TUGAS AKHIR – SS 141501

KLASIFIKASI FAKTOR-FAKTOR YANG MEMPENGARUHI KORBAN KECELAKAAN LALU LINTAS DI SURABAYA DENGAN PENDEKATAN REGRESI LOGISTIK MULTINOMIAL DAN *RANDOM FORESTS*

Ita Rakhmawati
NRP 1311100 058

Dosen Pembimbing
Dr.rer. pol. Heri Kuswanto, S.Si., M.Si.

JURUSAN STATISTIKA
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2015

FINAL PROJECT – SS 141501

CLASSIFICATION OF FACTORS AFFECTING TRAFFIC ACCIDENT VICTIMS IN SURABAYA USING MULTINOMIAL LOGISTIC REGRESSION AND *RANDOM FORESTS*

Ita Rakhmawati
NRP 1311 100 058

Supervisor
Dr. rer. pol. Heri Kuswanto, S.Si., M.Si.

DEPARTEMENT OF STATISTICS
Faculty of Mathematics and Natural Sciences
Institut Teknologi Sepuluh Nopember
Surabaya 2015

LEMBAR PENGESAHAN

KLASIFIKASI FAKTOR-FAKTOR YANG MEMPENGARUHI KORBAN KECELAKAAN LALU LINTAS DI SURABAYA DENGAN PENDEKATAN REGRESI LOGISTIK MULTINOMIAL DAN *RANDOM FORESTS*

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat Kelulusan
Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :
ITA RAKHMAWATI
1311 100 058

Disetujui oleh Pembimbing Tugas Akhir :

Dr. rer. pol. Heri Kuswanto, S.Si., M.Si.
NIP. 19820326 200312 1 004

Mengetahui
Ketua Jurusan Statistika FMIPA-ITS

Dr. Muhammad Mashuri, MT
NIP. 19620408 198701 1 001

SURABAYA, JANUARI 2015

KLASIFIKASI FAKTOR-FAKTOR YANG MEMPENGARUHI KORBAN KECELAKAAN LALU LINTAS DI SURABAYA DENGAN PENDEKATAN REGRESI LOGISTIK MULTINOMIAL DAN *RANDOM FORESTS*

Nama Mahasiswa : Ita Rakhmawati
NRP : 1311 100 058
Jurusan : Statistika
Dosen Pembimbing : Dr. rer. pol. Heri Kuswanto, S.Si., M.Si

Abstrak

Surabaya merupakan kota yang padat penduduk disebabkan karena Surabaya sendiri sudah termasuk dalam 5 kota besar yang berkembang cukup pesat. Maka dapat dipastikan jumlah pendatang melonjak naik. Kepadatan penduduk juga akan menyebabkan kepadatan kendaraan yang tinggi sehingga mempunyai resiko kecelakaan yang tinggi. Berdasarkan data polrestabes Surabaya, pada tahun 2013 ada 854 kejadian kecelakaan lalu lintas di Surabaya. Adapun korban meninggal, luka berat, luka ringan akibat kecelakaan lalu lintas ada 208 korban, 317 korban dan 680 korban pada tahun 2013. Tingginya angka kecelakaan di Surabaya dengan berbagai kategorinya menjadi dasar pentingnya dilakukan penelitian mengenai faktor yang mempengaruhi kategori kecelakaan tersebut. Penelitian ini menggunakan tujuh variabel prediktor yang diduga mempengaruhi korban kecelakaan lalu lintas. Data yang digunakan adalah data pada tahun 2013. Berdasarkan hasil analisis didapatkan variabel jenis kecelakaan, peran korban, kendaraan lawan, waktu dan usia berpengaruh signifikan terhadap korban kecelakaan lalu lintas. Ketepatan klasifikasi korban kecelakaan dengan regresi logistik multinomial sebesar 63,33 persen sedangkan random forests sebesar 58,33 persen sehingga metode yang baik digunakan pada kasus korban kecelakaan lalu lintas adalah regresi logistik multinomial.

Kata kunci : Korban Kecelakaan Lalu Lintas, Random Forests, Regresi Logistik Multinomial

CLASSIFICATION OF FACTORS AFFECTING TRAFFIC ACCIDENT VICTIMS IN SURABAYA USING MULTINOMIAL LOGISTIC REGRESSION AND RANDOM FORESTS

Name of Student : Ita Rakhmawati
NRP : 1311 100 058
Departement : Statistics
Supervisor : Dr. rer. pol. Heri Kuswanto, S.Si., M.Si

Abstract

Surabaya is a densely populated city due to Surabaya it self is included in the five major cities are growing quite rapidly. It is certain that the number of arrivals shot up. Population density will also cause a high density of vehicles that have a high risk of accidents. Based on data Polrestabes Surabaya in 2013 there were 854 traffic accidents in Surabaya. The fatalities, serious injuries, minor injuries caused by traffic accidents there are 208 victims, 317 victims and 680 victims in 2013. The high number of accidents in Surabaya with various categories became the basis of the importance of research on the factors influencing the accident categories. This research used seven predictor variables suspected to affect the victims of traffic accidents. The research use data in 2013. Based on the analysis results obtained variable types of accidents, the role of the victim, the opponent vehicle, time and age have a significant effect on the victims of traffic accidents. Classification accuracy of accident victims with multinomial logistic regression of 63.33 percent while the random forests of 58.33 percent, so a good method to use in case of traffic accident victims is multinomial logistic regression.

Keywords : Multinomial Logistics Regression, Random Forests, Traffic Accidents Victims

KATA PENGANTAR

Puji syukur kepada Allah SWT, yang telah memberikan rahmat sehingga penyusunan Tugas Akhir ini dapat terselesaikan tepat waktu. Tugas Akhir yang berjudul "***Klasifikasi Faktor-Faktor yang Mempengaruhi Korban Kecelakaan Lalu Lintas di Surabaya dengan Pendekatan Regresi Logistik Multinomial dan Random Forests***" ini disusun untuk memenuhi salah satu syarat kelulusan Program Studi S1 Jurusan Statistika FMIPA ITS.

Dengan terselesaikannya penyusunan Tugas Akhir ini, penulis mengucapkan terima kasih kepada:

1. Allah swt yang telah memberikan kemudahan dan kelancaran dalam menjalankan Tugas Akhir sampai dengan penyusunan laporan.
2. Dr. rer. pol. Heri Kuswanto, S.Si., M.Si. selaku dosen pembimbing yang selalu memberikan pengarahan kepada penulis selama penyusunan laporan Tugas Akhir.
3. Dr. Muhammad Mashuri, MT selaku Ketua Jurusan Statistika ITS.
4. Dra. Lucia Aridinanti, MS selaku Kaprodi S1 Jurusan Statistika ITS.
5. Ir. Dwi Atmono Agus Widodo, MiKom. dan Dr. Brodjol Sutijo Suprih Ulama, M.Si. selaku dosen penguji yang senantiasa memberikan kritik dan saran demi kesempurnaan Tugas Akhir ini.
6. Dr. Ir. Setiawan, MS selaku dosen wali yang telah memberikan pengarahan selama proses perkuliahan.
7. Dr. Vita Ratnasari, S.Si., M.Si atas semangat, perhatian, nasehat yang telah diberikan.
8. Seluruh dosen Statistika ITS dan dosen non Statistika ITS yang telah memberikan ilmu-ilmu yang tiada ternilai harganya dan segenap karyawan jurusan Statistika ITS.
9. Kasatlantas Polrestabes Surabaya, Bamin Unit Laka Satlantas Polrestabes Surabaya, serta seluruh staf administrasi Unit Laka Lantas Polrestabes Surabaya.

10. Bapak, Ibu, Mbak, Adik sebagai keluarga yang senantiasa memberikan dukungan baik moril maupun materil dan juga doa yang tiada henti.
11. Iin, Irma, Niken, Rivani sebagai sahabat yang selalu memberikan dorongan, bantuan dan semangat selama ini.
12. Eva Arum, Charisma, Olivia, Indana sebagai teman yang telah memberikan bantuan dalam menyelesaikan Tugas Akhir ini.
13. Mbak Dian, Mbak Hani, Mbak Oktiva, Mbak Wahyu, Mas Affandi, Mas Novri, Mas Jamal dan Mas Adi yang telah memberikan banyak pemahaman mengenai topik Tugas Akhir ini.
14. Seluruh teman-teman mahasiswa Statistika ITS khususnya S1 angkatan 2011 yang selalu memberikan semangat dan dorongan hingga terselesaikannya Tugas Akhir ini.
15. Semua pihak yang telah membantu dalam penulisan Tugas Akhir ini yang tidak dapat disebutkan satu per satu.

Penulis menyadari sepenuhnya bahwa laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis menerima kritik dan saran yang membangun bagi perbaikan di masa yang akan datang. Semoga laporan ini bermanfaat bagi penelitian selanjutnya.

Surabaya, Januari 2015

Penulis

DAFTAR ISI

(*Halaman ini sengaja dikosongkan*)

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

1.1 Latar Belakang

Berdasarkan UU RI No. 22 T ahun 2009 tentang Lalu Lintas dan Angkutan Jalan kecelakaan lalu lintas adalah suatu peristiwa di jalan yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan/atau kerugian harta benda. Menurut Organisasi Kesehatan Dunia (WHO), kecelakaan lalu-lintas adalah kejadian di mana sebuah kendaraan bermotor tabrakan dengan benda lain dan menyebabkan kerusakan. Kadang kecelakaan ini dapat mengakibatkan luka-luka atau kematian manusia atau binatang. Kecelakaan lalu lintas menelan korban jiwa sekitar 1,2 juta manusia setiap tahun (Raharjo, 2014). Banyaknya jumlah kendaraan bermotor yang meningkat dari tahun ke tahun merupakan faktor pendukung meningkatnya jumlah kecelakaan lalu lintas. Kecelakaan kendaraan bermotor ini dapat mempengaruhi pengendara menjadi truma akibat kecelakaan tersebut ataupun sampai meninggal. Dalam penelitian Fitriah (2012) menyebutkan bahwa 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. Dalam penelitian Indriani dan Indawati (2006), Songer *et al.* menyatakan bahwa *mortality rate* di negara-negara berkembang dimana kecelakaan lalu lintas menduduki peringkat delapan besar penyebab tingginya *mortality rate*.

Kota Surabaya merupakan Ibu Kota provinsi Jawa Timur yang merupakan kota padat penduduk. Hal ini disebabkan juga karena Surabaya sendiri sudah termasuk dalam 5 kota besar yang berkembang cukup pesat dalam hal pembangunan yang terpadat di Indonesia. Surabaya merupakan pusat bisnis, perdagangan,

industri, dan pendidikan di kawasan Indonesia timur. Maka dapat dipastikan jumlah pendatang dari luar wilayah Surabaya melonjak naik dan tidak dapat dibendung oleh pemerintah. Kepadatan penduduk di Surabaya juga akan menyebabkan kepadatan kendaraan yang tinggi sehingga mempunyai resiko kecelakaan yang cukup tinggi. Berdasarkan data Polrestabes Surabaya, pada tahun 2013 ada 854 kejadian kecelakaan lalu lintas di Surabaya. Adapun korban meninggal, luka berat, luka ringan akibat kecelakaan lalu lintas ada 208 korban, 317 korban dan 680 korban pada tahun 2013.

Tingginya angka kecelakaan di Surabaya dengan berbagai kategorinya menjadi dasar pentingnya dilakukan penelitian mengenai faktor yang mempengaruhi kategori kecelakaan tersebut. Beberapa penerapan metode statistika telah banyak diperkenalkan guna melakukan klasifikasi suatu variabel respon terhadap faktor-faktor yang mempengaruhinya seperti *Classification and R egression Trees* (CART) (Breiman, 1984), *Multivariate Adaptive Regression Spline* (MARS) (Friedman, 1991), regresi logistik dan *Random Forests* (Breiman, 2001).

Metode CART digunakan untuk menggambarkan hubungan antara variabel respon (variabel dependen atau tak bebas) dengan satu atau lebih variabel prediktor (variabel independen atau bebas). CART mempunyai kelebihan dibandingkan dengan metode klasifikasi lain yaitu hasilnya lebih mudah diinterpretasikan, lebih akurat dan lebih cepat penghitungannya, selain itu CART bisa diterapkan untuk himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilihan biner. Akan tetapi, CART juga memiliki kelemahan yaitu menghasilkan pohon yang kurang stabil karena CART sangat sensitif dengan data baru, bergantung dengan jumlah sampel. Jika sampel data *training* dan *testing* berubah maka pohon keputusan yang dihasilkan juga ikut berubah (Pratiwi & Zain, 2014).

MARS pertama kali diperkenalkan oleh Friedman (1991) di bidang data mining yang membahas beberapa keterbatasan RTA (*Regression Tree Analysis*) sehubungan dengan variabel kontinu. Prosedur MARS yaitu membangun model regresi yang fleksibel dengan menggunakan fungsi dasar untuk menyesuaikan *splines* terpisah untuk menentukan interval variabel prediktor yang jelas. Kedua variabel pada titik akhir dari interval, atau *knot*, ditemukan oleh prosedur pencarian lengkap menggunakan kelas khusus fungsi dasar. Pendekatan ini berbeda dari *splines* klasik di mana *knot* telah ditentukan dan ditempatkan secara merata. MARS dapat menemukan lokasi dan jumlah *kno*t yang dibutuhkan dalam beberapa tahap ke depan. Pertama, model ini *overfitted* dengan menghasilkan *knot* yang berlebihan dari yang diperlukan. Fungsi dasar dari MARS adalah untuk mengetahui hubungan antara variabel respon dan variabel prediktor. MARS memiliki kelebihan dibandingkan RTA di percabangan yang terputus pada *node* pohon digantikan oleh sebuah fungsi yang terus menerus dipandu oleh sifat lokal data. Oleh karena itu, MARS lebih baik dalam mendeteksi struktur data global dan linier sehingga outputnya lebih halus dan tidak kasar dan terputus-putus seperti pada RTA. Namun, MARS juga memiliki keterbatasan yaitu fungsi dasar yang kadang-kadang berlebihan dipandu oleh sifat lokal dari data, sehingga hasilnya tidak sesuai dan memilih nilai yang benar untuk parameter dapat menjadi rumit dan mungkin memerlukan beberapa langkah *trial and error* (Prasad, 2006).

Regresi logistik merupakan salah satu metode yang dapat digunakan untuk mencari hubungan variabel respon yang bersifat *dichotomous* (berskala nominal atau ordinal dengan dua kategori) atau *polychotomous* (mempunyai skala nominal atau ordinal dengan lebih dari dua kategori) dengan satu atau lebih variabel prediktor. Untuk variabel respon pada regresi logistik bersifat kontinyu atau kategorik (Agresti, 1900).

Penelitian terkait kecelakaan lalu lintas (*traffic accident*) juga pernah dilakukan oleh Indriani *et al*. (2006) yang membahas

mengenai hubungan korban kecelakaan lalu lintas dengan waktu, musim dan jenis kendaraan dan juga dijelaskan terkait konsep e*pidemiology triad* bahwa faktor-faktor yang mempengaruhi kecelakaan ada tiga yaitu kondisi pengendara (umur, jenis kelamin, pengalaman, alkohol, lelah), kondisi kendaraan (rusak, bentuk) dan lingkungan (kondisi jalan, keramaian lalu lintas, cuaca) dengan hasil analisis bahwa tingkat kecelakaan terbesar dengan kondisi korban meninggal terjadi pada keadaan waktu terang dengan mengendarai kendaraan roda dua dan pada musim penghujan. Penelitian yang dilakukan oleh Enache *et al*. (2009) yang membahas mengenai faktor yang mematikan dalam kecelakaan lalu lintas di Romania dengan hasil analisis bahwa selain dari faktor manusia, kendaraan, jalan dan lingkungan, kondisi mabuk teridentifikasi sebagai potensi yang mematikan saat berkendara. Penelitian yang pernah dilakukan Fitriah *et al*. (2012) mengenai faktor-faktor yang mempengaruhi korban kecelakaan dan klasifikasi korban kecelakaan lalu lintas di Surabaya meggunakan pendekatan regresi logistik ordinal dan *bagging* regresi logistik ordinal dengan hasil analisis bahwa jenis kecelakaan, peran korban, jenis kendaraan lawan dan usia mempengaruhi korban kecelakaan serta ketepatan klasifikasi menggunakan *bagging* regresi logistik ordinal lebih baik yaitu sebesar 56 persen daripada menggunakan regresi logistik ordinal yaitu sebesar 54,86 persen. Penelitian terkait kecelakaan juga dilakukan oleh Zhang *et al*. (2013) yang membahas mengenai hubungan antara resiko pelanggaran lalu lintas dengan keparahan korban kecelakaan di Cina dengan metode regresi logistik dengan hasil analisis bahwa faktor yang mempengaruhi kecelakaan lalu lintas di Cina adalah manusia, kendaraan, jalan dan lingkungan. Zhang *et al*. mengatakan bahwa resiko utama yang mengancam keselamatan jalan adalah pelanggaran lalu lintas.

Metode lebih baru yang dikembangkan untuk klasifikasi adalah *Random Forest* yang dikembangkan oleh Breiman (2001). Metode ini berusaha menangani beberapa kelemahan yang ada pada metode-metode sebelumnya seperti CART. Beberapa

penelitian yang sudah mengaplikasikan *Random Forests* pada analisis *driver* oleh Dewi *et al*. (2011) yang menyatakan bahwa *Random Forests* mampu memberikan akurasi yang tinggi dan stabil, pada *ecohydrological* oleh Peters *et al*. (2014) yang menyatakan bahwa *Random Forests* adalah model yang mampu meningkatkan hasil dari kebaikan model serta analisis metadata oleh Vaxjo (2014) yang menyatakan bahwa ketepatan prediksi *Random Forests* lebih baik yaitu sebesar 81.66% daripada regresi logistik yaitu sebesar 73.07%. Oleh sebab itu, akan diimplementasikan *Random Forests* pada kasus korban kecelakaan lalu lintas di Surabaya menggunakan variabel prediktor pada penelitian sebelumnya yaitu jenis kecelakaan, jenis kelamin, peran korban dalam kecelakaan, jenis kendaraan korban, jenis kendaraan lawan, waktu dan usia.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka permasalahan yang dibahas dalam penelitian ini adalah bagaimana ketepatan klasifikasi faktor-faktor yang mempengaruhi korban kecelakaan lalu lintas di Surabaya dengan pendekatan regresi logistik multinomial, pendekatan *Random Forests* dan bagaimana perbandingan hasil ketepatan klasifikasi dengan menggunakan pendekatan regresi logistik multinomial dan *Random Forests* pada faktor-faktor yang mempengaruhi korban kecelakaan lalu lintas di Surabaya.

1.3 Tujuan

Dalam penelitian ini tujuan yang ingin dicapai adalah mengetahui ketepatan klasifikasi faktor-faktor yang mempengaruhi korban kecelakaan lalu lintas di Surabaya dengan metode regresi logistik multinomial dan *Random Forests* serta membandingkan hasil ketepatan klasifikasi diantara kedua metode tersebut.

1.4 Manfaat Penelitian

Manfaat penelitian ini adalah mengembangkan wawasan ilmu pengetahuan yang berkaitan dengan “Regresi Logistik

Multinomial dan *Random Forests*". Selain itu dapat memberikan informasi kepada Unit Laka Satlantas Polrestabes Surabaya terkait faktor-faktor yang mempengaruhi kecelakaan di Surabaya tahun 2013 sehingga diharapkan Satlantas Polrestabes Surabaya memberikan perhatian yang lebih terhadap faktor-faktor tersebut dan dapat menjadi referensi untuk antisipasi jatuhnya korban kecelakaan lalu lintas di Surabaya.

1.5 Batasan Masalah

Batasan masalah pada penelitian ini adalah analisis korban kecelakaan lalu lintas di Surabaya berdasarkan pada data dari Unit Laka Satlantas Polrestabes Surabaya pada tahun 2013.

BAB II
TINJAUAN PUSTAKA

2.1 Regresi Logistik

Regresi logistik merupakan salah satu metode *dichotomus* (berskala nominal atau ordinal dengan dua kategori) atau *polychotomous* (mempunyai skala nominal atau ordinal dengan lebih dari dua kategori) dengan satu atau lebih variabel prediktor. Untuk variabel respon pada regresi logistik bersifat kontinyu atau kategorik (Agresti, 1990).

2.2 Regresi Logistik Multinomial

Regresi logistik multinomial merupakan regresi logistik yang digunakan saat variabel dependen mempunyai skala yang bersifat *polichotomous* atau multinomial. Skala multinomial adalah suatu pengukuran yang dikategorikan menjadi lebih dari dua kategori. Metode yang digunakan dalam penelitian ini adalah regresi logistik dengan variabel dependen berskala nominal dengan tiga kategori.

Mengacu pada regresi logistik *trichotomous* (Hosmer dan Lemeshow, 2000) untuk model regresi dengan variabel dependen berskala nominal tiga kategori digunakan kategori variabel hasil Y dikoding 0, 1, dan 2. Variabel Y terparameterisasi menjadi dua fungsi logit. Sebelumnya perlu ditentukan kategori hasil mana yang digunakan untuk membandingkan. Pada umumnya digunakan Y=0 sebagai pembanding. Untuk membentuk fungsi logit, akan dibandingkan Y=1 dan Y=2 terhadap Y=0. Bentuk model regresi logistik dengan *p* variabel prediktor seperti pada persamaan (2.1).

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)} \qquad (2.1)$$

Dengan menggunakan transformasi logit akan didapatkan dua fungsi logit.

$$g_1(x) = \ln\left[\frac{P(Y=1|x)}{P(Y=0|x)}\right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \ldots + \beta_{1p}x_p = x'\beta_1 \qquad (2.2)$$

$$g_2(x) = \ln\left[\frac{P(Y=2|x)}{P(Y=0|x)}\right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \ldots + \beta_{2p}x_p = x'\beta_2 \qquad (2.3)$$

Berdasarkan kedua fungsi logit tersebut maka didapatkan model regresi logistik *trichotomous* sebagai berikut.

$$\pi_0(x) = \frac{1}{1 + \exp g_1(x) + \exp g_2(x)} \qquad (2.4)$$

$$\pi_1(x) = \frac{\exp g_1(x)}{1 + \exp g_1(x) + \exp g_2(x)} \qquad (2.5)$$

$$\pi_2(x) = \frac{\exp g_2(x)}{1 + \exp g_1(x) + \exp g_2(x)} \qquad (2.6)$$

Dengan $P(Y = j|x) = \pi_j(x)$ untuk j=0,1,2 (Hosmer dan Lemeshow, 2000).

2.3 Estimasi Parameter

Banyak metode yang dapat digunakan untuk menaksir β salah satunya adalah metode *Maximum Likelihood Estimation* (MLE). Metode ini memperoleh dugaan maksimum likelihood bagi β dengan iterasi *Newton Raphson*. Pendugaan parameter maksimum merupakan penduga yang konsisten dan efisien untuk ukuran sampel yang besar. Estimasi maksimum *likelihood* merupakan pendekatan dari estimasi *Weighted Least Square* (WLS), dimana matrik pembobotnya berubah setiap putaran. Proses menghitung estimasi maksimum likelihood ini disebut juga sebagai *iteratif reweighted least square*.

