PEMODELAN POISSON *HIDDEN* MARKOV UNTUK PREDIKSI BANYAKNYA KECELAKAAN DI JALAN TOL JAKARTA-CIKAMPEK

NURHASANAH[1], B. SETIAWATY[2], F. BUKHARI[2]

Abstrak

Model Poisson *hidden* Markov digunakan untuk memodelkan banyaknya kecelakaan yang terjadi di jalan tol Jakarta-Cikampek pada tahun 2013-2014. Data banyaknya kecelakaan merupakan barisan observasi yang mengalami overdispersi dan bergantung pada penyebab kecelakaan yang diasumsikan tidak diamati secara langsung dan membentuk rantai Markov. Model Poisson *hidden* Markov dicirikan oleh parameternya. Pendugaan parameter model dilakukan dengan menggunakan metode Maksimum *Likelihood* yang perhitungannya menggunakan algoritme *Expectation Maximization*. Nilai dugaan parameter digunakan untuk membangkitkan barisan penduga kecelakaan. Keakuratan model diukur menggunakan *Mean Absolute Percentage Error* (MAPE). Menggunakan kriteria AIC diperoleh model Poisson *hidden* Markov 2 *state* sebagai model terbaik dengan nilai MAPE 34.0786% untuk prediksi satu waktu yang akan datang.

Kata Kunci: Model Poisson *hidden* Markov, rantai Markov, overdispersi, algoritme *Expectation Maximization*, AIC, MAPE

1. PENDAHULUAN

Jalan tol Jakarta-Cikampek merupakan jalan tol yang menghubungkan Jakarta dengan kota-kota lainnya di Pulau Jawa, sehingga jalan tol Jakarta-Cikampek merupakan salah satu infrastruktur transportasi nasional yang penting. Oleh karena itu, jalan tol Jakarta-Cikampek sering dipadati oleh berbagai jenis kendaraan. Namun, kepadatan kendaraan ini sering kali juga disebabkan oleh terjadinya kasus kecelakaan di tol sehingga kasus kecelakaan perlu menjadi perhatian khusus. Berkaitan dengan upaya mengantisipasi terjadinya kasus kecelakaan di tol, maka dapat dikembangkan suatu model yang dapat memprediksi banyaknya kecelakaan di jalan tol.

Model ini nantinya diharapkan dapat memprediksi banyaknya kecelakaan dalam kurun waktu tertentu. Hasil prediksi dari banyaknya kecelakaan ini dapat menjadi dasar untuk penanganan kasus kecelakaan pada lokasi ruas tol yang sering terjadi kecelakaan, seperti penentuan banyaknya mobil derek, alat pengangkut

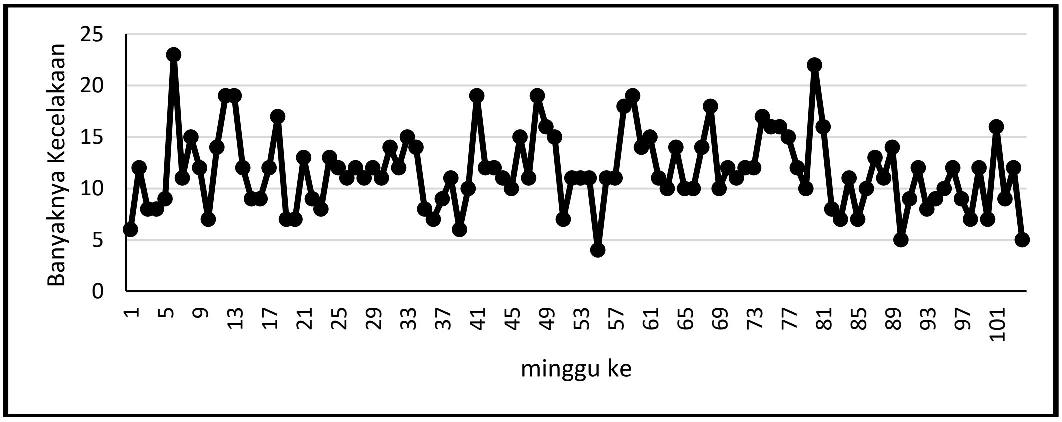
---

[1]Mahasiswa S2, Program Studi Matematika Terapan, Sekolah Pascasarjana IPB. E-mail: syarifhasanah@gmail.com

[2]Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Jalan Meranti Kampus IPB Dramaga Bogor, 16680.

kendaraan yang mengalami kecelakaan, banyaknya mobil ambulans yang bersiaga, banyaknya mobil patroli, maupun banyaknya mobil *rescue*, yakni mobil yang berisi peralatan yang digunakan dalam kecelakaan berat di jalan tol.

