



TUGAS AKHIR -SM141501

**PREDIKSI POLA KECELAKAAN KERJA PADA
PERUSAHAAN NON EKSTRAKTIF MENGGUNAKAN
ALGORITMA DECISION TREE: C4.5 DAN C5.0**

**YOLANDA RIZKITA PUTRI
NRP 1212 100 040**

**Dosen Pembimbing
Dr. Imam Mukhlash, S.Si, MT
Drs. Nurul Hidayat, M. Kom**

**JURUSAN MATEMATIKA
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2016**

“Halaman ini sengaja dikosongkan”



FINAL PROJECT -SM141501

***PREDICTION PATTERNS OF OCCUPATIONAL ACCIDENT
IN NON EXTRACTIVE COMPANY USING DECISION TREE
ALGORITHMS: C4.5 AND C 5.0***

YOLANDA RIZKITA PUTRI
NRP 1212 100 040

Supervisors

Dr. Imam Mukhlash, S.Si, MT
Drs. Nurul Hidayat, M. Kom

DEPARTMENT OF MATHEMATICS
Faculty of Mathematics and Natural Science
Sepuluh Nopember Institute of Technology
Surabaya 2016

“Halaman ini sengaja dikosongkan”

LEMBAR PENGESAHAN

**PREDIKSI POLA KECELAKAAN KERJA PADA
PERUSAHAAN NON EKSTRAKTIF MENGGUNAKAN
ALGORITMA DECISION TREE: C4.5 DAN C5.0**

***PREDICTION PATTERNS OF OCCUPATIONAL
ACCIDENT IN NON EXTRACTIVE COMPANY USING
DECISION TREE ALGORITHM: C4.5 AND C5.0***

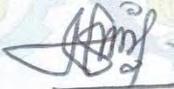
TUGAS AKHIR

**Diajukan Untuk Memenuhi Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Sains
Pada Bidang Studi Ilmu Komputer
Program Studi S-1 Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember Surabaya**

Oleh :
YOLANDA RIZKITA PUTRI
NRP. 1212 100 040

Menyetujui,

Dosen Pembimbing II,



Dts. Nurul Hidayat, M. Kom
NIP. 19630404 198903 1 002

Dosen Pembimbing I,



Dr. Imam Mukhlash, S.Si, MT
NIP. 19700831 199403 1 003

Mengetahui,

**Ketua Jurusan Matematika
FMIPA ITS**



Dr. Imam Mukhlash, S.Si, MT
NIP. 19700831 199403 1 003
Surabaya, 25 Juli 2016

PREDIKSI POLA KECELAKAAN KERJA PADA PERUSAHAAN NON EKSTRAKTIF MENGUNAKAN ALGORITMA DECISION TREE: C4.5 DAN C5.0

Nama : Yolanda Rizkita Putri
NRP : 1212 100 040
Jurusan : Matematika
Dosen Pembimbing : Dr. Imam Mukhlash, S.Si, MT
Drs. Nurul Hidayat, M. Kom

Abstrak

Kecelakaan kerja pada perusahaan sangat mungkin terjadi. Kemungkinan tersebut dapat muncul dari berbagai aspek, mulai dari aspek karyawan atau sistem perusahaan yang diterapkan. Salah satu tindakan preventif yang dapat dilakukan adalah memprediksi pola kecelakaan dari kecelakaan kerja yang terjadi. *Data mining* merupakan metode yang dapat memprediksi pola data menggunakan data histori. Salah satu metode yang dapat dipilih adalah klasifikasi. Klasifikasi bekerja untuk membentuk pola dari keterhubungan atribut yang ada pada data. Dalam Tugas Akhir ini, metode klasifikasi yang digunakan adalah *decision tree*. Pada *decision tree*, atribut pada data dapat membentuk suatu *rules* yang menggambarkan kemungkinan terjadinya kejadian. Saat ini telah berkembang algoritma *decision tree* yang dapat digunakan untuk memprediksi pola kecelakaan kerja, yaitu C4.5 dan C5.0. Hasil Algoritma C4.5 dan C5.0 berupa *rules* akan digunakan untuk proses klasifikasi. Selanjutnya proses testing digunakan untuk mengetahui performansi dari kedua algoritma. Performansi tersebut diukur dengan dua parameter yaitu akurasi dan waktu komputasi. Hasil pengujian menunjukkan akurasi Algoritma C4.5 lebih baik daripada Algoritma C5.0. Sedangkan waktu komputasi kedua algoritma hampir sama. Jadi dapat disimpulkan bahwa Algoritma C4.5 lebih baik dari Algoritma C5.0.

Kata Kunci —Kecelakaan Kerja, *Data Mining*, Klasifikasi, C4.5, C5.0.

“Halaman ini sengaja dikosongkan”

PREDICTION PATTERNS OF OCCUPATIONAL ACCIDENT IN NON EXTRACTIVE COMPANY USING DECISION TREE ALGORITHM: C4.5 AND C5.0

Name : Yolanda Rizkita Putri
NRP : 1212 100 040
Department : Matematika
Supervisors : Dr. Imam Mukhlash, S.Si, MT
Drs. Nurul Hidayat, M. Kom

Abstract

Occupational accidents at the company is very likely to occur. The possibility can arise from various aspects, ranging from aspects of the employee or the company's system. One of the preventive ways that can be done is by predicting the pattern of accident. Data mining is a method that can predict patterns of data using historical data. One method that can be selected is classification. Classification is working to establish a pattern of attribute connectivity of the data. In this Final Project, Classification method used are Decision Tree Algorithm. In that algorithm, these attributes can form a rules that describe the possibility of occurrence. Today there is a decision tree algorithm that can be used to predict the pattern of occupational accidents, namely C4.5 and C5.0. The result of C4.5 and C5.0 in the form of rules to be used for the classification process. Furthermore, the process of testing is used to determine performance of two algorithms. Performance will be measured by two parameters: accuracy and computation time. The test results showed accuracy from C4.5 is better than C5.0. While both algorithms computing time is almost same. So it can be concluded that C4.5 is better than C5.0.

Keywords—Occupational accident , Data mining, Classification, C4.5, C5.0.

“Halaman ini sengaja dikosongkan”

KATA PENGANTAR

Segala Puji bagi Allah SWT yang telah memberikan karunia, rahmat dan anugerah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul: **“Prediksi Pola Kecelakaan kerja Pada Perusahaan Non Ekstraktif Menggunakan Algoritma Decision Tree: C4.5 dan C5.0 ”** yang merupakan salah satu persyaratan akademis dalam menyelesaikan Program Studi S-1 pada Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya.

Tugas Akhir ini dapat diselesaikan dengan berkat kerjasama, bantuan, dan dukungan dari banyak pihak. Sehubungan dengan hal itu, penulis mengucapkan terima kasih kepada:

1. Dr. Imam Mukhlash, S.Si, MT selaku dosen pembimbing I serta Ketua Jurusan Matematika ITS yang senantiasa membimbing dengan sabar dan memberikan kritik dan saran dalam penyusunan Tugas Akhir ini.
2. Drs. Nurul Hidayat, M. Kom selaku pembimbing II yang senantiasa membimbing dan mengarahkan dalam penyusunan Tugas Akhir ini.
3. Drs. Mohammad Setijo Winarko, M. Si selaku Dosen Wali penulis.
4. Dr. Didik Khusnul Arif, S.Si., M.Si dan Drs. Iis Herisman, M.Sc. selaku selaku Kaprodi dan Sekretaris Kaprodi Jurusan Matematika yang mengkoordinir jadwal Tugas Akhir.
5. Seluruh jajaran dosen dan staf Jurusan Matematika ITS.
6. Ayah, ibu, adik tersayang yang senantiasa dengan ikhlas memberikan semangat, perhatian, kasih sayang, doa, motivasi dan nasihat yang sangat berarti bagi penulis untuk segera menyelesaikan Tugas Akhir ini.
7. Maya Nurlita Wulandari, Lailatul Mabadi Chaira, Devi Andriyani sebagai teman senasib sepenanggungan saat bimbingan ke dosen.

8. Komplotan “KambingLiarMakOnah” yang senantiasa mendampingi, mengingatkan serta saling mendoakan untuk tetap semangat sampai akhir.
9. Teman-teman seperjuangan 114 dan seluruh teman-teman Matematika Angkatan 2012 yang saya sayangi.

Tentu saja masih banyak pihak lain yang turut andil dalam penyelesaian Tugas Akhir ini yang tidak bisa penulis sebutkan satu persatu. Semoga Allah membalas dengan balasan yang lebih baik bagi semua pihak yang telah membantu penulis. *Amin ya rabbal ‘alamin.*

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan saran dan kritik dari pembaca. Akhir kata, semoga Tugas Akhir ini bermanfaat bagi semua pihak yang berkepentingan.

Surabaya, Juli 2016

Penulis

DAFTAR ISI

| | Halaman |
|---|---------|
| HALAMAN JUDUL | i |
| LEMBAR PENGESAHAN | v |
| ABSTRAK | vii |
| ABSTRACT | ix |
| KATA PENGANTAR | xi |
| DAFTAR ISI | xiii |
| DAFTAR GAMBAR | xvii |
| DAFTAR TABEL | xix |
| DAFTAR LAMPIRAN | xxi |
| | |
| BAB I. PENDAHULUAN | |
| 1.1 Latar Belakang Masalah | 1 |
| 1.2 Rumusan Masalah..... | 4 |
| 1.3 Batasan Masalah | 4 |
| 1.4 Tujuan Penelitian | 5 |
| 1.5 Manfaat Penelitian | 5 |
| 1.6 Sistematika Penulisan Tugas Akhir | 5 |
| BAB II. TINJAUAN PUSTAKA | |
| 2.1 Penelitian Terdahulu | 9 |
| 2.2 Data Mining | 12 |
| 2.3 <i>Knowledge Discovery in Database (KDD)</i> | 12 |
| 2.4 Klasifikasi | 14 |
| 2.5 <i>Decision Tree</i> | 16 |
| 2.6 Algoritma C4.5 | 17 |
| 2.7 Algoritma C5.0 | 19 |
| 2.8 <i>Information Gain</i> dan <i>Gain Ratio</i> | 22 |
| BAB III. METODOLOGI | |
| 3.1 Studi Literatur | 23 |
| 3.2 Pengumpulan Data | 23 |
| 3.3 Pengolahan Data | 25 |
| 3.3.1 <i>Pre Processing</i> | 25 |
| 3.3.2 <i>Data Mining</i> | 26 |

| | |
|---|----|
| 3.3.3 Analisis Pola dan Representasi Data..... | 27 |
| 3.4 Perancangan Interface dan Implementasi Sistem | 27 |
| 3.5 Evaluasi dan Kesimpulan..... | 27 |

BAB IV.PERANCANGAN DAN IMPLEMENTASI SISTEM

| | |
|---|----|
| 4.1 Analisis Sistem..... | 29 |
| 4.1.1 Analisis Sistem Perangkat Lunak | 29 |
| 4.1.2 Analisis Kebutuhan Sistem..... | 35 |
| 4.2 Perancangan Sistem | 36 |
| 4.2.1 Penjelasan Umum Sistem | 36 |
| 4.2.2 Perancangan Proses Algoritma | 38 |
| 4.2.2.1 Perancangan <i>Database</i> | 39 |
| 4.2.2.2 Perancangan Proses <i>Pre-processing</i> | 39 |
| 4.2.2.3 Perancangan Proses <i>Data Mining</i> . | 42 |
| 4.2.3 Perancangan Antar Muka Sistem..... | 51 |
| 4.2.3.1 Perancangan Halaman Utama | 51 |
| 4.2.3.2 Perancangan Halaman untuk Menampilkan Data Awal dan Tahap <i>Pre-processing</i> | 51 |
| 4.2.3.3 Perancangan Halaman Algoritma C4.5 | 52 |
| 4.2.3.4 Perancangan Halaman Algoritma C5.0 | 53 |
| 4.3 Implementasi Sistem..... | 54 |
| 4.3.1 Implementasi Pengambilan Data | 54 |
| 4.3.2 Implementasi Tahap <i>Pre-Processing</i> | 55 |
| 4.3.3 Implementasi Pembuatan <i>Tree</i> | 57 |
| 4.3.4 Implementasi Algoritma C4.5..... | 58 |
| 4.3.5 Implementasi Algoritma C5.0 | 59 |

BAB V. PENGUJIAN DAN PEMBAHASAN

| | |
|--|----|
| 5.1 Pre Processing dan Pembagian Data Uji..... | 63 |
| 5.2 Pengujian Algoritma C4.5 | 65 |
| 5.3 Pengujian Algoritma C5.0 | 71 |
| 5.4 Analisis Pola yang Ditemukan | 79 |

| | |
|-------------------------------------|-----------|
| BAB VI. KESIMPULAN DAN SARAN | |
| 6.1 Kesimpulan | 87 |
| 6.2 Saran | 87 |
| DAFTAR PUSTAKA | 89 |
| LAMPIRAN..... | 91 |

“Halaman ini sengaja dikosongkan”

DAFTAR GAMBAR

| | Halaman |
|-------------|--|
| Gambar 2.1 | Proses <i>Knowledge Discovery in Database</i> 14 |
| Gambar 3.1 | Diagram Alur Penelitian..... 28 |
| Gambar 4.1 | <i>Use Case Diagram</i> Perangkat Lunak <i>Data Mining</i> 30 |
| Gambar 4.2 | <i>Swimlane Diagram</i> Perangkat Lunak 31 |
| Gambar 4.3 | <i>Swimlane Diagram</i> pada <i>Pre-Processing</i> 33 |
| Gambar 4.4 | <i>Swimlane Diagram</i> pada Algoritma C4.5 34 |
| Gambar 4.5 | <i>Swimlane Diagram</i> pada Algoritma C5.0 35 |
| Gambar 4.6 | Diagram Alir Sistem 37 |
| Gambar 4.7 | Struktur <i>Database</i> 38 |
| Gambar 4.8 | Struktur Tabel Data Karyawan..... 38 |
| Gambar 4.9 | Data Mentah yang dikenai <i>Pre-Processing</i> 39 |
| Gambar 4.10 | Penolakan Masukan Data Duplikasi 40 |
| Gambar 4.11 | <i>Root</i> yang Dihasilkan dengan Algoritma C4.5. 48 |
| Gambar 4.12 | Antar Muka Halaman Utama 51 |
| Gambar 4.13 | Antar Muka Halaman Utama untuk Melakukan Proses Klasifikasi 52 |
| Gambar 4.14 | Antar Muka Halaman <i>Training</i> dengan Algoritma C4.5..... 52 |
| Gambar 4.15 | Antar Muka Halaman <i>Testing</i> dengan Algoritma C4.5..... 53 |
| Gambar 4.16 | Antar Muka Halaman <i>Training</i> dengan Algoritma C5.0..... 53 |
| Gambar 4.17 | Antar Muka Halaman <i>Testing</i> dengan Algoritma C5.0..... 54 |
| Gambar 5.1 | Hasil Tampilan Data dari <i>Database</i> 63 |
| Gambar 5.2 | Hasil <i>Pre processing</i> 64 |
| Gambar 5.3 | Hasil <i>Training</i> pada C4.5 dengan 95% Data.... 65 |
| Gambar 5.4 | Hasil <i>Testing</i> pada C4.5 dengan 5 % Data 66 |
| Gambar 5.5 | Hasil <i>Training</i> pada C4.5 dengan 90% Data.... 66 |
| Gambar 5.6 | Hasil <i>Testing</i> pada C4.5 dengan 10% Data 67 |
| Gambar 5.7 | Hasil <i>Training</i> pada C4.5 dengan 85% Data.... 67 |

| | | |
|-------------|---|----|
| Gambar 5.8 | Hasil <i>Testing</i> pada C4.5 dengan 15 % Data | 68 |
| Gambar 5.9 | Hasil <i>Training</i> pada C4.5 dengan 80% Data | 68 |
| Gambar 5.10 | Hasil <i>Testing</i> pada C4.5 dengan 20 % Data | 69 |
| Gambar 5.11 | Hasil <i>Training</i> pada C4.5 dengan 75% Data | 70 |
| Gambar 5.12 | Hasil <i>Testing</i> pada C4.5 dengan 25 % Data | 70 |
| Gambar 5.13 | Hasil <i>Training</i> pada C5.0 dengan 95% Data | 71 |
| Gambar 5.14 | Hasil <i>Testing</i> pada C5.0 dengan 5 % Data | 72 |
| Gambar 5.15 | Hasil <i>Training</i> pada C5.0 dengan 90% Data | 73 |
| Gambar 5.16 | Hasil <i>Testing</i> pada C5.0 dengan 10% Data | 73 |
| Gambar 5.17 | Hasil <i>Training</i> pada C5.0 dengan 85% Data | 74 |
| Gambar 5.18 | Hasil <i>Testing</i> pada C5.0 dengan 15 % Data | 74 |
| Gambar 5.19 | Hasil <i>Training</i> pada C5.0 dengan 80% Data | 75 |
| Gambar 5.20 | Hasil <i>Testing</i> pada C5.0 dengan 20 % Data | 76 |
| Gambar 5.21 | Hasil <i>Training</i> pada C5.0 dengan 75% Data | 77 |
| Gambar 5.22 | Hasil <i>Testing</i> pada C5.0 dengan 25 % Data | 77 |
| Gambar 5.23 | Grafik Perbandingan Keakuratan C4.5 dan C5.0 | 78 |

DAFTAR TABEL

| | Halaman |
|---|---------|
| Tabel 2.1 Perbandingan Hasil Penelitian | 11 |
| Tabel 4.1 Tabel Kebutuhan Sistem..... | 36 |
| Tabel 4.2 Hubungan Atribut Tipe Pekerjaan dan Waktu Per Hari..... | 41 |
| Tabel 4.3 Hubungan Atribut Lama Bekerja dan Umur..... | 41 |
| Tabel 4.4 Perbedaan Atribut saat Transformasi..... | 42 |
| Tabel 4.5 Perubahan Nilai Pada Atribut Umur | 42 |
| Tabel 4.6 Perhitungan Jumlah Kemunculan Kejadian..... | 43 |
| Tabel 4.7 Perhitungan <i>Entropy</i> dan <i>Information Gain</i> | 47 |
| Tabel 4.8 Perhitungan <i>Entropy</i> , <i>Information Gain</i> , dan <i>Gain Ratio</i> | 59 |
| Tabel 5.1 Pembagian Data yang Digunakan..... | 64 |
| Tabel 5.2 Perbandingan Hasil Algoritma C4.5 | 71 |
| Tabel 5.3 Perbandingan Hasil Algoritma C5.0 | 78 |
| Tabel 5.4 Pola yang Dihasilkan oleh Algoritma C4.5 | 79 |
| Tabel 5.5 Pola yang Dihasilkan oleh Algoritma C5.0 | 82 |

“Halaman ini sengaja dikosongkan”

DAFTAR LAMPIRAN

| | Halaman |
|------------------|---------|
| LAMPIRAN A | 91 |
| LAMPIRAN B | 109 |
| LAMPIRAN C | 123 |

“Halaman ini sengaja dikosongkan”

BAB I

PENDAHULUAN

Pada bab ini dijelaskan tentang latar belakang pembuatan Tugas Akhir, rumusan dan batasan permasalahan yang dihadapi dalam penelitian Tugas Akhir, tujuan dan manfaat pembuatan Tugas Akhir dan sistematika penulisan Tugas Akhir.

1.1. Latar Belakang Masalah

Kegiatan penggunaan teknologi dalam perusahaan (industri) di Indonesia beberapa tahun ini mengalami kemajuan. Budaya yang masih menggunakan teknologi yang sangat tradisional dan masih menggunakan tangan, telah diganti dengan memanfaatkan teknologi yang lebih maju. Dengan adanya perubahan ini, pihak perusahaan dihadapkan pada perubahan kebiasaan aturan. Aturan kerja yang berubah akan membawa dampak langsung bagi karyawan yang melakukan pekerjaan langsung dengan teknologi yang digunakan. Dengan demikian, resiko kerja yang dimiliki oleh perusahaan akan meningkat dan akhirnya akan menimbulkan terjadinya insiden kerja atau kecelakaan kerja.

Menurut Coling David A (1990) kecelakaan kerja adalah setiap kejadian tak terencana dan tidak terkontrol yang disebabkan oleh manusia, faktor situasi atau lingkungan atau merupakan kombinasi dari faktor-faktor tersebut yang mengganggu proses kerja yang mungkin berakibat atau tidak berakibat cedera, penyakit, kematian, kerusakan harta benda atau kejadian lain yang tidak diharapkan tetapi berpotensi untuk terjadinya hal tersebut [8].

Saat ini masalah kecelakaan kerja menjadi perhatian bagi Direktorat Jenderal Pembinaan dan Pengawasan Ketenagakerjaan. Tingkat kecelakaan kerja di Indonesia memang cukup tinggi dibanding dengan negara di Eropa dan Asia lainnya. Pada tahun 2010 tercatat tingkat kecelakaan kerja di Indonesia mencapai 98.711 kasus. Namun pada tahun 2011, angka tersebut

mengalami penurunan, yaitu 86.386 kasus. Hal terpenting dalam masalah kecelakaan kerja ialah bagaimana cara meminimalisir tingkat kecelakaan kerja yang terjadi pada suatu perusahaan. Untuk itu, tindakan preventif dan pengetahuan tentang aturan kerja sangat penting bagi karyawan perusahaan. Salah satu hal penting dalam menentukan tindakan preventif ini dengan mengetahui pola penyebab kecelakaan kerja yang telah terjadi sebelumnya. Salah satu cara yang dapat dilakukan untuk membuat pola kecelakaan kerja tersebut dapat dilakukan dengan *data mining*.

Data mining merupakan proses ekstraksi pola yang penting dari data dalam skala besar [4]. Penerapan *data mining* dalam menganalisis pola kecelakaan kerja diharapkan mampu menggali informasi dan menganalisis hasilnya untuk menentukan tindakan preventif yang dapat dilakukan oleh pihak perusahaan. Klasifikasi merupakan salah satu metode yang dapat digunakan untuk menemukan pola kecelakaan kerja [4]. Dari pola kecelakaan kerja yang didapatkan, kita dapat memprediksi kecelakaan kerja yang terjadi, sehingga pihak perusahaan dapat mengambil tindakan preventif yang tepat. Klasifikasi yang dapat digunakan untuk menentukan pola kecelakaan kerja ini bekerja dengan menentukan variabel penyebab kecelakaan kerja, menentukan variabel yang saling berhubungan dan menentukan kapasitas variabel tersebut dalam menentukan terjadinya kecelakaan kerja. Ketepatan dalam penentuan pola ini sangat penting saat mengambil keputusan tindakan preventif yang dipilih.

Rivas T dkk telah melakukan penelitian mengenai penjelasan dan prediksi pola kecelakaan kerja pada tahun 2011 [11]. Penelitian ini bertujuan untuk membandingkan beberapa algoritma dalam melihat keakuratan tiap-tiap algoritma yang digunakan. Studi kasus yang digunakan dalam penelitian ini adalah 2 perusahaan dengan bidang yang berbeda, yaitu sektor pertambangan dan konstruksi sipil. Variabel yang dipakai sebanyak 18 macam yang berasal dari faktor internal karyawan

dan sistem perusahaan itu sendiri. Hasil dari penelitian ini membuktikan bahwa metode Bayesian dengan Algoritma K2 adalah algoritma dengan keakuratan tertinggi (88,71%).

Pada penelitian lainnya, terdapat penerapan Algoritma C4.5 dan C5.0 yang menghasilkan keakuratan lebih tinggi. Penelitian Samuel A. Moore membandingkan 3 algoritma yang saling berhubungan yang berada di metode algoritma *decision tree*, yaitu C4.5, C5.0 dan J48, dimana C5.0 merupakan pengembangan dan bentuk komersil dari C4.5 dan J48 merupakan implementasi C4.5 pada platform *data mining* WEKA [7]. Penelitian ini berlandaskan penelitian Rulequest yang menyatakan bahwa algoritma *decision tree* menggunakan C5.0 memberikan hasil yang lebih baik dibandingkan C4.5, khususnya pada dataset yang panjang. Hasil dari penelitian ini menunjukkan bahwa algoritma *decision tree* dengan C4.5 memiliki performa yang lebih baik dibanding C5.0 dan J48, yaitu 90%.

Berdasarkan latar belakang tersebut, penulis ingin mengangkat topik serupa sebagai Tugas Akhir dan diberi judul “Prediksi Pola Kecelakaan Kerja Pada Perusahaan Non Ekstraktif Menggunakan Algoritma Decision Tree: C4.5 dan C5.0”. Pada penelitian Rivas, terdapat Algoritma ID3 yang merupakan cikal bakal dari Algoritma C4.5 [11]. Salah satu kelemahan algoritma ini yaitu hanya dapat menggunakan variabel diskrit. Selain ID3, ada juga Algoritma J48, yang merupakan implementasi Algoritma C4.5 pada platform WEKA dengan variabel diskrit dan kontinu yang direpresentasikan dengan *n-ary tree*. Jadi dapat disimpulkan bahwa kedua algoritma tersebut memiliki hubungan dengan Algoritma C4.5 yang saat ini banyak dipakai oleh peneliti dalam melakukan pengklasifikasian data. IEEE International Conference on Data Mining (ICDM) pada tahun 2006, menyatakan bahwa Algoritma C4.5 merupakan algoritma paling banyak dipakai dalam melakukan *klasifikasi* [13]. Hal ini karena performa C4.5 yang paling baik dalam kelas dan eranya. Algoritma C5.0 merupakan pengembangan dan versi komersil dari Algoritma C4.5 [5, 10].

Dataset untuk penelitian ini diambil dari PT. Ajinomoto Indonesia kantor cabang Mojokerto. Perusahaan ini termasuk industri non ekstraktif yang bergerak dibidang kuliner. Dibawah SOP (*Standart Operational Procedure*) Jepang, perusahaan ini menggunakan *rule mode* yang mementingkan keselamatan kerja karyawan. Dengan adanya pengarahan pihak perusahaan tentang pengetahuan dan ketepatan dalam menjalankan aturan kerja bagi karyawan, maka tingkat keselamatan kerja tinggi dan resiko terjadinya kecelakaan kerja menjadi rendah. Tingkat resiko kerja inilah yang menjadi menjadi tolok ukur keberhasilan perusahaan dalam menjaga kondisi.

Pada Tugas Akhir ini akan dibangun suatu perangkat lunak yang mampu menggali pola pada data menggunakan metode klasifikasi, yaitu algoritma *decision tree*. Selain itu, dapat mengetahui perbandingan tingkat keakuratan dari algoritma yang digunakan, yaitu C4.5 dan C5.0.

1.2. Rumusan Masalah

Berdasarkan uraian di atas, penulis merumuskan masalah yang timbul yaitu :

1. Bagaimana menggali pola kecelakaan kerja pada suatu perusahaan menggunakan *data mining* dengan metode klasifikasi?
2. Berapa banyak model prediksi yang dihasilkan dari klasifikasi menggunakan algoritma *decision tree* yaitu Algoritma C4.5 dan C5.0?
3. Bagaimana perbandingan keakuratan antara hasil yang diberikan oleh Algoritma C4.5 dan C5.0?

1.3. Batasan Masalah

Pada penelitian ini, dibuat batasan masalah sebagai berikut :

1. Objek yang digunakan pada penelitian ini adalah 11 atribut yang berasal dari 200 data kecelakaan kerja.
2. Dalam penelitian ini, terdapat 3 kategori objek, yaitu Event, Karyawan, dan Sistem Perusahaan.

3. Data yang digunakan pada penelitian ini adalah data dari kantor produksi PT. Ajinomoto Indonesia cabang Mojokerto.

1.4. Tujuan Penelitian

Tujuan Umum

Implementasi Algoritma C4.5 dan C5.0 dalam perangkat lunak untuk mendapatkan pola kecelakaan kerja sehingga dapat diambil keputusan dalam tindakan preventifnya.

