



TUGAS AKHIR - KS 184822

***PROPENSITY SCORE STRATIFICATION (PSS)
MENGUNAKAN MULTIVARIATE ADAPTIVE
REGRESSION SPLINE (MARS) PADA KASUS
DIABETES MELITUS TIPE 2***

ANE WAHYUNI
NRP 06211745000005

Dosen Pembimbing
Dr. Bambang Widjanarko Otok, M.Si

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019



TUGAS AKHIR - KS 184822

***PROPENSITY SCORE STRATIFICATION (PSS)
MENGUNAKAN MULTIVARIATE ADAPTIVE
REGRESSION SPLINE (MARS) PADA KASUS
DIABETES MELITUS TIPE 2***

ANE WAHYUNI
NRP 06211745000005

Dosen Pembimbing
Dr. Bambang Widjanarko Otok, M.Si

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019



FINAL PROJECT - KS 184822

***PROPENSITY SCORE STRATIFICATION (PSS)
USING MULTIVARIATE ADAPTIVE
REGRESSION SPLINE (MARS) IN CASE OF
DIABETES MELLITUS TYPE 2***

**ANE WAHYUNI
NRP 06211745000005**

**Supervisor
Dr. Bambang Widjanarko Otok, M.Si**

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCES
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019**

LEMBAR PENGESAHAN

***PROPENSITY SCORE STRATIFICATION (PSS)
MENGUNAKAN MULTIVARIATE ADAPTIVE
REGRESSION SPLINE (MARS) PADA KASUS DIABETES
MELITUS TIPE 2***

TUGAS AKHIR


Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Statistika
pada
Program Studi Sarjana Departemen Statistika
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember

Oleh :

ANE WAHYUNI
NRP. 06211745000005

Disetujui oleh Pembimbing Tugas Akhir :

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

()

Mengetahui
Kepala Departemen Statistika FMKSD-ITS



Dr. Suhanono
NIP. 19671029 199512 1 001

SURABAYA, JULI 2019

Halaman ini sengaja dikosongkan

***PROPENSITY SCORE STRATIFICATION (PSS)
MENGUNAKAN MULTIVARIATE ADAPTIVE
REGRESSION SPLINE (MARS) PADA KASUS DIABETES
MELITUS TIPE 2***

Nama Mahasiswa : Ane Wahyuni
NRP : 0621174500005
Departemen : Statistika
Pembimbing : Dr. Bambang Widjanarko O, M.Si

Abstrak

Diabetes Melitus (DM) merupakan penyakit gangguan metabolik menahun akibat pankreas tidak memproduksi cukup insulin atau tubuh tidak dapat menggunakan insulin yang diproduksi secara efektif. Berdasarkan laporan dari kementerian kesehatan, provinsi Jawa Timur memiliki angka prevalensi DM sebesar 2,1%. Angka ini lebih besar dari angka prevalensi di Indonesia yaitu 1,5%. Melihat tingginya kejadian DM di Jawa Timur, maka perlu dilakukan penelitian untuk mengetahui faktor-faktor penyebab terjadinya komplikasi DM. Faktor-faktor tersebut diduga memiliki hubungan antara satu dengan yang lainnya, sehingga perlu dilakukan analisis untuk mengurangi bias yang disebabkan oleh faktor confounding menggunakan Propensity Score Stratification (PSS). Metode yang digunakan untuk estimasi nilai propensity score adalah Multivariate Adaptive Regression Spline (MARS). Hasil analisis menunjukkan bahwa variabel yang terpilih sebagai confounding adalah aktivitas olahraga. Berdasarkan model MARS, diketahui bahwa variabel usia, jenis kelamin, riwayat genetik, lama menderita, obesitas, dan hipertensi memiliki pengaruh terhadap variabel aktivitas olahraga. Pada estimasi nilai ATE juga diketahui bahwa variabel aktivitas olahraga berpengaruh terhadap komplikasi penyakit DM. Sementara itu, besarnya bias yang dapat direduksi oleh metode PSS MARS adalah kurang dari 25%.

Kata Kunci : *Confounding, Diabetes Melitus, MARS, PSS.*

Halaman ini sengaja dikosongkan

***PROPENSITY SCORE STRATIFICATION (PSS) USING
MULTIVARIATE ADAPTIVE REGRESSION SPLINE
(MARS) IN CASE OF DIABETES MELLITUS TYPE 2***

Name of Student : Ane Wahyuni
NRP : 06211745000005
Department : Statistics
Supervisor : Dr. Bambang Widjanarko O, M.Si

Abstract

Diabetes Mellitus is a chronic metabolic disorder caused by the pancreas not producing enough insulin or the body cannot use insulin produced effectively. Based on reports from the Ministry of Health, East Java province has a DM prevalence rate of 2.1%. This figure is greater than the prevalence rate in Indonesia, which is 1.5%. Seeing the high incidence of DM in East Java, it is necessary to conduct research to know the factors that cause DM complications. These factors are thought to have a relationship between one another, so it is necessary to do an analysis to reduce the bias caused by the confounding factors using Propensity Score Stratification (PSS). The method that used to estimate the value of propensity score is Multivariate Adaptive Regression Spline (MARS). The results of the analysis is show that the variable chosen as confounding is sports activity. Based on MARS model, it is known that age, sex, genetic history, time suffering, obesity, and hypertension have influence on sports activity. In the estimation of ATE value, it is known that sports activity variable affects the complications of DM disease. Meanwhile, the magnitude of the bias that can be reduced by PSS MARS method is less than 25%.

