

## Algoritma CPAR untuk Analisa Data Kecelakaan

(Studi pada Kepolisian Daerah Sulawesi Tenggara)

Natalis Ransi<sup>\*1</sup>, Edi Winarko<sup>2</sup>

<sup>1</sup>Program Studi S2/S3 Ilmu Komputer, FMIPA UGM, Yogyakarta

<sup>2</sup>Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta

e-mail: <sup>1\*</sup>[natalis.ransi@gmail.com](mailto:natalis.ransi@gmail.com), <sup>2</sup>[ewinarko@ugm.ac.id](mailto:ewinarko@ugm.ac.id)

### Abstrak

*Kecelakaan lalu lintas (laka lantas) di Sulawesi Tenggara perlu mendapatkan penanganan yang efektif karena menyebabkan korban meninggal dunia yang terus meningkat setiap tahunnya. Salah satu langkah penanganan adalah analisis karakteristik laka lantas yang berhubungan dengan korban meninggal dunia. Analisis karakteristik laka lantas dapat dilakukan dengan pendekatan faktor penyebab kecelakaan, jenis kecelakaan, dan waktu kejadian.*

*Penelitian ini mengaplikasikan algoritma Classification based on Predictive Association Rules (CPAR) pada data mining untuk analisa karakteristik laka lantas. Algoritma CPAR menghasilkan Class Association Rules (CARs), selanjutnya CARs digunakan untuk mendeskripsikan karakteristik laka lantas yang berhubungan dengan korban meninggal dunia.*

*Hasil penelitian diperoleh bahwa faktor yang menyebabkan korban meninggal dunia pada kasus laka lantas adalah faktor manusia (berkendara dibawah pengaruh alkohol dan berkendara melebihi batas kecepatan) dan faktor lingkungan fisik (prasarana jalan yang rusak dan jalan dengan tikungan tajam). Jenis kecelakaan (tunggal dan depan-depan), waktu kejadian (tanggal 8-14, hari Senin dan Selasa, jam 13:00-18:59), jenis kendaraan (sepeda motor) dan merek kendaraan (Honda), berpotensi menimbulkan korban meninggal pada kasus laka lantas. Pengendara sepeda motor rentan menjadi korban pada kasus laka lantas. Pengujian akurasi menggunakan 10-fold cross validation Hasil pengujian menunjukkan bahwa rata-rata akurasi algoritma CPAR lebih tinggi yaitu 48,75% dibandingkan dengan algoritma PRM yaitu 41,13%.*

**Kata kunci**— data mining, algoritma CPAR, kecelakaan lalu lintas

### Abstract

*Traffic accident in Southeast Sulawesi needs to get treatment more effective. One of the handling is analysis of traffic accident characteristic and then it was related to the death. Analysis of traffic accident characteristics can be done with the approach factors the cause of the accident, the kind of an accident, and time genesis.*

*This Research apply CPAR algorithm on the data mining to analyze the characteristics of traffic accident. CPAR Algorithm produce Class Association Rules (CARs) that used to describe traffic accident characteristics related to the death.*

*Results of research, that the factors that caused the victim died in traffic accident is human factors (driving under the influence of alcohol and driving exceed the speed) and environmental factors physical (road infrastructure and damaged roads with elbow). Types of accidents (in the singular and home-front), time genesis (on 8-14, reported Monday and Tuesday, hours 1:00 pm-6:59 pm), the type of vehicle (motorcycle), potentially causing the death toll in the case laka then. Motorcycle drivers are prone to fall victim in that case laka then. Testing accuracy using 10-fold cross validation test result show that on average these accuracy algorithm CPAR 48.75%, higher than the algorithm PRM 41.13%.*

**Keywords**— data mining, CPAR algorithm, traffic accident

## 1. PENDAHULUAN

Kecelakaan lalu lintas (laka lintas) di Sulawesi Tenggara menyebabkan korban meninggal dunia yang terus meningkat setiap tahunnya [1]. Laka lintas turut berkontribusi pada penurunan kualitas sumber daya manusia, khususnya yang berakibat cacat tetap (luka berat) [2]. Oleh karena itu, laka lintas perlu mendapat penanganan yang lebih efektif.

Penanganan laka lintas dapat dilakukan dengan analisa data kecelakaan. Salah satu langkah pada analisa data kecelakaan adalah analisis karakteristik laka lintas yang berhubungan dengan korban meninggal dunia. Analisis karakteristik laka lintas dapat dilakukan dengan pendekatan faktor penyebab kecelakaan, jenis kecelakaan, lokasi kejadian, dan waktu kejadian [3].

Karakteristik kecelakaan dapat dicari dengan menggunakan teknik *data mining* [4,5,6]. Teknik *Data mining* adalah teknik analisis data untuk menemukan pola dan hubungan dari repositori data yang besar [7,8,9].

Karakteristik laka lintas yang berhubungan dengan korban meninggal dunia dapat didekati dengan menerapkan teknik *Associative Classification* pada *data mining*. *Associative Classification* merupakan salah satu teknik *data mining* yang mengintegrasikan teknik *classification* dan *association*, dimana pengintegrasian ini fokus pada pengolahan *special subset* dari *Association Rule*, yang lebih dikenal sebagai *Class Association Rules* (CARs). CARs digunakan untuk membangun *classifier* pada proses klasifikasi [9,10].