Kasus kecelakaan di tol dapat terjadi kapan saja dan tidak dapat dipastikan kapan terjadinya (*unpredictable*). Gambar 1 di bawah ini menunjukkan ketidakteraturan banyaknya kecelakaan di salah satu ruas tol.



Gambar 1 Data Banyaknya Kecelakaan per Minggu di Tol Jakarta-Cikampek Tahun 2013-2014 (Sumber Data : PT. Jasamarga)

Kasus kecelakaan berkaitan erat dengan faktor penyebab kecelakaan. Adapun faktor-faktor penyebab kecelakaan di tol antara lain, faktor cuaca, faktor hari raya, faktor hari libur, faktor manusia, faktor kendaraan dan lain-lain. Jika diasumsikan faktor penyebab kecelakaan ini tidak diamati secara langsung dan membentuk rantai Markov, maka data banyaknya kecelakaan dapat dimodelkan dengan model *hidden* Markov.

Sebaran Poisson biasa digunakan untuk memodelkan jumlah kecelakaan di jalan tol, namun salah satu asumsi yang harus dipenuhi pada sebaran Poisson adalah nilai ragam dan nilai rataan dari peubah responnya harus sama, hal ini tentunya tidak sesuai dengan data banyaknya kecelakaan di tol yang bersifat overdispersi, yakni nilai ragam lebih besar dari nilai rataannya. Pada data kecelakaan, penyebab kecelakaaan yang berbeda akan menyebabkan jumlah kecelakaan yang terjadi berbeda pula, sehingga pada data kecelakaan nilai rataan dari sebaran Poisson dapat berbeda sesuai dengan penyebab kejadiannya. Oleh karena itulah, maka untuk memodelkan banyaknya kecelakaan di tol digunakan model Poisson campuran, yakni model Poisson *hidden* Markov. Selanjutnya, akan ditentukan prediksi banyaknya kecelakaan, untuk meramalkan banyaknya kecelakaan yang terjadi di masa yang akan datang.

2. MODEL POISSON *HIDDEN* MARKOV

Model Poisson *hidden* Markov adalah model *hidden* Markov dengan waktu diskret yang terdiri atas pasangan $\{C_t, Y_t\}_{t\in\mathbb{N}}$ di mana $\{C_t\}_{t\in\mathbb{N}}$ merupakan penyebab kejadian yang tidak diamati secara langsung dan membentuk suatu rantai Markov, sedangkan $\{Y_t\}_{t\in\mathbb{N}}$ adalah proses observasinya yang bergantung pada $\{C_t\}_{t\in\mathbb{N}}$. Jika diasumsikan untuk setiap t, Y_t dengan diketahui C_t adalah peubah acak Poisson, maka pasangan $\{C_t, Y_t\}$ disebut model Poisson *hidden* Markov.

Karakteristik model Poisson *hidden* Markov dapat dicirikan sebagai berikut.

1. Diasumsikan $\{C_t\}_{t\in\mathbb{N}}$ adalah rantai Markov yang diskret, homogen, *ergodic*, dan tak tereduksi dengan ruang *state* $Q_C = \{1,2,\ldots,m\}$. Matriks peluang transisi *state* $\Gamma = [\gamma_{ij}]$, di mana Γ adalah matriks berukuran $m \times m$ dengan
$\gamma_{ij} = P(C_t = j|C_{t-1} = i) = P(C_2 = j|C_1 = i),\ \ 0 \le \gamma_{ij} \le 1\,, 1 \le i,j \le m,$
$$\sum_{j=1}^{m} \gamma_{ij} = 1\,, i \in Q_C\,. \tag{1}$$
2. Vektor peluang *state* awal $\delta = [\delta_i]$, di mana δ merupakan vektor berukuran $m \times 1$ dengan $\delta_i = P(C_1 = i), i = 1,2,\ldots,m$
$$\sum_{i\in Q_C} \delta_i = 1\,. \tag{2}$$
Karena rantai Markov $\{C_t\}_{t\in\mathbb{N}}$ diasumsikan rantai Markov yang *ergodic*, maka δ merupakan sebaran stasioner sehingga memenuhi persamaan $\delta' = \delta'\Gamma$.
3. Dalam model Poisson *hidden* Markov, sebaran bersyarat Y_t jika diketahui C_t berada pada *state* $i\ (i \in Q_C; t \in \mathbb{N})$ adalah Poisson dengan parameter λ_i. Matriks peluang dari proses observasi $\prod = [\pi_{yi}]$, dengan
$$\pi_{yi} = P(Y_t = y|C_t = i) = e^{-\lambda_i}\frac{\lambda_i{}^y}{y!}, y \in \{0,1,2,\ldots\}$$
$$\sum_{y=0}^{\infty} \pi_{yi} = 1\,, i \in Q_C\,. \tag{3}$$

