

Submission author:
Achmad Solichin

Check ID:
25776563

Check date:
25.07.2020 21:24:07 WIB

Check type:
Doc vs Internet

Report date:
25.07.2020 21:47:41 WIB

User ID:
85716

File name: **39-2020-Penerapan Metode Naive Bayes Classifier**

File ID: **30621054** Page count: **6** Word count: **3719** Character count: **24570** File size: **550.36 KB**

10% Matches

Highest match: **5.27%** with source <http://achmatim.net/2020/05/02/paper-penerapan-metode-naive-bayes-classifier-untuk-penju>.

10% Internet Matches 25

Page 8

No Library Sources Found

0% Quotes

No quotes found

89.9% Exclusions

Sources less than **8 words** were automatically excluded

89.9% Internet exclusions 2

Page 9

No library exclusions found

Replacement

No replaced characters found

Penerapan Metode Naive Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta

Ahmad Zainul Mafakhir ¹⁾, Achmad Solichin ^{2)*}

Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta
aahmad.ypi@gmail.com ¹⁾, achmad.solichin@budiluhur.ac.id ^{2)*}

Abstrak

Penjurusan siswa di suatu sekolah merupakan suatu proses penempatan siswa ke dalam jurusan tertentu sesuai dengan kemampuan dan keinginan siswa. Madrasah Aliyah Al-Falah (MA Al-Falah) merupakan sekolah yang sederajat dengan Sekolah Menengah Atas (SMA) dan memiliki tiga jurusan yaitu Ilmu-ilmu Keagamaan (IHK), Matematika dan Ilmu Alam (MIA), dan Ilmu-ilmu Sosial (IIS). Saat ini proses penjurusan di MA Al-Falah dilakukan dengan melakukan empat jenis tes yaitu matematika, bahasa Indonesia, bahasa Inggris, dan keagamaan terhadap siswa. Hasil tes tersebut dijadikan dasar penentuan jurusan siswa. Permasalahan yang dihadapi oleh pihak sekolah adalah kesulitan dalam mengklasifikasikan siswa berdasarkan hasil tes penjurusan tersebut. Saat ini teknik penjurusan yang dilakukan sekolah hanya mengandalkan pengolahan data dan pengurutan dengan Microsoft Excel. Pada penelitian ini, dilakukan penerapan metode Naive Bayes Classifier untuk melakukan klasifikasi nilai tes dan menghasilkan rekomendasi jurusan siswa. Sistem klasifikasi penjurusan siswa yang dikembangkan dapat membantu proses penjurusan siswa dengan lebih mudah, cepat, dan akurat. Berdasarkan pengujian yang telah dilakukan sistem penjurusan siswa dapat memberikan rekomendasi jurusan siswa dengan nilai akurasi sebesar 33,34%.

Kata kunci: penjurusan siswa, naive bayes, sistem klasifikasi

Abstract

[The Implementation of Naive Bayes Classifier Method for Student Majorization at Madrasah Aliyah Al-Falah Jakarta] Student majorization in a school is a process of placing students into certain majors according to their abilities and desires. Madrasah Aliyah Al-Falah (MA Al-Falah) is a school equivalent to High School and has three majors namely Religious Sciences (IHK), Mathematics and Natural Sciences (MIA), and Social Sciences (IIS). Currently the majors at MA Al-Falah are carried out by carrying out four types of tests namely mathematics, Indonesian, English, and religion on students. The test results are used as the basis for determining student majors. The problem faced by the school is the difficulty in classifying students based on the results of the majors test. Currently the majors that schools use rely solely on data processing and sequencing with Microsoft Excel. In this study, the Naive Bayes Classifier method was implemented to classify test scores and produce recommendations for student majors. The student majors classification system developed can help students' majors process more easily, quickly, and accurately. Based on the testing that has been done the student majors system can provide student majors recommendations with an accuracy value of 33,34%.

