

RANCANG BANGUN SISTEM PREDIKSI KELULUSAN MAHASISWA JURUSAN TEKNIK INFORMATIKA UNIVERSITAS PALANGKA RAYA

Agus Sehatman Saragih ¹⁾, Ade Chandra Saputra ²⁾

¹ Fakultas Teknik, Universitas Palangka Raya
email: assaragih@it.upr.ac.id

² Fakultas Teknik, Universitas Palangka Raya
email: adechandra@it.upr.ac.id

Abstract

The department is the leading part in the implementation of education from a college, so that it always conducts an evaluation to improve the quality and efficiency of higher education including the improvement of graduate quality. The length of student study is one of the reference variables of success level of the teaching process.

The graduation prediction system using the data mining classification method is Decision Tree C4.5. Data attributes used include; Gender, Religion, SKS, IPS, Graduated Semester, and TA Type. The Graduated Semester attribute is used as a predictive target attribute. Where the attribute value pass semester is made into 2 values that is 8-10 Semesters (≤ 5 Years) and 11-14 Semesters (> 5 Years). The prediction test was performed using k-fold cross-validation method and linear regression measurement.

The highest accuracy score on the prediction system was obtained in the 6th experiment and the 7th experiment was 61.54%. While for the lowest accuracy value obtained in the 5th experiment of 30.77%. From the value of R2 from experiment 1 to experiment 10 shows the highest value of 0.40 and the lowest 0.29. The value of R2 obtained is so small that it can further explain the result of prediction accuracy with decision tree C4.5 algorithm is very small value.

Keywords: *Predition, Decisision Tree C4.5, K-fold cross-validation, Regresi Linear*

1. PENDAHULUAN

Jurusan merupakan bagian terpenting dalam penyelenggaraan pendidikan dari sebuah perguruan tinggi, sehingga senantiasa melakukan evaluasi untuk meningkatkan mutu dan efisiensi perguruan tinggi termasuk peningkatan kualitas lulusan. Jurusan Teknik Informatika merupakan jurusan dibawah Fakultas Teknik Universitas Palangka Raya yang memiliki jumlah mahasiswa yang cukup banyak. Hal ini terlihat dari peningkatan jumlah calon mahasiswa baru pada setiap tahun ajaran. Permasalahan yang sering terjadi adalah masih banyaknya jumlah mahasiswa yang lulus dengan lama studi melampaui waktu yang telah ditetapkan dengan perolehan Indeks Prestasi Kumulatif (IPK) yang relatif rendah yang dapat mempengaruhi mutu lulusan Perguruan Tinggi. Seiring dengan terus bertambahnya jumlah mahasiswa di Jurusan Teknik Informatika maka jumlah data kemahasiswaan terus meningkat sehingga terjadi penumpukan data yang belum diolah dengan optimal untuk menggali informasi dan pengetahuan baru yang dapat digunakan

sebagai bahan pertimbangan pimpinan dalam proses pengambilan kebijakan dan keputusan. Data ini juga sebagai peringatan dini (*early warning*) bagi mahasiswa tertentu yang berdasarkan hasil prediksi dinyatakan berpotensi lulus dengan melampaui ketentuan lama studi ataupun berpotensi lulus dengan perolehan IPK yang rendah.

Hal inilah yang mendasari peneliti untuk merancang suatu sistem yang dapat melakukan prediksi terhadap mahasiswa baru yang mendaftar apakah berpeluang menyelesaikan kuliah tepat waktu atau tidak dengan menggunakan salah satu model data mining

2. KAJIAN LITERATUR

Ada beberapa penelitian yang terdahulu yang digunakan peneliti sebagai tinjauan pustaka dalam melakukan penelitian ini yaitu sebagai berikut :

Penelitian dengan judul perancangan aplikasi prediksi kelulusan tepat waktu pada STMIK Dipanegara Makasar. Pada penelitian tersebut prediksi dilakukan dengan

menggunakan algoritma *k-Nearest Neighbor* (*k-NN*) dan aplikasi web yang dibangun menggunakan bahasa pemrograman PHP 5.0 dan *database engine* MySQL. Atribut yang digunakan pada penelitian terdiri dari jenis kelamin, nilai ebtanas, jurusan dan provinsi SMA. Hasil pengujian dengan menerapkan algoritma *k-NN* dari kasus lama yaitu tahun wisuda 2004 s/d 2010 dan kasus baru tahun wisuda 2011 terhadap alumni diperoleh tingkat akurasi 83 %.[1]

Penelitian kedua adalah penelitian berupa implementasi algoritma C4.5 untuk memprediksi tingkat kelulusan mahasiswa. Pada penelitian tersebut dilakukan prediksi kelulusan pada mahasiswa alumni tahun 2007 dan 2008 sebagai data *training* dan data *testing* alumni tahun 2009. Atribut yang digunakan untuk melakukan prediksi terdiri dari IP Semester 1 s/d Semester 6, jenis kelamin, SMA, tipe kelulusan dan jumlah SKS pada semester 6. Hasil implementasi dan uji coba aplikasi tingkat akurasi hasil prediksi kelulusan sebesar 87.5 % dari data sebanyak 60 untuk data *training* dan data *testing* sebanyak 40.[2]

