ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 4, Nomor 3, Tahun 2015, Halaman 441-451
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

KETEPATAN KLASIFIKASI TINGKAT KEPARAHAN KORBAN KECELAKAAN LALU LINTAS MENGGUNAKAN METODE REGRESI LOGISTIK ORDINAL DAN *FUZZY K-NEAREST NEIGHBOR IN EVERY CLASS*

Candra Silvia[1], Yuciana Wilandari[2], Abdul Hoyyi[3]
[1]Mahasiswa Jurusan Statistika FSM UNDIP
[2,3]Staf Pengajar Jurusan Statistika FSM UNDIP
candrasilvia12@gmail.com[1], yuciana.wilandari@gmail.com[2], hoyyistat@gmail.com[3]

ABSTRACT

Traffic accident is an accidental event on the road involving vehicles with or without another road users which causes damage for the victims. Semarang has quite high number of traffic accidents, which in 2014 occured 801 cases of traffic accidents. Based on the government regulation number 43 of 1993 about highway infrastructure and traffic, the impact of traffic accidents can be classified based on victims conditions such as minor injuries, serious injuries, and died. In this research will discuss about the accuracy of severity traffic accidents victim classification in Semarang in 2014 using Ordinal Logistic Regression method and Fuzzy K-Nearest Neighbor in Every Class (FK-NNC). The result of Ordinal Logistic Regression method analysis produces the accuracy of classification value is 90,5405%, meanwhile Fuzzy K-Nearest Neighbor in Every Class method produces the accuracy of classification method is 89,19%.

Keywords: Traffic accidents, Ordinal Logistic Regression, *Fuzzy K-Nearest Neighbor in Every Class*

1. PENDAHULUAN

Kecelakaan lalu lintas merupakan kejadian di jalan yang tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain sehingga mengakibatkan kerugian bagi korbannya. Menurut Polrestabes Semarang jumlah kecelakaan lalu lintas tahun 2013 adalah sebanyak 957 kasus dengan korban meninggal dunia sebanyak 196, luka berat sebanyak 49, luka ringan sebanyak 1.221, dan total kerugian materi sebesar Rp 1.438.200.000,00. Sedangkan pada tahun 2014 jumlah kecelakaan lalu lintas sebanyak 801 kasus dengan korban meninggal dunia sebanyak 88, luka berat sebanyak 90, luka ringan sebanyak 970 dan total kerugian materi sebesar Rp1.424.650.000,00.

Angka kecelakaan pada tahun 2014 mengalami penurunan dari tahun 2013, akan tetapi angka tersebut dirasa masih cukup tinggi. Oleh karena itu diperlukan analisis lebih lanjut mengenai faktor-faktor yang mempengaruhi tingkat keparahan korban kecelakaan lalu lintas di Kota Semarang sehingga dengan mengetahui faktor-faktor yang mempengaruhi tingkat keparahan korban kecelakaan lalu lintas diharapkan dapat mengurangi angka kecelakaan lalu lintas pada tahun berikutnya. Salah satu analisis yang digunakan untuk mengetahui faktor-faktor yang mempengaruhi tingkat keparahan korban kecelakaan lalu lintas adalah metode regresi logistik ordinal. Selain mengetahui faktor-faktor yang mempengaruhi tingkat keparahan korban kecelakaan lalu lintas, dari regresi logistik ordinal juga dapat diketahui nilai ketepatan klasifikasinya. Analisis lain yang dapat digunakan untuk menghitung ketepatan klasifikasi atau akurasi dalam penelitian ini adalah metode data mining. Salah satu metode tersebut adalah *Fuzzy K-Nearest Neighbor in Every Class* (FK-NNC). Dengan demikian maka ketepatan klasifikasi yang terbaik dapat ditentukan dari kedua metode tersebut.

2. TINJAUAN PUSTAKA

2.1 Definisi Kecelakaan Lalu Lintas

Kecelakaan lalu lintas merupakan kejadian di jalan yang tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain sehingga mengakibatkan kerugian bagi korbannya. Berdasarkan Peraturan Pemerintah Nomor 43 tagun 1993, dampak kecelakaan lalu lintas dapat diklasifikasi berdasarkan kondisi korban menjadi tiga, yaitu:

a. Meninggal dunia adalah korban kecelakaan yang dipastikan meninggal dunia sebagai akibat kecelakaan lalu lintas dalam jangka waktu paling lama 30 hari setelah kecelakaan tersebut.
b. Luka berat adalah korban kecelakaan yang karena luka-lukanya menderita cacat tetap atau harus dirawat inap di rumah sakit dalam jangka waktu lebih dari 30 hari sejak kecelakaan.
c. Luka ringan adalah korban kecelakaan yang mengalami luka-luka yang tidak memerlukan rawat inap atau harus dirawat inap di rumah sakit kurang dari 30 hari.