Keywords: student majors, naive bayes, classification system

1. PENDAHULUAN

Keberadaan madrasah sebagai lembaga pendidikan Islam di Indonesia muncul sebagai kelanjutan logis lembaga pendidikan Islam sebelumnya. Di Indonesia pada saat ini terdapat pembagian beberapa tingkatan madrasah yaitu Raudhatul Athfal (RA), Madrasah Ibtidaiyah (MI), Madrasah Tsanawiyah (MTs), Madrasah Aliyah/Kejuruan (MA/MAK). Pembagian tersebut sudah disederajatkan oleh pemerintah dengan Taman Kanak-Kanak (TK), Sekolah Dasar (SD), Sekolah

Menengah Pertama (SMP), Sekolah Menengah Atas/Kejuruan (SMA/K).

Madrasah Aliyah Al-Falah (MA Al-Falah) merupakan salah satu sekolah yang berbasis Islam dalam proses belajar dan mengajarnya. MA Al-Falah saat ini memiliki pilihan jurusan/peminatan, yaitu Ilmu-ilmu Keagamaan (IHK), Matematika Ilmu Alam (MIA), dan Ilmu Ilmu Sosial (IIS). Terdapat tiga jurusan yang harus siswa-siswi Madrasah Aliyah (MA) Al-Falah pilih salah satunya dengan mekanisme yang sudah ditentukan pihak sekolah seperti tes tulis, wawancara, dan lain-lain. Hal tersebut yang nantinya

akan menjadi parameter menentukan klasifikasi siswa-siswi akan jurusan/peminatannya, walaupun siswa-siswi mempunyai ketertarikan atau keinginannya tersendiri namun pihak sekolah harus menyarankan sesuai dengan kemampuan para siswa-siswi didiknya agar tidak terbawakan oleh arus dan tidak merasa kesulitan di tengah-tengah proses belajar dan mengajar.

Penjurusan siswa merupakan suatu proses penempatan siswa ke dalam jurusan atau peminatan tertentu, sehingga siswa dapat menyerap semua mata pelajaran dengan optimal dan sesuai dari kemampuan yang dimilikinya. Penjurusan dilakukan karena terdapat perbedaan dari setiap individu yang membuat tingkat pemahaman setiap siswa/siswi dalam penerimaan informasi berbedabeda. Sulitnya dalam mengklasifikasikan jurusan siswa-siswi menjadi kendala bagi pihak sekolah dalam menghitung setiap kriteria para siswa-siswi dan menjadikan waktu yang kurang efisien karena harus menghitung satu per satu nilai tes yang sudah dilaksanakan sebelumnya.

Berdasarkan masalah yang dihadapi oleh pihak sekolah maka penelitian ini bertujuan untuk membuat suatu sistem klasifikasi penjurusan bagi setiap siswa yang akan mempermudah pihak sekolah dan juga mengefisienkan waktu pengklasifikasian. Pada penelitian ini diterapkan metode Naive Bayes untuk mengklasifikasikan data siswa ke dalam jurusan yang tepat. Metode *Naive Bayes Classifier* adalah salah satu algoritme pengklasifikasi dengan mencari nilai probabilitas dari setiap atribut.

Data mining adalah proses menemukan hubungan baru yang menemukan sebuah arti, pola dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan pengenalan pola seperti teknik statistik dan matematika. *Data mining* merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan analisis untuk penanganan permasalahan pengambilan informasi dari database [1].

Data mining terbagi beberapa kelompok pembahasan salah satunya adalah klasifikasi (*Classifier*). Dalam klasifikasi, terdapat target variable kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, sebagai berikut : pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Contoh lain klasifikasi dalam dunia bisnis dan penelitian, sebagai berikut :

1. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.
2. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk penyakit apa.
3. Menentukan pemilihan jurusan kelas pada Sekolah Menengah Atas (SMA).

Algoritme *Naive Bayes* merupakan suatu metode pengklasifikasian data dengan model statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*

dan dapat digunakan untuk memprediksi probabilitas keanggotaan pada suatu kelas [2]. Terdapat dua bagian untuk penggunaan algoritme *Naive Bayes Classifier*[3], sebagai berikut :

1. *Naive Bayes* untuk data kategorial.
2. *Naive Bayes* untuk data kontinu.

