

# SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN PROGRAM STUDI CALON PESERTA DIDIK MENGGUNAKAN METODE POHON KEPUTUSAN C4.5

(Studi Kasus : Prodi Universitas Amikom Yogyakarta)

Moh Royandi Azkia<sup>1)</sup>, Mayadi<sup>2)</sup>, Miftahul Huda<sup>3)</sup>, Kusri<sup>4)</sup>

<sup>1), 2), 3), 4)</sup> Magister Teknik Informatika, Universitas AMIKOM Yogyakarta  
Jl. Ring Road Utar, Condong Catur, Depok, Sleman, Yogyakarta 55281  
Email : royandiazkia@gmail.com<sup>1)</sup>, kusri@amikom.ac.id<sup>4)</sup>

## Abstrak

Efek dari kesalahan dalam memilih program studi akan berdampak buruk bagi kelangsungan proses belajar peserta didik. Untuk meminimalisir permasalahan tersebut, penulis bermaksud memberi usulan dengan membangun sistem pendukung keputusan untuk memilih program studi berdasarkan minat dan kemampuan calon peserta didik. Sistem yang dibangun menggunakan studi kasus pada program studi di Universitas Amikom Yogyakarta.

Sistem pendukung keputusan ini akan menerapkan model pohon keputusan dengan menggunakan algoritma C4.5 untuk membuat pohon keputusannya. Sedangkan data yang di mining berasal dari data mahasiswa Universitas Amikom Yogyakarta yang telah memilih program studi. Data-data tersebut sebelumnya dimining terlebih dahulu melewati proses preprocessing dan transformation agar lebih terstruktur yang siap diolah dengan algoritma C4.5. Algoritma C4.5 dinilai cocok untuk diterapkan pada sistem ini karena memiliki tingkat akurasi yang cukup efektif terhadap masalah seperti ini.

Dari hasil mining ditemukan faktor yang menjadi node pertama pada pohon keputusan yaitu minat. Untuk tingkat akurasi dari pohon keputusan yang dihasilkan sebesar 68%. Angkan tersebut didapat setelah melakukan pengujian dengan data testing menggunakan metode K-Fold Cross Validation.

**Kata kunci:** SPK, Klasifikasi, Pohon Keputusan, Algoritma C.45, Pemilihan Jurusan. K-Fold Cross Validation

## 1. Pendahuluan

Proses pendidikan pada hakikatnya adalah salah satu proses pembinaan sumber daya manusia yang ditekankan pada upaya pengembangan aspek-aspek pribadi peserta didik dari segi jasmani maupun rohani. Namun terdapat permasalahan yang sering dihadapi oleh peserta didik ketika baru saja menyelesaikan pendidikan sekolah menengah atas (SMA) – sederajat yang hendak melanjutkan pendidikan mereka ke perguruan tinggi. Para calon peserta didik ini dihadapkan dengan

pemilihan program studi yang akan mereka jalani sebagai bahan untuk meniti karir dimasa depan.

Pemilihan program studi ini sangat penting untuk diperhatikan karena menjadi salah satu faktor penyebab keterlambatan lulus dari mahasiswa itu sendiri. Kurangnya minat pada pilihan program studi akan menyebabkan peserta didik menjadi malas untuk kuliah dan motivasi belajar kurang maksimal sehingga membuat peserta didik tersebut sering bermalas-malasan yang secara tidak langsung akan berakibat pada keterlambatan kelulusan [1].

Penerapan sistem informasi untuk mengatasi masalah tersebut menjadi salah satu solusi yang dapat memberi hasil yang menjanjikan. Sebagaimana pada penerapan sistem informasi dibidang lain yang menunjukkan perubahan positif yang cukup signifikan. Contohnya penerapan sistem informasi transaksi pada sebuah perusahaan jual beli yang tentunya akan sangat membantu dalam hal pencatatan transaksi barang dan peramalan inventori.

