

**ANALISIS TINGKAT KECELAKAAN LALU LINTAS DI KLATEN DENGAN
MENGUNAKAN PERBANDINGAN 3 METODE DATA MINING**



**Disusun sebagai salah satu syarat menyelesaikan Program Studi Strata I
Pada Jurusan Informatika Fakultas Komunikasi dan Informatika**

Oleh:

PUTRI IKE NURMAWATI

L 200 140 119

**PROGRAM STUDI INFORMATIKA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

2018

HALAMAN PERSETUJUAN

**ANALISIS TINGKAT KECELAKAAN LALU LINTAS DI KLATEN
DENGAN MENGGUNAKAN PERBANDINGAN 3 METODE DATA MINING**

PUBLIKASI ILMIAH

oleh:

PUTRI IKE NURMAWATI

L 200 140 119

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing



Nurgiyatna, S.T., M.Sc., Ph.D

NIK. 881

HALAMAN PENGESAHAN

ANALISIS TINGKAT KECELAKAAN LALU LINTAS DI KLATEN DENGAN MENGUNAKAN PERBANDINGAN 3 METODE DATA MINING

OLEH
PUTRI IKE NURMAWATI
L 200 140 119

Telah dipertahankan di depan Dewan Penguji
Fakultas Komunikasi dan Informatika
Universitas Muhammadiyah Surakarta
Pada hari Senin, 05 November 2018
dan dinyatakan telah memenuhi syarat

Dewan Penguji:

1. Nurgiyatna, S.T., M.Sc., Ph.D

(Ketua Dewan Penguji)

2. Fatah Yasin Irsyad, S.T., M.T.

(Anggota I Dewan Penguji)

3. Devi Afriyanti Puspa P, S.Kom., M.Sc. (.....)

(Anggota II Dewan Penguji)

Publikasi ilmiah ini telah diterima sebagai salah satu persyaratan

Untuk memperoleh gelar sarjana

Tanggal

Mengetahui,


Dekan
Fakultas Komunikasi dan Informatika
Nurgiyatna, S.T., M.Sc., Ph.D
NIK: 831


Ketua Program Studi
Informatika
Dr. Heru Supriyono, M.Sc.
NIK: 970

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam naskah publikasi ini tidak terdapat karya yang diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya diatas, maka akan saya pertanggungjawabkan sepenuhnya.

Surakarta, 05 November 2018

Penulis



PUTRI IKE NURMAWATI

L 200 140 119



**UNIVERSITAS MUHAMMADIYAH SURAKARTA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: informatika@ums.ac.id

SURAT KETERANGAN LULUS PLAGIASI

No Surat 482/A.3-11.3/INF-FKI/XI/2018

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : Putri Ike Nurmawati
NIM : **L200140119**
Judul : **ANALISIS TINGKAT KECELAKAAN LALU LINTAS DI
KLATEN DENGAN MENGGUNAKAN PERBANDINGAN 3
METODE DATA MINING**
Program Studi : Informatika
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 12 Nopember 2018

Biro Skripsi Informatika

Ihsan Cahyo Utomo, S.Kom., M.Kom.



UNIVERSITAS MUHAMMADIYAH SURAKARTA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: informatika@ums.ac.id

https://ev.tumain.com/ups/carta/en_us/fo=1037287077&w=1&lang=en_us&u=1057550080

feedback studio ANALISIS TINGKAT KECELAKAAN LALU LINTAS DI KLATEN DENGAN MENGGUNAKAN PERBANDINGAN 3 METODE DA... /0 64 of 96

Preparing download...

ANALISIS TINGKAT KECELAKAAN LALU LINTAS DI KLATEN DENGAN MENGGUNAKAN PERBANDINGAN 3 METODE DATA MINING

Abstrak

Kecelakaan adalah kejadian yang melibatkan kendaraan dengan sesama pemakai jalan maupun tanpa pemakai jalan secara disengaja ataupun tidak disengaja yang berdampak pada korban jiwa dan kerugian secara materi, angka kecelakaan semakin meningkat. Data mining merupakan ekspansi data yang berjumlah besar yang mampu memberikan informasi baru terhadap suatu data. Penelitian ini dilakukan teknik *data mining* dengan tujuan untuk mengetahui tingkat kecelakaan lalu lintas di Kota Klaten memberikan informasi mengenai faktor pemicu kecelakaan kepada masyarakat sebagai bahan pertimbangan dan pengawasan khususnya pihak kepolisian sebagai tolak ukur mengurangi angka kecelakaan. Metode yang akan digunakan dalam penelitian ini diantaranya *Information Gain*, *Index Gini*, dan *Gain Ratio* dengan memakai atribut Usia, Jenis Kelamin, Waktu, Jumlah Kendaraan, Jumlah Korban, Tipe Jalan, Cuaca, Bentuk Jalan dan Situasi. Hasil kesimpulan dari penelitian ini, didapatkan nilai *accuracy*, *precision*, *recall*. Metode *Information Gain* dan *Index Gini* lebih unggul dengan nilai *accuracy* sebesar 70,94% daripada *Gain Ratio* 69,25%, metode *Index Gini* lebih akurat dibandingkan metode lain karena menurut nilai *accuracy*, *precision* dan *recall* mempunyai nilai yang lebih tinggi.

