

Klasifikasi Status Korban Kecelakaan Lalu Lintas Di Kota Samarinda Menggunakan Algoritma QUEST (*Quick, Unbiased, Efficient, Statistical Tree*)

Surya Prangga^{1✉}, Rito Goejantoro², Memi Nor Hayati³, Siti Mahmuda⁴, Dwi Husnul Mubiin⁵

^{1,2,3,4,5}Program Studi Statistika, FMIPA, Universitas Mulawarman, Kalimantan Timur, Indonesia

✉Email: suryapranggae@fmipa.unmul.ac.id

ABSTRAK

Kecelakaan lalu lintas merupakan kejadian yang sulit untuk diprediksi kapan dan dimana akan terjadi. Kasus kecelakaan lalu lintas di Indonesia merupakan terbilang cukup besar. Jika dilihat berdasarkan data dari Badan Pusat Statistik selama tahun 2019 sampai dengan 2021 rata-rata kecelakaan lalu lintas mencapai seratus ribu kasus setiap tahunnya. Khususnya di Kota Samarinda yang merupakan Ibukota Provinsi Kalimantan Timur menempati urutan tertinggi pada tahun 2020 dibandingkan beberapa kabupaten/kota lainnya yang berada di Provinsi Kalimantan Timur. Berdasarkan fakta tersebut kasus kecelakaan lalu lintas perlu untuk ditangani sehingga dapat meminimalisir korban kecelakaan. Salah satu teknik data mining yang digunakan untuk melihat pola kecelakaan lalu lintas yaitu dengan menggunakan metode klasifikasi berbasis pohon keputusan. Salah satu metode klasifikasi berbasis pohon keputusan adalah algoritma QUEST. Algoritma QUEST (*Quick, Unbiased, Efficient, and Statistical Tree*) sebagai metode berbasis pohon keputusan dapat digunakan untuk mengklasifikasikan status korban kecelakaan lalu lintas. Berdasarkan analisis data diperoleh hasil akurasi terbaik dari proporsi data 70:30 dalam mengklasifikasikan status korban kecelakaan lalu lintas yaitu sebesar 54,54% dengan nilai *F1-Score* sebesar 42,86%.

Kata kunci: Kecelakaan Lalu Lintas, Klasifikasi, Pohon Keputusan, QUEST.

ABSTRACT

Traffic accidents are difficult to predict in terms of when and where will occur. The number of traffic accident cases in Indonesia is relatively high. Regarding on data from the Central Statistics Agency (Badan Pusat Statistik) from 2019 to 2021, the average number of traffic accidents reaches one hundred thousand cases every year. Especially, in the Samarinda City, which is the capital of East Kalimantan Province, it ranked the highest in 2020 compared to several other regencies and cities within East Kalimantan Province. Considering these facts, traffic accident cases need to be addressed to minimize accident-related casualties. One data mining technique used to analyze traffic accident patterns is the decision tree-based classification method. One of the decision tree-based classification methods is QUEST algorithm. The QUEST algorithm (*Quick, Unbiased, Efficient, and Statistical Tree*) can be used to classify the status of traffic accident victims. Based on data analysis, the best accuracy to classify the status of traffic accident victims was obtained using a 70:30 data split, with an accuracy of 54.54% and an F1-Score of 42.86%.

Keywords: Traffic Accidents, Classification, Decision Trees, QUEST.

PENDAHULUAN

Kecelakaan lalu lintas merupakan kejadian dimana sebuah kendaraan bermotor bertabrakan dengan benda lain dan mengakibatkan kerusakan, bahkan mengakibatkan luka-luka atau kematian manusia atau binatang [1]. Kecelakaan lalu lintas juga merupakan kejadian yang sulit diprediksi kapan dan dimana akan terjadi [2]. Kecelakaan lalu lintas merupakan permasalahan yang dihadapi oleh setiap negara di seluruh dunia, tidak terkecuali di Indonesia.

Menurut data dari Badan Pusat Statistik yang bersumber dari data kepolisian Republik Indonesia, pada tahun 2019 tercatat total kecelakaan lalu lintas sejumlah 116.411 kemudian

terjadi penurunan di tahun 2020 menjadi 100.028 namun di tahun 2021 kembali terjadi kenaikan sebesar 103.645 sehingga besarnya kerugian materi yang ditaksir mencapai Rp. 246.653 juta rupiah [3]. Kemudian jika melihat kondisi lebih dalam yaitu di Kota Samarinda yang merupakan ibukota Provinsi Kalimantan Timur yang memiliki tingkat kecelakaan lalu lintas cukup tinggi.

