

## KLASIFIKASI MASA TUNGGU ALUMNI UNTUK MENDAPATKAN PEKERJAAN MENGGUNAKAN ALGORITMA C4.5 (Studi Kasus: Fasilkom Unsika)

**Febbyola Rezkika<sup>\*1</sup>, Betha Nurina Sari<sup>2</sup>, Agung Susilo Yuda Irawan<sup>3</sup>**

<sup>1,2,3</sup> Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang  
<sup>1,2,3</sup>Jl. HS. Ronggo Waluyo, Teluk Jambé Timur, Karawang. Telp. 081295094228

<sup>\*</sup>Email Corresponding Author: febbyola.rezkika17097@student.unsika.ac.id

### **Abstrak**

*Pelacakan alumni perguruan tinggi perlu dilakukan untuk mengetahui bagaimana status pekerjaan mereka saat ini, khususnya waktu tunggu dalam mendapatkan pekerjaan, sebagai salah indikator kualitas lulusan perguruan tinggi. Dukungan teknologi informasi seperti data mining dapat digunakan sebagai tools untuk menghasilkan suatu pengetahuan. Penelitian ini bertujuan untuk memprediksi waktu tunggu alumni mendapatkan pekerjaannya dengan menggunakan algoritma decision tree (C4.5) tunggal dan dibandingkan algoritma decision tree (C4.5) dengan fitur forward selection. Pengolahan data menggunakan algoritma C4.5 dengan bantuan software RapidMiner Studio. Hasilnya menunjukkan bahwa algoritma decision tree (C4.5) dengan fitur forward selection meraih performa terbaik dengan nilai accuracy 80,37%, precision 79,56%, recall 81,34%, f-measure 80,40% dan AUC 0.914 yang termasuk ke dalam kategori excellent classification. Dengan demikian, algoritma C4.5 berbasis Forward Selection terbukti dapat meningkatkan tingkat akurasi, dibandingkan dengan algoritma decision tree (C4.5) tunggal, yang ditandai dengan peningkatan nilai akurasi sebesar 25,93%*

**Kata kunci:** Data Mining, Decision Tree, C4.5, Klasifikasi, Waktu Tunggu Alumni

### **Abstract**

*Tracking college graduates needs to be done to find out how their current job status is, especially the waiting time to get a job, as an indicator of the quality of college graduates. Information technology support such as data mining can be used as tools to generate knowledge. This study aims to predict the waiting time for alumni to get a job using a single decision tree algorithm (C4.5) and compare the decision tree algorithm (C4.5) with the forward selection feature. Data processing uses the C4.5 algorithm with the help of RapidMiner Studio software. The results show that the decision tree algorithm (C4.5) with the forward selection feature achieves the best performance with 80.37% accuracy, 79.56% precision, 81.34% recall, 80.40% f-measure and 0.914 AUC which includes into the excellent classification category. Thus, the C4.5 algorithm based on Forward Selection is proven to increase the level of accuracy, compared to a single decision tree (C4.5) algorithm, which is characterized by an increase in the accuracy value of 25.93%.*

**Keywords:** Data Mining, Decision Tree, C4.5, Classification, Graduates Waiting Time

### **1. Pendahuluan**

Perguruan Tinggi (PT) adalah suatu lembaga pendidikan taraf akhir yang mencetak lulusan yang nantinya akan memasuki dunia kerja. Persaingan dunia global kerja yang semakin kompleks membawa setiap perguruan tinggi pada suatu konflik yang sama, yakni seberapa relevankah keluaran perguruan tinggi terhadap kebutuhan pengguna lulusan perguruan tinggi waktu ini. Hal penting yang dihadapi institusi pendidikan tinggi di Indonesia saat ini yaitu persaingan global. Akibat dari persaingan global tersebut muncul suatu tantangan dalam bidang pendidikan, termasuk perguruan tinggi. Mendapatkan pekerjaan merupakan salah satu perihalnya yang akan dicapai oleh alumni setelah menyelesaikan masa kuliahnya. Akan tetapi pada

prosesnya perlu memerlukan waktu tunggu dalam mendapatkan pekerjaan tersebut. Dimana salah satu tolak ukur keberhasilan suatu perguruan tinggi dalam bidang pendidikan adalah terserapnya alumni di dunia kerja [1]. Setiap perguruan tinggi tentu memiliki strategi yang berbeda guna meningkatkan mutululusannya untuk memenuhi kebutuhan dunia kerja. Oleh karena itu, untuk mengetahui lulusan mahasiswa dari suatu perguruan tinggi telah memperoleh pekerjaan dilihat dari data *tracer study*.

*Tracer study* diadakan tiap tahunnya oleh setiap perguruan tinggi guna memenuhi keperluan data akreditasi, perbaikan dalam proses pembelajaran serta pengembangan kurikulum yang diwajibkan oleh Dirjen Pembelajaran Kemahasiswaan. Alumni mahasiswa Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang dari tahun 2016 sampai tahun 2020 berjumlah 1.101 orang. Data *tracer study* diperoleh dari hasil kuesioner yang dibagikan kepada alumni yang nantinya diolah untuk mengetahui masa tunggu alumni dalam memperoleh pekerjaan setelah lulus. Pada dasarnya, semakin cepat alumni memperoleh pekerjaan bisa dikatakan mutu mahasiswa yang termasuk baik.