Keywords : *Confounding, Diabetes Melitus, MARS, PSS.*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur Alhamdulillah senantiasa penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “*Propensity Score Stratification (PSS) Menggunakan Multivariate Adaptive Regression Spline (MARS) Pada Kasus Diabetes Melitus Tipe 2*”. Keberhasilan dalam penyelesaian Tugas Akhir ini tidak lepas dari bantuan, arahan, bimbingan, serta dukungan dari berbagai pihak. Untuk itu, penulis mengucapkan terima kasih kepada:

1. Kedua orangtua dan keluarga yang telah memberikan do’a, dukungan, dan semangat yang tiada akhir untuk kesuksesan penulis.
2. Bapak Dr. Bambang Widjanarko Otok, M.Si selaku dosen pembimbing yang telah membimbing saya, memberikan segala masukan, waktu serta pengetahuan demi terselesaikannya Tugas Akhir ini.
3. Bapak Dr. Sutikno, M.Si dan Bapak Dr. Puhadi, M.Sc selaku dosen penguji yang telah memberikan banyak saran, kritik dan masukan demi kesempurnaan Tugas Akhir ini.
4. Bapak Dr. Suhartono selaku Ketua Departemen Statistika ITS yang telah memberikan fasilitas dan sarana dalam penyusunan Tugas Akhir ini.
5. Ibu Santi Wulan Purnami, M.Si, Ph.D selaku Ketua Prodi S1 Statistika dan segenap dosen maupun tenaga pendidik, yang telah mendidik penulis selama menuntut ilmu di Departemen Statistika ITS.
6. Teman-teman seperjuangan Lintas Jalur Statistika Angkatan 2017 yang telah membantu dan memberikan dukungan kepada penulis hingga terselesaikannya Tugas Akhir ini.
7. Pihak-pihak yang telah banyak membantu penulis dalam penyusunan Tugas Akhir ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kata sempurna, Oleh karena itu kritik dan saran yang bersifat membangun sangat diharapkan sebagai masukan dalam peneliti selanjutnya. Semoga Tugas Akhir ini dapat memberikan manfaat baik bagi penulis, pembaca, dan semua pihak.

Surabaya, Juni 2019

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
BAB I ENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 <i>Multivariate Adaptive Regression Spline</i> (MARS).....	7
2.1.1 Klasifikasi MARS Respon Biner.....	10
2.1.2 Ketepatan Klasifikasi.....	18
2.2 <i>Propensity Score</i>	19
2.2.1 Faktor <i>Confounding</i>	20
2.2.2 <i>Propensity Score</i> Menggunakan <i>Multivariate Adaptive Regression Splines</i> (MARS).....	21
2.2.3 <i>Propensity Score Stratification</i> (PSS).....	21
2.3 Uji <i>Balance</i> Kovariat.....	22
2.3.1 Uji <i>t</i>	22
2.3.2 Uji <i>z</i>	23
2.4 Estimasi <i>Average Treatment Effect</i> (ATE).....	23
2.5 Evaluasi <i>Propensity Score Stratification</i>	25
2.6 Diabetes Melitus.....	25

2.7 Faktor yang Mempengaruhi Komplikasi Penyakit DM Tipe 2	27
BAB III METODOLOGI PENELITIAN	31
3.1 Sumber Data.....	31
3.2 Variabel Penelitian.....	31
3.3 Data Penelitian	32
3.4 Langkah Analisis	32
3.5 Diagram Alir Penelitian	33
BAB IV ANALISIS DAN PEMBAHASAN	35
4.1 Karakteristik Pasien Diabetes Melitius (DM) Tipe 2.....	35
4.2 Penentuan Variabel <i>Confounding</i>	39
4.3 Estimasi Nilai <i>Propensity Score</i> dengan MARS.....	41
4.3.1 Pembentukan Model MARS	41
4.3.2 Ketepatan Klasifikasi Model MARS.....	49
4.3.3 Estimasi Nilai <i>Propensity Score</i>	49
4.4 Stratifikasi dan Pengujian <i>Balance</i> Kovariat	50
4.5 Estimasi <i>Average Treatment Effect</i> (ATE)	51
4.6 Perbandingan Bias Sebelum Strata dan Sesudah Strata...	52
BAB V KESIMPULAN DAN SARAN	55
5.1 Kesimpulan	55
5.2 Saran	55
DAFTAR PUSTAKA	57
LAMPIRAN	63

DAFTAR GAMBAR

Gambar 2.1	Kerangka Konsep Faktor-faktor yang Mempengaruhi Komplikasi Penyakit DM Tipe 2	27
Gambar 3.1	Diagram Alir Penelitian.....	33
Gambar 4.1	Persentase Pasien Komplikasi dan Tidak Komplikasi.....	36
Gambar 4.2	Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Jenis Kelamin	37
Gambar 4.3	Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Riwayat Genetik.....	37
Gambar 4.4	Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Obesitas.....	38
Gambar 4.5	Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Hipertensi	38
Gambar 4.2	Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Aktivitas Olahraga	39