*Associative Classification* membutuhkan algoritma dalam menghasilkan CARs. Salah satu algoritma dalam teknik *associative classification* adalah algoritma *Classification based on Predictive Association Rules* (CPAR). Keunggulan algoritma CPAR adalah memiliki tingkat akurasi yang tinggi dan kemudahan dalam hal interpretasi, dibandingkan dengan C4.5, *First Order Inductive Learner* (FOIL), *Repeated Incremental Pruning to Produce Error Reduction* (RIPPER), *Classification Based on Association* (CBA), dan *Classification based on Multiple Association Rules* (CMAR) [10,11,12,13,14, 15,16,17].

Penelitian ini mengaplikasikan algoritma CPAR untuk memperoleh karakteristik laka lintas yang berhubungan dengan korban meninggal dunia. Karakteristik yang diperoleh selanjutnya digunakan untuk membuat profil laka lintas di Kepolisian Resor Kabupaten Konawe, Kabupaten Muna dan Kota Bau-bau. Profil tersebut diharapkan dapat membantu Kepolisian untuk merumuskan kebijakan yang lebih efektif, sehingga korban meninggal dunia pada laka lintas berkurang [18,19,20].

## 2. METODE PENELITIAN

### 2.1 Data penelitian

Penelitian ini menggunakan data laka lintas yang diperoleh dari Aplikasi Manajemen Data Laka Lintas Direktorat Lalu Lintas Kepolisian Daerah Sulawesi Tenggara. Data laka lintas tahun 2010-2012, terdiri dari 25 variabel dan 138 item. Jumlah *records* adalah 10.568, merupakan hasil penggabungan tabel anatomi kecelakaan, kendaraan terlibat dan korban.

### 2.2 Algoritma *Classification based on Predictive Association Rules* (CPAR)

CPAR merupakan perluasan dari Algoritma *Predictive Rule Mining* (PRM), dan PRM itu sendiri merupakan perluasan dari *First Order Inductive Learner* (FOIL). CPAR memiliki keunggulan sebagai berikut: (1) CPAR menghasilkan aturan prediktif yang berkualitas tinggi yang jauh lebih kecil yang diambil secara langsung dari himpunan data; (2) untuk menghindari produksi *rules* yang berlebihan, CPAR menghasilkan tiap-tiap *rules* dengan mempertimbangkan himpunan aturan-aturan yang sudah di-*generate*; dan (3) ketika memprediksi label kelas dari sebuah *example*, CPAR menggunakan *rule k* terbaik yang dipenuhi oleh *example* yang bersangkutan

Pada CPAR ketika dalam proses membangun *rule*, selain memilih literal terbaik, semua literal yang memiliki selisih sedikit dengan *literal* terbaik juga dipilih, sehingga *rules* yang penting tidak akan terlewatkan. CPAR menghasilkan seperangkat *rules* yang lebih sedikit, dengan kualitas yang lebih tinggi dan tingkat *redundant* yang lebih kecil dibandingkan dengan *associative classification* yang lain seperti FOIL, PRM, dan *Clasification based on Association Rules* (CBA). Sehingga hasilnya, selain CPAR jauh lebih efisien dari segi waktu baik dalam proses menghasilkan *rule* maupun prediksi, CPAR juga mencapai akurasi setinggi yang dicapai oleh *associative classification* yang lain.

Pada algoritma CPAR, *Association Rule Mining* (ARM) digunakan untuk menghasilkan kandidat *rules*, yang memasukkan semua konjungsi dari literal yang memenuhi batas *support*. Kemudian sebuah *subset* dari *rules* tersebut dipilih dari kandidat-kandidat yang ada. *Subset* ini dibangun dengan kombinasi  $\delta$  *rules* terbaik dari setiap *example* ( $\delta = 1$  atau  $\delta = 4$ ). Hal ini dapat dianggap sebagai penyelidikan yang mendalam dalam menghasilkan *rules*. Algoritma PRM memiliki efisiensi yang lebih tinggi namun tingkat akurasinya lebih kecil dibandingkan dengan CPAR. Dalam algoritma PRM, setiap *rule* dihasilkan dari himpunan data yang tersisa. Misalkan sebuah *example*  $t$  yang berada di himpunan data yang tersisa dicakup oleh sebuah *rule*  $r$  yang baru saja dihasilkan. Kita tidak bisa memastikan apakah  $r$  merupakan *rule* terbaik untuk  $t$  karena: (1)  $r$  yang dihasilkan oleh algoritma *greedy*; dan (2)  $r$  bukan dihasilkan oleh himpunan data keseluruhan melainkan himpunan data tersisa. PRM juga menghasilkan sedikitnya sejumlah *rules* tertentu untuk tiap-tiap *example* (tergantung dari faktor pengurang berat  $\alpha$ ). Bagaimanapun juga, beberapa *rules* tersebut bukanlah *rules* terbaik untuk ilustrasi di atas.

Ketika memilih literal-literal selama proses pembangunan *rule*, PRM hanya memilih literal terbaik dan mengabaikan semua literal yang tersisa. Padahal sering kali ada beberapa literal yang memiliki nilai *gain* yang sama. Sehingga biasanya ada banyak *rules* dengan akurasi yang sama berdasarkan himpunan data yang tersisa. *Rule* terbaik yang dihasilkan pada himpunan data yang tersisa mungkin bukanlah *rule* terbaik berdasarkan himpunan data secara keseluruhan. Namun, PRM hanya memilih satu *rule* tersebut, yang akhirnya dapat mengarah pada kehilangan beberapa *rules* yang penting.