*Data mining* melakukan proses penggalian pengetahuan dan informasi yang berguna termasuk pola, asosiasi, perubahan anomali dan struktur signifikan dari data transaksi tersimpan dalam basis data, data *warehouse*, atau tempat penyimpanan informasi lainnya [2]. *Data mining* diartikan sebagai proses pencarian pola berupa pengetahuan yang tersembunyi [3]. Klasifikasi adalah metode *data mining* untuk memprediksi jenis atau kategori data *instance* berdasarkan sekumpulan atribut dalam data. Atribut yang digunakan bisa berupa kategorikal misal berupa bilangan *ordinal* (seperti urutan: kecil, sedang dan besar), nilai *integer* (jumlah kata dalam paragraf), nilai *rill* (seperti suhu). Algoritma C4.5 merupakan algoritma klasifikasi dengan metode *Decision Tree* (pohon keputusan) yang populer karena mempunyai kelebihan. Kelebihannya merupakan dapat mencerna informasi numerik (*kontinyu*) dan diskret, dapat menanggulangi nilai atribut yang hilang, serta menciptakan ketentuan yang mudah diinterpretasikan [4]. Dalam mengkonstruksi pohon algoritma C4.5 membaca seluruh sampel data latih dari *storage* dan memuatnya ke memori. Hal tersebut yang akan menjadi kelemahan algoritma ini dalam kategori skalabilitas yaitu algoritma ini hanya dapat digunakan jika data latih dapat disimpan secara keseluruhan pada waktu yang bersamaan di memori [5].

Dalam penelitian ini direncanakan akan menggunakan algoritma C4.5 berbasis *Feature Selection*. Penerapan algoritma dengan fitur seleksi diterapkan untuk ketepatan dalam memprediksi waktu tunggu alumni untuk mendapatkan pekerjaannya. Dengan menggunakan pendekatan ini diharapkan dapat meningkatkan nilai akurasi dan menghasilkan keputusan yang akurat. *Forward Selection* merupakan salah satu pilih fitur yang dilakukan sebelum pemrosesan dalam metode klasifikasi yang teruji efisien digunakan untuk memecahkan masalah pemilihan fitur yang relevan dalam sesuatu data. *Forward Selection* dijalankan menggunakan pemilihan fitur yang relevan untuk data yang mempengaruhi terhadap hasil klasifikasi serta meningkatkan efektifitas kerja metode klasifikasi [6].

Dari permasalahan diatas, maka akan dilakukan penelitian klasifikasi masa tunggu alumni untuk mendapatkan pekerjaan berdasarkan kompetensi menggunakan algoritma C4.5 dengan tahap *Knowledge Discovery In Database* (KDD). *Software* yang digunakan pada penelitian ini adalah *Rapid Miner Studio*. Data yang digunakan adalah data alumni Fasilkom UNSIKA pada tahun lulus 2016-2020 yang diperoleh dari hasil kuesioner yang dibagikan kepada alumni. Penelitian ini diharapkan dapat membantu pihak prodi dalam menyusun kurikulum dan membenahan pembelajaran sehingga menghasilkan lulusan yang berkualitas.

## 2. Tinjauan Pustaka

Terdapat penelitian sebelumnya yang sejalan dengan penelitian ini diantaranya penelitian yang sebelumnya yang berjudul Implementasi Naïve Bayes Untuk Prediksi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan [1]. Proses klasifikasi dengan algoritma Naive Bayes menggunakan aplikasi WEKA, dengan parameter pengujian 10 folds cross-validation. Berdasarkan hasil uji coba menunjukkan tingkat akurasi yang dihasilkan adalah sebesar 48.629%, dari 1240 data uji hanya 603 data yang berhasil diklasifikasi dengan benar. Sedangkan tingkat kesalahan ditunjukkan dengan MAE (Mean Absolute Error) yang bernilai 0.3. Selanjutnya penelitian sebelumnya yang berjudul Analisa dan Penerapan Metode Neural Network Dalam Mengidentifikasi Faktor-Faktor Masa Tunggu Kerja Lulusan [7]. Berdasarkan dari penelitian dengan menggunakan metode *neural network* diperoleh hasil *class precision* sebesar 87,76%,

dan prediksi masa tunggu mendapatkan pekerjaan kurang lebih sama dengan 6 bulan sebesar 43 dan masa tunggu lebih dari 6 bulan sebesar 6.

Selanjutnya penelitian yang sebelumnya yang berjudul Penerapan Algoritma C4.5 Untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta [8]. Berdasarkan pohon keputusan (*decision tree*) atribut yang paling berpengaruh terhadap jenis pekerjaan alumni adalah Fakultas. Hal itu terjadi karena fakultas sebagai akar dari pohon keputusannya, dan memiliki nilai *gain ratio* yang paling tinggi di antara atribut yang lain. Selanjutnya penelitian sebelumnya yang berjudul Klasifikasi Kompetensi Alumni Berdasarkan Masa Tunggu Alumni Untuk Mendapatkan Pekerjaan Menggunakan Metode Algoritma C4.5 [9]. Dari hasil metode algoritma c4.5 bahwa kompetensi yang paling berpengaruh atau yang merupakan akar (*root*) pohon keputusan yaitu, data 2015 adalah pengetahuan disiplin ilmu dengan tingkat akurasi 63,64%, data 2016 adalah manajemen waktu dengan tingkat akurasi 82,86%, data 2017 adalah kemampuan analisis dengan tingkat akurasi 69,60%, data 2018 adalah pengetahuan disiplin ilmu dengan tingkat akurasi 57,82% dan data 2019 adalah kemampuan komputer dengan tingkat akurasi 48,25%.

Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya, penelitian yang akan dilakukan saat ini yaitu penerapan algoritma C4.5 dalam pengklasifikasian masa tunggu alumni untuk mendapatkan pekerjaan berdasarkan kompetensi di Fasilkom UNSIKA menggunakan metodologi *Knowledge Discovery in Database* (KDD) dan proses pengolahan data dilakukan dengan *software* data mining yaitu *RapidMiner Studio*.