Menurut Munawar (2004) kecelakaan disebabkan oleh berbagai faktor, yaitu:

a. Manusia atau pemakai jalan
   Pemakai jalan adalah semua orang yang menggunakan fasilitas jalan secara langsung meliputi pengemudi, pejalan kaki dan pemakai jalan yang lain. Sifat pengemudi yang sangat berpengaruh dalam mengendalikan kendaraan adalah pribadinya, latihan dan sikap.
b. Kendaraan
   Kecelakaan dapat timbul karena perlengkapan kendaraan yang kurang bagus, kondisi penerangan kendaraan, mesin kendaraan, pengaman kendaraan dan lain-lain.
c. Jalan dan lingkungan
   Sifat-sifat jalan berpengaruh sebagai penyebab kecelakaan lalu lintas. Perbaikan terhadap kondisi jalan akan mempengaruhi pula terhadap karakteristik kecelakaan yang terjadi.

2.2 Model Regresi Logistik Ordinal

Menurut Agresti (2002) model regresi logistik termasuk dalam model linear umum (*Generalized Linear Models*). Model regresi logistik juga dapat disebut sebagai model logit. Model logit digunakan untuk memodelkan hubungan antara satu variabel respon yang bersifat kategori dan beberapa variabel bebas yang bersifat kategori maupun kontinu. Apabila variabel respon terbagi menjadi lebih dari dua kategori dan terdapat tingkatan dalam kategori tersebut (skala ordinal) maka dinamakan model regresi logistik ordinal.

Di dalam Agresti (2002) model untuk regresi logistik ordinal adalah model logit kumulatif (*cumulative logit models*). Pada model logit ini sifat ordinal dari respon Y dituangkan dalam peluang kumulatif. Misalkan variabel respon Y memiliki G buah kategori berskala ordinal dan $\mathbf{x}_i$ menyatakan vektor variabel prediktor pada pengamatan ke-i, $\mathbf{x}_i = \begin{bmatrix} x_{i1} & x_{i2} & \dots & x_{ip} \end{bmatrix}^T$ dengan $i = 1, 2, ..., n$, maka model logit kumulatif dinyatakan sebagai berikut:

$$\text{logit}\left[P\left(Y_i \le g \,|\, \mathbf{x}_i\right) \right] = \alpha_g + \mathbf{x}_i^T \boldsymbol{\beta} \quad , \quad g = 1, 2, ..., G-1$$

dimana $P\left(Y_i \le g \,|\, \mathbf{x}_i\right)$ adalah peluang kumulatif kurang dari atau sama dengan kategori ke-g jika diketahui $\mathbf{x}_i$, α_g merupakan parameter intersep dan memenuhi kondisi

$\alpha_1 \le \alpha_2 \le ... \le \alpha_{G-1}$ dan $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \beta_2 & ... & \beta_p \end{bmatrix}^T$ merupakan vektor koefisien regresi yang bersesuaian dengan $x_1, x_2, ..., x_p$.

Di dalam Agresti (2002) logit kumulatif didefinisikan sebagai:

$$\text{logit}\left[P\left(Y_i \le g \left| \mathbf{x}_i \right.\right)\right] = \ln\left[\frac{P\left(Y_i \le g \left| \mathbf{x}_i \right.\right)}{1 - P\left(Y_i \le g \left| \mathbf{x}_i \right.\right)} \right] \quad , g = 1, 2, ..., G-1$$

maka model regresi logistik ordinal dapat dinyatakan sebagai:

$$\text{logit}\left[P\left(Y_i \le g \left| \mathbf{x}_i \right.\right)\right] = \ln\left[\frac{P\left(Y_i \le g \left| \mathbf{x}_i \right.\right)}{1 - P\left(Y_i \le g \left| \mathbf{x}_i \right.\right)} \right] = \alpha_g + \mathbf{x}_i^T \boldsymbol{\beta} \quad , g = 1, 2, ..., G-1$$

Sehingga peluang untuk masing-masing kategori respon dapat dinyatakan sebagai:

$$\pi_g\left(\mathbf{x}_i\right) = \frac{\exp\left(\alpha_g + \mathbf{x}_i^T \boldsymbol{\beta}\right)}{1 + \exp\left(\alpha_g + \mathbf{x}_i^T \boldsymbol{\beta}\right)} - \frac{\exp\left(\alpha_{g-1} + \mathbf{x}_i^T \boldsymbol{\beta}\right)}{1 + \exp\left(\alpha_{g-1} + \mathbf{x}_i^T \boldsymbol{\beta}\right)}, \quad g = 1, 2, ..., G$$

