

PENERAPAN EQUAL-WIDTH INTERVAL DISCRETIZATION DALAM METODE NAIVE BAYES UNTUK MENINGKATKAN AKURASI PREDIKSI PEMILIHAN JURUSAN SISWA (STUDI KASUS: MAS PAB 2 HELVETIA, MEDAN)

IMPLEMENTATION OF EQUAL-WIDTH INTERVAL DISCRETIZATION IN NAIVE BAYES METHOD FOR INCREASING ACCURACY OF STUDENTS' MAJORS PREDICTION (CASE STUDY : MAS PAB 2 HELVETIA, MEDAN)

Alfa Saleh¹ dan Fina Nasari²

^{1,2} Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, Medan
Jl. K.L. Yos Sudarso KM 6.5 No.3-A Tanjung Mulia, Medan, 20241
E-mail: alfasoleh1@gmail.com¹, fina@potensi-utama.ac.id²

Naskah diterima 18 Agustus 2018, direvisi tanggal 27 September 2018, disetujui pada tanggal 12 Desember 2018

Abstract

The student majors is very important thing in developing students' academic skills and talents, it is required by the students are expected to hone the academic ability according to the field that is mastered and this is done so that each student can learn more in the Subjects that match the concentration that has been determined for each - students based on some predefined criteria. In this research has been tested by the method of Naive Bayes which aims to classify the students department based on the criteria that support. Where it is currently conducted with a case study on Madrasah Aliyah PAB 6 Helvetia students and obtained results from 100 student data with 90% accuracy rate. However, in order to improve the accuracy of the results of calcification, the researcher, the method used by using Unsupervised Discretization techniques that will transform numerical / continuous criteria into a categorical criterion. The result of the discretization on 100 data have been tested, it is proved that the results of the techniques used Discontented Disputes on the method of Naive Bayes rose from 90% to 93%.

Keywords : Data Mining, Naive Bayes, Equal-Width Interval Discretization, student major prediction

Abstrak

Jurusan siswa merupakan hal yang sangat penting dalam mengembangkan keterampilan dan bakat akademik siswa, hal ini dianggap perlu sebab siswa diharapkan mampu mengasah kemampuan akademis sesuai bidang yang dikuasai dan hal ini dilakukan agar setiap siswa dapat belajar lebih dalam pada mata pelajaran yang sesuai dengan konsentrasi yang telah ditentukan untuk masing-masing siswa berdasar beberapa kriteria yang telah ditetapkan. Pada penelitian ini telah dilakukan pengujian dengan metode Naive Bayes yang bertujuan untuk mengklasifikasikan jurusan siswa berdasarkan kriteria yang menunjang. pada penelitian ini dilakukan dengan studi kasus pada siswa Madrasah Aliyah Swasta PAB 6 Helvetia dan didapatkan hasil pengujian dari 100 data siswa dengan tingkat keakuratan 90%. Namun, untuk meningkatkan akurasi hasil kalsifikasi penentuan jurusan siswa ini, peneliti mengembangkan metode yang digunakan sebelumnya dengan menerapkan teknik Unsupervised Discretization yang akan mentransformasikan kriteria numerik/kontinyu menjadi kriteria kategorikal. Hasil dari diskritasi pada 100 data siswa yang diuji, terbukti bahwa hasil klasifikasi penerapan teknik Unsupervised Discretization pada metode Naive Bayes naik dari 90% menjadi 93%.

Kata Kunci : Data Mining, Naive Bayes, Equal-Width Interval Discretization, Memprediksi Jurusan Siswa

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. Hal ini 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 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, tempat peneliti melakukan penelitian tidak sepenuhnya efektif dikarenakan para siswa diberikan angket untuk menentukan jurusan apa yang mereka minati tanpa mempertimbangkan kriteria lain yang mungkin memiliki andil untuk menentukan kelayakan siswa tersebut dalam hal memilih jurusan. Padahal proses penentuan jurusan merupakan langkah penting dalam mempersiapkan siswa untuk berkonsentrasi pada bidang yang diminati saat melanjutkan pada jenjang pendidikan selanjutnya. Penelitian ini juga telah menjadi bahan penelitian untuk kategori sistem pendukung keputusan dalam menentukan jurusan di SMA sesuai dengan kemampuan siswa berdasarkan nilai semester, nilai potensi dan nilai pilihan siswa (Prawira & Hakim, 2011).

Penelitian lainnya seputar pemilihan jurusan juga peneliti temukan, di mana dalam proses pemilihan jurusan ini diimplementasikan metode 360 derajat untuk menentukan kesesuaian antara jurusan dan kriteria lain yang mendukung (Worang dkk, 2013). Pada penelitian sebelumnya, peneliti juga telah 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

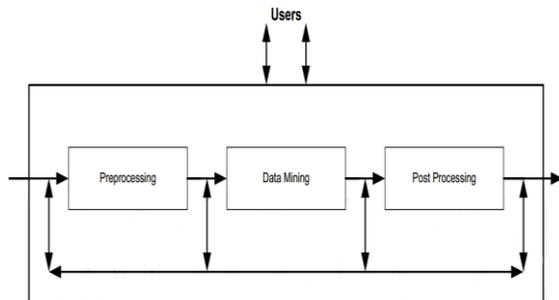
rata-rata mata pelajaran ilmu pengetahuan alam, nilai rata-rata ilmu pengetahuan sosial, rekomendasi guru kelas serta nilai angket yang diisi oleh siswa bersangkutan. Dari 100 data yang diuji menggunakan metode Naive Bayes, maka didapatkan nilai akurasi penentuan jurusan siswa sebesar 90% dengan error sebesar 10% (Saleh, 2014).

Metode Naive Bayes dipilih karena banyak diimplementasikan dalam berbagai bidang ilmu, seperti pada penelitian Xingxing Zhou (2016), Naive Bayes digunakan untuk mengklasifikasikan citra untuk meningkatkan akurasi diagnosis otak dengan menggunakan citra NMR, didapatkan hasil klasifikasi sensitivitas sebesar 94.5%, spesifisitas sebesar 91.70% dan keakurasiian keseluruhan sebesar 92.60 (Xingxing dkk,2015). Naive Bayes menjadi salah satu dari sepuluh 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 meningkat performa metode Naive Bayes sehingga persentase hasil keakuratan prediksi dapat meningkat dibandingkan sebelumnya. Penggunaan teknik *Unsupervised Discretization* dalam mentransformasikan kriteria/atribut numerik sangat baik sehingga pada penelitian ini metode tersebut akan diimplementasikan guna mendukung proses keoptimalisasian metode *Naive Bayes* (Kareem & Duaimi,2016).

Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data

berukuran besar (Mujib dkk, 2013). KDD dikarakteristikan sebagai proses yang terdiri dari beberapa tahap operasional : *Preprocessing*, *Data Mining* dan *Post Processing*. Berikut gambaran dari karakteristik KDD (Silvia dan Germano, 2009) pada gambar 1.



Gambar 1. Karakteristik KDD

a. Tahap *Preprocessing*.

Pada tahap *Preprocessing* memahami fungsi–fungsi yang berhubungan dengan penerimaan, organisasi dan untuk perawatan data, tahap ini sebagai tujuan penyusunan data untuk mengikuti tahap dalam *Data Mining*.

b. Tahap *Data Mining*.

Pada tahap *Data Mining* mendefinisikan teknik–teknik dan algoritma–algoritma yang digunakan oleh masalah di dalam pertanyaan, sebagai contoh dari teknik yang bisa digunakan dalam tahap ini seperti *Neural Network*, *Rough Set*, *Genetic Algorithms*, Statistik dan Probabilistik. Pemilihan teknik tergantung, dalam banyak hal, pada jenis tugas yang harus dikembangkan.

c. Tahap *Post Processing*.

Dalam tahap *Post Processing* pemeliharaan dari pengetahuan diperoleh selama tahap *Data Mining*. tahap ini tidak selalu dibutuhkan namun hal itu memungkinkan validasi kegunaan dari pengetahuan yang ditemukan.

Data Mining

Data Mining merupakan proses pengestraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan

algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data (Shyara dan Saroj, 2013).

Pengelompokan *Data Mining*

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan secara umum yaitu (Maclennan, Tang, & Crivat, 2011, p6) :

a. *Classification*

Fungsi dari *Classification* adalah untuk mengklasifikasikan suatu target *class* ke dalam kategori yang dipilih.

b. *Clustering*

Fungsi dari *Clustering* adalah untuk mencari pengelompokan atribut ke dalam segmentasi–segmentasi berdasarkan similaritas.

c. *Association*

Fungsi dari *Association* adalah untuk mencari keterkaitan antara atribut atau *item set*, berdasarkan jumlah item yang muncul dan *rule association* yang ada.

d. *Regression*

Fungsi dari *Regression* hampir mirip dengan klasifikasi. Fungsi dari *regression* adalah bertujuan untuk mencari prediksi dari suatu pola yang ada.

e. *Forecasting*

Fungsi dari *Forecasting* adalah untuk peramalan waktu yang akan datang berdasarkan trend yang telah terjadi di waktu sebelumnya.

f. *Sequence Analysis*

Fungsi dari *Sequence Analysis* adalah untuk mencari pola urutan dari rangkaian kejadian.

g. *Deviation Analysis*

Fungsi dari *Deviation Analysis* adalah untuk mencari kejadian langka yang sangat berbeda dari keadaan normal (kejadian abnormal).

Tahap-tahap Data Mining

Sebagai suatu rangkaian proses, *Data Mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*.

Tahap-tahap *Data Mining* adalah sebagai berikut (Mujib dkk, 2013):

a. Pembersihan data (*Data Cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.

b. Integrasi data (*Data Integration*).

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.

c. Seleksi data (*Data Selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

d. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.

e. Proses *Mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan *Data Mining*.

f. Evaluasi pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

g. Presentasi pengetahuan (*Knowledge Presentation*).

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas (Patil dan Sherekar, 2013). *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 (Mujib dkk, 2013). 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 (Pattekari dan Parveen, 2012). Adapun Persamaan dari *Naive Bayes* dapat dilihat pada persamaan 1 (Saleh, 2015):

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

variabel X merupakan Data dengan class yang belum diketahui, H merupakan Hipotesis data merupakan suatu class spesifik, $P(H|X)$ merupakan Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas), $P(H)$ merupakan Probabilitas hipotesis H (prior probabilitas), $P(X|H)$ merupakan Probabilitas X berdasarkan kondisi pada hipotesis H, kemudian $P(X)$ merupakan Probabilitas X.

Proses klasifikasi membutuhkan sejumlah petunjuk untuk menentukan kelas mana yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naive Bayes* pada persamaan 1 disesuaikan sebagai berikut :

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

Variabel C merepresentasikan kelas, sementara itu variabel $F_1 \dots F_n$ 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 persamaan 2 dapat pula ditulis secara sederhana sebagai berikut :

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

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, variabel k adalah parameter yang disediakan oleh pengguna. Prosesnya melibatkan pemilahan nilai-nilai yang teramati dari fitur/atribut kontinyu 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, variabel k adalah parameter yang diberikan oleh pengguna.

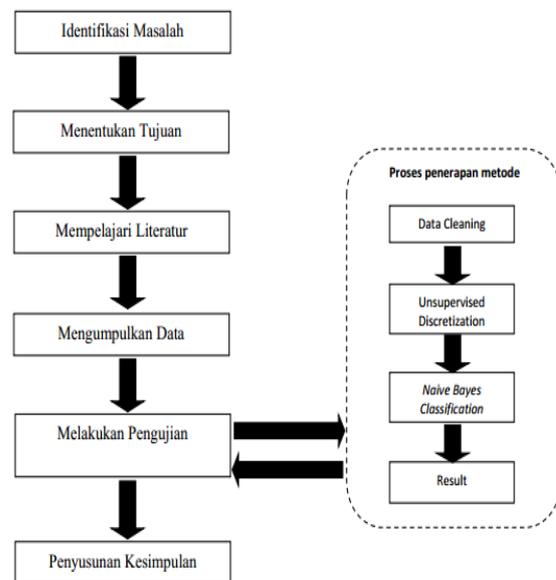
$$Interval = \frac{V_{max} - V_{min}}{k} \quad (4)$$

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

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.

METODE PENELITIAN

Dalam metode penelitian ini terdapat beberapa tahapan, berikut skema dalam melaksanakan penelitian dapat dilihat pada gambar 2 :



Gambar 2. Tahapan Penelitian

Berdasarkan gambar 2, berikut penjelasan dari tahapan penelitian tersebut :

1. Mengidentifikasi Masalah

Pada tahap ini dilakukan peninjauan untuk mengamati serta melakukan eksplorasi lebih dalam dan menggali permasalahan yang ada pada sistem yang berjalan saat ini. Tahap ini adalah langkah awal untuk menentukan rumusan masalah dari penelitian.

2. Menentukan Tujuan

Berdasarkan pemahaman dari permasalahan yang ada, langkah berikutnya adalah menentukan tujuan yang akan dicapai dalam penelitian ini. adapun tujuan penelitian ini adalah menerapkan *Unsupervised Discretization* pada metode Naive Bayes yang diharapkan mampu memprediksi jurusan siswa secara lebih akurat.

3. Mempelajari Literatur

Mempelajari literatur-literatur yang akan dipakai sebagai bahan referensi dalam penelitian ini. Adapun literatur yang dipakai adalah dari jurnal – jurnal ilmiah, modul pembelajaran dan buku tentang Data Mining khususnya yang berkaitan dengan *Unsupervised Discretization* dan dalam metode Naïve Bayes. Literatur – literatur ini akan menjadi pedoman untuk melakukan penelitian agar membantu dan memudahkan proses penelitian.

4. Mengumpulkan Data

Metode pengumpulan data dilakukan secara kolektif dengan meminta data hasil belajar siswa pada semester 1 dan 2 pada kelas XI serta kelas XII, menyebarkan Angket kepada siswa serta mendata hasil rekomendasi guru kelas. Data-data tersebut akan dijadikan acuan dalam menentukan variabel/kriteria yang mendukung.

5. Melakukan Pengujian

Pada tahap ini data yang telah dikumpulkan akan dianalisis. Analisis yang dilakukan adalah dengan mencari hubungan antara setiap variabel/kriteria yang mendukung dalam penentuan jurusan siswa menggunakan metode Naïve Bayes dan *Unsupervised Discretization* dengan tujuan untuk memberikan rekomendasi jurusan yang sesuai untuk siswa berdasarkan hasil belajar (nilai akademik) dan kriteria lain yang mendukung.

6. Membuat Laporan

Tahapan ini merupakan tahapan terakhir, menyusun laporan yang berkaitan dengan penentuan jurusan siswa di Madrasah Aliyah Swasta PAB 2 Helvetia.

HASIL DAN PEMBAHASAN

Adapun kriteria yang digunakan berdasarkan data yang telah dikumpulkan adalah seperti pada tabel 1. 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 angkat. Dua (2) di antaranya merupakan kriteria numerik atau kontinyu dan dua (2) kriteria kategorikal. Untuk meningkatkan tingkat akurasi metode naive bayes, maka dilakukan proses diskritasi menggunakan teknik *unsupervised discretization* pada kriteria yang bernilai numerik atau kontinyu menggunakan persamaan 4 dan 5. Hasil diskritasi kriteria numerik atau kontinyu terdapat pada tabel 2.

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

Tabel 2. Hasil Diskritasi Kriteria Numerik/Kontinyu

Kriteria Numerik/Kontinyu	
Nilai Rata-Rata Mata Pelajaran Eksakta	Nilai Rata-Rata Mata Pelajaran Non-Eksakta
<71.9125	<71.875
71.9125-73.825	71.875-73.75
73.825-75.7375	73.75-75.625
75.7375-77.65	75.625-77.5
77.65-79.5625	77.5-79.375
79.5625-81.475	79.375-81.25
81.475-83.3875	81.25-83.125
83.3875>	83.125>
Nilai k = 8	

Pada tabel 2, dapat dilihat hasil dari proses diskritisasi menggunakan teknik *Unsupervised Discretization*. Kriteria/atribut Nilai rata-rata mata pelajaran eksakta dan non-eksakta yang bertipe numerik atau kontinyu di transformasikan menjadi kriteria kategorikal dengan 8 jenis kategori. Kategori pertama adalah nilai rata-rata mata pelajaran eksakta yang berada di bawah 71.9125, kategori kedua adalah nilai rata-rata mata pelajaran eksakta antara 71.9125-73.825, kategori ketiga adalah

nilai rata-rata mata pelajaran eksakta antara 73.825-75.7375, kategori keempat adalah nilai rata-rata mata pelajaran eksakta antara 75.7375-77.65, kategori kelima adalah nilai rata-rata mata pelajaran eksakta antara 77.65-79.5625, kategori keenam adalah nilai rata-rata mata pelajaran eksakta antara 79.5625-81.475, kategori ketujuh adalah nilai rata-rata mata pelajaran eksakta antara 81.475-83.3875, dan kategori kedelapan adalah nilai rata-rata mata pelajaran eksakta di atas 81.475.

Selanjutnya, hasil diskritisasi kriteria nilai rata-rata mata pelajaran non-eksakta juga terbagi atas 8 kategori, Kategori pertama adalah nilai rata-rata mata pelajaran non-eksakta yang berada di bawah 71.875, kategori kedua adalah nilai rata-rata mata pelajaran non-eksakta antara 71.875-73.75, kategori ketiga adalah nilai rata-rata mata pelajaran non-eksakta antara 73.75-75.625, kategori keempat adalah nilai rata-rata mata pelajaran non-eksakta antara 75.625-77.5, kategori kelima adalah nilai rata-rata mata pelajaran non-eksakta antara 77.5-79.375, kategori keenam adalah nilai rata-rata mata pelajaran non-eksakta antara 79.375-81.25, kategori ketujuh adalah nilai rata-rata mata pelajaran non-eksakta antara 81.25-83.125, dan kategori kedelapan adalah nilai rata-rata mata pelajaran non-eksakta di atas 83.125.

Selanjutnya, menentukan nilai probabilitas setiap kriteria. Berikut nilai probabilitas kriteria nilai rata-rata mata pelajaran eksakta dapat dilihat pada tabel 3. Dapat dilihat pada tabel 3, terdapat sebanyak 51 siswa ditempatkan pada konsentrasi IPA dan 49 siswa lagi ditempatkan pada konsentrasi IPS. Berdasarkan data tersebut terdapat 3 siswa dengan nilai rata-rata mata pelajaran eksakta di bawah 71.9125 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.059, 1 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 71.9125-73.825 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.020, 12 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 73.825-75.7375

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

Nilai Rata-Rata Mata Pelajaran Eksakta	Probabilitas	
	IPA	IPS
<71.9125	0,059	0,245
71.9125-73.825	0,020	0,122
73.825-75.7375	0,235	0,245
75.7375-77.65	0,020	0,061
77.65-79.5625	0,039	0,020
79.5625-81.475	0,216	0,143
81.475-83.3875	0,137	0,102
83.3875>	0,275	0,061
Jumlah	0,51	0,49

ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.235, 1 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 75.7375-77.65 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.020, 2 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 77.65-79.5625 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.039, 11 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 79.5625-81.475 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.216, 7 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 81.475-83.3875 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.137, 14 siswa dengan nilai rata-rata mata pelajaran eksakta di atas 83.3875 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.275. Sementara itu, terdapat 12 siswa dengan nilai rata-rata mata pelajaran eksakta di bawah 71.9125 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.245, 6 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 71.9125-73.825 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.122, 12 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 73.825-75.7375 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.245, 3 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 75.7375-77.65 ditempatkan pada konsentrasi

IPS dan nilai probabilitasnya sebesar 0.061, 1 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 77.65-79.5625 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.020, 7 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 79.5625-81.475 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.143, 5 siswa dengan nilai rata-rata mata pelajaran eksakta diantara 81.475-83.3875 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.102, 3 siswa dengan nilai rata-rata mata pelajaran eksakta di atas 83.3875 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.061. Selanjutnya untuk Nilai probabilitas pada kriteria nilai rata-rata mata pelajaran non-eksakta dapat dilihat pada tabel 4.

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

Nilai Rata-Rata Mata Pelajaran Non-Eksakta	Probabilitas	
	IPA	IPS
<71.875	0,294	0,061
71.875-73.75	0,137	0,061
73.75-75.625	0,176	0,306
75.625-77.5	0,039	0,020
77.5-79.375	0,000	0,082
79.375-81.25	0,196	0,204
81.25-83.125	0,098	0,143
83.125>	0,059	0,122
Jumlah	0,51	0,49

Dari tabel 4, terdata sebanyak 51 siswa ditempatkan pada konsentrasi IPA dan 49 siswa ditempatkan pada konsentrasi IPS. Berdasarkan data tersebut terdapat 15 siswa dengan nilai rata-rata mata pelajaran non-eksakta di bawah 71.875 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.294, 7 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 71.875-73.75 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.137, 9 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 73.75-75.625 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.176, 2 siswa

dengan nilai rata-rata mata pelajaran non-eksakta diantara 75.625-77.5 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.039, tidak ada siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 77.5-79.375 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0, 10 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 79.375-81.25 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.196, 5 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 81.25-83.125 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.098, 3 siswa dengan nilai rata-rata mata pelajaran non-eksakta di atas 83.125 ditempatkan pada konsentrasi IPA dan nilai probabilitasnya sebesar 0.059. Sementara itu, Berdasarkan data tersebut terdapat 3 siswa dengan nilai rata-rata mata pelajaran non-eksakta di bawah 71.875 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.061, 3 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 71.875-73.75 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.061, 15 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 73.75-75.625 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.306, 1 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 75.625-77.5 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.020, 4 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 77.5-79.375 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.082, 10 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 79.375-81.25 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.204, 7 siswa dengan nilai rata-rata mata pelajaran non-eksakta diantara 81.25-83.125 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.143, 6 siswa dengan nilai rata-rata mata pelajaran non-eksakta di atas 83.125 ditempatkan pada konsentrasi IPS dan nilai probabilitasnya sebesar 0.122. Nilai probabilitas untuk kriteria rekomendasi dapat dilihat pada tabel 5.

Tabel 5. Nilai Probabilitas Kriteria Rekomendasi

Rekomendasi	Probabilitas	
	IPA	IPS
IPA	0,961	0,163
IPS	0,039	0,837
Jumlah	0.51	0.49

Jumlah keseluruhan siswa yang dijadikan data latih adalah 100 siswa yang telah direkomendasikan oleh wali kelas sebelumnya terdata sebanyak 51 siswa ditempatkan pada konsentrasi IPA dan 49 siswa lagi ditempatkan pada konsentrasi IPS. Berdasarkan data tersebut terdapat 49 siswa yang direkomendasikan masuk pada konsentrasi IPA dan ditempatkan pada konsentrasi IPA, sementara ada 2 siswa yang direkomendasikan masuk pada konsentrasi IPS namun ditempatkan pada konsentrasi IPA. Selanjutnya ada 8 siswa yang direkomendasikan masuk pada konsentrasi IPA namun ditempatkan pada konsentrasi IPS sementara itu ada 41 siswa yang direkomendasikan masuk pada konsentrasi IPS dan ditempatkan pada konsentrasi IPS. Dengan demikian bisa dihitung probabilitas siswa yang direkomendasikan masuk pada IPA dan ditempatkan pada konsentrasi IPA adalah sebesar 0.961 sementara probabilitas siswa yang direkomendasikan masuk pada IPS namun ditempatkan pada konsentrasi IPA adalah sebesar 0.039. Sedangkan probabilitas siswa yang direkomendasikan masuk pada konsentrasi IPA namun ditempatkan pada konsentrasi IPS adalah sebesar 0.163. lalu, probabilitas siswa yang direkomendasikan masuk pada konsentrasi IPS dan ditempatkan pada konsentrasi IPS adalah sebesar 0.837. Nilai probabilitas untuk kriteria angket dapat dilihat pada tabel 6.

Tabel 6. Nilai Probabilitas Kriteria Angket

Angket	Probabilitas	
	IPA	IPS
IPA	0,843	0,102
IPS	0,157	0,898
Jumlah	0.51	0.49

Jumlah keseluruhan siswa yang dijadikan data latih adalah 100 siswa yang telah diberikan angket sebelumnya terdata sebanyak 51 siswa ditempatkan pada konsentrasi IPA dan 49 siswa lagi ditempatkan pada konsentrasi IPS. Berdasarkan data tersebut terdapat 43 siswa yang memilih konsentrasi IPA dan ditempatkan pada konsentrasi IPA, sementara ada 8 siswa yang memilih konsentrasi IPS namun ditempatkan pada konsentrasi IPA. Selanjutnya ada 5 siswa yang memilih konsentrasi IPA namun ditempatkan pada konsentrasi IPS sementara itu ada 44 siswa yang memilih konsentrasi IPS dan ditempatkan pada konsentrasi IPS. Dengan demikian bisa dihitung probabilitas siswa yang memilih IPA dan ditempatkan pada konsentrasi IPA adalah sebesar 0.843, probabilitas siswa yang memilih IPS namun ditempatkan pada konsentrasi IPA adalah sebesar 0.157. Sedangkan probabilitas siswa yang memilih konsentrasi IPA namun ditempatkan pada konsentrasi IPS adalah sebesar 0.102 sementara probabilitas siswa yang memilih konsentrasi IPS dan ditempatkan pada konsentrasi IPS adalah sebesar 0.898.

Pengujian

Berikut pengujian dari penerapan *unsupervised discretization* pada metode *Naive Bayes* dengan menggunakan sampel 60 data yang akan dibandingkan dengan hasil penelitian sebelumnya, dapat dilihat pada tabel 7.

Tabel 7. Hasil Pengujian dengan 60 data

No	Weighted Average	TP Rate	FP Rate	Precision
1	Naive Bayes	0.900	0.10	0.900
2	Naive Bayes with Unsupervised Discretization	0.933	0.06	0.933

Dari tabel 7, hasil pengujian pada 60 data sampel, metode *Naive Bayes* berhasil mengklasifikasikan 54 data siswa dengan nilai presisi sebesar 0.90 atau persentasenya sebesar 90%, sementara hasil pengujian dari penerapan metode *Naive Bayes* yang dikombinasikan dengan teknik *unsupervised*

discretization berhasil mengklasifikasikan 56 data dengan nilai presisi sebesar 0.933 atau persentase keakuratan nya sebesar 93.3%. Sementara itu, pengujian juga dilakukan dengan 80 data, hasil pengujian terdapat pada tabel 8.

Tabel 8. Hasil Pengujian dengan 80 data

No	Weighted Average	TP Rate	FP Rate	Precision
1	Naïve Bayes	0.913	0.087	0.913
2	Naïve Bayes with Unsupervised Descretization	0.938	0.062	0.938

Hasil pengujian dari 80 data sampel, metode *Naive Bayes* berhasil mengklasifikasikan 73 data dengan nilai presisi sebesar 0.913 atau persentase keakuratan nya sebesar 91.3%, sementara hasil pengujian penerapan metode *naive bayes* yang dikombinasikan dengan teknik *unsupervised discretization* berhasil mengklasifikasikan 75 data dengan nilai presisi sebesar 0.938 atau persentase keakuratan nya sebesar 93.8%. pengujian selanjutnya dilakukan dengan 100 data, hasil pengujian terdapat pada tabel 9.

Tabel 9. Hasil Pengujian dengan 100 data

No	Weighted Average	TP Rate	FP Rate	Precision
1	Naïve Bayes	0.920	0.080	0.920
2	Naïve Bayes with Unsupervised Descretization	0.930	0.070	0.930

Hasil pengujian pada 100 data sampel, penerapan metode *naive bayes* berhasil mengklasifikasikan 92 data dengan nilai presisi sebesar 0.92 atau persentase keakuratan nya sebesar 92%, sementara hasil pengujian penerapan metode *naive bayes* yang dikombinasikan dengan teknik *unsupervised discretization* berhasil mengklasifikasikan 93 data dengan nilai presisi sebesar 0.93 atau persentase keakuratan nya sebesar 93%.

Beberapa pengujian yang telah dilakukan dengan jumlah data yang bervariasi berdasarkan nilai probabilitas setiap kriteria yang telah ditentukan sebelumnya, bertujuan untuk menganalisa hasil kerja penerapan teknik

Unsupervised Discretization pada metode *Naive Bayes*.

PENUTUP

Simpulan

Kesimpulan yang dapat dirangkum pada penelitian ini adalah penerapan *Unsupervised Discretization* pada metode *naive bayes* cukup berdampak pada hasil pengujian, sementara itu, kriteria yang digunakan untuk pengujian ini antara lain : data nilai rata-rata mata kuliah eksakta, data nilai rata-rata mata kuliah non eksakta, data rekomendasi dan data angket siswa. hasil akurasi metode *Naive Bayes* tanpa penerapan *Unsupervised Discretization* adalah sebesar 90% untuk 60 data sampel, 91.3% untuk 80 data sampel dan 92% untuk 100 data sampel. sementara hasil pengujian metode *Naive Bayes* dengan penerapan *Unsupervised Discretization* adalah sebesar 93.3% untuk 60 data sampel, 93.8% untuk 80 data sampel dan 93% untuk 100 data sampel. Dengan begitu, Terbukti penerapan *Unsupervised Discretization* pada metode *Naive Bayes* lebih baik dibandingkan tanpa penerapan *Unsupervised Discretization*.

Saran

Adapun saran yang dapat diuraikan untuk perbaikan pada penelitian ini adalah sebaiknya ada pengujian menggunakan metode data mining lain yang relevan untuk mendapatkan hasil akurasi yang lebih efektif serta diharapkan adanya penambahan variabel sebagai kriteria yang akan membantu dalam meningkatkan hasil akurasi metode sehingga dapat diterapkan dalam sebuah sistem yang mampu membantu pengambil keputusan dalam menentukan jurusan siswa.

Ucapan Terima Kasih

Peneliti mengucapkan banyak terima kasih kepada pihak Kementerian Riset Teknologi Dan Pendidikan Tinggi Republik Indonesia (Kemenristekdikti) yang telah membantu penelitian ini secara moril dan finansial.

DAFTAR PUSTAKA

- Prawira, T. Y., & Hakim, D. K. (2011). *Sistem Pendukung Keputusan Berbasis Web untuk Menentukan Penjurusan (IPA/IPS/Bahasa) pada SMA Islam Bumiayu*. JUITA: Jurnal Informatika, 1(4).
- Worang, S. G., Toera, 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).
- 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)
- Zhou, X., Wang, S., Xu, W., Ji, G., Phillips, P., Sun, P., & ZHANG, Y. (2015, April). *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.
- 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.
- Anderson, M.P., & Dubnicka, S.R. (2014). *A sequential naive Bayes classifier for DNA barcodes*. Statistical Applications in Genetics and Molecular Biology 13, 423-434.
- Zhang, Y., Wang, S., & Phillips, P., Ji, G.(2014). *Binary PSO with mutation operator for feature selection using decision tree applied to spam detection*. Knowledge-Based Systems 64, 22-31.
- Kotsiantis, S.(2014). *Integrating Global and Local Application of Naive Bayes Classifier*. International Arab Journal of Information Technology 11, 300-307.
- 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
- 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.
- Rissino, S., & Lambert-Torres, G. (2009). *Rough set theory—fundamental concepts, principals, data extraction, and applications*. In Data mining and knowledge discovery in real life applications. InTech.
- Shyara taruna R, & Saroj Hiranwal, (2013). *Enhanced Naive Bayes Algorithm for Intrusion Detection in Data Mining*. International Journal of Computer Science and information Technologies, Vol. 4.
- MacLennan, J., Tang, Z., & Crivat, B. (2011). *Data mining with Microsoft SQL server 2008*. John Wiley & Sons.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. jurnal EECCIS, 7(1), 59-64.
- Patil, T. R., & Sherekar, S. S. (2013). *Performance analysis of Naive Bayes and J48 classification algorithm for data classification*. International journal of computer science and applications, 6(2), 256-261.
- Pattekari, S. A., & Parveen, A. (2012). *Prediction system for heart disease using Naive Bayes*. International Journal of Advanced Computer and Mathematical Sciences, 3(3), 290-294.
- Saleh, A. (2015). *Implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga*. Creative Information Technology Journal, 2(3), 207-217.
- Joița, D. (2010). *Unsupervised static discretization methods in data mining*. Titu Maiorescu University, Bucharest, Roman