Jadi, model Poisson *hidden* Markov $\{C_t, Y_t\}_{t\in\mathbb{N}}$ dicirikan oleh parameter δ dan $\theta = (m, \Gamma, \lambda)$ dengan
$$\delta = [\delta_i], i \in Q_C$$
$$\Gamma = [\gamma_{ij}], i \in Q_C$$
$$\prod = [\pi_{yi}] = e^{-\lambda_i}\frac{\lambda_i{}^y}{y!}\,, y \in \{0,1,2,\ldots\}, i \in Q_C \text{ dan } \lambda = (\lambda_1,\ldots,\lambda_m)'. \tag{4}$$

3. PENDUGAAN PARAMETER

Parameter $\theta = (\Gamma, \lambda)$ akan diduga menggunakan metode Maksimum *likelihood*. Barisan observasi $\mathscr{y} = \{y_1, y_2, \ldots, y_T\}$ diasumsikan dibangkitkan oleh model Poisson *hidden* Markov. Didefinisikan fungsi *likelihood* $L_T(\theta)$ dengan parameter θ sebagai berikut.

$$\begin{aligned} L_T(\theta) &= P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T) \\ &= \sum_{i_1=1}^{m} \ldots \sum_{i_T=1}^{m} \delta_{i_1} \pi_{y_1 i_1} \prod_{t=2}^{T} \gamma_{i_{t-1} i_t} \pi_{y_t i_t} . \end{aligned}$$

Fungsi *likelihood* $L_T(\theta)$ di atas disebut fungsi *likelihood* tak lengkap. Definisikan pula fungsi *Likelihood* lengkap $L_T^c(\theta)$ sebagai berikut.

$$\begin{aligned} L_T^c(\theta) &= P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T, C_1 = i_1, C_2 = i_2, \ldots, C_T = i_T) \\ &= \delta_{i_1} \prod_{t=2}^{T} \gamma_{i_{t-1} i_t} \prod_{t=1}^{T} \pi_{y_t i_t} . \end{aligned} \tag{5}$$

.

Dengan menggunakan metode Maksimum *Likelihood* akan dicari $\hat{\theta}$ yang memaksimumkan nilai $L_T(\theta)$.

3.1 Algoritme *Expectation Maximization* (Algoritme EM)

Misalkan $\{P_\theta; \theta \in \Theta\}$ adalah himpunan ukuran peluang yang terdefinisi pada ruang parameter Θ, dengan Θ adalah himpunan semua nilai θ yang mungkin, yaitu

$$\Theta = \left\{ \begin{matrix} \theta \in (\Gamma, \lambda) : \ \Gamma = (\gamma_{ij})_{m \times m}, 0 \le \gamma_{ij} \le 1, \sum_{j=1}^{m} \gamma_{ij} = 1, i \in Q_C; \\ \lambda = (\lambda_i)_{m \times 1}, \qquad \lambda_i \ge 0 \end{matrix} \right\} \subset \mathbb{R}^{m^2+m}$$

Fungsi *likelihood* yang digunakan untuk menghitung penduga parameter θ berdasarkan informasi $\mathcal{Y} = \{y_1, y_2, \ldots, y_T\}$ adalah

$$L_T(\theta) = P_\theta(Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T)$$

dan Maksimum *Likelihood* Estimator (MLE) didefinisikan oleh

$$\hat{\theta} \in \operatorname*{argmax}_{\theta \in \Theta} L_T(\theta). \tag{6}$$