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1	Klasifikasi Respon Biner	18
Tabel 3.1	Variabel Penelitian	31
Tabel 3.2	Struktur Data	32
Tabel 4.1	Karakteristik Pasien Untuk Data Kontinu	35
Tabel 4.2	Pengujian Hubungan Antar Variabel X.....	40
Tabel 4.3	Trial and Error Pembentukan Model MARS.....	42
Tabel 4.4	Prediksi Nilai Z Menggunakan MARS	48
Tabel 4.5	Ketepatan Klasifikasi pada Model MARS	49
Tabel 4.6	Estimasi Nilai <i>Propensity Score</i> Menggunakan MARS.....	50
Tabel 4.7	Pembagian Strata dan Pengujian <i>Balance</i> Kovariat	51
Tabel 4.8	Hasil Estimasi ATE dan Standar Error PSS MARS.....	52
Tabel 4.9	<i>Percent Bias Reduction</i> PSS MARS	52

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

Lampiran 1	Surat Izin Penggunaan Data.....	63
Lampiran 2	Data Diabetes Melitus Tipe 2	64
Lampiran 3	Tabulasi Silang Antar Kovariat	67
Lampiran 4	<i>Syntax</i> dan <i>Output Chi-square</i>	72
Lampiran 5	<i>Syntax</i> dan <i>Output MARS</i>	76
Lampiran 6	<i>Syntax</i> dan <i>Output Estimasi Nilai Propensity</i>	86
	<i>Score</i>	86
Lampiran 7	Stratifikasi dan Pengujian <i>Balance</i> Kovariat	87
Lampiran 8	Estimasi Nilai ATE dan PBR untuk 2 Strata	99

Halaman ini sengaja dikosongkan

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes Melitus (DM) merupakan penyakit gangguan metabolik menahun akibat pankreas tidak memproduksi cukup insulin atau tubuh tidak dapat menggunakan insulin yang diproduksi secara efektif. Insulin adalah hormon yang mengatur keseimbangan kadar gula darah. Akibatnya terjadi peningkatan konsentrasi glukosa di dalam darah yang disebut juga dengan hiperglikemia (Kemenkes, 2013a). Penyakit DM terdiri dari dua tipe yaitu DM tipe 1 dan DM tipe 2. DM tipe 1 merupakan penyakit autoimun dimana sistem kekebalan tubuh menyerang dan menghancurkan sel beta di dalam pankreas yang berfungsi menghasilkan insulin. Akibatnya, produksi insulin dalam tubuh menjadi terhenti. Umumnya, penderita DM tipe 1 adalah anak-anak dan remaja. Pada DM tipe 2 pankreas masih bisa menghasilkan insulin, namun produksi insulin tidak efektif sehingga tubuh tidak mampu merespon sepenuhnya terhadap insulin. DM tipe 2 merupakan kasus yang paling banyak terjadi yaitu 90% dari semua kasus DM (International Diabetes Federation, 2017).

Penyakit DM masih menjadi masalah kesehatan yang perlu diperhatikan baik di dunia maupun di Indonesia. Secara global penyakit DM termasuk dalam 10 penyebab utama kematian. Menurut WHO (2016) pada tahun 2012 terdapat 1,5 juta kasus kematian di dunia yang disebabkan oleh diabetes, sedangkan pada tahun 2014 terdapat sebanyak 422 juta orang dewasa menderita penyakit diabetes. Wilayah yang memiliki kasus DM terbanyak adalah wilayah Asia Tenggara dan wilayah Pasifik Barat. IDF (2017) menunjukkan bahwa Indonesia menempati peringkat keenam dengan jumlah kasus DM terbanyak di dunia yaitu mencapai 10,3 juta kasus. Jumlah tersebut diprediksi meningkat hingga tahun 2045 sebanyak 16,7 juta kasus. Provinsi Jawa Timur merupakan salah satu provinsi di Indonesia yang memiliki angka

prevalensi DM yang cukup tinggi. Jawa timur menempati peringkat kelima setelah Yogyakarta, Jakarta, Sulawesi Utara, dan Kalimantan Timur. Data dari Kementerian Kesehatan menunjukkan bahwa Jawa Timur memiliki angka prevalensi DM sebesar 2,1%. Angka ini lebih besar dari angka prevalensi di Indonesia yaitu 1,5% (Kemenkes, 2013b).

Diabetes memang tidak dapat disembuhkan secara total, tetapi kadar gula darah pada penderita dapat dikendalikan agar tidak terjadi berbagai komplikasi penyakit (Holistic Health Solution, 2011). Beberapa komplikasi penyakit yang dapat ditimbulkan dari penyakit DM yaitu kerusakan pada jantung, pembuluh darah, mata, ginjal, serta meningkatkan risiko terjadinya stroke. Salah satu upaya yang dapat dilakukan untuk mencegah komplikasi pada penyakit DM adalah mengetahui faktor-faktor risiko apa saja yang mempengaruhinya. Faktor risiko penyakit DM dapat dikelompokkan menjadi dua, yaitu faktor resiko yang tidak dapat dimodifikasi dan yang dapat dimodifikasi. Faktor risiko yang tidak dapat dimodifikasi adalah ras dan etnik, umur, jenis kelamin, dan riwayat keluarga. Sedangkan faktor risiko yang dapat dimodifikasi adalah perilaku hidup yang kurang sehat, obesitas, kurangnya aktivitas fisik, hipertensi, dislipidemia, dan lain sebagainya (Kemenkes, 2013a).