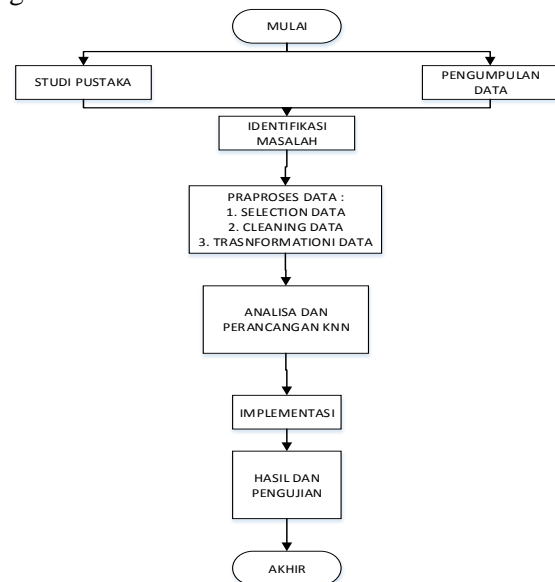
Penelitian ketiga adalah penelitian untuk melakukan prediksi mahasiswa berpotensi non-aktif menggunakan metode *decision tree* C4.5. Pada penelitian tersebut atribut yang digunakan adalah umur, kota asal, asal sekolah, IP semester 1-4 dan SKS dari semester 1-4. Hasil penelitian yang diperoleh pada percobaan ke-1 dengan ratio data *training* 90% diperoleh akurasi prediksi terbaik sebesar 92.79%. Pada percobaan ke-2 dengan ratio data *training* 20% diperoleh nilai akurasi terbaik 90.98% dan pada percobaan ke-3 dengan ratio data *training* 90% diperoleh nilai akurasi terbaik sebesar 97.60%. Kesimpulan yang diperoleh juga pada penelitian menyatakan bahwa variabel yang paling berpengaruh terhadap mahasiswa non-aktif adalah IP Semester dan SKS yang dikumpulkan dari semester 1-4.[3]

Berdasarkan penelitian-penelitian referensi di atas maka peneliti (2017) ingin mengimplementasikan algoritma *decision tree* C4.5 untuk melakukan prediksi kelulusan studi mahasiswa. Data studi kasus yang akan digunakan adalah data mahasiswa Jurusan Teknik Informatika Universitas Palangka Raya. *Decision tree* C4.5 dipilih sebagai algoritma untuk prediksi berdasarkan hasil

review pada keenam penelitian algoritma *decision tree* C4.5 memiliki tingkat akurasi lebih tinggi dan mudah dipahami. Data mahasiswa yang digunakan adalah data mahasiswa angkatan tahun 2008-2010. Atribut yang digunakan untuk melakukan prediksi terdiri atas SKS 1-3, IPS 1-3, jenis kelamin, agama dan jenis tugas akhir yang akan diambil. Hasil akhir penelitian ini adalah sebuah aplikasi *dekstop* untuk melakukan prediksi kelulusan studi mahasiswa sehingga dapat digunakan untuk melakukan prediksi dan dapat memberikan peringatan dini bagi pihak akademis perguruan tinggi untuk memberikan bimbingan awal bagi mahasiswa yang memiliki kecenderungan menyelesaikan studi lebih dari 10 semester.

3. METODE PENELITIAN

Tahapan-tahapan penelitian dalam mengembangkan sistem ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian Rancang Bangun Sistem Prediksi Kelulusan Mahasiswa dengan menggunakan Algoritma Decision Tree (C4.5)

3.1 Peubah dan Pengukuran

Data mahasiswa didapatkan dari pihak pengelola data akademik Fakultas Teknik Universitas Palangka Raya yaitu data mahasiswa Jurusan Teknik Informatika angkatan tahun 2008-2010 sebanyak 130 record data setelah dilakukan proses seleksi, validasi data dan transformasi data. Atribut data yang dipilih untuk melakukan prediksi

adalah Jenis Kelamin, Agama, SKS, IPS, Lulus Semester, dan Jenis TA. Atribut Lulus Semester dijadikan sebagai atribut target prediksi. Di mana nilai atribut lulus semester dibuat menjadi 2 nilai yaitu 8-10 Semester (≤ 5 Tahun) dan 11-14 Semester (> 5 Tahun).

Tabel 1. Atribut dan Nilai

No.	Atribut	Nilai (Kode)
1.	Jenis Kelamin	Laki-laki (L) dan Perempuan (P).
2.	Agama	ISLAM (A1), PROTESTAN (A2), KATOLIK (A3), HINDU (A4), dan BUDHA(A5).
3.	SKS 1-3	≤ 15 (SKS1), 16-20 (SKS2), dan ≥ 21 (SKS3).
4.	IPS 1-3	$\leq 2,50$ (IP1) , 2,51-3,00 (IP2) , 3,01-3,50 (IP3) dan $\geq 3,51$ (IP4).
5.	Jenis Tugas Akhir	SI WEB (JT1), SI DEKSTOP (JT2), WEB PROFIL (JT3), WEB E-GOV (JT4), GAME (JT5), JARINGAN (JT6), dan ANDROID (JT7).
6.	Lulus Semester	8-10 (LS1) dan 11-14 (LS4).

3.2 Teknik Pengumpulan dan Analisis Data

Teknik Pengumpulan dan Analisis Data dilakukan dengan studi pustaka dan observasi. Data yang digunakan pada sistem Sistem Prediksi Kelulusan Mahasiswa bersumber dari akademik Fakultas Teknik Universitas Palangka Raya yaitu data mahasiswa Jurusan Teknik Informatika angkatan tahun 2008-2010.

3.3 Rancangan Sistem

Sistem yang dikembangkan merupakan aplikasi Prediksi tingkat kelulusan mahasiswa Teknik Informatika Universitas Palangka Raya yang dikembangkan berbasis desktop. Metode pengembangan perangkat lunak menggunakan UML (*Unified Modeling Language*) yaitu *usecase diagram*, *activity diagram* dan *class diagram*.