Tujuan Khusus

1. Membuat sebuah perangkat lunak yang memuat Algoritma C4.5 dan C5.0 sehingga dapat membuat pola dari data training.
2. Mencari perbandingan tingkat keakuratan hasil pengklasifikasian dari Algoritma C4.5 dan C5.0 dalam perangkat lunak.

1.5. Manfaat Penelitian

Adapun manfaat Tugas Akhir ini adalah sebagai berikut :

1. Diperoleh informasi dan perbandingan mengenai pola kecelakaan kerja menggunakan metode *decision tree* yaitu C4.5 dan C5.0.
2. Diperoleh suatu informasi untuk mengambil tindakan preventif untuk terjadinya kecelakaan kerja pada perusahaan yang menjadi studi kasus.
3. Sebagai salah satu referensi penggunaan klasifikasi dalam penggalan pola dengan Algoritma C4.5 dan C5.0.

1.6. Sistematika Penulisan Tugas Akhir

Sistematika penulisan didalam Tugas Akhir ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini dijelaskan tentang latar belakang pembuatan Tugas Akhir, rumusan dan batasan permasalahan yang dihadapi dalam penelitian Tugas

Akhir, tujuan dan manfaat pembuatan Tugas Akhir dan sistematika penulisan Tugas Akhir.

BAB II KAJIAN TEORI

Pada bab ini dijelaskan tentang kajian teori dari referensi penunjang serta penjelasan permasalahan yang dibahas dalam Tugas Akhir ini, meliputi penelitian sebelumnya terkait Tugas Akhir, pengertian *Data Mining*, *Knowledge Discovery in Database* (KDD), Klasifikasi, *Decision Tree*, Algoritma C4.5, Algoritma C5.0, *Information Gain* dan *Gain Ratio*.

BAB III METODOLOGI PENELITIAN

Pada bab ini berisi metodologi atau urutan kegiatan yang dilakukan dalam menyelesaikan Tugas Akhir, meliputi studi literatur, pengumpulan data, analisis dan desain sistem, pembuatan program, uji coba dan evaluasi, hingga penulisan Tugas Akhir.

BAB IV PERANCANGAN DAN IMPLEMENTASI

Pada bab ini dijelaskan analisis data, perancangan gambaran umum sistem, perancangan algoritma, perancangan data, dan perancangan antar muka sistem. Sistem ini memiliki *input* berupa data mentah yang berasal dari *database* yang selanjutnya dilakukan tahap *pre-processing* untuk menghilangkan *noise* pada data yang berupa duplikasi data, *missing value data* dan transformasi data. Hasil dari *pre-processing* ini berupa data kategorikal yang siap digunakan sebagai *input* untuk menjalankan algoritma. Dalam menjalankan algoritma, tidak bisa dilakukan secara bersamaan. Akan tetapi, kedua algoritma dapat dijalankan dengan satu kali *pre-processing*. Hasil dari proses ini

adalah pola kecelakaan kerja berupa *rules* pada tiap algoritma, yang selanjutnya akan di uji sehingga memunculkan angka keakurasian pada tiap algoritma tersebut.

BAB V PENGUJIAN DAN PEMBAHASAN HASIL

Pada bab ini akan ditampilkan hasil *training* berupa *rules*, hasil testing berupa angka keakuratan, dan selanjutnya angka tersebut akan dibandingkan. Hasil pengujian inilah yang akan digunakan dalam perumusan kesimpulan dan saran.

BAB VI PENUTUP

Bab ini merupakan penutup, berisi tentang kesimpulan yang dapat diambil berdasarkan data yang ada dan saran yang selayaknya dilakukan bila Tugas Akhir ini dilanjutkan.

“Halaman ini sengaja dikosongkan”

BAB II TINJAUAN PUSTAKA

Pada bab ini dijelaskan tentang kajian teori dari referensi penunjang serta penjelasan permasalahan yang dibahas dalam Tugas Akhir ini, meliputi penelitian sebelumnya terkait Tugas Akhir, pengertian Data Mining, *Knowledge Discovery in Database* (KDD), Klasifikasi, *Decision Tree*, Algoritma C4.5, Algoritma C5.0, Information Gain dan Gain Ratio.

2.1 Penelitian Terdahulu

Penelitian mengenai kecelakaan kerja telah banyak dilakukan sebelumnya. Topik ini terlihat sangat menarik karena fleksibelitasnya dan banyak digunakan untuk inovasi baru. Dalam *data mining*, data kecelakaan kerja telah banyak digunakan sebagai objek penelitian. Selain menguji data kecelakaan kerja itu sendiri, data ini dapat digunakan untuk menguji algoritma-algoritma *data mining*, seperti membandingkan tingkat keakuratan dari algoritma.

Rivas T (2010) mengangkat topik kecelakaan kerja untuk membandingkan keakuratan 15 algoritma yang ada dalam *data mining* [11]. Variabel yang digunakan sebanyak 18 variabel yang dibagi kedalam 4 kategori, yaitu *event-related variables* (n=3), *worker-related variables* (n=9), *company-related variables* (n=5), *prediction variable*. Penelitian ini terdiri dari 2 tahap, yaitu inisialisasi variabel yang paling relevan dan menghitung kapasitas variabel tersebut untuk memprediksi terjadinya kecelakaan kerja.

Pada tahap pertama, menggunakan *cross validation* sebanyak 3 kali, yaitu dengan 10, 5 dan 3 grup. Fungsi atribut pilihan WEKA dikelompokkan menjadi 31 metode yang dikelompokkan menjadi 2 kategori berbeda yaitu *WrapperSubSetEval* (memilih variabel berdasarkan tingkat relasinya dengan variabel respon) dan *CfsSubSetEval* (memilih variabel berdasarkan penggabungan dari setiap variabel pada 19 model *data mining* yang berbeda). Pada tahap kedua, dihasilkan bobot relevansi untuk variabel

prediksi dari tahap pertama. Variabel yang telah memiliki relevansi kemudian dihitung persentase kapasitasnya terhadap penyebab kecelakaan kerja menggunakan 15 algoritma *data mining*.

Hasilnya menunjukkan bahwa Bayesian Network dengan Algoritma K2 memiliki keakuratan sebesar 88,71%. Setelah itu, keakuratan algoritma yang lain Bayesian dengan TAN Algorithm (83,87%), Bayesian dengan Hill-Climber Algorithm (85,48%), Naïve Bayes (83,87%), SVM (87,10%), Logistik Regression (72,58), ID3 Algorithm (77,42%), J48 Algorithm (87,10%), LMT Algorithm (82,26%), Decision Rule PART Algorithm (85,48%), dan Decision Rule OneR Algorithm (75,81%).

Dapat dilihat bahwa penelitian diatas menggunakan beberapa macam teknik klasifikasi, yaitu *Logostic Regression*, *Decision Tree*, *Bayesian Network*, dan *Support Vector Machine*. Rata-rata hasil yang ditunjukkan bahwa teknik *Bayesian Network* adalah teknik yang memiliki tingkat keakuratan tertinggi, yaitu 85,21%. Sementara yang lainnya, seperti *Logistic Regression* 72,58%, *Decision Tree* 82,26%, *Decision Rules* 80,64%, dan *SVM* 87.10%.

Penelitian mengenai perbandingan tingkat keakuratan algoritma pernah dilakukan oleh Moore Samuel A dari Universitas Fordham, USA [7]. Penelitian ini membandingkan C4.5, C5.0 dan J48 yang merupakan algoritma dari *decision tree*. Penelitian ini membuktikan bahwa penelitian dari Rulequest tidak benar, yaitu yang menyatakan bahwa C5.0 memiliki performa yang lebih baik dari C4.5. Hasil penelitian berdasarkan pada nilai rata-rata dari *F-measure* dan *error rate* dari masing-masing algoritma. Data yang digunakan merupakan data dari *UCI Repository*. Data set dibagi menjadi 4 macam, yaitu semua data set (semua data dipakai), *smaller data sets* (<1000), *larger data set* (>1000), *abs. diff* (absolute difference). Perbandingan ketiga algoritma dapat dilihat pada Tabel 2.1

Tabel 2.1. Perbandingan Hasil Penelitian [7]

| Algoritma | Terbaik | Tingkat Keakuratan | |
|--------------|---------|--------------------|-----|
| | | F-meas | ER |
| C4.5 vs C5.0 | C4.5 | .90 | .10 |
| C4.5 vs J48 | C4.5 | .68 | .32 |
| C5.0 vs J48 | C5.0 | .84 | .18 |

Dari tabel diatas, dapat dilihat bahwa hasil penelitian Moore tidak sesuai dengan penelitian Rulequest. Terlihat bahwa C4.5 memiliki performa lebih dibanding C5.0 untuk *dataset* ini.

Dalam beberapa penelitian, penggunaan C4.5 dan C5.0 telah banyak digunakan secara terpisah. Kemal Polat telah mencoba meneliti keakuratan Algoritma C4.5 [9]. Studi kasus penelitian ini adalah peningkatan keakuratan Algoritma C4.5 dalam masalah *multi-class*. Klasifikasi masalah *multi-class* yang ada saat ini antara lain kategorisasi dokumen, pengenalan gambar, klasifikasi sinyal, dan lain sebagainya. Topik yang diangkat dalam penelitian ini adalah *hybrid novel* [9]. Variabel yang diangkat dalam penelitian ini adalah dermatologi, segmentasi citra dan limfografi, dengan mengambil dataset dari *UCI Machine Learning Database*. Metode evaluasi yang digunakan yaitu menggunakan *10-fold cross validation* dan menggunakan *one-against-all approach* [9]. Untuk metode *10-fold cross validation*, keakuratan algoritma C4.5 di tiap-tiap topik sebesar 84.48%, 88.79%, dan 80.11%. Hasil tersebut telah meningkat saat memakai aturan *one-against-all approach* yaitu sebesar 96.71%, 95.18%, dan 87.95%. Sampai saat ini, penelitian ini telah banyak digunakan sebagai patokan penelitian sejenis. Metode ini dapat digunakan dalam banyak aplikasi pengenalan pola.

Tomasz Bujlow juga telah menggunakan Algoritma C5.0 [1]. Permasalahan yang diangkat adalah pemantauan kecepatan kinerja infrastruktur jaringan internet. Hal ini karena kecepatan pelayanan akan menentukan kualitas pelayanan ke pengguna. Saat ini metode yang digunakan adalah *Multi-hop Network* dan *Monitoring Backbone QoS* [1]. Karena banyak kelemahan dengan

metode ini, maka penelitian ini berusaha memecahkan masalah menggunakan Algoritma C5.0. Variabel yang digunakan adalah 7 aplikasi berbeda dalam 76.632 – 1.622.710 kasus tak terduga. Penelitian ini menunjukkan pengumpulan data lalu lintas internet, penyajian argumen dalam proses klasifikasi, pengenalan C5.0 dan cara kerjanya, serta evaluasi dan membandingkan hasilnya. Tingkat keakuratan yang didapat sebesar 99.3 %- 99.9%.

2.2 Data Mining

Data Mining merupakan proses ekstraksi pola yang penting dari data dalam jumlah besar. *Data mining* merupakan salah satu langkah dalam proses *Knowledge Discovery in Database* (KDD) untuk menemukan pola yang bermanfaat. *Data mining* juga didefinisikan sebagai suatu proses yang menggunakan berbagai perangkat analisis data untuk menemukan pola dan relasi data agar dapat digunakan untuk membuat prediksi dengan tepat. Terdapat dua kunci sukses dalam *data mining*, yaitu ketepatan formulasi masalah yang akan dipecahkan dan penggunaan data yang tepat. Setelah mendapatkan data yang akan digunakan untuk dianalisis, ketepatan dalam mentransformasikan dan mengolahnya sangat mempengaruhi ketepatan pengetahuan hasil penggalian data [4].

Beberapa karakteristik umum yang dimiliki data yang akan dianalisis adalah sebagai berikut [4]:

1. Ukuran Data Besar .
2. Data tidak lengkap, sehingga perlu proses *cleaning*.
3. Struktur data kompleks.
4. Merupakan kumpulan data yang heterogen.

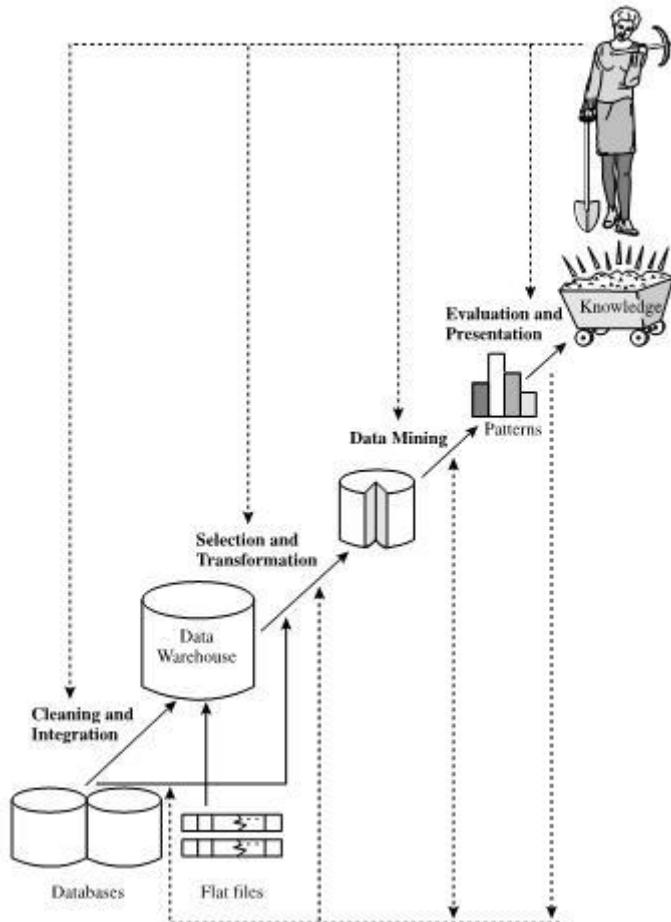
2.3 Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah proses menemukan informasi yang bermanfaat serta pola-pola yang ada dalam data. KDD merupakan sebuah proses yang terdiri dari serangkaian proses iteratif yang terurut dan *data mining*

merupakan salah satu langkah dalam KDD. Ada Beberapa langkah yang dilakukan dalam KDD [4]:

1. Pembersihan Data
Pembersihan data dilakukan untuk menghilangkan data yang tidak konsisten atau mengandung *noise*.
2. Integrasi Data
Proses integrasi data dilakukan untuk menggabungkan data dari berbagai basis data.
3. Seleksi Data
Proses seleksi data dilakukan dengan mengambil data yang relevan untuk proses analisis.
4. Transformasi Data
Proses ini dilakukan untuk mentransformasikan atau menggabungkan data ke dalam bentuk yang tepat agar proses *data mining* dapat dilakukan dengan lebih mudah.
5. *Data Mining*
Data Mining merupakan proses penting dimana metode-metode tertentu diaplikasikan untuk mengekstrak pola-pola unik dalam data.
6. Evaluasi Pola
Evaluasi pola diperlukan untuk mengidentifikasi pola-pola menarik yang ditemukan sebagai bahan dalam representasi pengetahuan.
7. Presentasi Pengetahuan
Penggunaan visualisasi dan teknik representasi untuk menunjukkan pengetahuan hasil penggalian data kepada pengguna.

Tahapan-tahapan proses KDD secara berurut dapat dilihat pada Gambar 2.1 berikut ini:



Gambar 2.1 *Proses Knowledge Discovery in Database [4]*

2.4 Klasifikasi

Klasifikasi adalah proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [4].

Proses klasifikasi data terdiri dari 2 tahap. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rules*. Proses kedua adalah tahap klasifikasi, dimana data *testing* digunakan untuk memperkirakan akurasi dari *rules* yang dihasilkan pada tahap *learning* [6].

Proses klasifikasi didasarkan pada empat komponen [3]:

- a. Kelas : variabel dependen yang berupa kategorikal yang merepresentasikan “label” yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempaa.
- b. *Predictor* : variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.
- c. *Training dataset* : suatu *dataset* yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
- d. *Testing dataset* : *dataset* yang berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

Berikut ini adalah algoritma klasifikasi dalam *data mining* yang paling populer yaitu adalah [3]:

- a. *Decision/* pohon klasifikasi
- b. *Bayesian classifiers/Naive Bayes classifiers*
- c. *Neural networks*
- d. *Statistical analysis*
- e. *Rough sets*
- f. *k-nearest neighbor classifier*
- g. *Rule-based methods*
- h. *Memory based reasoning*
- i. *Support vector machines.*

2.5 Decision Tree

Decision tree adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia [12]. *Decision tree* adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari *decision tree* adalah mengubah data menjadi pohon keputusan atau aturan-aturan keputusan. Pohon tersebut juga memperlihatkan faktor-faktor kemungkinan atau probabilitas yang akan mempengaruhi alternatif-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila mengambil alternatif keputusan tersebut.

Metode ini digunakan untuk memprediksi nilai diskrit dari fungsi target, yang mana fungsi pembelajaran dipresentasikan oleh sebuah *decision tree* [12]. *Decision tree* merupakan himpunan IF...THEN. Setiap *path* dalam *tree* dihubungkan sebuah aturan, dimana presi terdiri dari sekumpulan node - node yang ditemui, dan kesimpulan dari aturan terdiri atas kelas yang terhubung dengan *leaf* dari *path*

```

TreeGrowth (E,F)
1  if stopping_cond(E,F) = true then
2    leaf = createNode ().
3    leaf.label = Classify (E).
4    return leaf
5  Else
6    root = createNode()
7    root.test_cond = find_best_split(E,F)
8    let V = {v | v is a possible outcome of root.test_cond}
9    for each v ∈ V do
10      $E_v = \{e \mid \text{root.test\_cond}(e) = v \text{ and } e \in E\}$ .
11     child = TreeGrowth ( $E_v$ , F).
12     and child as desendent of root and label the aged (root
    → child) as v.
13   end foor
14 and if
15 return root.

```

Algoritma diatas menjelaskan pembangunan *tree* pada *decision tree*. Data masukan yang digunakan pada algoritma tersebut adalah himpunan data “E” dan “F”. “E” merupakan atribut yang akan diklasifikasikan kedalam atribut target “F”. Algoritma akan terus berjalan sampai batas yang ditentukan oleh *user*. Keluaran dari algoritma ini berupa *tree* yang merupakan representasi dari hasil klasifikasi data.

Terdapat 3 jenis node yang terdapat pada *decision tree* [12], yaitu:

1. *Root node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
2. *Internal Node*, merupakan *node* percabangan. Pada *node* ini terdapat percabangan. Pada *node* ini terdapat satu *input* dan memiliki *output* minimal dua.
3. *Leaf node* atau *terminal node*, merupakan *node* akhir. Pada *node* ini terdapat satu *input* dan tidak mempunyai *output*.

2.6 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan [5,10]. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti. *Structured Query Language* untuk mencari *record* pada kategori tertentu.

Contoh Algoritma C4.5 dapat dilihat dibawah ini [5].

```

Input: an attribute-prediktord dataset D
1   Tree = {}
2   if D is “pure” OR other stopping criteria met then
3     Terminate
4   end if
5   for all attribute  $a \in D$  do
6     compute information-theoretic criteria if we split on a

```

```

7   end for
8   abest = Best attribute according to above computed
      criteria
9   Tree = Create a decision node that tests abest in the root
10  Dv = Induced sub-datasets from D based on abest
11  for all Dv do
12    Tree v = C4.5(Dv)
13    Attach Tree v to the corresponding branch of Tree
14  end for
15  return Tree

```

Algoritma diatas menjelaskan jalannya Algoritma C4.5. Data masukan yang digunakan adalah dataset “D”. Secara ringkas, langkah kerja Algoritma C4.5 dapat digambarkan sebagai berikut [10]:

- Hitung setiap kemunculan kejadian pada tiap *prediktor* dari atribut yang digunakan terhadap *prediktor* dari atribut target.
- Hitung *entropy total* dan *entropy* untuk setiap *prediktor* atribut. *Entropy* merupakan kontroler pada atribut yang menyebabkan atribut tersebut digunakan dalam pembangunan *tree*. Hal ini karena tidak semua atribut akan digunakan pada tiap percabangan *tree*. Perhitungan *entropy* ini akan dibahas lebih lanjut pada subbab 2.8.
- Hitung *information gain* untuk tiap atribut. *Information gain* inilah yang akan digunakan untuk membuat *tree*. Pada iterasi pertama, atribut yang mempunyai nilai *information gain* tertinggi akan menjadi *root tree*. Perhitungan *information gain* ini akan dibahas lebih lanjut pada subbab 2.8.
- Bentuk simpul pertama yang berisi atribut yang memiliki *information gain* tertinggi.
- Ulangi perhitungan *information gain* sampai semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikutkan lagi dalam perhitungan nilai *information gain* pada iterasi berikutnya.

2.7 Algoritma C5.0

Algoritma C5.0 adalah salah satu algoritma *data mining* yang khususnya diterapkan pada *decision tree* [2]. C5.0 merupakan penyempurnaan algoritma sebelumnya yang dibentuk oleh Ross Quinlan pada tahun 1987, yaitu ID3 dan C4.5. Dalam algoritma ini pemilihan atribut diproses menggunakan *gain ratio*. Dalam memilih atribut untuk memisah objek dalam beberapa kelas harus dipilih atribut yang menghasilkan *gain ratio* paling besar. Atribut dengan nilai *gain ratio* terbesar akan dipilih sebagai *parent* bagi *node* selanjutnya.

C5.0 menghasilkan *tree* dengan jumlah cabang per *node* bervariasi [2]. Perbedaan yang dapat dilihat secara fisik yaitu *tree* yang akan dihasilkan oleh Algoritma C5.0 akan lebih ringkas jika dibandingkan dengan Algoritma C4.5. Karena itulah pembangunan *tree* pada algoritma ini lebih cepat dibandingkan Algoritma C4.5.

Langkah kerja pembangunan *tree* pada Algoritma C5.0 mirip dengan pembangunan Algoritma C4.5. Kemiripan tersebut meliputi perhitungan kemunculan kejadian, perhitungan *entropy* dan *information gain*. Jika pada Algoritma C4.5 berhenti sampai perhitungan *information gain*, maka pada Algoritma C5.0 akan melanjutkannya dengan perhitungan *gain ratio* dengan menggunakan *information gain* dan *entropy* yang telah ada. Perhitungan *gain ratio* akan dibahas pada Subbab 2.8.

2.8 Information Gain dan Gain Ratio

Ukuran *information gain* dan *gain ratio* digunakan untuk memilih atribut uji pada setiap *node* di dalam *tree*. Ukuran ini digunakan untuk memilih atribut atau *node* pada *tree*. Atribut dengan nilai *information gain* tertinggi akan terpilih sebagai *parent* bagi *node* selanjutnya pada Algoritma C4.5. Persamaan yang digunakan untuk *information gain* adalah [6]

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

dengan :

S = himpunan kasus

n = jumlah partisi |S|

p_i = proporsi dari |S_i| terhadap |S|

S adalah sebuah himpunan yang terdiri dari s data objek. Diketahui atribut *class* adalah m dimana mendefinisikan kelas-kelas di dalamnya, C_i (for i= 1, ..., m), S_i adalah jumlah objek pada S dalam *class* C_i. Untuk mengklasifikasikan objek yang digunakan maka diperlukan informasi dengan menggunakan aturan seperti di atas. Dimana p_i adalah proporsi kelas dalam output seperti pada kelas C_i dan diestimasikan dengan S_i / S.

Atribut A memiliki nilai tertentu {a₁, a₂, ..., a_v}. Atribut A dapat digunakan pada partisi S ke dalam v *subset*, {S₁, S₂, ..., S_v}, dimana S_j berisi objek pada S yang bernilai a_j pada A. Jika A dipilih sebagai atribut uji (sebagai contoh atribut terbaik untuk split), maka *subset* ini akan berhubungan pada cabang dari node himpunan S. S_{ij} adalah jumlah objek pada *class* C_i dalam sebuah *subset* S_j.

Untuk mendapatkan nilai *information gain* selanjutnya digunakan persamaan dibawah ini :

$$InformationGain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Dimana :

S = himpunan kasus

A = Atribut

n = jumlah artisi atribut A

|S_i| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

Perhitungan *gain ratio* untuk Algoritma C5.0 akan berjalan setelah perhitungan *information gain* diatas dilakukan. Perhitungan *gain ratio* selanjutnya menggunakan persamaan dibawah ini :

$$GainRatio = \frac{Information\ Gain\ (S, A)}{\sum_{i=1}^n Entropy\ (S_i)} \quad (3)$$

Dengan adanya perhitungan *gain ratio* inilah yang menjadikan pembangunan *tree* pada C5.0 lebih ringkas dibanding *tree* pada Algoritma C4.5.

Hasil dari pembangunan *tree*, akan direpresentasikan dalam bentuk *rules* sebagai pola kecelakaan kerja yang terbentuk. Selanjutnya pola ini dihitung keakuratannya pada tahap *testing*. Untuk menghitung keakuratan *rules* dari tiap algoritma, dapat menggunakan persamaan dibawah ini :

$$Akurasi = \frac{Nilai\ Kecocokan}{Jumlah\ Seluruh\ Kejadian} * 100\% \quad (4)$$

Akurasi juga dapat memperlihatkan kecocokan antara algoritma dan data yang digunakan sebagai studi kasus. Hal ini karena jenis data juga dapat mempengaruhi akurasi yang akan dihasilkan dari proses *data mining* yang diterapkan.

“Halaman ini sengaja dikosongkan”

BAB III METODE PENELITIAN

Bab ini membahas mengenai metodologi sistem yang digunakan untuk menyelesaikan Tugas Akhir. Pembahasan metodologi sistem diawali dengan penjelasan tentang objek penelitian, peralatan yang digunakan, dan tahap penelitian.

3.1. Studi Literatur

Studi Literatur ini dilakukan untuk identifikasi permasalahan dengan mencari referensi yang menunjang penelitian yang berupa Tugas Akhir, jurnal, buku, maupun artikel. Dalam Tugas Akhir ini, studi literatur yang dilakukan mengenai *data mining*, *classification*, *Decision Tree*, Algoritma C4.5, dan Algoritma C5.0.

3.2. Pengumpulan Data

Pengumpulan data merupakan tahap untuk mengumpulkan data yang diperlukan dalam pengerjaan Tugas Akhir ini. Data yang digunakan dalam Tugas Akhir ini berupa data sekunder, yang berupa data karyawan dan sistem perusahaan. Data sekunder adalah data yang diperoleh penulis secara langsung (dari tangan pertama). Data sekunder diambil di PT. Ajinomoto Indonesia kantor cabang Mojokerto.