Pada saat memilih literal, *gain* digunakan untuk mengukur informasi yang diperoleh dari menambahkan literal tersebut ke daftar *rules*. Misalkan ada contoh positif sebanyak  $|P|$  dan contoh negatif sebanyak  $|N|$  memenuhi *rule*  $r$ . Setelah literal  $p$  ditambahkan ke  $r$ , ada contoh positif sebanyak  $|P'|$  dan contoh negatif sebanyak  $|N'|$  yang memenuhi *rule*  $r$  yang baru. Nilai *gain* dari  $p$  didefinisikan seperti persamaan (1).

$$Gain(p) = |P'| \left( \log \frac{|P'|}{|P'| + |N'|} - \log \frac{|P|}{|P| + |N|} \right) \quad (1)$$

*Gain* ( $p$ ) = jumlah *bits* yang disimpan dalam mempresentasikan semua contoh positif dengan menambahkan  $p$  ke  $r$ .

$P$  = totalWeight( $P$ )

$N$  = totalWeight( $N$ )

$P'$  = totalWeight( $P'$ )

$N'$  = totalWeight( $N'$ )

CPAR, yang merupakan penyempurnaan dari PRM, berada pada garis pertengahan antara algoritma *exhaustive* dan algoritma *greedy*, dan mengkombinasikan keuntungan dari keduanya. CPAR membangun *rules* dengan menambahkan atribut satu persatu, mirip dengan PRM. Bedanya adalah PRM mengabaikan semua atribut yang tersisa (hanya yang terbaik yang dipilih), CPAR tetap menjaga atribut-atribut yang derajatnya dekat dengan atribut terbaik dalam proses pembangunan *rule*. Hal ini dilakukan pertama kali menghitung nilai *gain* untuk masing-masing atribut dan menerapkan *Gain Similarity Ratio* (GSR) ke nilai *gain* tersebut. Sehingga *Local Gain Threshold* (LGT) dapat diterapkan untuk memilih semua atribut yang *gain*-nya lebih tinggi dari nilai LGT. Adapun LGT dapat dicari menggunakan persamaan (2).

$$\text{LGT} = \text{bestGain} * \text{GSR} \quad (2)$$

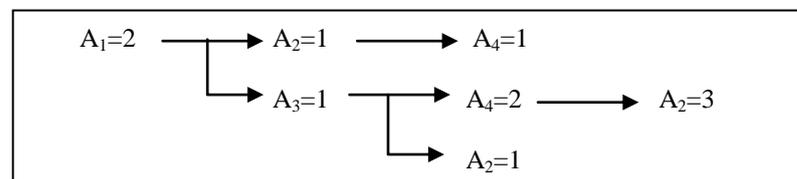
*bestGain* = gain tertinggi  
*GSR* = rasio kemiripan dari *gain*

Hasil LGT selanjutnya dipilih untuk pemrosesan lebih lanjut. Dengan demikian, CPAR dapat memilih lebih dari satu atribut pada saat yang bersamaan dan membangun *rules* secara simultan.

Berikut ini merupakan deskripsi detail dari CPAR dalam proses menghasilkan *rule*. Misalkan pada tahap tertentu dalam proses membangun sebuah *rule*, setelah menemukan literal *p* terbaik, literal *q* yang lain yang memiliki nilai *gain* yang sama dengan *p* (atau berbeda paling besar 1%) ditemukan. Disamping melanjutkan membangun *rule* dengan menambahkan *p* ke *r*, *q* juga ditambahkan ke *rule r* saat ini untuk membuat *rule r'* yang baru, yang dimasukkan ke dalam antrian. Tiap saat ketika *rule* baru dibangun, pertama antrian yang ada dicek lebih dahulu. Jika antrian tersebut tidak kosong, sebuah *rule* diambil dari antrian tersebut dan dinyatakan sebagai *rule* saat ini. Sehingga cara kerja ini membentuk pencarian terdalam terlebih dahulu (*depth-first-search*) dalam menghasilkan *rule*.

Contoh pada Gambar 1 merupakan pemisalan bagaimana CPAR menghasilkan *rules*. Setelah literal pertama ( $A_1 = 2$ ) dipilih, dua literal lainnya yakni ( $A_2 = 1$ ) dan ( $A_3 = 1$ ) ditemukan memiliki *gain* yang sama, yang nilainya lebih tinggi dibandingkan dengan literal-literal lain. Literal ( $A_2 = 1$ ) pertama-tama dipilih dan sebuah *rule* dihasilkan sepanjang arah tersebut. Setelah itu, *rule* ( $A_1 = 2, A_3 = 1$ ) dinyatakan sebagai *rule* saat ini. Ditemukan lagi dua literal dengan *gain* yang sama, ( $A_4 = 2$ ) dan ( $A_2 = 1$ ) dipilih dan sebuah *rule* dihasilkan sepanjang arah tersebut. Sehingga dengan cara demikian dihasilkan tiga buah *rule*.

1. ( $A_1 = 2, A_2 = 1, A_4 = 1$ ).
2. ( $A_1 = 2, A_3 = 1, A_4 = 2, A_2 = 3$ ).
3. ( $A_1 = 2, A_3 = 1, A_2 = 1$ ).