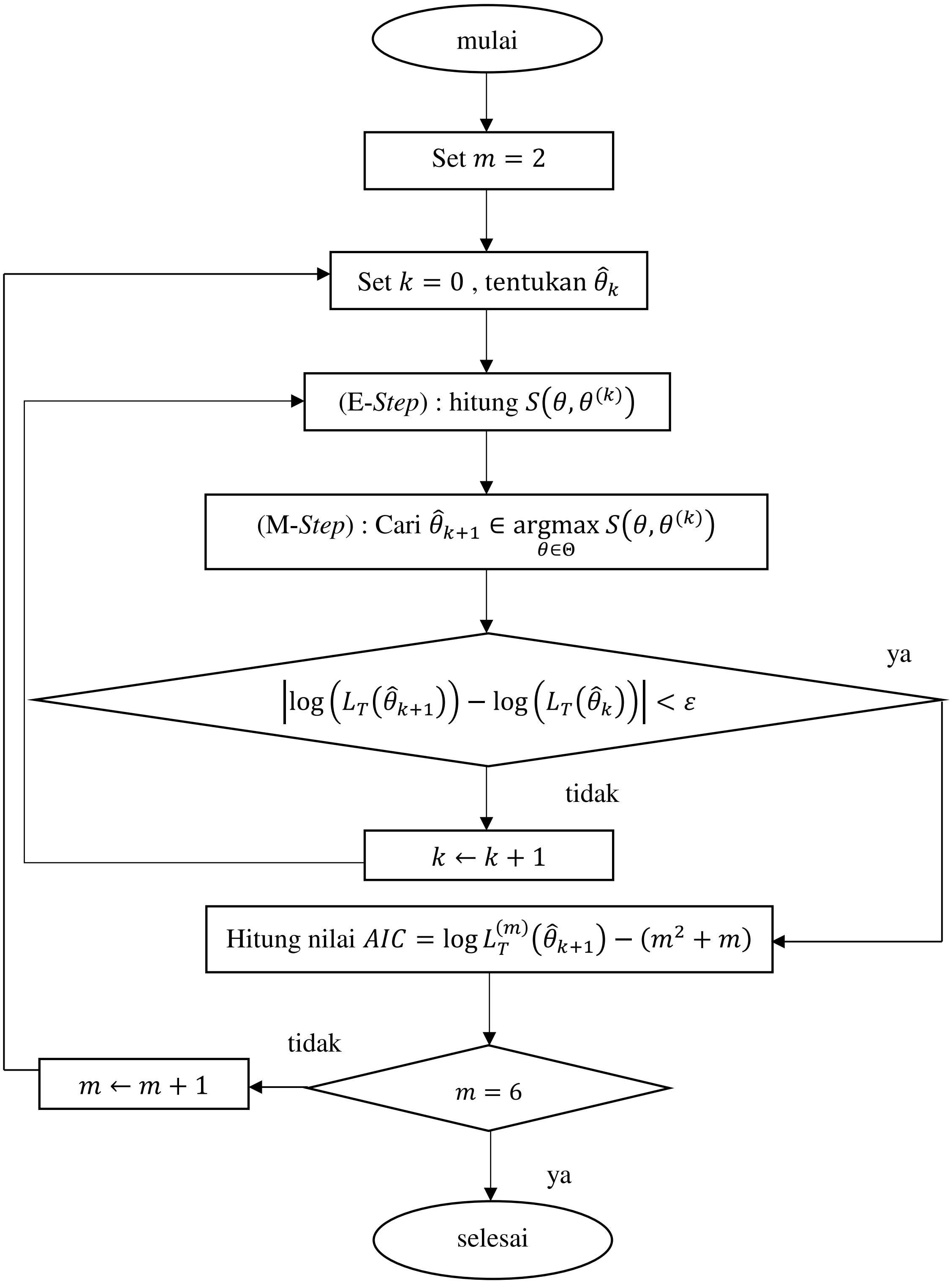
Umumnya MLE sulit dihitung secara langsung, oleh karena itu, diperlukan suatu metode aproksimasi berulang yakni algoritme *Expectation Maximization* (EM) untuk mencari nilai maksimum $\log\big(L_T(\theta)\big), \theta \in \Theta$, yang dirumuskan sebagai berikut

$$\log\big(L_T(\theta)\big) = S(\theta, \theta^*) - H(\theta, \theta^*) \tag{7}$$

dengan $S(\theta, \theta^*) = E_{\theta^*}\left[\log\left(L_T^c(\theta)\right) | \mathcal{Y}\right]$ dan $H(\theta, \theta^*) = E_{\theta^*}\left[\log\left(k(z|\mathcal{Y}, \theta)\right) | \mathcal{Y}\right]$.