Algoritme ini membangkitkan suatu fungsi yang memetakan *input* ke *output* yang diinginkan. Kualitas hasil pembelajaran sangat bergantung pada kesesuaian *input* dan *output* yang diberikan untuk menjadi *data training*. Dengan demikian, user sangat berperan penting dalam memvalidasi *input* dan *output* tersebut. Dengan kata lain bahwa yang dinamakan *Supervised Learning* adalah pembelajaran terawasi, *Supervised Learning* digunakan untuk masalah klasifikasi.

Klasifikasi berasal dari bahasa Inggris dari kata "*Classification*" dan kata ini berasal dari kata "*To Class*" yang berarti menggolongkan sebuah data menjadi satu. Klasifikasi adalah suatu proses untuk menemukan sekelompok dari beberapa data ke dalam beberapa kelas dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Beberapa algoritme klasifikasi yang sering digunakan antara lain *Naive Bayes*, *Bayesian Decision*, *k-Nearest Neighbor*, dan *Linear Discriminant Analysis* [4], [5]. Diantara algoritme klasifikasi tersebut, *Naive Bayes* merupakan salah satu metode yang banyak digunakan, antara lain untuk penjurusan siswa [6], rekomendasi jurusan kuliah [4], prediksi kelulusan tugas akhir [5], [7], dan untuk mendiagnosis penyakit gigi [8].

Saat ini sudah cukup banyak penelitian yang memanfaatkan algoritme data mining untuk membantu proses penjurusan siswa di sekolah. Penelitian oleh [6] melakukan penjurusan siswa di SMA Negeri 3 Boyolali. Penjurusan siswa dapat mengarahkan peserta didik agar lebih fokus dalam mengembangkan kemampuan diri dan minat yang dimiliki. Pemilihan jurusan yang tidak tepat dapat merugikan siswa terhadap minat dan karir mereka di masa mendatang. Metode yang digunakan adalah algoritme klasifikasi C4.5, *Naive Bayes*, dan *K-Means*. Dalam melakukan penjurusan, atribut yang digunakan terdiri dari jenis kelamin, minat, rata-rata nilai IPA, rata-rata nilai IPS, nilai psikotest IPA, nilai psikotest IPS, asal sekolah dan jurusan sekolah. Hasil penelitian menunjukkan bahwa algoritme C4.5 memiliki nilai akurasi yang lebih baik dalam melakukan penjurusan siswa dibanding dua metode lainnya. Secara umum, sistem berhasil mengklasifikasikan jurusan siswa berdasarkan kriteria yang diberikan.

Pada penelitian lainnya, diterapkan algoritme *Naive Bayes* yang digabungkan dengan metode *Weighted Product (WP)* untuk memberikan rekomendasi jurusan kuliah bagi lulusan SMA[4]. Sistem yang dikembangkan dapat membantu lulusan SMA dalam memilih jurusan secara tepat, sesuai minat dan sesuai kemampuan yang dimiliki siswa. Berdasarkan hasil pengujian yang dilakukan menunjukkan bahwa akurasi sistem dalam

memberikan rekomendasi sebesar 82%. Optimasi terhadap metode klasifikasi Naive Bayes dilakukan oleh Saleh dan Nasari pada [9] dengan menerapkan teknik *unsupervised discretization*. Teknik *unsupervised discretization* mentransformasikan kriteria bernilai numerik/kontinu menjadi kriteria kategorikal dan mengeliminasi satu kriteria yang dianggap tidak mempengaruhi keakuratan hasil pengujian. Optimasi berhasil dilakukan dengan peningkatan nilai akurasi sebesar 2,8% dibanding tanpa optimasi. Namun demikian, peningkatan tersebut masih kurang signifikan.

