

PENERAPAN ALGORITMA DATA MINING UNTUK PREDIKSI MINAT DAN KEMAMPUAN SISWA DALAM PEMILIHAN JURUSAN DI SMAN 12 TANGERANG SELATAN

Yudi Permana Wiyadi^{1*}, Yuda Pratama Wibawa², Eka Rini Yulia³

^{1,2}Universitas Pamulang, Tangerang Selatan

³Universitas Bina Sarana Informatika,

Keywords:

Algoritma C4.5, Naïve bayes, Penjurusan siswa, Data Mining, Teknologi

Correspondent Email:

dosen03323@unpam.ac.id

Abstrak. Penjurusan siswa yang dilakukan pada Sekolah Menengah Atas sesuai minat dan dilakukan sejak peserta didik mendaftar ke SMA/MA memiliki tujuan untuk memberikan kesempatan kepada peserta didik mengembangkan kompetensi sikap, kompetensi keterampilan peserta didik sesuai dengan minat, bakat, dan kemampuan akademik dalam sekelompok mata pelajaran keilmuan. Pada kurikulum yang baru penjurusan dilakukan pada saat awal masuk, yaitu pada kelas X. Perubahan kurikulum dimaksudkan agar memungkinkan penyesuaian program pendidikan pada satuan pendidikan dengan kondisi dan kekhasan potensi yang ada di daerah peserta didik (peraturan kemdikbud). Kemungkinan yang akan terjadi jika siswa mengalami kesalahan dalam penjurusan adalah rendahnya prestasi belajar siswa atau dapat menyebabkan terjadinya ketidakcocokan dengan jurusan yang sudah dipilih oleh siswa atau siswi sebelumnya.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract. School majors conducted in high school according to the interests and conducted since the students enroll to high school / MA has a goal to provide opportunities to learners develop the competence of attitudes, skills competence of learners in accordance with interests, talents, and academic ability in a group of scientific subjects. In the new curriculum, the curriculum is done at the beginning of admission, which is in class X. The curriculum changes are intended to enable the adjustment of educational programs in educational units with the conditions and peculiarities of potential that exist in the area of learners (kemdikbud regulations). The likelihood that will occur if students experience errors in the majors is the low student achievement or can cause a mismatch with the majors that have been selected by students or previous students.

1. PENDAHULUAN

Pada kurikulum yang baru penjurusan dilakukan pada saat awal masuk, yaitu pada kelas X. Perubahan kurikulum dimaksudkan agar memungkinkan penyesuaian program pendidikan pada satuan pendidikan dengan kondisi dan kekhasan potensi yang ada di daerah peserta didik [10][11]. Kemungkinan yang akan terjadi jika siswa mengalami kesalahan dalam penjurusan adalah rendahnya prestasi belajar siswa atau dapat menyebabkan

terjadinya ketidakcocokan dengan jurusan yang sudah dipilih oleh siswa atau siswi sebelumnya [10][11].

Dalam proses pendidikan di sekolah, perbedaan masing-masing siswa harus diperhatikan karena dapat menentukan baik buruknya prestasi belajar siswa. Tujuan sekolah yang mendasar adalah mengembangkan semua bakat dan kemampuan siswa selama proses pendidikan. Perbedaan individual antara siswa di sekolah di antaranya

meliputi perbedaan kemampuan kognitif, motivasi berprestasi, minat dan kreativitas. Dan dengan adanya perbedaan individu tersebut, maka fungsi pendidikan tidak hanya dalam proses belajar mengajar, tetapi juga meliputi bimbingan konseling, pemilihan dan penempatan jurusan untuk siswa harus sesuai dengan kapasitas individual yang dimiliki siswa tersebut [2].

Penelitian terdahulu yang terkait dengan topik yang sesuai dengan penulisan ini adalah penelitian yang dilakukan oleh Musthofa, dengan judul Pengembangan Sistem Pendukung Keputusan Penjurusan Bagi Siswa Baru Menggunakan Metode *Naive Bayes* yang melakukan klasifikasi penjurusan dengan menggunakan algoritma *Naive Bayes* [7]. Dari penelitian tersebut menghasilkan nilai *Accuracy* sebesar 77,48%. Penelitian yang dilakukan oleh Alfa Saleh, dengan judul Penerapan *Data Mining* Dalam Menentukan Jurusan Siswa untuk melihat keakuratan metode *Naive Bayes* dalam mengklasifikasikan jurusan siswa. Dari hasil penelitian tersebut mengukur tingkat akurasi dengan nilai sebesar 90% [13].

2. TINJAUAN PUSTAKA

Penjurusan siswa yang dilakukan pada Sekolah Menengah Atas sesuai minat dan dilakukan sejak peserta didik mendaftar ke SMA/MA memiliki tujuan untuk memberikan kesempatan kepada peserta didik mengembangkan kompetensi sikap, kompetensi keterampilan peserta didik sesuai dengan minat, bakat, dan kemampuan akademik dalam sekelompok mata pelajaran keilmuan [15].

Pemilihan jurusan merupakan salah satu tahapan penting dalam proses pendidikan di Sekolah Menengah Atas (SMA). Menurut Slameto [12], pemilihan jurusan adalah proses pengambilan keputusan yang dilakukan oleh siswa untuk menentukan bidang studi yang akan mereka tekuni sesuai dengan minat, bakat, kemampuan, dan tujuan masa depan. Pada umumnya, siswa di SMA dihadapkan pada dua pilihan utama, yaitu jurusan Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS). Keputusan ini memiliki dampak jangka panjang terhadap arah studi lanjutan dan peluang karier di masa

depan, sehingga memerlukan pertimbangan yang matang.

Pemilihan jurusan tidak hanya berkaitan dengan kemampuan akademik, tetapi juga melibatkan aspek psikologis dan sosial. Menurut Winkel [14], faktor internal seperti minat, motivasi, serta persepsi terhadap kemampuan diri memiliki pengaruh besar terhadap keputusan siswa. Siswa yang memahami potensi dan kecenderungan pribadinya cenderung mampu memilih jurusan yang sesuai dengan dirinya. Sebaliknya, siswa yang kurang mengenal diri sendiri sering kali mengalami kebingungan dalam menentukan jurusan, bahkan dapat memilih berdasarkan tekanan lingkungan atau sekadar mengikuti teman.

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*Machine Learning*) untuk menganalisis dan mengekstraksi pengetahuan (*Knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dan konsep-konsep yang akan dipelajari [5].

RapidMiner memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, *data preprocessing* dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi [1].

Evaluasi model klasifikasi didasarkan pada pengujian untuk memperkirakan obyek yang benar dan salah, urutan pengujian ditabulasikan dalam *confusion matrix* dimana kelas yang diprediksi ditampilkan dibagian atas matriks dan kelas yang diamati disisi kiri. Setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi [4].

Beberapa kelebihan dari pohon keputusan (*decision tree*)[3].

- a. Hasil analisa berupa diagram pohon yang sangat mudah dimengerti.
- b. Mudah untuk dibangun, serta membutuhkan data percobaan yang lebih

sedikit dibandingkan algoritma klasifikasi lainnya.

c. Mampu mengolah data nominal dan kontinyu.

d. Model yang dihasilkan dapat dengan mudah dimengerti, beberapa teknik klasifikasi yang berbeda seperti neural network menyajikan model dengan informasi logis yang tersirat, sehingga perlu dipelajari.

e. Menggunakan teknik statistik sehingga dapat divalidasi.

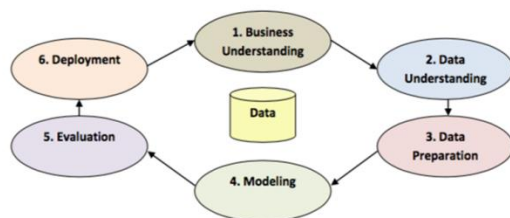
f. Cepat dan handal dalam mengolah dataset besar.

g. Akurasi yang dihasilkan mampu menandingi teknik klasifikasi yang lainnya.

Algoritma C4.5 diperkenalkan sebagai versi perbaikan dari ID3. Dalam ID3, induksi *decision tree* hanya bisa dilakukan pada fitur bertipe kategorikal (nominal dan ordinal), sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan. Perbaikan yang membedakan algoritma C4.5 dari ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) rule set. Algoritma C4.5 juga menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah node pada pohon yang diinduksi [9].

3. METODE PENELITIAN

Untuk penelitian Data Mining, telah ada metodologi standar yang disebut CRISP-DM atau Cross-Industry Standard Process for Data Mining. CRISP-DM merupakan hasil kolaborasi dari beberapa perusahaan, diantaranya Daimler-Benz, OHRA, NCR Corp, dan SPSS Inc. yang mulai dirintis sejak tahun 1999 [8].



Gambar 3.1 CRISP-DM (North, 2012)

CRISP-DM memiliki enam tahapan [8], yaitu:

1. Business Understanding

Pada tahapan pertama ini harus didefinisikan apa pengetahuan yang ingin didapatkan dalam bentuk pertanyaan-pertanyaan yang sifatnya

umum, misalnya bagaimana cara meningkatkan keuntungan, bagaimana cara mengantisipasi kesalahan cacat produk, dan sebagainya.

2. Data Understanding

Tahapan kedua ini bertujuan untuk mengumpulkan, mengidentifikasi, dan memahami aset data yang kita miliki. Data tersebut juga harus dapat diverifikasi kebenaran dan realibilitasnya.

3. Data Preparation

Tahapan ini meliputi banyak kegiatan, seperti membersihkan data, memformat ulang data, mengurangi jumlah data, dan sebagainya yang bertujuan untuk menyiapkan data agar konsisten sesuai format yang dibutuhkan.

4. Modelling

Model adalah representasi komputasi dari hasil pengamatan yang merupakan hasil dari pencarian dan identifikasi pola-pola yang terkandung pada data.

5. Evaluation

Evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil kita buat pada langkah sebelumnya.

6. Deployment

Deployment adalah saat dimana hasil dari seluruh tahapan sebelumnya digunakan secara nyata.

3.1 Tahapan CRISP-DM

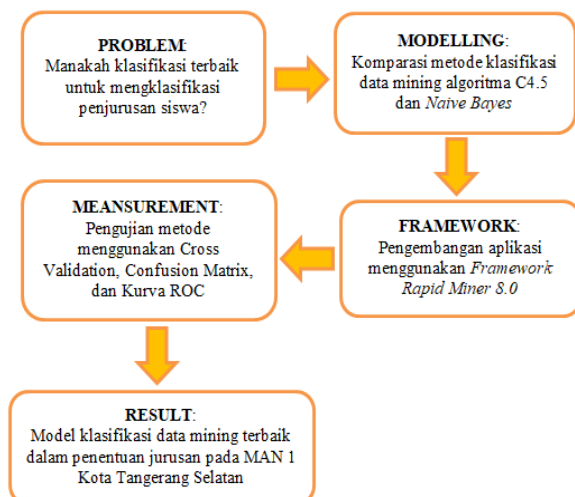
Metode penelitian yang digunakan dalam penelitian ini adalah dengan menggunakan metode eksperimen. Penelitian ini bertujuan untuk mengkomparasi dan mengevaluasi algoritma klasifikasi data mining pada penentuan penjurusan di SMAN 12 Kota Tangerang Selatan. Dalam mendesain metode penelitian eksperimen ini peneliti menggunakan metode penelitian standar yang digunakan pada data mining yaitu Cross Industry Standard Process for Data Mining (CRISP-DM) yaitu terdiri dari 6 fase dengan langkah-langkahnya adalah Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation dan Deployment [6].

3.2 Kerangka Pemikiran

Dalam menyelesaikan penelitian ini, maka penulis membuat sebuah kerangka pemikiran yang berguna sebagai pedoman atau acuan dalam penelitian ini sehingga penelitian ini dapat dilakukan secara konsisten.

Penelitian ini terdiri beberapa tahap seperti yang terlihat pada Gambar 3.2 kerangka pemikiran. Permasalahan problem (Problem) pada penelitian ini adalah apakah algoritma C4.5 dan Naive Bayes dapat diterapkan untuk penentuan jurusan pada SMAN 12 Kota Tangerang Selatan dan algoritma manakah yang akan memberikan model yang terbaik untuk klasifikasi penentu jurusan pada SMAN 12 Kota Tangerang Selatan.

Untuk itu dibuat model (modelling) yaitu dengan menggunakan algoritma C4.5 dan Naive Bayes untuk memecahkan permasalahan kemudian dilakukan pengujian terhadap kedua kinerja model tersebut. Setelah pengujian terhadap kedua model terbentuk, maka akan dilakukan pengujian dengan 10 fold cross validation. Hasil akurasi dari kedua model yang telah terbentuk akan diukur dengan menggunakan confusion matrix, sedangkan Area Under Curve (AUC) akan diukur dengan menggunakan ROC Curve. Untuk mengembangkan aplikasi (deployment) berdasarkan pada model yang telah dibuat maka digunakan tools Rapid Miner 8.0. Berikut penggambaran dari kerangka pemikiran yang telah dilakukan sebagai berikut:



Gambar 3.2 Kerangka Pemikiran

4. HASIL DAN PEMBAHASAN

4.1 Implementasi Metodologi

Sesuai dengan metodologi yang disampaikan di bab 3, berikut implementasi metodologi tersebut pada penelitian yang dilaksanakan:

4.1.1 Business Understanding

Tahap pertama dari CRISP-DM adalah Business Understanding. Pemahaman data siswa pada SMAN 12 Kota Tangerang Selatan bahwa dalam penentuan jurusan khusus nya siswa siswi kelas X yang baru masuk penentu jurusan tersebut hanya beracuan pada hasil uji psikotes saja, maka dari itu peneliti disini mengajukan atribut baru yaitu hasil uji tes masuk siswa siwi kelas X agar digunakan dengan hasil psikotes untuk menentukan hasil penjurusan siswa. Yang nantinya akan di ujikan dengan mengkomparasi dua algoritma yaitu C4.5 dan Naive Bayes.

4.1.2 Data Understanding

Tahap kedua dari CRISP-DM adalah Data Understanding. Pada tahap ini penulis meneliti data calon siswa baru SMAN 12 Kota Tangerang Selatan pada tahu pelajaran 2017/2018, data yang diambil adalah data hasil ujian tes masuk sekolah, dan juga data hasil dari psikotes siswa baru SMAN 12 Kota Tangerang Selatan. Untuk data hasil ujian tes masuk sekolah isinya adalah nilai-nilai mata pelajaran yang diujikan diantaranya adalah: B.Indonesia, Bahasa Inggris, Matematika, Fisika, Biologi dan juga Agama. Data yang digunakan dalam penelitian ini adalah data hasil uji tes masuk kelas X pada SMAN 12 Kota Tangerang Selatan dan juga data hasil Psikotes calon siswa kelas X tahun ajaran 2017-2018 dengan jumlah sebanyak 148 orang dari dua kelas IPA dan tiga kelas IPS. Dimana jumlah siswa laki-laknya sebanyak 60 siswa dan siswi perempuannya ada sebanyak 88 siswi. Jumlah siswa yang berada di kelas IPA ada sebanyak 83 orang dan 65 orang berada di kelas IPS. Berikut adalah data awal sebelum nilai angkanya di konversikan lihat Tabel 4.1 dibawah ini:

Tabel 4.1 Contoh Nilai Ujian Tes

N O	Ind o	Inggr is	MT K	Fisi ka	Biolo gi	Aga ma
1.	17	9	4	13	13	18
2.	13	7	9	9	12	14
3.	17	9	15	16	16	23
4.	17	5	14	17	17	21
5.	19	16	7	16	11	19
6.	17	13	5	16	16	21
7.	17	14	15	17	14	23

8.	15	8	11	10	14	21
9.	18	16	13	11	13	17
10.	17	10	6	17	17	20

Data siswa kelas X tahun pelajaran 2017/2018

Tabel 4.2 Tabel Jumlah Siswa

Kelas	Jumlah Siswa
X IPA 1	33 siswa
X IPA 2	31 siswa
X IPS 1	30 siswa
X IPS 2	28 siswa
X IPS 3	26 siswa

Tabel 4.2 di atas adalah jumlah data siswa kelas X tahun pelajaran 2017/2018 yang diterima di SMAN 12 Kota Tangerang Selatan kelas IPA dan IPS yang akan diolah datanya.

Data hasil psikotes kelas X tahun pelajaran 2017/2018

Tabel 4.3 Contoh tabel hasil psikotes

NO	Jenis Kelamin	Jurusan
1	P	IPA
2	L	IPA
3	L	IPA
4	P	IPA
5	P	IPA
6	P	IPS
7	P	IPS
8	P	IPS
9	L	IPS
10	L	IPS

4.1.3 Data Preparation

Tahapan yang ketiga dari CRISP-DM adalah Data Preparation, tahapan ketiga ini adalah mempersiapkan data sehingga dapat diproses pada tahapan berikutnya. Tahapan ini menyesuaikan atribut-atribut pada tabel data yang akan di proses menggunakan Framework Rapid Miner 8.0. Ditahapan sebelumnya, telah dijabarkan beberapa tabel yang akan diproses pada penelitian ini, ditahapan ini penulis membuat tabel baru yang menggabungkan tabel-tabel tersebut tadi kedalam satu tabel. Data pada Tabel 4.5 dibawah ini adalah hasil

penggabungan dari tabel-tabel data sebelumnya yang akan dimasukkan kedalam proses pengolahan menggunakan Framework Rapid Miner 8.0.

Tabel 4.4 Tabel data siswa

N O	In d o	Ing gris	M T K	Fis ika	Bio logi	Ag am a	J K	Jur usa n
1	17	9	15	15	16	23	P	IPA
2	17	5	14	17	17	21	L	IPA
3	19	16	7	16	11	19	P	IPA
4	17	13	5	16	16	21	P	IPA
5	17	14	15	17	14	23	P	IPA
6	18	11	10	12	14	21	L	IPS
7	16	10	9	12	9	20	P	IPS
8	18	9	9	12	12	14	P	IPS
9	13	7	7	13	13	20	L	IPS
10	20	13	9	15	6	16	L	IPS

Pada tabel data siswa di atas, peneliti mengolah kembali data tersebut dengan mengkonversi nilai-nilai yang redundan atau nilai yang terlalu beragam kedalam kelompok yang lebih kecil untuk mempermudah pembentukan model. Misal nilai dengan jumlah ≤ 2 dikategorikan cukup dan ≥ 17 dikategorikan sangat baik, Tabel 4.5 adalah tabel pengkategorian atribut bisa dilihat dibawah ini:

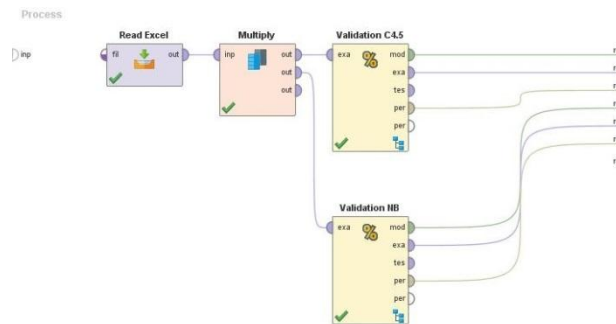
Tabel 4.5 Kategori atribut

Nilai Angka	Kategori
2 – 9	Cukup
10 – 16	Baik
17 – 24	Sangat Baik

4.1.4 Modelling

Pada tahap keempat dalam CRISP-DM adalah Modelling. Pada tahap ini, dataset yang telah dibuat pada tahap sebelumnya digunakan sebagai masukan untuk algoritma klasifikasi,

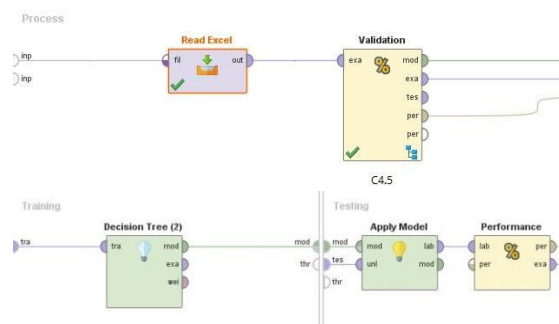
yakni dijadikan sebagai dataset training. Pada penelitian ini akan digunakan dua jenis algoritma klasifikasi yaitu C4.5 dan Naive Bayes. Berikut ini desain proses yang digunakan berikut keterangan nya:



Gambar 4.1 Desain proses perbandingan 2 algoritma (C4.5 dan Naive Bayes)

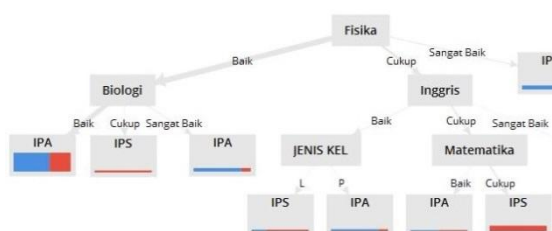
4.1.5 Evaluation

Tahap kelima dari metode CRISP-DM adalah Evaluation. bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil kita buat pada langkah sebelumnya. Untuk evaluasi digunakan 10-fold cross validation. Berikut ini desain proses yang digunakan.



Gambar 4.2 Pengujian Model Algoritma C4.5

Setelah dilakukan pengujian dengan model di atas, maka hasil yang terbentuk akan terlihat seperti pada Gambar 4.3 dibawah ini:



Gambar 4.3 Model Pohon Keputusan Algoritma C4.5

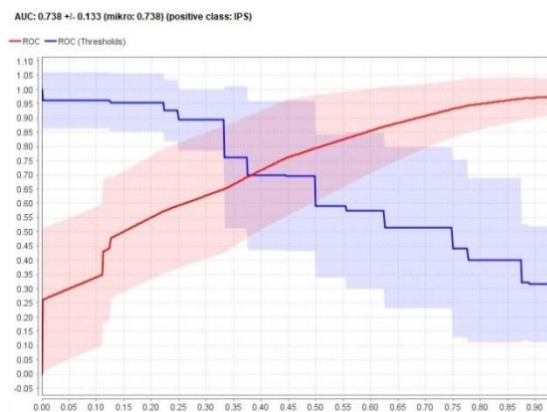
Berikut rule-rule dari model pohon keputusan algoritma C4.5, sebagai berikut:

1. R1 : Fisika = Baik dan Biologi = Baik maka IPA
2. R2 : Fisika = Baik dan Biologi = Cukup maka IPS
3. R3 : Fisika = Baik dan Biologi = Sangat Baik maka IPA
4. R4 : Fisika = Sangat Baik maka IPA
5. R5 : Fisika = Cukup dan Inggris = Baik dan Jenis Kel = L maka IPS
6. R6 : Fisika = Cukup dan Inggris = Baik dan Jenis Kel = P maka IPA
7. R7 : Fisika = Cukup dan Inggris = Cukup dan Matematika = Baik maka IPA
8. R8 : Fisika = Cukup dan Inggris = Cukup dan Matematika = Cukup maka IPS
9. R9 : Fisika = Cukup dan Inggris = Sangat Baik maka IPA

Berdasarkan hasil pengujian dengan menggunakan Rapid Miner maka diperoleh pohon keputusan dan rule-rule model seperti yang terlihat diatas dengan atribut nilai Nama, Indo, Inggris, MTK, Fisika, Biologi, Agama, IQ, Jenis kelamin dan Jurusan. Namun, pada pohon keputusan di atas tidak semua atribut muncul dikarenakan atribut tersebut mempunyai nilai gain yang kecil.

Dari hasil pengujian model yang telah dilakukan adalah untuk mendapatkan hasil akurasi dan Area Under Curve (AUC). Maka mendapatkan hasil grafik ROC dengan nilai Area Under Curve (AUC) sebesar 0,738 dengan performance akurasi yaitu Fair Classification, seperti yang terlihat pada Gambar 4.3 dibawah ini:

- a. ROC Curve (AUC)



Gambar 4.4 Nilai AUC dalam Algoritma C4.5

Dari hasil pengujian yang telah dilakukan dengan model Algoritma C4.5 memperoleh nilai accuracy sebesar 70,29%, seperti terlihat pada Tabel 4.6 dibawah ini:

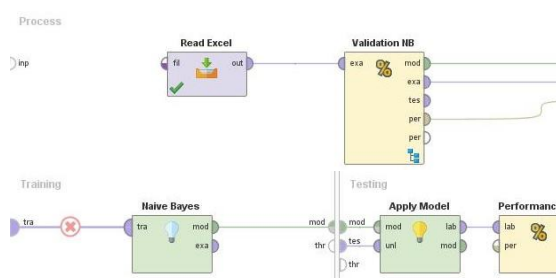
b. Akurasi

Tabel 4.6 Model untuk Algoritma C4.5

	true IPA	true IPS	class precision
pred. IPA	74	35	67.89%
pred. IPS	9	30	76.92%
class recall	89.16%	46.15%	

Jumlah True Positif (TP) adalah 74 record diklasifikasikan sebagai IPA terpilih dan False Negative (FN) sebanyak 9 record diklasifikasikan sebagai IPS terpilih. Berikutnya 35 record untuk False Positif diklasifikasikan sebagai IPA terpilih, dan 30 record untuk True Negative diklasifikasikan sebagai IPS terpilih. Berdasarkan Tabel 4.1 diatas tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma C4.5 adalah sebesar 70,29%.

Serupa dengan proses evaluation pada algoritma C4.5 di atas, untuk algoritma Naive Bayes juga dibuat evaluasinya seperti dibawah ini:

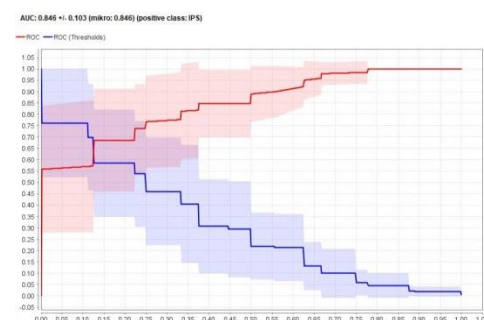


Gambar 4.5 Pengujian Model Algoritma Naive Bayes

5

Dari hasil pengujian model yang telah dilakukan adalah untuk mendapatkan hasil akurasi dan Area Under Curve (AUC). Maka mendapatkan hasil grafik ROC dengan nilai Area Under Curve (AUC) sebesar 0,846 dengan performance akurasi yaitu Good Classification, seperti yang terlihat pada Gambar 4.6 dibawah ini:

a. ROC Curve (AUC)



Gambar 4.6 Nilai AUC dalam Algoritma Naive Bayes

Dari hasil pengujian yang telah dilakukan dengan model Algoritma Naive Bayes memperoleh nilai accuracy sebesar 76,43%, seperti terlihat pada Tabel 4.7 dibawah ini:

b. Akurasi

Tabel 4.7 Model untuk Algoritma Naive Bayes

	true IPA	true IPS	class precision
pred. IPA	69	21	76.67%
pred. IPS	14	44	75.86%
class recall	83.13%	67.69%	

Jumlah True Positif (TP) adalah 69 record diklasifikasikan sebagai IPA terpilih dan False Negative (FN) sebanyak 21 record diklasifikasikan sebagai IPS terpilih. Berikutnya 21 record untuk False Positif diklasifikasikan sebagai IPA terpilih, dan 44 record untuk True Negative diklasifikasikan sebagai IPS terpilih. Berdasarkan Tabel 4.6 diatas tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma Naive Bayes adalah sebesar 76,43%.

4.1.6 Deployment

Tahap ini adalah tahap terakhir dalam standar pemodelan dalam data mining (CRISP-DM). Dalam tahap ini akan dilakukan pembuatan laporan berupa penulisan hasil penelitian tesis dan jurnal dari pendahuluan sampai kesimpulan dan penutup, serta akan dibuat juga Graphical User Interface (GUI) sehingga nantinya user yang menggunakan hasil penelitian ini bisa berinteraksi dan mengaplikasikannya dengan mudah.

4.2 Performance Komparasi

Berdasarkan hasil analisis dari masing-masing pengujian algoritma diatas maka dapat dirangkumkan hasilnya seperti pada Tabel 4.8 berikut ini:

Tabel 4.8 Perbandingan Performance Algoritma

	C4.5	Naive Bayes
AKURASI	70,29%	76,43%
AUC	0,738	0,846

Dari hasil perbandingan performance kedua algoritma diatas, maka hasil pengujian untuk Naive Bayes nilai akurasi lebih tinggi dibandingkan dengan algoritma C4.5. Nilai akurasi untuk model algoritma Naive Bayes sebesar 76,43% dan nilai akurasi untuk model algoritma C4.5 sebesar 70,29% dengan selisih akurasi sebesar 6,14%.

4.3 Analisis Hasil Komparasi

Melihat hasil perhitungan pada Tabel 4.7 di atas dan dengan menerapkan klasifikasi performance keakurasian Area Under Curve (AUC) maka hasil penelitian ini dapat dibagi menjadi dua pengklasifikasian yaitu Fair Clasification untuk algoritma C4.5 dengan AUC sebesar (0,738) dan Good Classification untuk algoritma Naive Bayes dengan nilai AUC sebesar (0,846).

Setelah dilakukan pengujian dengan 10 Fold Cross Validation pada penelitian ini maka dilakukan pengujian kembali dengan menggunakan pengujian T-Test untuk menguji kebenaran dan kepalsuan dari kedua model. Pada pengujian ini akan dilakukan perbandingan antara dua algoritma yaitu C4.5 dan juga Naive

Bayes. Untuk mendapatkan hasil perhitungan statistik T-Test pada algoritma C4.5 dan Naive Bayes maka akan terlihat pada Tabel 4.9 dibawah ini:

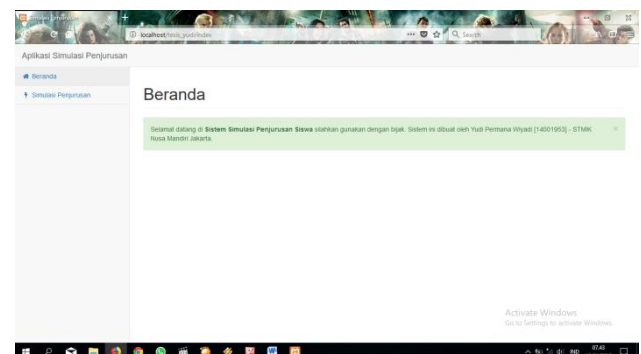
Tabel 4.9 Hasil pengujian T-Test C4.5 dan Naive Bayes

A	B	C
	0.729 +/- 0.089	0.764 +/- 0.124
0.729 +/- 0.089		0.469
0.764 +/- 0.124		

Berdasarkan pada Tabel 4.10 di atas, maka dapat dianalisis bahwa pada algoritma C4.5 dan Naive Bayes memiliki perbedaan nilai yang tidak signifikan, dan memiliki probabilitas >0,05 yaitu 0,469.

4.4 Pengembangan Hasil Analisis

Berdasarkan hasil analisis di atas, maka dari itu penulis mencoba membuat sebuah program Graphical User Interface (GUI) sehingga nantinya user yang menggunakan hasil penelitian ini bisa berinteraksi dan mengaplikasikan mudah. Berikut ini adalah tampilan awal Graphical User Interface (GUI) untuk klasifikasi jurusan siswa.



Gambar 4.7 Tampilan home sistem penjurusan siswa

Gambar 4.7 di atas menampilkan tampilan awal atau home dari sistem simulasi penjurusan siswa yang akan diterapkan di uji coba. Menampilkan tampilan beranda dan disana juga ada tombol untuk simulasi penjurusan, jika

tombol tersebut di klik maka akan menampilkan tampilan sebagai berikut:

Gambar 4.8 Tampilan simulasi penjurusan siswa

Gambar 4.8 di atas menampilkan sebuah form isian untuk simulasi penjurusan siswa. Didalamnya menampilkan atribut-atribut nilai mata pelajaran yang mempengaruhi jurusan yang akan menjadi pilihan siswa seperti: fisika, biologi, matematika dan bahasa inggris juga jenis kelamin sebagai penentu. Ketika semua field itu telah diisi oleh user, untuk melihat hasil penjurusannya user cukup mengklik tombol proses dibawah form isian. Setelah user mengklik tombol proses itu, maka nanti akan tampil hasil yang di peroleh sebagai berikut tampilannya.

Gambar 4.9 Tampilan hasil simulasi penjurusan siswa

KESIMPULAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan maka dapat disimpulkan bahwa: dari hasil pengujian yang telah dilakukan

menghasilkan, nilai akurasi untuk penentu jurusan siswa pada SMAN 12 Kota Tangerang Selatan dengan perbandingan dua algoritma klasifikasi data mining, dapat dibuktikan dengan hasil akurasi dan nilai AUC dari masing-masing algoritma yaitu untuk Naive Bayes akurasi = 76,43% dan nilai AUC = 0,846, sedangkan untuk algoritma C4.5 akurasi = 70,29% dan nilai AUC = 0,738.

5.2 Saran

Saran yang diberikan penulis untuk menyempurnakan hasil penelitian ini adalah:

- a. Dengan penerapan algoritma Naive Bayes diharapkan mampu untuk memberikan solusi bagi siswa-siswi SMAN 12 Kota Tangerang Selatan dalam menentukan jurusan yang sesuai dengan kemampuan, minat dan bakat siswa.
- a. b. Menambahkan beberapa algoritma optimasi seperti Partical Swarm Optimization (PSO) dan Genetic Algoritma (GA) juga algoritma yang lainnya untuk meningkatkan tingkat akurasi terutama dalam penentuan jurusan MAN sehingga dapat lebih bermanfaat bagi para siswa dalam pemilihan jurusan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] D. Aprilia, D. Baskoro, L. Ambarwati, Wicaksana, dan I. W. Simri, Belajar Data Mining dengan RapidMiner. Jakarta, Indonesia, 2013.
- [2] Bahar, "Penentuan Jurusan Sekolah Menengah Atas dengan Algoritma Fuzzy C-Means," Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia, 2011.
- [3] F. Gerunescu, Data Mining: Concepts, Models and Techniques. Berlin, Germany: Springer, 2011.
- [4] K. Hastuti, "Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif," 2012.
- [5] M. S. P. Hermawati, Manajemen Sumber Daya Manusia. Jakarta, Indonesia: Bumi Aksara, 2002.
- [6] D. T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining. Hoboken, NJ, USA: John Wiley & Sons, 2005.

- [7] M. Musthofa, "Pengembangan Sistem Pendukung Keputusan Penjurusan bagi Siswa Baru Menggunakan Metode Naive Bayes," Skripsi, Politeknik Negeri Malang, Malang, Indonesia, 2015.
- [8] M. North, Data Mining for the Masses. Global Text Project, 2012.
- [9] E. Prasetyo, Data Mining: Mengolah Data Menjadi Informasi Menggunakan MATLAB. Yogyakarta, Indonesia: Andi Offset, 2014.
- [10] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 81A Tahun 2013 tentang Implementasi Kurikulum. Jakarta, Indonesia: Kemendikbud, 2013.
- [11] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 64 Tahun 2014 tentang Peminatan pada Pendidikan Menengah. Jakarta, Indonesia: Kemendikbud, 2014.
- [12] A. Sulistiyo, "Penentuan Jurusan Sekolah Menengah Atas Menggunakan Metode K-Nearest Neighbor Classifier pada SMAN 16 Semarang," Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia, 2015.
- [13] A. Saleh, "Penerapan data mining dalam menentukan jurusan siswa," Seminar Nasional Informatika 2015, vol. 1, no. 1, pp. 351–355, 2015
- [14] I. H. Witten, E. Frank, dan M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.
- [15] A. Sulistiyo, "Penentuan jurusan sekolah menengah atas menggunakan metode K-Nearest Neighbor Classifier pada SMAN 16 Semarang," Universitas Dian Nuswantoro (UDINUS), Semarang, 2015.