Kata Kunci: Data Mining, Index Gini, Information Gain, Gain Ratio, Kecelakaan Lalu Lintas

Page: 1 of 15 Word Count: 2752 Text-only Report High Resolution On

Match Overview

21%

1	Submitted to Universita...	8%
2	eprints.ums.ac.id	4%
3	media.neliti.com	2%
4	Submitted to Chandra...	1%
5	waset.org	1%
6	repository.ums.ac.id	1%
7	eprints.ums.ac.id	1%

ANALISIS TINGKAT KECELAKAAN LALU LINTAS DI KLATEN DENGAN MENGUNAKAN PERBANDINGAN 3 METODE DATA MINING

Abstrak

Kecelakaan adalah kejadian yang melibatkan kendaraan dengan sesama pemakai jalan maupun tanpa pemakai jalan secara disengaja ataupun tidak disengaja yang berdampak pada korban jiwa dan kerugian secara materi, angka kecelakaan semakin meningkat. Data *mining* merupakan eksplorasi data yang berjumlah besar yang mampu memberikan informasi baru terhadap suatu data. Penelitian ini dilakukan teknik *data mining* dengan tujuan untuk mengetahui tingkat kecelakaan lalu lintas di Kota Klaten memberikan Informasi mengenai faktor pemicu kecelakaan kepada masyarakat sebagai bahan pertimbangan dan pengawasan khususnya pihak kepolisian sebagai tolak ukur mengurangi angka kecelakaan.. Metode yang akan digunakan dalam penelitian ini diantaranya *Information Gain*, *Index Gini*, dan *Gain Ratio* dengan memakai atribut Usia, Jenis Kelamin, Waktu, Jumlah Kendaraan, Jumlah Korban, Tipe Jalan, Cuaca, Bentuk Jalan dan Situasi. Hasil kesimpulan dari penelitian ini, didapatkan nilai *accuracy*, *precision*, *recall*. Metode *Information Gain* dan *Index Gini* lebih unggul dengan nilai *accuracy* sebesar 70.94% daripada *Gain Ratio* 69.25%, metode *Index Gini* lebih akurat dibandingkan metode lain karena menurut nilai *accuracy*, *precision* dan *recall* mempunyai nilai yang lebih tinggi.

Kata Kunci: *Data Mining*, *Index Gini*, *Information Gain*, *Gain Ratio*, *Kecelakaan Lalu Lintas*

Abstract

Accidents are events involving vehicles with road users or road users intentionally or unintentionally which have resulted in loss of life and material loss, the number of accidents is increasing. Data mining is an exploration of large amounts of data capable of providing new information to a data. This research conducted data mining techniques with the aim to determine the level of traffic accidents in the City of Klaten providing information about the factors that trigger accidents to the public as a matter of consideration and supervision, especially the police as a benchmark to reduce accident rates. The methods to be used in this study include *Information Gain*, *Gini Index*, and *Gain Ratio* using the Age attribute, Gender, Time, Number of Vehicles, Number of Victims, Type of Road, Weather, Form of Road and Situation. The results of the conclusions from this study, obtained values of *accuracy*, *precision*, *recall*. The *Information Gain* and *Gini Index* method is superior with *accuracy* value of 70.94% rather than *Gain Ratio* of 69.25%, the *Index Gini* method is more accurate than other methods because according to the values of *accuracy*, *precision* and *recall* have a higher value.

Keywords: *Data Mining*, *Index Gini*, *Information Gain*, *Gain Ratio*, *Traffic Accident*

1. PENDAHULUAN

Teknologi informasi semakin berkembang cepat mengikuti perkembangan jaman yang pesat seperti saat ini, sehingga berdampak terhadap kebutuhan informasi yang tepat, cepat dan akurat. Pemanfaatan teknologi informasi menyebar luas dalam berbagai bidang salah satunya pada bidang kepolisian mengingat manfaat serta peran pentingnya dalam mengolah sebuah informasi

baru, dalam pasal 93 Peraturan Pemerintah Nomor 43 tahun 1993 ayat 1 yang menjelaskan tentang definisi kecelakaan lalu lintas yaitu sebuah kejadian yang tidak disengaja yang dialami oleh sesama pemakai jalan maupun tidak yang menimbulkan korban jiwa mulai dari luka ringan hingga luka berat atau bahkan menyebabkan kehilangan nyawa seseorang atau biasa disebut dengan meninggal dunia selain itu mengakibatkan kerugian materi.

Terdapat 3400 orang meninggal dunia pada setiap harinya menurut data *World Health Organization*. Berdasarkan data dari Satuan Lalu Lintas Polisi Resort Klaten (Satlantas Polres Klaten) mencatat setiap hari selalu terjadi kecelakaan lalu lintas. Pada tahun 2016 terhitung dari bulan Januari hingga bulan September terdapat 608 kasus kecelakaan lalu lintas di Kota Klaten yang menyebabkan 127 orang kehilangan nyawa, tercatat orang yang mengalami luka berat sebanyak 2 orang, dan sebanyak 1049 orang mengalami luka ringan. Kecelakaan tersebut mengakibatkan kerugian material sebesar Rp. 1.068.250.000,-. Pada tahun 2017 terhitung dari bulan Januari hingga bulan September 2017 terdapat 470 kasus kecelakaan yang menyebabkan korban yang mengalami luka ringan tercatat 808 kasus dengan keterangan meninggal dunia tercatat 129 orang. Rata-rata korban yang mengalami kecelakaan adalah berusia 16 sampai 25 tahun.

Akibat dari kecelakaan lalu lintas maka dibutuhkan penanganan sungguh-sungguh, karena mengingat angka kecelakaan dari tahun ke tahun semakin meningkat dan mengakibatkan kerugian yang cukup besar sehingga kasus ini termasuk salah satu permasalahan yang serius. Informasi yang akurat mengenai faktor yang memicu terjadinya kecelakaan lalu lintas sangat dibutuhkan oleh masyarakat sebagai bahan pertimbangan dan pengawasan khususnya pihak kepolisian dengan tujuan untuk mengurangi angka kecelakaan (Arumsari, dkk, 2016).

