

Penerapan Algoritma *Decision Tree* C45 untuk Klasifikasi Penjurusan Siswa

Bramefio Qibran Husaini^{1*)}, Jemakmun²⁾

¹⁾ Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Bina Darma

²⁾ Program Studi Sistem Informatika, Fakultas Ilmu Komputer, Universitas Bina Darma

Correspondence author: qibranb@gmail.com, Banyuasin, Sumatera Selatan, Indonesia

DOI: <https://doi.org/10.37012/jtik.v9i1.1512>

Abstrak

SMAN 1 Air Kumbang merupakan sekolah yang penyimpanan datanya telah terkomputerisasi, salah satunya data penjurusan siswa. Penjurusan siswa SMAN 1 Air Kumbang dilakukan pada awal tahun pelajaran sesuai dengan kurikulum 2013 yang terdiri dari 2 jurusan yaitu IPA dan IPS. Penjurusan bertujuan untuk menyalurkan kemampuan, minat dan bakat siswa SMA tersebut. Dalam menentukan penjurusan siswa selama ini belum diterapkannya proses klasifikasi jurusan dengan akurasi nya juga belum diketahui. Penelitian ini bertujuan untuk mengklasifikasi siswa berdasarkan jurusan dengan menerapkan metode data mining menggunakan algoritma decision tree C4.5 dan juga aplikasi Rapidminer. Klasifikasi ini menggunakan atribut kontrol yaitu nilai rata-rata rapor SMP yang meliputi mata pelajaran B. Indonesia, B. Inggris, Matematika dan IPA, serta rekomendasi guru BK SMP. Berdasarkan pengujian yang dilakukan terhadap siswa tahun masuk ajaran 2022/2023 dengan sampel data sebesar ± 400 data dengan atribut rekomendasi guru BK SMP yang mempengaruhi keputusan penjurusan siswa. Hasil diperoleh akurasi sebesar 84.27% dan *classification error* sebesar 15.73%. Dari hasil akurasi sebesar 84.27% berada pada nilai rentang 0,80-0,90 dan hasil ini masuk kategori **Good Classification**. Dapat disimpulkan bahwa menentukan penjurusan dengan menggunakan data mining algoritma decision tree C4.5 dapat mempercepat dan akurat dalam pengambilan keputusan penjurusan siswa SMAN 1 Air Kumbang.

Kata Kunci: *Data Mining*, Klasifikasi, *Decision Tree*, *Algoritma C45*

ABSTRACT

SMAN 1 Air Kumbang is a school that has computerized data storage, one of which is student majors data. The majors for students of SMAN 1 Air Kumbang are carried out at the beginning of the school year according to the 2013 curriculum which consists of 2 majors, namely Natural Sciences and Social Sciences. The major aims to channel the abilities, interests and talents of these high school students. In determining the majors of students so far the process of classifying majors has not been implemented with the accuracy also unknown. This study aims to classify students based on majors by applying the data mining method using the C4.5 decision tree algorithm and also the Rapidminer application. This classification uses control attributes, namely the average value of junior high school report cards which include subjects B. Indonesian, B. English, Mathematics and Science, as well as recommendations for junior high school counseling teachers. Based on tests carried out on students in the 2022/2023 school year with a data sample of ± 400 data with the attributes of the recommendations of Middle School Guidance and Guidance teachers that influence student major decisions. The results obtained were an accuracy of 84.27% and a classification error of 15.73%. From the results of the accuracy of 84.27% is in the range of 0.80-0.90 and these results are included in the Good Classification category. It can be concluded that determining majors using data mining decision tree algorithm C4.5 can speed up and accurately make major decisions for students of SMAN 1 Air Kumbang.

Keywords: *Data Mining*, *Classification*, *Decision Tree*, *C45 Algorithm*

PENDAHULUAN

Sekolah adalah media Pendidikan yang paling utama di Indonesia dalam mengembangkan masyarakat untuk dapat memiliki kehidupan yang lebih baik lagi. Di Indonesia ada banyak sekolah yang tersebar di seluruh Indonesia untuk mendukung tingkat Pendidikan masyarakat yang lebih baik lagi. Sumber daya manusia sebagai pendukung dalam memberikan Pendidikan telah dipersiapkan oleh pemerintah melalui guru Honor dan PNS. Hal ini juga merupakan bagian dari tanggung jawab pemerintahan dalam menyediakan Pendidikan yang layak untuk seluruh kalangan masyarakat di Indonesia.

Tujuan dari jurusan ini yakni untuk membuat pelajaran yang ditawarkan kepada siswa lebih fokus ke masa depan, karena tidak jarang siswa ceroboh dalam memilih jurusan yang akan diambil dan kendala lain datang dari dalam atau luar siswa itu sendiri. Kecendrungan yang selalu ada sekarang, siswa bingung atau ragu dalam memilih minat dan bakatnya juga memilih jurusan apa (Hernadewita, Sambani, & Nuaeni, 2018). Faktor dari luar sering kali berpengaruh dalam memilih jurusan untuk siswa oleh orang tua mereka dan juga paksaan itu biasanya karena harapan orang tua terhadap pekerjaan anak-anak mereka di masa depan (Subiyanto, 2017).

Usaha untuk menyelenggarakan pendidikan yang berkualitas pada suatu instansi sekolah dan perguruan tinggi banyak siswa/i ditingkat kejuruan masih sulit untuk menentukan jurusan yang diminatin sehingga pemilihan jurusan tidak sesuai dengan minat dari siswa itu sendiri. Maka pada akhirnya siswa tersebut banyak yang putus sekolah dan kurang bersemangat untuk belajar karena tidak sesuai dengan minat. Seperti yang terjadi pada SMA N 1 Air Kumbang yang terdiri dari jurusan IPA dan IPS. Untuk mengetahui siswa yang berminat dalam memilih jurusan yang diinginkan maka perlu data siswa setiap kelas dari nilai masing-masing siswa. Sehingga berdasarkan nilai tersebut bisa di kelompokkan siswa yang banyak memilih jurusan yang sama dan mengetahui siswa yang sesuai minat untuk memilih jurusan tersebut.