### 3. Metodologi

Penelitian ini dilakukan dengan metodologi *Knowledge Discovery in Database* (KDD). *Knowledge Discovery in Database* (KDD) adalah sebuah proses mencari sebuah ilmu yang berguna dari sekumpulan data. Proses ini bersifat interaktif dan iteratif, meliputi sejumlah langkah yang melibatkan pengguna dalam membuat keputusan dan dapat mengulang pada beberapa langkah sebelumnya [10]. Tahapan yang terdapat pada *Knowledge Discovery in Database* (KDD) adalah sebagai berikut [11] : *data cleaning, data integration, data selection, data transformation, data mining, interpretasi/ evaluasi, knowledge* (pengetahuan).

Langkah selanjutnya yaitu melakukan klasifikasi menggunakan algoritma *decision tree* (C4.5). Algoritma C4.5 merupakan metode pembuatan pohon keputusan berdasarkan data latih yang sebelumnya disediakan. Algoritma C4.5 merupakan salah satu bentuk pengembangan ID3. Beberapa pengembangan yang dilakukan C4.5 mencakup berpotensi untuk proses data lanjutan dan pemangkasannya. Secara umum algoritma C4.5 akan membangun pohon keputusan sebagai berikut [9] :

1. Pilih atribut sebagai root (akar)

$$\text{Gain (S,A)} = \text{Entropy (s)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy (s}_i) \tag{1}$$

dengan :

- Si : Himpunan kasus pada partisi ke i
- A : Variabel
- n : Jumlah partisi atribut A
- |Si| : Jumlah kasus pada partisi ke i
- |S| : Jumlah kasus dalam S

Untuk menghitung Entropy dapat dilihat pada persamaan (2)

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

dengan:

- S : Himpunan Kasus
- n : Jumlah partisi S
- Pi : Proporsi dari Si terhadap S

$$\text{Presentase} = \frac{\text{Jumlah prediksi benar}}{\text{Jumlah data}} \times 100\% \tag{3}$$

2. Buat cabang untuk setiap skor
3. Membagi kasus di cabang
4. Ulangi proses untuk setiap cabang sehingga semua kasus di cabang tersebut memiliki nilai yang sama.

## 4. Hasil dan Pembahasan

### 4.1 Hasil

Data yang digunakan pada penelitian ini adalah data primer yang diperoleh dari alumni mahasiswa Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang. Data yang digunakan adalah data lulusan alumni dari tahun 2016 sampai tahun 2020. Tahap ini dimulai dari pengumpulan data yang dilakukan dengan membagikan kuesioner (*google form*). Terdapat 462 *record* data dengan 19 atribut terdiri dari Nama, NPM, Tempat Lahir, Tanggal Lahir, Jenis Kelamin, Angkatan, Tahun Lulus, IPK, Lama Studi, Riwayat Organisasi/Kepanitiaan, Pengalaman Magang/Kerja, Memiliki Sertifikasi Keahlian, Status Pekerjaan, Lama Waktu Mendapatkan Pekerjaan, Bidang Pekerjaan, Jenis Perusahaan, Posisi Pekerjaan, Kota Tempat Bekerja, dan Kesesuaian Bidang Pekerjaan dengan TI. Tampilan data yang telah diperoleh dapat dilihat pada Tabel 1.