Berdasarkan latar belakang tersebut, maka dalam penelitian ini akan dilakukan proses *data mining* untuk mengetahui tingkat kecelakaan lalu lintas di Kota Klaten dengan harapan dapat mengurangi angka kecelakaan lalu lintas yang dari tahun ke tahun semakin meningkat. *Data mining* adalah eksplorasi dari analisa secara otomatis mengenai data-data yang berjumlah besar untuk menemukan pola maupun aturan yang berarti, dengan tujuan *data mining* tersebut dapat membantu memberikan informasi baru yang penting dari data tersebut. (Ahmed, Kawsar dan Tasnuba Jesmin, 2014). Metode yang digunakan dalam penelitian ini adalah *Information gain*, *Index Gini* dan *Gain Ratio*. Metode *Information Gain* merupakan metode yang digunakan

untuk membentuk pola pohon keputusan dengan memakai teori dari *entropy* informasi dimana mencari nilai *gain* tertinggi untuk mempersempit atribut yang dipilih dan mengevaluasi nilai atribut dengan cara mengukur informasi (Ayyapan G, dkk, 2017). Metode *Index Gini* adalah metode yang mempunyai operasi matematis paling sederhana dan digunakan untuk pengukuran *split* dimana mencari nilai *split* terkecil (Mandyartha, 2015). Metode yang ketiga metode *Gain Ratio* adalah metode yang dimodifikasi dari metode *Information Gain* untuk mengurangi bias terhadap atribut multi-nilai dengan mempertimbangkan jumlah dan ukuran cabang saat menemukan sebuah atribut (Defiyanti, 2008). Ketiga metode tersebut dianalisa untuk dibandingkan dengan tujuan mengetahui dan membuktikan metode yang paling baik, unggul dan akurat melalui perbandingan nilai *accuracy*, *precession* dan *recall*.

2. METODE

2.1 Penentuan Atribut

Proses data mining ini membutuhkan atribut yang sesuai dengan tujuan dari penelitian ini, sebagai berikut :

Tabel 1. Atribut yang digunakan

Y	Tingkat Kecelakaan
X1	Usia
X2	Jenis Kelamin
X3	Waktu
X4	Jumlah Kendaraan
X5	Jumlah Korban
X6	Tipe Jalan
X7	Cuaca
X8	Bentuk Jalan
X9	Situasi

2.2 Pengumpulan Data

Penelitian ini membutuhkan data yang harus diolah untuk dapat menghasilkan sebuah informasi, dengan menggunakan metode berdasarkan data kecelakaan lalu lintas dari Satuan Lalu Lintas Resort Klaten (Polres Klaten). Serta melakukan wawancara secara langsung kepada pihak kepolisian. Data yang digunakan dalam penelitian ini adalah data set kecelakaan lalu lintas di Kota Klaten dan sekitarnya, data tersebut digunakan sebagai data *training* dalam penelitian ini.

2.2.1 Data Training

Data *training* adalah kumpulan dari beberapa data *sample* atau data pelatihan yang akan digunakan untuk memprediksi kelas data baru yang belum pernah ada sehingga menemukan sebuah pola (Patel, Bhaskar N ,dkk 2012). Data *training* yang digunakan dalam penelitian ini adalah data set kecelakaan lalu lintas yang diperoleh dari Satuan Lalu Lintas Resort Klaten (Polres Klaten). Jumlah data dalam penelitian ini ada 296 data.

2.3 Analisis Data

Tahap analisis data melakukan pemilihan data yang sesuai dengan atribut yang telah ditentukan, pengolahan data untuk diproses kedalam tabel-tabel untuk dikelompokkan dan melakukan analisis sesuai dengan metode yang ditentukan. Persamaan ini nantinya akan digunakan untuk mengetahui tingkat kecelakaan lalu lintas melalui nilai dari variabel Y yang dipengaruhi dari variabel X. Analisis *data mining* mempunyai bagian-bagian yang harus dilakukan untuk mencapai hasil yang sesuai dengan tujuan yang telah diinginkan.

2.4 Implementasi Data Mining

2.4.1 Penggunaan Metode Information Gain

Ayyapan, G, dkk (2017) menyatakan rumus *information gain* terdapat pada persamaan 1,2

Rumus Entropy :

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

Rumus *information gain* :

$$Gain (S, A) = Entropi (S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi (S_i) \quad (2)$$

Keterangan :

S : himpunan kasus

Pi : proporsi Si terhadap S

Si :jumlah sampel untuk nilai v

2.4.2 Penggunaan Metode Decision Tree Algoritma Index Gini

Dai,Qing-yun, dkk (2015) rumus *index gini* dinyatakan pada persamaan 3

Rumus gini Split :

$$Gini(A) = 1 - \sum_{k=1}^C P_k^2 \quad (3)$$

Keterangan

K : Kelas atribut

C : Jumlah kelas variabel Y

P_k : Proporsi jumlah kelas dalam atribut K terhadap kelas dalam variabel Y

2.4.3 Penggunaan Metode *Gain Ratio*

Defiyanti (2008) menyatakan rumus *gain ratio* terdapat pada persamaan 4,5

Rumus *Gain Ratio* :

$$Gain Ratio(S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \quad (4)$$

Rumus Split info (x):

$$SplitInfo(S,A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (5)$$

Keterangan :

S : data sampel yang digunakan untuk training

A : atribut

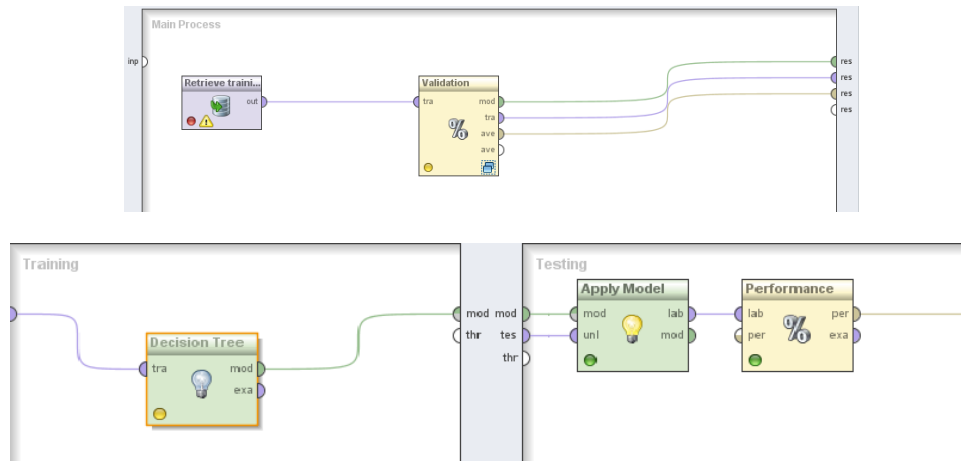
Gain(S,A): information gain pada atribut A

SplitInfo(S,A): split information pada atribut A

3. HASIL DAN PEMBAHASAN

Penelitian yang telah dilakukan menghasilkan perbandingan metode *Index Gini*, *Information Gain*, dan *Gain Ratio* untuk mengetahui tingkat kecelakaan lalu lintas di Kota Klaten menggunakan semua atribut yang telah ditentukan dalam penelitian ini.