Penelitian mengenai penyakit DM telah dilakukan oleh Fahmiyah (2016), dimana hasil penelitian menunjukkan bahwa variabel yang secara signifikan mempengaruhi kadar gula darah puasa pasien DM tipe 2 adalah kadar HDL, LDL, dan Trigliserida. Sa'diyah (2017) juga pernah melakukan penelitian mengenai penyakit DM dan hasil penelitian menunjukkan bahwa terdapat empat variabel yang berpengaruh terhadap DM tipe 2 yaitu jenis kelamin, *Body Mass Index* (BMI), usia, dan aktivitas olahraga. Selanjutnya penelitian yang dilakukan oleh Isnaini (2017) mengenai pemodelan kasus DM tipe 2 dengan metode probit biner, dimana pemodelan tersebut menghasilkan empat variabel yang signifikan yaitu tekanan darah, rokok, serat, dan pengawet. Penelitian yang telah dilakukan ini pada umumnya

tidak mempertimbangkan adanya hubungan antar faktor sehingga estimasi menjadi bias dan menyebabkan kesimpulan yang diperoleh tidak akurat. Salah satu analisis yang dapat mengatasi efek bias adalah *propensity score*. *Propensity score* dapat mereduksi dimensi, dimana jika vektor x memiliki banyak kovariat yang direpresentasikan dalam banyak dimensi, maka akan direduksi menjadi satu skor dimensi yaitu nilai *propensity score* (Guo dan Fraser, 2010).

Estimasi nilai *propensity score* dapat dilakukan dengan menggunakan *Multivariate Adaptive Regression Splines* (MARS) karena mudah dalam estimasi dan interpretasi serta tidak memerlukan asumsi bahwa fungsinya harus bersifat linier (Adityaningrum, 2017). Penelitian mengenai MARS pernah dilakukan oleh Sumarno (2014) yaitu klasifikasi menggunakan metode regresi logistik dan MARS. Hasil penelitian menunjukkan bahwa klasifikasi dengan metode MARS tanpa interaksi secara umum menghasilkan tingkat ketepatan yang lebih baik jika dibandingkan dengan metode regresi logistik. Selain itu Anam, Sugiman, dan Sunarmi (2017) juga pernah melakukan penelitian menggunakan MARS pada data kelompok rumah tangga miskin dan tidak miskin. Hasil penelitian menunjukkan bahwa ketepatan klasifikasi dengan metode MARS adalah sebesar 98%.

Salah satu metode *propensity score* yang dapat digunakan untuk mengatasi efek bias adalah *Propensity Score Stratification* (PSS). PSS bertujuan untuk mengelompokan subjek atau pengamatan ke dalam kelas-kelas yang sama berdasarkan estimasi nilai *propensity score* (Austin, 2011). Penelitian mengenai PSS telah pernah dilakukan sebelumnya oleh Akolo (2016) yaitu perbandingan *propensity score stratification bootstrap* dengan *propensity score stratification* tanpa *bootstrap* menggunakan regresi logistik biner. Pada penelitian tersebut diperoleh kesimpulan bahwa PSS bootstrap lebih baik dibandingkan PSS data asli karena memberikan reduksi bias yang lebih besar. Akbar (2017) juga melakukan penelitian PSS pada kasus ISPA dan diperoleh hasil bahwa variabel yang berpengaruh signifikan

terhadap ISPA adalah status perokok pasif dan status imunisasi anak. Selanjutnya penelitian mengenai PSS pada kasus HIV telah dilakukan oleh Ernawati (2018) dimana hasil penelitian menunjukkan bahwa variabel perlakuan pengobatan terapi ARV merupakan variabel yang berpengaruh terhadap variabel infeksi oportunistik pada pasien HIV/AIDS.

Tingginya jumlah kasus penyakit DM yang ada di Jawa Timur akan membuat peluang risiko terjadinya komplikasi juga tinggi. Oleh karena itu, perlu dilakukan penelitian untuk mencegah risiko tersebut dengan mengetahui faktor-faktor yang menyebabkan komplikasi pada penyakit DM. Faktor-faktor penyebab komplikasi DM pada penelitian ini didasarkan pada penelitian yang pernah dilakukan oleh Hasanah (2018) yaitu usia, jenis kelamin, riwayat genetik, lama menderita, obesitas, hipertensi, dan aktivitas olahraga. Faktor-faktor tersebut diduga memiliki hubungan antara satu dengan yang lainnya. Variabel yang menyebabkan adanya hubungan antar faktor disebut juga dengan variabel *confounding*. Pada penelitian ini variabel yang diduga sebagai *confounding* adalah variabel obesitas, hipertensi, dan aktivitas olahraga. Untuk itu, pada penelitian ini dilakukan analisis menggunakan *propensity score stratification* agar dapat mengurangi bias yang terjadi akibat adanya hubungan antar faktor.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka rumusan masalah dalam penelitian ini adalah.

1. Bagaimana hasil penentuan variabel *confounding*?
2. Bagaimana hasil pemodelan MARS untuk mengestimasi nilai *propensity score*?
3. Bagaimana hasil estimasi nilai ATE dan *Percent Bias Reduction* (PBR) dari metode PSS MARS?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah.