Menurut Paroli & Spezia [5], untuk mencari $\theta \in \Theta$ yang memaksimumkan $\log\big(L_T(\theta)\big)$ cukup dengan mencari $\theta \in \Theta$ yang memaksimumkan $S(\theta, \theta^*)$, hal ini karena $\frac{\partial H(\theta, \theta^*)}{\partial \theta} = 0$.

Langkah-langkah dalam algoritme EM adalah sebagai berikut.



Gambar 2 Diagram alir proses pendugaan parameter dengan Algoritme EM

Pada E-*Step* dihitung nilai $S(\theta,\theta^*)$ dengan cara menghitung nilai $E_{\theta^*}(\log L_T^C(\theta)\ |\mathcal{Y})$. Nilai ekspektasi dari fungsi *likelihood* lengkap $\log L_T^C(\theta)$, yang merupakan nilai *likelihood* yang berbasis pada barisan observasi $\mathcal{Y}$ dan nilai $\hat{\theta}$ sebelumnya.

Pada M-*Step* akan dicari nilai maksimum dari $S(\theta,\theta^*), \theta \in \Theta$ yang telah dihitung pada E-*Step*. Berdasarkan persamaan $\hat{\theta}_{\boldsymbol{k+1}} \in \underset{\theta\in\Theta}{\operatorname{argmax}}\, S(\theta,\theta^*)$ maka $S(\hat{\theta}_{\boldsymbol{k+1}},\hat{\theta}_{\boldsymbol{k}}) \geq S(\theta,\hat{\theta}_{\boldsymbol{k}}), \forall\theta \in \Theta$, sehingga diperoleh nilai $S(\theta,\theta^*)$ yang tak turun dan Dempster *et al.* [1] dalam Teorema 1 telah membuktikan bahwa dalam algoritme EM untuk $k = 0,1,2,\dots$ akan dihasilkan fungsi *likelihood* yang tak turun, yakni

$$\log\left(L_T(\hat{\theta}_{\boldsymbol{k+1}})\right) \geq \log\left(L_T(\hat{\theta}_{\boldsymbol{k}})\right). \tag{8}$$

Dalam Paroli *et al.* [6], jika algoritme konvergen pada iterasi ke-$(k+1)$, maka dapat dikatakan $\left(\hat{\theta}_{\boldsymbol{k+1}}; \log\left(L_T(\hat{\theta}_{\boldsymbol{k+1}})\right)\right)$ adalah titik stasioner dan $\hat{\theta}_{\boldsymbol{k}}$ adalah penduga maksimum lokal dari fungsi *likelihood* untuk parameter θ. Dalam model Poisson *hidden* Markov, parameter Poisson $\lambda_i, (i = 1,2,\dots,m)$ harus positif dan terbatas, tetapi titik stasioner yang konvergen pada algoritme EM belum tentu merupakan titik maksimum global, maka untuk mengidentifikasi titik yang maksimum global diperlukan penentuan titik awal.

a. Re-estimasi Parameter

Pada tahap ini, akan dimaksimumkan peluang observasi $P(Y_1 = y_1, Y_2 = y_2, \dots, Y_T = y_T)$ untuk memperoleh nilai parameter model Poisson *hidden* Markov $\theta = (m, \Gamma, \lambda)$ yang dapat dengan baik mendeskripsikan rangkaian observasi yang terjadi. Formula re-estimasi dapat diperoleh dengan cara memaksimumkan nilai $S(\theta,\hat{\theta}_{\boldsymbol{k}}), \theta \in \Theta$ untuk setiap iterasinya, sehingga akan diperoleh nilai *likelihood* yang tak turun. Fungsi $S(\theta,\hat{\theta}_{\boldsymbol{k}})$ pada langkah E pada iterasi ke-$(k+1)$ pada algoritme EM adalah