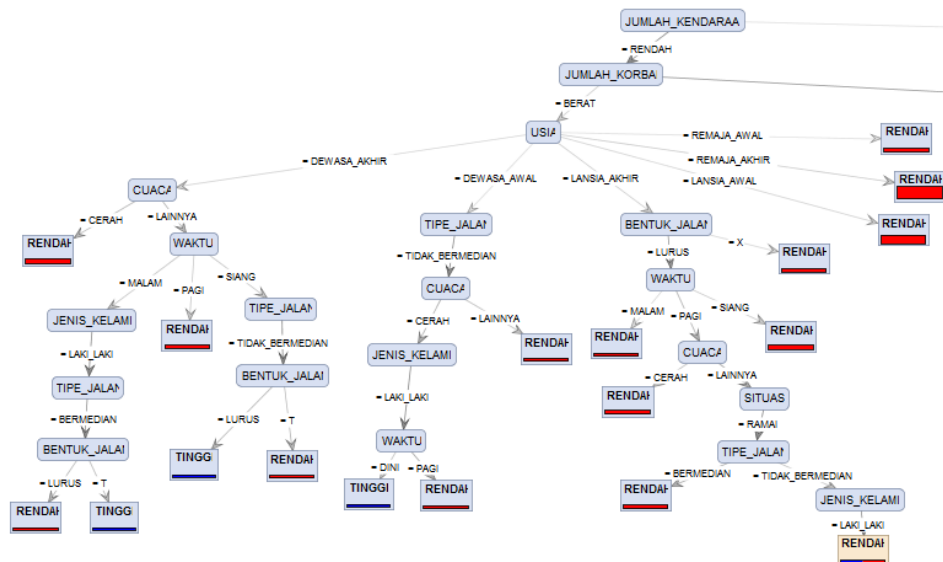
Berikut adalah hasil dari penelitian yang telah dilakukan:



Gambar 1. Rancangan Proses Penerapan *Decision Tree*

a. Hasil Implementasi *Decision Tree* Algoritma *Index Gini* dengan *Rapid Miner 5*

Proses *data mining* metode *Index Gini* menggunakan *Rapid Miner 5* menghasilkan bentuk pohon keputusan ditunjukkan pada gambar 3.

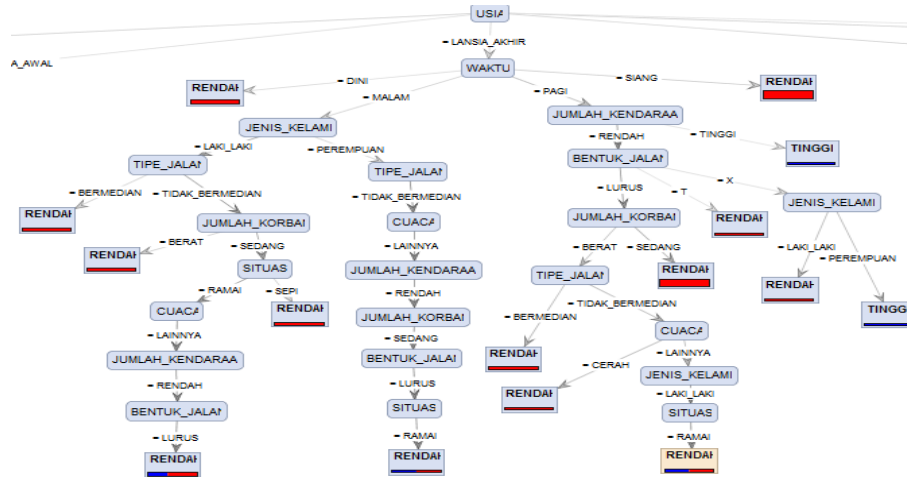


Gambar 2. Hasil Pohon Keputusan Algoritma *Index Gini*

Pada gambar 2 dapat diambil sebuah kesimpulan bahwa Jumlah Kendaraan merupakan variabel yang paling mempengaruhi tingkat kecelakaan, dikarenakan Jumlah Kendaraan menempati posisi *root node* sebagai simpul akar dari sebuah pohon keputusan, dengan dilengkapi oleh variabel lain yang dapat mempengaruhi tingkat kecelakaan.

b. Hasil Implementasi *Decision Tree* Algoritma *Information Gain* dengan *Rapid Miner 5*

Pengolahan *data mining* algoritma *Information Gain* yang telah diproses menggunakan *Rapid Miner 5* menghasilkan pola pohon keputusan yang ditunjukkan pada gambar 3.

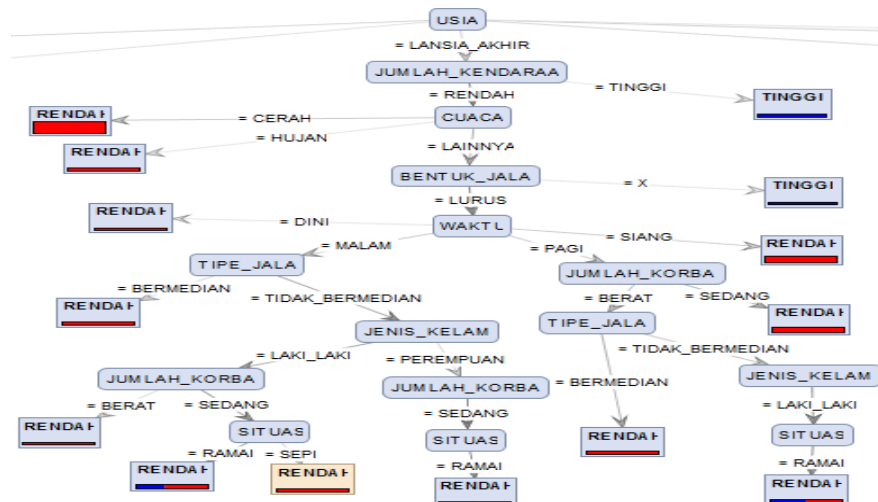


Gambar 3. Hasil Pohon Keputusan Algoritma *Information Gain*

Hasil dari gambar 3 dapat disimpulkan bahwa variabel Waktu disebut sebagai *root node* yaitu variabel yang paling berpengaruh dalam sebuah bentuk pohon keputusan, dilengkapi dengan variabel lainnya yang juga mempengaruhi tingkat.