Di dalam Agresti (2002) penaksiran parameter regresi logistik ordinal dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Menurut Hosmer dan Lemeshow (2000) prinsip dari metode MLE adalah mengestimasi vektor parameter $\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 & \alpha_2 & ... & \alpha_{G-1} & \beta_1 & \beta_2 & ... & \beta_p \end{bmatrix}^T$ dengan cara memaksimumkan fungsi *likelihood.* Untuk mempermudah perhitungan, maka dilakukan transformasi *ln* pada fungsi *likelihood* sehingga terbentuk fungsi *ln-likelihood*. Estimasi parameter melalui metode MLE adalah dengan melakukan turunan parsial fungsi *ln-likelihood* terhadap parameter yang akan diestimasi kemudian disamadengankan nol. Turunan parsial pertama dari fungsi *ln-likelihood* yang akan diestimasi merupakan fungsi yang nonlinear terhadap parameter. Estimasi parameter dari persamaan regresi yang nonlinear tidak mudah jika menggunakan metode kemungkinan maksimum dan memerlukan metode yang bersifat iterasi untuk memperoleh estimasi parameternya. Menurut Agresti (2002) metode iterasi yang digunakan adalah metode iterasi *Newton Raphson*.

Pengujian parameter model regresi logistik ordinal dapat dilakukan secara keseluruhan maupun individu serta uji kesesuaian model.

a. Uji Rasio Likelihood (Uji Keseluruhan)

Menurut Hosmer dan Lemeshow (2000) uji keseluruhan digunakan untuk mengetahui apakah variabel bebas yang terdapat dalam model berpengaruh nyata atau tidak secara keseluruhan. Hipotesis yang digunakan:

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

H_1 : paling sedikit ada satu $\beta_k \ne 0$ dengan $k = 1, 2, ..., p$

Statistik uji yang digunakan yaitu uji rasio *likelihood* adalah:

$$G^2 = -2 \ln \left(\frac{\textit{likelihood tanpa variabel bebas}}{\textit{likelihood dengan variabel bebas}}\right)$$

H_0 ditolak apabila nilai $G^2 > \chi^2_{(\alpha, p)}$ atau nilai signifikansi < α.

b. Uji Wald (Uji Parameter Secara Individu)

Menurut Hosmer dan Lemeshow (2000) uji wald untuk mengetahui signifikansi parameter terhadap variabel respon. Hipotesis yang digunakan:

$H_0 : \beta_k = 0$

$H_1 : \beta_k \ne 0$ dengan $k = 1, 2, ..., p$

Statistik uji yang digunakan yaitu: $W_k = \left[\frac{\hat{\beta}_k}{\widehat{SE}(\hat{\beta}_k)}\right]^2$

dengan $\hat{\beta}_k$ merupakan penaksir parameter β_k dan standar error $\hat{\beta}_k$ diperoleh dari $SE(\hat{\beta}_k) = \sqrt{Vâr(\hat{\beta}_k)}$. H_0 ditolak apabila nilai $W_k > \chi^2_{(\alpha,1)}$ atau nilai signifikansi < α .

c. Uji Kesesuaian Model (*Goodness of Fit*)

Menurut Hosmer dan Lemeshow (2000) uji kesesuaian model digunakan untuk menilai apakah model sesuai atau tidak. Hipotesis yang digunakan:

H_0: Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

H_1: Model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi)

Statistik uji yang digunakan yaitu: $D = -2\sum_{i=1}^{n}\sum_{g=1}^{G}\left[y_{ig} \ln\left(\frac{\hat{\pi}_{ig}}{y_{ig}}\right)\right]$

dengan $\hat{\pi}_{ig} = \hat{\pi}_g(x_i)$ merupakan peluang observasi ke-*i* pada kategori ke-*g*, *df* = J-(p+1) dimana J merupakan jumlah kovariat. H_0 ditolak jika $D > \chi^2_{(\alpha,\, df)}$

2.3 *Fuzzy K-Nearest Neighbor in Every Class*

Menurut Prasetyo (2012a) metode *Fuzzy K-Nearest Neighbor in Every Class* (FK-NNC) menggunakan sejumlah K tetangga terdekat pada setiap kelas dari sebuah data uji. Setiap data uji x_i harus dicarikan K tetangga terdekat pada setiap kelas menggunakan formula jarak seperti berikut:

$$d(x_i,x_j) = \left(\sum_{l=1}^{N} |x_{il} - x_{jl}|^t\right)^{\frac{1}{t}}$$

dimana N adalah dimensi (jumlah fitur) data. Dan t adalah penentu jarak yang digunakan. Dalam artikel ini digunakan t = 2 yang biasa disebut dengan jarak Euclidean.