$$\begin{aligned} S(\theta,\hat{\theta}_{\boldsymbol{k}}) &= E_{\hat{\theta}_{\boldsymbol{k}}}(\log L_T^C(\theta)|\mathcal{Y}) \\ &= \sum_{i=1}^{m} \frac{{\alpha_1}^{(k)}(i){\beta_1}^{(k)}(i)}{\sum_{l=1}^{m} {\alpha_t}^{(k)}(l){\beta_t}^{(k)}(l)} \log\delta_i \\ &\quad + \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\sum_{t=1}^{T-1} {\alpha_t}^{(k)}(i){\gamma_{ij}}^{(k)}{\pi_{y_{t+1}j}}^{(k)}{\beta_{t+1}}^{(k)}(j)}{\sum_{l=1}^{m} {\alpha_t}^{(k)}(l){\beta_t}^{(k)}(l)} \log\gamma_{ij} \\ &\quad + \sum_{i=1}^{m} \frac{\sum_{t=1}^{T} {\alpha_t}^{(k)}(i){\beta_t}^{(k)}(i)}{\sum_{l=1}^{m} {\alpha_t}^{(k)}(l){\beta_t}^{(k)}(l)} \log\pi_{y_t i}\,. \end{aligned} \tag{9}$$

Penduga maksimum *likelihood* γ_{ij} yang didapat pada iterasi ke-$(k+1)$ dengan algoritme EM, yaitu

$$\widehat{\gamma_{ij}}^{(k+1)} = \frac{\sum_{t=1}^{T-1} \alpha_t^{(k)}(i)\gamma_{ij}^{(k)}\pi_{y_{t+1}j}^{(k)}\beta_{t+1}^{(k)}(j)}{\sum_{t=1}^{T-1} \alpha_t^{(k)}(i)\beta_t^{(k)}(i)}, \tag{10}$$

untuk setiap *state* i dan *state* $j, j \neq i$ pada rantai Markov $\{C_t\}_{t\in\mathbb{N}}$, sedangkan penduga maksimum *likelihood* λ_i yang didapat pada iterasi ke-$(k+1)$, yaitu

$$\widehat{\lambda_i}^{(k+1)} = \frac{\sum_{t=1}^{T} \alpha_t^{(k)}(i)\beta_t^{(k)}(i)y_t}{\sum_{t=1}^{T} \alpha_t^{(k)}(i)\beta_t^{(k)}(i)}, \tag{11}$$

untuk setiap *state* i pada rantai Markov $\{C_t\}_{t\in\mathbb{N}}$.

Setelah mencari nilai dugaan parameter Γ dan λ, paramater lain yang perlu diduga adalah banyaknya *state* rantai Markov yang tidak diamati, yakni nilai $m = m^*$. Menurut Leroux & Puterman [4] nilai m^* yang dipilih adalah nilai m yang memaksimumkan $\log L_T^{(m)}(\theta) - a_{m,T}$, di mana $\log L_T^{(m)}(\theta)$ adalah nilai fungsi *loglikelihood* model Poisson *Hidden* Markov m *state* dan $a_{m,T}$ adalah penalti yang bergantung pada m *state* dan T observasi.

Pada penelitian ini digunakan kriteria pemilihan nilai m^*, yakni AIC (*Akaike Information Criterion*). Pada kriteria AIC, $a_{m,T} = d_m$, di mana d_m adalah banyaknya parameter yang diduga pada algoritme EM, yakni $m^2 + m$ sehingga

$$AIC = \log L_T^{(m)}(\theta) - (m^2 + m)\,. \tag{12}$$

Berdasarkan nilai AIC, maka akan diperoleh nilai m yang optimum untuk pendugaan parameter model. Menggunakan nilai m yang optimum dan nilai parameter model yang bersesuaian akan dibangkitkan barisan data dugaan $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T$. Akurasi model diukur menggunakan MAPE.

$$\begin{aligned}\hat{Y}_T &= E\left[Y_t = y_t | \mathcal{Y}_{t=1}^{t-1}; \hat{\theta}\right] \\ &= \sum_{Y_t} Y_t P\left(Y_t = y_t | \mathcal{Y}_{t=1}^{t-1}; \hat{\theta}\right) \\ &= \sum_{i=1}^{m} E\left(Y_t = y_t | C_t = i; \mathcal{Y}_{t=1}^{t-1}; \hat{\theta}\right) P(C_t = i) = \sum_{i=1}^{m} \lambda_i\, \delta_i\end{aligned}$$

b. Akurasi Model

Mean Absolute Percentage Error (MAPE) adalah rataan persentase kesalahan absolut pada tiap periode dibagi dengan nilai observasi yang nyata untuk periode tersebut. Dalam Hu & Ho [2], rumus MAPE adalah sebagai berikut

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\% \tag{13}$$

dengan $\hat{Y}_t$ merupakan nilai data dugaan pada waktu t, sedangkan Y_t merupakan nilai data observasi pada waktu t dan T menunjukkan banyaknya data.