Atribut data yang akan dipakai dalam Tugas Akhir ini antara lain:

- Id kejadian** : Id kejadian kecelakaan kerja merupakan waktu kunci primer sebagai identitas kejadian.
- Nama** : Nama karyawan yang mengalami kecelakaan
- Jenis kelamin** : Jenis kelamin yang mengalami kecelakaan. Nilai dari atribut ini yaitu
 1. Laki- laki (L)

2. Perempuan (P)
- Umur : Umur karyawan saat mengalami kecelakaan kerja. Atribut ini akan dijadikan nilai kategorikal pada tahap transformasi, menjadi :
1. Muda
 2. Tua
- Kewarganegaraan : Asal negara karyawan yang mengalami kecelakaan kerja
- Tipe pekerjaan : Departemen karyawan yang mengalami kecelakaan kerja. Nilai dari data ini yaitu :
1. GRD A
 2. GRD B
 3. GRD C
- Waktu per hari : Waktu per hari pada tiap departemen. Ada 2 nilai dalam atribut ini, yaitu :
1. 08:00:00,00
 2. 09:00:00,00
- Lama bekerja : Waktu kerja karyawan selama di perusahaan. Nilai dalam atribut ini juga akan dikategorikan menjadi,
1. Junior
 2. Senior
- Status : Status karyawan saat ia menjadi bagian dari departemen, yaitu :
1. Regular
 2. Fuji
 3. Swabina
 4. ISS
 5. Kontraktor
- Pelatihan : Riwayat pelatihan yang dimiliki oleh karyawan, yaitu YES atau NO
- Kecelakaan Kerja : Kecelakaan kerja dibagi menjadi 3 jenis, yaitu :

1. *On work*
2. *Traffic on commute*
3. *Traffic inside fact*

Klasifikasi kecelakaan kerja diatas, berdasarkan data pada perusahaan. Definisi dari istilah tersebut yaitu :

- *On Work* : Kecelakaan kerja yang terjadi saat karyawan sedang dalam proses pekerjaan dan berada dalam area perusahaan
- *Traffic Inside Fact* : Kecelakaan yang terjadi selama di area perusahaan yang tidak termasuk *On Work*
- *Traffic On Commute* : Kecelakaan kerja yang terjadi saat karyawan berada di luar area perusahaan, contoh : waktu berangkat kerja, waktu pulang kerja, dinas luar.

3.3. Pengolahan Data

Pengolahan data yang dimaksud adalah mengolah data sekunder yang berupa tabel *excel* menjadi data yang telah menghasilkan pola sebagai tujuan dari Tugas Akhir ini. Untuk menjadikan data sekunder menjadi pola akhir, ada beberapa tahap yang harus dilalui. Tahap-tahap yang harus dilalui ini sesuai dengan KDD yang telah dijelaskan pada Bab 2. Dalam Tugas Akhir ini, tahap-tahap KDD yang harus dilalui antara lain :

3.3.1. *Pre-Processing*

Proses ini merupakan proses yang dilakukan untuk membuat data mentah menjadi *taining data* (data yang dapat digunakan). Untuk Tugas Akhir ini, *pre-processing* yang dilakukan ada 2 tahap, yaitu :

1. **Pembersihan Data (*Data Cleaning*)** merupakan proses menghapus data *noise* (pengganggu) dan mengisi data yang hilang. Contoh *noise* yang dapat dihilangkan yaitu *missing value data*. *Missing value data* dapat dihilangkan jika tidak ada perubahan atau efek penghilangannya.

Data yang dipakai dalam proses ini masih berupa format .excel. Proses ini menghilangkan data dengan atribut yang tidak lengkap.

2. **Transformasi Data** merupakan proses pengubahan data ke dalam format tertentu untuk dapat diproses ke dalam *data mining*. Data yang sudah dibersihkan dari *noise*, akan diubah formatnya menjadi format yang dipakai dalam *data mining* agar bisa diproses.

3.3.2. *Data Mining*

Proses ini merupakan proses ekstraksi pola-pola penting dalam basis data yang akan dianalisis. Hal ini dilakukan setelah *pre-processing* selesai. Dalam Tugas Akhir ini, metode data mining yang dilakukan adalah C4.5 dan C5.0. *Pre-processing* yang dilakukan pada kedua algoritma sama. Dalam tahap ini data dibagi menjadi 2 macam, yaitu data *testing* dan data *training*.

Terdapat beberapa tahap dalam proses ini, yaitu:

Perancangan Algoritma

Dalam tahap ini, penulis akan membangun sebuah *tree* dari kedua algoritma secara terpisah. Data yang digunakan adalah data set yang telah melalui *pre-procesing*. Keluaran dari tahap ini adalah pola kecelakaan kerja berupa *rules* yang nantinya akan membentuk kelas-kelas berlabel jenis kecelakaan kerja.

Pengujian Algoritma

Dalam tahap ini akan menguji keakuratan *rules* yang dihasilkan oleh tahap perancangan algoritma. Pengujian dilakukan dengan mencocokkan pola yang sudah terbentuk (*rules*) dengan data testing yang telah dipilih sebelumnya. Keluaran dari tahap ini adalah angka keakurasian pada tiap pengujian pola.

Analisis dan Hasil Pola

Tahap ini akan menganalisis pola yang telah terbentuk dan dilakukan beberapa kali pengujian sampai didapat hasil yang paling akurat.

Setelah melakukan tahap-tahap data mining diatas dan akan diketahui hasil polanya, sehingga tujuan pertama dari Tugas

Akhir ini tercapai. Jika proses ini berhasil, maka akan didapat 2 angka prosentasi yang nantinya dapat dianalisis untuk mencapai tujuan kedua.

3.3.3. Analisis Pola dan Representasi Data

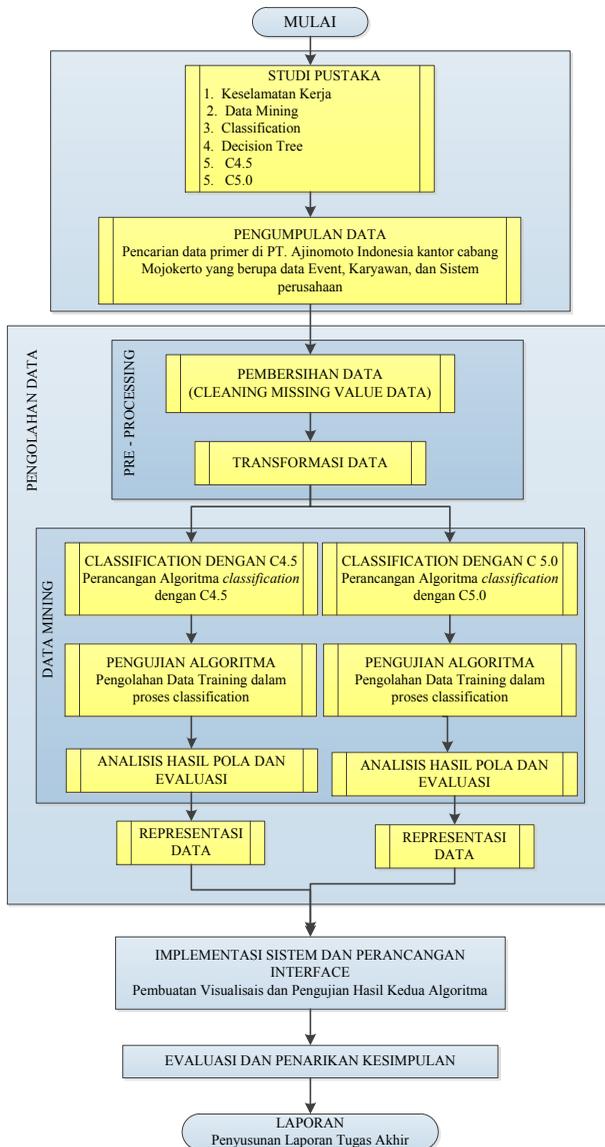
Tahap ini merupakan proses penarikan kesimpulan dari informasi atau pola yang sudah diperoleh dari proses *data mining*.

3.4. Perancangan Interface dan Implementasi Sistem

Pada tahap ini, penulis membuat visualisasi dari hasil *data mining* tersebut agar mudah dipahami oleh pembaca. Pengujian akan dilakukan beberapa kali sampai dipastikan tidak ada eror. Pengujian ini akan mempengaruhi pengambilan keputusan untuk membandingkan keakuratan kedua algoritma yang dipakai.

3.5. Evaluasi dan Kesimpulan

Setelah dilakukan analisis dan pembahasan maka dapat ditarik suatu kesimpulan dan saran sebagai masukan untuk pengembangan penelitian lebih lanjut.



Gambar 3.1. Diagram Alur Penelitian

BAB IV

PERANCANGAN DAN IMPLEMENTASI SISTEM

Pada bab ini dijelaskan analisis data, perancangan gambaran umum sistem, perancangan algoritma, perancangan data, dan perancangan antar muka sistem. Sistem ini memiliki *input* berupa data mentah yang berasal dari *database* yang selanjutnya dilakukan tahap *pre-processing* untuk menghilangkan *noise* pada data yang berupa duplikasi data, *missing value data* dan transformasi data. Hasil dari *pre-processing* ini berupa data kategorikal yang siap digunakan sebagai *input* untuk menjalankan algoritma. Dalam menjalankan algoritma, tidak bisa dilakukan secara bersamaan. Akan tetapi, kedua algoritma dapat dijalankan dengan satu kali *pre-processing*. Hasil dari proses ini adalah pola kecelakaan kerja berupa *rules* pada tiap algoritma, yang selanjutnya akan di uji sehingga memunculkan angka keakurasian pada tiap algoritma tersebut.

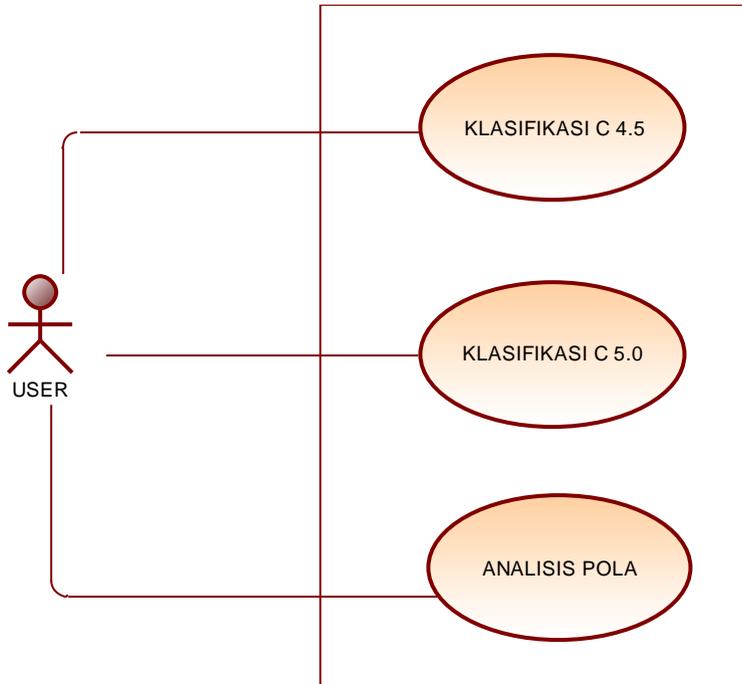
4.1 Analisis Sistem

Dalam sub bab ini akan dijelaskan proses analisis dalam membangun perangkat lunak, mulai dari *Use Case Diagram*, *Activity Diagram*, diagram alir sistem, diagram *pre-processing*, proses klasifikasi yang meliputi proses *training* dan *testing* pada Algoritma C4.5 dan C5.0. Pengujian hasil akan dilakukan dengan membagi data (*split data*) hingga mendapatkan keakuratan tertinggi.

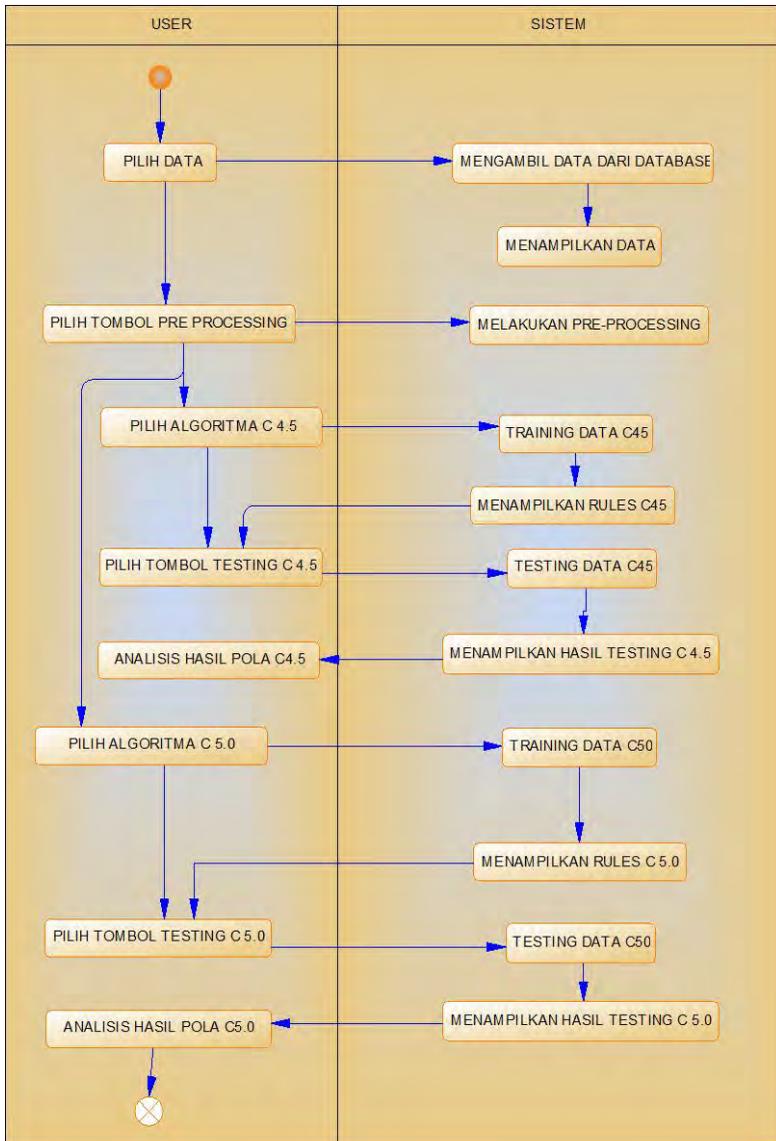
4.1.1 Analisis Sistem Perangkat Lunak

Perangkat lunak yang akan dibangun dapat membantu pihak-pihak dalam menganalisis data dalam keadaan kecil maupun besar. Klasifikasi ini dapat memprediksi pola dari data histori. Selain untuk memprediksi data kecelakaan kerja, sistem perangkat lunak juga bisa membantu untuk prediksi data lainnya seperti, penyakit, kredit, dan dapat juga sebagai pendukung keputusan untuk sistem pakar.

Secara umum *Use Case Diagram* dari sistem ini disajikan pada Gambar 4.1, *Swimlane Diagram* pada Gambar 4.2 menunjukkan jalannya sistem perangkat lunak.



Gambar 4.1. *Use Case Diagram* Perangkat Lunak Data Mining



Gambar 4.2. *Swimlane Diagram* Perangkat Lunak

Sistem perangkat lunak yang dibangun ini memiliki beberapa tahapan sebagai berikut :

a. *Pre-processing*

Data masukan yang berupa atribut-atribut tidak bisa langsung digunakan dalam menjalankan algoritma. Untuk itu *pre-processing* berguna untuk menyiapkan data agar menjadi data siap pakai. Tahap *pre-processing* sendiri bekerja untuk menghilangkan *noise* pada data. Dalam perangkat lunak ini, ada 3 tahapan *pre-processing* yaitu membersihkan duplikasi data, *missing value data*, dan transformasi data.

Duplikasi data berarti terdapat 2 data yang sama. Data yang seperti ini akan ditangani dengan menghilangkan salah satu data dengan pencocokan *primary key*. Setelah dijamin tidak ada data yang sama, tahap selanjutnya adalah menangani *missing value data*.

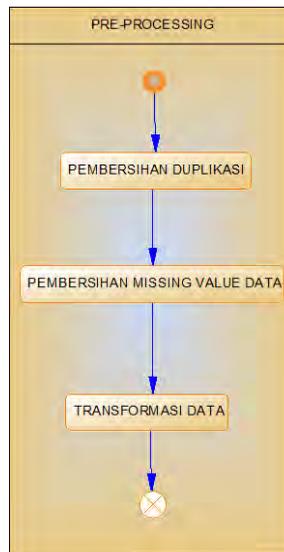
Data yang seperti ini berarti mempunyai nilai yang kosong pada salah satu atribut atau lebih. Untuk menangani *missing value data*, terdapat 2 langkah yaitu menghitung modus dan rata-rata. Untuk atribut yang bersifat numerik, menangani *missing value data* dengan menghitung rata-rata. Dari nilai yang ada, akan dihitung nilai rata-ratanya dan akan digunakan untuk mengisi data pada atribut yang kosong. Penanganan *missing value data* pada data yang bersifat kategorikal akan menggunakan modus. Dari setiap atribut akan dicari data yang paling sering muncul. Setelah ditemukan frekuensi kemunculan dari atribut tersebut, maka data itu akan dijadikan masukan pada atribut data yang kosong.

Transformasi data akan dilakukan setelah tahap penanganan *missing value data* dilakukan. Transformasi data dilakukan untuk memilih data yang dibutuhkan dalam pembentukan algoritma.

Transformasi dilakukan karena tidak semua atribut dalam data digunakan untuk membentuk algoritma. Atribut yang dipandang berpengaruh yang akan dipilih, dan atribut yang tidak berpengaruh akan dihilangkan dalam proses selanjutnya. Selain untuk mengurangi atribut yang tidak dibutuhkan, transformasi data juga digunakan untuk mengubah format data. Format data

yang rapi dan terstruktur akan memudahkan untuk pembangunan algoritma kedepannya.

Perubahan format data ini biasanya terjadi pada data numerik yang mempunyai variasi yang banyak. Dari variasi ini akan dibuat range dengan label yang berbeda. Label ini akan bersifat kategorikal. Data yang bersifat kategorikal adalah data dengan format yang lebih jelas dan terstruktur nilainya, sehingga mempermudah pengolahan data dalam tahap *training*. Data hasil transformasi inilah yang dapat digunakan untuk input pada proses algoritma selanjutnya.

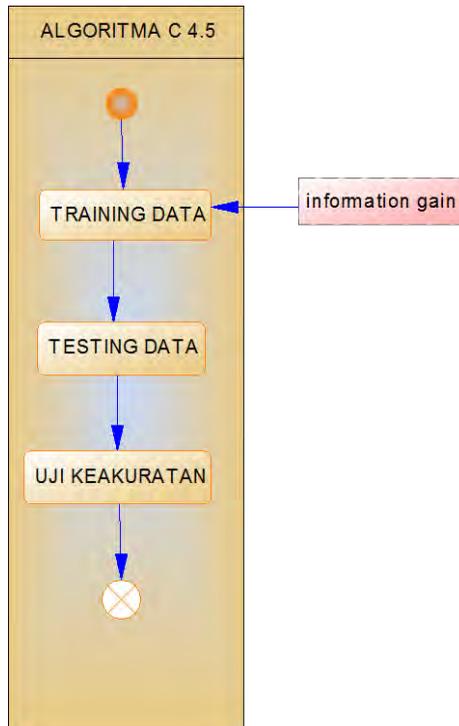


Gambar 4.3. Swimlane Diagram pada *Pre-Processing*

b. Algoritma C4.5

Tahap algoritma ini dilakukan setelah ada data hasil transformasi. Pembentukan *tree* pada algoritma ini berdasarkan *information gain* pada tiap iterasinya. Perhitungan *information gain* melihat kemunculan tiap *prediktor* pada tiap atribut terhadap semua kemunculan kejadian kecelakaan kerja. Iterasi akan terus berjalan sampai tidak ada data yang dapat dipecah.

Pembangunan *tree* ini merupakan tahap *training* data. Output dari tahap *training* adalah pola kecelakaan kerja berupa *rules* yang akan digunakan untuk mengklasifikasikan data pada tahap *testing*.

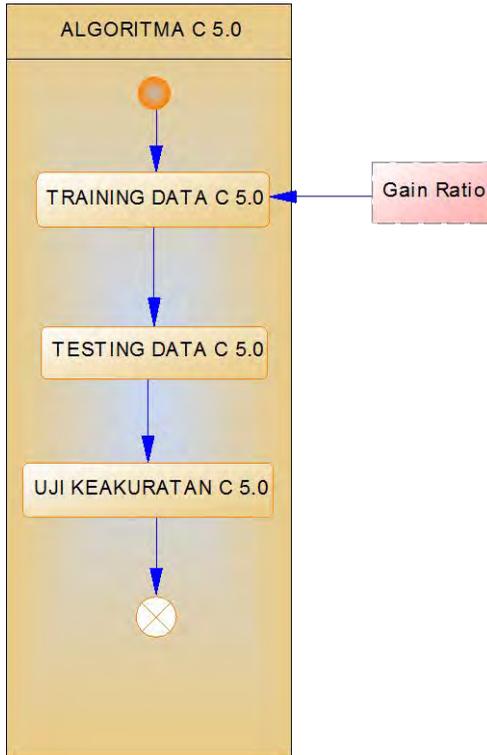


Gambar 4.4. Swimlane Diagram pada Algoritma C4.5

c. Algoritma C5.0

Pembangunan *tree* pada Algoritma C5.0 merupakan proses yang sama pada Algoritma C4.5. Saat Algoritma C4.5 cukup menggunakan *information gain* untuk pembangunan *tree*, C5.0 masih membutuhkan *gain ratio* untuk pembangunannya. *Gain ratio* dihitung berdasarkan *information gain* dan *entropy* yang telah dihitung sebelumnya. Tahap perhitungan *entropy* dan

information gain menggunakan rumus dan langkah yang sama dengan Algoritma C4.5.



Gambar 4.5. Swimlane Diagram pada Algoritma C5.0

4.1.2 Analisis Kebutuhan Sistem

Perangkat lunak ini dibangun menggunakan software NetBeans IDE 8.1 dan DBMS mySQL baik dari desain *interface*, sistem *toolbox computer vision* dan pengolahan data.

Tabel 4.1. Tabel Kebutuhan Sistem

| | |
|-----------------|---|
| Perangkat Keras | Prosesor : Intel® Core™ i3 – 5010U CPU @2.10GHz 2.10 GHz |
| | RAM Memory : 2.00 GB |
| Perangkat Lunak | Sistem Operasi : Windows 8 64 - bit |
| | Tools : NetBeans IDE 8.1 |

4.2 Perancangan Sistem

Setelah analisis sistem, kemudian dilanjutkan dengan perancangan sistem. Perancangan sistem tersebut meliputi perancangan data sistem, perancangan *class* sistem, perancangan proses algoritma sistem, dan perancangan antarmuka sistem.

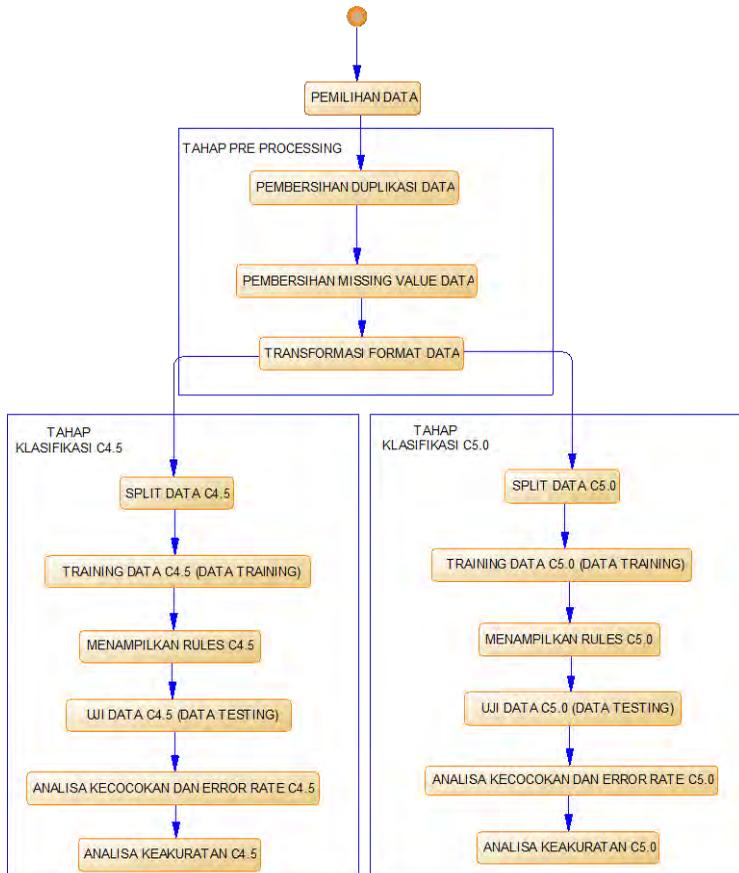
4.2.1 Penjelasan Umum Sistem

Proses *data mining* berjalan saat pemilihan data dilakukan. Data kecelakaan kerja yang berada pada database akan dipanggil sebagai masukan utama. Setelah pemanggilan sukses, kemudian data akan di *pre-processing* untuk menghilangkan *noise* dan mengubah format pada atribut. Setelah data yang siap diolah terbentuk, tahap selanjutnya adalah pembangunan algoritma. Tahap *data mining* pertama adalah membagi dataset menjadi 2 bagian (*split data*), yaitu data *training* dan data *testing*. Selanjutnya, pembangunan Algoritma C4.5 dilakukan terlebih dahulu menggunakan *information gain* sebagai tahap *training*. Hasil dari tahap *training* adalah pola kecelakaan (*rules*) yang merupakan representasi dari lintasan (*path*) dalam *tree*. Tahap *testing* akan dilakukan untuk mengetahui keakuratan pola yang terbentuk pada algoritma ini.

Karena pembentukan *tree* pada Algoritma C4.5 hampir sama dengan Algoritma C5.0, maka C5.0 menggunakan *information gain* dan *entropy* pada Algoritma C4.5 untuk menghitung *gain ratio*. Hasil dari *gain ratio* digunakan untuk membangun *tree*. Hasil *tree* pada Algoritma C5.0 berupa pola kecelakaan kerja (*rules*), yang nantinya akan digunakan sebagai input pada proses *testing*. *Testing* pada algoritma ini sama dengan Algoritma C4.5, *output* yang akan dihasilkan berupa angka keakuratan algoritma.

Setelah melakukan *testing* pada kedua algoritma, *user* dapat membandingkan dan menganalisis keakuratan pola kecelakaan kerja pada kedua algoritma. Selanjutnya, *user* dapat menarik kesimpulan terhadap hasil Algoritma C4.5 dan C5.0.

Berikut ini akan ditampilkan Diagram Alir Sistem pada perangkat lunak *data mining*.



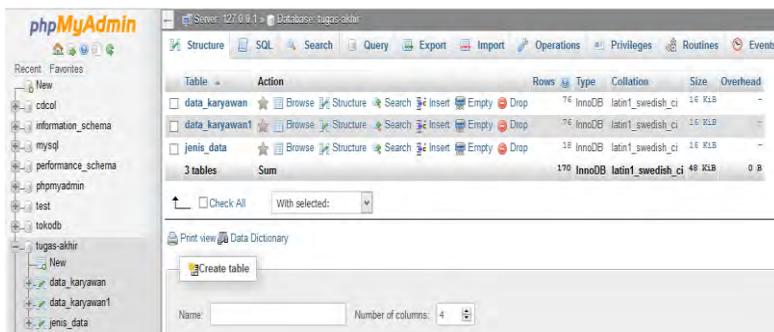
Gambar 4.6. Diagram Alir Sistem

4.2.2 Perancangan Algoritma

Perancangan algoritma ini secara garis besar di bagi dalam dua tahapan, yaitu tahap *pre-processing* dan *data mining*. Tahapan ini merupakan alur karena proses *data mining* harus dilakukan setelah *pre-processing* sesuai KDD. *Pre-processing* dilakukan untuk mempermudah pengolahan data karena dalam tahap ini data akan disempurnakan dan disiapkan untuk proses *data mining* selanjutnya.

4.2.2.1 Perancangan Database

DBMS yang digunakan dalam pembuatan perangkat lunak ini adalah *MySql*. *DBMS* akan menampung semua data *input* yang berupa data mentah. Pada saat *pre-processing*, data yang diambil adalah data pada Tabel “*data_karyawan*”. *Output* dari *pre-processing* juga kembali disimpan pada *database*, namun dengan tabel berbeda.