Bila variabel respon pengamatan mempunyai tiga kategori maka akan ada tiga kemungkinan *outcome* dan mempunyai distribusi *trichotomous* sehingga fungsi likelihoodnya adalah sebagai berikut.

$$l(\beta)=\prod_{i=1}^{n}\left[\pi_0(x_i)^{y_{0i}}\,\pi_1(x_i)^{y_{1i}}\,\pi_2(x_i)^{y_{2i}}\right] \qquad (2.7)$$

Dengan $\sum_{j=0}^{2} y_{ij}=1$

$$L(\beta)=\sum_{i=1}^{n}\left\{Y_{1i}g_1(x_i)+Y_{2i}g_2(x_i)\right\}-\ln(1+\exp(g_1(x_i))+\exp(g_2(x_i)) \qquad (2.8)$$

Dengan mendifferensialkan fungsi pada persamaan (2.8) akan dihitung parameter-parameternya melalui persamaan (2.9) berikut.

$$\frac{\partial L(\beta)}{\partial \beta_{jk}}=\sum_{i=1}^{n} x_{ki}(Y_{ij}-\pi_{ij}) \qquad (2.9)$$

untuk j=1, 2, ..., g dan k=0, 1, 2, ..., p

Berdasarkan teori *maximum likelihood*, untuk mengestimasi varians kovarian diperoleh melalui turunan ke dua fungsi likelihoodnya.

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk}\partial \beta_{jk}}=-\sum_{i=1}^{n} x_{k'i}x_{ki}\pi_{ji}\left(1-\pi_{ji}\right) \qquad (2.10)$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk}\partial \beta_{j'k}}=-\sum_{i=1}^{n} x_{k'i}x_{ki}\pi_{ij}\pi_{j'i} \qquad (2.11)$$

untuk j, j'=1, 2 dan k, k'=0, 1, 2, ...,p (Hosmer dan Lemeshow, 2000).

2.4 Pengujian Parameter

Untuk menguji signifikansi koefisien β dari model yang telah diperoleh, maka dilakukan uji parsial dan uji serentak.

2.4.1 Uji Serentak

Pengujian ini dilakukan untuk mengetahui apakah model telah tepat (signifikan) dan untuk memeriksa kemaknaan koefisien β secara keseluruhan dengan hipotesis sebagai berikut.

$H_0 : \beta_1=\beta_2=\ldots=\beta_p=0$

H_1 : paling sedikit ada satu $\beta_i \neq 0$ dengan $i = 1, 2, ..., p$
Statistik uji yang digunakan adalah statistik uji G atau *likelihood ratio test.*

$$G = 2\left\{\sum_{i=1}^{n}\left[y_i \ln\left(\hat{\pi}_i\right) + \left(1 - y_i\right)\ln(1 - \hat{\pi}_i)\right] - \left[n_1 \ln\left(n_1\right) + n_0 \ln\left(n_0\right) - n\ln(n)\right]\right\} \quad (2.12)$$

Dengan n_1 = banyaknya observasi yang berkategori 1 dan n_2 = banyaknya observasi yang berkategori 0. Daerah penolakan Ho adalah jika $G > \chi^2_{(\alpha,v)}$ dengan $db = v$.

2.4.2 Uji Parsial

Pengujian ini dilakukan untuk mengetahui signifikansi parameter terhadap variabel respon. Pengujian signifikansi parameter menggunakan uji Wald dengan hipotesis sebagai berikut.
$H_0 : \beta_i = 0$
$H_1 : \beta_i \neq 0$ dengan $i = 1, 2, ..., p$
Perhitungan statistik uji Wald adalah sebagai berikut.

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (2.13)$$

Daerah penolakan Ho adalah jika $|W| > Z_{\alpha/2}$ atau $W^2 > \chi^2_{(v,\alpha)}$ dengan derajat bebas v (Hosmer dan Lemeshow, 2000).

2.5 Interpretasi Model

Interpretasi model pada regresi logistik multinomial menggunakan nilai *odds ratio* (ψ). Nilai *odds ratio* (ψ) digunakan untuk menunjukkan kecenderungan hubungan suatu variabel X terhadap variabel Y. Jika nilai *odds ratio* (ψ)<1, maka antara variabel prediktor dan variabel respon terdapat hubungan negatif setiap kali perubahan nilai variabel bebas (x) dan jika *odds ratio* (ψ)>1 maka antara variabel prediktor dengan variabel respon terdapat hubungan positif setiap kali perubahan nilai variabel bebas (x) (Agresti, 1990).

$$\psi = \frac{N_{11}/N_{21}}{N_{12}/N_{22}} = \frac{N_{11}N_{22}}{N_{12}N_{21}} \qquad (2.14)$$

Keterangan:

N_{11} = banyak kejadian kelompok pertama dengan kategori 1 terhadap kejadian kelompok kedua dengan kategori 1

N_{12} = banyak kejadian kelompok pertama dengan kategori 1 terhadap kejadian kelompok kedua dengan kategori 2

N_{21} = banyak kejadian kelompok pertama dengan kategori 2 terhadap kejadian kelompok kedua dengan kategori 1

N_{22} = banyak kejadian kelompok pertama dengan kategori 2 terhadap kejadian kelompok kedua dengan kategori 2

2.6 Uji Kesesuaian Model

Uji kesesuaian model dilakukan untuk mengetahui apakah model dengan variabel dependen tersebut merupakan model yang sesuai. Uji kesesuaian model dapat menggunakan statistik uji *Chi-Square* dengan hipotesis sebagai berikut.

H_0 : model sesuai (tidak ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model)

H_1 : model tidak sesuai (ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model)

Perhitungan statistik uji *Chi-Square* sebagai berikut.

$$\chi^2 = \sum_{k=1}^{g} \frac{\left(o_k - n'_k \bar{\pi}_k\right)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \qquad (2.15)$$

dimana

o_k $= \sum_{j=1}^{n'_k} y_j$ (jumlah pengamatan pada grup ke-k)

$\bar{\pi}_k$ $= \sum_{j=1}^{n'_k} \frac{m_j \pi_j}{n'_k}$ (rata-rata taksiran probabilitas)

g = jumlah grup (kombinasi kategori dalam model serentak)

m_j = banyaknya observasi yang memiliki nilai $\hat{\pi}_j$

n'_k = banyak observasi pada grup ke-k

Pengambilan keputusan didasarkan pada tolak H_0 jika $\chi^2_{hitung} \geq \chi^2_{(db,\alpha)}$ dengan *db=g-2*.

2.7 *Random Forests*

Pada pembahasan *random forests* akan dijelaskan terkait CART, pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi, penentuan pohon klasifikasi optimal, pengertian *random forests*, karakteristik *random forests* dan algoritma *random forests*.

2.7.1 CART (*Classification And Regression Trees*)

CART adalah pendekatan model nonparametrik yang dapat menjelaskan variabel respon yang dipengaruhi oleh variabel prediktor yang berisifat kontinu maupun kategorik. Data dependent tergantung dari partisi serangkaian *node* yang bercabang ke kanan dan ke kiri dapat disebut simpul anak (*child nodes)* yang berasal dari simpul utama (*parent node)*. Setelah partisi telah berhenti, *child nodes* disebut sebagai *terminal nodes* (Zheng et al., 2009).

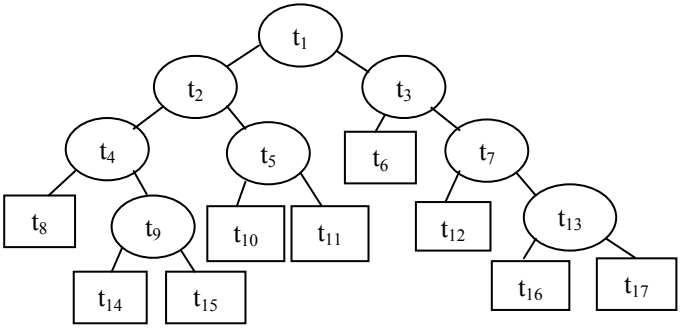
Ilustrasi pohon klasifikasi ditunjukkan pada Gambar 2.1. Simpul awal yang merupakan variabel terpenting dalam menduga kelas amatan disebut sebagai simpul utaman (*parent node)* dengan notasi t_1, simpul dalam (*internal nodes)* dinotasikan dengan t_2, t_3, t_4, t_5, t_7, t_9 dan t_{13}, serta simpul akhir yang disebut sebagai simpul akhir (*terminal nodes)* dinotasikan dengan t_6, t_8, t_{10}, t_{11}, t_{12}, t_{14}, t_{15}, t_{16} dan t_{17} dimana setelahnya tidak ada lagi pemilihan. Setiap simpul berada pada kedalaman (*depth*) tertentu dimana t_1 berada kedalaman 1, t_2 dan t_3 berada pada kedalaman 2, dan begitu seterusnya hingga t_{14}, t_{15}, t_{16} dan t_{17} yang berada pada kedalaman 5.

2.7.2 Pembentukan Pohon Klasifikasi

Proses pembentukan pohon klasifikasi terdiri atas 3 tahapan, yaitu :

a. Pemilihan (*Classifier*)

Data yang digunakan pada tahap ini adalah sampel data *training*/*learning* (*L)* yang kemudian dipilah berdasarkan aturan-



Gambar 2.1 Ilustrasi Pohon Klasifikasi

pemilihan dan kriteria *goodness of split*. Himpunan bagian yang dihasilkan dari proses pemilihan harus lebih homogen dibandingkan simpul induknya. Hal ini dapat dilakukan dengan mendefinisikan fungsi keheterogenan simpul (*impurity atau i(t)*). Fungsi heterogenitas yang umum digunakan adalah Indeks Gini. Metode ini memiliki kelebihan yaitu proses perhitungan yang sederhana dan relatif cepat, serta mudah dan sesuai untuk diterapkan dalam berbagai kasus (Breiman *et al*., 1993). Fungsi Indeks Gini dituliskan dalam persamaan (2.16).

$$i(t) = \sum_{i,j=1} p(j|t)\, p(i|t), i \neq j \qquad (2.16)$$

dengan $p(j|t)$ adalah proporsi kelas j pada simpul t dan $p(i|t)$ adalah proporsi kelas i pada simpul t. Setelah dilakukan pemilihan dari semua kemungkinan pemilah, maka tahapan berikutnya adalah menentukan kriteria *goodness of split (ϕ(s,t))* untuk mengevaluasi pemilah dari pemilah s pada simpul t. *Goodness of split (ϕ(s,t))* didefinisikan sebagai penurunan heterogenitas sebagai berikut.

$$\phi(s,\Delta) = , i(s\ t) = i(t) - p_L i(t_L) - (p_R) i\ t_R \qquad (2.17)$$

dengan

$i(t)$ = fungsi heterogenitas pada simpul t

p_L = proporsi pengamatan simpul kiri

p_R = proporsi pengamatan menuju simpul kanan

$i(t_L)$ = fungsi heterogenitas pada simpul anak kiri
$i(t_R)$ = fungsi heterogenitas pada simpul anak kanan

$$\Delta i\left(s^*, t_1\right) = \max_{s \subset S} \Delta i(s,t) \qquad (2.18)$$

Pemilah yang menghasilkan ϕ(s,t) lebih tinggi merupakan pemilah terbaik karena mampu mereduksi heterogenitas lebih tinggi. Pengembangan pohon dilakukan dengan pencarian pemilah yang mungkin pada simpul t_1 yang kemudian akan dipilah menjadi t_2 dan t_3 oleh pemilah s^* dan begitu seterusnya.

b. Penentuan Simpul Terminal

Suatu simpul *t* akan menjadi simpul terminal atau tidak, akan dipilih kembali bila pada simpul *t* tidak terdapat penurunan keheterogenan secara berarti atau adanya batasan minimum *n* seperti halnya terdapat satu pengamatan pada tiap simpul anak. Jumlah minimum dalam suatu terminal akhir umumnya adalah 5, dan apabila hal itu terpenuhi maka pengembangan pohon dihentikan (Breiman et al., 1993).

c. Penandaan Label Kelas

Penandaan label kelas pada *terminal nodes* dilakukan berdasarkan aturan jumlah terbanyak. Label kelas simpul terminal *t* adalah j_0 yang memberi nilai dugaan kesalahan pengklasifikasian simpul *t* terbesar. Proses pembentukan pohon klasifikasi berhenti saat terdapat hanya satu pengamatan dalam tiap-tiap simpul anak atau adanya batasan minimum *n*, semua pengamatan dalam tiap simpul anak identik dan adanya batasan jumlah level/kedalaman pohon maksimal.

$$p\left(j_0|t\right) = \max_j p(j \mid t) = \max_j \frac{N_j(t)}{N(t)} \qquad (2.19)$$

dengan $N_j(t)$ merupakan banyaknya amatan kelas *j* pada *terminal nodes t,* dan *N(t)* merupakan jumlah total pengamatan dalam *terminal node t.* Label kelas untuk *terminal node t* adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul *t* paling kecil sebesar $r(t) = 1 - \max_j p(j|t)$.

2.7.3 Pemangkasan Pohon Klasifikasi

Bagian pohon yang kurang penting dilakukan pemangkasan sehingga didapatkan pohon klasifikasi yang optimal. Pemangkasan didasarkan pada suatu penilaian ukuran sebuah pohon tanpa mengorbankan kebaikan ketepatan melalui pengurangan simpul pohon sehingga dicapai ukuran pohon yang layak. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon y ang layak tersebut adalah *cost complexity minimum* (Lewis, 2000).

$$R_{\alpha}\left(t\right) = R\left(t\right) + \alpha \left|\tilde{T}\right| \tag{2.20}$$

dimana

$R(T)$ = *resubtitusion estimate* (proporsi kesalahan pada sub pohon)

α = kompleksitas parameter (*complexity parameter*)

$\left|\tilde{T}\right|$ = ukuran banyaknya simpul terminal pohon T

2.7.4 Penentuan Pohon Klasifikasi Optimal

Ukuran pohon yang terlalu besar akan menyebabkan nilai *cost complexity* yang tinggi karena struktur data yang digambarkan cenderung kompleks sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti yang cukup kecil. Bila $R(T)$ dipilih sebagai penduga terbaik, maka akan cenderung dipilih pohon yang besar, sebab pohon yang semakin besar akan membuat nilai $R(T)$ semakin kecil. Terdapat dua macam penduga untuk mendapatkan pohon klasifikasi optimal yaitu penduga sampel uji (*test sample estimate*) dan penduga validasi silang lipat v (*cross validation v-fold estimate*).

a. Penduga Sampel Uji (*Test Sample Estimate*)

Penduga sampel uji digunakan ketika data berukuran besar. Prosedur *test sample estimate* diawali dengan membagi data *learning* menjadi dua bagian yaitu L_1 dan L_2. Pengamatan dalam L_1 digunakan untuk membentuk pohon T, sedangkan pengamatan

dalam L_2 digunakan untuk menduga $R^{ts}(T_t)$. Persamaan *test sample estimate* adalah sebagai berikut.

$$R^{ts}\left(T_t\right)=\frac{1}{N_2}\sum_{(x_n,j_n)\in L_2} X(d(x_n)\neq j_n) \qquad (2.21)$$

dengan N_2 adalah jumlah pengamatan dalam L_2 dan $X(.)$ bernilai 0 jika pertanyaan dalam tanda kurung salah dan bernilai 1 jika pertanyaan dalam tanda kurung benar. Pohon klasifikasi yang optimum dipilih T^* dengan $R^{ts}(T^*)=min\ R^{ts}(T_t)$.

b. Penduga Validasi Silang Lipat V (*Cross Validation V-Fold Estimate*)

Penduga pengganti ini sering digunakan apabila pengamatan yang ada tidak cukup besar. Pengamatan dalam L dibagi secara random menjadi V bagian yang saling lepas dengan ukuran kurang lebih sama besar untuk setiap kelas. Pohon $T^{(v)}$ dibentuk dari sampel *learning* ke v dengan $v=1,2,...,V$. Dimisalkan $d^{(v)}(x)$ adalah hasil pengklasifikasian, maka penduga sampel uji untuk $R(T_t^{(v)})$ adalah sebagai berikut.

$$R\left(T_t^{(v)}\right)=\frac{1}{N_v}\sum_{(x_n,j_n)\in L_v}^{N} X(d^{(v)}(x_n)\neq j_n) \qquad (2.22)$$

dengan $N_v \cong N/V$ adalah jumlah pengamatan dalam L_v.

Selanjutnya, dilakukan prosedur yang sama dengan menggunakan semua pengamatan dalam L untuk membentuk deret pohon T_t. Penduga validasi silang lipat v untuk $T_t^{(v)}$ adalah

$$R^{cv}\left(T_t\right)=\frac{1}{V}\sum_{v=1_v}^{V} R^{cv}(T_t^{(v)}) \qquad (2.23)$$

Pohon klasifikasi yang optimum dipilih T^* dengan $R^{cv}(T^*)=\min_t R^{cv}(T_t)$.

2.7.5 Pengertian *Random Forests*

Random forests adalah suatu metode klasifikasi yang terdiri dari gabungan pohon klasifikasi (CART) yang saling independen yang berasal dari distribusi yang sama melalui proses *voting* (jumlah terbanyak) untuk memperoleh prediksi klasifikasi.

Random forests merupakan pengembangan dari metode *ensemble* yang pertama kali dikembangkan oleh Leo Breiman (2001) yang digunakan untuk meningkatkan ketepatan klasifikasi. Bila dalam proses *bagging* digunakan *resampling bootstrap* untuk membangkitkan pohon k lasifikasi dengan banyak versi yang kemudian mengkombinasikannya untuk memperoleh prediksi akhir, maka dalam *random forests* proses pengacakan untuk membentuk pohon k lasifikasi tidak hanya dilakukan untuk data sampel saja melainkan juga pada pengambilan variabel prediktor. Sehingga, proses ini akan menghasilkan kumpulan pohon klasifikasi dengan ukuran dan bentuk yang berbeda-beda. Hasil yang diharapkan adalah suatu kumpulan pohon klasifikasi yang memiliki korelasi kecil antar pohon. Korelasi yang kecil akan menurunkan hasil kesalahan prediksi *Random Forests* (Breiman, 2001).

2.7.6 Karakteristik *Random Forests*

Menurut Breiman tahun 2001, salah satu kekuatan yang dimiliki oleh *random forests* adalah dapat meminimumkan korelasi yang dapat menurunkan hasil kesalahan prediksi *random forests*. Hasil *random forests* memberikan keakuratan sebaik Adabost. Karakteristik *Random Forests* adalah sebagai berikut.

a. Keakuratan akurasi sebaik Adaboost dan kadang-kadang lebih baik dari Adaboost.
b. *Random Forests* relatif kuat untuk mengatasi *outlier* dan pengganggu yang lain.
c. *Random Forests* prosesnya lebih cepat daripada *bagging* atau *boostin*g.
d. *Random Forests* berguna dalam hal mengestimasi *error*, kekuatan, korelasi dan variabel yang penting.
e. *Random Forests simple* selain itu juga mudah.

Adaboost (*Adaptive Boosting*) merupakan salah satu metode *ensemble* seperti *bagging* dan *random forests*. *Bagging* dan *random forests* mendapatkan banyak pohon dari anak gugus data yang berbeda-beda hasil dari proses *bootstrap*. Akan tetapi kalau adaboost, setiap kali pembuatan pohon, data yang digunakan tetap

seperti semula tetapi memiliki sebaran bobot yang berbeda dalam setiap iterasi. Penggunaan bobot juga dilakukan pada saat proses penggabungan dugaan akhir dari banyak pohon yang dihasilkan (Sartono *et al*., 2010).

2.7.7 Algoritma *Random Forests*

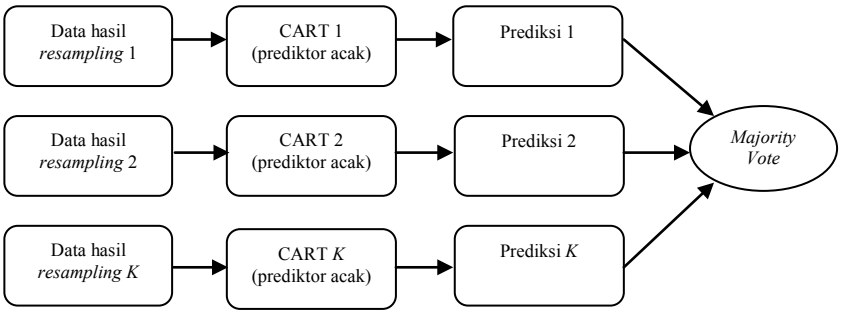
Secara umum, pengembangan *random forests* yang dilakukan dari proses *bagging* yaitu terletak pada proses pemilihan pemilah. Pada *random forests* pemilihan pemilah hanya melibatkan beberapa variabel prediktor yang terambil secara acak. Algoritma *Random Forests* dijelaskan sebagai berikut.

a. Mengambil *n* data sampel dari *dataset* awal dengan menggunakan teknik *resampling bootstrap* dengan pengembalian.
b. Menyusun pohon klasifikasi dari setiap *dataset* hasil *resampling bootstrap*, dengan penentuan pemilah terbaik didasarkan pada variabel prediktor yang diambil secara acak. Jumlah variabel yang diambil secara acak dapat ditentukan melalui perhitungan $\log_2(M+1)$ dimana *M* adalah banyak variabel prediktor (Breiman, 2001) atau $\sqrt{p}$ dimana *p* adalah banyak variabel prediktor (Genuer et al., 2009) atau $\sqrt{d}$ dimana *d* adalah banyak variabel prediktor (Sartono et al., 2010).
c. Melakukan prediksi klasifikasi data sampel berdasarkan pohon klasifikasi yang terbentuk.
d. Mengulangi langkah a-c hingga diperoleh sejumlah pohon klasifikasi yang diinginkan. Perulangan dilakukan sebanyak *K* kali.
e. Melakukan prediksi klasifikasi data sampel akhir dengan mengkombinasikan hasil prediksi pohon klasifikasi yang diperoleh berdasarkan aturan *majority vote.*

Dalam analisis dengan menggunakan metode *random forests* dimulai dari pengambilan data dengan teknik *resampling bootstrap*. *Bootstrap* adalah suatu metode yang dapat bekerja tanpa membutuhkan asumsi distribusi karena sampel asli

digunakan sebagai populasi. *Bootstrap* pertama kali diperkenalkan oleh Efron (1979) yang digunakan untuk mencari distribusi sampling dari suatu estimator dengan prosedur *resampling* dengan pengembalian dari data asli (Sungkono, 2013). Berikut adalah algoritma dari *resampling bootstrap*.

a. Mengkonstruksi distribusi empiris $\hat{F}_n$ dari suatu sampel dengan memberikan probabilitas 1/*n* pada setiap X_i dimana *i*=1, 2, ..., *n*.
b. Mengambil sampel *bootstrap* berukuran *n* secara random dengan pengembalian dari distribusi empiris $\hat{F}_n$ sebut sebagai sampel *bootstrap* pertama X^{*1}.
c. Menghitung stastistik $\hat{\theta}$ yang diinginkan dari sampel *bootstrap* X^{*1} sebut sebagai $\hat{\theta}_1{}^*$.
d. Mengulangi langkah 2 d an 3 h ingga *B* kali diperoleh $\hat{\theta}_1{}^*, \hat{\theta}_2{}^*, \ldots, \hat{\theta}_B{}^*$.
e. Mengkontruksi suatu distribusi probabilitas dari $\hat{\theta}_B{}^*$ dengan memberikan probabilitas *1/B* pada setiap $\hat{\theta}_1{}^*, \hat{\theta}_2{}^*, \ldots, \hat{\theta}_B{}^*$. Distribusi tersebut merupakan estimator *bootstrap* untuk distribusi sampling $\hat{\theta}$ dan dinotasikan dengan $\hat{F}^*$.
f. Pendekatan estimasi *bootstrap* adalah $\hat{\theta}^* = \sum_{b=1}^{B} \hat{\theta}_b{}^* \frac{1}{B}$.