Menurut Lewis [4], interpretasi dari nilai MAPE dapat dilihat dari TABEL 1.

Tabel 1
Interpretasi Nilai MAPE

MAPE (%)	Interpretasi
<10	Peramalan yang sangat akurat
10-20	Peramalan yang baik
20-50	Peramalan yang layak
>50	Peramalan yang tidak akurat

4. PEMODELAN POISSON *HIDDEN* MARKOV

4.1 Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data banyaknya kasus kecelakaan yang terjadi di sepanjang jalan tol Jakarta-Cikampek selama kurun waktu Januari 2013 sampai Desember 2014. Periode waktu observasi yang digunakan pada data ini adalah minggu, sehingga terdapat 104 barisan observasi dengan rataan sebesar 11.7019 dan ragamnya sebesar 13.9005. Jadi, data observasi yang diperoleh bersifat overdispersi. Data yang digunakan pada penelitian ini diperoleh dari PT. Jasamarga.

4.2 Pemodelan Poisson *Hidden* Markov Pada Data Banyaknya Kecelakaan di Jalan Tol Jakarta-Cikampek

Barisan data banyaknya kecelakaan pada jalan tol Jakarta-Cikampek diasumsikan dibangkitkan oleh model Poisson *hidden* Markov. Diasumsikan barisan data banyaknya kecelakaan dibangkitkan berdasarkan penyebab kejadian yang membentuk rantai Markov yang tidak diamati secara langsung. Faktor-faktor penyebab terjadinya kecelakaan diasumsikan sebagai suatu *state* dari suatu rantai Markov $\{C_t\}_{t\in\mathbb{N}}$. Barisan data banyaknya kecelakaan merupakan data observasi $\{Y_t\}_{t\in\mathbb{N}}$ di mana sebaran bersyarat $Y_t(t \in \mathbb{N})$ jika diketahui $C_t(t \in \mathbb{N})$ berada pada *state* $i(i \in Q_C)$ adalah Poisson (λ_i). Jadi, pasangan $\{C_t, Y_t\}_{t\in\mathbb{N}}$ merupakan model Poisson *hidden* Markov.

Data banyaknya kecelakaan di jalan tol Jakarta-Cikampek merupakan barisan observasi dengan periode observasi per minggu dengan t menunjukkan minggu ke-t, Y_t merupakan banyaknya kecelakaan pada minggu ke-t sehingga Y_t bersifat diskret. C_t merupakan faktor penyebab kecelakaan pada minggu ke-t, C_t adalah rantai Markov yang tidak diamati secara langsung. Banyaknya kecelakaan pada minggu ke-t bergantung pada faktor penyebab kecelakaan pada minggu ke-t, atau dengan kata lain Y_t bergantung pada C_t.

4.3 Hasil Komputasi

Dari algoritme pemrograman yang telah disusun sebelumnya, dibuat program komputasi menggunakan *software Mathematica* 10.0. Pada *software Mathematica* 10.0, telah tersedia paket untuk pemodelan *Hidden* Markov, sehingga memudahkan untuk proses perhitungan pendugaan parameter model yang akan dicari.

Pendugaan parameter model dilakukan mulai dari *state* 2 sampai dengan *state* 5. Berdasarkan hasil program, diperoleh nilai *loglikelihood* dan nilai *Mean Absolute Percentage Error* (MAPE) dari setiap *state*, yang dapat dilihat pada Tabel 2.

Tabel 2

Nilai *Loglikelihood* dan nilai AIC dari setiap *state*

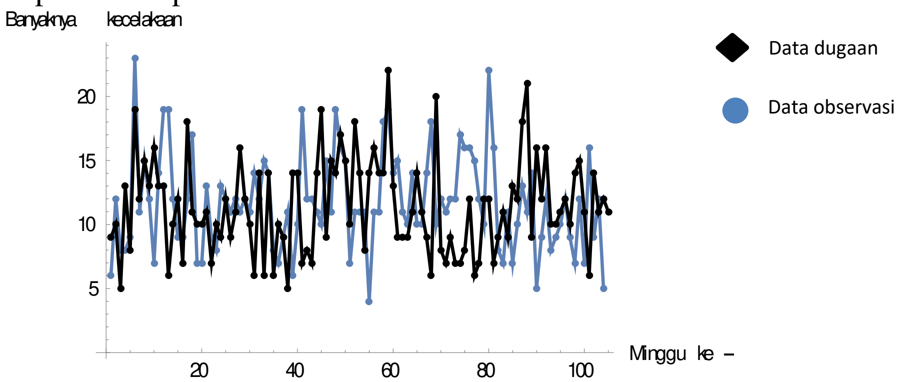
m *state*	Nilai *Loglikelihood*	AIC
2	-279.295	$\mathbf{-285.295}$
3	-278.151	-290.151
4	-277.592	-297.592
5	-276.618	-306.618
6	-276.503	-318.503