Data mining merupakan salah satu ilmu pengetahuan yang sangat berkembang dalam pengolahan data menjadi informasi yang berguna. Pengelompokan data siswa yang berminat dan tidak berminat untuk memilih jurusannya dibangku Sekolah merupakan salah satu langkah untuk mempermudah pendidik untuk melaksanakan kegiatan proses pembelajaran dan menghasilkan peserta didik yang berkualitas. Proses pembelajaran yang dilakukan

pendidik yaitu dengan melakukan perencanaan proses pembelajaran, pelaksanaan proses pembelajaran, penilaian hasil pembelajaran, dan pengawasan proses pembelajaran untuk terlaksananya proses pembelajaran yang efektif dan efisien. Namun dari hal itu tidak termasuk penyelesaian masalah untuk mendapatkan didikan yang bersemangat untuk belajar. Tapi bergantung bagi siswa itu sendiri, semangat dan niat.

Data Mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan (Haryati, Sudarsono, & Suryanan, 2015). Data mining merupakan bagian dari *Knowledge Discovery in Database* (KDD) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan data, transformasi, data mining dan evaluasi hasil (Nas, 2020). Dengan menggunakan data mining maka informasi berharga di dalam kumpulan data dapat digali. Dengan menggunakan data mining, data-data minat calon mahasiswa baru dapat diolah dengan sebuah algoritma. Didalam data mining terdapat algoritma-algoritma yang mampu menganalisis sebuah data. Dalam penelitian ini, algoritma yang digunakan adalah algoritma C4.5. Algoritma C4.5 merupakan metode klasifikasi dan prediksi yang digunakan untuk membentuk *Decision Tree* berdasarkan data training (Rismayanti, 2015).

Decision tree adalah struktur *flowchart* yang memiliki *Tree* (Pohon), dimana setiap simpul internal menandakan suatu tes atribut, setiap cabang merepresentasikan hasil tes dan simpul daun merepresentasikan kelas atau distribusi kelas. Berdasarkan uraian diatas maka peneliti merasa tertarik untuk mengambil judul penelitian Penerapan Algoritma *Decision Tree C45* Untuk menentukan penjurusan siswa.

METODE

Metode penelitian yang di pakai pada penelitian ini adalah *Knowledge Discovery in Database* (KDD), untuk menghasilkan informasi sesuai dengan urutan yang sudah ditentukan. Pengertian KDD lebih spesifik menyatakan istilah *Knowledge Discovery in Database* (KDD) dan data mining secara bergantian untuk memberi penjelasan mengenai proses penggalian informasi yang tersembunyi di dalam suatu kumpulan data yang besar. Tetapi dua istilah tersebut mempunyai konsep yang berbeda, tetapi memiliki kerkaitan anara satu sama lain dan salah satu tahap dalam proses KDD adalah data mining (Arta, Indrawan, & Dantes 2017).

Hal pertama yang perlu dilakukan dalam analisis kebutuhan sistem adalah menentukan dan mengungkapkan kebutuhan sistem. Tahapan awal yang harus dilakukan oleh analis dalam tahap konsep yaitu menentukan kebutuhan perangkat keras dan perangkat lunak antara lain:

1) Kebutuhan Perangkat Keras.

Spesifikasi perangkat keras yang digunakan pada saat analisis sistem yaitu:

- a. *Processor core i5 (3,26 GHz, 8 MB L3 cache).*
- b. *2 GB DDR3 Memory.*
- c. *Harddisk 320 GB HDD.*
- d. *Smartphone Android*

2) Kebutuhan Perangkat Lunak.

Perangkat lunak yang dibutuhkan dalam pembangunan sistem yaitu:

- a. *Microsoft Windows 10*
- b. *Rapidminer 9*
- c. *Microsoft Office*

Data mining yaitu suatu proses penggalian data data dari sebuah informasi yang sangat penting. Data Mining juga merupakan suatu proses untuk menggali pola-pola dari data. Pola-pola itu didapatkan dari berbagai jenis basis data seperti basis data relasional, data warehouse, data transaksi, dan data berorientasi objek. Penggunaan data mining dapat membantu para pebisnis dalam pengambilan keputusan secara cepat dan tepat (Nurajizah, 2019).

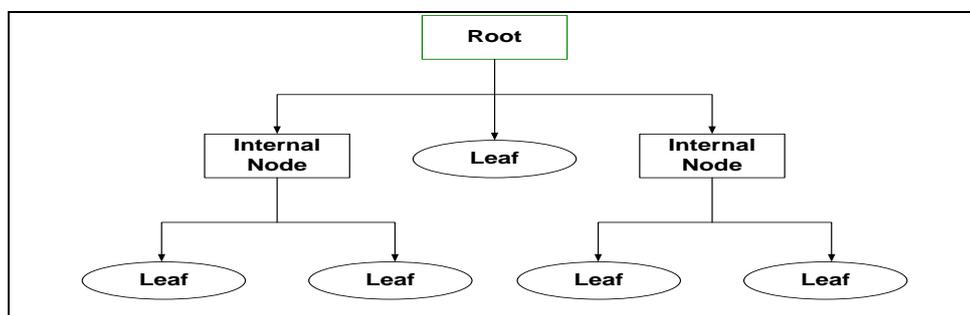
Data mining merupakan sebuah inti dari proses *Knowledge Discovery in Database* (KDD), meliputi dugaan algoritma yang mengeksplor data, membangun model dan menemukan pola yang belum diketahui. KDD bersifat otomatis, dapat didefinisikan sebagai pengorganisasian proses untuk pengidentifikasian yang benar, berguna dan penemuan pola dari kumpulan data yang besar dan kompleks (Muningsih & Kiswati, 2015).