Gambar 2.2 Algoritma *Random Forests*

2.7.8 Ilustrasi *Random Forests*

Ilustrasi *random forests* berikut menggunakan data sebanyak 10. Ilustrasi yang terdiri dari variabel respon (Y) dengan tiga kategori dan sebanyak tujuh variabel prediktor.

Tabel 2.1 Data Sampel untuk Ilustrasi *Random Forests*

Data Ilustrasi Ke-	Data Awal				Data Resampling			
	Y	X_1	...	X_7	Y	X_1	X_5	X_7
1	2	0		18	0	0	3	34
2	1	1		13	0	0	1	26
3	1	0		13	1	3	0	18
4	0	4		34	2	3	2	21
5	1	1		26	2	2	2	36
6	2	3		18	2	4	0	18
7	0	0		36	1	1	3	13
8	2	4		21	0	1	3	13
9	2	2		18	0	2	1	18
10	1	2		36	1	3	0	21

Tahapan *random forests* diawali dengan melakukan *resampling* dengan pengembalian pada data ilustrasi awal. Data hasil *resampling* tersebut kemudian digunakan untuk membentuk pohon klasifikasi (CART). Namun, pemilah terbaik dipilih berdasarkan variabel prediktor yang diambil secara acak (misal p=3). Selain itu, pemilihan dilakukan hingga diperoleh pohon maksimal tanpa dilakukan pemangkasan pada pohon. Dari pohon yang terbentuk tersebut kemudian dilakukan prediksi klasifikasi-dengan memberikan label kelas pada setiap *terminal node* yang dihasilkan. Proses ini diulang hingga sejumlah K replikasi sesuai keinginan peneliti. Bila hasil prediksi dari semua replikasi telah diperoleh, maka selanjutnya dilakukan voting mayoritas (*majority vote*) untuk menentukan prediksi klasifikasi akhir yang kemudian akan digunakan untuk menghitung kesalahan klasifikasinya. Berikut diberikan contoh hasil prediksi dengan menggunakan sebanyak empat kali replikasi pengambilan data dengan pengembalian (K=4). Hasil prediksi ditunjukkan pada Tabel 2.2 berikut.

Tabel 2.2 Hasil Prediksi Kelas dengan *Random Forests*

Data Ilustrasi ke-	Y	X_1	...	X_7	Prediksi CART 1	Prediksi CART 2	Prediksi CART 3	Prediksi CART 4	Prediksi Kelas *
1	2	0		18	1	0	0	2	2
2	1	1		13	2	1	1	1	1
3	1	0		13	1	2	1	1	1
4	0	4		34	0	1	2	1	1
5	1	1		26	0	0	1	0	0
6	2	3		18	0	1	1	1	1
7	0	0		36	1	1	2	2	2
8	2	4		21	2	1	1	1	1
9	2	2		18	1	1	1	1	1
10	1	2		36	1	2	2	1	1

*Hasil *majority vote* dari prediksi CART 1, CART 2, CART 3 dan CART 4

Apabila dibandingkan antara hasil pengamatan dengan prediksi, ada 6 pengamatan yang salah diklasifikasikan yaitu pengamatan ke-4, 5, 6, 7, 8, 9 s edangkan 4 pe ngamatan yang lainnya tepat diklasifikasikan.

2.8 Ketepatan Klasifikasi

Menurut Johnson (2007) untuk menghitung ketepatan klasifikasi pada hasil pengelompokkan digunakan *apparent error rate* (APER). Nilai APER menyatakan representasi proporsi sampel yang salah diklasifikasikan. Dalam penelitian kali ini digunakan respon multinomial tiga kategori sehingga penentuan kesalahan klasifikasi dapat dihitung dari tabel klasifikasi berikut.

Tabel 2.3 Tabel Klasifikasi Respon Multinomial Tiga Kategori

Aktual	Prediksi			Total
	1	2	3	
1	n_{11}	n_{12}	n_{13}	$N_{1.}$
2	n_{21}	n_{22}	n_{23}	$N_{2.}$
3	n_{31}	n_{32}	n_{33}	$N_{3.}$
Total	$N_{.1}$	$N_{.2}$	$N_{.3}$	N

Nilai APER dihitung sebagai berikut.

$$APER(\%) = \frac{n_{12} + n_{13} + n_{23} + n_{21} + n_{31} + n_{32}}{N} x100\% \quad (2.24)$$

$$Ketepatan\ klasifikasi\ = 1 - APER \quad (2.25)$$

keterangan

n_{11} : jumlah observasi dari kelas 1 yang tepat diprediksi sebagai kelas 1

n_{21} : jumlah observasi dari kelas 2 yang salah diprediksi sebagai kelas 1

n_{31} : jumlah observasi dari kelas 3 yang salah diprediksi sebagai kelas 1

n_{12} : jumlah observasi dari kelas 1 yang salah diprediksi sebagai kelas 2

n_{22} : jumlah observasi dari kelas 2 yang tepat diprediksi sebagai kelas 2

n_{32} : jumlah observasi dari kelas 3 yang salah diprediksi sebagai kelas 2

n_{13} : jumlah observasi dari kelas 1 yang salah diprediksi sebagai kelas 3

n_{23} : jumlah observasi dari kelas 2 yang salah diprediksi sebagai kelas 3

n_{33} : jumlah observasi dari kelas 3 yang tepat diprediksi sebagai kelas 3

$N_{1.}$: jumlah observasi dari kelas 1

$N_{2.}$: jumlah observasi dari kelas 2

$N_{3.}$: jumlah observasi dari kelas 3

N : jumlah observasi

2.9 Kecelakaan Lalu Lintas

Menurut Undang-Undang Republik Indonesia No. 22 Tahun 2009 Tentang Lalu Lintas dan Angkutan Jalan bahwa kecelakaan lalu lintas adalah suatu peristiwa di jalan yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan/atau kerugian harta benda.

Pengertian kecelakaan lalu lintas menurut Raharjo (2014) adalah kejadian di mana sebuah kendaraan bermotor bertabrakan dengan benda lain dan menyebabkan kerusakan. Kadang kecelakaan ini dapat mengakibatkan luka-luka atau kematian manusia atau binatang. Kecelakaan lalu lintas menelan korban jiwa sekitar 1,2 juta manusia setiap tahun menurut WHO. Ada tiga faktor utama yang menyebabkan terjadikannya kecelakaan, pertama adalah faktor manusia, kedua adalah faktor kendaraan dan yang terakhir adalah faktor jalan. Kombinasi dari ketiga faktor itu bisa saja terjadi, antara manusia dengan kendaraan misalnya berjalan melebihi batas kecepatan yang ditetapkan kemudian ban pecah yang mengakibatkan kendaraan mengalami kecelakaan. Di samping itu masih ada faktor lingkungan, cuaca yang juga bisa berkontribusi terhadap kecelakaan.

1. Faktor Manusia

Faktor manusia merupakan faktor yang paling dominan dalam kecelakaan. Hampir semua kejadian kecelakaan didahului dengan pelanggaran rambu-rambu lalu lintas. Pelanggaran dapat terjadi karena sengaja melanggar, ketidaktahuan terhadap aturan yang berlaku ataupun tidak melihat ketentuan yang diberlakukan atau pura-pura tidak tahu. Selain itu manusia sebagai pengguna jalan raya sering sekali lalai bahkan ugal-ugalan dalam mengendarai kendaraan, tidak sedikit angka kecelakaan lalu lintas diakibatkan karena membawa kendaraan dalam keadaan mabuk, mengantuk, dan mudah terpancing oleh ulah pengguna jalan lainnya yang dapat memancing gairah untuk balapan.

Dalam penelitiannya Fitriah (2012), faktor manusia yang dicatat kepolisian meliputi jenis kelamin korban, usia korban, profesi korban, dan peran korban dalam berkendara. Maksud dari peran korban dalam berkendara adalah posisi korban saat terjadi kecelakaan, apakah termasuk sebagai pengemudi, penumpang, pejalan kaki, penyeberang jalan, dan lain-lain. Jenis tabrakan yang dialami oleh korban juga merupakan salah satu faktor manusia yang diduga mampu mempengaruhi tingkat keparahan korban kecelakaan lalu lintas. Berdasarkan jenis tabrakannya,

kecelakaan lalu lintas dibagi menjadi tabrak depan (tabrak depan-depan), tabrak belakang (tabrak depan-belakang), tabrak samping (tabrak samping-depan dan tabrak samping-samping) dan kecelakaan karena lepas kendali atau hilang kendali. Berikut ini penjelasan dari jenis tabrakan yang kemungkinan terjadi.

a. Tabrakan Belakang (TB)
   Tabrakan belakang adalah jenis tabrakan antara dua kendaraan yang tengah melaju satu arah sehingga salah satu kendaraan menabrak bagian belakang kendaraan lainnya.
b. Tabrakan Depan (TD)
   Tabrakan depan adalah jenis tabrakan antara dua kendaraan yang tengah berlawanan arah sehingga bagian kendaraan yang satu menabrak bagian depan kendaraan lainnya.
c. Tabrakan Samping (TS)
   Tabrakan samping adalah jenis tabrakan antara dua kendaraan yang tengah melaju dimana bagian samping kendaraan yang satu menabrak bagian yang lain.
d. Hilang Kendali (HK)
   Hilang kendali adalah kecelakaan yang terjadi saat pengemudi tidak dapat menguasai kendaraannya.
e. Lain-lain
   Kecelakaan yang bukan termasuk dalam tabrakan belakang, tabrakan depan, tabrakan samping dan hilang kendali.

2. Faktor Kendaraan

Faktor kendaraan yang paling sering adalah kelalaian perawatan yang dilakukan terhadap kendaraan. Untuk mengurangi faktor kendaraan perawatan dan perbaikan kendaraan diperlukan, di samping itu adanya kewajiban untuk melakukan pengujian kendaraan bermotor secara reguler.

3. Faktor Jalan

Faktor jalan terkait dengan kecepatan, rencana jalan, geometrik jalan, pagar pengaman di daerah pegunungan, ada tidaknya media jalan, jarak pandang dan kondisi permukaan jalan. Jalan yang rusak/berlubang sangat membahayakan pemakai jalan terutama bagi pemakai sepeda.

4. Faktor Cuaca

Faktor cuaca seperti hujan juga mempengaruhi unjuk kerja kendaraan seperti jarak pengereman menjadi lebih jauh, jalan menjadi lebih licin, jarak pandang juga terpengaruh karena penghapus kaca tidak bisa bekerja secara sempurna atau lebatnya hujan mengakibatkan jarak pandang menjadi lebih pendek. Asap dan kabut juga bisa mengganggu jarak pandang, terutama di daerah pegunungan.

(Halaman ini sengaja dikosongkan)

BAB III
METODOLOGI PENELITIAN

3.1 Sumber Data

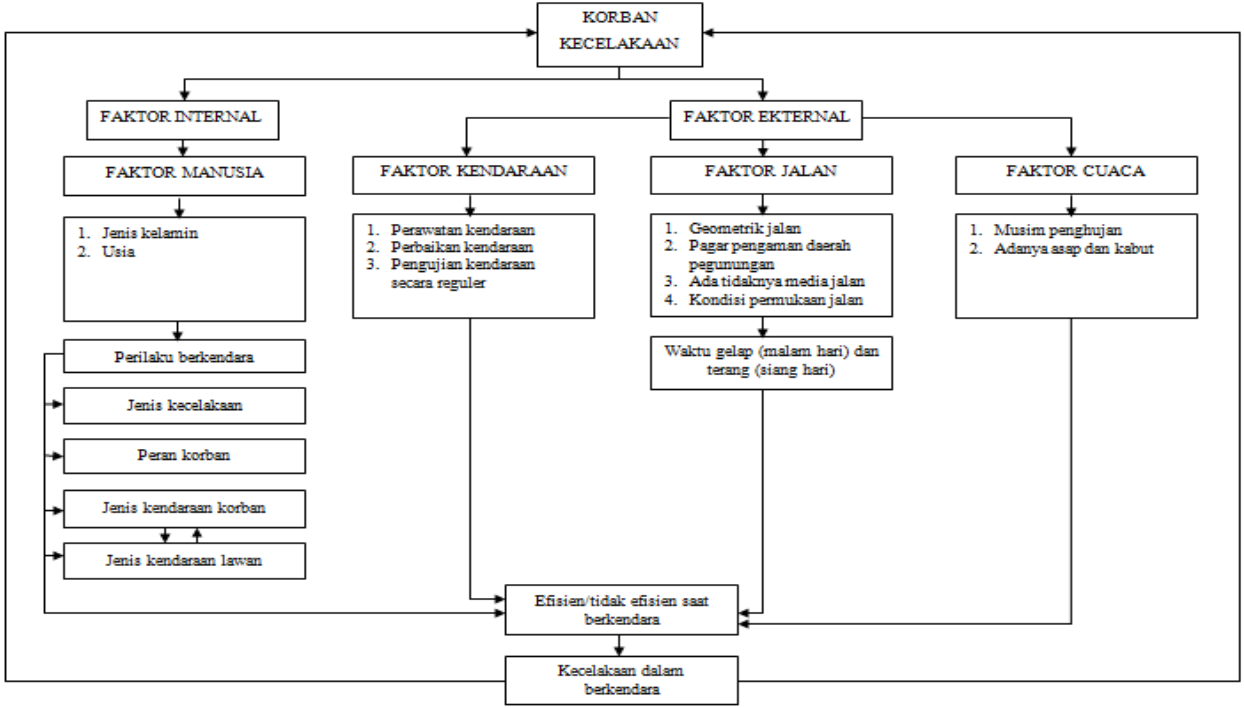
Data yang digunakan pada penelitian ini adalah data sekunder mengenai data korban kecelakaan lalu lintas di Surabaya pada tahun 2013. D ata sekunder ini didapatkan dari Unit Laka Satlantas Polrestabes Surabaya Jln. Dukuh Kupang No. 16 Surabaya. Data korban kecelakaan lalu lintas di Surabaya ada 854 kejadian dan ada 1205 korban kecelakaan.

3.2 Kerangka Konsep

Kerangka konsep dalam suatu penelitian sangatlah diperlukan untuk mengetahui ketepatan dalam menentukan variabel respon dan variabel prediktor. Dalam kerangka konsep akan dijelaskan mengenai hubungan antara variabel respon (variabel yang dipengaruhi) dan variabel prediktor(variabel yang mempengaruhi). Kerangka konsep pada penelitian ini dapat dilihat pada Gambar 3.1.

Pada dasarnya yang mempengaruhi korban kecelakaan lalu lintas ada dua yaitu faktor internal dan faktor ekternal. Faktor internal meliputi faktor manusia sedangkan faktor ekternal meliputi faktor kendaraan, faktor jalan dan faktor cuaca. Faktor manusia terdiri dari jenis kelamin dan usia dimana dari jenis kelamin dan usia seseorang, menyebabkan perilaku berkendara yang berbeda-beda, dari perilaku berkendara tersebut akan menyebabkan efisien tidaknya seseorang dalam berkendara. Apabila perilaku berkendara yang kurang baik akan menyebabkan kecelakaan dalam berkendara. Kecelakaan dalam berkendara akan menyebabkan korban kecelakaan.

Untuk faktor eksternal terdiri dari faktor kendaraan, faktor jalan dan faktor cuaca. Faktor kendaraan meliputi perawatan kendaraan, perbaikan kendaran, pengujian kendaraan secara reguler. Apabila hal tersebut tidak dilakukan maka akan menyebabkan efisien tidaknya seseorang dalam berkendara yang



Gambar 3.1 Kerangka Konsep Penelitian Korban Kecelakaan Lalu Lintas di Surabaya

akan menyebabkan kecelakaan dalam berkendara sehingga menyebabkan korban kecelakaan begitu juga dengan faktor jalan dan cuaca.

3.3 Variabel Penelitian

Variabel penelitian yang digunakan pada penelitian ini menggunakan 1 variabel respon dan 7 variabel prediktor. Variabel responnya adalah korban kecelakaan lalu lintas yang terdiri dari 3 kategori yaitu korban meninggal, korban luka berat dan korban luka ringan. Variabel prediktor meliputi 6 variabel yang berskala nominal dan 1 variabel yang berskala rasio. Variabel prediktor yang berskala nominal terdiri dari jenis kecelakaan, jenis kelamin, peran korban dalam kecelakaan, jenis kendaraan korban, jenis kendaraan lawan dan waktu sedangkan variabel prediktor yang berskala rasio yaitu usia korban yang dapat dilihat pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

No	Variabel	Skala/Kategori
1	Korban kecelakaan lalu lintas (Y)	Nominal Y(0) = korban meninggal Y(1) = korban luka berat Y(2) = korban luka ringan
2	Jenis kecelakaan (X_1)	Nominal $X_1(0)$ = Tabrakan Belakang (TB) $X_1(1)$ = Tabrakan Depan (TD) $X_1(2)$ = Tabrakan Samping (TS) $X_1(3)$ = Hilang Kendali (HK) $X_1(4)$ = Lain-lain
3	Jenis kelamin (X_2)	Nominal $X_2(0)$ = laki-laki $X_2(1)$ = perempuan
4	Peran korban dalam kecelakaan (X_3)	Nominal $X_3(0)$ = pengendara $X_3(1)$ = penumpang kendaraan selain pengendara $X_3(2)$ = pengguna jalan non penumpang kendaraan (penyebrang jalan, pejalan kaki dll)

Tabel 3.1. Variabel Penelitian (*Lanjutan*)

No	Variabel	Skala/Kategori
5	Jenis kendaraan korban (X_4)	Nominal $X_4(0)$ = sepeda motor (kendaraan bermotor roda dua atau tiga) $X_4(1)$ = kendaraan roda empat $X_4(2)$ = k endaraan dengan lebih dari empat roda $X_4(3)$ = lain-lain (pejalan kaki, sepeda angin, becak, atau kendaraan bukan bermotor lainnya)
6	Jenis Kendaraan Lawan (X_5)	Nominal $X_5(0)$ = sepeda motor (kendaraan bermotor roda dua atau tiga) $X_5(1)$ = kendaraan roda empat $X_5(2)$ = k endaraan dengan lebih dari empat roda $X_5(3)$ = lain-lain (pejalan kaki, sepeda angin, becak, atau kendaraan bukan bermotor lainnya)
7	Waktu Kejadian (X_6)	Nominal $X_6(0)$ = waktu terang (06.00 – 18.00 WIB) $X_6(1)$ = waktu gelap (selain waktu terang)
8	Usia Korban (X_7)	Rasio

3.4 Struktur Data Penelitian

Data yang digunakan dalam penelitian ini adalah data korban kecelakaan lalu lintas di Surabaya pada tahun 2013. Data yang digunakan adalah data jumlah korban kecelakaan sebanyak 1205 korban. Struktur data penelitian ini ditampilkan pada Tabel 3.2.

Tabel 3.2. Struktur Data Penelitian

Subjek	Status	Variabel Prediktor						
		X_1	X_2	X_3	X_4	X_5	X_6	X_7
Korban 1	K_1							
Korban 2	K_2							
⋮	⋮							
Korban 1205	K_{1205}							

Struktur data pada penelitian ini, subjeknya adalah korban kecelakaan lalu lintas di Surabaya. Pada masing-masing korban yang terdiri ada 1205 korban dikategorikan sebagai korban meninggal atau korban luka berat atau korban luka ringan yang tertera pada Tabel 3.2 yang dinamakan status. Untuk variabel prediktor yang terdiri dari X_1, X_2, X_3, X_4, X_5, X_6, X_7 artinya bahwa X_1 adalah variabel jenis kecelakaan, X_2 adalah variabel jenis kelamin, X_3 adalah variabel peran korban dalam kecelakaan, X_4 adalah variabel jenis kendaraan korban, X_5 adalah variabel jenis kendaraan lawan, X_6 adalah variabel waktu dan X_7 adalah variabel usia korban.

3.5 Metode Analisis

Analisis yang digunakan dalam penelitian ini adalah analisis regresi logistik multinomial dan *random forests* yang digunakan untuk pengklasifikasian jenis keparahan korban kecelakaan lalu lintas di Surabaya. Berikut tahapan analisis dalam penelitian ini.

1. Melakukan pra-pemrosesan data penelitian.
   Pra-pemrosesan data dilakukan meliputi pengkodingan data sesuai dengan kategori yang telah ditentukan.
2. Membagi data menjadi dua bagian yaitu data *training* dan *testing*. Data dibagi menurut kombinasi data *training* dan data *testing* dengan proporsi sebesar 75%:25%, 80%:20%, 85%:15%, 90%:10%, 95%:5%. Data *training* dan *testing* dipilih secara random dari hasil *resampling*.

3. Melakukan analisis regresi logistik multinomial melalui tahapan berikut.
   a. Melakukan uji serentak terhadap variabel-variabel yang berpengaruh terhadap kecelakaan lalu lintas di Surabaya
   b. Melakukan uji parsial terhadap variabel-variabel yang berpengaruh terhadap kecelakaan lalu lintas di Surabaya
   c. Membuat persamaan model regresi logistik multinomial dan interpretasi model
   d. Uji kesesuain model
   e. Menghitung ketepatan klasifikasi hasil bentukan dengan menggunakan data *training* dan mengevaluasinya dengan cara menjalankan data *testing* pada persamaan regresi logistik yang terbentuk
   f. Membandingkan hasil ketepatan klasifikasi dari setiap kombinasi data *training* dan data *testing*.
4. Melakukan analisis CART untuk membentuk pohon klasifikasi optimal dengan menggunakan data *training* dengan mengevaluasi menggunakan data *testing* dan dilakukan analisis replikasi CART pada analisis *random forests*.
5. Melakukan analisis klasifikasi *random forests* melalui tahapan berikut.
   a. Mengambil *n* sampel *bootstrap* dengan pengembalian dari data *training*
   b. Menentukan jumlah variabel prediktor yang akan dilakukan pengambilan secara acak dalam proses penentuan pemilah saat pembentukan pohon klasifikasi
      $$\log_2(M+1) \text{ atau } \sqrt{p}$$
   c. Membentuk pohon k lasifikasi dimana pemilihan *node* terbaik dilakukan berdasarkan variabel-variabel prediktor yang diambil secara acak
   d. Melakukan prediksi klasifikasi untuk data *training*
   e. Mengulangi langkah b-d hingga *K* kali replikasi
   f. Melakukan voting mayoritas (*majority vote*) hasil prediksi klasifikasi dari *K* kali replikasi pembentukan pohon klasifikasi

g. Menghitung ketepatan klasifikasi data *training*
h. Menghitung ketepatan klasifikasi data *testing*
i. Mengulangi langkah b-h dengan mencobakan kombinasi jumlah pohon (*K*) yang berbeda, yaitu sebesar 50, 100, 250, 500 dan 1000.