Pada kasus ini sistem informasi yang di jadikan sebagai usulan untuk menyelesaikan masalah adalah sistem pendukung keputusan. Scott morton mendefinikasi sistem pendukung keputusan sebagai sistem berbasis komputer interaktif, yang bertujuan membantu proses pengambilan keputusan dalam menyelesaikan masalah-masalah tidak terstruktur menggunakan data dan berbagai model. Konsep kerja sistem ini yaitu dengan memadukan sumber daya intelektual dari individu dengan kapabilitas komputer untuk menentukan kualitas keputusan [2].

Salah satu model dari sistem pendukung keputusan adalah Pohon keputusan. Pohon keputusan (*Decision Tree*) digunakan dalam metode clasification dan pengelompokan. *Decision Tree* memecahkan masalah menjadi subset yang semakin diskrit, bekerja dari infomasi yang lebih umum kemudian ke yang semakin spesifik [3]. Sedangkan algoritma untuk membentuk *Decision Tree* bermacam-macam seperti C4.5, C5.0, ID3, CHART [4].

Hssina B dkk, melakukan penelitian pada *Decision Tree* untuk menguji perbandingan antara algoritma ID3 dan C4.5 dengan tujuan mencari algoritma terbaik untuk machine learning. Pada proses eksperimen yang dilakukan dengan membandingkan ID3/C4.5, C4.5/C5.0 dan C5.0/CART menghasilkan kesimpulan bahwa algoritma C4.5 adalah algoritma yang paling cocok untuk machine learning [5].

Mazid MM dkk, pernah mengusulkan untuk memperbaiki algoritma klasifikasi *rulebase* C4.5 dengan tujuan meningkatkan akurasi dalam mengklasifikasi. Pada percobaan ini dilakukan dengan menekan pengurangan *input space* dengan menggunakan *entropy* dan beberapa rumus koefisien korelasi, metode yang diusulkan ini menunjukkan kinerja yang lebih baik untuk setiap file data. Namun, secara menyeluruh hasil dari setiap perbaikan yang dilakukan menunjukkan kinerja yang tidak lebih baik dari algoritma C4.5 asli. Hal itu serupa seperti hasil perbaikan yang telah dilakukan oleh Spearman dan Kendall. Artinya algoritma C4.5 asli merupakan sebuah algoritma yang sudah teruji akurasi [6].

Kemudian dalam hubungan dengan masalah yang diangkat pada makalah ini, Zhiwu Liu dan Xiuzhi Zhang, telah melakukan penelitian menggunakan algoritma C4.5 untuk mengetahui faktor-faktor yang mempengaruhi kebiasaan belajar dan perilaku yang buruk pada siswa sehingga kedepannya bisa diperbaiki. Selain itu juga untuk melihat efek dari pengajaran kepada oleh guru. Pada penelitiannya mereka memanfaatkan sifat prediksi yang dimiliki *Decision Tree*. Dengan demikian algoritma C4.5 dinilai cocok untuk mengatasi masalah ini [7].

Sangat pentingnya mempertimbangkan minat, bakat, proyeksi pertumbuhan dalam memilih karir. Hal ini sering terlihat pada siswa memiliki catatan akademis mereka yang buruk karena memilih karir tanpa mempertimbangkan kemampuan yang akan menyebabkan pemborosan waktu dan uang. Untuk mengatasi hal tersebut, Katore L.S dkk membangun sebuah sistem pendukung keputusan siswa dalam memilih karir yang tepat sesuai dengan psikologis siswa itu sendiri. Dengan memanfaatkan algoritma C4.5. Keakuratan sistem yang dicapai adalah 86%.[8]. Sedangkan perbedaan dengan penelitian yang dilakukan pada makalah ini adalah dari segi pengambilan variable yang didasarkan pemilihan jurusan dengan melihat minat dan kemampuan calon peserta didik.

Dengan uraian diatas disimpulkan bahwa untuk mengatasi permasalahan tersebut. Perlu dibangun sebuah sistem pendukung keputusan menggunakan model pohon keputusan dengan algoritma C4.5 dalam menentukan jurusan/program studi peserta didik sesuai minat dan bakatnya masing-masing.

