

PENGUNAAN TEKNIK *UNSUPERVISED DISCRETIZATION* PADA METODE NAIVE BAYES DALAM MENENTUKAN JURUSAN SISWA MADRASAH ALIYAH

Alfa Saleh¹, Fina Nasari²

^{1,2}Universitas Potensi Utama

Email: ¹alfa@potensi-utama.ac.id, ² fina@potensi-utama.ac.id

(Naskah masuk: 15 Maret 2018, diterima untuk diterbitkan: 28 Juli 2018)

Abstrak

Pemilihan jurusan bagi siswa merupakan langkah positif yang dilakukan untuk memfokuskan siswa sesuai dengan potensi yang dimiliki, hal ini dianggap penting karena dengan adanya jurusan, siswa diharapkan mampu mengembangkan kemampuan akademis sesuai bidang yang dikuasai. Pada penelitian sebelumnya, telah dilakukan pengujian dengan metode *Naive Bayes* yang bertujuan untuk mengklasifikasikan jurusan siswa berdasarkan kriteria yang menunjang dengan studi kasus pada siswa Madrasah Aliyah Swasta PAB 6 Helvetia, dan didapatkan hasil pengujian dari 100 data siswa dengan tingkat keakuratan 90%. pada penelitian ini, dilakukan optimalisasi metode yang digunakan sebelumnya dengan menerapkan teknik *Unsupervised Discretization* yang akan mentransformasikan kriteria numerik/kontinu menjadi kriteria kategorikal dan mengeliminasi satu kriteria yang dianggap tidak mempengaruhi keakuratan hasil pengujian, dengan begitu keakuratan hasil klasifikasi dapat meningkat. Dari 120 data siswa yang diuji, terbukti bahwa hasil klasifikasi penerapan teknik *unsupervised discretization* pada metode *naive bayes* naik dari 90% menjadi 92.8%.

Kata kunci: *Naive Bayes, Unsupervised Discretization, Klasifikasi Jurusan Siswa*

APPLIED OF UNSUPERVISED DISCRETIZATION TECHNIQUE IN NAIVE BAYES METHOD IN DETERMINING MADRASAH ALIYAH STUDENTS' MAJOR

Abstract

Selection of majors for students is a positive step that is done to focus students in accordance with their potential, it is considered important because with the majors, students are expected to develop academic ability according to the controlled field. In previous research, Naive Bayes method has been tested to classify the students department based on the supportive criterias (case study on Madrasah Aliyah PAB 6 Helvetia), and the test result of 100 students data, the classification accuracy is about 90% . in this study, optimizaton is done with a method used earlier by applying Unsupervised Discretization techniques that would transform numerical / continuous criteria into categorical criteria and eliminating one criterion that is considered not affect the accuracy of test results. thus the accuracy of classification results could increase. 120 students data is tested, it is evident that the results of the classification of the application of unsupervised discretization techniques on the Naive Bayes method rose from 90% to 92.8%.

Keywords: *Naive Bayes, Unsupervised Discretization, Students' Major*

1. PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi berkembang begitu pesat bahkan telah merambah ke semua aspek kehidupan. Mulai dari bisnis, sosial hingga pendidikan. Peran pendidikan sangat penting dalam mendukung perkembangan teknologi yang hampir sudah merambah ke semua bidang. Hal ini tentu juga berpengaruh pada penentuan jurusan bagi siswa SMA/MA sederajat, yang mana penentuan jurusan siswa tersebut merupakan proses untuk memfokuskan siswa dalam bidang konsentrasi tertentu, hal ini dilakukan agar setiap siswa dapat mempelajari lebih dalam mata pelajaran – mata

pelajaran yang sesuai dengan konsentrasi yang telah ditentukan untuk siswa tersebut. Yang menjadi masalah ialah sistem yang tengah berjalan pada sekolah Madrasah Aliyah Swasta PAB 2 Helvetia Medan, tidak sepenuhnya efektif dikarenakan para siswa diberikan angket untuk menentukan jurusan apa yang diminati tanpa mempertimbangkan kriteria lain yang mungkin memiliki andil untuk menentukan kelayakan siswa tersebut dalam hal memilih jurusan. Padahal proses penentuan jurusan untuk siswa merupakan langkah penting dalam mempersiapkan siswa untuk berkonsentrasi pada bidang yang siswa tersebut minati saat harus melanjutkan pada jenjang pendidikan

