ID: 20

Kajian Model *Hidden Markov* dengan Empat *State*

Study of Hidden Markov Models with Four States

Rini Cahyandari[1*], Elis Asri NF[2], Aep Saepuloh[3]
[1,2,3] Department of Mathematics, UIN Sunan Gunung Djati Bandung, Indonesia
A.H Nasution 105 Cibiru Bandung
rini_cahyandari@uinsgd.ac.id[1*]

Abstrak – *Proses stokastik adalah himpunan variabel acak* $\{X(t), t \in T\}$. *Dimana state pada masa yang akan datang dari proses itu tidak tergantung pada masa yang telah lalu dan hanya tergantung pada masa sekarang saja, proses ini disebut proses Markov. Sebuah proses Markov dengan ruang parameter yang diskrit yang berada pada suatu keadaan yang diskrit, proses ini disebut rantai Markov. Hidden Markov model merupakan sebuah metode statistik berdasarkan model Markov sederhana yang memodelkan sistem serta membaginya dalam dua state yaitu state tersembunyi dan state terobservasi dimana pada dasarnya hidden Markov model keadaannya tidak dapat terlihat langsung meskipun parameter model diketahui model tersebut tetap tersembunyi. Kajian model hidden Markov dengan empat state dimana pengamatan yang pernah ada mengenai hidden Markov model pada kartu kredit dengan data yang berbeda, dari keempat kelompok profil pengeluaran menggunakan algoritma K-Means clustering peluang proses berada pada rendahnya pengeluaran kartu kredit yang dikaji secara mendalam dalam suatu rangkaian proses dan kejadian, sehingga analisa kejadian-kejadian di waktu-waktu mendatang dapat diketahui secara matematis.*

Kata Kunci*: Proses stokastik, proses Markov, rantai Markov, hidden Markov model*

Abstract – *A stochastic process is a set of random variables* $\{X(t), t \in T\}$. *Where the future state of the process does not depend on the past and only depends on the present, this process is called a Markov process. A Markov process with a discrete parameter space that exists in a discrete state, this process is called a Markov chain. The hidden Markov model is a statistical method based on a simple Markov model which models the system and divides it into two states, namely the hidden state and the observed state, where basically the hidden Markov model cannot be seen directly even though the model parameters are known to be hidden. Study of the hidden Markov model with four states where previous observations regarding the hidden Markov model on credit cards with different data, from the four groups of spending profiles using the K-Means clustering algorithm, the opportunity for the process is low credit card spending which is studied in depth in a series processes and events, so that the analysis of future events can be known mathematically.*

Keywords: *Stochastic process, Markov process, Markov chain, hidden Markov model.*

1. Pendahuluan

Proses stokastik adalah himpunan variabel acak $\{X(t), t \in T\}$, dimana langkah *state* pada masa yang akan datang dari proses ini tidak tergantung pada masa yang telah lalu dan hanya tergantung pada masa sekarang saja, proses ini disebut proses Markov. Sebuah proses Markov dengan ruang parameter yang diskrit yang berada pada suatu keadaan yang diskrit, proses ini disebut rantai Markov. *Hidden Markov model* merupakan sebuah metode statistik berdasarkan model Markov sederhana yang memodelkan sistem serta membaginya dalam dua *state* yaitu *state* tersembunyi dan *state* terobservasi dimana pada dasarnya *hidden Markov model* keadaannya tidak dapat terlihat langsung meskipun parameter model diketahui model tersebut tetap tersembunyi.

Kajian *hidden Markov model* pengamatan tentu akan dikaji secara mendalam dalam suatu rangkaian proses dan kejadian, sehingga analisa kejadian-kejadian di waktu-waktu mendatang dapat diketahui secara matematis.

Penelitian ini melanjutkan penelitian [6] dengan data yang berbeda, beberapa metode digunakan dalam *hidden Markov model* mulai dari algoritma *Forward-Backward*, *Viterbi*, dan *Baum-Welch*.