1. Mendapatkan variabel yang memiliki potensi sebagai *confounding*.
2. Mendapatkan model MARS untuk estimasi nilai *propensity score*.
3. Mendapatkan nilai ATE dan besar bias yang tereduksi dari metode PSS MARS.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah memberikan tambahan wawasan mengenai penerapan metode *propensity score stratification* menggunakan metode MARS serta dapat digunakan sebagai usaha pencegahan komplikasi penyakit DM tipe 2 melalui faktor-faktor yang mempengaruhi komplikasi penyakit tersebut.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah.

1. Variabel *confounding* yang digunakan hanya satu.
2. Metode yang digunakan untuk estimasi nilai *propensity score* adalah MARS.

Halaman ini sengaja dikosongkan

BAB II

TINJAUAN PUSTAKA

2.1 *Multivariate Adaptive Regression Spline (MARS)*

MARS merupakan salah satu pendekatan regresi nonparametrik multivariat yang berguna untuk mengatasi permasalahan data yang berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor sebanyak $3 \leq n \leq 20$ (Friedman, 1991). Selain itu model MARS mampu menghasilkan prediksi variabel respon yang akurat. Metode MARS menjadi populer karena tidak menentukan tipe khusus seperti hubungan (linear, kuadratik, dan kubik) diantara variabel prediktor dan respon pada proses pembentukan model MARS tidak memerlukan asumsi (Otok, Guritno, Subanar, dan Haryatmi, 2006).

Beberapa istilah yang perlu diperhatikan dalam pemodelan MARS adalah sebagai berikut.

1. Knot

Knot merupakan suatu nilai/titik tempat perubahan pola apabila suatu garis regresi tidak bisa menjelaskan keseluruhan data yang ada dari variabel prediktor. Knot merupakan akhir dari sebuah garis regresi dan juga awal dari garis regresi yang lain (Nash dan Bradford, 2001).

2. Basis Fungsi

Basis fungsi merupakan suatu fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor. Basis fungsi bisa memiliki lebih dari satu variabel yang merupakan fungsi dari tiap garis regresi yang dihasilkan. Maksimum basis fungsi yang diperbolehkan adalah 2 sampai 4 kali jumlah variabel prediktornya.

3. Interaksi

Interaksi merupakan hubungan korelasi antar variabel dengan maksimum interaksi (MI) adalah 1, 2, dan 3. Jika MI lebih dari tiga maka akan menghasilkan model yang lebih kompleks.

4. Minimum Observasi

Minimum Observasi (MO) merupakan jumlah pengamatan paling minimal antar knot sebesar 0, 1, 2, 3, 5, dan 10.

Secara umum estimator model MARS dapat ditulis pada persamaan berikut.

$$\hat{f}(\mathbf{x}) = \hat{\alpha}_0 + \sum_{m=1}^M \hat{\alpha}_m \prod_{k=1}^{K_m} \left[S_{km} (x_{v(k,m)} - t_{km}) \right]_+ \quad (2.1)$$

dengan,

$\hat{\alpha}_0$ = koefisien konstanta basis fungsi B_0

$\hat{\alpha}_m$ = koefisien dari basis fungsi ke- m

M = maksimum basis fungsi

m = banyaknya basis fungsi

K_m = banyaknya interaksi pada basis fungsi m

k = banyaknya interaksi

S_{km} = nilainya 1 atau -1 jika data berada di sebelah kanan atau kiri titik knot

$x_{v(k,m)}$ = variabel prediktor

t_{km} = nilai knot dari variabel prediktor $x_{v(k,m)}$

Berdasarkan persamaan (2.1) model MARS dapat dituliskan sebagai berikut.

$$y_i = \alpha_0 + \sum_{m=1}^M \alpha_m B_{im}(\mathbf{x}_i) + \varepsilon_i \quad (2.2)$$

dengan $B_{im}(\mathbf{x}_i) = \prod_{k=1}^{K_m} \left[S_{km} (x_{v(k,m)i} - t_{km}) \right]_+$. Sehingga jika ditulis dalam bentuk matriks dapat menjadi

$$\mathbf{y} = \mathbf{B}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \quad (2.3)$$

dengan,

$$\mathbf{y} = (y_1, y_2, y_3, \dots, y_n)^T, \quad \boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_M)^T, \quad \boldsymbol{\varepsilon} = (\varepsilon_0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$$

$$B = \begin{pmatrix} 1 & \prod_{k=1}^{K_1} [S_{k1}(x_{v(k,1)1} - t_{k1})] & \cdots & \prod_{k=1}^{K_m} [S_{kM}(x_{v(k,M)1} - t_{kM})] \\ 1 & \prod_{k=1}^{K_1} [S_{k1}(x_{v(k,1)2} - t_{k1})] & \cdots & \prod_{k=1}^{K_m} [S_{kM}(x_{v(k,M)2} - t_{kM})] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} [S_{k1}(x_{v(k,1)n} - t_{k1})] & \cdots & \prod_{k=1}^{K_m} [S_{kM}(x_{v(k,M)n} - t_{kM})] \end{pmatrix}$$

Persamaan (2.1) dapat dijabarkan sebagai berikut.

$$\begin{aligned} \hat{f}(x) &= \hat{\alpha}_0 + \sum_{m=1}^M \hat{\alpha}_m [S_{1m}(x_{v(1,m)} - t_{1m})]_+ \\ &+ \sum_{m=1}^M \hat{\alpha}_m [S_{1m}(x_{v(1,m)} - t_{1m})]_+ \cdot [S_{2m}(x_{v(2,m)} - t_{2m})]_+ \\ &+ \sum_{m=1}^M \hat{\alpha}_m [S_{1m}(x_{v(1,m)} - t_{1m})]_+ \cdot [S_{2m}(x_{v(2,m)} - t_{2m})]_+ \cdot [S_{3m}(x_{v(3,m)} - t_{3m})]_+ \\ &+ \cdots \end{aligned} \quad (2.4)$$