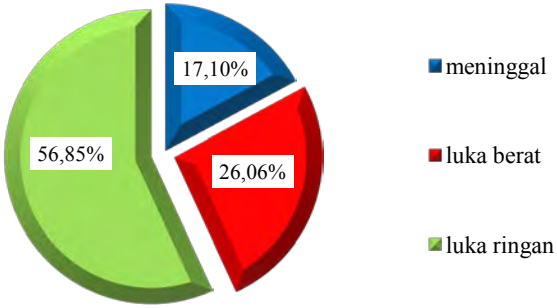
6. Membandingkan tingkat akurasi klasifikasi antara regresi logistik multinomial dengan *Random Forests*. Klasifikasi yang mempunyai 1-APER terbesar dipilih sebagai tingkat klasifikasi yang lebih baik dalam melakukan klasifikasi.

(Halaman ini sengaja dikosongkan)

BAB IV
ANALISIS DAN PEMBAHASAN

4.1 Karakteristik Korban Kecelakaan Lalu Lintas di Surabaya

Analisis dan pembahasan mengenai statistika deskriptif terhadap keparahan korban kecelakaan, jenis kecelakaan, jenis kelamin, peran korban dalam kecelakaan, jenis kendaraan korban, jenis kendaraan lawan, waktu dan usia di Surabaya pada tahun 2013 diharapkan dapat berguna bagi masyarakat dan Unit Laka Satlantas Polrestabes Surabaya sebagai referensi antisipasi jatuhnya korban kecelakaan lalu lintas di Surabaya.

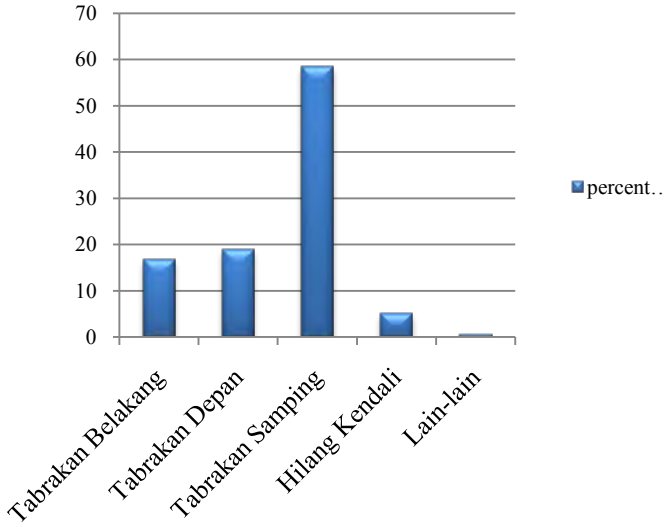


Gambar 4.1 Keparahan Korban Kecelakaan Lalu Lintas

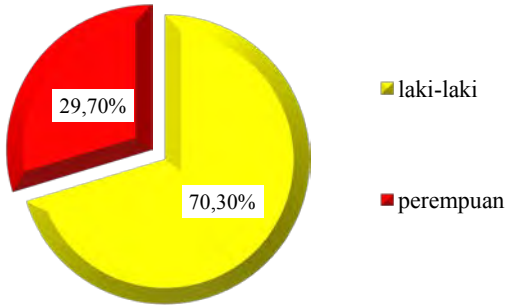
Gambar 4.1 m enunjukkan persentase keparahan korban kecelakaan lalu lintas. Keparahan korban kecelakaan lalu lintas terbagi menjadi tiga kategori yaitu korban meninggal, korban luka berat dan korban luka ringan. Jenis keparahan korban kecelakaan lalu lintas di Surabaya yang persentasenya tertinggi adalah korban luka ringan yaitu sebesar 56 persen sedangkan persentase tertinggi kedua adalah korban luka berat yaitu sebesar 26,06 persen kemudian korban meninggal sebesar 17,10 persen.

Jenis kecelakaan korban lalu lintas di Surabaya dengan jumlah persentase tertinggi adalah tabrakan samping, kemudian tabrakan depan, tabrakan belakang, hilang kendali dan lain-lain dengan persentase masing-masing sebesar 58,15 persen, 19 persen, 16,8 persen, 5,1 persen dan 0,6 persen yang dapat dilihat

pada Gambar 4.2. Korban kecelakaan lalu lintas dengan jumlah terbesar adalah korban dengan jenis kelamin laki-laki. Hal ini ditunjukkan dengan persentase tertinggi pada Gambar 4.3 y aitu laki-laki sebesar 70,30 persen sedangkan perempuan sebesar 29,70 persen.

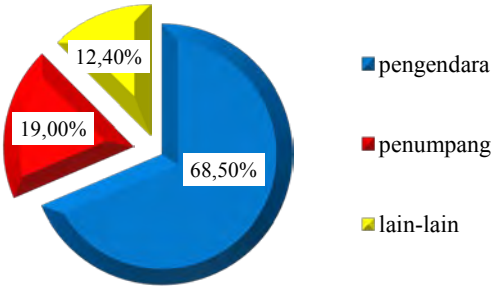


Gambar 4.2 Jenis Kecelakaan Korban Kecelakaan Lalu Lintas

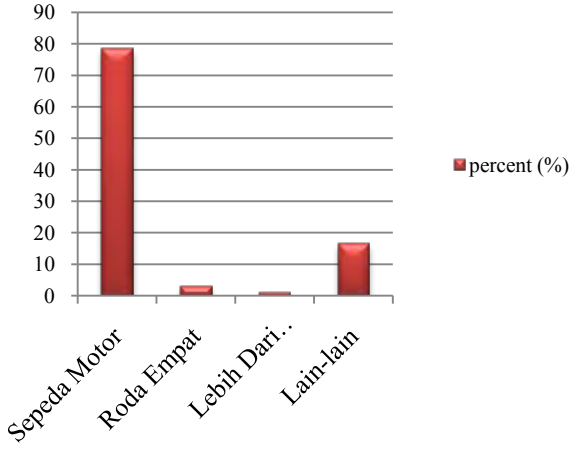


Gambar 4.3 Jenis Kelamin Korban Kecelakaan Lalu Lintas

Dalam berkendara di lalu lintas ada berbagai macam pola tingkah laku dari pengendara meliputi pengendara, penumpang, lain-lain yang termasuk pejalan kaki dan penyebrang jalan. Pada tahun 2013, korban kecelakaan yang terbanyak adalah korban sebagai pengendara yaitu sebesar 68,50 persen, korban kecelakaan sebagai penumpang sebesar 19,00 persen, korban kecelakaan sebagai pejalan kaki dan penyebrang jalan sebesar 12,40. Hal ini disebabkan karena pengendara yang posisinya di depan sehingga kalau terjadi kecelakaan sasaran yang terkena utama dengan lawan adalah pengendara.

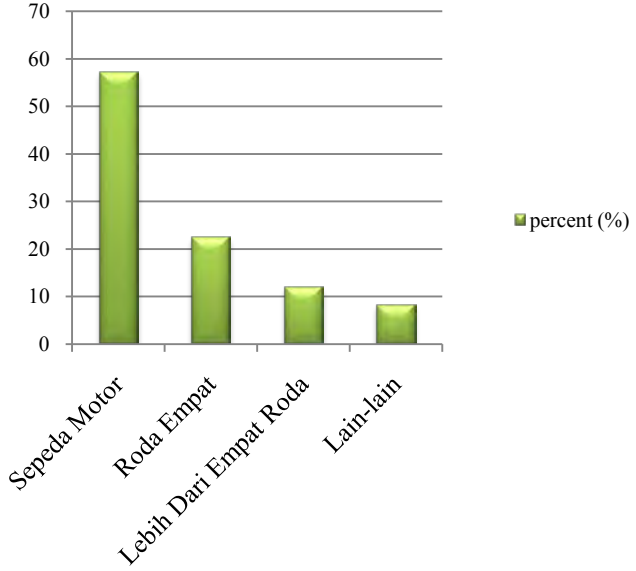


Gambar 4.4 Peran Korban Kecelakaan Lalu Lintas



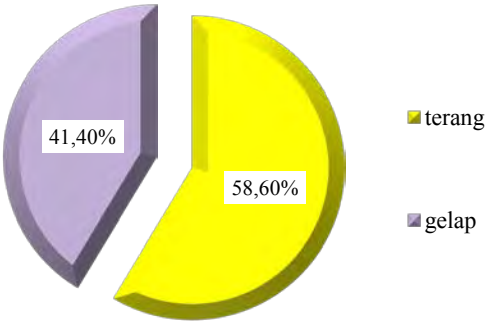
Gambar 4.5 Jenis Kendaraan Korban Kecelakaan Lalu Lintas

Lalu lintas merupakan gerak kendaraan dan orang di ruang lalu lintas jalan. Kendaraan dalam lalu lintas beraneka ragam meliputi kendaraan bermotor, kendaraan tidak bermotor, pejalan kaki dan lain sebagainya. Dengan berbagai macam kendaraan dalam lalu lintas, maka tidak dapat dipungkiri hal-hal yang tidak diinginkan juga terjadi misalnya kecelakaan lalu lintas. Kecelakaan lalu lintas di Surabaya yang terjadi pada tahun 2013 banyak memakan korban. Korban kecelakaan terbesar adalah dengan jenis kendaraan korban sepeda motor yaitu sebesar 78,60 persen sama halnya dengan jenis kendaraan lawan yang dapat dilihat pada Gambar 4.5 dan kendaraan lawan sebesar 57,3 persen yang dapat dilihat pada Gambar 4.6.

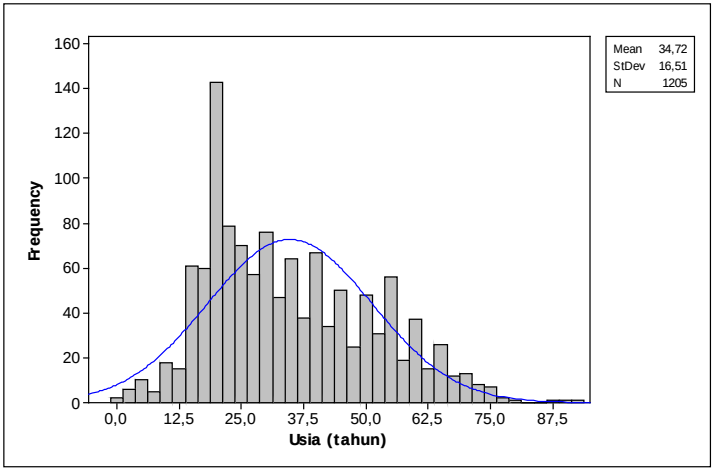


Gambar 4.6 Jenis KendaraanLawan Kecelakaan Lalu Lintas

Waktu kejadian kecelakaan tertinggi terjadi pada waktu siang hari (terang) pada pukul 06.00 WIB sampai 18.00 WIB dikarenakan pada waktu siang hari lebih banyak orang berkendara daripada diam di rumah. Hal ini ditunjukkan pada Gambar 4.7 bahwa persentase kecelakaan terjadi pada waktu terang sebesar 58,60 persen.



Gambar 4.7 Waktu Kejadian Kecelakaan Lalu Lintas



Gambar 4.8 Histogram Usia Korban Kecelakaan Lalu Lintas

Usia korban kecelakaan lalu lintas di Surabaya pada tahun 2013 yang terbesar adalah korban di usia kisaran antara 18,75 tahun sampai 21,25 tahun, yang dapat diambil titik tengahnya yaitu 20 tahun. Jadi, korban kecelakaan lalu lintas di Surabaya pada tahun 2013 terbanyak kurang lebih berusia 20 tahun yang terdapat pada histogram usia korban kecelakaan lalu lintas pada Gambar 4.8. Rata-rata (mean) usia korban kecelakaan lalu lintas di Surabaya adalah 35 tahun artinya jumlah korban kecelakaan

lalu lintas di Surabaya 50 persen berusia dibawah 35 tahun dan 50 persen berusia diatas 35 tahun. Varian usia korban kecelakaan lalu lintas sebesar 272,69 artinya begitu banyak anekaragam usia korban kecelakaan lalu lintas di Surabaya dimana semakin tinggi varian maka semakin tinggi pula keanekaragamnya. Usia korban kecelakaan lalu lintas yang paling rendah berusia 0,5 tahun dan yang tertinggi berusia 93 tahun.

4.2 Analisis Regresi Logistik Multinomial untuk Klasifikasi Korban Kecelakaan Lalu Lintas di Surabaya

Data pengamatan yang digunakan untuk mengklasifikasikan jenis keparahan korban kecelakaan lalu lintas di Surabaya adalah sebesar 1205 pengamatan korban kecelakaan, data tersebut dapat dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Pembagian data menjadi dua bagian bertujuan untuk mendapatkan suatu model yang optimal dengan melakukan percobaan klasifikasi menggunakan beberapa kombinasi data *training* dan data *testing*. Kombinasi data *training* dan data *testing* yang dicobakan dalam penelitian ini adalah 75%:25%, 80%:20%, 85%:15%, 90% :10%, 95%:5%.

Hasil klasifikasi yang diperoleh dari tiap-tiap kombinasi selanjutnya dilakukan perbandingan terhadap ketepatan klasifikasi *training* dan *testing*. Kombinasi proporsi data *training* dan *testing* terbaik dipilih berdasarkan nilai ketepatan klasifikasi yang paling besar. Data *training* digunakan untuk mengestimasi parameter model dan data *testing* digunakan untuk mencerminkan kebaikan model dalam mengklasifikasikan data baru.

Analisis regresi logistik multinomial dilakukan dengan tahapan awal uji serentak, kemudian uji parsial hingga diperoleh suatu model. Berikut diberikan penjelasan untuk masing-masing tahapan analisis klasifikasi mengggunakan regresi logistik multinomial menggunakan kombinasi data *training* dan *testing* sebesar 95%:5%.

4.2.1 Uji Serentak Terhadap Variabel-variabel Yang Berpengaruh Terhadap Korban Kecelakaan

Langkah awal dalam analisis regresi logistik multinomial adalah uji serentak terhadap variabel-variabel yang berpengaruh terhadap korban kecelakaan lalu lintas. Pada analisis ini

digunakan data *training* sebesar 1145 data dan data *testing* sebesar 60 data.

Tabel 4.1 Uji Serentak Regresi Logistik Multinomial

Variabel	*Chi-Square*	*df*	*P-value*
Jenis Kecelakaan	24,581	8	0,002
Jenis Kelamin	0,377	2	0,828
Peran Korban	16,163	4	0,003
Kendaraan Korban	4,350	6	0,629
Kendaraan Lawan	41,879	6	0,000
Waktu	9,484	2	0,009
Usia	13,686	2	0,001

Variabel prediktor yang berpengaruh secara signifikan terhadap korban kecelakaan lalu lintas adalah jenis kecelakaan, peran korban dalam kecelakaan, kendaraan lawan, waktu dan usia. Hal ini ditunjukkan dengan *p-value* yang lebih kecil dari 0,05 dimana *p-value* variabel jenis kecelakaan sebesar 0,002, *p-value* variabel peran korban sebesar 0,003, *p-value* variabel kendaraan lawan sebesar 0,000, *p-value* variabel waktu sebesar 0,009 dan *p-value* variabel usia sebesar 0,001 yang dapat dilihat pada Tabel 4.1.

4.2.2 Uji Parsial Terhadap Variabel-variabel Yang Berpengaruh Terhadap Korban Kecelakaan

Langkah kedua setelah dilakukan uji serentak, selanjutnya dilakukan uji parsial terhadap variabel-variabel yang berpengaruh terhadap korban kecelakaan. Berdasarkan dari hasil analisis uji serentak didapatkan bahwa variabel yang berpengaruh adalah jenis kecelakaan, peran korban, jenis kendaraan lawan, waktu dan usia. Akan tetapi, tidak diketahui kategori dari lima variabel yang berpengaruh terhadap korban kecelakaan sehingga dilakukan uji parsial untuk mengetahui variabel prediktor yang mempengaruhi korban kecelakaan lalu lintas dengan berbagai macam kategorinya. Berikut uji parsial untuk mengetahui variabel yang berpengaruh terhadap kecelakaan lalu lintas ditampilkan pada Tabel 4.2.

Tabel 4.2 Uji Parsial Regresi Logistik Multinomial

Variabel Respon	Variabel Prediktor	B	Wald	*df*	*P-value*	Exp(B)
Korban Meninggal	Konstan	-0,851	0,492	1	0,483	
	Jenis Kecelakaan (0)	-0,562	0,246	1	0,620	0,570
	Jenis Kecelakaan (1)	-0,623	0,301	1	0,583	0,537
	Jenis Kecelakaan (2)	-0,818	0,534	1	0,465	0,441
	Jenis Kecelakaan (3)	1,026	0,749	1	0,387	2,789
	Jenis Kelamin (0)	0,128	0,363	1	0,547	1,137
	Peran Korban (0)	-0,651	1,904	1	0,168	0,521
	Peran Korban (1)	-1,066	4,290	1	0,038	0,345
	Kendaraan Korban (0)	0,011	0,001	1	0,979	1,011
	Kendaraan Korban (1)	-0,713	1,295	1	0,255	0,490
	Kendaraan Korban (2)	-0,677	0,705	1	0,401	0,508
	Kendaraan Lawan (0)	-0,536	1,705	1	0,192	0,585
	Kendaraan Lawan (1)	-0,244	0,344	1	0,558	0,783
	Kendaraan Lawan (2)	1,155	6,984	1	0,008	3,173
	Waktu (0)	0,543	9,263	1	0,002	1,721
	Usia	0,020	13,077	1	0,000	1,021

Tabel 4.2 Uji Parsial Regresi Logistik Multinomial (*Lanjutan*)

Variabel Respon	Variabel Prediktor	B	Wald	*df*	*P-value*	Exp(B)
Korban Luka Berat	Konstan	0,821	0,598	1	0,439	
	Jenis Kecelakaan (0)	-0,946	0,890	1	0,345	0,388
	Jenis Kecelakaan (1)	-0,772	0,596	1	0,440	0,462
	Jenis Kecelakaan (2)	-1,030	1,080	1	0,299	0,357
	Jenis Kecelakaan (3)	-1,106	1,006	1	0,316	0,331
	Jenis Kelamin (0)	0,045	0,072	1	0,789	1,046
	Peran Korban (0)	-0,699	3,378	1	0,066	0,497
	Peran Korban (1)	-1,361	10,364	1	0,001	0,256
	Kendaraan Korban (0)	-0,115	0,113	1	0,737	0,892
	Kendaraan Korban (1)	-0,727	1,652	1	0,199	0,483
	Kendaraan Korban (2)	-0,309	0,160	1	0,689	0,734
	Kendaraan Lawan (0)	0,002	0,000	1	0,996	1,002
	Kendaraan Lawan (1)	0,077	0,042	1	0,837	1,080
	Kendaraan Lawan (2)	0,559	1,836	1	0,175	1,749
	Waktu (0)	0,062	0,177	1	0,674	1,064
	Usia	0,002	0,196	1	0,658	1,002

Setelah dilakukan uji serentak, selanjutnya dilakukan uji parsial pada masing-masing variabel prediktor dengan berbagai

kategorinya. Berdasarkan variabel yang signifikan pada uji parsial yang ditunjukkan dengan *p-value* yang lebih kecil dari 0,05. Selanjutnya, dari variabel yang signifikan akan dilakukan interpretasi model menggunakan nilai *odds ratio*. Nilai *odds ratio* sebesar 3,173 pada variabel kendaraan lawan (2) artinya resiko seseorang dengan kendaraan lawan lebih dari empat roda menjadi korban meninggal daripada korban luka ringan sebesar 3,173 kali lipat lebih besar dari seseorang dengan kendaraan lawan dengan kategori lain-lain (bukan kendaraan bermotor seperti becak, sepeda, pejalan kaki). Nilai *odds ratio* sebesar 1,721 pada variabel waktu (0) artinya resiko seseorang dengan waktu kejadian kecelakaan siang hari (terang) menjadi korban meninggal daripada korban luka ringan sebesar 1,721 kali lipat lebih besar dari seseorang dengan waktu kejadian kecelakaan malam hari (gelap). Nilai *odds ratio* sebesar 1,021 pada variabel usia artinya semakin tinggi usia seseorang, maka kecenderungan seseorang tersebut akan menjadi korban meninggal 1,021 k ali lipat dari korban luka ringan. Nilai *odds ratio* sebesar 0,256 pada variabel peran korban (1) artinya resiko seseorang yang berperan sebagai penumpang menjadi korban luka berat daripada korban luka ringan sebesar 0,256 kali lipat lebih besar dari seseorang yang berperan sebagai pengguna jalan bukan penumpang kendaraan (penyebrang jalan, pejalan kaki, dll) atau resiko seseorang yang berperan sebagai pengguna jalan bukan penumpang kendaraan (penyebrang jalan, pejalan kaki, dll) menjadi korban luka berat daripada korban luka ringan sebesar (1/0,256=3,906) kali lipat lebih besar dari seseorang yang berperan sebagai penumpang. Berdasarkan faktor-faktor yang mempengaruhi korban kecelakaan lalu lintas yang telah dijelaskan di atas dapat dijadikan sebagai masukan (saran) bagi Unit Laka Satlantas Polrestabes Surabaya dan bagi pengendara lalu lintas.

4.2.3 Model Regresi Logistik Multinomial dalam Kasus Korban Kecelakaan Lalu Lintas

Tahapan selanjutnya, setelah dilakukan uji serentak dan uji parsial adalah membuat persamaan model regresi logistik multinomial yang digunakan untuk pengklasifikasian baik data

training maupun data *testing* yang selanjutnya dilakukan interpretasi terhadap model tersebut. Berdasarkan nilai parameter pada analisis uji parsial yang ditampilkan pada Tabel 4.2 didapatkan persamaan logit untuk korban meninggal dan luka berat sebagai berikut.

$$\begin{aligned} g_0(x) = -0{,}851 &- 0{,}562x_1(0) - 0{,}623x_1(1) - 0{,}818x_1(2) \\ &+ 1{,}026x_1(3) + 0{,}128x_2(0) - 0{,}651x_3(0) \\ &- 1{,}066x_3(1) + 0{,}011x_4(0) - 0{,}713x_4(1) \\ &- 0{,}677x_4(2) - 0{,}536x_5(0) - 0{,}244x_5(1) \\ &+ 1{,}155x_5(2) + 0{,}543x_6(0) + 0{,}020x_7 \end{aligned}$$

$$\begin{aligned} g_1(x) = 0{,}821 &- 0{,}946x_1(0) - 0{,}772x_1(1) - 1{,}030x_1(2) \\ &- 1{,}106x_1(3) + 0{,}045x_2(0) - 0{,}699x_3(0) \\ &- 1{,}361x_3(1) - 0{,}115x_4(0) - 0{,}727x_4(1) \\ &- 0{,}309x_4(2) + 0{,}002x_5(0) + 0{,}077x_5(1) \\ &+ 0{,}559x_5(2) + 0{,}062x_6(0) + 0{,}002x_7 \end{aligned}$$

Selanjutnya persamaan $g_0(x)$ dan $g_1(x)$ disubtitusikan pada model regresi logistik multinomial berikut.

$$\pi_0(x) = \frac{\exp g_0(x)}{1 + \exp g_0(x) + \exp g_1(x)} = \frac{e^{g_0(x)}}{1 + e^{g_0(x)} + e^{g_1(x)}}$$

$$\pi_1(x) = \frac{\exp g_1(x)}{1 + \exp g_0(x) + \exp g_1(x)} = \frac{e^{g_1(x)}}{1 + e^{g_0(x)} + e^{g_1(x)}}$$

$$\pi_2(x) = \frac{1}{1 + \exp g_0(x) + \exp g_1(x)} = \frac{1}{1 + e^{g_0(x)} + e^{g_1(x)}}$$

dimana $\pi_0(x)$ persamaan regresi logistik untuk korban meninggal, $\pi_1(x)$ untuk korban luka berat dan $\pi_2(x)$ untuk korban luka ringan.

4.2.4 Uji Kesesuaian Model

Uji kesesuaian model dilakukan untuk mengetahui model regresi logistik multinomial yang terbentuk pada kasus korban kecelakaan lalu lintas sudah sesuai atau belum yang akan digunakan dalam analisis pengklasifikasian korban kecelakaan lalu lintas. Uji kesesuain model ditampilkan pada Tabel 4.3.

Tabel 4.3 *Goodness of Fit* Data Korban Kecelakaan

	Chi-Square	*df*	*P-value*
Pearson	1816,220	1740	0,099

Model regresi logistik multinomial yang telah terbentuk sudah sesuai yaitu tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model. Hal ini ditunjukkan dengan *p-value* sebesar 0,099 yang nilainya lebih besar dari 0,05. Artinya, bahwa model regresi logistik dari data *training* dapat digunakan dalam analisis pengklasifikasian korban kecelakaan lalu lintas.

4.2.5 Klasifikasi Korban Kecelakaan Lalu Lintas

Klasifikasi korban kecelakaan lalu lintas digunakan untuk mengetahui ketepatan model yang telah terbentuk dalam mengestimasi parameter model menggunakan data *training* dan untuk mengetahui ketepatan dalam mengklasifikasikan data baru menggunakan data *testing*. Berikut klasifikasi korban kecelakaan lalu lintas.

Tabel 4.4 Klasifikasi Korban Kecelakaan Lalu Lintas

Observasi		Prediksi			Total	1-APER (%)
		0	1	2		
Data *Training*	0	43	12	141	196	57,80
	1	16	17	268	301	
	2	24	22	602	648	
Data *Testing*	0	2	0	8	10	63,33
	1	1	0	12	13	
	2	1	0	36	37	

Ketepatan klasifikasi untuk data *training* sebesar 57,80 persen sedangkan ketepatan klasifikasi untuk data *testing* sebesar 63,33 persen dapat dilihat pada Tabel 4.4. Model regresi logistik multinomial untuk data korban kecelakaan lalu lintas mempunyai ketepatan 57,80 persen dalam mengestimasi parameter model sedangkan model regresi logistik multinomial untuk data korban kecelakaan lalu lintas mempunyai ketepatan 63,33 persen dalam mengklasifikasikan data baru.

4.2.6 Pemilihan Kombinasi Data *Training* dan Data *Testing* Terbaik dalam Analisis Regresi Logistik Multinomial

Proses analisis regresi logistik multinomial yang telah dijelaskan sebelumnya dijalankan pula pada kombinasi data *training* dan *testing* lainnya yaitu 75%:25%, 80%:20%, 85%:15% dan 90%:10%. Berdasarkan hasil pengolahan analisis regresi logistik multinomial, maka diperoleh hasil ketepatan klasifikasi (1-APER) untuk masing-masing kombinasi data *training* dan *testing* sebagai berikut.

Tabel 4.5 Perbandingan *Total Accuracy Rate* Beberapa Kombinasi Data

Kombinasi Data *Training* dan *Testing*	*Total Accuracy Rate* (1-APER) (dalam %)	
	Data *Training*	Data *Testing*
75%:25%	58,20	57,14
80%:20%	57,80	58,92
85%:15%	59,50	55,80
90%:10%	57,80	58,33
95%:5% *	57,80	63,33

*kombinasi data *training* dan *testing* terpilih

Tabel 4.5 menunjukkan bahwa kombinasi data *training* dan data *testing* yang mampu memberikan ketepatan klasifikasi yang tinggi adalah pada kombinasi data *training* sebesar 95 persen dan data *testing* sebesar 5 persen. Ketepatan klasifikasi yang dipilih dari beberapa kombinasi data *training* dan *testing* adalah ketepatan klasifikasi pada data *training* dan data *testing* yang paling tinggi. Apabila pada data *training* memberikan nilai ketepatan klasifikasi yang paling tinggi akan tetapi pada data *testing* tidak memberikan nilai ketepatan klasifikasi yang paling tinggi begitu juga sebaliknya sehingga yang pilih adalah kombinasi data *training* dan *testing* yang memberikan nilai ketepatan klasifikasi tertinggi pada ketepatan akurasi data *testing*.

Ketepatan klasifikasi untuk data *testing* yang diperoleh adalah sebesar 63,33 persen yang merupakan nilai tertinggi diantara semua kombinasi data. Untuk ketepatan klasifikasi data *training* diperoleh ketepatan klasifikasi sebesar 57,80 persen. Oleh karena itu, hasil analisis yang digunakan untuk menjelaskan

jenis keparahan korban kecelakaan lalu lintas di Surabaya adalah hasil klasifikasi yang menggunakan kombinasi data *training* dan *testing* terbaik, yaitu data *training* sebesar 95 persen dan data *testing* sebesar 5 persen.

4.3 Analisis CART dan *Random Forests* untuk Klasifikasi Korban Kecelakaan Lalu Lintas di Surabaya

Random forests merupakan salah satu metode *ensemble* yang berguna untuk meningkatkan akurasi klasifikasi dari sebuah pemilah tunggal yang tidak stabil dengan cara mengkombinasikan banyak pemilah dari suatu metode yang sama (CART) melalui proses *voting.* Jadi, hasil klasifikasi menggunakan *random forests* berasal dari pengulangan metode CART sehingga perlu dilakukan analisis menggunakan metode CART terlebih dahulu.

4.3.1 Analisis CART untuk Klasifikasi Korban Kecelakaan Lalu Lintas di Surabaya

Sebelum dilakukan analisis *random forests*, perlu dilakukan analisis CART karena analisis CART adalah metode yang mendasari dari analisis *random forests* itu sendiri. Pada analisis CART digunakan data sebesar 1205 korban kecelakaan sama halnya yang digunakan pada analisis regresi logistik multinomial. Data dapat dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Kombinasi data *training* dan *testing* yang dicobakan dalam penelitian ini adalah 75%:25%, 80%:20%, 85%:15%, 90%:10% dan 95%:5%.

Analisis CART diawali dengan langkah pembentukan pohon klasifikasi yang maksimal. Berikut penjelasan untuk masing-masing tahapan analisis klasifikasi CART dengan menggunakan kombinasi data *training* dan *testing* sebesar 95%:5%. Serangkaian tahapan analisis CART berikut juga dilakukan untuk kombinasi *training* dan *testing* yang lain yaitu 75%:25%, 80%:20%, 85%:15% dan 90%:10%.

a. Pembentukan Pohon Klasifikasi Maksimal

Tahapan awal yang dilakukan untuk membentuk pohon klasifikasi adalah dengan menentukan variabel pemilah dan nilai variabel (*threshold*). Variabel pemilah dan *treshold* dipilih dari beberapa kemungkinan pemilah dari masing-masing variabel.

Kemungkinan banyaknya pemilah untuk variabel prediktor kategorik menggunakan banyak kategori dari variabel tersebut. Perhitungan banyaknya kemungkinan pemilah ditampilkan pada Tabel 4.6.

Tabel 4.6 Perhitungan Kemungkinan Jumlah Pemilah dari Setiap Variabel

Variabel	Nama Variabel	Skala Data	Banyaknya Kategori (Nilai Amatan Sampel)	Kemungkinan Pemilah
X_1	Jenis Kecelakaan	Nominal	5	$2^{5-1}-1=15$ pemilah
X_2	Jenis Kelamin	Nominal	2	$2^{2-1}-1=1$ pemilah
X_3	Peran Korban	Nominal	3	$2^{3-1}-1=3$ pemilah
X_4	Jenis Kendaraan Korban	Nominal	4	$2^{4-1}-1=7$ pemilah
X_5	Jenis Kendaraan Lawan	Nominal	4	$2^{4-1}-1=7$ pemilah
X_6	Waktu Kejadian	Nominal	2	$2^{2-1}-1=1$ pemilah
X_7	Usia Korban	Rasio	82	82-1=81 pemilah

Dari berbagai kemungkinan pemilah dari tiap variabel, selanjutnya dihitung Indeks Gini yang merupakan ukuran keheterogenan simpul. Indeks Gini lebih sering digunakan karena alasan kesederhanaan dalam proses perhitungan. Cara kerja Indeks Gini adalah melakukan pemilihan simpul dengan berfokus pada masing-masing simpul kanan atau kiri. Hasil perhitungan Indeks Gini kemudian digunakan utuk menentukan *goodness of split* dari masing-masing pemilah. Pemilah yang terpilih adalah variabel pemilah dan nilai variabel (*threshold*) yang memiliki nilai *goodness of split* tertinggi. Pemilah yang terpilih merupakan

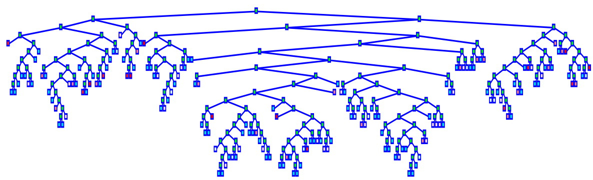
variabel yang terpenting dalam mengklasifikasikan data pengamatan. Besarnya kontribusi variabel sebagai pemilah baik pemilah utama maupun pengganti pada pohon klasifikasi maksimal yang terbentuk ditunjukkan melalui suatu angka skor yang ditampilkan pada Tabel 4.7.

Tabel 4.7 Skor Variabel Terpenting dari Pohon Klasifikasi Maksimal

<table>
<tr><th>Variabel</th><th>Skor Variabel</th><th></th></tr>
<tr><td>X7</td><td>100,00</td><td>||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>X1</td><td>43,02</td><td>|||||||||||||||||</td></tr>
<tr><td>X5</td><td>31,72</td><td>|||||||||||||</td></tr>
<tr><td>X4</td><td>26,38</td><td>||||||||||</td></tr>
<tr><td>X3</td><td>26,9</td><td>||||||||||</td></tr>
<tr><td>X2</td><td>17,85</td><td>|||||||</td></tr>
<tr><td>X6</td><td>11,82</td><td>||||</td></tr>
</table>

Tabel 4.7 menunjukkan bahwa semua variabel menjadi pembangun pohon klasifikasi. Akan tetapi, berdasarkan skor yang dihasilkan diketahui bahwa variabel yang terpenting dan menjadi pemilah utama dalam mengklasifikasikan korban kecelakaan lalu lintas adalah usia korban (X1) karena memiliki skor paling tinggi yaitu sebesar 100. Selain itu, terdapat beberapa variabel yang juga berpengaruh dalam melakukan pemilihan yaitu jenis kecelakaan sebesar 43,02 begitu juga seterusnya sampai waktu kejadian (X6).

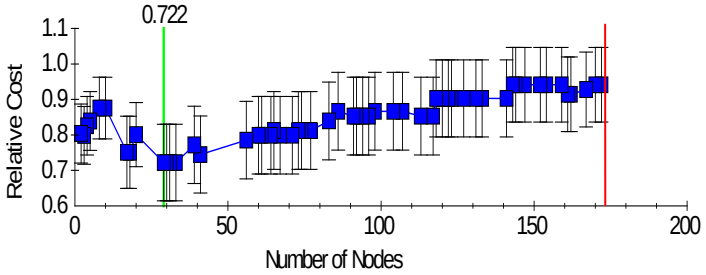
Hasil penyekatan rekursif secara biner dari data pengamatan yang digunakan akan menghasilkan pohon klasifikasi yang berukuran relatif besar dan tingkat kedalaman (*depth*) yang tinggi. Pohon klasifikasi tersebut disebut sebagai pohon klasifikasi maksimal yang ditunjukkan pada Gambar 4.9.



Gambar 4.9 Konstruksi Pohon Klasifikasi Maksimal untuk Korban Kecelakaan Lalu Lintas

b. Pemangkasan Pohon Klasifikasi Maksimal (*Pruning*)

Untuk mempermudah proses analisis, pohon klasifikasi maksimal yang dihasilkan kemudian dilakukan pemangkasan secara iteratif berdasarkan kriteria *test set relative cost*. Setiap hasil pemangkasan memiliki nilai *relative cost* tertentu, sehingga kemudian dipilih hasil pemangkasan dengan nilai *relative cost* yang minimum.



Gambar 4.10 Plot *Relative Cost* dan Jumlah *Terminal Nodes* dalam Klasifikasi Korban Kecelakaan Lalu Lintas

Berdasarkan Gambar 4.10, garis hijau menunjukkan pohon klasifikasi optimal sedangkan garis merah menujukkan klasifikasi maksimal. Pohon klasifikasi maksimal yang terbentuk terdiri dari 172 *terminal nodes* dan *relative cost* sebesar 0,941±0,105 yang dapat dilihat pada Tabel 4.8. Pemangkasan pohon di lakukan secara iteratif berdasarkan *test set relative cost* yang minimum. Tabel 4.8 menunjukkan bahwa nilai *test set relative cost* yang minimum adalah pada saat jumlah pohon 46 dan *terminal nodes* 29 sehingga dapat dikatakan bahwa pohon k lasifikasi optimal yang terbentuk terdiri dari 29 *terminal nodes*.

Tabel 4.8 Pembentukan Pohon Klasifikasi (*Tree Sequence*)

Tree	*Terminal Nodes*	*Test Set Relative Cost*	*Resubtitution Relative Cost*	*Complexity Parameter*
1	172	0,941±0,105	0,416	0,000
46*	29	0,722±0,108	0,657	0,003
47	20	0,801±0,091	0,697	0,003
48	18	0,751±0,102	0,707	0,003
49	17	0,751±0,102	0,712	0,003

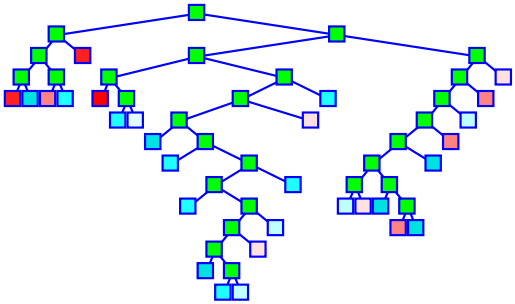
Tabel 4.8 Pembentukan Pohon Klasifikasi (*Tree Sequence*) (*Lanjutan*)

Tree	*Terminal Nodes*	*Test Set Relative Cost*	*Resubtitution Relative Cost*	*Complexity Parameter*
50	10	0,876±0,087	0,749	0,004
51	8	0,876±0,087	0,760	0,004
52	5	0,839±0,083	0,788	0,006
53	4	0,826±0,082	0,800	0,008
54	3	0,799±0,082	0,831	0,021
55	2	0,804±0,083	0,878	0,031
56	1	1,000±0,000	1,000	0,081

*pohon klasifikasi optimal

c. Pemilihan Pohon Klasifikasi Optimal

Hasil pemangkasan yang diperoleh dari Gambar 4.10 selanjutnya digunakan untuk memilih pohon k lasifikasi yang optimal. Pohon klasifikasi yang optimal dengan jumlah *terminal nodes* 29, *test set relative cost* sebesar 0,722±0,108.



Gambar 4.11 Konstruksi Pohon Klasifikasi Optimal untuk Korban Kecelakaan Lalu Lintas

Setiap simpul pada pohon klasifikasi yang terbentuk telah dilakukan pelabelan kelas menjadi tiga yaitu merah tua bila suatu simpul diklasifikasikan sebagai korban meninggal, merah muda bila suatu simpul diklasifikasikan sebagai korban luka berat dan biru bila diklasifikasikan sebagai korban luka ringan yang ditunjukkan pada Gambar 4.11.

Pohon klasifikasi optimal dipengaruhi oleh semua variabel prediktor. Akan tetapi, urutan variabel terpenting dalam pohon

klasifikasi optimal adalah usia korban, jenis kecelakaan, jenis kendaraan lawan, jenis kendaraan korban, peran korban dalam kecelakaan, waktu kejadian kecelakaan dan jenis kelamin korban yang ditunjukkan pada Tabel 4.9.

Tabel 4.9 Skor Variabel Terpenting dari Pohon Klasifikasi Optimal

Variabel	Skor Variabel	
X7	100,00	||||||||||||||||||||||||||||||||||||||||
X1	91,13	|||||||||||||||||||||||||||||||||||||
X5	84,33	||||||||||||||||||||||||||||||||||
X4	62,99	|||||||||||||||||||||||||
X3	35,44	||||||||||||||
X6	13,10	|||||
X2	7,89	||

Secara keseluruhan dapat diketahui bahwa terdapat sebanyak 11 simpul terminal yang diklasifikasikan sebagai korban meninggal, 9 s impul terminal yang diklasifikasikan sebagai korban luka berat dan sebanyak 9 simpul terminal yang diklasifikasikan sebagai korban luka ringan yang ditampilkan pada Lampiran E. Karakteristik dari beberapa simpul terminal yang paling kuat mengklasifikasikan data menjadi salah satu dari ketiga kategori kecelakaan lalu lintas (dilihat dari nilai persentase terbesar dari hasil klasifikasi tiap terminal nodes ke dalam salah satu kategori korban kecelakaan) dijelaskan sebagai berikut.

1. Simpul terminal 6 adalah korban dengan kategori meninggal yaitu berusia kurang dari 52 tahun, dengan mengendarai kendaraan roda empat dan lebih dari empat roda dengan kendaraan lawan menggunakan sepeda motor dan lain-lain (sepeda, becak, pejalan kaki).
2. Simpul terminal 9 adalah korban dengan kategori luka ringan yaitu berusia kurang dari 48 tahun, dengan mengendarai kendaraan sepeda motor dan lain-lain (sepeda, becak, pejalan kaki) dengan kendaraan lawan menggunakan sepeda motor dan lain-lain (sepeda, becak, pejalan kaki) juga.
3. Simpul terminal 24 adalah korban dengan kategori luka berat yaitu berusia 52 sampai 71 tahun dengan mengendarai

kendaran lain-lain (sepeda, becak, pejalan kaki) dengan jenis kecelakaan tabrakan belakang, tabrakan samping, tabrakan depan, hilang kendali dan terjadi pada waktu malam hari (gelap).

4.3.2 Analisis *Random Forests* untuk Klasifikasi Korban Kecelakaan Lalu Lintas di Surabaya

CART adalah suatu metode pengklasifikasian yang mampu bekerja pada data yang berdimensi tinggi. Metode ini masih memiliki kelemahan dimana bila terjadi sedikit perubahan pada data *training* maka pohon yang dihasilkan akan memiliki perbedaan yang cukup besar atau dengan kata lain pohon klasifikasi yang dihasilkan tidak stabil sehingga perlu dicobakan suatu solusi yang mampu mengatasi keterbatasan tersebut. Salah satu cara yang dapat digunakan adalah dengan menjalankan suatu metode *ensemble* yaitu *random forests*. *Random forests* adalah suatu metode *ensemble* yang dapat meningkatkan akurasi prediksi dari pohon klasifikasi yang dihasilkan melalui proses pembentukan pohon dan prediksi pada beberapa kali pengambilan sampel dan memiliki perhitungan komputasi yang cepat. Algoritma *random forests* yang digunakan pada kombinasi data *training* dan *testing* adalah semua kombinasi yaitu 75%:25%, 80%:20%, 85%:15%, 90%:10% dan 95%:5%.

Dalam penelitian ini parameter kontrol yang ditentukan adalah variabel prediktor yang diambil secara acak dari tujuh variabel amatan yang digunakan sebagai pembentuk pohon dalam setiap pemilihan adalah tiga variabel yang diperoleh dari hasil akar banyaknya variabel prediktor yaitu 2,65 variabel yang dapat dibulatkan menjadi tiga variabel atau $\log_2 8 = 3$. Selain itu, jumlah pohon yang dibentuk akan dicobakan pada kombinasi 50, 100, 250, 500 dan 1000 pohon.

Dalam *random forests*, dilakukan pembentukan pohon klsifikasi hingga *K* kali replikasi. Setiap pohon klasifikasi (CART) akan menghasilkan prediksi kelas keparahan korban kecelakaan lalu lintas yang diharapkan untuk tiap observasi. Hasil prediksi dari sejumlah kombinasi pohon dilakukan *majority vote* hasil klasifikasi terhadap jenis keparahan korban kecelakaan lalu lintas. Apabila suatu korban kecelakaan lalu lintas diprediksi

lebih banyak sebagai korban meninggal maka korban kecelakaan lalu lintas tersebut diklasifikasikan ke dalam kategori 0 yaitu korban meninggal. Apabila suatu korban kecelakaan lalu lintas diprediksi lebih banyak sebagai korban luka berat maka korban kecelakaan lalu lintas tersebut diklasifikasikan ke dalam kategori 1 yaitu korban luka berat. Apabila suatu korban kecelakaan lalu lintas diprediksi lebih banyak sebagai korban luka ringan maka korban kecelakaan lalu lintas tersebut diklasifikasikan ke dalam kategori 2 yaitu korban luka ringan. Hasil klasifikasi dari masing-masing kombinasi pohon yang dicobakan ditampilkan pada tabel 4.10.

Tabel 4.10 menunjukkan bahwa kombinasi data *training* dan *testing* yang terpilih adalah 80%:20% dan jumlah pohon sebesar 250. Hal itu disebabkan pada kombinasi data *training* dan *testing* 80%:20% dan jumlah pohon s ebesar 250 mempunyai ketepatan klasifikasi yang tertinggi yaitu sebesar 59,34 persen. Selanjutnya, dilakukan ketepatan klasifikasi pada kombinasi data *training* dan *testing* 95%:5% karena kombinasi terpilih dari regresi logistik adalah kombinasi data *training* dan *testing* sebesar 95%:5% yang akan digunakan sebagai pembanding. Kombinasi data *training* dan *testing* 95%:5% menggunakan *random forests* yang terbaik adalah dengan banyak pohon klasifikasi sebesar 50 karena mempunyai ketepatan akurasi tertinggi untuk data *testing*.

Berdasarkan Tabel 4.11 menyatakan bahwa dari 1145 data korban kecelakaan lalu lintas sebagai data *training*, terdapat sebanyak 116 korban kecelakaan yang tepat diklasifikasikan sebagai korban meninggal, terdapat sebanyak 161 k orban kecelakaan yang tepat diklasifikasikan sebagai korban luka berat dan terdapat sebanyak 624 korban kecelakaan yang tepat diklasifikasikan sebagai korban luka ringan. Secara keseluruhan didapatkan *apparent error rate* (APER) sebesar 21,31 persen dan *total accuracy rate* (1-APER) sebesar 78,69 persen untuk data *training*.

