



TESIS - SS142501

**ESTIMASI *PROPENSITY SCORE MATCHING*
BERDASARKAN PENDEKATAN *MULTIVARIATE
ADAPTIVE REGRESSION SPLINES***

STUDI KASUS: DIABETES MELLITUS (DM)

AMANDA ADITYANINGRUM
NRP. 1315 201 024

Dosen Pembimbing:
Dr. Bambang Widjanarko Otok, M.Si
Dr. Dra. Kartika Fitriyasari, M.Si

PROGRAM MAGISTER
PROGRAM PASCA SARJANA STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2017



TESIS - SS142501

PROPENSITY SCORE MATCHING ESTIMATION USING MULTIVARIATE ADAPTIVE REGRESSION SPLINES

CASE OF STUDY: DIABETES MELLITUS (DM)

**AMANDA ADITYANINGRUM
NRP. 1315 201 024**

**SUPERVISOR:
Dr. Bambang Widjanarko Otok, M.Si
Dr. Dra. Kartika Fitriasari, M.Si**

**PROGRAM MAGISTER
PROGRAM PASCA SARJANA STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2017**

**ESTIMASI PROPENSITY SCORE MATCHING BERDASARKAN
PENDEKATAN MULTIVARIATE ADAPTIVE REGRESSION
SPLINES**

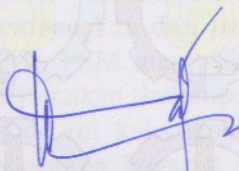
(Studi Kasus: Diabetes Mellitus)

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar Magister Sains (M.Si)
di
Institut Teknologi Sepuluh Nopember
Oleh:

**AMANDA ADITYANINGRUM
NRP. 1315 201 024**

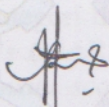
Tanggal Ujian : 13 Januari 2017
Periode Wisuda : Maret 2017

Disetujui oleh:



1. Dr. Bambang Widjanarko Otok, M.Si
NIP. 19681124 199412 1 001

(Pembimbing I)



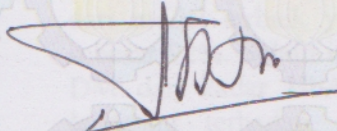
2. Dr. Dra. Kartika Fitriyanti, M.Si
NIP. 19691212 199303 2 002

(Pembimbing II)



3. Dr. Sutikno, M.Si
NIP. 19710313 199702 1 001

(Penguji)



4. Dr. Drs. Agus Suharsono, MS
NIP. 19580823 198403 1 003

(Penguji)

**an. Direktur Program Pascasarjana
Asisten Direktur**

Direktur Program Pasca Sarjana,

Prof. Dr. Ir. H. Widjaja, M.Eng.
NIP. 19611021 198603 1 001

Prof. Ir. Djauhar Manfaat, M.Sc., Ph.D.
NIP.19601202 198701 1 001



**ESTIMASI *PROPENSITY SCORE MATCHING*
MENGUNAKAN PENDEKATAN *MULTIVARIATE
ADAPTIVE REGRESSION SPLINES (MARS)*
STUDI KASUS: DIABETES MELITUS (DM)**

Nama mahasiswa : Amanda Adityaningrum
NRP : 1315 201 024
Pembimbing I : Dr. Bambang Widjanarko Otok, M.Si
Pembimbing II : Dr. Dra. Kartika Fitriasari, M.Si

ABSTRAK

Propensity Score (PS) merupakan probabilitas bersyarat mendapatkan perlakuan tertentu berdasarkan kovariat yang diamati. Salah satu metode yang dikembangkan dari *propensity score* adalah metode *Propensity Score Matching (PSM)*. PSM merupakan metode yang dilakukan dengan menyeimbangkan atau menyamakan kelompok subjek penelitian dengan metode matching. Dengan metode ini kelompok *treatment* dipasangkan dengan kelompok *non-treatment* berdasarkan kovariat yang diamati. Dalam analisis studi observasi, metode ini digunakan untuk mengurangi bias dalam estimasi efek perlakuan pada data yang bersifat observasi karena adanya faktor *confounding*. Jika *confounding* terdiri atas dua kategori dan fungsinya tidak linier, maka *Multivariate Adaptive Regression Splines (MARS)* merupakan model pendekatan yang tepat untuk estimasi nilai *propensity score* karena mudah dalam estimasi dan interpretasinya serta tidak memerlukan asumsi bahwa fungsinya harus bersifat linier. Metode PSM dengan MARS pada penelitian ini diterapkan pada data pasien penderita Diabetes Melitus (DM) tipe-2 yang dirawat di RS Dunda tahun 2015 dengan kasus Neuropati Diabetik Perifer (NDP). Tujuan penelitian adalah mengkaji estimasi *Propensity Score* berdasarkan *Multivariate Adaptive Regression Splines* dengan respon biner dalam mendapatkan estimasi ATT pada kasus DM khususnya Neuropati Diabetik Perifer (NDP). Hasil estimasi parameter dengan MLE adalah $\lambda = \delta_L(\mathbf{B}^T \boldsymbol{\beta})$. Akurasi dari model terbaik yang diperoleh dari estimasi PSM dengan MARS adalah sebesar 92%. Selain itu, dari model juga diketahui bahwa lama pasien menderita DM dan kadar HbA1C merupakan variabel yang berpengaruh secara tidak langsung terhadap status NDP pasien, sedangkan umur, jenis kelamin, dislipidemia, dan obesitas dari pasien merupakan variabel yang berpengaruh langsung terhadap status NDP pasien.

Keywords: *Confounding, Diabetes Melitus, MARS, Propensity Score Matching, Studi Observasi.*

(Halaman sengaja dikosongkan)

**PROPENSITY SCORE MATCHING ESTIMATION
USING MULTIVARIATE ADAPTIVE REGRESSION SPLINES
(MARS)
CASE OF STUDY: DIABETES MELLITUS (DM)**

Student Name : Amanda Adityaningrum
Student Identity Number : 1315 201 024
Supervisor : Dr. Bambang Widjanarko Otok, M.Si
Co. Supervisor : Dr. Dra. Kartika Fitriasari, M.Si

ABSTRACT

Propensity score (PS) is the conditional probability to get certain treatments involving the observed covariates. One method that was developed from the propensity score is the Propensity Score Matching (PSM). PSM is a method to estimate the effect of treatment by balancing the subject group using matching methods. This method help matching subjects in *treatment* group with subjects in *non-treatment* group based on the observed covariates. In observational studies, this method used to reduce bias in the estimation of the impact of treatment on observational data for their confounding factors. If confounding is binary and its function is not linear, then Multivariate Adaptive Regression Splines (MARS) with binary response is one of the methods to estimate the value of propensity score because of its easiness in terms of estimation and interpretation without having a linear function as for the assumption its model. PSM methods using MARS in this study applied to the data of patients with diabetes mellitus (DM) type-2 who were treated in Dunda Hospital at 2015. This study aims to review the estimate of propensity score based on Multivariate Adaptive Regression Splines with response continue to get estimate of ATT in case of diabetic neuropathy. The result of parameter estimation with MLE is $\lambda = \delta_L(\mathbf{B}^T \boldsymbol{\beta})$. The accuracy of the best model of PSM estimation using MARS is 92%. In addition, the model shown that the variables which are not directly influence to the NDP are duration of DM and levels of HbA1C of the patient, while age, gender, dyslipidemia and obesity of the patients are variables which are directly influence to the NDP.

Keywords: *Observational Study, Propensity Score Matching, Confounding, MARS, Diabetes Mellitus.*

(Halaman sengaja dikosongkan)

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT atas rahmat dan karunia-Nya kepada penulis, serta sholawat kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan Tesis dengan judul “Estimasi *Propensity Score Matching* berdasarkan pendekatan *Multivariate Regression* pada kasus Diabetes Mellitus”. Tesis ini dimaksudkan untuk memberikan sumbangsih bagi pengambil keputusan dalam memprediksi faktor-faktor yang mempengaruhi Diabetes Mellitus dimasa yang akan datang.

Sebagai rasa syukur, pada kesempatan ini perkenankanlah saya menyampaikan terima kasih serta penghargaan yang setinggi-tingginya teristimewa kepada yang terhormat Bapak Dr. Bambang Widjanarko Otok dan Ibu Dr. Dra. Kartika Fitriasari, M.Si selaku pembimbing yang dengan penuh kesabaran dan kesungguhan telah memberikan petunjuk, bimbingan dan arahan serta memberikan dukungan kepada penulis mulai dari perencanaan penelitian sampai pada penulisan tesis.

Pada kesempatan ini pula perkenankanlah penulis menyampaikan ucapan terima kasih yang tulus kepada yang terhormat :

1. Bapak Dr. Suhartono, M.Sc dan Bapak Dr. rer. pol. Heri Kuswanto, M.Si selaku Ketua Jurusan dan Ketua Program Studi Pasca Sarjana Statistika-FMIPA ITS Surabaya, beserta para Asisten Direktur dan Staf yang telah memberikan kesempatan dan kelancaran untuk mengikuti Program Magister pada Program Studi Pasca Sarjana Statistika-FMIPA ITS Surabaya.
2. Bapak Dr. Sutikno, M.Si dan Bapak Dr. Drs. Agus Suharsono, M selaku dosen penguji yang telah memberikan arahan dan masukan kepada penulis dalam menyelesaikan tesis.
3. Teman-teman seperjuangan, senasib dan sepenanggungan mahasiswa Program Studi Pasca Sarjana Statistika-FMIPA ITS Surabaya angkatan 2015 (khususnya Alvionita, Nisa, Agnes dan Riza) yang telah banyak

memberikan bantuan dan sumbangan pikiran serta semangat persahabatan yang tulus dalam belajar sehingga penulis dapat menyelesaikan tesis ini.

Akhirnya kepada Ayah H. Agust Gani SE dan Ibu dan Hj. Dr. Dra. Herlina Jusuf, M.Kes tersayang, serta kakak Dion Teguh Pratomo SH. MH dan adik Intan Tiara Kartika, S.IP tersayang, terima kasih atas pengertian, kasih sayang, kesetiaan, ketabahan dan kesabaran serta pengorbanannya dalam memberikan semangat, dorongan serta doa restunya selama penulis menyelesaikan studi. *Thank you for everything that I can't even give anything to repay all the kindness. All I can do is to feel gratitude because of these.* Semoga apa yang telah diberikan merupakan ibadah yang nantinya akan memperoleh imbalan dari Yang Maha Kuasa.

Penulis menyadari bahwa tesis ini masih kurang dari sempurna, sehingga dengan kerendahan hati penulis mengharapkan kritik dan saran yang sifatnya membangun. Semoga tesis ini akan bermanfaat.

Surabaya, Januari 2017

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	iii
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xii
DAFTAR TABEL	xv
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Permasalahan	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah	6
BAB 2. TINJAUAN PUSTAKA	7
2.1 <i>Propensity Score</i>	7
2.2 <i>Propensity Score Matching</i>	8
2.3 <i>Postmatching</i>	9
2.3.1 Uji <i>balance</i> kovariat	10
2.3.2 Estimasi <i>Average Treatment of Treated (ATT)</i>	11
2.4 Faktor <i>Confounding</i>	12
2.5 Konsep Klasifikasi	13
2.6 Metode Pengujian <i>k-fold Cross Validation</i>	14
2.7 Model <i>Multivariate Adaptive Regression Splines</i>	15
2.8 Estimasi Model MARS Respon Biner (MARS Klasifikasi).....	17
2.9 Diabetes Melitus (DM) Tipe-2 dan Komplikasinya	18
2.9.1 Neuropati Diabetik Perifer (NDP).....	19

2.9.2	Faktor yang Mempengaruhi Neuropati Diabetik Perifer (NDP)...	19
BAB 3.	METODOLOGI PENELITIAN	23
3.1	Sumber Data	23
3.2	Kerangka Konsep Penelitian	23
3.3	Variabel Penelitian	24
3.4	Metode Penelitian	26
3.5	Diagram Alir Penelitian	29
BAB 4.	HASIL DAN PEMBAHASAN	33
4.1	Estimasi Nilai <i>Propensity Score</i> menggunakan <i>Multivariate Adaptive Splines</i> dengan metode <i>Maximum Likelihood Estimation</i>	33
4.2	Analisis <i>Propensity Score Matching</i> (PSM) pada Kasus Neuropati Diabetik Perifer (NDP) Pasien Diabetes Melitus (DM) tahun 2015 - 2016.....	38
4.2.1	Gambaran Umum Karakteristik Pasien DM Tipe-2.....	39
4.2.2	Analisis <i>Propensity Score Matching</i> dengan MARS	42
4.2.2.1	Pemilihan Variabel Confounding.....	42
4.2.2.2	Estimasi Nilai <i>Propensity Score</i> dengan <i>Multivariate Adaptive Regression Splines</i>	46
4.2.2.3	Analisis <i>Matching</i>	52
4.2.2.4	Analisis <i>Postmatching</i>	53
a.	Pengujian <i>balance kovariat</i>	53
b.	Estimasi <i>Average Treatment of Treated</i>	55
BAB 5.	KESIMPULAN DAN SARAN	59
5.1	Kesimpulan	59
5.2	Saran	60
	DAFTAR PUSTAKA	61
	LAMPIRAN	65
	BIOGRAFI PENULIS	87

DAFTAR GAMBAR

	Halaman
Gambar 3.1 Kerangka Konsep	26
Gambar 3.2 Diagram Alir Penelitian	31
Gambar 4.1 Penentuan Model Terbaik <i>Propensity Score</i> dengan <i>Multivariate Adaptive Regression Splines</i>	51

(Halaman sengaja dikosongkan)

DAFTAR TABEL

	Halaman
Tabel 2.1	Kontingensi Ekspektasi Dua Arah 14
Tabel 2.2	Matriks Konfusi 15
Tabel 3.1	Struktur Data Penelitian 25
Tabel 3.2	Variabel Penelitian 27
Tabel 4.1	Karakteristik Pasien berdasarkan Umur (X_1) dan Lama Menderita DM (X_4) 39
Tabel 4.2	Karakteristik Pasien berdasarkan Status NDP untuk Data Kategorik 39
Tabel 4.3	Pengujian Hubungan Variabel yang Berpotensi sebagai <i>Confounding</i> dengan Kovariat (X) Lainnya 43
Tabel 4.3	Pengujian Hubungan Variabel yang Berpotensi sebagai <i>Confounding</i> dengan Kovariat (X) Lainnya 43
Tabel 4.4	Hasil <i>Trial</i> dan <i>Error</i> pembentukan Model MARS antara Kovariat (X) dan Kadar AUS (X_7) 48
Tabel 4.5	Hasil Matriks Konfusi 52
Tabel 4.6	Hasil Pengujian <i>Balance</i> 55
Tabel 4.7	Hasil Estimasi ATT dan Standar <i>Error</i> 56
Tabel 4.8	Model Hubungan Variabel Kadar AUS dengan Pasien NDP 57

(Halaman sengaja dikosongkan)

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Penelitian merupakan kajian yang mendetail tentang suatu masalah spesifik, atau isu dengan menggunakan metode ilmiah. Pada umumnya penelitian terbagi menjadi dua kelompok, yaitu eksperimental dan observasi. Salah satu standar penelitian eksperimental adalah peneliti dapat mengubah subjek atau variabel dengan cara apapun. Dengan kata lain selain mengamati, peneliti memiliki kontrol terhadap variabel penelitian. Berbeda dengan penelitian observasi (studi observasi), peneliti hanya bertindak sebagai pengamat, dimana peneliti tidak memiliki *non-treatment* pada penelitian tersebut, sehingga tidak memungkinkan untuk mengubah subjek atau variabel dari penelitian. Ada beberapa studi observasi yang biasanya dijumpai dalam kehidupan yakni: survey, sensus, record administrasi dan data lainnya yang tidak memungkinkan dilakukan pengacakan (Guo & Fraser, 2010).

Pada studi observasi di bidang kesehatan, inferensi dari efek *treated* melibatkan spekulasi tentang satu kelompok yang menerima *treatment*. Misalnya pada studi observasi terdapat 2 *treatment*, maka perbandingan antara *treatment A* dan *treatment B* disebut *Causal Effect* (efek kausal). Karena satu kelompok hanya diberikan satu *treatment*, yaitu *treatment A* atau *B*, dimana yang diamati hanyalah satu *treatment* (bukan keduanya) maka perbandingan dari kedua *treatment* menyiratkan adanya perbedaan spekulasi (Rosenbaum and Rubin, 1983). Untuk membuat inferensi kausal, pemilihan subjek dan pemberian *treatment* secara *random* (acak) sangat disarankan agar asumsi interpendensi terpenuhi sehingga bias dapat diminimalisir. Akan tetapi pada studi observasi, *treatment* secara acak tidak memungkinkan. Akibatnya pada studi observasi sulit untuk membuat inferensi kausal karena peneliti kesulitan untuk membuat perbedaan antara hasil dari kelompok *non-treatment* dan kelompok *treatment*. Hal inilah yang memotivasi para statistikawan untuk menemukan metode baru yang dapat menangani permasalahan ini.

Metode *propensity score* pertama kali diperkenalkan oleh Rosenbaum dan Rubin pada tahun 1983. *Propensity score* didefinisikan sebagai probabilitas bersyarat untuk menerima intervensi berdasarkan pada karakteristik-karakteristik sebelum intervensi tersebut. Metode ini merupakan metode penyesuaian statistik yang dapat digunakan untuk menganalisis data dari desain studi observasi dimana desain pemberian *treatment* melalui pengacakan untuk kelompok *treatment* atau *non-treatment* tidak mungkin dilakukan. Peneliti dapat menggunakan *propensity score* pada statistik untuk menyeimbangkan atau menyamakan kelompok subjek penelitian sehingga dapat mengurangi bias akibat pemberian *treatment* yang tidak *random*. Bias tersebut akan tereduksi ketika hasil perbandingan antara kelompok *non-treatment* dan *treatment* yang sedekat mungkin. Kata “terdekat” diukur dalam ukuran karakteristik-karakteristik yang dapat diobservasi (Rosenbaum and Rubin, 1983).

Ada beberapa metode berbasis *propensity score* yakni, *Propensity Score Matching* (PSM), *Propensity Score Stratification* (PSS), *covariate adjustment propensity score* dan *Propensity Score Weighting* (PSW). Rossi dan Freedman (1989) dalam (Guo & Fraser, 2010) menjelaskan bahwa Metode PSM menitikberatkan bagaimana menyeimbangkan data kelompok *treatment* dan *non-treatment* dengan melakukan pencocokan berdasarkan kovariat yang diamati. Sedangkan PSS merupakan prosedur dimana observasi dibagi menjadi beberapa kelompok atas dasar kovariat, dan PSW menitikberatkan pada pemberian bobot untuk setiap observasi berdasarkan kovariat yang diamati atau berdasarkan estimasi nilai dari *propensity score* (Guo & Fraser, 2010).

Beberapa penelitian tentang estimasi nilai *propensity score* yaitu, D’Agostino (1998) menggunakan analisis diskriminan untuk estimasi *propensity score*, McCaffrey (2004) menggunakan model *generalized boosted*, Setouguchi (2008) menggunakan studi simulasi pada *neural network* dan *classification tree*, Farida (2015) menggunakan regresi logistik pada kasus data HIV/AIDS serta Ingka (2016) menggunakan regresi logistik biner dengan bootstrap pada kasus Diabetes Melitus. Sebagian besar metode tersebut memiliki kelemahan seperti datanya harus bersifat nominal (analisis diskriminan) atau adanya kompleksitas dan permasalahan dalam interpretasi, serta menghadapi banyak kerumitan dari

hasil algoritma yang kompleks dan implementasinya (*generalized boosted, neural network, classification tree*) dan bergantung pada asumsi bahwa fungsi harus bersifat linier (regresi logistik).

Multivariate Adaptive Regression Splines (MARS) adalah salah satu metode regresi nonparametric. MARS merupakan Metode untuk menyelesaikan masalah regresi dan klasifikasi dengan tujuan memprediksi variabel respon bersifat biner atau kontinu berdasarkan beberapa variabel. Metode ini merupakan kombinasi *recursive partition regression* dengan *spline truncated*. Metode ini berguna untuk mengatasi permasalahan data berdimensi tinggi (yakni melibatkan jumlah observasi dan jumlah variabel prediktor yang besar, 3-20 variabel) dan tidak bergantung pada asumsi bahwa fungsi harus bersifat linier sehingga mampu menghasilkan prediksi variabel respon yang akurat, serta mampu mengatasi kelemahan *recursive partition regression* dengan menghasilkan model yang kontinu pada knot, yang didasarkan pada nilai *generalized cross validation* (GCV) minimum (Leathwick, Rowe, Richardson, Elith, Hastie, 2005).

Selain penelitian tentang estimasi nilai *propensity score*, telah banyak pula dikembangkan penelitian tentang metode berbasis *propensity score*, Littnerova *et al.*, (2013) menggunakan PSM untuk mereduksi pengaruh bias pada penelitian tentang *mortality*, D'Agostino (1998) menggunakan PSM dan PSS untuk mereduksi bias pada perbandingan kelompok *treatment* dan kelompok *non-treatment* untuk kasus dalam bidang obat-obatan, Austin (2011) membandingkan 4 metode *propensity score* (PSM, PSS, *covariate adjustment propensity score* dan PSW) untuk mereduksi perbedaan sistematis antara kelompok *treatment* dan control pada kasus *Smoking Cessation Counseling on Mortality*, dimana diperoleh kesimpulan bahwa metode PSM merupakan metode terbaik dalam mereduksi perbedaan sistematis untuk kasus tersebut. Berdasarkan beberapa penelitian diatas dapat disimpulkan bahwa PSM merupakan metode terbaik untuk mereduksi pengaruh bias pada studi observasi.

Saat ini perhatian penyakit tidak menular semakin meningkat. Dari sepuluh penyebab utama kematian, dua diantaranya adalah penyakit tidak menular. Salah satu diantaranya adalah Diabetes Melitus (DM). DM merupakan penyakit tidak menular yang prevalensinya cukup tinggi di dunia. Neuropati diabetik perifer (NDP) merupakan komplikasi utama yang sering ditemukan pada

pasien DM khususnya pasien DM tipe-2 (Bansal, Gudala, Muthyala, Esam, Nayakallu, & Bhansali, 2014).

Kadar Asam Urat Serum (AUS) tinggi berhubungan dengan kejadian komplikasi NDP pada penderita DM. Hubungan kadar AUS tinggi dengan DM khususnya komplikasi NDP dijelaskan dalam penelitian Darsana (2014) bahwa ada korelasi positif antara kadar AUS dengan kejadian NDP. Oleh karena itu, kadar AUS merupakan faktor resiko dari kejadian NDP.

Penelitian tentang NDP sudah banyak dilakukan. Misalnya Darsana (2014) meneliti tentang korelasi positif antara kadar AUS dengan NDP menggunakan analisis deskriptif dan uji *coefficient contingency*, Bruce & Young (2008) meneliti tentang prevalensi dan faktor resiko NDP menggunakan uji nonparametrik Kruskal-Wallis dan regresi logistik multivariate, Qilsy & Ardiansyah (2012) meneliti tentang hubungan antara hiperglikemia, usia dan lama menderita DM dengan angka kejadian neuropati diabetika menggunakan uji korelasi spearman, uji chi-square dan multivariate logistik biner.

Penelitian mengenai NDP diatas mengindikasikan banyak sekali faktor yang mempengaruhi kejadian NDP pada penderita DM tipe-2. Penelitian-penelitian tersebut dilakukan tanpa memperhatikan kemungkinan adanya kombinasi yang kuat antara faktor-faktor yang mempengaruhi kejadian NDP. Padahal pada studi observasi belum tentu semua data yang terkumpul diamati secara benar. Kemungkinan ada faktor yang tidak teramati, seperti variabel confounding, atau kemungkinan terjadinya interaksi antar faktor-faktor sebelumnya dapat menyebabkan hasil estimasi yang bias dan kesimpulan yang tidak akurat. Bias tersebut terjadi karena kelompok *treatment* dan *non-treatment* yang tidak seimbang. Oleh karena itu dibutuhkan penelitian yang menggunakan metode untuk mereduksi bias pada studi observasi, khususnya faktor *confounding*. Beberapa penelitian sebelumnya berusaha menangani faktor *confounding* dengan randomisasi (pengacakan) namun dalam bidang kesehatan hal ini tidak bisa dilakukan. Maka dibutuhkan suatu metode yang dapat menangani efek bias yang ditimbulkan oleh faktor *confounding* ini. Salah satu metode yang bisa menangani *confounding* adalah metode *propensity score matching* (PSM).

Mengingat semakin tingginya prevalensi kejadian DM di Indonesia yang berdampak pada tingginya prevalensi NDP, maka perlu dilakukan penelitian untuk mengkaji lebih jauh faktor-faktor yang mempengaruhi terjadinya Neuropati Diabetik Perifer (NDP) ini. Berdasarkan uraian tersebut maka tulisan ini mengkaji dan memaparkan bagaimana estimasi nilai *propensity score* dengan metode MARS untuk mendapatkan pengaruh faktor-faktor yang mempengaruhi NDP pada pasien DM tipe-2 yang dirawat di Rumah Sakit Dunda pada tahun 2015-2016.