Decision tree atau pohon keputusan merupakan metode klasifikasi yang sangat kuat dan populer. *Decision tree* dapat mengubah data yang sangat besar menjadi sebuah pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami, karenanya *decision tree* merupakan teknik klasifikasi yang mudah untuk dipelajari dan sangat populer digunakan. *Decision tree* merupakan teknik yang digunakan

untuk melakukan klasifikasi terhadap sekumpulan objek atau data dengan representasi struktur pohon. Pohon keputusan umumnya digunakan untuk eksplorasi pola pada sekumpulan data serta menemukan hubungan antara sejumlah variabel input dengan variabel output (Agustin, Kusriani, & Luthfi, 2017).

Adapun bagan *decision tree* terdiri dari tiga bagian yaitu :

1. *Root Node* : Merupakan *node* yang terletak paling atas dari pohon keputusan.
2. *Internal Node* : Merupakan *node* percabangan, dimana hanya terdapat satu input dan menghasilkan minimal dua output.
3. *Leaf Node* : Merupakan *node* akhir, hanya mempunyai satu input dan tidak memiliki output.



Gambar 1. Struktur Pohon Keputusan

Algoritma pohon keputusan C4.5 digunakan untuk membangun pohon keputusan sesuai dengan situasi aktual dan faktor non-intelijen pada prestasi belajar bahasa Inggris mahasiswa. Hasil tes menunjukkan bahwa aturan prediksi dapat secara akurat memprediksi rata-rata siswa dan siswa yang kurang berprestasi dalam prestasi belajar bahasa Inggris. Ini dapat digunakan sebagai prediktor yang baik untuk memenuhi persyaratan bantuan pengajaran dan digunakan untuk siswa dengan prestasi belajar Bahasa Inggris dengan nilai rata-rata atau buruk (Sun, Lang, Fujita, & Li, 2018).

Metode pohon keputusan dengan algoritma C4.5 dibangun metodologi seleksi penalaran kategori dengan meningkatkan akuisisi pada koefisien keseimbangan menggunakan elemen pengetahuan yang temporer. Metode ini merupakan kunci manajemen temporer dibagi dengan pemilihan model dua tingkat melalui penggunaan model hirarkis. Metode ini merupakan proses konstruksi model teratas berdasarkan penalaran kategoris dan proses pemilihan model yang mendasarinya berdasarkan algoritma pohon keputusan C4.5.

Dari hasil tersebut, metode ini meningkatkan akurasi dan ketepatan waktu (Han, Kamber, & Pei, 2019).

Menurut (Muhamad, Windarto, & Suhada, 2019) Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera berikut:

$$Gain(S, A) = Entropy(S) \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai Gain adalah dengan mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar :

$$Entropy(S) \sum_{i=1}^n -p_i * \log_2 p_i \quad (2)$$

S : Himpunan Kasus

n : Jumlah partisi S

p_i : Proporsi dari S_i

Evaluasi dilakukan untuk mengetahui tingkat performa dari hasil klasifikasi yang berupa *classification rule*. Kriteria yang dinilai paling tepat dalam mengukur tingkat performa dari *classification rule* adalah prediksi akurasi yang secara rinci menjelaskan kinerja *classifier* (Andriani, 2017).

Kurva ROC yaitu Confusion Matrix menampilkan nilai akurasi berupa angka persentase ketepatan classification rule dalam mengklasifikasi data yang dilengkapi dengan nilai *precision* atau *confidence* dan nilai *recall* atau *sensitivity* (Andriani, 2017). Kurva ROC menunjukkan nilai akurasi dengan menggunakan grafik dua dimensi. Kurva ROC disebut juga dengan AUC yaitu *The area under curve*. Nilai yang ditunjukkan pada kurva ROC dikategorikan dalam beberapa kategori yang ditunjukkan pada Tabel 1. (Andriani, 2017).

Tabel 1. Kategori Nilai Kurva Roc

Nilai Akurasi	Kategori
0,90-1,00	<i>Excellent Classification</i>
0,80-0,90	<i>Good Classification</i>
0,70-0,80	<i>Fair Classification</i>
0,60-0,70	<i>Poor Classification</i>
0,50-0,60	<i>Failure</i>

Rapidminer merupakan sebuah alat atau perangkat lunak untuk mengelola *data mining*. Aplikasi ini menjadi salah satu solusi dalam melaksanakan analisa *data mining*, analisis peramalan serta *text mining*. Aplikasi ini memakai bermacam-macam teknik prediksi dan deskriptif untuk menyampaikan pengetahuan untuk pemakai sampai bisa menghasilkan keputusan yang sesuai dengan apa yang diharapkan (Purwanto & A., 2018). *RapidMiner* dikembangkan pada *open core model*, dengan menggunakan *RapidMiner Basic Edition* pengguna dapat mengunduhnya dibawah lisensi AGPL (Arunadevi, Ramya, & Raja, 2018).

HASIL DAN PEMBAHASAN

Dalam penelitian ini bahan yang digunakan untuk diolah sebagai bahan analisis adalah data siswa di tiga tahun terakhir yang ada pada SMA Negeri 1 Air Kumbang. Data yang sudah melalui tahap transformasi yang dapat dilihat pada Tabel 2. Berdasarkan tabel 2 tersebut, dilakukan perhitungan klasifikasi penjurusan siswa menggunakan *Decision Tree*, maka didapatkan perhitungan seperti Tabel 3. Tahapan pada Algoritma *Decision Tree C45* ini adalah menentukan calon noktah keputusan dengan cara menyusun calon cabang (*candidate split*) yang dilakukan terhadap seluruh atribut prediktor yaitu Rata-rata nilai IPA, dan Rata-rata nilai IPS, sehingga terbentuk daftar calon cabang mutakhir seperti yang ditunjukkan pada Tabel 4.