Tabel 1. Data Alumni Fasilkom Unsika

No	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Nama	NPM (isi jika ingat)	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Angkatan	Tahun Lulus	IPK	Lama Studi	Riwayat Organisasi/Kepanitiaan	Memiliki Sertifikasi Keahlian (Kon)	Status Pekerjaan	Lama Waktu Mendapatkan Pekerjaan	Bidang Pekerjaan	Jenis Perusahaan	Posisi Pekerjaan	Kota Tempat Bekerja	Kesesuaian Bidang Pekerjaan dengan TI	
2	Deny Ansyad	1241177004223	Karawang	17/05/1994	Laki-Laki	2012	2016	3,28	4 Tahun	3	Ada	Tidak	Bekerja	0 - 2 Bulan	gawai Swas	Lokai/ Daerah	Programmer	Karawang	3
3	ochamad Saepud	1241177004228	Karawang	16/01/1993	Laki-Laki	2012	2016	3,25	4 Tahun	1	Tidak	Tidak	Bekerja	3 - 6 Bulan	gawai Swas	Lokai/ Daerah	Honorer	Karawang	2
4	fan Maulana zidmi	1241177004230	tegay	12/05/1994	Laki-Laki	2012	2016	3,27	4 Tahun	4	Ada	Tidak	Bekerja	0 - 2 Bulan	gawai Swas	Nasional (Multi Na)	Staff	karawang	3
5	Karnadi	1241177004291	Karawang	03/04/1994	Laki-Laki	2012	2016	3,52	4,5 Tahun	2	Ada	Ada	Bekerja	0 - 2 Bulan	gawai Swas	Nasional	Software Engineer	Jakarta	3
6	rdi Yoga Prasety	1141177004143	Karawang	07/05/1992	Laki-Laki	2011	2016	3,04	4,5 Tahun	1	Tidak	Tidak	Bekerja	7 - 12 Bulan	gawai Swas	Nasional (Multi Na)	Staff Program	Bekasi	3
7	Octaviana Hermi	1241177004122	Kota Bandung	22/10/1994	erempua	2012	2016	3,52	4,5 Tahun	5	Ada	Tidak	belum Bekerj	0 - 2 Bulan	belum Bekerja				
8	ea wanda framuz	1141177004156	Karawang	06/01/1994	Laki-Laki	2011	2016	2,89	5 Tahun	1	Tidak	Ada	Bekerja	7 - 12 Bulan	gawai Swas	Nasional	Staff	Karawang	1
9	Tri widia dutamz	1341177004084	Banten	28/06/1995	erempua	2013	2017	2,99	4 Tahun	1	Tidak	Ada	Bekerja	0 - 2 Bulan	gawai Swas	Nasional	Staff IT	Karawang	3
10	Kevin Pratama	1341177004081	Bandung	27/08/1995	Laki-Laki	2013	2017	2,97	4 Tahun	1	Ada	Ada	Bekerja	0 - 2 Bulan	gawai Swas	Nasional (Multi Na)	Engineer Staff	Karawang	1
11	Avian Abdul Jaba	1341177004107	Bandung	06/10/1994	Laki-Laki	2013	2017	3,4	4 Tahun	4	Ada	Ada	Bekerja	0 - 2 Bulan	gawai Swas	Lokai/ Daerah	Staff IT	Karawang	3
12	Boby Setyadi	1341177004053	Bekasi	14/10/1995	Laki-Laki	2013	2017	3,21	4 Tahun	1	Tidak	Tidak	Bekerja	3 - 6 Bulan	gawai Swas	Nasional	IT	Jakarta Timur	3
13	Muhammad Afif	1341177004111	Karawang	24/09/1995	Laki-Laki	2013	2017	3,33	4 Tahun	1	Tidak	Ada	Bekerja	3 - 6 Bulan	gawai Swas	Nasional (Multi Na)	port & purchasing	kabupaten bi	2
14	usuf Eka Wicaksa	1341177004017	Kediri	25/05/1995	Laki-Laki	2013	2017	3,37	4 Tahun	4	Tidak	Ada	Bekerja	7 - 12 Bulan	gawai Swas	Lokai/ Daerah	Programmer	Kab. Karawang	3
15	Yuni Apriani	1341177004042	Karawang	08/04/1993	erempua	2013	2017	3,15	4 Tahun	1	Tidak	Tidak	Bekerja	0 - 2 Bulan	gawai Swas	Lokai/ Daerah	Operator packing	Karawang	2
16	Tia Mutiasih	1341177004094	Bekasi	11/10/1994	erempua	2013	2017	3,25	4 Tahun	1	Tidak	Tidak	belum Bekerj	7 - 12 Bulan	belum Bekerja				
17	EMILA RAHMAH	1341177004044	CREBON	16/10/1994	erempua	2013	2017	3,05	4 Tahun	1	Ada	Ada	Bekerja	3 - 6 Bulan	gawai Swas	Lokai/ Daerah	HRD	Karawang	2
18	Eryan Fauzan	1341177004064	Bekasi	10/07/1995	Laki-Laki	2013	2017	3,43	4 Tahun	5	Tidak	Tidak	Bekerja	0 - 2 Bulan	gawai Swas	Lokai/ Daerah	Development Lea	Karawang	3
19	Adi Fauzi Kusmar	1341177004059	Karawang	16/06/1993	Laki-Laki	2013	2017	3,34	4 Tahun	1	Ada	Tidak	Bekerja	0 - 2 Bulan	awal Negeri	Lokai/ Daerah	sional Pranata Kon	Karawang	3
20	Imad wawan seti	1341177004253	karawang	17/04/1995	Laki-Laki	2013	2017	3,24	4 Tahun	3	Tidak	Ada	belum Bekerj	3 - 6 Bulan	gawai Swas	Nasional	customer service	karawang	1
21	Uri Saruni, S. Kom	1321177004002	Kota Serang	12/12/1981	Laki-Laki	2013	2017	3,46	4 Tahun	2	Ada	Ada	Bekerja	0 - 2 Bulan	gawai Swas	Lokai/ Daerah	arehouse Manag	Kabupaten Serang	2
22	Saraswati	1341177004129	upatan Pekalon	05/06/1995	erempua	2013	2017	3,21	4 Tahun	2	Ada	Ada	Bekerja	3 - 6 Bulan	gawai Swas	Lokai/ Daerah	Guru Honorer	ibupaten Karawa	2
23	rul Yanuari Sunar	1341177004010	Karawang	30/01/1996	erempua	2013	2017	3,25	4 Tahun	3	Ada	Tidak	Bekerja	3 - 6 Bulan	gawai Swas	Lokai/ Daerah	Guru	Karawang	3
24	Selvi Handayani	1341177004039	Karawang	12/05/1995	erempua	2013	2017	3,12	4 Tahun	1	Ada	Ada	belum Bekerj	7 - 12 Bulan	gawai Swas	Lokai/ Daerah	Belum bekerja	irawang Jawa ba	1

Pada penelitian ini dari 19 atribut hanya 14 atribut yang digunakan, dapat dilihat pada Tabel 2.

Tabel 2. Atribut Terpilih

No	Atribut Terpilih	Keterangan
1.	Nama	ID
2.	Jenis Kelamin	Atribut Utama
3.	Angkatan	Atribut Utama
4.	Tahun Lulus	Atribut Utama
5.	IPK	Atribut Utama
6.	Lama Studi	Atribut Utama
7.	Riwayat Organisasi/Kepanitiaan	Atribut Utama
8.	Pengalaman Magang/Kerja	Atribut Utama
9.	Memiliki Sertifikasi Keahlian	Atribut Utama
10.	Lama Waktu Mendapatkan Pekerjaan	Kelas Label
11.	Bidang Pekerjaan	Atribut Utama
12.	Jenis Perusahaan	Atribut Utama
13.	Kota Tempat Bekerja	Atribut Utama
14.	Kesesuaian Bidang Pekerjaan dengan TI	Atribut Utama

Pada Tabel 2, "Lama Waktu Mendapatkan Pekerjaan" adalah atribut yang akan dijadikan kelas label. Dari seluruh atribut yang digunakan, 3 atribut diantaranya akan dilakukan transformasi data, dapat dilihat pada Tabel 3.