selanjutnya. penelitian ini juga kiranya telah menjadi bahan penelitian untuk kategori Sistem pendukung keputusan dalam menentukan jurusan di sekolah menengah atas sesuai dengan kemampuan siswa berdasarkan nilai semester, nilai potensi dan nilai pilihan siswa (Prawira & Hakim, 2011). Penelitian lainnya seputar pemilihan jurusan telah diteliti, di mana dalam proses pemilihan jurusan ini diimplementasikan metode 360 derajat guna menentukan kesesuaian antara jurusan dan kriteria lain yang mendukung (Worang dkk, 2013). Pada penelitian sebelumnya, juga telah dilakukan penelitian dengan melakukan proses mining untuk menggali informasi seputar penentuan jurusan siswa menggunakan metode Naive Bayes, dari hasil penelitian tersebut diuji 100 data siswa berdasarkan beberapa kriteria meliputi nilai rata-rata mata pelajaran ilmu pengetahuan alam, nilai rata-rata ilmu pengetahuan sosial, rekomendasi guru kelas serta nilai angket yang diisi oleh siswa bersangkutan. Dari 100 data yang diuji menggunakan metode Naive Bayes, maka didapatkan nilai akurasi penentuan jurusan siswa sebesar 90% dengan error sebesar 10% (Saleh, 2014). Metode naive bayes dipilih karena banyak diimplementasikan dalam berbagai bidang ilmu, seperti pada penelitian Xingxing Zhou (2016), metode naive bayes digunakan untuk mengklasifikasikan citra untuk meningkatkan akurasi diagnosis otak dengan menggunakan citra NMR, di mana didapatkan hasil klasifikasi sensitivitas sebesar 94.5%, spesifisitas sebesar 91.70% dan keakuratan keseluruhan sebesar 92.60 (Xingxing dkk, 2016). Naive bayes menjadi salah satu dari sepuluh (10) algoritma data mining terbaik karena kesederhanaan dan efisiensi, hal ini terbukti dari kinerja Naive bayes dalam mengklasifikasikan teks (Jiang, Wang & Zhang, 2016). Selain itu, Naive Bayes dikenal luas sebagai metode klasifikasi probabilistik yang sederhana dan efektif (Anderson & Dubnicka, 2014), dan kinerjanya sebanding dengan atau lebih tinggi dari pada pohon keputusan (decision tree) (Zhang, 2014) dan jaringan syaraf tiruan (Kotsiantis, 2014).

Namun, peneliti ingin mengembangkan penelitian sebelumnya dengan menerapkan *Unsupervised Discretization* (Palaniappan & Hong, 2008) untuk meningkatkan performa metode Naive Bayes sehingga persentase hasil keakuratan prediksi dapat meningkat dibandingkan sebelumnya. Di mana Penggunaan teknik *Unsupervised Discretization* dalam mentransformasikan kriteria/atribut numerik sangat baik (Kareem & Duaimi, 2016).

2. METODOLOGI PENELITIAN

2.1. Penelitian Terkait

Pada penelitian sebelumnya (Erfan dkk, 2012), penelitian dilakukan untuk menentukan jurusan siswa yang dinilai kurang efektif

disebabkan sistem pemilihannya masih manual, sehingga pada penelitian tersebut peneliti mempertimbangkan perlu adanya penggunaan metode untuk mengelompokkan siswa dalam proses pemilihan jurusan khusus bagi siswa kelas X (sepuluh). Penelitian ini menggunakan algoritma Fuzzy C-Means untuk mengelompokkan data siswa SMA berdasarkan Nilai mata pelajaran inti sebagai proses penjurusan. Adapun hasil pengelompokan tersebut dinilai lebih akurat dengan hasil akurasi 83.33%.