Tabel 2. Hasil Transformasi Data Siswa

	A	B	C	D	E	F	G
1	No	NO. PENDAFTARAN	NISN	NAMA LENGKAP	IPA	IPS	Pilihan Jurusan
2	1	PPDB-001	47882258	Tedi Irwansyah	78,8	81,6	IPA
3	2	PPDB-002	53784900	A Ahsanul Wafa	78,2	82,2	IPA
4	3	PPDB-003	36140642	Giga Alfais	85,4	86,6	IPA
5	4	PPDB-005	51371255	Dara Mentari Indani	87,2	86,8	IPA
6	5	PPDB-006	50715881	Reza Restu Ningtias	79,8	77,2	IPA
7	6	PPDB-007	53090601	Desma Dwi Sara	84,8	86,4	IPS
8	9	PPDB-015	39804486	Cindy Maretha Yusman	76,4	79,6	IPS
9	10	PPDB-017	56941043	Salwa Afiqah	89	90,4	IPS
10	11	PPDB-018	51378760	Ilham Mahfudz	89,2	89,6	IPA
11	12	PPDB-019	60072303	Riena Karina	87,6	88,4	IPA
12	13	PPDB-020	45195788	Desti Yana Juwita	84,2	84,4	IPA
13	14	PPDB-021	50713090	DEDI ATMI SUGANDA	85,4	82,4	IPA
14	15	PPDB-024	46554769	Chandika Adestia	87,6	87,6	IPA
15	16	PPDB-025	60030211	NABILA AZZAHRA	76,2	77,8	IPS
16	17	PPDB-027	45095650	RENI RAHMAWATI	79,2	80,2	IPA
17	18	PPDB-031	59525675	Juli Yansa	83	87,6	IPA
18	19	PPDB-033	50732350	Anggie Travela	87,8	87,4	IPA
19	20	PPDB-034	55403438	M. FURQON NAJIR	76,2	79,8	IPS
434	437	PPDB-441	50731597	Muhammad Sidik	76,8	75,4	IPS
435	438	PPDB-512	65164358	Azzahra	85,8	86,2	IPA
436	439	PPDB-830	42765253	Jabarudin Ahmad	82	78,8	IPA
437	440	PPDB-298	50718326	Wildan Arosyid	83,4	82	IPS
438	441	PPDB-845	45072083	IIN NOVIANI	77,2	76	IPS
439	442	PPDB-311	4510819	Ridho Armanda	73,6	73,8	IPS
440	443	PPDB-683	45751587	Mulan Aulia Putri	76	79,4	IPS
441	444	PPDB-196	59939365	REVALDO	78,2	75,8	IPS
442	445	PPDB-403	52303251	TOMI INDRA SAPUTRA	81,2	81	IPS
443	446	PPDB-187	51234235	Lukvan Ansori	78	80,8	IPA
444	447	PPDB-871	57218490	Safingudin	75,6	75,4	IPS
445	449	PPDB-313	50732376	ANJAR SULASIH	76,6	76,6	IPA
446	451	PPDB-793	44733575	ANANG RISMANTO	72,2	72,2	IPA
447	452	PPDB-843	45421660	SANDI WIRATAMA	76	79,4	IPS

Tabel 3. Daftar Calon Mutakhir

Cabang	Cabang Kiri	Cabang Kanan
2	Rata-rata nilai IPA ≤ 65	Rata-rata nilai IPA > 65
3	Rata-rata nilai IPA ≤ 80	Rata-rata nilai IPA > 80
4	Rata-rata nilai IPS ≤ 70	Rata-rata nilai IPS > 70
5	Rata-rata nilai IPS ≤ 85	Rata-rata nilai IPS > 85

Data yang telah diolah pada Tabel 3 selanjutnya di seleksi berdasarkan data latihan sehingga menghasilkan data calon mutakhir pada Tabel 4.

Tabel 4. Daftar Calon Mutakhir

Cabang	Cabang Kiri	Cabang Kanan
1	0	15
2	6	9
3	0	15
4	8	17

Data diatas adalah data latih yang diambil dari data record sebanyak 15 record data. Selanjutnya menghitung *candidate split purity left* (PL) dan *purity right* (PR) sehingga menghasilkan nilai yang ada pada Tabel 5.

Tabel 5. Nilai *candidate split purity left* (PL) dan *purity right* (PR)

Cabang	Cabang Kiri	Cabang Kanan
1	0/15=0.	15/15=1
2	6/15=0.4	9/15=0.6
3	0/15=0	15/15=1
4	8/15=0.5333	7/15=0.4667

Selanjutnya melakukan penentuan pada label jurusan menggunakan persamaan untuk jurusan IPA dan IPS. Sehingga menghasilkan perhitungan pada Tabel 6.

Tabel 6. Label Jurusan

Cabang	Jurusan	Nilai (j/T_L)	Nilai (j/T_R)
1	IPA	0/0 = 0	11/15 = 0.625
	IPS	0/0= 0	1/15 = 0.5
2	IPA	4/6 = 0.6666	7/9 = 0.7778
	IPS	1/6 = 0.1666	0/9 = 0
3	IPA	0/0 = 0	11/15 = 0.7333
	IPS	0/0 = 0	1/15 = 0.6667
4	IPA	6/8 = 0,75	5/7 = 0.7143
	IPS	1/8 = 0.125	0/7 = 0

Selanjutnya mencari nilai $Q (s/t)$ dengan cara mencari selisih nilai $P (j/TL)$ dan $P(j/TR)$ pada setiap label atau jurusan. Sehingga menghasilkan nilai $Q (s/t)$ seperti ditunjukkan pada Tabel 7.