3.3 Hasil Implementasi Metode *Decision Tree* Algoritma *Gain Ratio* dengan *Rapid Miner 5*

Hasil dari proses data mining algoritma *Gain Ratio* dengan *Rapid Miner 5* ditunjukkan pada gambar 4



Gambar 4. Hasil Pohon Keputusan Algoritma *Gain Ratio*

Hasil yang ditunjukkan pada gambar 4 dapat disimpulkan bahwa variabel Usia disebut sebagai *root node* dengan diikuti Variabel Jumlah Kendaraan sebagai internal node pertama.

3.4 Interpretasi *Index Gini*

Hasil yang diperoleh bahwa Jumlah Kendaraan adalah sebagai *root node* maka terbentuk sebuah aturan yang ada dalam *decision tree Index Gini* sebagai berikut :

```
JUMLAH_KENDARAAN = RENDAH
|
|   JUMLAH_KORBAN = BERAT
|   |   USIA = DEWASA_AKHIR
|   |   |   CUACA = CERAH: RENDAH {TINGGI=0, RENDAH=6}
|   |   |   CUACA = LAINNYA
|   |   |   |   WAKTU = MALAM
|   |   |   |   |   JENIS_KELAMIN = LAKI_LAKI
|   |   |   |   |   |   TIPE_JALAN = BERMEDIAN
|   |   |   |   |   |   |   BENTUK_JALAN = LURUS: RENDAH {TINGGI=0, RENDAH=1}
|   |   |   |   |   |   |   BENTUK_JALAN = T: TINGGI {TINGGI=1, RENDAH=0}
|   |   |   |   |   |   WAKTU = PAGI: RENDAH {TINGGI=0, RENDAH=2}
|   |   |   |   |   |   WAKTU = SIANG
|   |   |   |   |   |   |   TIPE_JALAN = TIDAK_BERMEDIAN
|   |   |   |   |   |   |   BENTUK_JALAN = LURUS: TINGGI {TINGGI=1, RENDAH=0}
|   |   |   |   |   |   |   BENTUK_JALAN = T: RENDAH {TINGGI=0, RENDAH=1}
|   |   |   |   |   |   USIA = DEWASA_AWAL
|   |   |   |   |   |   |   TIPE_JALAN = TIDAK_BERMEDIAN
|   |   |   |   |   |   |   CUACA = CERAH
|   |   |   |   |   |   |   |   JENIS_KELAMIN = LAKI_LAKI
|   |   |   |   |   |   |   |   WAKTU = DINI: TINGGI {TINGGI=1, RENDAH=0}
```

Gambar 5. Hasil Aturan Algoritma *Index Gini*

c. Interpretasi *Information Gain*

Hasil yang menunjukkan bahwa variabel Usia sebagai *root node*, sehingga membentuk sebuah aturan *decision tree* yang ditunjukkan pada gambar 6.

```
USIA = DEWASA_AKHIR
|
|   JENIS_KELAMIN = LAKI_LAKI
|   |   WAKTU = DINI: RENDAH {TINGGI=0, RENDAH=5}
|   |   WAKTU = MALAM
|   |   |   JUMLAH_KORBAN = BERAT
|   |   |   |   BENTUK_JALAN = LURUS: RENDAH {TINGGI=0, RENDAH=3}
|   |   |   |   BENTUK_JALAN = T: TINGGI {TINGGI=1, RENDAH=0}
|   |   |   |   JUMLAH_KORBAN = SEDANG: RENDAH {TINGGI=0, RENDAH=12}
|   |   |   WAKTU = PAGI
|   |   |   |   TIPE_JALAN = BERMEDIAN
|   |   |   |   |   BENTUK_JALAN = LURUS
|   |   |   |   |   |   CUACA = CERAH
|   |   |   |   |   |   |   JUMLAH_KORBAN = BERAT: RENDAH {TINGGI=0, RENDAH=1}
|   |   |   |   |   |   |   JUMLAH_KORBAN = SEDANG
|   |   |   |   |   |   |   |   JUMLAH_KENDARAAN = RENDAH
|   |   |   |   |   |   |   |   |   SITUASI = RAMAI: RENDAH {TINGGI=1, RENDAH=2}
|   |   |   |   |   |   |   |   CUACA = LAINNYA: RENDAH {TINGGI=0, RENDAH=3}
|   |   |   |   |   |   |   BENTUK_JALAN = X
|   |   |   |   |   |   |   CUACA = LAINNYA
|   |   |   |   |   |   |   |   JUMLAH_KENDARAAN = RENDAH
```