Tabel 3. Transformasi Data

Atribut		Nilai	Keterangan	Bobot
Kota Bekerja	Tempat	Karawang	Karawang	110
		Bekasi	Luar Karawang	160
		Purwakarta		
		Jakarta		
Lama Studi		DII	3.5-4 Tahun	139
		3.5-4 Tahun		
		4.5 Tahun	>4 Tahun	131
		4.8 Tahun		
Lama Mendapatkan Pekerjaan	Waktu	Tahun	0-2 Bulan	121
		DII		
		0-2 Bulan	≥3 Bulan	149
		3-6 Bulan		
7-12 Bulan	>12 Bulan			

Dalam membuat pohon keputusan menghitung terlebih dahulu kelas label dari hasil pengumpulan data yang dilakukan telah didapatkan. Diketahui dari data *testing* berjumlah 462 dan data *training* sebanyak 270 data. Dari data tersebut diketahui bahwa sebanyak 121 data "0-2 Bulan" dan 149 data "≥3 Bulan". Berikutnya proses pengolahan data menggunakan algoritma C4.5 untuk melihat pohon keputusan dan *rule*.

- Langkah pertama adalah menghitung nilai entropy total berdasarkan jumlah data "0-2 Bulan" dan jumlah "≥3 Bulan" berdasarkan uraian tersebut menggunakan persamaan:

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

Sehingga didapat nilai entropy keseluruhan sebesar:

$$\begin{aligned} \text{Entropy}(S) &= ((-121/270) \cdot \log_2(121/270) + (-149/270) \cdot \log_2(149/270)) \\ &= 0,992228333 \end{aligned}$$

- Hitung nilai entropy dan nilai gain masing-masing atribut. Nilai gain tertinggi akan menjadi *root node* dari pohon keputusan yang dibuat. Nilai entropy dapat dihitung menggunakan rumus yang sama seperti mencari gain total. Setelah ditemukan nilai entropy maka nilai gain dapat ditemukan dengan persamaan :

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

Tabel 4. Hasil Perhitungan Entropy dan Gain

Atribut		Jumlah Kasus	0-2 Bulan	≥3 Bulan	Entropy	Information Gain
TOTAL		270	121	149	0,992228333	
Jenis Kelamin	LAKI-LAKI	218	99	119	0,993920016	0,000438646
	PEREMPUAN	52	22	30	0,98285869	
						<b>0,153438</b>
Angkatan	2010	2	0	2	0	0,016974
	2011	14	9	5	0,940285959	
	2012	40	29	11	0,848548178	
	2013	58	39	19	0,91244118	
	2014	65	17	48	0,829059069	
	2015	41	19	22	0,996134484	
	2016	50	8	42	0,634309555	
Tahun Lulus	2016	6	4	2	0,918295834	
	2017	29	14	15	0,999142104	
	2018	69	28	41	0,974240833	

Atribut		Jumlah Kasus	0-2 Bulan	≥3 Bulan	Entropy	Information Gain
IPK	2019	85	46	39	0,995102271	0,063030
	2020	81	30	51	0,950956048	
	<3	65	46	19	0,871683617	
	>3	205	75	130	0,947435136	
Lama Studi	3.5 Tahun – 4 Tahun	139	42	97	0,883913177	0,066966
	>4 Tahun	131	79	52	0,969136403	
Kota Tempat Bekerja	Karawang	160	83	77	0,998985367	0,021371
	Luar Karawang	110	38	72	0,929942935	
Riwayat Organisasi	1	96	54	42	0,988699408	0,031105
	2	45	16	29	0,938932011	
	3	66	29	37	0,989375583	
	4	37	10	27	0,84185219	
	5	26	12	14	0,995727452	
Pengalaman Magang/Kerja	Ada	195	95	100	0,999525689	0,011722
	Tidak	75	26	49	0,931055823	
Memiliki Sertifikasi	Ada	157	73	84	0,996456059	0,001148
	Tidak	113	48	65	0,983611636	
Bidang Pekerjaan	Pegawai Swasta	254	105	149	0,978244179	0,071954
	Pegawai Negeri Sipil	16	16	0	0	
Jenis Perusahaan	Lokal	137	64	73	0,99688469	0,001155
	Nasional	90	39	51	0,987137774	
	Internasional	43	18	25	0,980798365	
Kesesuaian Bidang Pekerjaan	1	70	39	31	0,9905577	0,016299
	2	96	35	61	0,946422051	
	3	104	47	57	0,993320422	

Berdasarkan hasil perhitungan *entropy* dan *gain* pada Tabel 4, nilai *gain* tertinggi terdapat pada atribut “Angkatan” yaitu dengan nilai 0,153438. Oleh karena itu, atribut “Angkatan” akan menjadi *node root* dari pohon keputusan. Setelah dilakukan hasil perhitungan *entropy* dan *gain*, maka pohon keputusan yang terbentuk akan terlihat seperti Gambar 1.



Gambar 1. Pohon Keputusan

### 4.2 Pembahasan

Pada tahap evaluasi ini dilakukan perbandingan hasil rata-rata dari percobaan sebanyak 10 kali menggunakan *Cross Validation* menggunakan algoritma C4.5 dan juga fitur seleksi *forward selection*.

Tabel 5. Skenario Uji dengan *Cross Validation*

Skenario	Nilai <i>k-fold</i>
1.	2 folds
2.	3 folds
3.	4 folds
4.	5 folds
5.	6 folds
6.	7 folds
7.	8 folds
8.	9 folds
9.	10 folds

Nilai *Accuracy*

$$= \frac{TP+TN}{TP+TN+FP+FN} \times 100$$

Nilai *Recall*

$$= \frac{TP}{TP+FN} \times 100$$

Nilai *F-Measure*

$$= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

a) Perbandingan Hasil Evaluasi *Decision Tree* (C4.5)

Perbandingan evaluasi dilandasi pada penggunaan parameter uji diantaranya nilai *Accuracy*, *Precision*, *Recall*, *F-Measure* dan *AUC*.