Tabel 7. Perhitungan Nilai $Q (s/t)$

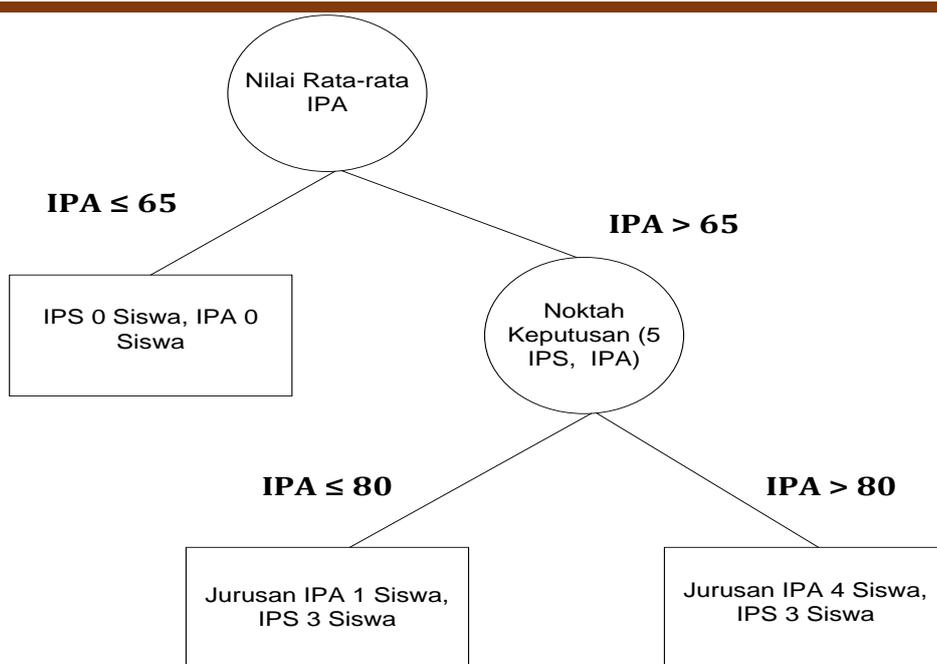
Cabang	Jurusan	Nilai ($j T_L$)	Nilai ($j T_R$)	SelisihAbsolute	$Q (s/t)$
1	IPA	$0/0 = 0$	$11/15 = 0.625$	0.625	1.125
	IPS	$0/0 = 0$	$1/15 = 0.5$	0.5	
2	IPA	$4/6 = 0.6666$	$7/9 = 0.7778$	0.7112	0,8778
	IPS	$1/6 = 0.1666$	$0/9 = 0$	0.1666	
3	IPA	$0/0 = 0$	$11/15 = 0.7333$	0.7333	1.4
	IPS	$0/0 = 0$	$1/15 = 0.6667$	0.6667	
4	IPA	$6/8 = 0,75$	$5/7 = 0.7143$	0.0357	0.1607
	IPS	$1/8 = 0.125$	$0/7 = 0$	0.125	

Berdasarkan Tabel 7 diatas, selanjutnya dihitung nilai kesesuaian (*goodness*) calon cabang $\phi(s|t)$ dengan mengalikan nilai 2PLPR dan $Q(s|t)$. Nilai 2PLPR didapat dari $(2PL \times PR)$ kalikan 2. Hasil perhitungannya ditunjukkan pada Tabel 8.

Tabel 8. Perhitungan Nilai P_L dan P_R

No	$2P_1P_2$	$Q(s t)$.	$\phi(s t)$
1	0	1.125	0
2	0.48	0.8778	0,4213
3	0	1.4	0
4	0.4978	0.1607	0,0799

Calon cabang tertinggi berada pada nomor calon cabang ke 1 yaitu **0.4213** berdasarkan Tabel 8 diatas. Nilai yang dihasilkan pada calon cabang kiri Rata-rata nilai IPA ≤ 80 dan calon cabang kanan Rata-rata nilai IPA > 80 sehingga calon cabang inilah yang dipilih sebagai *root node* pada tahap ini. Hal ini akan menghasilkan jumlah kesesuaian hasil prediksi jurusan yang memilih jurusan IPA dan IPS. Perhitungan iterasi terakhir ditunjukkan pada Gambar 2 di bawah ini, yang menghasilkan pohon keputusan.



Gambar 2. Pohon keputusan

Selanjutnya dilakukan pengklasifikasian data uji seperti tabel pada Tabel 9

Tabel 9. Data Uji

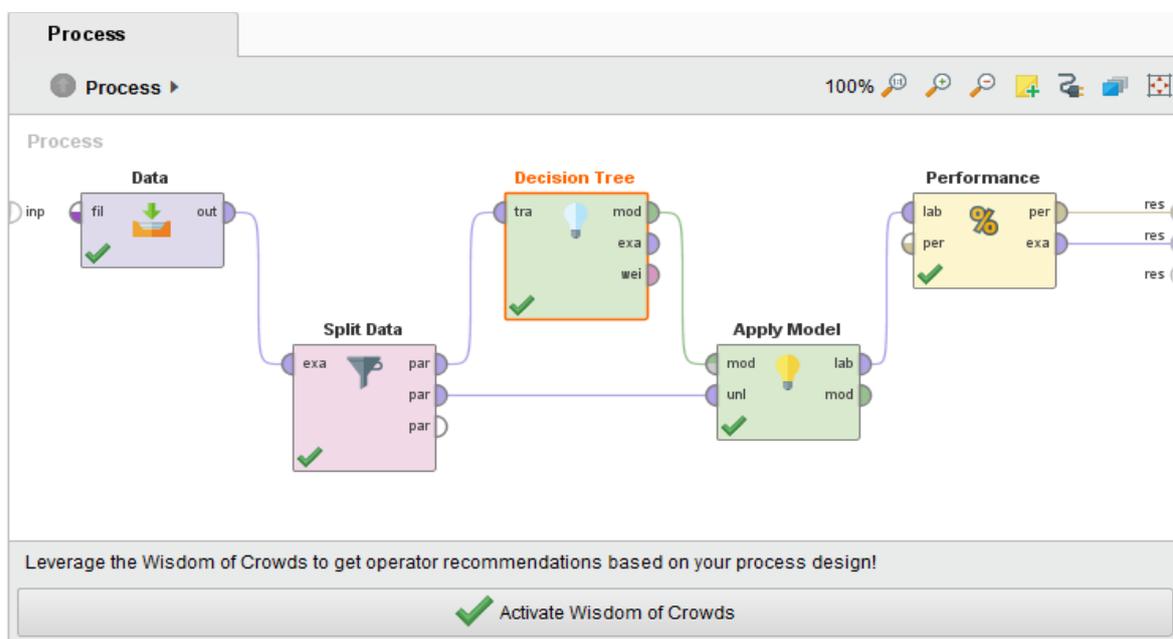
No	NAMA LENGKAP	IPA	IPS	Pilihan Jurusan	Prediksi
1	Andri Muhammad	80	54	IPA	?