## 2. Pembahasan

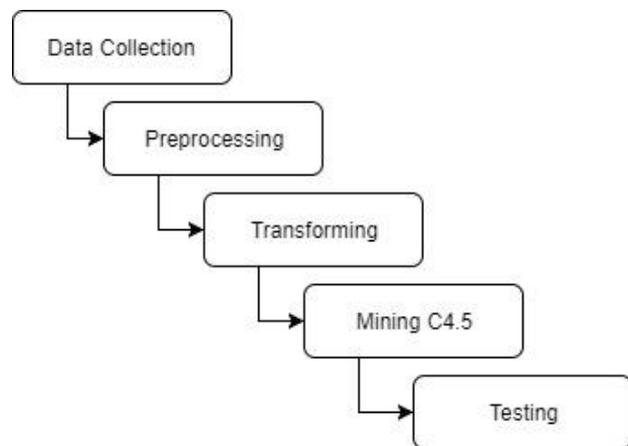
### Studi Kasus

Penelitian ini menggunakan studi kasus program studi S1 pada Universitas Amikom Yogyakarta. Program studi tersebut antara lain Informatika, Sistem Informasi, Teknologi Informasi, Teknik Komputer, Arsitektur, Perencanaan Wilayah dan Kota, Geografi, Kewirausahaan, Ekonomi, Akuntansi, Ilmu Pemerintahan, Ilmu Komunikasi, Hubungan Internasional.

Tetapi pada penelitian ini program studi yang diambil sebagai tujuan rekomendasi sistem pendukung keputusan adalah program studi Informatika dan Sistem Informasi.

### Metode Penelitian

Metode penelitian yang digunakan pada makalah ini adalah eksperimen dengan alur seperti Gambar 1.



Gambar 1. Alur Penelitian

### Data Collection

Pada tahap ini dilakukan pengumpulan data yang akan digunakan untuk proses mining. Dari pengumpulan data diperoleh mahasiswa sebesar 91 record. Pada Gambar 2 merupakan bentuk data asli yang didapat langsung dari hasil survei lapangan.

### Preprocessing

Pada tahap ini merupakan awal dalam mengolah data input sebelum memasuki proses tahapan utama. Disini data akan diproses dari bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan proses mining yang lebih lanjut. Pada tahap ini juga data di bagi menjadi dua bagian yaitu kelompok data training dan data testing untuk keperluan pengujian.

Perubahan yang data yang dimaksud adalah pada kesesuaian dengan atribul yang dipakai dalam proses mining. Pada Tabel 1 adalah contoh preprocessing yang dilakukan pada data yang digunakan.

Tabel 1. Preprocessing Data

Data Asli	Data Siap Mining
Teknik Komputer Jaringan	TKJ
Rekayasa Perangkat Lunak	RPL
Audio Visual, Musik	Multimedia
Ilmu Pengetahuan Alam	IPA
Ilmu Pengetahuan Sosial	IPS

Setelah melalui proses preprocessing data yang awal berjumlah 91 record di saring menjadi 80 buah record karena data mahasiswa yang mengambil prodi selain informatika dan sistem informasi kurang dari 10 record. Penghilangan data tersebut karena dinilai akan mengurangi efisiensi pohon keputusan yang terbentuk karena terlalu minimnya data pada prodi yang lain.

**Transforming**

Tahap ini merupakan tahap optional, artinya proses transforming ini akan dijalankan jika dibutuhkan saja. Pada tahap ini dilakukan klasifikasi terhadap data dengan tujuan untuk mempermudah proses mining. Contoh proses transforming pada Table 2.

Table 2. Contoh Transforming Nilai Matematika

Nilai	Transforming
n <= 50	Rendah
50 < n > 80	Sedang
80 <= n	Tinggi

**Algoritma C4.5**

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk sebuah pohon keputusan. Algoritma ini merupakan pengembangan dari algoritma ID3 [4].