Menurut Prasetyo (2012a), jarak atau ukuran ketidakmiripan suatu data kategorik ordinal digunakan rumus sebagai berikut:

$$d = |x_{il} - x_{jl}| / (q - 1)$$

nilainya dipetakan ke tipe integer 0 sampai q − 1, dimana q adalah banyaknya kategorik. Sedangkan ukuran ketidakmiripan suatu data rasio adalah:

$$d = |x_{il} - x_{jl}|$$

Selanjutnya menurut Prasetyo (2012a), jarak data uji x_i ke semua K tetangga dari setiap kelas ke-g dijumlahkan. Formula yang digunakan adalah

$$S_{ig} = \sum_{r=1}^{K} d_r\left(x_i, x_j\right)^{\frac{-2}{m-1}}$$

Nilai d_r sebagai akumulasi jarak data uji x_i ke K tetangga dalam kelas ke-g dilakukan sebanyak G kelas. Nilai m di sini merupakan pangkat bobot yang menunjukkan banyak kelas (*weight exponent*). Selanjutnya, akumulasi jarak data uji x_i ke setiap kelas digabungkan, disimbolkan D. Formula yang digunakan adalah

$$D_i = \sum_{g=1}^{G} S_{ig}$$

Untuk mendapatkan nilai u_{ig}, nilai keanggotaan data uji x_i pada setiap kelas ke-*g* (ada G kelas), menggunakan rumus:

$$u_{ig} = \frac{S_{ig}}{D_i}$$

Untuk menentukan kelas hasil prediksi data uji x_i, dipilih kelas dengan nilai keanggotaan terbesar dari data x_i. Formula yang digunakan adalah

$$y' = max_{g=1}^{G}(u_{ig})$$

dengan: y' = kelas prediksi, G = banyak kelas

2.4 Ketepatan Klasifikasi

Ketepatan klasifikasi yang dipakai pada penelitian ini adalah APER (*Apparent Error Rate*). Menurut Johnson dan Wichern (2007) APER adalah ukuran evaluasi yang digunakan untuk melihat peluang kesalahan klasifikasi yang dihasilkan oleh suatu fungsi klasifikasi. Sehingga untuk mencari nilai ketepatannya dapat menggunakan 1-APER.

3. METODOLOGI PENELITIAN

Sumber data yang digunakan dalam penelitian ini adalah data sekunder yaitu data kecelakaan lalu lintas di kota Semarang yang bersumber dari Satlantas Polrestabes (Satuan Lalu Lintas Polisi Resor Kota Besar) Semarang. Data ini diambil dari Januari 2014 sampai Desember 2014. Variabel penelitian yang dianalisis terdiri dari variabel respon dan variabel bebas. Secara ringkas variabel-variabel yang digunakan dalam penelitian dapat disajikan dalam Tabel 1.

Tabel 1. Variabel Penelitian yang Digunakan

No	Nama Variabel	Keterangan
1.	Tingkat keparahan korban kecelakaan lalu lintas (Y)	Y = 1 (korban luka ringan) Y = 2 (korban luka berat) Y = 3 (korban meninggal dunia)
2.	Jenis kecelakaan (X_1)	1 = Tabrak belakang 2 = Tabrak depan 3 = Tabrak samping 4 = Hilang kendali 5 = Lain-lain
3.	Jenis kelamin (X_2)	1 = Laki-laki 2 = Perempuan
4.	Usia (X_3)	
5.	Peran korban dalam kecelakaan (X_4)	1 = Pengguna kendaraan 2 = Pengguna jalan non pengguna kendaraan (penyebrang jalan, pejalan kaki, dan lain-lain)
6.	Jenis kendaraan korban (X_5)	1 = Lain-lain (pejalan kaki, sepeda angin, becak, atau kendaraan bukan bermotor lainnya) 2 = Sepeda motor (kendaraan bermotor roda dua atau tiga) 3 = Kendaraan roda empat
7.	Jenis kendaraan lawan (X_6)	1 = Lain-lain (pejalan kaki, sepeda angin, becak, atau kendaraan bukan bermotor lainnya) 2 = Sepeda motor (kendaraan bermotor roda dua atau tiga) 3 = Kendaraan roda empat 4 = Kendaraan dengan lebih dari empat roda
8	Waktu kecelakaan (X_7)	1 = Padat kendaraan (antara pukul 06.00 WIB – 08.00 WIB, antara pukul 12.00 WIB – 13.30 WIB, antara pukul 16.00 WIB – 18.00 WIB) 2 = Sepi kendaraan (selain waktu padat kendaraan)

Tahapan-tahapan analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Mengumpulkan data kecelakaan lalu lintas yang akan digunakan dalam penelitian
2. Membagi semua data menjadi 2 bagian, berupa training 90% dan testing 10%