Gambar 1 Beberapa *rules* yang dihasilkan oleh CPAR [10]

CPAR (sebagaimana FOIL dan PRM) mengambil *input* sebuah himpunan data *binary valued* (yang terpisah oleh spasi) *R* dan memproduksi sebuah himpunan data dari *Classification Association Rules* (CARs) [21]. Gambar 2 dan Gambar 3 menunjukkan proses Algoritma CPAR secara garis besar.

```

method: startCPAR
parameters: none
-----
Generate an empty global attributes array A
For each c in C
  generate global P and N examples arrays
  generate global PN array
  determine minimum total Weight threshold

  while (total weight P > minimum total weight threshold)
    A' <-- A, N' <-- N, P' <-- P, PN' <-- PN
    if no attributes exist with weightings that can
      produce a gain above minimum break
    cparGeneration ( {}, c )
  end loop
end loop

```

Gambar 2 Metode *startCPAR* [21]

```

method: cparGeneration/2
parameters: parameter ante (antecedent so far)
              and cons (consequent) for current rule
global access to: A, A', N, N', P, P', PN, PN'
-----
for each a in A
  if (a not in antecedent1) calculate gain using
                          information in PN array and add to attribute array
end loop
i = "available" column in A with best gain
if (A[i][0] <= MIN_BEST_GAIN)
  add antecedent->c to rule list
  for all records in P reduce weighting by decay factor
  and adjust PN array accordingly

  return

gainThreshold = bestGain*GAIN_SIMILARITY_RATIO
for each a in A
  if a available and a.gain > gainThreshold
    tempP' <-- P', tempN' <-- N', tempA' <-- A', tempNP <-- NP
    add a' to antecedent
    remove examples from N' and P' that do not contain
    antecedent and adjust NP array accordingly

  if (N' == {})
    add antecedent->c to c rule list
    for all records in P reduce weighting bu decay factor
  else cparGeneration(antecedent,c)
  P' <-- tempP, N' <-- tempN, A' <-- tempA', NP' <-- tempNP'
end loop

```

Gambar 3 Method *CPARGeneration* [21]

Minimum gain constant = 0.7, decay factor = 1/3 dan bobot baru (*new weightings*) dihitung menggunakan persamaan (3).

$$newWeighting = oldweighting * decayFactor \quad (3)$$

## 2.2 Evaluasi Rule dalam CPAR

Sebelum membuat prediksi, setiap *rule* yang dihasilkan perlu dievaluasi untuk menentukan tingkat keakuratan prediksinya. Untuk sebuah *rules*  $r = "p_1 \wedge p_2 \wedge \dots \wedge p_n \rightarrow c"$ , didefinisikan tingkat akurasi yang diharapkan (*expected accuracy*) sebagai kemungkinan bahwa sebuah *example* memenuhi aturan dari *r* yang dimiliki oleh kelas *c*, atau *prob* ( $t \in c \mid tuple\ t\ memenuhi\ aturan\ dari\ r$ ).

Digunakan *Laplace Accuracy* untuk memperkirakan tingkat akurasi dari *rule* yang dihasilkan, yang didefinisikan pada persamaan (4).

$$LaplaceAccuracy = \frac{(n_c + 1)}{n_{total} + k} \quad (4)$$

Dengan *k* merupakan jumlah kelas yang ada,  $n_{total}$  merupakan jumlah total *examples* yang memenuhi semua *rule antecedent* yang dihasilkan, dimana diantaranya terdapat  $n_c$  yang merupakan *examples* yang dimiliki oleh kelas *c*, yang merupakan kelas yang diprediksi.

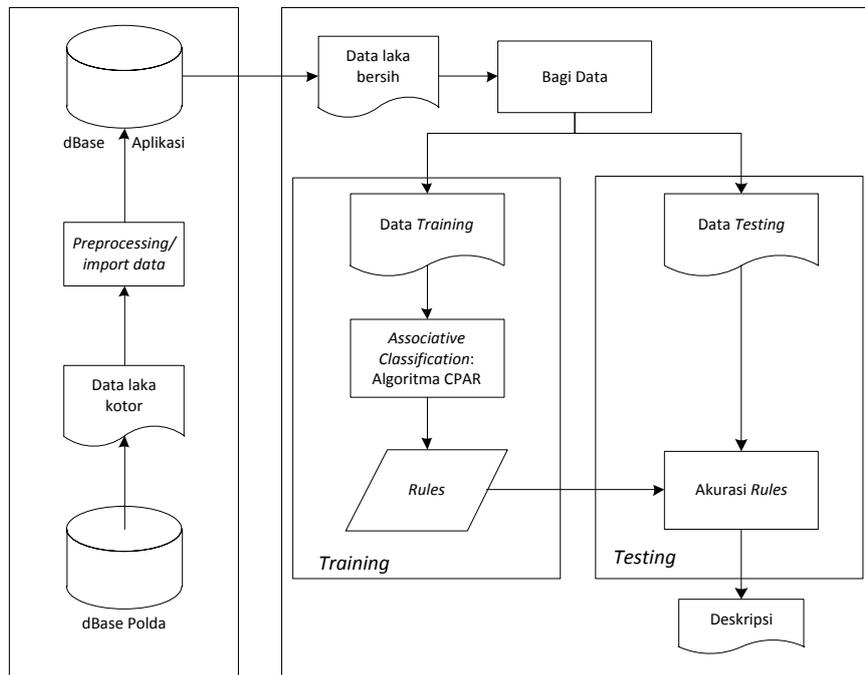
## 2.3 Klasifikasi dalam CPAR

Pada CPAR digunakan *k rules* terbaik setiap kelas untuk melakukan prediksi. Klasifikasi pada CPAR dilakukan dengan langkah sebagai berikut:

1. Memilih semua *rules* yang tercakup pada contoh data;
2. Dari *rules* yang dipilih pada langkah pertama, dipilih *k rules* untuk setiap kelas; dan

Membandingkan rata-rata nilai *Laplace Accuracy* dari *k rules* terbaik untuk setiap kelas dan dipilih kelas dengan nilai *Laplace Accuracy* tertinggi sebagai kelas prediksi.