Tahapan pembentukan pohon keputusan C4.5 sebagai berikut :

1. Menyiapkan data training
2. Menentukan akar pohon.
3. Hitung nilai Gain dengan persamaan (1).

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

Keterangan :

- S = Himpunan Kasus
- A = Atribut
- n = Jumlah Partisi Atribut A
- |Si| = Jumlah Kasus pada Partisi ke-i
- |S| = Jumlah Kasus dalam S

Untuk menghitung nilai Gain dibutuhkan nilai Entropy, persamaan (2) menghitung nilai Entropy sebagai berikut :

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

Keterangan :

- S = Himpunan Kasus
- A = Atribut
- n = Jumlah Partisi S
- pi = Proporsi dari Si terhadap S

4. Ulangi proses ke-2 untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama, maka proses perulangan berhenti.

**Pemilihan Variabel**

Adapun variable yang menjadi tolak ukur untuk melihat minat dan bakat calon peserta didik beserta faktor pendukung yang di ambil berdasarkan kesesuaian dengan kriteria program studi yang ada pada Universitas Amikom Yogyakarta sebagai berikut :

1. Jenis Kelamin

Di Amikom Yogyakarta jenis kelamin menjadi salah satu faktor yang dapat mempengaruhi peserta didik dalam memilih jurusan. Hal tersebut dapat dilihat dari presentase mahasiswa dalam memilih prodi Teknik Informatika yang hampir 80% adalah laki-laki, begitu pun sebaliknya untuk jurusan Sistem Informasi peserta didik berjenis kelamin perempuan mendominasi hampir 65%.

1	nama	jenis_kelamin	asa_sekolah	jurusan	k_berhitung	k_logika	k_menggambar	c_berbahasa	minat	saran_ortu	prodi
2	Heri Heri	Laki-laki	SMK	TKJ	Tinggi	Sedang	Sedang	Sedang	Teknologi	Tidak ada saran	Informatika
3	Heri Heri	Laki-laki	RMA	IPA	Tinggi	Tinggi	Sedang	Relebr	Teknologi	Infomatika	Informatika
4	Heri Heri	Perempuan	SMK	RP	Sedang	Sedang	Sedang	Sedang	Teknologi	Sistem Informasi	Sistem Informasi
5	Heri Heri	Laki-laki	Sekolah Menengah Atas	IPA	Sedang	Sedang	Rendah	Tinggi	Teknologi	Tidak ada saran	Informatika
6	Heri Heri	Perempuan	Sekolah Menengah Atas	Multimedia	Sedang	Sedang	Sedang	Tinggi	Teknologi	Tidak ada saran	Informatika
7	Heri Heri	Perempuan	SMK	Multimedia	Sedang	Sedang	Sedang	Sedang	Komunikasi	Akademik	Sistem Informasi
8	Heri Heri	Perempuan	SMK	Multimedia	Tinggi	Sedang	Rendah	Sedang	Komunikasi	Tidak ada saran	Sistem Informasi
9	Heri Heri	Laki-laki	Sekolah Menengah Atas	IPA	Tinggi	Tinggi	Rendah	Relebr	Teknologi	Platoni	Informatika
10	Heri Heri	Laki-laki	SMK	TKJ	Sedang	Sedang	Sedang	tinggi	Bahasa / sng (perancis)	Tidak ada saran	Sistem Informasi
11	Heri Heri	Laki-laki	MA	IPA	Tinggi	Tinggi	Sedang	Rendah	Teknologi	Sistem Informasi	Sistem Informasi
12	Heri Heri	Perempuan	SMK	RP	Sedang	Sedang	Rendah	Relebr	Teknologi	Infomatika	Informatika
13	Heri Heri	Perempuan	Sekolah Menengah Atas	IPS	Sedang	Sedang	Hendah	tinggi	Komunikasi	Kawruasaan	Kewruasaan
14	Heri Heri	Perempuan	MA	IPA	Sedang	Sedang	Sedang	Sedang	Komunikasi	Sistem Informasi	Sistem Informasi
15	Heri Heri	Perempuan	SMK	RP	Sedang	Sedang	Sedang	Relebr	Teknologi	Sistem Informasi	Sistem Informasi
16	Heri Heri	Perempuan	SMK	IPA	Sedang	Sedang	Hendah	Sedang	Teknologi	Sistem Informasi	Sistem Informasi
17	Heri Heri	Perempuan	SMK	RP	Sedang	Sedang	Tinggi	Tinggi	Teknologi	Sistem Informasi	Sistem Informasi
18	Heri Heri	Laki-laki	SMK	TKJ	Sedang	Sedang	Sedang	Relebr	Teknologi	Infomatika	Informatika
19	Heri Heri	Laki-laki	MA	IPA	tinggi	Sedang	Sedang	tinggi	Teknologi	Informatika	Informatika
20	Heri Heri	Laki-laki	MA	IPA	Sedang	Sedang	Sedang	Sedang	Teknologi	Informatika	Informatika
21	Heri Heri	Perempuan	RMA	IPS	Sedang	Sedang	Tinggi	Tinggi	Teknologi	Sistem Informasi	Sistem Informasi
22	Heri Heri	Perempuan	SMK	IPS	Sedang	Sedang	tinggi	tinggi	Teknologi	Tidak ada saran	Sistem Informasi
23	Heri Heri	Laki-laki	MA	IPS	Sedang	Sedang	Rendah	Sedang	Teknologi	Sistem Informasi	Sistem Informasi
24	Heri Heri	Perempuan	SMK	Pertanian	Sedang	Sedang	Rendah	Rendah	Komunikasi	Sistem Informasi	Sistem Informasi
25	Heri Heri	Perempuan	SMK	IPA	Sedang	Hendah	Hendah	Sedang	Manajemen	Informatika	Informatika
26	Heri Heri	Laki-laki	MA	IPA	Sedang	Sedang	Rendah	Sedang	Teknologi	Teorik Komputer	Informatika
27	Heri Heri	Perempuan	Sekolah Menengah Atas	IPS	Sedang	Rendah	Relebr	Relebr	Teknologi	Tidak ada saran	Teknologi Informasi
28	Heri Heri	Perempuan	Sekolah Menengah Atas	IPS	Rendah	Rendah	Rendah	Rendah	Ekonomi	Ilmu Pemerintahan	Ilmu Pemerintahan