3.3.1 Use Case Diagram

Pada sistem yang akan dirancang hanya ada 1 aktor yaitu Admin seperti yang terlihat pada Tabel 2 dan untuk setiap *use case* beserta deskripsinya dapat dilihat pada Tabel 3.

Tabel 2. Definisi Aktor

No.	Aktor	Deskripsi
1.	Admin	Orang yang menganalisis hasil prediksi kelulusan berdasarkan data <i>training</i> yang didapatkan dari akademis Fakultas Teknik dan melakukan uji prediksi terhadap data <i>testing</i> .

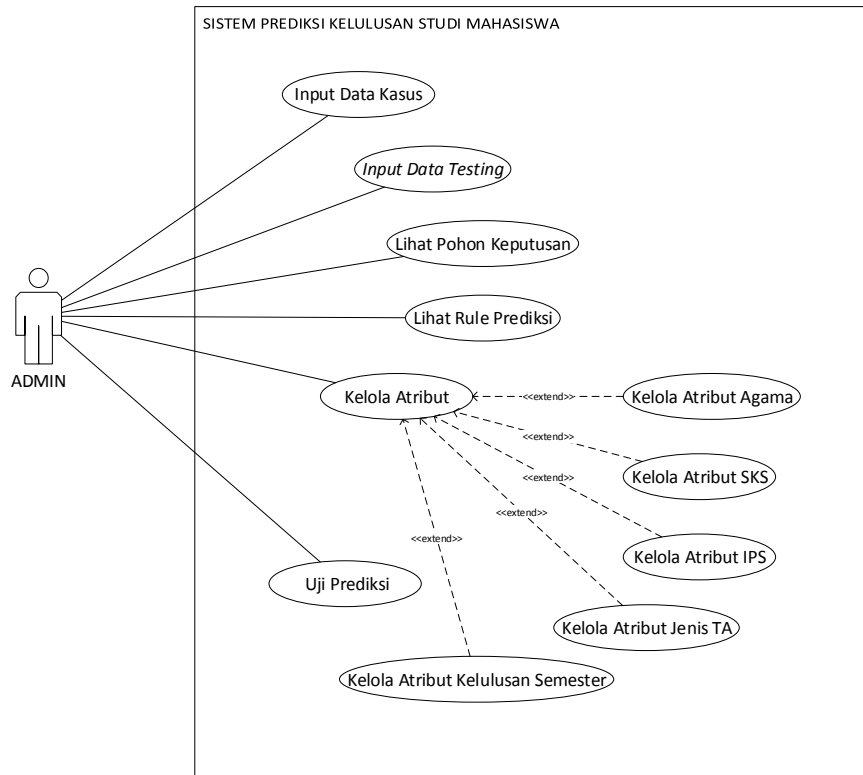
Tabel 3. Deskripsi Use case

No.	Use case	Deskripsi
1.	Input Data Kasus	Untuk memasukkan data yang digunakan sebagai data kasus/data <i>training</i> .
2.	Input Data <i>Testing</i>	Untuk memasukkan data <i>testing</i> yang akan dilakukan prediksi kelulusannya.
3.	Lihat Pohon Keputusan	Untuk melihat pohon keputusan yang dibentuk dari data <i>training</i> yang sudah diproses. Pohon keputusan menggambarkan <i>rule</i> dan pola kelulusan yang akan digunakan untuk melakukan prediksi.
4.	Lihat <i>Rule</i> Prediksi	Untuk melihat <i>rule</i> yang akan digunakan melakukan prediksi.
5.	Kelola Atribut	Untuk mengelola nilai data atribut yang digunakan pada data kasus, data <i>testing</i> dan data uji prediksi seperti atribut agama, sks, ips, jenis TA, dan kelulusan semester.
6.	Uji Prediksi	Untuk melakukan pengujian hasil prediksi data <i>testing</i> sehingga diketahui tingkat akurasi prediksi

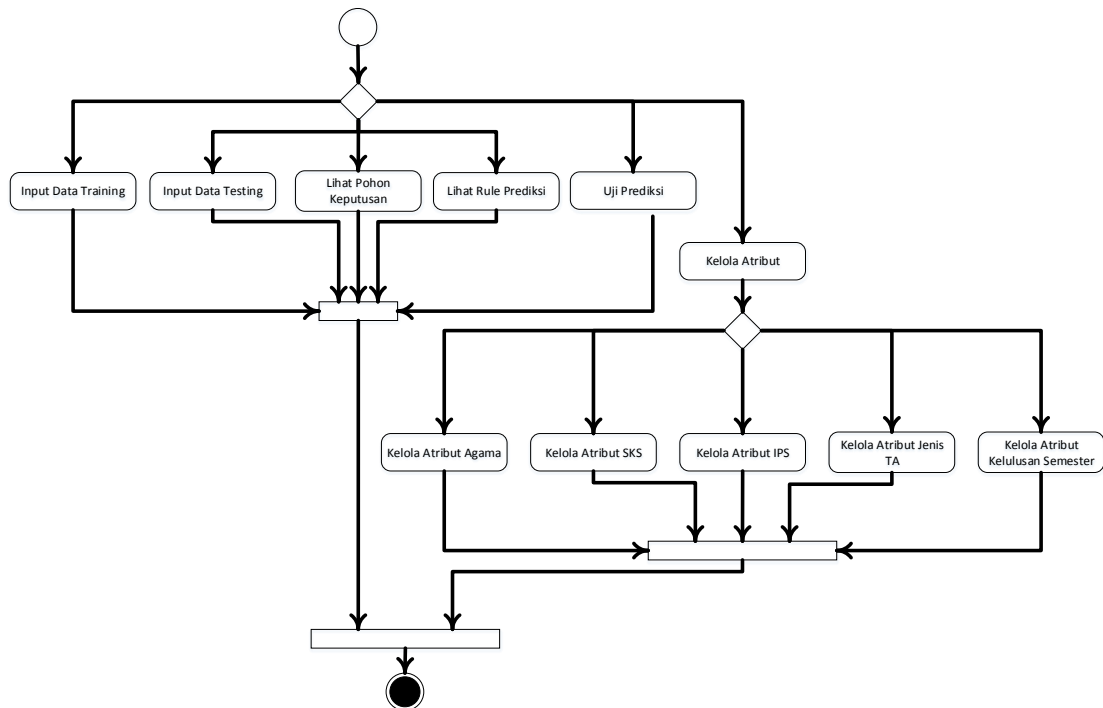
Pada sistem prediksi hanya ada satu pengguna aplikasi yaitu Admin, kemudian untuk proses-proses yang dilakukan ada 6 *use case* utama yaitu Input Data Kasus, Input Data *Testing*, Lihat Pohon Keputusan, Lihat *Rule* Prediksi, Kelola Atribut dan Uji Prediksi. Usecase diagram pada gambar 2.