## 2.4 Garis Besar Sistem



Gambar 4 Garis besar sistem

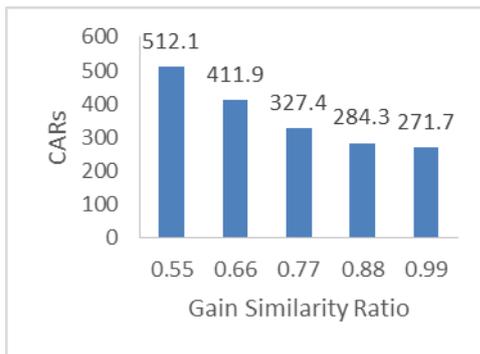
Gambar 4 memperlihatkan proses *import* data dari aplikasi manajemen data Direktorat Lalu Lintas Polisi Daerah Sulawesi Tenggara. Pada proses *import* data ini yang diambil adalah data anatomi kecelakaan, data kendaraan yang terlibat, dan data korban kecelakaan. Data tersebut selanjutnya dilakukan *preprocessing* berupa pembersihan, seleksi, dan tranformasi data. Proses pembersihan dilakukan secara terpisah dari sistem *preprocessing* yang dirancang, pembersihan data dilakukan dengan *text command* menggunakan aplikasi phpMyAdmin). Hasil akhir *preprocessing* adalah himpunan data yang disimpan di basis data aplikasi. Data selanjutnya digunakan untuk proses *associative classification* dengan algoritma CPAR.

Selanjutnya data dibagi dalam bentuk data *training* dan *testing* dengan masing-masing ukuran sesuai nilai *fold* yang ditentukan pada penelitian ini mengukaan *ten cross validation* untuk menguji validasi dari algoritma yang diterapkan. Untuk data *training* selanjutnya diolah menggunakan metode *associative classification* menggunakan algoritma CPAR yang kemudian menghasilkan *model/rules*. Dari *model/rules* ini selanjutnya digunakan sebagai uji data *testing* yang ada sehingga akan menghasilkan nilai akurasi dari *rules* yang dihasilkan

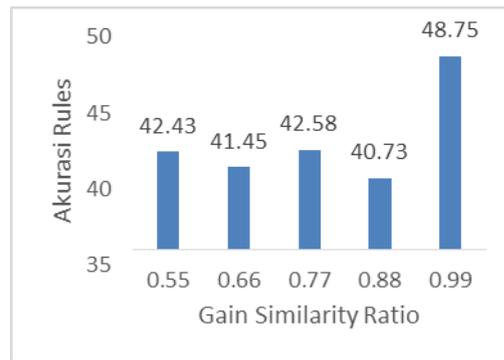
## 3. HASIL DAN PEMBAHASAN

### *Pengujian jumlah CARs dan Akurasi Rules*

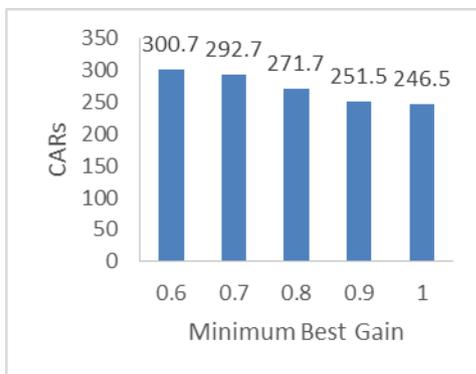
Dari hasil pengujian dengan parameter *Gain Similarity Ratio* 0,55, 0,66, 0,77, 0,88 dan 0,99 diperoleh bahwa perubahan nilai *Gain Similarity Ratio* mempengaruhi rata-rata *Class Association Rules* (CARs) dan nilai akurasi yang dihasilkan. Gambar 1 memperlihatkan histogram jumlah CARs dan Gambar 2 memperlihatkan histogram rerata akurasi *model/rules* yang dihasilkan.



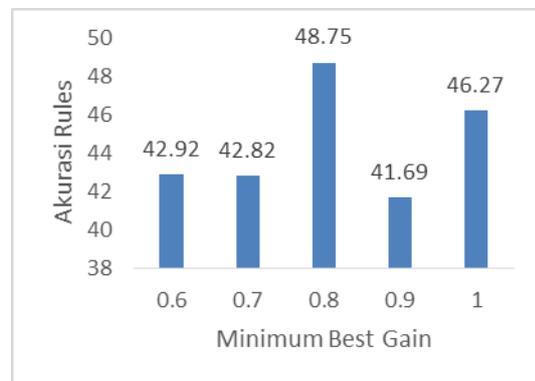
Gambar 5 Histogram rerata CARs dengan perubahan parameter Gain Similarity Ratio



Gambar 6 Histogram rerata akurasi rules dengan perubahan Gain Similarity Ratio



Gambar 7 Histogram rerata CARs dengan perubahan parameter Minimum Best Gain



Gambar 8 Histogram rerata akurasi rules dengan perubahan parameter Minimum Best Gain

Gambar 5 memperlihatkan histogram hasil uji rerata CARs yang diperoleh dengan perubahan nilai *Gain Similarity Ratio*. Berdasarkan Gambar 5 diperoleh bahwa semakin nilai *Gain Similarity Ratio* mendekati 5 maka CARs yang dihasilkan semakin sedikit. Hal ini mengkonfirmasi bahwa jumlah CARs yang diperoleh bergantung pada *Gain Similarity Ratio*.