Gambar 6. Hasil Aturan Algoritma *Information Gain*

d. Interpretasi *Gain Ratio*

Hasil proses yang menunjukkan bahwa variabel Usia adalah *root node*, maka aturan pohon keputusan yang terbentuk sebagai berikut :

```

USIA = DEWASA_AKHIR
|  JENIS_KELAMIN = LAKI_LAKI
|  |  WAKTU = DINI: RENDAH {TINGGI=0, RENDAH=5}
|  |  WAKTU = MALAM
|  |  |  JUMLAH_KORBAN = BERAT
|  |  |  |  BENTUK_JALAN = LURUS: RENDAH {TINGGI=0, RENDAH=3}
|  |  |  |  BENTUK_JALAN = T: TINGGI {TINGGI=1, RENDAH=0}
|  |  |  |  JUMLAH_KORBAN = SEDANG: RENDAH {TINGGI=0, RENDAH=12}
|  |  WAKTU = PAGI
|  |  |  TIPE_JALAN = BERMEDIAN
|  |  |  |  BENTUK_JALAN = LURUS
|  |  |  |  |  CUACA = CERAH
|  |  |  |  |  |  JUMLAH_KORBAN = BERAT: RENDAH {TINGGI=0, RENDAH=1}
|  |  |  |  |  |  JUMLAH_KORBAN = SEDANG
|  |  |  |  |  |  JUMLAH_KENDARAAN = RENDAH
|  |  |  |  |  |  |  SITUASI = RAMAI: RENDAH {TINGGI=1, RENDAH=2}
|  |  |  |  |  |  |  CUACA = LAINNYA: RENDAH {TINGGI=0, RENDAH=3}
|  |  |  |  |  BENTUK_JALAN = X
|  |  |  |  |  CUACA = LAINNYA
|  |  |  |  |  |  JUMLAH_KENDARAAN = RENDAH

```

Gambar 7. Hasil Aturan Algoritma *Gain Ratio*

e. Implementasi Perhitungan *Decision Tree* Algoritma *Index Gini*

i. Menentukan *root node* (simpul akar)

Tabel 2. Hasil perhitungan penentuan *root node*

Atribut	Nilai <i>gini split</i>
X1= Usia	0.25409
X2= Jenis Kelamin	0.25766
X3= Tipe Jalan	0.25713
X4= Cuaca	0.25764
X5= Jumlah Kendaraan	0.24252
X6= Jumlah Korban	0.25265
X7= Waktu	0.25516
X8= Bentuk Jalan	0.25721
X9= Situasi	0.25614

Tabel 2 membuktikan nilai *gini split* terkecil adalah Jumlah Kendaraan sehingga atribut Jumlah Kendaraan menempati posisi *root node*.

Fakta menunjukkan :

$$\begin{aligned}\text{Gini (s) tinggi} &= 1 - \left(\frac{5}{7}\right)^2 + \left(\frac{2}{7}\right)^2 \\ &= 0.4081\end{aligned}$$

$$\begin{aligned}\text{Gini (s) rendah} &= 1 - \left(\frac{40}{289}\right)^2 + \left(\frac{249}{289}\right)^2 \\ &= 0.2385\end{aligned}$$

$$\begin{aligned}\text{Split} &= \left(\frac{7}{296}\right) * 0.4081 + \left(\frac{289}{296}\right) * 0.2385 \\ &= 0.24252\end{aligned}$$

f. Implementasi Perhitungan *Decision Tree* Algoritma *Information Gain*

i. Menentukan *root node* (simpul akar)

Tabel 3. Hasil perhitungan penentuan root node

Atribut	Nilai <i>Gain</i>
X1= Usia	0.0105
X2= Jenis Kelamin	0.00046
X3= Tipe Jalan	0.00201
X4= Cuaca	0.00053
X5= Jumlah Kendaraan	0.02814
X6= Jumlah Korban	0.0172
X7= Waktu	0.00745
X8= Bentuk Jalan	0.00165
X9= Situasi	0.00417

Tabel 11, membuktikan nilai *gain* tertinggi adalah Jumlah Kendaraan sehingga atribut Jumlah Kendaraan menempati posisi root node.

Fakta menunjukkan :

$$\begin{aligned}\text{Entropy (S)} &= - \left(\frac{45}{296}\right) \log_2 \left(\frac{45}{296}\right) - \left(\frac{251}{296}\right) \log_2 \left(\frac{251}{296}\right) \\ &= 0.61489\end{aligned}$$

$$\text{Entropy (tinggi)} = - \left(\frac{5}{7}\right) \log_2 \left(\frac{5}{7}\right) - \left(\frac{2}{7}\right) \log_2 \left(\frac{2}{7}\right)$$

$$= 0.86312$$

$$\begin{aligned}\text{Entropy (rendah)} &= -\left(\frac{40}{289}\right) \log_2 \left(\frac{40}{289}\right) - \left(\frac{249}{289}\right) \log_2 \left(\frac{249}{289}\right) \\ &= 0.58006\end{aligned}$$

$$\begin{aligned}\text{Gain (S,A)} &= 0.61489 - \left(\frac{7}{296}\right) * 0.86312 + \left(\frac{289}{296}\right) * 0.58006 \\ &= \mathbf{0.02814}\end{aligned}$$

g. Implementasi Perhitungan *Decision Tree* Algoritma *Gain Ratio*

i. Menentukan *root node* (simpul akar)

Tabel 4. Hasil perhitungan penentuan *root node*

Atribut	Nilai <i>Gain</i>
X1= Usia	0.00437
X2= Jenis Kelamin	0.0006
X3= Tipe Jalan	0.002278
X4= Cuaca	0.00051
X5= Jumlah Kendaraan	0.17428
X6= Jumlah Korban	0.02212
X7= Waktu	0.00398
X8= Bentuk Jalan	0.06548
X9= Situasi	0.01026

Tabel 4, membuktikan atribut Jumlah Kendaraan sebagai *root node* karena memiliki nilai *gain* tertinggi.