Secara umum persamaan (2.4) dapat dituliskan sebagai berikut.

$$\hat{f}(x) = \hat{\alpha}_0 + \sum_{i=1}^v f_i(x_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^v f_{ij}(x_i, x_j) + \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^v f_{ijk}(x_i, x_j, x_k) + \cdots \quad (2.5)$$

Persamaan (2.5) menunjukkan bahwa penjumlahan suku pertama meliputi semua basis fungsi untuk satu variabel. Penjumlahan suku kedua meliputi semua basis fungsi untuk interaksi antara dua variabel. Penjumlahan suku ketiga meliputi semua basis fungsi untuk interaksi antara tiga variabel, dan seterusnya.

Pemodelan MARS ditentukan berdasarkan *trial and error* pada kombinasi BF, MI, dan MO untuk mendapatkan nilai GCV yang minimum (Nisa' dan Budiantara, 2012). Pemilihan model pada MARS dapat menggunakan metode *stepwise (forward dan backward)*. Pemilihan model dengan menggunakan *forward stepwise* dilakukan untuk mendapatkan jumlah basis fungsi maksimum, sedangkan pada *backward stepwise* dilakukan

pemilihan basis fungsi yang dihasilkan dari *forward stepwise* dengan meminimumkan nilai *Generalized Cross Validation* (GCV). Model terbaik dipilih berdasarkan nilai GCV yang paling minimum. Fungsi GCV minimum didefinisikan sebagai berikut.

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(\mathbf{x}_i)]^2}{\left[1 - \frac{C(M)}{N}\right]^2} \quad (2.6)$$

dengan,

N : banyaknya pengamatan

y_i : variabel respon

\mathbf{x}_i : variabel prediktor

$\hat{f}_M(\mathbf{x}_i)$: nilai taksiran variabel respon pada pengamatan ke- i

$C(M)$: jumlah parameter dalam model = $Trace[\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T] + 1$

\mathbf{B} : matriks basis fungsi

2.1.1 Klasifikasi MARS Respon Biner

Klasifikasi pada model MARS didasarkan pada analisis regresi. Jika variabel respon terdiri dari dua nilai, maka dikatakan sebagai regresi dengan respon biner (Cox dan Snell, 1989), sehingga dapat digunakan model probabilitas dengan persamaan sebagai berikut.

$$P(Z = 1 | X = \mathbf{x}) = \pi(\mathbf{x}) = \frac{e^{f(\mathbf{x})}}{1 + e^{f(\mathbf{x})}}$$

dan

$$P(Z = 0 | X = \mathbf{x}) = 1 - \pi(\mathbf{x}) = \frac{1}{1 + e^{f(\mathbf{x})}}$$

dengan $f(\mathbf{x}) = z = \text{logit } \pi(\mathbf{x})$. Model MARS untuk klasifikasi dapat dinyatakan sebagai berikut.

$$z = \text{logit } \pi(\mathbf{x}) = \ln \left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{K_m} \left[S_{km} (x_{v(k,m)} - t_{km}) \right]_+ \quad (2.7)$$

Variabel z merupakan variabel respon yang bersifat dikotomus (memiliki 2 kategori) yaitu 1 untuk kategori “sukses” dan 0 untuk kategori “gagal”. Distribusi Bernoulli digunakan untuk variabel random biner dengan peluang kejadian sukses $P(Z=1)=\pi$ dan peluang kejadian gagal $P(Z=0)=1-\pi$. Dengan demikian, variabel random Z dikatakan berdistribusi Bernoulli, $z_i \sim \text{Bernoulli}(1, \pi(\mathbf{x}_i))$.

Persamaan (2.7) dapat dijabarkan sebagai berikut

$$\ln\left(\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)}\right) = \alpha_0 + \alpha_1 b_{i1} + \dots + \alpha_M b_{iM} \quad (2.8)$$

$$\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)} = \exp(\alpha_0 + \alpha_1 b_{i1} + \dots + \alpha_M b_{iM}) = \exp(\boldsymbol{\alpha}'\mathbf{b}_i) \quad (2.9)$$

Dari persamaan (2.9) diperoleh persamaan sebagai berikut

$$\begin{aligned} \pi(\mathbf{x}_i) &= (1-\pi(\mathbf{x}_i))\exp(\boldsymbol{\alpha}'\mathbf{b}_i) \\ \pi(\mathbf{x}_i) &= \exp(\boldsymbol{\alpha}'\mathbf{b}_i) - (\pi(\mathbf{x}_i)\exp(\boldsymbol{\alpha}'\mathbf{b}_i)) \\ \pi(\mathbf{x}_i) + (\pi(\mathbf{x}_i)\exp(\boldsymbol{\alpha}'\mathbf{b}_i)) &= \exp(\boldsymbol{\alpha}'\mathbf{b}_i) \\ \pi(\mathbf{x}_i)(1 + \exp(\boldsymbol{\alpha}'\mathbf{b}_i)) &= \exp(\boldsymbol{\alpha}'\mathbf{b}_i) \\ \pi(\mathbf{x}_i) &= \frac{\exp(\boldsymbol{\alpha}'\mathbf{b}_i)}{(1 + \exp(\boldsymbol{\alpha}'\mathbf{b}_i))} \end{aligned} \quad (2.10)$$

dengan,

$$\boldsymbol{\alpha}' = (\alpha_0 \quad \alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_M); \mathbf{b}_i = \begin{pmatrix} 1 \\ b_{i1} \\ b_{i2} \\ \vdots \\ b_{iM} \end{pmatrix}$$