2. BAHAN DAN METODE

2.1. Kerangka Pikir Penelitian

Dalam melakukan penelitian, langkah-langkah terstruktur diperlukan agar tujuan penelitian dapat tercapai secara baik. Tahapan penelitian menjadi referensi dalam melakukan penelitian. Gambar 1 menunjukkan kerangka pikir penelitian. Berdasarkan masalah penelitian yang dijelaskan pada bagian sebelumnya, dilakukan studi literatur berbagai penelitian sebelumnya yang terkait dengan penerapan algoritma klasifikasi dalam penjurusan siswa. Setelah melakukan studi literatur, dipilih algoritma klasifikasi Naive Bayes karena dinilai paling sesuai untuk kasus penjurusan siswa di MA Al-Falah.

Dengan memanfaatkan data siswa kelas X tahun ajaran 2015/2016 dan tahun ajaran 2018/2019, selanjutnya dikembangkan sebuah prototype aplikasi penjurusan siswa yang menerapkan algoritma Naive Bayes. Selanjutnya, untuk mengetahui performa penerapan metode Naive Bayes, dilakukan pengujian baik terhadap hasil klasifikasinya maupun terhadap fungsionalitas aplikasi yang dihasilkan.



Gambar 1. Kerangka pikir penelitian

2.2. Dataset Penelitian

Pengumpulan data dilakukan dengan mengumpulkan data jurusan siswa-siswi kelas X (sepuluh) Madrasah Aliyah Al-Falah Jakarta tahun ajaran 2015/2016 dan 2018/2019 dengan jumlah total data yang dikumpulkan sebesar 197 data siswa yang akan menjadi data latih (*training*). Tabel 1 menyajikan statistik deskriptif dari data siswa yang digunakan pada penelitian ini.

Data siswa-siswi yang dikumpulkan merupakan data tes penjurusan yang telah dilaksanakan. Kriteria penjurusan terdiri dari nilai matematika, nilai bahasa Indonesia, nilai bahasa

Inggris, nilai agama, rata-rata UN, dan rata-rata rapor. Data nilai yang didapat pada mulanya berbentuk numerik maka dirubah menjadi nilai dalam bentuk skala seperti terlihat pada Tabel 2.

Tabel 1. Data Siswa

Kelas Siswa	Tahun Ajaran	Jumlah Siswa Per/Kelas
Agama	2015-2016	37
IIS	2015-2016	35
MIA	2015-2016	27
Agama	2018-2019	34
IIS	2018-2019	35
MIA	2018-2019	29
TOTAL		197

Tabel 2. Skala Penilaian

Skala Nilai	Nama Skala
90-100	Sangat Baik (SB)
70-89	Baik (B)
0-69	Cukup (C)

2.3. Perhitungan Naive Bayes

Metode klasifikasi Naive Bayes termasuk metode klasik yang menggunakan konsep statistik probabilitas dalam melakukan klasifikasi data. Naive Bayes menggunakan teorema probabilitas dengan formula dasar menggunakan Persamaan (1).

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

yang mana X adalah bukti dan H adalah hipotesis, $P(H|X)$ adalah probabilitas bahwa hipotesis H benar untuk bukti X atau dengan kata lain $P(H|X)$ merupakan probabilitas posterior H dengan syarat X, $P(H|X)$ adalah probabilitas bahwa bukti X benar untuk hipotesis H atau probabilitas posterior X dengan syarat H, $P(H)$ adalah probabilitas posterior hipotesis H, dan $P(X)$ adalah prior bukti X.

Pada penelitian ini, digunakan dataset nilai tes siswa yang berupa data numerik dalam rentang 0-100. Sementara itu, metode Naive Bayes tidak cocok untuk data yang bersifat numerik atau yang memiliki rentang yang sangat banyak. Oleh karena itu, langkah pertama penerapan metode Naive Bayes adalah perubahan nilai numerik menjadi nilai kategorikal. Nilai yang semula berada pada rentang 0-100 diskalakan menjadi tiga skala yaitu Sangat Baik (SB), Baik (B), dan Cukup (C) seperti disajikan pada Tabel 2.

Tabel 3 menyajikan contoh cuplikan dataset penelitian yang digunakan pada penelitian ini. Berdasarkan aturan skala penilaian pada Tabel 2, nilai-nilai pada Tabel 3 diubah menjadi tiga skala dan ditampilkan pada Tabel 4.