Gambar 4.7. Struktur Database

| # | Name | Type | Collation | Attributes | Null | Default | Extra | Action |
|----|------------------|-------------|-------------------|------------|------|---------|-------|--|
| 1 | id | varchar(15) | latin1_swedish_ci | | No | | | Change Drop Primary Unique Index Spatial Fulltext More |
| 2 | Nama | varchar(20) | latin1_swedish_ci | | No | None | | Change Drop Primary Unique Index Spatial Fulltext More |
| 3 | jenis_kelamin | char(1) | latin1_swedish_ci | | No | None | | Change Drop Primary Unique Index Spatial Fulltext More |
| 4 | Umur | int(2) | | | No | None | | Change Drop Primary Unique Index Spatial Fulltext More |
| 5 | kewarganegaraan | varchar(10) | latin1_swedish_ci | | Yes | NULL | | Change Drop Primary Unique Index Spatial Fulltext More |
| 6 | tipe_pekerjaan | varchar(5) | latin1_swedish_ci | | No | None | | Change Drop Primary Unique Index Spatial Fulltext More |
| 7 | waktu_per_hari | time(2) | | | Yes | NULL | | Change Drop Primary Unique Index Spatial Fulltext More |
| 8 | jama_bekerja | int(4) | | | Yes | NULL | | Change Drop Primary Unique Index Spatial Fulltext More |
| 9 | status | text | latin1_swedish_ci | | No | None | | Change Drop Primary Unique Index Spatial Fulltext More |
| 10 | pelatihan | char(5) | latin1_swedish_ci | | Yes | NULL | | Change Drop Primary Unique Index Spatial Fulltext More |
| 11 | kecelakaan_serja | varchar(20) | latin1_swedish_ci | | No | None | | Change Drop Primary Unique Index Spatial Fulltext More |

Gambar 4.8. Struktur Tabel Data Karyawan

4.2.2.2 Perancangan Tahap *Pre-processing*

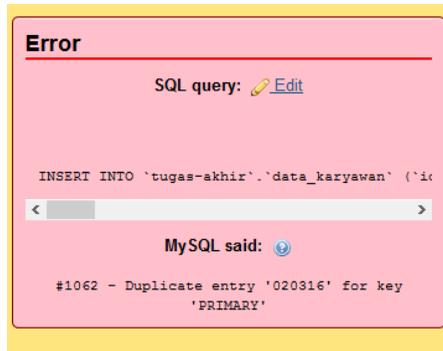
Awal dari proses ini adalah pemanggilan data dari *database*. Data yang ditampilkan merupakan data real yang kemungkinan masih memiliki *noise*. Setelah data ditampilkan, maka *pre-processing* dapat dijalankan dengan 3 tahap, yaitu menangani duplikasi data, *missing value data*, dan transformasi data.

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|-------------|----|---|----|-----------|-------|---------|----|------------|-----|---------|-------------|
| 1 | 2031613145 | MM | L | 40 | INDONESIA | GRD C | 00:00:0 | 20 | ISS | NO | TRAFFIC | INSIDE FACT |
| 2 | 2051412261 | R | L | 24 | INDONESIA | GRD B | 00:00:0 | 2 | SWABINA | YES | TRAFFIC | ON COMMUTE |
| 3 | 2101412431 | Y | P | 25 | INDONESIA | GRD B | 00:00:0 | 3 | SWABINA | NO | TRAFFIC | INSIDE FACT |
| 4 | 3011311211 | H | L | 50 | INDONESIA | GRD A | 00:00:0 | 26 | KONTRAKTOR | YES | TRAFFIC | INSIDE FACT |
| 5 | 3021612401 | LL | L | 24 | INDONESIA | GRD B | 00:00:0 | 2 | FUJI | YES | TRAFFIC | ON COMMUTE |
| 6 | 4051312049 | KL | L | 41 | INDONESIA | GRD B | 00:00:0 | 19 | REGULER | YES | TRAFFIC | INSIDE FACT |
| 7 | 4081113170 | DG | L | 27 | INDONESIA | GRD C | 00:00:0 | 3 | SWABINA | YES | TRAFFIC | INSIDE FACT |
| 8 | 5011413803 | GD | L | 31 | INDONESIA | GRD C | 00:00:0 | 7 | SWABINA | NO | ON WORK | |
| 9 | 7031311139 | J | L | 39 | INDONESIA | GRD A | 00:00:0 | 17 | SWABINA | NO | ON WORK | |
| 10 | 8021512313 | DD | L | 24 | INDONESIA | GRD B | 00:00:0 | 2 | FUJI | NO | TRAFFIC | ON COMMUTE |
| 11 | 8031311212 | I | L | 23 | INDONESIA | GRD A | 00:00:0 | 3 | FUJI | NO | ON WORK | |
| 12 | 8041413320 | GT | L | 29 | INDONESIA | GRD C | 00:00:0 | 5 | KONTAKTOR | NO | TRAFFIC | INSIDE FACT |
| 13 | 8061511091 | HH | L | 23 | INDONESIA | GRD A | 00:00:0 | 3 | FUJI | NO | TRAFFIC | ON COMMUTE |
| 14 | 8081213079 | E | L | 35 | INDONESIA | GRD C | 00:00:0 | 13 | REGULER | NO | ON WORK | |
| 15 | 9031412097 | IY | L | 49 | INDONESIA | GRD B | 00:00:0 | 29 | REGULER | YES | TRAFFIC | INSIDE FACT |
| 16 | 9041412930 | EB | L | 34 | INDONESIA | GRD B | 00:00:0 | 12 | FUJI | YES | TRAFFIC | ON COMMUTE |
| 17 | 9061512054 | II | L | 31 | INDONESIA | GRD B | 00:00:0 | 10 | REGULER | NO | ON WORK | |
| 18 | 9121512205 | KK | L | 30 | INDONESIA | GRD B | 00:00:0 | 9 | REGULER | YES | TRAFFIC | ON COMMUTE |
| 19 | 1004131541 | IK | L | 31 | INDONESIA | GRD C | 00:00:0 | 10 | REGULER | YES | TRAFFIC | ON COMMUTE |
| 20 | 1008151139 | JJ | L | 24 | INDONESIA | GRD A | 00:00:0 | 2 | REGULER | NO | TRAFFIC | ON COMMUTE |
| 21 | 10101412210 | Z | L | 24 | INDONESIA | GRD B | 00:00:0 | 4 | FUJI | NO | TRAFFIC | ON COMMUTE |
| 22 | 11091311012 | M | L | 30 | INDONESIA | GRD A | 00:00:0 | 7 | SWABINA | YES | TRAFFIC | ON COMMUTE |

Gambar 4.9. Data Mentah yang Dikenai *Pre-Processing*

a. Pembersihan Duplikasi Data

Tahap ini ditangani dengan menggunakan *primary key* yang ada pada data. *Primary key* pada suatu tabel tidak diijinkan ada yang sama. Oleh karena itu, saat *input record* baru yang mempunyai *primary key* yang sama, maka record tersebut akan ditolak. Untuk data karyawan kecelakaan kerja ini, yang menjadi *primary key* adalah atribut “id_kejadian”. Contoh untuk kejadian duplikasi data tersebut seperti Gambar 4.10 dibawah ini.



Gambar 4.10. Penolakan Masukan Data yang Mengandung Duplikasi

b. Penanganan *Missing Value Data*

Tahap ini menggunakan 2 cara, yaitu dengan menghitung modus dan rata-rata pada atribut. Pencarian modus akan dikenakan pada atribut „tipe_pekerjaan“. *Prediktor* dari atribut ini berupa *string*, yaitu „GRD A“, „GRD B“, dan „GRD C“. Dari semua *record* yang ada, akan dicari kemungkinan muncul terbesar. *Prediktor* terbesar akan mengisi *record* yang kosong. Pencarian modus ini berlaku untuk atribut kategorikal.

Atribut „tipe_pekerjaan“ mempunyai relasi dengan atribut „waktu_per_hari“. Jika ada *missing value data* pada atribut „waktu_per_hari“, maka pengisian akan bergantung ada *prediktor* di kolom „tipe_pekerjaan“. Hubungan prediktor kedua atribut terlihat pada Tabel 4.2 dibawah ini

Tabel 4.2. Hubungan Atribut „tipe_pekerjaan“ dan „waktu_per_hari“

| „tipe_pekerjaan“ | „waktu_per_hari“ |
|------------------|------------------|
| GRD A | 08:00:00.00 |
| GRD B | 09:00:00.00 |
| GRD C | 09:00:00.00 |

Atribut „jama_bekerja“ juga bergantung pada atribut „umur“. Ketentuan hubungan kedua atribut ini terlihat pada Tabel 4.3.

Tabel 4.3. Hubungan Atribut „jama_bekerja“ dan „umur“

| Klasifikasi Umur | Lama Bekerja |
|------------------|--------------|
| umur1 | Junior |
| umur2 | Junior |
| umur3 | Senior |
| umur4 | Senior |

c. Transformasi Data

Transformasi data pada perangkat lunak ini akan mereduksi beberapa atribut dan mengubah format pada beberapa atribut. Atribut yang direduksi dianggap tidak memiliki pengaruh pada potensi terjadinya kecelakaan kerja. Perubahan jumlah atribut dapat dilihat pada Tabel 4.4.

Tabel 4.4. Perbedaan Atribut saat Transformasi

| Sebelum Transformasi | Sesudah Transformasi |
|----------------------|----------------------|
| 1. Id | 1. Id |
| 2. Nama | 2. Jenis Kelamin |
| 3. Jenis kelamin | 3. Umur |
| 4. Umur | 4. Tipe pekerjaan |
| 5. Kewarganegaraan | 5. Lama bekerja |
| 6. Tipe pekerjaan | 6. Status |
| 7. Waktu per hari | 7. Pelatihan |
| 8. Lama bekerja | 8. Kecelakaan Kerja |
| 9. Status | |
| 10. Pelatihan | |
| 11. Kecelakaan Kerja | |

Atribut yang tereduksi dianggap tidak mempengaruhi pembangun *tree*. Atribut “nama” direduksi karena hanya

menampilkan nama karyawan yang mengalami kecelakaan kerja, sehingga tidak berpengaruh pada pembangun *tree*. Atribut “Kewarganegaraan” hanya memiliki 1 nilai pada *dataset* “data_karyawan”, yaitu Indonesia. Karena nilai pada atribut ini homogen, maka atribut ini tidak berpengaruh sehingga dapat dihilangkan. Atribut “waktu_per_hari” telah bergantung pada “tipe_pekerjaan”. Sehingga nilai pada atribut “waktu_per_hari ” dapat diwakilkan pada atribut “tipe_pekerjaan”.

Pada tahap transformasi data ini atribut “umur” akan diubah menjadi data kategorikal. Perubahan ini bergantung pada aturan perusahaan yang mengklasifikasikan umur karyawan berdasarkan *range* per 10 tahun. Klasifikasi tersebut dapat dilihat pada Tabel 4.5.

Tabel 4.5. Perubahan Nilai Pada Atribut “umur”

| Klasifikasi Umur | Umur (tahun) |
|------------------|--------------|
| Umur1 | <25 |
| Umur2 | 25 – 35 |
| Umur2 | 35 – 45 |
| Umur4 | 45 - 55 |

Transformasi juga berguna untuk mengelompokkan *record* yang mempunyai nilai yang bervariasi. Perubahan jenis *prediktor* ini berlaku untuk atribut “umur” dan “lama_bekerja”. Aturan dari perubahan jenis atribut ini berdasarkan rata-rata dari *record* yang ada.

4.2.2.3 Perancangan Proses Data Mining

Proses *data mining* adalah pembentukan *tree* sebagai awal dari Algoritma C4.5 dan C5.0. Sebelum membangun *tree* dengan kedua algoritma, *dataset* akan dibagi menjadi 2 bagian, yaitu data *training* dan data *testing*. Dalam pembentukan *tree*, data yang

digunakan adalah data *training*. Sedangkan data *testing* digunakan untuk menghitung keakuratan pola kecelakaan kerja yang dihasilkan. Langkah awal dari proses pembangunan *tree* adalah menghitung kemunculan kejadian untuk setiap *prediktor* tiap atribut. Dari tabel hasil transformasi, akan dihitung atribut untuk perhitungan *node* 1 (*root*) . Untuk pembuatan tabel dalam perhitungan *root*, dapat dilihat pada Tabel 4.6 dibawah ini.

Tabel 4.6. Perhitungan Jumlah Kemunculan Kejadian

| <i>Node</i> | | | Jml Kasus | <i>On Work</i> | <i>Traffic on Commute</i> | <i>Traffic Inside Fact</i> |
|-------------|----------------|---------|-----------|----------------|---------------------------|----------------------------|
| 1 | Total | | 200 | 35 | 113 | 52 |
| | Jenis Kelamin | L | 195 | 35 | 112 | 48 |
| | | P | 5 | 0 | 1 | 4 |
| | Umur | Umur1 | 42 | 6 | 31 | 5 |
| | | Umur2 | 69 | 19 | 36 | 14 |
| | | Umur3 | 61 | 6 | 33 | 22 |
| | | Umur4 | 28 | 4 | 13 | 11 |
| | Tipe Pekerjaan | GRD A | 88 | 14 | 61 | 13 |
| | | GRD B | 66 | 8 | 40 | 18 |
| | | GRD C | 46 | 13 | 12 | 21 |
| | Lama Bekerja | Junior | 111 | 27 | 65 | 19 |
| | | Senior | 89 | 8 | 48 | 33 |
| | Status | ISS | 7 | 0 | 1 | 6 |
| | | Reguler | 65 | 15 | 37 | 13 |

| | | | | | | |
|--|-----------|------------|-----|----|----|----|
| | | Fuji | 60 | 7 | 47 | 6 |
| | | Swabina | 52 | 11 | 26 | 15 |
| | | Kontraktor | 16 | 2 | 2 | 12 |
| | Pelatihan | Yes | 101 | 5 | 64 | 32 |
| | | No | 99 | 30 | 49 | 20 |

Setelah diketahui kemunculan setiap *prediktor* seperti yang terlihat pada tabel diatas, kemudian dicari nilai *entropy*. Perhitungan *entropy* pada Algoritma C4.5 sama dengan algoritma C5.0, dengan Persamaan (1).

Persamaan diatas berlaku pada semua atribut, termasuk atribut target „kecelakaan kerja“. *Entropy* pada atribut „kecelakaan kerja“ akan menjadi *entropy* total. Berikut perhitungan *entropy* dalam pemilihan *root*.

$$\begin{aligned} Entropy (Total) &= - \left[\frac{35}{200} \log_2 \frac{35}{200} + \frac{113}{200} \log_2 \frac{113}{200} + \frac{52}{200} \log_2 \frac{52}{200} \right] \\ &= 1.4107 \end{aligned}$$

$$\begin{aligned} Entropy (L) &= - \left[\frac{35}{195} \log_2 \frac{35}{195} + \frac{112}{195} \log_2 \frac{112}{195} + \frac{48}{195} \log_2 \frac{48}{195} \right] \\ &= 1.4021 \end{aligned}$$

$$Entropy (P) = - \left[\frac{0}{5} \log_2 \frac{0}{5} + \frac{1}{5} \log_2 \frac{1}{5} + \frac{4}{5} \log_2 \frac{4}{5} \right] = NaN$$

$$\begin{aligned} Entropy (umur1) &= - \left[\frac{6}{42} \log_2 \frac{6}{42} + \frac{31}{42} \log_2 \frac{31}{42} + \frac{5}{42} \log_2 \frac{5}{42} \right] \\ &= 1.0899 \end{aligned}$$

$$Entropy (umur2) = - \left[\frac{19}{69} \log_2 \frac{19}{69} + \frac{36}{69} \log_2 \frac{36}{69} + \frac{14}{69} \log_2 \frac{14}{69} \right]$$

$$= 1.4689$$

$$\begin{aligned} \text{Entropy (umur3)} &= - \left[\frac{6}{61} \log_2 \frac{6}{61} + \frac{33}{61} \log_2 \frac{33}{61} + \frac{22}{61} \log_2 \frac{22}{61} \right] \\ &= 1.3392 \end{aligned}$$

$$\text{Entropy (umur4)} = - \left[\frac{4}{28} \log_2 \frac{4}{28} + \frac{13}{28} \log_2 \frac{13}{28} + \frac{11}{28} \log_2 \frac{11}{28} \right] = 1.4445$$

$$\text{Entropy (GRD A)} = - \left[\frac{14}{88} \log_2 \frac{14}{88} + \frac{61}{88} \log_2 \frac{61}{88} + \frac{11}{88} \log_2 \frac{11}{88} \right] = 1.1960$$

$$\text{Entropy (GRD B)} = - \left[\frac{8}{66} \log_2 \frac{8}{66} + \frac{40}{66} \log_2 \frac{40}{66} + \frac{18}{66} \log_2 \frac{18}{66} \right] = 1.3181$$

$$\text{Entropy (GRD C)} = - \left[\frac{13}{46} \log_2 \frac{13}{46} + \frac{12}{46} \log_2 \frac{12}{46} + \frac{21}{46} \log_2 \frac{21}{46} \right] = 1.5374$$

$$\begin{aligned} \text{Entropy (Junior)} &= - \left[\frac{27}{111} \log_2 \frac{27}{111} + \frac{65}{111} \log_2 \frac{65}{111} + \frac{19}{111} \log_2 \frac{19}{111} \right] \\ &= 1.3841 \end{aligned}$$

$$\text{Entropy (Senior)} = - \left[\frac{8}{89} \log_2 \frac{8}{89} + \frac{48}{89} \log_2 \frac{48}{89} + \frac{33}{89} \log_2 \frac{33}{89} \right] = 1.3236$$

$$\text{Entropy (ISS)} = - \left[\frac{0}{7} \log_2 \frac{0}{7} + \frac{1}{7} \log_2 \frac{1}{7} + \frac{6}{7} \log_2 \frac{6}{7} \right] = \text{NaN}$$

$$\begin{aligned} \text{Entropy (Reguler)} &= - \left[\frac{15}{65} \log_2 \frac{15}{65} + \frac{37}{65} \log_2 \frac{37}{65} + \frac{13}{65} \log_2 \frac{13}{65} \right] \\ &= 1.4153 \end{aligned}$$

$$\text{Entropy (Fuji)} = - \left[\frac{7}{60} \log_2 \frac{7}{60} + \frac{47}{60} \log_2 \frac{47}{60} + \frac{6}{60} \log_2 \frac{6}{60} \right] = 0.9698$$

$$\begin{aligned} \text{Entropy (Swabina)} &= - \left[\frac{11}{52} \log_2 \frac{11}{52} + \frac{26}{52} \log_2 \frac{26}{52} + \frac{15}{52} \log_2 \frac{15}{52} \right] \\ &= 1.4914 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Kontraktor)} &= - \left[\frac{2}{16} \log_2 \frac{2}{16} + \frac{2}{16} \log_2 \frac{2}{16} + \frac{12}{16} \log_2 \frac{12}{16} \right] \\ &= 1.0613 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Yes)} &= - \left[\frac{5}{101} \log_2 \frac{5}{101} + \frac{64}{101} \log_2 \frac{64}{101} + \frac{32}{101} \log_2 \frac{32}{101} \right] \\ &= 1.1571 \end{aligned}$$

$$\text{Entropy (No)} = - \left[\frac{30}{99} \log_2 \frac{30}{99} + \frac{49}{99} \log_2 \frac{49}{99} + \frac{20}{99} \log_2 \frac{20}{99} \right] = 1.5508$$

Setelah perhitungan *entropy* seperti diatas, maka dilakukan perhitungan *information gain*. Perhitungan *information gain* pada kedua algoritma juga sama. Perhitungan *information gain* ini menggunakan Persamaan (2). Berikut perhitungan *Information Gain* dalam penentuan *root*. Nilai pada *entropy* (S) yang dipakai adalah *Entropy Total*.

$$\begin{aligned} \text{InformationGain (JK, Total)} &= 1,4107 - \left[\frac{195}{200} * 1.4021 + \frac{5}{200} \text{NaN} \right] \\ &= 0.0437 \end{aligned}$$

$$\text{InformationGain (Umur, Total)} = 1,4107 -$$

$$\left[\frac{42}{200} * 1.0899 + \frac{69}{200} * 1.4689 + \frac{61}{200} * 1.3392 + \frac{28}{200} * 1.4445 \right] = 0.0644$$

$$\text{InformationGain (Tipe Pekerjaan, Total)} = 1,4107 -$$

$$\left[\frac{88}{200} * 1.1960 + \frac{66}{200} * 1.3181 + \frac{46}{200} * 1.5374 \right] = 0.0959$$

$$\text{InformationGain (Lama Bekerja, Total)} = 1,4107 -$$

$$\left[\frac{111}{200} * 1.3841 + \frac{89}{200} * 1.3236 \right] = 0.0535$$

$$\begin{aligned} \text{InformationGain}(\text{Status}, \text{Total}) &= 1,4107 - \\ &\left[\frac{7}{200} * \text{NaN} + \frac{65}{200} * 1.4135 + \frac{60}{200} * 0.9698 + \frac{52}{200} * 1.4914 \right. \\ &\quad \left. + \frac{16}{200} * 1.0613 \right] = 0.1871 \end{aligned}$$

$$\text{InformationGain}(\text{Pelatihan}, \text{Total}) = 1,4107 -$$

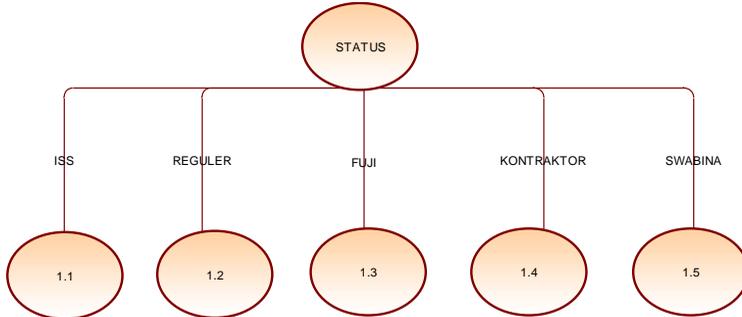
$$\left[\frac{101}{200} * 1.1571 + \frac{99}{200} * 1.5508 \right] = -0.1503$$

Perhitungan *information gain* seperti diatas yang digunakan untuk membuat *node* 1 (*root*) pada Algoritma C4.5. Sehingga Tabel 4.6 diatas berubah menjadi Tabel 4.7 dibawah ini.

Tabel 4.7. Perhitungan *Entropy* dan *Information Gain*

| Node | | | Entropy | Information Gain |
|------------|----------------|---------|---------|------------------|
| 1 | Total | | 1.4107 | |
| | Jenis Kelamin | L | 1.4021 | 0.0437 |
| | | P | NaN | |
| | Umur | Umur1 | 1.0899 | 0.0644 |
| | | Umur2 | 1.4689 | |
| | | Umur3 | 1.3392 | |
| | | Umur4 | 1.4445 | |
| | Tipe Pekerjaan | GRD A | 1.1960 | 0.0959 |
| | | GRD B | 1.3181 | |
| | | GRD C | 1.5374 | |
| | Lama Bekerja | Junior | 1.3841 | 0.0535 |
| | | Senior | 1.3236 | |
| | Status | ISS | NaN | 0.1871 |
| | | Reguler | 1.4153 | |
| | | Fuji | 0.9698 | |
| Swabina | | 1.4914 | | |
| Kontraktor | | 1.0613 | | |
| Pelatihan | Yes | 1.1571 | -0.1503 | |
| | No | 1.5508 | | |

Node 1 akan mengambil atribut dengan nilai *information gain* tertinggi. Dari perhitungan diatas, dapat dilihat bahwa atribut “Status” memiliki nilai *Information Gain* tertinggi. Sehingga, dalam perhitungan tersebut, atribut “Status” akan menjadi *root* dalam pembentukan *tree*.



Gambar 4.11. *Root* yang dihasilkan dengan Algoritma C4.5

Perhitungan *information gain* ini akan dilakukan sampai tidak ada atribut yang dapat dipecah menjadi lebih kecil lagi. Penentuan keputusan tiap cabang akan dilihat pada nilai *entropy*. Saat nilai *entropy* pada cabang bernilai 0, maka cabang tersebut akan menemui keputusan. Pemilihan keputusan dapat dilihat pada kemunculan *prediktor* terbanyak terhadap atribut target.

Pada Algoritma C5.0, perhitungan *node* akan dilakukan berdasarkan perhitungan *gain ratio*. Untuk perhitungan ini, dapat menggunakan Persamaan (3). Berikut perhitungan *Gain Ratio* dalam penentuan *root*.

$$\text{GainRatio (Jenis Kelamin)} = \frac{0.0437}{1.4021} = 0.0313$$

$$\text{GainRatio (Umur)} = \frac{0.0644}{1.0899 + 1.4689 + 1.3392 + 1.4445} = 0.0121$$

$$\text{GainRatio (Tipe Pekerjaan)} = \frac{0.0959}{1.1960 + 1.3181 + 1.5374} = 0.0237$$

$$\text{GainRatio (Lama Bekerja)} = \frac{0.0535}{1.3841 + 1.3236} = 0.0198$$

$$\begin{aligned} \text{GainRatio (Status)} &= \frac{0.1871}{0.5917 + 1.4153 + 0.9698 + 1.4914 + 1.0612} \\ &= 0.0338 \end{aligned}$$

$$\text{GainRatio (Pelatihan)} = \frac{-0.1503}{1.1571 + 1.5508} = -0.0555$$

Dapat dilihat bahwa disinilah perbedaan Algoritma C4.5 dan C5.0 terlihat. Tabel perhitungan *gain ratio* Algoritma C5.0 terlihat pada Tabel 4.8 dibawah ini.

Tabel 4.8. Perhitungan *Entropy*, *Information Gain*, dan *Gain Ratio*

| <i>Node</i> | | | <i>Entropy</i> | <i>Information Gain</i> | <i>Gain Ratio</i> |
|-------------|----------------|--------|----------------|-------------------------|-------------------|
| 1 | Total | | 1.4107 | | |
| | Jenis Kelamin | L | 1.4021 | 0.0437 | 0.0313 |
| | | P | NaN | | |
| | Umur | Umur1 | 1.0899 | 0.0644 | 0.0121 |
| | | Umur2 | 1.4689 | | |
| | | Umur3 | 1.3392 | | |
| | | Umur4 | 1.4445 | | |
| | Tipe Pekerjaan | GRD A | 1.1960 | 0.0959 | 0.0237 |
| | | GRD B | 1.3181 | | |
| | | GRD C | 1.5374 | | |
| | Lama Bekerja | Junior | 1.3841 | 0.0535 | 0.0198 |
| Senior | | 1.3236 | | | |
| Status | ISS | NaN | 0.1871 | 0.0338 | |

| | | | | | |
|--|-----------|------------|--------|---------|---------|
| | | Reguler | 1.4153 | | |
| | | Fuji | 0.9698 | | |
| | | Swabina | 1.4914 | | |
| | | Kontraktor | 1.0613 | | |
| | Pelatihan | Yes | 1.1571 | -0.1503 | -0.0198 |
| | | No | 1.5508 | | |

Pada Tabel 4.8 terlihat bahwa atribut “Status” memiliki nilai *gain ratio* tertinggi, sehingga *root* pada pembangun *tree* Algoritma C5.0 yaitu “Status”. Sehingga *root* yang dihasilkan Algoritma C4.5 sama dengan C5.0.

Perhitungan *gain ratio* juga akan berhenti saat tidak ada lagi atribut yang dapat dipecah menjadi lebih kecil. Pengambilan *node* 1 (*root*) akan memilih atribut dengan nilai *gain ratio* tertinggi.