Penelitian ini membahas mengenai pemilihan jurusan siswa Madrasah Aliyah menggunakan metode klasifikasi Naive Bayes, di mana metode tersebut telah digunakan untuk memprediksi jurusan siswa dengan memanfaatkan historikal data siswa yang meliputi nilai kumulatif mata pelajaran pengetahuan alam atau eksakta (matematika, fisika, kimia dan biologi), nilai kumulatif mata pelajaran pengetahuan social atau non-eksakta (Ekonomi, Sejarah, Geografi dan Sosiologi), Rekomendasi wali kelas dan Angket yang diisi oleh siswa. Selain itu pada penelitian ini juga menerapkan teknik *Unsupervised Discretization* ke dalam metode *Naive Bayes*, dimana pada penelitian ini, setiap kriteria numerik yang digunakan pada penelitian sebelumnya akan ditransformasi menggunakan teknik diskritisasi, sehingga menghasilkan data kategorikal sesuai dengan interval data yang optimal serta jumlah data yang diuji bervariasi untuk melihat optimalisasi teknik *Unsupervised Discretization* pada jumlah data yang dinamis, serta bertujuan untuk menganalisa seberapa jauh penerapan teknik diskritisasi tersebut mampu dalam meningkatkan hasil akurasi prediksi penerapan metode naive bayes untuk kasus pemilihan jurusan siswa.

2.2. Naive Bayes

Persamaan dari teorema *Bayes* adalah (Prasojo dkk, 2017) :

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

Di mana :

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

Untuk menjabarkan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas mana

yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naive Bayes* di atas disesuaikan sebagai berikut :

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)} \quad (2)$$

Di mana Variabel C merepresentasikan kelas, sementara itu variabel $F1 \dots Fn$ merepresentasikan karakteristik petunjuk yang digunakan untuk proses klasifikasi. sehingga rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, yang juga dikenal dengan istilah *prior*), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut :

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (3)$$

2.3 Unsupervised Discretization

Di antara metode *unsupervised Discretization*, terdapat beberapa metode yang sederhana. 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 (Joita, 2010). 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 kontinu dan menemukan nilai minimum (V_{min}) dan maksimum (V_{max}). 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.

$$Interval = \frac{V_{max} - V_{min}}{k} \quad (4)$$

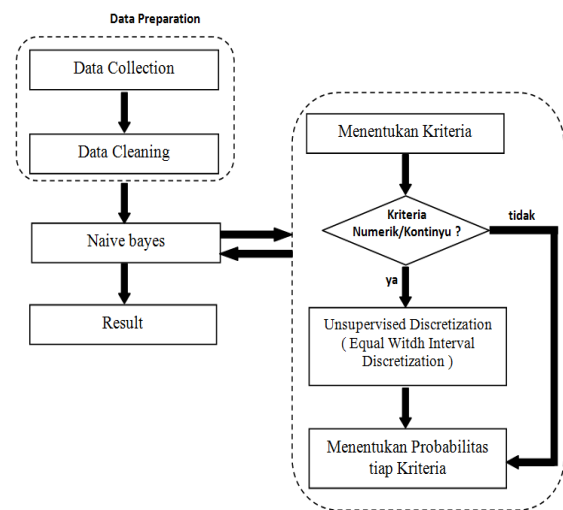
$$Boundaries = V_{min} + (i \times Interval) \quad (5)$$

Kemudian batas-batasnya dapat dibangun untuk $i = 1 \dots k-1$ dengan menggunakan persamaan di atas. Jenis diskretisasi ini tidak bergantung pada struktur data multi-relasional. 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.

3. HASIL DAN PEMBAHASAN

3.1. Pembahasan

Dalam metode *Naive Bayes*, data String(kategorikal) yang bersifat konstan dibedakan dengan data numerik yang bersifat kontinu, perbedaan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria baik itu kriteria dengan nilai data string maupun kriteria dengan nilai data numerik. Adapun tahapan penerapan metode *Naive Bayes* pada penelitian ini dapat dilihat pada gambar 1 berikut.



Gambar 1. Tahapan Penerapan Unsupervised Discretization pada Metode Naive Bayes

A. Data Collection

Data yang akan dijadikan sebagai data latih (*data training*) adalah data akademik siswa, di mana jumlah sampel data siswa yang diambil sebanyak 120 data, yang kemudian akan dilakukan proses *cleaning* guna menentukan data apa saja yang dianggap penting dan mempengaruhi hasil klasifikasi.