Menurut Badan Pusat Statistik Provinsi Kalimantan Timur tahun 2020 yang bersumber dari Kepolisian Daerah Provinsi Kalimantan Timur, bahwa Kota Samarinda menempati posisi tertinggi pada kasus kecelakaan lalu lintas selama 3 tahun berturut-turut dari 2018 sampai dengan 2020 yaitu dengan jumlah kasus sebesar 143 kasus, 109 kasus dan 96 kasus. Berdasarkan beberapa fakta tersebut menunjukkan bahwa kasus kecelakaan lalu lintas masih menjadi permasalahan dan memerlukan solusi penyelesaian, minimal dalam meminimalisir korban kecelakaan lalu lintas khususnya di Kota Samarinda. Informasi faktor-faktor yang diduga memengaruhi kecelakaan lalu lintas dapat digunakan untuk mendapatkan pola atau informasi yang bermanfaat terkait status (tingkat keparahan) korban kecelakaan. Harapannya dengan mengetahui pola atau klasifikasi status kecelakaan tersebut dapat digunakan untuk membuat kebijakan oleh instansi terkait dalam meminimalisir jumlah korban kecelakaan lalu lintas.

Analisis Statistika menawarkan berbagai metode yang dapat digunakan untuk mencari pola atau *pattern* dalam menangani permasalahan klasifikasi. Diantara teknik yang dapat digunakan adalah metode berbasis pohon keputusan. Salah satu algoritma pohon keputusan yang dapat dijadikan sebagai alternatif untuk mengklasifikasikan status korban kecelakaan adalah algoritma QUEST (*Quick, Unbiased, Efficient, Statistical Tree*) [4]. Terdapat beberapa penelitian terdahulu yang menggunakan metode klasifikasi berbasis pohon dalam mengklasifikasikan korban kecelakaan lalu lintas. Diantaranya seperti yang dilakukan oleh Susiana yang berjudul Penerapan Metode *Boosting* pada CART untuk Mengklasifikasikan Korban Kecelakaan lalu lintas di Kota Palu. Adapun faktor-faktor yang dinilai berupa jenis kecelakaan, jenis kelamin, usia, peran korban, jenis kendaraan, waktu kecelakaan dan jenis pelanggaran. Hasil ketepatan klasifikasi dengan menggunakan CART yaitu sebesar 77,9% dan setelah dilakukan *Boosting* meningkat menjadi 82% [5].

Penelitian lainnya yang dilakukan oleh Franseda yang berjudul Integrasi metode *Decision Tree* dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas. Hasil akurasi tertinggi yang diperoleh mencapai 71,12% [6]. Berdasarkan beberapa uraian tersebut peneliti tertarik untuk mengimplementasikan metode pohon keputusan QUEST untuk melihat performa algoritma dalam mengklasifikasikan status korban kecelakaan lalu lintas berupa ketepatan klasifikasi terbaik serta untuk melihat faktor-faktor apa saja yang memengaruhi kecelakaan di Kota Samarinda.

UJI INDEPENDENSI (*CHI-SQUARE*)

Persoalan yang ingin diketahui dari uji independensi adalah apakah ada hubungan antara 2 unsur/variabel. Setiap unsur terdiri atas beberapa kategori. Kedua unsur tersebut ditampilkan dalam suatu tabel kontingensi yang terdiri atas beberapa baris dan beberapa kolom [7].

ANALISIS VARIANSI

Analisis variansi adalah suatu prosedur untuk menguji perbedaan rata-rata beberapa populasi (lebih dari dua). Konsep analisis variansi didasarkan pada konsep distribusi F dan biasanya dapat diaplikasikan untuk berbagai macam kasus maupun dalam analisis hubungan antara berbagai variabel yang diamati [8].

UJI KESAMAAN VARIANSI (*LEVENE'S TEST*)

Uji Levene digunakan untuk menguji kesamaan variansi dari beberapa populasi. Uji ini merupakan alternatif dari uji Bartlett. Uji Levene tidak membutuhkan syarat kenormalan data [9].

ALGORITMA *QUEST*

QUEST merupakan kependekan dari *Quick, Unbiased, Efficient, Statistical Tree*. Pohon *QUEST* dikatakan *binary decision tree* karena merupakan algoritma pohon keputusan yang dibangun dengan memisahkan *node* ke dalam dua *child nodes* yang dilakukan secara berulang-ulang (*rekursif*), diawali dengan *root node* yang berisi *training sample*. *QUEST* dikembangkan oleh Wei-Yin Loh (*University of Wisconsin-Madison*) dan Yu-Shan Shih (*National Chung Cheng University, Taiwan*) pada tahun 1997 dari metode pohon klasifikasi non-biner *FACT* [10]. Proses pertumbuhan pohon *QUEST* terdiri dari tiga bagian yakni, algoritma pemilihan variabel prediktor penyekat, algoritma penentuan simpul penyekat, dan algoritma penghentian pembentukan pohon.

PENGUKURAN PERFORMA KLASIFIKASI

Pada umumnya pengukuran performa klasifikasi dilakukan dengan menggunakan bantuan matriks konfusi (*confusion matrix*). Matriks konfusi adalah tabel pencatat hasil kerja klasifikasi. Matriks konfusi pada permasalahan klasifikasi digunakan untuk membantu perhitungan nilai akurasi, sensitivitas dan *F1-Score*. Adapun formulanya sebagai berikut.

$$\text{akurasi} = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (1)$$

$$\text{sensitivitas} = \frac{TP}{(TP+FN)} \quad (2)$$

$$\text{presisi} = \frac{TP}{(TP+FP)} \quad (3)$$

$$F1 - \text{Score} = \frac{2 \times \text{Presisi} \times \text{Sensitivitas}}{\text{Presisi} + \text{Sensitivitas}} \quad (4)$$

METODE PENELITIAN

Data pada penelitian merupakan bentuk data sekunder berupa data kasus kecelakaan lalu lintas di Kota Samarinda yang diperoleh dari Kantor Kepolisian Resor Kota Samarinda. Populasi penelitian ini merupakan korban kecelakaan lalu lintas di Kota Samarinda. Sampel pada penelitian ini adalah korban kecelakaan lalu lintas di Kota Samarinda Tahun 2020 sampai dengan 2021. Teknik pengambilan sampel menggunakan *sampling purposive* dimana pada proses pengambilan sampel peneliti mempertimbangkan kebaruan dan kelengkapan data. Adapun jumlah variabel untuk menjelaskan karakteristik dan kondisi korban kecelakaan lalu lintas di Kota Samarinda menggunakan tujuh variabel. Lebih jelasnya dapat dilihat pada Tabel 1.

Variabel yang digunakan sebagian besar berupa variabel kategorik dan bersifat biner. Tahapan analisis data pada penelitian ini secara ringkas mengikuti langkah-langkah berikut.

1. Melakukan analisis statistika deskriptif terhadap karakteristik dan kondisi korban kecelakaan lalu lintas di Kota Samarinda
2. Melakukan pembagian proporsi data *training* dan *testing* dengan menerapkan pengambilan sampel acak. Adapun proporsi data yang digunakan diantaranya 70:30, 80:20 dan 90:10.
3. Melakukan analisis klasifikasi dengan algoritma *QUEST* menggunakan data *training*.

4. Mengukur performa hasil klasifikasi data *testing* dengan menggunakan formula (1), (2), (3) dan (4).
5. Mendapatkan performa hasil klasifikasi terbaik.
6. Membuat pohon keputusan dan aturan klasifikasi berdasarkan performa terbaik.

Tabel 1. Variabel Operasional

Variabel	Jenis Data	Kodifikasi
Status Kecelakaan (Y)	Kategorik	0 = Meninggal dunia 1 = Tidak meninggal dunia
Usia (X_1)	Kategorik	0 = Usia kanak-kanak dan muda (0-35 tahun) 1 = Usia dewasa dan lansia (> 36 tahun)
Jenis kelamin (X_2)	Kategorik	0 = Laki-laki 1 = Perempuan
Peran Korban (X_3)	Kategorik	0 = Pengemudi 1 = Penumpang dan lainnya
Kendaraan Korban (X_4)	Kategorik	0 = Kendaraan bermotor roda dua atau tiga 1 = Kendaraan roda empat atau lebih dan lainnya
Status Jalan (X_5)	Kategorik	0 = Nasional dan Provinsi 1 = Kota dan Desa
Kondisi Cuaca (X_6)	Kategorik	0 = Hujan atau gerimis 1 = Cerah

HASIL DAN PEMBAHASAN

Setelah dilakukan pengolahan terhadap data korban kecelakaan lalu lintas di Kota Samarinda Tahun 2020 – 2021 sehingga diperoleh hasil sebagai berikut.

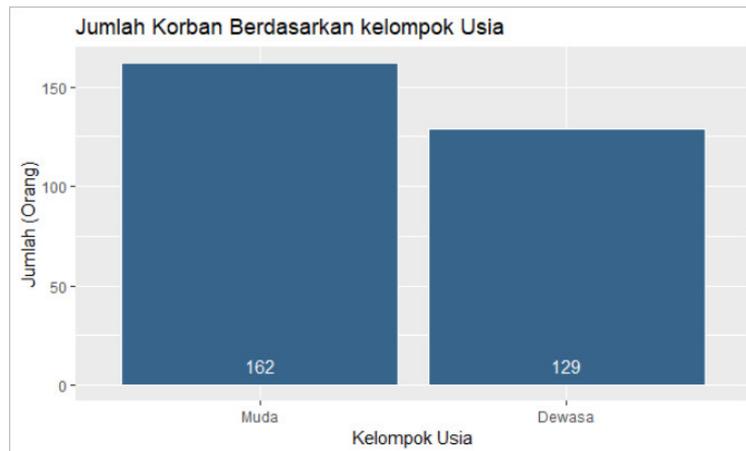
Statistika Deskriptif

Analisis statistika deskriptif bertujuan untuk memberikan gambaran secara umum terkait karakteristik korban kecelakaan lalu lintas di Kota Samarinda Tahun 2020 – 2021. Terdapat beberapa variabel yang dapat mewakili karakteristik korban kecelakaan lalu lintas di Kota Samarinda diantaranya status kecelakaan, usia, jenis kelamin, peran korban, kendaraan korban dan status jalan. Berikut visualisasi dengan menggunakan diagram batang untuk menjelaskan gambaran masing-masing variabel.



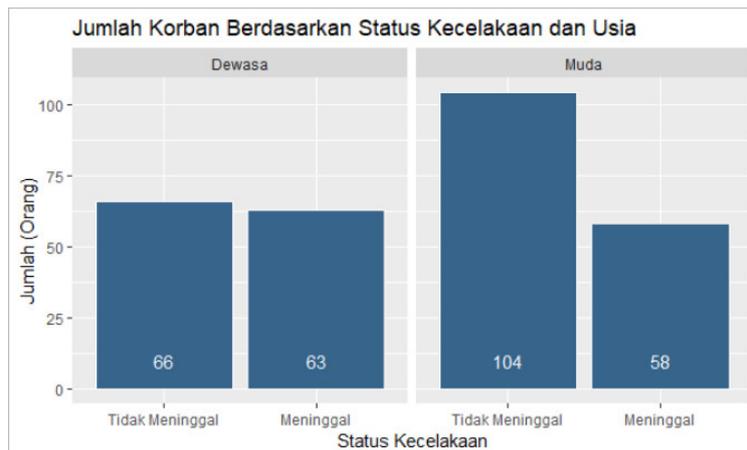
Gambar 1. Status Kecelakaan

Berdasarkan Gambar 1, dapat dilihat bahwa dari 291 korban kecelakaan lalu lintas di Kota Samarinda yang terjadi pada tahun 2019 sampai dengan 2021 terdapat 170 korban kecelakaan dengan status tidak meninggal dan sisanya 121 korban tercatat dengan status meninggal. Jika dipersentasekan terdapat sekitar 41,58% korban kecelakaan lalu lintas dengan status meninggal dunia. Apabila melihat korban berdasarkan kelompok usia seperti yang diperlihatkan pada Gambar 2.

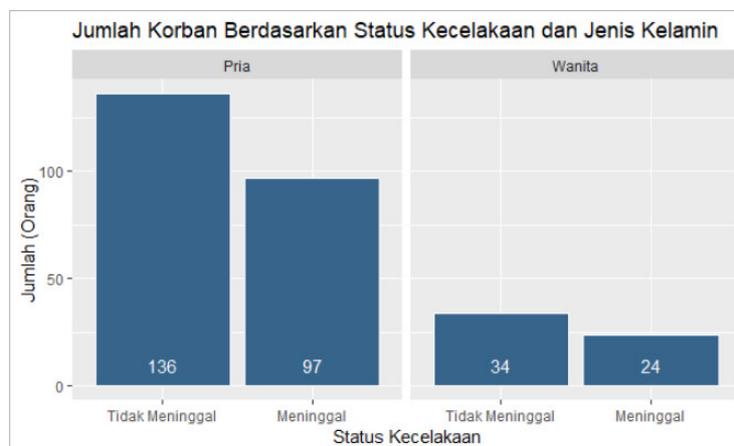


Gambar 2. Korban berdasarkan kelompok usia

dapat dilihat bahwa korban kecelakaan lalu lintas di Kota Samarinda yang tercatat pada Tahun 2019 – 2021 didominasi oleh kelompok usia muda. Jumlah yang tercatat sebanyak 162 orang atau sekitar 55,7% dari total keseluruhan korban kecelakaan. Jika dilihat lebih mendalam pada Gambar 3 dari 162 orang dari kelompok usia muda terdapat 58 orang yang tercatat sebagai korban meninggal atau sekitar 19,9% dari total korban kecelakaan. Sisanya sekitar 35,7% tidak meninggal.



Gambar 3. Jumlah korban berdasarkan status kecelakaan dan kelompok usia

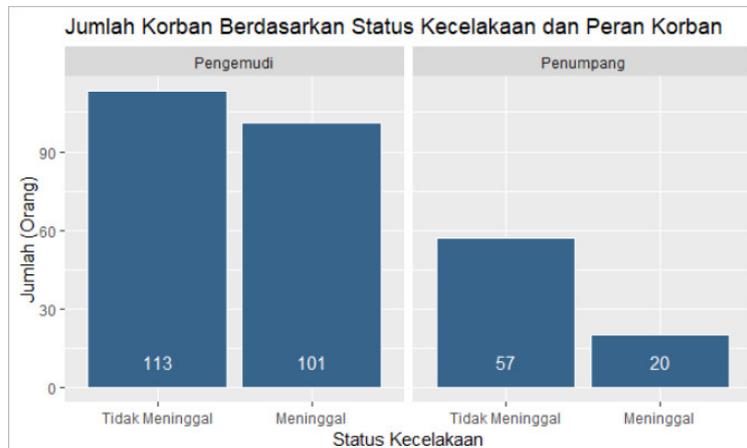


Gambar 4. Korban berdasarkan status kecelakaan dan jenis kelamin

Selain itu, jika dilihat status kecelakaan berdasarkan jenis kelamin yang terlihat pada Gambar 4 menunjukkan bahwa tingkat kecelakaan didominasi oleh para pria. Dari 233 pria

terdapat sekitar 97 pria yang berstatus meninggal atau sekitar 41,6%. Sementara korban kecelakaan dari kalangan wanita yang berstatus meninggal sebanyak 24 orang. jika dipersentasekan kurang lebih sebanyak 10,3% dari keseluruhan korban kecelakaa.

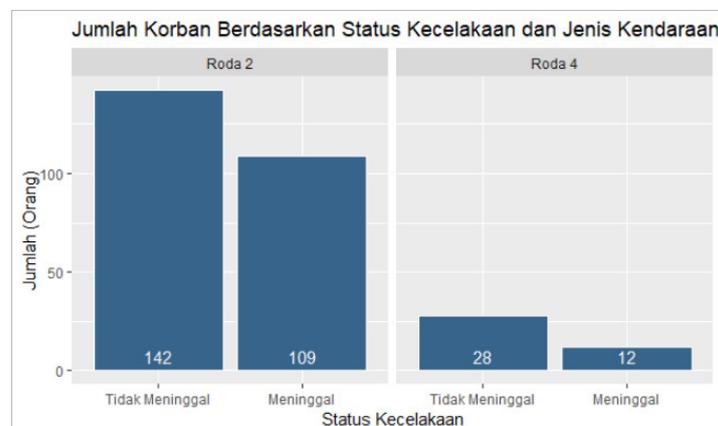
Karakteristik lainnya yaitu dilihat berdasarkan peran korban seperti yang terlihat pada Gambar 5.



Gambar 5. Korban kecelakaan berdasarkan peran

Dapat dilihat pada Gambar 5 di atas bahwa korban kecelakaan lalu lintas di Samarinda mayoritas berperan sebagai pengemudi. Dari 214 total pengemudi 101 diantaranya meninggal atau sekitar 47,2% dan sisanya sebesar 52,8% tercatat sebagai korban kecelakaan lalu lintas yang tidak meninggal.

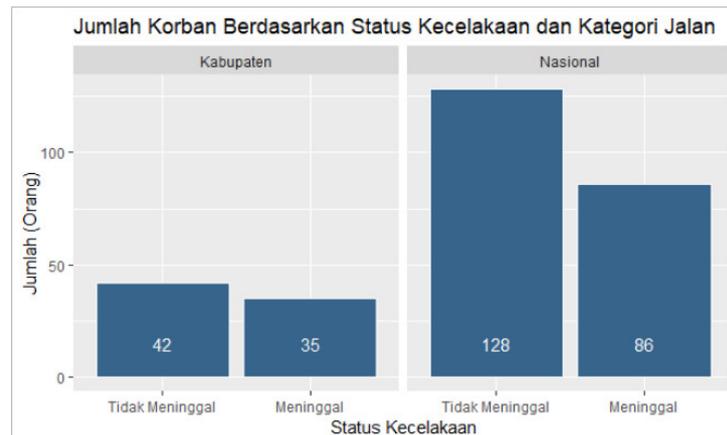
Karakteristik selanjutnya yaitu dilihat dari jenis kendaraan yang diperlihatkan pada Gambar 6 berikut.



Gambar 6. Korban kecelakaan berdasarkan jenis kendaraan

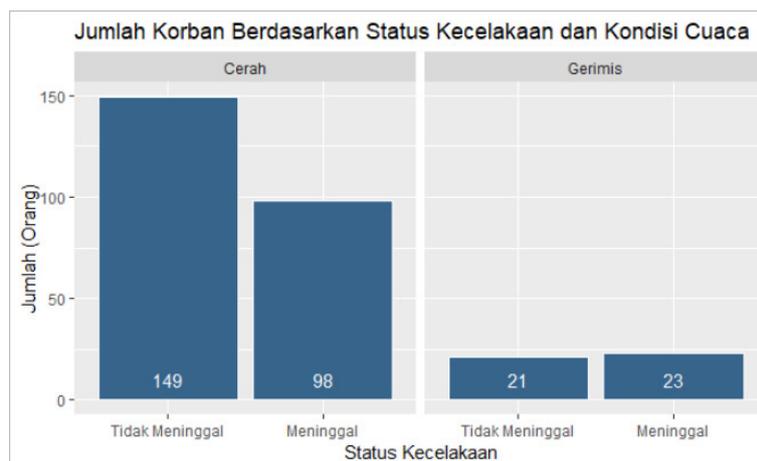
Dari grafik tersebut terlihat bahwa mayoritas korban kecelakaan merupakan pengguna kendaraan roda dua yaitu sebanyak 251 korban. 109 diantaranya atau sekitar 43,4% tercatat sebagai korban meninggal. Hal ini dapat menjadi *early warning* bagi para pengguna kendaraan roda dua agar tetap berhati-hati dalam berkendara karena memiliki peluang yang cukup besar sebagai korban kecelakaan lalu lintas.

Karakteristik berikutnya berkaitan dengan kategori jalan. Jumlah korban kecelakaan berdasarkan status kecelakaan dan kategori jalan diperlihatkan pada gambar 7 berikut.



Gambar 7. Korban kecelakaan berdasarkan Kategori jalan

Berdasarkan Gambar 7 dapat dilihat bahwa kategori jalan nasional menjadi kategori jalan yang memiliki intensitas yang tinggi terjadinya kecelakaan lalu lintas. Berdasarkan data tersebut dari 214 korban kecelakaan lalu lintas di jalan nasional 86 diantaranya meninggal atau sekitar 40,1%. Kemudian karakteristik korban kecelakaan yang terakhir berkaitan dengan kondisi cuaca. Jumlah korban kecelakaan berdasarkan kondisi cuaca dapat dilihat pada Gambar 8 berikut.

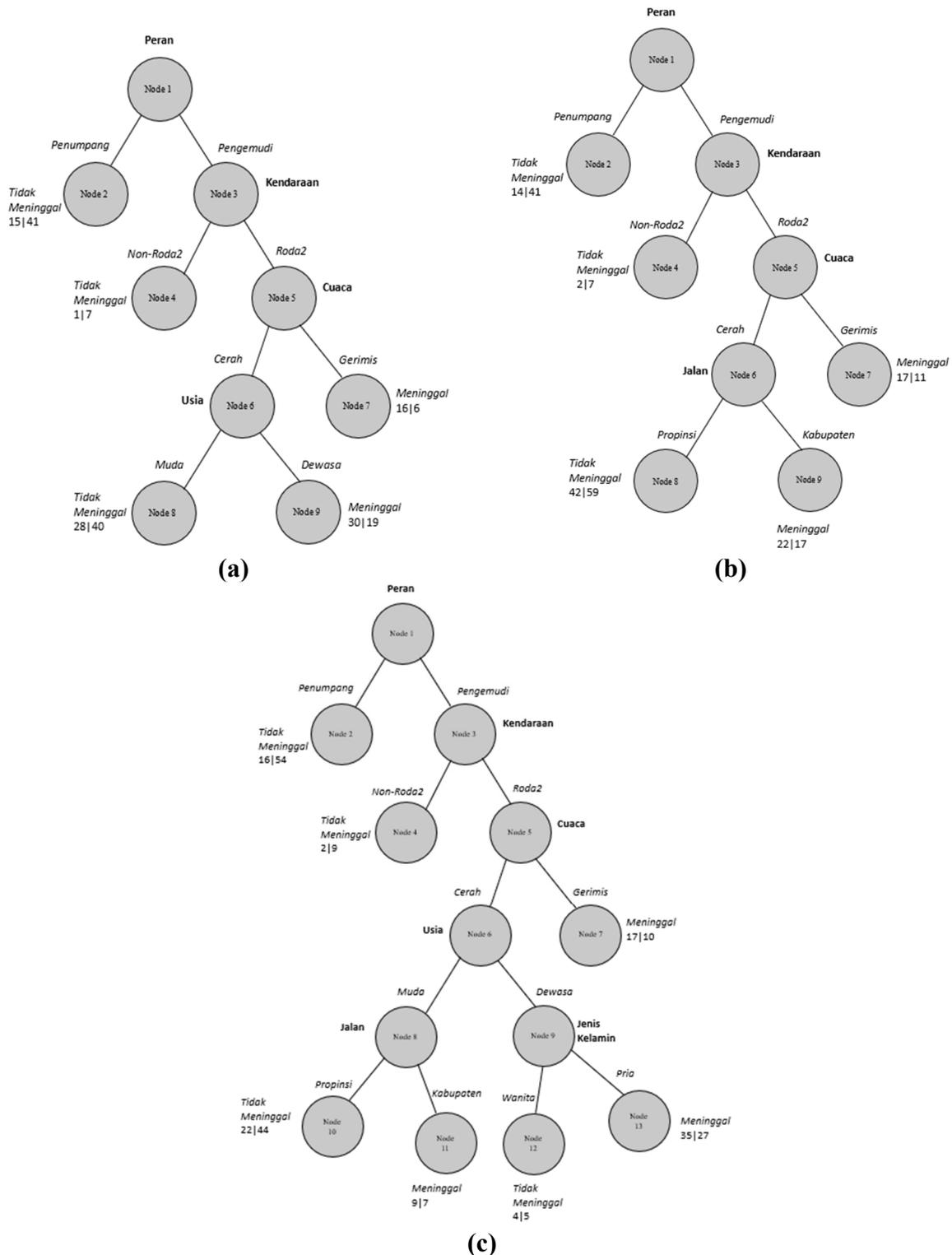


Gambar 8. Korban kecelakaan berdasarkan kondisi jalan

Gambar 8 terlihat bahwa kondisi cuaca yang cerah menjadi hal yang patut diwaspadai ketika berkendara. Tercatat bahwa dari 291 korban kecelakaan lalu lintas, terdapat sebanyak 247 korban kecelakaan yang terjadi ketika kondisi cuaca cerah dan 98 korban atau sekitar 33,7% diantaranya dinyatakan meninggal.

Pohon Keputusan Algoritma *QUEST*

Pohon keputusan yang dihasilkan dengan menggunakan algoritma *QUEST* disajikan pada Gambar 9 berikut.



Gambar 9. (a) Pohon keputusan dari proporsi 70:30, (b) pohon keputusan dari proporsi 80:20, (c) pohon keputusan dari proporsi 90:10

Perbandingan Performa Klasifikasi

Setelah mendapatkan hasil pohon keputusan dari ketiga proporsi data maka perlu dilakukan pengujian performa dengan menggunakan bantuan tabel *confusion matrix* untuk masing-masing pohon keputusan. Berikut merupakan tabel *confusion matrix* dari masing-masing proporsi diperlihatkan pada Tabel 2.

Tabel 2. *Confusion matrix*

70:30				80:20				90:10			
Prediksi				Prediksi				Prediksi			
Aktual	MD	TM	Total	Aktual	MD	TM	Total	Aktual	MD	TM	Total
MD	15	16	31	MD	5	19	24	MD	4	12	16
TM	24	33	57	TM	12	23	35	TM	7	7	14
Total	39	49	88	Total	17	42	59	Total	11	19	30

Ukuran performa klasifikasi yang digunakan adalah akurasi, sensitivitas, presisi, dan *F1-Score*. Hasil performa klasifikasi ditunjukkan pada Tabel 3.

Tabel 3. Perbandingan Kinerja Klasifikasi

Ukuran	70:30	80:20	90:10
Akurasi	54,54%	52,54%	36,67%
Presisi	38,46%	29,41%	36,36%
Sensitivitas	48,39%	20,83%	25%
Spesifisitas	57,90%	65,71%	50%
<i>F1-Score</i>	42,86%	24,39%	29,63%
<i>Missclassification cost</i>	45,45%	52,54%	63,30%

Hasil kinerja tersebut menunjukkan bahwa besarnya akurasi yang diperoleh dari model proporsi 70:30 lebih besar dibandingkan dengan proporsi lainnya, akan tetapi secara keseluruhan nilai akurasi masih terbilang rendah apabila digunakan sebagai acuan dalam memprediksi kelas dari data yang baru. Ukuran presisi digunakan untuk mengukur seberapa reliabel model dalam mengklasifikasikan data ke dalam label positif yang sebenarnya data tersebut adalah memang positif. Sehingga jika kita melihat ukuran presisi dari ketiga proporsi, maka proporsi 70:30 dapat memberikan nilai yang lebih unggul dibandingkan dengan proporsi lainnya. Meskipun ukuran tersebut masih terbilang sangat rendah dari yang diharapkan.