1.2 Permasalahan

Multivariate Adaptive Regression Splines (MARS) adalah metode untuk menyelesaikan masalah regresi dan klasifikasi dengan tujuan memprediksi variabel respon berdasarkan beberapa variabel. Selain itu, jika variabel respon bersifat biner dan melibatkan dimensi data yang besar yakni dengan jumlah observasi dan jumlah variabel prediktor (3-20 variabel) yang cukup banyak, serta fungsinya tidak linier, maka akan lebih baik jika digunakan MARS respon biner (MARS klasifikasi). Adanya hubungan antara faktor-faktor yang mempengaruhi Neuropati Diabetik Perifer (NDP) seringkali dapat menyebabkan hasil estimasi yang bias dan kesimpulan yang tidak akurat karena adanya faktor *confounding* yang tidak teridentifikasi. Beberapa penelitian sebelumnya berusaha menangani faktor *confounding* dengan randomisasi (pengacakan) namun dalam studi observasi pada bidang kesehatan hal ini tidak bisa dilakukan. Oleh karena itu, dibutuhkan suatu metode yang dapat menangani efek bias yang biasanya terjadi pada studi observasi. Salah satu metode yang bisa menangani bias adalah metode *Propensity Score Matching* (PSM). Berdasarkan uraian tersebut maka tulisan ini mengkaji dan bagaimana estimasi nilai *propensity score matching* berdasarkan MARS dan mengkaji bagaimana analisis *propensity score matching* pada kasus penderita diabetes mellitus tipe-2.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan diatas, maka tujuan yang ingin dicapai dalam penelitian ini adalah

1. Mengkaji estimasi *propensity score matching* berdasarkan MARS dengan respon biner
2. Menerapkan analisis *propensity score matching* pada kasus penderita diabetes mellitus

1.4 Manfaat Penelitian

Manfaat yang ingin dicapai dari hasil penelitian ini adalah sebagai berikut.

1. Menambah wawasan keilmuan dalam pengembangan dan penerapan analisis *propensity score matching* dan MARS khususnya dalam bidang kesehatan.
2. Memberikan informasi terkait faktor-faktor yang mempengaruhi NDP, sehingga dapat dijadikan bahan pertimbangan dalam mengambil kebijakan untuk menyusun langkah strategis yang lebih efektif guna meminimalisir kejadian NDP.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Data yang digunakan adalah data sekunder hasil rekam medis pasien DM tipe-2 di Rumah Sakit Dunda tahun 2015-2016.
2. Metode estimasi nilai *propensity score* yang digunakan adalah metode MARS.
3. Metode *propensity score* yang digunakan adalah *propensity score matching*.
4. *Software* yang digunakan adalah program R.

BAB 2

TINJAUAN PUSTAKA

2.1 *Propensity Score*

Propensity Score Analysis diperkenalkan oleh Rosenbaum dan Rubin tahun 1983 dalam jurnalnya yang berjudul “*The central role of the propensity score in observational studies for causal effects*”. Analisis *propensity score* adalah metode statistika inovatif yang berkembang pesat dan berguna untuk mengevaluasi efek *treatment* ketika menggunakan data observasi (Guo dan Fraser, 2010). Untuk observasi i ($i = 1, \dots, n$), *propensity score* didefinisikan sebagai probabilitas bersyarat dari *treatment* tertentu ($Z_i = 1$) versus *non-treatment* ($Z_i = 0$) berdasarkan pada karakteristik kovariat x_i yang diobservasi dimana pengacakan tidak bisa dilakukan (Rosenbaum dan Rubin, 1983).

Menurut Guo dan Fraser (2010) nilai *propensity score* $e(\mathbf{x}_i)$ didefinisikan sebagai berikut.

$$e(\mathbf{x}_i) = P(Z_i = 1 | X_i = x_i) \quad (2.1)$$

Pada kasus percobaan acak, status *treatment* Z_i independen tanpa syarat pada variabel respon Y_i . Untuk data observasi non-acak, asumsi independen tersebut tidak dapat dicapai karena adanya faktor *confounding* yang tidak terdeteksi. Akibatnya, perbandingan sederhana dari hasil rata-rata antara unit *treatment* dan unit *non-treatment* tidak akan secara umum mengungkapkan efek kausal atau dengan kata lain kesimpulan yang diperoleh tidak akurat karena adanya seleksi bias (Li, Graham dan Majumdar, 2013). *Propensity score* merupakan metode yang tepat untuk mengatasi permasalahan seleksi bias tersebut. Keuntungan dari *propensity score* dibandingkan dengan penyesuaian multivariabel adalah pemisahan faktor *confounding* dan analisis langkah pengaruh *treatment* (Littnerova et al., 2013).

Menurut Yanovitzky, Zanutto, & Hornik (2005), langkah pertama yang dilakukan dalam *propensity score* yaitu memilih kovariat sebagai *confounder* untuk estimasi *propensity score*. Proses pemilihan kovariat sebagai *confounder*

untuk estimasi *propensity score* pada umumnya didasarkan pada teori sebelumnya atau bukti empiris tentang hubungan antar variabel.

Model *propensity score* pada umumnya ada tiga yakni: *propensity score matching* (PSM), *propensity score stratification* (PSS), dan *propensity score weighting* (PSW). Jika diterapkan dengan tepat, maka model-model ini dapat membantu memecahkan masalah seleksi bias dan memberikan estimasi yang valid dari efek rata-rata *treatment* (Guo dan Fraser, 2010).

Selain memecahkan masalah seleksi bias, salah satu keuntungan dari *propensity score matching* dan *stratification* adalah mereduksi dimensi, dimana jika vektor x memiliki banyak kovariat yang direpresentasikan dalam banyak dimensi maka pendekatan *propensity* mampu mereduksi semua dimensi menjadi satu skor dimensi, yakni nilai *propensity score* (Guo dan Fraser, 2010).

2.2 *Propensity Score Matching*

Propensity score matching mengacu pada bagaimana menyeimbangkan data kelompok *treatment* dan *non-treatment* dengan melihat nilai yang mirip pada hasil estimasi *propensity score* dan melakukan pencocokan berdasarkan kovariat yang diamati, kemudian membuang unit yang tidak cocok (Rubin, 2001). *Propensity score matching* biasanya digunakan untuk membandingkan dua subjek kelompok, tapi metode ini juga dapat digunakan untuk analisis dengan lebih dari dua kelompok.

Menurut Guo dan Fraser (2010) *propensity score matching* model dijelaskan dalam tiga tahap proses analisis. Langkah pertama yaitu mencari variabel atau kovariat yang diduga sebagai penyebab ketidakseimbangan antara kelompok *treatment* (misalnya diberi simbol 1) dan kelompok *non-treatment* (misalnya diberi simbol 0). Model *propensity score* selalu dimulai dengan estimasi peluang bersyarat dari yang menerima *treatment* dengan analisis data menggunakan estimasi MARS. Analisis pada langkah ini adalah untuk mengidentifikasi kovariat yang diamati yang dapat menyebabkan adanya seleksi bias. Secara definisi, *propensity score* adalah probabilitas bersyarat dari peserta yang menerima *treatment* pada penelitian berdasarkan kovariat yang diamati. Sehingga tidak hanya peserta dalam kelompok *treatment*, namun dalam kelompok

non-treatment juga mungkin memiliki nilai *propensity score* yang tidak nol. Dengan kata lain, *propensity score* merupakan skor seimbang (*balance score*) yang mewakili kovariat, dimana sepasang peserta kelompok *treatment* dan *non-treatment* memiliki nilai yang hampir sama.

Langkah kedua yaitu pencocokkan (*matching*) dengan metode *Nearest Neighbor Matching*, setelah diperoleh *balance score* dari estimasi *propensity score*, peneliti kemudian menggunakan nilai (*score*) tersebut untuk mencocokkan peserta pada kelompok *treatment* dengan peserta pada kelompok *non-treatment*. Dengan *matching* biasanya peserta dari kelompok *non-treatment* akan dikurangi, sehingga ada beberapa peserta yang bahkan tidak digunakan sama sekali karena tidak memiliki nilai yang cocok dengan peserta pada kelompok *non-treatment*, oleh sebab itu *matching* juga sering disebut dengan istilah *resampling*.

Misalkan P_1 dan P_0 merupakan nilai *propensity score* dari kelompok *treatment* dan *non-treatment*, I_1 adalah himpunan peserta pada kelompok *treatment* dan I_0 adalah himpunan peserta pada kelompok *non-treatment*. *Neighborhood* $A(P_i)$ memuat setiap peserta pada kelompok *non-treatment* 0 ($0 \in I_0$) sebagai pasangan dari peserta pada kelompok *treatment* 1 ($1 \in I_1$), jika nilai mutlak dari *propensity score* adalah yang paling minimum diantara semua pasangan nilai *propensity score* antar 1 dan 0 , maka

$$A(P_i) = \min_0 \|P_1 - P_0\|, \quad 0 \in I_0 \quad (2.2)$$

Setelah seorang peserta dari kelompok *non-treatment* (0) dipasangkan ke kelompok *treatment* (1), peserta ini kemudian dikeluarkan dari I_0 .

Langkah ketiga dari *propensity score matching*, yaitu melakukan *postmatching* dengan nilai *propensity score*. Pada tahap ini peneliti akan menghitung nilai *balance* kovariat dan nilai ATT dengan membandingkan perbedaan rata-rata dari hasil antara kelompok *treatment* dan *non-treatment* untuk seluruh sampel seperti yang dilakukan pada penelitian eksperimental.

2.3 *Postmatching*

Jika estimasi dari nilai *propensity score* telah diperoleh dan *matching* telah dilakukan, maka langkah selanjutnya adalah melakukan analisis *postmatching* dari

propensity score. Pada analisis *postmatching* juga akan dilakukan evaluasi dari *propensity score matching*, seperti uji keseimbangan (*balance*) kovariat dan estimasi *Average Treatment of Treated* (ATT) atau estimasi efek perlakuan (Tu dan Zhou, 2002).

2.3.1 Uji *balance* kovariat

Salah satu cara untuk mengecek keseimbangan kovariat pada *confounders* antar kelompok *treatment* dan *non-treatment* dari *propensity score matching* yaitu membandingkan variasi statistik seperti rata-rata, median, variansi, uji kesamaan dua rata-rata, uji kesamaan dua proporsi pada setiap kovariat (Li, 2012). Pada *propensity score* evaluasi ini digunakan untuk mengecek keseimbangan kovariat pada *confounders* (menguji perbedaan) antara kelompok *treatment* dan *non-treatment*. Untuk menguji perbedaan tersebut digunakan uji-t (uji kesamaan rata-rata) dan uji-z (uji kesamaan dua proporsi). Uji-t digunakan untuk menguji perbedaan dua kelompok data kontinu, sedangkan uji-z digunakan untuk data kategori.

2.3.1.1 Uji-t

Untuk data kontinu yang menjadi perbandingan antara kelompok *treatment* dan kelompok *non-treatment* adalah rata-rata dengan menggunakan uji-t. Sebelum melakukan uji-t, terlebih dahulu dilakukan pengujian kesamaan variansi dengan uji Bartlett. Uji Bartlett merupakan uji chi-square dengan derajat bebas $k-1$, dimana k adalah jumlah kategorik (atau kelompok) dari variabel independen. Misalkan S_j^2 menotasikan variansi dari sampel n_j dari populasi j ($j=1, \dots, k$). Hipotesis yang digunakan adalah sebagai berikut:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

H_1 : setidaknya ada dua variansi populasi yang tidak sama (misalnya

$$\sigma_i^2 \neq \sigma_j^2 \text{ untuk } i, j)$$

Statistik uji yang digunakan pada uji Bartlett adalah sebagai berikut (Arsham, 2010).

$$B = \frac{(N-k) \ln \left(\frac{\sum_{i=1}^k (n_i - 1) s_i^2}{N-k} \right) - \sum_{i=1}^k (n_i - 1) \ln (s_i^2)}{1 + \frac{1}{3(k-1) \left[\left(\sum_{i=1}^k \frac{1}{n_i - 1} \right) - \frac{1}{N-k} \right]}} \quad (2.3)$$

dengan N merupakan jumlah dari seluruh sampel.

Setelah melakukan pengujian kesamaan variansi, maka langkah selanjutnya, yaitu menguji kesamaan dua rata-rata untuk data kontinu dengan menggunakan uji-t. Hipotesis yang digunakan pada uji-t adalah sebagai berikut.

$H_0 : \mu_1 = \mu_0$ (tidak ada perbedaan antara kelompok *treatment* dan kelompok *non-treatment*)

$H_1 : \mu_1 \neq \mu_0$ (ada perbedaan antara kelompok *treatment* dan *non-treatment*)

dengan statistik uji dan kriteria keputusannya adalah sebagai berikut.

- Untuk varians yang homogen (variansi sama)

$$t_{hit} = \frac{\bar{x}_1 - \bar{x}_0}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_0}}} \quad (2.4)$$

dengan

$$s = \sqrt{\frac{(n_1 - 1) s_1^2 + (n_0 - 1) s_0^2}{n_1 + n_0 - 2}} \quad (2.5)$$

Tolak H_0 jika $|t_{hit}| > t_{\frac{\alpha}{2}, df}$; $df = n_T + n_C - 2$ atau p-value $< \alpha$. \bar{x}_1 dan \bar{x}_0

merupakan mean dari kelompok *treatment* dan kelompok kontrol, n_1 dan n_0 masing-masing merupakan banyaknya observasi pada kelompok *treatment* dan kelompok *non-treatment*, s_1^2 dan s_0^2 merupakan varians kelompok *treatment* dan kelompok *non-treatment*, dan s merupakan simpangan baku gabungan.

- Untuk varians tidak homogen (variansi berbeda)

$$t_{hit} = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_0^2}{n_0}}} \quad (2.6)$$

dimana

$$df = \frac{\left[\left(\frac{s_1^2}{n_1} \right) + \left(\frac{s_0^2}{n_0} \right) \right]^2}{\left[\left(\frac{s_1^2}{n_1} \right) / (n_1 - 1) \right] + \left[\left(\frac{s_0^2}{n_0} \right) / (n_0 - 1) \right]} \quad (2.7)$$

Tolak H_0 jika $|t_{hit}| > t_{\frac{\alpha}{2}, df}$ atau p-value $< \alpha$. \bar{x}_1 dan \bar{x}_0 merupakan mean dari kelompok *treatment* dan kelompok *non-treatment*, n_1 dan n_0 masing-masing merupakan banyaknya observasi pada kelompok *treatment* dan kelompok *non-treatment*, s_1^2 dan s_0^2 merupakan varians kelompok *treatment* dan kelompok *non-treatment*, dan s merupakan simpangan baku gabungan.

2.3.1.2 Uji-z

Untuk data kategori yang menjadi perbandingan antara kelompok *treatment* dan kelompok *non-treatment* adalah proporsi dengan menggunakan uji-z. Hipotesis yang digunakan pada uji-z adalah sebagai berikut.

$H_0 : P_1 = P_0$ (Tidak ada perbedaan proporsi antara kelompok *treatment* dan kelompok *non-treatment*)

$H_1 : P_1 \neq P_0$ (Ada perbedaan proporsi antara kelompok *treatment* dan kelompok *non-treatment*)

dengan statistik ujinya adalah sebagai berikut.

$$Z = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{\hat{p}(1-\hat{p}) \left[\frac{1}{n_1} + \frac{1}{n_0} \right]}}; \quad \hat{p} = \frac{x_1 + x_0}{n_1 + n_0}; \quad \hat{p}_1 = \frac{x_1}{n_1}; \quad \hat{p}_0 = \frac{x_0}{n_0} \quad (2.8)$$

Tolak H_0 jika $|Z_{hit}| > Z_{\alpha/2}$ atau p-value $< \alpha$. \hat{p} adalah proporsi gabungan, dengan \hat{p}_1 adalah proporsi kelompok *treatment* dan \hat{p}_0 adalah proporsi kelompok *non-*

treatment, serta n_1 adalah proporsi kelompok *treatment* dan n_0 adalah proporsi kelompok *non-treatment*.

2.3.2 Estimasi *Average Treatment for Treated* (ATT)

Setelah uji *balance* kovariat dilakukan, langkah selanjutnya adalah estimasi *Average Treatment for Treated* (ATT) atau estimasi perlakuan terhadap *treatment*. Estimasi ini merupakan hal penting dalam *propensity score* karena dengan metode ini peneliti dapat memperoleh estimasi perlakuan terhadap *treatment* (ATT) yang tidak bias. Efek perlakuan $\hat{\theta}$ (ATT), standar error dan statistik ujinya dapat dihitung dengan menggunakan persamaan sebagai berikut.

$$\hat{\theta} = \sum_{k=1}^K \frac{n_{1k} + n_{0k}}{n_1 + n_0} (\hat{p}_{1k} - \hat{p}_{0k}) \quad (2.9)$$

$$SE(\hat{\theta}) = \sqrt{\sum_{k=1}^K \left(\frac{n_{1k} + n_{0k}}{n_1 + n_0} \right) \left(\frac{s_{1k}^2}{n_{1k}} + \frac{s_{0k}^2}{n_{0k}} \right)} \quad (2.10)$$

n_1 dan n_0 merupakan jumlah subjek kelompok *treatment* dan *non-treatment*, n_{1k} jumlah subjek kelompok *treatment* kelas ke- k , n_{0k} jumlah subjek kelompok *non-treatment* kelas ke- k , respon kelompok *non-treatment* kelas ke- k , \hat{p}_{1k} dan \hat{p}_{0k} adalah proporsi respon dari kelompok *treatment* dan *non-treatment* kelas ke- k , s_{1k}^2 dan s_{0k}^2 adalah varians respon untuk kelompok *treatment* dan *non-treatment* kelas ke- k , $k = 1, \dots, K$.

Berdasarkan persamaan (2.9) dan (2.10) diperoleh statistik uji untuk estimasi $\hat{\theta}$ adalah sebagai berikut

$$T = \frac{\hat{\theta}}{SE(\hat{\theta})} \quad (2.11)$$

$\hat{\theta}$ merupakan rata-rata efek perlakuan terhadap kelompok *treated* (ATT), $SE(\hat{\theta})$ merupakan standar error ATT. Daerah kritis untuk taraf signifikansi α adalah tolak H_0 jika $|T_{hit}| > T_{\alpha/2}$ atau p-value $< \alpha$.

2.4 Faktor *Confounding*

Istilah *Confounding* berasal dari bahasa latin *cunfundere* yang berarti *to mix together*. *Confounding* merupakan *distorsi* (gangguan) dalam menaksir pengaruh paparan terhadap kejadian penyakit/*outcome* akibat tercampurnya pengaruh faktor luar. Faktor luar (*extraneous variable*) ini yang dapat mempengaruhi faktor resiko dan penyakit secara bersama-sama dalam proses timbulnya penyakit. Faktor inilah yang disebut faktor *confounding*. Faktor *confounding* biasanya ada jika perbandingan antara kelompok *treatment* dan *non-treatment* tidak seimbang. Kriteria variabel *confounding* yaitu: (1) Merupakan faktor resiko bagi kasus yang diteliti; (2) Mempunyai hubungan dengan variabel bebas lainnya (Bruce, Pope, dan Stanistreet, 2008).

Untuk mengidentifikasi faktor yang berpotensi sebagai *confounding* dapat dilakukan dengan dua cara yakni: (1) Variabel tersebut diidentifikasi sebagai *confounder* pada penelitian sebelumnya; (2) Faktor yang memenuhi kriteria *confounding* (Kamangar, 2012). Proses pemilihan *confounder* dapat berdasarkan teori yang kemudian ditunjukkan hubungan antara variabel tersebut dengan uji empiris. Untuk memeriksa hubungan antar variabel secara empiris dapat menggunakan tabel kontingensi ekspektasi dua arah.

Tabel 2.1 Tabel Kontingensi Ekspektasi Dua Arah (Agresti, 1990)

Variabel X_1	Variabel X_2		Total
	1	2	
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

Hipotesis yang digunakan untuk menguji hubungan dua variabel pada tabel 2.1 (Agresti, 1990) adalah sebagai berikut.

$H_0: P(X_1X_2) = P(X_1)P(X_2)$ (Variabel X_1 dan X_2 independen/tidak terdapat hubungan yang signifikan antar variabel)

$H_1: P(X_1X_2) \neq P(X_1)P(X_2)$ (Variabel X_1 dan X_2 dependen/terdapat hubungan yang signifikan antar variabel)

Statistik uji yang digunakan adalah uji *chi-square* dengan rumus dibawah ini:

$$\chi^2 = \sum_{r=1}^R \sum_{c=1}^C \frac{(n_{rc} - e_{rc})^2}{e_{rc}}; e_{rc} = \frac{n_r \cdot n_c}{n}, r = 1, 2, \dots, R; c = 1, 2, \dots, C \quad (2.12)$$

dengan r adalah baris, c adalah kolom, n_{rc} adalah frekuensi pengamatan sel ke- rc , e_{rc} adalah frekuensi harapan sel ke- rc , dimana daerah kritiknya adalah H_0 ditolak jika $\chi_{hit}^2 > \chi_1^2$ atau $p\text{-value} < \alpha$

Ada beberapa cara untuk mengontrol *confounding*, yakni mengontrol pada tahap design penelitian atau pada tahap analisa. Pada tahap design dapat dilakukan restriksi berupa membatasi obyek penelitian, matching dan randomisasi. Sedangkan pada tahap analisa dapat dilakukan stratifikasi dan analisis multivariate dengan teknis regresi (Bruce *et al.*, 2008).

2.5 Konsep Klasifikasi

Andaikan terdapat set data dengan 2 kelas, yaitu kelas A dan kelas B. Data dari dua kelas tersebut dapat dipisahkan dengan degan suatu garis keputusan (*decision line*). Garis inilah yang memberikan jawaban ketika ada data yang baru. Garis lurus yang memisahkan dua kelas tersebut berperan untuk membagi wilayah fitur menjadi 2 wilayah atau lebih yang berbeda kelas disebut dengan *klasifikator*.

Sebuah sistem yang melakukan klasifikasi diharapkan dapat mengklasifikasikan semua set data dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi.

Tabel 2.2 Tabel Matriks Konfusi (Prasetyo, 2012)

f_{ij}		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas =1	f_{11}	f_{10}
	Kelas =0	f_{01}	f_{00}

Setiap sel f_{ij} dalam matriks menyatakan jumlah data dari kelas i yang hasil prediksinya masuk ke kelas j . Misalnya, sel f_{11} adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1, sedangkan f_{10} adalah data dalam kelas 1 yang dipetakan secara salah ke kelas 0.

Berdasarkan isi matriks konfusi, dapat diketahui jumlah data dari masing-masing kelas yang diprediksi secara benar, yaitu $(f_{11} + f_{00})$ dan data yang diklasifikasikan salah, yaitu $(f_{10} + f_{01})$. Kuantitas matriks konfusi dapat diringkas menjadi dua nilai, yaitu akurasi dan laju error. Dengan mengetahui jumlah data yang diklasifikasikan secara benar, maka dapat diketahui akurasi hasil prediksi, jumlah data yang diklasifikasikan secara salah, dan laju error (kesalahan prediksi) dari prediksi yang dilakukan. Data kuantitas ini digunakan sebagai metrik kinerja klasifikasi (Prasetyo, 2012). Untuk menghitung akurasi digunakan formula sebagai berikut.

$$\text{Akurasi} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}}$$

Untuk menghitung error digunakan formula sebagai berikut (Prasetyo, 2012).

$$\text{Error} = 1 - \text{Akurasi}$$

2.6 Metode Pengujian *K-Fold Cross Validation*

Nilai akurasi dan error dapat diperoleh dengan menguji set data yang ada. Salah satu metode pengujian yang digunakan, yaitu metode validasi silang (*Cross Validation*). Pada metode ini, setiap data digunakan dalam jumlah yang sama untuk pelatihan (*training*) dan tepat satu kali untuk pengujian (*testing*). Ilustrasinya adalah sebagai berikut, misalnya terdapat suatu set data yang dipecah menjadi dua bagian dengan ukuran yang sama. Kemudian dipilih satu bagian untuk *training*, sedangkan lainnya untuk *testing*. Selanjutnya adalah menukar peran, bagian yang tadinya menjadi set data *train* sekarang ditukar menjadi set data *test*, begitu pula sebaliknya. Pendekatan seperti ini disebut *two-fold cross-validation*. Total *error* didapatkan dengan menjumlahkan *error* yang diperoleh dari dua kali proses tersebut. Pada ilustrasi ini, setiap data berkesempatan satu kali menjadi data *test* dan satu kali menjadi data *train*.

Bentuk umum metode ini disebut dengan *k-fold cross validation*, yang memecah set data menjadi k bagian set dengan ukuran yang sama. Setiap kali *running*, satu pecahan berperan sebagai data *train*, sedangkan yang lainnya menjadi set data *test*. Prosedur tersebut dilakukan sebanyak k kali sehingga setiap data berkesempatan menjadi data *test* tepat satu kali dan menjadi data *train* sebanyak $k-1$ kali. Total *error* diperoleh dengan menjumlahkan semua *error* yang didapatkan dari k kali proses (Tan, Steinbach, Kumar, 2006).