B. Data Cleaning

Berbeda dengan penelitian sebelumnya, pada penelitian ini, terdapat sebuah kriteria yang dieliminasi dikarenakan kriteria tersebut tidak begitu berpengaruh pada hasil akurasi klasifikasi metode *Naive Bayes*. adapun data yang akhirnya digunakan pada penelitian ini adalah data nilai mata pelajaran eksakta, nilai mata pelajaran non-eksakta, rekomendasi dari wali kelas, dan angket yang diisi oleh siswa.

C. Menentukan Kriteria

Adapun kriteria yang digunakan berdasarkan data yang telah dikumpulkan adalah seperti pada tabel 1 berikut :

Tabel 1. Kriteria

NO	Kriteria	Jenis Kriteria	Nilai Kriteria
1	Nilai rata-rata mata pelajaran eksakta	Numerik/Kontinyu	0 - 100
2	Nilai rata-rata mata pelajaran non-eksakta	Numerik/Kontinyu	0 - 100
3	Rekomendasi	Kategorikal	IPA,IPS
4	Angket	Kategorikal	IPA,IPS

Terdapat empat (4) kriteria yang digunakan dalam penelitian ini, yaitu nilai rata-rata mata pelajaran eksakta, nilai rata-rata mata pelajaran non-eksakta, rekomendasi dan angket. Dua (2) di antaranya merupakan kriteria numerik/kontinyu dan dua (2) kriteria kategorikal. Untuk meningkatkan tingkat akurasi metode naive bayes, maka dilakukan proses diskritisasi menggunakan teknik unsupervised discretization pada kriteria yang bernilai numerik/kontinyu, tujuannya adalah untuk mentransformasikan kriteria yang bernilai numerik/kontinyu menjadi kriteria kategorikal menggunakan persamaan 4 dan 5. Berikut tabel 2 hasil diskritisasi kriteria numerik/kontinyu.

Tabel 2. Hasil Diskritisasi Kriteria Numerik/Kontinyu

Kriteria Numerik/Kontinyu	
Nilai Rata-Rata Mata Pelajaran Eksakta	Nilai Rata-Rata Mata Pelajaran Non-Eksakta
<71.53	<71.5
71.53-73.06	71.5-73
73.06-74.59	73-74.5
74.59-76.12	74.5-76
76.12-77.65	76-77.5
77.65-79.18	77.5-79
79.18-80.71	79-80.5
80.71-82.24	80.5-82
82.24-83.77	82-83.5
83.77>	83.5>
Nilai $k = 10$	

Pada tabel 2 di atas, dapat dilihat hasil dari proses diskritisasi menggunakan teknik *Unsupervised Discretization*. Di mana kriteria/atribut Nilai Rata-rata mata pelajaran eksakta dan non-eksakta yang bertipe numerik atau kontinyu di transformasikan menjadi kriteria kategorikal dengan 10 jenis kategori. Penerapan *Unsupervised Discretization* dilakukan untuk menyederhanakan data numerik yang dinilai tidak begitu optimal dalam mengklasifikasikan data siswa. Ini merupakan kelebihan dari proses diskritisasi pada penelitian ini.

D. Probabilitas Setiap Kriteria

Telah ditetapkan beberapa kriteria yang dijadikan acuan dalam mengklasifikasikan jurusan siswa menggunakan teknik *Unsupervised Discretization* pada metode *Naive Bayes*. Langkah selanjutnya, menentukan nilai probabilitas setiap kriteria. Berikut nilai probabilitas kriteria nilai rata-rata mata pelajaran eksakta dapat dilihat pada tabel 3.

Tabel 3. Nilai Probabilitas Kriteria Nilai Rata-Rata Mata Pelajaran Eksakta

Nilai Rata-Rata Mata Pelajaran Non-Eksakta	Probabilitas	
	IPA	IPS
<71.5	0.033	0.142
71.5-73	0.008	0.025
73-74.5	0.050	0.067
74.5-76	0.075	0.092
76-77.5	0.000	0.008
77.5-79	0.017	0.017
79-80.5	0.075	0.075
80.5-82	0.092	0.050
82-83.5	0.025	0.008
83.5>	0.125	0.017
Jumlah	0.5	0.5

Dapat dilihat pada tabel 3 di atas, nilai probabilitas setiap keputusan (IPA/IPS) ditentukan dari nilai rata-rata siswa setiap kategorinya. Selanjutnya untuk Nilai probabilitas pada kriteria nilai rata-rata mata pelajaran non-eksakta dapat dilihat pada tabel 4 berikut.