Tabel 6. Perbandingan Evaluasi *Decision Tree* (C4.5)

Skenario	K-folds	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>AUC</i>
1	2	<b>60,74%</b>	<b>62,28%</b>	<b>52,99%</b>	<b>57,26%</b>	<b>0.741</b>
2	3	57,78%	57,94%	54,48%	56,16%	0.628
3	4	58,52%	58,33%	57,46%	57,89%	0.736
4	5	58,15%	58,14%	55,97%	57,34%	0.707
5	6	59,26%	59,84%	54,48%	57,03%	0.684
6	7	59,25%	59,68%	55,22%	57,36%	0.698
7	8	53,87%	59,54%	58,21%	58,87%	0.650
8	9	57,41%	57,98%	51,49%	54,54%	0.648
9	10	54,44%	54,55%	49,25%	51,76%	0.627
Rata-rata		57,71%	58,70%	54,39%	56,47%	0.627

Berdasarkan Tabel 6. di atas, maka skenario ke-1 dengan *2-folds Cross Validation* merupakan skenario terbaik dengan nilai *Accuracy* sebesar 60,74%, *Precision* 62,28%, *Recall* 52,99%, *F-Measure* 57,26% dan *AUC* 0.741 yang termasuk ke dalam kategori *fair classification*. Sedangkan skenario ke-9 dengan *10-folds Cross Validation* dengan nilai *Accuracy* sebesar 54,44%, *Precision* 54,55%, *Recall* 49,25%, *F-Measure* 51,76% dan *AUC* 0.627 yang termasuk ke merupakan skenario terburuk dalam kategori *poor classification*.

b) Perbandingan Hasil Evaluasi *Decision Tree* (C4.5) dengan *Forward Selection*

Perbandingan evaluasi didasari pada penggunaan parameter uji diantaranya nilai *Accuracy*, *Precision*, *Recall*, *F-Measure*, dan *AUC*

Tabel 7. Perbandingan Evaluasi *Decision Tree + Forward Selection*

Skenario	K-folds	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>AUC</i>
1	2	80,37%	79,56%	81,34%	80,44%	0.914
2	3	80,37%	79,56%	81,34%	80,44%	0.914
3	4	80,37%	79,56%	81,34%	80,44%	0.914
4	5	80,37%	79,56%	81,34%	80,44%	0.914
5	6	80,37%	79,56%	81,34%	80,44%	0.914
6	7	80,37%	79,56%	81,34%	80,44%	0.914
7	8	80,37%	79,56%	81,34%	80,44%	0.914
8	9	80,37%	79,56%	81,34%	80,44%	0.914
9	10	80,37%	79,56%	81,34%	80,44%	0.914
Rata-rata		80,37%	79,56%	81,34%	80,44%	<b>0.914</b>

Berdasarkan Tabel 7. di atas, memiliki nilai evaluasi yang sama yaitu *Accuracy* sebesar 80,37%, *Precision* 79,56%, *Recall* 81,34%, *F-Measure* 80,44% dan *AUC* 0.914 yang termasuk ke dalam kategori *excellent classification*.

c) Hasil Fitur Seleksi *Forward Selection*

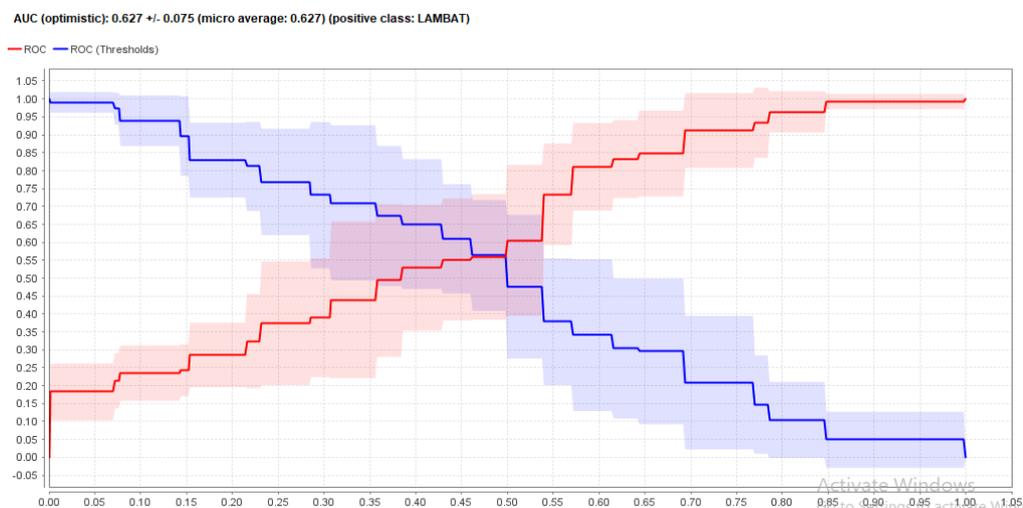
Tabel 8. Hasil Fitur *Forward Selection*

No.	Atribut	Weight
1.	Lama Studi	0
2.	Kota Tempat Bekerja	1
3.	Jenis Kelamin	0
4.	Angkatan	1
5.	Tahun Lulus	1
6.	IPK (Indeks Prestasi Kumulatif)	1
7.	Riwayat Organisasi/Kepanitiaan	1
8.	Pengalaman Magang	0
9.	Memiliki Sertifikasi Keahlian	1
10.	Bidang Pekerjaan	0
11.	Jenis Perusahaan	1
12.	Kesesuaian Bidang Pekerjaan dengan TI	0

