

PENERAPAN ALGORITMA C4.5 PADA PREDIKSI WAKTU TUNGGU KERJA ALUMNI TEKNIK INFORMATIKA

by Arfan Pontoh T3114051

Submission date: 01-Dec-2019 09:59AM (UTC+0700)

Submission ID: 1224129955

File name: 5_PADA_PREDIKSI_WAKTU_TUNGGU_KERJA_ALUMNI_TEKNIK_INFORMATIKA.pdf (1.5M)

Word count: 15220

Character count: 89676

**PENERAPAN ALGORITMA C4.5 PADA PREDIKSI
WAKTU TUNGGU KERJA ALUMNI
TEKNIK INFORMATIKA**

**Oleh
ARFAN PONTOH
T3114051**

SKRIPSI



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2019**

JUDUL SKRIPSI

**PENERAPAN ALGORITMA C4.5 PADA PREDIKSI
WAKTU TUNGGU KERJA ALUMNI
TEKNIK INFORMATIKA**

**Oleh
ARFAN PONTOH
T3114051**

SKRIPSI



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2019**

PERSETUJUAN SKRIPSI

**PENERAPAN ALGORITMA C4.5 PADA PREDIKSI
WAKTU TUNGGU KERJA ALUMNI
TEKNIK INFORMATIKA**

Oleh
ARFAN PONTOH
T3114051

2
SKRIPSI

Untuk memenuhi salah satu syarat ujian
guna memperoleh gelar Sarjana
Program Studi Teknik Informatika,
ini telah disetujui oleh Tim Pembimbing

Gorontalo, November 2019

Pembimbing I

Pembimbing II

Budy Santoso, S.Kom., M.Eng

Yusrianto Malago, S.Kom., M.Kom

PENGESAHAN SKRIPSI

**PENERAPAN ALGORITMA C4.5 PADA PREDIKSI
WAKTU TUNGGU KERJA ALUMNI
TEKNIK INFORMATIKA**

**Oleh
ARFAN PONTOH
T3114051**

2
Diperiksa oleh Panitia Ujian Strata Satu (S1)
Universitas Ichsan Gorontalo

1. Ketua Penguji
Amiruddin, S.Kom., M.Kom
2. Anggota
Irma Surya Kumala Idris, S.Kom., M.Kom
3. Anggota
Hastuti, S.Kom., M.Kom
4. Anggota
Budy Santoso, S.Kom., M.Eng
5. Anggota
Yusrianto Malago, S.Kom., M.Kom

PERNYATAAN SKRIPSI

7

Dengan ini saya menyatakan bahwa:

1. Karya tulis (skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan dari Tim Pembimbing.
3. Dalam karya tulis (skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, November 2019

Yang membuat pernyataan,

Arfan Pontoh

ABSTRAK

Semakin banyaknya jumlah lulusan tidak sejalan dengan meningkatnya jumlah lapangan pekerjaan khususnya yang terkait dengan rumpun ilmu teknik informatika yang dapat di tunjukan melalui kurang relevanya jenis pekerjaan yang diperoleh lulusan dengan disiplin ilmunya. Oleh karena itu kegiatan *tracer study* terhadap alumni teknik informatika perlu ditingkatkan, khususnya terkait dengan waktu tunggu kerja mereka. Namun masih sulit untuk memperoleh informasi mengenai waktu tunggu kerja alumni tanpa menggunakan sistem yang tepat dan akurat. Algoritma C4.5 dapat digunakan untuk memprediksi waktu tunggu kerja alumni teknik informatika untuk mempermudah pekerjaan jurusan teknik informatika. Dengan demikian penelitian ini bertujuan untuk mengembangkan sistem prediksi waktu tunggu kerja alumni teknik informatika berbasis algoritma C4.5, sehingga dapat diimplementasikan dalam kegiatan *tracer study* oleh suatu jurusan teknik informatika. Hasil penelitian menunjukkan bahwa diperoleh sistem prediksi waktu tunggu kerja alumni teknik informatika berbasis algoritma C4.5 yang tidak kompleks dengan alur logika program yang sudah benar berdasarkan nilai $V(G) = 3$ dari pengujian *White Box* dan sudah bebas dari kesalahan komponen sistem berdasarkan pengujian *Black Box*, sehingga sistem ini dapat diimplementasikan untuk keperluan *tracer study* pada suatu jurusan teknik informatika di perguruan tinggi.

Kata kunci: C4.5, sistem prediksi waktu tunggu kerja, alumni teknik informatika.

ABSTRACT

The increasing number of graduates is not in line with the increasing number of jobs, especially those related to the Informatics Engineering family, which can be demonstrated through the lack of relativity of the types of jobs obtained by graduates with scientific discipline. Therefore tracer study activities of alumni of informatics engineering need to be improved, especially related to the waiting time of their work. However, it is still difficult to obtain information on waiting time for alumni work without using an appropriate and accurate system. Algortima C4.5 can be used to predict the waiting time of informatics engineering alumni work to facilitate the work of informatics engineering majors. Thus this study aims to develop a predictive waiting system for information technology alumni based on the C4.5 algorithm, so that it can be implemented in tracer study activities by an informatics engineering department. The results showed that a predictive waiting system for information technology alumni based on C4.5 algorithm that is not complex with the correct program logic based on the value of $V(G) = 3$ from the White Box test and is free from error system components based on Black testing Box, so that this system can be implemented for the purpose of tracer study in a major in informatics engineering in tertiary institutions.

Keywords: C4.5, System of work waiting time prediction, fresh graduate of computer science.

KATA PENGANTAR

Alhamdulillah, penulis panjatkan kehadiran Allah Subhanahu wa Ta'ala atas berkat rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan usulan penelitian ini untuk memenuhi salah satu syarat penyusunan Skripsi Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo. Salam dan taslim kepada junjungan kita, Nabi Muhammad Shallallahu 'Alaihi wa Sallam atas perjuangan beliau yang telah mengantarkan kita dari alam kebodohan ke alam yang penuh ilmu pengetahuan. Usulan penelitian ini penulis beri judul: **"Penerapan Algoritma C4.5 pada Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika."**

Penulis menyadari sepenuhnya bahwa usulan penelitian ini tidak mungkin terwujud tanpa bantuan dan dorongan dari berbagai pihak, baik bantuan moril maupun materil. Untuk itu, dengan segala keikhlasan dan kerendahan hati, penulis mengucapkan terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Muhammad Ichsan Gaffar, M.Ak., selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke, M.Si., selaku Rektor Universitas Ichsan Gorontalo;
3. Ibu Zohrahayaty, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman S. Panna, M.Kom., selaku Wakil Dekan I Bidang Akademik;
5. Ibu Irma Surya Kumala, M.Kom., selaku Wakil Dekan II Bidang Kepegawaian, Administrasi Umum, dan Keuangan;
6. Bapak Sudirman Melangi, M.Kom., selaku Wakil Dekan III Bidang Kemahasiswaan.
7. Bapak Irvan Abraham Salihi, M.Kom., selaku Ketua Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
8. Bapak Budy Santoso, S.Kom., M.Eng., selaku Dosen Pembimbing Utama yang telah membimbing penulis dalam penyusunan usulan penelitian ini;

9. Bapak Yusrianto Malago, M.Kom., selaku Dosen Pembimbing Pedamping yang telah membimbing penulis dalam penyusunan usulan penelitian ini;
10. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;
11. Kedua Orang Tua dan keluarga atas segala kasih sayang, jerih payah dan doa restunya dalam membesarkan, mendidik penulis, serta memberikan bantuan dan dukungan moril yang sangat besar kepada penulis;
12. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan dukungan moril yang sangat besar kepada penulis;
13. Semua pihak yang ikut membantu dalam penyelesaian usulan penelitian ini yang tak sempat penulis sebutkan satu-persatu.

Semoga Allah SWT melimpahkan balasan atas jasa-jasa mereka dan kepada penulis. Selanjutnya, penulis menyadari sepenuhnya bahwa apa yang telah dicapai ini masih jauh dari kesempurnaan dan masih banyak terdapat kekurangan. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran yang konstruktif.

Akhirnya penulis berharap semoga hasil yang telah dicapai ini dapat mendukung program pemerintah dalam mencerdaskan kehidupan bangsa serta bermanfaat bagi pengembangan ilmu pengetahuan dan teknologi khususnya di bidang ilmu komputer maupun bermanfaat bagi masyarakat.

Gorontalo, November 2019

Arfan Pontoh

DAFTAR ISI

JUDUL SKRIPSI	i
PERSETUJUAN SKRIPSI.....	ii
PENGESAHAN SKRIPSI.....	iii
PERNYATAAN SKRIPSI.....	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
BAB II LANDASAN TEORI	6
2.1 Tinjauan Studi	6
2.2 Tinjauan Pustaka	7
2.2.1 Tracer Study	7
2.2.2 Data Mining.....	7
2.2.3 Prediksi	9
2.2.4 Algoritma C 4.5	10
2.2.5 Confusion Matrix.....	13
2.2.6 Pengembangan Sistem	14
2.2.7 Desain Model.....	15
2.2.8 White Box Testing.....	19
2.2.9 Black Box Testing	22
2.2.10 Perangkat Lunak Pendukung	22

2.3	Kerangka Pemikiran	23
BAB III	METODE PENELITIAN	24
3.1	Objek Penelitian	24
3.2	Pengumpulan Data.....	24
3.3	Pengembangan Sistem.....	25
3.3.1	Analisis Sistem	26
3.3.2	Desain Sistem	27
3.3.3	Konstruksi Sistem.....	27
3.3.4	Pengujian Sistem	28
BAB IV	HASIL PENELITIAN	29
4.1	Analisis Sistem	29
4.1.1	Analisis Sistem yang Berjalan	29
4.1.2	Analisis Sistem yang Diusulkan	30
4.2	Desain Model.....	31
4.2.1	Diagram Konteks.....	31
4.2.2	Diagram Berjenjang.....	32
4.2.3	Diagram Arus Data Level 1	33
4.2.4	Diagram Arus Data Level 2.....	34
4.2.5	Kamus Data	34
4.3	Desain Sistem	36
4.3.1	Desain Output.....	36
4.3.2	Desain Input.....	38
4.3.3	Desain Basis Data.....	41
4.3.4	Desain Teknologi.....	42
BAB V	PEMBAHASAN	43
5.1	Pengujian Sistem	43
5.1.1	White Box Testing.....	43
5.1.2	Black Box Testing	45
5.2	Pembahasan Aplikasi.....	45
5.2.1	Homepage.....	45
5.2.2	Halaman Login	46
5.2.3	Halaman User	46

5.2.4	Halaman Training	47
5.2.5	Halaman Prediksi.....	47
5.3	Penyelesaian Manual C4.5	48
BAB VI KESIMPULAN DAN SARAN		61
6.1	Kesimpulan.....	61
6.2	Saran.....	61
DAFTAR PUSTAKA		62
LAMPIRAN		64

DAFTAR GAMBAR

Gambar 2.1 Bidang Ilmu Data Mining	9
Gambar 2.2 Siklus Hidup Pengembangan Sistem.....	15
Gambar 2.3 Contoh Bagan Alir White Box	20
Gambar 2.4 Contoh Grafik Alir White Box.....	20
Gambar 2.5 Kerangka Pemikiran.....	23
Gambar 3.1 Sistem yang berjalan	25
Gambar 3.2 Sistem yang Diusulkan.....	26
Gambar 4.1 Bagan Alir Dokumen: Analisis Sistem yang Berjalan	29
2 Gambar 4.2 Bagan Alir Dokumen: Sistem yang Diusulkan	30
Gambar 4.3 Diagram Konteks.....	31
Gambar 4.4 Digram Berjenjang	32
Gambar 4.5 DAD Level 1	33
Gambar 4.6 DAD Level 2	34
Gambar 4.7 Desain Output Daftar Alumni	37
Gambar 4.8 Desain Output Hasil Prediksi	37
Gambar 4.9 Desain Output Laporan Prediksi	27 38
Gambar 4.10: Desain Input Homepage	39
Gambar 4.11 Desain Input Page Login	39
Gambar 4.12 Desain Input Page User	39
Gambar 4.13 Desain Input Page Training.....	40
Gambar 4.14 Desain Input Page Prediction	40
Gambar 5.1 Flowgraph Proses Prediksi Berbasis C4.5.....	44
Gambar 5.2 Homepage.....	45
Gambar 5.3 Halaman Login.....	46
Gambar 5.4 Halaman User	46
Gambar 5.5 Halaman Training.....	47
Gambar 5.6 Halaman Prediksi	48

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	6
Tabel 2.2 Dataset Contoh Penerapan Algoritma C4.5	11
Tabel 2.3 Contoh Prediksi untuk Perhitungan Nilai Gain dan Entropy	11
Tabel 2.4 Tabel Confusion Matrix	13
Tabel 2.5 Bagan Alir Sistem	16
Tabel 2.6 Elemen–Elemen dari DFD dan Lambangnya	19
Tabel 3.1 Atribut/Variabel Input.....	24
² Tabel 4.1 Kamus Data User	35
Tabel 4.2 Kamus Data Alumni.....	35
Tabel 4.3 Kamus Data Prediksi.....	36
³ Tabel 4.4 Desain Output Secara Umum	36
Tabel 4.5 Desain Input Secara Umum	38
Tabel 4.6 Desain Basis Data Secara Umum.....	41
Tabel 4.7 Struktur Data User	41
Tabel 4.8 Struktur Data Alumni.....	41
Tabel 4.9 Struktur Data Prediksi.....	42
Tabel 5.1 Basis Path Proses Prediksi Berbasis C4.5	44
Tabel 5.2 Hasil Pengujian Black Box	45
Tabel 5.3 Dataset Waktu Tunggu Kerja Alumni Teknik Informatika	48

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Keberhasilan lulusan Perguruan Tinggi (PT) dalam memasuki dunia kerja merupakan salah satu indikator *outcome* pembelajaran dan relevansi PT bagi masyarakat. Dengan demikian, selain memberikan kompetensi yang lebih relevan untuk lulusan, PT bertanggung jawab pula memfasilitasi dan menjembatani lulusan memasuki dunia kerja. Oleh karena itu, perlu dilakukan penelusuran terhadap lulusan PT (Aptikom, 2015).

Alumni merupakan produk dari suatu institusi pendidikan. Kualitas alumni menunjukkan kualitas dari institusi pendidikan tersebut. Fakta tersebut semakin terasa, khususnya untuk alumni perguruan tinggi. Hal ini dikarenakan alumni perguruan tinggi secara langsung akan bersentuhan dengan dunia kerja. Setiap Universitas menghasilkan alumni namun banyaknya alumni yang tentunya akan mewarnai kehidupan dalam masyarakat di setiap wilayah. Untuk membekali agar alumni memiliki kesiapan yang cukup dalam menghadapi dunia kerja, setiap Universitas perlu mengupayakan beragam cara yang efektif dan efisien, tidak hanya pada kesiapan alumni tersebut dalam mengarungi dunia baru (dunia kerja). Bagaimana menjaga hubungan baik antar alumni, maupun antara alumni dengan almamaternya. Pada umumnya, ketika seorang mahasiswa telah lulus, hubungan ini menjadi renggang atau bahkan putus sama sekali. Tentunya dengan hubungan yang tetap terjaga akan membawa banyak manfaat baik dari sisi alumni maupun institusi pendidikan. Berdasarkan kenyataan tersebut, maka perlu suatu sistem alumni yang dapat digunakan dalam mengelola alumni, sehingga tercipta alumni-alumni yang berkualitas dengan tetap menjaga hubungan baik dengan almamater mereka (Arfianto, 2014).

Tracer Study (studi pelacakan) merupakan studi yang fokus utamanya untuk menggali informasi mengenai transisi dari dunia kerja pendidikan tinggi ke dunia kerja (termasuk/masa tunggu kerja), situasi pekerjaan, dan pandangan *stakeholders* (pemangku kepentingan) terhadap lulusan PT (Aptikom, 2015). *Tracer Study* yang dilakukan kepada alumni 2 tahun setelah lulus dapat dilakukan sebagai survei

pertama dan dianggap sebagai alumni yang *fresh graduate* (Dekker and Vleeshouwers, 2013).

Semakin banyaknya jumlah lulusan yang tidak sejalan dengan meningkatnya jumlah lapangan pekerjaan khususnya yang terkait dengan rumpun ilmu Teknik Informatika yang dapat di tunjukan melalui kurang relevannya jenis pekerjaan yang diperoleh lulusan dengan disiplin ilmunya (Hastuti, 2012), kurangnya lowongan pekerjaan di bidang *technopreneurship*, dan kurangnya kepercayaan *stakeholders* terhadap *programmer* lokal. Ketidakseimbangan ini dapat menyebabkan persaingan di dunia kerja yang semakin ketat bagi lulusan tersebut (kamagi dan Hansun, 2014), semakin kurangnya daya serap *stakeholders* terhadap lulusan tersebut, semakin sulitnya lulusan tersebut memperoleh pekerjaan yang relevan dengan disiplin ilmunya (Teknik Informatika), dan bahkan semakin meningkatnya jumlah pengangguran dari lulusan tersebut. Untuk itulah, program *Treacer study* terhadap *Fresh Graduate* Teknik Informatika di Gorontalo Khususnya waktu tunggu kerja mereka perlu dilakukan. Selama ini, program tersebut belum pernah dilakukan dengan tepat. *Treacer Study* yang masi dilakukan secara manual juga tentunya membutuhkan waktu, tenaga, dan biaya yang cukup besar. Dengan demikian, diperlukan suatu alat bantu yang dapat memudahkan dan memfasilitasi kegiatan tersebut, namun tentunya terlebih dahulu dilakukan analisis data untuk menemukan pendekatan/ metode yang tepat dan akurat dalam menyelesaikan masalah tersebut.

Pendekatan *Data Mining* dengan menggunakan metode-metode *machine learning* semakin berkembang dan penerapannya pun semakin luas karena karakter cerdasnya yang semakin dibutuhkan, keahliannya dalam menganalisis data, kemampuannya dalam menggali informasi yang tersembunyi pada data untuk menemukan pengetahuan baru, dan termasuk kemampuannya dalam melakukan prediksi (Jantawan dan Tsai, 2013). Pendekatan ini juga diusulkan untuk menganalisis data pendidikan (edukasi), menjawab pertanyaan-pertanyaannya, dan penelitiannya (Mishra dkk, 2016). Salah satu pendekatan/peran dalam *Data Mining* adalah klasifikasi, metode-metode *machine learning*, seperti algoritma: C4.5, *Naive Bayes* (NB), *K-Nearest Neighbor* (K-NN), *Artificial Neural Network* (ANN) –

Backpropagation (BP), dan *support Vector Machine (SVN)* paling sering digunakan (Prasetyo, 2014). Dengan pendekatan tersebut, prediksi waktu tunggu kerja *Fresh graduate* dapat dilakukan sehingga dapat memudahkan/ mendukung kegiatan *Tracer Study*.