Fakta menunjukkan :

$$\begin{aligned}\text{Split Info (S,A)} &= -\left(\frac{7}{296}\right) \log_2 \left(\frac{7}{296}\right) + \left(\frac{289}{296}\right) \log_2 \left(\frac{289}{296}\right) \\ &= 0.16146\end{aligned}$$

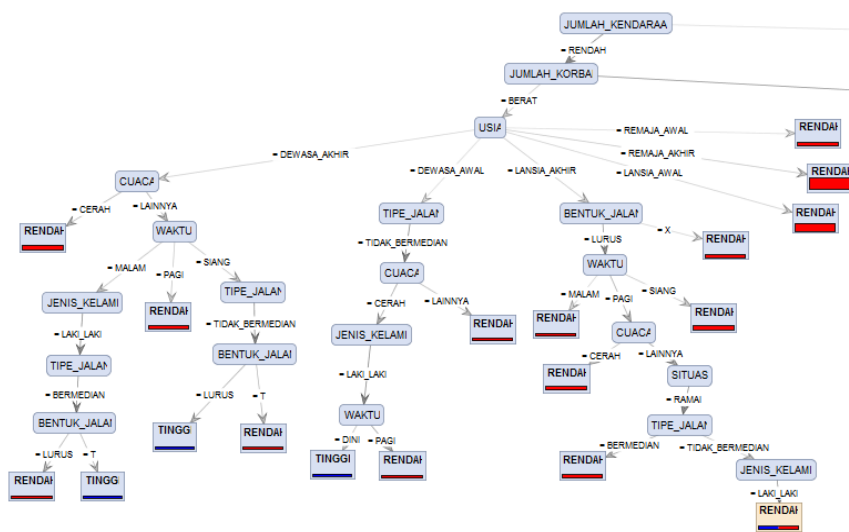
$$\begin{aligned}\text{Gain Ratio (S,A)} &= \text{Gain (S,A)} - \text{Split Info (S,A)} \\ &= 0.02814 - 0.16146 = \mathbf{0.17428}\end{aligned}$$

h. Hasil Penelitian

Pada hasil penelitian ini, melakukan perhitungan untuk mengetahui nilai dari masing-masing metode kemudian melakukan perbandingan dengan hasil dari perhitungan *Rapid Miner* dan setelah melakukan analisis terhadap semua metode berdasarkan nilai *Accuracy*, *Precision*, dan *Recall* untuk mengetahui keakuratan dari metode *Gini Index*, *Gain Ratio*, dan *Information Gain*.

3.10. Perbedaan Hasil Implementasi Algoritma *Index Gini*

Perbandingan yang dilakukan setelah melakukan perhitungan menghasilkan perbedaan. Perbedaan hasil *Index Gini* Jumlah Korban “rendah” yang dihasilkan dari analisis implementasi *rapid Miner* dengan implementasi perhitungan sebagai berikut :



Gambar 8. Hasil rapid miner

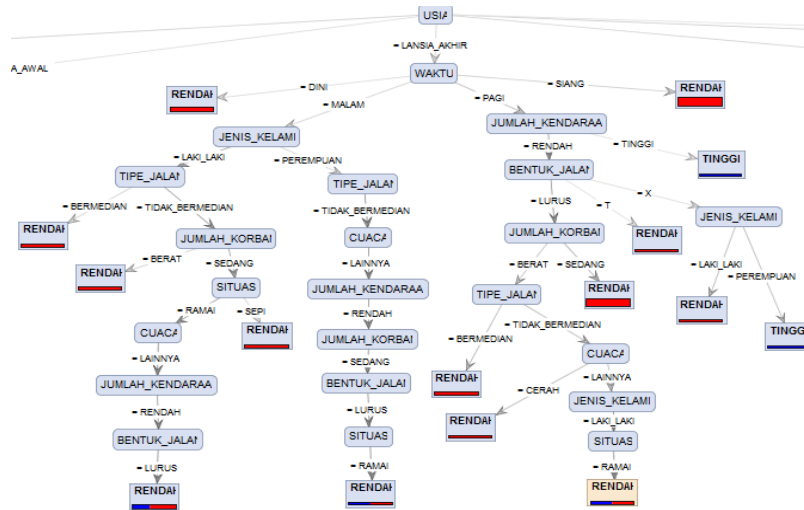


Gambar 9. Hasil perhitungan

Berdasarkan perbedaan hasil implementasi pada gambar 9 dan 10 menunjukkan bahwa posisi internal node ketiga sampai posisi leaf node jumlah korban “rendah” antara hasil *rapid miner* dengan hasil perhitungan tidak sama.

i. Perbedaan hasil Implementasi Algoritma *Information Gain*

Perbandingan yang dilakukan setelah melakukan perhitungan menghasilkan perbedaan. Perbedaan hasil *Information Gain* Jumlah Korban “rendah” yang dihasilkan dari analisis implementasi *rapid Miner* dengan implementasi perhitungan ditunjukkan pada gambar 10 dan 11.



Gambar 10. Hasil rapid miner

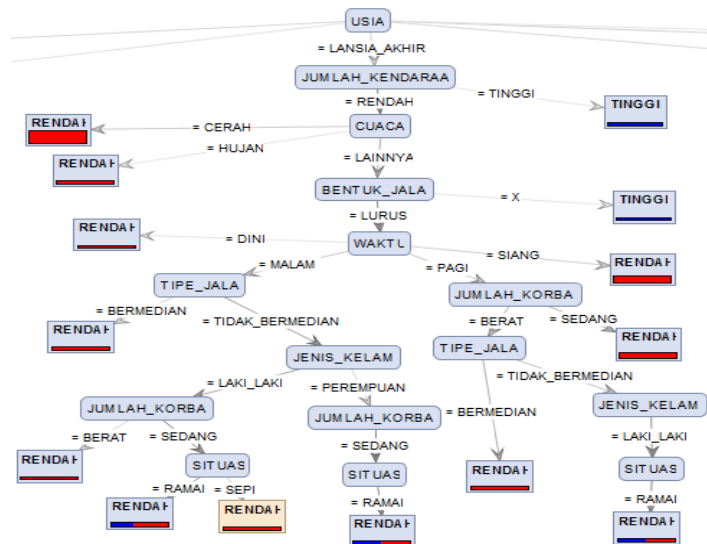


Gambar 11. Hasil perhitungan

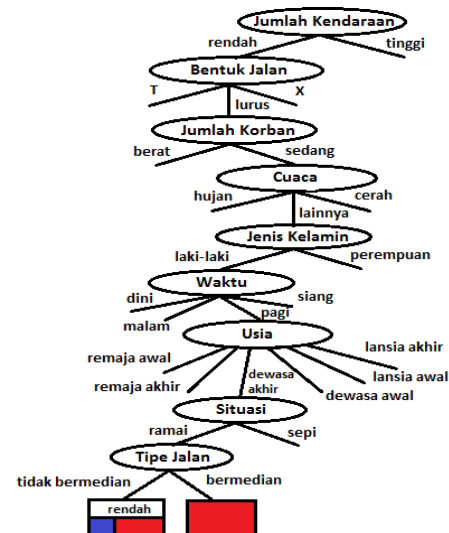
Berdasarkan perbedaan hasil implementasi pada gambar 10 dan 11 menunjukkan bahwa posisi *root node* sampai posisi *leaf node* antara hasil *rapid miner* dengan hasil perhitungan tidak sama.