3. Menentukan model awal dari metode Regresi Logistik Ordinal menggunakan data training
4. Melakukan uji Rasio Likelihood atau uji secara keseluruhan terhadap data training untuk mengetahui apakah variabel independen yang terdapat dalam model berpengaruh nyata atau tidak secara keseluruhan
5. Melakukan uji Wald terhadap data training untuk mengetahui signifikansi parameter terhadap variabel respon
6. Melakukan uji Kesesuaian model terhadap data training untuk mengetahui apakah model sesuai atau tidak
7. Menentukan model akhir dari metode Regresi Logistik Ordinal
8. Menghitung ketepatan klasifikasi atau akurasi menggunakan model akhir regresi logistik ordinal pada data testing
9. Melakukan pengolahan data menggunakan metode *Fuzzy K-Nearest Neighbor in Every Class* sesuai dengan model akhir dari Regresi Logistik Ordinal
10. Mencari K tetangga terdekat pada kelas 1, kelas 2, dan kelas 3
11. Menghitung nilai S_{ig} sebagai akumulasi jarak K tetangga terdekat pada kelas 1, kelas 2, dan kelas 3
12. Menghitung nilai D_i sebagai akumulasi semua jarak dari GxK tetangga
13. Menghitung nilai u_{ig} sebagai nilai keanggotaan data pada kelas 1, kelas 2 maupun kelas 3
14. Menentukan nilai keanggotaan terbesar untuk dijadikan kelas hasil prediksi data tingkat keparahan korban kecelakaan lalu lintas tersebut
15. Menghitung nilai ketepatan klasifikasi atau akurasi terhadap data tingkat keparahan korban kecelakaan lalu lintas
16. Membandingkan antara ketepatan klasifikasi data dengan Regresi Logistik Ordinal dan data diolah dengan *Fuzzy K-Nearest Neighbor in Every Class*
17. Memilih ketepatan klasifikasi yang paling tinggi

4. HASIL DAN PEMBAHASAN

a. Model Regresi Logistik Ordinal

1. Model awal tahap pertama

$$\text{Logit}[P(Y_i \leq 1|X_i)] = 1{,}438 - 1{,}174\,X_1(1) - 0{,}652\,X_1(2) - 0{,}811\,X_1(3) + 0{,}566\,X_1(4) + 0{,}089\,X_2(1) - 0{,}016\,X_3 - 0{,}533\,X_4(1) - 2{,}085\,X_5(1) - 0{,}750\,X_5(2) + 4{,}326\,X_6(1) + 4{,}018\,X_6(2) + 2{,}629\,X_6(3) + 0{,}499\,X_7(1)$$

$$\text{Logit}[P(Y_i \leq 2|X_i)] = 2{,}329 - 1{,}174\,X_1(1) - 0{,}652\,X_1(2) - 0{,}811\,X_1(3) + 0{,}566\,X_1(4) + 0{,}089\,X_2(1) - 0{,}016\,X_3 - 0{,}533\,X_4(1) - 2{,}085\,X_5(1) - 0{,}750\,X_5(2) + 4{,}326\,X_6(1) + 4{,}018\,X_6(2) + 2{,}629\,X_6(3) + 0{,}499\,X_7(1)$$

a. Uji Rasio Likelihood (Uji Keseluruhan)

Hipotesis : $H_0 : \beta_1 = \beta_2 = ... = \beta_{13} = 0$ (Model tidak signifikan)

H_1 : Paling sedikit ada salah satu dari $\beta_k \neq 0$ dengan k=1,2,...,13 (Model signifikan)

Taraf Signifikansi : $\alpha = 5\%$

Statistik Uji : $G^2 = -2 \ln \left(\frac{likelihood\ tanpa\ variabel\ bebas}{likelihood\ dengan\ variabel\ bebas}\right) = 90{,}409.$

Kriteria Uji : H_0 ditolak jika $G^2 > \chi^2_{(0,05;13)}$, dimana nilai $\chi^2_{(0,05;13)}$ adalah 22,36

Keputusan : Karena nilai $G^2 = 90{,}409 > (\chi^2_{(0,05;13)}) = 22{,}36$ maka H_0 ditolak.

Sehingga dapat disimpulkan bahwa model signifikan.

b. Uji Wald

Hipotesis : $H_0 : \beta_k = 0$ (parameter tidak signifikan atau variabel bebas tidak memiliki hubungan yang kuat dengan variabel respon)

$H_1 : \beta_k \neq 0$ dengan k = 1,2,....,13 (parameter signifikan atau variabel bebas memiliki hubungan yang kuat dengan variabel respon)

Taraf Signifikansi : α = 5%

Statistik Uji : $W_k = \left[\frac{\hat{\beta}_k}{\widehat{SE}(\hat{\beta}_k)}\right]^2$

Hasil uji wald dapat dilihat pada Tabel 2.