Ukuran sensitivitas digunakan untuk mengukur seberapa sensitif model dapat mendeteksi data berlabel positif. Jika kita melihat hasil sensitivitas maka proporsi 70:30 mampu memberikan nilai yang lebih unggul dibandingkan dua proporsi lainnya. Sehingga dapat dikatakan bahwa sensitivitas model dari proporsi 70:30 lebih baik dari model dua proporsi lainnya.

Ukuran spesifisitas adalah kebalikan dari sensitivitas yang digunakan untuk mengukur keandalan model untuk mendeteksi data berlabel negatif. Jika melihat hasil spesifisitas maka model dari proporsi 80:20 yang mampu memberikan hasil terbaik dibandingkan dengan model dari dua proporsi lainnya.

Ukuran *F1-Score* merupakan ukuran kinerja dengan membandingkan rata-rata presisi dan *recall* (sensitivitas). Jika mengacu pada hasil *F1-Score* pada tabel 3 menunjukkan bahwa model dari proporsi 70:30 dapat memberikan nilai yang lebih baik dibandingkan dengan model pada dua proporsi lainnya. Ukuran terakhir berkaitan dengan biaya yang dikeluarkan ketika model melakukan kesalahan dalam klasifikasi sehingga jika kita melihat hasil *missclassification cost* maka model dari proporsi 70:30 memberikan biaya yang lebih kecil dibandingkan dengan model dari dua proporsi lainnya ketika melakukan kesalahan klasifikasi.

Adapun jumlah aturan klasifikasi yang terbentuk berdasarkan struktur pohon keputusan dari algoritma QUEST adalah sebanyak 5 aturan, diantaranya adalah:

1. Jika Peran korban sebagai **penumpang** maka Status Korban Kecelakaan **tidak meninggal**.
2. Jika Peran korban sebagai **pengemudi**, dan jenis kendaraan adalah **non-roda 2** maka status korban kecelakaan **tidak meninggal**.

3. Jika Peran korban sebagai **pengemudi**, jenis kendaraan **roda 2**, cuaca **cerah**, usia **muda**, maka status korban kecelakaan **tidak meninggal**.
4. Jika Peran korban sebagai **pengemudi**, jenis kendaraan **roda 2**, cuaca **cerah**, usia **dewasa**, maka status korban kecelakaannya **meninggal**.
5. Jika Peran korban sebagai **pengemudi**, jenis kendaraan **roda 2**, cuaca **gerimis**, maka status korban kecelakaannya **meninggal**.

KESIMPULAN

Berdasarkan hasil analisis diperoleh kesimpulan bahwa model terbaik dihasilkan oleh proporsi 70:30 dengan nilai akurasi sebesar 54,54% dengan nilai *F1-Score* sebesar 42,86%. Model pohon keputusan dari proporsi 70:30 juga memberikan informasi yang berkaitan dengan variabel-variabel yang berpengaruh dalam menentukan status korban kecelakaan yaitu peran korban, jenis kendaraan, kondisi cuaca dan usia korban.

UCAPAN TERIMA KASIH

Ucapan terima kasih peneliti sampaikan kepada Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman yang telah memberikan bantuan dana melalui Skim Pembiayaan Penelitian dan Pengabdian Kepada Masyarakat pendanaan PNPB Tahun Anggaran 2023 sehingga penelitian ini dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] WHO, "World Health day: Road safety is no accident," 2004.
- [2] A. D. Saputra, "Studi Tingkat Kecelakaan Lalu Lintas Jalan di Indonesia Berdasarkan Data KNKT Dari Tahun 2007-2016," *Warta Penelitian Perhubungan*, vol. 29, no. 2, 2017.
- [3] BPS, Statistik Transportasi Darat, Jakarta: Badan Pusat Statistik, 2021.
- [4] W.-Y. Loh and Y.-S. Shih, "Split Selection Methods for Classification Trees," *Statistica Sinica*, vol. 7, pp. 815-840, 1997.
- [5] L. Susiana, I. T. Utami and J. , "Penerapan Metode Boosting Pada Cart Untuk Mengklasifikasikan Korban Kecelakaan Lalu Lintas Di Kota Palu," *Natural Science Journal of Science and Technology*, vol. 8, pp. 106-109, 2019.
- [6] A. Franseda, "Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas," *Jurnal Sistem dan Teknologi Informasi*, vol. 8, 2020.
- [7] A. Fauzy, Statistik Industri, Jakarta: Erlangga, 2008.
- [8] J. Supranto, Statistik: Teori dan Aplikasi Jilid 2, Jakarta: Erlangga, 2001.
- [9] J. L. Gastwirth, Y. R. Gel and W. Miao, "The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice," *Statistical Science*, vol. 24, no. 3, pp. 343-360, 2009.
- [10] W. Y. Loh and N. Vanichsetakul, "Tree-Structured Classification Via Generalized Discriminant Analysis," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 715-725, 1988.