PROSES KEPUTUSAN MARKOV DENGAN METODE *THE POLICY IMPROVEMENT ALGORITHM*

(Skripsi)

Oleh
Nafisatutaliah

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

ABSTRACT

MARKOV DECISION PROCESSES BY THE POLICY IMPROVEMENT ALGORITHM METHOD

By

Nafisatutaliah

Markov Decision Process is the decision making technique in the optimal policy selection of several states and decisions given where the cost incurred is considered. One of the methods can be used in the markov decision processes is The Policy Improvement Algorithm Method which evaluates each states and decisions in the process of policy improvement. The method can be used after it is known that the markov decisions processes has a recursive equation that satisfies the assumption to be a steady-states probability to obtain an equation approximation for the step of value determination to be used in the process of policy improvement. Application in case studies is needed to better understand that the policy optimal selection by the method can applied to the same of transition probability and cost at each decisions or the differents of transition probability and cost at each decisions, and also the same transition probability with different costs at each decisions.

Keywords: Markov Decision Processes, The Policy Improvement Algorithm, The Optimal Policy

ABSTRAK

PROSES KEPUTUSAN MARKOV DENGAN METODE *THE POLICY IMPROVEMENT ALGORITHM*

Oleh

Nafisatutaliah

Proses Keputusan Markov merupakan teknik pengambilan keputusan dalam pemilihan kebijakan yang optimal dari beberapa *state* dan keputusan yang diberikan dimana biaya yang terjadi dipertimbangkan. Salah satu metode yang dapat digunakan dalam proses keputusan Markov adalah metode *The Policy Improvement Algorithm* yang mengevaluasi setiap *state* dan keputusan dalam proses perbaikan kebijakan. Metode tersebut dapat digunakan setelah diketahui bahwa proses keputusan Markov memiliki persamaan rekursif yang memenuhi asumsi untuk menjadi peluang *steady-state* sehingga diperoleh aproksimasi persamaan untuk langkah penentuan nilai yang dipergunakan dalam proses perbaikan kebijakan. Penerapan dalam studi kasus diperlukan untuk lebih memahami bahwa pemilihan kebijakan optimal dengan metode tersebut dapat berlaku untuk peluang transisi dan biaya yang sama pada setiap keputusan atau peluang transisi dan biaya yang berbeda pada setiap keputusan, dan juga peluang transisi yang sama dengan biaya yang berbeda pada setiap keputusan.

Kata kunci: Proses Keputusan Markov, *The Policy Improvement Algorithm*, Kebijakan Optimal

PROSES KEPUTUSAN MARKOV DENGAN
METODE *THE POLICY IMPROVEMENT ALGORITHM*

Oleh

Nafisatutaliah

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA SAINS

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

Judul Skripsi : **PROSES KEPUTUSAN MARKOV DENGAN METODE *THE POLICY IMPROVEMENT ALGORITHM***

Nama Mahasiswa : Nafisatutaliah

Nomor Pokok Mahasiswa : 1317031056

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Dian Kurniasari, S.Si., M.Sc.	**Prof. Dra. Wamiliana, M.A., Ph.D.**
NIP 19690305 199603 2 001	NIP 19631108 198902 2 001

2. Ketua Jurusan Matematika

Prof. Dra. Wamiliana, M.A., Ph.D.
NIP 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Dian Kurniasari, S.Si., M.Sc.** ....

Sekretaris : **Prof. Dra. Wamiliana, M.A., Ph.D.** ....

Penguji
Bukan Pembimbing : **Prof. Drs. Mustofa Usman, M.A., Ph.D.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Warsito, S.Si., D.E.A., Ph.D.
NIP 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **02 Februari 2018**

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan di bawah ini:

Nama	: Nafisatutaliah
Nomor Pokok Mahasiswa	: 1317031056
Judul	: PROSES KEPUTUSAN MARKOV DENGAN METODE *THE POLICY IMPROVEMENT ALGORITHM*
Jurusan	: Matematika

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, Febuari 2018

Penulis,

Nafisatutaliah
NPM. 1317031056

RIWAYAT HIDUP

Penulis dilahirkan di Bandar Lampung pada tanggal 21 Juni 1995, sebagai anak pertama dari empat bersaudara, dari pasangan Bapak Mustaqim dan Ibu Muslihatun. Penulis telah menyelesaikan pendidikan di Sekolah Dasar (SD) Negeri 1 Sawah Lama pada tahun 2007, Sekolah Menengah Pertama (SMP) Negeri 1 Bandar Lampung pada tahun 2010, dan Madrasah Aliyah (MA) Negeri 2 Bandar Lampung pada tahun 2013.

Pada tahun 2013 penulis terdaftar sebagai Mahasiswa Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung (UNILA) melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN). Selama menjadi mahasiswa, penulis pernah mengikuti organisasi mahasisa tingkat Jurusan dan Fakultas, yaitu Anggota Gematika periode 2013-2014, Anggota muda ROIS periode 2013-2014, Anggota Garuda pada periode 2013-2014, Anggota Medinfo BEM FMIPA UNILApada 2014-2015, Wakil Bendahara Himpunan Mahasiswa Matematika (HIMATIKA) pada periode 2014-2015 dan Bendahara Umum HIMATIKA pada periode 2015-2016. Selain itu, pada tahun 2016 penulis telah melakukan Kerja Praktik (KP) di Kantor Badan Pusat Statistik Provinsi Lampung dan Kuliah Kerja Nyata (KKN) di Desa Gedung Ratu, Kecamatan Anak Ratu Aji, Kabupaten Lampung Tengah, Provinsi Lampung.

MOTTO

Always keep calm and be patient to fight the world

May be now blue, but future must be colourful

Dunia adalah proses, proses adalah pengalaman, dan pengalaman selalu menjadi pengajar terbaik dalam hidup

Always remember The Lord of Mankind and The Prophet of The Last

PERSEMBAHAN

Dengan penuh rasa syukur dan terima kasih yang dalam karya ini dipersembahkan untuk :

Allah SWT dan Nabi Muhammad SAW yang selalu menjadi penunjuk hidup dikala senang ataupun susah

Bapak, Mamak, dan adik-adik tercinta yang selalu menyemangati, mengingatkan, dan memahami kala duka maupun bahagia

Dosen-dosen pembimbing dan pembahas yang selalu penyabar

Teman-teman pejuang skripsi yang selalu ada dan menyemangati di setiap langkah

Almamater tercinta Universitas Lampung

SANWACANA

Puji syukur kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga Skripsi yang berjudul "Proses Keputusan Markov dengan Metode *The Policy Improvement Algorithm*" dapat diselesaikan dengan baik. Shalawat serta salam tidak lupa selalu disanjung agungkan kepada suri tauladan umat Islam Nabi Muhammad SAW.

Penulisan skripsi ini tentu masih jauh dari kata sempurna. Namun, semoga skripsi ini dapat bermanfaat bagi penulis, khususnya, maupun pembaca, pada umumnya. Terselesaikannya Skripsi ini tentu mendapatkan bantuan, kerjasama, dan dukungan berbagai pihak. Pada kesempatan kali ini ucapan terima kasih setulus hati diberikan kepada:

1. Ibu Dian Kurniasari, S.Si., M.Sc. selaku Dosen Pembimbing I yang telah banyak memberikan bimbingan, kritik, saran, dan dukungan dalam penyelesaian skripsi ini
2. Ibu Prof. Dra. Wamiliana, M.A., Ph.D. selaku Dosen Pembimbing II dan Ketua Jurusan Matematika Fakultas Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung yang telah banyak memberikan bimbingan, kritik, saran, dan dukungan dalam penyelesaian skripsi ini

3. Bapak Prof. Drs. Mustofa Usman, M.A., Ph.D. selaku Dosen Penguji yang telah banyak memberikan kritik dan saran yang mendidik
4. Bapak Drs. Rudi Ruswandi, M.Si. selaku pembimbing akademik yang telah memberikan saran dan pembelajaran selama di perkuliahan
5. Bapak Prof. Warsito, S.Si., D.E.A., Ph.D. selaku Dekan Fakultas Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung
6. Seluruh dosen, staff, dan karyawan Jurusan Matematika Fakultas Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung
7. Orang tua, serta adik-adik yang selalu mendoakan dan memberikan dukungan baik moril maupun materil
8. Teman tersayang Aul, Dita, Galuh, Lia, Hanifah, Imel, Nina, Rifa, dan Tiyas yang telah mendukung dan mendoakan
9. Teman-teman satu bimbingan skripsi yang telah memberi dukungan moril
10. Teman-teman seperjuangan Angkatan 2013 selama perkuliahan
11. Keluarga besar HIMATIKA FMIPA Universitas Lampung
12. Seluruh pihak yang telah berperan dalam penyelesaian skripsi ini yang tidak dapat disebutkan satu per satu.

Bandar Lampung, Febuari 2018

Penulis,

Nafisatutaliah

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang

Dalam perkembangannya, banyak permasalahan-permasalahan yang terjadi dalam berbagai bidang yang telah diselesaikan dengan menggunakan statistika. Salah satu kajian statistika yang saat ini telah banyak digunakan dan sedang berkembang ialah Rantai Markov. Rantai Markov merupakan suatu teknik matematika yang biasa digunakan untuk membuat suatu model berbagai macam sistem dengan mengamati transisi yang terjadi pada suatu kejadian. Oleh karena itu, suatu model Rantai Markov akan diperoleh untuk membantu menyelesaikan permasalahan Rantai Markov tersebut.

Model Rantai Markov, yang merupakan model suatu sistem yang mengamati perubahan-perubahan yang terjadi dari variabel dinamis, akan sangat membantu dalam pengambilan keputusan untuk memperkirakan kejadian di waktu mendatang dengan mengamati perubahan-perubahan yang terjadi pada saat ini. Keputusan yang diambil tentu akan berpengaruh terhadap kebijakan yang akan dipergunakan, misalkan dalam perusahaan. Suatu perusahaan tidak mungkin hanya memiliki satu kebijakan sehingga keputusan yang akan diambil tentu harus mempertimbangkan beberapa kebijakan dan risiko yang akan terjadi jika mengambil kebijakan tersebut. Akibatnya, terdapat banyak pilihan alternatif keputusan yang memerlukan teknik-teknik tertentu dalam pengambilan keputusan agar diperoleh hasil yang optimal. Proses keputusan Markov (*Markov Decision Process)* dapat menjelaskan model dinamika dalam pengambilan keputusan tersebut (Ustazila, 2014).

Pada proses keputusan Markov setiap langkah dipilih tindakan tertentu dan tindakan tersebut akan menimbulkan biaya sesuai dengan tindakan yang dipilih. Kemudian biaya yang optimal akan diperoleh jika kebijakan yang akan diambil juga optimal. Salah satu metode yang dapat digunakan agar memperoleh kebijakan yang optimal adalah metode *The Policy Improvement Algorithm*. Metode ini akan memperbaiki kebijakan secara iteratif dengan mempertimbangkan biaya minimum yang akan dikeluarkan hingga mencapai kondisi yang optimal. Penelitian ini akan membahas deskripsi konsep analisis proses keputusan Markov dengan menggunakan metode *The Policy Improvement Algorithm* dan penerapannya pada beberapa studi kasus untuk menentukan kebijakan optimal pada studi kasus tersebut.

1.2 Tujuan Penelitian

Adapun tujuan dari penulisan skripsi ini adalah

1. Mendeskripsikan konsep analisis proses keputusan Markov dengan metode *The Policy Improvement Algorithm*

2. Menentukan solusi kebijakan optimal dengan menggunakan metode *The Policy Improvement Algorithm* pada beberapa studi kasus.

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Menambah wawasan dan memperkaya literatur dalam bidang statistika terutama yang berhubungan dengan proses keputusan Markov
2. Mampu mengaplikasikan ilmu yang diperoleh di perkuliahan untuk memecahkan masalah yang ada dalam kehidupan sehari-hari.

II. TINJAUAN PUSTAKA

Pada bab ini akan dijabarkan beberapa istilah atau materi yang dipergunakan dalam penelitian dengan penjelasan sebagai berikut:

2.1 Proses Stokastik

Proses stokastik adalah kumpulan dari peubah acak $\{X_t\ ;\ t \in \mathbb{N}\}$, dimana t sebagai waktu, yang didefinisikan sebagai *state* dari proses pada waktu t. $\mathbb{N}$ disebut indeks dari proses atau ruang parameter, yang apabila $\mathbb{N}$ merupakan himpunan terbilang proses stokastik tersebut dapat dikatakan sebagai proses waktu-diskrit. Himpunan dari semua nilai yang mungkin pada peubah acak X_t dapat disebut sebagai ruang *state* dalam proses dan dinotasikan dengan S (Ross, 2007).

2.2 *State*

State adalah kondisi yang merupakan peubah acak X_t, dimana jika suatu peubah acak berada pada *state* tersebut maka dapat berpindah ke *state* lainnya.. Himpunan atau kumpulan dari *state-state* tersebut membentuk ruang *state* dan dinyatakan dengan S, dimana S = {0,1, 2, ..., M} (Cox dan Miller, 1965).

2.3 Rantai Markov

Rantai Markov adalah salah satu bentuk dari proses stokastik yang memenuhi sifat Markov, yaitu peluang kejadian atau peubah acak X pada waktu $t + 1$ hanya akan dipengaruhi oleh kejadian X pada waktu t dan tidak akan dipengaruhi oleh kejadian sebelum waktu t atau dapat dinyatakan dengan

$$P\{X_{t+1} = j | X_0 = h, \dots, X_t = i\} = P\{X_{t+1} = j | X_t = i\} \quad (2.3.1)$$

(Kijima, 1997).

Dengan kata lain, sifat Markov tersebut dapat dinyatakan sebagai peluang bersyarat terhadap suatu kejadian di masa mendatang yang tidak dipengaruhi oleh kejadian di masa lalu, tetapi hanya dipengaruhi oleh kejadian saat ini (Hillier dan Lieberman, 2001). Oleh karena itu, Rantai Markov biasa digunakan untuk membuat suatu model berbagai macam sistem dengan mengamati transisi yang terjadi pada suatu kejadian sehingga diperoleh model Rantai Markov.

Dalam teori probabilitas, Model Rantai Markov adalah model stokastik yang digunakan untuk memodelkan sistem yang berubah-ubah secara random di mana diasumsikan bahwa kondisi masa depan tergantung hanya pada keadaan sekarang dan bukan pada urutan peristiwa yang mendahuluinya (mengasumsikan properti Markov). Umumnya, asumsi ini memungkinkan penalaran dan perhitungan, yang jika menggunakan model lain mungkin akan lebih sulit diselesaikan (Zada, 2016).

Berikut merupakan berbagai istilah yang dipergunakan dan berkaitan dengan Rantai Markov:

2.3.1 Peluang Transisi Rantai Markov

Peluang bersyarat $P\{X_{n+1} = j | X_t = i\}$ disebut sebagai peluang transisi dan dinotasikan dengan $p_{i,j}$ dimana untuk setiap i dan j,

$$P\{X_{n+1} = j | X_n = i\} = P\{X_1 = j | X_0 = i\} \qquad ; n = 0,1, \dots \quad (2.4.1)$$

disebut sebagai peluang transisi (satu langkah), dan untuk setiap i, j, dan n (t = 0,1,2,...),

$$P\{X_{n+t} = j | X_t = i\} = P\{X_t = j | X_0 = i\} \qquad ; t = 0,1, \dots \quad (2.4.2)$$

akan disebut peluang transisi n-langkah.

Karena $p_{i,j}$ merupakan peluang bersyarat maka peluang tersebut harus positif dan karena dalam prosesnya terjadi transisi ke beberapa *state* maka peluang bersyarat $p_{i,j}$ tersebut harus memenuhi sifat

(1) $p_{i,j}{}^{(n)} \geq 0 \qquad ; i, j \in S; \ n = 0,1,2, \dots$ (2.4.3)

(2) $\sum_{j \in S} p_{i,j}{}^{(n)} = 1 \quad ; i \in S; \ n = 0,1,2, \dots$ (2.4.4)

dengan bentuk matriks dari peluang transisi sebagai berikut

$$p_{i,j}{}^{(n)} = \begin{bmatrix} P_{0,0}{}^{(n)} & \cdots & P_{0,M}{}^{(n)} \\ \vdots & \ddots & \vdots \\ P_{M,0}{}^{(n)} & \cdots & P_{M,M}{}^{(n)} \end{bmatrix}$$

untuk n = 0, 1, 2, ... (Hillier dan Lieberman, 2001).

2.3.2 *Irreducible*

Suatu Rantai Markov dapat dikatakan *irreducible* jika ruang *state*-nya membentuk satu kelas komunikasi dinamakan *irreducible*, sebaliknya dinamakan *reducible* (Privault, 2013).

2.3.3 *Recurrent*

Suatu Rantai Markov dikatakan *recurrent* jika, dimulai dari *state* i, rantai akan kembali ke *state* i dalam waktu (acak) terhingga, dengan peluang 1. Dengan kata lain, *state* $i \in S$ *recurrent* jika

$$p_{i,i} := P(T_i^r < \infty \,|X_0 = i) = P(X_n = i \,|X_0 = i) = 1$$

dimana $n \geq 1$ (Privault, 2013).

Suatu Rantai Markov yang *recurrent* dapat dikatakan *positive recurrent* apabila diasumsikan bahwa ruang *state* S dari Rantai Markov $(X_n)_{n\in\mathbb{N}}$ terhingga sehingga untuk semua *state* yang *recurrent* dalam S merupakan *positive recurrent*. Oleh karena itu, dimisalkan $(X_n)_{n\in\mathbb{N}}$ Rantai Markov yang *irreducible* dengan ruang *state* terhingga S maka semua *state* dari $(X_n)_{n\in\mathbb{N}}$ merupakan *positive recurrent* (Privault, 2013).

2.3.4 Distribusi Stasioner

Suatu distribusi peluang pada S, yaitu keluarga $\pi = (\pi_i)_{i\in S}$ di $[0,1]$ sedemikian sehingga

$$\sum_{i \in S} \pi_i = 1$$

dikatakan stasioner jika, dimulai X_0 pada waktu 0 dengan berdistribusi $(\pi_i)_{i \in S}$, terbukti bahwa distribusi X_1 tetap $(\pi_i)_{i \in S}$ pada waktu 1. Dengan kata lain, $(\pi_i)_{i \in S}$ stasioner untuk Rantai Markov dengan matriks transisi P

$$P(X_0 = i) = \pi_i \quad , i \in S$$

dan

$$P(X_1 = i) = P(X_0 = i) = \pi_i \quad , i \in S$$

Hal ini juga berarti

$$\pi_j = P(X_1 = j) = \sum_{i \in S} P(X_1 = j | X_0 = i)\, P(X_0 = i) = \sum_{i \in S} p_{i,j}\, \pi_i$$

dimana distribusi π adalah stasioner jika dan hanya jika *invariant* dengan matriks P, yang berarti

$$\pi = \pi P$$

dimana perkalian P pada sisi kanan tidak berlaku pada sisi kiri.

Secara umum, dengan asumsi X_n berada dalam distribusi stasioner π pada waktu n, diperoleh

$$\pi_j = P(X_{n+1} = j) = \sum_{i \in S} P(X_{n+1} = j | n = i)\, P(X_n = i)$$

$$= \sum_{i \in S} p_{i,j}\, P(X_n = i)$$

$$= \sum_{i \in S} p_{i,j}\, \pi_i$$

$$= \pi P \quad . \qquad i, j \in S$$

Hal tersebut dikarenakan matriks transisi $(X_n)_{n\in\mathbb{N}}$ merupakan waktu yang homogen sebab peluang transisi $p_{i,j}$ tidak bergantung pada waktu n. Oleh karena itu,

$$P(X_n = j) = \pi_j, \quad j \in S \quad \rightarrow \quad P(X_{n+1} = j) = \pi_j, \quad j \in S$$

dan diinduksi pada $n \geq 0$ sehingga

$$P(X_n = j) = \pi_j, \quad j \in S, \quad n \geq 1.$$

Rantai $(X_n)_{n\in\mathbb{N}}$ tetap dalam distribusi yang sama π pada seluruh waktu $n \geq 1$ dengan catatan telah dimulai dengan distribusi stasioner π pada waktu $n = 0$ (Privault, 2013).

.

2.3.5 *Limiting Distribution*

Rantai Markov $(X_n)_{n\in\mathbb{N}}$ dikatakan mempunyai *limiting distribution* jika

$$\lim_{n\to\infty} P(X_n = j|X_0 = i)$$

ada untuk setiap $i, j \in S$ dan dari distribusi peluang pada S yaitu

$$\lim_{n\to\infty} \sum_{k\in S} P(X_n = j|X_0 = i) = 1.$$

Ketika matriks transisi P adalah *regular* (yaitu memiliki matriks yang kuat dimana koefisiennya tidak nol) pada ruang *state* berhingga S = {0, 1, . . . , N}, rantai tersebut memuat *limiting distribution* $\pi = (\pi_i)_{i\in S}$ yang diberikan oleh

$$\pi_j = \lim_{n\to\infty} P(X_n = j|X_0 = i) \qquad ; 0 \leq i, j \leq N$$

(Privault, 2013).

2.3.6 Ergodik

Jika dalam suatu Rantai Markov diberikan *state* $i \in S$, misalkan himpunan bilangan bulat $\{n \geq 1 : [P^n]_{i,i} > 0\}$. Periode dari *state* $i \in S$ merupakan pembagi persekutuan terbesar dari $\{n \geq 1 : [P^n]_{i,i} > 0\}$. Suatu *state* akan dikatakan aperiodik apabila memiliki periode 1, yang merupakan kasus khusus jika $p_{i,i} > 0$. Selain itu, *recurrent state* $i \in S$ dikatakan ergodik jika memenuhi asumsi *positive recurrent* dan aperiodik (Privault, 2013).

2.3.7 *Steady-State*

Rantai Markov yang ergodik dan *irreducible* dapat menunjukkan bahwa $\lim_{n\to\infty} p_{i,j}{}^{(n)}$ ada dan independen terhadap i. Selanjutnya

$$\lim_{n\to\infty} p_{i,j}{}^{(n)} = \pi_j$$

dimana π_j secara khusus memenuhi persamaan *steady-state*

$$(1)\quad \pi_j = \sum_{i\in S} \pi_i p_{ij} \qquad ; j \in S \tag{2.7.1}$$

$$(2)\quad \sum_{i\in S} \pi_j = 1. \tag{2.7.2}$$

Notasi π_j dikatakan peluang *steady-state* dari Rantai Markov. Peluang *steady-state* merupakan peluang untuk menemukan proses dalam keadaan tertentu, misalkan j, setelah sebagian besar transisi terjadi cenderung kepada nilai π_j, tidak berpengaruh terhadap distribusi peluang keadaan awal. Peluang *steady-state* bukan berarti proses akan berhenti pada satu *state*. Sebaliknya, proses masih berlanjut untuk menentukan transisi dari satu *state* ke *state* yang lainnya dan setiap n-langkah peluang transisi dari *state* i ke *state* j tetap $p_{i,j}$ (Hillier dan Lieberman, 2001).

2.4 Proses Keputusan Markov

Proses keputusan Markov (Markov *Decision Process / MDP)* awalnya diperkenalkan oleh Andrey Markov, seorang matematikawan Rusia pada awal abad ke-20. Proses keputusan Markov berguna untuk mempelajari berbagai masalah optimisasi yang dipecahkan melalui *Dynamic Programming.* Dengan mempertimbangkan peluang dari perubahan keadaan yang terjadi pada sistem, pada dasarnya proses keputusan Markov merupakan perluasan dari Rantai Markov sehingga harus memenuhi syarat. Selain itu, pengambilan suatu keputusan memungkinkan adanya tindakan yang akan dipilih pada sistem akan mempengaruhi biaya rata-rata yang akan digunakan. Oleh karena itu, proses keputusan Markov dapat didefinisiskan sebagai proses stokastik yang mendeskripsikan evolusi atau perubahan sistem dinamis yang dikontrol oleh barisan keputusan atau tindakan (Derman, 1970).

Pada dasarnya proses keputusan Markov dibangun oleh kebijakan, *state*, tindakan, biaya rata-rata yang digunakan, dan peluang transisi yang merupakan komponen cukup penting dalam menentukan keputusan yang akan dipilih. *State* sendiri dibangun berdasarkan observasi pengamatan yang dilakukan pada sistem. Dimisalkan bahwa sistem diobservasi pada waktu t = 0, 1, ... , dan diklasifikasikan menjadi salah satu dari sejumlah *state* yang dilabeli 0,1, ..., M. Kemudian dimisalkan $\{X_t ; t = 0, 1, ...\}$ yang menotasikan barisan dari *state* yang diobservasi. Setelah setiap observasi, salah satu keputusan (tindakan) k (yang terhingga) yang mungkin, dilabeli 1, 2, ..., K, diambil. Selanjutnya, dimisalkan $\{\Delta_t ; t = 0, 1, ...\}$ menotasikan barisan keputusan sebenarnya yang dibuat.

Suatu kebijakan, dinotasikan R, diasumikan sebagai suatu aturan yang mengatur keputusan $d_i(R)$ saat sistem dalam *state* i, i = 0, 1, ..., M. Demikian R sebenarnya dicirikan oleh nilai $\{ d_0(R), d_1(R), ..., d_M(R)\}$. Perlu diperhatikan bahwa deskripsi ini menunjukkan bahwa kapanpun *state* berada di *state* i, keputusan yang dibuat sama untuk semua nilai dari t. Kebijakan yang memiliki sifat ini disebut kebijakan stasioner (Hillier dan Lieberman, 2001).

Penjelasan lebih lanjutnya untuk permisalan di atas, yaitu diketahui $X_0, X_1, \dots$ barisan *state* yang diobservasi dan $\Delta_0, \Delta_1, \dots$ barisan keputusan. Kelas C dari semua keputusan kebijakan R mengandung fungsi

$$D_k(X_0, \Delta_0, \dots, X_t) = P(\Delta_t = d_k | X_0, \Delta_0, \dots, X_t)$$

untuk $k = 1, \dots, K; t = 0, 1, \dots$. Pada setiap $X_0, \Delta_0, \dots, X_t,\ t = 0, 1, \dots$ diasumsikan

$$D_k(X_0, \Delta_0, \dots, X_t) \geq 0, \qquad\qquad k = 1, \dots, K$$

dan

$$\sum_{k=1}^{K} D_k(X_0, \Delta_0, \dots, X_t) = 1.$$

Dari hal tersebut diasumsikan bahwa

$$P(X_{t+1} = j | X_0, \Delta_0, \dots, X_t = i) = \sum_{k=1}^{K} w_{ij}(k) D_k(X_0, \Delta_0, \dots, X_t = i)$$

untuk $i, j = 0, \dots, M; t = 0, 1, \dots$ dimana $\{w_{ij}(k)\}$ diasumsikan

$$w_{ij}(k) \geq 0, \qquad i, j \in S; k = 1, \dots, K$$

dan

$$\sum_{j \in S} w_{ij}(k) = 1, \qquad i \in S; k = 1, \dots, K.$$

Diketahui $w_{ij}(k)$ menjadi matriks stokastik untuk setiap k merupakan kemungkinan dalil yang diberlakukan oleh berbagai keputusan, dan merupakan fungsi dari pengamatan *state* terakhir sistem. Kemudian dimisalkan C' kelas dari kebijakan R dimana

$$D_k(X_0, \Delta_0, \dots, X_t = i) = D_{ik}, \qquad i \in S; k = 1, \dots, K,$$

independen terhadap $X_0, \dots, X_{t-1}, \Delta_{t-1}$ dan t. Jika $R \in C'$, barisan $\{X_t\}, t = 0, 1, \dots$ merupakan Rantai Markov dengan peluang transisi yang stasioner $\{p_{ij}(k)\}$ dimana

$$p_{ij}(k) = \sum_{j \in S} w_{ij}(k) D_{ik}$$

(Derman, 1962).

Dari hal tersebut diketahui bahwa pergerakan setiap *state* tersebut mengakibatkan adanya perpindahan dari *state* yang satu ke *state* yang lain dan mengakibatkan adanya ukuran kemungkinan atau ketidakpastian dari perpindahan *state* tersebut yang merupakan peluang transisi *state*. Pengaruh dari peluang transisi yang berkembang dari waktu ke waktu dan sejumlah barisan keputusan yang dibuat bergantung pada *state* awal, X_0. Dimisalkan kapanpun sistem berada di *state* i dan keputusan $d_i(R) = k$ dibuat, sistem berpindah ke *state* j, dengan peluang transisi $p_{ij}(k)$, untuk setiap i, j = 0, 1, ..., M dan k = 1, 2, ..., K. Dengan demikian jika diberikan kebijakan R yang mengikuti, proses stokastik yang dihasilkan adalah

Rantai Markov dengan matriks transisi yang diketahui (tergantung pada kebijakan yang dipilih). Diasumsikan bahwa Rantai Markov yang berkaitan dengan setiap matriks transisi *irreducible*.

Selanjutnya diberikan distribusi $P\{X_0 = i\}$ selama awal *state* dari sistem dan kebijakan R, sistem berkembang dari waktu ke waktu sesuai peluang transisi *state* dan barisan keputusan yang dibuat (tindakan yang diambil). Khususnya, ketika sistem berada pada *state* i dan keputusan $d_i(R) = k$ dibuat maka peluang bahwa sistem berada di *state* j pada periode waktu pengamatan berikutnya diberikan oleh $p_{ij}(k)$. Situasi ini menghasilkan barisan keputusan dari pengamatan *state* X_0, X_1, ... dan barisan keputusan yang dibuat Δ_0, Δ_1, Δ_2, ... Barisan pengamatan *state* dan keputusan yang dibuat ini disebut Proses Keputusan Markov (Hillier dan Lieberman, 2001).

2.5 Metode *The Policy Improvement*

Berikut merupakan prosedur iteratif yang memperbaiki setiap iterasi dan berakhir setelah sejumlah iterasi yang terbatas dengan kebijakan yang optimal.

Diberikan sebarang $R_1 \in C_D$. Kemudian $v_i(R_1), i \in S$ memenuhi persamaan

$$v_i(R_1) = C_{iR_1} + \sum_j p_{ij}(R_1)\, v_j(R_1), \qquad i \in S. \tag{2.5.1}$$

Diberikan E_1 yang menunjukkan himpunan tindakan pada i dimana $C_{ik} + \sum_j p_{ij}(k)\, v_j(R_1)$ sangat kurang dari sisi kanan (2.5.1). Didefinisikan $R_2 \in C_D$

sebagai berikut: pada satu atau lebih *state i*, dimana E_i bukan himpunan kosong, ditentukan tindakan k pada E_i, sedangkan pada *state* lainnya dilakukan tindakan yang ditentukan oleh R_1. Mengacu pada turunan R_1 dari R_2 sebagai iterasi perbaikan kebijakan .

Fakta bahwa iterasi adalah perbaikan yang telah ditetapkan sebagai berikut:

TEOREMA 2.3.

Jika E_i bukan himpunan kosong untuk setidaknya satu *state* i, maka

$$v_i(R_2) \leq v_i(R_1), \quad i \in S$$

dengan pertidaksamaan yang ketat yang ada pada setiap *i* dimana $R_2 \neq R_1$ (Derman, 1970).

Bukti:

Dengan definisi iterasi perbaikan kebijakan,

$$v_i(R_1) \geq C_{iR_2} + \sum_j p_{ij}(R_2)\, v_j(R_1), \qquad i \in S \tag{2.5.2}$$

dengan pertidaksamaan yang ketat yang ada pada setiap *i*. Diberikan $\{p_{ij}{}^{(t)}(R_2)\}$ yang menunjukkan probabilitas transisi t-langkah di bawah R_2. Kemudian dari (2.9.2), dilakukan perkalian oleh $p_{ij}{}^{(t)}(R_2)$ dan penjumlahan selama *i*, untuk $t =$ 0,1, ..., diperoleh

$$\sum_i p_{ji}{}^{(t)}(R_2)\, v_j(R_1) \geq \sum_i p_{ji}{}^{(t)}(R_2)\, C_{iR_2} + \sum_l p_{jl}{}^{(t+1)}(R_2)\, v_l(R_1), \quad j \in S. \tag{2.5.3}$$

Pada t = 0, persamaan (2.9.2) dan (2.9.3) identik. Pada penjumlahan (2.9.3) selama $t = 0,1, \ldots$, karena $v_j(R_2) = \sum_{t=0}^{\infty} \sum_i p_{ji}{}^{(t)}(R_2)\, C_{iR_2}$, diperoleh

$$\sum_{t=0}^{\infty} \sum_{i} p_{ji}{}^{(t)}(R_2)\, v_j(R_1) \geq v_j(R_2) + \sum_{t=1}^{\infty} \sum_{l} p_{jl}{}^{(t+1)}(R_2)\, v_l(R_1), \quad j \in S$$

dengan pertidaksamaan yang ketat, karena persyaratan pada t = 0, pada masing-masing j dimana $R_2 \neq R_1$. Pada penguraian bentuk kedua di sisi kanan dari bentuk kiri, dan karena keduanya berbeda hanya jika j = i (karena $p_{ij}{}^{(0)} = \delta_{ij}$), diperoleh $v_j(R_1) \geq v_j(R_2), i \in S$ pertidaksamaan yang ketat untuk masing-masing j dimana $R_2 \neq R_1$. Dengan demikian teorema tersebut terbukti.

Mengacu pada urutan iterasi perbaikan kebijakan sebagai prosedur perbaikan kebijakan diperoleh bahwa prosedur perbaikan kebijakan berakhir, setelah sejumlah iterasi yang terbatas, pada kebijakan yang optimal. Singkatnya, prosedur perbaikan kebijakan memberikan urutan kebijakan monoton (selalu memperbaiki) konvergen dan mencapai jumlah iterasi yang terbatas atas kebijakan yang optimal. Dalam hal ini melibatkan pemecahan sistem linear (2.5.1) (Derman, 1970).

.

2.6 Proses Keputusan Markov dengan Metode *The Policy Improvement Algorithm*

Metode *The Policy Improvement Algorithm* berguna untuk menentukan kebijakan opimal dengan cepat dan sangat efisien karena biasanya kebijakan optimal dicapai dengan jumlah iterasi yang relatif kecil. Suatu biaya yang diharapkan C_{ik} bergantung pada *state* yang diamati dalam sistem dan keputusan yang dibuat terjadi. Diketahui sistem berpindah ke *state* baru j pada periode waktu yang diamati selanjutnya, dengan peluang transisi yang diberikan $p_{ij}(k)$. Jika suatu biaya bergantung pada *state* awal dan perpindahan yang terjadi. Dinotasikan $q_{ij}(k)$ biaya yang diharapkan terjadi ketika sistem dalam *state* i dan keputusan k yang dibuat dan kemudian berpindah pada *state* j pada pengamatan periode waktu selanjutnya sehingga

$$C_{ik} = \sum_{j \in S} p_{i,j}(k)\, q_{ij}(k).$$

Ketika suatu sistem dioperasikan di bawah kebijakan R, terdapat nilai $g(R), v_0(R), v_1(R), \dots, v_M(R)$ yang memenuhi

$$g(R) + v_i(R) = C_{ik} + \sum_{j \in S} p_{i,j}(k)\, v_j(R), \qquad i \in S$$

kemudian akan diberikan justifikasi heuristik dari hubungan ini dan interpretasi untuk nilai-nilai ini.

Dinotasikan $v_i^n(R)$ total biaya harapan dari suatu sistem dimulai dari keadaan i (mulai periode waktu yang diamati pertama kali) dan dikembangkan untuk n periode waktu. $v_i^n(R)$ memiliki dua komponen: C_{ik}, biaya yang dikeluarkan selama periode waktu diamati pertama kali, dan $\sum_{j \in S} p_{i,j}(k)\, v_j^{n-1}(R)$, biaya harapan dari sistem yang dikembangkan selama $n-1$ periode waktu yang tersisa. Kedua komponen tersebut dapat dinyatakan dalam persamaan rekursif

$$v_i^n(R) = C_{ik} + \sum_{j \in S} p_{i,j}(k)\, v_j^{n-1}(R), \qquad i \in S$$

dimana $v_i^1(R) = C_{ik}$ untuk setiap *i* yang diperoleh. Hal tersebut dapat digunakan untuk menyelidiki perilaku dari total biaya yang diharapkan $v_i^n(R)$ untuk *n* yang bertambah besar. Diketahui bahwa biaya rata-rata harapan (jangka panjang) per satuan waktu mengikuti setiap kebijakan R dapat dinyatakan

$$g(R) = \sum_{i \in S} \pi_i \, C_{ik}$$

yang independen terhadap *state* awal *i*. Oleh karena itu, $v_i^n(R)$ bertindak kurang lebih seperti $n\, g(R)$ untuk n yang besar. Sebenarnya, jika fluktuasi yang kecil diabaikan, $v_i^n(R)$ dapat dinyatakan sebagai jumlah dari dua komponen

$$v_i^n(R) \approx n\, g(R) + v_i(R)$$

dimana komponen pertama independen terhadap *state* awal dan komponen kedua bergantung kepada *state* awal. Dengan demikian, $v_i(R)$ dapat diinterpretasikan sebagai efek dari biaya harapan total dalam *state* awal *i*. Akibatnya,

$$v_i^n(R) - v_j^n(R) \approx v_i(R) - v_j(R)$$

sehingga $v_i(R) - v_j(R)$ merupakan ukuran terhadap *state* awal *i* bukan *state j*.

Misalkan n bertambah besar, dapat disubtitusikan $v_i^n(R) = n\, g(R) + v_i(R)$ dan $v_j^{n-1}(R) = (n-1)\, g(R) + v_j(R)$ ke dalam persamaan rekursif $v_i^n(R)$ sehingga diperoleh

$$g(R) + v_i(R) = C_{ik} + \sum_{j \in S} p_{i,j}(k)\, v_j(R), \qquad i \in S$$

sehingga nilai ini memenuhi persamaan yang diberikan. Perlu diperhatikan bahwa sistem ini memiliki M+1 persamaan dengan persamaan M+2 yang tidak diketahui sehingga salah satu dari variabel ini dapat dipilih sembarang. Berdasarkan konvensi, $v_M(R)$ akan dipilih sama dengan nol. Oleh karena itu, dengan menyelesaikan persamaan linier, dapat diperoleh $g(R)$.

Pada prinsipnya, semua kebijakan dapat dienumerasi dan kebijakan yang meminimumkan $g(R)$ dapat ditentukan. Namun, pada sejumlah state dan keputusan yang moderat teknik ini tidak praktis sehingga digunakan algoritma untuk mengevaluasi kebijakan dan menemukan yang optimal tanpa pengenumerasian lengkap, seperti yang dijelaskan berikut ini.

Metode *The Policy Improvement Algorithm* dimulai dengan pemilihan sebarang kebijakan R_1. Kemudian dibuktikan sistem persamaan untuk menemukan nilai $g(R), v_0(R), v_1(R), \dots, v_{M-1}(R)$ [dengan $v_M(R) = 0$]. Tahap ini disebut *penentuan nilai* atau *nilai determinasi*. Kebijakan yang lebih baik, dinotasikan R_2, yang kemudian dibangun. Tahap ini disebut *perbaikan kebijakan*. Kedua tahap ini merupakan suatu iterasi terhadap algoritma. Penggunaan kebijakan baru R_2, dilakukan iterasi lainnya. Iterasi ini akan terus berjalan hingga dua iterasi yang beruntun menyebabkan kebijakan yang identik, yang menandakan bahwa kebijakan optimal telah diperoleh.

Rinciannya diuraikan di bawah:

Inisialisasi: Pilih sebarang kebijakan awal R1 dengan n=1.

Iterasi n:

- *Langkah 1: Penentuan nilai:* Untuk kebijakan R_n, digunakan $p_{ij}(k)$, C_{ik}, dan $v_M(R_n) = 0$ membuktikan sistem persamaan *M+1*

$$g(R_n) = C_{ik} + \sum_{j \in S} p_{i,j}(k)\, v_j(R_n) - v_i(R_n), \qquad i \in S \tag{2.9}$$

untuk semua *M+1* nilai yang tidak diketahui dari $g(R), v_0(R), v_1(R), \ldots, v_{M-1}(R)$.

- *Langkah 2 : Perbaikan Kebijakan:* Menggunakan nilai-nilai saat ini dari $v_i(R_n)$, dihitung untuk kebijakan R_n, menemukan alternatif kebijakan R_{n+1}, sehingga untuk setiap *state* i, $d_i(R_{n+1}) = k$ adalah keputusan yang meminimalkan

$$C_{ik} + \sum_{j \in S} p_{i,j}(k)\, v_j(R_n) - v_i(R_n) \tag{2.10}$$

yaitu, untuk setiap *state* *i*,

$$\underset{k=1,2,\ldots,k}{\text{Minimize}}\left[C_{ik} + \sum_{j \in S} p_{i,j}(k)\, v_j(R_n) - v_i(R_n)\right] \tag{2.11}$$

dan kemudian mengatur $d_i(R_{n+1})$ sama dengan nilai meminimalkan *k*. Prosedur ini mendefinisikan kebijakan baru R_{n+1}.

Tes optimalitas: Arus kebijakan R_{n+1} adalah optimal jika kebijakan ini identik dengan kebijakan Rn. Jika ya, berhenti. Jika tidak, ulang *n= n+1* dan melakukan iterasi lain.

Dua sifat utama dari algoritma ini adalah

1. $g(R_{n+1}) \leq g(R_n), \qquad untuk\ n = 1, 2, \ldots$
2. Algoritma berakhir dengan kebijakan yang optimal dalam jumlah terbatas iterasi

(Hillier dan Lieberman, 2001).

III. METODE PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada Semester Genap Tahun Ajaran 2016/2017 di Jurusan Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Metode Penelitian

Penelitian ini dilakukan secara studi literatur dan sistematis yang diperoleh dari buku-buku, jurnal-jurnal, atau media lain yang dapat menunjang proses penulisan skripsi ini.

Adapun langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Mendeskripsikan konsep analisis proses keputusan Markov dengan menggunakan metode *The Policy Improvement Algorithm* secara studi literatur
2. Menerapkan proses keputusan Markov dengan menggunakan metode *The Policy Improvement Algorithm* secara sistematis pada studi kasus dengan langkah sebagai berikut :
   a. Merumuskan suatu model stokastik yang melibatkan state dan tindakan atau keputusan dari studi kasus yang ada

b. Menentukan model Markov dan peluang transisi dari studi kasus yang akan digunakan dalam perhitungan menggunakan software Octave 4.2.1.

c. Menentukan biaya rata-rata yang diharapkan, pada saat transisi jika sistem saat ini berada di state i dan keputusan k, yang akan digunakan dalam perhitungan menggunakan software Octave 4.2.1.

d. Menentukan kebijakan optimal menggunakan proses keputusan Markov metode *The Policy Improvement Algorithm* dengan langkah-langkahnya sebagai berikut:

   a) Memilih sebarang kebijakan awal (R_1) untuk menentukan peluang transisi $(p_{ij}(k))$ dan biaya yang diharapkan (C_{ik}) sesuai kebijakan awal yang dipilih untuk digunakan dalam perhitungan menggunakan software Octave 4.2.1;

   b) Melakukan langkah penentuan nilai. Dengan menggunakan software Octave 4.2.1 akan diketahui penentuan nilai sesuai dengan kebijakan yang dipilih menggunakan persamaan

$$g(R_n) = C_{ik} + \sum_{j=0}^{M} p_{ij}(k)\, v_j(R_1) - v_i(R_1), \qquad i = 0, 1, 2, \dots, M$$

   dimana $v_M(R_n) = 0$;

   c) Melakukan langkah perbaikan kebijakan. Perbaikan kebijakan dilakukan menggunakan software Octave 4.2.1 yang menentukan alternatif kebijakan untuk kebijakan selanjutnya pada setiap *state* i yang memenuhi

$$Z = \underset{k=1,2,\dots,k}{\text{Minimize}} \left[C_{ik} + \sum_{j=0}^{M} p_{ij}(k)\, v_j(R_n) - v_i(R_n) \right];$$

d) Melakukan pengujian optimalitas. Dengan menggunakan software Octave 4.2.1 akan diketahui kebijakan optimal dari pengujian tersebut. Pengujian optimalitas dilakukan dengan mempertimbangkan kebijakan baru yang terbentuk pada setiap state setelah dilakukan perbaikan kebijakan. Apabila kebijakan baru identik dengan kebijakan sebelumnya maka iterasi akan dihentikan. Sebaliknya, apabila kebijakan belum identik maka akan dilakukan perbaikan kebijakan dengan kembali ke langkah b).

V. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dijabarkan pada bab sebelumnya maka dapat diambil kesimpulan sebagai berikut:

1. Proses keputusan Markov merupakan perluasan dari rantai Markov yang digunakan untuk memperoleh biaya yang optimal dengan mempertimbangkan keputusan k yang dipilih di bawah kebijakan R. Biaya yang optimal tersebut diperoleh dengan menggunakan metode *The Policy Improvement Algorithm* yang meminimumkan total biaya yang diharapkan terjadi dimana dalam perhitungannya melibatkan peluang *steady-state* dan perilaku asimtotik dari persamaan rekursif total biaya yang diharapkan.
2. Dari studi kasus I mengenai pemilihan kebijakan yang optimal dari konfigurasi layanan yang akan digunakan pada awal periode dapat disimpulkan bahwa kebijakan yang optimal terjadi apabila apabila belum terdapat pelanggan yang tiba dipilih tipe konfigurasi “lambat”, sedangkan pelanggan sebanyak satu maka dipilih tipe konfigurasi “lambat” dan apabila terdapat pelanggan yang datang sebanyak dua maka dipilih tipe konfigurasi “cepat”.
3. Dari studi kasus II mengenai pemilihan kebijakan yang optimal saat memilih lahan parkir yang paling tepat agar dapat meminimalisir biaya apabila terjadi

penyok pada mobil dapat disimpulkan bahwa kebijakan yang optimal terjadi apabila mobil dalam keadaan baik diparkirkan di jalanan dengan mengambil satu ruang dan apabila mobil dalam keadaan penyok pemilik memilih untuk memperbaikinya.

4. Dari studi kasus II mengenai pemilihan kebijakan yang optimal dalam menentukan investasi yang paling tepat dapat disimpulkan bahwa kebijakan yang optimal terjadi apabila pergerakan naik (turun) pasar terjadi pada angka 11.000 dan 12.000 maka dana akan diinvestasikan pada reksa dana Go-Go Fund dan apabila pergerakan naik (turun) pasar terjadi pada angka 13.000 maka dana akan diinvestasikan pada reksa dana Go-Slow Fund.
5. Pemilihan kebijakan optimal menggunakan proses keputusan Markov dengan metode *The Policy Improvement Algorithm* dapat berlaku untuk peluang transisi dan biaya yang berbeda pada setiap keputusan ataupun peluang transisi yang berbeda dengan biaya yang sama untuk setiap keputusan, serta peluang transisi yang sama pada setiap keputusan tetapi memiliki biaya yang berbeda pada setiap *state* dan keputusan. Namun, apabila jumlah keputusan yang berlaku berbeda untuk masing-masing state maka perlu dilakukan permisalan sesuai dengan *state* dan keputusan yang berlaku.

DAFTAR PUSTAKA

Bellman, R. E., dan Dreyfus, S. E. 1962. *Applied Dynamic Programming*. Princeton University Press, New Jersey.

Cox, D. R. dan Miller, H. D. 1965. *The Theory of Stochastic Processes.* Methuen, London.

Derman, Cyrus. 1970. *Finite State Markovian Decision Processes.* Academic Press. Inc., New York.

Derman, Cyrus. 1962. “On Sequential Decisions and Markov Chains.” *Management Science* 9(1): 16–24.

Hillier, F. S. and Lieberman, G. J. 2001. *Introduction to Operation Research*. 7th ed. McGraw Hill Companies Inc., New York.

Howard, R. A. 1960. *Dynamic Programming and Markov Processes.* Technology Press of Massachusetts Institute of Technology and John Wiley & Sons, Inc., New York.

Kijima, M. 1997. *Markov Process for Stochastic Modelling*. Chapman & Hall, London.

Privault, N. 2013. *Understanding Markov Chain: Examples and Applications*. Springer Science & Business Media, Singapore.

Ross, S. 2007. *Introduction to Probability Model*. John Wiley & Sons Inc. New York.

Ustazila, B. 2014. Penyelesaian Model Tahap Terhingga dan Takhingga pada Proses Keputusan Markov dan Aplikasinya di Bidang Pertanian (Skripsi). Institut Pertanian Bogor, Bogor.

Zada, T. Muhammad Shah. 2016. Model Markov untuk Pengambilan Keputusan Medis (Skripsi). Universitas Sumatera Utara, Medan.