Belum ditemukan penelitian yang menerapkan metode-metode *Machine Learning* tersebut pada prediksi waktu tunggu kerja *Fresh Graduate Teknik Informatika*. Namun penerapan untuk menganalisis data mahasiswa atau lulusan PT sudah cukup banyak penelitiannya. Penerapan dan komparasi metode-metode tersebut untuk prediksi kemungkinan kerja lulusan mengidentifikasi bahwa algoritma C4.5 yang terbaik (Rasiman dkk, 2008). Begitu pun pada prediksi mahasiswa *Droup out*, namun tidak jauh berbeda dengan NB dan ANN-BP. Sedangkan pada prediksi kelayakan kerja, algoritma NB dengan akurasi 99.77% yang terbaik kinerjanya. Penerapan ANN-BP memiliki tingkat keberhasilan 64% dalam memetakan komposisi NEM terhadap prestasi akademik lulusan . Akurasi C4.5 pada prediksi tingkat kelulusan mahasiswa, yaitu 87.5%. Akurasi NB pada prediksi waktu studi mahasiswa, yaitu 66%, sedangkan pada kinerja akademik mahasiswa, yaitu 70% (Widodo dan Hernawai, 2013).

Penelitian yang lainnya melakukan analisis korelasi antara beberapa variabel-variabel bebas yang mempengaruhi waktu tunggu kerja *fresh graduate* (variabel terikat). Variabel jenis kelamin, umur, status nikah, dan tempat tinggal dapat mempengaruhi waktu tunggu kerja *fresh graduate* (Septiana, 2013). Sementara itu, dengan *Logistic Regression (LR)*, variabel jumlah anggota keluarga, status SLTA, jenis pekerjaan yang dipilih, dan motivasi dalam mencari kerja yang signifikan mempengaruhi waktu tunggu kerja *fresh graduate* (Putranto & Mashuri, 2012). Sedangkan pencarian kerja melalui internet tidak mempersingkat waktu mencari kerja, Variabel-variabel tersebut ditambah variabel-variabel pada kuesioner *tracer study* dari (prasetyo, 2014), dapat dijadikan sebagai variabel input dalam penelitian ini.

Label class pada waktu tunggu kerja *fresh graduate* (variabel output) dapat ditentukan menggunakan teknik pemusatan data *mean* dan *median*, dimana salah satu labelnya adalah ≤ 4 bulan namun kekurangan teknik *median* adalah nilainya

yang relatif tidak stabil. Sedangkan pelaksanaan *tracer study* terhadap Jurusan Statistika di Institut Teknologi Sepuluh November, membagi waktu tunggu kerja lulusan menjadi 3 label, yaitu: < 3 bulan (cepat), 3-5 bulan (rata-rata), dan > 5 bulan (lama) (Ristekdikti, 2016). *Tracer study* pada Program Studi Pendidikan Teknik Informatika di Universitas Negeri Yogyakarta mengidentifikasi bahwa sebagian besar alumni memperoleh pekerjaan pertama, yaitu < 6 bulan. Itu sejalan dengan standar waktu tunggu kerja, yaitu ≤ 6 bulan, termasuk dalam kategori cepat. Jadi, penelitian ini dapat menggunakan 2 label, yaitu: cepat (≤ 6 bulan) dan lama (> 6 bulan) (Rasiman dan Nurhadi, 2008).

Berdasarkan penerapan tersebut, maka penelitian ini akan melakukan pemodelan dari metode *Machine learning* C4.5 untuk waktu tunggu kerja *Frash Graduate* Teknik Informatika di Gorontalo, kemudian dievaluasi kinerjanya dengan menggunakan Confusion Matrix dan ROC Curve untuk mengukur *accuracy*, *precision*, dan *recall*. Selanjutnya tiap-tiap model tersebut dikomparasi kinerjanya, menggunakan metode *T-Test*. Setelah model terbaik diperoleh, dilanjutkan dengan *development* (mereayasa alat bantu). Alat bantu tersebut akan diuji pula efisiensinya dengan menggunakan metode *white Box Testing* dan *Black Box Testing*, serta efektifitasnya dengan menggunakan metode *User Acceptance Testing*.

Dari latar belakang diatas penulisan ingin merancang sebuah sistem prediksi waktu tunggu kerja alumni teknik informatika menggunakan algoritma C4.5. Diharapkan sistem tersebut dapat mempermudah kegiatan *tracer study*.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah di uraikan, maka identifikasi masalah dalam penelitian ini adalah sulitnya untuk mengetahui waktu tunggu kerja alumni tanpa menggunakan sistem yang tepat dan akurat.

1.3 ² Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan dalam penelitian ini dapat dirumuskan masalah penelitian, yaitu: bagaimana kinerja dan efektifitas sistem prediksi waktu tunggu kerja alumni teknik informatika berbasis C4.5 yang dapat diimplementasi kan untuk mempermudah kegiatan *tracer study*?

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengimplementasikan algoritma C 4.5 pada prediksi waktu tunggu kerja alumni teknik informatika di Gorontalo dan mengembangkan sistemnya sehingga dapat diimplementasikan dalam kegiatan *tracer study*.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah:

1. Manfaat Teoritis: Memberikan kontribusi terhadap pengembangan Ilmu Komputer, khususnya pada bidang *Data Mining*, berupa penerapan algoritma C4.5 untuk prediksi waktu tunggu kerja alumni teknik informatika.
2. Manfaat Praktis: Sumbangan pemikiran, karya bahan pertimbangan, atau solusi bagi perguruan tinggi untuk kegiatan *tracer study*.

BAB II LANDASAN TEORI

2.1 Tinjauan Studi

Penelitian terkait dengan penerapan algoritma C4.5 dan prediksi waktu tunggu kerja mahasiswa ditunjukkan pada Tabel 2.1 berikut ini.

Tabel 2.1 Penelitian Terkait

Tahun	Peneliti	Masalah	Metode	Akurasi
2014	Sunjana	Tidak diketahuinya pola status nasabah untuk analisis perusahaan dalam menentukan calon nasabah dimasa yang akan datang	Decision Tree C4.5 Atribut yang digunakan penghasilan, premi dasar, cara pembayaran, dan mata uang	Status nasabah tidak lancar berjumlah 11,25 % dan nasabah yang lancar 6,75 % dengan persentase atribut premi dasar dan penghasilan maka dapat diketahui rata-rata status nasabah
2015	Nurma Jayanti, et al	Sulitnya menentukan rasio keuangan bank antara bank yang dikategorikan bangkrut dan tidak bangkrut	algoritma C4.5 dengan metode klasifikasi pohon keputusan	evaluasi indeks yang ditunjukkan pada area <i>classifier output</i> meliputi banyaknya jumlah yang diprediksi benar sebesar 94,44%, yang diprediksi salah sebesar 5,56%
2015	Firmansyah	Hasil analisis pengajuan kredit yang tidak akurat	Algoritma C4.5 dengan atribut jumlah tanggungan, level golongan, level pinjaman, dan jangka waktu	Hasil akurasi 90 %

2.2 Tinjauan Pustaka

2.2.1 Tracer Study

Pengertian *Tracer study* (studi pelacakan) merupakan studi yang fokus utamanya untuk menggali informasi mengenai lulusan yang sudah bekerja maupun yang belum bekerja (Rasiman, Cahyono, Sulianto, & Nurhadi, 2008) Informasi yang dimaksud adalah waktu tunggu tunggu kerja dan pandangan *stakeholders* (pemangku kepentingan) terhadap lulusan untuk menentukan profil lulusan dan kopetensinya PT (RISTEKDIKTI, 2016) (APTIKOM, 2015). *Tracer Study* (studi pelacakan) yang dilakukan kepada alumni 2 tahun setelah lulus (RISTEKDIKTI, 2016) dapat dilakukan sebagai survei pertama dan dianggap sebagai alumni yang *fresh graduate* (APTIKOM, 2015).

Tracer Study sudah dijadikan salah satu syarat kelengkapan akreditasi oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT), juga merupakan kelengkapan dalam dokumen Evaluasi Diri yang diperlukan dalam pengajuan proposal melalui kemendikbud. Meskipun demikian dalam kenyataannya, *tracer study* yang dilakukan masih sangat bervariasi dari segi kualitas. Pemanfaatan informasi yang diperoleh juga belum optimal. Akreditasi dilakukan di level fakultas atau program studi dengan demikian gambaran di tingkat universitas atau antar universitas belum tersedia.

2.2.2 Data Mining

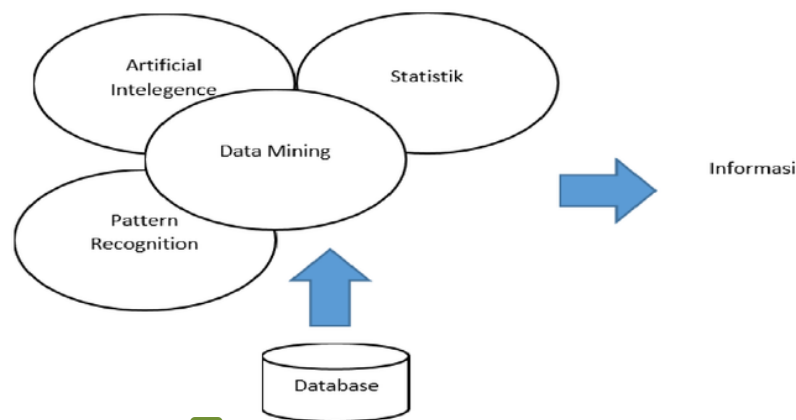
Pada *Data mining* merupakan cabang ilmu yang termasuk masih baru tetapi telah menghasilkan keuntungan yang cukup besar saat ini. Manfaat yang dirasakan dengan penggunaan *Data Mining* sudah tidak diragukan lagi. *Data Mining* sudah sering dimanfaatkan untuk mendeteksi kejadian-kejadian yang ganjil seperti penyakit tertentu, transaksi yang mencurigakan, hingga mendeteksi telepon yang dilakukan oleh pihak yang bermaksud menipu, seperti penyalahgunaan kartu kredit. Pihak pasar swalayan pun saat ini sudah menggunakan teknik *Data Mining* untuk mengatur peletakan *stand* penjualan agar mempermudah pembeli dalam pergerakannya di dalamnya.

Saat ini sudah banyak vendor-vendor yang mengkhususkan diri dalam bidang *Data Mining* seperti SPSS, Microsoft, Oracle, dan lain-lain. Produknya pun beragam dari bidang ekonomi seperti *Business Intelligent* (BI), *Business Performance Management* (BPM), *Customer Relationship Management* (CRM), dan sebagainya hingga bidang kesehatan seperti *Health Informatics*, *e-Health*, dan sebagainya. Tentu saja bidang yang lain juga telah lama menerapkan *Data Mining* seperti Telekomunikasi, Meteorologi, Biologi dan lain-lain (Widodo, Handayanto, & Herlawati, 2013).

⁵ Nama data mining sebenarnya mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis, hingga medis (Gonunescu, 2011). Data mining dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lama, dan belum melewati masa ‘remaja’ maka data mining masih diperdebatkan posisi bidang pengetahuan yang memilikinya. Maka, Daryl Pregibon menyatakan bahwa “data mining adalah campuran dari statistic, kecerdasan buatan, dan riset basis data” yang masih berkembang (Gonunescu, 2011).

⁹ Munculnya data mining didasarkan pada jumlah data yang tersimpan dalam basis data semakin besar. Misalnya dalam sebuah supermarket, ada berapa transaksi pelanggan yang terjadi dalam sehari dan ada berapa juta data yang sudah tersimpan dalam sebulan. Dalam perusahaan, ada berapa juta data yang sudah tersimpan dari setiap kegiatan produksi untuk setiap produk yang dibuat dalam beberapa tahun. Contoh lain, jika anda mempunyai kartu kredit, mungkin anda sering menerima surat penawaran barang atau jasa. Jika bank mempunyai 1.000.000 nasabah dan biaya pengiriman surat pernasabah adalah 500 rupiah, maka biaya yang harus dikeluarkan bank adalah 500 juta rupiah, padahal nasabah yang mungkin benar-benar membeli hanya sekitar 15%. Akibatnya, ada pembuangan biaya sekitar 85% dari 500 juta atau sekitar 425 juta; sungguh sia-sia. Jika perusahaan dapat memanfaatkan data-data yang ada sehingga hanya nasabah yang berpotensi untuk membeli saja yang dikirim surat, maka biaya pengiriman tersebut dapat ditekan.

18 Yang menjadi pertanyaan untuk data-data dalam perusahaan yang semakin menggunung jumlahnya dari waktu ke waktu adalah mau diapakan data tersebut, apakah hanya untuk dibuat laporan akhir tahun kemudian dibuang? Apakah hanya akan dikubur dalam gudang data dan tidak diapa-apakan? Tentu saying sekali jika data-data tersebut tidak dimanfaatkan untuk kepentingan perusahaan atau instansi-institusi yang berkepentingan (Prasetyo, Data Mining : Mengolah Data Menjadi Informasi Menggunakan Matlab, 2014).



22 Gambar 2.1 Bidang Ilmu Data Mining

2.2.3 Prediksi

17 Prediksi terbagi menjadi dua fase, yaitu pelatihan dan pengujian. Pada fase pelatihan, sebagian data yang telah diketahui kelasnya (data latih) digunakan untuk membentuk model. Pada fase pengujian model yang sudah terbentuk diuji dengan sebagian data lainnya (data uji) untuk mengetahui akurasi dari model tersebut. Selanjutnya model dapat digunakan untuk memprediksi kelas data yang belum diketahui (Kristina Pastianti, 2011). Menurut (puput hariono, 2015) 6 klasifikasi adalah sebuah proses untuk menemukan model yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui. Didalam klasifikasi diberikan sejumlah *record* yang dinamakan *training set*, yang terdiri dari beberapa atribut, atribut dapat berupa kontinyu ataupun kategori, salah satu atribut menunjukan kelas

atau *record*. Untuk mendapatkan model, kita harus melakukan analisis terhadap data latih (*training set*). Sedangkan data uji (*test set*) digunakan untuk mengetahui tingkat akurasi dari model yang telah dihasilkan, model itu sendiri bisa berupa aturan “jika-maka”, ada beberapa teknik klasifikasi yang digunakan antara lain: *decision tree*, *rule based*, *neural network*, *support vector machine*, *naive bayes*, dan *nearest neighbor*.

Menurut (Kevin Martha, 2013) klasifikasi merupakan suatu teknik data mining yang melihat sifat dari atribut, dari kelompok data yang telah didefinisikan. Teknik ini dapat digunakan untuk memberi pengetahuan pada data baru dengan memanipulasi data yang ada yang telah diklasifikasikan dan dengan menggunakan hasilnya untuk memberikan pengetahuan atau sejumlah aturan. Aturan tersebut digunakan data baru untuk dapat diklasifikasikan terhadap suatu kategori atau kelas tertentu.

2.2.4 Algoritma C 4.5

Algoritma C4.5 adalah pengembangan dari algoritma ID3. Oleh karena pengembangan tersebut algoritma C4.5 mempunyai prinsip dasar kerja yang sama dengan algoritma ID3. Hanya saja dalam algoritma C4.5 (Chih-Chiang dan Minegishi, et.al., 2009., Karegowda, et.al, 2010)

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

$$\text{Gain (S, A)} = \text{Entropy (S)} - \sum_{i=1}^n \text{Entropy (S}_i\text{)}$$

Di mana:

S = Himpunan kasus

A = Atribut

I = Jumlah Partisi Atribut

|S_i| = Jumlah Kasus pada partisi ke i

|S| = Jumlah Kasus dalam S

Secara umum algoritma C4.5 dapat dibangun menggunakan pohon keputusan (Kusrini and Lutfhi, 2010) dengan cara:

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.

Sebagai contoh, untuk membentuk pohon keputusan dapat dilakukan dengan mencari nilai Entropy dan Gain terlebih dahulu. Misalkan ditampilkan 14 dataset pengujian untuk klasifikasi C4.5 (Kusrini dan Luthfi, 2010) sebagai berikut

Tabel 2.2 Dataset Contoh Penerapan Algoritma C4.5

No	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Sumber : (Kusrini dan Luthfi, 2010)

Untuk dapat menentukan nilai-nilai Gain dan Entropy dari masing-masing atribut di atas, maka terlebih dikonversikan data ke dalam bentuk table klasifikasi yang lebih detail sebagai berikut :

Tabel 2.3 Contoh Prediksi untuk Perhitungan Nilai Gain dan Entropy

Node	Kriteria	Kategori	Jumlah Kasus (s)	No (s1)	Yes (s2)	Entropy	Gain
1	Total		14	4	10	0.863	
	Outlook						0.258
		Cloudy	4	0	4	0	
		Rainy	5	1	4	0.721	
		Sunny	5	3	2	0.970	
	Temperature						0.183
		Cool	4	0	4	0	
		Hot	4	2	2	1	
		Mild	6	2	4	0.918	
	Humidity						0.370

		High	7	4	3	0.985	
		Normal	7	0	7	0	
	Windy						0.005
		False	8	2	6	0.811	
		True	6	4	2	0.918	

Sumber : (Kusrini dan Luthfi, E.T. 2010)

Setelah tabel tersebut dibuat, maka langkah selanjutnya adalah melakukan perhitungan dengan mencari nilai Entropy menggunakan persamaan (1) Setelah menentukan nilai Entropy langkah selanjutnya yaitu mencari nilai Gain menggunakan persamaan (2). Langkah pertama yang harus dilakukan yaitu dengan mencari nilai Entropy total sebagai berikut:

Baris Total Kolom Entropy pada table 2 di hitung dengan rumus:

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

Entropy (Total)

$$= ((-4/14) * -1,807) + ((-10/14) * (-0,385))$$

$$= 0,516 + 0,346$$

$$= 0,863$$

Langkah selanjutnya, mencari nilai Entropy untuk masing-masing atribut, misalnya pada Outlook (Cloudy, Rainy dan Sunny) sebagai berikut:

$$\text{Entropy Outlook (Cloudy)} = (-0/4 * \text{Log}_2(0/4) + (-4/4) * \text{Log}_2(4/4) = 0 + 0 = 0$$

$$\text{Entropy Outlook (Rainy)} = (-1/5) * \text{Log}_2(1/5) + (-4/5) * \text{Log}_2(4/5) = 0,464 + 0,257 = 0,721$$

$$\text{Entropy Outlook (Sunny)} = (-3/5) * \text{Log}_2(3/5) + (-2/5) * \text{Log}_2(2/5) = 0,442 + 0,528 = 0,970$$

Perhitungan selanjutnya adalah menghitung nilai Gain total untuk Outlook dengan menggunakan persamaan (2) berdasarkan nilai Entropy dari masing-masing atributnya sebagai berikut:

$$\text{Gain (S, A)} = \text{Entropy (S)} - \sum_{i=1}^n$$

$$\text{Information Gain (Total, Outlook)} = \frac{|S_i|}{|S|}$$

$$= 0,863 - ((4/14 * 0) + (5/14 * 0,721) + (5/14 * 0,970))$$

$$= 0,863 - (0 + 0,240 + 0,3467) = 0,863 - 0,604 = 0,258$$

Perhitungan nilai Entropy dan gain dilakukan sampai atribut terakhir sehingga dapat diperoleh nilai-nilai Gain dan Entropy tertentu. Hasil dari nilai Gain kemudian dibandingkan dengan nilai Gain lainnya untuk mencari nilai tertinggi sehingga Gain yang merupakan nilai tertinggi akan dijadikan acuan untuk melakukan proses perhitungan kembali pada proses pembentukan struktur pohon keputusan (Kusrini dan Luthf, 2010)

2.2.5 Confusion Matrix

Confusion Matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Confusion matrix memberikan keputusan yang diperoleh penilaian performance klasifikasi berdasarkan objek dengan benar atau salah (gurunesu). Confusion matrix berisi informasi actual (actual) dan prediksi (predicted) pada sistem klasifikasi.

Tabel 2.4 Tabel Confusion Matrix

	actual	Actual
	positive	Negative
predicted positive	TP	FP
predicted negative	FN	TN

Recall	=	TP/ (TP + FN)
Precision	=	TP/ (TP + FP)
True Positive Rate	=	TP/ (TP + FN)
False Positive Rate	=	FP/ (FP + TN)

Keterangan:

1. True Positive (tp) = proporsi positif dalam data set yang diklasifikasikan positif.
2. True Negative (tn) = proporsi negative dalam data set yang diklasifikasikan negative.
3. False Positive (fp) = proporsi negative dalam data set yang diklasifikasikan positive.

4. False Negative (fn) = proporsi negative dalam data set yang diklasifikasikan

2.2.6 Pengembangan Sistem

⁵
Software (perangkat lunak) berhubungan dengan: (1) Perintah (program komputer) yang bila dieksekusi memberikan fungsi dan unjuk kerja seperti yang diinginkan; (2) Struktur data yang memungkinkan program memanipulasi informasi secara proporsional; dan (3) dokumen yang menggambarkan operasi dan kegunaan program. “*Software engineering* adalah suatu disiplin rekayasa (rancang-bangun) yang terkait dengan semua aspek produksi perangkat lunak (Pressman, *Software Engineering: A Practitioners Approach*, 2002)”. Sedangkan menurut IEEE (Institute of Electrical and Electronic Engineers):

“*Software engineering* adalah: (1) Aplikasi dari sebuah pendekatan yang bersifat kuantitatif, disiplin, dan sistematis bagi pengembangan, operasi, dan pemeliharaan perangkat lunak; (2) Studi tentang pendekatan-pendekatan seperti pada (1) (Pressman, *Software Engineering: A Practitioners Approach*, 2012).”

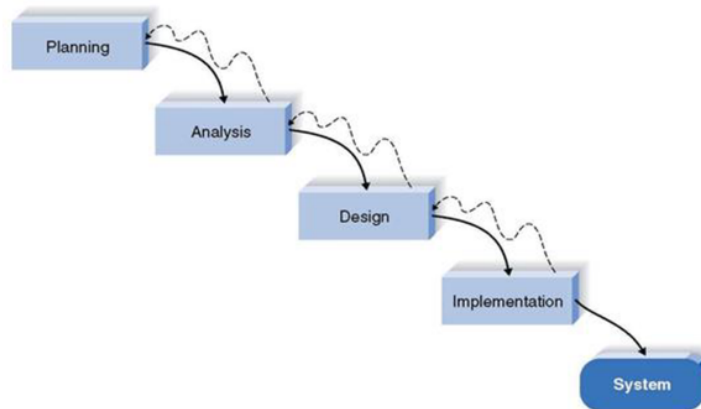
Dengan demikian bidang penelitian *software engineering* mengacu pada kedua hal tersebut.

Perbedaan utama paradigma berorientasi objek dibandingkan konvensional adalah pada penyatuan proses/fungsi dan data ke dalam bentuk yang terenkapsulasi, sedangkan paradigma konvensional memisahkan data dengan proses. Data merupakan sekumpulan atribut yang di-enkapsulasi (dikemas/dibungkus) bersama algoritma (operasi/metode/servis) untuk melakukan suatu proses berdasarkan pesan (stimulan/message/event) yang masuk ke objek/*class*. Pesan ini merupakan sarana *interface* antar-objek (Pressman, *Software Engineering: A Practitioners Approach*, 2012).

Objek-objek merupakan realitas yang unik sehingga tidak efisien bila dideskripsikan satu persatu sementara mereka memiliki beberapa kesamaan. Misal kerbau dengan sapi adalah dua objek yang berbeda, tetapi keduanya adalah pemakan rumput. Oleh sebab itu diciptakan kelas yang memuat karakteristik yang sama dari kerbau dan sapi. Dengan demikian kelas ini dapat menginformasikan berbagai binatang pemakan rumput (tetapi tidak bisa menyebutkan binatang-

binatang di kebun binatang). Kelas ini dengan mudah dikembangkan untuk merekrut objek/binatang-binatang lain pemakan rumput secara efisien tanpa mengulang-ulang deskripsi yang sama. Bahkan dapat dibuat hirarkhinya sebagai superkelas atau subkelas yang masing-masing memiliki warisan (*inheritance*) dari kelas di atasnya. Sedangkan *Polymorphism* diperlukan guna memperluas sistem (*software engineering*) berbasis objek yang ada secara efisien, yakni dengan nama yang sama tetapi operasinya berbeda (Pressman, *Software Engineering: A Practitioners Approach*, 2012).

Sistem Development Life Cycle (SDLC) berorientasi objek dengan pendekatan *waterfall* dapat ditunjukkan pada gambar berikut ini.



Sumber: (Alan Dennis, 2010)




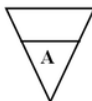
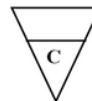

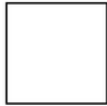

Gambar 2.2 Siklus Hidup Pengembangan Sistem



2.2.7 ² Desain Model




Tahap desain model terbagi menjadi dua, yaitu desain model secara umum dan terinci. Tahap desain model secara umum berupa desain sistem secara fisik dan logika. Desain fisik dapat digambarkan dengan bagan alir sistem dan bagan alir dokumen, dan desain secara logika digambarkan dengan diagram arus data (DAD). Pada tahap desain model terinci, model akan mendefinisikan secara rinci urutan-urutan langkah dari masing-masing proses yang digambarkan di DAD. Urut-urutan langkah proses ini diwakili oleh suatu program komputer.

² Bagan alir sistem merupakan bagan yang menunjukkan arus pekerjaan secara keseluruhan dari sistem. ¹³ Bagan alir sistem digambar dengan simbol-simbol berikut.

Tabel 2.5 Bagan Alir Sistem

NAMA SIMBOL	SIMBOL	KETERANGAN
Dokumen		Menunjukkan Simbol Dokumen input dan output baik untuk proses manual, mekanik atau komputer
Kegiatan Manual		Menunjukkan pekerjaan Manual
Simpan Offline		File non komputer yang diarsip urut angka (<i>Numerical</i>)
		File non komputer yang diarsip urut Huruf (<i>Alphabetical</i>)
		File non komputer yang diarsip urut Tanggal (<i>Cronological</i>)
Kartu Plong		Menunjukkan input /output yang menggunakan kartu plong (<i>punched card</i>)
Operasi Luar		Simbol operasi yang dilakukan diluar proses operasi komputer
Proses		¹⁵ mbol kegiatan proses dari operasi program komputer

NAMA SIMBOL	SIMBOL	KETERANGAN
Pengurutan Offline		Proses pengurutan data diluar proses komputer
Pita Magnetik		Menunjukkan <i>Input</i> / <i>Ouput</i> menggunakan pita magnetic
Hardisk		Menunjukkan <i>Input</i> / <i>Ouput</i> menggunakan Harddisk
Diskette		Menunjukkan <i>Input</i> / <i>Ouput</i> menggunakan Diskette
Drum Magnetik		Menunjukkan <i>Input</i> / <i>Ouput</i> menggunakan drum magnetic
Pita Kertas Berlubang		Menunjukkan <i>Input</i> / <i>Ouput</i> menggunakan pita kertas berlubang
Keyboard		Menunjukkan <i>Input</i> / <i>Ouput</i> menggunakan Keyboard
Pita Kontrol		Menunjukkan Penggunaan pita kontrol (<i>control tape</i>) dalam batch control tetap untuk mencocokkan diproses batchprocessing
Hubungan Komunikasi		Menunjukkan proses transmisi
³ Display		Menunjukkan <i>output</i> yang ditampilkan dilayar

NAMA SIMBOL	SIMBOL	KETERANGAN
² Garis Alir		Menunjukkan Arus dari proses
Penjelasan		Menunjukkan penjelasan dari suatu proses
Penghubung		Menunjukkan penghubung halaman yang masih sama atau ke halaman lain

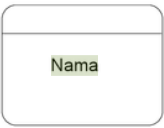
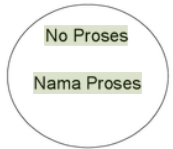




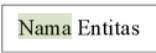
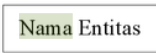
(Sumber : Jogiyanto, 2013)

Untuk memudahkan penggambaran suatu sistem yang ada atau sistem yang baru yang akan dikembangkan secara logika tanpa memperhatikan lingkungan fisik di mana data tersebut mengalir atau lingkungan fisik di mana data tersebut akan disimpan, maka digunakannya Diagram Arus Data (DAD) atau *Data Flow Diagram* (DFD). Dalam menggambarkan sistem perlu dilakukan pembentukan simbol, berikut ini simbol-simbol yang sering digunakan dalam DAD :

1. *External entity* (kesatuan luar) atau *boundary* (batas sistem), kesatuan luar merupakan kesatuan di lingkungan luar sistem yang dapat berupa orang, organisasi atau sistem lain yang berada di lingkungan luarnya yang akan memberikan input serta menerima *output* dari sistem.
2. *Data flow* (arus data), arus data ini menunjukkan arus atau aliran data yang dapat berupa masukan untuk sistem atau hasil dari proses sistem.
3. *Process* (proses), suatu proses adalah kegiatan atau kerja yang dilakukan orang, mesin atau komputer dari hasil suatu arus data yang masuk ke dalam proses untuk dihasilkan arus data yang akan keluar dari proses.
4. *Data store* (simpanan data), simbol-simbol dari Diagram Arus Data yang dapat dilihat pada tabel 2.6 berikut ini.

10

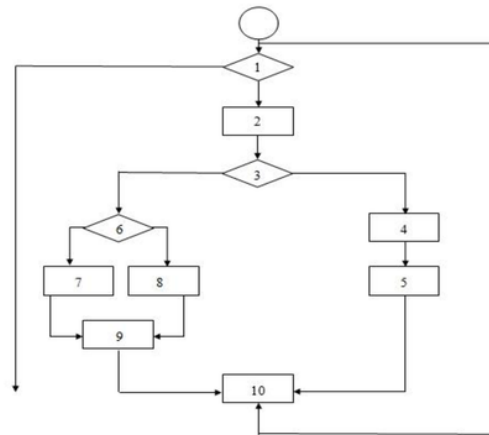
Tabel 2.6 Elemen–Elemen dari DFD dan Lambangnya

Elemen Data Flow Diagram	Field Tipikal yang biasa digunakan	Simbol Gene and Sarson	Simbol De Marco and Jourdan
Setiap Proses memiliki: Nomor Nama Deskripsi proses Satu/lebih output Data flow Satu/lebih input Flow	Label (Nama) Type (Proses) Deskripsi Nomor Proses		
Setiap Data Flow memiliki: Nama Deskripsi Satu/lebih koneksi Ke suatu proses	Label Deskripsi Alias Komposisi (Deskripsi dari elemen data)		
Setiap Data Store memiliki: Nomor Nama Deskripsi Satu/lebih input Data flow Satu/lebih output	Label (Nama) Type Deskripsi Alias Komposisi Catatan		
Setiap Entitas Eksternal memiliki: Nama Deskripsi	Label, Type Deskripsi Alias Deskripsi Entitas		

Sumber: Al Fatah (2010)

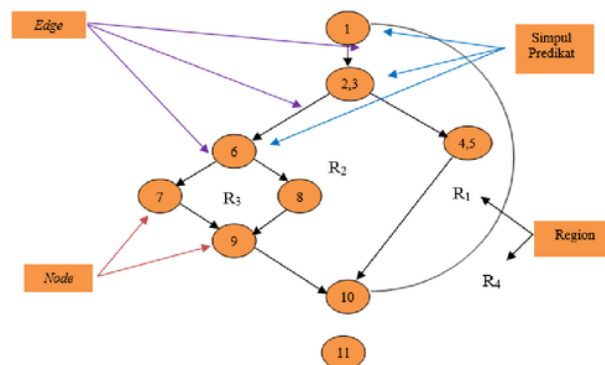
2.2.8 White Box Testing

Pengujian *White box*, adalah metode pengujian yang menggunakan struktur kontrol desain produser untuk ²test case. Yang memberikan jaminan bahwa semua jalur independen pada suatu modul telah digunakan paling tidak satu kali, menggunakan semua keputusan logis pada sisi true dan false, mengeksekusi semua loop pada batasan mereka dan pada batas operasional mereka, ³dan menggunakan strujtur dan internal untuk menjamin validitasnya. (Sumber: Roger S. Pressman, 2012:536)



Gambar 2.3 Contoh Bagan Alir White Box

Bagan alir digunakan untuk menggambarkan struktur control program dan untuk menggambarkan grafik alir, harus memperhatikan representasi desain procedural pada bagan alir. Pada gambar dibawah ini, grafik alir memetakan bagan alir tersebut ke dalam grafik alir yang sesuai (dengan mengasumsikan bahwa tidak ada kondisi senyawa yang diisikan dalam diamond keputusan bagan alir tersebut). Masing-masing lingkaran, yang disebut simpul grafik alir, mempresentasikan satu atau lebih statemen procedural. Urutan kotak proses dan permata keputusan dapat memetakan simpul tunggal. Anak panah tersebut yang disebut edges atau links, merepresentasikan aliran control dan analog dengan anak panah bagan alir. Edge harus berhenti pada suatu simpul, meskipun pada simpul tersebut tidak merepresentasikan statemen prosedural. (Sumber: Roger S. Pressman, 2012:536)



Gambar 2.4 Contoh Grafik Alir White Box

Kompleksitas siklomatis adalah metrics perangkat lunak yang memberikan pengukuran kuantitatif terhadap kompleksitas logis suatu program. Bila metriks ini digunakan dalam konteks metode pengujian basis path, maka nilai yang terhitung untuk kompleksitas siklomatis menentukan jumlah jalur independen. Jalur independen adalah jalur yang melalui program yang mengintroduksi sedikitnya satu rangkain statemen proses baru atau suatu kondisi baru. Bila dinyatakan dengan terminologi grafik alir, jalur independen harus bergerak sepanjang paling tidak satu edge yang tidak dilewatkan sebelum jalur tersebut ditemukan. Sebagai contoh, serangkaian jalur independen untuk grafik alir yang ditunjukkan pada gambar di atas adalah:

Jalur 1 : 1 - 11

Jalur 2 : 1 - 2 - 3 - 4 - 5 - 10 - 1 - 11

Jalur 3 : 1 - 2 - 3 - 6 - 8 - 9 - 10 - 1 - 11

Jalur 4 : 1 - 2 - 3 - 6 - 7 - 9 - 10 - 1 - 11

Jalur 1,2,3, dan 4 yang ditentukan diatas terdiri dari sebuah basisi set untuk grafik alir pada gambar 2.3. Bagaimana kita tahu banyaknya jalur yang dicari? Komputasi kompleksitas siklomatis memberikan jawaban. Fondasi kompleksitas siklomatis adalah teori grafik, dan memberi kita metrics perangkat lunak yang sangat berguna. Kompleksitas dihitung dalam salah satu dari tiga cara berikut:

1. Jumlah region grafik alir sesuai dengan kompleksitas siklomatis.
2. Kompleksitas siklomatis, $V(G)$, untuk grafik G ditentukan sebagai $V(G) = E - N + 2$ dimana E adalah jumlah edge grafik alir dan N adalah jumlah simpul grafik alir.
3. Kompleksitas siklomatis, $V(G)$, untuk grafik alir G juga ditentukan sebagai $V(G) = P + 1$, dimana P adalah jumlah simpul predikat yang diisikan dalam grafik alir.

Pada gambar tersebut, grafik alir, kompleksitas siklomatis dapat dihitung dengan menggunakan masing-masing dari algoritma yang ditulis diatas:

1. Grafik alir mempunyai 4 region
2. $V(G) = 11 \text{ edge} - 9 \text{ simpul} + 2 = 4$

3. $V(G) = 3$ simpul yang diperkirakan $+ 1 = 4$ Dengan demikian, kompleksitas siklomatis dari grafik alir pada gambar 2.3.

Yang lebih penting, nilai untuk $V(G)$ member kita batas atas untuk jumlah jalur independen yang membentuk basisi set, dan implikasinya, batas atas.

2.2.9 Black Box Testing

Selanjutnya *software* diuji pula dengan metode *Black Box Testing* yang focus pada keperluan fungsional dari *software* dan berusaha untuk menemukan kesalahan dalam beberapa kategori, diantaranya:

1. Fungsi-fungsi yang salah atau hilang.
2. Kesalahan interface.
3. Kesalahan dalam struktur data atau akses basis data external.
4. Kesalahan performa.
5. Kesalahan inisialisasi dan terminasi.

Jika sudah tidak ada kesalahan-kesalahan tersebut, maka sistem dinyatakan efisien dari segi kesalahan komponen-komponen sistem.

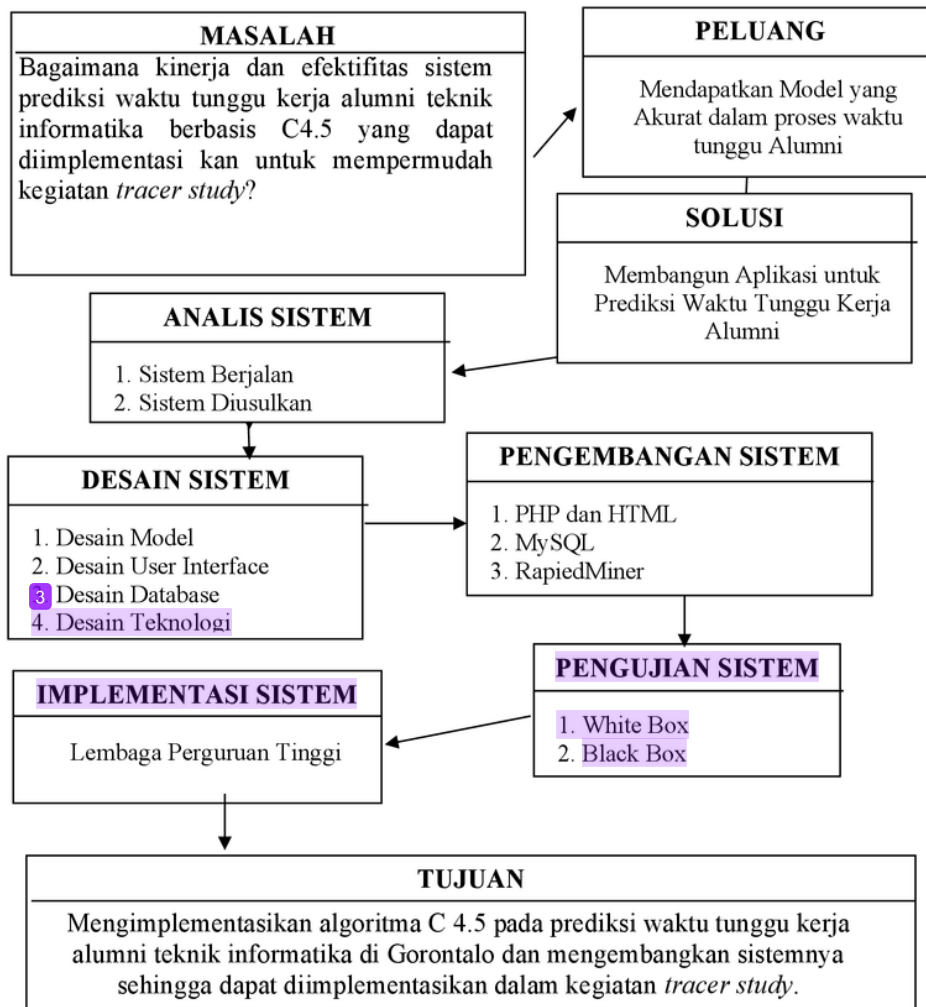
2.2.10 Perangkat Lunak Pendukung

Perangkat lunak pendukung untuk membangun sistem ini, seperti pada tabel dibawah ini:

Tabel 2.7 Perangkat lunak pendukung

No	TOOLS	24 KEGUNAAN
1	Xampp	Digunakan sebagai server yang berdiri sendiri (localhost), yang terdiri atas program Apache HTTP Server, MySQL database, dan penerjemah bahasa yang ditulis dengan bahasa pemrograman PHP dan Perl untuk Pengembangan sistem Penerapan Algoritma C4.5 Pada Prediksi Waktu Tunggu Kerja Alumni.
2.	MySQL	Digunakan untuk mengolah data sistem prediksi Waktu Tunggu Kerja Alumni.

2.3 Kerangka Pemikiran



Gambar 2.5 Kerangka Pemikiran

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Objek dan penelitian ini adalah waktu tunggu kerja untuk *treacer study* Alumni Teknik Informatika dan menggunakan metode penelitian eksperimen. Penelitian ini dimulai pada bulan Desember 2018 sampai April 2019.

3.2 Pengumpulan Data

Dalam melakukan pengumpulan data, ada beberapa cara yang digunakan, di antaranya adalah:

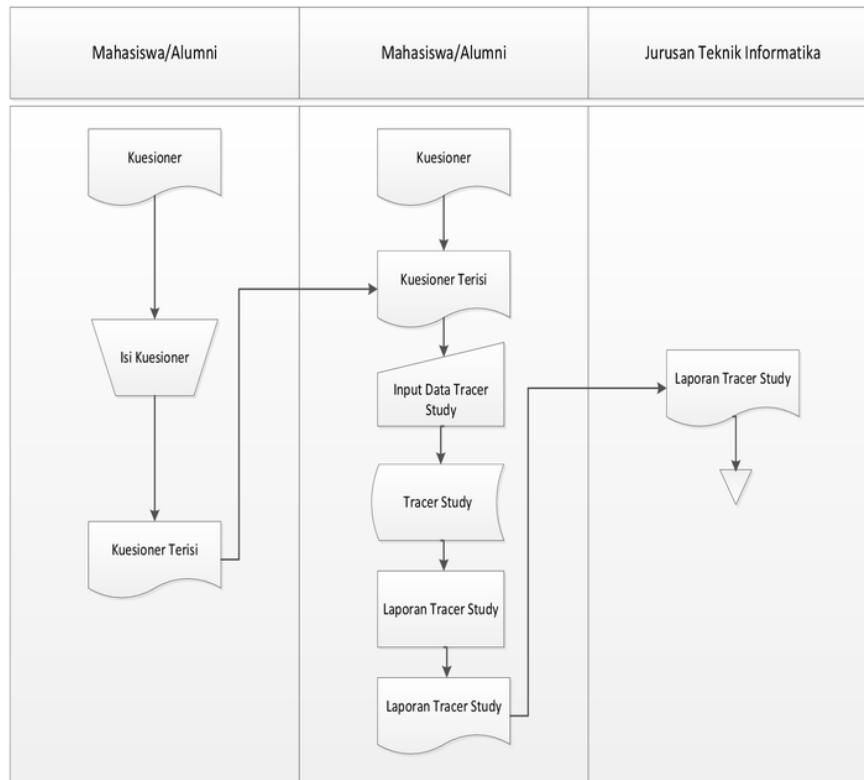
1. Observasi: dilakukan pengamatan langsung dilapangan mengenai data Alumni yang didapat dilokasi penelitian.
2. Wawancara: dilakukan kepada pihak yang terkait yakni Alumni Teknik Informatika.
3. Kuesioner: dibuat untuk dibagikan kepada Alumni Teknik Informatika yang ada di Perguruan Tinggi Negeri/Swasta Gorontalo.
4. Dokumentasi: teknik ini digunakan untuk mengumpulkan data sekunder yaitu, buku-buku, papers, dan sebagainya.

Label pada atribut/variabel *output/class*, yaitu: cepat (≤ 6 bulan), lama (> 6 bulan) dan Belum Bekerja. Sedangkan atribut/variabel *input* ditunjukkan pada Tabel 3.1 berikut ini.

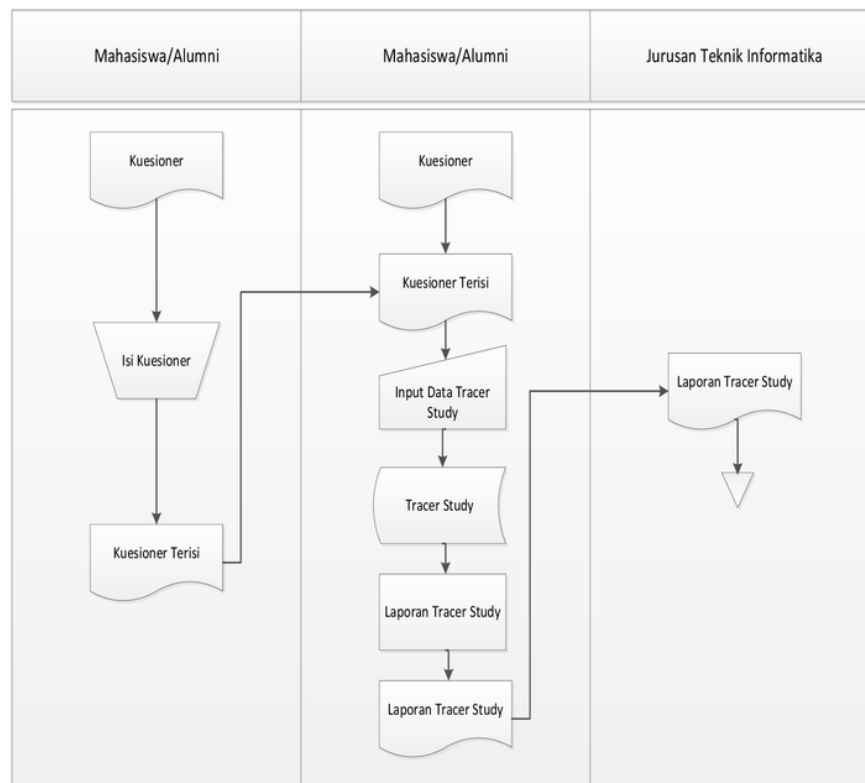
Tabel 3.1 Atribut/Variabel Input

No.	Attribut / Variabel	Value	Tipe Data
1	Tahun Lulus	2010 - 2018	Numerik
2	Jenis Kelamin	L, P	Binominal
3	Usia Saat Ini	20 - 50	Numerik
4	Tempat Tinggal (Kota, Luar Kota)	Opsi: 1 – 2	Nominal
5	Status Pernikahan (Menikah, Belum Menikah)	Opsi: 1 – 2	Nominal
6	Jumlah Anggota Keluarga	0 – 15 Orang	Numerik
7	Konsentrasi	Opsi: 1 – 5	Ordinal
8	Keluarga	Opsi: 1 – 13	Nominal
9	Thaun Lulus	Opsi: 1 – 4	Ordinal
10	IPK	Opsi: 1 – 5	Ordinal
11	Prediksi	Opsi: 1 – 5	Ordinal

3.3 Pengembangan Sistem



Gambar 3.1 Sistem yang berjalan



Gambar 3.2 Sistem yang Diusulkan

3.3.1 Analisis Sistem

Pada tahap ini dilakukan analisis terhadap sistem yang akan direkayasa meliputi :

1. **Analisis Sistem Berjalan.** Pada tahap ini dilakukan analisis kebutuhan dan masalah dalam perekayasa sistem yang akan dibuat, kemudian menetapkan sistem yang akan direkayasa dalam penelitian ini. Pengembangan sistem juga harus memperhatikan representasi pengetahuan dari pakar, sehingga sistem pakar yang akan digunakan sesuai dengan kebutuhan setiap pengguna.
2. **Analisis Sistem yang Diusulkan.** Pada tahap ini dilakukan pendalaman tentang kejelasan sasaran, kejelasan tujuan dari sistem informasi pengolahan data, kejelasan sistem yang akan direkayasa serta bimbingan teknis penggunaan sistem..

3. ² **Sumber Data.** Sumber data yang digunakan pada tahap ini adalah data primer, yaitu data yang diperoleh secara langsung dengan menggunakan teknik representasi pengetahuan.
4. ² **Alat.** Alat yang digunakan pada tahap ini adalah *flowchart*, diagram konteks, dan diagram alir data.

3.3.2 Desain Sistem

Desain sistem menggunakan pendekatan berorientasi objek yang digambarkan dalam bentuk:

1. *Architecture Design*, menggunakan alat bantu Visio dalam bentuk:
 - a. Model jaringan dari sistem *stand alone*.
 - b. Spesifikasi *hardware* dan *software* yang direkomendasikan.
2. *Interface Design*, menggunakan alat bantu Visio dalam bentuk:
 - a. Mekanisme User
 - b. Mekanisme Navigasi
 - c. Mekanisme Input
 - d. Mekanisme Output
3. *Data Design*, menggunakan alat bantu Visio dalam bentuk:
 - a. Format data yang digunakan *Relational Database (SQL)*
 - b. Struktur Data
 - c. *Databae Diagram*
4. *Program Design*, menggunakan alat bantu visio dalam bentuk:
 - a. Sistem yang berjalan dan Sistem yang diusulkan
 - b. Diagram Arus Data
 - c. Kamus Data
 - d. Rancangan Sistem Umum dan Terinci

3.3.3 Konstruksi Sistem

³ Pada tahap ini menerjemahkan hasil pada tahap analisis dan desain ke dalam kode-kode program komputer kemudian membangun aplikasinya. Model yang

telah diperoleh kemudian dibuatkan aplikasinya. ⁴ Alat bantu yang digunakan pada tahap ini adalah XAMPP dengan Bahasa Pemrograman PHP. Alat bantu *database* yang digunakan *MySql*.

3.3.4 Pengujian Sistem

Software yang telah direkayasa kemudian diuji dengan metode *White Box Testing* pada kode program proses penerapan metode/modelnya. Kode program tersebut dibuatkan *flowchart* programnya, kemudian ³ dipetakan kedalam bentuk *flowgraph* (bagan alir kontrol) yang tersusun dari beberapa *node* dan *edge*. Berdasarkan *flowgraph*, ditentukan jumlah *Region* dan *Cyclomatic Complexity* (CC). apabila *independent path* = $V(G) = (CC) = Region$, di mana setiap *path* hanya dieksekusi sekali dan sudah benar, maka sistem dinyatakan efisien dari segi kelayakan logika pemrograman.

¹² Selanjutnya *software* diuji pula dengan metode *Black Box Testing* yang focus pada keperluan fungsional dari *software* dan berusaha untuk menemukan kesalahan dalam beberapa kategori, diantaranya:

1. Fungsi-fungsi yang salah atau hilang.
2. Kesalahan *interface*.
3. Kesalahan dalam struktur data atau akses basis data external.
4. Kesalahan performa.
5. Kesalahan inisialisasi dan terminasi.

Jika sudah tidak ada kesalahan-kesalahan tersebut, maka sistem dinyatakan efisien dari segi kesalahan komponen-komponen sistem.

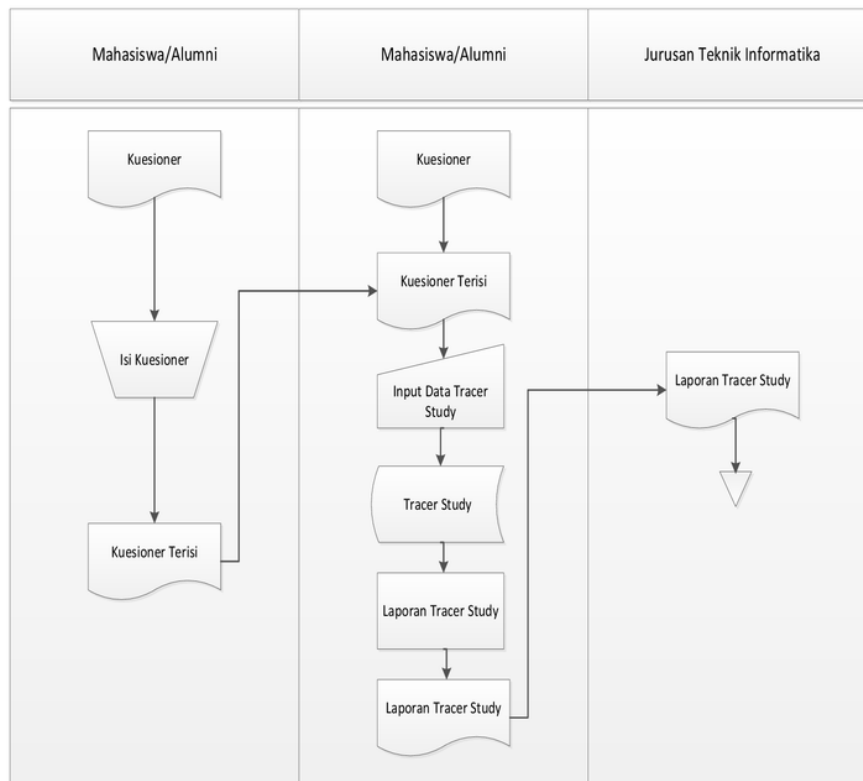
BAB IV HASIL PENELITIAN

4.1 Analisis Sistem

2

4.1.1 Analisis Sistem yang Berjalan

Analisis sistem yang berjalan ditunjukkan dalam bentuk bagan alir dokumen pada Gambar 4.1 berikut ini.

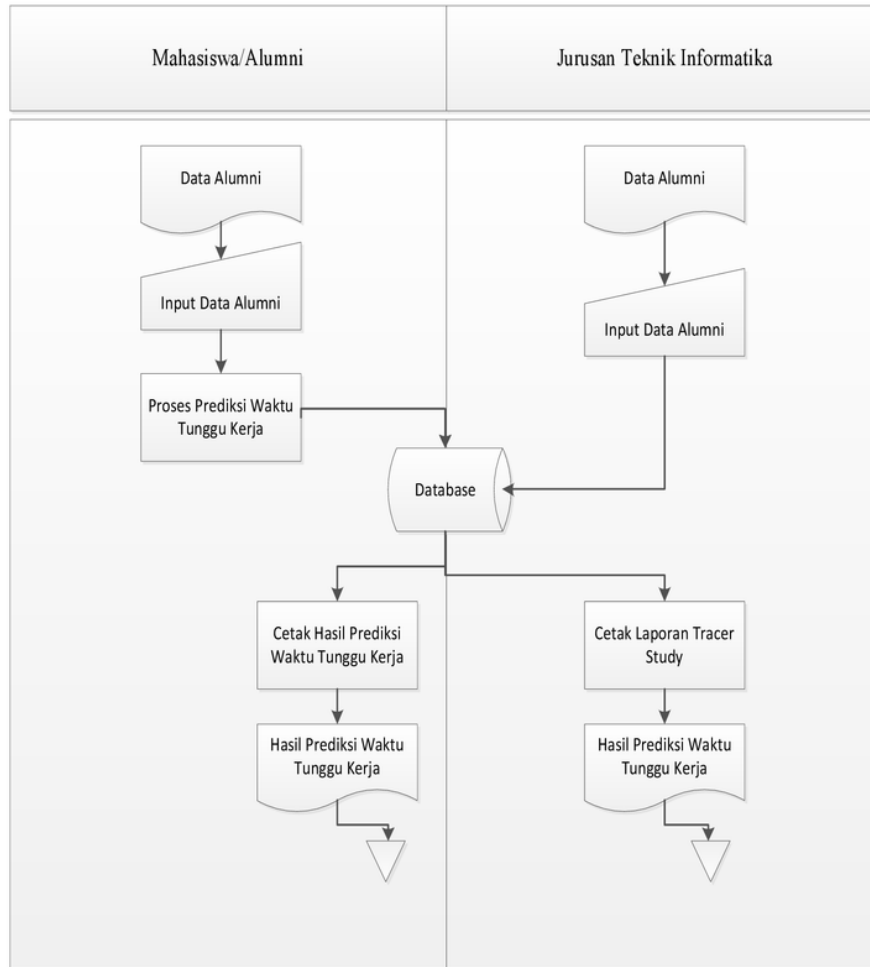


Gambar 4.1 Bagan Alir Dokumen: Analisis Sistem yang Berjalan

Sistem yang berjalan terdiri dari dua entitas, yaitu alumni dan jurusan. Sistem yang berjalan menggunakan teknik pengumpulan data kuesioner dalam melakukan *tracer study*, termasuk terkait dengan waktu tunggu kerja alumni. Kuesioner disebar kemudian diisi oleh alumni yang selanjutnya akan diproses secara manual oleh jurusan untuk memperoleh informasi tentang alumni.

4.1.2 Analisis Sistem yang Diusulkan

Analisis sistem yang diusulkan ditunjukkan dalam bentuk bagan alir dokumen pada Gambar 4.2 berikut ini.



Gambar 4.2 Bagan Alir Dokumen: Sistem yang Diusulkan

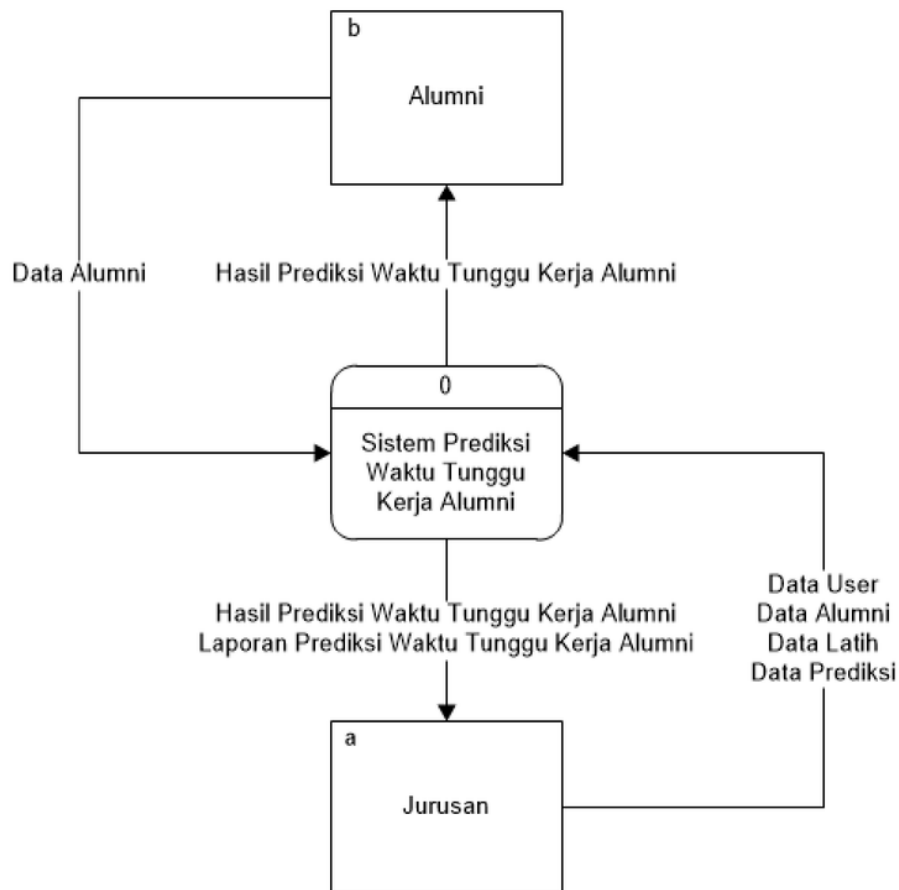
Pada sistem yang diusulkan telah menggunakan aplikasi berbasis web untuk memperoleh informasi mengenai waktu tunggu kerja alumni, sehingga memudahkan jurusan dalam melakukan salah satu kegiatan *tracer study*.

4.2 Desain Model

Hasil desain model dari sistem yang diusulkan disajikan dalam bentuk diagram konteks, diagram arus data, dan kamus data yang ditunjukkan pada berbagai gambar-gambar dan tabel-tabel berikut ini.

4.2.1 Diagram Konteks

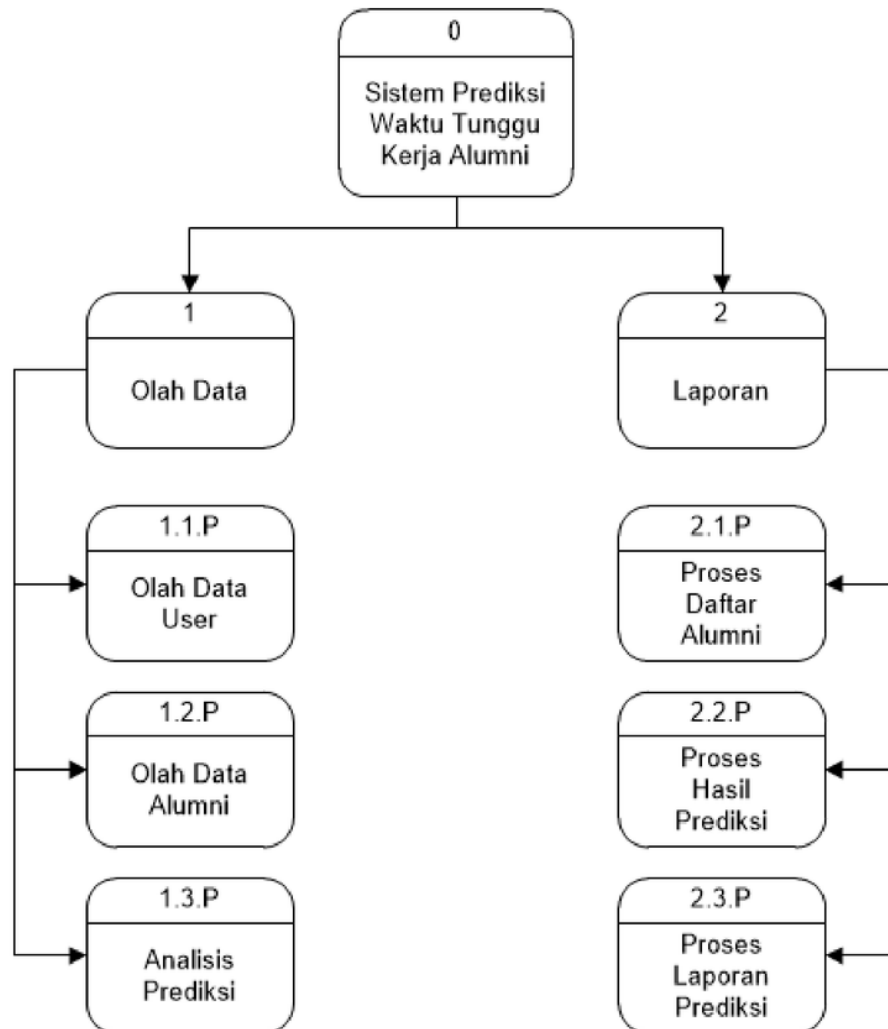
Diagram Konteks Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 ditunjukkan pada Gambar 4.3 berikut ini.



Gambar 4.3 Diagram Konteks

4.2.2 Diagram Berjenjang

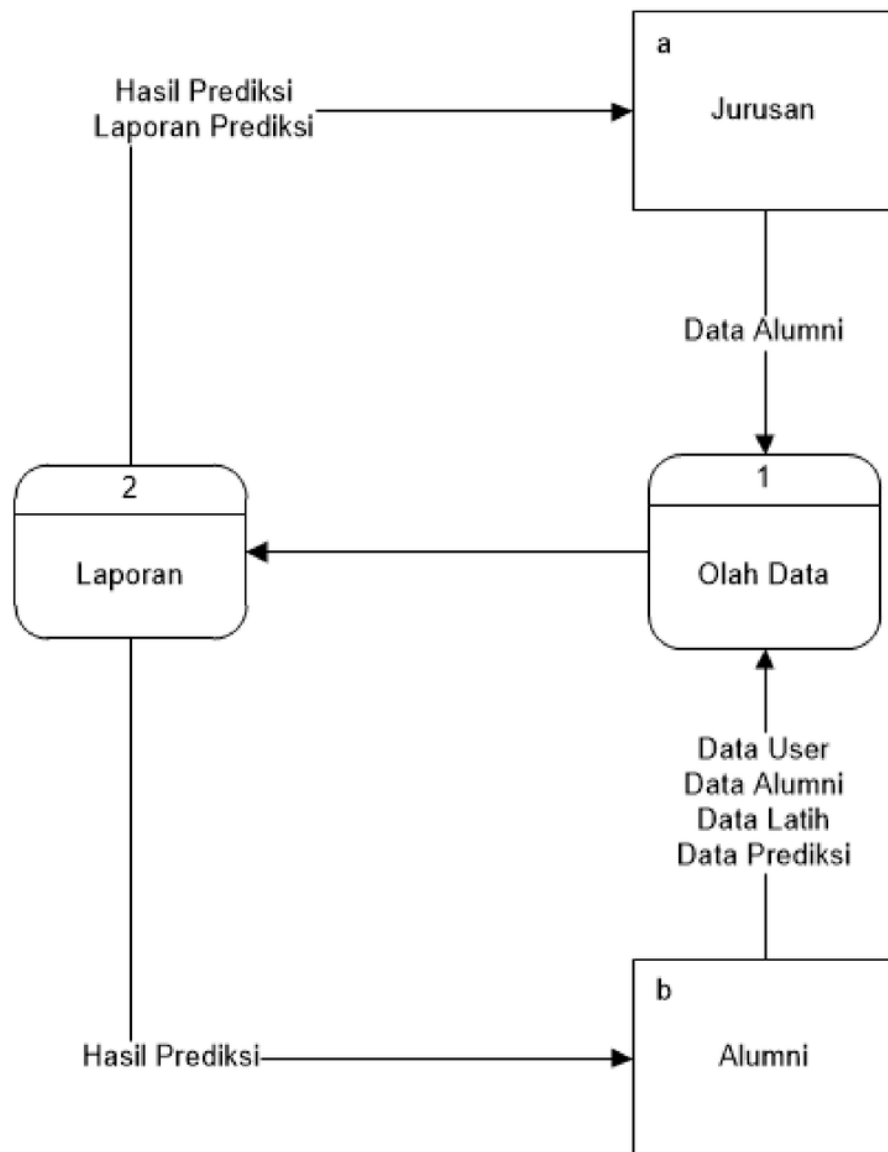
Diagram Berjenjang Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 ditunjukkan pada Gambar 4.4 berikut ini.



Gambar 4.4 Diagram Berjenjang

4.2.3 Diagram Arus Data Level 1

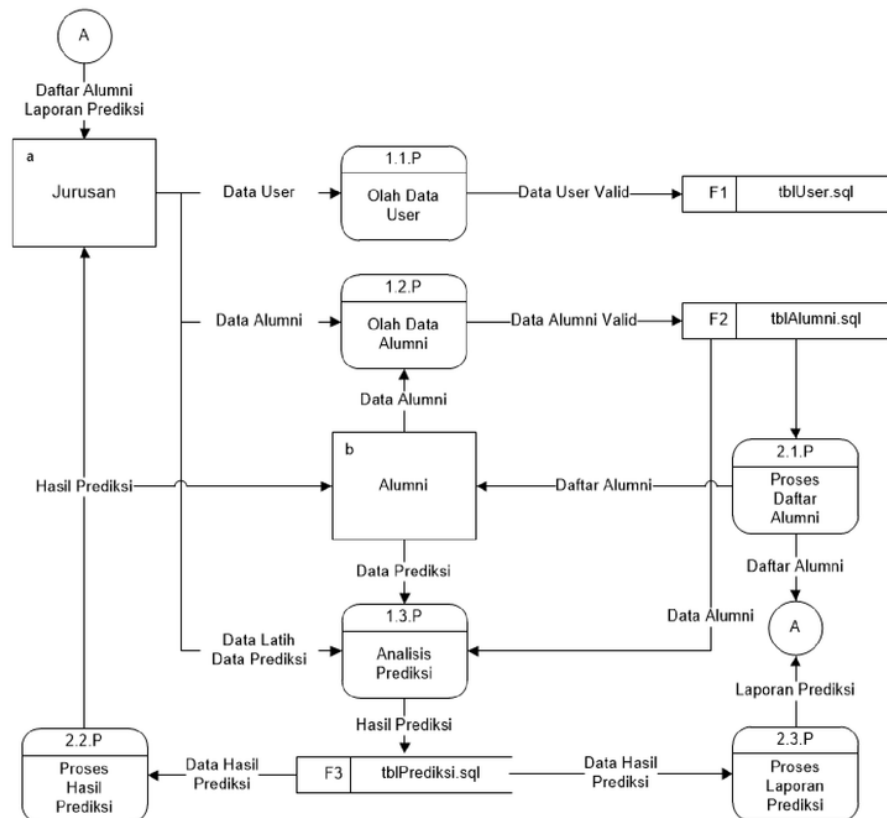
Diagram Arus Data Level 1 Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 ditunjukkan pada Gambar 4.5 berikut ini.



Gambar 4.5 DAD Level 1

4.2.4 Diagram Arus Data Level 2

Diagram Arus Data Level 2 Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 ditunjukkan pada Gambar 4.6 berikut ini.



Gambar 4.6 DAD Level 2

4.2.5 Kamus Data

Kamus Data dari Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 merupakan katalog tentang data dan kebutuhan-kebutuhan informasi yang digunakan untuk merancang input, database dan output sistem. Kamus data ini dibuat berdasarkan arus data yang mengalir pada DAD yang ditunjukkan pada tabel-tabel berikut ini.

Tabel 4.1 Kamus Data User

Nama Arus Data : Data User				
Alias : -				
Bentuk Data : .sql				
Penjelasan : Merupakan data pengguna (admin dan operator) aplikasi				
Periode : Setiap kali ada penginputan, perubahan, dan penghapusan data user				
Arus : a – 1.1.P – F1				
Volume : Tidak tentu				
Struktur Data :				
No	Item	Tipe	Ukuran	Keterangan
1.	userId	Char	3	ID pengguna
2.	userName	Varchar	100	Nama pengguna
3.	userNamaLengkap	Varchar	100	Nama lengkap pengguna
4.	userPassword	Varchar	100	Password pengguna
5.	userJenisKelamin	Varchar	100	Jenis kelamin pengguna
6.	userStatusAdmin	Varchar	100	Status admin pengguna

Tabel 4.2 Kamus Data Alumni

Nama Arus Data : Data Alumni				
Alias : -				
Bentuk Data : .sql				
Penjelasan : Merupakan data alumni jurusan teknik informatika				
Periode : Setiap kali ada penginputan, perubahan, dan penghapusan data alumni				
Arus : a,b – 1.2.P – F2 – 2.1.P – a,b				
Volume : Tidak tentu				
Struktur Data :				
No	Item	Tipe	Ukuran	Keterangan
1.	alumniId	Char	10	ID alumni
2.	alumniNama	Varchar	100	Nama alumni
3.	alumniPT	Varchar	100	Asal perguruan tinggi alumni
4.	alumniJinkel	Varchar	100	Jenis kelamin alumni
5.	alumniUmur	Integer	4	Umur alumni
6.	alumniStatusKawin	Varchar	100	Status pernikahan alumni
7.	alumniKeluarga	Varchar	100	Jumlah anggota keluarga alumni
8.	alumniKonsentrasi	Varchar	100	Konsentrasi saat kuliah
9.	alumniTahunLulus	Integer	4	Tahun lulus kuliah
10.	alumniIPK	Float	4	IPK alumni

Tabel 4.3 Kamus Data Prediksi

Nama Arus Data : Data Latih, Data Prediksi, Data Hasil Prediksi				
Alias : -				
Bentuk Data : .sql				
Penjelasan : Merupakan data untuk analisis prediksi waktu tunggu kerja alumni				
Periode : Setiap kali ada penginputan, perubahan, dan penghapusan data				
Arus : a,b – 1.3.P – F3 – 2.2.P, 2.3.P – a,b				
Volume : Tidak tentu				
Struktur Data :				
No	Item	Tipe	Ukuran	Keterangan
1.	preNomor	Char	10	Nomor prediksi
2.	alumniId	Char	10	ID alumni
3.	userId	Char	3	ID user
4.	preJinkel	Varchar	100	Jenis kelamin alumni
5.	preUmur	Integer	4	Umur alumni
6.	preStatusKawin	Varchar	100	Status pernikahan alumni
7.	preKeluarga	Varchar	100	Jumlah anggota keluarga alumni
8.	preKonsentrasi	Varchar	100	Konsentrasi saat kuliah
9.	preTahunLulus	Integer	4	Tahun lulus kuliah
10.	preIPK	Float	4	IPK alumni
11.	preClassPrediksi	Varchar	10	Label class prediksi

4.3 Desain Sistem

4.3.1 Desain Output

Desain output secara umum dari Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 ditunjukkan pada Tabel 4.4 berikut ini.

Tabel 4.4 Desain Output Secara Umum

No	Nama	Tipe	Format	Media	Alat	Distribusi	Periode
1.	Daftar Alumni	Internal	Tabel	Kertas, Layar	Monitor, Printer	Jurusan Alumni	-
2.	Hasil Prediksi	Internal	Tabel	Kertas, Layar	Monitor, Printer	Jurusan Alumni	-
3.	Laporan Prediksi	Internal	Tabel	Kertas, Layar	Monitor, Printer	Jurusan	-

Desain output secara terinci dari Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 ditunjukkan pada Gambar 4.7, Gambar 4.8, dan Gambar 4.9 berikut ini.

[Page Header]			
Daftar Alumni			@Filter
[Report Header]			
@Perguruan_Tinggi			
@alamat; @tlp;			
[Group Header]			
@Asal Perguruan Tinggi			
ID	Nama	...	IPK
ID ke-1	Nama ke-1	...	IPK ke-1
ID ke-2	Nama ke-2	...	IPK ke-2
...
ID ke-n	Nama ke-n	...	IPK ke-n
[Group Footer]			
[Details Footer]			
[Report Footer]			
			Ketua Jurusan
			@Nama_Ketua_Jurusan
[Page Footer]			
Dicetak Oleh : @User; Pada Tgl MM/dd/yyyy			Page N to M

Gambar 4.7 Desain Output Daftar Alumni

[Page Header]			
Hasil Prediksi Waktu Tunggu Kerja Alumni Berbasis C4.5			
[Report Header]			
@Perguruan_Tinggi			
@alamat; @tlp;			
[Details]			
Nomor	:	@preNomor	
ID Alumni	:	@alumniId	
Asal Perguruan Tinggi	:	@alumniPT	
Jenis Kelamin	:	@preJinkel	
Umur	:	@preUmur	
Staus Kawin	:	@preStatusKawin	
Jumlah Anggota Keluarga	:	@preKeluarga	
Konsentrasi Jurusan	:	@preKonsentrasi	
Tahun Lulus	:	@preTahunLulus	
IPK	:	@preIPK	
Waktu Tunggu Kerja	:	@preClassPrediksi	
[Report Footer]			Ketua Jurusan
			@Nama_Ketua_Jurusan
[Page Footer]			
Dicetak Oleh : @User; Pada Tgl MM/dd/yyyy			Page N to M

Gambar 4.8 Desain Output Hasil Prediksi

[Page Header]			
Laporan Prediksi Waktu Tunggu Kerja Alumni Berbasis C4.5			@Filter
[Report Header]			
@Perguruan_Tinggi			
@alamat; @tlp;			
[Group Header]			
@preClassPrediksi			
Nomor	ID Alumni	...	Waktu Tunggu Kerja
No ke-1	ID ke-1	...	Waktu Tunggu Kerja ke-1
No ke-2	ID ke-2	...	Waktu Tunggu Kerja ke-2
...
No ke-n	ID ke-n	...	Waktu Tunggu Kerja ke-n
[Group Footer]			
[Details Footer]			
[Report Footer]			
Ketua Jurusan			
@Nama_Ketua_Jurusan			
[Page Footer]			
Dicetak Oleh : @User; Pada Tgl MM/dd/yyyy			Page N to M

Gambar 4.9 Desain Output Laporan Prediksi

4.3.2 Desain Input

Desain input secara umum dari Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 ditunjukkan pada Tabel 4.5 berikut ini.

Tabel 4.5 Desain Input Secara Umum

No	Nama	Tipe	Format	Media	Alat	Diakses Oleh	Periode
1.	Homepage	Internal	Page	Layar	Keyboard, Mouse	Semua User	-
2.	Page Login	Internal	Page	Layar	Keyboard, Mouse	Semua User	-
3.	Page User	Internal	Page	Layar	Keyboard, Mouse	Jurusan	-
4.	Page Training	Internal	Page	Layar	Keyboard, Mouse	Jurusan	-
5.	Page Prediction	Internal	Page	Layar	Keyboard, Mouse	Jurusan, Alumni	-

Desain input secara terinci dari Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 ditunjukkan pada Gambar 4.10, Gambar 4.11, Gambar 4.12, Gambar 4.13, dan Gambar 4.14 berikut ini.

Search

Barner Slide

User Training Prediction Logout

PREDIKSI WAKTU TUNGGU KERJA ALUMNI TEKNIK INFORMATIKA

Page

Status

6
Gambar 4.10: Desain Input Homepage

LOGIN

Your Email or User Name:

Your Password

Login

Gambar 4.11 Desain Input Page Login

DATA USER

[Input Data Baru](#)

ID	Nama Lengkap	User Name	Password	Jenis Kelamin	Status Admin	Aksi
						[edit] [hapus]
						[edit] [hapus]
						...
						[edit] [hapus]

Gambar 4.12 Desain Input Page User

DATA TRAINING

Unggah Data:

Choose File

 UPLOAD

[Tambah Data Training](#)

No	Nama	...	Waktu Tunggu Kerja	Aksi
				[edit] [hapus] [edit] [hapus] [edit] [hapus] [edit] [hapus]

Gambar 4.13 Desain Input Page Training

INPUT DATA PREDICTION

Nomor :
 Nama :
 Asal Perguruan Tinggi :
 Jenis Kelamin :
 Umur :
 Status Kawin :
 Jumlah Anggota Keluarga :
 Konsentrasi Jurusan :
 Tahun Lulus :
 IPK :

Hapus Item
 Prediksi

Gambar 4.14 Desain Input Page Prediction

4.3.3 Desain Basis Data

Desain basis data secara umum dari Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis C4.5 ditunjukkan pada Tabel 4.6 berikut ini.

Tabel 4.6 Desain Basis Data Secara Umum

Kode	Nama	Tipe	Media	Organisasi	Field Kunci
F1	tblUser.sql	Master	Server	Index	userId
F2	tblAlumni.sql	Master	Server	Index	alumniId
F3	tblPrediksi.sql	Transaksi	Server	Index	preNomor

Desain basis data secara terinci dari Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 disajikan dalam bentuk struktur data yang ditunjukkan pada Tabel 4.7, Tabel 4.8, dan Tabel 4.9 berikut ini.

Tabel 4.7 Struktur Data User

Nama : Data User Tipe : Master Primary Key : userId Foreign Key : - Media : Server Fungsi : Merupakan data pengguna aplikasi Struktur Data :				
No	Atribut	Tipe	Ukuran	Keterangan
1.	userId	Char	3	ID pengguna
2.	userName	Varchar	100	Nama pengguna
3.	userNameLengkap	Varchar	100	Nama lengkap pengguna
4.	userPassword	Varchar	100	Password pengguna
5.	userJenisKelamin	Varchar	100	Jenis kelamin pengguna
6.	userStatusAdmin	Varchar	100	Status admin pengguna

Tabel 4.8 Struktur Data Alumni

Nama : Data Alumni Tipe : Master Primary Key : alumniId Foreign Key : - Media : Server Fungsi : Merupakan data alumni teknik informatika Struktur Data :				
No	Item	Tipe	Ukuran	Keterangan
1.	alumniId	Char	10	ID alumni
2.	alumniNama	Varchar	100	Nama alumni
3.	alumniPT	Varchar	100	Asal perguruan tinggi alumni
4.	alumniJinkel	Varchar	100	Jenis kelamin alumni
5.	alumniUmur	Integer	4	Umur alumni
6.	alumniStatusKawin	Varchar	100	Status pernikahan alumni
7.	alumniKeluarga	Varchar	100	Jumlah anggota keluarga alumni
8.	alumniKonsentrasi	Varchar	100	Konsentrasi saat kuliah
9.	alumniTahunLulus	Integer	4	Tahun lulus kuliah
10.	alumniIPK	Float	4	IPK alumni

Tabel 4.9 Struktur Data Prediksi

Nama : Data Prediksi Tipe : Transaksi Primary Key : preNomor Foreign Key : alumniId, userId Media : Server Fungsi : Merupakan data prediksi waktu tunggu kerja alumni teknik informatika Struktur Data :				
No	Item	Tipe	Ukuran	Keterangan
1.	preNomor	Char	10	Nomor prediksi
2.	alumniId	Char	10	ID alumni
3.	userId	Char	3	ID user
4.	preJinkel	Varchar	100	Jenis kelamin alumni
5.	preUmur	Integer	4	Umur alumni
6.	preStatusKawin	Varchar	100	Status pernikahan alumni
7.	preKeluarga	Varchar	100	Jumlah anggota keluarga alumni
8.	preKonsentrasi	Varchar	100	Konsentrasi saat kuliah
9.	preTahunLulus	Integer	4	Tahun lulus kuliah
10.	preIPK	Float	4	IPK alumni
11.	preClassPrediksi	Varchar	10	Label class prediksi

4.3.4 Desain Teknologi

Minimum Spesifikasi *Hardware* dan *Software*:

1. Processor Intel Atom atau Pentium 4 atau Dual Core;
2. 1 GB RAM;
3. 800 x 600 Resolusi Display;
4. Ms. Windows 7 untuk *OS*;
5. MySQL Server 5;
6. Web Browser.

Rekomendasi Spesifikasi *Hardware* dan *Software*:

1. Processor Core2 Duo keatas;
2. 2 GB RAM atau lebih;
3. 1366 x 768 Resolusi Display;
4. Ms. Windows 10 untuk *OS*;
5. MySQL Server 5;
6. Web Browser.

BAB V PEMBAHASAN

5.1 Pengujian Sistem

Metode pengujian sistem yang digunakan untuk menguji kinerja Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 adalah *White Box Testing* dan *Black Box Testing*.

5.1.1 White Box Testing

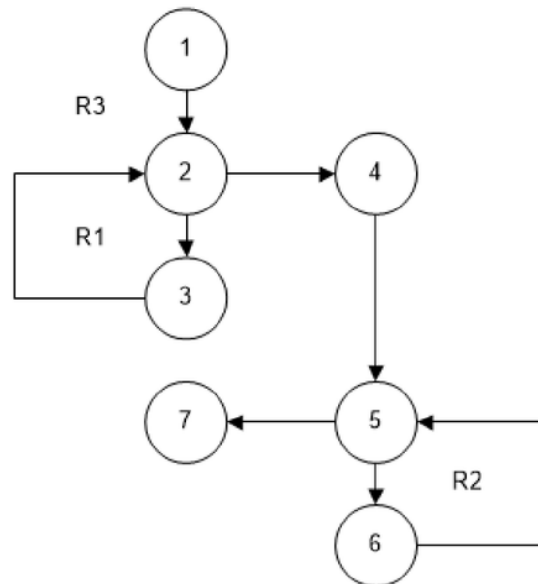
Kode program yang diuji adalah kode program pada proses prediksi berbasis C4.5. Berikut ini adalah *pseudocode* kode program tersebut.

```

init ..... (1)
while ($row ← mysql_fetch_array($query)) { ..... (2)
    $nm ← $row['nama_lengkap']; ..... (3)
    $ap ← $row['asal_perguruan']; ..... (3)
    $kelamin ← $row['jenis_kelamin']; ..... (3)
    $usia ← $row['umur']; ..... (3)
    $kawin ← $row['status_kawin']; ..... (3)
    $famili ← $row['keluarga']; ..... (3)
    $kons ← $row['konsentrasi']; ..... (3)
    $thn ← $row['tahun_lulus']; ..... (3)
    6 $nilai ← $row['ipk']; //echo $nm; }; ..... (3)
    $sql2 ← mysql_query("SELECT count(no) as no FROM training "); ..... (4)
    while ($dt2 ← mysql_fetch_array($sql2)) { ..... (5)
        S ← select count(jumlah_data); ..... (6)
        S1 ← select count(jumlah_data_≤6_bulan); ..... (6)
        S2 ← select count(jumlah_data_>6_bulan); ..... (6)
        X1 ← select count(x=variabel and l=labelVariabel1); ..... (6)
        X2 ← select count(x=variabel and l=labelVariabel2); ..... (6)
        X11 ← select count(x=variabel and l=labelVariabel1 and y=labelClass1); ..... (6)
        X12 ← select count(x=variabel and l=labelVariabel1 and y=labelClass2); ..... (6)
        X21 ← select count(x=variabel and l=labelVariabel2 and y=labelClass1); ..... (6)
        X22 ← select count(x=variabel and l=labelVariabel2 and y=labelClass2); ..... (6)
        ES ← -((log2(S1/S)*(S1/S))+(log2(S2/S)*(S2/S))); ..... (7)
        EX11 ← -((log2(X11/X1)*(X11/X1))+(log2(X12/X1)*(X12/X1))); ..... (7)
        EX12 ← -((log2(X21/X2)*(X21/X2))+(log2(X22/X2)*(X22/X2))); ..... (7)
        EX1 ← -((log2(X1/S)*(X1/S))+(log2(X2/S)*(X2/S))); ..... (7)
        GiX1 ← ((X1/S)*EX11)+((X2/S)*EX12); ..... (7)
        GX1 ← ES - GiX1; ..... (7)
        GRX1 ← GX1 / EX1; ..... (7)
    }
}

```

Selanjutnya *flowgraph* dari kode program proses prediksi berbasis C4.5 di atas tersebut ditunjukkan pada Gambar 5.1 berikut ini.



Gambar 5.1 Flowgraph Proses Prediksi Berbasis C4.5

Dari *flowgraph* proses prediksi berbasis C4.5, didapatkan:

Diketahui: $\text{Region (R)} = 2$
 $\text{Node (N)} = 7$
 $\text{Edge (E)} = 8$
 $\text{Predikat Node (P)} = 2$

Rumus: $V(G) = (E - N) + 2$ atau $V(G) = P + 1$

Penyelesaian: $V(G) = (8 - 7) + 2 = 3$
 $V(G) = 2 + 1 = 3$

Tabel 5.1 Basis Path Proses Prediksi Berbasis C4.5

NO	PATH	INPUT	OUTPUT	KET.
1.	1-2-3-2-4-5-7	Data prediksi dan umlah data setiap atribut.	Label class prediksi	OK
2.	1-2-4-5-6-5-7	Data prediksi dan entropy.	Label class prediksi	OK
3.	1-2-3-2-4-5-6-5-7	Data prediksi, jumlah data setiap atribut, entropy.	Label class prediksi	OK

5.1.2 Black Box Testing

Hasil pengujian Black Box Testing ditunjukkan pada Tabel 5.2 berikut ini.

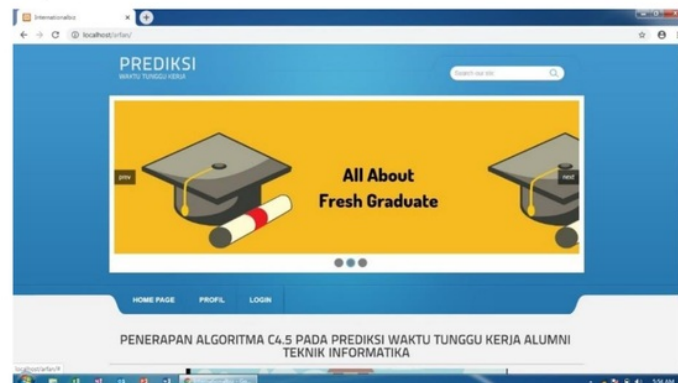
Tabel 5.2 Hasil Pengujian Black Box

NO	INPUT / EVENT	FUNGSI	HASIL	HASIL
1.	Login	Menginput password pengguna aplikasi untuk dapat menggunakan aplikasi.	7. Jika password salah, maka ulangi memilih user 8. Jika password benar, maka tutup login dan masuk ke homepage.	Sesuai
3.	Menu-menu	Menampilkan data user, dan training, dan data prediction untuk mengolah data-data tersebut.	Data user, training, dan prediction tampil dan siap untuk diolah.	Sesuai
4.	Input, edit, dan hapus data user.	Menginput, merubah, dan menghapus data user.	9. Data bertambah 10. Data berubah 11. Data terhapus	Sesuai
5.	Input, edit, dan hapus data training.	Menginput, merubah, dan menghapus data training.	12. Data bertambah 13. Data berubah 14. Data terhapus	
6.	Input dan proses prediksi.	Menginput dan memproses prediksi	Data terinput, diproses menggunakan C4.5, dan hasil prediksi ditampilkan	Sesuai

5.2 Pembahasan Aplikasi

5.2.1 Homepage

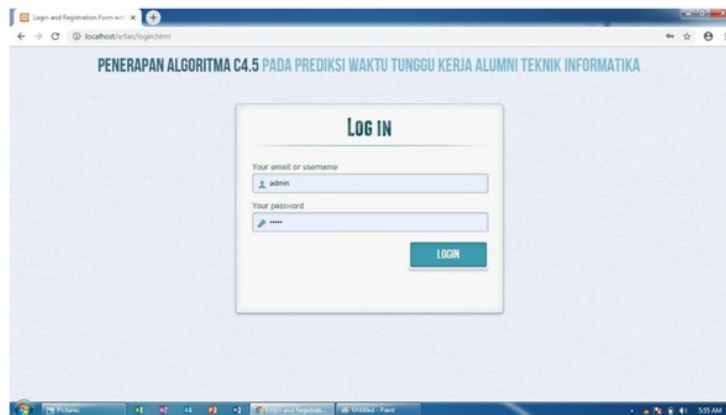
Homepage merupakan halaman utama aplikasi ini berisi menu-menu yang terdiri dari menu *user*, *training*, *prediction*, dan *logout*. Tampilan *homepage* ditunjukkan pada Gambar 5.2 berikut ini.



Gambar 5.2 Homepage

5.2.2 Halaman Login

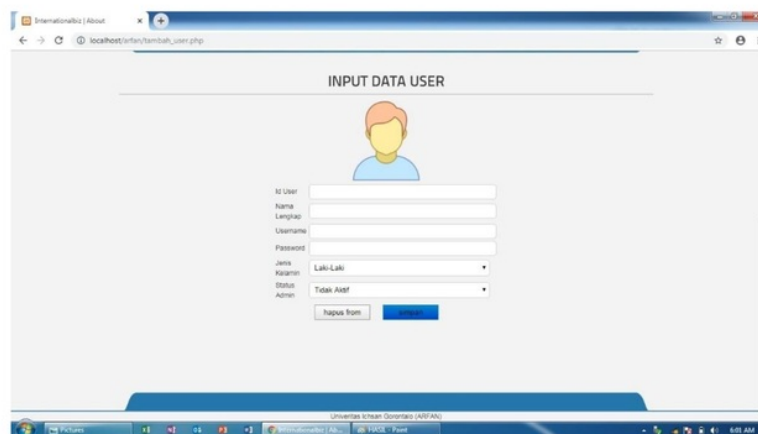
Halaman *login* digunakan untuk menginput ID dan *password* pengguna aplikasi agar dapat menggunakan aplikasi ini. Tampilan halaman login ditunjukkan pada Gambar 5.3 berikut ini.



Gambar 5.3 Halaman Login

5.2.3 Halaman User

Halaman *user* digunakan untuk mengolah (menambah, merubah, dan menghapus) data pengguna aplikasi. Tampilan halaman *user* ditunjukkan pada Gambar 5.4 berikut ini.



Gambar 5.4 Halaman User

5.2.4 Halaman Training

Halaman *training* digunakan untuk mengolah (menambah, merubah, dan menghapus) data latih yang akan dianalisis untuk memproses suatu data prediksi menggunakan algoritma C4.5. Tampilan halaman *training* ditunjukkan pada Gambar 5.5 berikut ini.

No.	Nama Lengkap	Asal Perguruan Tinggi	Jenis Kelamin	Umur	Status Kawan	Kejuruan	Keahlian	Tahun Lulus	Waktu Tunggu 1-5 6-10 >10
1	Adi, Miki Limon	Universitas Khairi Gorontalo	Laki-laki	25	Menantan	5	Demogramen Web	2012	4
2	Adi, T. Anan	Universitas Khairi Gorontalo	Laki-laki	27	Menantan	5	Demogramen Visual	2013	4
3	Bearta Eka Aprila Adi	Universitas Khairi Gorontalo	Laki-laki	27	Menantan	7	Framework Programming	2013	3
4	Djuma Polungge	Universitas Khairi Gorontalo	Perempuan	27	Menantan	5	Desain Grafis Dan Animasi	2013	4
5	Jelly Maharam	Universitas Khairi Gorontalo	Perempuan	23	Menantan	4	IT Audit	2010	3
6	Paksa Tuludu	Universitas Khairi Gorontalo	Perempuan	27	Menantan	4	Pengolahan Data	2013	4
7	Hapsah Umar Al	Universitas Khairi Gorontalo	Laki-laki	27	Menantan	5	Cryptography	2013	4
8	Hamda Gede	Universitas Khairi Gorontalo	Perempuan	27	Menantan	6	Linux Operating System	2013	3
9	Harwan Kabulu	Universitas Khairi Gorontalo	Laki-laki	27	Belum Menant	4	Logika Dan Pemrograman	2013	4
10	Jafar Yahya	Universitas Khairi Gorontalo	Laki-laki	26	Belum Menant	2	Microprocessor	2010	4

Gambar 5.5 Halaman Training

5.2.5 Halaman Prediksi

Halaman prediksi digunakan untuk menginput data yang akan diprediksi menggunakan algoritma C4.5 dan menampilkan hasil prediksinya. Tampilan halaman prediksi ditunjukkan pada Gambar 5.6 berikut ini.

The screenshot shows a web browser window with the address bar displaying 'localhost/infany/tambah_testing.php'. The page title is 'INPUT DATA TESTING'. Below the title is a placeholder for a profile picture. The form contains the following fields and values:

- No: 12
- Nama Lengkap: (empty)
- Asal perguruan tinggi: Universitas Ichsan Gorontalo
- Jenis Kelamin: Laki-Laki
- Umur: 22
- Status kawin: Menikah
- Keluarga: 2
- Konsentrasi: pemrograman web
- Tahun lulus: 2010
- IPK: 3

At the bottom of the form are two buttons: 'hapus from' and 'simpan'.

Gambar 5.6 Halaman Prediksi

5.3 Penyelesaian Manual C4.5

Tabel 5.3 berikut ini merupakan *dataset* waktu tunggu kerja alumni teknik informatika. Atribut No, Nama dan Asal Perguruan Tinggi alumni tidak diproses, namun hanya sebagai ID.

Tabel 5.3 Dataset Waktu Tunggu Kerja Alumni Teknik Informatika

No.	JENIS KELAMIN	UMUR	STATUS	KELUARGA	KONSENTRASI	TAHUN LULUS	IPK	CLASS
1.	Laki-laki	(>25)	Menikah	(<=5)	Pemrograman	(<=2012)	(<=3,5)	(<= 6 Bulan)
2.	Laki-Laki	(>25)	Menikah	(>5)	Pemrograman	(>2012)	(>3,5)	(<= 6 Bulan)
...
96	Perempuan	(<=25)	Menikah	(<=5)	Jaringan	(<=202)	(.3,5)	(> 6 Bulan)

Formula *Entropy* variabel output (*class*):

$$E(S) = - \sum_{i=1}^m P(W_i|S) \log_2 P(W_i|S)$$

$E(S)$: menyatakan *Entropy* dari *class*.

m : menyatakan banyaknya label *class*.

$P(W_i|S)$: menyatakan proporsi label *class* ke- i (W_i) dalam semua data latih yang diproses dalam node (S), yaitu jumlah data milik label *class* ke- i (W_i) dibagi dengan jumlah seluruh data latih yang diproses dalam node (S).

Formula *Entropy* variabel input (atribut):

$$E(X_j) = - \sum_{i=1}^k P(X_i|S) \text{Log}_2 P(X_i|S)$$

$E(X_j)$: menyatakan *Entropy* dari atribut ke- j (X_j).

k : menyatakan banyaknya label atribut ke- j .

$P(X_i|S)$: menyatakan proporsi label ke- i (X_i) atribut ke- j dalam semua data latih yang diproses dalam node (S), yaitu jumlah data milik label ke- i (X_i) atribut ke- j dibagi dengan jumlah seluruh data latih yang diproses dalam node (S).

Formula *Entropy* label dari atribut:

$$E(X_{j,l}) = - \sum_{i=1}^m P(V_i|S) \text{Log}_2 P(V_i|S)$$

$E(X_{j,l})$: menyatakan *Entropy* dari label ke- l atribut ke- j ($X_{j,l}$).

m : menyatakan banyaknya label *class*.

$P(V_i|S)$: menyatakan proporsi label *class* ke- i dari label ke- l atribut ke- j (V_i) dalam semua data latih milik atribut ke- j yang diproses dalam node (S), yaitu jumlah data milik label *class* ke- i dari label ke- l atribut ke- j (V_i) dibagi dengan jumlah data milik atribut ke- j yang diproses dalam node (S).

Formula *Information Gain* variabel input (atribut):

$$GI(X_j) = \sum_{i=1}^k P(X_i|S) E(S_i)$$

$GI(X_j)$: menyatakan *Information Gain* dari atribut ke- j (X_j).

k : menyatakan banyaknya label atribut ke- j .

$P(X_i|S)$: menyatakan proporsi label ke- i (X_i) atribut ke- j dalam semua data latih yang diproses dalam node (S), yaitu jumlah data milik label ke- i (X_i) atribut ke- j dibagi dengan jumlah seluruh data latih yang diproses dalam node (S).

$E(S_i)$: menyatakan *Entropy* dari label ke- i atribut ke- j (S_i).

Formula *Gain* variabel input (atribut):

$$G(X_j) = E(S) - GI(X_j)$$

$G(X_j)$: menyatakan *Gain* dari atribut ke- j (X_j).

$E(S)$: menyatakan *Entropy* dari *class*.

$GI(X_j)$: menyatakan *Information Gain* dari atribut ke- j .

Formula *Gain Ratio* variabel input (atribut):

$$GR(X_j) = \frac{\text{Max}(G(X_j))}{E(X_j)}$$

$GR(X_j)$: menyatakan *Gain Ratio* dari atribut ke- j (X_j).

$\text{Max}((G(X_j)))$: menyatakan *Gain* dari atribut ke- j yang maksimum.

$E(X_j)$: menyatakan *Entropy* dari atribut ke- j .

Langkah pertama yang dilakukan adalah dengan melakukan diskretisasi menggunakan pendekatan binerisasi (*binning*) terhadap atribut-atribut bertipe numerik, yaitu pada atribut Umur, Jumlah Anggota Keluarga, Tahun Lulus, dan IPK. Hasil diskretisasi atribut-atribut tersebut, yaitu:

Umur = $\{ \leq 25; > 25 \}$

Jumlah Anggota Keluarga = $\{ \leq 5; > 5 \}$

Tahun Lulus = $\{ \leq 2012; > 2012 \}$

IPK = $\{ \leq 3,5; > 3,5 \}$

Pemodelan C4.5 berdasarkan data tersebut adalah sebagai berikut.

Node Akar:

Entropy class:

Jumlah data (S) = 96

Jumlah data label *class* ≤ 6 Bulan ($S_{\leq 6 \text{Bulan}}$) = 60

Jumlah data label *class* > 6 Bulan ($S_{> 6 \text{Bulan}}$) = 36

$$E(S) = - \left(\left(\left(\frac{60}{96} \right) \log_2 \left(\frac{60}{96} \right) \right) + \left(\left(\frac{36}{96} \right) \log_2 \left(\frac{36}{96} \right) \right) \right) = 0,9544$$

Entropy, Information Gain, Gain, dan Gain Ratio pada atribut Jenis Kelamin (X_1):

Jumlah data $X_{1, \text{Laki-laki}}$ = 58

Jumlah data $X_{1, \text{Perempuan}}$ = 38

Jumlah data $X_{1, \text{Laki-laki}}$ untuk $S_{\leq 6 \text{Bulan}}$ = 39

Jumlah data $X_{1, \text{Laki-laki}}$ untuk $S_{> 6 \text{Bulan}}$ = 19

Jumlah data $X_{1, \text{Perempuan}}$ untuk $S_{\leq 6 \text{Bulan}}$ = 21

Jumlah data $X_{1, \text{Perempuan}}$ untuk $S_{>6 \text{Bulan}}$ = 17

$$E(X_{1, \text{Laki-laki}}) = - \left(\left(\frac{39}{58} \log_2 \frac{39}{58} \right) + \left(\frac{19}{58} \log_2 \frac{19}{58} \right) \right) = 0,9124$$

$$E(X_{1, \text{Perempuan}}) = - \left(\left(\frac{21}{38} \log_2 \frac{21}{38} \right) + \left(\frac{17}{38} \log_2 \frac{17}{38} \right) \right) = 0,9920$$

$$E(X_1) = - \left(\left(\frac{58}{96} \log_2 \frac{58}{96} \right) + \left(\frac{38}{96} \log_2 \frac{38}{96} \right) \right) = 0,9685$$

$$GI(X_1) = \left(\frac{58}{96} 0,9124 \right) + \left(\frac{38}{96} 0,9920 \right) = 0,9439$$

$$G(X_1) = 0,9544 - 0,9439 = 0,0105$$

Pada node akar, atribut yang memiliki nilai *Gain* tertinggi adalah Tahun Lulus (X_6), yaitu 0,1616, sehingga untuk memperoleh *Gain Ratio* dari atribut Jenis Kelamin (X_1) adalah sebagai berikut:

$$GR(X_1) = \frac{0,1616}{0,9685} = 0,1668$$

Secara lengkap, seluruh proses pada node akar ditunjukkan pada tabel berikut ini.

Atribut	Label Atribut	Total	(≤ 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
		96	60	36	0,9544				
JENIS KELAMIN	Laki-laki	58	39	19	0,9124	0,9685	0,9439	0,0105	0,1668
	Perempuan	38	21	17	0,9920				
UMUR	(≤25)	58	35	23	0,9689	0,9685	0,9522	0,0022	0,1668
	(>25)	38	25	13	0,9268				
STATUS MENIKAH	Menikah	40	27	13	0,9097	0,9799	0,9489	0,0055	0,1649
	Belum Menikah	56	33	23	0,9769				
KELUARGA	(≤5)	64	41	23	0,9422	0,9183	0,9529	0,0015	0,1759
	(>5)	32	19	13	0,9745				
KONSENTRASI	Pemrograman	50	35	15	0,8813	0,9987	0,9356	0,0189	0,1618
	Jaringan dan Lainnya	46	25	21	0,9945				
TAHUN LULUS	(≤2012)	26	25	1	0,2352	0,8427	0,7929	0,1616	0,1917
	(>2012)	70	35	35	1,0000				
IPK	(≤3,5)	40	30	10	0,8113	0,9799	0,9192	0,0352	0,1649
	(>3,5)	56	30	26	0,9963				

Atribut yang memiliki nilai *Gain* tertinggi adalah Tahun Lulus (X_6), sehingga menjadi node akar [1], maka pohon keputusan yang dibentuk dari node akar adalah:

IF [1] Tahun Lulus = ≤ 2012 AND [2]

IF [1] Tahun Lulus = > 2012 AND [3]

Node [1] → [2]:

Data yang digunakan pada node ini adalah Tahun Lulus = ≤ 2012.

Entropy class:

$$\text{Jumlah data (S)} = 26$$

$$\text{Jumlah data label class } \leq 6 \text{ Bulan (S}_{\leq 6 \text{Bulan}}) = 25$$

$$\text{Jumlah data label class } > 6 \text{ Bulan (S}_{> 6 \text{Bulan}}) = 1$$

$$E(S) = - \left(\left(\left(\frac{25}{26} \log_2 \right) \frac{25}{26} \right) + \left(\left(\frac{1}{26} \log_2 \right) \frac{1}{26} \right) \right) = 0,2352$$

Entropy, Information Gain, Gain, dan Gain Ratio pada atribut Jenis Kelamin (X₁):

$$\text{Jumlah data X}_{1,\text{Laki-laki}} = 16$$

$$\text{Jumlah data X}_{1,\text{Perempuan}} = 10$$

$$\text{Jumlah data X}_{1,\text{Laki-laki}} \text{ untuk S}_{\leq 6 \text{Bulan}} = 16$$

$$\text{Jumlah data X}_{1,\text{Laki-laki}} \text{ untuk S}_{> 6 \text{Bulan}} = 0$$

$$\text{Jumlah data X}_{1,\text{Perempuan}} \text{ untuk S}_{\leq 6 \text{Bulan}} = 9$$

$$\text{Jumlah data X}_{1,\text{Perempuan}} \text{ untuk S}_{> 6 \text{Bulan}} = 1$$

$$E(X_{1,\text{Laki-laki}}) = - \left(\left(\left(\frac{16}{16} \log_2 \right) \frac{16}{16} \right) + \left(\left(\frac{0}{16} \log_2 \right) \frac{0}{16} \right) \right) = 0$$

$$E(X_{1,\text{Perempuan}}) = - \left(\left(\left(\frac{9}{10} \log_2 \right) \frac{9}{10} \right) + \left(\left(\frac{1}{10} \log_2 \right) \frac{1}{10} \right) \right) = 0,4690$$

$$E(X_1) = - \left(\left(\left(\frac{16}{26} \log_2 \right) \frac{16}{26} \right) + \left(\left(\frac{10}{26} \log_2 \right) \frac{10}{26} \right) \right) = 0,9612$$

$$GI(X_1) = \left(\frac{16}{26} 0 \right) + \left(\frac{10}{26} 0,4690 \right) = 0,1804$$

$$G(X_1) = 0,2352 - 0,1804 = 0,0548$$

Pada node ini, atribut yang memiliki nilai *Gain* tertinggi adalah Jenis Kelamin (X₁), yaitu 0,1804, sehingga untuk memperoleh *Gain Ratio* dari atribut Jenis Kelamin (X₁) adalah sebagai berikut:

$$GR(X_1) = \frac{0,1804}{0,9612} = 0,0570$$

Secara lengkap, seluruh proses pada node ini ditunjukkan pada tabel berikut ini.

Node 2		Total	(≤ 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	26	25	1	0,2352				
JENIS KELAMIN	Laki-laki	16	16	0	0,0000	0,9612	0,1804	0,0548	0,0570
	Perempuan	10	9	1	0,4690				
UMUR	(≤ 25)	16	15	1	0,3373	0,9612	0,2076	0,0276	0,0570

	(>25)	10	10	0	0,0000				
STATUS MENIKAH	Menikah	11	10	1	0,4395	0,9829	0,1859	0,0493	0,0558
	Belum Menikah	15	15	0	0,0000				
KELUARGA	(<=5)	21	20	1	0,2762	0,7063	0,2231	0,0121	0,0776
	(>5)	5	5	0	0,0000				
KONSENTRASI	Pemrograman	10	10	0	0,0000	0,9612	0,2076	0,0276	0,0570
	Jaringan dan Lainnya	16	15	1	0,3373				
TAHUN LULUS	(<=2012)	26	25	1	0,2352	0,0000	0,2352	0,0000	0,0000
	(>2012)	0	0	0	0,0000				
IPK	(<=3,5)	15	15	0	0,0000	0,9829	0,1859	0,0493	0,0558
	(>3,5)	11	10	1	0,4395				

Atiribut yang memiliki nilai *Gain* tertinggi adalah Jenis Kelamin (X_1), sehingga menjadi node [2], yang mana label Laki-Laki variabel Jenis Kelamin berada semua pada label *class* ≤ 6 Bulan, maka pohon keputusan yang dibentuk dari node 2 adalah:

IF [1] Tahun Lulus = ≤ 2012 AND [2] Jenis Kelamin = Laki-Laki THEN ≤ 6 Bulan

IF [1] Tahun Lulus = ≤ 2012 AND [2] Jenis Kelamin = Perempuan AND [5]

Secara lengkap, setiap proses dari node-node ditunjukkan pada tabel-tabel berikut.

Node [1] \rightarrow [3]:

Node 3		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	70	35	35	1,0000				
JENIS KELAMIN	Laki-laki	42	23	19	0,9934	0,9710	0,9902	0,0098	0,0629
	Perempuan	28	12	16	0,9852				
UMUR	(<=25)	42	20	22	0,9984	0,9710	0,9975	0,0025	0,0629
	(>25)	28	15	13	0,9963				
STATUS MENIKAH	Menikah	29	17	12	0,9784	0,9787	0,9848	0,0152	0,0624
	Belum Menikah	41	18	23	0,9892				
KELUARGA	(<=5)	43	21	22	0,9996	0,9620	0,9994	0,0006	0,0635
	(>5)	27	14	13	0,9990				
KONSENTRASI	Pemrograman	40	25	15	0,9544	0,9852	0,9389	0,0611	0,0620
	Jaringan dan Lainnya	30	10	20	0,9183				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	1,0000	0,0000	0,0000
	(>2012)	70	35	35	1,0000				
IPK	(<=3,5)	25	15	10	0,9710	0,9403	0,9839	0,0161	0,0649
	(>3,5)	45	20	25	0,9911				

Atiribut yang memiliki nilai *Gain* tertinggi adalah Konsentrasi (X_5), sehingga menjadi node [3].

Node [1] \rightarrow [2] \rightarrow [5]:

Node 5		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	10	9	1	0,4690				
JENIS KELAMIN	Laki-laki	0	0	0	0,0000	0,0000	0,4690	0,0000	0,0000
	Perempuan	10	9	1	0,4690				
UMUR	(<=25)	7	6	1	0,5917	0,8813	0,4142	0,0548	0,3052
	(>25)	3	3	0	0,0000				
STATUS MENIKAH	Menikah	5	4	1	0,7219	1,0000	0,3610	0,1080	0,2690
	Belum Menikah	5	5	0	0,0000				

KELUARGA	(<=5)	8	7	1	0,5436	0,7219	0,4349	0,0341	0,3726
	(>5)	2	2	0	0,0000				
KONSENTRASI	Pemrograman	3	3	0	0,0000	0,8813	0,4142	0,0548	0,3052
	Jaringan dan Lainnya	7	6	1	0,5917				
TAHUN LULUS	(<=2012)	10	9	1	0,4690	0,0000	0,4690	0,0000	0,0000
	(>2012)	0	0	0	0,0000				
IPK	(<=3,5)	8	8	0	0,0000	0,7219	0,2000	0,2690	0,3726
	(>3,5)	2	1	1	1,0000				

Atiribut yang memiliki nilai *Gain* tertinggi adalah IPK (X_7), sehingga menjadi node [5].

Node [1] → [3] → [6]:

Node 6		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	40	25	15	0,9544				
JENIS KELAMIN	Laki-laki	26	17	9	0,9306	0,9341	0,9497	0,0047	0,0186
	Perempuan	14	8	6	0,9852				
UMUR	(<=25)	22	14	8	0,9457	0,9928	0,9539	0,0005	0,0175
	(>25)	18	11	7	0,9641				
STATUS MENIKAH	Menikah	20	14	6	0,8813	1,0000	0,9370	0,0174	0,0174
	Belum Menikah	20	11	9	0,9928				
KELUARGA	(<=5)	26	16	10	0,9612	0,9341	0,9539	0,0005	0,0186
	(>5)	14	9	5	0,9403				
KONSENTRASI	Pemrograman	40	25	15	0,9544	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	0	0	0	0,0000				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	40	25	15	0,9544				
IPK	(<=3,5)	10	10	0	0,0000	0,8113	0,0000	0,0000	0,0214
	(>3,5)	30	15	15	1,0000				

Atiribut yang memiliki nilai *Gain* tertinggi adalah Status Menikah (X_3), sehingga menjadi node [6].

Node [1] → [3] → [7]:

Node 7		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	30	10	20	0,9183				
JENIS KELAMIN	Laki-laki	16	6	10	0,9544	0,9968	0,9118	0,0065	0,0071
	Perempuan	14	4	10	0,8631				
UMUR	(<=25)	20	6	14	0,8813	0,9183	0,9112	0,0071	0,0078
	(>25)	10	4	6	0,9710				
STATUS MENIKAH	Menikah	9	3	6	0,9183	0,8813	0,9183	0,0000	0,0081
	Belum Menikah	21	7	14	0,9183				
KELUARGA	(<=5)	17	5	12	0,8740	0,9871	0,9118	0,0065	0,0072
	(>5)	13	5	8	0,9612				
KONSENTRASI	Pemrograman	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	30	10	20	0,9183				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	30	10	20	0,9183				
IPK	(<=3,5)	15	5	10	0,9183	1,0000	0,9183	0,0000	0,0071
	(>3,5)	15	5	10	0,9183				

Atiribut yang memiliki nilai *Gain* tertinggi adalah Umur (X_2), sehingga menjadi node [7].

Node [1] → [3] → [6] → [10]:

Node 10		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	20	14	6	0,8813				
JENIS KELAMIN	Laki-laki	10	8	2	0,7219	1,0000	0,8464	0,0349	0,0349

	Perempuan	10	6	4	0,9710				
UMUR	(<=25)	5	4	1	0,7219	0,8113	0,8692	0,0121	0,0430
	(>25)	15	10	5	0,9183				
STATUS MENIKAH	Menikah	20	14	6	0,8813	0,0000	0,0000	0,0000	0,0000
	Belum Menikah	0	0	0	0,0000				
KELUARGA	(<=5)	15	10	5	0,9183	0,8113	0,8692	0,0121	0,0430
	(>5)	5	4	1	0,7219				
KONSENTRASI	Pemrograman	20	14	6	0,8813	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	0	0	0	0,0000				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	20	14	6	0,8813				
IPK	(<=3,5)	5	5	0	0,0000	0,8113	0,0000	0,0000	0,0430
	(>3,5)	15	9	6	0,9710				

Atiribut yang memiliki nilai *Gain* tertinggi adalah Jenis Kelamin (X_1), sehingga menjadi node [10].

Node [1] → [3] → [6] → [11]:

Node 11		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	20	11	9	0,9928				
JENIS KELAMIN	Laki-laki	16	9	7	0,9887	0,7219	0,9910	0,0018	0,0336
	Perempuan	4	2	2	1,0000				
UMUR	(<=25)	17	10	7	0,9774	0,6098	0,9685	0,0242	0,0397
	(>25)	3	1	2	0,9183				
STATUS MENIKAH	Menikah	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Belum Menikah	20	11	9	0,9928				
KELUARGA	(<=5)	11	6	5	0,9940	0,9928	0,9927	0,0001	0,0244
	(>5)	9	5	4	0,9911				
KONSENTRASI	Pemrograman	20	11	9	0,9928	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	0	0	0	0,0000				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	20	11	9	0,9928				
IPK	(<=3,5)	5	5	0	0,0000	0,8113	0,0000	0,0000	0,0299
	(>3,5)	15	6	9	0,9710				

Atiribut yang memiliki nilai *Gain* tertinggi adalah Umur (X_2), sehingga menjadi node [11].

Node [1] → [3] → [7] → [12]:

Node 12		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	20	6	14	0,8813				
JENIS KELAMIN	Laki-laki	12	4	8	0,9183	0,9710	0,8755	0,0058	0,0060
	Perempuan	8	2	6	0,8113				
UMUR	(<=25)	20	6	14	0,8813	0,0000	0,0000	0,0000	0,0000
	(>25)	0	0	0	0,0000				
STATUS MENIKAH	Menikah	1	0	1	0,0000	0,2864	0,0000	0,0000	0,0203
	Belum Menikah	19	6	13	0,8997				

KELUARGA	(<=5)	11	3	8	0,8454	0,9928	0,8782	0,0031	0,0058
	(>5)	9	3	6	0,9183				
KONSENTRASI	Pemrograman	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	20	6	14	0,8813				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	20	6	14	0,8813				
IPK	(<=3,5)	10	3	7	0,8813	1,0000	0,8813	0,0000	0,0058
	(>3,5)	10	3	7	0,8813				

Atiribut yang memiliki nilai *Gain* tertinggi adalah Jenis Kelamin (X_1), sehingga menjadi node [12].

Node [1] → [3] → [7] → [13]:

Node 13		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	10	4	6	0,9710				
JENIS KELAMIN	Laki-laki	4	2	2	1,0000	0,9710	0,9510	0,0200	0,0206
	Perempuan	6	2	4	0,9183				
UMUR	(<=25)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>25)	10	4	6	0,9710				
STATUS MENIKAH	Menikah	8	3	5	0,9544	0,7219	0,9635	0,0074	0,0277
	Belum Menikah	2	1	1	1,0000				
KELUARGA	(<=5)	6	2	4	0,9183	0,9710	0,9510	0,0200	0,0206
	(>5)	4	2	2	1,0000				
KONSENTRASI	Pemrograman	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	10	4	6	0,9710				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	10	4	6	0,9710				
IPK	(<=3,5)	5	2	3	0,9710	1,0000	0,9710	0,0000	0,0200
	(>3,5)	5	2	3	0,9710				

Atiribut yang memiliki nilai *Gain* tertinggi adalah Keluarga (X_4), sehingga menjadi node [13].

Node [1] → [3] → [6] → [10] → [14]:

Node 14		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	10	8	2	0,7219				
JENIS KELAMIN	Laki-laki	10	8	2	0,7219	0,0000	0,0000	0,0000	0,0000
	Perempuan	0	0	0	0,0000				
UMUR	(<=25)	4	3	1	0,8113	0,9710	0,7145	0,0074	0,0076
	(>25)	6	5	1	0,6500				
STATUS MENIKAH	Menikah	10	8	2	0,7219	0,0000	0,0000	0,0000	0,0000
	Belum Menikah	0	0	0	0,0000				
KELUARGA	(<=5)	8	6	2	0,8113	0,7219	0,0000	0,0000	0,0103
	(>5)	2	2	0	0,0000				
KONSENTRASI	Pemrograman	10	8	2	0,7219	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	0	0	0	0,0000				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	10	8	2	0,7219				
IPK	(<=3,5)	2	2	0	0,0000	0,7219	0,0000	0,0000	0,0103
	(>3,5)	8	6	2	0,8113				

Atiribut yang memiliki nilai *Gain* tertinggi adalah Umur (X_2), sehingga menjadi node [14].

Node [1] → [3] → [6] → [10] → [15]:

Node 15		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	10	6	4	0,9710				
JENIS KELAMIN	Laki-laki	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000

	Perempuan	10	6	4	0,9710				
UMUR	(<=25)	1	1	0	0,0000	0,4690	0,0000	0,0000	0,0124
	(>25)	9	5	4	0,9911				
STATUS MENIKAH	Menikah	10	6	4	0,9710	0,0000	0,0000	0,0000	0,0000
	Belum Menikah	0	0	0	0,0000				
KELUARGA	(<=5)	7	4	3	0,9852	0,8813	0,9651	0,0058	0,0066
	(>5)	3	2	1	0,9183				
KONSENTRASI	Pemrograman	10	6	4	0,9710	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	0	0	0	0,0000				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	10	6	4	0,9710				
IPK	(<=3,5)	3	3	0	0,0000	0,8813	0,0000	0,0000	0,0066
	(>3,5)	7	3	4	0,9852				

11

Atiribut yang memiliki nilai *Gain* tertinggi adalah Keluarga (X_4), sehingga menjadi node [15].

Node [1] → [3] → [6] → [11] → [16]:

Node 16		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Variabel	Label Variabel	17	10	7	0,9774				
JENIS KELAMIN	Laki-laki	13	8	5	0,9612	0,7871	0,9704	0,0071	0,0090
	Perempuan	4	2	2	1,0000				
UMUR	(<=25)	17	10	7	0,9774	0,0000	0,0000	0,0000	0,0000
	(>25)	0	0	0	0,0000				
STATUS MENIKAH	Menikah	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Belum Menikah	17	10	7	0,9774				
KELUARGA	(<=5)	8	5	3	0,9544	0,9975	0,9738	0,0036	0,0071
	(>5)	9	5	4	0,9911				
KONSENTRASI	Pemrograman	17	10	7	0,9774	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	0	0	0	0,0000				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	17	10	7	0,9774				
IPK	(<=3,5)	4	4	0	0,0000	0,7871	0,0000	0,0000	0,0090
	(>3,5)	13	6	7	0,9957				

21

Atiribut yang memiliki nilai *Gain* tertinggi adalah Jenis Kelamin (X_1), sehingga menjadi node [16].

Node [1] → [3] → [7] → [12] → [18]:

Node 18		Total	(<= 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Label	Label Variabel	12	4	8	0,9183				
JENIS KELAMIN	Laki-laki	12	4	8	0,9183	0,0000	0,0000	0,0000	0,0000
	Perempuan	0	0	0	0,0000				
UMUR	(<=25)	12	4	8	0,9183	0,0000	0,0000	0,0000	0,0000
	(>25)	0	0	0	0,0000				
STATUS MENIKAH	Menikah	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Belum Menikah	12	4	8	0,9183				
KELUARGA	(<=5)	7	2	5	0,8631	0,9799	0,9080	0,0102	0,0105
	(>5)	5	2	3	0,9710				
KONSENTRASI	Pemrograman	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	12	4	8	0,9183				
TAHUN LULUS	(<=2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	12	4	8	0,9183				
IPK	(<=3,5)	5	2	3	0,9710	0,9799	0,9080	0,0102	0,0105
	(>3,5)	7	2	5	0,8631				

11

Atiribut yang memiliki nilai *Gain* tertinggi adalah IPK (X_7), sehingga menjadi node [18].

Node [1] → [3] → [7] → [12] → [19]:

Node 19		Total	(≤ 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Label	Label Variabel	8	2	6	0,8113				
JENIS KELAMIN	Laki-laki	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Perempuan	8	2	6	0,8113				
UMUR	(≤25)	8	2	6	0,8113	0,0000	0,0000	0,0000	0,0000
	(>25)	0	0	0	0,0000				
STATUS MENIKAH	Menikah	1	0	1	0,0000	0,5436	0,0000	0,0000	0,0289
	Belum Menikah	7	2	5	0,8631				
KELUARGA	(≤5)	4	1	3	0,8113	1,0000	0,8113	0,0000	0,0157
	(>5)	4	1	3	0,8113				
KONSENTRASI	Pemrograman	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	8	2	6	0,8113				
TAHUN LULUS	(≤2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	8	2	6	0,8113				
IPK	(≤3,5)	5	1	4	0,7219	0,9544	0,7956	0,0157	0,0165
	(>3,5)	3	1	2	0,9183				

Atribut yang memiliki nilai *Gain* tertinggi adalah IPK (X_7), sehingga menjadi node [19].

Node [1] → [3] → [6] → [11] → [16] → [26]:

Node 26		Total	(≤ 6 Bulan)	(> 6 Bulan)	Entropy	Entropy Variabel	Information Gain	Gain	Gain Ratio
Label	Label Variabel	13	8	5	0,9612				
JENIS KELAMIN	Laki-laki	13	8	5	0,9612	0,0000	0,0000	0,0000	0,0000
	Perempuan	0	0	0	0,0000				
UMUR	(≤25)	13	8	5	0,9612	0,0000	0,0000	0,0000	0,0000
	(>25)	0	0	0	0,0000				
STATUS MENIKAH	Menikah	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	Belum Menikah	13	8	5	0,9612				
KELUARGA	(≤5)	6	4	2	0,9183	0,9957	0,9543	0,0069	0,0069
	(>5)	7	4	3	0,9852				
KONSENTRASI	Pemrograman	13	8	5	0,9612	0,0000	0,0000	0,0000	0,0000
	Jaringan dan Lainnya	0	0	0	0,0000				
TAHUN LULUS	(≤2012)	0	0	0	0,0000	0,0000	0,0000	0,0000	0,0000
	(>2012)	13	8	5	0,9612				
IPK	(≤3,5)	3	3	0	0,0000	0,7793	0,0000	0,0000	0,0089
	(>3,5)	10	5	5	1,0000				

Atribut yang memiliki nilai *Gain* tertinggi adalah Keluarga (X_4), sehingga menjadi node [26].

Dengan demikian, pohon yang dibentuk atau *rule* dari model C4.5 untuk prediksi waktu tunggu kerja alumni teknik informatika adalah sebagai berikut.

1. IF [1] Tahun Lulus IS ≤ 2012 AND [2] Jenis Kelamin IS Laki-Laki THEN Waktu Tunggu IS ≤ 6 Bulan
2. IF [1] Tahun Lulus IS ≤ 2012 AND [2] Jenis Kelamin IS Perempuan AND [5] IPK IS $\leq 3,5$ THEN Waktu Tunggu IS ≤ 6 Bulan
3. IF [1] Tahun Lulus IS ≤ 2012 AND [2] Jenis Kelamin IS Perempuan AND [5] IPK IS $> 3,5$ THEN Waktu Tunggu IS ≤ 6 Bulan (50%) OR > 6 Bulan (50%)

4. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Pemrograman AND [6] Status Nikah IS Belum Menikah AND [11] Umur IS > 25 THEN Waktu Tunggu IS \leq 6 Bulan (33%) OR > 6 Bulan (67%)
5. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Jaringan dan Lainnya AND [7] Umur IS > 25 AND [13] Keluarga IS \leq 5 THEN Waktu Tunggu \leq 6 Bulan (33%) OR > 6 Bulan (67%)
6. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Jaringan dan Lainnya AND [7] Umur IS > 25 AND [13] Keluarga IS > 5 THEN Waktu Tunggu \leq 6 Bulan (50%) OR > 6 Bulan (50%)
7. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Pemrograman AND [6] Status Nikah IS Menikah AND [10] Jenis Kelamin IS Laki-laki AND [14] Umur IS \leq 25 THEN Waktu Tunggu IS \leq 6 Bulan (75%) OR > 6 Bulan (25%)
8. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Pemrograman AND [6] Status Nikah = Menikah AND [10] Jenis Kelamin IS Laki-Laki AND [14] Umur IS > 25 THEN Waktu Tunggu IS \leq 6 Bulan (83%) OR > 6 Bulan (17%)
9. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Pemrograman AND [6] Status Nikah IS Menikah AND [10] Jenis Kelamin IS Perempuan AND [15] Keluarga IS \leq 5 THEN Waktu Tunggu IS \leq 6 Bulan (57%) OR > 6 Bulan (43%)
10. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Pemrograman AND [6] Status Nikah IS Menikah AND [10] Jenis Kelamin IS Perempuan AND [15] Keluarga IS > 5 THEN Waktu Tunggu \leq 6 Bulan (67%) OR > 6 Bulan (33%)
11. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Pemrograman AND [6] Status Nikah IS Belum Menikah AND [11] Umur IS \leq 25 AND [16] Jenis Kelamin IS Perempuan THEN Waktu Tunggu \leq 6 Bulan (50%) OR > 6 Bulan (50%)
12. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Jaringan dan Lainnya AND [7] Umur IS \leq 25 AND [12] Jenis Kelamin IS Laki-Laki AND [18] IPK IS \leq 3,5 THEN Waktu Tunggu \leq 6 Bulan (40%) OR > 6 Bulan (60%)
13. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Jaringan dan Lainnya AND [7] Umur IS \leq 25 AND [12] Jenis Kelamin IS Laki-laki AND [18] IPK IS > 3,5 THEN Waktu Tunggu \leq 6 Bulan (29%) OR > 6 Bulan (71%)

14. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Jaringan dan Lainnya AND [7] Umur IS ≤ 25 AND [12] Jenis Kelamin IS Perempuan AND [19] IPK IS $\leq 3,5$ THEN Waktu Tunggu ≤ 6 Bulan (20%) OR > 6 Bulan (80%)
15. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Jaringan dan Lainnya AND [7] Umur IS ≤ 25 AND [12] Jenis Kelamin IS Perempuan AND [19] IPK IS > 3,5 THEN Waktu Tunggu IS ≤ 6 Bulan (33%) OR > 6 Bulan (67%)
16. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Pemrograman AND [6] Status Nikah IS Belum Menikah AND [11] Umur IS ≤ 25 AND [16] Jenis Kelamin IS Laki-Laki AND [26] Keluarga IS ≤ 5 THEN Waktu Tunggu IS ≤ 6 Bulan (67%) OR > 6 Bulan (33%)
17. IF [1] Tahun Lulus IS > 2012 AND [3] Konsentrasi IS Pemrograman AND [6] Status Nikah IS Belum Menikah AND [11] Umur IS ≤ 25 AND [16] Jenis Kelamin IS Laki-Laki AND [26] Keluarga IS > 5 THEN Waktu Tunggu IS ≤ 6 Bulan (57%) OR > 6 Bulan (43%)

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, maka dapat disimpulkan bahwa diperoleh Sistem Prediksi Waktu Tunggu Kerja Alumni Teknik Informatika Berbasis Algoritma C4.5 yang tidak kompleks dengan alur logika program yang sudah benar berdasarkan nilai $V(G) = 3$ dari pengujian *White Box* dan sudah bebas dari kesalahan komponen sistem berdasarkan pengujian *Black Box*, sehingga sistem ini dapat diimplementasikan untuk keperluan *tracer study* pada suatu jurusan teknik informatika di perguruan tinggi. Sementara itu, model C4.5 dari prediksi waktu tunggu kerja alumni teknik informatika menghasilkan 17 pohon keputusan (*rule*).

6.2 Saran

Setelah melakukan penelitian ini, maka ada adapun saran yang dapat penulis sampaikan adalah:

1. Penerapan algoritma lainnya (berbeda).
2. Pengumpulan data yang lebih banyak dengan varian-varian data yang berbeda.
3. Melakukan eksperimen terhadap beberapa metode *machine learning* untuk memperoleh model *machine learning* yang terbaik untuk prediksi waktu tunggu kerja alumni teknik informatika.

DAFTAR PUSTAKA

- Arfianto, S.2014. Sistem informasi Tracer Study Alumni UAD Studi Kasus Pada Lembaga dan Penelitian UAD. Yogyakarta: UAD
- APTIKOM. (2015). Kerangka Kualifikasi Nasional Indonesia (KKNI) Rumpun Ilmu Informatika Dan Komputer. APTIKOM.
- ²³ Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting students drop out : a case study. Proceedings of the 2nd International Conference on Educational Data Mining (pp. 41-50). EDM: Cordoba.
- Firmansyah, Penerapan Algoritma Klasifikasi C4.5 untuk Penentuan kelayakan Pemberian Kredit Koperasi. Nusa Mandiri, 2015
- Hastuti, K. (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif. Seminar Nasional Teknologi Informasi & Komunikasi Terapan (SEMANTIK) (pp. 241-249). Semarang: INFRM.
- Jantawan, B., & Tsai, C.-F. (2013). The Application of Data Mining to Build Classification Model for Predicting Graduate Employment. International Journal of Computer Science and Information Security (IJCSIS), Vol. 11, No. 10.
- ¹⁵ Jogiyanto, HM., 2013, Analisis dan Desain Sistem Informasi Pendekatan Terstruktur Edisi IV. Yogyakarta, Andi Offset.
- Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. ULTIMATICS, Vol. 6, No. 1, 15-20.
- Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2004). Preventing Student Dropout in Distance Learning Using Machine Learning Techniques. KES 2003 (pp. 267-274). Berlin: Springer.
- ²⁸ Mishra, T., Kumar, D., & Gupta, S. (2016). Students Employability Prediction Model through Data Mining. International Journal of Applied Engineering Research, Vol. 11, No. 4, 2275-2282.
- Muis Nanja dan Purwanto, (2015). "Metode K- Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Harga Komoditi Lada". Jurnal Pseudocode, Volume 2 Nomor 1, Februari 2015, ISSN 2355 – 5920.
- Nuraeni, Y. (2009). Penerapan Jaringan Syaraf Tiruan untuk Mengukur Tingkat Korelasi antara NEM dengan IPK Kelulusan Mahasiswa. TELKOMNIKA, Vol. 7, No. 3, 195-200.
- ⁵ Nurma Jayanti, Sulisty Puspitodjati, Tety Elida (2015). Teknik Klasifikasi Pohon Keputusan untuk Memprediksi Kebangkrutan Bank Berdasarkan Rasio Keuangan Bank. Depok : KOMMIT

- Prasetyo, E. (2014). Data Mining : Mengolah Data Menjadi Informasi Menggunakan Matlab. Gresik: Andi Offset.
- Pressman, R. S. (2012). Software Engineering: A Practitioners Approach. Mc Graw-Hill.
- Rasiman, Cahyono, A. N., Sulianto, J., & Nurhadi. (2008). Penelusuran Lulusan Program Studi Pendidikan Matematika IKIP PGRI Semarang Melalui Studi Pelacakan (Tracer Study) Sebagai Umpan Balik Penyempurnaan Kurikulum Tahun 2008. Jurnal Media Penelitian Pendidikan, Vol. 2, No. 2, 162-173.
- RISTEKDIKTI. (2016). Panduan Hibah Tracer Study. RISTEKDIKTI.
- Septiana, L. (2013). Penerapan Neural Network Berbasis Particle Swarm Optimization untuk Seleksi Atribut Penentuan Mahasiswa Drop Out. Pilar Nusa Mandiri, Vol. 9, No.2 , 104-112.
- Sunjana, Klasifikasi Nasabah Sebuah Asuransi Menggunakan Algoritma C4.5. Yogyakarta : 2014.
- Widodo, P. P., Handayanto, R. T., & Herlawati. (2013). Penerapan Data Mining Dengan Matlab. Bandung: Rekayasa Sains.

LAMPIRAN

PENERAPAN ALGORITMA C4.5 PADA PREDIKSI WAKTU TUNGGU KERJA ALUMNI TEKNIK INFORMATIKA

ORIGINALITY REPORT

25%

SIMILARITY INDEX

23%

INTERNET SOURCES

5%

PUBLICATIONS

20%

STUDENT PAPERS

PRIMARY SOURCES

1

e-journal.uajy.ac.id

Internet Source

4%

2

www.scribd.com

Internet Source

4%

3

Submitted to LL Dikti IX Turnitin Consortium

Student Paper

3%

4

Submitted to Universitas Diponegoro

Student Paper

2%

5

docplayer.info

Internet Source

2%

6

Submitted to Universitas Muria Kudus

Student Paper

2%

7

es.scribd.com

Internet Source

1%

8

jurnal.untan.ac.id

Internet Source

1%

9

Submitted to Universitas Putera Batam

10

Submitted to Politeknik Negeri Bandung

Student Paper

1 %

11

core.ac.uk

Internet Source

1 %

12

Submitted to Universitas Krisnadwipayana -
Faculty of Administration

Student Paper

<1 %

13

eprints.radenfatah.ac.id

Internet Source

<1 %

14

www.wayanfm.lecture.ub.ac.id

Internet Source

<1 %

15

nonosun.staf.upi.edu

Internet Source

<1 %

16

Submitted to Sriwijaya University

Student Paper

<1 %

17

id.123dok.com

Internet Source

<1 %

18

edrianhadinata.wordpress.com

Internet Source

<1 %

19

id.scribd.com

Internet Source

<1 %

20	Internet Source	<1 %
21	de.slideshare.net Internet Source	<1 %
22	repository.widyatama.ac.id Internet Source	<1 %
23	link.springer.com Internet Source	<1 %
24	riyanardian.blogspot.com Internet Source	<1 %
25	lib.dr.iastate.edu Internet Source	<1 %
26	eprints.umm.ac.id Internet Source	<1 %
27	Submitted to Universitas Dian Nuswantoro Student Paper	<1 %
28	ephjournal.com Internet Source	<1 %

Exclude quotes On

Exclude matches

< 25 words

Exclude bibliography On