Validasi dengan menggunakan data *testing* terdapat sebanyak 3 korban kecelakaan yang tepat diklasifikasikan sebagai

Tabel 4.10 Perbandingan Ketepatan Klasifikasi *Random Forests* pada Beberapa Kombinasi Data dan Jumlah Pohon

Kombinasi Data *Training* dan *Testing*	Kombinasi Jumlah Pohon									
	K=50		*K*=100		*K*=250*		*K*=500		*K*=1000	
	Training	*Testing*	*Training*	*Testing*	*Training*	*Testing*	*Training*	*Testing*	*Training*	*Testing*
75%:25%	79,87	52,16	79,98	53,82	80,09	52,16	79,98	52,49	80,53	52,82
80%:20%*	79,25	56,02	77,59	58,1	78,42	59,34	78,31	58,51	79,36	57,26
85%:15%	77,93	53,59	78,91	52,49	79,00	53,04	78,22	53,59	78,91	53,59
90%:10%	78,6	58,68	79,34	56,2	79,61	55,37	79,43	55,38	79,15	57,85
95%:5%	78,69	58,33	80,17	53,33	78,78	55,00	79,13	55,00	79,65	56,67

*jumlah kombinasi data dan jumlah pohon terpilih

Tabel 4.11 Ketepatan Klasifikasi *Random Forests* pada Kombinasi 50

Observasi		Prediksi			Total	APER (%)	1-APER (%)
		0	1	2			
Data *Training*	0	116	11	10	137	21,31	78,69
	1	13	161	14	188		
	2	67	129	624	820		
Data *Testing*	0	3	1	1	5	41,67	58,33
	1	0	1	5	6		
	2	7	11	31	49		

korban meninggal, terdapat sebanyak 1 korban kecelakaan yang tepat diklasifikasikan sebagai korban luka berat dan terdapatsebanyak 31 korban kecelakan yang tepat diklasifikasikan sebagai korban luka ringan. Secara keseluruhan didapatkan *apparent error rate* (APER) dan *total accuracy rate* (1-APER) untuk data *testing* sebesar 41,67 persen dan 58,33 persen.

4.4 Perbandingan Hasil Klasifikasi Regresi Logistik Multinomial dan *Random Forests*

Metode yang digunakan dalam pengklasifikasian korban kecelakaan lalu lintas di Surabaya adalah regresi logistik multinomial dan *random forests*. Kriteria yang digunakan untuk membandingkan antara kedua metode tersebut adalah *apparent error rate* (APER) dan *total accuracy rate* (1-APER) dari masing-masing data *training* dan data *testing*. Metode klasifikasi yang terbaik dipilih sebagai metode pengklasifikasian korban kecelakaan lalu lintas di Surabaya dengan melihat nilai APER yang terkecil dan (1-APER) yang terbesar.

Tabel 4.12 Perbandingan Hasil Klasifikasi Regresi Logistik Multinomial dan *Random Forests*

Metode Klasifikasi	Data	APER (%)	1-APER (%)
Regresi Logistik Multinomial	*Training*	42,20	57,80
	Testing	36,67	63,33
Random Forests	*Training*	21,31	78,69
	Testing	41,67	58,33

Metode klasifikasi yang mempunyai nilai APER yang terkecil dan 1-APER terbesar untuk data *training* adalah metode *random forests* akan tetapi untuk data *testing* nilai APER yang terkecil dan 1-APER terbesar adalah metode regresi logistik multinomial. Hal ini disebabkan karena distribusi data korban kecelakaan lalu lintas di Surabaya tidak *balance* antara korban meninggal, korban luka berat dan korban luka ringan dalam kecelakaan lalu lintas di Surabaya dimana jumlah korban terbesar adalah jumlah korban luka ringan sehingga pada saat prediksi klasifikasi menggunakan regresi logistik multinomial, hasil

prediksinya lebih banyak pada korban luka ringan yang menyebabkan regresi logistik multinomial lebih baik daripada *random forests*. Jadi dapat disimpulkan bahwa metode klasifikasi terbaik untuk kasus pengklasifikasian keparahan jenis korban kecelakaan lalu lintas di Surabaya adalah metode regresi logistik karena hasil ketepatan data *testing* lebih tinggi dibandingkan metode *random forests*.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan yang diperoleh dari penelitian ini adalah sebagai berikut.

1. Berdasarkan hasil deskripsi mengenai korban kecelakaan lalu lintas di Surabaya diketahui bahwa pada tahun 2013 korban yang paling tinggi adalah korban luka ringan dengan jenis kecelakaan tabrakan samping yang berjenis kelamin laki-laki sebagai pengendara menggunakan sepeda motor dengan kendaraan lawan juga menggunakan sepeda motor yang terjadi pada waktu terang dengan kisaran usia kurang lebih 20 tahun.
2. Klasifikasi korban kecelakaan lalu lintas menggunakan regresi logistik multinomial yang menggunakan kombinasi data *training* dan *testing* sebesar 75%:25%, 80%:20%, 85%:15%, 90%:10% dan 95%:5% bahwa yang memberikan ketepatan klasifikasi tertinggi pada ketepatan klasifikasi data *testing* adalah kombinasi data *training* dan *testing* sebesar 95%:5% dengan ketepatan untuk data *training* sebesar 57,80 persen dan ketepatan data *testing* sebesar 63,33 persen. Variabel yang berpengaruh secara signifikan terhadap korban kecelakaan lalu lintas adalah jenis kecelakaan, peran korban dalam kecelakaan, kendaraan lawan, waktu kejadian dan usia korban.
3. Pada analisis klasifikasi korban kecelakaan menggunakan *random forests* yang menggunakan kombinasi data *training* dan *testing* sebesar 95%:5% dengan berbagai kombinasi pohon dimana yang memberikan ketepatan klasifikasi pada data *testing* terbesar adalah pada kombinasi pohon sebanyak 50. Ketepatan klasifikasi dengan kombinasi pohon sebanyak 50 unt uk data *training* sebesar 78,69 persen dan data *testing* sebesar 58,33 persen.

4. Perbandingan kedua metode regresi logistik multinomial dan *random forests* dalam pengklasifikasian korban kecelakaan lalu lintas di Surabaya memberikan hasil ketepatan klasifikasi yang berbeda baik untuk data *training* maupun data *testing*. Ketepatan klasifikasi untuk data *training* lebih besar dengan menggunakan *random forests* yaitu sebesar 78,69 sedangkan ketepatan klasifikasi dengan menggunakan data *testing* lebih besar menggunakan regresi logistik yaitu sebesar 63,33 persen.

5.2 Saran

Saran yang dapat diberikan dari hasil penelitian ini adalah sebagai berikut.

1. Untuk dijadikan sebagai rekomendasi kepada peneliti selanjutnya yaitu penambahan faktor-faktor yang mempengaruhi korban kecelakaan lalu lintas di Surabaya serta penambahan data untuk mendapatkan hasil klasifikasi yang lebih informatif.
2. Berdasarkan hasil analisis bahwa jumlah tertinggi yang menjadi korban kecelakaan lalu lintas berusia kisaran 20 tahun diharapkan pihak Unit Laka Satlantas Polrestabes Surabaya mewajibkan mempunyai SIM bagi pihak pengendara. Korban kecelakaan juga sering terjadi pada waktu siang hari sehingga pihak Unit Laka Satlantas Polrestabes Surabaya lebih memperketat peraturan mengenai lalu lintas dengan cara diberikan sanksi apabila telah melanggar peraturan lalu lintas. Selain itu, untuk tahun selanjutnya pihak Unit Laka Satantas Polrestabes Surabaya mencatat *safety riding* para korban kecelakaan lalu lintas di Surabaya karena *safety riding* juga merupakan faktor yang dominan mempengaruhi korban kecelakaan lalu lintas. Misalkan pemakaian helm bagi pengendara roda dua dan pemakaian sabuk pengaman bagi pengendara roda empat atau lebih.
3. Untuk para pengendara di lalu lintas, sebaiknya lebih diperhatikan lagi bagi pengendara menggunakan sepeda

motor dengan usia kisaran 20 tahun untuk lebih berhati-hati dalam berkendara karena korban kecelakaan tertinggi pada usia kisaran 20 tahun. Hal itu disebabkan, usia kisaran 20 tahun, usia dimana seseorang suka mencoba-coba hal baru yang terkadang juga membahayakan diri sendiri seperti seorang pembalap, ugal-ugalan dan lain-lain. Selain itu, kondisi kendaraan juga perlu diperhatikan misalnya dalam hal spion, lampu kendaraan, rem dan lain-lain karena faktor kendaraan juga dapat mempengaruhi kecelakaan.

(*Halaman ini sengaja dikosongkan*)

DAFTAR LAMPIRAN

LAMPIRAN A

Data Pengamatan Korban Kecelakaan Lalu Lintas di Surabaya

Data Pengamatan Korban Kecelakaan Lalu Lintas untuk Statistika Deskriptif dan Analisis Regresi Logistik Multinomial

No	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	2	2	0	0	0	0	0	18
2	1	2	0	0	0	0	0	13
3	1	2	1	1	0	0	0	13
4	0	1	0	0	0	0	0	34
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1204	2	0	1	0	0	1	1	52
1205	2	2	1	1	0	0	0	19

Keterangan :

Y: Jenis keparahan korban kecelakaan

X_1 : Jenis kecelakaan korban

X_2 : Jenis kelamin korban

X_3 : Peran korban dalam kecelakaan

X_4 : Jenis kendaraan korban

X_5 : Jenis kendaraan lawan

X_6 : Waktu kejadian kecelakaan

X_7 : Usia korban

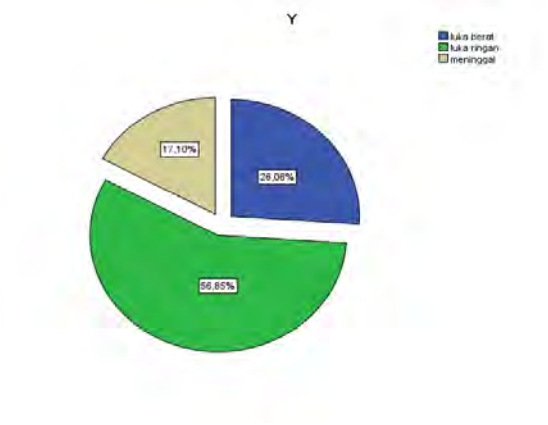
Data Pengamatan Korban Kecelakaan Lalu Lintas untuk Analisis *Random Forests*

No	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	c	c	a	a	a	a	a	18
2	b	c	a	a	a	a	a	13
3	b	c	b	b	a	a	a	13
4	a	b	a	a	a	a	a	34
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1204	c	a	b	a	a	b	b	52
1205	c	c	b	b	a	a	a	19

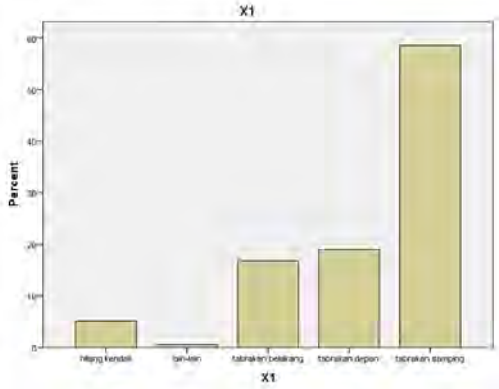
LAMPIRAN B

Hasil Deskriptif Karakteristik Korban Kecelakaan Lalu Lintas Surabaya

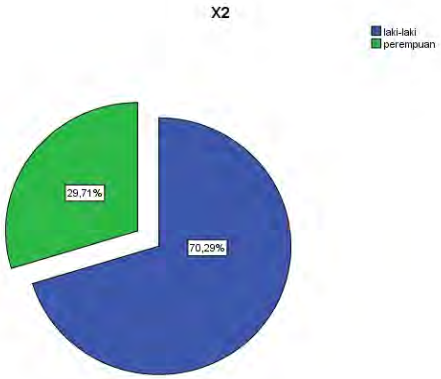
***Pie Chart* Jenis Keparahan Korban Kecelakaan Lalu Lintas di Surabaya**



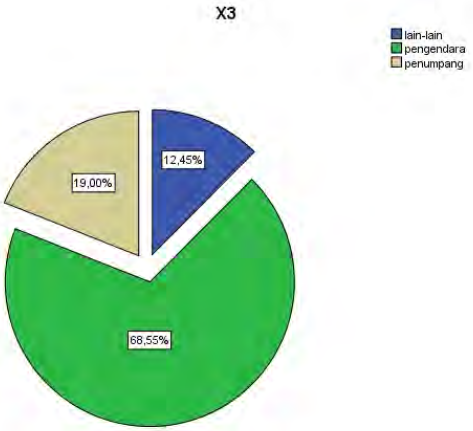
***Pie Chart* Jenis Kecelakaan Korban Kecelakaan Lalu Lintas di Surabaya**



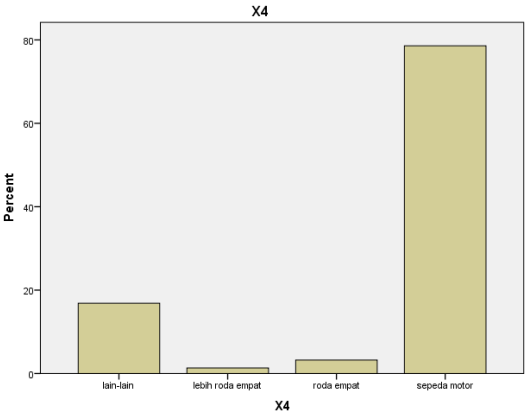
***Pie Chart* Jenis Kelamin Korban Kecelakaan Lalu Lintas di Surabaya**



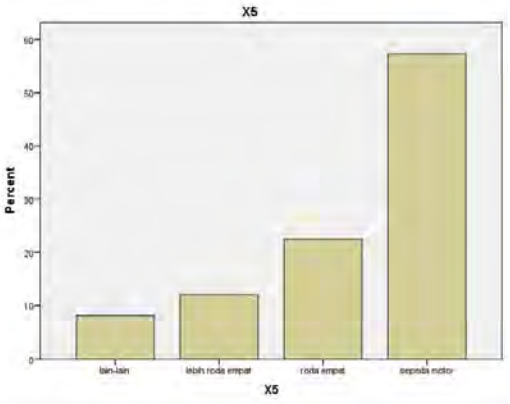
***Pie Chart* Peran Korban Kecelakaan Lalu Lintas di Surabaya**



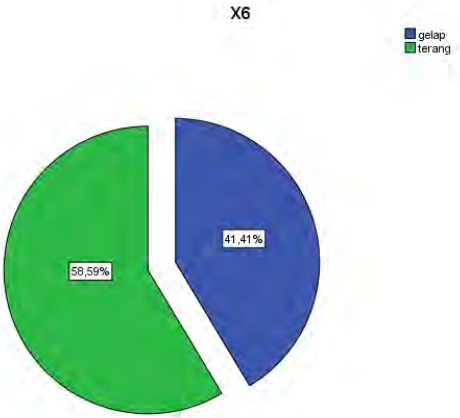
***Pie Chart* Jenis Kendaraan Korban Kecelakaan Lalu Lintas di Surabaya**



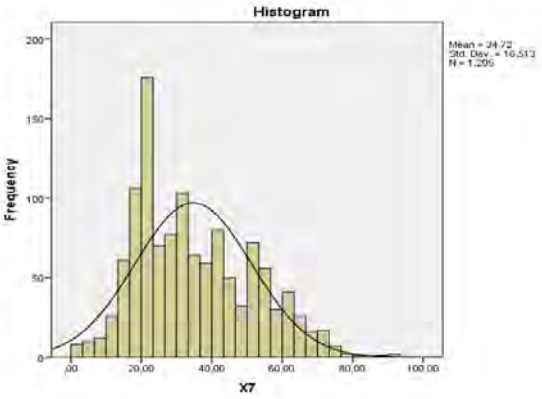
***Pie Chart* Jenis Kendaraan Lawan Kecelakaan Lalu Lintas di Surabaya**



***Pie Chart* Waktu Kejadian Kecelakaan Lalu Lintas di Surabaya**



Histogram Usia Korban Kecelakaan Lalu Lintas di Surabaya



LAMPIRAN C

Output* Regresi Logistik Multinomial dari Percobaan Kombinasi Proporsi Data *Training* dan Data *Testing

***Output* Regresi Logistik Multinomial untuk Proporsi Data *Training* 75% dan Data *Testing* 25%**

Case Processing Summary

		N	Marginal Percentage
y	,00	158	17,5%
	1,00	228	25,2%
	2,00	518	57,3%
x1	,00	148	16,4%
	1,00	174	19,2%
	2,00	525	58,1%
	3,00	50	5,5%
	4,00	7	0,8%
x2	,00	632	69,9%
	1,00	272	30,1%
x3	,00	619	68,5%
	1,00	168	18,6%
	2,00	117	12,9%
x4	,00	708	78,3%
	1,00	28	3,1%
	2,00	14	1,5%
	3,00	154	17,0%
x5	,00	520	57,5%
	1,00	206	22,8%
	2,00	105	11,6%
	3,00	73	8,1%
x6	,00	363	40,2%
	1,00	541	59,8%
Valid		904	100,0%
Missing		0	
Total		904	
Subpopulation		726[a]	

a. The dependent variable has only one value observed in 646 (89,0%) subpopulations.

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1617,528			
Final	1502,715	114,813	30	,000

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	1470,550	1420	,171
Deviance	1382,164	1420	,759

Pseudo R-Square

Cox and Snell	,119
Nagelkerke	,139
McFadden	,065

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	1502,715[a]	,000	0	.
x7	1514,599	11,884	2	,003
x1	1523,141	20,426	8	,009
x2	1503,561	,846	2	,655
x3	1508,052	5,338	4	,254
x4	1507,416	4,701	6	,583
x5	1533,796	31,081	6	,000
x6	1511,240	8,525	2	,014

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Parameter Estimates

y[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
	Intercept	-,843	1,203	,491	1	,484			
	x7	,022	,006	11,705	1	,001	1,022	1,009	1,035
	[x1=,00]	-,494	1,112	,197	1	,657	,610	,069	5,394
	[x1=1,00]	-,519	1,112	,218	1	,641	,595	,067	5,264
	[x1=2,00]	-,663	1,091	,370	1	,543	,515	,061	4,374
	[x1=3,00]	,742	1,163	,408	1	,523	2,101	,215	20,522
	[x1=4,00]	0[b]	.	.	0	.	.	.	.
	[x2=,00]	,190	,233	,667	1	,414	1,210	,766	1,911
	[x2=1,00]	0[b]	.	.	0	.	.	.	.
	[x3=,00]	-,325	,535	,368	1	,544	,723	,253	2,064
,00	[x3=1,00]	-,645	,584	1,218	1	,270	,525	,167	1,649
	[x3=2,00]	0[b]	.	.	0	.	.	.	.
	[x4=,00]	-,220	,488	,204	1	,652	,802	,308	2,088
	[x4=1,00]	-,507	,692	,537	1	,464	,602	,155	2,338
	[x4=2,00]	-,494	,868	,323	1	,570	,610	,111	3,347
	[x4=3,00]	0[b]	.	.	0	.	.	.	.
	[x5=,00]	-,826	,455	3,301	1	,069	,438	,179	1,067
	[x5=1,00]	-,706	,460	2,356	1	,125	,494	,201	1,216
	[x5=2,00]	,748	,482	2,408	1	,121	2,113	,821	5,432
	[x5=3,00]	0[b]	.	.	0	.	.	.	.
	[x6=,00]	,538	,198	7,431	1	,006	1,713	1,163	2,524

1,00	[x6=1,00]	0[b]	.	.	0	.	.	.	.
	Intercept	,977	1,073	,828	1	,363			
	x7	,004	,005	,668	1	,414	1,004	,994	1,015
	[x1=,00]	-,945	,992	,907	1	,341	,389	,056	2,716
	[x1=1,00]	-,507	,987	,264	1	,607	,602	,087	4,166
	[x1=2,00]	-1,002	,976	1,053	1	,305	,367	,054	2,488
	[x1=3,00]	-1,540	1,124	1,876	1	,171	,214	,024	1,942
	[x1=4,00]	0[b]	.	.	0	.	.	.	.
	[x2=,00]	-,035	,186	,035	1	,852	,966	,671	1,391
	[x2=1,00]	0[b]	.	.	0	.	.	.	.
	[x3=,00]	-,125	,435	,082	1	,774	,883	,376	2,071
	[x3=1,00]	-,627	,485	1,673	1	,196	,534	,207	1,381
	[x3=2,00]	0[b]	.	.	0	.	.	.	.
	[x4=,00]	-,518	,390	1,761	1	,185	,596	,277	1,280
	[x4=1,00]	-1,424	,736	3,741	1	,053	,241	,057	1,019
	[x4=2,00]	-,329	,821	,160	1	,689	,720	,144	3,600
	[x4=3,00]	0[b]	.	.	0	.	.	.	.
	[x5=,00]	-,472	,419	1,271	1	,259	,623	,274	1,417
	[x5=1,00]	-,444	,431	1,059	1	,304	,642	,276	1,494
	[x5=2,00]	,074	,473	,024	1	,876	1,077	,426	2,722
	[x5=3,00]	0[b]	.	.	0	.	.	.	.
	[x6=,00]	-,044	,170	,068	1	,794	,957	,685	1,335
	[x6=1,00]	0[b]	.	.	0	.	.	.	.

a. The reference category is: 2,00.
b. This parameter is set to zero because it is redundant.

Classification

Observed	Predicted			
	,00	1,00	2,00	Percent Correct
,00	34	3	121	21,5%
1,00	12	11	205	4,8%
2,00	20	17	481	92,9%
Overall Percentage	7,3%	3,4%	89,3%	58,2%

***Output* Regresi Logistik Multinomial untuk Proporsi Data *Training* 80% dan Data *Testing* 20%**

Case Processing Summary

		N	Marginal Percentage
y	,00	171	17,7%
	1,00	251	26,0%
	2,00	542	56,2%
x1	,00	157	16,3%
	1,00	191	19,8%
	2,00	562	58,3%
	3,00	50	5,2%
	4,00	4	0,4%
x2	,00	682	70,7%
	1,00	282	29,3%
x3	,00	669	69,4%
	1,00	186	19,3%
	2,00	109	11,3%
x4	,00	772	80,1%
	1,00	30	3,1%
	2,00	13	1,3%
	3,00	149	15,5%
x5	,00	546	56,6%
	1,00	221	22,9%
	2,00	116	12,0%
	3,00	81	8,4%
x6	,00	412	42,7%
	1,00	552	57,3%
Valid		964	100,0%
Missing		0	
Total		964	
Subpopulation		763[a]	

a. The dependent variable has only one value observed in 671 (87,9%) subpopulations.