Gambar 2. Bentuk Data Asli

Variabel jenis kelamin ini terdiri dari 2 data value yaitu Laki-laki dan Perempuan.

2. Asal Sekolah

Pemilihan prodi berdasarkan asal sekolah tentu berpengaruh terhadap minat calon peserta didik. Sebagai contoh : A merupakan lulusan dari SMK X (Kejuruan Komputer). Jika A melanjutkan prodi ke perguruan tinggi presentase memilih jurusan yang berhubungan erat dengan komputer pasti akan dipilih.

Variabel asal sekolah ini terdiri dari 3 data value yaitu SMA, SMK, Sekolah Menengah Atas Agama.

3. Jurusan semasa Sekolah

Di Amikom Yogyakarta jurusan semasa sekolah juga berpengaruh terhadap pemilihan prodi. Hal itu karena disebagian prodi pada memiliki syarat tertentu untuk bisa memilih prodi. Contoh : Calon peserta didik yang boleh mengambil prodi informatika, jika berasal dari SMK harus yang berjurusan RPL dan TKJ, sedangkan yang dari SMA harus yang berjurusan IPA.

Variabel jurusan ini terdiri dari 9 data value yaitu IPA, IPS, RPL, TKJ, Multimedia, Perkantoran, Otomotif, Pertambangan, Tarbiyah.

4. Kemampuan Berhitung

5. Kemampuan Logika

6. Kemampuan Menggambar

7. Kemampuan Berbahasa

Untuk poin 4,5,6,7 ditujukan untuk melihat bakat peserta didik. Karna ke-empat variable tersebut dinilai sangat berpengaruh terhadap setiap program studi di Amikom yogyakarta.

Variabel kemampuan berhitung, logika, menggambar dan berbahasa ini terdiri dari 3 data value yaitu Rendah, Sedang dan Tinggi.

8. Minat Bidang

Variable ini digunakan untuk melihat langsung minat yang diinginkan oleh calon peserta didik.

Variabel minat bidang ini terdiri dari 3 data value yaitu Teknologi, Arsitektur, Ekonomi, Komunikasi, Biologi, Sosial dan Politik dan Geografi.

9. Jurusan Saran Orang Tua

Saran orang tua juga menjadi pertimbangan dalam memilih prodi. Hal tersebut dikarenakan adanya kemungkinan calon peserta didik sudah di arahkan oleh orang tuanya untuk masa depan yang cerah tentunya. Karna orang tua pasti memilih yang terbaik untuk anaknya.