Langkah selanjutnya adalah menghitung probabilitas dari masing-masing kelas. Berdasarkan data latih pada contoh Tabel 4, diperoleh probabilitas setiap jurusan yaitu: IIK = $10/20 = 0,5$; MIA = $4/20 = 0,2$; dan IIS = $6/20 = 0,3$.

Tabel 3. Cuplikasi dataset nilai siswa yang masih dalam bentuk nilai numerik

Siswa	MTK	INDO	ING	AGM	UN	RPT	JUR
S1	73	84	82	80	82,50	84,13	IJK
S2	76	85	79	80	77,95	80,25	IJK
S3	83	75	68	83	81,50	82,17	MIA
S4	70	82	85	79	78,00	81,29	IIS
S5	69	75	92	85	78,50	78,03	IJK
S6	76	81	80	85	83,75	79,55	IJK
S7	71	70	75	80	78,50	82,17	IJK
S8	74	83	70	90	83,50	80,15	IJK
S9	65	75	70	85	76,00	78,65	IJK
S10	73	81	75	72	85,25	80,15	IIS
S11	77	83	75	70	89,25	89,65	IJK
S12	84	90	85	92	80,25	85,35	MIA
S13	80	86	82	85	79,50	79,14	MIA
S14	72	80	70	75	78,00	82,23	IIS
S15	69	82	73	80	84,50	78,56	IIS
S16	71	82	80	92	78,25	80,15	IJK
S17	71	80	75	85	81,75	84,17	IJK
S18	81	79	80	82	80,25	82,26	MIA
S19	69	85	83	80	77,00	76,15	IIS
S20	64	75	73	79	78,50	69,52	IIS

Tabel 4. Nilai numerik diubah menjadi nilai kategorikal dengan aturan skala penilaian.

Siswa	MTK	INDO	ING	AGM	UN	RPT	JUR
S1	B	B	B	B	B	B	IJK
S2	B	B	B	B	B	B	IJK
S3	B	B	C	B	B	B	MIA
S4	B	B	B	B	B	B	IIS
S5	C	B	SB	B	B	B	IJK
S6	B	B	B	B	B	B	IJK
S7	B	B	B	B	B	B	IJK
S8	B	B	B	SB	B	B	IJK
S9	C	B	B	B	B	B	IJK
S10	B	B	B	B	B	B	IIS
S11	B	B	B	B	B	B	IJK
S12	B	SB	B	SB	B	B	MIA
S13	B	B	B	B	B	B	MIA
S14	B	B	B	B	B	B	IIS
S15	C	B	B	B	B	B	IIS
S16	B	B	B	SB	B	B	IJK
S17	B	B	B	B	B	B	IJK
S18	B	B	B	B	B	B	MIA
S19	C	B	B	B	B	B	IIS
S20	C	B	B	B	B	C	IIS

Sebagai ilustrasi proses perhitungan metode Naive Bayes, dicontohkan sebuah data siswa yang akan diberikan rekomendasi jurusannya dengan kriteria sebagai berikut:

S21 = {MTK=B; INDO=B; ING=C; AGM=SB; UN=B; RPT=B} maka perhitungan probabilitas bersyarat untuk setiap kelas sebagai berikut:

- Jurusan MIA:
 - $P(\text{MTK}=\text{B} \mid \text{JUR}=\text{MIA}) = 4/4 = 1,0$
 - $P(\text{INDO}=\text{B} \mid \text{JUR}=\text{MIA}) = 3/4 = 0,75$
 - $P(\text{ING}=\text{C} \mid \text{JUR}=\text{MIA}) = 1/4 = 0,25$
 - $P(\text{AGM}=\text{SB} \mid \text{JUR}=\text{MIA}) = 1/4 = 0,25$
 - $P(\text{UN}=\text{B} \mid \text{JUR}=\text{MIA}) = 4/4 = 1$
 - $P(\text{RPT}=\text{B} \mid \text{JUR}=\text{MIA}) = 4/4 = 1,0$
- Jurusan IIS:
 - $P(\text{MTK}=\text{B} \mid \text{JUR}=\text{IIS}) = 3/6 = 0,5$
 - $P(\text{INDO}=\text{B} \mid \text{JUR}=\text{IIS}) = 6/6 = 1,0$
 - $P(\text{ING}=\text{C} \mid \text{JUR}=\text{IIS}) = 0/6 = 0$
 - $P(\text{AGM}=\text{SB} \mid \text{JUR}=\text{IIS}) = 0/6 = 0$

- $P(\text{UN}=\text{B} \mid \text{JUR}=\text{IIS}) = 6/6 = 1$
- $P(\text{RPT}=\text{B} \mid \text{JUR}=\text{IIS}) = 5/6 = 0,83$

- Jurusan IJK:
 - $P(\text{MTK}=\text{B} \mid \text{JUR}=\text{IJK}) = 8/10 = 0,8$
 - $P(\text{INDO}=\text{B} \mid \text{JUR}=\text{IJK}) = 10/10 = 1,0$
 - $P(\text{ING}=\text{C} \mid \text{JUR}=\text{IJK}) = 0/10 = 0$
 - $P(\text{AGM}=\text{SB} \mid \text{JUR}=\text{IJK}) = 2/10 = 0,2$
 - $P(\text{UN}=\text{B} \mid \text{JUR}=\text{IJK}) = 10/10 = 1,0$
 - $P(\text{RPT}=\text{B} \mid \text{JUR}=\text{IJK}) = 10/10 = 1,0$

Langkah selanjutnya adalah melakukan perhitungan HMAP (*Hypothesis Maximum Appropri Probability*) yang pada dasarnya melakukan perhitungan probabilitas kemunculan data uji terhadap masing-masing kelas. Kesimpulan kelas diperoleh dari nilai terbesar dari probabilitas tersebut. Pada kasus di atas, perhitungan nilai probabilitas kemunculan data uji (S21) terhadap setiap kelas adalah:

$$P(\text{JUR}=\text{"MIA"}) = 1,0 * 0,75 * 0,25 * 0,25 * 1 * 1 * 1 * 0,2 = \mathbf{0,009375}$$

$$P(\text{JUR}=\text{"IIS"}) = 0,5 * 1 * 0 * 0 * 1 * 0,83 * 0,3 = 0$$

$$P(\text{JUR}=\text{"IJK"}) = 0,8 * 1,0 * 0 * 0,2 * 1 * 1 * 1 * 0,5 = 0$$

Berdasarkan perhitungan di atas, nilai probabilitas terbesar diperoleh untuk jurusan "MIA" sehingga siswa direkomendasikan untuk masuk ke jurusan MIA.

2.4. Rancangan Prototype Aplikasi Penjurusan Siswa

Pada penelitian ini dikembangkan sebuah prototype aplikasi penjurusan siswa yang bertujuan untuk melakukan pengujian penerapan metode Naive Bayes dalam memberikan rekomendasi jurusan siswa. Prototype dikembangkan menggunakan bahasa pemrograman PHP dan perangkat pangkalan data MySQL. Keduanya dipilih atas dasar kemudahan dan rendahnya biaya dalam implementasi di MA Al-Falah.



Gambar 2. Rancangan layar halaman login

Gambar 2 menyajikan rancangan layar halaman login ke aplikasi. Pada halaman login, pengguna diminta untuk memasukkan username dan password yang telah terdaftar sebelumnya. Jika login berhasil, maka akan ditampilkan halaman utama aplikasi. Sementara itu, Gambar 3 menunjukkan salah satu fitur utama aplikasi, yaitu proses klasifikasi atau penjurusan siswa. Data siswa yang akan diuji dapat dimasukkan satu persatu atau diunggah dalam bentuk

file. Hasil rekomendasi jurusan ditampilkan pada tabel rekomendasi yang terletak halaman yang sama.