Estimasi parameter $\boldsymbol{\alpha}$ dapat dilakukan dengan metode *Maximum Likelihood Estimation* (MLE). Penaksiran parameter tersebut dapat dilakukan dengan memaksimumkan fungsi *ln likelihood* dengan cara menurunkan fungsi *ln likelihood* terhadap $\boldsymbol{\alpha}$. Sedangkan penaksiran varian dan kovarian diperoleh dari

turunan kedua fungsi \ln *likelihood*. Berikut ini merupakan fungsi *likelihood* yang diperoleh.

$$\begin{aligned}
 L(\boldsymbol{\alpha}) &= \prod_{i=1}^n \pi(\mathbf{x}_i)^{z_i} (1 - \pi(\mathbf{x}_i))^{1-z_i} \\
 &= \prod_{i=1}^n \pi(\mathbf{x}_i)^{z_i} (1 - \pi(\mathbf{x}_i))^{-z_i} (1 - \pi(\mathbf{x}_i)) \\
 &= \prod_{i=1}^n \left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right)^{z_i} (1 - \pi(\mathbf{x}_i))
 \end{aligned} \tag{2.11}$$

Selanjutnya memaksimalkan fungsi \ln *likelihood* sebagai berikut.

$$\begin{aligned}
 l(\boldsymbol{\alpha}) &= \ln(L(\boldsymbol{\alpha})) \\
 &= \ln \left(\prod_{i=1}^n \left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right)^{z_i} (1 - \pi(\mathbf{x}_i)) \right) \\
 &= \sum_{i=1}^n \ln \left[\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right)^{z_i} (1 - \pi(\mathbf{x}_i)) \right] \\
 &= \sum_{i=1}^n \left[\ln \left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right)^{z_i} + \ln(1 - \pi(\mathbf{x}_i)) \right] \\
 &= \sum_{i=1}^n \left[z_i \ln \left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right) + \ln(1 - \pi(\mathbf{x}_i)) \right] \\
 &= \sum_{i=1}^n \left[z_i \ln \left(\frac{\frac{\exp(\mathbf{a}'\mathbf{b}_i)}{1 + \exp(\mathbf{a}'\mathbf{b}_i)}}{1 - \frac{\exp(\mathbf{a}'\mathbf{b}_i)}{1 + \exp(\mathbf{a}'\mathbf{b}_i)}} \right) + \ln \left(1 - \frac{\exp(\mathbf{a}'\mathbf{b}_i)}{1 + \exp(\mathbf{a}'\mathbf{b}_i)} \right) \right] \\
 &= \sum_{i=1}^n \left[z_i \ln \left(\frac{\frac{\exp(\mathbf{a}'\mathbf{b}_i)}{1 + \exp(\mathbf{a}'\mathbf{b}_i)}}{\frac{1 + \exp(\mathbf{a}'\mathbf{b}_i) - \exp(\mathbf{a}'\mathbf{b}_i)}{1 + \exp(\mathbf{a}'\mathbf{b}_i)}} \right) + \ln \left(\frac{1 + \exp(\mathbf{a}'\mathbf{b}_i) - \exp(\mathbf{a}'\mathbf{b}_i)}{1 + \exp(\mathbf{a}'\mathbf{b}_i)} \right) \right]
 \end{aligned}$$

$$\begin{aligned}
l(\boldsymbol{\alpha}) &= \sum_{i=1}^n \left[z_i \ln(\exp(\boldsymbol{\alpha}' \mathbf{b}_i)) + \ln\left(\frac{1}{1 + \exp(\boldsymbol{\alpha}' \mathbf{b}_i)}\right) \right] \\
&= \sum_{i=1}^n \left[z_i \ln(\exp(\boldsymbol{\alpha}' \mathbf{b}_i)) + \ln(1) - \ln(1 + \exp(\boldsymbol{\alpha}' \mathbf{b}_i)) \right] \\
&= \sum_{i=1}^n \left[z_i (\boldsymbol{\alpha}' \mathbf{b}_i) - \ln(1 + \exp(\boldsymbol{\alpha}' \mathbf{b}_i)) \right] \\
&= \sum_{i=1}^n \left[z_i \sum_{m=0}^M (\alpha_m b_{im}) - \ln\left(1 + \exp\left(\sum_{m=0}^M (\alpha_m b_{im})\right)\right) \right]
\end{aligned} \tag{2.12}$$

Untuk menunjukkan fungsi maksimum, maka diturunkan fungsi \ln *likelihood* terhadap α_m dan disamadengankan nol.