Kriteria Uji : H_0 ditolak jika $W_k > \chi^2_{(0,05;1)}$

Tabel 2. Uji Wald Tahap Pertama

Variabel Bebas	Wald	sig.	$\chi^2_{(0,05;1)}$	Keputusan
$[X_1=1]$	1,112	0,292	3,84	H_0 diterima
$[X_1=2]$	0,330	0,566	3,84	H_0 diterima
$[X_1=3]$	0,542	0,462	3,84	H_0 diterima
$[X_1=4]$	0,162	0,687	3,84	H_0 diterima
$[X_2=1]$	0,087	0,768	3,84	H_0 diterima
X_3	4,330	0,037	3,84	H_0 ditolak
$[X_4=1]$	0,914	0,339	3,84	H_0 diterima
$[X_5=1]$	4,152	0,042	3,84	H_0 ditolak
$[X_5=2]$	0,741	0,389	3,84	H_0 diterima
$[X_6=1]$	14,625	0,000	3,84	H_0 ditolak
$[X_6=2]$	53,550	0,000	3,84	H_0 ditolak
$[X_6=3]$	27,089	0,000	3,84	H_0 ditolak
$[X_7=1]$	3,234	0,072	3,84	H_0 diterima

Sehingga dapat disimpulkan variabel bebas yang signifikan dan memiliki hubungan yang kuat dengan variabel respon adalah X_3, X_5, X_6.

2. Model awal tahap kedua

Logit $[P(Y_i \leq 1|X_i)] = 0,635 - 0,017\ X_3 - 1,788\ X_5(1) - 0,872\ X_5(2) + 4,213\ X_6(1) + 3,834\ X_6(2) + 2,412\ X_6(3)$

Logit $[P(Y_i \leq 2|X_i)] = 1,501 - 0,017\ X_3 - 1,788\ X_5(1) - 0,872\ X_5(2) + 4,213\ X_6(1) + 3,834\ X_6(2) + 2,412\ X_6(3)$

a. Uji Rasio Likelihood (Uji Keseluruhan)

Hipotesis : $H_0 : \beta_1 = \beta_2 = ... = \beta_6 = 0$ (Model tidak signifikan)

H_1 : Paling sedikit ada salah satu dari $\beta_k \neq 0$ dengan k = 1, 2, ...,6 (Model signifikan)

Taraf Signifikansi : α = 5%

Statistik Uji : $G^2 = -2 \ln \left(\frac{likelihood\ tanpa\ variabel\ bebas}{likelihood\ dengan\ variabel\ bebas}\right) =$ 79,433.

Kriteria Uji : H_0 ditolak jika $G^2 > \chi^2_{(0,05;6)}$ dimana nilai $\chi^2_{(0,05;6)}$ adalah 12,59

Keputusan : Karena nilai $G^2 = 79,433 > (\chi^2_{(0,05;6)}) = 12,59$ maka H_0 ditolak.

Sehingga dapat disimpulkan bahwa model signifikan.

b. Uji Wald

Hipotesis : $H_0 : \beta_k = 0$ (parameter tidak signifikan atau variabel bebas tidak memiliki hubungan yang kuat dengan variabel respon)

$H_1 : \beta_k \neq 0$ dengan k = 1,2,...,6 (parameter signifikan atau variabel bebas memiliki hubungan yang kuat dengan variabel respon)

Taraf Signifikansi : $\alpha = 5\%$

Statistik Uji : $W_k = \left[\frac{\widehat{\beta}_k}{\widehat{SE}(\widehat{\beta}_k)}\right]^2$

Hasil uji wald dapat dilihat pada Tabel 3.

Kriteria Uji : H_0 ditolak jika $W_k > \chi^2_{(0,05;1)}$

Tabel 3. Uji Wald Tahap Kedua

Variabel Bebas	Wald	sig.	$\chi^2_{(0,05;1)}$	Keputusan
X_3	4,902	0,027	3,84	H_0 ditolak
$[X_5=1]$	3,834	0,050	3,84	H_0 diterima
$[X_5=2]$	1,041	0,308	3,84	H_0 diterima
$[X_6=1]$	14,625	0,000	3,84	H_0 ditolak
$[X_6=2]$	53,550	0,000	3,84	H_0 ditolak
$[X_6=3]$	27,089	0,000	3,84	H_0 ditolak

Sehingga dapat disimpulkan variabel bebas yang signifikan dan memiliki hubungan yang kuat dengan variabel respon adalah X_3 dan X_6.