Tabel 4. Nilai Probabilitas Kriteria Nilai Rata-Rata Mata Pelajaran Non-Eksakta

Nilai Rata-Rata Mata Pelajaran Eksakta	Probabilitas	
	IPA	IPS
<71.53	0.033	0.142
71.53-73.06	0.008	0.025
73.06-74.59	0.050	0.067
74.59-76.12	0.075	0.092
76.12-77.65	0.000	0.008
77.65-79.18	0.017	0.017
79.18-80.71	0.075	0.075
80.71-82.24	0.092	0.050
82.24-83.77	0.025	0.008
83.77>	0.125	0.017
Jumlah	0.5	0.5

Nilai probabilitas untuk kriteria rekomendasi dapat dilihat pada tabel 5.

Tabel 5. Nilai Probabilitas Kriteria Rekomendasi

Rekomendasi	Probabilitas	
	IPA	IPS
IPA	0.483	0.075
IPS	0.017	0.425
Jumlah	0.5	0.5

Berdasarkan tabel 5 di atas, nilai probabilitas kriteria rekomendasi untuk jurusan IPA adalah 0.483 dan 0.017, sementara nilai probabilitas untuk jurusan IPS sebesar 0.075 dan 0.425. Nilai probabilitas untuk kriteria angket dapat dilihat pada tabel 6.

Tabel 6. Nilai Probabilitas Kriteria Angket

Angket	Probabilitas	
	IPA	IPS
IPA	0.417	0.058
IPS	0.083	0.442
Jumlah	0.5	0.5

Selanjutnya, nilai probabilitas kriteria Angket untuk jurusan IPA adalah 0.417 dan 0.083, sementara nilai probabilitas kriteria Angket untuk jurusan IPS sebesar 0,058 dan 0.442

3.2 Pengujian

Berikut pengujian dari penerapan *unsupervised discretization* pada metode *Naive Bayes* dengan beberapa sampel data yang diuji. Adapun hasil pengujian menggunakan sampel 40 data dapat dilihat pada tabel 7 berikut.

Tabel 7. Hasil Pengujian dengan 40 data

No	Weighted Average	TP Rate	FP Rate	Precision
1	Naïve Bayes	0.950	0.050	0.950
2	Naïve Bayes with Unsupervised Descretization	0.950	0.050	0.950

Dari hasil pengujian pada 40 data sampel, penerapan metode *naive bayes* berhasil mengklasifikasikan 38 data dengan nilai presisi sebesar 95%, sementara hasil pengujian dari penerapan metode *naive bayes* yang dikombinasikan dengan teknik *unsupervised discretization* berhasil mengklasifikasikan 38 data dengan nilai presisi sebesar 95%. Sementara itu, pengujian juga dilakukan dengan 60 data, hasil pengujian dapat dilihat pada tabel 8 berikut.

Tabel 8. Hasil Pengujian dengan 60 data

No	Weighted Average	TP Rate	FP Rate	Precision
1	Naïve Bayes	0.933	0.067	0.933
2	Naïve Bayes with Unsupervised Descretization	0.967	0.031	0.969

hasil pengujian pada 60 data sampel, penerapan metode *naive bayes* berhasil mengklasifikasikan 56 data dengan nilai presisi sebesar 93.3%, sementara hasil pengujian penerapan metode *naive bayes* yang dikombinasikan dengan teknik *unsupervised discretization* berhasil mengklasifikasikan 57 data dengan nilai presisi sebesar 96.9%. pengujian selanjutnya dilakukan dengan 90 data, hasil pengujian dapat dilihat pada tabel 9 berikut.

Tabel 9. Hasil Pengujian dengan 80 data

No	Weighted Average	TP Rate	FP Rate	Precision
1	Naïve Bayes	0.925	0.075	0.925
2	Naïve Bayes with Unsupervised Descretization	0.938	0.061	0.940

hasil pengujian pada 80 data sampel, penerapan metode *naive bayes* berhasil mengklasifikasikan 74 data dengan nilai presisi sebesar 92.5%, sementara hasil pengujian penerapan metode *naive bayes* yang dikombinasikan dengan teknik *unsupervised discretization* berhasil mengklasifikasikan 75 data dengan nilai presisi sebesar 94%. Pengujian dilakukan kembali dengan jumlah data yang lebih besar yaitu 100 data, hasil pengujian dapat dilihat pada tabel 10 berikut.

Tabel 10. Hasil Pengujian dengan 100 data

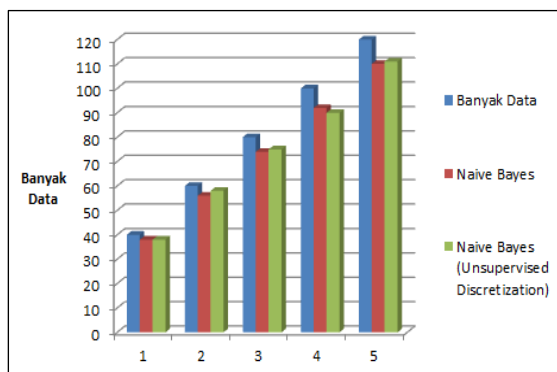
No	Weighted Average	TP Rate	FP Rate	Precision
1	Naïve Bayes	0.920	0.80	0.920
2	Naïve Bayes with Unsupervised Descretization	0.900	0.10	0.900

hasil pengujian pada 100 data sampel, penerapan metode *naive bayes* berhasil mengklasifikasikan 92 data dengan nilai presisi sebesar 92%, sementara hasil pengujian penerapan metode *naive bayes* yang dikombinasikan dengan teknik *unsupervised discretization* berhasil mengklasifikasikan 90 data dengan nilai presisi sebesar 90%. Terakhir pengujian dilakukan dengan 120 data, hasil pengujian dapat dilihat pada tabel 11 berikut.

Tabel 11. Hasil Pengujian dengan 120 data

No	Weighted Average	TP Rate	FP Rate	Precision
1	Naïve Bayes	0.917	0.083	0.917
2	Naïve Bayes with Unsupervised Descretization	0.925	0.075	0.928

hasil pengujian pada 120 data sampel, penerapan metode naive bayes berhasil mengklasifikasikan 110 data dengan nilai presisi sebesar 91.7%, sementara hasil pengujian penerapan metode naive bayes yang dikombinasikan dengan teknik unsupervised discretization berhasil mengklasifikasikan 111 data dengan nilai presisi sebesar 92.8%. beberapa pengujian yang telah dilakukan dengan jumlah data yang bervariasi. berdasarkan nilai probabilitas setiap kriteria yang telah ditentukan sebelumnya, dimana proses diskritisasi mengambil peran cukup penting dalam pengujian tersebut, perubahan kriteria numerik menjadi kriteria kategorikal dengan memanfaatkan teknik diskritisasi (Unsupervised Discretization) berdampak pada hasil akurasi, hal ini sejalan dengan tujuan penelitian untuk menganalisa hasil kerja penerapan teknik *Unsupervised Discretization* pada metode *Naive Bayes*. adapun grafik hasil pengujian dengan beberapa banyak data sebelumnya dapat dilihat pada gambar 2 berikut :



Gambar 2. Hasil Pengujian Penerapan Unsupervised Discretization pada metode Naive Bayes

4. KESIMPULAN

Penerapan teknik *Unsupervised Discretization* pada metode *Naive Bayes* yang melibatkan beberapa kriteria seperti data nilai mata pelajaran eksakta, nilai mata pelajaran non-eksakta, rekomendasi dari wali kelas, dan angket dari siswa. Dari keempat kriteria tersebut, terdapat 2 kriteria numerik (Nilai mata pelajaran eksakta dan nilai mata pelajaran non-eksakta) dan 2 kriteria kategorikal (angket dan rekomendasi). 120 Data siswa berdasarkan kriteria tersebut telah diproses lalu dipisahkan antara kriteria kategorikal dan kriteria numerik. Untuk meningkatkan akurasi dari metode naive bayes, maka dilakukan proses transformasi, di mana kriteria dengan jenis numerik atau kontinyu diubah menjadi kriteria kategorikal dengan memanfaatkan teknik diskritisasi (*Unsupervised Discretization*). Proses ini ternyata mampu mengoptimalkan penerapan metode naive bayes. Hasilnya, terlihat pada penelitian sebelumnya, hasil akurasi dari penerapan metode Naive Bayes tidak begitu optimal meskipun cukup baik. Pada penelitian ini terbukti penerapan teknik

Unsupervised Discretization pada metode *Naive Bayes* tersebut mampu mengoptimalkan akurasi metode Naive Bayes pada penelitian sebelumnya dengan nilai akurasi 90% meningkat menjadi 92.8%.