2.7 Model Multivariate Adaptive Regression Splines (MARS)

MARS merupakan salah satu pemodelan regresi non parametrik yang dikembangkan oleh Friedman pada tahun 1990 dengan menggunakan algoritma *recursive partitioning* yang dimodifikasi. Pada proses pembentukan model MARS, diawali dengan menentukan tiga hal berikut.

1. Basis fungsi (BF) merupakan fungsi yang didefinisikan dari setiap region. Friedman (1991) menyarankan jumlah maksimum fungsi basis (BF) adalah 2-4 kali jumlah variabel prediktor.
2. Maksimum interaksi (MI) merupakan banyaknya interaksi yang bisa terjadi pada model. Jumlah maksimum interaksi (MI) adalah 1,2, dan 3, karena menurut Friedman (1991) jika maksimum interaksi yang digunakan lebih dari 3, maka nilai GCV akan semakin meningkat dan model yang digunakan akan semakin kompleks. Jika maksimum interaksi (MI) yang digunakan sebesar 1, artinya tidak ada interaksi antar variabel dalam model. Jika maksimum interaksi (MI) yang digunakan adalah 2, maka dalam model akan terjadi interaksi antara 2 variabel. Begitupun jika maksimum interaksi (MI) yang digunakan sebesar 3, berarti banyaknya interaksi dalam model yang paling banyak terjadi adalah antara 3 variabel.
3. Minimum observasi (MO) merupakan jumlah pengamatan paling minimal antar knot. Jumlah maksimum interaksi (MI) adalah 1,2, 3, 4 dan 5 karena diatas itu nilai GCV akan semakin meningkat.

Penentuan basis fungsi (BF), maksimum interaksi, dan minimum observasi akan berpengaruh terhadap model MARS yang akan terbentuk. Menurut Friedman

(1991) estimator model *Multivariate Adaptive Regression Splines* (MARS) dapat dituliskan sebagai berikut.

$$\hat{f}(x) = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} \left[S_{km} \cdot (x_{v(k,m)} - t_{km})_+ \right] \quad (2.13)$$

dengan β_0 merupakan konstanta, β_m merupakan konstanta dari fungsi basis ke- m , m merupakan banyaknya fungsi basis (*nonconstant basis function*), K_m merupakan derajat interaksi, s_{km} merupakan nilai ± 1 , $x_{v(k,m)}$ merupakan variabel prediktor, t_{km} merupakan nilai *knot* dari variabel prediktor $x_{v(k,m)}$.

Model MARS dari persamaan (2.13) dapat juga dituliskan sebagai berikut:

$$\text{logit } \lambda = \ln \left(\frac{\lambda}{1-\lambda} \right) = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} \left[S_{km} \cdot (x_{v(k,m)} - t_{km})_+ \right] \quad (2.14)$$

dan dalam bentuk matriks menjadi

$$\text{logit } \lambda = \ln \left(\frac{\lambda}{1-\lambda} \right) = \mathbf{B}\beta \quad (2.15)$$

dengan

$$\beta = (\beta_0, \beta_1, \dots, \beta_M)^T$$

$$\mathbf{B} = \begin{bmatrix} 1 & \prod_{k=1}^{K_m} [S_{1m} \cdot (x_{1(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_m} [S_{Mm} \cdot (x_{1(M,m)} - t_{Mm})_+] \\ 1 & \prod_{k=1}^{K_m} [S_{1m} \cdot (x_{2(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_m} [S_{Mm} \cdot (x_{2(M,m)} - t_{Mm})_+] \\ \vdots & \vdots & \dots & \vdots \\ 1 & \prod_{k=1}^{K_m} [S_{1m} \cdot (x_{n(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_m} [S_{Mm} \cdot (x_{n(M,m)} - t_{Mm})_+] \end{bmatrix}$$

Persamaan (2.15) disebut sebagai model MARS klasifikasi (respon biner) dan dalam bentuk fungsi probabilitas dapat dinyatakan sebagai

$$\lambda = \delta_L(\mathbf{B}^T \beta) \quad (2.16)$$

Untuk memperoleh estimasi β digunakan metode estimasi maksimum likelihood.

Menurut Friedman (1991), hal yang harus diperhatikan dalam pemilihan model yang optimum (terbaik) dalam model mars adalah model yang memiliki nilai GCV terkecil dibandingkan dengan model-model lain. Pada model MARS,

pemilihan model MARS dengan metode *stepwise*. *Forward stepwise* dilakukan untuk mendapatkan fungsi dengan jumlah basis fungsi maksimum. Kriteria pemilihan basis fungsi pada *forward* adalah dengan meminimumkan *Average Sum Square Residual* (ASR). Untuk memenuhi konsep parsimoni (model sederhana) dilakukan *backward stepwise* yaitu memilih basis fungsi yang dihasilkan dari *forward stepwise* dengan meminimumkan nilai *Generalized Cross Validation* (GCV). Nilai GCV menurut Freidman dan Silverman (1989) didefinisikan sebagai berikut.

$$GCV(M) = \frac{\left(\frac{1}{N}\right) \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{\left[\frac{1 - C(M)}{N}\right]^2} \quad (2.17)$$

dengan N merupakan banyak observasi, M merupakan banyak *subregion*, $C(M)$ merupakan nilai kompleksitas model yang berupa matriks fungsi basi berukuran $M \times N$, dan $C(M)$ merupakan $\text{Trace}(\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T) + 1$.

2.8 Estimasi Model MARS Respon Biner (MARS Klasifikasi)

Model MARS menghasilkan model yang kontinu pada knots dengan basis

fungsi $\beta(x) = \prod_{k=1}^{K_m} \left[S_{km} \cdot (x_{v(k,m)} - t_{km})_+ \right]$. Estimasi β pada persamaan (2.16)

digunakan metode estimasi maksimum likelihood. Menurut Otok (2010), langkah-langkah estimasi parameter model MARS respon biner adalah sebagai berikut.

1. Menuliskan model MARS respon biner seperti pada persamaan (2.13).
2. Memasukkan model MARS respon biner pada persamaan *propensity*

$$\text{score (2.1), sehingga diperoleh } e(\mathbf{x}_i) = \frac{\exp(\hat{f}(x))}{1 + \exp(\hat{f}(x))}, \hat{f}(x) \text{ kemudian}$$

dinotasikan dengan z .

3. Mengambil sampel random sebanyak n , $y \sim \text{Ber}(1, \delta_L(\mathbf{B}_i \boldsymbol{\beta}))$ dan dinyatakan dalam model $z = \delta_L(\mathbf{B}\boldsymbol{\beta}) + b$
4. Membentuk fungsi *likelihood* $L(\boldsymbol{\beta})$

5. Memaksimumkan $L(\boldsymbol{\beta})$ atau \ln *likelihood* dari $L(\boldsymbol{\beta})$ yang dinotasikailah dengan $l(\boldsymbol{\beta})$
6. Mencari turunan pertama $l(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ dan menyamakannya dengan 0 untuk memperoleh estimator maksimum *likelihood* dari $\boldsymbol{\beta}$ ($\hat{\boldsymbol{\beta}}$)
7. Mencari turunan kedua $l(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$, sehingga diperoleh matriks Hessian \mathbf{H} yang digunakan untuk membuktikan bahwa $\hat{\boldsymbol{\beta}}$ adalah estimator yang memaksimumkan fungsi *likelihood*.

2.9 Diabetes Melitus (DM) Tipe-2 dan Komplikasinya

Diabetes melitus adalah penyakit metabolisme yang merupakan suatu kumpulan gejala yang timbul pada seseorang karena adanya peningkatan kadar glukosa darah di atas nilai normal. Penyakit ini disebabkan gangguan metabolisme glukosa akibat kekurangan insulin baik secara absolut maupun relatif. Ada 2 tipe diabetes melitus yaitu diabetes tipe 1/diabetes juvenile yaitu diabetes yang umumnya didapat sejak masa kanak-kanak dan diabetes tipe 2 yaitu diabetes yang didapat setelah dewasa (Kemenkes, 2013). Pada penderita DM tipe-2 terjadi hiperinsulinemia tetapi insulin tidak bisa membawa glukosa masuk ke dalam jaringan karena terjadi resistensi insulin yang merupakan turunya kemampuan insulin untuk merangsang pengambilan glukosa oleh jaringan perifer dan untuk menghambat produksi glukosa oleh hati. Komplikasi yang biasa muncul pada pasien DM tipe-2 adalah komplikasi mikrovaskuler seperti diabetik retinopati, diabetik nefropati, dan diabetik neuropati, serta komplikasi makrovaskuler berupa penyakit jantung koroner, dan stroke (Ndraha, 2014).

2.9.1 Neuropati Diabetik Perifer (NDP)

Istilah “neuropati” merupakan terminologi yang sangat luas, dimana saraf tepi mengalami gangguan fungsi yang disebabkan berbagai faktor antara lain metabolik, trauma, jebakan, penyakit defisiensi, keracunan, gangguan imunologis dan genetik. Neuropati merupakan salah satu komplikasi DM yang sering dijumpai. Neuropati Diabetik (ND) adalah keadaan dimana saraf tepi mengalami

gangguan fungsi akibat kerusakan seluler maupun molekuler yang etiologinya karena penyakit DM. Polineuropati diabetik menggambarkan keterlibatan banyak saraf tepi dan distribusinya umumnya bilateral simetris meliputi gangguan motorik, sensoris, maupun otonom (Tesfaye,2004). Angka kejadian neuropati ini meningkat bersamaan dengan lamanya menderita penyakit Diabetes Melitus dan bertambahnya usia penderita.

2.9.2 Faktor yang Mempengaruhi Neuropati Diabetik Perifer (NDP)

Menurut Veves dan Malik (2007) faktor-faktor yang mempengaruhi NDP adalah lama menderita DM, merokok, umur, kadar HbA1C, jenis kelamin, peningkatan massa tubuh (obesitas), penggunaan alkohol CAGE tinggi, kadar albumin serum yang rendah, pengobatan insulin, hyperglikemia. Sedangkan menurut Bruce dan Young (2008) faktor yang mempengaruhi NDP adalah glukosa, lipid dan asam amino, merokok, lama DM, pengukuran antropometri seperti berat, lingkar pinggang dan lingkar pinggul serta informasi demografi berupa jenis kelamin, usia. Namun, tidak semua faktor tersebut dijelaskan pada penelitian ini.

Berikut ini disajikan teori tentang faktor-faktor yang mempengaruhi Neuropati Diabetik Perifer (NDP) berdasarkan variabel yang ada pada data sekunder hasil rekam medis pasien penderita DM tipe-2 di Rumah Sakit Dunda.

- **Umur**

Berdasarkan penelitian yang dilakukan oleh Bansal, *et al.* (2014) diketahui bahwa umur pasien DM berpengaruh signifikan terhadap kejadian NDP dimana prevalensi NDP akan meningkat seiring bertambahnya umur pasien DM. Hal ini didukung juga oleh penelitian yang dilakukan Tesfaye, Stevens, Stephenson, Fuller, Plater, Tirgoviste, Nuber, Pozza, Ward, dan the EURODIAB IDDM (1996) dimana ditemukan pula hal yang serupa yakni umur mempunyai hubungan dengan kejadian NDP. Ini berarti bahwa umur merupakan salah satu faktor resiko terjadinya NDP.

- **Jenis Kelamin**

Berdasarkan penelitian Bansal, *et al.* (2014) diketahui bahwa pasien DM laki-laki dan perempuan memiliki resiko yang sama untuk menderita NDP,

hal ini berarti bahwa jenis kelamin tidak berpengaruh terhadap kejadian NDP. Akan tetapi, hal ini bertentangan dengan pendapat Veves dan Malik (2007) dan penelitian Bruce dan Young. (2008). Veves dan Malik (2007) dalam bukunya menyatakan bahwa jenis kelamin merupakan salah satu faktor resiko terjadinya NDP sedangkan Bruce dan Young. (2008) dalam penelitiannya menemukan bahwa jenis kelamin ternyata berpengaruh signifikan terhadap kejadian NDP, dimana diketahui bahwa perempuan mempunyai resiko 2,689 kali lebih besar untuk terkena NDP dibanding laki-laki.

- **Dislipidemia**

Dislipidemia adalah kelainan metabolisme lipid yang ditandai peningkatan kolesterol total, kolesterol LDL, trigliserida di atas nilai normal serta penurunan kolesterol HDL di dalam darah. Berdasarkan pendapat Veves dan Malik (2007) dan penelitian yang dilakukan oleh Tesfaye, *et al.* (1996) diketahui bahwa dislipidemia merupakan salah satu faktor resiko dari kejadian NDP. Hal ini didukung juga oleh penelitian yang dilakukan Bansal, *et al.* (2014), dalam penelitiannya Bansal, *et al.* menjelaskan bahwa dislipidemia berpengaruh signifikan terhadap kejadian NDP dimana pasien DM yang dislipidemia mempunyai resiko 0,43 kali lebih besar untuk terkena NDP dibandingkan pasien yang tidak dislipidemia.

- **Lama Menderita DM**

Berdasarkan penelitian yang dilakukan oleh Lestari, Hidayat, dan Risanti (2015) diperoleh bahwa terdapat hubungan yang signifikan antara lama menderita DM dengan proporsi kejadian neuropati diabetik dengan pengaruh sebesar 51,42%. Hal serupa ditemukan pula oleh Bansal *et al.* (2014) dimana dalam penelitiannya diketahui bahwa resiko pasien yang menderita DM selama 6-10 tahun untuk terkena NDP adalah sebesar 1,33 kali, sedangkan pasien yang menderita DM lebih dari 15 tahun mempunyai resiko terkena NDP sebesar 8,03 kali. Hal ini berarti bahwa semakin lama seseorang menderita DM maka resiko untuk terkena NDP semakin besar.

- **Obesitas**

Menurut Mayer (dalam Wijayanti 2013) obesitas merupakan keadaan patologis karena penimbunan lemak berlebihan daripada yang diperlukan untuk fungsi tubuh. Obesitas merupakan salah satu faktor resiko penyebab terjadinya penyakit degeneratif seperti DM dan penyakit jantung koroner (Wijayanti, 2013). Metode yang banyak digunakan untuk mengukur tingkat obesitas adalah indeks massa tubuh (IMT). Kriteria IMT untuk masing-masing negara berbeda.

- **Kadar HbA1C**

HbA1c merupakan ikatan molekul glukosa pada hemoglobin secara non-enzimatik melalui proses glikasi post translasi. Hemoglobin yang terglikasi terlihat dalam beberapa asam amino HbA yang terdiri dari HbA1a, HbA1b dan HbA1c. Berdasarkan pendapat Veves dan Malik (2007) dan penelitian Bruce dan Young. (2008) diketahui bahwa kadar HbA1C merupakan salah satu faktor resiko yang berpengaruh signifikan terhadap kejadian NDP. Hal ini didukung juga oleh penelitian yang dilakukan Bansal, *et al.* (2014) dimana dalam penelitiannya dijelaskan bahwa kadar HbA1C berpengaruh signifikan terhadap kejadian NDP. Pasien DM yang kadar HbA1C $> 6,5\%$ mempunyai resiko 1,03 kali lebih besar untuk terkena NDP dibandingkan pasien yang kadar HbA1C $\leq 6,5\%$.

- **Kadar Asam Urat Serum (AUS)**

Asam urat merupakan asam lemah yang terdapat dalam cairan tubuh. Menurut Sadr, Namayandeh, Moadares, dan Rafiei (2009) faktor-faktor yang mempengaruhi kadar asam urat serum adalah umur, jenis kelamin, obesitas, dislipidemia, dan merokok. Sedangkan menurut Conen, Wietlisbach, Bovet, Shamlaye, Riesen, Paccaud, dan Burnier (2004) kadar asam urat serum dipengaruhi oleh genetik DM, konsumsi alkohol yang berlebih, obesitas, dislipidemia, umur dan jenis kelamin. Faktor-faktor di atas dapat meningkatkan kadar asam urat secara signifikan. Peningkatan kadar asam urat ini dapat menimbulkan komplikasi yang salah satunya adalah neuropati diabetik. Berdasarkan penelitian yang dilakukan oleh

Darsana (2014) diketahui bahwa ada korelasi positif antara kadar asam urat serum tinggi dengan neuropati diabetik perifer pada penderita DM tipe-2. Hal ini berarti semakin tinggi kadar AUS penderita DM tipe-2 semakin tinggi pula kemungkinan untuk menderita NDP. Oleh karena ini, kadar asam urat serum merupakan salah satu faktor resiko dari kejadian NDP.

BAB 3

METODOLOGI PENELITIAN

3.1 Sumber Data

Dalam penelitian ini sumber data yang digunakan adalah data sekunder berupa rekam medis pasien penderita Diabetes Melitus tipe-2 yang dirawat di ruang rawat inap Rumah Sakit Dunda tahun 2015-2016.

Tabel 3.1 Struktur Data Penelitian

Pasien ke (i)	Variabel Respon (y_i)	Variabel Prediktor ($x_{i,p}$)			
		$x_{i,1}$	$x_{i,2}$...	$x_{i,7}$
1	y_1	$x_{1,1}$	$x_{1,2}$...	$x_{1,7}$
2	y_2	$x_{2,1}$	$x_{2,2}$...	$x_{2,7}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
145	y_{145}	$x_{145,1}$	$x_{145,2}$...	$x_{145,7}$

dengan

y_i = status NDP pada pasien ke- i , $i = 1,2,\dots,145$; bernilai 0 untuk pasien tidak NDP serta bernilai 1 untuk pasien dengan NDP.

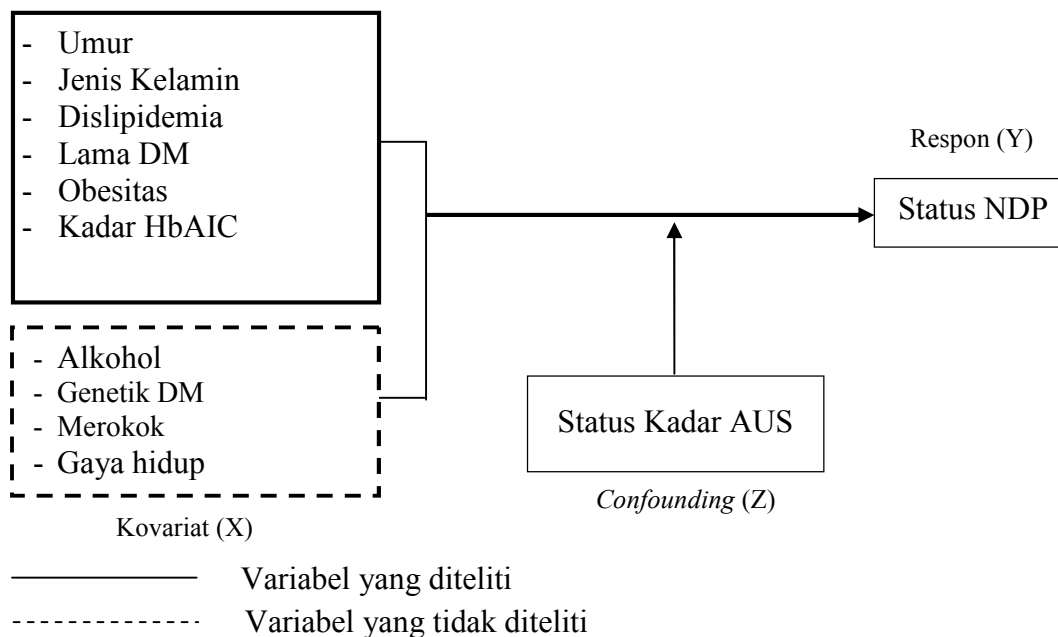
$x_{i,p}$ = nilai variabel prediktor pada pasien ke- i prediktor ke- p , $i = 1,2,\dots,145$;

$p = 1,2,\dots,7$

3.2 Kerangka Konsep Penelitian

Berdasarkan teori Neuropati Diabetik Perifer (NDP) dan faktor-faktor yang dijelaskan pada tinjauan pustaka diketahui bahwa faktor-faktor yang mempengaruhi kejadian NDP, yaitu faktor umur jenis kelamin, dislipidemia, obesitas, status kadar HbA1C, status kadar AUS (Veves dan Malik, 2007) dan lama DM (Bruce dan Young, 2008). Akan tetapi menurut teori sebelumnya, faktor-faktor yang mempengaruhi

status kadar AUS yaitu faktor umur, jenis kelamin, dislipidemia, obesitas (Sadr et al., 2009). Faktor-faktor ini mempengaruhi terjadinya NDP tergantung pada status kadar AUS dari penderita. Semakin tinggi status kadar AUS penderita DM tipe-2, semakin tinggi pula kemungkinan untuk menderita NDP (Darsana, 2014). Dari penelitian diatas, maka sebagai hipotesis awal variabel *confounding* yang akan digunakan dalam penelitian ini adalah status kadar AUS. Faktor-faktor diatas merupakan faktor resiko dari kejadian NDP. Untuk lebih jelasnya, kerangka konsep yang digunakan dalam penelitian ini disajikan pada Gambar 3.1. Lebih lanjut, akan dilakukan uji hubungan dengan mengganti variabel *confounding* sebelumnya dengan variabel kovariat lainnya.



Gambar 3.1 Kerangka Konsep Faktor-Faktor yang Berhubungan dengan Neuropati Diabetik Perifer (NDP) pasien DM menurut Darsana (2014) setelah dimodifikasi

3.3 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari dua jenis, yaitu variabel respon (*dependen*) dan variabel prediktor (*independen*). Variabel respon (Y)

pada penelitian adalah Status NDP. Variabel status NDP terdiri dari dua kategori yakni tidak NDP dan NDP. Sedangkan variabel prediktor sebanyak 7 dimana satu diantaranya akan dijadikan variabel *confounding*. Definisi Operasional dari variabel yang digunakan dalam penelitian ini ditampilkan pada Tabel 3.2.

Tabel 3.2 Variabel Penelitian

Nama Variabel	Deskripsi Variabel	Kategori	Skala
Status NDP (y)	NDP dibagi menjadi stadium 0, stadium 1, stadium 2, dan stadium 3	0 = Tidak NDP (tidak NDP memenuhi kriteria stadium 0) 1 = NDP (NDP memenuhi kriteria stadium 1,2, dan 3)	Nominal
Umur (x_1)	Umur pasien pada saat data diambil berdasarkan KTP (thn)	-	Rasio
Jenis kelamin (x_2)	Jenis kelamin dari penderita Diabetes Melitus	0 = laki-laki 1 = perempuan	Nominal
Dislipidemia (x_3)	Kelainan metabolisme lemak yang ditandai oleh peningkatan maupun penurunan lipid dalam plasma	0 = Tidak dislipidemia (kenaikan kadar kolesterol total ≤ 200 mg/dL) 1 = Dislipidemia (kenaikan kadar kolesterol total > 200 mg/dL)	Nominal
Lama menderita DM (x_4)	Waktu dalam hitungan tahun sejak penderita didiagnosis menderita DM tipe 2 yang diketahui dari wawancara kepada pasien, keluarga, dan rekam medis penderita	-	Rasio
Obesitas (x_5)	Keadaan fisik yang diukur dengan metode antropometri dengan mengukur indeks massa tubuh (IMT) IMT dihitung dengan membagi berat badan (kg) dengan tinggi badan kuadrat (m^2).	0 = Tidak obesitas (jika $IMT \leq 25kg/m^2$) 1 = Obesitas (jika $IMT > 25kg/m^2$)	Ordinal

Tabel 3.2 Lanjutan Variabel Penelitian

Nama Variabel	Deskripsi Variabel	Kategori	Skala
Status kadar HbA1C (x_6)	Bentuk glikolisasi dari hemoglobin yang dapat digunakan sebagai indikator dari toleransi glukosa dan regulasi glukosa pada penderita DM	0 = Rendah/normal (HbA1C \leq 7%) 1 = Tinggi (HbA1C $>$ 7%)	Ordinal
Status kadar AUS (x_7)	Status kadar AUS yang diperiksa dengan metode <i>enzymatic colorimetric</i> dengan alat <i>Synchron CX9 Pro</i> buatan <i>Beckham Coulter</i>	0 = Rendah/normal (kadar AUS \leq 5,5 mg/dl) 1 = Tinggi (kadar AUS $>$ 5,5 mg/dL)	Ordinal

3.4 Metode Penelitian

Metode penelitian yang digunakan sebagai tahapan untuk mencapai tujuan penelitian dijabarkan sebagai berikut :

1. Mendapatkan estimasi *propensity score* dengan menggunakan MARS dengan respon biner.
 - a. Menuliskan model MARS dengan respon biner dalam bentuk fungsi probabilitas, seperti pada persamaan (2.10).
 - b. Memasukkan model MARS respon biner pada persamaan *propensity score* (2.1) sehingga diperoleh

$$\hat{e}(x) = \frac{\exp(\hat{f}(x))}{1 + \exp(\hat{f}(x))}, \hat{f}(x) \text{ kemudian dinotasikan dengan } z.$$

- c. Mengambil sampel random sebanyak n , $y \sim \text{Ber}(1, \delta_L(\mathbf{B}_i\boldsymbol{\beta}))$ dan dinyatakan dalam model $z = \delta_L(\mathbf{B}\boldsymbol{\beta}) + b$
- d. Membentuk fungsi *likelihood* $L(\boldsymbol{\beta})$

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n P(z = z_i)$$

- e. Memaksimumkan $L(\boldsymbol{\beta})$ atau \ln *likelihood* dari $L(\boldsymbol{\beta})$ yang dinotasikani sebagai

$$l(\boldsymbol{\beta}) = \ln[L(\boldsymbol{\beta})]$$

- f. Mencari turunan pertama $l(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ dan menyamakan dengan 0 untuk memperoleh estimator maksimum *likelihood* dari $\boldsymbol{\beta}$ ($\hat{\boldsymbol{\beta}}$)

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$$

- g. Mencari turunan kedua $l(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_a \partial \boldsymbol{\beta}_b}$$

Sehingga diperoleh matriks Hessian \mathbf{H} yang digunakan untuk membuktikan bahwa $\hat{\boldsymbol{\beta}}$ adalah estimator yang memaksimumkan fungsi *likelihood*.