3. Model awal tahap ketiga

Logit $[P(Y_i \leq 1|X_i)] = 0{,}315 - 0{,}025\ X_3 + 3{,}937\ X_6(1) + 3{,}262\ X_6(2) + 2{,}048\ X_6(3)$

Logit $[P(Y_i \leq 2|X_i)] = 1{,}167 - 0{,}025\ X_3 + 3{,}937\ X_6(1) + 3{,}262\ X_6(2) + 2{,}048\ X_6(3)$

a. Uji Rasio Likelihood (Uji Keseluruhan)

Hipotesis : $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (Model tidak signifikan)

H_1 : Paling sedikit ada salah satu dari $\beta_k \neq 0$ dengan k=1,2,3,4 (Model signifikan)

Taraf Signifikansi : $\alpha = 5\%$

Statistik Uji : $G^2 = -2 \ln \left(\frac{likelihood\ tanpa\ variabel\ bebas}{likelihood\ dengan\ variabel\ bebas}\right) = 70{,}526.$

Kriteria Uji : H_0 ditolak jika $G^2 > \chi^2_{(0,05;4)}$ dimana nilai $\chi^2_{(0,05;4)}$ adalah 9,49

Keputusan : Karena nilai $G^2 = 70{,}526 > (\chi^2_{(0,05;4)}) = 9{,}49$ maka H_0 ditolak.

Sehingga dapat disimpulkan bahwa model signifikan.

b. Uji Wald

Hipotesis : $H_0 : \beta_k = 0$ (parameter tidak signifikan atau variabel bebas tidak memiliki hubungan yang kuat dengan variabel respon)

$H_1 : \beta_k \neq 0$ dengan k = 1, 2, 3, 4 (parameter signifikan atau variabel bebas memiliki hubungan yang kuat dengan variabel respon)

Taraf Signifikansi : $\alpha = 5\%$

Statistik Uji : $W_k = \left[\frac{\widehat{\beta}_k}{\widehat{SE}(\widehat{\beta}_k)}\right]^2$

Hasil uji wald dapat dilihat pada Tabel 4.

Kriteria Uji : H_0 ditolak jika $W_k > \chi^2_{(0,05;1)}$ atau nilai signifikansi < 5% (α)

Tabel 4. Uji Wald Tahap Ketiga

Variabel Bebas	Wald	sig.	$\chi^2_{(0,05;1)}$	Keputusan
X_3	11,844	0,001	3,84	H_0 ditolak
$[X_6=1]$	12,711	0,000	3,84	H_0 ditolak
$[X_6=2]$	50,092	0,000	3,84	H_0 ditolak
$[X_6=3]$	22,071	0,000	3,84	H_0 ditolak

Sehingga dapat disimpulkan variabel bebas yang signifikan dan memiliki hubungan yang kuat dengan variabel respon adalah X_3 dan X_6.

c. Uji Kesesuaian Model

Hipotesis : H_0 : Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

H_1 : Model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi)

Taraf Signifikansi : α = 5%

Statistik Uji : $D = -2\sum_{i=1}^{n}\sum_{g=1}^{G}\left[y_{ig} \ln\left(\frac{\hat{\pi}_{ig}}{y_{ig}}\right)\right] = 248{,}533$

Kriteria Uji : H_0 ditolak jika nilai Deviance > $\chi^2_{(0,05;328)}$

Keputusan : Karena nilai Deviance = 248,533 < $\chi^2_{(0,05;\,328)}$ = 371,23 maka H_0 diterima.

Sehingga dapat disimpulkan bahwa model akhir tahap ketiga sesuai atau tidak ada perbedaan antara hasil observasi dengan hasil prediksi.

4. Ketepatan Klasifikasi

Tabel 5. APER Metode Regresi Logistik Ordinal

		prediksi		
		luka ringan [Y=1]	luka berat [Y=2]	meninggal dunia [Y=3]
observasi	luka ringan [Y=1]	66	0	0
	luka berat [Y=2]	2	0	1
	meninggal dunia [Y=3]	4	0	1

$$APER = \frac{0+0+2+1+4+0}{66+0+0+2+0+1+4+0+1} \times 100\% = 9{,}4595\ \%$$

Sehingga dapat disimpulkan bahwa nilai ketepatan klasifikasinya sebesar (1 - APER) yaitu 90,5405 %

b. Metode *Fuzzy K-Nearest Neighbor in Every Class* (FK-NNC)

Metode FK-NNC pada penelitian ini digunakan nilai K masing-masing sebesar 1, 2, 3, 4, 5, 6, 7, 8, 9. Contoh perhitungan jarak untuk data training terhadap data testing untuk K=3 adalah sebagai berikut:

a. Untuk d_1 yaitu jarak untuk semua data training terhadap data testing 1, maka perhitungannya sebagai berikut:

Data training 1 terhadap data testing 1

$d(x_1, x_1) = \sqrt{|11-15|^2 + ((|3-3|)/3)^2} = 4$

.

.