Gambar 3. Rancangan halaman penjurusan siswa (pengujian)

3. HASIL DAN PEMBAHASAN

3.1. Prototype Aplikasi Penjurusan Siswa

Prototype aplikasi penjurusan siswa MA Al-Falah telah dikembangkan menggunakan bahasa pemrograman PHP dan perangkat pangkalan data MySQL. Platform aplikasi dikembangkan berbasis web sehingga dapat diakses dengan mudah melalui peramban yang terkoneksi ke jaringan. Gambar 4 menyajikan tampilan halaman pengelolaan data training yang mana pengguna dapat melihat, mencari, menambah, mengubah, dan menghapus data training.



Gambar 4. Tampilan master data training siswa

Untuk mempermudah pengguna dalam memahami data training, aplikasi juga menyajikan statistik deskriptif dari data seperti terlihat pada Gambar 5. Data disajikan dalam bentuk grafik sehingga pengguna dapat melihat komposisi data

untuk setiap kelas. Sementara itu, Gambar 6 menunjukkan contoh hasil rekomendasi jurusan siswa berdasarkan data masukan yang dimiliki oleh siswa. Selain hasil rekomendasi jurusan, perhitungan klasifikasi juga disajikan untuk menjamin validitas proses perhitungan.



Gambar 5. Halaman grafik statistik deskriptif



Gambar 6. Halaman rekomendasi jurusan siswa

3.2. Pengujian Performa Algoritme

Pada penelitian ini dilakukan pengujian kualitas algoritme klasifikasi Naïve Bayes dalam memberikan rekomendasi jurusan siswa di MA Al-Falah. Data uji yang digunakan sebanyak 39 data atau 20% dari total data latih. Adapun ukuran pengujian untuk mengetahui performa algoritme adalah akurasi, presisi, dan recall yang dihitung dengan tabel confusion matrix [10], [11]. Hasil tabel confusion matrix ditunjukkan pada Tabel 5.

Tabel 5. Hasil pengujian performa algoritme

A C T U A L	PREDICTION		
	KELAS	MIA	IIS
MIA	0	11	2
IIS	0	7	6
IIK	3	4	6

Berdasarkan Tabel 5, selanjutnya dihitung nilai akurasi, presisi, dan recall seperti disajikan pada Tabel 6. Secara keseluruhan nilai akurasi algoritme klasifikasi Naïve Bayes adalah 33,34, nilai presisi sebesar 21,18 dan nilai recall sebesar 33,77.

Tabel 6. Hasil perhitungan nilai akurasi, presisi, dan recall

Ukuran	MIA	IIS	IIK	Rata-rata
Akurasi	33,34	33,34	33,34	33,34
Presisi	0,00	31,82	18,18	21,18
Recall	0,00	53,85	30,77	33,77

Berdasarkan hasil pengujian yang telah dilakukan, menunjukkan bahwa performa algoritme klasifikasi Naive Bayes dalam mengklasifikasikan jurusan siswa tidak terlalu baik. Secara keseluruhan, nilai akurasinya hanya sekitar 33,34. Berdasarkan analisis peneliti, buruknya nilai akurasi disebabkan karena proses pengubahan nilai numerik menjadi nilai kategorikal. Nilai asli yang berada dalam rentang 0-100 disederhanakan menjadi tiga nilai saja yaitu Sangat Baik, Baik, dan Cukup. Penyederhanaan nilai menyebabkan terdapat detail informasi yang terkandung di dalam data menjadi hilang. Oleh karena itu, dalam penelitian selanjutnya sebaiknya proses penyederhanaan tersebut tidak dilakukan sehingga diperoleh nilai performa yang lebih baik.

4. KESIMPULAN

Berdasarkan hasil pengujian dan uji coba terhadap metode dan pratype yang telah dikembangkan diperoleh kesimpulan bahwa metode Naive Bayes dapat digunakan untuk mengklasifikasikan jurusan siswa. Namun demikian, proses pengubahan nilai numerik menjadi nilai kategorikal menyebabkan hasil akurasi, presisi, dan recall menjadi rendah. Penyederhanaan menyebabkan detail informasi menjadi hilang. Pada penelitian selanjutnya, peningkatan performa algoritme Naive Bayes perlu dilakukan antara lain dengan menerapkan teknik Unsupervised Discretization seperti diusulkan dalam [9] atau menerapkan metode Fuzzy [12] sehingga proses kategorisasi menjadi lebih akurat.

5. DAFTAR PUSTAKA

- [1] Y. Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [2] S. Winanta, Y. Oslan, and G. Santoso, "Implementasi Metode Bayesian Dalam Penjurusan Di Sma Bruderan Purworejo Studi Kasus: Sma Bruderan Purworejo," *Eksis*, vol. 06, no. 2, pp. 21–28, 2013.
- [3] Suyanto, *Machine Learning Tingkat Dasar Dan Lanjut*, 1st ed. Bandung: Informatika Bandung, 2018.
- [4] R. Fitriawanti, I. Cholissodin, and R. K. Dewi, "Klasifikasi dan Rekomendasi Jurusan Kuliah Bagi Pelajar SMA Menggunakan Algoritme Naive Bayes -WP," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 4914–4922, 2018.
- [5] A. Solichin, "Comparison of Decision Tree, Naive Bayes and K- Nearest Neighbors for Predicting Thesis Graduation," in *The 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI 2019)*, 2019.
- [6] Y. S. Nugroho and S. N. Haryati, "Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 1, no. 1, pp. 1–6, 2015.
- [7] D. Himawan, "Aplikasi Data Mining Menggunakan Algoritma ID3 Untuk Mengklasifikasi Kelulusan Mahasiswa Pada Universitas Dian Nuswantoro Semarang," Universitas Dian Nuswantoro, 2011.
- [8] Yuliyana and A. S. R. M. Sinaga, "Sistem Pakar Diagnosa Penyakit Gigi Menggunakan Metode Naive Bayes," *Fountain Informatics J.*, vol. 4, no. 1, pp. 4–8, 2019.
- [9] A. Saleh and F. Nasari, "Penggunaan Teknik Unsupervised Discretization Pada Metode Naive Bayes Dalam Menentukan Jurusan Siswa Madrasah Aliyah," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, pp. 353-360, 2018.
- [10] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
- [11] A. Solichin, "Mengukur Kualitas Citra Hasil Steganografi," *Achmatim.Net*, 2015. [Online]. Available: <http://achmatim.net/2015/04/16/mengukur-kualitas-citra-hasil-steganografi/>. [Accessed: 20-Jun-2016].
- [12] G. Y. Tütüncü and N. Kayaalp, "Journal of Computational and Applied An Aggregated Fuzzy Naive Bayes Data Classifier," *J. Comput. Appl. Math.*, vol. 286, pp. 17–27, 2015.

Matches

Internet matches

25

2	http://achmatim.net/2020/05/02/paper-penerapan-metode-naive-bayes-classifier-untuk-penjurusan-siswa-pada-madrasah	3 Sources	5.27%
3	http://journals.ums.ac.id/index.php/khif/article/view/1175		1.59%
4	https://journal.portalgaruda.org/index.php/EECSI/article/download/1991/1445	4 Sources	0.91%
5	http://repository.usu.ac.id/bitstream/handle/123456789/52750/Chapter%20II.pdf?sequence=4&isAllowed=y	5 Sources	0.86%
6	https://achmatim.net/publikasi	2 Sources	0.75%
7	http://jurnal.upnyk.ac.id/index.php/semnasif/article/download/833/734	4 Sources	0.73%
8	http://journal.widyatama.ac.id/index.php/jitter/article/download/251/215	2 Sources	0.32%
9	http://eltikom.poliban.ac.id/index.php/eltikom/article/download/86/23	2 Sources	0.3%
10	http://baper.if.uinsgd.ac.id/article/detail/0087		0.27%
11	http://eprints.dinus.ac.id/13314/1/jurnal_13886.pdf		0.24%

Exclusions

Internet exclusions

2

https://ejournal.unida.gontor.ac.id/index.php/FIJ/article/download/4007/pdf_36

2 Sources 99.9%