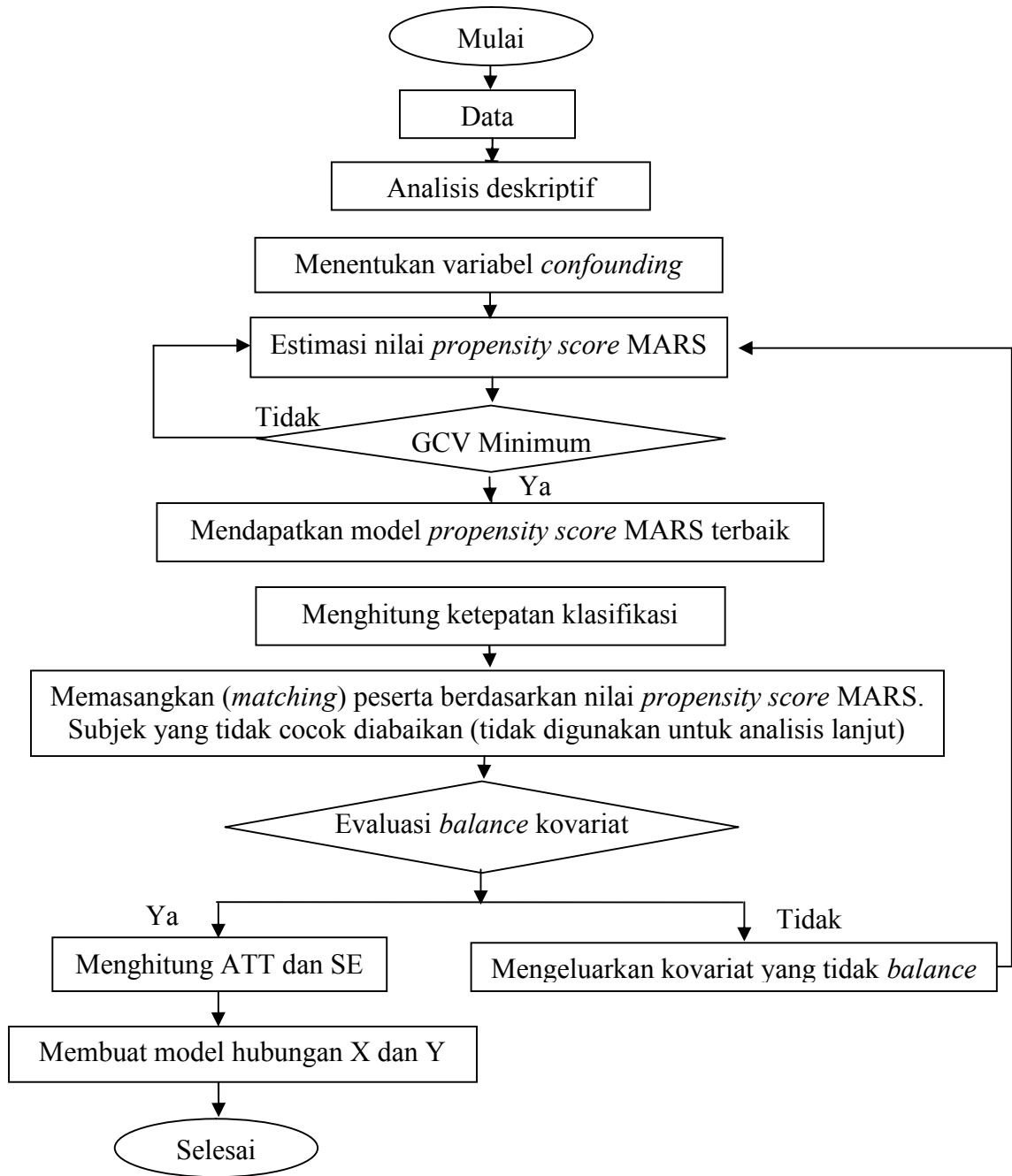
2. Aplikasi metode *propensity score matching* digunakan pada data DM dengan langkah-langkah sebagai berikut:
- Melakukan statistik deskriptif pada data.
 - Menentukan variabel *confounding*, selanjutnya variabel *confounding* dinotasikan Z .
 - Menghitung nilai estimasi *propensity score* MARS yang telah diperoleh pada langkah (1), sehingga diperoleh *balance score* dimana nilai (*score*) tersebut akan digunakan pada analisis *matching*.
 - Menghitung ketepatan klasifikasi dari nilai estimasi *propensity score* MARS.
 - Melakukan analisis *matching* dengan metode *Nearest Neighbor Matching*, yang dilakukan dengan memasang (*matching*) pasien pada kelompok *treatment* (atau kelompok 1) dengan pasien pada kelompok *non-treatment* (atau kelompok 0) berdasarkan nilai estimasi *propensity score* MARS

yang diperoleh pada langkah (2d), kelompok *non-treatment* yang tidak mempunyai pasangan, tidak diikuti pada analisis selanjutnya.

- f. Melakukan analisis *postmatching* dari hasil pasangan *treatment* dan *non-treatment* yang diperoleh pada langkah (2e). Dari hasil tersebut dihitung ATT dan SE dengan membandingkan perbedaan rata-rata dari hasil antara kelompok *treatment* dan *non-treatment* untuk seluruh sampel seperti yang dilakukan pada penelitian eksperimental. Serta menguji *balance* pada setiap kovariat.
- g. Membuat model hubungan kovariat X dan Y.

3.5 Diagram Alir Penelitian

Berikut adalah diagram alir langkah-langkah analisis yang dilakukan dalam penelitian ini



Gambar 3.2 Diagram Alir Penelitian

(Halaman sengaja dikosongkan)

BAB 4

HASIL DAN PEMBAHASAN

4.1 Estimasi Nilai *Propensity Score* menggunakan *Multivariate Adaptive Regression Splines* dengan Metode *Maximum Likelihood Estimation*

Pada dasarnya estimasi nilai *propensity score* menggunakan *multivariate adaptive regression splines* bergantung pada tipe data dari variabel *confounding*. Jika variabel *confounding* terdiri dari 2 kategori maka estimasi yang cocok adalah menggunakan *multivariate adaptive regression splines* respon biner (MARS klasifikasi), tapi jika variabel *confounding* terdiri dari 3 kategori atau lebih maka estimasi yang cocok adalah *multivariate adaptive regression splines* respon kontinu. Variabel *confounding* yang digunakan dalam penelitian ini terdiri dari 2 kategori, sehingga nilai *propensity score* diestimasi menggunakan MARS respon biner. Tahapan estimasi sesuai dengan tahapan (1) pada metode penelitian.

Nilai *propensity score* merupakan nilai yang menyatakan peluang subjek untuk mendapatkan suatu *treatment* berdasarkan karakteristik dari kovariat yang diamati sebelumnya. Untuk mendapatkan nilai *propensity* maka parameter dari nilai *propensity* harus diketahui. Parameter nilai *propensity score* diestimasi dengan pendekatan metode *Maximum Likelihood Estimation* (MLE). Berikut ini tahapan dalam estimasi parameter dengan pendekatan MLE.

Pertimbangkan y suatu variabel acak berdistribusi Bernoulli, $y \sim \text{Bernoulli}(1, \lambda)$ dengan $y \in \{0, 1\}$ dan $\mathbf{x} \in \mathfrak{R}^p$, maka $P(Z_i = 1) = \lambda$ dan $P(Y_i = 0) = 1 - \lambda$. $\mathbf{x} \in \mathfrak{R}^p$ adalah vektor dari p variabel prediktor. Misalkan ditulis sebagai, $y = \lambda + b$ maka $b = y - \lambda$, sehingga

$$E(b) = E(y - \lambda) = E(y) - E(\lambda) = \lambda - \lambda = 0 \quad (4.1)$$

dan

$$\text{Var}(b) = \text{Var}(y - \lambda) = \lambda(1 - \lambda) \quad (4.2)$$

Lemma 1: Jika hubungan dengan model logistik $\delta_L : \mathbb{R} \rightarrow (0,1)$, $\delta_L = \lambda = \left(\frac{e^z}{1+e^z} \right)$ maka invers dari δ_L merupakan transformasi logit, sehingga logit $\lambda = \ln\left(\frac{\lambda}{1-\lambda}\right) = z$. Bukti:

$$\begin{aligned}\delta_L &= \pi(x) \\ &= \frac{e^z}{1+e^z} \\ &= \frac{e^z / e^z}{(1+e^z) / e^z} \\ \delta_L &= \frac{1}{1+e^{-z}}\end{aligned}$$

Turunan pertama dari δ_L yang dinotasikan dengan δ'_L adalah:

$$\begin{aligned}\delta'_L(z) &= \frac{\partial}{\partial z} \left(\frac{e^z}{1+e^z} \right) \\ &= \frac{e^z(1+e^z) - (e^z)^2}{(1+e^z)^2} \\ \delta'_L(z) &= \frac{e^z}{(1+e^z)^2} > 0, \quad \forall z \in \mathfrak{R}\end{aligned}\tag{4.3}$$

Invers dari δ_L , yaitu:

$$\begin{aligned}\lambda &= \frac{e^z}{1+e^z} \\ \lambda[1+e^z] &= e^z \\ \lambda + \lambda e^z &= e^z \\ \lambda &= e^z - \lambda e^z \\ \lambda &= [1-\lambda]e^z \\ \frac{\lambda}{1-\lambda} &= e^z \\ \ln\left(\frac{\lambda}{1-\lambda}\right) &= z\end{aligned}$$

Jadi, invers dari δ_L dapat dikatakan sebagai tranformasi logit, yaitu:

$$\text{logit } \lambda = \ln\left(\frac{\lambda}{1-\lambda}\right) = z \quad (4.4)$$

Selanjutnya jika Lemma 1 terpenuhi dan $z = \hat{f}(x)$, yaitu:

$$z = \hat{f}(x) = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} \left[S_{km} \cdot \left(x_{v(k,m)} - t_{km} \right)_+ \right] \quad (4.5)$$

maka dapat ditulis dalam model

$$\text{logit } \lambda = \ln\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} \left[S_{km} \cdot \left(x_{v(k,m)} - t_{km} \right)_+ \right] \quad (4.6)$$

dan dalam bentuk matriks menjadi

$$\text{logit } \lambda = \ln\left(\frac{\lambda}{1-\lambda}\right) = \mathbf{B}\beta \quad (4.7)$$

dengan

$$\beta = (\beta_0, \beta_1, \dots, \beta_M)^T$$

$$\mathbf{B} = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} \left[S_{1m} \cdot \left(x_{1(1,m)} - t_{1m} \right) \right] & \dots & \prod_{k=1}^{K_M} \left[S_{Mm} \cdot \left(x_{1(M,m)} - t_{Mm} \right)_+ \right] \\ 1 & \prod_{k=1}^{K_1} \left[S_{1m} \cdot \left(x_{2(1,m)} - t_{1m} \right) \right] & \dots & \prod_{k=1}^{K_M} \left[S_{Mm} \cdot \left(x_{2(M,m)} - t_{Mm} \right)_+ \right] \\ \vdots & \vdots & \dots & \vdots \\ 1 & \prod_{k=1}^{K_1} \left[S_{1m} \cdot \left(x_{n(1,m)} - t_{1m} \right) \right] & \dots & \prod_{k=1}^{K_M} \left[S_{Mm} \cdot \left(x_{n(M,m)} - t_{Mm} \right)_+ \right] \end{bmatrix}$$

Persamaan (4.7) disebut sebagai model MARS klasifikasi (respon biner) dan dalam bentuk fungsi probabilitas dapat dinyatakan sebagai

$$\lambda = \delta_L(\mathbf{B}^T \beta) \quad (4.8)$$

Untuk memperoleh estimasi β digunakan metode estimasi maksimum likelihood.

Langkah pertama dalam metode estimasi maksimum likelihood, yaitu mengambil sampel random sebanyak n . Misalkan variabel random Y , $y \sim \text{Bernoulli}(1, \delta_L(\mathbf{B}\beta))$ dan dinyatakan dalam model, $y = \delta_L(\mathbf{B}\beta) + b$ maka $b = y - \delta_L(\mathbf{B}\beta)$. Jika fungsi δ_L diterapkan pada vektor $z = (z_1, \dots, z_1)^T \in R^n$

mempunyai interpretasi sebagai $\delta_L = (\delta_L(z_1), \dots, \delta_L(z_n))^T$. Analog persamaan (4.1) dan (4.2) diperoleh,

$$\begin{aligned}
E(b) &= 0 \\
a_i^2 &= \text{Var}(b_i) \\
&= \text{Var}(y - \delta_L(\mathbf{B}_i^T \boldsymbol{\beta})) \\
&= \delta_L(\mathbf{B}_i^T \boldsymbol{\beta})(1 - \delta_L(\mathbf{B}_i^T \boldsymbol{\beta})) \\
&= \frac{e^{\mathbf{B}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}} \left[1 - \frac{e^{\mathbf{B}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}} \right] \\
&= \frac{e^{\mathbf{B}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}} \left[\frac{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}} - e^{\mathbf{B}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}} \right] \\
&= \frac{e^{\mathbf{B}_i^T \boldsymbol{\beta}}}{(1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}})^2} = \delta'_L(\mathbf{B}_i^T \boldsymbol{\beta})
\end{aligned}$$

Langkah selanjutnya adalah membentuk fungsi *likelihood* $L(\boldsymbol{\beta})$ atau fungsi distribusi probabilitas bersama yang sebagai berikut:

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^n P(z = z_i) \\
&= \prod_{i=1}^n \binom{1}{z_i} [\delta_L(\mathbf{B}_i^T \boldsymbol{\beta})]^{z_i} [1 - \delta_L(\mathbf{B}_i^T \boldsymbol{\beta})]^{1-z_i} \\
&= \prod_{i=1}^n \frac{1}{z_i!} [\delta_L(\mathbf{B}_i^T \boldsymbol{\beta})]^{z_i} [1 - \delta_L(\mathbf{B}_i^T \boldsymbol{\beta})]^{1-z_i} \\
&= \prod_{i=1}^n \left(\frac{\delta_L(\mathbf{B}_i^T \boldsymbol{\beta})}{1 - \delta_L(\mathbf{B}_i^T \boldsymbol{\beta})} \right)^{z_i} [1 - \delta_L(\mathbf{B}_i^T \boldsymbol{\beta})] \\
&= \prod_{i=1}^n \left(\frac{\frac{e^{\mathbf{B}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}}}{\frac{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}} - e^{\mathbf{B}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}}} \right)^{z_i} \left(\frac{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}} - e^{\mathbf{B}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}} \right) \\
&= \prod_{i=1}^n (e^{\mathbf{B}_i^T \boldsymbol{\beta}})^{z_i} \left(\frac{1}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}} \right) \\
&= \prod_{i=1}^n \frac{e^{\mathbf{B}_i^T \boldsymbol{\beta} z_i}}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}}
\end{aligned}$$

Selanjutnya adalah memaksimumkan $L(\boldsymbol{\beta})$ atau fungsi \ln *likelihood* dari $L(\boldsymbol{\beta})$, sehingga diperoleh

$$\begin{aligned}
\ln L(\boldsymbol{\beta}) &= \ln \left(\prod_{i=1}^n \frac{e^{\mathbf{B}_i^T \boldsymbol{\beta} z_i}}{1 + e^{\mathbf{B}_i^T \boldsymbol{\beta}}} \right) \\
&= \sum_{i=1}^n \ln(\exp(\mathbf{B}_i^T \boldsymbol{\beta} z_i)) - \sum_{i=1}^n \ln(\exp(\mathbf{B}_i^T \boldsymbol{\beta})) \\
&= \sum_{i=1}^n \mathbf{B}_i^T \boldsymbol{\beta} z_i - \sum_{i=1}^n \ln(\exp(\mathbf{B}_i^T \boldsymbol{\beta})) \\
\ln L(\boldsymbol{\beta}) &= \mathbf{z}^T \mathbf{B} \boldsymbol{\beta} - \sum_{i=1}^n \ln(\exp(\mathbf{B}_i^T \boldsymbol{\beta})) \tag{4.9}
\end{aligned}$$

Langkah selanjutnya, yaitu mencari turunan pertama dari fungsi \ln *likelihood*nya ($l(\boldsymbol{\beta})$) terhadap $\boldsymbol{\beta}$ dan untuk menunjukkan estimator maksimum, maka diturunkan fungsi log-likelihood terhadap $\boldsymbol{\beta}$, yaitu sebagai berikut:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} \ln L(\boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}} \left(\mathbf{z}^T \mathbf{B} \boldsymbol{\beta} - \sum_{i=1}^n \ln(\exp(\mathbf{B}_i^T \boldsymbol{\beta})) \right) \\
&= \mathbf{B}^T \mathbf{z} - \sum_{i=1}^n \frac{\exp(\mathbf{B}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{B}_i^T \boldsymbol{\beta})} \mathbf{B}_i \\
&= \mathbf{B}^T \mathbf{z} - \mathbf{B}^T \sigma_L(\mathbf{B} \boldsymbol{\beta}) \\
&= \mathbf{B}^T \left[\mathbf{z} - \sigma_L(\mathbf{B} \boldsymbol{\beta}) \right]
\end{aligned}$$

Jika turunan pertama disamadengankan dengan nol, maka $\hat{\boldsymbol{\beta}}$ adalah estimator maksimum *likelihood* dari $\boldsymbol{\beta}$, yaitu:

$$\mathbf{B}^T \left[\mathbf{z} - \sigma_L(\mathbf{B} \hat{\boldsymbol{\beta}}) \right] = 0 \tag{4.10}$$

Untuk membuktikan $\hat{\boldsymbol{\beta}}$ adalah estimator yang memaksimumkan persamaan (4.7) maka dicari turunan kedua dari fungsi \ln *likelihood*nya ($l(\boldsymbol{\beta})$) pada persamaan (4.9) terhadap $\boldsymbol{\beta}$, sehingga diperoleh:

$$\begin{aligned}
\frac{\partial^2}{\partial \boldsymbol{\beta}_a \partial \boldsymbol{\beta}_b} \ln L(\boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}_a} \left(\mathbf{B}^T [z - \sigma_L(\mathbf{B}\boldsymbol{\beta})] \right) \\
&= \frac{\partial}{\partial \boldsymbol{\beta}_a} \left(\sum_{i=1}^n \mathbf{B}_{ib} z_i - \sum_{i=1}^n \mathbf{B}_{ib} \sigma_L(\mathbf{B}_i^T \boldsymbol{\beta}) \right) \\
&= - \sum_{i=1}^n \mathbf{B}_{ib} z_i \frac{\partial}{\partial \boldsymbol{\beta}_b} \left(\sigma_L(\mathbf{B}_i^T \boldsymbol{\beta}) \right) \\
&= - \sum_{i=1}^n \mathbf{B}_{ia} \mathbf{B}_{ib} \sigma_L'(\mathbf{B}_i^T \boldsymbol{\beta}) \\
&= - \mathbf{B}^T D(\mathbf{B})
\end{aligned}$$

Sehingga diperoleh matriks Hessian $H_{\ln L}(\boldsymbol{\beta}) \in \mathfrak{R}^p$, yaitu:

$$H_{\ln L}(\boldsymbol{\beta}) = -\mathbf{B}^T D(\mathbf{B})$$

dimana $D(\boldsymbol{\beta}) = d_{ij} \in \mathfrak{R}^n$ adalah matriks diagonal yang didefinisikan sebagai,

$$d_{ij} = \begin{cases} \delta'_L(\mathbf{B}_i^T \boldsymbol{\beta}), & \text{jika } i = j \\ 0 & \text{jika } i \neq j \end{cases}$$

dan $H_{\ln L}(\boldsymbol{\beta})$ semi definit untuk setiap $\boldsymbol{\beta} \in \mathfrak{R}^p$. Sehingga diperoleh:

$$\mathbf{u}^T H_{\ln L}(\boldsymbol{\beta}) \mathbf{u} = -\mathbf{u}^T \mathbf{B}^T D(\boldsymbol{\beta}) \mathbf{B} \mathbf{u} = -\sum_{i=1}^n (\mathbf{B}_i^T \mathbf{u})^2 \delta'_L(\mathbf{B}_i^T \boldsymbol{\beta}) \quad (4.11)$$

Jadi, turunan pertama selalu positif, sedangkan dari persamaan (4.11) diperoleh bahwa $\mathbf{u}^T H_{\ln L}(\boldsymbol{\beta}) \mathbf{u} \leq 0$ untuk semua $\mathbf{u} \in \mathfrak{R}^p$ dan $\boldsymbol{\beta} \in \mathfrak{R}^p$. Maka dapat disimpulkan bahwa $\hat{\boldsymbol{\beta}}$ adalah estimator yang memaksimumkan.

4.2 Analisis Propensity Score Matching (PSM) pada Kasus Neuropati Diabetik Perifer (NDP) Pasien Diabetes Melitus (DM) tahun 2015-2016

Analisis PSM diawali dengan mengetahui gambaran umum tentang karakteristik pasien DM tipe-2 yang digunakan dalam penelitian ini. Setelah itu dilakukan tahapan analisis PSM, yakni pemilihan variabel *confounding* sampai memodelan hubungan antara kovariat X dengan status NDP (Y).

4.2.1 Gambaran Umum Karakteristik Pasien DM Tipe-2

Analisis deskriptif merupakan tahap awal eksplorasi data yang dilakukan untuk mendapatkan gambaran umum dari data penelitian. Karakteristik responden dapat dilihat dari deskriptif pada masing-masing variabel. Variabel yang digunakan dalam penelitian ini, terdiri atas dua yaitu variabel data kontinu yakni umur (X_1) dan lama menderita DM (X_4) dan variabel data kategorik yakni jenis kelamin (X_2), dislipidemia (X_3), obesitas (X_5), status kadar HbA1C (X_6), status kadar AUS (X_7). Untuk data kontinu dideskripsikan berdasarkan rata-rata (mean), standar deviasi, nilai minimum dan nilai maksimum. Untuk data kategori karakteristik pasien dideskripsikan berdasarkan status NDP (Y). Karakteristik pasien untuk masing-masing variabel ditampilkan pada Tabel 4.1 dan Tabel 4.2.

Tabel 4.1 Karakteristik Pasien berdasarkan Umur (X_1) dan Lama Menderita DM (X_2)

Kovariat (X)	Mean (tahun)	Standar Deviasi (tahun)	Minimum (tahun)	Maksimum (tahun)
Umur (X_1)	59,848	8,972	33	79
Lama Menderita DM (X_4)	5,138	3,904	1	20

Tabel 4.1 memberikan gambaran tentang umur pasien dan lama pasien menderita DM tipe-2. Berdasarkan Tabel 4.1 diketahui bahwa rentang umur pasien DM yakni dari 33 tahun sampai dengan umur 79 tahun, dimana rata-rata umur pasien adalah 59,85 tahun dan standar deviasi 8,97. Selain itu, rentang lama pasien menderita DM adalah 1 tahun sampai 20 tahun dengan rata-rata 5,14 tahun dan standar deviasi 3,90.

Tabel 4.2 Karakteristik Pasien berdasarkan status NDP untuk Data Kategorik

Kovariat (X)	Status NDP		Total (%)
	Tidak Ada	Ada	
Jenis Kelamin (X_2)			
Laki-laki	25	33	40%
Perempuan	27	60	60%

Tabel 4.2 Karakteristik Pasien berdasarkan status NDP untuk Data Kategorik (Lanjutan)

Kovariat (X)	Status NDP		Total (%)
	Tidak Ada	Ada	
Dislipidemia (X ₃)			
Tidak	34	54	61%
Ya	18	39	39%
Obesitas (X ₅)			
Tidak	28	56	58%
Ya	24	37	42%
Kadar HbA1C (X ₆)			
Rendah	34	42	52%
Tinggi	18	51	48%
Kadar AUS (X ₇)			
Rendah	37	41	54%
Tinggi	15	52	46%
Total (%)	36%	64%	

Tabel 4.2 memberikan informasi tentang karakteristik pasien berdasarkan status NDP untuk data kategori. Berdasarkan Tabel 4.2 diketahui bahwa sebagian besar pasien DM tipe-2 terindikasi memiliki Neuropati Diabetik Perifer (NDP). Hal ini dapat dilihat dari jumlah proporsi pasien yang terindikasi NDP adalah 64% dibandingkan dengan proporsi yang tidak terindikasi NDP.