Dilihat pada Tabel 8. bahwa nilai *weight* yang bernilai 1 artinya signifikan atau berpengaruh sedangkan nilai *weight* yang bernilai 0 artinya tidak signifikan atau tidak berpengaruh. Dalam hal ini bahwa atribut yang berpengaruh dalam untuk prediksi waktu tunggu alumni Fasilkom Unsika mendapatkan pekerjaan adalah “Kota Tempat Bekerja”, “Angkatan”, “Tahun Lulus”, “IPK”, “Riwayat Organisasi/Kepanitiaan”, “Memiliki Sertifikasi Keahlian”, dan “Jenis Perusahaan”.

d) Hasil Evaluasi Model Menggunakan *ROC (Receiver Operating Characteristic)*

Selain *confusion matrix*, kurva *Receiver Operating Characteristic (ROC)* dihasilkan oleh *RapidMiner Studio* dapat dilihat pada Gambar 2. dan Gambar 3.

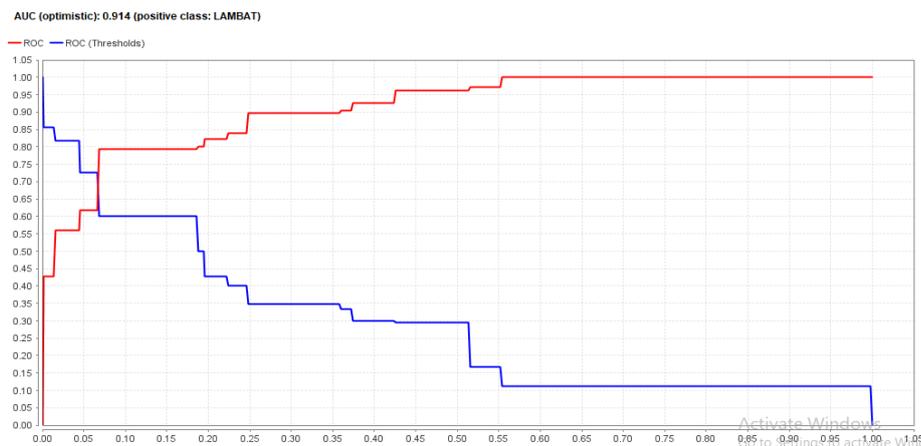


Gambar 2. Grafik AUC Algoritma C4.5

Tingkat akurasi menurut Gorunescu (2011) dapat diinterpretasikan sebagai berikut.

1. 0.90 – 1.00 = *excellent classification*
2. 0.80 – 0.90 = *good classification*
3. 0.70 – 0,80 = *fair classification*
4. 0.60 – 0.70 = *poor classification*
5. 0.50 – 0.60 = *failure classification*

Dari Gambar 2. di atas diketahui bahwa nilai AUC algoritma *decision tree (C4.5)* yang dihasilkan sebesar 0.627 sehingga hasil klasifikasi penelitian ini masuk ke dalam kategori *poor classification*.



Gambar 3. Grafik AUC Algoritma C4.5 Dengan Fitur *Forward Selection*

Tingkat akurasi menurut Gorunescu (2011) dapat diinterpretasikan sebagai berikut.

1. 0.90 – 1.00 = *excellent classification*
2. 0.80 – 0.90 = *good classification*
3. 0.70 – 0,80 = *fair classification*
4. 0.60 – 0.70 = *poor classification*
5. 0.50 – 0.60 = *failure classification*

Dari Gambar 3. di atas diketahui bahwa nilai AUC algoritma *Decision Tree* (C4.5) dengan fitur *forward Selection* yang dihasilkan sebesar 0.914 sehingga hasil klasifikasi penelitian ini masuk ke dalam kategori *excellent classification*.

## 5. Kesimpulan

Berdasarkan pada penelitian ini, pengujian klasifikasi masa tunggu alumni untuk mendapatkan pekerjaan berdasarkan algoritma *decision tree* (C4.5) hasil dari semua pengujian yang telah dilakukan dengan 9 skenario uji menggunakan *k-fold cross validation*. Diketahui bahwa pada pemodelan *decision tree* (C4.5) dihasilkan performa terbaik pada skenario uji ke-1 dengan *2 folds cross validation* dengan menghasilkan nilai *accuracy* sebesar 60,74%, *precision* 62,28%, *recall* 52,99%, *f-measure* 57,26% dan *AUC* 0.741 yang termasuk ke dalam kategori *fair classification*. Sedangkan pada pemodelan *decision tree* dengan fitur seleksi *forward selection* menghasilkan nilai yang sama dari 9 pengujian evaluasi yaitu *accuracy* sebesar 80,37%, *precision* 79,56%, *recall* 81,34%, *f-measure* 80,44% dan *AUC* 0.914 yang termasuk ke- dalam kategori *excellent classification*.