3.3.2 Activity Diagram

Pada Gambar 3 dapat dilihat *activity diagram* yang menggambarkan urutan aktivitas proses yang dilakukan pada aplikasi prediksi kelulusan studi mahasiswa berdasarkan *use case* yang dibuat sebelumnya.



Gambar 2. Use Case Diagram Aplikasi Prediksi Kelulusan Studi Mahasiswa

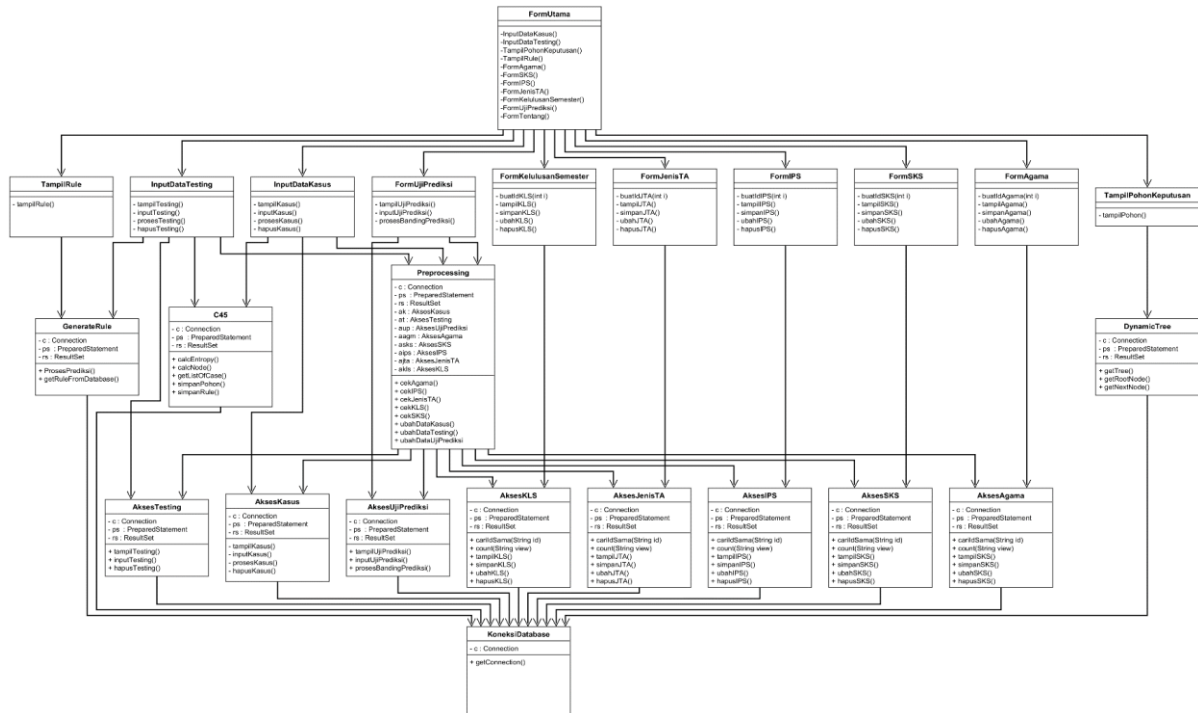


Gambar 3. Activity Diagram Aplikasi Prediksi Kelulusan Studi Mahasiswa

3.1.3 Class Diagram

Rancangan *class diagram* yang akan digunakan untuk membangun Aplikasi

Prediksi Kelulusan Studi Mahasiswa dapat dilihat pada Gambar 4.