Tabel 4.7. dan 4.8 dicetak setiap menambahkan cabang baru. Tahap inilah yang dinamakan *fase training* dengan *decision tree*. Setelah *tree* terbentuk, maka *rules* dapat direpresentasikan dengan bentuk *list*. *Rules* yang dihasilkan merupakan lintasan (*path*) *tree* yang diambil pada setiap cabang terakhir. Lintasan dalam *tree* merupakan hubungan antar atribut. Jadi pola kecelakaan kerja merupakan relasi-relasi atribut dalam *tree* yang akan menemui keputusan pada *leaf node*.

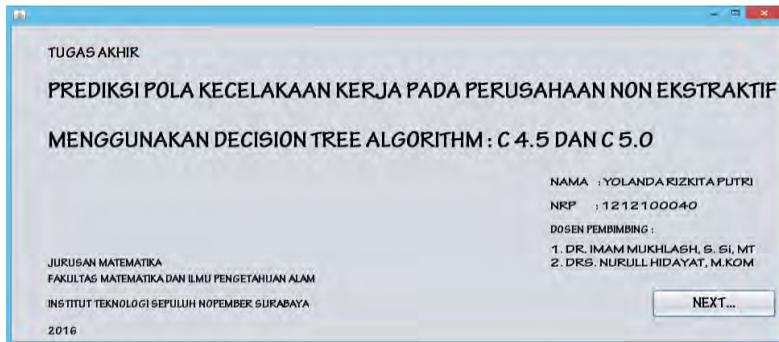
Rules yang dihasilkan dalam *fase training*, akan diuji pada tahap klasifikasi (tahap uji). Uji keakuratan dihitung dengan mencocokkan *rules* dengan data *testing*. Perhitungan yang dilakukan adalah dengan mencocokkan data *testing* dengan pola kecelakaan kerja yang telah terbentuk. Pencocokan ini berlangsung selama kondisi “IF.. THEN”. Jika kondisi “IF.. THEN.” pada data *testing* dan pola sama, maka akan menambah 1 poin kecocokan. Poin-poin kecocokan tersebut kemudian akan dihitung untuk mendapatkan nilai akurasi. Perhitungan tersebut dapat dilakukan dengan Persamaan (4).

4.2.3 Perancangan Antar Muka Sistem

Desain antar muka sistem dibutuhkan agar pengguna dengan mudah mengoperasikan perangkat lunak yang dibangun.

4.2.3.1 Perancangan Halaman Utama

Halaman utama ini memiliki desain antar muka yang ditunjukkan pada Gambar 4.12. Pada halaman ini dilakukan pemilihan data dari *database*, pemilihan tahap *pre-processing*, Algoritma C4.5, Algoritma C5.0.



Gambar 4.12. Antar Muka Halaman Awal.

4.2.3.2 Perancangan Halaman untuk Menampilkan Data Awal dan Tahap *Pre Processing*

Pada halaman ini ditampilkan *button* untuk menampilkan data dari *database*. Data yang diambil adalah data awal atau data mentah.

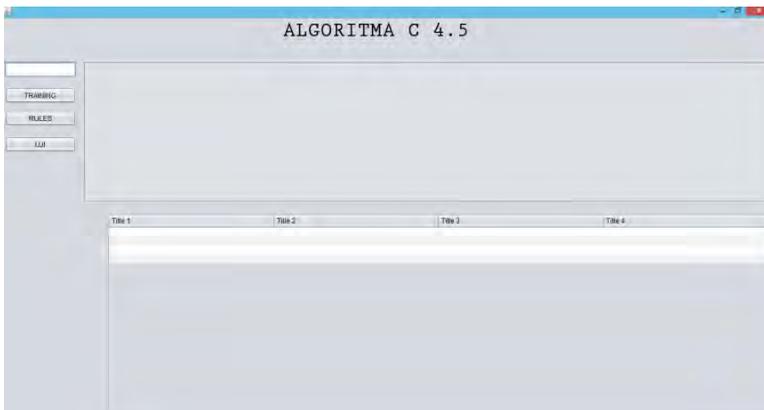
Setelah data ditampilkan, maka tahap *pre-processing* dapat dilakukan dengan memilih *button preprocessing*.



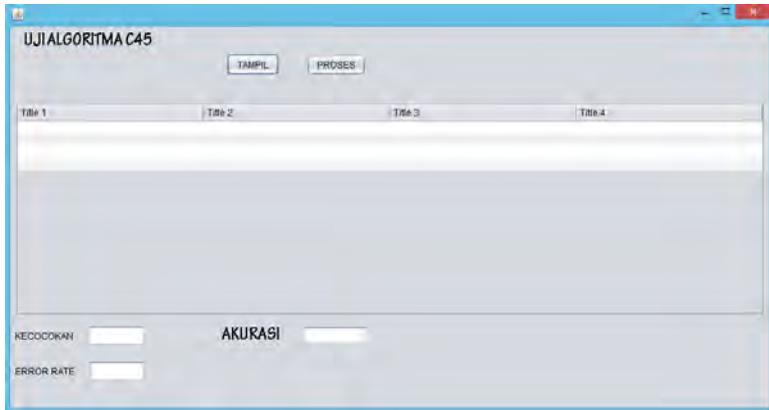
Gambar 4.13. Antar Muka Halaman Utama untuk Melakukan Proses Klasifikasi

4.2.3.3 Perancangan Halaman Algoritma C4.5

Antar muka untuk Algoritma C4.5 terdiri dari antar muka *training* dan uji. Antar muka *training* memungkinkan *user* untuk membagi data menjadi data *training* dan *testing*, dan antar muka *testing* memungkinkan *user* untuk melakukan *testing* pada *rules* hasil *training*. Pada antar muka *training*, *user* dapat melihat pola kecelakaan kerja yang terbentuk sebagai hasil *training*.



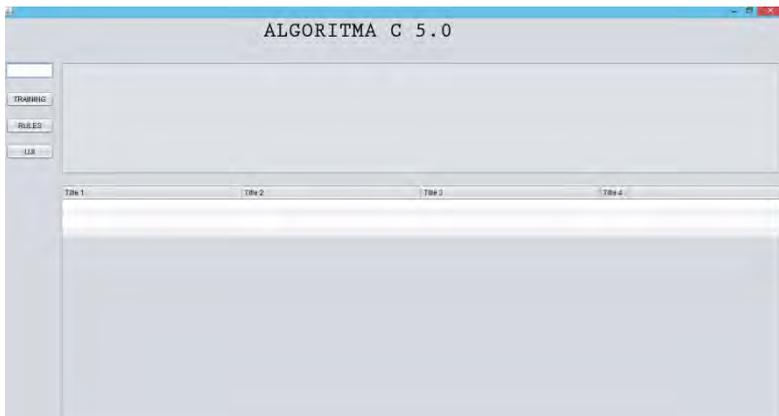
Gambar 4.14. Antar Muka Halaman Training dengan Algoritma C4.5



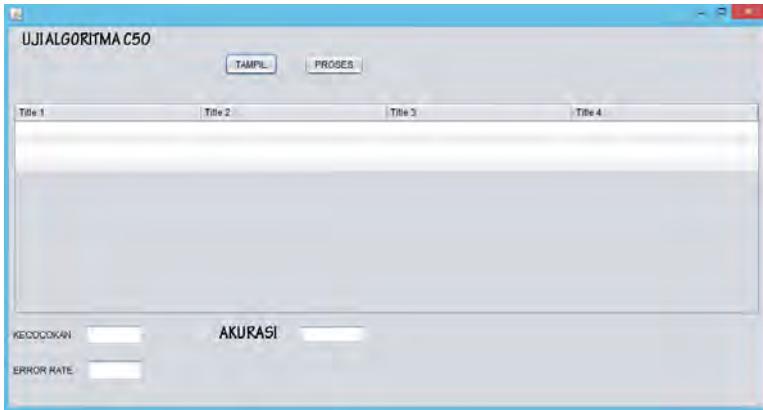
Gambar 4.15. Antar Muka Halaman Testing dengan Algoritma C4.5

4.2.3.4 Perancangan Halaman Algoritma C5.0

Antar muka untuk Algoritma C5.0 terdiri dari antar muka *training* dan uji. Antar muka *training* memungkinkan *user* untuk membagi data menjadi data *training* dan *testing*, dan antar muka *testing* memungkinkan *user* untuk melakukan *testing* pada *rules* hasil *training*.



Gambar 4.16. Antar Muka Halaman Training dengan Algoritma C5.0



Gambar 4.17. Antar Muka Halaman Testing dengan Algoritma C5.0

4.3 Implementasi Sistem

Pada subbab ini ditampilkan implementasi sistem, mulai dari pengambilan data awal, sampai pengujian Algoritma C4.5 dan C5.0

4.3.1 Implementasi Pengambilan Data

Pengambilan data dari *database* akan ditampilkan pada halaman utama dalam bentuk Tabel. *Query* yang digunakan untuk menampilkan data seperti dibawah ini.

```
tampildata();
String sql4 = "DELETE FROM data_karyawan_testing";
cn.executeUpdate(sql4);
String sql ="SELECT * FROM data_karyawan";
Statement stat = conn.createStatement();
ResultSet res = stat.executeQuery(sql);
res.last();
jum = res.getRow();
res.beforeFirst();
JOptionPane.showMessageDialog(rootPane, jum +" data
karyawan berhasil ditampilkan!");
jButton3.enable(true);}
```

```

catch (Exception e){
JOOptionPane.showMessageDialog(null,
"koneksi gagal");
System.out.println(e.getMessage());
}

```

4.3.2 Implementasi Tahap *Pre-Processing*

Data yang telah ditampilkan akan disiapkan untuk tahap *data mining*. Data akan dibersihkan dari *noise* dengan pengecekan duplikasi, *missing prediktor* dan transformasi data.

Pembersihan data dilakukan dengan *query* dibawah ini :

```

koneksi();
String sql1="UPDATE data_karyawan1 SET
"+atr+"'"+modus(atr)+"' where "+atr+"='";
cn.executeUpdate(sql1);
}
catch(Exception e){
JOOptionPane.showMessageDialog(null, "koneksi gagal");
System.out.println(e.getMessage());
}
}
public void missvalnum(String atr){
try{
koneksi();
String sql1="UPDATE data_karyawan1 SET
"+atr+"="+ratarata(atr)+" where "+atr+"=0";
cn.executeUpdate(sql1);
}
catch(Exception e){
JOOptionPane.showMessageDialog(null, "koneksi gagal");
System.out.println(e.getMessage());
}
}

```

Transformasi data dilakukan dengan *query* dibawah ini:

```

sql0="select id,waktu_per_hari,tipe_pekerjaan from
data_karyawan";
ResultSetrs1=conn.createStatement().executeQuery(sql0
);
while(rs1.next()){
if("GRD A".equals(rs1.getString("tipe_pekerjaan"))){
String sql00="update data_karyawan set
waktu_per_hari='08:00:00.00' where
id='"+rs1.getString("id")+"'";
cn.executeUpdate(sql00);
} else if("GRD
B".equals(rs1.getString("tipe_pekerjaan"))){
String sql00="update data_karyawan set
waktu_per_hari='09:00:00.00' where
id='"+rs1.getString("id")+"'";
cn.executeUpdate(sql00);
}else{
String sql00="update data_karyawan set
waktu_per_hari='09:00:00.00' where
id='"+rs1.getString("id")+"'";
cn.executeUpdate(sql00);
}} else if("GRD
B".equals(rs1.getString("tipe_pekerjaan"))){
String sql00="update data_karyawan set
waktu_per_hari='09:00:00.00' where
id='"+rs1.getString("id")+"'";
cn.executeUpdate(sql00);
}else{
String sql00="update data_karyawan set
waktu_per_hari='09:00:00.00' where
id='"+rs1.getString("id")+"'";
cn.executeUpdate(sql00);
} }
} }

```

4.3.3 Implementasi Pembuatan *Tree*

Pembuatan *tree* dilakukan dengan *perhitungan information gain* untuk C4.5 dan *gain ratio* untuk C5.0. Perhitungan ini dilakukan dengan mengambil nilai tertinggi pada tiap iterasi dan dilakukan sampai tidak ada lagi data yang dipecah. Untuk pembuatan cabang pada kedua algoritma sama, yaitu dengan menggunakan *query* dibawah ini.

```

KemunculanKejadian newBranch = new Kemunculan
Kejadian (datasetBaru.get (maxGain) .getNama ());
for (int j = 0; j < datasetBaru.get (maxGain) . Get
AtributKemunculanKejadian ().size (); j++) {
newBranch.SetAtributKemunculanKejadian (datasetBaru
.get (maxGain) .getAtributKemunculanKejadian ().get (j
).getNama ());
newBranch.getAtributKemunculanKejadian ().get (j) .se
tJumlahKejadian ((ArrayList<JumlahKejadian>)
datasetBaru.get (maxGain) .getAtributKemunculanKejad
ian ().get (j) .getJumlahKejadian ().clone ());
newBranch.getAtributKemunculanKejadian ().get (j) .se
tJumlah (datasetBaru.get (maxGain) .getAtributKemuncu
lanKejadian ().get (j) .getJumlah ());
double newEntropy = newBranch.getAtribut
KemunculanKejadian ().get (j) .hitungEntropyAtribut (
);
newBranch.setEntropyTotal (newEntropy);
newBranch.hitungGain ();}
branch.getAtributKemunculanKejadian ().get (i) .setCa
bang (newBranch);
newBranch.setRoot (branch.getAtributKemunculanKejad
ian ().get (i)); branch.getAtribut
KemunculanKejadian ().get (i) .setParent (branch);
for (int j = 0; j < datasetBaru.size (); j++) {
datasetBaru.get (j) .setEntropyTotal (datasetBaru.get (
maxGain) .hitungEntropyTotal ()); }
datasetBaru.remove (maxGain);

```

4.3.4 Implementasi Algoritma C4.5

Implementasi ini terdiri dari *training* dan *testing*. *Training* merupakan *query* untuk membentuk *tree*. Pembentukan *tree* dilakukan dengan menghitung *entropy* dan *information gain*. Dari proses inilah, pola kecelakaan kerja terbentuk. Pola kecelakaan kerja diimplementasikan dalam bentuk *list rules*.

Perhitungan *entropy* dilakukan dengan *query* seperti dibawah ini

```
int jumlah = 0;
this.EntropyTotal = 0;
for (int i = 0; i < this.Total.size(); i++) {
    jumlah+=this.Total.get(i).getJumlah();
}
for (int i = 0; i < this.Total.size(); i++) {
    this.EntropyTotal += ((-
1)*((double)this.Total.get(i).getJumlah()/jumlah)
* Math.log ((( double)this.Total.get(i). getJumlah
()/jumlah))/Math.log(2);
}
return this.EntropyTotal;
```

Perhitungan *information gain* dilakukan dengan *query* seperti dibawah ini

```
int jumlah = 0;
this.Gain = 0;
for (int i = 0; i < Atributes.size(); i++) {
    jumlah += Atributes.get(i).getJumlah();}
for (int i = 0; i < Atributes.size(); i++) {
    this.Gain += ((double )Atributes.get(i).
getJumlah()/jumlah) * Atributes.get(i). hitung
Entropy Attribute();}
double gain = (this.EntropyTotal - this.Gain);
this.Gain = gain;
return this.Gain;
```

Setelah melakukan *training*, *testing* dilakukan untuk mengetahui keakuratan *rules* yang terbentuk. Data yang akan ditesing adalah sebagian data awal yang telah dipecah untuk menjadi data *testing*. *Testing* akan menghitung jumlah kecocokan *rules* terhadap data testing, *error rate*, dan keakuratan.

Error rate adalah jumlah dari *rules* yang tidak cocok dengan data testing. Keakuratan akan dihitung dari hasil kecocokan dan *error rate*. Perhitungan keakuratan dapat dilakukan dengan *query* dibawah ini.

```
tampildatal();
String sql ="SELECT * FROM data_karyawan_testing";
Statement stat = conn.createStatement();
ResultSet res = stat.executeQuery(sql);
res.last();
jum = res.getRow();
res.beforeFirst();
JOptionPane.showMessageDialog(rootPane, jum + " data
karyawan berhasil ditampilkan!");
}
catch (Exception e){
JOptionPane.showMessageDialog(null, "koneksi
gagal");
System.out.println(e.getMessage());
}
double hitung = 100 * cocok/jum;
jTextField1.setText(""+cocok);
jTextField2.setText(""+(jum-cocok));
jTextField3.setText(hitung+" %");
```

4.3.5 Implementasi Algoritma C5.0

Implementasi ini terdiri dari *training* dan *testing*. *Training* merupakan *query* untuk membentuk *tree*. Pembentukan *tree* dilakukan dengan menghitung *entropy* dan *gain ratio*. Dari proses inilah, pola kecelakaan kerja terbentuk. Pola kecelakaan diimplementasikan dalam bentuk *list rules*.

Perhitungan *entropy* dilakukan dengan *query* seperti dibawah ini

```
int jumlah = 0;
this.EntropyTotal = 0;
for (int i = 0; i < this.Total.size(); i++) {
jumlah+=this.Total.get(i).getJumlah();}
for (int i = 0; i < this.Total.size(); i++) {
this.EntropyTotal += ((-
1)*((double)this.Total.get(i).getJumlah()/jumlah))
* Math.log ((( double)this.Total.get(i). getJumlah
()/jumlah))/Math.log(2);}
return this.EntropyTotal;
```

Perhitungan *gain ratio* dilakukan dengan *query* seperti dibawah ini

```
public double hitungGain50(){
int jumlah = 0;
this.Gain = 0;
int jument = 0;
for (int i = 0; i < Atributes.size(); i++) {
jumlah += Atributes.get(i).getJumlah();}
for (int i = 0; i < Atributes.size(); i++) {
this.Gain += ((double)Atributes.get(i) .getJumlah()
/jumlah) * Atributes.get(i). hitungEntropy
Attribute(); }
for (int i = 0; i < Atributes.size(); i++) {
jument +=
Atributes.get(i).hitungEntropyAttribute();}
double gain=(this.EntropyTotal -this.Gain) / j u
ment;
this.Gain = gain;
return this.Gain; }
```

Setelah melakukan *training*, *testing* dilakukan untuk mengetahui keakuratan pola kecelakaan kerja (*rules*) yang terbentuk. *Testing* yang dilakukan juga sama dengan *testing* pada

Algoritma C5.0. Data yang akan ditesting adalah sebagian data awal yang telah dipecah untuk menjadi data *testing*. *Testing* akan menghitung jumlah kecocokan *rules* terhadap data testing, *error rate*, dan keakuratan.

Error rate adalah jumlah dari *rules* yang tidak cocok dengan data *testing*. Keakuratan akan dihitung dari hasil kecocokan dan *error rate*. Perhitungan keakuratan dapat dilakukan dengan query dibawah ini.

```
tampildata1();
String sql ="SELECT * FROM data_karyawan_testing";
Statement stat = conn.createStatement();
ResultSet res = stat.executeQuery(sql);
res.last();
jum = res.getRow();
res.beforeFirst();
JOptionPane.showMessageDialog(rootPane, jum +" data
karyawan berhasil ditampilkan!");
}
catch (Exception e){
JOptionPane.showMessageDialog(null,"koneksi
gagal");
System.out.println(e.getMessage());
}
double hitung = 100 * cocok/jum;
    jTextField1.setText(""+cocok);
    jTextField2.setText(""+(jum-cocok));
    jTextField3.setText(hitung+" %");
```

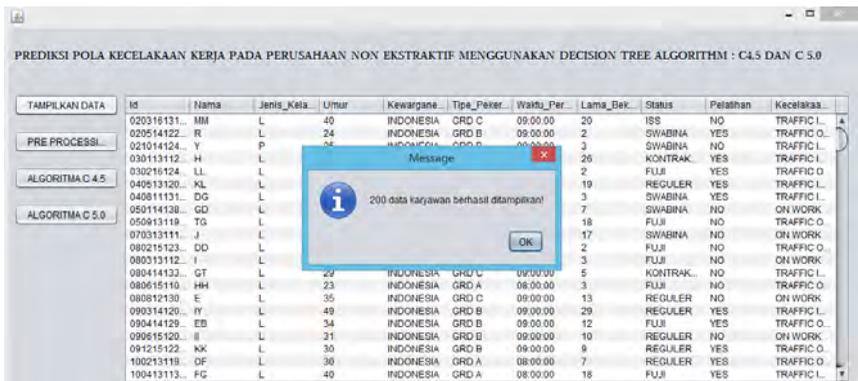
“Halaman ini sengaja dikosongkan”

BAB V PENGUJIAN DAN PEMBAHASAN

Pada bab ini berisi metodologi atau urutan kegiatan yang dilakukan dalam menyelesaikan Tugas Akhir, meliputi studi literatur, pengumpulan data, analisis dan desain sistem, pembuatan program, uji coba dan evaluasi, hingga penulisan Tugas Akhir.

5.1 *Pre Processing* dan Pembagian Data Uji

Pre Processing dilakukan diawal menjalankan perangkat lunak ini. Hasil dari tahap *pre processing* ini terlihat pada Gambar 5.1 dan 5.2 dibawah ini.



| TAMPILKAN DATA | Id | Nama | Jenis_Kelua | Umur | Kewargana | Tipe_Dalvar | Waktu_Per | Lama_Bek | Status | Pelatihan | Kecelakaan |
|----------------|-----------|------|-------------|------|-----------|-------------|-----------|----------|---------|-----------|------------|
| | 020316131 | MM | L | 40 | INDONESIA | GRD C | 09:00:00 | 20 | ISS | NO | TRAFFIC I |
| | 020514422 | RR | L | 24 | INDONESIA | GRD B | 09:00:00 | 2 | SWABINA | YES | TRAFFIC O |
| | 021014124 | Y | P | 34 | INDONESIA | GRD B | 09:00:00 | 3 | SWABINA | NO | TRAFFIC I |
| | 030113124 | H | L | | | | | 26 | KONTRAK | YES | TRAFFIC I |
| | 030216124 | LL | L | | | | | 2 | FUJI | YES | TRAFFIC O |
| | 040513120 | KL | L | | | | | 19 | REGULER | YES | TRAFFIC L |
| | 040811131 | DG | L | | | | | 3 | SWABINA | YES | TRAFFIC I |
| | 050114138 | GD | L | | | | | 7 | SWABINA | NO | ON WORK |
| | 050913119 | TG | L | | | | | 18 | FUJI | NO | TRAFFIC O |
| | 070313111 | J | L | | | | | 17 | SWABINA | NO | ON WORK |
| | 080215123 | DD | L | | | | | 2 | FUJI | NO | TRAFFIC O |
| | 080313112 | I | L | | | | | 3 | FUJI | NO | ON WORK |
| | 080414133 | GT | L | 29 | INDONESIA | GRD C | 09:00:00 | 5 | KONTRAK | NO | TRAFFIC I |
| | 080615110 | HH | L | 23 | INDONESIA | GRD A | 08:00:00 | 3 | FUJI | NO | TRAFFIC O |
| | 080812130 | E | L | 35 | INDONESIA | GRD C | 09:00:00 | 13 | REGULER | NO | ON WORK |
| | 090314120 | FF | L | 49 | INDONESIA | GRD B | 09:00:00 | 29 | REGULER | YES | TRAFFIC L |
| | 090414129 | EB | L | 34 | INDONESIA | GRD B | 09:00:00 | 12 | FUJI | YES | TRAFFIC O |
| | 090615120 | II | L | 31 | INDONESIA | GRD B | 09:00:00 | 10 | REGULER | NO | ON WORK |
| | 091215122 | KK | L | 30 | INDONESIA | GRD B | 09:00:00 | 9 | REGULER | YES | TRAFFIC O |
| | 100213119 | OF | L | 30 | INDONESIA | GRD A | 08:00:00 | 7 | REGULER | YES | TRAFFIC O |
| | 100413113 | FG | L | 40 | INDONESIA | GRD A | 08:00:00 | 18 | FUJI | YES | TRAFFIC I |

Gambar 5.1 Hasil Tampilan Data dari Database

PREDIKSI POLA KECELAKAAN KERJA PADA PERUSAHAAN NON EKSTRAKTIV MENGGUNAKAN DECISION TREE ALGORITHM : C4.5 DAN C5.0

| TAMPAKAN DATA | Jenis_Kelamin | Umur | Tipe_Penyerapan | Lama_Betapa | Status | Penilaian | Kecelakaan_Kerja |
|----------------|---------------|-------|-----------------|-------------|------------|-----------|---------------------|
| PRE PROCESSING | L | umur3 | GRD C | Senior | ISS | NO | TRAFFIC INSIDE FACT |
| ALGORITMA C4.5 | L | umur1 | GRD B | Junior | SIWABINA | NO | TRAFFIC ON COMMUTE |
| ALGORITMA C5.0 | L | umur1 | GRD B | Junior | SIWABINA | NO | TRAFFIC INSIDE FACT |
| | L | umur4 | GRD A | Senior | KONTRAKTOR | YES | TRAFFIC INSIDE FACT |
| | L | umur1 | GRD B | Junior | FUJI | YES | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD C | Senior | REGULER | YES | TRAFFIC INSIDE FACT |
| | L | umur2 | GRD C | Junior | SIWABINA | YES | TRAFFIC INSIDE FACT |
| | L | umur2 | GRD C | Junior | SIWABINA | NO | TRAFFIC ON COMMUTE |
| | L | umur3 | GRD A | Senior | FUJI | NO | TRAFFIC ON COMMUTE |
| | L | umur1 | GRD A | Senior | SIWABINA | NO | TRAFFIC ON COMMUTE |
| | L | umur1 | GRD B | Junior | FUJI | NO | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD C | Senior | KONTRAKTOR | NO | TRAFFIC INSIDE FACT |
| | L | umur1 | GRD C | Junior | FUJI | NO | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD C | Senior | REGULER | NO | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD C | Junior | EGULER | NO | TRAFFIC INSIDE FACT |
| | L | umur2 | GRD C | Junior | EGULER | YES | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD C | Junior | FUJI | YES | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD C | Junior | EGULER | NO | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD C | Junior | EGULER | YES | TRAFFIC ON COMMUTE |
| | L | umur3 | GRD A | Senior | REGULER | YES | TRAFFIC ON COMMUTE |
| | L | umur3 | GRD A | Senior | FUJI | NO | TRAFFIC INSIDE FACT |
| | L | umur3 | GRD B | Senior | REGULER | YES | TRAFFIC INSIDE FACT |
| | L | umur3 | GRD B | Senior | REGULER | YES | TRAFFIC INSIDE FACT |
| | L | umur2 | GRD C | Junior | REGULER | YES | TRAFFIC ON COMMUTE |
| | L | umur4 | GRD A | Senior | REGULER | YES | TRAFFIC ON COMMUTE |
| | L | umur3 | GRD A | Senior | SIWABINA | NO | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD A | Senior | SIWABINA | YES | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD B | Junior | FUJI | NO | TRAFFIC ON COMMUTE |
| | L | umur3 | GRD A | Senior | SIWABINA | YES | TRAFFIC ON COMMUTE |
| | L | umur1 | GRD A | Junior | REGULER | NO | TRAFFIC ON COMMUTE |
| | L | umur1 | GRD B | Junior | FUJI | NO | TRAFFIC ON COMMUTE |
| | L | umur4 | GRD A | Senior | FUJI | YES | TRAFFIC ON COMMUTE |
| | L | umur1 | GRD A | Senior | FUJI | YES | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD A | Junior | SIWABINA | YES | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD B | Junior | REGULER | YES | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD B | Junior | REGULER | YES | TRAFFIC ON COMMUTE |
| | L | umur2 | GRD B | Junior | FUJI | YES | TRAFFIC ON COMMUTE |

Gambar 5.2 Hasil Pre Processing

Uji coba pada perangkat lunak dalam penelitian Tugas Akhir ini dilakukan dengan beberapa pembagian data *testing* dan data *training*. Data *training* adalah data yang digunakan untuk membentuk pola kecelakaan kerja. Sedangkan data *testing* adalah data yang digunakan untuk menguji pola kecelakaan kerja yang telah dihasilkan oleh tahap *training*. Dalam perangkat lunak ini dibagi beberapa jenis pembagian seperti Tabel 5.1

Tabel 5.1 Pembagian Data yang Digunakan.

| No | Data Training | Data Testing |
|----|---------------|--------------|
| 1 | 95% | 5% |
| 2 | 90% | 10% |
| 3 | 85% | 15% |
| 4 | 80% | 20% |
| 5 | 75% | 25% |

5.2 Pengujian Algoritma C4.5

Pengujian pada Algoritma C4.5 dilakukan sebanyak 5 kali dengan pembagian data seperti pada Tabel 5.1. Pembagian data tersebut bertujuan untuk mengetahui pembagian data yang paling baik dalam melakukan klasifikasi. Ukuran baik tersebut yaitu menghasilkan angka akurasi tertinggi.