Menurut jenis kelamin, untuk data ini diperoleh bahwa pasien DM perempuan lebih banyak terindikasi memiliki NDP (60 pasien) dibandingkan pasien DM laki-laki (33 pasien), sedangkan pasien laki-laki lebih banyak tidak terindikasi NDP (25 pasien) dibandingkan pasien perempuan (27 pasien).

Selain jenis kelamin, salah satu faktor yang dianggap berpengaruh terhadap status NDP adalah dislipidemia. Dislipidemia merupakan kelainan metabolisme lipid yang ditandai peningkatan kolesterol total, kolesterol LDL, trigliserida di atas nilai normal serta penurunan kolesterol HDL di dalam darah. Kategori dislipidemia yang digunakan dalam penelitian ini adalah kenaikan kadar kolestrol total dalam darah dimana untuk kolestrol total ≤ 200 mg/dL dikategorikan tidak ada dislipidemia sedangkan kolestrol total > 200 mg/dL dikategorikan ada dislipidemia. Berdasarkan Tabel 4.2 diketahui bahwa sebagian

besar pasien yang jadi subjek penelitian ternyata ada dislipidemia. Untuk pasien yang terindikasi NDP banyaknya pasien masing-masing 54 pasien untuk yang tidak dislipidemia dan 39 pasien untuk yang dislipidemia. Ini berarti bahwa pasien DM yang tidak dislipidemia sebagian besar terindikasi NDP untuk data ini.

Variabel selanjutnya yang menjadi perhatian adalah obesitas. Obesitas merupakan keadaan patologis karena penimbunan lemak berlebihan daripada yang diperlukan untuk fungsi tubuh. Berdasarkan Tabel 4.2 diketahui bahwa banyaknya pasien yang terindikasi NDP dan obesitas lebih banyak dibandingkan dengan pasien yang NDP dan tidak obesitas. Hal ini dapat dilihat dari banyaknya pasien terindikasi NDP dan obesitas yakni 37 pasien dan tidak obesitas sebanyak 56 pasien. Ini berarti bahwa pasien DM yang terindikasi NDP sebagian besar tidak obesitas untuk data ini.

Komponen yang terpenting dari glikasi hemoglobin dalam penyakit DM adalah HbA1C. HbA1C digunakan sebagai patokan utama untuk pengendalian penyakit DM karena HbA1C dapat menggambarkan kadar gula darah dalam rentang 1 – 3 bulan karena usia sel darah merah yang terikat oleh molekul glukosa adalah 120 hari. Berdasarkan Tabel 4.2 diketahui bahwa sebagian besar pasien memiliki kadar HbA1C tinggi. Menurut kategori kadar HbA1C, pasien DM yang memiliki kadar HbA1C rendah lebih banyak terindikasi NDP dibandingkan pasien DM yang memiliki kadar HbA1C tinggi, dengan banyaknya pasien masing-masing adalah 41 pasien dan 52 pasien.

Variabel kategori yang terakhir adalah kadar Asam Urat Serum atau disingkat kadar AUS. Asam urat serum adalah jumlah asam urat dalam serum darah, dinyatakan dalam satuan mg/dl yang diperoleh dari standar pemeriksaan laboratorium. Peningkatan asam urat serum atau hiperurisemia merupakan hal yang umum ditemukan pada penderita DM tipe-2. Berdasarkan Tabel 4.2 diketahui bahwa sebagian besar pasien memiliki kadar AUS yang tinggi. Menurut kategori kadar AUS diketahui bahwa pasien DM yang memiliki kadar AUS tinggi lebih banyak terindikasi memiliki NDP dibandingkan pasien DM yang memiliki kadar AUS yang rendah, dengan banyaknya pasien masing-masing sebanyak 52 pasien dan 41 pasien. Hal ini sesuai dengan teori yang mengatakan bahwa

semakin tinggi kadar AUS pasien DM maka semakin tinggi pula kemungkinan pasien menderita NDP.

4.2.2 Analisis *Propensity Score Matching* dengan MARS

Pada bagian ini dibahas tentang langkah-langkah analisis *Propensity Score Matching* mulai dari tahap awal yakni pemilihan variabel *confounding* sampai tahap interpretasi.

4.2.2.1 Pemilihan Variabel *Confounding*

Langkah awal dalam analisis *Propensity Score Matching* adalah memilih kovariat sebagai variabel *confounder*. Variabel *confounder* dipilih hasil uji empiris berupa hubungan antar variabel. Statistik uji yang dapat digunakan untuk memeriksa hubungan antar variabel adalah uji *chi-square* dengan hipotesis sebagai berikut.

H₀: Tidak terdapat hubungan yang signifikan antar variabel

H₁: Terdapat hubungan yang signifikan antar variabel

Statistik uji yang digunakan adalah uji *chi-square* sesuai persamaan (2.9) dengan daerah kritik: H₀ ditolak jika $\chi_{hit}^2 > \chi_1^2$ atau $p\text{-value} < \alpha$.

Berdasarkan teori Neuropati Diabetik Perifer (NDP) dan faktor-faktor yang dijelaskan pada tinjauan pustaka diketahui bahwa faktor-faktor yang mempengaruhi kejadian NDP, yaitu faktor umur jenis kelamin, dislipidemia, obesitas, status kadar HbA1C, status kadar AUS (Veves dan Malik, 2007) dan lama DM (Bruce dan Young, 2008). Akan tetapi menurut teori sebelumnya, faktor-faktor yang mempengaruhi status kadar AUS yaitu faktor umur, jenis kelamin, dislipidemia, obesitas (Sadr et al., 2009). Dari penelitian diatas diketahui bahwa variabel-variabel tersebut merupakan faktor resiko dari kejadian NDP. Lebih lanjut, akan dilakukan uji empiris untuk melihat hubungan dari tiap variabel sebagai *confounding* dengan variabel kovariat lainnya.

Hipotesis yang digunakan untuk mengetahui hubungan variabel yang diduga sebagai *confounding* dengan kovariat X adalah sebagai berikut.

H_0 : Tidak terdapat hubungan signifikan antar variabel yang berpotensi sebagai *confounding* X_m dan variabel kovariat lainnya X ; $m = 1, 2, 3, 4, 5, 6$

H_1 : Terdapat hubungan signifikan antar variabel yang berpotensi sebagai *confounding* X_m dan variable kovariat lainnya; $m = 1, 2, 3, 4, 5, 6$

Berdasarkan Tabel 4.3 dan dengan menggunakan statistik uji chi-square pada taraf signifikansi $\alpha = 5\%$, diperoleh hasil pengujian seperti yang ditampilkan pada Tabel 4.3. Hasil pengujian selengkapnya dapat dilihat di lampiran 7.

Tabel 4.3 Pengujian Hubungan Variabel yang Berpotensi sebagai *Confounding* dengan Kovariat (X) lainnya

Variabel	χ^2	<i>p-value</i>	Keputusan
X_1 (Umur)			
$X_2 * X_1$	2.175	0.140	Gagal Tolak H_0
$X_3 * X_1$	3.626	0.057	Gagal Tolak H_0
$X_4 * X_1$	0.181	0.671	Gagal Tolak H_0
$X_5 * X_1$	1.592	0.207	Gagal Tolak H_0
$X_6 * X_1$	0.535	0.464	Gagal Tolak H_0
$X_7 * X_1$	0.360	0.549	Gagal Tolak H_0
X_2 (Jenis kelamin)			
$X_1 * X_2$	2.175	0.140	Gagal Tolak H_0
$X_3 * X_2$	4.052	0.044	Tolak H_0^{**}
$X_4 * X_2$	0.062	0.804	Gagal Tolak H_0
$X_5 * X_2$	1.528	0.216	Gagal Tolak H_0
$X_6 * X_2$	0.226	0.635	Gagal Tolak H_0
$X_7 * X_2$	1.184	0.277	Gagal Tolak H_0
X_3 (Dislipidemia)			
$X_1 * X_3$	3.626	0.057	Gagal Tolak H_0
$X_2 * X_3$	4.052	0.044	Tolak H_0^{**}
$X_4 * X_3$	0.822	0.365	Gagal Tolak H_0
$X_5 * X_3$	0.124	0.725	Gagal Tolak H_0
$X_6 * X_3$	0.408	0.523	Gagal Tolak H_0
$X_7 * X_3$	0.824	0.364	Gagal Tolak H_0
X_4 (Lama menderita DM)			
$X_1 * X_4$	0.181	0.671	Gagal Tolak H_0
$X_2 * X_4$	0.062	0.804	Gagal Tolak H_0
$X_3 * X_4$	0.822	0.365	Gagal Tolak H_0
$X_5 * X_4$	0.155	0.694	Gagal Tolak H_0
$X_6 * X_4$	20.814	0.000	Tolak H_0^{**}
$X_7 * X_4$	22.503	0.000	Tolak H_0^{**}

** $p\text{-value} < \alpha$

Tabel 4.3 Pengujian Hubungan Variabel yang Berpotensi sebagai *Confounding* dengan Kovariat (X) lainnya (Lanjutan)

Variabel	χ^2	df	<i>p-value</i>
X ₅ (Obesitas)			
X ₁ *X ₅	1.592	0.207	Gagal Tolak H ₀
X ₂ *X ₅	1.528	0.216	Gagal Tolak H ₀
X ₃ *X ₅	0.124	0.725	Gagal Tolak H ₀
X ₄ *X ₅	0.155	0.694	Gagal Tolak H ₀
X ₆ *X ₅	1.840	0.175	Gagal Tolak H ₀
X ₇ *X ₅	0.160	0.689	Gagal Tolak H ₀
X ₆ (Kadar HbA1C)			
X ₁ *X ₆	0.535	0.464	Gagal Tolak H ₀
X ₂ *X ₆	0.226	0.635	Gagal Tolak H ₀
X ₃ *X ₆	0.408	0.523	Gagal Tolak H ₀
X ₄ *X ₆	20.814	0.000	Tolak H ₀ **
X ₅ *X ₆	1.840	0.175	Gagal Tolak H ₀
X ₇ *X ₆	81.802	0.000	Tolak H ₀ **
X ₇ (Kadar AUS)			
X ₁ *X ₇	0.360	0.549	Gagal Tolak H ₀
X ₂ *X ₇	1.184	0.277	Gagal Tolak H ₀
X ₃ *X ₇	0.824	0.364	Gagal Tolak H ₀
X ₄ *X ₇	22.503	0.000	Tolak H ₀ **
X ₅ *X ₇	0.160	0.689	Gagal Tolak H ₀
X ₆ *X ₇	81.802	0.000	Tolak H ₀ **

**p-value < α

Tabel 4.3 memberikan informasi tentang uji hubungan antara variabel yang berpotensi sebagai *confounding* dengan kovariat lainnya. Untuk variabel jenis kelamin (X₂) sebagai variabel *confounding*, terdapat satu hubungan yang signifikan antara variabel yang berpotensi sebagai *confounding* (X₂) dengan kovariat lain, yaitu variabel dislipidemia (X₃). Hal yang sama terjadi ketika variabel dislipidemia (X₃) dipilih menjadi variabel *confounding*, hanya terdapat satu hubungan yang signifikan antara variabel yang berpotensi sebagai *confounding* dengan kovariat lain, yaitu variabel jenis kelamin (X₂).

Berbeda dengan hasil untuk variabel umur (X₁) dan variabel Obesitas (X₅) sebagai variabel *confounding*. Hasilnya menunjukkan bahwa tidak terdapat satupun hubungan yang signifikan antara variabel obesitas dengan kovariat lainnya. Sedangkan untuk variabel lama menderita DM (X₄), Kadar HbA1C (X₆),

dan Kadar AUS (X_7), ketiganya memiliki hasil yang hampir mirip. Ketiga variabel ini masing-masing terdapat dua hubungan yang signifikan antara variabel tersebut sebagai variabel *confounding* dengan kovariat lainnya, yaitu kadar HbA1C (X_6) dan kadar AUS (X_7) untuk variabel lama menderita DM (X_4), lama menderita DM (X_4) dan kadar AUS (X_7) untuk variabel kadar HbA1C (X_6), serta lama menderita DM (X_4) dan HbA1C (X_6) untuk variabel kadar kadar AUS (X_7).

Dari uji empiris variabel yang berpotensi sebagai *confounding* adalah variabel yang memiliki lebih banyak hubungan signifikan antara variabel tersebut saat menjadi *confounding*. Oleh sebab itu, variabel lama menderita DM (X_4), Kadar HbA1C (X_6), dan Kadar AUS (X_7) dipilih sebagai variabel yang berpotensi menjadi variabel *confounding*.

Selain dilakukan pengujian hubungan kovariat dengan variabel yang berpotensi sebagai *confounder*, dilakukan juga pengujian hubungan antara variabel yang berpotensi sebagai *confounder* dengan variabel status NDP (Y) untuk menentukan satu variabel *confounding* pada penelitian ini. Hipotesis yang digunakan adalah sebagai berikut.

H_0 : Tidak terdapat hubungan yang signifikan antar variabel yang berpotensi sebagai *confounding* X_m dan variable Y; $m = 1, 2, 3, 4, 5, 6$

H_1 : Terdapat hubungan yang signifikan antar variabel yang berpotensi sebagai *confounding* X_m dan variable Y; $m = 1, 2, 3, 4, 5, 6$

Berdasarkan Tabel 4.3 dan dengan menggunakan statistik uji chi-square pada taraf signifikansi $\alpha = 5\%$, diperoleh hasil pengujian seperti yang ditampilkan pada Tabel 4.3.

Tabel 4.3 Pengujian Hubungan Variabel yang Berpotensi sebagai *Confounding* dan Status NDP (Y)

Variabel	χ^2	<i>p-value</i>	Keputusan
Y*X4	9.036	0.003	Tolak H_0^{**}
Y*X6	5.469	0.019	Tolak H_0^{**}
Y*X7	9.831	0.002	Tolak H_0^{**}

** p -value $< \alpha$

Berdasarkan Tabel 4.3 diketahui bahwa ketiga variabel yang berpotensi sebagai variabel *confounding* yakni lama menderita DM (X_4), Kadar HbA1C (X_6),

dan Kadar AUS (X_7) sama-sama punya hubungan yang signifikan dengan status NDP. Namun dari ketiga variabel ini, variabel Kadar AUS (X_7) memiliki nilai *p-value* terkecil, sehingga variabel kadar AUS dipilih sebagai variabel *confounding* (z) dengan parameter θ untuk mengetahui seberapa besar pengaruh kadar AUS terhadap status NDP. Ini sesuai dengan teori, dimana pada referensi penelitian sebelumnya yakni penelitian yang dilakukan oleh Darsana (2014) membuktikan bahwa terdapat korelasi positif antara kadar AUS dengan status NDP.

4.2.2.2 Estimasi Nilai *Propensity Score* dengan *Multivariate Adaptive Regression Splines*

Setelah penentuan variabel *confounding* (z) maka langkah selanjutnya dari *Propensity Score Matching* (PSM) adalah estimasi nilai *propensity score*. Nilai *propensity* diestimasi menggunakan model MARS karena variabel *confounding* yang terpilih terdiri atas dua kategori dan jumlah variabel yang dimiliki cukup banyak (6 variabel). Nilai *propensity score* merupakan nilai yang menyatakan seberapa besar peluang seseorang untuk terkena suatu perlakuan berdasarkan kovariat yang diamati. Nilai *propensity score* dengan MARS dituliskan sebagai berikut:

$$e(\mathbf{x}_i) = \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}} = \frac{e^{(z)}}{1 + e^{(z)}} \quad (4.12)$$

dimana z diperoleh dari persamaan (4.4), yaitu

$$\text{logit } \lambda = \ln\left(\frac{\lambda}{1 - \lambda}\right) = z$$

Nilai *propensity score* ini akan diperoleh jika model MARS terbaik sudah dipeoleh.

Pada proses pembentukan model MARS, tiga hal yang perlu diperhatikan yaitu basis fungsi (BF), maksimum interaksi (MI), dan minimum observasi (MO). Basis fungsi (BF) merupakan fungsi yang didefinisikan dari setiap region, pada umumnya BF yang digunakan adalah dua sampai empat kali jumlah variabel prediktor (Friedman 1991). Variabel prediktor yang diduga mempengaruhi Neuropati Diabetik Perifer (NDP) adalah sebanyak 6 variabel, sehingga

banyaknya basis fungsi yang akan dikombinasikan dalam pembentukan model adalah 12, 18, dan 24.

Maksimum interaksi (MI) merupakan banyaknya interaksi yang bisa terjadi pada model. Maksimum interaksi yang digunakan dalam penelitian ini adalah 1, 2, dan 3, karena menurut Friedman (1991) jika maksimum interaksi yang digunakan lebih dari 3, maka nilai GCV akan semakin meningkat dan model yang digunakan akan semakin kompleks. Jika MI yang digunakan sebesar 1, artinya tidak ada interaksi antar variabel dalam model. Jika MI yang digunakan adalah 2, maka dalam model akan terjadi interaksi antara 2 variabel. Begitupun jika MI yang digunakan sebesar 3, berarti banyaknya interaksi dalam model yang paling banyak terjadi adalah antara 3 variabel, dan seterusnya.

Minimum observasi (MO) merupakan jumlah pengamatan paling minimal antar knot. Dalam penelitian ini MO yang digunakan adalah 0, 1, 2, 3, 4, dan 5 karena diatas itu nilai GCV akan semakin meningkat. Tahapan pembentukan model MARS dilakukan dengan *trial and error* untuk semua kombinasi nilai BF, MI, dan MO yang telah ditentukan sebelumnya. Banyaknya model yang mungkin berdasarkan kombinasi tersebut adalah sebanyak 54 model. Dari masing-masing pemodelan ini akan dihasilkan nilai GCV dan diperoleh variabel prediktor yang masuk dalam model. Hasil pemodelan MARS dengan semua kemungkinan model disajikan pada Tabel 4.4.

Metode pengujian yang digunakan adalah *k-fold cross validation*, dengan k merupakan variabel partisi yang diinginkan. Umumnya k yang digunakan dalam penelitian adalah 10. Karena jika semakin besar k yang dipilih maka komputasi yang dilakukan akan mahal untuk mengulang prosuder sebanyak N (banyaknya data) kali, sehingga dapat membuat variasi nilai kinerja menjadi sangat tinggi (Prasetyo, 2012).

Tabel 4.4 Hasil *Trial and Error* pembentukan Model MARS antara Kovariat (X) dan Kadar AUS (X₇)

Nomor Model	Kombinasi			GCV	R ²
	BF	MI	MO		
1	12	1	0	0.0928	0.6266
2	12	1	1	0.0934	0.6243
3	12	1	2	0.0934	0.6243
4	12	1	3	0.0934	0.6243
5	12	1	4	0.0950	0.6177
6	12	1	5	0.0959	0.6141
7	12	2	0	0.0829	0.6664
8	12	2	1	0.0820	0.6701
9	12	2	2	0.0818	0.6711
10	12	2	3	0.0809	0.6745
11	12	2	4	0.0862	0.6534
12	12	2	5	0.0830	0.6659
13	12	3	0	0.0774	0.6887
14	12	3	1	0.0779	0.6867
15	12	3	2	0.0781	0.6858
16	12	3	3	0.0786	0.6838
17	12	3	4	0.0801	0.6777
18	12	3	5	0.0774	0.6887
19	18	1	0	0.0864	0.6524
20	18	1	1	0.0918	0.6306
21	18	1	2	0.0919	0.6303
22	18	1	3	0.0910	0.6341
23	18	1	4	0.0882	0.6452
24	18	1	5	0.0903	0.6368
25	18	2	0	0.0750	0.6981
26	18	2	1	0.0766	0.6918
27	18	2	2	0.0763	0.6931
28	18	2	3	0.0755	0.6964
29	18	2	4	0.0793	0.6809
30	18	2	5	0.0786	0.6837
31	18	3	0	0.0713	0.7133
32	18	3	1	0.0696	0.7201
33	18	3	2	0.0697	0.7196
34	18	3	3	0.0728	0.7072
35	18	3	4	0.0744	0.7005
36	18	3	5	0.0722	0.7094

Tabel 4.4 Hasil Trial and Error Pembentukan Model MARS antara Kovariat (X) dan Kadar AUS (X₇) (Lanjutan)

Nomor Model	Kombinasi			GCV	R ²
	BF	MI	MO		
37	24	1	0	0.0842	0.6611
38	24	1	1	0.0821	0.6696
39	24	1	2	0.0851	0.6575
40	24	1	3	0.0830	0.6659
41	24	1	4	0.0850	0.6579
42	24	1	5	0.0877	0.6472
43	24	2	0	0.0703	0.7172
44	24	2	1	0.0715	0.7125
45	24	2	2	0.0724	0.7086
46	24	2	3	0.0709	0.7149
47	24	2	4	0.0753	0.6972
48	24	2	5	0.0743	0.7009
49	24	3	0	0.0636	0.7442
50	24	3	1	0.0636	0.7441**
51	24	3	2	0.0643	0.7414
52	24	3	3	0.0688	0.7232
53	24	3	4	0.0670	0.7305
54	24	3	5	0.0671	0.7302

**Nilai GCV terkecil dan R² terbesar

Berdasarkan Kombinasi BF, MI dan MO yang terkecil, maka model terbaik adalah model ke 50 dengan BF, MI, dan MO berturut-turut sebesar 24, 3, 1. Hasil pengujian selengkapnya dapat dilihat di lampiran 9 dan 10. Dari hasil tersebut dipilih model ke 50, karena berdasarkan kombinasi tersebut maka didapatkan model nilai *propensity score* MARS adalah sebagai berikut:

$$\begin{aligned} \text{logit } \lambda &= \ln\left(\frac{\lambda}{1-\lambda}\right) = z \\ \frac{\lambda}{1-\lambda} &= e^z \\ \lambda &= e^z (1-\lambda) \\ e(\mathbf{x}_7) &= \frac{e^{(z)}}{1+e^{(z)}} = \frac{e^{\hat{f}(\mathbf{x}_7)}}{1+e^{\hat{f}(\mathbf{x}_7)}} \end{aligned} \tag{4.13}$$

dengan

$$\begin{aligned}
z = & 0,141 + 0,521X_{6,1} + 0,143 \max(0, X_1 - 63) - 0,004 \max(0, 65 - X_1) \\
& - 0,239 \max(0, X_1 - 65) + 0,440 \max(0, X_1 - 70) - 0,448 \max(0, X_1 - 72) \\
& - 0,023 \max(0, 6 - X_4) - 0,020 \max(0, X_4 - 6) - 0,103 \max(0, 62 - X_1)(X_{6,1}) \\
& - 0,034 \max(0, X_1 - 62)(X_{6,1}) + 0,108 \max(0, 65 - X_1)(X_{6,1}) \\
& + 0,037 \max(0, 4 - X_4)(X_{6,1}) - 0,018 \max(0, X_4 - 4)(X_{6,1}) \\
& + 0,059 \max(0, X_1 - 62)(X_{5,1})(X_{6,1}) - 0,219(X_{2,1}) \max(0, 4 - X_4)(X_{6,1})
\end{aligned}$$

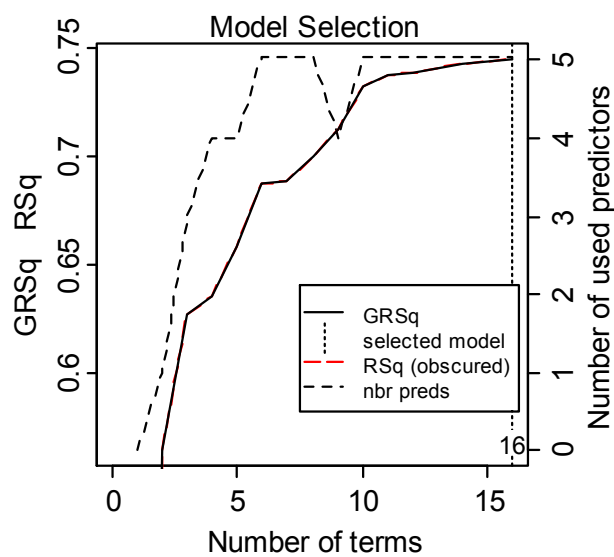
Interpretasi dari koefisien-koefisien fungsi basis pada model MARS diatas adalah

- $BF_1 = 0,141 + 0,521X_{6,1}$
Artinya, resiko seorang pasien diabetes mellitus mengalami indikasi NDP ($X_7 = 1$) pada pasien dengan kadar HbA1C tinggi ($X_6 = 1$) adalah sebesar 1/1,939 atau 0,516 kali lebih besar dibandingkan dengan pasien dengan kadar HbA1C trendah ($X_6 = 0$).
- $BF_2 = 0,143 \max(0, X_1 - 63)$
Artinya, 0,143 akan bermakna jika umur (X_1) pasien diabetes melitus lebih tinggi dari 63 tahun. Setiap kenaikan basis fungsi (BF_2) dapat menambah resiko seorang pasien diabetes mellitus terindikasi NDP sebesar 0,143.
- $BF_3 = -0,004 \max(0, 65 - X_1)$
Artinya, 0,004 akan bermakna jika umur (X_1) pasien diabetes melitus lebih rendah dari 65 tahun. Setiap kenaikan basis fungsi (BF_3) mengurangi resiko seorang pasien diabetes mellitus terindikasi NDP sebesar 0,004
- $BF_9 = -0,103 \max(0, 62 - X_1)(X_{6,1})$
Artinya, 0,103 akan bermakna jika umur (X_1) pasien diabetes melitus lebih rendah dari 65 tahun, kadar HbA1C-nya tinggi. Setiap kenaikan basis fungsi (BF_9) dapat mengurangi resiko seorang pasien diabetes mellitus terindikasi NDP sebesar 0,103.
- $BF_{14} = 0,059 \max(0, X_1 - 62)(X_{5,1})(X_{6,1})$
Artinya, 0,059 akan bermakna jika lama menderita DM (X_4) pasien diabetes melitus lebih tinggi dari 62 tahun, pasien memiliki obesitas, dan kadar HbA1C-nya tinggi. Setiap kenaikan basis fungsi (BF_{14}) dapat

menambah resiko seorang pasien diabetes mellitus terindikasi NDP sebesar 0,059.