Gambar 6 menunjukkan bahwa rerata akurasi tertinggi diperoleh dengan nilai *Gain Similarity Ratio* = 0,99, yaitu sebesar 48,75%. Dengan demikian *rules item* yang terbentuk dengan tingkat kemiripan nilai *gain* 0,9, memiliki nilai akurasi yang lebih baik dibanding dengan *rules item* yang dibentuk dengan tingkat kemiripan dibawah 0,99.

Hasil pengujian dengan parameter *Minimum Best Gain* 0.6, 0.7, 0.8, 0.9 dan 1 diperoleh bahwa perubahan nilai tersebut mempengaruhi jumlah CARs dan nilai akurasi yang dihasilkan. Gambar 7 memperlihatkan histogram rerata CARs dan Gambar 8 memperlihatkan histogram rerata akurasi *model/rules* yang dihasilkan.

Gambar 7 memperlihatkan hasil uji rerata CARs yang diperoleh dengan perubahan nilai *Minimum Best Gain*. Berdasarkan Gambar 7 diperoleh bahwa semakin *Minimum Best Gain* mendekati 1 maka jumlah CARs semakin kecil. Hal ini mengkonfirmasi bahwa jumlah *candidate rules* yang diperoleh bergantung pada *Minimum Best Gain*.

Gambar 8 diperoleh bahwa rerata akurasi tertinggi diperoleh dengan nilai *Minimum Best Gain* = 0,8 yaitu sebesar 48,75%. Dengan demikian *rules item* yang bentuk dengan nilai *gain* 0,8 memiliki nilai akurasi yang lebih baik dibanding dengan *rules item* yang dibentuk dengan nilai *gain* dibawah maupun diatas 0,8.

Dengan demikian, dari hasil pengujian yang dilakukan dapat disimpulkan bahwa *model/rules* terbaik diperoleh dengan menggunakan parameter *Gain Similarity Ratio* = 0,99, *Minimum Best Gain* = 0,8, dengan nilai rerata *candidate rules* 271,7 dan rerata akurasi sebesar 48,75%.

#### *Karakteristik Laka Lantas*

Bagian ini diberikan karakteristik laka lantas untuk beberapa Kepolisian Resor di wilayah Kepolisian Daerah Sulawesi Tenggara (Kabupaten Konawe, Kabupaten Muna dan Kota Bau-Bau). Karakteristik laka lantas merupakan CARs yang menarik. CARs yang menarik adalah CARs yang memiliki *Laplace Accuracy* (LA) di atas 0,75. Tabel 1 memperlihatkan 8 *rules antecedent* Kepolisian Resor Kabupaten Konawe, Tabel 2 memperlihatkan 3 *rules antecedent* Kepolisian Resor Kabupaten Muna, dan Tabel 3 memperlihatkan 7 *rules antecedent* Kepolisian Resor Kota Bau-Bau

Tabel 1 Rules Antecedent Kepolisian Resor Kabupaten Konawe

No	<i>Rule antecedent</i>	LA
1	HARI = AWAL PEKAN dan JENIS_LOKASI = KAWASAN PEMUKIMAN dan JENIS_LAKA = TUNGGAL dan POSISI_KORBAN = PENGENDARA	0,78
2	TANGGAL = MINGGU KEDUA dan JAM = SORE dan KONDISI_JALAN = TIKUNG TAJAM	0,78
3	JAM = SORE dan JENIS_LAKA = DEPAN DEPAN dan KONDISI_JALAN = TIKUNG TAJAM	0,78
4	JAM = SORE dan KONDISI_JALAN = TIKUNG TAJAM dan JENIS_KENDARAAN = SEPEDA MOTOR	0,78
5	JAM = SORE dan KONDISI_JALAN = TIKUNG TAJAM dan MERK_KENDARAAN = HONDA	0,78
6	JAM = SORE dan KONDISI_JALAN = TIKUNG TAJAM dan PENDIDIKAN_KORBAN = SMA	0,78
7	JAM = SORE dan KONDISI_JALAN = TIKUNG TAJAM dan POSISI_KORBAN = PENGENDARA	0,78
8	JENIS_JALAN = NASIONAL dan KONDISI_JALAN = TIKUNG TAJAM dan PEKERJAAN KORBAN = PNS	0,75

#### *Karakteristik Laka Lantas di Kabupaten Konawe*

Faktor utama penyebab laka lantas yang berpotensi menimbulkan korban meninggal dunia di Wilayah Resor Konawe adalah lingkungan fisik, yaitu kondisi sarana dan prasarana jalan. Hal tersebut sesuai dengan fakta bahwa di Wilayah Resor Konawe terdapat beberapa jalan dengan tikungan/tanjakan/turunan tajam. Jalan dengan tikungan/tanjakan/turunan tajam membuat pengendara sering kali tidak tepat mengambil badan jalan.

Jenis laka lantas, yaitu tunggal dan depan-depan turut berpotensi menimbulkan korban meninggal dunia. Jenis laka lantas tunggal sering terjadi dikawasan pemukiman, sedangkan jenis laka lantas depan-depan sering terjadi di jalan nasional. Waktu kejadian laka lantas (antara tanggal 8 sampai 14, hari Senin dan Selasa, jam antara 13:00 sampai 18:59 berpotensi menimbulkan korban meninggal dunia. Selain itu, diperoleh pengetahuan bahwa jenis kendaraan sepeda motor dan merk Honda berpotensi menimbulkan korban meninggal dunia.

Tabel 2 Rules Antecedent Kepolisian Resor Kabupaten Muna

No	Rule antecedent	LA
1	JENIS_LAKA = TUNGGAL DAN JENIS_KENDARAAN = SEPEDA MOTOR DAN POSISI_KORBAN = PENGENDARA	0,86
2	JENIS_LAKA = TUNGGAL DAN POSISI_KORBAN = PENGENDARA	0,80
3	UMUR_KORBAN = DEWASA 2 DAN PENDIDIKAN_KORBAN = TIDAK SEKOLAH	0,80

Tabel 3 Rules Antecedent Jajaran Resor Bau-Bau

No	Rule antecedent	LA
1	KERUGIAN_MATERIL = BESAR dan KONDISI_PENGE MUDI = PENGARUH ALKOHOL dan UMUR_KORBAN = DEWASA 1 dan POSISI_KORBAN = PENGENDARA	0,82
2	HARI = AWAL PEKAN dan JAM = PAGI dan KONDISI_PENGE MUDI = PENGARUH ALKOHOL	0,78
3	JENIS_LAKA = TUNGGAL dan UMUR_PENGE MUDI = DEWASA 1 dan POSISI_KORBAN = PENGENDARA	0,77
4	KERUGIAN_MATERIL = KECIL dan KONDISI_JALAN = RUSAK dan JENIS_KENDARAAN = MOBIL BEBAN	0,77
5	JENIS_LAKA = TUNGGAL dan KERUGIAN_MATERIL = KECIL dan JENIS_KENDARAAN = SEPEDA MOTOR dan UMUR_PENGE MUDI = DEWASA 1	0,75
6	JAM = SIANG dan KONDISI_ALAM = KABUT dan KONDISI_PENGE MUDI = PENGARUH ALKOHOL	0,75
7	KONDISI_ALAM = KABUT dan KONDISI_PENGE MUDI = PENGARUH ALKOHOL dan SIM_PENGE MUDI = TANPA SIM	0,75

CARs dapat menjadi rujukan bagi Kepolisian untuk melakukan pendidikan masyarakat tentang lalu lintas (DIKMAS Lantas), yaitu rujukan untuk menentukan prioritas jajaran dan peserta DIKMAS Lantas. Selain itu, CARs dapat menjadi rujukan bagi Dinas Perhubungan dan Dinas Pekerjaan Umum untuk merancang sarana dan prasarana (*geometric design* dan konstruksi) jalan yang aman dan prioritas lokasi perbaikan jalan.

#### Pengujian dengan Algoritma PRM

Tabel 4 memperlihatkan hasil pengujian dengan algoritma PRM menggunakan perangkat lunak LUCK-KDD dari Universitas Liverpool dan CPAR menggunakan aplikasi yang dikembangkan berdasarkan waktu, *candidate rules* yang terbentuk dan akurasi.

Tabel 4 Perbandingan algoritma PRM dan algoritma CPAR

Algoritma	Waktu (detik)	Rerata CARs	Rerata Akurasi (%)
PRM	6,97	239,7	41,13
CPAR	11,77	271,7	48,75

Tabel 4 menunjukkan bahwa algoritma CPAR lebih lama dalam membangun CARs yaitu 11,77 detik dibandingkan dengan algoritma PRM dengan waktu 6,97 detik. Algoritma CPAR menghasilkan rata-rata CARs lebih banyak yaitu 271,7 dibandingkan algoritma PRM dengan rata-rata 238,7. Selanjutnya algoritma CPAR menghasilkan akurasi lebih tinggi yaitu 48,75% dibandingkan dengan PRM sebesar 41,13%. Hal tersebut terjadi karena pada algoritma CPAR memasukkan parameter *Gain Similarity Ratio* (GSR) pada saat membangun CARs. Parameter GSR pada algoritma CPAR mampu mencari CARs yang mungkin saja dilewati oleh algoritma PRM.

#### 4. KESIMPULAN

Kesimpulan pada penelitian ini adalah sebagai berikut:

1. Algoritma CPAR dapat digunakan untuk menghasilkan karakteristik laka lantas di Wilayah Ditlantas Polda Sultra. Algoritma CPAR menghasilkan CARs laka lantas berbentuk “if-then”. Setiap CARs mempresentasikan faktor-faktor yang berhubungan dengan korban meninggal dunia.
2. Faktor yang menyebabkan korban meninggal dunia pada kasus laka lantas adalah faktor manusia yaitu berkendara dibawah pengaruh alkohol dan berkendara melebihi batas kecepatan dan faktor lingkungan fisik (prasarana jalan yang rusak dan kondisi jalan dengan tikungan tajam). Kombinasi kedua faktor tersebut berpotensi menyebabkan korban meninggal dunia pada kasus laka lantas.
3. Waktu kejadian, jenis dan merek kendaraan, jenis kecelakaan berpotensi menimbulkan korban meninggal dunia pada kasus laka lantas.
4. Pengendara roda dua (sepeda motor) rentan menjadi korban laka lantas.
5. Akurasi algoritma CPAR lebih memiliki tingkat akurasi lebih tinggi dibandingkan dengan algoritma PRM. Hal tersebut terlihat dari akurasi algoritma PRM sebesar 41,13% sedangkan algoritma CPAR sebesar 48,75%. Algoritma CPAR menghasilkan rata-rata CARs yang lebih banyak yaitu 271,7 dibandingkan algoritma PRM yaitu 239,7. Waktu yang dibutuhkan algoritma PRM lebih cepat yaitu 6,97 detik dibandingkan algoritma CPAR yaitu 11,77 detik.

#### 5. SARAN

Saran pada penelitian adalah sebagai berikut:

1. Algoritma lain perlu dicoba untuk mendapatkan akurasi *rules* yang lebih tinggi.
2. Perbandingan dengan ilmu budaya perlu dilakukan, yaitu tentang hubungan budaya berkendara dengan laka lantas di suatu daerah.
3. Untuk analisa pola laka lantas di Indonesia sebaiknya dilakukan setiap Wilayah Kepolisian, karena karakteristik laka lantas berbeda setiap daerah.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Direktur Eksekutif Program HPEQ Universitas Haluoleo yang telah memberi dukungan finansial terhadap penelitian ini.

## DAFTAR PUSTAKA

- [1] Sigit, S., 2011, Laporan Akuntabilitas Kinerja Kepolisian Daerah Sulawesi Tenggara Tahun 2010, <http://humas.polri.go.id>, diakses tanggal 23 Desember 2013.
- [2] Herukoco, 2013, *Kamseltibcar Lantas Ditlintas Polisi Daerah Sulawesi Tenggara tahun 2013*, Primatama, Kendari - Indonesia.
- [3] Muhammad, I., dan Iman, S., 2004, *Penanganan Lokasi Rawan Kecelakaan*. Pusat Penelitian dan Pengembangan Prasarana Transportasi Badan Penelitian dan Pengembangan Kimpraswil. Departemen Pemukiman dan Prasarana Wilayah Republik Indonesia.
- [4] Mona, K. A. E., dan Hesham, R. E., 2004, Data Mining of Traffic Accidents in Miami-Dade County, *Proc. IIE Annual Conference*. Hal. 1-6.
- [5] Karolien G., Isabelle, T., dan Geert, W., 2005, Understanding Spatial Concentrations of Road Accidents Using Frequent Item Sets, *Accident Analysis and Prevention* Vol. 37, Hal. 787-799.
- [7] Kumar, V., Tan, P. N. dan Steinbach, M., 2001, *Data Mining*, CRC Press, LLC, New York, USA
- [8] Turban, E., Aronson, J. E. dan Liang, T. P., 2005, *Decision Support Systems and Intelligent Systems – 7th Ed*, Prabantini, D., Penerbit Andi, Yogyakarta.
- [9] Han, J., Kamber, M. dan Pei, J., 2012, 2012, *Data Mining: Concepts and Techniques Second Edition*, Morgan Kaufmann Publishers, San Francisco.
- [10] Xiaoxin, Y., and Jiawei, H., 2003, CPAR: Classification based on Predictive Association Rules, Proc. 2003 SIAM Int. Conf. on Data Mining (SDM'03), San Fransisco, CA, May 2003.
- [11] Quinlan, J.R., 1993, C4.5 Programs for Machine Learning, Morgan Kaufmann.
- [12] Quinlan, J. R. dan Cameron-Jones, R. M., 1993, FOIL A midterm report. In Proc. 1993 European Conf. Machine Learning, Hal. 3-20, Vienna, Austria.
- [13] Cohen, W., 1995, Fast effective rule induction. In ICML'95, Hal. 115-123, Tahoe City, CA.
- [14] Liu, B., Hsu, W. dan Ma, Y., 1998, Integrating classification and association rule mining. In KDD'98, Hal. 80-86, New York, USA.
- [15] Li, W., Han, J. dan Pei, J., 2001, CMAR: Accurate and efficient classification based on multiple class-association rules. In ICDM'01, Hal. 369-376, San Jose, CA, Nov.2001.
- [16] Rahmat H., Mustapha, A., Said, M.S.M. dan Amit N.A., 2011, Classification Assessment based on Accuracy, Compactness, and Speed of C4.5 and CPAR A Comparative Study, ICGST AIML-11 Conference, Hal.121-125, Dubai, UAE.
- [17] Prafulla, G., dan Durga, T., 2011, Performance Comparison of Rule Based Classification Algorithms, *International Journal of Computer Science and Informatics*, Volume-I, Issue-II, Hal. 37-42.
- [18] Benoit, D., Geert, W., dan Koen, V., 2008, Traffic Accident Segmentation by Means of Latent Class Clustering, *Accident Analysis and Prevention* Vol. 40 (2008), Hal. 1257-1266.
- [19] Cheng, C.W., Lin, C.C. dan Leu, S.S., 2009, Use of Association Rules to Explore Cause-Effect Relationships in Occupational Accidents in the Taiwan Construction Industry, *Safety Science*, Vol. 48, Hal. 436-444.
- [20] Pakgohar, A., Tabrizi, R., S., Khalili M. dan Esmaeili, A., 2011, The Role of Human Factor in Incidence and Severity of Road Crashes based on the CART and LR Regression: a Data Mining Approach, *Procedia Computer Science* Vol. 3, Hal.764-769.

- [21] Frans, C., 2004, The LUCS-KDD Implementations of CPAR (Classification Based on Predictive Association Rules), Department of Computer Science The University of Liverpool.