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1726,379			
Final	1602,108	124,271	30	,000

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	1543,823	1494	,180
Deviance	1462,569	1494	,715

Pseudo R-Square

Cox and Snell	,121
Nagelkerke	,141
McFadden	,066

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	1602,108[a]	,000	0	.
x7	1610,208	8,100	2	,017
x1	1626,636	24,528	8	,002
x2	1602,992	,884	2	,643
x3	1614,934	12,826	4	,012
x4	1609,008	6,900	6	,330
x5	1640,358	38,250	6	,000
x6	1608,925	6,817	2	,033

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Parameter Estimates

y[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
	Intercept	-1,494	1,602	,870	1	,351			
	x7	,017	,006	7,902	1	,005	1,017	1,005	1,030
	[x1=,00]	,435	1,512	,083	1	,774	1,545	,080	29,890
	[x1=1,00]	,072	1,510	,002	1	,962	1,074	,056	20,712
	[x1=2,00]	,011	1,495	,000	1	,994	1,011	,054	18,922
	[x1=3,00]	2,040	1,580	1,667	1	,197	7,690	,348	170,147
	[x1=4,00]	0[b]	.	.	0	.	.	.	.
	[x2=,00]	,217	,233	,864	1	,353	1,242	,786	1,962
	[x2=1,00]	0[b]	.	.	0	.	.	.	.
	[x3=,00]	-,639	,501	1,630	1	,202	,528	,198	1,408
,00	[x3=1,00]	-1,010	,546	3,426	1	,064	,364	,125	1,061
	[x3=2,00]	0[b]	.	.	0	.	.	.	.
	[x4=,00]	-,199	,443	,200	1	,654	,820	,344	1,955
	[x4=1,00]	-,774	,674	1,319	1	,251	,461	,123	1,728
	[x4=2,00]	-1,239	,821	2,274	1	,132	,290	,058	1,450
	[x4=3,00]	0[b]	.	.	0	.	.	.	.
	[x5=,00]	-,453	,464	,955	1	,328	,635	,256	1,578
	[x5=1,00]	-,139	,474	,086	1	,770	,870	,344	2,203
	[x5=2,00]	1,301	,495	6,891	1	,009	3,672	1,390	9,696
	[x5=3,00]	0[b]	.	.	0	.	.	.	.
	[x6=,00]	,475	,194	5,998	1	,014	1,607	1,099	2,350

1,00	[x6=1,00]	0[b]	.	.	0	.	.	.	.
	Intercept	,550	1,339	,169	1	,681			
	x7	,003	,005	,268	1	,605	1,003	,993	1,013
	[x1=,00]	-,451	1,265	,127	1	,722	,637	,053	7,611
	[x1=1,00]	-,436	1,260	,120	1	,729	,647	,055	7,637
	[x1=2,00]	-,761	1,251	,371	1	,543	,467	,040	5,420
	[x1=3,00]	-,622	1,385	,202	1	,653	,537	,036	8,105
	[x1=4,00]	0[b]	.	.	0	.	.	.	.
	[x2=,00]	,063	,183	,118	1	,731	1,065	,744	1,523
	[x2=1,00]	0[b]	.	.	0	.	.	.	.
	[x3=,00]	-1,121	,466	5,798	1	,016	,326	,131	,812
	[x3=1,00]	-1,597	,502	10,117	1	,001	,202	,076	,542
	[x3=2,00]	0[b]	.	.	0	.	.	.	.
	[x4=,00]	,338	,424	,637	1	,425	1,403	,611	3,220
	[x4=1,00]	-,258	,649	,158	1	,691	,773	,217	2,757
	[x4=2,00]	-1,306	1,165	1,256	1	,262	,271	,028	2,657
	[x4=3,00]	0[b]	.	.	0	.	.	.	.
	[x5=,00]	-,112	,397	,080	1	,777	,894	,411	1,946
	[x5=1,00]	,029	,410	,005	1	,944	1,029	,461	2,300
	[x5=2,00]	,623	,453	1,892	1	,169	1,865	,767	4,533
	[x5=3,00]	0[b]	.	.	0	.	.	.	.
	[x6=,00]	-,025	,162	,024	1	,876	,975	,710	1,340
	[x6=1,00]	0[b]	.	.	0	.	.	.	.

a. The reference category is: 2,00.

b. This parameter is set to zero because it is redundant.

Classification

Observed	Predicted			
	,00	1,00	2,00	Percent Correct
,00	44	5	122	25,7%
1,00	19	10	222	4,0%
2,00	23	16	503	92,8%
Overall Percentage	8,9%	3,2%	87,9%	57,8%

***Output* Regresi Logistik Multinomial untuk Proporsi Data *Training* 85% dan Data *Testing* 15%**

Case Processing Summary

		N	Marginal Percentage
y	,00	177	17,3%
	1,00	264	25,8%
	2,00	583	56,9%
x1	,00	181	17,7%
	1,00	186	18,2%
	2,00	604	59,0%
	3,00	46	4,5%
	4,00	7	0,7%
x2	,00	726	70,9%
	1,00	298	29,1%
x3	,00	709	69,2%
	1,00	188	18,4%
	2,00	127	12,4%
x4	,00	811	79,2%
	1,00	30	2,9%
	2,00	12	1,2%
	3,00	171	16,7%
x5	,00	580	56,6%
	1,00	235	22,9%
	2,00	132	12,9%
	3,00	77	7,5%
x6	,00	416	40,6%
	1,00	608	59,4%
Valid		1024	100,0%
Missing		0	
Total		1024	
Subpopulation		821[a]	

a. The dependent variable has only one value observed in 737 (89,8%) subpopulations.

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1833,413			
Final	1684,956	148,457	30	,000

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	1712,181	1610	,038
Deviance	1554,149	1610	,837

Pseudo R-Square

Cox and Snell	,135
Nagelkerke	,157
McFadden	,074

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	1684,956[a]	,000	0	.
x7	1697,587	12,631	2	,002
x1	1708,437	23,480	8	,003
x2	1686,544	1,587	2	,452
x3	1697,489	12,533	4	,014
x4	1694,510	9,554	6	,145
x5	1732,660	47,704	6	,000
x6	1696,341	11,385	2	,003

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Parameter Estimates

y[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
	Intercept	-,926	1,267	,535	1	,465			
	x7	,020	,006	10,893	1	,001	1,020	1,008	1,032
	[x1=,00]	-,458	1,167	,154	1	,695	,632	,064	6,224
	[x1=1,00]	-,498	1,168	,182	1	,670	,608	,062	5,998
	[x1=2,00]	-,843	1,153	,534	1	,465	,430	,045	4,126
	[x1=3,00]	1,049	1,234	,723	1	,395	2,855	,254	32,068
	[x1=4,00]	0[b]	.	.	0	.	.	.	.
	[x2=,00]	,203	,223	,831	1	,362	1,225	,792	1,895
	[x2=1,00]	0[b]	.	.	0	.	.	.	.
	[x3=,00]	-,726	,531	1,871	1	,171	,484	,171	1,369
,00	[x3=1,00]	-,893	,569	2,464	1	,116	,409	,134	1,249
	[x3=2,00]	0[b]	.	.	0	.	.	.	.
	[x4=,00]	-,010	,482	,000	1	,984	,991	,385	2,546
	[x4=1,00]	-1,431	,761	3,533	1	,060	,239	,054	1,063
	[x4=2,00]	-,921	,912	1,020	1	,313	,398	,067	2,379
	[x4=3,00]	0[b]	.	.	0	.	.	.	.
	[x5=,00]	-,536	,445	1,448	1	,229	,585	,245	1,401
	[x5=1,00]	-,157	,450	,122	1	,727	,855	,354	2,065
	[x5=2,00]	1,315	,471	7,797	1	,005	3,724	1,480	9,372
	[x5=3,00]	0[b]	.	.	0	.	.	.	.
	[x6=,00]	,619	,190	10,673	1	,001	1,857	1,281	2,693

1,00	[x6=1,00]	0[b]	.	.	0	.	.	.	.
	Intercept	,711	1,095	,421	1	,516			
	x7	-,001	,005	,056	1	,812	,999	,989	1,009
	[x1=,00]	-,746	1,020	,535	1	,464	,474	,064	3,502
	[x1=1,00]	-,572	1,018	,316	1	,574	,564	,077	4,152
	[x1=2,00]	-,944	1,008	,877	1	,349	,389	,054	2,807
	[x1=3,00]	-,995	1,154	,744	1	,388	,370	,038	3,548
	[x1=4,00]	0[b]	.	.	0	.	.	.	.
	[x2=,00]	,192	,179	1,152	1	,283	1,212	,853	1,722
	[x2=1,00]	0[b]	.	.	0	.	.	.	.
	[x3=,00]	-,516	,407	1,603	1	,205	,597	,269	1,327
	[x3=1,00]	-1,223	,458	7,124	1	,008	,294	,120	,723
	[x3=2,00]	0[b]	.	.	0	.	.	.	.
	[x4=,00]	-,466	,364	1,638	1	,201	,628	,308	1,281
	[x4=1,00]	-,991	,628	2,492	1	,114	,371	,108	1,270
	[x4=2,00]	-,821	,912	,811	1	,368	,440	,074	2,629
	[x4=3,00]	0[b]	.	.	0	.	.	.	.
	[x5=,00]	,109	,400	,074	1	,786	1,115	,509	2,444
	[x5=1,00]	,136	,412	,109	1	,741	1,146	,511	2,572
	[x5=2,00]	,768	,448	2,943	1	,086	2,156	,896	5,186
	[x5=3,00]	0[b]	.	.	0	.	.	.	.
	[x6=,00]	,022	,160	,020	1	,889	1,023	,748	1,398
	[x6=1,00]	0[b]	.	.	0	.	.	.	.

a. The reference category is: 2,00.

b. This parameter is set to zero because it is redundant.

Classification

Observed	Predicted			
	,00	1,00	2,00	Percent Correct
,00	47	11	119	26,6%
1,00	19	27	218	10,2%
2,00	23	25	535	91,8%
Overall Percentage	8,7%	6,2%	85,2%	59,5%

***Output* Regresi Logistik Multinomial untuk Proporsi Data *Training* 90% dan Data *Testing* 10%**

Case Processing Summary

		N	Marginal Percentage
Y	,00	177	16,3%
	1,00	289	26,7%
	2,00	618	57,0%
X1	,00	176	16,2%
	1,00	213	19,6%
	2,00	633	58,4%
	3,00	55	5,1%
	4,00	7	0,6%
X2	,00	757	69,8%
	1,00	327	30,2%
X3	,00	740	68,3%
	1,00	208	19,2%
	2,00	136	12,5%
X4	,00	846	78,0%
	1,00	37	3,4%
	2,00	15	1,4%
	3,00	186	17,2%
X5	,00	623	57,5%
	1,00	249	23,0%
	2,00	128	11,8%
	3,00	84	7,7%
X6	,00	452	41,7%
	1,00	632	58,3%
Valid		1084	100,0%
Missing		0	
Total		1084	
Subpopulation		850[a]	

a. The dependent variable has only one value observed in 744 (87,5%) subpopulations.

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1906,010			
Final	1775,399	130,611	30	,000

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	1726,699	1668	,155
Deviance	1613,084	1668	,829

Pseudo R-Square

Cox and Snell	,114
Nagelkerke	,133
McFadden	,062

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	1775,399[a]	,000	0	.
X7	1787,949	12,550	2	,002
X1	1792,588	17,189	8	,028
X2	1775,482	,083	2	,959
X3	1786,728	11,330	4	,023
X4	1781,163	5,764	6	,450
X5	1820,319	44,920	6	,000
X6	1783,456	8,058	2	,018

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Parameter Estimates

Y[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
	Intercept	-,598	1,215	,243	1	,622			
	X7	,021	,006	12,336	1	,000	1,021	1,009	1,033
	[X1=,00]	-,513	1,131	,206	1	,650	,599	,065	5,498
	[X1=1,00]	-,486	1,130	,185	1	,667	,615	,067	5,633
	[X1=2,00]	-,789	1,115	,500	1	,479	,454	,051	4,040
	[X1=3,00]	,684	1,181	,335	1	,563	1,982	,196	20,077
	[X1=4,00]	0[b]	.	.	0	.	.	.	.
	[X2=,00]	,010	,217	,002	1	,964	1,010	,660	1,545
	[X2=1,00]	0[b]	.	.	0	.	.	.	.
	[X3=,00]	-,732	,492	2,215	1	,137	,481	,183	1,261
,00	[X3=1,00]	-1,001	,532	3,540	1	,060	,368	,130	1,043
	[X3=2,00]	0[b]	.	.	0	.	.	.	.
	[X4=,00]	-,018	,445	,002	1	,969	,983	,411	2,351
	[X4=1,00]	-,657	,638	1,061	1	,303	,518	,148	1,811
	[X4=2,00]	-,924	,868	1,133	1	,287	,397	,072	2,175
	[X4=3,00]	0[b]	.	.	0	.	.	.	.
	[X5=,00]	-,766	,434	3,120	1	,077	,465	,199	1,088
	[X5=1,00]	-,366	,436	,707	1	,400	,693	,295	1,628
	[X5=2,00]	1,061	,458	5,372	1	,020	2,889	1,178	7,086
	[X5=3,00]	0[b]	.	.	0	.	.	.	.
	[X6=,00]	,508	,187	7,402	1	,007	1,662	1,153	2,397

1,00	[X6=1,00]	0[b]	.	.	0	.	.	.	.	
	Intercept	,848	1,067	,631	1	,427				
	X7	,004	,005	,657	1	,418	1,004	,994	1,013	
	[X1=,00]	-,859	1,002	,735	1	,391	,424	,059	3,020	
	[X1=1,00]	-,640	,999	,410	1	,522	,527	,074	3,735	
	[X1=2,00]	-,931	,989	,885	1	,347	,394	,057	2,742	
	[X1=3,00]	-1,165	1,101	1,120	1	,290	,312	,036	2,700	
	[X1=4,00]	0[b]	.	.	0	.	.	.	.	
	[X2=,00]	,049	,170	,082	1	,775	1,050	,752	1,465	
	[X2=1,00]	0[b]	.	.	0	.	.	.	.	
	[X3=,00]	-,623	,381	2,679	1	,102	,536	,254	1,131	
	[X3=1,00]	-1,170	,424	7,628	1	,006	,310	,135	,712	
	[X3=2,00]	0[b]	.	.	0	.	.	.	.	
	[X4=,00]	-,241	,340	,504	1	,478	,786	,404	1,529	
	[X4=1,00]	-1,047	,595	3,099	1	,078	,351	,109	1,126	
	[X4=2,00]	-,510	,778	,430	1	,512	,601	,131	2,758	
	[X4=3,00]	0[b]	.	.	0	.	.	.	.	
	[X5=,00]	-,165	,381	,188	1	,664	,848	,402	1,789	
	[X5=1,00]	-,045	,391	,014	1	,907	,956	,444	2,056	
	[X5=2,00]	,503	,430	1,370	1	,242	1,654	,712	3,839	
	[X5=3,00]	0[b]	.	.	0	.	.	.	.	
	[X6=,00]	-,001	,152	,000	1	,996	,999	,741	1,347	
	[X6=1,00]	0[b]	.	.	0	.	.	.	.	

a. The reference category is: 2,00.

b. This parameter is set to zero because it is redundant.

Classification

Observed	Predicted			
	,00	1,00	2,00	Percent Correct
,00	37	15	125	20,9%
1,00	15	23	251	8,0%
2,00	24	27	567	91,7%
Overall Percentage	7,0%	6,0%	87,0%	57,8%

***Output* Regresi Logistik Multinomial untuk Proporsi Data *Training* 95% dan Data *Testing* 5%**

Case Processing Summary

		N	Marginal Percentage
Y	,00	196	17,1%
	1,00	301	26,3%
	2,00	648	56,6%
X1	,00	193	16,9%
	1,00	214	18,7%
	2,00	671	58,6%
	3,00	60	5,2%
	4,00	7	0,6%
X2	,00	808	70,6%
	1,00	337	29,4%
X3	,00	784	68,5%
	1,00	218	19,0%
	2,00	143	12,5%
X4	,00	895	78,2%
	1,00	39	3,4%
	2,00	16	1,4%
	3,00	195	17,0%
X5	,00	656	57,3%
	1,00	256	22,4%
	2,00	138	12,1%
	3,00	95	8,3%
X6	,00	475	41,5%
	1,00	670	58,5%
Valid		1145	100,0%
Missing		0	
Total		1145	
Subpopulation		886[a]	

a. The dependent variable has only one value observed in 775 (87,5%) subpopulations.

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	2019,339			
Final	1871,817	147,522	30	,000

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	1816,220	1740	,099
Deviance	1698,238	1740	,759

Pseudo R-Square

Cox and Snell	,121
Nagelkerke	,141
McFadden	,066

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	1871,817[a]	,000	0	.
X7	1885,504	13,686	2	,001
X1	1896,399	24,581	8	,002
X2	1872,194	,377	2	,828
X3	1887,980	16,163	4	,003
X4	1876,168	4,350	6	,629
X5	1913,697	41,879	6	,000
X6	1881,302	9,484	2	,009

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Parameter Estimates

Y[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
	Intercept	-,851	1,214	,492	1	,483			
	X7	,020	,006	13,077	1	,000	1,021	1,009	1,032
	[X1=,00]	-,562	1,134	,246	1	,620	,570	,062	5,260
	[X1=1,00]	-,623	1,134	,301	1	,583	,537	,058	4,953
	[X1=2,00]	-,818	1,120	,534	1	,465	,441	,049	3,959
	[X1=3,00]	1,026	1,185	,749	1	,387	2,789	,274	28,435
	[X1=4,00]	0[b]	.	.	0	.	.	.	.
	[X2=,00]	,128	,213	,363	1	,547	1,137	,749	1,725
	[X2=1,00]	0[b]	.	.	0	.	.	.	.
	[X3=,00]	-,651	,472	1,904	1	,168	,521	,207	1,315
,00	[X3=1,00]	-1,066	,514	4,290	1	,038	,345	,126	,944
	[X3=2,00]	0[b]	.	.	0	.	.	.	.
	[X4=,00]	,011	,425	,001	1	,979	1,011	,440	2,326
	[X4=1,00]	-,713	,626	1,295	1	,255	,490	,144	1,673
	[X4=2,00]	-,677	,807	,705	1	,401	,508	,105	2,469
	[X4=3,00]	0[b]	.	.	0	.	.	.	.
	[X5=,00]	-,536	,410	1,705	1	,192	,585	,262	1,308
	[X5=1,00]	-,244	,416	,344	1	,558	,783	,346	1,772
	[X5=2,00]	1,155	,437	6,984	1	,008	3,173	1,348	7,470
	[X5=3,00]	0[b]	.	.	0	.	.	.	.
	[X6=,00]	,543	,178	9,263	1	,002	1,721	1,213	2,442

1,00	[X6=1,00]	0[b]	.	.	0	.	.	.	.
	Intercept	,821	1,062	,598	1	,439			
	X7	,002	,005	,196	1	,658	1,002	,993	1,011
	[X1=,00]	-,946	1,002	,890	1	,345	,388	,054	2,770
	[X1=1,00]	-,772	,999	,596	1	,440	,462	,065	3,279
	[X1=2,00]	-1,030	,991	1,080	1	,299	,357	,051	2,490
	[X1=3,00]	-1,106	1,103	1,006	1	,316	,331	,038	2,874
	[X1=4,00]	0[b]	.	.	0	.	.	.	.
	[X2=,00]	,045	,167	,072	1	,789	1,046	,753	1,452
	[X2=1,00]	0[b]	.	.	0	.	.	.	.
	[X3=,00]	-,699	,380	3,378	1	,066	,497	,236	1,047
	[X3=1,00]	-1,361	,423	10,364	1	,001	,256	,112	,587
	[X3=2,00]	0[b]	.	.	0	.	.	.	.
	[X4=,00]	-,115	,342	,113	1	,737	,892	,456	1,742
	[X4=1,00]	-,727	,566	1,652	1	,199	,483	,160	1,465
	[X4=2,00]	-,309	,774	,160	1	,689	,734	,161	3,348
	[X4=3,00]	0[b]	.	.	0	.	.	.	.
	[X5=,00]	,002	,364	,000	1	,996	1,002	,491	2,043
	[X5=1,00]	,077	,375	,042	1	,837	1,080	,518	2,252
	[X5=2,00]	,559	,412	1,836	1	,175	1,749	,779	3,924
	[X5=3,00]	0[b]	.	.	0	.	.	.	.
	[X6=,00]	,062	,148	,177	1	,674	1,064	,796	1,422
	[X6=1,00]	0[b]	.	.	0	.	.	.	.

a. The reference category is: 2,00.

b. This parameter is set to zero because it is redundant.

Classification

Observed	Predicted			
	,00	1,00	2,00	Percent Correct
,00	43	12	141	21,9%
1,00	16	17	268	5,6%
2,00	24	22	602	92,9%
Overall Percentage	7,2%	4,5%	88,3%	57,8%

LAMPIRAN D
Ouput CART dari Kombinasi Data *Training* 95% dan Data *Testing* 5%

Output Statistika Deskriptif Variabel Penelitian

```
LEARNING AND TEST (*) SAMPLE VARIABLE STATISTICS
 ================================================

                          CLASS
 VARIABLE                     0           1           2
OVERALL
 ------------------------------------------------------------
----------
        Y     MEAN|       0.000       1.000       2.000
1.395
                SD|       0.000       0.000       0.000
0.763
                 N|     196.000     301.000     648.000
1145.000
               SUM|       0.000     301.000    1296.000
1597.000
             * MEAN|       0.000       1.000       2.000
1.450
              * SD|       0.000       0.000       0.000
0.769
               * N|      10.000      13.000      37.000
60.000
             * SUM|       0.000      13.000      74.000
87.000
       X1     MEAN|       1.694       1.508       1.509
1.541
                SD|       0.970       0.823       0.825
0.853
                 N|     196.000     301.000     648.000
1145.000
               SUM|     332.000     454.000     978.000
1764.000
             * MEAN|       1.300       1.231       1.622
1.483
              * SD|       1.059       0.832       0.681
0.792
               * N|      10.000      13.000      37.000
60.000
             * SUM|      13.000      16.000      60.000
89.000
```

```
       X2     MEAN|       0.240        0.292        0.312
0.294
              SD|       0.428        0.456        0.464
0.456
               N|     196.000      301.000      648.000
1145.000
             SUM|      47.000       88.000      202.000
337.000
            * MEAN|       0.300        0.462        0.324
0.350
             * SD|       0.483        0.519        0.475
0.481
              * N|      10.000       13.000       37.000
60.000
             * SUM|       3.000        6.000       12.000
21.000
       X3     MEAN|       0.469        0.495        0.406
0.440
              SD|       0.740        0.794        0.646
0.705
               N|     196.000      301.000      648.000
1145.000
             SUM|      92.000      149.000      263.000
504.000
            * MEAN|       0.500        0.538        0.351
0.417
             * SD|       0.707        0.776        0.676
0.696
              * N|      10.000       13.000       37.000
60.000
             * SUM|       5.000        7.000       13.000
25.000
       X4     MEAN|       0.658        0.757        0.461
0.573
              SD|       1.198        1.282        1.032
1.138
               N|     196.000      301.000      648.000
1145.000
             SUM|     129.000      228.000      299.000
656.000
            * MEAN|       0.300        0.692        0.324
0.400
             * SD|       0.949        1.316        0.944
1.028
              * N|      10.000       13.000       37.000
60.000
             * SUM|       3.000        9.000       12.000
24.000
       X5     MEAN|       1.163        0.611        0.625
0.714
```

```
                SD|       1.148        0.890        0.912
0.972
                 N|     196.000      301.000      648.000
1145.000
               SUM|     228.000      184.000      405.000
817.000
            * MEAN|       1.400        0.615        0.514
0.683
              * SD|       1.174        0.768        0.837
0.930
               * N|      10.000       13.000       37.000
60.000
             * SUM|      14.000        8.000       19.000
41.000
       X6     MEAN|       0.480        0.608        0.606
0.585
                SD|       0.501        0.489        0.489
0.493
                 N|     196.000      301.000      648.000
1145.000
               SUM|      94.000      183.000      393.000
670.000
            * MEAN|       0.600        0.769        0.541
0.600
              * SD|       0.516        0.439        0.505
0.494
               * N|      10.000       13.000       37.000
60.000
             * SUM|       6.000       10.000       20.000
36.000
       X7     MEAN|      38.189       35.546       33.432
34.802
                SD|      18.174       17.180       15.678
16.610
                 N|     196.000      301.000      648.000
1145.000
               SUM|    7485.000    10699.450    21664.000
39848.450
            * MEAN|      35.400       30.462       33.595
33.217
              * SD|      12.765       15.273       15.032
14.586
               * N|      10.000       13.000       37.000
60.000
             * SUM|     354.000      396.000     1243.000
1993.000
 ------------------------------------------------------------
----------
```

Output Tree Sequance

TREE SEQUENCE
=============

Dependent variable: Y

Terminal Tree	Nodes	Test Set Relative Cost	Resubstitution Relative Cost	Complexity Parameter
1	172	0.941 +/- 0.105	0.416	0.000
46**	29	0.722 +/- 0.108	0.657	0.003
47	20	0.801 +/- 0.091	0.697	0.003
48	18	0.751 +/- 0.102	0.707	0.003
49	17	0.751 +/- 0.102	0.712	0.003
50	10	0.876 +/- 0.087	0.749	0.004
51	8	0.876 +/- 0.087	0.760	0.004
52	5	0.839 +/- 0.083	0.788	0.006
53	4	0.826 +/- 0.082	0.800	0.008
54	3	0.799 +/- 0.082	0.831	0.021
55	2	0.804 +/- 0.083	0.878	0.031
56	1	1.000 +/- 0.000	1.000	0.081

Output Ketepatan Akurasi Data *Training* dan *Testing*

LEARNING SAMPLE CLASSIFICATION TABLE
====================================

Actual Class	Predicted Class 0	1	2	Actual Total
0	133.00	13.00	50.00	196.00
1	71.00	97.00	133.00	301.00
2	126.00	78.00	444.00	648.00
PRED. TOT.	330.00	188.00	627.00	1145.00
CORRECT	0.679	0.322	0.685	
SUCCESS IND.	0.507	0.059	0.119	
TOT. CORRECT	0.589			

==========
LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE
==========

Actual Class	Predicted Class 0	1	2	Actual Total
0	0.679	0.066	0.255	1.000
1	0.236	0.322	0.442	1.000
2	0.194	0.120	0.685	1.000

```
TEST SAMPLE CLASSIFICATION TABLE
 ================================

 Actual        Predicted Class                                     Actual
 Class                       0            1            2            Total
 -----------------------------------------------------------------
 0                        6.00         1.00         3.00          10.00
 1                        2.00         4.00         7.00          13.00
 2                        8.00         5.00        24.00          37.00
 -----------------------------------------------------------------
 PRED. TOT.              16.00        10.00        34.00          60.00
 CORRECT                 0.600        0.308        0.649
 SUCCESS IND.            0.433        0.091        0.032
 TOT. CORRECT            0.567
 -----------------------------------------------------------------


 ============================================
 TEST SAMPLE CLASSIFICATION PROBABILITY TABLE
 ============================================

 Actual        Predicted Class                                     Actual
 Class                       0            1            2            Total
 -----------------------------------------------------------------
 0                       0.600        0.100        0.300          1.000
 1                       0.154        0.308        0.538          1.000
 2                       0.216        0.135        0.649          1.000
 -----------------------------------------------------------------
```

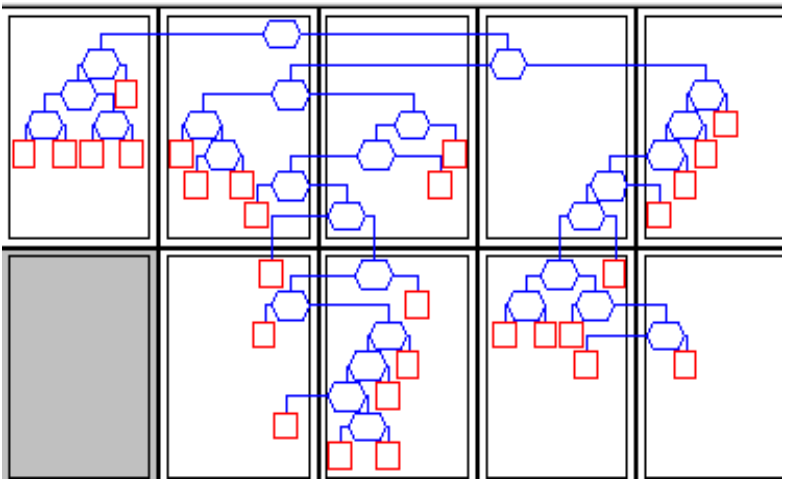
Output Variable Importance Pohon Klasifikasi Maksimal

Variable	Score	
X7	100.00	||||||||||||||||||||||||||||||||||||||||||
X1	43.02	||||||||||||||||||
X5	31.72	||||||||||||||
X4	26.38	|||||||||||
X3	26.09	|||||||||||
X2	17.85	||||||||
X6	11.82	|||||

Output Variable Importance Pohon Klasifikasi Optimal

Variable	Score	
X7	100.00	|||||||||||||||||||||||||||||||||||||||||
X1	91.13	||||||||||||||||||||||||||||||||||||||
X5	84.33	|||||||||||||||||||||||||||||||||||
X4	62.99	||||||||||||||||||||||||||
X3	35.44	|||||||||||||||
X6	13.10	||||||
X2	7.89	||

LAMPIRAN E
Konstruksi Pohon Klasifikasi Optimal



Lampiran F
Program R untuk Pembagian Data *Training* dan Data *Testing*

```
datakec<-read.table("D:/datakec.txt",header=TRUE)  #load the data

#Split the data frame
splitDataFrame<-function(dataframe,seed=null,n=trainSize){
 if(!is.null(seed))set.seed(seed)
 index<-1:nrow(dataframe)
 trainindex<-sample(index,n)
 trainset<-dataframe[trainindex,]
 testset<-dataframe[-trainindex,]
 list(trainset=trainset,testset=testset)
 }
# Training Data 75% and Testing Data 25%
 split<-splitDataFrame(datakec,NULL,round(nrow(datakec)*0.75))
 train75<-split$trainset
 test25<-split$testset
write.csv(train75, "D:\\train75.csv")
write.csv(test25, "D:\\test25.csv")

# Training Data 80% and Testing Data 20%
 split<-splitDataFrame(datakec,NULL,round(nrow(datakec)*0.80))
 train80<-split$trainset
 test20<-split$testset
write.csv(train80, "D:\\train80.csv")
write.csv(test20, "D:\\test20.csv")

# Training Data 85% and Testing Data 15%
 split<-splitDataFrame(datakec,NULL,round(nrow(datakec)*0.85))
 train85<-split$trainset
 test15<-split$testset
write.csv(train85, "D:\\train85.csv")
write.csv(test15, "D:\\test15.csv")

# Training Data 90% and Testing Data 10%
 split<-splitDataFrame(datakec,NULL,round(nrow(datakec)*0.90))
 train90<-split$trainset
```

```
 test10<-split$testset
write.csv(train90, "D:\\train90.csv")
write.csv(test10, "D:\\test10.csv")

# Training Data 95% and Testing Data 5%
 split<-splitDataFrame(datakec,NULL,round(nrow(datakec)*0.95))
 train95<-split$trainset
 test5<-split$testset
write.csv(train95, "D:\\train95.csv")
write.csv(test5, "D:\\test5.csv")
```

LAMPIRAN G

Program R untuk *Random Forests* Menggunakan Kombinasi Data *Training* dan Data *Testing* Terpilih (95%:5%)

```
#grow tree for RF (50 combination)
train75$x7<-as.numeric(train75$x7)
test25$x7<-as.numeric(test25$x7)
fit.rf<-randomForest(y~.,data=train95,mtry=3,ntree=50)

#Accuracy Prediction and calculation of Training data set
fitrf.pred1<-predict(fit.rf,train95[-1146,])
write.csv(fitrf.pred1, "D:\\rfdatakectrain95.csv")
crosstabrf1<-table(predicted=fitrf.pred1,observed=train95[-1146,"y"])
APERrf1<-1-(sum(diag(crosstabrf1))/sum(crosstabrf1))
accuracyrf1<-(sum(diag(crosstabrf1))/sum(crosstabrf1))
print(crosstabrf1)
print(APERrf1)
print(accuracyrf1)

#Accuracy prediction and calculate of Testing dataset
fitrf.pred2<-predict(fit.rf,test5[-61,])
write.csv(fitrf.pred2,"D:\\rfdatakectest5.csv")
crosstabrf2<-table(predicted=fitrf.pred2,observed=test5[-61,"y"])
APERrf2<- 1-(sum(diag(crosstabrf2))/sum(crosstabrf2))
accuracyrf2<-(sum(diag(crosstabrf2))/sum(crosstabrf2))
print(crosstabrf2)
print(APERrf2)
print(accuracyrf2)

#grow tree for RF (100 combination)
fit.rf<-randomForest(y~.,data=train95,mtry=3,ntree=100)

#Accuracy Prediction and calculation of Training data set
fitrf.pred1<-predict(fit.rf,train95[-1146,])
write.csv(fitrf.pred1, "D:\\rfdatakectrain95.csv")
crosstabrf1<-table(predicted=fitrf.pred1,observed=train95[-1146,"y"])
APERrf1<-1-(sum(diag(crosstabrf1))/sum(crosstabrf1))
accuracyrf1<-(sum(diag(crosstabrf1))/sum(crosstabrf1))
```

```
print(crosstabrf1)
print(APERrf1)
print(accuracyrf1)

#Accuracy prediction and calculate of Testing dataset
fitrf.pred2<-predict(fit.rf,test5[-61,])
write.csv(fitrf.pred2,"D:\\rfdatakectest5.csv")
crosstabrf2<-table(predicted=fitrf.pred2,observed=test5[-61,"y"])
APERrf2<- 1-(sum(diag(crosstabrf2))/sum(crosstabrf2))
accuracyrf2<-(sum(diag(crosstabrf2))/sum(crosstabrf2))
print(crosstabrf2)
print(APERrf2)
print(accuracyrf2)

#grow tree for RF (250 combination)
fit.rf<-randomForest(y~.,data=train95,mtry=3,ntree=250)

#Accuracy Prediction and calculation of Training data set
fitrf.pred1<-predict(fit.rf,train95[-1146,])
write.csv(fitrf.pred1, "D:\\rfdatakectrain95.csv")
crosstabrf1<-table(predicted=fitrf.pred1,observed=train95[-1146,"y"])
APERrf1<-1-(sum(diag(crosstabrf1))/sum(crosstabrf1))
accuracyrf1<-(sum(diag(crosstabrf1))/sum(crosstabrf1))
print(crosstabrf1)
print(APERrf1)
print(accuracyrf1)

#Accuracy prediction and calculate of Testing dataset
fitrf.pred2<-predict(fit.rf,test5[-61,])
write.csv(fitrf.pred2,"D:\\rfdatakectest5.csv")
crosstabrf2<-table(predicted=fitrf.pred2,observed=test5[-61,"y"])
APERrf2<- 1-(sum(diag(crosstabrf2))/sum(crosstabrf2))
accuracyrf2<-(sum(diag(crosstabrf2))/sum(crosstabrf2))
print(crosstabrf2)
print(APERrf2)
print(accuracyrf2)

#grow tree for RF (500 combination)
fit.rf<-randomForest(y~.,data=train95,mtry=3,ntree=500)
```

```
#Accuracy Prediction and calculation of Training data set
fitrf.pred1<-predict(fit.rf,train95[-1146,])
write.csv(fitrf.pred1, "D:\\rfdatakectrain95.csv")
crosstabrf1<-table(predicted=fitrf.pred1,observed=train95[-1146,"y"])
APERrf1<-1-(sum(diag(crosstabrf1))/sum(crosstabrf1))
accuracyrf1<-(sum(diag(crosstabrf1))/sum(crosstabrf1))
print(crosstabrf1)
print(APERrf1)
print(accuracyrf1)

#Accuracy prediction and calculate of Testing dataset
fitrf.pred2<-predict(fit.rf,test5[-61,])
write.csv(fitrf.pred2,"D:\\rfdatakectest5.csv")
crosstabrf2<-table(predicted=fitrf.pred2,observed=test5[-61,"y"])
APERrf2<- 1-(sum(diag(crosstabrf2))/sum(crosstabrf2))
accuracyrf2<-(sum(diag(crosstabrf2))/sum(crosstabrf2))
print(crosstabrf2)
print(APERrf2)
print(accuracyrf2)

#grow tree for RF (1000 combination)
fit.rf<-randomForest(y~.,data=train95,mtry=3,ntree=1000)

#Accuracy Prediction and calculation of Training data set
fitrf.pred1<-predict(fit.rf,train95[-1146,])
write.csv(fitrf.pred1, "D:\\rfdatakectrain95.csv")
crosstabrf1<-table(predicted=fitrf.pred1,observed=train95[-1146,"y"])
APERrf1<-1-(sum(diag(crosstabrf1))/sum(crosstabrf1))
accuracyrf1<-(sum(diag(crosstabrf1))/sum(crosstabrf1))
print(crosstabrf1)
print(APERrf1)
print(accuracyrf1)

#Accuracy prediction and calculate of Testing dataset
fitrf.pred2<-predict(fit.rf,test5[-61,])
write.csv(fitrf.pred2,"D:\\rfdatakectest5.csv")
crosstabrf2<-table(predicted=fitrf.pred2,observed=test5[-61,"y"])
APERrf2<- 1-(sum(diag(crosstabrf2))/sum(crosstabrf2))
```

```
accuracyrf2<-(sum(diag(crosstabrf2))/sum(crosstabrf2))
print(crosstabrf2)
print(APERrf2)
print(accuracyrf2)
```

LAMPIRAN H
Output* R untuk *Random Forests

***Output* R untuk *Random Forests* pada 50 Kombinasi**

```
> #Accuracy Prediction and calculation of Training data set
> print(crosstabrf1)
          observed
predicted   a   b   c
        a 116  11  10
        b  13 161  14
        c  67 129 624
> print(APERrf1)
[1] 0.2131004
> print(accuracyrf1)
[1] 0.7868996

> #Accuracy prediction and calculate of Testing dataset
> print(crosstabrf2)
          observed
predicted  a  b  c
        a  3  1  1
        b  0  1  5
        c  7 11 31
> print(APERrf2)
[1] 0.4166667
> print(accuracyrf2)
[1] 0.5833333
```

***Output* R untuk *Random Forests* pada 100 Kombinasi**

```
> #Accuracy Prediction and calculation of Training data set
> print(crosstabrf1)
        observed
predicted   a   b   c
       a 124   7   9
       b  15 168  13
       c  57 126 626
> print(APERrf1)
[1] 0.1982533
> print(accuracyrf1)
[1] 0.8017467

> #Accuracy prediction and calculate of Testing dataset
> print(crosstabrf2)
        observed
predicted  a  b  c
       a  3  1  2
       b  0  1  7
       c  7 11 28
> print(APERrf2)
[1] 0.4666667
> print(accuracyrf2)
[1] 0.5333333
```

***Output* R untuk *Random Forests* pada 250 Kombinasi**

```
> #Accuracy Prediction and calculation of Training data set
> print(crosstabrf1)
        observed
predicted   a   b   c
        a 121   9  12
        b  11 159  14
        c  64 133 622
> print(APERrf1)
[1] 0.2122271
> print(accuracyrf1)
[1] 0.7877729

> #Accuracy prediction and calculate of Testing dataset
> print(crosstabrf2)
        observed
predicted  a  b  c
        a  3  1  2
        b  0  1  6
        c  7 11 29
> print(APERrf2)
[1] 0.45
> print(accuracyrf2)
[1] 0.55
```

***Output* R untuk *Random Forests* pada 500 Kombinasi**

```
> #Accuracy Prediction and calculation of Training data set
> print(crosstabrf1)
          observed
predicted   a   b   c
        a 125   8  10
        b  10 156  13
        c  61 137 625
> print(APERrf1)
[1] 0.2087336
> print(accuracyrf1)
[1] 0.7912664

> #Accuracy prediction and calculate of Testing dataset
> print(crosstabrf2)
          observed
predicted  a  b  c
        a  3  1  2
        b  0  1  6
        c  7 11 29
> print(APERrf2)
[1] 0.45
> print(accuracyrf2)
[1] 0.55
```

***Output* R untuk *Random Forests* pada 1000 Kombinasi**

```
> #Accuracy Prediction and calculation of Training data set
> print(crosstabrf1)
        observed
predicted   a   b   c
        a 125   9  11
        b  12 162  12
        c  59 130 625
> print(APERrf1)
[1] 0.2034934
> print(accuracyrf1)
[1] 0.7965066

> #Accuracy prediction and calculate of Testing dataset
> print(crosstabrf2)
        observed
predicted  a  b  c
        a  3  1  2
        b  0  2  6
        c  7 10 29
> print(APERrf2)
[1] 0.4333333
> print(accuracyrf2)
[1] 0.5666667
```

DAFTAR PUSTAKA

Agresti, A. (1990). *Categorical Data Analysis*. John Wiley and Sons. New York.

Breiman, L. (2001). *Random Forests. Machine Learning,* 45, 5-32.

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J.. (1993). *Classification and Regression Trees*. New York : Chapman Hall.

Dewi, N. K., Syafitri, U. D. & Mulyadi, S. Y. (2011). *Penerapan Metode Random Forest dalam Driver Analysis, Forum Statistika dan Komputasi*, Vol. 16, No.1, 0853-811.

Enache, A., Chatzinikolaou, F., Enache, F. & Enache, B. (2009). *The Analysis of Lethal Traffic Accidents and Risk Factors, Legal Medicine* 11.S327-S330.

Fitriah, W. W, Mashuri, M. & Irhamah. (2012). *Faktor-Faktor yang Mempengaruhi Keparahan Korban Kecelakaan Lalu Lintas di Kota Surabaya dengan Pendekatan Bagging Regresi Logistik Ordinal, Jurnal SAINS dan S eni ITS*. Vol.1, No. 1, ISSN:2301-928X.

Geneur, R., Poggi, J., M. & Malot, C., T. (2009). *Variable Selection using Random Forests*. France : Laboratoire de Mathematiques, Universite Paris-Sud 11, B at. 425, 9 1405 Orsay.

Hosmer, D. W., & Lemeshow. 2000. *Applied Logistic Regression*. USA : John Wiley and Sons.

Indriani, D. & Indawati, R. (2006). *Model Hubungan dan estimasi Tingkat Kecelakaan Lalu Lintas, Berita Kedokteran Masyarakat*. Vol. 22, No. 3. Hal 100-106.

Johnson, R. A. & Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*, Edisi ke-7.

Lewis, R. J. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis, Annual Meeting of the Society for Academic Emergency Medicine*. California : San Francisco.

Prasad, A. M, Iverson, L. R & Liaw, A. (2006). *Newer Classification and Regression Tree Techniques : Bagging and Random Forests for Ecological Prediction, Ecosystems* 9:181-199, DOI: 10.1007/s10021-005-0054-1.

Peters, J., Baets, B. D., Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. D. & Huybrechts, W. (2007). *Random Forest as a Tool for Ecohydrological Distribution Modelling, Ecological Mod*eling 207. 304-318.

Pratiwi, F. E & Zain, I. (2014). *Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Trees) di Provinsi Sulawesi Utara, Jurnal Sains dan Seni POMITS*, Vol.3, No.1, 2337-3520.

Raharjo, R. (2014). *Tertib Berlalu Lintas*. Yogyakarta : Shafa Media.

Republik Indonesia. (2009). *Undang-Undang Republik Indonesia Nomor 22 Tahun 2009 Tentang Lalu Lintas dan Angkutan Jalan*. Bandung : Fokusmedia.

Sartono, B. & Syafitri, U. D. (2010). *Metode Pohon Gabungan : Solusi Pilihan untuk Mengatasi Kelemahan Pohon Regresi dan Klasifikasi Tunggal, Forum Statistika dan Komputasi, Vol. 15, No.1, ISSN : 0853-8115.*

Sungkono, J. (2013). *Resampling Bootstrap Pada R, Magistra No. 84, Th. XXV, ISSN : 0215-9511*.

Vaxjo, K. (2014). *Evaluation of Logistic Regression and Random Forest Classification Based on Prediction Accuracy and Metadata Analysis*. Institutionen for Matematik.

Zhang, G., Yau, K. K. W. & Chen, G. (2013). *Risk Factors Associated with Traffic Violations and Accident Severity in China, Accident Analysis and Preventation,* Vol. 59, Hal.18-25.

Zheng, H., Chen, L., Han, X., Zhao, X. & Ma, Yan. (2009). *Classification and Regression Tree (CART) for Analysis of Soybean Yield Variability Among Fields in Northeast China : The Importance of Phosporus Application Rates*

under Drought Conditions, Agriculture, Ecosystems and Environment 132, 98-105.

(Halaman ini sengaja dikosongkan)

BIODATA PENULIS

Ita Rakhmawati, lahir di Jombang pada tanggal 01 O ktober 1993, merupakan anak kedua dari tiga bersaudara. Penulis mulai menempuh jenjang pendidikan formal di bangku RA/TK Darussalam dan dilanjutkan hingga ke MI Darussalam, MTsN Jogoroto, SMAN 3 Jombang dan pada tahun 2011 terdaftar sebagai mahasiswa di Jurusan Statistika ITS melalui jalur masuk SNMPTN Undangan Bidik Misi. Pengalaman perkuliahan selama 4 tahun, penulis mengisinya dengan mengikuti beberapa organisasi, kepanitian, kegiatan *training*, penyaji makalah, olimpiade, kerja *part time*. Organisasi yang pernah saya ikuti adalah FORSIS ITS sebagai staff keputrian periode 2011/2012, sebagai guru pengajar di kampung binaan yang diadakan oleh JMMI ITS periode 2012/2013. Kepanitian yang pernah saya ikuti adalah panitia STASION (*Statistics Competition*), panitia *FMIPA In Art*, panitia ONDIF 1 (Orientasi Dakwah Islam FORSIS), panitia *ESQ Training*, panitia Konferensi Nasional Matematika XVII. Selain kegiatan di kampus, penulis juga kerja *part time* di luar kampus yaitu sebagai guru les privat SD (semua mata pelajaran), SMP (Matematika dan IPA) dan SMA (Matematika). Penulis menerima segala saran, kritik, pertanyaan serta komentar yang membangun dari pembaca dapat melalui email atau No. Hp yaitu ita_1or93@yahoo.co.id atau 085790243969.