Berikut ini hasil pengujian dan pembahasan pola kecelakaan kerja dan angka keakuratan yang dihasilkan Algoritma C4.5.

Pembagian Data 95% : 5%

Pada pembagian ini data *training* sebanyak 95% dari data hasil *pre-processing* (200 data). Sehingga data *testing* sebanyak 10 data. Hasil dari *training* dan *testing* terdapat pada Gambar 5.3. dan 5.4. di bawah ini.

The screenshot shows the ALGORITMA C 4.5 interface. The main window displays the following content:

ALGORITMA C 4.5

95

-----Pembagian pada cabang 05-----
 Gain jenis_alamat 1 0.0
 Gain Umur 0.0
 Gain tpx_pelayanan 0.859024216483541
 Gain jmla_bekena 0.0
 Gain pedahan 0.0
 Cabang terbaginya adalah: Yes_>0503040

-----Pembagian keputusan pada cabang GRD C adalah TRAFFIC INSIDE FACT

-----Pembagian pada cabang REGULER-----
 Gain jenis_alamat 0.0
 Gain Umur 0.1477810567084295
 Gain tpx_pelayanan 0.1035642094442029
 Gain jmla_bekena 0.1332840325884087

| RULES | DECISION |
|----------------------------------|---------------------|
| 05 -> GRD C | TRAFFIC INSIDE FACT |
| 05B -> GRD A | TRAFFIC ON COMMUTE |
| REGULER -> YES -> umur2 -> GRD B | TRAFFIC ON COMMUTE |
| REGULER -> YES -> umur4 -> GRD B | TRAFFIC INSIDE FACT |
| REGULER -> YES -> umur4 -> GRD C | TRAFFIC INSIDE FACT |
| REGULER -> YES -> umur4 -> GRD A | TRAFFIC ON COMMUTE |
| REGULER -> NO -> umur1 | TRAFFIC ON COMMUTE |
| REGULER -> NO -> umur2 -> GRD C | ON WORK |
| REGULER -> NO -> umur3 -> GRD B | TRAFFIC ON COMMUTE |
| REGULER -> NO -> umur3 -> GRD C | TRAFFIC ON COMMUTE |
| REGULER -> NO -> umur4 -> GRD B | ON WORK |
| REGULER -> NO -> umur4 -> GRD C | ON WORK |
| REGULER -> NO -> umur4 -> GRD A | TRAFFIC ON COMMUTE |
| FLUJ -> GRD B -> umur1 | TRAFFIC ON COMMUTE |
| FLUJ -> GRD B -> umur2 -> YES | TRAFFIC ON COMMUTE |
| FLUJ -> GRD B -> umur3 | TRAFFIC ON COMMUTE |
| FLUJ -> GRD B -> umur4 | TRAFFIC ON COMMUTE |
| FLUJ -> GRD C -> umur2 | TRAFFIC INSIDE FACT |
| FLUJ -> GRD C -> umur3 | TRAFFIC ON COMMUTE |
| FLUJ -> GRD C -> umur4 -> YES | TRAFFIC ON COMMUTE |
| FLUJ -> GRD A -> umur1 -> YES | TRAFFIC ON COMMUTE |
| FLUJ -> GRD A -> umur2 -> NO | TRAFFIC ON COMMUTE |

Gambar 5.3 Hasil *Training* pada C4.5 dengan 95% Data

The screenshot shows a window titled "UJI ALGORITMA C4.5" with two buttons: "TAMPIL" and "PROSES". Below the buttons is a table with 10 columns: ID, Jenis_Kelamin, Umur, Tipe_Pekerjaan, Lama_Bekerja, Status, Pelatihan, Kecelakaan_Ka, and Klasifikasi. The table contains 10 rows of data. Below the table, there are input fields for "KECOCOKAN" (7), "AKURASI" (70.0 %), and "ERROR RATE" (3).

| ID | Jenis_Kelamin | Umur | Tipe_Pekerjaan | Lama_Bekerja | Status | Pelatihan | Kecelakaan_Ka | Klasifikasi |
|-------------|---------------|-------|----------------|--------------|------------|-----------|-----------------|-------------|
| 02031912145 | L | umur3 | GRD C | Senior | ISS | NO | TRAFFIC INSD... | cocok |
| 02051412291 | L | umur1 | GRD B | Junior | SWABINA | YES | TRAFFIC DN C... | cocok |
| 02101412431 | P | umur1 | GRD B | Junior | SWABINA | NO | TRAFFIC INSD... | cocok |
| 03011314241 | L | umur4 | GRD A | Senior | KONTRAKTOR | YES | TRAFFIC INSD... | cocok |
| 03021612401 | L | umur1 | GRD B | Junior | FLJL | YES | TRAFFIC DN C... | cocok |
| 04051312049 | L | umur3 | GRD B | Senior | REGULER | YES | TRAFFIC INSD... | tidak cocok |
| 04081113170 | L | umur2 | GRD C | Junior | SWABINA | YES | TRAFFIC INSD... | tidak cocok |
| 05011413803 | L | umur2 | GRD C | Junior | SWABINA | NO | DN WORK | cocok |
| 05091118460 | L | umur2 | GRD A | Senior | FLJL | NO | TRAFFIC DN C... | cocok |
| 07031311130 | L | umur3 | GRD A | Senior | SWABINA | NO | DN WORK | tidak cocok |

Gambar 5.4 Hasil *Testing* pada C4.5 dengan 5% Data

Pola kecelakaan kerja yang terbentuk sebanyak 38 pola (*rules*). Hasil keakuratan yang ditunjukkan pada *testing* sebesar 70%. Hasil pola kecelakaan kerja selengkapnya disajikan pada lampiran

Pembagian Data 90% : 10%

Pada pembagian ini data *training* sebanyak 90% dari data hasil *pre-processing* (200 data). Sehingga data *testing* sebanyak 20 data. Hasil dari *training* dan *testing* terdapat pada Gambar 5.5. dan 5.6. di bawah ini.

The screenshot shows a window titled "ALGORITMA C 4.5" with buttons for "TRAINING", "RULES", and "JLR". Below the buttons, there are fields for "ID" and "No". The main area displays training data and a list of rules. The rules are organized into two columns: "RULES" and "DECISION".

| RULES | DECISION |
|-------------------------------|--------------------|
| ISS → GRD C | TRAFFIC INSD F/ACT |
| ISS → GRD A | TRAFFIC DN COMMUTE |
| REGULER → YES → GRD B → umur2 | TRAFFIC DN COMMUTE |
| REGULER → YES → GRD B → umur4 | TRAFFIC INSD F/ACT |
| REGULER → YES → GRD C → umur4 | TRAFFIC INSD F/ACT |
| REGULER → YES → GRD A → umur4 | TRAFFIC DN COMMUTE |
| REGULER → NO → umur1 | TRAFFIC DN COMMUTE |
| REGULER → NO → umur4 → GRD C | DN WORK |
| REGULER → NO → umur3 → GRD B | TRAFFIC DN COMMUTE |
| REGULER → NO → umur3 → GRD A | TRAFFIC DN COMMUTE |
| REGULER → NO → umur4 → GRD B | DN WORK |
| REGULER → NO → umur4 → GRD A | TRAFFIC DN COMMUTE |
| FLJL → GRD B → umur2 → YES | TRAFFIC DN COMMUTE |
| FLJL → GRD B → umur2 | TRAFFIC DN COMMUTE |
| FLJL → GRD C → umur2 | TRAFFIC DN COMMUTE |
| FLJL → GRD C → umur2 | TRAFFIC INSD F/ACT |
| FLJL → GRD A → umur1 → YES | TRAFFIC DN COMMUTE |
| FLJL → GRD A → umur1 → YES | TRAFFIC DN COMMUTE |

Gambar 5.5 Hasil *Training* pada C4.5 dengan 90% Data

The screenshot shows a software window titled "UJI ALGORITMA C45". It contains a table with 10 columns: ID, Jenis_kelamin, Umur, Tipe_Pekerjaan, Lama_Deteksi, Status, Pelatihan, Kecelakaan_Ke..., and Klasifikasi. Below the table, there are input fields for "KECOCOKKAN" (set to 14) and "ERROR RATE" (set to 10). The "AKURASI" (Accuracy) is displayed as 46.0%.

| ID | Jenis_kelamin | Umur | Tipe_Pekerjaan | Lama_Deteksi | Status | Pelatihan | Kecelakaan_Ke... | Klasifikasi |
|--------------|---------------|-------|----------------|--------------|------------|-----------|------------------|-------------|
| 020316131465 | L | umur3 | GRD C | Senior | ISS | NO | TRAFFIC INSID... | cocok |
| 020514122911 | L | umur1 | GRD B | Junior | SWABIRIA | YES | TRAFFIC ON C... | cocok |
| 021014124231 | P | umur1 | GRD B | Junior | SWABIRIA | NO | TRAFFIC INSID... | cocok |
| 030113112111 | L | umur4 | GRD A | Senior | KONTRAKTOR | YES | TRAFFIC INSID... | cocok |
| 030216124011 | L | umur1 | GRD B | Junior | FLUJ | YES | TRAFFIC ON C... | cocok |
| 040513120449 | L | umur3 | GRD B | Senior | REGULER | YES | TRAFFIC INSID... | tidak cocok |
| 040811131710 | L | umur2 | GRD C | Junior | SWABIRIA | YES | TRAFFIC INSID... | tidak cocok |
| 050114138033 | L | umur2 | GRD C | Junior | SWABIRIA | NO | ON WORK | cocok |
| 050513119460 | L | umur3 | GRD A | Senior | FLUJ | NO | TRAFFIC ON C... | cocok |
| 070313111339 | L | umur3 | GRD A | Senior | SWABIRIA | NO | ON WORK | tidak cocok |
| 080215123133 | L | umur1 | GRD B | Junior | FLUJ | NO | TRAFFIC ON C... | cocok |
| 080313112112 | L | umur1 | GRD A | Junior | FLUJ | NO | ON WORK | tidak cocok |
| 080411413320 | L | umur2 | GRD C | Junior | KONTRAKTOR | NO | TRAFFIC INSID... | tidak cocok |
| 080515110911 | L | umur1 | GRD A | Junior | FLUJ | NO | TRAFFIC ON C... | tidak cocok |
| 080812130719 | L | umur2 | GRD C | Junior | REGULER | NO | ON WORK | cocok |
| 080914130607 | L | umur4 | GRD A | Senior | REGULER | YES | TRAFFIC INSID... | cocok |

Gambar 5.8 Hasil Testing pada C4.5 dengan 15% Data

Pola kecelakaan kerja yang dihasilkan sebanyak 38 pola (*rules*). Hasil keakuratan yang ditunjukkan pada *testing* sebesar 46%. Hasil pola kecelakaan kerja selengkapnya disajikan pada lampiran.

Pembagian Data 80% : 20%

Pada pembagian ini data *training* sebanyak 80% dari data hasil *pre-processing* (200 data). Sehingga data *testing* sebanyak 40 data. Hasil dari *training* dan *testing* terdapat pada Gambar 5.9. dan 5.10. di bawah ini.

The screenshot shows a software window titled "ALGORITMA C 4.5". It displays a list of rules under the heading "RULES" and a corresponding "DECISION" column. The rules are based on attributes like 'ISS', 'GRD', 'umur', and 'FLUJ'. The decision outcomes include 'TRAFFIC INSIDE FACT', 'TRAFFIC ON COMMITTE', 'ON WORK', and 'TRAFFIC ON COMMITTE'.

| RULES | DECISION |
|----------------------------------|---------------------|
| ISS → GRD C | TRAFFIC INSIDE FACT |
| ISS → GRD A | TRAFFIC ON COMMITTE |
| REGULER → umur1 | TRAFFIC ON COMMITTE |
| REGULER → umur2 → YES → GRD B | TRAFFIC ON COMMITTE |
| REGULER → umur3 → NO → GRD C | ON WORK |
| REGULER → umur3 → GRD B → NO | TRAFFIC ON COMMITTE |
| REGULER → umur4 → GRD B → YES | TRAFFIC INSIDE FACT |
| REGULER → umur4 → GRD C → YES | ON WORK |
| REGULER → umur4 → GRD C → NO | TRAFFIC INSIDE FACT |
| REGULER → umur4 → GRD A | TRAFFIC ON COMMITTE |
| FLUJ → GRD B → umur1 | TRAFFIC ON COMMITTE |
| FLUJ → GRD B → umur2 → YES | TRAFFIC ON COMMITTE |
| FLUJ → GRD B → umur2 → NO | ON WORK |
| FLUJ → GRD B → umur4 | TRAFFIC ON COMMITTE |
| FLUJ → GRD C → umur2 | TRAFFIC INSIDE FACT |
| FLUJ → GRD C → umur3 | TRAFFIC ON COMMITTE |
| FLUJ → GRD A → umur1 → YES | TRAFFIC ON COMMITTE |
| FLUJ → GRD A → umur4 → NO → FLUJ | TRAFFIC ON COMMITTE |

Gambar 5.9 Hasil Training pada C4.5 dengan 80% Data

UJI ALGORITMA C4.5

TAMBAH PROSES

| ID | Jenis_Kelamin | Umur | Tipe_Pelatihan | Lama_Bekerja | Status | Pelatihan | Kecelakaan_Kerja | Klasifikasi |
|-------------|---------------|------|----------------|--------------|------------|-----------|---------------------|-------------|
| 02031915145 | L | umr3 | GRD C | Senior | ISE | NO | | |
| 02051412021 | L | umr1 | GRD B | Junior | SWABHA | YES | TRAFFIC ON COMMUTE | 0000 |
| 02071912421 | P | umr1 | GRD B | Junior | SWABHA | NO | TRAFFIC INSIDE FACT | 0000 |
| 02091312121 | L | umr4 | GRD A | Senior | KONTRAKTOR | YES | TRAFFIC INSIDE FACT | 0000 |
| 02021912401 | L | umr1 | GRD B | Junior | FLJ | YES | TRAFFIC ON COMMUTE | 0000 |
| 04051310389 | L | umr3 | GRD B | Senior | REGULER | YES | TRAFFIC INSIDE FACT | 8000 cocok |
| 04081312170 | L | umr2 | GRD C | Junior | SWABHA | YES | TRAFFIC INSIDE FACT | 8000 cocok |
| 05011413803 | L | umr2 | GRD C | Junior | SWABHA | NO | ON WORK | 0000 |
| 02091311840 | L | umr3 | GRD A | Senior | FLJ | NO | TRAFFIC ON COMMUTE | 0000 |
| 07031311520 | L | umr3 | GRD A | Senior | SWABHA | NO | ON WORK | 8000 cocok |
| 08021512213 | L | umr1 | GRD B | Junior | FLJ | NO | TRAFFIC ON COMMUTE | 0000 |
| 08031512122 | L | umr1 | GRD A | Junior | FLJ | NO | ON WORK | 8000 cocok |
| 08041413200 | L | umr2 | GRD C | Junior | KONTRAKTOR | NO | TRAFFIC INSIDE FACT | 8000 cocok |
| 08081510381 | L | umr1 | GRD A | Junior | FLJ | NO | TRAFFIC ON COMMUTE | 8000 cocok |
| 08081213079 | L | umr2 | GRD C | Junior | REGULER | NO | ON WORK | 0000 |
| 09021512007 | L | umr2 | GRD B | Junior | REGULER | YES | TRAFFIC INSIDE FACT | 0000 |
| 09041512004 | L | umr2 | GRD B | Junior | FLJ | YES | TRAFFIC ON COMMUTE | 0000 |
| 09121512005 | L | umr2 | GRD B | Junior | REGULER | NO | ON WORK | 8000 cocok |
| 09121512005 | L | umr2 | GRD B | Junior | REGULER | YES | TRAFFIC ON COMMUTE | 0000 |
| 10021911964 | L | umr2 | GRD A | Junior | REGULER | YES | TRAFFIC ON COMMUTE | 8000 cocok |
| 10041311381 | L | umr3 | GRD A | Senior | FLJ | YES | TRAFFIC INSIDE FACT | 8000 cocok |
| 10041312001 | L | umr3 | GRD B | Senior | REGULER | YES | TRAFFIC INSIDE FACT | 8000 cocok |
| 10041312001 | L | umr3 | GRD B | Senior | REGULER | YES | TRAFFIC INSIDE FACT | 8000 cocok |
| 10041513341 | L | umr2 | GRD C | Junior | REGULER | YES | TRAFFIC ON COMMUTE | 8000 cocok |
| 10051511700 | L | umr4 | GRD A | Senior | REGULER | YES | TRAFFIC ON COMMUTE | 0000 |
| 10051511608 | L | umr3 | GRD A | Senior | SWABHA | YES | TRAFFIC ON COMMUTE | 8000 cocok |
| 10051511618 | L | umr3 | GRD A | Senior | SWABHA | YES | TRAFFIC ON COMMUTE | 8000 cocok |
| 10081412430 | L | umr2 | GRD B | Junior | FLJ | NO | TRAFFIC ON COMMUTE | 8000 cocok |
| 10020413451 | L | umr3 | GRD A | Senior | SWABHA | YES | TRAFFIC ON COMMUTE | 8000 cocok |

NECESSARY 21 AKURASI 62.0 %

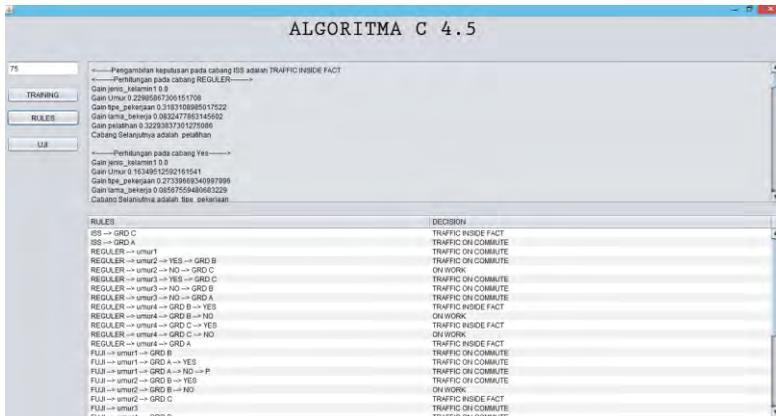
ERROR RATE 19

Gambar 5.10 Hasil *Testing* pada C4.5 dengan 20% Data

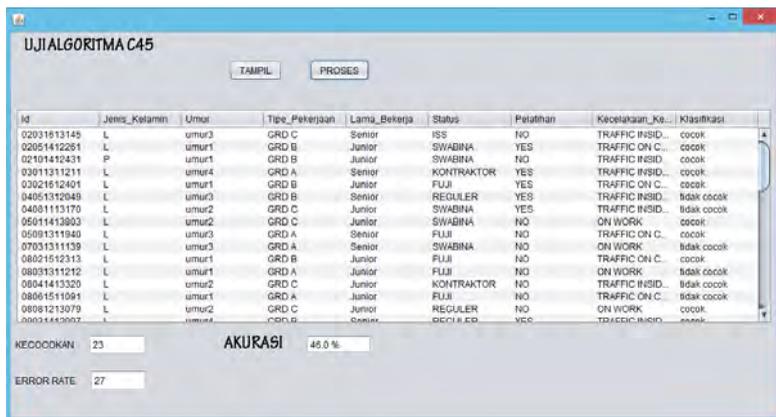
Pola kecelakaan kerja yang dihasilkan sebanyak 38 pola (*rules*). Hasil keakuratan yang ditunjukkan pada *testing* sebesar 52%. Hasil pola kecelakaan kerja selengkapnya disajikan pada lampiran.

Pembagian Data 75% : 25%

Pada pembagian ini data *training* sebanyak 75% dari data hasil *pre-processing* (200 data). Sehingga data *testing* sebanyak 50 data. Hasil dari *training* dan *testing* terdapat pada Gambar 5.11. dan 5.12. di bawah ini.



Gambar 5.11 Hasil *Training* pada C4.5 dengan 75% Data



Gambar 5.12 Hasil *Testing* pada C4.5 dengan 25% Data

Pola kecelakaan kerja yang dihasilkan sebanyak 38 pola (*rules*). Hasil keakuratan yang ditunjukkan pada *testing* sebesar 46%. Hasil pola kecelakaan kerja selengkapnya disajikan pada lampiran.

Dari hasil *testing* yang telah dilakukan, terlihat bahwa angka keakuratan tertinggi saat pembagian data sebesar 95%. Perbandingan hasil keakuratan, terlihat pada Tabel 5.2

Tabel 5.2. Tabel Hasil Algoritma C4.5

| Pembagian Data | Pola yang Dihasilkan | Hasil |
|----------------|----------------------|-------|
| 95% - 5% | 38 pola | 70% |
| 90% - 10% | 38 pola | 60% |
| 85% - 15% | 38 pola | 46% |
| 80% - 20 % | 38 pola | 52% |
| 75% - 25 % | 38 pola | 46% |

5.3 Pengujian Algoritma C5.0

Pengujian pada Algoritma C5.0 dilakukan beberapa kali dengan pembagian data seperti pada Tabel 5.1. Pembagian data tersebut bertujuan untuk mengetahui pembagian data yang paling baik dalam melakukan klasifikasi. Ukuran baik ini yaitu menghasilkan angka akurasi tertinggi.

Berikut ini hasil pengujian dan pembahasan pola kecelakaan kerja dan angka keakuratan yang dihasilkan algoritma C5.0.

Pembagian Data 95% : 5%

Pada pembagian ini data *training* sebanyak 95% dari data hasil *pre-processing* (200 data). Sehingga data *testing* sebanyak 10 data. Hasil dari *training* dan *testing* terdapat pada Gambar 5.13. dan 5.14. di bawah ini.

**Gambar 5.13** Hasil *Training* pada C5.0 dengan 95% Data

UJIALGORITMA C5.0

TAMPL PROSES

| Id | Jenis_Kelamin | Umur | Tipe_Pekerjaan | Lama_Belajar | Status | Pelatihan | Kecelakaan_Kerja | Klasifikasi |
|-------------|---------------|-------|----------------|--------------|------------|-----------|---------------------|-------------|
| 02031013146 | L | umur3 | GRD C | Senior | ISB | NO | TRAFFIC INSIDE FACT | cocok |
| 02051412261 | L | umur1 | GRD B | Junior | SIWABNA | YES | TRAFFIC ON COMMUTE | cocok |
| 02101412431 | P | umur1 | GRD B | Junior | SIWABNA | NO | TRAFFIC INSIDE FACT | cocok |
| 03011311211 | L | umur4 | GRD A | Senior | KONTRAKTOR | YES | TRAFFIC INSIDE FACT | cocok |
| 03021012401 | L | umur1 | GRD B | Junior | FLUB | YES | TRAFFIC ON COMMUTE | cocok |
| 04051312040 | L | umur3 | GRD B | Senior | REGULER | YES | TRAFFIC INSIDE FACT | tidak cocok |
| 04081113170 | L | umur2 | GRD C | Junior | SIWABNA | YES | TRAFFIC INSIDE FACT | tidak cocok |
| 05011413803 | L | umur2 | GRD C | Junior | SIWABNA | NO | ON WORK | cocok |
| 0501111360 | L | umur3 | GRD A | Senior | FLUB | NO | TRAFFIC ON COMMUTE | tidak cocok |
| 07031311130 | L | umur1 | GRD A | Senior | SIWABNA | NO | ON WORK | tidak cocok |

GELOMBOK: 6 AKURASI: 60.0 %

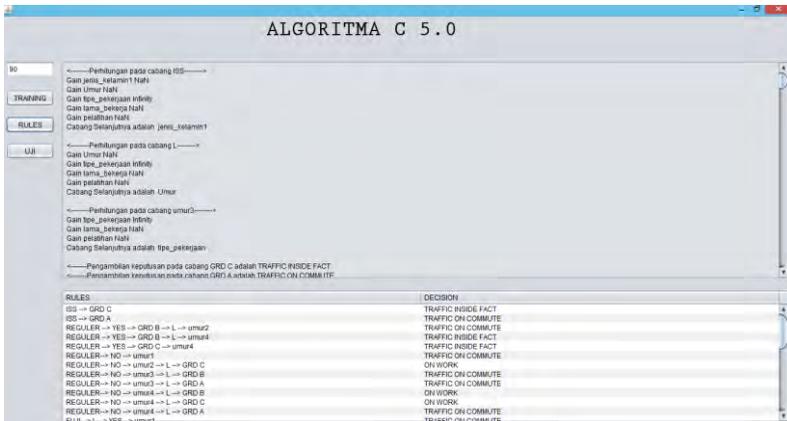
ERROR RATE: 4

Gambar 5.14 Hasil *Testing* pada C5.0 dengan 5% Data

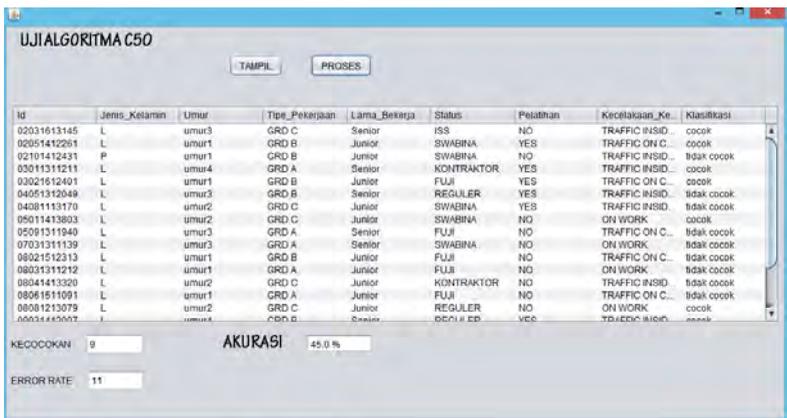
Pola kecelakaan kerja yang dihasilkan sebanyak 35 pola (*rules*). Hasil keakuratan yang ditunjukkan pada *testing* sebesar 60%. Hasil pola kecelakaan kerja selengkapnya disajikan pada lampiran.

Pembagian Data 90% : 10%

Pada pembagian ini data *training* sebanyak 90% dari data hasil *pre-processing* (200 data). Sehingga data *testing* sebanyak 20 data. Hasil dari *training* dan *testing* terdapat pada Gambar 5.15 dan 5.16. di bawah ini.