Data training 666 terhadap data testing 1

$d(x_{666}, x_1) = \sqrt{|17-15|^2 + ((|4-3|)/3)^2} = 2{,}0276$

b. Untuk d_{74} yaitu jarak untuk semua data training terhadap data testing 74, maka perhitungannya sebagai berikut:

Data training 1 terhadap data testing 74

$d(x_1, x_{74}) = \sqrt{|11-65|^2 + ((|3-2|)/3)^2} = 54{,}001$

.

.

Data training 666 terhadap data testing 74

$d(x_{666}, x_{74}) = \sqrt{|17-65|^2 + ((|4-2|)/3)^2} = 48{,}0046$

Mengambil 3 tetangga terdekat pada setiap kelas, diambil 3 tetangga terdekat untuk setiap d, mulai dari d_1, d_2, hingga d_{74}. Pengambilan tetangga terdekat yaitu dipilih nilai jarak data training terhadap data testing yang terkecil, misal untuk d_1, tiga tetangga terdekat pada kelas 1 adalah 0,0001; 0,0001; 0,3333 kemudian tiga tetangga terdekat pada kelas 2 adalah 2; 2,0276; 3 dan tiga tetangga terdekat pada kelas 3 adalah 1,0541; 1,0541; 2

Menghitung nilai S_{ig} sebagai akumulasi semua jarak 3 tetangga terdekat pada kelas 1, 2, dan 3. Rumus yang digunakan adalah:

$$S_{ig} = \sum_{r=1}^{3} d_r(x_i, x_j)^{-1}$$

Misal untuk d_1:

$S_{11} = 0{,}0001^{-1} + 0{,}0001^{-1} + 0{,}3333^{-1} = 20003$

$S_{12} = 2^{-1} + 2{,}0276^{-1} + 3^{-1} = 1{,}3265$

$S_{13} = 1{,}0541^{-1} + 1{,}0541^{-1} + 2^{-1} = 2{,}3974$

Menghitung nilai D_i sebagai akumulasi semua jarak, misal untuk d_1:

$D_1 = 20003 + 1{,}3265 + 2{,}3974 = 20006{,}72$

Menghitung nilai keanggotaan u_{ig} pada kelas 1, kelas 2, dan kelas 3. Kemudian menentukan nilai keanggotaan terbesar untuk dijadikan kelas hasil prediksi:

$$u_{ig} = \frac{S_{ig}}{D_i}$$

Misal untuk d_1:

$u_{11} = \frac{20003}{20006{,}72} = 0{,}9998$

$u_{12} = \frac{1{,}3265}{20006{,}72} = 6{,}63 \times 10^{-5}$

$u_{13} = \frac{2{,}3974}{20006{,}72} = 0{,}00012$

karena nilai u_{11} lebih besar dari u_{12} dan u_{13}, maka data testing diprediksi masuk ke kelas 1.

Berdasarkan contoh perhitungan dapat diketahui nilai ketepatan klasifikasi sebagai berikut:

Tabel 6. Hasil Ketepatan Klasifikasi FK-NNC

K	Ketepatan Klasifikasi FK-NNC	APER
1	89,19%	10,81%
2	85,14%	14,86%
3	85,14%	14,86%
4	74,32%	25,68%
5	74,32%	25,68%
6	71,62%	28,38%
7	70,27%	29,73%
8	68,92%	31,08%
9	66,22%	33,78%

5. KESIMPULAN

Berdasarkan hasil dan pembahasan maka diperoleh kesimpulan sebagai berikut:

1. Hasil yang diperoleh dari analisis regresi logistik ordinal menunjukkan bahwa variabel yang mempengaruhi tingkat keparahan korban kecelakaan lalu lintas di Kota Semarang adalah variabel usia (X_3) dan variabel jenis kendaraan lawan (X_6).
2. Model regresi logistik ordinal yang terbentuk dapat digunakan untuk menghitung ketepatan klasifikasi tingkat keparahan korban kecelakaan lalu lintas yaitu sebesar 90,5405%. Sedangkan hasil analisis FK-NNC menunjukkan bahwa pada K = 1 telah diperoleh ketepatan klasifikasi tingkat keparahan korban kecelakaan lalu lintas sebesar 89,19%.

6. DAFTAR PUSTAKA

Agresti, A. 2002. *Categorical Data Analysis Second Edition*. John Wiley and Sons. New York.

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

Johnson, R. A. dan Wichern, D. W., 2007. *Applied Multivariate Statistical Analysis.* Prentice Hall. New Jersey.

Munawar, A. 2004. *Manajemen Lalu Lintas Perkotaan*. Beta Offset. Yogyakarta.

Peraturan Pemerintah Nomor 43 Tahun 1993 tentang Prasarana Jalan Raya dan Lalu Lintas.

Prasetyo, E. 2012a. *Data Mining Konsep dan Aplikasi Menggunakan Matlab.* Andi. Yogyakarta.

Undang-Undang Republik Indonesia Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan.