discretization Approach, as an effort to improve the internal quality assurance system, especially the standards for the admission process for new students in determining study programs at the Harapan Polytechnic with Tegal. Where in the process of accepting new students, planning, implementing, evaluating, and monitoring have been carried out as a form of implementing the Internal Quality Assurance System (SPMI). In this study, the data used was data on the results of the admission of prospective new students from all study programs. These data include data on the administrative completeness of the requirements of prospective new students, as well as data on the value of the results of the new student admission test. The data used is data for 1 academic year 2019/2020. From this data, training and testing will be carried out using Rapidminer 9, a classification of lecturer teaching performance will be obtained.*

**Keywords** : Naive Bayes, Unsupervised Discretization, Decision Tree, Classification, Interests.

### ABSTRAK

Perguruan Tinggi merupakan tempat penyelenggara pendidikan yang bertujuan menghasilkan sumber daya manusia yang berkualitas dan mampu menghadapi persaingan kerja yang semakin ketat. Maka dari dalam proses rekrutmen atau proses penerimaan calon mahasiswa baru harus mempertimbangkan berbagai prosedur yang bertujuan untuk dapat mengarahkan calon mahasiswa baru dalam menentukan program studi yang akan ditempuh oleh calon mahasiswa baru. Adapun hal sudah ditempuh dalam proses penerimaan mahasiswa baru antara lain dari nilai hasil ujian nasional, nilai raport, nilai ujian sekolah dan nilai test penerimaan mahasiswa baru, serta proses penerimaan dari jalur prestasi dan bidik misi. Dari hal – hal tersebut harus ditingkatkan kinerjanya merupakan faktor penunjang agar proses transformasi keilmuan pendidikan kepada mahasiswa dapat dilakukan dengan baik. Tujuan dari penelitian ini adalah untuk mendapatkan klasifikasi dalam menentukan program studi calon mahasiswa baru dengan mengoptimalkan metode *Naïve Bayes dan Decision Tree* dengan Pendekatan *Unsupervised Discretization*, sebagai upaya dalam peningkatan sistem penjaminan mutu internal khususnya standar proses penerimaan mahasiswa baru dalam menentukan program studi pada Politeknik Harapan Bersama Tegal. Dimana dalam proses penerimaan mahasiswa baru ini telah dilakukan perencanaan, pelaksanaan, evaluasi, dan monitoring sebagai bentuk penerapan Sistem Penjamin Mutu Internal (SPMI). Dalam penelitian ini, data yang digunakan adalah data hasil penerimaan calon mahasiswa baru dari seluruh program studi. Data tersebut antara lain data kelengkapan administratif persyaratan calon mahasiswa baru, serta data nilai hasil test penerimaan mahasiswa baru. Data yang

digunakan yakni data selama 1 tahun akademik 2019/2020. Dari data tersebut akan dilakukan *training* dan *testing* dengan menggunakan Rapidminer 9, maka akan didapatkan klasifikasi kinerja pengajaran dosen.

**Kata Kunci** : *Naive Bayes; Unsupervised Discretization; Decision Tree ; Klasifikasi; Minat;*

## 1. PENDAHULUAN

Perguruan tinggi merupakan suatu tempat atau organisasi satuan pendidikan, yang menyelenggarakan pendidikan di jenjang pendidikan tinggi, penelitian dan pengabdian kepada masyarakat. Perguruan Tinggi merupakan wadah bagi masyarakat kampus. Sebagai suatu organisasi maka perguruan tinggi mempunyai struktur, aturan penyelesaian tugas, yang mencakup pembagian tugas antar kelompok fungsional dan antar warga dalam kelompok yang sama, rencana kegiatan, dan tujuan. Tujuan dibimbing oleh asas dan membimbing rencana kegiatan. Struktur dan aturan penyelesaian tugas menjadi prasarana pencapaian tujuan dan sekaligus mencerminkan asas.

Pemilihan program studi juga sangat penting untuk calon mahasiswa baru, karena menentukan program studi yang sesuai dengan apa yang diikuti oleh calon mahasiswa baru merupakan salah satu hal yang dapat dinilai bahwa calon mahasiswa tersebut akan berkompeten di bidang yang diharapkan, sehingga kompetensi calon mahasiswa baru tersebut mana kala sudah lulus dari perguruan tinggi tersebut sesuai dengan kompetensi yang ada [1].

Seiring perkembangan dalam bidang penelitian mengenai menentukan program studi bagi calon mahasiswa baru dilakukan dengan menerapkan konsep data mining yang terkomputerisasi. Penggunaan data mining ini, khususnya pada ranahnya klasifikasi dapat digunakan untuk dijadikan alat bantu dalam menunjang keputusan berdasarkan data pendidikan, untuk mengatasi tantangan yang sulit dan untuk meningkatkan kompetensi serta kualitas pendidikan calon mahasiswa baru dalam memilih atau menentukan program studi yang diharapkan [2].

Menentukan program studi bagi perguruan tinggi adalah hal yang sangat penting juga, karena berkaitan erat dengan pilihan dan minat masyarakat. Peminat adalah kemampuan yang cenderung meningkatkan fokus pada hal tertentu. Di Indonesia, banyak perguruan tinggi yang sudah menyediakan berbagai macam pilihan program studi yang diminati masyarakat. Politeknik Harapan Bersama Tegal merupakan salah satu perguruan tinggi swasta yang menyediakan berbagai macam program studi [2].

Klasifikasi merupakan topik yang diangkat dalam penelitian ini. Selain digunakan untuk melakukan klasifikasi pada pengujian data yang bersifat numerik, klasifikasi juga dapat digunakan untuk data yang

bersifat kategorikal. Selain kategori dan numerical, klasifikasi juga dapat digunakan untuk mengidentifikasi gambar – gambar yang sesuai dengan kriteria yang diinginkan [3].

Naïve Bayes adalah model pertama yang digunakan untuk klasifikasi pada penelitian ini. Naïve Bayes dapat bertugas untuk menetapkan label yang relevan. Tugas ini dilakukan dengan memperluas pengklasifikasian satu label seperti pengklasifikasi Naïve Bayes. Secara khusus, klasifikasi Naïve Bayes dapat melakukan evaluasi pada data uji eksperimental dan pemodelan prediktif [4].

Sedangkan untuk model kedua yang digunakan untuk klasifikasi pada penelitian ini adalah Decision Tree . Metode ini merupakan salah satu metode yang diusulkan karena fleksibel yang memungkinkan untuk mengidentifikasi kondisi pada suatu kriteria. Metode Decision Tree membuat kumpulan data pelatihan dan pengujian saat membuat dan memvalidasi pohon keputusan, untuk meminimalkan beban komputasi [5].

Selain dua metode diatas, metode lain yang digunakan untuk penelitian ini adalah Unsupervised Discretization. Ini merupakan salah satu metode pendekatan yang sederhana yang dipakai untuk mengoptimalkan dua metode yaitu metode Naïve Bayes dan Decision Tree. Dari kedua metode tersebut akan dicari manakah yang terbaik dari keduanya setelah dilakukan optimalisasikan [6].

Oleh karena itu dalam penelitian ini bermaksud ingin melakukan penelitian yang dimana penelitian tersebut adalah pengembangan dari penelitian sebelumnya, yaitu tentang penentuan program studi untuk calon mahasiswa baru dengan menggunakan metode Naïve Bayes dan optimalisasi dengan pendekatan Unsupervised Discretization. Oleh karena itu maka penelitian ini diberi judul Optimalisasi Metode Naive Bayes Dan Decision Tree Untuk Menentukan Program Studi Bagi Calon Mahasiswa Baru Dengan Pendekatan Unsupervised Discretization.

## 2. KERANGKA TEORI

### 2.1. Metode Naïve Bayes

Menurut Alfa Saleh dalam penelitiannya tentang Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga yang menjelaskan definisi Naive Bayes yang merupakan sebuah model pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas

dengan menjumlahkan frekuensi dan komposisi nilai dari dataset yang diberikan. Algoritma yang digunakan menggunakan teorema dari Bayes dan mengasumsikan bahwa semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variable kelas [7].

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. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

## 2.2. Metode Decision Tree

Algoritma Decision Tree merupakan suatu metode pengklasifikasian yang menggunakan contoh pohon, menyatakan node yang menggambarkan tiap atribut, yang mana daun menggambarkan tiap kelas, juga setiap cabangnya menggambarkan nilai dari tiap kelas. Node akar menyatakan node yang berada paling atas dari pohon. Setiap node ini menggambarkan node pembagi, yang mana tiap node ini merupakan satu masukan dan memiliki sedikitnya dua keluaran [8].

Leaf node adalah node terakhir, hanya mempunyai satu masukan, dan tidak mempunyai keluaran. Pohon keputusan pada tiap leaf node menyatakan label tiap kelas. Pohon keputusan pada tiap cabangnya menyatakan keadaan yang harus diisi dan tiap puncak pohonnya menggambarkan nilai kelas data [9].

## 2.3. Metode Unsupervised Discretization

Metode Unsupervised Discretization merupakan salah satu metode yang sederhana. Salah satu metode sederhana lainnya seperti (equal-width Interval Discretization dan equal frequency Interval Discretization) dan yang lebih canggih, berdasarkan analisis clustering, seperti k-means discretization. Rentang kontinu dibagi menjadi sub rentang oleh lebar yang ditentukan pengguna atau Frekuensi. Namun pada penelitian ini, peneliti menggunakan teknik Equal-width interval Discretization, yaitu metode diskretisasi paling sederhana yang membagi rentang nilai yang teramati pada setiap fitur/atribut, di mana k adalah parameter yang disediakan oleh pengguna. Prosesnya melibatkan pemilahan nilai-nilai yang teramati dari fitur/atribut kontinyu dan menemukan nilai minimum (Vmin) dan maksimum (Vmax). Interval dapat dihitung

dengan membagi kisaran nilai yang teramati untuk variabel ke dalam k yang berukuran sama dengan menggunakan rumus berikut, dimana k adalah parameter yang diberikan oleh pengguna [10][3].

$$\text{Interval} = \frac{V_{max} - V_{min}}{k}$$

$$\text{Boundaries} = \{V_{min} + (i \times \text{Interval})\}$$

Kemudian batas-batasnya dapat dibangun untuk  $i = 1 \dots k-1$  dengan menggunakan persamaan di atas. Jenis diskretisasi ini tidak bergantung pada struktur data multirelasional. Namun, metode discretization ini sensitif terhadap outlier yang dapat secara drastis mengurangi kisaran. Keterbatasan metode ini diberikan oleh distribusi poin data yang tidak merata: beberapa interval mungkin berisi lebih banyak titik data daripada yang lain.

## 2.4. Pustaka Rujukan

Pada penelitian terkait tentang pemilihan program studi yaitu dilakukan Arif Sofanudin yakni tentang Implementasi Algoritma Naive Bayes Untuk Penjurusan Di Ma Al Mahrusiyah Lirboyo. Pada penelitian ini dilatarbelakangi berdasarkan hasil pengamatan yang dilakukan yang pada saat itu seorang siswa naik kelas, dari kelas X naik ke kelas XI. Siswa tersebut mengalami persoalan baru yaitu dalam menentukan bidang jurusan. Ada kemungkinan kesalahan yang dilakukan oleh siswa tersebut dalam penjurusan, yaitu rendahnya prestasi belajar siswa atau terjadinya ketidakcocokan dalam memilih jurusan sehingga mempengaruhi ketika akan masuk ke jenjang perguruan tinggi. Dari berbagai hal tersebut adapun solusi yang ditawarkan dengan memanfaatkan teknologi dan data mining dengan membangun system untuk menentukan jurusan menggunakan program PHP dengan Teknik mengimplementasikan pengelompokan data yang diperoleh dari hasil data mining MA Al Mahrusiyah Lirboyo dengan metode Naive Bayes. Dengan menggunakan metode implementasi Naive Bayes dan program PHP maka guru yang ada dalam MA Al Mahrusiyah Lirboyo dapat semakin mudah dalam menentukan jurusan berdasarkan nilai prestasi dan minat sebagai acuan sedangkan bagi siswa tersebut tidak perlu khawatir dalam memilih jurusan karena sudah disesuaikan dengan nilai prestasi siswa itu sendiri dan berdasarkan kemampuannya [2].

Penelitian tentang penentuan atau pemilihan jurusan yang dilakukan oleh Alfa Saleh dan Fina Nasari yang berjudul Penggunaan Teknik Unsupervised Discretization Pada Metode Naive Bayes Dalam

Menentukan Jurusan Siswa Madrasah Aliyah pada tahun 2017. Pada penelitian ini melakukan optimalisasi metode Naive Bayes dengan menerapkan Teknik Unsupervised Discretization yang akan ditransformasikan kriteria numerik atau kontinyu menjadi kriteria kategorikal dan mengimplementasikan satu kriteria yang dianggap tidak mempunyai pengaruh terhadap keakuratan terhadap hasil pengujian, maka dengan begitu keakuratan hasil dari klasifikasi dapat meningkat. Dari hasil pengujian 120 data siswa memperoleh bukti bahwa penerapan Teknik Unsupervised Discretization pada metode Naive Bayes naik dari 90% menjadi 92,8% [11].

Penelitian berikutnya dilakukan oleh Jia Wu, Shirui Pan, Xingquan Zhu, Zhihua Cai, Peng Zhang, Chengqi Zhang tentang Self-adaptive attribute weighting for Naive Bayes classification. Pada penelitian ini metodenya adalah Artificial Immune System (AIS) untuk mencari bobot nilai atribut adaptif, klasifikasi yang digunakan adalah Naive Bayes. Metode yang diusulkan yaitu AIS atau AISWNB dimana metode tersebut digunakan untuk mencari bobot atribut agar mencapai nilai yang optimal dengan menggunakan teori imunitas. Hasil yang diperoleh dari penelitian ini, bahwa AISWNB dapat memperoleh bobot nilai atribut yang baik selama proses pengujian, dari eksperimen 36 data set yang digunakan dan 6 data set Field Code Changed klasifikasi gambar menunjukkan bahwa AISWNB secara signifikan mengungguli rekan-rekannya dalam nilai akurasi klasifikasi, estimasi probabilitas kelas, dan kinerja peringkat kelas [12].

Penelitian selanjutnya dilakukan oleh Tzu-Tsung Wong yaitu tentang A hybrid discretization method for Naive Bayesian classifier. Pada penelitian ini diusulkan ukuran nonparametrik untuk mengevaluasi tingkat ketergantungan antara atribut kontinyu dan kelas. Ukuran nonparametrik kemudian untuk mengembangkan metode hibrida untuk mendiskritisasi atribut kontinyu sehingga nilai akurasi klasifikasi Naive Bayes dapat ditingkatkan. Metode Hybrid ini diuji pada 20 data set, dan hasilnya menunjukkan bahwa mendiskritisasi atribut kontinyu dalam data set dengan berbagai metode umumnya dapat memiliki prediksi nilai akurasi yang lebih tinggi [10].

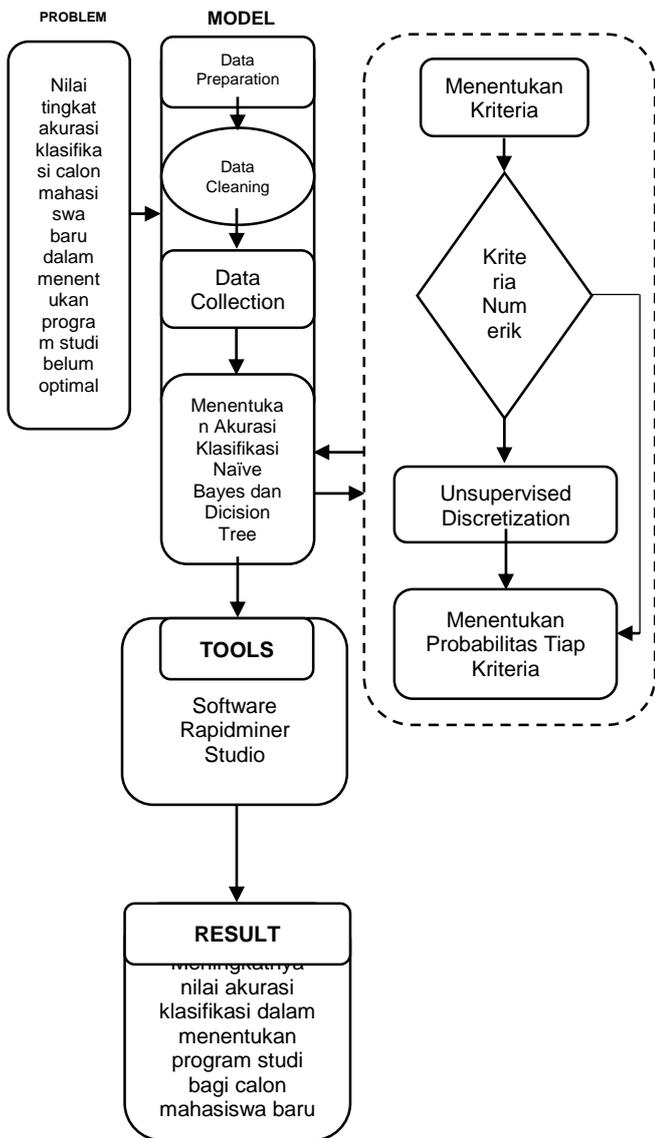
Penelitian berikutnya dilakukan oleh R.F. Roper, S. Renooij, L.C. van der Gaag pada tahun 2017 yang berjudul Discretizing environmental data for learning Bayesian-network classifiers. Pada penelitian ini membahas tentang prediksi keberadaan berbagai macam spesies burung di Andalusia dari data penggunaan lahan. Pada penelitian ini model yang diusulkan adalah membandingkan antara model

Bayesian-network classifiers dan logistic-regression. Data set yang digunakan adalah data seimbang dan kurang seimbang, model yang dipelajari adalah dari data kontinyu asli dan dari data setelah diskritisasi yang mempunyai tujuan akhir bahwa membandingkan empat metode diskritisasi yang berbeda antara lain Equal Frequency, Equal Width, Chi-Merge dan MDLP. Dari hasil percobaan data set spesies ini menunjukkan bahwa klasifikasi Naive Bayes sederhana lebih banyak disukai dari pada model regresi logistik dan metode Chi-Merge yang related tidak dikenal adalah metode yang disukai untuk mendiskritisasi data lingkungan ini [13].

Penelitian berikutnya yang berkaitan dengan perbandingan dua metode yang sama yaitu metode Naive Bayes dan metode Decision Tree ada pada penelitian yang dilakukan oleh Dewan Md. Farid, Li Zhang, Chowdhury Mofizur Rahman, M.A.Hossain, Rebecca Strachan. Penelitian ini memperkenalkan dua metode yang digunakan untuk meningkatkan klasifikasi pada Decision Tree dan Naive Bayes untuk klasifikasi masalah kelas jamak. Hasil eksperimen menunjukkan bahwa metode yang diusulkan telah menghasilkan hasil yang mengesankan dalam klasifikasi masalah kelas jamak yang menantang di kehidupan nyata. Mereka juga dapat secara otomatis mengekstrak kumpulan data pelatihan yang paling berharga dan mengidentifikasi atribut yang paling efektif untuk deskripsi instance dari database pelatihan kompleks yang berisikan dengan dimensi atribut yang besar [5].

### 3. METODE

Metode yang diusulkan seperti terlihat pada bagan di bawah ini:



Gambar 1. Metode Penelitian

Pada gambar diatas adalah kerangka pemikiran yang diusulkan pada penelitian ini. Dataset pada penelitian adalah data calon mahasiswa baru Politeknik Harapan Bersama Tegal, dataset berupa tersebut file excel. Kemudian data tersebut dilakukan preposisi agar dapat dilakukan klasifikasi menggunakan metode Naive Bayes dan Decision Tree.

Tahapan penelitian yang akan dilakukan preposisi data yakni pengolahan data mentah dari data penerimaan mahasiswa baru. Dalam metode Naive Bayes dan Decision Tree, data String atau kategorikal yang bersifat konstan dibedakan menjadi dua macam yaitu data numerik yang bersifat kontinyu, sehingga perbedaan yang dihasilkan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria, baik kriteria dengan nilai data string atau kriteria dengan data numerik.

#### 4. HASIL DAN PEMBAHASAN

##### 4.1. Hasil Menggunakan Metode Naive Bayes

Tahapan penelitian yang akan dilakukan preposisi data yakni pengolahan data mentah dari data penerimaan calon mahasiswa baru tahun akademik 2019 / 2020 dengan jumlah dataset 2256. Dalam metode Naive Bayes, data String atau kategorikal yang bersifat konstan dibedakan menjadi dua macam yaitu data numerik yang bersifat kontinyu, sehingga perbedaan yang dihasilkan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria, baik kriteria dengan nilai data string atau kriteria dengan data numerik. Adapun tahapan yang dilakukan adalah sebagai berikut:

1. Data Collection, yaitu data yang dijadikan sebagai data training dan testing, dalam hal ini adalah data calon mahasiswa baru. Kriteria untuk menentukan data training dan data testing adalah untuk data training 70% dan data testing 30%.
2. Data Cleaning, yaitu pada tahap ini terdapat sebuah kriteria yang dieliminasi dikarenakan kriteria tersebut tidak berpengaruh pada hasil akurasi klasifikasi metode Naive Bayes. Dari jumlah dataset sebesar 2256, setelah dilakukan eliminasi terhadap dataset tersebut, maka jumlah datasetnya sebesar 1671. Sedangkan jumlah atribut yang digunakan sebanyak 23.
3. Menentukan Kriteria, pada tahap menentukan kriteria ini digunakan kriteria data yang berdasarkan data yang telah terkumpul.
4. Menentukan Probabilitas Setiap Kriteria, pada tahap ini menentukan kriteria yang dijadikan acuan dalam mengklasifikasi program studi bagi calon mahasiswa baru.
5. Pengujian, pada tahap ini merupakan tahap penerapan Unsupervised Discretization pada metode Naive Bayes dengan beberapa data yang siap diujikan.

Adapun 20 contoh data yang telah siap untuk dilakukan tahapan berikutnya adalah 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...	DANY Y...	Reguler...	Gelomb...	JLSETIA...	L	BREBES	Dec 3, 1...	20.759	ISLAI

Gambar 2. Contoh DataSet Yang Akan Di Training Dan Testing

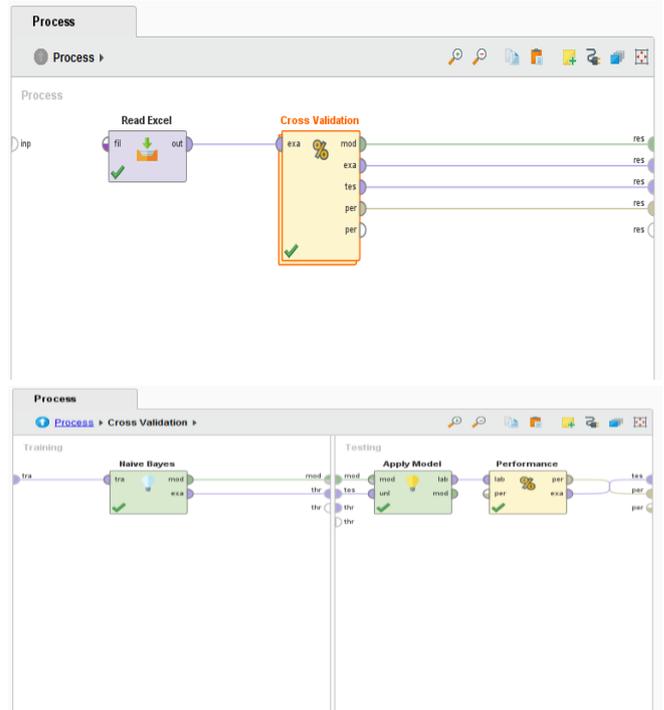
Data tersebut diatas kemudian siap untuk training dan testing dengan metode Optimalisasi Naive Bayes dan pendekatan Unsupervised Discretization. Tetapi sebelum memasuki tahap Optimalisasi Naive Bayes menggunakan Rapidminer, terlebih dahulu harus dilakukan metode Naive Bayes dengan menggunakan rapidminer. 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 terlihat pada tabel berikut:

Tabel 1. Hasil Eksperimen Menggunakan Naive Bayes

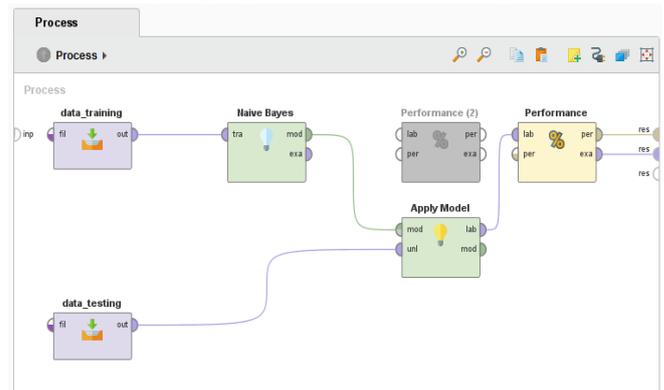
Test	Metod	Number Of Cross Validation	Sampling Type	Accuracy
1	Naive Bayes	5	Automatic	18.07%
	Naive Bayes	5	Linear	15.80%
	Naive Bayes	5	Saffled	19.27%
	Naive Bayes	5	Stratified	18.07%
	Naive Bayes	5	Sampling	18.07%
2	Naive Bayes	10	Automatic	19.33%
	Naive Bayes	10	Linear	15.44%
	Naive Bayes	10	Saffled	18.67%
	Naive Bayes	10	Stratified	19.33%
	Naive Bayes	10	Sampling	18.67%
3	Naive Bayes	15	Automatic	18.79%
	Naive Bayes	15	Linear	15.15%
	Naive Bayes	15	Saffled	18.20%
	Naive Bayes	15	Stratified	18.79%
	Naive Bayes	15	Sampling	18.79%
4	Naive Bayes	20	Automatic	18.74%
	Naive Bayes	20	Linear	14.54%
	Naive Bayes	20	Saffled	18.56%
	Naive Bayes	20	Stratified	18.74%
	Naive Bayes	20	Sampling	18.74%

Dari beberapa pengujian diatas memperoleh nilai akurasi paling tinggi 19.33%, nilai akurasi tersebut merupakan nilai yang belum sesuai dengan apa 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 di bawah ini:



Gambar 3. Hasil Desain Eksperimen Dengan Cross Validation Pada Naive Bayes

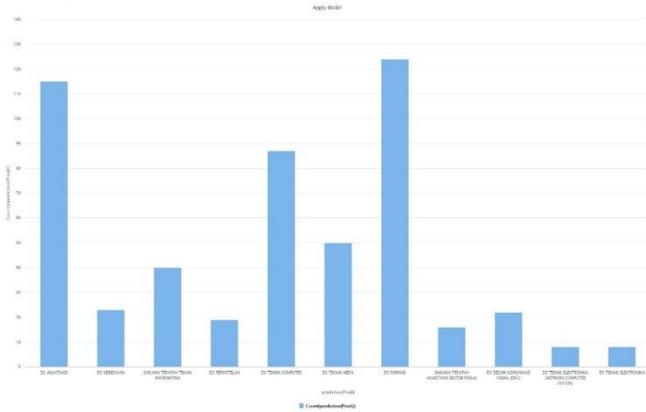
Pada eksperimen berikutnya dengan desain eksperimen yang berbeda dengan yang terdahulu mendapatkan nilai akurasi 96.68%. Adapun desain eksperimen yang digunakan adalah sebagai berikut:



Gambar 4. Hasil Desain Eksperimen Data Training Dan Testing Pada Naive Bayes

Desain eksperimen diatas menggambarkan bahwa dataset yang digunakan dipecah menjadi dua bagian dengan rasio 70% digunakan untuk data training dengan jumlah record 1158 data dan 30% data testing dengan jumlah record 513 data dari jumlah total record dari dataset 1671. Dan hasil akurasi yang diperoleh dari desain eksperimen diatas dengan menggunakan metode Naive Bayes adalah nilai akurasinya 96.68%. Nilai akurasi tersebut masih belum menggunakan optimalisasi pedekatan Unsupervised Discretization. Nilai akurasi tersebut dapat dilihat melalui grafik

penentuan program studi pilihan mahasiswa, yaitu sebagai berikut:



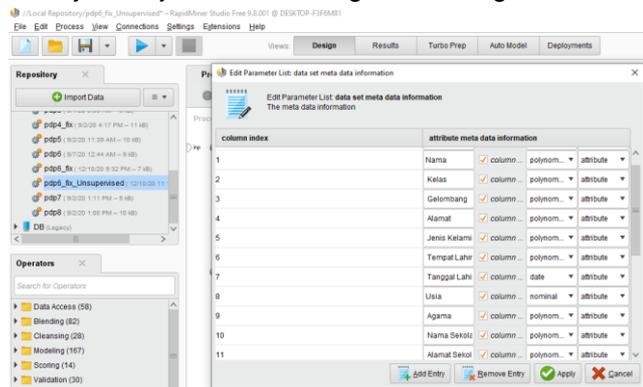
Gambar 5. Grafik Hasil Eksperimen Pada Naive Bayes

Dari hasil grafik eksperimen diatas dapat disimpulkan bahwa pemilihan program studi bagi calon mahasiswa baru menunjuk pada program studi DIII Farmasi.

#### 4.2. Hasil Menggunakan Metode Naive Bayes Dan Unsupervised Discretization

Sedangan untuk optimalisasi dengan pendekatan Unsupervised Discretization masih menggunakan desain eksperimen seperti tampilan diatas tetapi disini menitikberatkan pada transformasi kriteria numerik / kontinyu menjadi kriteria kategorikal dan menghilangkan satu kriteria yang dianggap tidak berpengaruh pada keakuratan hasil pengujian, dengan keakurasian hasil klasifikasi dapat meningkat.

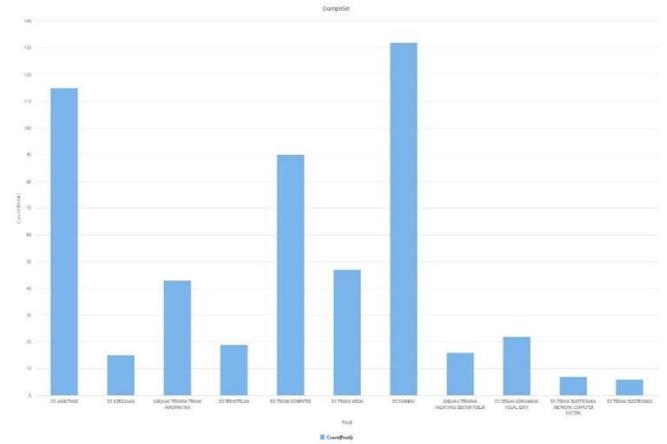
Adapun tampilan perubahan kriteria numerik / kontinyu menjadi kriteria kategorikal sebagai berikut:



Gambar 6. Hasil Perubahan Kriteria Numerik Menjadi Kriteria Kategorikal

Desain eksperimen diatas menggambarkan bahwa dataset yang digunakan dipecah menjadi dua bagian dengan rasio 70% digunakan untuk data training dengan jumlah record 1158 data dan 30% data testing dengan jumlah record 513 data dari jumlah total record dari dataset 1671. Dan hasil akurasi yang diperoleh dari

desain eksperimen diatas dengan menggunakan metode Naive Bayes dengan optimalisasi pedekatan Unsupervised Discretization adalah nilai akurasinya 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 penentuan program studi pilihan mahasiswa, yaitu sebagai berikut:

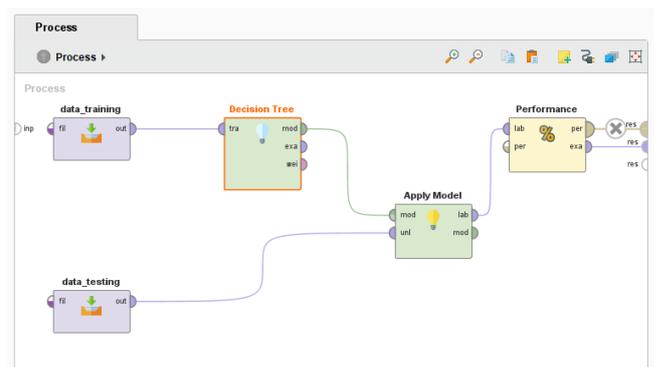


Gambar 7. Grafik Hasil Eksperimen Pada Naive Bayes Dan Unsupervised Discretization

Dari hasil grafik eksperimen diatas dapat disimpulkan bahwa pemilihan program studi bagi calon mahasiswa baru masih menunjuk pada Program Studi DIII Farmasi.

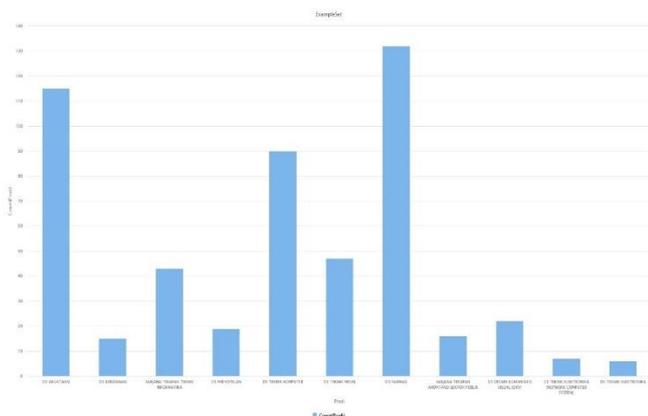
#### 4.3. Hasil Menggunakan Metode Decision Tree

Pada pengujian menggunakan metode Decision Tree terdapat beberapa kriteria yang digunakan antara lain gain\_rasio, information\_gain, gini\_index, dan accuracy. Kriteria tersebut merupakan salah satu dari beberapa operator yang nantinya akan menghasilkan memperkirakan seberapa akurat suatu model yang akan tampil. Adapun desain pengujian menggunakan metode Decision Tree adalah sebagai berikut:



Gambar 8. Hasil Desain Eksperimen Data Training Dan Testing Pada Decision Tree

Desain eksperimen diatas menggambarkan bahwa dataset yang digunakan dipecah menjadi dua bagian dengan rasio 70% digunakan untuk data training dengan jumlah record 1158 data dan 30% data testing dengan jumlah record 513 data dari jumlah total record dari dataset 1671. Pengujian data training dan testing di atas menggunakan metode Decision Tree memperoleh hasil akurasi 51.56%. Nilai akurasi ini diperoleh dengan berbagai macam pengujian antara lain dengan nama kriteria gini\_index dengan nilai maximum depth 20, Apply Pruning Confidence 0.1 dan 0.2, dan Apply Prepruning 0.01 dan 0.02. Jika pengujian dilakukan dengan nilai diatas nilai maximum depth 20, misalkan 50 dan nilai Apply Pruning Confidence, nilai Apply Prepruning nilainya tetap maka hasil yang diperoleh untuk nilai juga sama sebesar 51.56%. Nilai akurasi tersebut dapat dilihat melalui grafik penentuan program studi pilihan mahasiswa, yaitu sebagai berikut:



Gambar 9. Grafik Hasil Eksperimen Pada Decision Tree

Dari hasil grafik eksperimen diatas dapat disimpulkan bahwa pemilihan program studi bagi calon mahasiswa baru masih menunjuk pada Program Studi DIII Farmasi.

#### 4.4. Hasil Menggunakan Metode Decision Tree Dan Unsupervised Discretization

Pada pengujian data training dan data testing menggunakan metode Decision Tree dan optimalisasi pendekatan Unsupervised Discretization mengalami penurunan nilai akurasi. Padahal untuk desain eksperimen sama dengan metode Naive Bayes dan Pendekatan metode Unsupervised Discretization serta menggambarkan bahwa dataset yang digunakan dipecah menjadi dua bagian dengan rasio 70% digunakan untuk data training dengan jumlah record 1158 data dan 30% data testing dengan jumlah record 513 data dari jumlah total record dari dataset 1671.

Hasil yang diperoleh dari desain eksperimen menggunakan metode ini adalah 49.80%.

#### 4.5. Pembahasan

Jika dibandingkan antara penelitian terdahulu dengan penelitian yang sedang dikerjakan maka hasilnya lebih baik dari hasil penelitian terdahulu dibandingkan dengan sekarang, perbandingan hasil penelitian terdahulu dengan sekarang dapat dilihat pada tabel berikut:

Tabel 1. Perbandingan Hasil Penelitian

No	Metode	Hasil Akurasi
1.	Naive Bayes	96.68%
2.	Naive Bayes + Unsupervised Discretization	97.66%
3.	Decision Tree	51.56%
4.	Decision Tree + Unsupervised Discretization	49.80%

Dari tabel diatas menunjukkan bahwa antara penelitian terdahulu dengan penelitian yang sekarang lebih baik penelitian terdahulu, karena dilihat dari hasil akurasi yang diperoleh antara penelitian yang terdahulu memperoleh nilai akurasi 96.68%, dan setelah dilakukan optimalisasi dengan pendekatan metode Unsupervised Discretization memperoleh nilai akurasi sebesar 97.66%, sedangkan penelitian yang sekarang mengalami penurunan nilai akurasi, dengan menggunakan metode Decision Tree tanpa melakukan optimalisasi memperoleh nilai akurasi sejumlah 51.56% dan setelah dilakukan optimalisasi metode Decision Tree menggunakan pendekatan metode Unsupervised Discretization memperoleh nilai akurasi sejumlah 49.80%.

Eksperimen diatas menunjukkan bahwa untuk melakukan klasifikasi minat dalam menentukan program studi untuk calon mahasiswa baru menggunakan metode Naive Bayes, metode Naive Bayes dan Pendekatan Unsupervised Discretization dengan metode Decision Tree, metode Decision Tree dan pendekatan Unsupervised Discretization yang lebih baik digunakan adalah metode Naive Bayes dan pendekatan Unsupervised Discretization, karena hasil akurasi yang diperoleh sejumlah 97.66%.

## 5. SIMPULAN

Dari pembahasan diatas maka dapat diambil beberapa kesimpulan antara lain:

1. Klasifikasi minat dalam menentukan program studi bagi calon mahasiswa baru dapat dilakukan dengan beberapa metode, antara lain dengan

- metode Naïve Bayes, Naïve Bayes dan optimalisasi dengan pendekatan Unsupervised Discretization, Decision Tree, Decision Tree dan optimalisasi dengan pendekatan Unsupervised Discretization.
2. Hasil penelitian yang dilakukan dari tahap awal hingga pengujian dengan menerapkan metode klasifikasi Naïve Bayes memperoleh hasil akurasi sebesar 96.68%, metode klasifikasi Naïve Bayes dan pendekatan Unsupervised Discretization mendapatkan hasil akurasi sebesar 97.66%, sedangkan untuk metode pembandingan dalam penelitian ini adalah metode klasifikasi Decision Tree memperoleh hasil akurasi sebesar 51.56%, sedangkan selanjutnya metode klasifikasi Decision Tree dan pendekatan Unsupervised Discretization memperoleh hasil akurasi sebesar 49.80%.
  3. Hasil akhir dari metode klasifikasi yang terbaik dalam penelitian ini yaitu untuk menentukan minat dalam penentuan program studi bagi calon mahasiswa adalah metode Naïve Bayes dan pendekatan Unsupervised Discretization, karena nilai akurasi yang diperoleh sebesar 97.66%.

#### UCAPAN TERIMA KASIH

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.
- [2] M. A. Al and M. Lirboyo, "Program studi sistem informasi fakultas teknik universitas nusantara PGRI Kediri 2017," 2017.
- [3] 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.
- [4] 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.
- [5] L. Vanfretti and V. S. N. Arava, "Electrical Power and Energy Systems Decision tree-based classification of multiple operating conditions for power system voltage stability assessment," *Electr. Power Energy Syst.*, vol. 123, no. April, p. 106251, 2020.
- [6] M. Shanmugapriya, H. K. Nehemiah, R. S. Bhuvaneshwaran, K. Arputharaj, and J. Jabez Christopher, "Unsupervised Discretization: An Analysis of Classification Approaches for Clinical Datasets," *Res. J. Appl. Sci. Eng. Technol.*, vol. 14, no. 2, pp. 67–72, 2017.
- [7] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," vol. 2, no. 3, pp. 207–217, 2015.
- [8] I. Sutoyo, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 217, 2018.
- [9] N. F. Romdhoni, K. Usman, and B. Hidayat, "Deteksi Kualitas Kacang Kedelai Melalui Pengolahan Citra Digital dengan Metode Gray-Level Co-Occurrence Matrix ( GLCM ) dan Klasifikasi Decision Tree," *Pros. Semin. Nas. Ris. dan Inf. Sci.*, vol. 2, pp. 132–137, 2020.
- [10] T. Wong, "A hybrid discretization method for naive Bayesian classifiers," *Pattern Recognit.*, vol. 45, no. 6, pp. 2321–2325, 2012.
- [11] A. Saleh, F. Nasari, U. P. Utama, and K. J. Siswa, "Penggunaan Teknik Unsupervised Discretization Pada Metode Naive Bayes Dalam Menentukan Jurusan Siswa Applied Of Unsupervised Discretization Technique In Naive Bayes Method In Determining Madrasah Aliyah Students ' MAJOR," vol. 5, no. 3, pp. 353–360, 2018.
- [12] 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.
- [13] R. F. Roperio, S. Renooij, and L. C. van der Gaag, "Discretizing environmental data for learning Bayesian-network classifiers," *Ecol. Modell.*, vol. 368, pp. 391–403, 2018.