5. UCAPAN TERIMA KASIH

diucapkan terima kasih kepada pihak Kementerian Riset Teknologi Dan Pendidikan Tinggi Republik Indonesia (KEMENRISTEKDIKTI) yang telah membantu penuh penelitian ini.

DAFTAR PUSTAKA

- ANDERSON, M. P., & DUBNICKA, S. R., 2014. A sequential naive Bayes classifier for DNA barcodes. *Statistical applications in genetics and molecular biology*, 13(4), 423-434.
- JIANG, L., LI, C., WANG, S., & ZHANG, L., 2016. Deep feature weighting for naive Bayes and its application to text classification. *Engineering Applications of Artificial Intelligence*, 52, 26-39.
- JOITA, D., 2010. Unsupervised static discretization methods in data mining. Titu Maiorescu University, Bucharest, Romania.
- KAREEM, I. A., & DUAIMI, M. G., 2014. Improved accuracy for decision tree algorithm based on unsupervised discretization. *International Journal of Computer Science and Mobile Computing*, 3(6), 176-183.
- KOTSIANTIS, S., 2014. Integrating Global and Local Application of Naive Bayes Classifier. *Inter-national Arab Journal of Information Technology* 11, 300-307.
- MUHAMAD, H., PRASOJO, C. A., SUGIANTO, N. A., SURTININGSIH, L., & CHOLISSODIN, I., 2017. Optimasi Naive Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 4(3), 180-184.
- PALANIAPPAN, S., & HONG, T. K., 2008. Discretization of continuous valued dimensions in OLAP data cubes. *International Journal of Computer Science and Network Security*, 8(11), 116-126.
- PRAWIRA, T. Y., HAKIM, D. K., 2011. Sistem Pendukung Keputusan berbasis Web untuk Menentukan Penjurusan (IPA/IPS/Bahasa) pada SMA Islam Bumiayu, JUITA ISSN : 2086-9398 Vol. 1 Nomor 4, November 2011.

- PUTRANTO, E. A., & SALAMAH, U,. 2016. Sistem Pendukung Keputusan Penjurusan Siswa Kelas X SMA Negeri 2 dengan Metode Fuzzy C-Means dengan Penggunaan Daya Dukung Minat. *ITSMART: Jurnal Teknologi dan Informasi*, 1(2), 68-73.
- SALEH, A,. 2014. Klasifikasi Metode Naive Bayes Dalam Data Mining Untuk Menentukan Konsentrasi Siswa (Studi Kasus Di MAS PAB 2 Medan). Konferensi Nasional Pengembangan Teknologi Informasi dan Komunikasi (KeTIK).
- WORANG, S. G., TOEERA, N. K., LAVINIA, S., & TANAAMAH, A. R. 2013. Penerapan Metode 360 Derajat dalam Sistem Pendukung Keputusan Penentuan Jurusan SMA Berbasis Web (Studi Kasus: SMA Negeri 1 Salatiga). In Seminar Nasional Aplikasi Teknologi Informasi (SNATI) (Vol. 1, No. 1).
- ZHANG, Y., WANG, S., PHILLIPS, P., JI, G., 2014. Binary PSO with mutation operator for feature se-lection using decision tree applied to spam detection. *Knowledge-Based Systems* 64, 22–31.
- ZHOU, X., WANG, S., XU, W., JI, G., PHILLIPS, P., SUN, P., & ZHANG, Y,. 2015. Detection of pathological brain in MRI scanning based on wavelet-entropy and naive Bayes classifier. In *International Conference on Bioinformatics and Biomedical Engineering* (pp. 201-209). Springer, Cham.

Halaman ini sengaja dikosongkan