## DAFTAR REFERENSI

- [1] Adnyana I. M. B. Implementasi *Naive Bayes* Untuk Memprediksi Waktu Tunggu Alumni Dalam Memeroleh Pekerjaan. *Seminar Nasional Teknologi & Sains (SAINTEKS)*. 2020: 131-134.
- [2] Han, J, Kamber, M, Pei, J. *Data mining Concepts and Techniques*. Waltham: Morgan Kaufmann. 2012
- [3] Bahar, B. Model Pengujian Akurasi Berbasis Empiris Pada Algoritma A-Priori. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 2019, 8(2): 45-56.
- [4] Wulandari. N. Analisis Minat Belajar Siswa Pada Lembaga Pendidikan Indonesia Amerika Perumnas 3 Bekasi Dengan Metode C4.5. *Jurnal Rekayasa Informasi*. 2019, 8(1): 22-31.
- [5] Fitriani E. Perbandingan Algoritma C4.5 dan *Naive Bayes* Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *SISTEMASI: Jurnal Sistem Informasi*. 2020, 9(1): 103-115.
- [6] Fanani M. R. Algoritma Naive Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa. *Jurnal DISPROTEK*. 2020, 11(1):13-22.
- [7] Khoirudin, Hadi S, Nugroho A. Analisa dan Penerapan Metode Neural Networks Dalam Mengidentifikasi Faktor-Faktor Masa Tunggu Kerja Lulusan. *Pengembangan Rekayasa dan Teknologi*. 2020, 16(1): 17-22.
- [8] Asroni, Respati B, Riyadi S. Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta. *Semesta Teknika*. 2018, 21(2): 158-165.
- [9] Cahyaningtyas R, Luqman. R, Yolana I. H. Klasifikasi Kompetensi Alumni Berdasarkan Masa Tunggu Alumni untuk Mendapatkan Pekerjaan Menggunakan Metode Algoritma C4.5. *KILAT*. 2017, 9(2): 297-310.
- [10] Apriliana, Natalis R, Jumadil N. Implementasi Text Mining Klasifikasi Skripsi Menggunakan Metode *Naive Bayes Classifier*. *Semantik*. 2017, 3(2): 187–94.
- [11] Aziz M. F, Defiyanti, S, Sari, B. N. Perbandingan Algoritma Cart dan K-Nearest Neighbor Untuk Prediksi Luas Lahan Panen Tanaman Padi di Kabupaten Karawang. *Jurnal TAM (Technology Acceptance Model)*. 2018, 9(2): 74- 78.
- [12] Amrinda G.D. Analisis Klasifikasi Waktu Tunggu Kerja Dengan Metode *Support Vector Machine* dan *Naive Bayes Classification*. Skripsi MIPA. Universitas Islam Indonesia, Yogyakarta. 2018
- [13] Anisa, Mesran. Analisa Pola Pekerjaan Lulusan STMIK Budi Darma Menerapkan Metode C4.5. *Konferensi Nasional Teknologi Informasi dan Komputer*, 2018, (2): 446-449.
- [14] Ardiansyah D, Walim. Algoritma C4.5 Untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswa SMP Dengan Menggunakan Aplikasi *Rapid Miner*. *Jurnal Inkofar*. 2018, 1(2): 5-12.
- [15] Bode A. *K-Nearest Neighbor* Dengan *Feature Selection* Menggunakan *Backward Selection* Untuk Prediksi Harga Komoditi Kopi Arabika. *Jurnal Ilmiah*. 2017, 9(2): 188-195.
- [16] Chusna, H. A, Rumiaty, A. T. Penerapan Metode K-Means dan Fuzzy C-Means untuk Pengelompokan Sekolah Menengah Pertama (SMP) di Indonesia Berdasarkan Standar Nasional Pendidikan (SNP). *JURNAL SAINS DAN SENI ITS*. 2020, 9(2): 216-223.
- [17] Eka S, Murtiadi A, Muh. Ikhsan Alif. S. Analisis Kesesuaian Kompetensi Alumni Uin Alauddin Terhadap Kepuasan Stakeholder. *Jurnal Minds: Manajemen Ide dan Inspirasi*, 2016, 3(1): 27-57
- [18] Fahriah S, Wiktasari. Algoritma C4.5 Berbasis Forward Selection Untuk Klasifikasi Bidang Minat Studi Mahasiswa Teknik Informatika. *Jurnal Teknik Elektro Terapan*. 2021, 10(1): 1-6.
- [19] Larose, D. T. *Discovering Knowledge In Data An Introduction*. A John Wiley & Sons, Inc. 2005.
- [20] Mardi Y. Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*. 2019, 2(2): 213-219.
- [21] N. Nuraeni. Penentuan Kelayakan Kredit Dengan Algoritma *Naive Bayes Classifier*: Studi Kasus Bank Mayapada Mitra Usaha Cabang Pgc. *Teknik Komputer*, 2017, 3(1): 9–15.
- [22] Nanja M, Purwanto. Metode *K-Nearest Neighbor* Berbasis *Forward Selection* Untuk Prediksi Harga Komoditi Lada. *Jurnal Pseudocode* 2015, 2(1): 53-64.

- [23] Novakovic, J. D., Veljovic, A., Ilic, S. S., Papic, Z., & Tomovic, M. Evaluation of Classification Models in Machine Learning. *Theory and Applications of Mathematics & Computer Science*. 2017, 7(1): 39-46.
- [24] Nugroho M. F, Wibowo S. Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma *Naive Bayes*. *Jurnal informatika UPGRIS*. 2017, 3(1): 63-70.
- [25] Purba, S. D, Sirait, P, Arwin. Credit Card Risk Classification Using KNearest Neighbor Weighted Algorithm Based on Forward Selection. *Jurnal Mantik*. 2020, 4(3): 1551-1559.
- [26] Takdirillah, R. Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Penjualan Bisnis Ritel. *Edumatic: Jurnal Pendidikan Informatika*. 2020, 4(1): 37-46.
- [27] Uruilal, I, Supit, C. J, Jansen, T. Prediksi Banjir Di Sungai Ranowanko Kecamatan Amurang Kabupaten Minahasa Selatan. *Jurnal Sipil Statik*. 2020. 8(2): 167-174.
- [28] Wahyono H, Dwiza R. Prediksi Calon Pendorong Darah Potensial Dengan Algoritma *Naive Bayes*, *K-Nearest Neighbors* dan *Decision Tree C4.5*. *JURIKOM (Jurnal Riset Komputer)*. 2020, 7(1): 7-14.
- [29] Wahyono, Nugroho A. Penerapan Algoritma C4.5 Untuk Prediksi Tingkat Kompetensi Karyawan PT Multistarda Arah Sarana. *JOUTICA*. 2018, 3(1): 145-150