Variabel saran orang tua ini terdiri dari 12 data value yaitu Informatika, Sistem Informasi, Teknik Komputer, Teknologi Informasi, Ekonomi, Akutansi, Kewirausahaan, Arsitektur, Perencanaan Wilayah dan Tata Kota, Ilmu Pemerintahan, Hubungan Internasional dan Tidak ada saran.

Variable-variabel tersebut kemudian dihitung berdasarkan data training untuk membuat pohon keputusan yang dipakai dalam rekomendasi pemilihan jurusan oleh calon peserta didik baru.

Data Training

Pada Gambar 3 menunjukkan bentuk data training yang siap untuk dimining dengan algoritma C4.5 setelah melalui tahap preprocessing dan transforming. Jumlah data training yang digunakan untuk membentuk pohon keputusan berjumlah 80% dari total data asli. Artinya dari 80 record data, 65 diantaranya digunakan sebagai data training.

NO	NAMA	JK	ASAL SEKOLAH	JURUSAN	BERHITUNG	LOGIKA	MENGGAMBAR	BERBAHASA	MINAT	SARAN	PRODI
1		Laki-laki	SMK	Multimedia	Sedang	Sedang	Tinggi	Sedang	Teknologi	Tidak ada saran	Informatika
2		Laki-laki	SMK	KPL	Sedang	Sedang	Tinggi	Sedang	Teknologi	Tidak ada saran	Informatika
3		Laki-laki	SMA	IPS	Rendah	Rendah	Rendah	Sedang	Ekonomi	Ekonomi	Ekonomi
4		Laki-laki	SMA	IPS	Rendah	Sedang	Tinggi	Sedang	Arsitektur	Informatika	Informatika
5		Perempuan	SMA	IPS	Rendah	Tinggi	Sedang	Tinggi	Arsitektur	Arsitektur	Informatika
6		Laki laki	SMA	IPA	Tinggi	Tinggi	Sedang	Rendah	Teknologi	Sistem Informasi	Sistem Informasi
7		Laki-laki	SMK	RPL	Rendah	Tinggi	Rendah	Sedang	Teknologi	Teknik Komputer	Sistem Informasi
8		Laki-laki	SMK	KPL	Sedang	Tinggi	Rendah	Sedang	Teknologi	Tidak ada saran	Informatika
9		Perempuan	Sekolah Menengah Atas Agama	IPS	Rendah	Rendah	Rendah	Rendah	Ekonomi	Ilmu Pemerintahan	Sistem Informasi
10		Perempuan	SMA	IPS	Rendah	Sedang	Sedang	Sedang	Ekonomi	Teknologi Informasi	Informatika
11		Laki laki	SMA	IPS	Sedang	Sedang	Rendah	Sedang	Teknologi	Sistem Informasi	Sistem Informasi
12		Laki laki	SMK	Otomotif	Rendah	Sedang	Sedang	Rendah	Teknologi	Teknik Komputer	Informatika
13		Perempuan	SMK	Multimedia	Sedang	Sedang	Sedang	Sedang	Komunikasi	Ilmu Komunikasi	Sistem Informasi
14		Laki-laki	SMK	TKJ	Sedang	Sedang	Tinggi	Sedang	Teknologi	Informatika	Informatika
15		Perempuan	SMA	IPA	Sedang	Rendah	Rendah	Sedang	Manajemen	Informatika	Informatika

Gambar 3. Data Siap di Mining

### Data Testing

Data testing yang digunakan untuk mengukur akurasi dari pohon keputusan yang terbentuk adalah 20% dari data training yang dipilih secara random. Artinya dari 80 record data, 15 diantaranya digunakan sebagai data testing.

### Hasil Pohon Keputusan

Berikut merupakan pohon keputusan yang terbentuk setelah dilakukan proses mining pada data training, di temukan bahwa yang menjadi node pertama adalah minat.

minat = Teknologi (Informatika = 29, Sistem Informasi = 5) : Informatika

minat = Arsitektur (Informatika = 7, Sistem Informasi = 1) : Informatika

minat = Ekonomi (Informatika = 1, Sistem Informasi = 5) : ?

| asal\_sekolah = Sekolah Menengan Atas Agama (Informatika = 0, Sistem Informasi = 0) : Kosong

| asal\_sekolah = SMA (Informatika = 1, Sistem Informasi = 5) : Sistem Informasi

| asal\_sekolah = SMK (Informatika = 0, Sistem Informasi = 0) : Kosong

minat = Sosial dan Politik (Informatika = 0, Sistem Informasi = 3) : Sistem Informasi

minat = Komunikasi (Informatika = 3, Sistem Informasi = 4) : ?

| k\_berbahasa = Tinggi (Informatika = 3, Sistem Informasi = 1) : Informatika

| k\_berbahasa = Rendah (Informatika = 0, Sistem Informasi = 1) : Sistem Informasi

| k\_berbahasa = Sedang (Informatika = 0, Sistem Informasi = 2) : Sistem Informasi

minat = Bahasa Asing (Informatika = 0, Sistem Informasi = 0) : Kosong

### Testing

Untuk melihat seberapa akurat kemampuan prediksi yang dimiliki oleh pohon keputusan yang dihasilkan. Pada tahap ini dilakukan pengujian dengan cara mencocokkan dengan menggunakan pohon keputusan dengan data testing yang sudah disediakan sebelumnya. Proses pengujian ini dilakukan secara manual dengan hasil pada Tabel 3.

**Tabel 3. Potongan Record Data Hasil Manual Testing**

Asal Sekolah	Berbahasa	Minat	Prodi	
SMK	Sedang	Teknologi	Informatika	BENAR
SMA	Sedang	Teknologi	Informatika	BENAR
SMK	Sedang	Teknologi	Sistem Informasi	SALAH
Sekolah	Tinggi	Teknologi	Informatika	BENAR

Menengah Atas Agama				
Sekolah Menengah Atas Agama	Tinggi	Teknologi	Informatika	BENAR
SMK	Sedang	Komunikasi	Sistem Informasi	BENAR
SMK	Sedang	Komunikasi	Sistem Informasi	BENAR
Sekolah Menengah Atas Agama	Sedang	Teknologi	Informatika	BENAR
SMK	Tinggi	Bahasa Asing	Sistem Informasi	Kosong
SMA	Rendah	Teknologi	Sistem Informasi	SALAH
SMK	Sedang	Teknologi	Informatika	BENAR
SMA	Sedang	Komunikasi	Sistem Informasi	BENAR
SMK	Sedang	Teknologi	Sistem Informasi	SALAH
SMA	Sedang	Teknologi	Sistem Informasi	SALAH
SMK	Tinggi	Teknologi	Sistem Informasi	SALAH

Sedangkan untuk melihat tingkat akurasi yang dihasilkan oleh pohon keputusan yang terbentuk digunakan metode *K-Fold Cross Validation* dengan iterasi sebanyak 10 kali ( $K=10$ ). Metode pengujian tersebut merupakan metode standar yang dilakukan untuk memprediksi *error rate*. Data training dan testing dibagi secara random menjadi 10 bagian dengan perbandingan 80% data untuk data training dan 20% data untuk data testing. Setelah diimplementasikan setiap bagian dengan sistem yang telah dibangun barulah dihitung *error rate* yang muncul pada tiap bagian. Selanjutnya dari *error rate* tersebut dihitung rata-ratanya untuk mendapatkan nilai *error rate* secara keseluruhan. Rincian keberanan perhitungan *error rate* dapat dilihat pada Tabel 4.

**Tabel 4. K-Fold Cross Validation**

Iterasi	Benar	Salah	Kosong	Akurasi (%) $\frac{\sum \text{Benar}}{\sum \text{Benar} + \sum \text{Salah}} \times 100$
1	10	4	1	71,42857143
2	7	6	2	53,84615385
3	9	5	1	64,28571429
4	10	5	0	66,66666667
5	8	3	4	72,72727273
6	10	3	2	76,92307692

**Tabel Lanjutan. (Tabel 4)**

7	8	6	1	57,14285714
8	8	5	2	61,53846154
9	8	3	4	72,72727273
10	12	2	1	85,71428571
<b>Rata-rata Akurasi (%)</b>				<b>68,3000333</b>

Dari pengujian dengan menggunakan *K-Fold Cross Validation* didapatkan hasil perhitungan rata-rata seperti pada Tabel 4, bahwa akurasi yang dihasilkan oleh sistem pendukung keputusan pemilihan program studi menggunakan algoritma C4.5 ini sebesar 68%. Perhitungan tersebut dilakukan hanya untuk keputusan yang bernilai benar dan salah saja, untuk yang bernilai kosong tidak dihitung karna sistem tidak memberi keputusan.