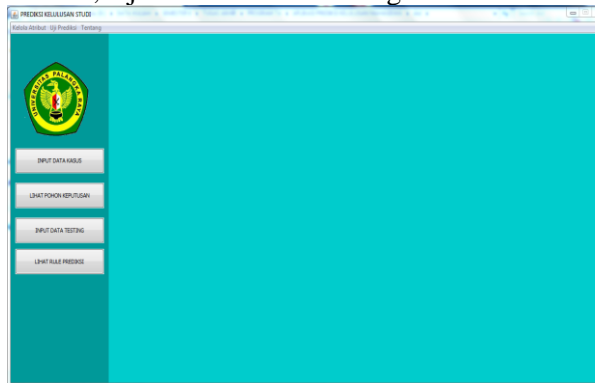
Gambar 4. Class Diagram Aplikasi Prediksi Kelulusan Studi Mahasiswa

4. HASIL DAN PEMBAHASAN

4.1 Implementasi Sistem

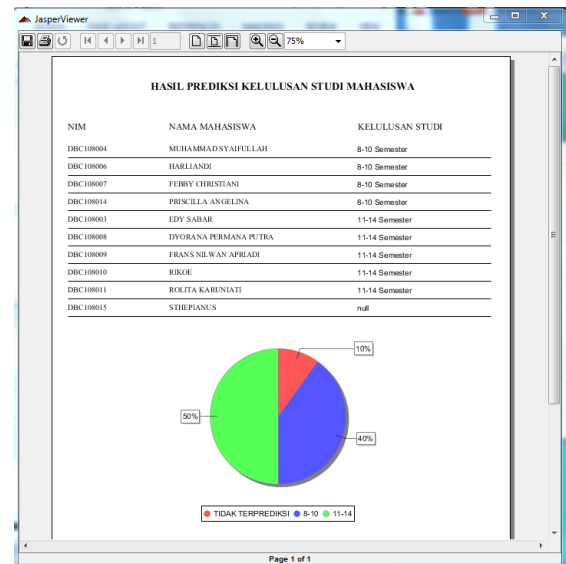
Implementasi merupakan proses yang dilakukan untuk merepresentasikan hasil perancangan ke dalam pemrograman. Implementasi dilakukan untuk menghasilkan aplikasi yang mampu melakukan prediksi kelulusan studi mahasiswa dan menampilkan atribut yang paling berpengaruh terhadap kelulusan studi mahasiswa.

Implementasi antarmuka tampilan awal aplikasi saat pertama kali dijalankan dapat dilihat pada Gambar 5. Dari Tampilan Awal dapat dilihat pada panel utama aplikasi terdapat menu Input Data Kasus Lihat Pohon Keputusan, Input Data *Testing* dan Lihat *Rule* Prediksi. Pada *menu bar* atas terdapat Kelola Atribut, Uji Prediksi dan Tentang.



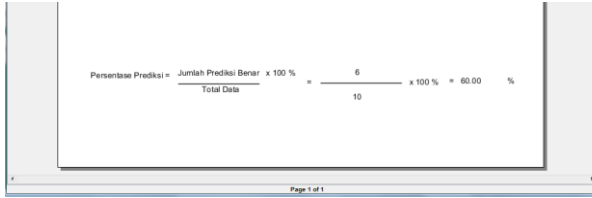
Gambar 5. Implementasi Tampilan Awal Aplikasi

Proses uji akurasi prediksi dapat dijalankan dengan mengklik tombol Uji Prediksi, kemudian aplikasi akan memberikan konfirmasi saat proses sudah selesai dan langsung menampilkan *report* hasil uji akurasi prediksi seperti yang terlihat pada Gambar 6.



Gambar 6. Proses Uji Akurasi Prediksi Selesai

Pada Gambar 7 menunjukkan persentase akurasi *report* hasil uji akurasi prediksi sebesar 60%. Hasil perhitungan dilakukan dengan membandingkan nilai prediksi yang sesuai dengan data uji akurasi dibagi total data yang dibandingkan.



Gambar 7. Report Hasil Uji Akurasi Prediksi

4.2 Pengujian Sistem

4.2.1 Pengujian Algoritma *Decision Tree* C4.5

Pada perbandingan hitungan manual data kasus yang digunakan ada sebanyak 10 data seperti yang terlihat pada Tabel 4.

Tabel 4. Data Kasus

NIM	NAMA	JK	AGAMA	SKS 1	SKS 2	SKS 3	IPS 1	IPS 2	IPS 3	JT	LS
DBC110057	BAGUS IN	Laki-laki	ISLAM	24	25	23	3,75	3,52	3,50	ANDROID	10
DBC109009	JHONI PR	Laki-laki	PROTESTA	18	19	23	2,08	3,11	2,96	SI WEB	12
DBC109011	RIA ARISK	Perempuan	PROTESTA	21	17	20	2,13	2,63	3,00	WEB E-GC	12
DBC109012	AGNESTRY	Laki-laki	PROTESTA	21	16	12	2,17	2,47	2,12	SI WEB	12
DBC110074	YOLANTYA	Perempuan	PROTESTA	21	22	21	2,96	2,68	3,21	SI WEB	8
DBC11079	SELVIA LE	Perempuan	ISLAM	22	25	21	3,21	2,76	3,29	SI WEB	9
DBC110080	PRIHTIAN	Perempuan	ISLAM	24	25	21	3,33	2,86	3,43	SI WEB	9
DBC108056	YENIRA A	Perempuan	PROTESTA	18	19	19	2,08	2,95	2,74	SI DEKSTC	13
DBC108094	VIRMAN S	Laki-laki	ISLAM	20	11	9	2,08	1,55	1,92	SI WEB	14
DBC108096	REDIANSY	Laki-laki	ISLAM	14	10	10	1,96	2,06	1,75	WEB PROF	14

Data pada Tabel 4 adalah contoh data kasus yang masih belum dilakukan *preprocessing*. Data tersebut nantinya saat akan masuk ke dalam aplikasi akan di rubah nilai setiap atribut sesuai kategori atributnya masing-masing.

Perubahan nilai atribut dilakukan untuk memudahkan proses perhitungan terutama untuk data nilai SKS dan IPS menjadi rentang nilai kategorial. Pengkategorian nilai tersebut disesuaikan dengan kategori yang ada pada akademik. Hasil perubahan nilai atribut dapat dilihat pada Tabel 5.

Tabel 5. Data Kasus Hasil Preprocessing

NIM	NAMA	JK	AGAMA	SKS 1	SKS 2	SKS 3	IPS 1	IPS 2	IPS 3	JT	LS
DBC110057	BAGUS IN	L	A1	SKS3	SKS3	SKS3	IP4	IP4	IP4	JT7	L1
DBC109009	JHONI PR	L	A2	SKS2	SKS2	SKS3	IP1	IP2	IP2	JT1	L2
DBC109011	RIA ARISK	P	A2	SKS3	SKS2	SKS2	IP1	IP2	IP2	JT4	L2
DBC109012	AGNESTRY	L	A2	SKS3	SKS3	SKS1	IP1	IP1	IP1	JT1	L2
DBC110074	YOLANTYA	P	A2	SKS3	SKS3	SKS3	IP2	IP2	IP3	JT1	L1
DBC11079	SELVIA LE	P	A1	SKS3	SKS3	SKS3	IP2	IP2	IP3	JT1	L1
DBC110080	PRIHTIAN	P	A1	SKS3	SKS3	SKS3	IP3	IP2	IP3	JT1	L1
DBC108056	YENIRA A	P	A2	SKS2	SKS2	SKS2	IP1	IP2	IP2	JT2	L2
DBC108094	VIRMAN S	L	A1	SKS2	SKS1	SKS1	IP1	IP1	IP1	JT1	L2
DBC108096	REDIANSY	L	A1	SKS1	SKS1	SKS1	IP1	IP1	IP1	JT3	L2

Hitung nilai entropi, *gain*, *split info* dan *gain ratio*. Atribut simpul dipilih dengan nilai *gain ratio* paling besar[4].

Perhitungan awal yang perlu dilakukan adalah menghitung entropi total dari semua atribut dengan menggunakan Rumus.

$$Entropy(Total) = \left(-\frac{6}{10} * \log_2 \frac{6}{10}\right) + \left(-\frac{4}{10} * \log_2 \frac{4}{10}\right) = 0,44217936 + 0,528771238$$

$$Entropy(Total) = 0,970950594$$

Kemudian hitung nilai entropi Jenis Kelamin (JK) dengan nilai atribut L dan P dengan Rumus.

$$Entropy(L) = \left(-\frac{4}{5} * \log_2 \frac{4}{5}\right) + \left(-\frac{1}{5} * \log_2 \frac{1}{5}\right) = 0,257542476 + 0,464385619$$

$$Entropy(L) = 0,721928095$$

$$Entropy(P) = \left(-\frac{2}{5} * \log_2 \frac{2}{5}\right) + \left(-\frac{3}{5} * \log_2 \frac{3}{5}\right) = 0,528771238 + 0,442179356$$

$$Entropy(P) = 0,970950594$$

Selanjutnya perhitungan *gain* atribut Jenis Kelamin (JK) dapat dilakukan dengan menggunakan Rumus 2.2.

$$Gain(Total, JK) = 0,970950594$$

$$- \left(\left(\frac{5}{10} * 0,721928095 \right) + \left(\frac{5}{10} * 0,970950594 \right) \right)$$

$$Gain(Total, JK) = 0,12451125$$

Sementara perhitungan *split info* Jenis Kelamin (JK) dilakukan dengan menggunakan Rumus .

$$SplitInfo(Total, JK)$$

$$= \left(-\frac{|5|}{|10|} * \log_2 \frac{|5|}{|10|} \right) + \left(-\frac{|5|}{|10|} * \log_2 \frac{|5|}{|10|} \right)$$

$$SplitInfo(Total, JK) = 0,5 + 0,5 = 1$$

Dari perhitungan *gain* dan *split info* Jenis Kelamin di atas maka dapat dihitung *gain ratio* atribut JK dengan Rumus.

$$GainRatio(Total, JK) = \frac{0,12451125}{1} = 0,12451125$$

Jadi, nilai *gain ratio* untuk atribut jenis kelamin adalah 0,12451125

Tabel 6. Perhitungan Untuk Membentuk Pohon Keputusan

PEMBENTUKAN POHON KEPUTUSAN (LEVEL 0)										
NODE	TOTAL	JML KASUS (S)	11-14 (S1)	8-10 (S2)	Entropi S1	Entropi S2	Entropi Total	Gain	SPLITINFO	Gain Ratio
JENIS KELAMIN										
LAKI-LAKI (L)	5	4	1	0,25754248	0,464385619	0,721928095			0,124511	0,12451125
PEREMPUAN (P)	5	2	3	0,52877124	0,442179356	0,970950594				0,5
AGAMA										
ISLAM (A1)	5	2	3	0,52877124	0,442179356	0,970950594			0,124511	0,12451125
PROTESTAN (A2)	5	4	1	0,25754248	0,464385619	0,721928095				0,5
KATOLIK (A3)	0	0	0	0	0	0				0
HINDU (A4)	0	0	0	0	0	0				0
SKS 1										
<=15 (SKS1)	1	1	0	0	0	0			0,419973	1,2954619
16-20 (SKS2)	3	3	0	0	0	0				0,32418793
>=21 (SKS3)	6	2	4	0,52832083	0,389975	0,918295834				0,4421794
SKS 2										
<=15 (SKS2)	2	2	0	0	0	0			0,970951	1,5219281
16-20 (SKS2)	4	4	0	0	0	0				0,637974026
>=21 (SKS3)	4	0	4	0	0	0				0,5287712
SKS 3										
<=15 (SKS3)	3	3	0	0	0	0			0,609987	1,4854753
16-20 (SKS2)	3	2	0	0	0	0				0,5210897
>=21 (SKS3)	5	1	4	0,46438562	0,257542476	0,721928095				0,5
IPS 1										
<=2,50 (IP1)	6	6	0	0	0	0			0,970951	1,5709506
2,51-3,00 (IP2)	1	0	1	0	0	0				0,4421794
3,01-3,50 (IP3)	2	0	2	0	0	0				0,3321928
>=3,51 (IP4)	1	0	1	0	0	0				0,3321928
IPS 2										
<=2,50 (IP1)	3	3	0	0	0	0			0,485475	1,6854753
2,51-3,00 (IP2)	3	2	3	0,52877124	0,442179356	0,970950594				0,5210897
3,01-3,50 (IP3)	1	1	0	0	0	0				0,5
>=3,51 (IP4)	1	0	1	0	0	0				0,3321928
IPS 3										
<=2,50 (IP1)	3	3	0	0	0	0			0,970951	1,5709506
2,51-3,00 (IP2)	3	3	0	0	0	0				0,5210897
3,01-3,50 (IP3)	4	0	4	0	0	0				0,5287712
>=3,51 (IP4)	0	0	0	0	0	0				0
JENIS TA										
SI WEB (JT1)	6	3	3	0,5	0,5	1			0,370951	1,7709506
SI DESKTOP (JT2)	1	1	0	0	0	0				0,4421794
WEB PROFIL (JT3)	1	1	0	0	0	0				0,3321928
WEB E-GOV (JT4)	1	1	0	0	0	0				0,3321928
GAME (JT5)	0	0	0	0	0	0				0
JARINGAN (JT6)	0	0	0	0	0	0				0
ANDROID (JT7)	1	0	1	0	0	0				0,3321928
MAX GAIN										0,637974026

Berdasarkan hasil perhitungan yang terlihat pada Tabel 6 atribut yang terpilih sebagai atribut akar adalah SKS_2 dengan *gain ratio* = 0,637974026.

Masing-masing nilai atribut SKS_2 memiliki nilai entropi = 0 sehingga sudah positif memprediksikan kasus yaitu SKS1 (0-15) dan SKS2 (16-20) dengan kelulusan studi 11-14 Semester, sementara SKS3 (21-26) mengarahkan ke kelulusan studi 8-10 Semester. Proses pembentukan pohon keputusan berhenti pada atribut SKS_2.

4.2.2 Pengujian Hasil Prediksi

Pengujian hasil prediksi dilakukan dengan menggunakan *k-fold cross-validation* yang membagi data menjadi 10 bagian ($k=10$) [5]. Total data yang digunakan sebanyak 130 data mahasiswa Jurusan Teknik Informatika 2008-2010. Dari total data tersebut untuk percobaan 1-10 digunakan 13 data sebagai data *testing* dan 117 data sebagai data kasus. Hasil persentase akurasi prediksi dari percobaan 1-10 dapat dilihat pada Tabel 7.

Tabel 7. Hasil Pengujian Akurasi Prediksi.

Percobaan ke-	Akurasi Prediksi	Jumlah Prediksi Benar / Total Data	Atribut Akar
1	53,85 %	7/13	IPS_1
2	53,85 %	7/13	IPS_1
3	53,85 %	7/13	SKS_2
4	53,85 %	7/13	SKS_2
5	30,77 %	4/13	SKS_2
6	61,54 %	8/13	IPS_1
7	61,54 %	8/13	IPS_1
8	46,15 %	6/13	IPS_1
9	46,15 %	6/13	IPS_1
10	46,15 %	6/13	IPS_1
Rata-rata Persentase Prediksi			50,77 %

Pada setiap percobaan yang telah dilakukan didapatkan rata-rata akurasi prediksi sebesar 50,77 % dan atribut akar yang paling sering muncul adalah IPS_1 (nilai Indeks Prestasi pada semester 1). Nilai akurasi paling tinggi didapatkan pada percobaan ke-6 dan percobaan ke-7 sebesar 61,54 %. Sementara untuk nilai akurasi paling rendah didapatkan pada percobaan ke-5 sebesar 30,77 %.

4.2.2 Pengujian Estimasi Nilai dan Persamaan Linier Prediksi

Analisis regresi adalah teknik statistik untuk permodelan dan investigasi hubungan dua atau lebih variabel [6]. Nilai R^2 menunjukkan koefisien determinasi, untuk mengukur kebaikan suai (*goodness of fit*) dari persamaan regresi. Nilai R^2 terletak antara 0 – 1, dan kecocokan model dikatakan lebih baik kalau R^2 semakin mendekati 1. Pada Tabel 8 dapat dilihat persamaan linier berganda untuk data kasus setiap percobaan beserta nilai R^2 .

Tabel 8. Persamaan Linier Prediksi Percobaan 1-10

Percobaan ke-	Persamaan Linier	Nilai R^2
1	$Y = 15,95 + 0,04X_1 - 0,12X_2 - 0,05X_3 - 0,50X_4 + 0,28X_5 - 0,67X_6$	0,38
2	$Y = 16,46 + 0,02X_1 - 0,08X_2 - 0,03X_3 - 0,62X_4 + 0,04X_5 - 0,64X_6$	0,35
3	$Y = 15,95 + 0,03X_1 - 0,11X_2 - 0,02X_3 - 0,18X_4 + 0,01X_5 - 0,95X_6$	0,33
4	$Y = 16,01 + 0,02X_1 - 0,12X_2 - 0,07X_3 - 0,51X_4 + 0,43X_5 - 0,51X_6$	0,39
5	$Y = 15,83 + 0,03X_1 - 0,12X_2 - 0,04X_3 - 0,39X_4 + 0,25X_5 - 0,65X_6$	0,35
6	$Y = 15,71 + 0,02X_1 - 0,09X_2 - 0,02X_3 - 0,39X_4 + 0,24X_5 - 0,90X_6$	0,29
7	$Y = 16,31 + 0,01X_1 - 0,08X_2 - 0,01X_3 - 0,38X_4 + 0,13X_5 - 0,98X_6$	0,37
8	$Y = 16,17 + 0,05X_1 - 0,12X_2 - 0,08X_3 - 0,73X_4 + 0,40X_5 - 0,59X_6$	0,40
9	$Y = 16,14 + 0,01X_1 - 0,09X_2 - 0,05X_3 - 0,04X_4 + 0,18X_5 - 0,69X_6$	0,35
10	$Y = 15,92 + 0,02X_1 - 0,11X_2 - 0,04X_3 - 0,45X_4 + 0,23X_5 - 0,67X_6$	0,34

Ket : Y : Variabel nilai Kelulusan Semester

X_1 : Variabel SKS_1

X_2 : Variabel SKS_2

X_3 : Variabel SKS_3

X_4 : Variabel IPS_1

X_5 : Variabel IPS_2

X_6 : Variabel IPS_3

Koefisien determinasi (R^2) sering digunakan untuk mengetahui seberapa besar kemampuan semua variabel bebas (X_1-X_6) dalam menjelaskan variabel terikatnya. Pada percobaan 1, diperoleh nilai R^2 sebesar 0,38 berarti kemampuan variabel bebas untuk menjabarkan variabel terikatnya sebesar 38%

dan menyatakan terdapat nilai sebesar 62% untuk variabel terikat dijelaskan oleh faktor-faktor lainnya. Dari nilai R^2 dari percobaan 1 hingga percobaan 10 menunjukkan nilai tertinggi sebesar 0,40 dan terendah 0.29. Nilai R^2 yang didapatkan sangat kecil sehingga dapat semakin menjelaskan hasil akurasi prediksi dengan algoritma *decision tree C4.5* bernilai sangat kecil. Jadi, dengan menggunakan nilai R^2 dari persamaan regresi linier berganda dapat dinyatakan bahwa kemampuan variabel SKS 1-3 dan IPS 1-3 masih belum cukup untuk memprediksikan nilai Kelulusan Semester mahasiswa dan perlu adanya tambahan variabel bebas (atribut prediksi) yang dapat menggambarkan pola kelulusan mahasiswa.

5. KESIMPULAN

Berdasarkan hasil penelitian pada tugas akhir dengan judul “Prediksi Kelulusan Studi Mahasiswa Dengan *Decision Tree C4.5*”, maka dapat ditarik kesimpulan sebagai berikut.

1. Pengujian hasil prediksi dilakukan dengan menggunakan *k-fold cross-validation* dengan nilai $k=10$. Dari percobaan yang telah dilakukan didapatkan rata-rata akurasi prediksi sebesar 50,77 % dan atribut akar yang paling sering muncul adalah IPS_1 (nilai Indeks Prestasi pada semester 1). Nilai akurasi paling tinggi didapatkan pada percobaan ke-6 dan percobaan ke-7 sebesar 61,54 %. Sementara untuk nilai akurasi paling rendah didapatkan pada percobaan ke-5 sebesar 30,77 %.
2. Nilai R^2 yang dihasilkan setiap percobaan juga sangat kecil (< 0.90), sehingga menunjukkan bahwa kemampuan variabel atribut prediksi dalam memprediksikan atribut kelulusan semester mahasiswa sangat kurang, sehingga memerlukan atribut prediksi tambahan.
3. Berdasarkan rata-rata akurasi prediksi dengan nilai sebesar 50,77 % dapat disimpulkan algoritma *decision tree*

C4.5 pada penelitian ini, masih belum optimal untuk melakukan prediksi kelulusan studi mahasiswa. Keterbatasan atribut prediksi yang digunakan sebagai parameter prediksi, terutama penggunaan atribut jenis kelamin dan agama yang tidak tepat membuat tingkat akurasi prediksi menjadi lebih rendah dibandingkan penelitian lain yang menggunakan algoritma *decision tree C4.5* yang bisa mencapai akurasi prediksi di atas 70% .

4. Berdasarkan perhitungan menggunakan metode Regresi Linear dari percobaan 1 hingga percobaan 10 menunjukkan nilai tertinggi sebesar 0,40 dan terendah 0.29.

6. REFERENSI

- [1] Mustafa, Muhammad Syukri, dan I Wayan Simpen. t.t. *Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining*. Makassar : STMIK Dipanegara.
- [2] Kamagi, David Hartanto, dan Seng Hansun. 2014. *Implementasi Data Mining dengan Algoritma C4.5 Untuk Memprediksi Tingkat Kelulusan Mahasiswa*. Tangerang : Universitas Multimedia Nusantara.
- [3] Untari, D. t.t. *Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5*. Semarang : Universitas Dian Nuswantoro.
- [4] Kusriani dan Luthfi E.T. 2009. *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- [5] Refaeilzadeh, Payam, Lei Tang, dan Huan Liu. 2008. *Cross-Validation*. Arizona : Arizona State University.
- [6] Santosa, Budi (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta : Graha Ilmu.