Berdasarkan nilai yang ditunjukkan pada TABEL 2, nilai *loglikelihood* semakin meningkat sesuai dengan penambahan jumlah *state*. Namun, untuk kriteria model yang terbaik didasarkan pada nilai AIC. Model Poisson *hidden* Markov dengan 2 *state* atau $m = 2$ memiliki nilai AIC terbesar dibandingkan dengan *state* lainnya, sehingga untuk nilai $m = 2$ merupakan nilai m yang optimum untuk pendugaan parameter model.

Dari nilai pendugaan parameter untuk $m = 2$, diperoleh nilai dugaan parameter sebagai berikut

$$\hat{\Gamma} = \begin{bmatrix} 0.8479 & 0.1521 \\ 0.1674 & 0.8326 \end{bmatrix}, \hat{\delta} = \begin{bmatrix} 0.9999 \\ 2.4417 \times 10^{-46} \end{bmatrix}, \hat{\lambda} = \begin{bmatrix} 10.0968 \\ 13.5148 \end{bmatrix}$$

Dari parameter yang didapat akan dibangkitkan barisan data dugaan, yang dapat dilihat pada Gambar 2.



Gambar 2 Perbandingan Data Dugaan Model Poisson *Hidden* Markov 2 *State* dengan Data Observasi

4.4 Prediksi Banyaknya Kecelakaan di Jalan Tol Jakarta-Cikampek

Setelah melakukan pemodelan terhadap data banyaknya kecelakaan, berikutnya akan dilakukan prediksi. Proses prediksi dilakukan dalam 2 tahap, tahap pertama melakukan prediksi untuk data observasi dan tahap kedua melakukan prediksi untuk periode data yang akan datang. Data hasil prediksi diperoleh dengan membangkitkan data berdasarkan nilai parameter model terbaik dan dihitung nilai kesalahannya dengan MAPE. Nilai kesalahan dan jumlah *state* terbaik dari masing-masing periode prediksi dapat dilihat pada Tabel 3.

Tabel 3
Nilai kesalahan dari data prediksi untuk setiap periode prediksi

Periode Prediksi	**MAPE (%)**
1 minggu ke depan	34.0786
2 minggu ke depan	34.3781
3 minggu ke depan	34.5761
4 minggu ke depan	34.8461

5. SIMPULAN

Model Poisson *hidden* Markov digunakan untuk memodelkan banyaknya kecelakaan yang terjadi per minggu di jalan tol Jakarta-Cikampek dari bulan Januari 2013 sampai dengan bulan Desember 2014. Berdasarkan nilai AIC diperoleh model Poisson *hidden* Markov 2 *state* sebagai model terbaik. Menggunakan model terbaik dilakukan prediksi banyaknya kecelakaan sampai empat periode ke depan dan diperoleh nilai MAPE berkisar 34%. Dari nilai MAPE yang diperoleh mengindikasikan bahwa hasil prediksi banyaknya kecelakaan termasuk peramalan yang layak, tapi belum cukup baik.

DAFTAR PUSTAKA

[1] Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm, *Journal Royal Statistical Society* Series B. 39: 1-38.

[2] Hu TY, Ho WM. 2010. Travel time prediction for urban networks : the comparisons of simulation-based and time-series models, *Proceedings 17th ITS World Congress*. 1-11.

[3] Leroux BG, Puterman ML. 1992. Maximum penalized likelihood estimation for independent and Markov dependent mixture models. *Biometrics* 48: 545-558.

[4] Lewis CD. 1982. *Industrial and Business Forecasting Methods*. London: Butterworths.

[5] Paroli R, Spezia L. 1999. Gaussian hidden Markov models : parameters estimation and applications to air pollution data. *Serie E. P. N. 94* Instituto di Statistica, Universita Catollica S. C Milano.

[6] Paroli R, Redaelli G, Spezia L. 2000. Poisson hidden Markov models for time series of overdispersed insurance counts. *Astin Colloquium.* 461-474.