

Faktor-Faktor yang Mempengaruhi Keparahan Korban Kecelakaan Lalu Lintas di Kota Surabaya dengan Pendekatan Bagging Regresi Logistik Ordinal

Wahyu Wulan Fitriah, Muhammad Mashuri, dan Irhamah

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111

E-mail: m_mashuri@statistika.its.ac.id

Abstrak—Kecelakaan lalu lintas adalah salah satu faktor pembunuh terbesar di Indonesia maupun di dunia. Hal ini memerlukan suatu penanganan agar korban kecelakaan tidak semakin memuncak. Oleh karena itu dilakukan penelitian untuk mengetahui faktor-faktor yang mempengaruhi keparahan korban kecelakaan lalu lintas di Surabaya pada tahun 2011 dengan menggunakan analisis regresi logistik ordinal dimana keparahan korban kecelakaan dibagi menjadi tiga, yaitu meninggal, luka berat dan luka ringan. Berdasarkan hasil penelitian diperoleh gambaran bahwa sebagian besar korban kecelakaan lalu lintas mengalami luka berat selanjutnya diikuti oleh korban luka ringan dan korban meninggal. Pada analisis regresi logistik ordinal diperoleh variabel prediktor yang signifikan mempengaruhi keparahan korban kecelakaan antara lain adalah jenis kecelakaan, peran korban, kendaraan lawan dan usia korban. Ketepatan klasifikasi yang dihasilkan untuk data training sebesar 50,08% dan untuk data testing sebesar 54,86%. Selanjutnya, data dianalisis dengan menggunakan bagging regresi logistik ordinal dengan harapan mampu meningkatkan ketepatan klasifikasi. Hasil analisis bagging regresi logistik ordinal pada data kombinasi 1 menunjukkan bahwa pada percobaan ke-10 untuk replikasi bootstrap 80 kali diperoleh ketepatan klasifikasi terbesar untuk data training sebesar 51,79% dan untuk data testing sebesar 56%. Hasil ini menunjukkan jikapenggunaan bagging memberikan peningkatan ketepatan klasifikasi.

Kata kunci : bagging, kecelakaan lalu lintas, ketepatan klasifikasi, regresi logistik ordinal.

I. PENDAHULUAN

KECELAKAAN lalu lintas merupakan masalah yang membutuhkan penanganan serius mengingat besarnya kerugian yang diakibatkannya. Indonesia merupakan salah satu negara dengan tingkat kecelakaan yang cukup tinggi. Menurut Dinas Perhubungan, kecelakaan lalu lintas menjadi penyebab kematian nomor tiga di Indonesia setelah serangan jantung dan stroke. Sementara itu Organisasi Kesehatan Dunia (WHO) meramalkan pada tahun 2030 kecelakaan lalu lintas akan menjadi faktor pembunuh manusia paling besar kelima di dunia [1]. Berdasarkan data POLRI pada tahun 2011, tercatat jika kematian akibat kecelakaan lalu lintas pada tahun 2010 mencapai 31.186 jiwa. Rata-rata 84 orang meninggal setiap harinya atau 3-4 orang meninggal setiap jamnya akibat kecelakaan lalu lintas. Korban dari kecelakaan tersebut sebesar 67% berada pada usia produktif (22-50) [2].

Surabaya sebagai salah satu kota besar di Indonesia memiliki tingkat kecelakaan lalu lintas yang cukup tinggi dibandingkan daerah lain. Menurut Direktur Lalu Lintas Polda Jatim pada tahun 2011, polrestabes Surabaya menempati urutan pertama dengan tingkat kejadian laka lintas tertinggi dibandingkan dengan polres lain [3].

Penelitian sebelumnya mengenai kecelakaan pernah dilakukan oleh Sitanggang pada tahun 2008. Penelitian tersebut menganalisis hubungan antara jumlah kecelakaan lalu lintas dengan faktor-faktor yang mempengaruhinya dengan menggunakan regresi linier berganda [4]. Penelitian lain tentang kecelakaan lalu lintas dilakukan yang menganalisis tentang pola tingkat keparahan korban kecelakaan lalu lintas di Surabaya dengan menggunakan regresi logistik multomial [5]. Model yang dihasilkan pada penelitian tersebut memiliki ketepatan klasifikasi sebesar 47,7% [5]. Ketepatan klasifikasi ini terbilang masih rendah karena masih kurang dari 50% sehingga harus diperbaiki karena keakuratan prediksi atau ketepatan dalam klasifikasi merupakan kriteria yang pertama kali dinilai pada sebuah model yang dihasilkan dari suatu metode atau algoritma klasifikasi.

Tugas akhir ini mengimplementasikan metode *Bootstrap Aggregating* (Bagging) pada kasus kecelakaan lalu lintas di Surabaya pada tahun 2011 dimana pola hubungan keparahan korban kecelakaan lalu lintas dengan faktor yang mempengaruhinya dianalisis dengan menggunakan regresi logistik ordinal sehingga bisa didapatkan faktor-faktor yang mempengaruhi keparahan korban kecelakaan. Metode *bootstrap aggregating* (bagging) adalah salah satu metode yang dapat digunakan untuk meningkatkan tingkat akurasi prediksi dan ketepatan dalam klasifikasi. Analisis ordinal ini digunakan karena respon terdiri dari tiga kategori yaitu meninggal, luka berat dan luka ringan dimana antar kategori dianggap memiliki tingkatan. Sedangkan penggunaan bagging digunakan dengan harapan mampu meningkatkan ketepatan klasifikasi sehingga bisa memperbaiki klasifikasi yang telah dilakukan sebelumnya [5].

II. TINJAUAN PUSTAKA

A. Regresi Logistik

Regresi logistik digunakan jika variabel respon bersifat kategorik (nominal atau ordinal) dengan variabel-variabel prediktor bersifat kontinu maupun kategorik.

Bentuk khusus dari model regresi logistik adalah sebagai berikut.

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad (1)$$

Regresi logistik ordinal adalah suatu metode analisis yang digunakan untuk mengetahui hubungan antara variabel respon dengan variabel prediktor dimana variabel responnya memiliki lebih dari dua kategori dan dalam setiap kategori memiliki tingkatan [6]. Model logit dalam regresi logistik ordinal disebut dengan *cumulative logit models*. Model logit dari regresi logistik ordinal tersebut adalah sebagai berikut.

$$\text{Logit } \psi_j = \ln \left[\frac{P(Y \leq j | \mathbf{x}_i)}{P(Y > j | \mathbf{x}_i)} \right] = \alpha_j + \sum_{k=1}^p \beta_k \mathbf{x}_{ik} \quad (2)$$

Dimana $j = 0, 1, 2, 3, \dots, J - 1$ dan J adalah banyaknya kategori pada variabel respon. Sedangkan k adalah banyaknya variabel prediktor. Persamaan dari peluang kumulatif kategori respon ke- j sebagai berikut.

$$P(Y \leq j | \mathbf{x}_i) = \frac{\exp(\alpha_j + \sum_{k=1}^p \beta_k \mathbf{x}_{ik})}{1 + \exp(\alpha_j + \sum_{k=1}^p \beta_k \mathbf{x}_{ik})} \quad (3)$$

Pada penelitian ini terdapat tiga kategori respon yaitu kategori 0, 1, dan 2 sehingga akan terbentuk dua fungsi logit dan dua fungsi peluang kumulatif. Berdasarkan kedua peluang kumulatif maka akan didapatkan peluang untuk masing-masing kategori respon sebagai berikut.

$$\begin{aligned} P(Y = 0 | \mathbf{x}_i) &= \varphi_0(\mathbf{x}) = P(Y \leq 0 | \mathbf{x}_i) \\ P(Y = 1 | \mathbf{x}_i) &= \varphi_1(\mathbf{x}) = P(Y \leq 1 | \mathbf{x}_i) - \varphi_0(\mathbf{x}) \\ P(Y = 2 | \mathbf{x}_i) &= \varphi_2(\mathbf{x}) = 1 - P(Y \leq 1 | \mathbf{x}_i) \end{aligned} \quad (4)$$

Pada penelitian ini metode yang digunakan untuk menaksir nilai β adalah Metode *Maksimum Likelihood Estimation* (MLE). Prinsip metode ini adalah menaksir nilai β dengan memaksimalkan fungsi *likelihood*. Pada penelitian ini, kategori variabel respon ada tiga yaitu 0, 1, dan 2 sehingga secara berturut-turut fungsi *likelihood* dan *log-likelihood*-nya adalah sebagai berikut.

$$l(\beta) = \prod_{i=1}^n \varphi_0(\mathbf{x}_i)^{y_{0i}} \varphi_1(\mathbf{x}_i)^{y_{1i}} \varphi_2(\mathbf{x}_i)^{y_{2i}} \quad (5)$$

$$L(\beta) = \sum_{i=1}^n y_{0i} \ln[\varphi_0(\mathbf{x}_i)] + y_{1i} \ln[\varphi_1(\mathbf{x}_i)] + y_{2i} \ln[\varphi_2(\mathbf{x}_i)] \quad (6)$$

Parameter-parameter diperoleh dengan menurunkan persamaan $L(\beta)$ terhadap parameter yang ingin diketahui dan selanjutnya turunan pertama tersebut disamadengankan nol. Selanjutnya berdasarkan teori *maximum likelihood*, taksiran

varian kovarian didapatkan dari turunan kedua fungsi *log-likelihood*-nya. Nilai parameter diestimasi dengan menggunakan metode numerik karena persamaannya bersifat nonlinier. Metode numerik tersebut adalah metode iterasi *Newton-Raphson* [10].

Setelah mendapatkan parameter maka selanjutnya adalah menguji signifikansi dari parameter yang telah diestimasi tersebut. Dalam model regresi logistik terdapat dua jenis pengujian yaitu pengujian secara parsial dan pengujian secara serentak. Dari estimasi model regresi logistik yang telah diperoleh, selanjutnya dicari seberapa besar kesesuaian model dalam menjelaskan variabel respon.

Interpretasi dalam regresi logistik menggunakan nilai *odds ratio* yang menunjukkan perbandingan tingkat kecenderungan dari kategori yang ada dalam satu variabel prediktor. *Odds ratio* untuk $Y \leq j$ terhadap $Y > j$ yang dihitung pada dua nilai (misal $x=a$ dan $x=b$) adalah sebagai berikut.

$$\psi(a, b) = \frac{\frac{P(Y \leq j | x = a)}{P(Y > j | x = a)}}{\frac{P(Y \leq j | x = b)}{P(Y > j | x = b)}} = \frac{\exp(\beta_{0j} + \beta_i(a))}{\exp(\beta_{0j} + \beta_i(b))} = \exp[\beta_i(a - b)] \quad (17)$$

Sehingga jika $a - b = 1$ maka $\psi = \exp(\beta_i)$.

Odds ratio untuk variabel prediktor kontinu memerlukan unit perubahan sebesar c , sehingga *odds ratio* diperoleh dengan $\exp(c\beta_i)$ [9].

Pengukuran yang biasa dipakai untuk melihat kesalahan klasifikasi adalah *apparent error rate* (APER).

B. Bagging (Bootstrap aggregating)

Bagging merupakan metode yang dapat memperbaiki hasil dari algoritma klasifikasi *machine learning* [4]. Bagging merupakan salah satu metode yang berdasar pada *ensemble method*, yaitu metode yang menggunakan kombinasi dari beberapa model. Bagging prediktor adalah metode yang digunakan untuk membangkitkan *multiple versions* dari prediktor dan menggunakannya untuk mendapatkan kumpulan prediktor. *Multiple versions* dibentuk dengan replikasi *bootstrap* dari sebuah data percobaan [7].

Algoritma pada bagging merujuk pada algoritma *Bootstrap* yaitu pengambilan sampel dengan pengembalian pada data *training* sehingga terbentuk variasi data baru. Peningkatan ketepatan klasifikasi dengan bagging memberikan hasil yang baik jika model yang dihasilkan dari perulangan berbeda satu sama lain. Oleh karena itu model yang sesuai untuk bagging adalah model yang tidak stabil dimana perubahan kecil pada data *training* mengakibatkan perubahan besar pada parameter model [8].

Banyaknya replikasi *bootstrap* sampling yang optimal telah banyak direkomendasikan oleh beberapa peneliti. Sutton (2004) merekomendasikan untuk melakukan replikasi sebanyak 25 atau 50 kali. Peningkatan akurasi akan terjadi

jika banyaknya replikasi ditingkatkan dari 50 ke 100 kali dan jika replikasinya ditingkatkan menjadi yang lebih dari 100 kali akan menghasilkan akurasi yang tidak lebih besar dibandingkan replikasi 100 kali [9]. Untuk memperoleh hasil yang lebih baik, maka replikasi *bootstrap* dilakukan sebanyak mungkin yaitu sebanyak 50 sampai 200 kali [10].

C. Kecelakaan Lalu Lintas

Menurut Kadiyali (1975) dalam [11], kecelakaan adalah tabrakan, *overtuning* atau slip yang terjadi di jalan terbuka pada lalu lintas umum yang menyebabkan luka, kematian / fatal, kerusakan pada kendaraan atau kerugian material. Kriteria korban kecelakaan lalu lintas menurut pasal 93 pada Peraturan Pemerintah No. 43 tahun 1993 tentang prasarana dan lalu lintas jalan yaitu :

- a. Korban meninggal
- b. Korban luka berat
- c. Korban luka ringan

Penyebab kecelakaan antara lain adalah: manusia, sarana dan prasarana (kendaraan dan jalan), alam atau lingkungan [1].

1. Faktor Manusia

Faktor manusia yang dicatat oleh kepolisian, meliputi jenis kelamin korban, usia korban, profesi korban, dan peran korban dalam berkendara.

2. Faktor Kendaraan

Faktor kendaraan yang paling sering terjadi adalah ban pecah, rem tidak berfungsi sebagaimana seharusnya, mesin yang tiba-tiba mati, lampu mati, dan berbagai penyebab lainnya.

3. Faktor Jalan

Jalan merupakan bagian dari sistem transportasi darat yang memegang peranan penting untuk kelancaran transportasi. Selain peran tersebut, jalan juga memegang peran penting sebagai salah satu penyebab kecelakaan lalu lintas.

4. Faktor Alam atau Lingkungan

Faktor lingkungan ini bisa berupa pengaruh cuaca yang tidak menguntungkan, kondisi lingkungan jalan, benda-benda asing yang ada di jalan dan lain sebagainya.

III. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder dari laporan Unit Laka Satlantas Polrestabes Surabaya selama tahun 2011. Variabel respon dalam penelitian ini adalah keparahan korban kecelakaan lalu lintas, dimana keparahan ini dibagi menjadi 3 kategori, yaitu:

- 0 = korban meninggal
- 1 = korban luka berat
- 2 = korban luka ringan

Variabel Prediktor yang dipakai dalam penelitian ini berdasarkan pada penelitian sebelumnya serta berdasarkan data yang tersedia di Unit Laka Satlantas Polrestabes Surabaya. Variabel-variabel tersebut adalah sebagai berikut.

1. Jenis kecelakaan (X_1)

- 1=Tabrakan belakang (TB)
- 2=Tabrakan depan (TD)

3=Tabrakan samping (TS)

4=Hilang kendali (HK)

5=Lain-lain

2. Jenis Kelamin (X_2)

1=Laki-Laki

2=Perempuan

3. Peran korban dalam kecelakaan (X_3)

1=Pengendara

2=Penumpang kendaraan selain pengendara

3=Pengguna jalan non penumpang kendaraan (penyeberang jalan, pejalan kaki, dll)

4. Lokasi Kecelakaan (X_4)

1=Surabaya Pusat

2=Surabaya Timur

3=Surabaya Barat

4=Surabaya Utara

5=Surabaya Selatan

5. Jenis kendaraan korban (X_5)

1=Sepeda motor (kendaraan bermotor roda dua atau tiga)

2=Kendaraan roda empat

3=Kendaraan dengan lebih dari empat roda

4=Lain-lain (sepeda angin, becak, atau kendaraan bukan bermotor lainnya)

6. Jenis kendaraan lawan (X_6)

1=Sepeda motor (kendaraan bermotor roda dua atau tiga)

2=Kendaraan roda empat

3=Kendaraan dengan lebih dari empat roda

4=Lain-lain (sepeda angin, becak, atau kendaraan bukan bermotor lainnya)

7. Jenis Kepadatan Jalan (X_7)

1=Padat kendaraan (antara pukul 06.00 WIB – 08.00 WIB, antara pukul 12.00 WIB – 13.30 WIB, antara pukul 16.00 WIB – 18.00 WIB)

2=Sepi (selain waktu padat)

8. Usia korban (X_8)

Langkah-langkah analisis yang dilakukan dalam penelitian ini dimulai dengan melakukan analisis statistika deskriptif. Selanjutnya menentukan model regresi logistik ordinal untuk mendapatkan faktor-faktor yang berpengaruh signifikan terhadap keparahan korban kecelakaan lalu lintas di Surabaya. Analisis ini diawali dengan membagi data menjadi data *training* dan data *testing*. Analisis yang terakhir adalah menggunakan bagging (*bootstrap aggregating*). Bagging dilakukan pada data pengamatan dimana banyaknya replikasi yang akan dilakukan ditentukan terlebih dahulu. Setiap sampel *bootstrap* selanjutnya diregresikan dengan menggunakan regresi logistik ordinal sehingga model bagging regresi logistik ordinal akhir didapat dari rata-rata setiap parameter pada setiap replikasi. Signifikansi parameter dihitung dengan menggunakan selang kepercayaan. Setelah didapatkan parameter-parameter yang signifikan, langkah terakhir adalah menghitung ketepatan klasifikasi dari model bagging regresi logistik ordinal.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Korban Kecelakaan Lalu Lintas di Surabaya Tahun 2011

Karakteristik korban kecelakaan lalu lintas di Surabaya untuk kawasan hukum Jajaran Polrestabes Surabaya pada tahun 2011 adalah bahwa dari 872 data kecelakaan yang telah diambil, ada sebesar 48,05% yang mengalami luka berat. Selanjutnya diikuti oleh korban yang mengalami luka ringan yaitu sebesar 30,62% dan korban meninggal sebesar 21,33%. Sebagian besar korban mengalami tabrak samping yaitu sebesar 38,53%. Sebanyak 71,1% korban berjenis kelamin laki-laki, hampir tiga kali dari korban yang berjenis kelamin perempuan. Kebanyakan dari korban berperan sebagai pengendara yaitu sebesar 68%. Lokasi yang paling sering terjadi kecelakaan adalah Surabaya Selatan yaitu sebesar 39,3%. Kendaraan korban sebagian besar adalah sepeda motor yaitu sebesar 77,9%. Sama halnya dengan kendaraan korban, kendaraan yang paling banyak digunakan oleh lawan juga sepeda motor yaitu 52,9%. Hal itu mengingat jika sepeda motor adalah kendaraan yang paling umum digunakan oleh masyarakat di Indonesia saat ini. Diketahui pula jika kecelakaan lalu lintas banyak terjadi saat jalan dalam keadaan sepi. Selain itu berdasarkan usia korban diketahui bahwa dari 872 korban, median usia yang menjadi korban kecelakaan adalah korban yang berusia 30 tahun.

Jika dihitung nilai ekspektasi atau nilai harapan, maka pada variabel kendaraan korban yaitu pada kategori mobil roda 4 dan mobil roda lebih dari 4 memiliki nilai harapan yang sangat kecil yaitu kurang dari 5. Oleh karena itu, kategori tersebut harus digabung agar nilai ekspektasinya bisa lebih besar. Namun jika kategori kendaraan dengan roda 4 dan roda lebih dari 4 digabung, nilai ekspektasi dari penggabungan itu masih terlalu kecil sehingga masih harus digabung dengan kategori lain. Oleh karena itu kategori selain sepeda motor akan dijadikan satu kategori yaitu menjadi kategori jenis kendaraan lainnya.

B. Analisis Regresi Logistik Ordinal

Sebelum dilakukan analisis dengan menggunakan regresi logistik ordinal terlebih dahulu data dibagi menjadi dua yaitu data *training* dan data *testing*. Pada penelitian ini pembagian data terdiri dari dua kombinasi. Kombinasi 1 terdiri dari 80% data *training* dan 20% data *testing*. Selanjutnya kombinasi 2 terdiri dari 90% data *training* dan 10% data *testing*. Pada analisis regresi logistik ordinal terdapat dua pengujian yaitu pengujian secara serentak dan pengujian secara parsial.

Tabel 1 Hasil Uji Serentak Kedua Kombinasi

	Kombinasi 1	Kombinasi 2
G	144,711	152,787
P-Value	0,000	0,000

Berdasarkan Tabel 1 dapat diketahui jika kedua kombinasi menghasilkan *p-value* sebesar 0,000. Hasil tersebut mengindikasikan jika hipotesis nol ditolak karena *p-value* < 0,1 sehingga dapat dikatakan jika pada masing-masing kombinasi paling tidak ada satu variabel prediktor yang berpengaruh

secara signifikan terhadap variabel respon sehingga perlu dilakukan pengujian parsial untuk melihat variabel mana yang signifikan. Pengujian parsial untuk masing-masing kombinasi data dijelaskan secara rinci berikut ini.

a. Pengujian Parsial Kombinasi 1

Berdasarkan Tabel 2, variabel prediktor yang berpengaruh signifikan terhadap keparahan korban kecelakaan untuk data kombinasi pertama adalah variabel X_1 , X_3 , X_6 , dan X_8 .

Berdasarkan estimasi parameter yang ditunjukkan pada Tabel 2, maka fungsi logitnya dapat ditulis sebagai berikut.

$$\text{Logit 1} = -2,632 + 4,563X_1(4) + 0,943X_1(5) - 0,784X_3(2) + 0,509X_6(2) + 1,494X_6(3) + 0,016X_8$$

$$\text{Logit 2} = 4,563X_1(4) + 0,943X_1(5) - 0,784X_3(2) + 0,509X_6(2) + 1,494X_6(3) + 1,494X_6(3) + 0,016X_8$$

Tabel 2 Uji Signifikansi Parameter Secara Parsial untuk Data Kombinasi 1

Variabel Prediktor	β	P-value	Odds Ratio
Konstanta(1)	-2,632	0,000*	
Konstanta(2)	-0,203	0,546	
$X_1(2)$	-0,070	0,748	0,93
$X_1(3)$	0,085	0,679	1,09
$X_1(4)$	4,563	0,000*	95,84
$X_1(5)$	0,944	0,002*	2,57
$X_2(2)$	-0,142	0,405	0,87
$X_3(2)$	-0,784	0,000*	0,46
$X_3(3)$	-0,129	0,725	0,88
$X_4(2)$	-0,039	0,890	0,96
$X_4(3)$	0,402	0,130	1,49
$X_4(4)$	0,406	0,150	1,50
$X_4(5)$	0,273	0,241	1,31
$X_5(2)$	0,497	0,126	1,64
$X_6(2)$	0,509	0,005*	1,66
$X_6(3)$	1,494	0,000*	4,46
$X_6(4)$	-1,075	0,120	0,34
$X_7(2)$	-0,059	0,715	0,94
X_8	0,016	0,002*	1,02

Berdasarkan nilai *odds ratio* tersebut, maka interpretasi yang dapat diberikan adalah sebagai berikut.

- Jenis Kecelakaan, jika dilihat dari *odds ratio* pada variabel jenis kecelakaan hilang kendali maka orang yang mengalami jenis tabrakan hilang kendali ataupun jenis tabrakan lainnya, memiliki resiko meninggal dunia lebih besar dibanding orang yang mengalami tabrak belakang.
- Peran Korban, jika dilihat dari *odds ratio* pada variabel ini maka orang yang berperan sebagai pengendara memiliki resiko meninggal dunia 2 kali lipat lebih besar dibanding orang yang berperan sebagai penumpang.
- Kendaraan Lawan, jika dilihat dari *odds ratio* pada variabel ini maka orang yang bertabrakan dengan mobil roda 4 dan mobil dengan roda lebih dari 4 memiliki resiko meninggal yang lebih besar daripada orang yang bertabrakan dengan sepeda motor.
- Usia, jika dilihat dari variabel usia maka orang yang berusia lebih tua satu tahun memiliki peluang untuk meninggal dunia 1,02 kali dari orang yang lebih muda satu tahun dengan syarat faktor resiko lain yang dimiliki adalah sama. Dengan kata lain, orang yang berusia lebih tua memiliki peluang sedikit lebih besar untuk meninggal

dunia dibanding dengan orang yang lebih muda satu tahun.

b. Pengujian Parsial Kombinasi 2

Berdasarkan Tabel 3 tersebut diketahui jika variabel prediktor yang berpengaruh signifikan terhadap keparahan korban kecelakaan untuk data kombinasi 2 adalah variabel $X_1, X_3, X_4, X_6,$ dan X_8 .

Berdasarkan estimasi parameter yang ditunjukkan pada Tabel 3, maka fungsi logitnya dapat ditulis sebagai berikut.

$$\text{Logit 1} = -2,672 + 4,073X_1(4) + 0,966X_1(5) - 0,911X_3(2) + 0,418X_4(3) + 0,519X_6(2) + 1,348X_6(3) + 0,017X_8$$

$$\text{Logit 2} = 4,073X_1(4) + 0,966X_1(5) - 0,911X_3(2) + 0,418X_4(3) + 0,519X_6(2) + 1,348X_6(3) + 0,017X_8$$

Tabel 3 Uji Signifikansi Parameter Secara Parsial untuk Data Kombinasi 2

Variabel Prediktor	β	P-value	Odds Ratio
Konstanta(1)	-2,672	0,000*	
Konstanta(2)	-0,236	0,457	
$X_1(2)$	-0,049	0,814	0,95
$X_1(3)$	0,108	0,585	1,11
$X_1(4)$	4,073	0,000*	58,76
$X_1(5)$	0,966	0,001*	2,63
$X_3(2)$	-0,100	0,540	0,90
$X_3(3)$	-0,911	0,000*	0,40
$X_3(4)$	0,091	0,788	1,10
$X_4(2)$	-0,064	0,809	0,94
$X_4(3)$	0,419	0,096*	1,52
$X_4(4)$	0,165	0,533	1,18
$X_4(5)$	0,160	0,465	1,17
$X_5(2)$	0,292	0,325	1,34
$X_6(2)$	0,519	0,003*	1,68
$X_6(3)$	1,348	0,000*	3,85
$X_6(4)$	-0,665	0,293	0,51
$X_7(2)$	0,020	0,897	1,02
X_8	0,017	0,000*	1,02

Berdasarkan nilai *odds ratio*, maka interpretasi yang dapat diberikan adalah sebagai berikut.

- Jenis Kecelakaan, jika dilihat dari *odds ratio* pada variabel jenis kecelakaan hilang kendali maka orang yang mengalami jenis tabrakan hilang kendali ataupun jenis tabrakan lainnya, memiliki resiko meninggal dunia lebih besar dibanding orang yang mengalami tabrak belakang.
- Peran Korban, jika dilihat dari *odds ratio* pada variabel ini maka orang yang berperan sebagai pengemudi memiliki resiko meninggal dunia 2 kali lipat lebih besar dibanding orang yang berperan sebagai penumpang.
- Lokasi Kecelakaan, jika dilihat dari *odds ratio* pada variabel ini maka orang yang kecelakaan di Surabaya Barat memiliki resiko meninggal dunia hampir 2 kali lipat lebih besar dibanding orang yang kecelakaan di Surabaya Pusat.
- Kendaraan Lawan, jika dilihat dari *odds ratio* pada variabel ini maka orang yang bertabrakan dengan mobil roda 4 ataupun mobil dengan roda lebih dari 4 memiliki resiko meninggal yang lebih besar daripada orang yang bertabrakan dengan sepeda motor.
- Usia, jika dilihat dari variabel usia maka orang yang berusia lebih tua satu tahun memiliki peluang untuk meninggal dunia 1,02 kali dari orang yang lebih muda satu tahun

dengan syarat faktor resiko lain yang dimiliki adalah sama. Dengan kata lain, orang yang berusia lebih tua memiliki peluang sedikit lebih besar untuk meninggal dunia dibanding dengan orang yang lebih muda satu tahun.

Setelah didapatkan estimasi parameter dan telah dibentuk model logit maka selanjutnya dilakukan pengujian kesesuaian model.

Tabel 4 Uji Kesesuaian Model

	Kombinasi 1		Kombinasi 2	
	Chi-Square	P-value	Chi-Square	P-value
Pearson	1348,65	0,279*	1498,12	0,305

Keterangan : *Gagal tolak hipotesis nol pada $\alpha = 10\%$

Berdasarkan Tabel 4 diketahui jika *p-value* uji Pearson chi-square dari kedua kombinasi data sudah lebih dari $\alpha=0,1$ sehingga keputusannya adalah gagal tolak hipotesis nol. Hal ini berarti model yang didapatkan dari masing-masing kombinasi data telah sesuai.

Kedua kombinasi data, sama-sama telah memenuhi segala pengujian yang telah dilakukan sebelumnya sehingga nilai ketepatan klasifikasi menjadi kriteria yang bisa menentukan kombinasi data yang menghasilkan model yang paling sesuai. Ketepatan klasifikasi dari masing-masing kombinasi data ditampilkan pada Tabel 5.

Berdasarkan Tabel 5 diketahui bahwa dari persentase ketepatan klasifikasi pada data *training* maka kombinasi data yang ketepatan klasifikasinya tertinggi adalah data kombinasi 2 yaitu sebesar 52,48%. Sedangkan jika dilihat dari persentase ketepatan klasifikasi pada data *testing* maka kombinasi data yang menghasilkan ketepatan klasifikasi tertinggi adalah data kombinasi 1 yaitu sebesar 54,86%. Penentuan model yang paling sesuai adalah berdasarkan besarnya ketepatan klasifikasi pada data *testing* karena data *testing* merupakan suatu data yang dianggap baru yang digunakan untuk mengecek suatu model yang dihasilkan oleh data *training*. Oleh karena itu model regresi logistik yang paling sesuai dari kedua kombinasi data tersebut adalah model regresi logistik yang dihasilkan oleh data kombinasi 1.

Tabel 5 Ketepatan Klasifikasi dari Kedua Kombinasi Data

	Kombinasi 1	Kombinasi 2
Data Training	51,08%	52,48%
Data Testing	54,86%	47,13%

C. Bagging (Bootstrap Aggregating)

Sama halnya dengan regresi logistik ordinal, pembentukan model bagging ini juga akan menggunakan data *training* dari setiap kombinasi data. Kemudian data direplikasi bootstrap sebanyak 50 dan 80 kali dimana pada setiap replikasi akan dilakukan 10 kali percobaan bagging. Hasil ketepatan klasifikasi dari model bagging regresi logistik ordinal untuk keparahan korban kecelakaan di Surabaya pada setiap replikasi dan setiap percobaan untuk data kombinasi 1 ditampilkan pada Tabel 6.

Tabel 6 Hasil Ketepatan Klasifikasi dengan Bagging untuk Data Kombinasi 1

Percobaan	Replikasi 50		Replikasi 80		Replikasi 100	
	Training (%)	Testing (%)	Training (%)	Testing (%)	Training (%)	Testing (%)
1	52,37	55,43	51,79	54,86	52,37	53,71
2	51,65	55,43	52,37	54,29	51,51	53,71
3	51,36	54,86	51,79	54,86	51,79	56,00
4	51,51	54,86	52,08	54,29	51,94	53,71
5	51,51	53,14	52,22	55,43	52,37	54,86
6	51,79	54,86	51,65	55,43	52,08	54,29
7	51,94	54,86	51,36	55,43	52,37	54,29
8	51,36	53,71	51,94	54,29	52,51	54,29
9	51,94	54,86	51,94	54,29	51,79	55,43
10	51,65	54,86	51,79	56,00	51,79	54,86
Rata-Rata	51,71	54,69	51,89	54,91	52,05	54,51

Berdasarkan Tabel 7 diketahui bahwa dari ketiga jenis replikasi *bootstrap* dan ke-10 percobaan bagging pada setiap replikasi, ketepatan klasifikasi terbesar dihasilkan oleh 80 replikasi *bootstrap* percobaan yang ke-10. Ketepatan klasifikasi pada data *training* dan data *testing* yang dihasilkan masing-masing adalah 51,79% dan 56%. Model bagging ini dapat meningkatkan ketepatan klasifikasi data *testing* dari model data tunggal yang telah didapatkan sebelumnya yaitu dari 54,86% menjadi 56%.

Estimasi parameter model bagging merupakan nilai rata-rata estimasi parameter dari setiap replikasi *bootstrap* yang telah diuji signifikansinya dengan selang kepercayaan. Pada 80 replikasi *bootstrap* percobaan 10 diperoleh ketepatan klasifikasi terbesar yang meningkatkan ketepatan klasifikasi pada model tunggal. Fungsi logit yang diperoleh untuk klasifikasi adalah sebagai berikut.

$$\begin{aligned} \text{Logit 1} &= -2,752 - 0,052X_1(2) + 0,099X_1(3) + 5,981X_1(4) + \\ & 1,063X_1(5) - 0,152X_2(2) - 0,808X_3(2) - 0,123X_3(3) + \\ & 0,424X_4(3) + 0,399X_4(4) + 0,29X_4(5) + 0,524X_5(2) + \\ & 0,497X_6(2) + 1,515X_6(3) - 2,273X_6(4) + 0,016X_8 \\ \text{Logit 2} &= -0,264 - 0,052X_1(2) + 0,099X_1(3) + 5,981X_1(4) + \\ & 1,063X_1(5) - 0,152X_2(2) - 0,808X_3(2) - 0,123X_3(3) + \\ & 0,424X_4(3) + 0,399X_4(4) + 0,29X_4(5) + 0,524X_5(2) + \\ & 0,497X_6(2) + 1,515X_6(3) - 2,273X_6(4) + 0,016X_8 \end{aligned}$$

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Karakteristik korban kecelakaan lalu lintas di Surabaya untuk kawasan hukum Jajaran Polrestabes Surabaya pada tahun 2011 adalah bahwa dari 872 data kecelakaan yang telah diambil sebagian besar korban mengalami luka berat yang diikuti oleh korban luka ringan dan korban meninggal. Pada analisis regresi logistik ordinal, didapatkan kombinasi data yang menghasilkan model yang paling sesuai yaitu data kombinasi 1. Variabel yang berpengaruh terhadap keparahan korban kecelakaan pada data kombinasi 1 adalah jenis kecelakaan, peran korban dalam kecelakaan, kendaraan lawan dan usia korban. Ketepatan klasifikasi yang dihasilkan untuk data *training* sebesar 51,08% dan untuk data *testing* sebesar 54,86%.

Hasil analisis bagging regresi logistik ordinal pada data kombinasi 1 menunjukkan bahwa pada percobaan ke-10 untuk

replikasi *bootstrap* 80 kali diperoleh ketepatan klasifikasi untuk data *training* sebesar 51,79% dan untuk data *testing* sebesar 56%. Hasil ini menunjukkan jika bagging telah meningkatkan ketepatan klasifikasi untuk data kombinasi 1.

B. Saran

Penggunaan metode bagging pada penelitian ini sudah mampu meningkatkan ketepatan klasifikasi namun masih belum cukup signifikan. Hal ini kemungkinan karena metode yang digunakan masih belum sesuai dengan data ataupun karena variabel-variabel prediktor yang digunakan masih belum lengkap. Oleh karena itu, pada penelitian berikutnya diharapkan untuk bisa menggunakan variabel yang lebih lengkap. Selain itu, penggunaan metode lain yang bisa meningkatkan ketepatan klasifikasi juga baik untuk dilakukan seperti dengan metode *boosting* atau metode non parametrik seperti *Multivariate Adaptive Regression Spline* (MARS).

DAFTAR PUSTAKA

- [1] Media Indonesia (2011). Pemerintah Luncurkan Aksi Keselamatan Jalan [Online]. Available: <http://www.mediaindonesia.com/read/2011/06/17/234834/35/5/Pemerintah-Luncurkan-Aksi-Keselamatan-Jalan>.
- [2] Pikiran Rakyat (2011). Tiap Jam 3-4 Orang Meninggal Akibat Kecelakaan Lalu Lintas [Online]. Available: <http://www.pikiran-rakyat.com/node/144627>
- [3] Z. Effendi (2011). Angka Kecelakaan di Surabaya Masih Tinggi [Online]. Available: [<http://us.surabaya.detik.com/read/2011/03/23/011506/1599124/466/angka-kecelakaan-di-surabaya-masih-tak-tertandingi>].
- [4] J. Sitanggang, "Analisis Faktor-Faktor yang Mempengaruhi Jumlah Kecelakaan Lalu Lintas di Kotamadya Medan," Medan: Universitas Sumatera Utara. (belum dipublikasikan).
- [5] L. N. Afidah, "Pola Tingkat Keparahan Korban Kecelakaan Lalu Lintas di Surabaya dengan Menggunakan Regresi Logistik Multinomial," Surabaya: Institut Teknologi Sepuluh Nopember (belum dipublikasikan).
- [6] D. W. Hosmer dan Lemenshow, *Applied Logistic Regression*, USA : John Wiley and Sons (2000).
- [7] L. Breiman, "Bagging Predictors," *Machine Learning*, Vol. 24. (1996) 123-140.
- [8] Z. Barutcuoglu dan E. Alpaydin, *A Comparison of Model Aggregation Methods for Regression*. Berlin: Springer-Verlag (2003).
- [9] T. Hastie, R. Tibshirani, dan J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, New York: Springer-Verlag (2001).
- [10] B. Efron dan R. J. Tibshirani, *An Introduction to the Bootstrap*. New York : Chapman and Hall (1993).
- [11] A. A. Samekto, " Studi tentang Karakteristik Korban Kecelakaan Lalu Lintas di Kota Semarang," *Jurnal Sains dan Teknik Maritim*, Vol. VII No 2. (2009).