2. Metode Penelitian

Kajian model *hidden Markov* dengan empat *State* disini membahas teori dasar proses stokastik, proses Markov, rantai Markov, *hidden Markov model* beserta aplikasi algoritma *Forward-Backward*, *Viterbi*, dan *Baum-Welch*.

Algoritma *forward* didefinisikan variabel *forward*

$$\alpha_t(i) = \left(O_1 = v_{m_1}, \dots, O_T = v_{m_t}, Q_t = s_i \middle| \lambda\right)$$

yaitu peluang barisan observasi $O_1, \dots, O_T$ dan *state* s_i pada waktu t jika diberikan λ. Secara induktif $P(O|\lambda)$ dapat dihitung dengan algoritma sebagai berikut:

a. Inisialisasi

$$\alpha_t(i) = \pi_i . b_i(O_1),\ 1 \le i \le N \quad (1)$$

b. Induksi

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i). a_{ij}\right]. b_j(O_{t+1}),\ 1 \le t \le T-1,\ 1j \le N \quad (2)$$

c. Akhir

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \quad (3)$$

Algoritma *backward* didefinisikan sebagai,

$$\beta_t(i) = P(O_{t+1}, \dots, O_T | Q_t, \lambda)$$

yaitu peluang dari barisan observasi parsial $O_{t+1}, \dots, O_T$ diberikan *state* s_i pada waktu t dan model λ. Selanjutnya, $\beta_t(i)$ dapat diselesaikan secara induksi sebagai berikut:

a. Inisialisasi

$$\beta_t(i) = 1\ , 1 \le i \le N \quad (4)$$

b. Induksi

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j),\ \ t = T-1, \dots, 1,\ 1 \le i \le N \quad (5)$$

c. Terminasi

$$P(O|\lambda) = \sum_{i=1}^{N} b_i(1) \pi(i) \beta_1(i) \quad (6)$$

Algoritma viterbi digunakan dua variabel bantu, yaitu:

1) $\delta_t(i) = \max P\left(Q_1 = s_{1_t}, ., Q_t = s_i, O_1 = v_{m_1}, \dots, O_T = v_{m_t} \middle| \lambda\right)$
dengan induksi diperoleh :
$\delta_{t+1}(j) = \max \left[\delta_t(i). a_{ij}\right] b_j(O_{t+1})$

2) $\Psi_t(i) = \arg\max P\left(Q_1 = s_{1_t}, ., Q_t = s_i, O_1 = v_{m_1}, \dots, O_T = v_{m_t} \middle| \lambda\right)$

ISSN (p): 2985 - 4903
ISSN (e): 2986 - 2477

Langkah-langkah dalam algoritma viterbi untuk menentukan barisan *state* terbaik yaitu:

a) Inisialisasi

Set t = 2

$$\delta_1(i) = \pi_i . b_i(O_1),\ 1 \le i \le N \tag{7}$$

$$\Psi_t(i) = 0$$

b) Induksi

$$\delta_t(j) = b_j(O_t) \max_{1 \le i \le N} [\delta_t(i). a_{ij}], 1 \le j \le N \tag{8}$$

$$\Psi_t(j) = \underset{1 \le i \le N}{\operatorname{argmax}} [\delta_t(i). a_{ij}], 1 \le j \le N$$

c) *Update t*

Set t = t+1

Kembali ke langkah (b) jika $t \le T$.

Lainnya, masuk ke langkah (d).

d) Terminasi

$$P^* = \max_{1 \le j \le N} [\delta_T(j)] \tag{9}$$

$$Q^* = \underset{1 \le j \le N}{\operatorname{argmax}} [\delta_T(j)]$$

e) Barisan *state* terbaik digunakan proses *backtracking*

Inisialisasi

Set *t* = T-1

Backtracking

$$Q_T^* = \Psi_{t+1}(Q_{t+1}^*) \tag{10}$$

Update t

Set *t = t-1*

Kembali ke langkah (1) jika $t \ge T$.

Untuk menyelesaikan masalah algoritma *Baum-Welch*, didefinisikan empat variabel, yaitu variabel *forward*, *backward*, $\xi_t(i,j)$ dan $\gamma_t(i)$. Variabel *forward* dan *backward* akan digunakan dalam perhitungan variabel $\xi_t(i,j)$ dan $\gamma_t(i)$. $\xi_t(i,j)$ didefinisikan sebagai

$$\xi_t(i,j) = P(Q_t, Q_{t+1} | O, \lambda)$$

yaitu peluang proses pada saat t berada pada *state* s_i dan pada saat t+1 berada pada *state* s_j, jika diberikan barisan O dan model λ. Persamaan $\beta_t(i)$ bisa ditulis:

$$\xi_t(i,j) = \frac{P(Q_t = s_i, Q_{t+1} = s_j, O | \lambda)}{P(O|\lambda)}$$

atau dapat diekspresikan dengan variabel *forward-backward* sebagai

$$\xi_t(i,j) = \frac{\alpha_t(i).\beta_{t+1}(j).b_j(O_{t+1}).a_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i).a_{ij}.\beta_{t+1}(j)b_j(O_{t+1})} \tag{11}$$

variabel keempat yakni variabel $\gamma_t(i)$ didefinisikan sebagai:

$$\gamma_t(i) = P(Q_t | O, \lambda)$$

yaitu peluang proses berada pada *state* s_i saat waktu t, jika diberikan barisan observasi O, dan model λ. Persamaan $\gamma_t(i)$ dapat diekspresikan dengan variabel *forward-backward* sebagai berikut:

ISSN (p): 2985 - 4903
ISSN (e): 2986 - 2477

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)}$$

3. Hasil dan Pembahasan

Studi kasus yang akan dibahas seperti pada [6] dengan menerapkan *hidden Markov model* pada kartu kredit dengan data yang berbeda. Misalkan terdapat sebuah penelitian di sebuah bank dengan menggunakan kartu kredit yang mengangkat metode *hidden Markov model* untuk mendeteksi penyalahgunaan kartu kredit, dimana data yang diambil sebanyak 45 data observasi dan data dikelompokkan menjadi 4 kelompok profil pengeluaran dengan algoritma *K-Means Clustering*, yaitu rendah (0, Rp. 200.000), sedang (Rp. 200.000, Rp. 500.000), tinggi (Rp. 500.000, Rp. 1.000.000), dan tinggi sekali (Rp. 1.000.000, limit kartu kredit). Kemudian, data input diubah menjadi 1 = rendah, 2 = sedang, 3 = tinggi dan 4 = tinggi sekali. Penyebab kejadian besarnya pengeluaran kartu kredit seseorang disebabkan oleh suasana hati pemegang kartu kredit yaitu selau berbelanja, sering berbelanja, kadang-kadang berbelanja dan tidak pernah berbelanja.

Langkah yang pertama dalam menyelesaikan kasus ini diantaranya adalah:

1. Menentukan nilai N (jumlah *state* dalam model).

Keadaan tersembunyi yaitu selalu, sering, kadang-kadang dan tidak pernah. sehingga nilai N = 4.

2. Menentukan nilai M (jumlah pengamatan setiap *state*).

Keadaan terobservasi yaitu rendah, sedang, tinggi, dan tinggi sekali, sehingga nilai M = 4.

3. Menentukan nilai *Probability Distribution Vector* (π) atau nilai himpunan distribusi awal $\pi = \pi(i)$.

$$\pi = \begin{bmatrix} 10/45 \\ 10/45 \\ 12/45 \\ 13/45 \end{bmatrix} = \begin{bmatrix} 0{,}22 \\ 0{,}22 \\ 0{,}26 \\ 0{,}28 \end{bmatrix}$$

4. Menentukan nilai *Transition Probability Matrix* atau matriks peluang transisi $(A = a_{ij})$ yang akan menghasilkan matriks A dengan ordo $N \times N$.

$$\text{A} = \begin{bmatrix} 0{,}5 & 0{,}2 & 0{,}1 & 0{,}2 \\ 0{,}6 & 0{,}1 & 0{,}2 & 0{,}1 \\ 0{,}3 & 0{,}2 & 0{,}1 & 0{,}4 \\ 0{,}2 & 0{,}3 & 0{,}4 & 0{,}1 \end{bmatrix}$$

5. Menentukan nilai *Emission Probability Matrix* atau nilai matriks peluang bersyarat $(B = b_i(k))$ yang akan menghasilkan matriks B dengan ordo $N \times M$.

$$\text{B} = \begin{bmatrix} 0{,}1 & 0{,}2 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}3 & 0{,}4 & 0{,}1 \\ 0{,}4 & 0{,}3 & 0{,}1 & 0{,}2 \\ 0{,}6 & 0{,}1 & 0{,}1 & 0{,}2 \end{bmatrix}$$

6. Menghitung peluang observasi dengan menggunakan algoritma *Forward-Backward.*

ISSN (p): 2985 - 4903
ISSN (e): 2986 - 2477

Tabel 1. Hasil Perhitungan Algoritma *Forward* $(\alpha_t(i))$

T	1	2	3	4
$\alpha_t(1)$	0,021	0,02106	0,002404	0,001155
$\alpha_t(2)$	0,046	0,00846	0,002586	0,01248
$\alpha_t(3)$	0,1	0,01914	0,004442	0,0001373
$\alpha_t(4)$	0,186	0,01348	0,008437	0,00336

$$P(O = selalu, sering, kadang - kadang\ dan\ tidak\ pernah|\lambda) = \sum_{i=1}^{N} \alpha_t(i) = 0{,}01$$

Tabel 2. Hasil Perhitungan Algoritma *Backward* $(\beta_t(i))$

T	4	3	2	1
$\beta_t(1)$	1	0,21	0,0581	0,221148
$\beta_t(2)$	1	0,22	0,0478	0,012122
$\beta_t(3)$	1	0,19	0,0851	0,011216
$\beta_t(4)$	1	0,26	0,0634	0,011834

$$P(O|\lambda) = \sum_{i=1}^{N} b_i(1)\pi(i)\beta_1(i) = 0{,}01$$

Sehingga peluang selalu, sering, kadang-kadang dan tidak pernah dalam penggunaan kartu kredit adalah 0,01.

7. Menentukan barisan keadaan tersembunyi dengan menggunakan algoritma *Viterbi.*

Tabel 3. Hasil Perhitungan $\xi_t(i,j)$

T	1
$\{ \xi_t(1,1)\}$	0,012782
$\{ \xi_t(1,2)\}$	0,002103
$\{ \xi_t(1,3)\}$	0,003744
$\{ \xi_t(1,4)\}$	0,005579
$\{ \xi_t(2,1)\}$	0,025754
$\{ \xi_t(2,2)\}$	0,006078
$\{ \xi_t(2,3)\}$	0,031069
$\{ \xi_t(2,4)\}$	0,021258

8. Menaksir parameter *hidden Markov model* dengan menggunakan *algoritma Baum-Welch*.

Tabel 4. Hasil Perhitungan $\gamma_t(i)$

T	1
$\gamma_t(1)$	0,024208
$\gamma_t(2)$	0,084159

Menggunakan hasil dari perhitungan-perhitungan tersebut dapat dicari penaksir parameter HMM yaitu $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$. Penaksiran inilah yang nantinya akan menghasilkan $P(O|\hat{\lambda}) \geq P(O|\lambda)$. Berikut hasil perhitungan untuk penaksir parameter *hidden Markov model*:

$$\hat{\pi} = \begin{bmatrix} \gamma_1(1) \\ \gamma_2(2) \end{bmatrix} = \begin{bmatrix} 0{,}024208 \\ 0{,}084159 \end{bmatrix}$$

ISSN (p): 2985 - 4903
ISSN (e): 2986 - 2477

Nilai ini merupakan taksiran peluang awal, artinya agar nilai $P(O|\hat{\lambda}) \geq P(O|\lambda)$ terpenuhi, maka peluang proses berada pada rendahnya pengeluaran kartu kredit adalah 0,024208 dan 0,084159.

$$\hat{A} = \begin{bmatrix} \frac{\sum_{t=1}^{T} \xi_t(1,1)}{\sum_{t=1}^{T} \gamma_t(1)} & \frac{\sum_{t=1}^{T} \xi_t(1,2)}{\sum_{t=1}^{T} \gamma_t(1)} \\ \frac{\sum_{t=1}^{T} \xi_t(2,1)}{\sum_{t=1}^{T} \gamma_t(2)} & \frac{\sum_{t=1}^{T} \xi_t(2,2)}{\sum_{t=1}^{T} \gamma_t(2)} \end{bmatrix} = \begin{bmatrix} 0{,}663 & 0{,}105 \\ 0{,}187 & 0{,}278 \\ 0{,}128 & 0{,}075 \\ 0{,}383 & 0{,}265 \end{bmatrix}$$

Matriks tersebut merupakan penaksir untuk matriks transisi A. Matriks $\hat{A}$ menggambarkan bahwa untuk mencapai nilai $P(O|\hat{\lambda}) \geq P(O|\lambda)$ maka peluang transisi dari keadaan banyak sedikitnya pengeluaran yang menggunakan kartu kredit selalu ke selalu adalah sebesar 0,663 ,selalu ke sering adalah sebesar 0,105,sering ke selalu adalah sebesar 0,187, sering ke selalu adalah sebesar 0,278, kadang-kadang ke selalu adalah sebesar 0,128, kadang-kadang ke sering adalah sebesar 0,075, tidak pernah ke selalu adalah sebesar 0,383 dan tidak pernah ke sering adalah sebesar 0,265.

4. Kesimpulan

Kajian *hidden Markov model* dengan empat *state* dari keempat kelompok profil dapat disimpulkan bahwa penyalahgunaan kartu kredit rendah sehingga penyalahgunaan kartu kredit terkontol.

Referensi

[1] Jamaludin, A. 2015. *Pengenalan Lafal Hukum Nun Mati atau Tanwin Menggunakan Hidden Markov Model*. Skripsi. UIN Sunan Gunung Djati Bandung.

[2] Koosasi H A, Sarno R, dan Munif A. 2017. *Deteksi Fraud Menggunakan Metode Model Markov Tersembunyi Pada Proses Bisnis*. Jurnal Teknik Informatika Institut Teknologi Sepuluh Nopember Surabaya, Vol.6 , No.1. ISSN:2337-3539.

[3] Mahmudi dan Ardi. 2016. *Prediksi Pergerakan Nilai Tukar Rupiah Terhadap Dolar Amerika Serikat Menggunakan Hidden Markov Model (HMM)*. Jurnal Program Studi Matematika. Fakultas Sains dan Teknologi. Universitas Islam Negeri Syarif Hidayatullah Jakarta.

[4] Cahyandari, R dan Nursolihah, R. 2015. *Penerapan Model Markov Tersembunyi untuk Mengetahui Persentase Kecocokan dari Deoxyribonucleic Acid pada Pohon Filogenetik Ursidae (Beruang)*. Statistika, Vol.15, No.2, p 73-86.

[5] Ross, S M.1996. *Stochastic Processes*. United State Of America ; John Wiley and Sons, Inc.

[6] Utari P, B Setiawaty , dan N K K Ardana. 2012. "Aplikasi Model Hidden Markov Diskret untuk Mendeteksi Penyalahgunaan Kartu Kredit". Institute Pertanian Bogor. JMA, Vol.11, No.1, p.21-30.

[7] Walpole, R E dan Raymond H M.1995. Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan, Edisi Keempat, terjemahan RK Sembiring. Bandung : Penerbit ITB.

[8] W. Jones, Peter and Smith, Peter. 2010. Stochastic Processes An Introduction (Second Edition). CRC Press

ISSN (p): 2985 - 4903
ISSN (e): 2986 - 2477