Berdasarkan kesimpulan *rule*, rata-rata nilai IPA merupakan nilai yang diprioritaskan terlebih dahulu untuk menentukan jurusan. Siswa yang bernama **Kurniawan Sandy** tidak memenuhi syarat dikarenakan rata-rata nilai $IPS \leq 70$ sehingga tidak memenuhi masuk jurusan IPS, rata-rata nilai $IPA > 78$ sehingga memenuhi masuk jurusan IPA. Sehingga disimpulkan bahwa jurusan yang cocok untuk **Kurniawan Sandy** adalah IPA. Hasilnya ditunjukkan pada Tabel 10.

Tabel 10. Data Hasil Uji Perhitungan

No	NAMA LENGKAP	IPA	IPS	Pilihan Jurusan	Hasil
1	Andri Muhammad	80	54	IPA	IPA

Data yang telah diolah kemudian akan dilakukan perhitungan menggunakan RapidMiner, tahapan dimulai dari konfigurasi data ke RapidMiner dengan mengatur dataset baru kemudian pembentukan model. Tampak seperti Gambar 3.



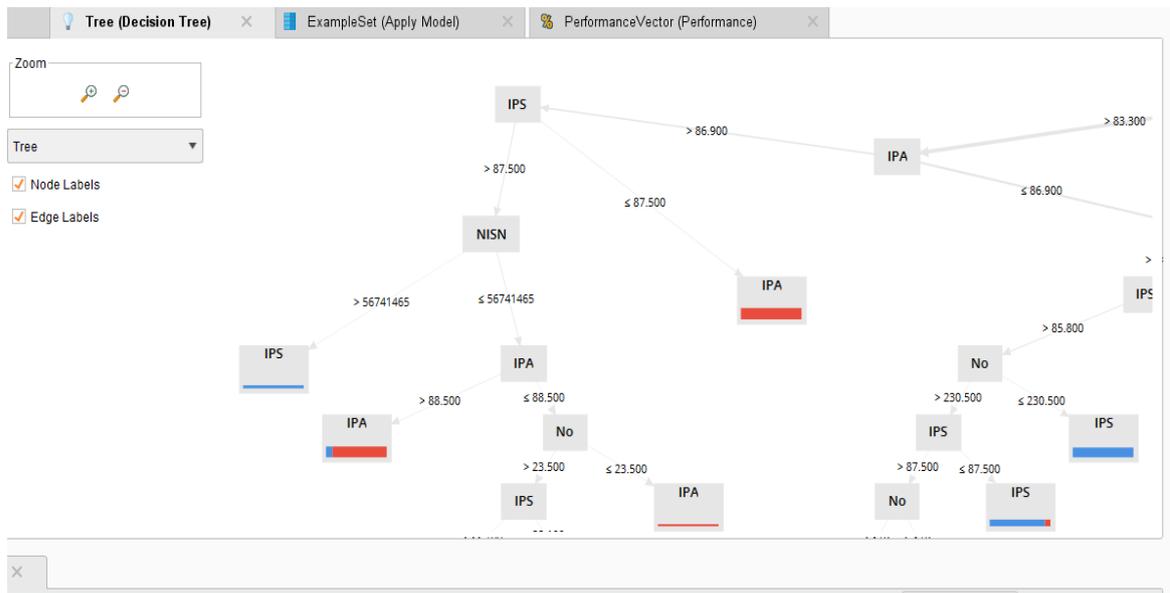
Gambar 3. Pembentukan Model RapidMiner

Dataset dibagi menjadi dua, yaitu training, dan testing. Dataset yang ada dalam testing inilah yang akan diolah, sehingga menghasilkan keluaran yang bisa dijadikan sebagai hasil analisa pada algoritma *Decision Tree*.

Row No.	Pilihan Jurusan	prediction(Pilihan J...	confidence(L...	confidence(L...	No	NO. PENDAF...	NISN	NAMA LENG...	IPA	IPS
1	IPS	IPS	1	0	9	PPDB-015	39804486	Cindy Mareth...	76.400	79.600
2	IPS	IPS	1	0	12	PPDB-019	60072303	Riena Karina	87.600	88.400
3	IPS	IPS	0.960	0.040	17	PPDB-027	45095650	RENI RAHIMA...	79.200	80.200
4	IPS	IPS	0.968	0.032	18	PPDB-031	59525675	Juli Yansa	83	87.600
5	IPA	IPA	0	1	19	PPDB-033	50732350	Anggie Travela	87.800	87.400
6	IPS	IPS	0.968	0.032	32	PPDB-061	54417177	Trisna Muallifah	83.200	85.400
7	IPA	IPA	0	1	47	PPDB-082	43860357	ANISA NUR A...	78.600	72
8	IPS	IPA	0	1	49	PPDB-084	56380560	RAHMANING...	79	79.600
9	IPA	IPA	0	1	70	PPDB-134	50732341	Dina Nur Ind...	88.200	86.400
10	IPS	IPS	0.950	0.050	83	PPDB-151	54457456	IRPAN MUZA...	75.400	77.600
11	IPS	IPS	0	0	86	PPDR-155	51378625	Indah Putri T.	87.200	88