Gambar 5.15 Hasil *Training* pada C5.0 dengan 90% Data

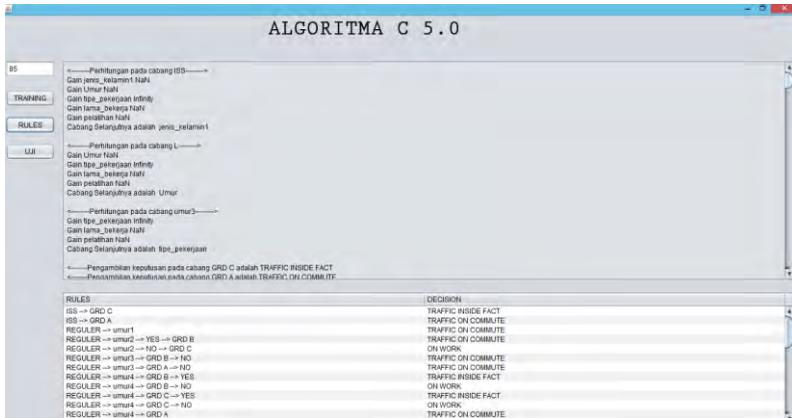


Gambar 5.16 Hasil *Testing* pada C5.0 dengan 10% Data

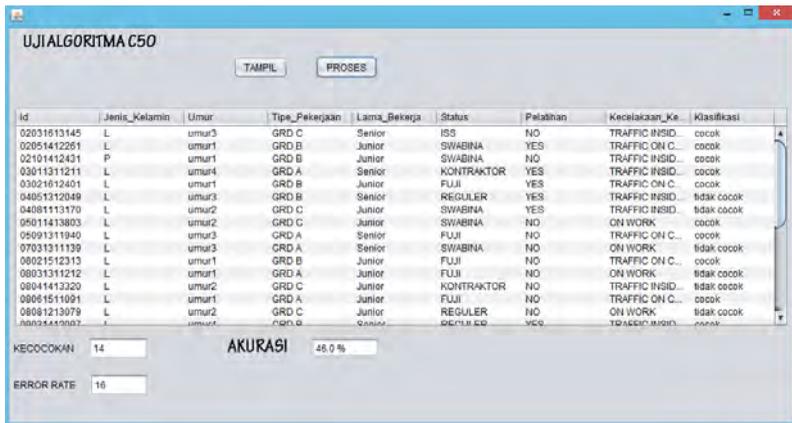
Pola kecelakaan kerja yang dihasilkan sebanyak 34 pola (*rules*). Hasil keakuratan yang ditunjukkan pada *testing* sebesar 45%. Hasil pola kecelakaan kerja selengkapnya disajikan pada lampiran.

Pembagian Data 85% : 15%

Pada pembagian ini data *training* sebanyak 85% dari data hasil *pre-processing* (200 data). Sehingga data *testing* sebanyak 30 data. Hasil dari *training* dan *testing* terdapat pada Gambar 5.17. dan 5.18. di bawah ini.



Gambar 5.17 Hasil *Training* pada C5.0 dengan 85% Data



Gambar 5.18 Hasil *Testing* pada C5.0 dengan 15% Data

Pola kecelakaan kerja yang dihasilkan sebanyak 32 pola (*rules*). Hasil keakuratan yang ditunjukkan pada *testing* sebesar 46%. Hasil pola kecelakaan kerja selengkapnya disajikan pada lampiran.

Pembagian Data 80% : 20%

Pada pembagian ini data *training* sebanyak 80% dari data hasil *pre-processing* (200 data). Sehingga data *testing* sebanyak 40 data. Hasil dari *training* dan *testing* terdapat pada Gambar 5.19. dan 5.20. di bawah ini.



Gambar 5.19 Hasil *Training* pada C5.0 dengan 80% Data

UJIALGORITMA C50

TAMPII PROSES

| Id | Jenis_Kelamin | Umur | Tipe_Pekerjaan | Lama_Bekerja | Status | Pelatihan | Kecelakaan_Ke... | Klasifikasi |
|-------------|---------------|-------|----------------|--------------|------------|-----------|------------------|-------------|
| 02031613145 | L | umur3 | GRD C | Senior | ISS | NO | TRAFFIC INSID... | cocok |
| 02051412261 | L | umur1 | GRD B | Junior | SWABINA | YES | TRAFFIC ON C... | cocok |
| 02101412431 | P | umur1 | GRD B | Junior | SWABINA | NO | TRAFFIC INSID... | cocok |
| 03011311211 | L | umur4 | GRD A | Senior | KONTRAKTOR | YES | TRAFFIC INSID... | cocok |
| 03021612401 | L | umur1 | GRD B | Junior | FUJI | YES | TRAFFIC ON C... | cocok |
| 04051312048 | L | umur3 | GRD B | Senior | REGULER | YES | TRAFFIC INSID... | tidak cocok |
| 04081113170 | L | umur2 | GRD C | Junior | SWABINA | YES | TRAFFIC INSID... | tidak cocok |
| 05011413803 | L | umur2 | GRD C | Junior | SWABINA | NO | ON WORK | cocok |
| 05091311940 | L | umur3 | GRD A | Senior | FUJI | NO | TRAFFIC ON C... | cocok |
| 07031311139 | L | umur3 | GRD A | Senior | SWABINA | NO | ON WORK | tidak cocok |
| 08021512313 | L | umur1 | GRD B | Junior | FUJI | NO | TRAFFIC ON C... | cocok |
| 08031311212 | L | umur1 | GRD A | Junior | FUJI | NO | ON WORK | tidak cocok |
| 08041413320 | L | umur2 | GRD C | Junior | KONTRAKTOR | NO | TRAFFIC INSID... | tidak cocok |
| 08061511091 | L | umur1 | GRD A | Junior | FUJI | NO | TRAFFIC ON C... | tidak cocok |
| 08081213079 | L | umur2 | GRD C | Junior | REGULER | NO | ON WORK | cocok |
| 08091413002 | L | umur4 | GRD B | Senior | REGULER | YES | TRAFFIC INSID... | tidak cocok |

KECOCOKAN 13 AKURASI 32.0 %

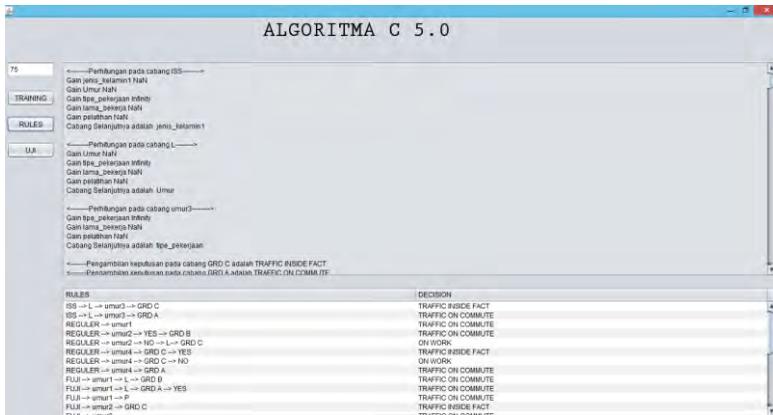
ERROR RATE 27

Gambar 5.20 Hasil *Testing* pada C5.0 dengan 20% Data

Pola kecelakaan kerja yang dihasilkan sebanyak 32 pola (*rules*). Hasil keakuratan yang ditunjukkan pada *testing* sebesar 32%. Hasil pola kecelakaan kerja selengkapnya disajikan pada lampiran.

Pembagian 75% : 25%

Pada pembagian ini data *training* sebanyak 75% dari data hasil *pre-processing* (200 data). Sehingga data *testing* sebanyak 50 data. Hasil dari *training* dan *testing* terdapat pada Gambar 5.21. dan 5.22. di bawah ini.



Gambar 5.21 Hasil *Training* pada C5.0 dengan 75% Data

| Id | Jenis_Kelamin | Umur | Tipe_Pekerjaan | Lama_Bekerja | Status | Pembelahan | Kecelakaan_Ke... | Klasifikasi |
|-------------|---------------|-------|----------------|--------------|------------|------------|------------------|-------------|
| 02031613145 | L | umur3 | GRD C | Senior | ISS | NO | TRAFFIC INSID | cocok |
| 02051412261 | L | umur1 | GRD B | Junior | SWABINA | YES | TRAFFIC ON C... | cocok |
| 02101412431 | P | umur1 | GRD B | Junior | SWABINA | NO | TRAFFIC INSID... | cocok |
| 03011311211 | L | umur4 | GRD A | Senior | KONTRAKTOR | YES | TRAFFIC INSID... | tidak cocok |
| 03021912491 | L | umur1 | GRD B | Junior | FUJI | YES | TRAFFIC ON C... | cocok |
| 04051312049 | L | umur3 | GRD B | Senior | REGULER | YES | TRAFFIC INSID | tidak cocok |
| 04081113170 | L | umur2 | GRD C | Junior | SWABINA | YES | TRAFFIC INSID | tidak cocok |
| 05011413893 | L | umur2 | GRD C | Junior | SWABINA | NO | ON WORK | cocok |
| 05091311940 | L | umur3 | GRD A | Senior | FUJI | NO | TRAFFIC ON C... | cocok |
| 07031311139 | L | umur3 | GRD A | Senior | SWABINA | NO | ON WORK | tidak cocok |
| 08021512313 | L | umur1 | GRD B | Junior | FUJI | NO | TRAFFIC ON C... | cocok |
| 08031311212 | L | umur1 | GRD A | Junior | FUJI | NO | ON WORK | tidak cocok |
| 08041413203 | L | umur2 | GRD C | Junior | KONTRAKTOR | NO | TRAFFIC INSID | tidak cocok |
| 08061511091 | L | umur1 | GRD A | Junior | FUJI | NO | TRAFFIC ON C... | tidak cocok |
| 08081213079 | L | umur2 | GRD C | Junior | REGULER | NO | ON WORK | cocok |
| 09031413007 | L | umur4 | GRD B | Senior | REGULER | YES | TRAFFIC INSID | tidak cocok |

KECOCOKAN: 16 AKURASI: 32.0% ERROR RATE: 34

Gambar 5.22 Hasil *Testing* pada C5.0 dengan 25% Data

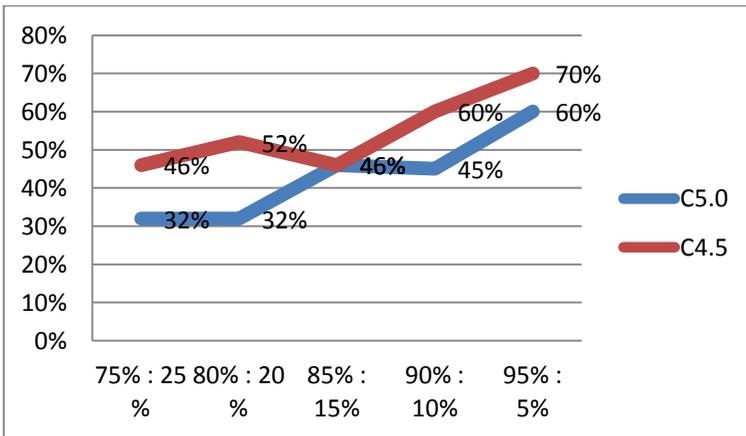
Pola kecelakaan kerja yang dihasilkan sebanyak 27 pola (*rules*). Hasil keakuratan yang ditunjukkan pada *testing* sebesar 32%. Hasil pola kecelakaan kerja selengkapnya disajikan pada lampiran.

Dari hasil *testing* yang telah dilakukan, terlihat bahwa angka keakuratan tertinggi saat pembagian data sebesar 95%. Perbandingan hasil keakuratan, terlihat pada Tabel 5.3

Tabel 5.3. Tabel Hasil Algoritma C5.0

| Pembagian Data | Pola yang Dihasilkan | Hasil |
|----------------|----------------------|-------|
| 95% - 5% | 35 pola | 60% |
| 90% - 10% | 34 pola | 45% |
| 85% - 15% | 32 pola | 46% |
| 80% - 20 % | 32 pola | 32% |
| 75% - 25 % | 27 pola | 32% |

Pada uji coba yang dilakukan pada kedua algoritma, terlihat bahwa prosentasi keakuratan cenderung menurun jika pembagian data pada data *training* lebih diperbesar. Semakin besar nilai prosentasi data *training*, maka prosentasi keakuratan semakin kecil. Perbandingan keakuratan kedua algoritma dapat terlihat pada Gambar 5.23.



Gambar 5.23 Grafik Perbandingan Keakuratan C4.5 dan C5.0

5.4 Analisis Pola yang Ditemukan

Hasil uji coba dengan Algoritma C4.5 dan C5.0 menghasilkan jumlah pola yang berbeda, tergantung pada data *training* yang digunakan. Berikut ini contoh pola yang dihasilkan oleh kedua algoritma dengan pembagian data sebesar 95% : 5%.

Tabel 5.4 Tabel Pola yang dihasilkan oleh Algoritma C4.5

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|---|-------------------------------------|
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'ISS' and 'TIPE-PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' AND 'PELATIHAN' = 'NO' and 'UMUR' = | TRAFFIC ON COMMUTE |

| | |
|---|---------------------|
| 'umur3' and 'TIPE PEKERJAAN' = 'GRD B' | |
| IF 'STATUS' = 'REGULER' AND 'PELATIHAN' = 'NO' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' AND 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD B' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' AND 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur2' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur1' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' | TRAFFIC ON COMMUTE |

| | |
|---|---------------------|
| = 'umur1' and 'PELATIHAN' = 'NO' and 'JK' = 'P' | |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and L | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur3' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur1' | ON WORK |

| | |
|---|---------------------|
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' - 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' - 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' - 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'NO' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |

Tabel 5.5 Tabel Pola yang dihasilkan oleh Algoritma C5.0

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|---|-------------------------------------|
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and L and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' and 'JK' = 'P' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' and 'JK' = 'P' and 'TIPE | TRAFFIC ON COMMUTE |

| | |
|--|---------------------|
| PEKERJAAN' = 'GRD A' | |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur2' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'YES' and 'JK' = 'L' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'YES' and 'JK' = 'L' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'YES' and 'JK' = 'L' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'YES' and 'JK' = 'L' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |

| | |
|---|---------------------|
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD B' and 'JK' = 'L' and 'LAMA BEKERJA' = 'SENIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'UMUR' = 'umur2' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD A' and 'LAMA BEKERJA' = 'JUNIOR' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD A' and 'LAMA BEKERJA' = 'SENIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'L' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'SWABINA' and | TRAFFIC ON COMMUTE |

| | |
|--|---------------------|
| 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' and 'JK' = 'L' and 'LAMA BEKERJA' = 'JUNIOR' | |
| IF 'STATUS' = 'KONTRAKTOR' and umur 1 | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and umur 3 and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and umur 3 and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD A' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and umur 3 and 'JK' = 'L' and 'PELATIHAN' = 'NO' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and umur 4 and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and umur 4 and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |

Pola diatas dihasilkan dari relasi atribut yang membentuk *tree*. Relasi tersebut mulai dari *root* sampai *node* terakhir atau *leaf node*. *Leaf node* merepresentasikan keputusan atau jenis kecelakaan kerja yang terjadi. Salah satu pembacaan pola kecelakaan kerja seperti dibawah ini ;

```
IF "Status" = "REGULER" && "Pelatihan" = "YES" && "Tipe
Pekerjaan" = "GRD B" THEN "Kecelakaan Kerja" = "TRAFFIC
ON COMMUTE"
```

Dari pola yang dihasilkan pada uji coba Algoritma C4.5 dan C5.0, atribut yang paling berpengaruh terhadap terjadinya kecelakaan kerja adalah atribut "Status". Atribut "Status" memiliki nilai *information gain* dan *gain ratio* tertinggi. Nilai tertinggi pada atribut ini dipengaruhi oleh kemunculan atribut pada jumlah kasus yang terjadi. Selain itu, banyaknya *predictor*

pada atribut ini juga mempengaruhi nilai dari *entropy*. Semakin tinggi nilai *entropy* pada suatu atribut, maka semakin berpengaruh atribut tersebut pada pembangunan *tree*. Jadi dapat disimpulkan bahwa atribut “Status ” pada 200 data kecelakaan kerja ini adalah atribut yang paling berpengaruh dalam terjadinya kecelakaan kerja.

Uji coba dengan pembagian data yang lain pada Algoritma C4.5 dan C5.0 secara lengkap pada lampiran.

LAMPIRAN A

Hasil Pola Kecelakaan Kerja yang Didapat dengan Algoritma C4.5

Pola kecelakaan kerja yang dihasilkan pada Algoritma C4.5 merupakan representasi dari *tree* yang dibangun. Pola tersebut merupakan lintasan (*path*) *tree* dari *root* hingga *leaf node*. Jadi dapat disimpulkan bahwa pola kecelakaan kerja yang dihasilkan merupakan relasi-relasi dari atribut yang menemui keputusan pada *node* terakhir.

Relasi atribut merupakan kondisi IF... THEN, yang merupakan aturan dari *decision tree*. Dapat diambil salah satu contoh pola yang dihasilkan yaitu

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (THEN Condition) |
|---|----------------------------|
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |

Cara pembacaan pola pada tabel diatas sebagai berikut :

IF "Status" = "REGULER" && "Pelatihan" = "YES" && "Tipe Pekerjaan" = "GRD B" THEN "Kecelakaan Kerja" = "TRAFFIC ON COMMUTE"

Pola yang dihasilkan pada tiap cabang berbeda karena adanya perhitungan *entropy*. Sehingga banyaknya atribut pada tiap cabang berbeda-beda. Berikut ini disajikan pola kecelakaan kerja yang dihasilkan pada Algoritma C4.5 dengan 5 kali percobaan. Pembagian data yang digunakan berdasarkan Tabel 5.1.

LAMPIRAN A (LANJUTAN)

Pola Kecelakaan Kerja dengan Pembagian Data 90% : 10%

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|---|-------------------------------------|
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur4' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur4' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and | TRAFFIC ON COMMUTE |

LAMPIRAN A (LANJUTAN)

| | |
|---|------------------------|
| 'PELATIHAN' = 'NO' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' | |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD B' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur2' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur1' and | TRAFFIC ON COMMUTE |

LAMPIRAN A (LANJUTAN)

| | |
|--|------------------------|
| 'PELATIHAN' = 'YES' | |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur1' and 'PELATIHAN' = 'NO' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and L | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |

LAMPIRAN A (LANJUTAN)

| | |
|---|------------------------|
| and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' | |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur1' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'NO' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |

LAMPIRAN A (LANJUTAN)

Pola Kecelakaan Kerja dengan Pembagian Data 85% : 15%

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|---|-------------------------------------|
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE | TRAFFIC INSIDE FACT |

LAMPIRAN A (LANJUTAN)

| | |
|---|------------------------|
| PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'YES' | |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur2' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur1' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |

LAMPIRAN A (LANJUTAN)

| | |
|--|------------------------|
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur1' and 'PELATIHAN' = 'NO' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'L' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |

LAMPIRAN A (LANJUTAN)

| | |
|--|------------------------|
| and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' | |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur1' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and GRD A and 'PELATIHAN' = 'YES' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and GRD A and 'PELATIHAN' = 'NO' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |

LAMPIRAN A (LANJUTAN)

Pola Kecelakaan Kerja dengan Pembagian Data 80% : 20%

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|---|-------------------------------------|
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE | TRAFFIC INSIDE FACT |

LAMPIRAN A (LANJUTAN)

| | |
|---|------------------------|
| PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'YES' | |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur2' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur1' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |

LAMPIRAN A (LANJUTAN)

| | |
|--|------------------------|
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur1' and 'PELATIHAN' = 'NO' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'TIPE PEKERJAAN' = 'GRD A' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'L' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |

LAMPIRAN A (LANJUTAN)

| | |
|---|---------------------|
| and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' | |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur1' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and GRD A and 'PELATIHAN' = 'YES' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and GRD A and 'PELATIHAN' = 'NO' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |

LAMPIRAN A (LANJUTAN)

Pola Kecelakaan Kerja dengan Pembagian Data 75% : 25%

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|---|-------------------------------------|
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'ISS' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE | ON WORK |

LAMPIRAN A (LANJUTAN)

| | |
|--|------------------------|
| PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'NO' | |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur1' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur1' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur1' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'NO' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD B' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and | TRAFFIC ON COMMUTE |

LAMPIRAN A (LANJUTAN)

| | |
|--|------------------------|
| 'UMUR' = 'umur3' | |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'L' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur3' and 'PELATIHAN' = 'NO' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |

LAMPIRAN A (LANJUTAN)

| | |
|---|------------------------|
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur1' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'NO' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' | TRAFFIC INSIDE FACT |

LAMPIRAN A (LANJUTAN)

“Halaman ini sengaja dikosongkan”

LAMPIRAN B

Hasil Pola Kecelakaan Kerja yang Didapat Dengan Algoritma C5.0

Pola kecelakaan kerja yang dihasilkan pada Algoritma C5.0 merupakan representasi dari *tree* yang dibangun. Pola tersebut merupakan lintasan (*path*) *tree* dari *root* hingga *leaf node*. Jadi dapat disimpulkan bahwa pola kecelakaan kerja yang dihasilkan merupakan relasi-relasi dari atribut yang menemui keputusan pada *node* terakhir.

Relasi atribut merupakan kondisi IF... THEN, yang merupakan aturan dari *decision tree*. Dapat diambil salah satu contoh pola yang dihasilkan yaitu

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|--|-------------------------------------|
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |

Cara pembacaan pola pada tabel diatas sebagai berikut :

IF "Status" = "REGULER" && "Pelatihan" = "YES" && "Jenis Kelamin" = "L" && "Tipe Pekerjaan" = "GRD B" THEN "Kecelakaan Kerja" = "TRAFFIC ON COMMUTE"

Pola yang dihasilkan pada tiap cabang berbeda karena adanya perhitungan *entropy*. Sehingga banyaknya atribut pada tiap cabang berbeda-beda. Berikut ini disajikan pola kecelakaan kerja yang dihasilkan pada Algoritma C5.0 dengan 5 kali percobaan. Pembagian data yang digunakan berdasarkan Tabel 5.1.

LAMPIRAN B (LANJUTAN)

Pola Kecelakaan Kerja dengan Pembagian Data 90% : 10%

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|---|-------------------------------------|
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD B' and 'LAMA BEKERJA' = 'SENIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' | TRAFFIC INSIDE FACT |

LAMPIRAN B (LANJUTAN)

| | |
|---|------------------------|
| = 'umur2' | |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and GRD A and 'LAMA BEKERJA' = 'SENIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'L' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'SWABINA' and GRD A and 'PELATIHAN' = 'YES' and 'JK' = 'L' and 'LAMA BEKERJA' = 'JUNIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur1' | ON WORK |

LAMPIRAN B (LANJUTAN)

| | |
|---|---------------------|
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and GRD A | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'NO' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'JK' = 'L' and GRD A | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'JK' = | TRAFFIC INSIDE |

LAMPIRAN B (LANJUTAN)

| | |
|---|---------------------|
| 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD C' | FACT |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD B' and 'LAMA BEKERJA' = 'SENIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'PELATIHAN' = 'NO' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur2' | TRAFFIC INSIDE FACT |

Pola Kecelakaan Kerja dengan Pembagian Data 85% : 15%

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|--|-------------------------------------|
| IF 'STATUS' = 'ISS' and 'JK' = 'L' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'ISS' and 'JK' = 'L' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur2' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |

LAMPIRAN B (LANJUTAN)

| | |
|---|------------------------|
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur2' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and | TRAFFIC ON COMMUTE |

LAMPIRAN B (LANJUTAN)

| | |
|--|---------------------|
| 'TIPE PEKERJAAN' = 'GRD B' | |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD A' and 'LAMA BEKERJA' = 'JUNIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'NO' and 'LAMA BEKERJA' = 'JUNIOR' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'JK' = 'L' and 'PELATIHAN' = 'NO' and 'LAMA BEKERJA' = 'SENIOR' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'L' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = | ON WORK |

LAMPIRAN B (LANJUTAN)

| | |
|--|------------------------|
| 'NO' | |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' and 'JK' = 'L' and 'LAMA BEKERJA' = 'JUNIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur1' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD A' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'NO' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |

Pola Kecelakaan Kerja dengan Pembagian Data 80% : 20%

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|--|--|
| IF 'STATUS' = 'ISS' and 'JK' = 'L' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'ISS' and 'JK' = 'L' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |

LAMPIRAN B (LANJUTAN)

| | |
|---|------------------------|
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' and 'JK' = 'L' and 'LAMA BEKERJA' = 'JUNIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD C' and 'UMUR' = 'umur4' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur2' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'PELATIHAN' = 'NO' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |

LAMPIRAN B (LANJUTAN)

| | |
|---|------------------------|
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur1' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur1' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'L' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |

LAMPIRAN B (LANJUTAN)

| | |
|--|------------------------|
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD A' and 'JK' = 'L' and 'LAMA BEKERJA' = 'JUNIOR' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur1' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD A' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'NO' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |

LAMPIRAN B (LANJUTAN)

Pola Kecelakaan Kerja dengan Pembagian Data 75% : 25%

| Pola Kecelakaan Kerja (<i>IF Condition</i>) | Keputusan (<i>THEN Condition</i>) |
|---|-------------------------------------|
| IF 'STATUS' = 'ISS' and 'JK' = 'L' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'ISS' and 'JK' = 'L' and 'UMUR' = 'umur3' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur1' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur2' and 'PELATIHAN' = 'NO' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'YES' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD C' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'REGULER' and 'UMUR' = 'umur4' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur1' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' | TRAFFIC ON |

LAMPIRAN B (LANJUTAN)

| | |
|---|------------------------|
| = 'umur1' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' | COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur2' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur3' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'FUJI' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'L' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur1' and 'JK' = 'P' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur2' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD B' and 'UMUR' = 'umur4' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'UMUR' = 'umur4' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD C' and 'JK' = 'L' and 'PELATIHAN' = 'NO' | ON WORK |
| IF 'STATUS' = 'SWABINA' and 'TIPE PEKERJAAN' = 'GRD A' and 'PELATIHAN' = 'YES' and 'JK' = | TRAFFIC ON COMMUTE |

LAMPIRAN B (LANJUTAN)

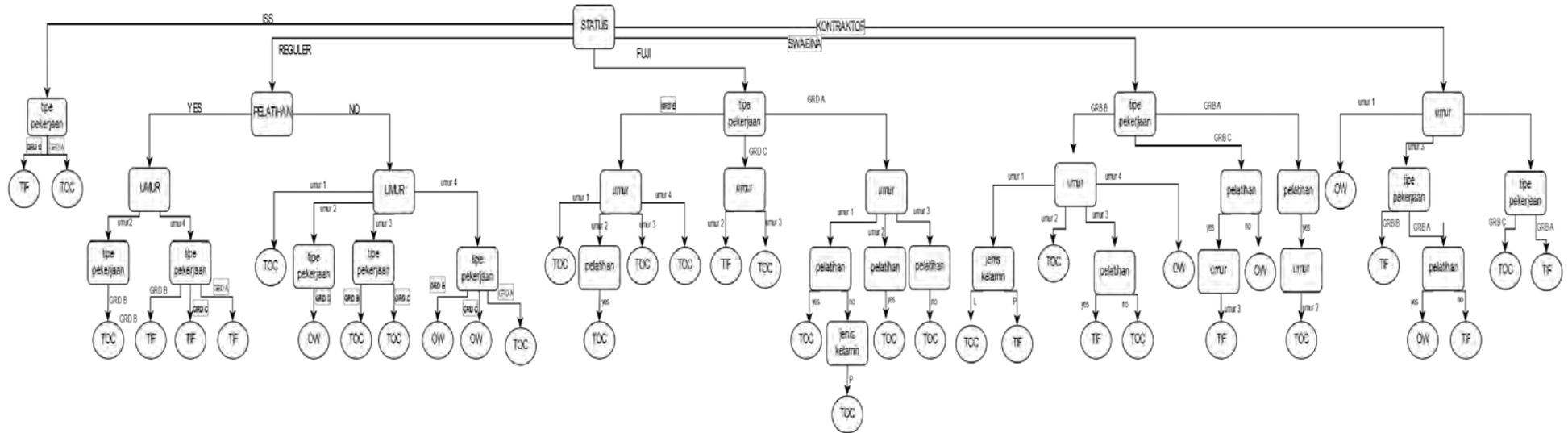
| | |
|--|------------------------|
| 'L' and 'LAMA BEKERJA' = 'JUNIOR' | |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur1' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD A' | ON WORK |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur3' and 'JK' = 'L' and 'PELATIHAN' = 'YES' and 'TIPE PEKERJAAN' = 'GRD A' | TRAFFIC INSIDE FACT |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD C' | TRAFFIC ON COMMUTE |
| IF 'STATUS' = 'KONTRAKTOR' and 'UMUR' = 'umur4' and 'JK' = 'L' and 'TIPE PEKERJAAN' = 'GRD B' | TRAFFIC INSIDE FACT |

LAMPIRAN C

Hasil Tree Pada Tahap Training

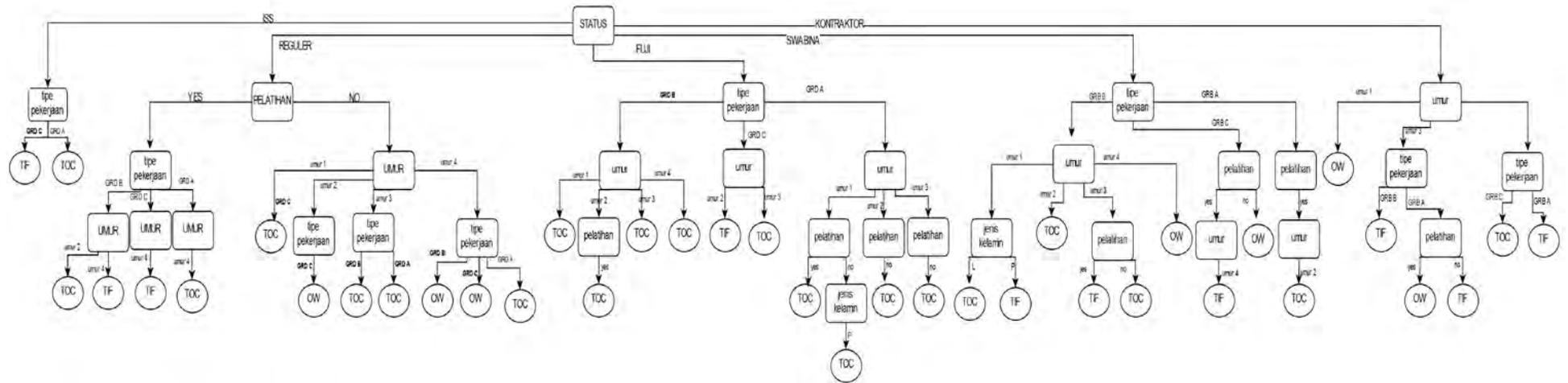
Pada lampiran ini, akan disajikan hasil *training* dalam bentuk *tree*. Pada lampiran A dan B, pola yang ditampilkan merupakan representasi dari lintasan *tree* dari *root* hingga *leaf node*. *Tree* yang akan disajikan merupakan hasil uji coba pada Algoritma C4.5 dan C5.0 dengan pembagian data *training* dan *testing* sesuai dengan Tabel 5.1.