ii. Perbedaan hasil Implementasi Algoritma Gain Ratio

Perbandingan yang dilakukan setelah melakukan perhitungan menghasilkan perbedaan. Perbedaan hasil Gain Ratio Jumlah Korban “rendah” yang dihasilkan dari analisis implementasi Rapid Miner dengan implementasi perhitungan ditunjukkan pada gambar 12 dan 13.



Gambar 12. Hasil rapid miner



Gambar 13. Hasil perhitungan

Berdasarkan perbedaan hasil implementasi pada gambar 11 dan 12 menunjukkan bahwa posisi *root node* sampai posisi *leaf node* antara hasil *rapid miner* dengan hasil perhitungan tidak sama.

Musthafa (2015) sebuah model klasifikasi menunjukkan hasil benar dan salah maka disebut *confusion matrix* dengan menggunakan perhitungan *Accuracy*, *Precesion*, dan *Recall* ditunjukkan pada tabel 5.

Tabel 5. *Confusion Matrix*

Komponen	Pengertian	Rumus
Accuracy	Tingkat kedekatan antara prediksi dengan actual	$TP / (TP+FP) \times 100\%$
Precision	Tingkat ketepatan antara informasi dan hasil sistem	$TP / (TP+FN) \times 100\%$
Recall	Tingkat keberhasilan sistem menemukan informasi	$(TP+TN)/total\ sampel \times 100\%$

Tabel 6. Hasil perbandingan

Komponen	<i>Gini Index</i>	<i>Gain Ratio</i>	<i>Information Gain</i>
Accuracy	70.94%	69.25%	70.94%
Precision	19.40%	14.28%	15.25%
Recall	28.88%	20.00%	20.00%

Hasil perbandingan yang terdapat pada tabel 6 maka dapat disimpulkan bahwa menurut dari nilai *Accuracy* metode *Information Gain* dan *Index Gini* lebih unggul dengan nilai sebesar 70.94%, menurut nilai *Precision* metode *Index Gini* lebih baik dari metode lain dan menurut nilai *Recall* metode *Information Gain* dan *Gain Ratio* memiliki nilai lebih tinggi sebesar 20.00%.

4. PENUTUP

Kesimpulan yang dapat diambil dari hasil penelitian ini adalah :

1. Jumlah Korban adalah disebut sebagai posisi *root node* dikarenakan pada perhitungan dari semua metode variabel Jumlah Korban selalu berada diposisi paling atas, yang artinya merupakan variabel yang paling berpengaruh dalam pola pohon keputusan dipenelitian ini.
2. Menurut nilai *accuracy* metode *Information Gain*, *Index Gini* lebih tepat digunakan pada penelitian ini karena mempunyai nilai sama sebesar 70.94%, berdasarkan nilai *precision* metode *Information Gain* lebih baik dibandingkan dengan metode lain karena memiliki nilai *precesion* paling tinggi yaitu sebesar 19.40%, berdasarkan nilai *Recall* metode *Index*

Gini memiliki nilai paling tinggi yaitu sebesar 28.88%, namun pada metode *Gain Ratio* dan *Information Gain* memiliki nilai yang sama.

3. Adanya perbedaan hasil implementasi rasi miner dengan perhitungan pada ketiga metode yaitu *Index Gini*, *Information Gain*, dan *Gain Ratio*.

DAFTAR PUSTAKA

- Ahmed, Kawsar; Tasnuba Jesmin. 2014 .“*Comparative Analysis of Data Mining Classification Algorithms in Type-2 Diabetes Prediction Data Using WEKA Approach*.” International Jurnal. Sci Eng, Vol.7(2):155-160, October
- Arumsari, Nanda Dewi; Arief Laila Nugraha; Moehammad Awaluddin. 2016. “*Pemodelan Daerah Rawan Kecelakaan Dengan Menggunakan Cluster Analysis*”. Jurnal Geodesi Undip
- Ayyappan, G; C. Nalini; A. Kumaravel. 2017. “*Efficient Mining For Social Networks Using Information Gain Ratio Based On Academic Dataset*”. IJCIET, Vol. 8, Issue 1, January 2017.
- Dai, Qing-yun; Chun-ping Zhang; Hao Wu.2016. “ *Research of decision tree classification algorithm in data mining*”. International journal of database theory and application. Vol .9,No. 5 pp1-8
- Defiyanti, Sofi. 2008. ”*Perbandingan Kinerja Algoritma ID3 dan C4.5 Dalam Klasifikasi SPAM-MAIL*”. Skripsi. Sistem Informasi. Universitas Gunadarma
- Desphande, S. P.; Thakare, V. M.. 2010. “*Data Mining System And Application : A Review*”. IJDPS, Vol. 1, No. 1, September 2010.
- Mandiyartha; Kurniawan, Eka Prakarsa; Perdana, Muchammad; Setya, Rizal. 2015. “*Identifikasi Sel Darah Merah Bertumpuk Menggunakan Pohon Keputusan Fuzzy Berbasis Gini Index*”. Jurnal Buana Informatika, Vol. 6, No. 1.
- Patel, Bhaskar N.; Satish G. Prajapati ; Dr. Kamaljit I. Lakhtaria. 2012. “*Efficient Classification of Data Using Decision Tree*”. International Jurnal of Data Mining, Vol.2, No. 1, March