- $BF_{15} = -0,219(X_{2,1})\max(0,4 - X_4)(X_{6,1})$

Artinya, 0,219 akan bermakna jika pasien diabetes melitus berjenis kelamin perempuan, lama menderita DM-nya (X_4) lebih rendah dari 4 tahun, dan kadar HbA1C-nya tinggi. Setiap kenaikan basis fungsi (BF_{15}) dapat mengurangi resiko seorang pasien diabetes mellitus terindikasi NDP sebesar 0,219.



Gambar 4.1 Penentuan Model Terbaik *Propensity Score* dengan Multivariate Adaptive Regression Splines

Dari Gambar 4.1 diperoleh bahwa model yang terpilih memiliki 16 term dengan 6 prediktor. Nilai R-square dan GR-square berada pada garis yang sama karena dalam penelitian ini penalty yang digunakan adalah 0. Jika penalti bertambah, maka garis GR-square akan bertambah sesuai dengan model dari parameter.

Dari model terbaik *propensity score* dengan MARS yang terpilih diperoleh matriks konfusi seperti pada Tabel 4.5. Hasil pengujian selengkapnya dapat dilihat di lampiran 12.

Tabel 4.5 Hasil Matriks Konfusi (Prasetyo, 2012)

f_{ij}		Kelas hasil prediksi (j)	
		Kelas = 1	Kelas = 0
Kelas asli (i)	Kelas = 1	59	3
	Kelas = 0	8	75

Dari Tabel 4.5 maka didapatkan akurasi dari model *propensity score* dengan MARS adalah sebagai berikut:

$$\text{Akurasi} = \frac{59 + 75}{59 + 8 + 3 + 75} = 0,924 = 92\% \quad (4.14)$$

Nilai akurasi yang diperoleh dari persamaan (4.14) adalah sebesar 92%, dimana 75 pasien dengan kadar AUS rendah yang diprediksi secara benar sebagai pasien DM dengan kadar AUS rendah dan 3 pasien dengan kadar AUS tinggi diprediksi secara salah sebagai pasien DM dengan kadar AUS rendah, serta 59 pasien dengan kadar AUS tinggi secara benar sebagai pasien DM dengan kadar AUS tinggi dan 8 pasien dengan kadar AUS rendah secara salah diprediksi sebagai pasien DM dengan kadar AUS tinggi.

4.2.2.3 Analisis *Matching*

Langkah kedua dalam *propensity score matching* adalah analisis *matching* (pencocokkan), dimana pada tulisan ini digunakan metode *Nearest Neighbor Matching*. Nilai estimasi *propensity score* (nilai prediksi dari kovariat X_7) biasanya disebut dengan *balance score*, dimana nilai (*score*) tersebut dilakukan untuk mencocokkan peserta pada kelompok *treatment* dengan peserta pada kelompok *non-treatment*.

Hasil analisis *matching* pada lampiran 14 diketahui bahwa dari 145 pasien DM hanya 67 pasien dengan kadar AUS rendah (kelompok *non-treatment*) yang dipasangkan ke pasien dengan kadar AUS tinggi (kelompok *treatment*). Pasien dengan kadar AUS rendah yang tidak memiliki pasangan ini kemudian dikeluarkan dari himpunan pasien dengan kadar AUS rendah (kelompok *non-treatment*) dan tidak diikutkan pada analisis selanjutnya.

4.2.2.4 Analisis *Postmatching*

Jika estimasi dari nilai *propensity score* telah diperoleh dan *matching* telah dilakukan, maka langkah selanjutnya adalah melakukan analisis *postmatching* dengan stratifikasi dari *propensity score*. Pada analisis *postmatching* akan dilakukan evaluasi dari *propensity score matching*, seperti uji keseimbangan (*balance*) kovariat dan estimasi *Average Treatment of Treated* (ATT) atau estimasi efek perlakuan (Tu dan Zhou, 2002).

a. Pengujian *balance* kovariat

Pada *propensity score* evaluasi ini digunakan untuk mengecek keseimbangan kovariat pada *confounders* antara kelompok *treatment* dan *non-treatment*. Pada penelitian ini digunakan uji kesamaan dua rata-rata dan uji kesamaan proporsi untuk melihat perbedaan antara kelompok *treatment* dan kelompok *non-treatment*.

Sebelum melakukan pengujian *balance*, terlebih dahulu dilakukan pengujian kesamaan variansi dengan uji Bartlett. Hipotesis yang digunakan adalah sebagai berikut:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

$$H_1 : \text{setidaknya ada dua variansi populasi yang tidak sama (misalnya } \sigma_i^2 \neq \sigma_j^2 \text{ untuk } i,j)$$

Untuk taraf signifikansi $\alpha = 5\%$ dan menggunakan uji Bartlett sesuai dengan persamaan (2.3). Hasil uji Bartlett menunjukkan bahwa nilai *p-value* lebih kecil dari taraf signifikansi $\alpha = 5\%$ ($0,000 < 0,05$), sehingga H_0 ditolak. Ini berarti setidaknya ada dua variansi populasi dari kelompok *treatment* dan *non-treatment* yang tidak sama. Hasil pengujian selengkapnya dapat dilihat di lampiran 14.

Setelah melakukan pengujian kesamaan variansi, maka langkah selanjutnya, yaitu menguji kesamaan dua rata-rata. Pengujian *balance* dibedakan menjadi dua, yakni pengujian untuk variabel data kategori dan variabel data kontinu. Untuk data kategori digunakan uji-z untuk menguji beda proporsi antara kelompok *treatment* dan kelompok *non-treatment*. Sedangkan untuk data kontinu digunakan uji-t untuk menguji perbedaan mean antara kelompok *treatment* dan kelompok

non-treatment. Hipotesis yang digunakan untuk menguji perbedaan proporsi dan mean pada variabel data kategori dan data kontinu adalah sebagai berikut.

1. Hipotesis uji beda mean untuk variabel umur (X_1)
 - H_0 : tidak ada perbedaan rata-rata umur pasien DM antara kelompok perlakuan dan kelompok *non-treatment*
 - H_1 : ada perbedaan rata-rata umur pasien DM antara kelompok perlakuan dan kelompok *non-treatment*
2. Hipotesis uji beda proporsi untuk variabel jenis kelamin (X_2)
 - H_0 : tidak ada perbedaan proporsi jenis kelamin pasien DM antara kelompok perlakuan dan kelompok *non-treatment*
 - H_1 : ada perbedaan proporsi jenis kelamin pasien DM antara kelompok perlakuan dan kelompok *non-treatment*
3. Hipotesis uji beda proporsi untuk variabel dislipidemia (X_3)
 - H_0 : tidak ada perbedaan proporsi dislipidemia pasien DM antara kelompok perlakuan dan kelompok *non-treatment*
 - H_1 : ada perbedaan proporsi dislipidemia pasien DM antara kelompok perlakuan dan kelompok *non-treatment*
4. Hipotesis uji beda mean untuk variabel lama DM (X_4)
 - H_0 : tidak ada perbedaan rata-rata lama pasien menderita DM antara kelompok perlakuan dan kelompok *non-treatment*
 - H_1 : ada perbedaan rata-rata lama pasien menderita DM antara kelompok perlakuan dan kelompok *non-treatment*
5. Hipotesis uji beda proporsi untuk variabel obesitas (X_5)
 - H_0 : tidak ada perbedaan proporsi kejadian obesitas pasien DM antara kelompok perlakuan dan kelompok *non-treatment*
 - H_1 : ada perbedaan proporsi kejadian obesitas pasien DM antara kelompok perlakuan dan kelompok *non-treatment*
6. Hipotesis uji beda proporsi untuk variabel kadar HbA1C (X_6)
 - H_0 : tidak ada perbedaan proporsi kadar HbA1C pasien DM antara kelompok perlakuan dan kelompok *non-treatment*
 - H_1 : ada perbedaan proporsi kadar HbA1C pasien DM antara kelompok perlakuan dan kelompok *non-treatment*

Untuk taraf signifikansi $\alpha = 10\%$ dan menggunakan uji-z sesuai dengan persamaan (2.8) dan uji-t untuk varians tidak homogen sesuai dengan persamaan (2.6) diperoleh hasil pengujian *balance* untuk setiap kovariat seperti pada Tabel 4.6. Hasil selengkapnya dapat dilihat pada lampiran 14.

Tabel 4.6 Hasil pengujian *balance*

Variabel	balance	<i>p-value</i>	Variabel	balance	<i>p-value</i>
X ₁	Sebelum	0,094	X ₄	Sebelum	0,000
	Setelah	0,142**		Setelah	0,000
X ₂	Sebelum	0,155	X ₅	Sebelum	0,715
	Setelah	0,412**		Setelah	0,866**
X ₃	Sebelum	0,218	X ₆	Sebelum	0,000
	Setelah	0,239**		Setelah	0,000

***p-value* > α

Dari Tabel 4.6 diperoleh bahwa setelah melakukan analisis *propensity score matching* dengan MARS, variabel Lama menderita Diabetes Melitus (X₄) dan kadar HbA1C (X₆) tidak *balance*. Hal ini berarti bahwa kedua variabel tersebut menunjukkan adanya perbedaan rata-rata antar kelompok *treatment* dan *non-treatment*. Adanya perbedaan rata-rata antara kelompok *treatment* dan *non-treatment* menyebabkan kesimpulan dari studi observasi menjadi bias karena tidak adanya randomisasi. Sehingga dapat disimpulkan bahwa variabel Lama menderita Diabetes Melitus (X₄) dan kadar HbA1C (X₆) merupakan variabel yang bukan berpengaruh secara langsung terhadap status NDP pasien DM. Dari pengujian *balance* juga diketahui bahwa variabel Umur (X₁), Jenis Kelamin (X₂), Dislipidemia (X₃) dan Obesitas (X₅) tidak menunjukkan adanya perbedaan rata-rata antar kelompok *treatment* dan *non-treatment*. Sehingga dapat disimpulkan bahwa keempat variabel ini merupakan variabel yang berpengaruh secara langsung terhadap status NDP pasien DM.

b. Estimasi *Average Treatment of Treated*

Setelah pengujian *balance* dilakukan, langkah selanjutnya dari *Propensity Score Matching* (PSM) dengan MARS adalah estimasi nilai *Treatment for Treated* (ATT) dan standar error. Estimasi nilai ATT merupakan hal penting dalam

propensity score karena hal ini merupakan tujuan utama digunakan metode *propensity*. Estimasi nilai ATT dan standar error dilakukan dengan menggunakan persamaan (2.10).

Estimasi ATT dilakukan untuk mengetahui seberapa besar pengaruh kadar AUS (Z) terhadap kejadian NDP (Y) pada saat pengaruh dari variabel lain (kovariat) sudah direduksi. Karena dari hasil pengujian *balance* diperoleh bahwa variabel Lama menderita Diabetes Melitus (X₄) dan kadar HbA1C (X₆) merupakan variabel yang tidak *balance* maka perlu dilakukan kembali PSM dengan MARS tanpa mengikutsertakan kedua variabel tersebut. Hal ini bertujuan untuk melihat nilai signifikansi terkecil untuk estimasi ATT. Setelah dilakukan kembali PSM dengan MARS tanpa mengikutsertakan variabel Lama menderita Diabetes Melitus (X₄) dan kadar HbA1C (X₆), maka langkah selanjutnya adalah menguji kembali *balance* dari hasil PSM dengan MARS. Hasil pengujian selengkapnya dapat dilihat pada lampiran 15 dan 16.

Dari pengujian tersebut diperoleh bahwa seluruh variabel telah *balance*. Hal ini berarti bahwa variabel Umur (X₁), Jenis Kelamin (X₂), Dislipidemia (X₃) dan Obesitas (X₅) menunjukkan tidak adanya perbedaan rata-rata antar kelompok *treatment* dan *non-treatment*.

Setelah pengujian *balance* dilakukan, langkah selanjutnya adalah menghitung nilai *Average Treatment for Treated* (ATT) dan standar error untuk masing-masing persamaan. Hasil estimasi ditunjukkan pada Tabel 4.7 berikut.

Tabel 4.7 Hasil Estimasi ATT dan Standar *Error*

Variabel yang masuk	ATT θ	SE (ATT)	T	<i>p-value</i>
X ₁ , X ₂ , X ₃ , X ₄ , X ₅ , X ₆	0,254	0,080	3,0002	0,0015**
X ₁ , X ₂ , X ₃ , X ₅	0,239	0,034	3,2539	0,0011**

***p-value* < α

Tabel 4.7 memberikan informasi bahwa masing-masing estimasi memberikan hasil yang berbedan namun memberikan kesimpulan signifikansi yang sama yakni variabel kadar AUS (Z) berpengaruh secara signifikan terhadap kejadian Neuropati Diabetik Perifer (NDP) pada pasien DM. Tabel 4.7 juga

memberikan informasi bahwa nilai signifikansi terkecil dimiliki oleh hasil estimasi untuk persamaan kedua, karena variabel Lama menderita Diabetes Melitus (X_4) dan kadar HbA1C (X_6) yang menyebabkan penelitian tidak *balance* telah dikeluarkan dari model.

Berdasarkan Tabel 4.7 diketahui pula bahwa efek kadar asam urat tinggi terhadap status NDP pasien DM adalah sebesar 0,239 atau peluang pasien DM yang memiliki kadar asam urat tinggi untuk terindikasi NDP adalah 0,239 kali lebih besar dibandingkan pasien yang memiliki kadar asam urat rendah.

Setelah diperoleh estimasi nilai ATT dan standar error, langkah selanjutnya adalah membuat model hubungan antara variabel *confounding z* atau variabel kadar AUS (X_7) dengan variabel respon atau kejadian NDP (Y). Model hubungan dapat ditunjukkan pada Tabel 4.8.

Tabel 4.8 Model hubungan variabel kadar AUS dengan pasien NDP

Kovariat (X_7)	Status NDP	
	Tidak Ada	Ada
Rendah	23%	29%
Tinggi	12%	35%
Total	36%	64%

Tabel 4.8 memberikan informasi tentang hubungan variabel kadar AUS dengan pasien NDP. Berdasarkan Tabel 4.8 diketahui bahwa sebagian besar pasien memiliki kadar AUS yang tinggi. Hal ini dapat dilihat dari jumlah proporsi pasien yang terindikasi NDP dengan kadar AUS tinggi adalah sebesar 35% dibandingkan dengan proporsi pasien yang terindikasi NDP dengan kadar AUS rendah. Menurut kategori kadar AUS diketahui bahwa pasien DM yang memiliki kadar AUS tinggi lebih banyak terindikasi memiliki NDP dibandingkan pasien DM yang memiliki kadar AUS yang rendah, dengan banyaknya pasien masing-masing sebanyak 51 pasien dan 42 pasien. Hal ini sesuai dengan teori yang mengatakan bahwa semakin tinggi kadar AUS pasien DM maka semakin tinggi pula kemungkinan pasien menderita NDP.

(Halaman sengaja dikosongkan)

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis data dan pembahasan, maka kesimpulan yang diperoleh:

1. Hasil estimasi *propensity score* menggunakan MARS klasifikasi (respon biner) diperoleh

$$\lambda = \delta_L(\mathbf{B}^T \beta)$$

2. Analisis *Propensity Score Matching* (PSM) memberikan beberapa kesimpulan, yaitu:
 - Nilai akurasi yang diperoleh dari persamaan diatas adalah sebesar 92%,
 - Dari 145 pasien DM hanya 67 pasien dengan kadar AUS rendah (kelompok *non-treatment*) yang dipasangkan ke pasien dengan kadar AUS tinggi (kelompok *treatment*).
 - Dengan pengujian *balance* kovariat dengan *propensity score matching* menggunakan MARS diperoleh bahwa variabel lama menderita DM (X_4) dan kadar HbA1C (X_6) bukan merupakan variabel yang berpengaruh secara langsung terhadap status NDP pasien DM, sedangkan variabel umur (X_1), Jenis Kelamin (X_2), Dislipidemia (X_3), dan Obesitas (X_5) merupakan variabel yang berpengaruh secara langsung terhadap status NDP pasien DM.
 - Dari estimasi ATT diperoleh bahwa bahwa semakin tinggi kadar AUS (X_7) pasien DM maka semakin tinggi pula kemungkinan pasien menderita NDP. Karena kadar AUS merupakan variabel *confounding*, maka faktor-faktor pada penelitian ini dapat diwakili oleh satu faktor saja, yaitu kadar AUS.

5.2 Saran

Hal yang dapat disarankan untuk penelitian selanjutnya, yaitu:

1. Pada penelitian ini estimasi MARS dilakukan pada metode *propensity score* yang digunakan adalah metode *propensity score matching*, pada penelitian selanjutnya lebih baik jika melakukan *propensity score stratification* atau *propensity score weighting*, sehingga tidak hanya sebatas menyeimbangkan kelompok *treatment* dan *non-treatment* saja.
2. Pada penelitian ini hanya menggunakan metode *Nearest Neighbor Matching* untuk *Propensity score matching* dengan MARS, maka akan lebih baik jika pada penelitian selanjutnya, peneliti menggunakan metode matching lainnya, seperti metode *Nearest Neighbor Matching* dengan *caliper* atau *Full Matching*, sehingga tidak banyak anggota dari kelompok *non-treatment* yang terbuang pada penelitian.

DAFTAR PUSTAKA

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley and Sons, Inc.
- Arsham, H. (2010). Bartlett's Test. https://www.researchgate.net/publication/252322443_Bartlett's_Test
- Austin. (2011). *A Tutorial and Case Study in Propensity Score Analysis: Research*, 46: pp: 135-147.
- Bansal, D., Gudala, K., Muthyala, H., Esam, H.P, Nayakallu, R. & Bhansali, A. (2014). *Prevalence and Risk Factors of Development of Peripheral Diabetic Neuropathy in Type 2 DM in a Tertiary Care Setting*. *Journal Diabetes Invest*, Vol. 5: pp. 714–721.
- Bruce, N., Pope, D., & Stanistreet, D. (2008). *Quantitative Methods for Health research; A Practical Interactive Guide to Epidemiology and Statistics*. John Wiley & Sons, Ltd UK.
- Bruce, S.G., & Young, T.K. (2008). *Prevalence and Risk Factors for Neuropathy in a Canadian First Nation Community*. *Diabetes Care* vol. 31, pp. 1837-1841.
- Conen, D., Wietlisbach, V., Bovet, P., Shamlaye, C., Riesen, W., Paccaud, F., & Burnier, M. (2004). *Prevalence of Hyperuricemia and Relation of Serum Uric Acid with Cardiovascular risk factors in a developing country*. *Journal BMC Public Health*, Vol. 4, pp. 1-9.
- Darsana, I.N. (2014). *Korelasi Positif Kadar Asam Urat Serum Tinggi dengan Neuropati Diabetik Perifer pada Penderita DM Tipe-2 di RSUP Sanglah Denpasar*. TESIS UNUD.
- D' Agostino, R.B. (1998). *Tutorial in Biostatistics Propensity Score Method for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group*. 17, pp. 2265-2281.
- Friedman, J.H. and Silverman, B.W. (1989), *Flexible Parsimony Smoothing And Additive Modeling*. *Technometrics*, 31, 3 – 39.

- Friedman, J.H. (1991). *Multivariate Adaptive Regression Splines (with discussion)*. *Ann. Statist.* 19: 1 – 141.
- Guo, S. & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: Sage Publications.
- Kamangar, F. (2012). *Confounding Variables in Epidemiologic Studies: Basics and Beyond*. *Arch Iran Med.* 2012; 15(8): 508 – 516.
- Kemenkes. (2013). *Riset Kesehatan Dasar 2013*. Badan Penelitian dan Pengembangan Kesehatan Kemenkes RI.
- Leathwick JR, Rowe D, Richardson J, Elith J, Hastie T. 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology* 50: 2034–2052.
- Lestari, D., Hidayat, N., & Risanti, E.D. (2015). *Hubungan antara Lama Menderita DM tipe-2 dengan terjadinya Neuropati Sensorik Diabetik di RSUD Salatiga*. Skripsi FK UMS.
- Li, H., Graham, D.J., & Majumdar, A. (2013). *The Impacts of Speed Cameras on Road Accidents: An Application of Propensity Score Matching Methods*. *Accident Analysis and Prevention*, 60, 148-57.
- Littnerova, S., Jarkovsky, J., Parenica, J., Pavlik, T., Spinar, J., & Dusek, L. (2013). *Why to use Propensity Score in Observational Studies? Case Study Based on Data from the Czech Clinical Database AHEAD 2006-09*, *cor et Vasa*, 55(4), pp. 383-390.
- McCaffrey, D.F., Ridgeway, G., & Moral, A.R. (2004). *Propensity Score Estimation with Boosted Regression for Evaluating Causal Effect in Observational Studies*. *Psychological Method*, 9(4), pp 403.
- Ndraha, S. (2014). *Diabetes Melitus dan Tata Laksana Terkini*. Leading Article Vol. 27, No.2.
- Otok, W, B. (2010). *Pendekatan Multivariate Adaptive Regression Spline (MARS) pada Pengelompokan Zona Musim Suatu Wilayah*. *Statistika*, Vol. 10 No. 2, 107-120.
- Pan, W., & Bai, H. (2015). *Propensity Score Analysis: Fundamental and Developments*. New York: Gulford Press.

- Prasetyo, E. (2012). *DATA MINING: Mengolah Dat Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi
- Rosenbaum, P.R., & Rubin, D.B. (1983). *The Central Role of the Propensity Score in Observational Studies for Causal Effects*. *Journal Biometrika*, vol.70, No.1, pp 41-55.
- Tesfaye, S. 2004. Epidemiology and Etiology of Diabetic Peripheral Neuropathies. *Ad Stud Med*, 4: 1-8.
- Tesfaye, S., Stevens, L.K., Stephenson, J.M, Fuller, J.H, Plater, M., Tirgoviste, C.I., Nuber, A., Pozza, G, Ward, J.D., & the EURODIAB IDDM Group. (1996). *Prevalence of Diabetic Peripheral Neuropathy and Its Relation to Glycaemic Control and Potential Risk Factors: the EURODIAB IDDM Complication Study*. *Journal Diabetologia*, Vol. 39, pp. 1377-1384.
- Tu, W., & Zhou, X.H. (2002). *A Bootstrap Confidence Interval Procedure for the Treatment Effect Using Propensity Score Subclassification*. *Health Services & Outcomes Research Methodology* 3: 135–147.
- Veves, A., & Malik, R.A. (2007). *Diabetic Neuropathy Clinical Management Second Edition*. Totowa: Humana Press.
- WHO. (2001). *Indicators for the Minimum Data Set Project on Ageing: A Critical Review in Sub-Saharan Africa*. Dar es Salaam, United Republic of Tanzania.
- Wijayanti, D.N. (2013). *Analisis Faktor Penyebab Obesitas dan Cara Mengatasi Obesitas pada Remaja*. Skripsi UNNES.
- Yanovitzky, I., Zanutto, E., & Hornik, R. (2005). *Estimating Causal Effects of Public Health Education Campaigns using Propensity Score Methodology*. *Journal Elsevier Evaluation and Program Planning* 28 (2005) pp. 209–220.

(Halaman sengaja dikosongkan)

LAMPIRAN

Lampiran 1. Data Kasus Diabetes Melitus (DM)

No	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	1	0	0	1	1	1	1	1
2	1	0	1	0	0	1	1	0
3	1	1	0	0	1	1	1	1
4	1	1	0	0	0	0	0	1
5	1	1	1	0	0	0	0	1
6	1	1	0	1	0	0	0	1
7	1	0	1	0	0	0	0	0
8	1	0	1	1	0	1	1	1
9	0	1	0	0	0	0	0	1
10	0	1	0	0	1	0	0	1
11	0	0	0	0	0	1	1	0
12	0	1	0	0	1	0	0	1
13	1	1	0	1	1	1	1	1
14	1	0	0	0	0	0	0	0
15	1	1	0	0	1	0	1	1
16	1	1	0	0	0	0	0	1
17	1	0	1	0	0	0	0	1
18	0	0	0	0	1	0	0	0
19	1	1	0	0	0	0	0	0
20	1	1	0	0	0	0	0	0
21	1	0	1	0	1	0	1	1
22	1	1	1	0	0	1	1	1
23	1	0	0	0	0	1	0	1
24	1	0	0	0	0	1	1	1
25	1	1	1	0	1	0	0	1
.
.
.
.
141	1	1	1	0	1	0	0	1
142	1	1	1	1	0	1	1	1
143	0	0	1	0	0	1	1	1
144	1	1	0	0	1	0	1	1
145	1	0	0	0	0	0	0	1

Lampiran 2. Surat Izin Penggunaan Data



PEMERINTAH KABUPATEN GORONTALO
RSUD Dr. M.M. DUNDA LIMBOTO

Jl. Achmad A. Wahab No.53 Limboto Telp. (0435)881455-880047 Kode Pos 96219
Fax (0435) 881095 Website : www.rsudunda.com

SURAT KETERANGAN
NOMOR : 070 / 1841 / RSUD-DUNDA


Yang bertanda tangan dibawah ini :

N a m a : **dr. H. SUPANDI M. ABDULLAH. M.Kes**
N I P : 19621123 199103 1 006
Jabatan : Direktur

Menerangkan dengan benar bahwa :

N a m a : **AMANDA ADITYANINGRUM**
Pekerjaan : Mahasiswa Pasca Sarjana
Jurusan Statistika ITS Surabaya

Adalah benar-benar telah mengambil data DM pada bulan Agustus 2016 di RSUD Dr. M.M Dunda Limboto Kabupaten Gorontalo untuk digunakan sebagai data penelitian
Demikian surat keterangan ini dibuat untuk digunakan seperlunya.

Limboto, 01 Agustus 2016
DIREKTUR

dr. H. SUPANDI M. ABDULLAH. M.Kes
NIP. 19621123 199103 1 006

Lampiran 3. Sintaks tabulasi silang dan pengujian hubungan kovariat X dengan status NDP (Y) menggunakan *software R*

```

tableYX1 = table(X1,Y)
colnames(tableYX1)=c("Tidak Ada NDP", "Ada NDP")
rownames(tableYX1)=c("muda", "tua")
tableYX2 = table(X2,Y)
colnames(tableYX2)=c("Tidak Ada NDP", "Ada NDP")
rownames(tableYX2)=c("laki laki", "perempuan")
tableYX3 = table(X3,Y)
colnames(tableYX3)=c("Tidak Ada NDP", "Ada NDP")
rownames(tableYX3)=c("Tidak Dislipidemia", "Dislipidemia")
tableYX4 = table(X4,Y)
colnames(tableYX4)=c("Tidak Ada NDP", "Ada NDP")
rownames(tableYX4)=c("Lama", "Baru")
tableYX5 = table(X5,Y)
colnames(tableYX5)=c("Tidak Ada NDP", "Ada NDP")
rownames(tableYX5)=c("Tidak Obesitas", "Obesitas")
tableYX6 = table(X6,Y)
colnames(tableYX6)=c("Tidak Ada NDP", "Ada NDP")
rownames(tableYX6)=c("Kadar HbA1C tinggi", "Kadar HbA1C rendah")
tableYX7 = table(X7,Y)
colnames(tableYX7)=c("Tidak Ada NDP", "Ada NDP")
rownames(tableYX7)=c("Kadar AUS rendah", "Kadar AUS tinggi")
rbind(tableYX1,tableYX2,tableYX3,tableYX4,tableYX5,tableYX6,tableYX7)

```

Lampiran 4. Output tabulasi silang antara kovariat X dengan status NDP (Y)

	Tidak Ada NDP	Ada NDP
muda	8	12
tua	44	81
laki laki	25	33
perempuan	27	60
Tidak Dislipidemia	34	54
Dislipidemia	18	39
Lama	48	66
Baru	4	27
Tidak Obesitas	28	56
Obesitas	24	37
Kadar HbA1C tinggi	34	42
Kadar HbA1C rendah	18	51
Kadar AUS rendah	37	41
Kadar AUS tinggi	15	52

Lampiran 5. Sintaks pengujian hubungan antara variabel yang berpotensi sebagai *confounding* dengan kovariat X dalam menentukan variabel *confounding*

```

#Chisquare X1
tableX2X1 = table(X2,X1)
testX2X1 = chisq.test(tableX2X1, correct=FALSE)
tableX3X1 = table(X3,X1)
testX3X1 = chisq.test(tableX3X1, correct=FALSE)
testX4X1 = chisq.test(tableX4X1, correct=FALSE)
tableX4X1 = table(X4,X1)
testX4X1 = chisq.test(tableX4X1, correct=FALSE)

```

```

tableX5X1 = table(X5,X1)
testX5X1 = chisq.test(tableX5X1, correct=FALSE)
tableX6X1 = table(X6,X1)
testX6X1 = chisq.test(tableX6X1, correct=FALSE)
tableX7X1 = table(X7,X1)
testX7X1 = chisq.test(tableX7X1, correct=FALSE)
tableY1 = table(X7,Y)
testY1 = chisq.test(tableY1, correct=FALSE)
variabelX1=rbind("X2*X1","X3*X1","X4*X1","X5*X1","X6*X1","X7*X1",
"Y*X1")
xsquareX1=rbind(testX2X1$statistic, testX3X1$statistic,
testX4X1$statistic, testX5X1$statistic, testX6X1$statistic,
testX7X1$statistic, testY1$statistic)
pvalueX1=rbind(testX2X1$p.value, testX3X1$p.value,
testX4X1$p.value, testX5X1$p.value, testX6X1$p.value,
testX7X1$p.value, testY1$p.value)
confoundX1=data.frame(variabelX1, xsquareX1, pvalueX1)

#Chisquare X2
tableX1X2 = table(X1,X2)
testX1X2 = chisq.test(tableX1X2, correct=FALSE)
tableX3X2 = table(X3,X2)
testX3X2 = chisq.test(tableX3X2, correct=FALSE)
tableX4X2 = table(X4,X2)
testX4X2 = chisq.test(tableX4X2, correct=FALSE)
tableX5X2 = table(X5,X2)
testX5X2 = chisq.test(tableX5X2, correct=FALSE)
tableX6X2 = table(X6,X2)
testX6X2 = chisq.test(tableX6X2, correct=FALSE)
tableX7X2 = table(X7,X2)
testX7X2 = chisq.test(tableX7X2, correct=FALSE)
tableY2 = table(Y,X2)
testY2 = chisq.test(tableY2, correct=FALSE)
variabelX2=rbind("X1*X2","X3*X2","X4*X2","X5*X2","X6*X2","X7*X2",
"Y*X2")
xsquareX2=rbind(testX1X2$statistic, testX3X2$statistic,
testX4X2$statistic, testX5X2$statistic, testX6X2$statistic,
testX7X2$statistic, testY2$statistic)
pvalueX2=rbind(testX1X2$p.value, testX3X2$p.value,
testX4X2$p.value, testX5X2$p.value, testX6X2$p.value,
testX7X2$p.value, testY2$p.value)
confoundX2=data.frame(variabelX2, xsquareX2, pvalueX2)

#Chisquare X3
tableX1X3 = table(X1,X3)
testX1X3 = chisq.test(tableX1X3, correct=FALSE)
tableX2X3 = table(X2,X3)
testX2X3 = chisq.test(tableX2X3, correct=FALSE)
tableX4X3 = table(X4,X3)
testX4X3 = chisq.test(tableX4X3, correct=FALSE)
tableX5X3 = table(X5,X3)
testX5X3 = chisq.test(tableX5X3, correct=FALSE)
tableX6X3 = table(X6,X3)
testX6X3 = chisq.test(tableX6X3, correct=FALSE)
tableX7X3 = table(X7,X3)
testX7X3 = chisq.test(tableX7X3, correct=FALSE)
tableY3 = table(Y,X3)

```

```

testY3      = chisq.test(tableY3, correct=FALSE)
variabelX3=rbind("X1*X3", "X2*X3", "X4*X3", "X5*X3", "X6*X3", "X7*X3",
"Y*X3")
xsquareX3=rbind(testX1X3$statistic,          testX2X3$statistic,
testX4X3$statistic,      testX5X3$statistic,      testX6X3$statistic,
testX7X3$statistic, testY3$statistic)
pvalueX3=rbind(testX1X3$p.value,          testX2X3$p.value,
testX4X3$p.value,      testX5X3$p.value,      testX6X3$p.value,
testX7X3$p.value, testY3$p.value)
confoundX3=data.frame(variabelX3, xsquareX3, pvalueX3)

#Chisquare X4
tableX1X4  = table(X1,X4)
testX1X4   = chisq.test(tableX1X4, correct=FALSE)
tableX2X4  = table(X2,X4)
testX2X4   = chisq.test(tableX2X4, correct=FALSE)
tableX3X4  = table(X3,X4)
testX3X4   = chisq.test(tableX3X4, correct=FALSE)
tableX5X4  = table(X5,X4)
testX5X4   = chisq.test(tableX5X4, correct=FALSE)
tableX6X4  = table(X6,X4)
testX6X4   = chisq.test(tableX6X4, correct=FALSE)
tableX7X4  = table(X7,X4)
testX7X4   = chisq.test(tableX7X4, correct=FALSE)
tableY4    = table(Y,X4)
testY4     = chisq.test(tableY4, correct=FALSE)
variabelX4=rbind("X1*X4", "X2*X4", "X3*X4", "X5*X4", "X6*X4", "X7*X4",
"Y*X4")
xsquareX4=rbind(testX1X4$statistic,          testX2X4$statistic,
testX3X4$statistic,      testX5X4$statistic,      testX6X4$statistic,
testX7X4$statistic, testY4$statistic)
pvalueX4=rbind(testX1X4$p.value,          testX2X4$p.value,
testX3X4$p.value,      testX5X4$p.value,      testX6X4$p.value,
testX7X4$p.value, testY4$p.value)
confoundX4=data.frame(variabelX4, xsquareX4, pvalueX4)

#Chisquare X5
tableX1X5  = table(X1,X5)
testX1X5   = chisq.test(tableX1X5, correct=FALSE)
tableX2X5  = table(X2,X5)
testX2X5   = chisq.test(tableX2X5, correct=FALSE)
tableX3X5  = table(X3,X5)
testX3X5   = chisq.test(tableX3X5, correct=FALSE)
tableX4X5  = table(X4,X5)
testX4X5   = chisq.test(tableX4X5, correct=FALSE)
tableX6X5  = table(X6,X5)
testX6X5   = chisq.test(tableX6X5, correct=FALSE)
tableX7X5  = table(X7,X5)
testX7X5   = chisq.test(tableX7X5, correct=FALSE)
tableY5    = table(Y,X5)
testY5     = chisq.test(tableY5, correct=FALSE)
variabelX5=rbind("X1*X5", "X2*X5", "X3*X5", "X4*X5", "X6*X5", "X7*X5",
"Y*X5")
xsquareX5=rbind(testX1X5$statistic,          testX2X5$statistic,
testX3X5$statistic,      testX4X5$statistic,      testX6X5$statistic,
testX7X5$statistic, testY5$statistic)
pvalueX5=rbind(testX1X5$p.value,          testX2X5$p.value,

```

```

testX3X5$p.value,          testX4X5$p.value,          testX6X5$p.value,
testX7X5$p.value, testY5$p.value)
confoundX5=data.frame(variabelX5, xsquareX5, pvalueX5)

#Chisquare X6
tableX1X6 = table(X1,X6)
testX1X6 = chisq.test(tableX1X6, correct=FALSE)
tableX2X6 = table(X2,X6)
testX2X6 = chisq.test(tableX2X6, correct=FALSE)
tableX3X6 = table(X3,X6)
testX3X6 = chisq.test(tableX3X6, correct=FALSE)
tableX4X6 = table(X4,X6)
testX4X6 = chisq.test(tableX4X6, correct=FALSE)
tableX5X6 = table(X5,X6)
testX5X6 = chisq.test(tableX5X6, correct=FALSE)
tableX7X6 = table(X7,X6)
testX7X6 = chisq.test(tableX7X6, correct=FALSE)
tableY6 = table(Y,X6)
testY6 = chisq.test(tableY6, correct=FALSE)
variabelX6=rbind("X1*X6","X2*X6","X3*X6","X4*X6","X5*X6","X7*X6",
"Y*X6")
xsquareX6=rbind(testX1X6$statistic,          testX2X6$statistic,
testX3X6$statistic,          testX4X6$statistic,          testX5X6$statistic,
testX7X6$statistic, testY6$statistic)
pvalueX6=rbind(testX1X6$p.value,          testX2X6$p.value,
testX3X6$p.value,          testX4X6$p.value,          testX5X6$p.value,
testX7X6$p.value, testY6$p.value)
confoundX6=data.frame(variabelX6, xsquareX6, pvalueX6)

#Chisquare X7
tableX1X7 = table(X1,X7)
testX1X7 = chisq.test(tableX1X7, correct=FALSE)
tableX2X7 = table(X2,X7)
testX2X7 = chisq.test(tableX2X7, correct=FALSE)
tableX3X7 = table(X3,X7)
testX3X7 = chisq.test(tableX3X7, correct=FALSE)
tableX4X7 = table(X4,X7)
testX4X7 = chisq.test(tableX4X7, correct=FALSE)
tableX5X7 = table(X5,X7)
testX5X7 = chisq.test(tableX5X7, correct=FALSE)
tableX6X7 = table(X6,X7)
testX6X7 = chisq.test(tableX6X7, correct=FALSE)
tableY7 = table(Y,X7)
testY7 = chisq.test(tableY7, correct=FALSE)
variabelX7=rbind("X1*X7","X2*X7","X3*X7","X4*X7","X5*X7","X6*X7",
"Y*X7")
xsquareX7=rbind(testX1X7$statistic,          testX2X7$statistic,
testX3X7$statistic,          testX4X7$statistic,          testX5X7$statistic,
testX6X7$statistic, testY7$statistic)
pvalueX7=rbind(testX1X7$p.value,          testX2X7$p.value,
testX3X7$p.value,          testX4X7$p.value,          testX5X7$p.value,
testX6X7$p.value, testY7$p.value)
confoundX7=data.frame(variabelX7, xsquareX7, pvalueX7)

confound=cbind(confoundX1,confoundX2,confoundX3,confoundX4,confou
ndX5,confoundX6,confoundX7)
confound

```

Lampiran 7. Output pengujian hubungan antara variabel yang berpotensi sebagai *confounding* dengan kovariat X dalam menentukan variabel *confounding*

	variabelX1	X.squared	pvalueX1
1	X2*X1	2.1750000	0.140269516
2	X3*X1	3.6261563	0.056878035
3	X4*X1	0.1809423	0.670564712
4	X5*X1	1.5917740	0.207072861
5	X6*X1	0.5353166	0.464380207
6	X7*X1	0.3595867	0.548735871
7	Y*X1	9.8308620	0.001716079
	variabelX2	X.squared	pvalueX2
1	X1*X2	2.17500000	0.14026952
2	X3*X2	4.05186736	0.04412251
3	X4*X2	0.06154499	0.80407047
4	X5*X2	1.52810304	0.21639737
5	X6*X2	0.22581363	0.63464549
6	X7*X2	1.18382447	0.27657878
7	Y*X2	2.20378412	0.13767241
	variabelX3	X.squared	pvalueX3
1	X1*X3	3.6261563	0.05687803
2	X2*X3	4.0518674	0.04412251
3	X4*X3	0.8219823	0.36460123
4	X5*X3	0.1235733	0.72519077
5	X6*X3	0.4078360	0.52306990
6	X7*X3	0.8241679	0.36396439
7	Y*X3	0.7490843	0.38676630
	variabelX4	X.squared	pvalueX4
1	X1*X4	0.18094228	6.705647e-01
2	X2*X4	0.06154499	8.040705e-01
3	X3*X4	0.82198228	3.646012e-01
4	X5*X4	0.15471126	6.940732e-01
5	X6*X4	20.81357566	5.062286e-06
6	X7*X4	22.50328746	2.097843e-06
7	Y*X4	9.03596540	2.647192e-03
	variabelX5	X.squared	pvalueX5
1	X1*X5	1.5917740	0.2070729
2	X2*X5	1.5281030	0.2163974
3	X3*X5	0.1235733	0.7251908
4	X4*X5	0.1547113	0.6940732
5	X6*X5	1.8404410	0.1748993
6	X7*X5	0.1601940	0.6889779
7	Y*X5	0.5551036	0.4562398
	variabelX6	X.squared	pvalueX6
1	X1*X6	0.5353166	4.643802e-01
2	X2*X6	0.2258136	6.346455e-01
3	X3*X6	0.4078360	5.230699e-01
4	X4*X6	20.8135757	5.062286e-06
5	X5*X6	1.8404410	1.748993e-01
6	X7*X6	81.8018136	1.504404e-19
7	Y*X6	5.4688591	1.935826e-02
	variabelX7	X.squared	pvalueX7
1	X1*X7	0.3595867	5.487359e-01
2	X2*X7	1.1838245	2.765788e-01
3	X3*X7	0.8241679	3.639644e-01

4	X4*X7	22.5032875	2.097843e-06
5	X5*X7	0.1601940	6.889779e-01
6	X6*X7	81.8018136	1.504404e-19
7	Y*X7	9.8308620	1.716079e-03

Lampiran 8. Sintaks estimasi *Propensity Score* dengan *Multivariate Adaptive Regression Splines* untuk mendapatkan model terbaik menggunakan *software R*

```

analysis.mars = function(data){
data = data

psmars12.1.0      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 1, minspan = 0, endspan=0, penalty = -1, data = mydata)
psmars12.1.1      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 1, minspan = 1, endspan=1, penalty = -1, data = mydata)
psmars12.1.2      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 1, minspan = 2, endspan=2, penalty = -1, data = mydata)
psmars12.1.3      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 1, minspan = 3, endspan=3, penalty = -1, data = mydata)
psmars12.1.4      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 1, minspan = 4, endspan=4, penalty = -1, data = mydata)
psmars12.1.5      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 1, minspan = 5, endspan=5, penalty = -1, data = mydata)

psmars12.2.0      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 2, minspan = 0, endspan=0, penalty = -1, data = mydata)
psmars12.2.1      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 2, minspan = 1, endspan=1, penalty = -1, data = mydata)
psmars12.2.2      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 2, minspan = 2, endspan=2, penalty = -1, data = mydata)
psmars12.2.3      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 2, minspan = 3, endspan=3, penalty = -1, data = mydata)
psmars12.2.4      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 2, minspan = 4, endspan=4, penalty = -1, data = mydata)
psmars12.2.5      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 2, minspan = 5, endspan=5, penalty = -1, data = mydata)

psmars12.3.0      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 3, minspan = 0, endspan=0, penalty = -1, data = mydata)
psmars12.3.1      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 3, minspan = 1, endspan=1, penalty = -1, data = mydata)
psmars12.3.2      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 3, minspan = 2, endspan=2, penalty = -1, data = mydata)
psmars12.3.3      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 3, minspan = 3, endspan=3, penalty = -1, data = mydata)
psmars12.3.4      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 3, minspan = 4, endspan=4, penalty = -1, data = mydata)
psmars12.3.5      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 12, degree
= 3, minspan = 5, endspan=5, penalty = -1, data = mydata)

psmars18.1.0      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 18, degree
= 1, minspan = 0, endspan=0, penalty = -1, data = mydata)
psmars18.1.1      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 18, degree
= 1, minspan = 1, endspan=1, penalty = -1, data = mydata)
psmars18.1.2      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 18, degree

```



```

psmars24.2.5      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 24, degree
= 2, minspan = 5, endspan=5, penalty = -1, data = mydata)

psmars24.3.0      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 24, degree
= 3, minspan = 0, endspan=0, penalty = -1, data = mydata)
psmars24.3.1      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 24, degree
= 3, minspan = 1, endspan=1, penalty = -1, data = mydata)
psmars24.3.2      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 24, degree
= 3, minspan = 2, endspan=2, penalty = -1, data = mydata)
psmars24.3.3      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 24, degree
= 3, minspan = 3, endspan=3, penalty = -1, data = mydata)
psmars24.3.4      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 24, degree
= 3, minspan = 4, endspan=4, penalty = -1, data = mydata)
psmars24.3.5      = earth(X7 ~ X1+X2+X3+X4+X5+X6, nk = 24, degree
= 3, minspan = 5, endspan=5, penalty = -1, data = mydata)

nilai.nk = c(psmars12.1.0$nk, psmars12.1.1$nk, psmars12.1.2$nk,
psmars12.1.3$nk, psmars12.1.4$nk, psmars12.1.5$nk,
psmars12.2.0$nk,          psmars12.2.1$nk,          psmars12.2.2$nk,
psmars12.2.3$nk, psmars12.2.4$nk, psmars12.2.5$nk,
psmars12.3.0$nk,          psmars12.3.1$nk,          psmars12.3.2$nk,
psmars12.3.3$nk, psmars12.3.4$nk, psmars12.3.5$nk,
psmars18.1.0$nk,          psmars18.1.1$nk,          psmars18.1.2$nk,
psmars18.1.3$nk, psmars18.1.4$nk, psmars18.1.5$nk,
psmars18.2.0$nk,          psmars18.2.1$nk,          psmars18.2.2$nk,
psmars18.2.3$nk, psmars18.2.4$nk, psmars18.2.5$nk,
psmars18.3.0$nk,          psmars18.3.1$nk,          psmars18.3.2$nk,
psmars18.3.3$nk, psmars18.3.4$nk, psmars18.3.5$nk,
psmars24.1.0$nk,          psmars24.1.1$nk,          psmars24.1.2$nk,
psmars24.1.3$nk, psmars24.1.4$nk, psmars24.1.5$nk,
psmars24.2.0$nk,          psmars24.2.1$nk,          psmars24.2.2$nk,
psmars24.2.3$nk, psmars24.2.4$nk, psmars24.2.5$nk,
psmars24.3.0$nk,          psmars24.3.1$nk,          psmars24.3.2$nk,
psmars24.3.3$nk, psmars24.3.4$nk, psmars24.3.5$nk)

nilai.degree = c(1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,3,3,
1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,3,3,
1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,3,3)

nilai.minspan = c(0,1,2,3,4,5,0,1,2,3,4,5,0,1,2,3,4,5,0,1,2,3,4,
5,0,1,2,3,4,5,0,1,2,3,4,5,0,1,2,3,4,5,0,1,2,3,4,5,0,1,2,3,4,5)

nilai.GCV      =      c(psmars12.1.0$gcv,          psmars12.1.1$gcv,
psmars12.1.2$gcv,          psmars12.1.3$gcv,          psmars12.1.4$gcv,
psmars12.1.5$gcv,
psmars12.2.0$gcv,          psmars12.2.1$gcv,          psmars12.2.2$gcv,
psmars12.2.3$gcv, psmars12.2.4$gcv, psmars12.2.5$gcv,
psmars12.3.0$gcv,          psmars12.3.1$gcv,          psmars12.3.2$gcv,
psmars12.3.3$gcv, psmars12.3.4$gcv, psmars12.3.5$gcv,
psmars18.1.0$gcv,          psmars18.1.1$gcv,          psmars18.1.2$gcv,
psmars18.1.3$gcv, psmars18.1.4$gcv, psmars18.1.5$gcv,
psmars18.2.0$gcv,          psmars18.2.1$gcv,          psmars18.2.2$gcv,
psmars18.2.3$gcv, psmars18.2.4$gcv, psmars18.2.5$gcv,
psmars18.3.0$gcv,          psmars18.3.1$gcv,          psmars18.3.2$gcv,
psmars18.3.3$gcv, psmars18.3.4$gcv, psmars18.3.5$gcv,
psmars24.1.0$gcv,          psmars24.1.1$gcv,          psmars24.1.2$gcv,
psmars24.1.3$gcv, psmars24.1.4$gcv, psmars24.1.5$gcv,

```

```

psmars24.2.0$gcv,      psmars24.2.1$gcv,      psmars24.2.2$gcv,
psmars24.2.3$gcv, psmars24.2.4$gcv, psmars24.2.5$gcv,
psmars24.3.0$gcv,      psmars24.3.1$gcv,      psmars24.3.2$gcv,
psmars24.3.3$gcv, psmars24.3.4$gcv, psmars24.3.5$gcv)

nilai.RS      =      c(psmars12.1.0$rsq,      psmars12.1.1$rsq,
psmars12.1.2$rsq,      psmars12.1.3$rsq,      psmars12.1.4$rsq,
psmars12.1.5$rsq,
psmars12.2.0$rsq,      psmars12.2.1$rsq,      psmars12.2.2$rsq,
psmars12.2.3$rsq, psmars12.2.4$rsq, psmars12.2.5$rsq,
psmars12.3.0$rsq,      psmars12.3.1$rsq,      psmars12.3.2$rsq,
psmars12.3.3$rsq, psmars12.3.4$rsq, psmars12.3.5$rsq,
psmars18.1.0$rsq,      psmars18.1.1$rsq,      psmars18.1.2$rsq,
psmars18.1.3$rsq, psmars18.1.4$rsq, psmars18.1.5$rsq,
psmars18.2.0$rsq,      psmars18.2.1$rsq,      psmars18.2.2$rsq,
psmars18.2.3$rsq, psmars18.2.4$rsq, psmars18.2.5$rsq,
psmars18.3.0$rsq,      psmars18.3.1$rsq,      psmars18.3.2$rsq,
psmars18.3.3$rsq, psmars18.3.4$rsq, psmars18.3.5$rsq,
psmars24.1.0$rsq,      psmars24.1.1$rsq,      psmars24.1.2$rsq,
psmars24.1.3$rsq, psmars24.1.4$rsq, psmars24.1.5$rsq,
psmars24.2.0$rsq,      psmars24.2.1$rsq,      psmars24.2.2$rsq,
psmars24.2.3$rsq, psmars24.2.4$rsq, psmars24.2.5$rsq,
psmars24.3.0$rsq,      psmars24.3.1$rsq,      psmars24.3.2$rsq,
psmars24.3.3$rsq, psmars24.3.4$rsq, psmars24.3.5$rsq)

hasil = data.frame(nilai.nk,      nilai.degree,      nilai.minspan,
nilai.GCV, nilai.RS)
print(hasil)
}

analisis.mars (data)

```

Lampiran 9. Output model dari estimasi *Propensity Score* dengan *Multivariate Adaptive Regression Splines* menggunakan *software R*

	nilai.nk	nilai.degree	nilai.minspan	nilai.GCV	nilai.RS
1	12	1	0	0.09282482	0.6265515
2	12	1	1	0.09337786	0.6243265
3	12	1	2	0.09337786	0.6243265
4	12	1	3	0.09337786	0.6243265
5	12	1	4	0.09501367	0.6177454
6	12	1	5	0.09592003	0.6140990
7	12	2	0	0.08292236	0.6663906
8	12	2	1	0.08199773	0.6701106
9	12	2	2	0.08175769	0.6710763
10	12	2	3	0.08091160	0.6744802
11	12	2	4	0.08615641	0.6533795
12	12	2	5	0.08304693	0.6658894
13	12	3	0	0.07737413	0.6887120
14	12	3	1	0.07786745	0.6867273
15	12	3	2	0.07809173	0.6858250
16	12	3	3	0.07858309	0.6838482
17	12	3	4	0.08009968	0.6777467
18	12	3	5	0.07737413	0.6887120
19	18	1	0	0.08639300	0.6524277
20	18	1	1	0.09180924	0.6306373
21	18	1	2	0.09188931	0.6303152

22	18	1	3	0.09095150	0.6340882
23	18	1	4	0.08817969	0.6452396
24	18	1	5	0.09028547	0.6367677
25	18	2	0	0.07504121	0.6980977
26	18	2	1	0.07661008	0.6917859
27	18	2	2	0.07627159	0.6931477
28	18	2	3	0.07546627	0.6963876
29	18	2	4	0.07932143	0.6808777
30	18	2	5	0.07862204	0.6836915
31	18	3	0	0.07127032	0.7132686
32	18	3	1	0.06956467	0.7201306
33	18	3	2	0.06968772	0.7196356
34	18	3	3	0.07279021	0.7071538
35	18	3	4	0.07444433	0.7004990
36	18	3	5	0.07223169	0.7094008
37	24	1	0	0.08422785	0.6611384
38	24	1	1	0.08213608	0.6695539
39	24	1	2	0.08512939	0.6575114
40	24	1	3	0.08304106	0.6659131
41	24	1	4	0.08502447	0.6579335
42	24	1	5	0.08768591	0.6472261
43	24	2	0	0.07028718	0.7172239
44	24	2	1	0.07145458	0.7125272
45	24	2	2	0.07241895	0.7086475
46	24	2	3	0.07087450	0.7148610
47	24	2	4	0.07525652	0.6972315
48	24	2	5	0.07434761	0.7008882
49	24	3	0	0.06357701	0.7442199
50	24	3	1	0.06361298	0.7440752
51	24	3	2	0.06428353	0.7413775
52	24	3	3	0.06879288	0.7232357
53	24	3	4	0.06697877	0.7305341
54	24	3	5	0.06705446	0.7302296

Lampiran 10. Output model terbaik dari estimasi *Propensity Score* dengan *Multivariate Adaptive Regression Splines* menggunakan *software R*

	coefficients
(Intercept)	0.14121367
X61	0.52077163
h(X1-63)	0.14295438
h(65-X1)	-0.00396286
h(X1-65)	-0.23807566
h(X1-70)	0.43952664
h(X1-72)	-0.44804180
h(6-X4)	-0.02333911
h(X4-6)	-0.02001233
h(62-X1) * X61	-0.10274837
h(X1-62) * X61	-0.03435814
h(65-X1) * X61	0.10827679
h(4-X4) * X61	0.03739791
h(X4-4) * X61	0.01848360
h(X1-62) * X51 * X61	0.05899064
X21 * h(4-X4) * X61	-0.21940575

Selected 16 of 16 terms, and 5 of 6 predictors

```

Termination condition: Reached nk 24
Importance: X61, X21, X4, X1, X51, X31-unused
Number of terms at each degree of interaction: 1 8 5 2
GCV 0.06357701   RSS 9.218667   GRSq 0.7442199   RSq 0.7442199
CVRSq 0.4691906

Note: the cross-validation sd's below are standard deviations
across folds

Cross validation:   nterms 16.60 sd 0.52   nvars 5.10 sd 0.32

      CVRSq   sd      ClassRate   sd      MaxErr   sd
      0.469 0.345          0.855 0.101          1.33 0.9

0.1412137
+ 0.5207716 * X61
+ 0.1429544 * pmax(0, X1 - 63)
- 0.003962856 * pmax(0, 65 - X1)
- 0.2380757 * pmax(0, X1 - 65)
+ 0.4395266 * pmax(0, X1 - 70)
- 0.4480418 * pmax(0, X1 - 72)
- 0.02333911 * pmax(0, 6 - X4)
- 0.02001233 * pmax(0, X4 - 6)
- 0.1027484 * pmax(0, 62 - X1) * X61
- 0.03435814 * pmax(0, X1 - 62) * X61
+ 0.1082768 * pmax(0, 65 - X1) * X61
+ 0.03739791 * pmax(0, 4 - X4) * X61
+ 0.0184836 * pmax(0, X4 - 4) * X61
+ 0.05899064 * pmax(0, X1 - 62) * X51 * X61
- 0.2194057 * X21 * pmax(0, 4 - X4) * X61

```

Lampiran 11. Sintaks klasifikasi untuk model terbaik dari estimasi *Propensity Score* dengan *Multivariate Adaptive Regression Splines* menggunakan *software R*

```

X7_predict = predict(psmars, mydata)
colnames(X7_predict)= c("X7 prediksi")
X7_predict_c= predict(psmars, mydata, type="class")
colnames(X7_predict_c)= c("X7 prediksi kelas")
predicted = cbind(X7_predict,X7_predict_c)
tabel_akurasi=table(X7_predict_c, mydata$X7)
akurasi=sum(diag(tabel_akurasi))/sum(tabel_akurasi)
error=1-akurasi

```

Lampiran 12. Output klasifikasi untuk model terbaik dari estimasi *Propensity Score* dengan *Multivariate Adaptive Regression Splines* menggunakan *software R*.

```

      X7 prediksi X7 prediksi kelas
[1,] 0.995188712          1
[2,] 0.973556409          1
[3,] 0.975113357          1
[4,] 0.086609743          0
[5,] 0.048402166          0
[6,] 0.109312774          0
[7,] 0.163509657          0

```

[8,]	1.002758717	1
[9,]	-0.055615039	0
[10,]	-0.031399876	0
[11,]	0.967356637	1
[12,]	0.003827798	0
[13,]	1.052643834	1
[14,]	-0.051509218	0
[15,]	0.059307780	0
[16,]	0.210187872	0
[17,]	0.215305616	0
[18,]	-0.023474165	0
[19,]	0.063270636	0
[20,]	0.380444216	0
[21,]	0.546970324	1
[22,]	0.988858297	1
[23,]	0.943873240	1
[24,]	0.928248876	1
[25,]	0.118419488	0
[26,]	0.012191537	0
[27,]	0.074876821	0
[28,]	0.032005817	0
[29,]	1.022853200	1
[30,]	0.078246004	0
[31,]	1.046373426	1
[32,]	1.008836817	1
[33,]	1.010476033	1
[34,]	0.212721892	0
[35,]	0.086609743	0
[36,]	-0.111477862	0
[37,]	0.094097427	0
[38,]	0.043456357	0
[39,]	0.238644723	0
[40,]	-0.071823068	0
[41,]	0.004265826	0
[42,]	0.973202698	1
[43,]	0.163509657	0
[44,]	1.015172712	1
[45,]	1.035488162	1
[46,]	0.959497604	1
[47,]	0.437873563	0
[48,]	0.089696544	0
[49,]	0.929814436	1
[50,]	0.058431725	0
[51,]	0.948569920	1
[52,]	0.588512174	1
[53,]	0.047419212	0
[54,]	0.105985995	0
[55,]	-0.070600125	0
[56,]	0.403783324	0
[57,]	0.018803167	0
[58,]	1.005889837	1
[59,]	0.466672990	0
[60,]	0.094097427	0
[61,]	1.033885775	1
[62,]	1.021324468	1
[63,]	0.831399509	1
[64,]	-0.015548453	0

[65,]	0.881786834	1
[66,]	1.058758763	1
[67,]	0.427122431	0
[68,]	1.001450951	1
[69,]	0.531645028	1
[70,]	0.031567789	0
[71,]	0.202564967	0
[72,]	0.105985995	0
[73,]	0.957994513	1
[74,]	-0.048483961	0
[75,]	0.061956553	0
[76,]	1.037090550	1
[77,]	1.024492416	1
[78,]	1.023000511	1
[79,]	0.063270636	0
[80,]	1.007529052	1
[81,]	0.386375656	0
[82,]	0.038617446	0
[83,]	0.063270636	0
[84,]	0.059307780	0
[85,]	0.118419488	0
[86,]	1.022853200	1
[87,]	0.121399390	0
[88,]	0.868636016	1
[89,]	0.875671906	1
[90,]	0.987292737	1
[91,]	0.664492193	1
[92,]	0.048402166	0
[93,]	0.722901933	1
[94,]	0.881786834	1
[95,]	0.039055473	0
[96,]	0.226999155	0
[97,]	0.876685058	1
[98,]	1.015172712	1
[99,]	0.210187872	0
[100,]	0.204356237	0
[101,]	0.066795464	0
[102,]	0.308341372	0
[103,]	0.182100599	0
[104,]	-0.174228522	0
[105,]	0.881786834	1
[106,]	-0.023912192	0
[107,]	0.928248876	1
[108,]	0.090134571	0
[109,]	0.032005817	0
[110,]	0.520686109	1
[111,]	0.999701253	1
[112,]	0.664492193	1
[113,]	0.403783324	0
[114,]	1.023000511	1
[115,]	0.046543157	0
[116,]	0.313773606	0
[117,]	-0.031837904	0
[118,]	-0.001504223	0
[119,]	0.403783324	0
[120,]	0.163509657	0
[121,]	0.957932044	1

```

[122,] 0.052127224      0
[123,] 1.010549688      1
[124,] 0.039931528      0
[125,] 0.979784338      1
[126,] 0.012191537      0
[127,] 1.055701298      1
[128,] 0.956366485      1
[129,] 0.959528838      1
[130,] 0.490012097      0
[131,] 0.063270636      0
[132,] 1.019943047      1
[133,] -0.141840391     0
[134,] -0.011585597     0
[135,] 0.313773606      0
[136,] 0.928544300      1
[137,] 0.926097926      1
[138,] 0.735334677      1
[139,] 1.033812119      1
[140,] 0.031567789      0
[141,] 0.016592421      0
[142,] 1.018451143      1
[143,] 0.978284322      1
[144,] 0.118419488      0
[145,] 0.261983830      0

      X7
X7_predict_c 0 1
              0 75 8
              1 3 59

akurasi
[1] 0.9241379

error
[1] 0.07586207

```

Lampiran 13. Sintaks pengujian kesamaan variansi dengan uji Bartlett menggunakan *software R*

```

x=list(data$X1, data$X2, data$X3, data$X4, data$X5, data$X6)
x7=data$X7
x7=as.factor(X7)
bartlett.test(x,x7)

```

Lampiran 14. Output pengujian kesamaan variansi dengan uji Bartlett menggunakan *software R*

```

      Bartlett test of homogeneity of variances

data:  x
Bartlett's K-squared = 2193.6, df = 5, p-value < 2.2e-16

```

Lampiran 15. Sintaks analisis *balance* dan *postmatching* untuk mendapatkan estimasi ATT pada *Propensity Score Matching* menggunakan *software R*

```
att.psmars = Match(Y=Y, Tr=X7, X=psmars$fitted, M=1,
estimand="ATT", replace = FALSE, BiasAdjust = FALSE)
summary(att.psmars)
bal.psmars = MatchBalance(X7_predict_c ~ X1+X2+X3+X4+X5+X6,
match.out = att.psmars, nboots = 5000)
```

Lampiran 14. Output *balance* dan *postmatching* untuk estimasi ATT pada *Propensity Score Matching* menggunakan *software R*

```
***** (V1) X1 *****
                                Before Matching      After Matching
mean treatment.....           58.435                59.015
mean control.....            60.904                61.134
std mean diff.....           -30.346                -26.109

mean raw eQQ diff.....         2.8548                2.1194
med  raw eQQ diff.....          3                      2
max  raw eQQ diff.....          6                      4

mean eCDF diff.....            0.067106              0.054457
med  eCDF diff.....            0.054023              0.044776
max  eCDF diff.....            0.22367                0.1791

var ratio (Tr/Co).....          0.7389                1.048
T-test p-value.....            0.094219              0.1416
KS Bootstrap p-value..         0.0278                0.1544
KS Naive p-value.....           0.057394              0.23277
KS Statistic.....              0.22367                0.1791

***** (V2) X21 *****
                                Before Matching      After Matching
mean treatment.....           0.53226              0.55224
mean control.....             0.6506                0.62687
std mean diff.....           -23.526              -14.895

mean raw eQQ diff.....         0.1129                0.074627
med  raw eQQ diff.....          0                      0
max  raw eQQ diff.....          1                      1

mean eCDF diff.....            0.059172              0.037313
med  eCDF diff.....            0.059172              0.037313
max  eCDF diff.....            0.11834              0.074627

var ratio (Tr/Co).....          1.0997                1.0571
T-test p-value.....            0.15529              0.41168

***** (V3) X31 *****
                                Before Matching      After Matching
mean treatment.....           0.45161              0.43284
mean control.....             0.3494                0.34328
std mean diff.....            20.373                17.939

mean raw eQQ diff.....         0.1129                0.089552
```


med raw eQQ diff.....	0	0
max raw eQQ diff.....	1	1
mean eCDF diff.....	0.051108	0.044776
med eCDF diff.....	0.051108	0.044776
max eCDF diff.....	0.10222	0.089552
var ratio (Tr/Co).....	1.094	1.0889
T-test p-value.....	0.21849	0.23868
***** (V4) X4 *****		
	Before Matching	After Matching
mean treatment.....	6.8387	6.597
mean control.....	3.8675	3.5373
std mean diff.....	67.204	70.135
mean raw eQQ diff.....	3.0323	3.0597
med raw eQQ diff.....	3	2
max raw eQQ diff.....	9	13
mean eCDF diff.....	0.16528	0.17817
med eCDF diff.....	0.10902	0.14925
max eCDF diff.....	0.39973	0.41791
var ratio (Tr/Co).....	2.3246	3.8931
T-test p-value.....	1.2367e-05	9.6691e-07
KS Bootstrap p-value..	< 2.22e-16	< 2.22e-16
KS Naive p-value.....	2.3746e-05	1.6563e-05
KS Statistic.....	0.39973	0.41791
***** (V5) X51 *****		
	Before Matching	After Matching
mean treatment.....	0.40323	0.40299
mean control.....	0.43373	0.41791
std mean diff.....	-6.1691	-3.0201
mean raw eQQ diff.....	0.032258	0.014925
med raw eQQ diff.....	0	0
max raw eQQ diff.....	1	1
mean eCDF diff.....	0.015255	0.0074627
med eCDF diff.....	0.015255	0.0074627
max eCDF diff.....	0.030509	0.014925
var ratio (Tr/Co).....	0.98381	0.98901
T-test p-value.....	0.71478	0.86626
***** (V6) X61 *****		
	Before Matching	After Matching
mean treatment.....	0.98387	0.8806
mean control.....	0.096386	0.13433
std mean diff.....	698.81	228.42
mean raw eQQ diff.....	0.8871	0.74627
med raw eQQ diff.....	1	1
max raw eQQ diff.....	1	1

```

mean eCDF diff.....      0.44374                0.37313
med  eCDF diff.....      0.44374                0.37313
max  eCDF diff.....      0.88749                0.74627

var ratio (Tr/Co).....    0.18296                0.90421
T-test p-value.....      < 2.22e-16            < 2.22e-16

Before Matching Minimum p.value: < 2.22e-16
Variable Name(s): X4 X61  Number(s): 4 6

After Matching Minimum p.value: < 2.22e-16
Variable Name(s): X4 X61  Number(s): 4 6

Estimate...  0.25373
SE.....     0.07994
T-stat.....  3.174
p.val.....  0.0015034

Original number of observations..... 145
Original number of treated obs..... 67
Matched number of observations..... 67
Matched number of observations (unweighted). 67

```

Lampiran 15. Output model terbaik dari estimasi *Propensity Score* dengan *Multivariate Adaptive Regression Splines* ketika variabel Lama menderita Diabetes Melitus (X_4) dan kadar HbA1C (X_6) tidak masuk dalam model dengan menggunakan *software R*

```

coefficients
(Intercept)      1.13230623
X21              -0.11821759
X51              -2.32962678
h(51-X1)         -0.05857378
h(X1-51)         -0.10766931
h(X1-55)         0.07945728
h(X1-51) * X51   0.27110877
h(62-X1) * X51   0.10394346
h(X1-62) * X51  -0.81262956
h(X1-63) * X51   0.99048708
h(X1-64) * X51  -0.47406069
h(X1-51) * X31 * X51 -0.02288409
h(X1-62) * X31 * X51 -0.30564656
h(X1-64) * X31 * X51  0.43933440

Selected 14 of 14 terms, and 4 of 4 predictors
Termination condition: Reached nk 24
Importance: X1, X51, X31, X21
Number of terms at each degree of interaction: 1 5 5 3
GCV 0.1620439  RSS 23.49636  GRSq 0.3480727  RSq 0.3480727  CVRSq
-0.1364061

Note: the cross-validation sd's below are standard deviations
across folds

Cross validation:  nterms 13.80 sd 0.42  nvars 3.70 sd 0.48

```

CVRSq	sd	ClassRate	sd	MaxErr	sd
-0.136	0.39	0.628	0.15	-2.04	1.2
1.132306					
-	0.1182176	*	X21		
-	2.329627	*	X51		
-	0.05857378	*	pmax(0, 51 - X1)		
-	0.1076693	*	pmax(0, X1 - 51)		
+	0.07945728	*	pmax(0, X1 - 55)		
+	0.2711088	*	pmax(0, X1 - 51) * X51		
+	0.1039435	*	pmax(0, 62 - X1) * X51		
-	0.8126296	*	pmax(0, X1 - 62) * X51		
+	0.9904871	*	pmax(0, X1 - 63) * X51		
-	0.4740607	*	pmax(0, X1 - 64) * X51		
-	0.02288409	*	pmax(0, X1 - 51) * X31 * X51		
-	0.3056466	*	pmax(0, X1 - 62) * X31 * X51		
+	0.4393344	*	pmax(0, X1 - 64) * X31 * X51		

Lampiran 16. Output *balance* dan *postmatching* untuk estimasi ATT pada *Propensity Score Matching* ketika variabel Lama menderita Diabetes Melitus (X_4) dan kadar HbA1C (X_6) tidak masuk dalam model menggunakan *software R*

***** (V1) X1 *****		
	Before Matching	After Matching
mean treatment.....	58.237	59.015
mean control.....	60.953	61.075
std mean diff.....	-35.483	-25.374
mean raw eQQ diff.....	3.8305	2.4478
med raw eQQ diff.....	4	2
max raw eQQ diff.....	10	5
mean eCDF diff.....	0.090279	0.060487
med eCDF diff.....	0.055183	0.037313
max eCDF diff.....	0.36106	0.1791
var ratio (Tr/Co).....	0.62764	0.84272
T-test p-value.....	0.061654	0.16511
KS Bootstrap p-value..	< 2.22e-16	0.1594
KS Naive p-value.....	0.00021814	0.23277
KS Statistic.....	0.36106	0.1791
***** (V2) X21 *****		
	Before Matching	After Matching
mean treatment.....	0.44068	0.55224
mean control.....	0.7093	0.61194
std mean diff.....	-53.647	-11.916
mean raw eQQ diff.....	0.25424	0.059701
med raw eQQ diff.....	0	0
max raw eQQ diff.....	1	1
mean eCDF diff.....	0.13431	0.029851
med eCDF diff.....	0.13431	0.029851
max eCDF diff.....	0.26862	0.059701

```

var ratio (Tr/Co).....      1.2019                1.0413
T-test p-value.....      0.0013335            0.46602

***** (V3) X31 *****
                                Before Matching          After Matching
mean treatment.....      0.37288              0.43284
mean control.....      0.40698              0.34328
std mean diff.....      -6.9907              17.939

mean raw eQQ diff.....    0.033898            0.089552
med  raw eQQ diff.....    0                0
max  raw eQQ diff.....    1                1

mean eCDF diff.....      0.017048            0.044776
med  eCDF diff.....      0.017048            0.044776
max  eCDF diff.....      0.034095            0.089552

var ratio (Tr/Co).....    0.97414              1.0889
T-test p-value.....      0.68154              0.21985

***** (V4) X51 *****
                                Before Matching          After Matching
mean treatment.....      0.40678              0.40299
mean control.....      0.43023              0.38806
std mean diff.....      -4.7337              3.0201

mean raw eQQ diff.....    0.016949            0.014925
med  raw eQQ diff.....    0                0
max  raw eQQ diff.....    1                1

mean eCDF diff.....      0.011726            0.0074627
med  eCDF diff.....      0.011726            0.0074627
max  eCDF diff.....      0.023453            0.014925

var ratio (Tr/Co).....    0.98973              1.0131
T-test p-value.....      0.78038              0.85322

Before Matching Minimum p.value: < 2.22e-16
Variable Name(s): X1  Number(s): 1

After Matching Minimum p.value: 0.1594
Variable Name(s): X1  Number(s): 1

Estimate...  0.23881
SE.....  0.073391
T-stat.....  3.2539
p.val.....  0.0011385

Original number of observations.....  145
Original number of treated obs.....  67
Matched number of observations.....  67
Matched number of observations (unweighted).  67

```

Lampiran 17. Sintaks model hubungan variabel kadar AUS (X_7) dan variabel NDP (Y) dengan *Propensity Score Matching* berdasarkan *Multivariate Adaptive Regression Splines* menggunakan software R

```
tableY_X7 = table(X7_predict_c,Y)
colnames(tableY_X7)=c("Tidak Ada NDP", "Ada NDP")
rownames(tableY_X7)=c("Kadar AUS rendah", "Kadar AUS tinggi")
```

Lampiran 18. Output model hubungan variabel kadar AUS (X_7) dan variabel NDP (Y) dengan *Propensity Score Matching* berdasarkan *Multivariate Adaptive Regression Splines* menggunakan software R

X7_predict_c	Y	
	Tidak Ada NDP	Ada NDP
Kadar AUS rendah	38	45
Kadar AUS tinggi	14	48

BIOGRAFI PENULIS



Penulis bernama Amanda Adityaningrum, dilahirkan di Provinsi Gorontalo, 27 Mei 1992, merupakan anak kedua dari tiga bersaudara. Penulis telah menempuh pendidikan formal yaitu, SDN 53 Kota Gorontalo, SMPN 2 Kota Gorontalo, SMAN 3 Kota Gorontalo. Pada tahun 2010 penulis melanjutkan studi program Sarjana Statistika di Universitas Hasanuddin, kemudian pada tahun 2015 penulis melanjutkan studi program Magister Statistika. Pembaca yang ingin memberikan saran, kritik, serta pertanyaan untuk penulis mengenai Tesis ini dapat dikirim melalui *email*: adityaningrum@gmail.com.