Berikut ini beberapa *tree* dari uji coba yang telah dilakukan pada Tugas Akhir ini.

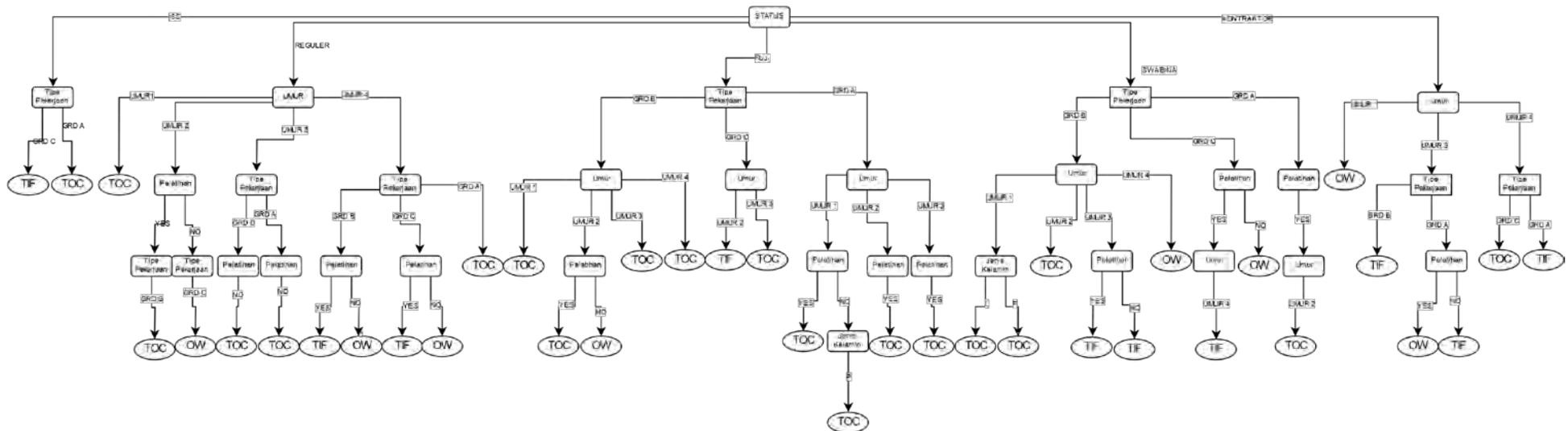


Tree diatas merupakan hasil uji coba pada Algoritma C4.5 dengan pembagian data *training* dan *testing* sebesar 95%:5%

LAMPIRAN C

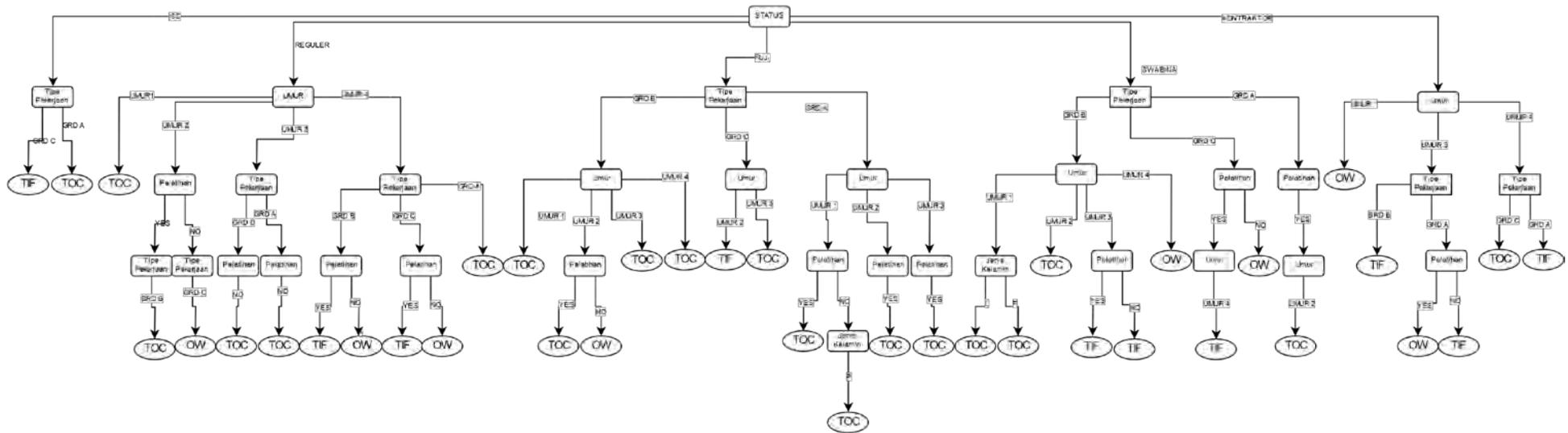


Tree diatas merupakan hasil uji coba pada Algoritma C4.5 dengan pembagian data *training* dan *testing* sebesar 90%:10%



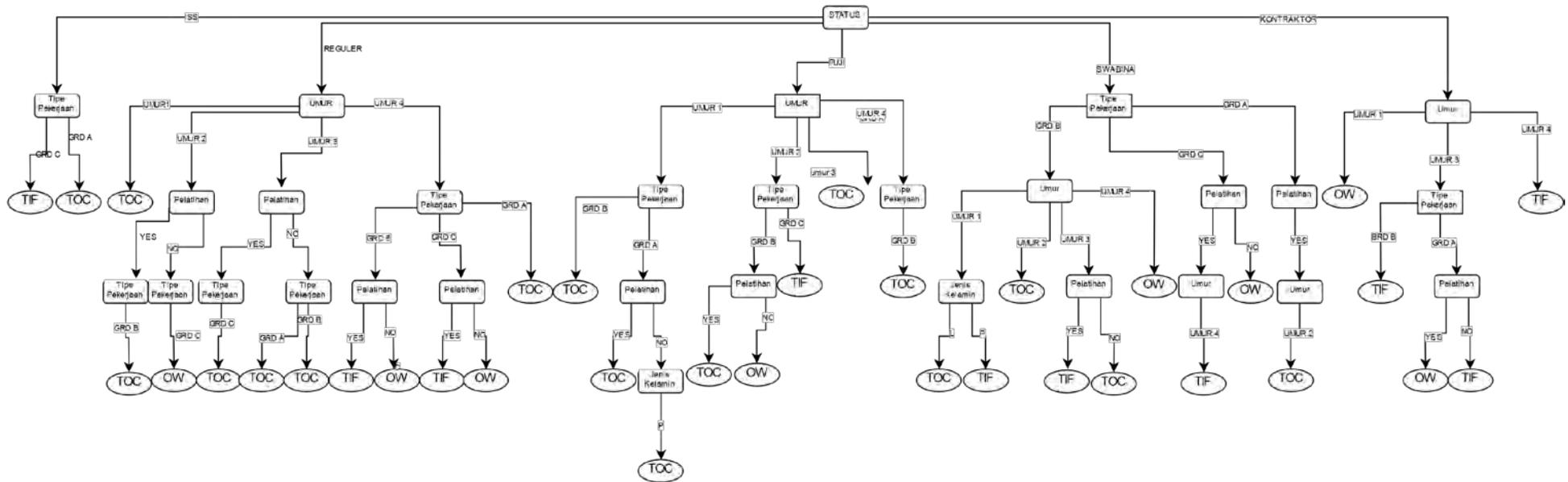
Tree diatas merupakan hasil uji coba pada Algoritma C4.5 dengan pembagian data *training* dan *testing* sebesar 85% : 15%

LAMPIRAN C



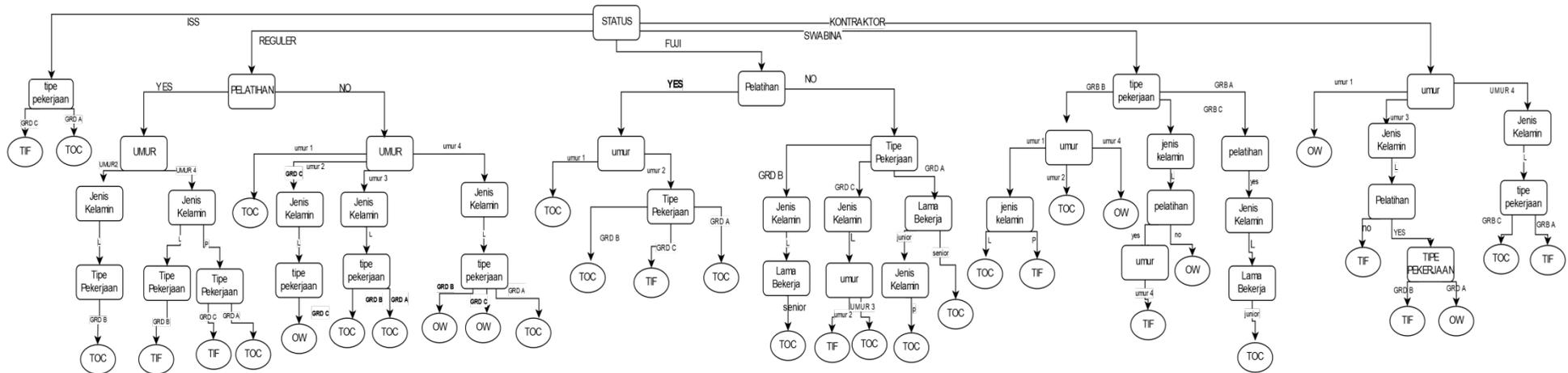
Tree diatas merupakan hasil uji coba pada Algoritma C4.5 dengan pembagian data *training* dan *testing* sebesar 80% : 20%

LAMPIRAN C

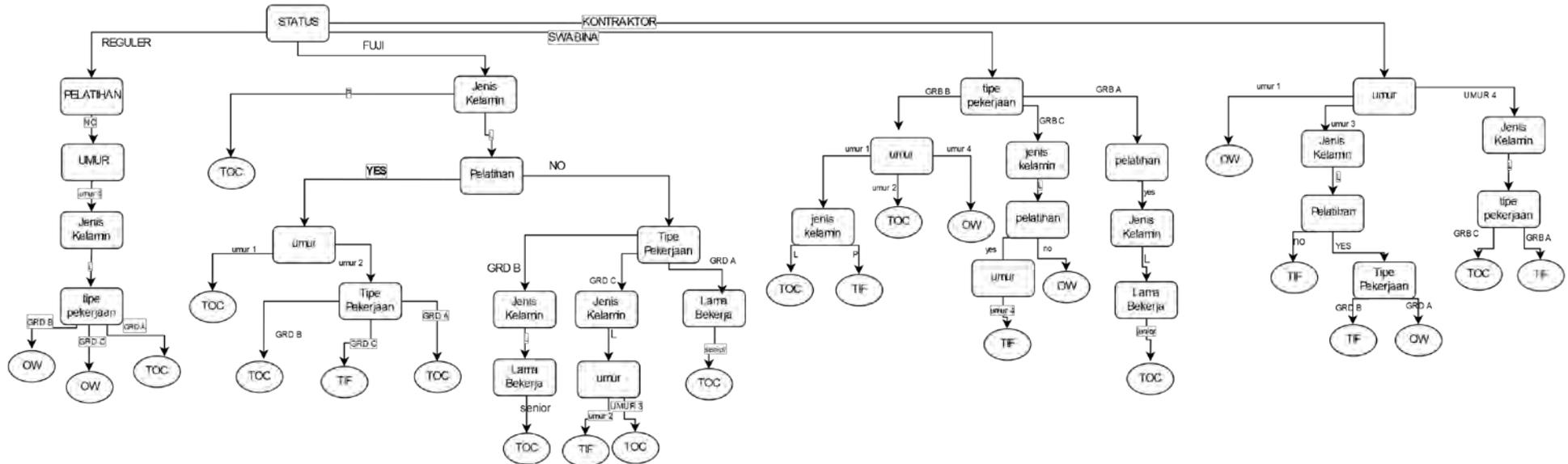


Tree diatas merupakan hasil uji coba pada Algoritma C4.5 dengan pembagian data *training* dan *testing* sebesar 75% : 25%

LAMPIRAN C

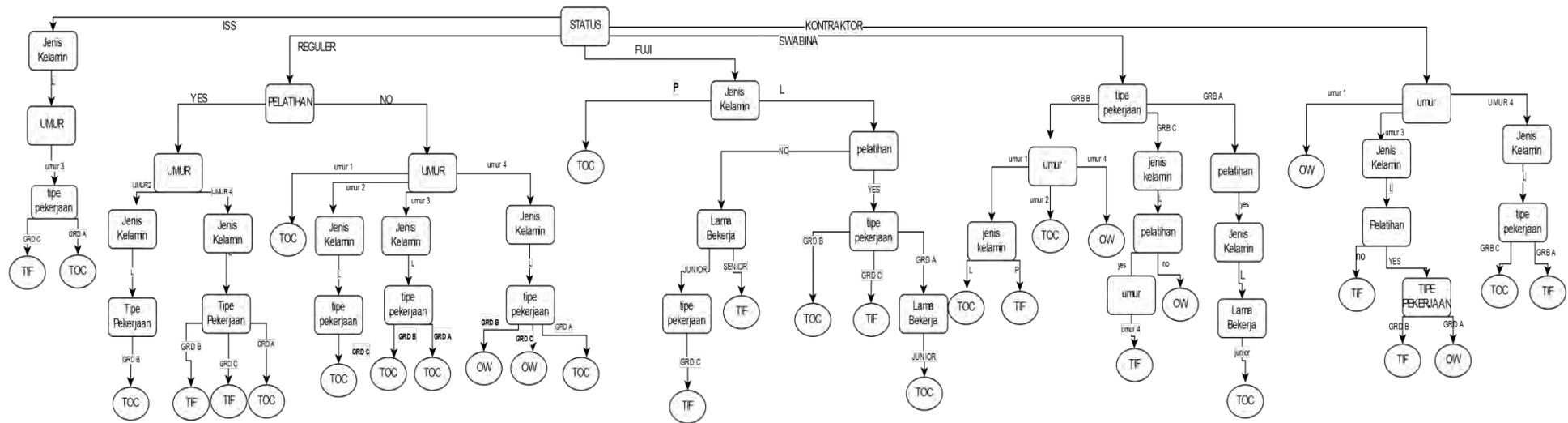


Tree diatas merupakan hasil uji coba pada Algoritma C5.0 dengan pembagian data *training* dan *testing* sebesar 95% : 5%



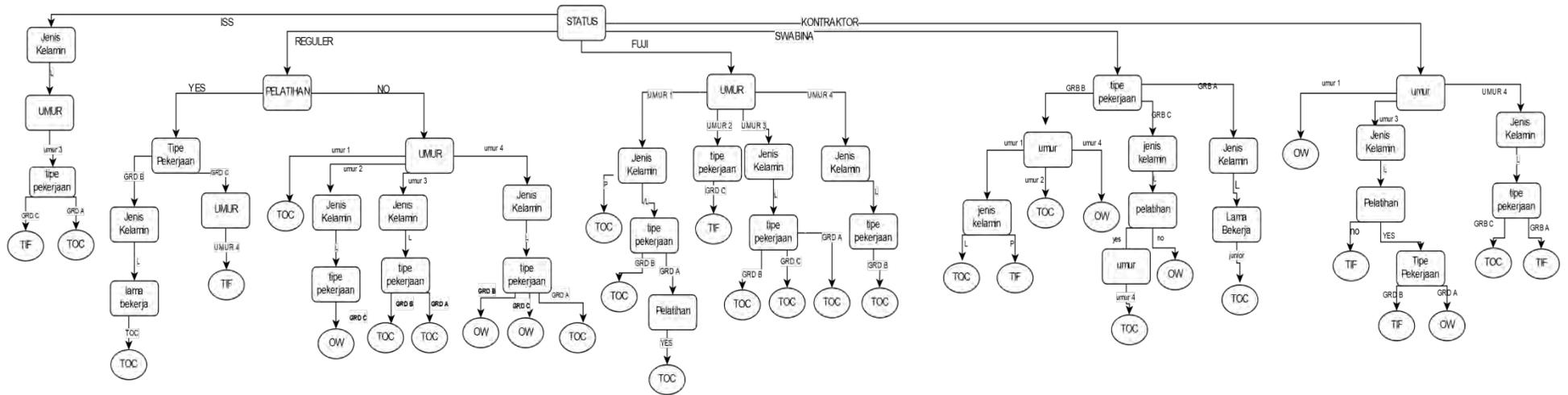
Tree diatas merupakan hasil uji coba pada Algoritma C5.0 dengan pembagian data *training* dan *testing* sebesar 90% : 10%

LAMPIRAN C

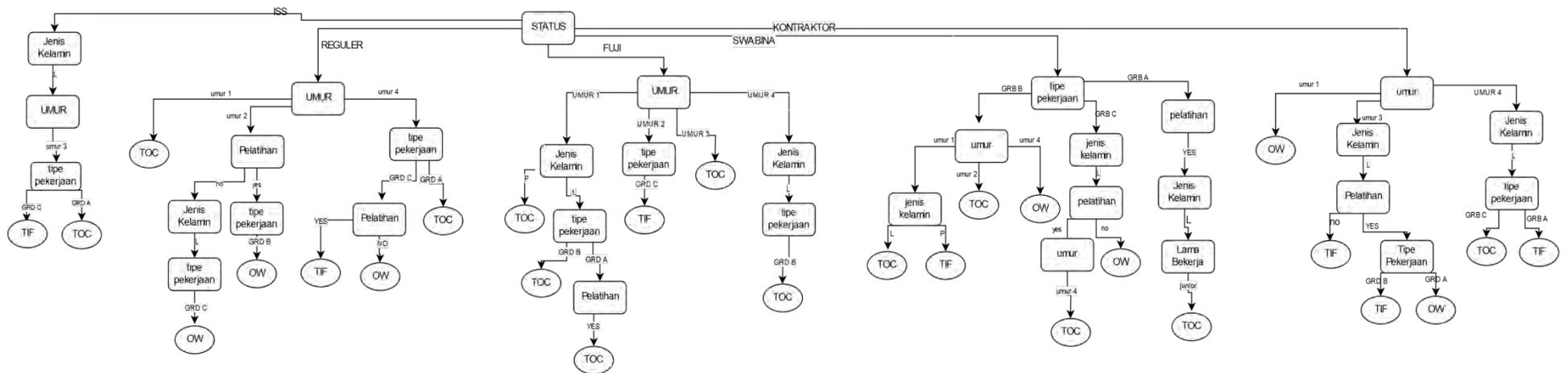


Tree diatas merupakan hasil uji coba pada Algoritma C5.0 dengan pembagian data *training* dan *testing* sebesar 85% : 15%

LAMPIRAN C



Tree diatas merupakan hasil uji coba pada Algoritma C5.0 dengan pembagian data *training* dan *testing* sebesar 80% : 20%



Tree diatas merupakan hasil uji coba pada Algoritma C5.0 dengan pembagian data *training* dan *testing* sebesar 75% : 25 %

BAB VI

KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan yang dihasilkan berdasarkan penelitian yang telah dilaksanakan. Di samping itu, pada bab ini juga dimasukkan beberapa saran yang dapat digunakan jika penelitian ini ingin dikembangkan.

6.1 Kesimpulan

Berdasarkan analisis terhadap hasil pengujian yang telah dilakukan terhadap Algoritma C4.5 dan C5.0 dengan 200 data kecelakaan kerja, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Perangkat lunak dapat melakukan *pre-processing*, pembangunan *tree* dan pengujian algoritma sehingga didapatkan pola prediksi kecelakaan kerja beserta angka keakuratannya.
2. *Rules* sebagai representasi dari pola kecelakaan kerja yang dihasilkan oleh Algoritma C5.0 lebih ringkas daripada *rules* yang dihasilkan oleh C4.5, yaitu 32 pola kecelakaan kerja.
3. Hasil keakuratan pola yang dihasilkan oleh Algoritma C4.5 lebih tinggi dibanding dengan Algoritma C5.0, yaitu 70%.

6.2 Saran

Berdasarkan hasil yang dicapai pada Tugas Akhir ini, ada beberapa hal yang penulis sarankan untuk pengembangan selanjutnya yaitu:

1. Data yang dipakai dalam perangkat lunak ini masih sedikit dan bervariasi. Dalam perhitungan *data mining*, sebaiknya data harus dalam jumlah yang banyak karena mempengaruhi pola dan hasil keakuratan yang didapatkan. Untuk penelitian selanjutnya dapat menggunakan data yang lebih banyak.
2. Perangkat lunak ini belum bisa menampilkan hasil *decision tree* yang menarik. Untuk penelitian selanjutnya dapat menggunakan representasi data dalam bentuk *tree* atau *list*

rules yang lebih baik sehingga memudahkan dalam testing data.

3. Pengembangan perangkat lunak ini dapat dilakukan dengan menambah menu input data baru demi manfaat yang lebih baik.

DAFTAR PUSTAKA

- [1] Bujlow, Tomasz, Tahir Riaz dkk. (2012). “A Method for Classification of Network Traffic on C5.0 Machine Learning Algorithm”. **ICNC’12 : 2012 International Conference on Computing, Networking and Communication (ICNC)**. (pp 237-241). IEEE
- [2] Dunham, MH. (2003). “Data Mining Introductory and Advanced Topics”. **Upper Saddle River, NJ : Pearson Education, Inc**
- [3] Gorunescu, Florin. (2011). “Data Mining: Concepts, Models, and Techniques”. **Verlag Berlin Heidelberg. Springer**
- [4] Han, J, Micheline K, Jian Pei. (2012). “Data Mining Concept and Techniques 3nd Edition”. **Morgan Kaufman Publisher. USA**
- [5] I. H. Witten and E. Frank. (2005). “Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition”. **Morgan Kaufmann, San Francisco**
- [6] Kantardzic, M. (2003). “Data Mining Concept Models, Methods, and Algorithm”. **New Jersey, USA : A John Wiley & Sons**
- [7] Moore, Samuel. A, Daniel M. D Addario, dkk. (2011). “Are Decision Trees Always Greener on the Open (Source) Side of the Fence?”. **Sciweaver**
- [8] Permatasari, Anggun. (2009). “Investigasi Kecelakaan Penyabrangan Perlintasan KRL UI-Margonda Depok”. **Tugas Akhir Fakultas Kesehatan Masyarakat. Universitas Indonesia**
- [9] Polat, Kemal, Gunes, Salih. (2009). “A Novel Hybrid Intelligent Method Based on C4.5 Decision Tree Classifier and One-Against-All Approach for Multi-Class Classification Problem”. **Elsevier**. (pp 1587-1592).

- [10] Quinlan, J. R. (1993). "C4.5: Programs for Machine Learning. San Mateo" **CA : Morgan Kaufmann**
- [11] Rivas, T, M. Paz, dkk. (2011). "Explaining and Predicting Workplace Accident Using Data-Mining Techniques". **Elsivier**. (pp 739-747).
- [12] Tan, Pang-Ning, Michael Steinbach, Vipin Kumar. (2006). "Introduction of Data Mining". **Boston : Pearson – Addison Wesley**
- [13] Wu, Xindong, Kumar, Vipin, dkk. (2008). "Top 10 Algorithms in Data Mining". **Knowl Inf System (2008) 14 : 1-37. Springer-Verlag London Limited**

BIODATA PENULIS



Penulis memiliki nama lengkap Yolanda Rizkita Putri, lahir di Mojokerto pada tanggal 15 Januari 1994. Penulis bertempat tinggal di Ds. Pohkecik No. 135 RT.04/RW.01 Kec. Dlanggu, Kab. Mojokerto. Pendidikan formal yang pernah ditempuh yaitu SD Negeri Pohkecik I, SMP Negeri 1 Kota Mojokerto, dan SMAN 1 Sooko Mojokerto. Kemudian, penulis melanjutkan studi di jurusan Matematika ITS, dengan bidang minat ilmu komputer.

Dalam bidang minat ini penulis mulai mengenal bahasa pemrograman diantaranya adalah C, C++, Java, PHP-MySQL, dan MATLAB. Semasa menempuh jenjang pendidikan S-1, penulis juga aktif dalam kegiatan non-akademis diantaranya aktif di organisasi kemahasiswaan Matematika ITS, Badan Eksekutif Mahasiswa ITS dan mengikuti beberapa kepanitiaan acara besar yang ada di ITS. Selama penulisan Tugas Akhir ini Penulis tidak lepas dari kekurangan, untuk itu penulis mengharapkan kritik, saran, dan pertanyaan mengenai Tugas Akhir ini yang dapat dikirimkan melalui *e-mail* ke yolanda.riezkieta@gmail.com.