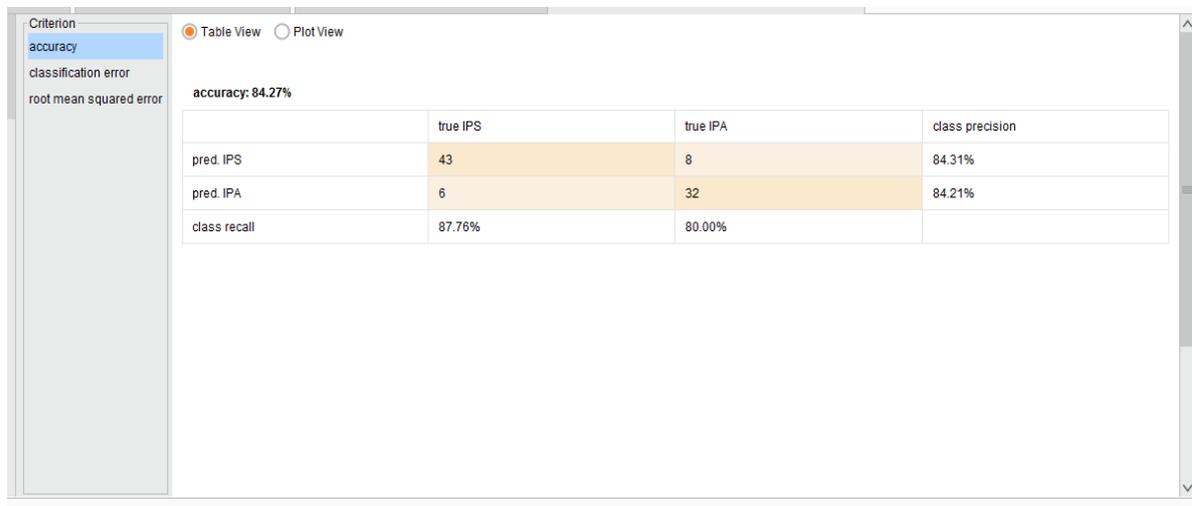
Gambar 4. Hasil Pengolahan *Decision Tree* C45 Model RapidMiner

Kemudian untuk hasil pohon keputusan pada model *Decision Tree* akan tampak terlihat seperti Gambar 5 di bawah ini:



Gambar 5. Hasil Pohon Keputusan

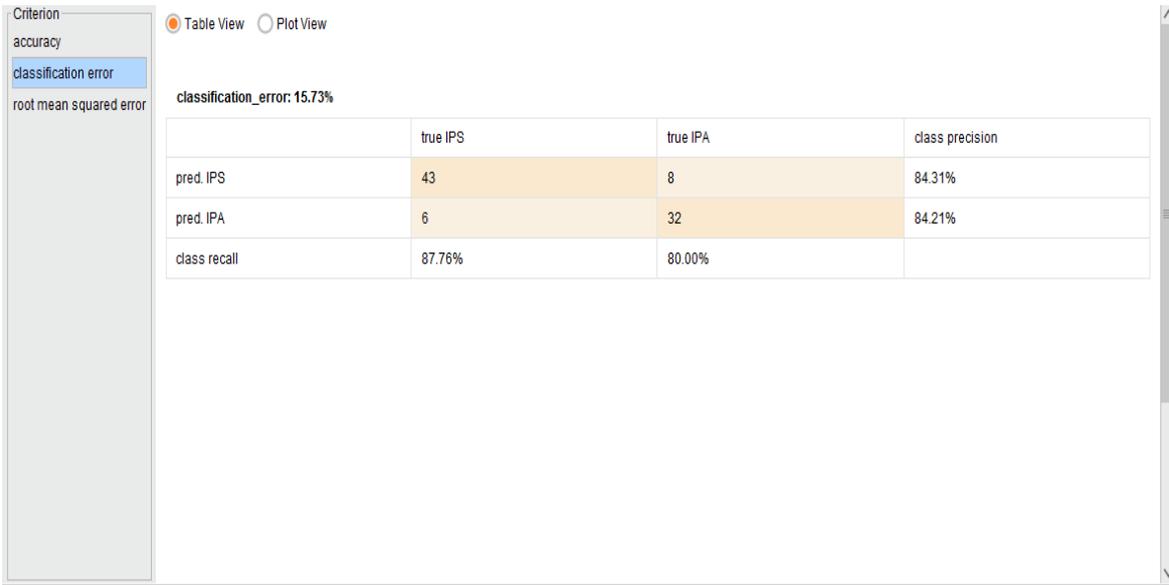
Pada tahapan ini dilakukan pengukuran tingkat akurasi, tujuan dari mengukur akurasi adalah untuk melihat berapa nilai akurasi pada prediksi. Berikut ini adalah hasil proses prediksi untuk pengukuran tingkat akurasi.



Gambar 6. Nilai Accuracy

Nilai *accuracy* yang didapatkan dari prediksi untuk mengukur tingkat akurasi adalah sebesar 84.27%. Selanjutnya dilakukan pengukuran tingkat klasifikasi kesalahan, tujuan dari

mengukur klasifikasi adalah untuk melihat jarak nilai prediksi dengan kenyataannya. Berikut ini adalah proses prediksi untuk pengukuran tingkat kesalahan.

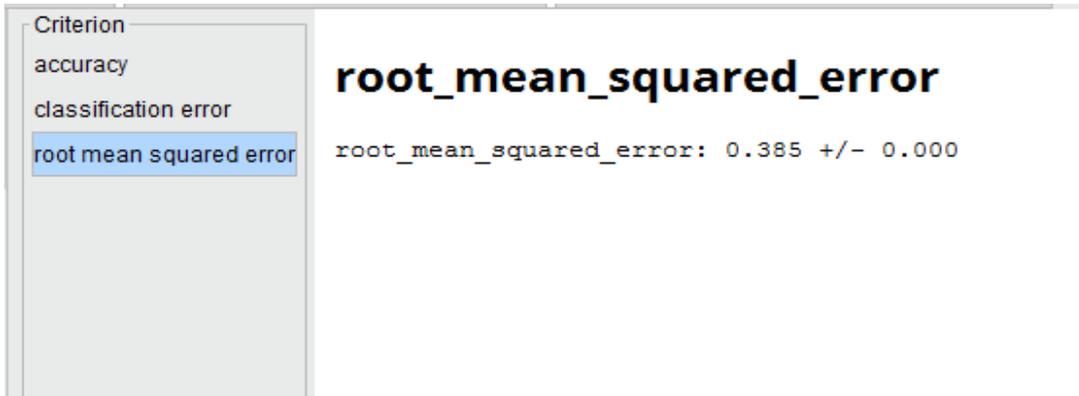


The screenshot shows a software interface for analyzing classification errors. On the left, a 'Criterion' menu is open, listing 'accuracy', 'classification error', and 'root mean squared error'. The 'classification error' option is selected. The main area displays 'classification_error: 15.73%' and a table with the following data:

	true IPS	true IPA	class precision
pred. IPS	43	8	84.31%
pred. IPA	6	32	84.21%
class recall	87.76%	80.00%	

Gambar 7. Hasil *Classification Error*

Nilai *classification error* yang didapatkan dari prediksi untuk mengukur klasifikasi adalah sebesar 15.73%. Kemudian mengukur RSME, tujuan RSME adalah untuk melihat jarak nilai prediksi dengan kenyataannya. Berikut ini adalah proses prediksi untuk pengukuran tingkat kesalahan.