### 3. Kesimpulan

Dari hasil mining dengan sistem pendukung keputusan menggunakan algoritma C.45 yang telah dibangun, menghasilkan pohon keputusan dengan tingkat akurasi sebesar 68%. Angka tersebut didapatkan setelah melakukan pengujian dengan metode *K-Fold Cross Validation* dengan nilai  $K = 10$ .

Sebagai saran untuk penelitian kedepan untuk meningkatkan akurasi pohon keputusan dibutuhkan data training dan data testing dengan jumlah yang lebih banyak lagi. Sedangkan untuk pengambilan variabel dapat disesuaikan dengan studi kasus yang diambil untuk penelitian berikutnya.

### Daftar Pustaka

- [1] F.R. Wahyudi, "Faktor-Faktor Penyebab Mahasiswa Mahasiswa S1 Pendidikan Teknik Bangunan Tidak Lulus Tepat Waktu", *Skripsi*, Program Studi Pendidikan Teknik Bangunan, Fakultas Teknik, Universitas Negeri Malang, 2012
- [2] Turban E, Aronson J.E, Liang T.P., *Decision Support Systems and Intelligent Systems*, 7th Ed. Prentice-Hall: New Delhi, 2005
- [3] S. Gollapudi, *Practical Machine Learning*, Packt Publishing, 2016
- [4] Kusriani, E.T. Luthfi, *Algoritma Data Mining*, Andi Offset: Yogyakarta, 2009
- [5] Hssina B, Merbouha A, Ezzikouri H, Erritali M., "A Comparative Study of Decision Tree ID3 and C4.5." in *International Journal of Advanced Computer Science and Applications*, ResearchGate, Juli 2014
- [6] Mazid M.M, Ali A.B.M.S, Tickle K.S., "Improved C4.5 Algorithm for Rule Based Classification". In *Proceedings Of The 9th Wseas International Conference On Applied Computer Science*, wseas.us, 2010
- [7] Z. Liu, X. Zhang, "Prediction and Analysis for Students' Marks Based on Decision Tree Algorithm", in *Third International Conference on Intelligent Networks and Intelligent Systems (ICINIS)*, IEEE Xplore, 2010
- [8] Katore L.S, Ratnaparkhi B.S, Umale J.S., "Novel Professional Career prediction and recommendation method for individual through analytics on personal Traits using C4.5 Algorithm". In *Proceedings Of Global Conference on Communication Technologies*, IEEE Xplore, 2015

### Biodata Penulis

**Moh Royandi Azkia**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika Universitas Amikom Yogyakarta, lulus tahun 2015. Saat ini sedang menempuh pendidikan di Magister Teknik Informatika Universitas Amikom Yogyakarta.

**Mayadi**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK Bumigora Mataram,

lulus tahun 2016. Saat ini sedang menempuh pendidikan di Magister Teknik Informatika Universitas Amikom Yogyakarta.

**Miftahul Huda**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK Wicidia Samarinda, lulus tahun 2015. Saat ini sedang menempuh pendidikan di Magister Teknik Informatika Universitas Amikom Yogyakarta.

**Kusriani**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Ilmu Komputer Universitas Gadjah Mada, lulus tahun 2002. Memperoleh gelar Master Komputer (M.Kom), Jurusan Ilmu Komputer Universitas Gadjah Mada, lulus tahun 2006. Memperoleh gelar Doktor (Dr), Jurusan Ilmu Komputer Universitas Gadjah Mada, lulus tahun 2010. Saat ini menjadi Dosen dan Direktur Program Pascasarjana di Universitas Amikom Yogyakarta.