The screenshot shows a software interface for analyzing the root mean squared error. On the left, a 'Criterion' menu is open, listing 'accuracy', 'classification error', and 'root mean squared error'. The 'root mean squared error' option is selected. The main area displays 'root_mean_squared_error' in large text and 'root_mean_squared_error: 0.385 +/- 0.000' below it.

Gambar 8. Nilai *Root Mean Square Error*

Nilai *root mean square error* yang didapatkan dari prediksi untuk mengukur tingkat kesalahan adalah sebesar 0.385.

Dilakukan pengukuran tingkat performa dari model *Decision Tree*. Performa diukur berdasarkan *Confusion Matrix*. Berikut ini adalah proses hasil *Confusion Matrix*.

```
ConfusionMatrix:  
True:  IPS    IPA  
IPS:   43     8  
IPA:   6      32  
root_mean_squared_error: 0.385 +/- 0.000
```

Gambar 9. *Confusion Matrix*

Dari hasil proses pengklasifikasian data ini dapat dilihat nilai *accuracy*, dan *classification error* dari setiap kelas dapat dilihat tingkat kemampuan pemrosesan sistem dalam mencari tingkat keputusan dalam menentukan jurusan bagi siswa yang SMA Negeri 1 Air Kumbang.

KESIMPULAN DAN REKOMENDASI

Dari hasil proses pengklasifikasian data ini dapat dilihat nilai *accuracy*, dan *classification error* dari setiap kelas. Dapat dilihat tingkat kemampuan pemrosesan sistem dalam mencari tingkat keputusan dalam menentukan jurusan bagi siswa SMA Negeri 1 Air Kumbang. Dengan nilai *accuracy* yang didapatkan dari prediksi untuk mengukur tingkat akurasi adalah sebesar 84.27% dan *classification error* sebesar 15.73%. Dari hasil akurasi sebesar 84.27% berada pada nilai rentang 0,80-0,90 dan hasil ini masuk kategori **Good Classification**, maka peneliti menyimpulkan bahwa hasil prediksi jurusan siswa menggunakan RapidMiner dengan metode Algoritma C45 dapat dikatakan baik.

REFERENSI

- Agustin, Y. H., Kusriani, K., & Luthfi, E. T. (2017). Klasifikasi Penerimaan Mahasiswa Baru Menggunakan Algoritma C4.5 dan Adaboost (Studi Kasus: STMIK XYZ). *CSRID (Computer Science Research and Its Development)*, 9(1), 1-11.
- Andriani, A. (2017). Hubungan Mutu Pelayanan Kesehatan dengan Kepuasan Pasien di Ruang Poliklinik Umum Puskesmas Bukittinggi. *Journal Endurance*, 2(1), 45-52.

- Arunadevi, J., Ramya, S., & Raja, M. (2018). A Study of Classification Algorithms using Rapidminer. *International Journal of Pure and Applied Mathematics*, 119(12), 15977-15988.
- Han, J., Kamber, M., & Pei, J. (2019). *Data Mining: Concept And Techniques 3rd Ed.* Morgan Kaufman Publisher.
- Haryati, S., Sudarsono, A., & Suryanan, E. (2015). Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5. *Jurnal Media Infotama*, 11(2), 130-138.
- Hernadewita, Sambani, E., & Nuaeni, F. (2018). Penerapan Algoritma C4.5 untuk Klasifikasi Pola Penjurusan di Sekolah Menengah Kejuruan (SMK) Kota Tasikmalaya. *CSRID (Computer Sci. Res. Its Dev Journal)*, 9(3), 144.
- Muhamad, M., Windarto, A. P., & Suhada, S. (2019). Penerapan Algoritma C4.5 pada Klasifikasi Potensi Siswa Drop Out. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3(1).
- Muningsih, E., & Kiswati, S. (2015). Penerapan metode K-means untuk clustering produk online shop dalam penentuan stok barang. *Bianglala Informatika*, 3(1).
- Nas, D. (2020). Data Mining Pengelompokan Bidang Keahlian Mahasiswa Menggunakan Algoritma K-Means (Studi Kasus: Universitas CIC Cirebon). *Syntax: Jurnal Informatika*, 09(2), 1-14.
- Nurajizah, S. (2019). Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori. *INOVTEK Polbeng-Seri Informatika*, 4(1), 35-44.
- Purwanto, A., & A., D. E. (2018). Perbandingan Minat Siswa SMU pada Metode Klasifikasi Menggunakan 5 Algoritma. *Ikraith-Informatika*, 2(1), pp. 43-47.
- Rismayanti. (2015). Implementasi Algoritma C4.5 untuk Menentukan Penerima Beasiswa di STT Harapan Medan. *Jurnal Media Infotama*, 11(2), 130-138.
- Subiyanto. (2017). Sistem Rekomendasi Penjurusan Sekolah Menengah Kejuruan dengan Algoritma C4.5. *Jurnal Kependidikan*, 1(1), 139-149.
- Sun, J., Lang, J., Fujita, H., & Li, H. (2018). Imbalanced enterprise credit evaluation with DTE-SBD: Decision Tree Ensemble Based on SMOTE and Bagging with Differentiated Sampling Rates. *Information Sciences*, 76-91.