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\alpha})}{\partial \alpha_m} &= \frac{\partial \sum_{i=1}^n \left[z_i \sum_{m=0}^M \alpha_m b_{im} - \ln\left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right) \right]}{\partial \alpha_m} \\
0 &= \sum_{i=1}^n \left[\frac{\partial z_i \sum_{m=0}^M \alpha_m b_{im}}{\partial \alpha_m} - \frac{\partial}{\partial \alpha_m} \ln\left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right) \right] \\
0 &= \sum_{i=1}^n \left[z_i b_{im} - \left[\frac{1}{\left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right)} \frac{\partial}{\partial \alpha_m} \left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right) \right] \right] \\
0 &= \sum_{i=1}^n \left[z_i b_{im} - \left[\frac{1}{\left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right)} (b_{im}) \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right) \right] \right]
\end{aligned}$$

$$\begin{aligned}
0 &= \sum_{i=1}^n \left[z_i b_{im} - \frac{\exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)}{\left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right)} (b_{im}) \right] \\
0 &= \sum_{i=1}^n z_i b_{im} - \sum_{i=1}^n (b_{im}) \frac{\exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)}{\left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right)} \quad (2.13) \\
0 &= \sum_{i=1}^n z_i b_{im} - \sum_{i=1}^n (b_{im}) \pi(x_i)
\end{aligned}$$

Estimasi varian kovarian diperoleh dari turunan kedua fungsi \ln *likelihood*. Turunan keduanya adalah sebagai berikut.

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\alpha})}{\partial \alpha_m \partial \alpha_{m^*}} &= \frac{\partial \left[\frac{\partial l(\boldsymbol{\alpha})}{\partial \alpha_m} \right]}{\partial \alpha_{m^*}} \\
&= \frac{\partial \left[\sum_{i=1}^n z_i b_{im} - \sum_{i=1}^n b_{im} \frac{\exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)}{1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)} \right]}{\partial \alpha_{m^*}} \\
&= - \sum_{i=1}^n b_{im} \frac{\partial}{\partial \alpha_{m^*}} \left[\frac{\exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)}{1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)} \right]
\end{aligned}$$

Misalkan:

$$\begin{aligned}
 u &= \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right) & v &= 1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right) = 1 + u \\
 u' &= b_{im} \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right) & v' &= 0 + b_{im} \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right) \\
 &= b_{im} \cdot u & &= b_{im} \cdot u
 \end{aligned}$$

sehingga,

$$\begin{aligned}
 \frac{\partial^2 l(\boldsymbol{\alpha})}{\partial \alpha_m \partial \alpha_{m^*}} &= -\sum_{i=1}^n b_{im} \left[\frac{u' v - u v'}{v^2} \right] \\
 &= -\sum_{i=1}^n b_{im} \left[\frac{(b_{im} \cdot u)(1+u) - (u)(b_{im} \cdot u)}{(1+u)^2} \right] \\
 &= -\sum_{i=1}^n b_{im} \left[\frac{b_{im} \cdot u [(1+u) - u]}{(1+u)^2} \right] \\
 &= -\sum_{i=1}^n b_{im} b_{im} \left[\frac{u}{(1+u)^2} \right]
 \end{aligned} \tag{2.14}$$

Substitusikan nilai u ke dalam persamaan (2.14) dan diperoleh hasil sebagai berikut.

$$\begin{aligned}
 \frac{\partial^2 l(\boldsymbol{\alpha})}{\partial \alpha_m \partial \alpha_{m^*}} &= -\sum_{i=1}^n b_{im} b_{im} \left[\frac{\exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)}{\left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right)^2} \right] \\
 &= -\sum_{i=1}^n b_{im} b_{im} \left[\frac{\exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)}{\left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right)} \left(\frac{1}{\left(1 + \exp\left(\sum_{m=0}^M \alpha_m b_{im}\right)\right)} \right) \right] \\
 &= -\sum_{i=1}^n b_{im} b_{im} \pi(\mathbf{x}_i) (1 - \pi(\mathbf{x}_i))
 \end{aligned} \tag{2.15}$$

Karena hasil turunan pertama pada persamaan (2.13) tidak *close form*, maka nilai α diestimasi dengan metode iterasi Newton Raphson. Rumus metode iterasi Newton-Raphson adalah sebagai berikut.

$$\alpha^{(t+1)} = \alpha^{(t)} - \left(\mathbf{H}(\alpha)^{(t)} \right)^{-1} \mathbf{g}(\alpha)^{(t)} \quad ; t = 1, 2, \dots, \text{sampai konvergen}$$

dengan $\mathbf{H}(\alpha)^{(t)}$ merupakan matriks Hessian dan $\mathbf{g}(\alpha)^{(t)}$ merupakan vektor *gradient* yang dapat dituliskan pada persamaan berikut.

$$\mathbf{H}(\alpha)^{(t)} = \begin{bmatrix} h_{00} & h_{01} & \cdots & h_{0M} \\ h_{10} & h_{11} & \cdots & h_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M0} & h_{M1} & \cdots & h_{MM} \end{bmatrix}; h_{mm} = \frac{\partial^2 l(\alpha)}{\partial \alpha_m \partial \alpha_m} \quad (2.16)$$

$$\mathbf{g}(\alpha)^{(t)} = \left(\frac{\partial l(\alpha)}{\partial \alpha_0}, \frac{\partial l(\alpha)}{\partial \alpha_1}, \dots, \frac{\partial l(\alpha)}{\partial \alpha_M} \right) \quad (2.17)$$

untuk setiap langkah iterasi ke-t, berlaku.

$$\begin{aligned} h_{00}^{(t)} &= \left. \frac{\partial^2 l(\alpha)}{\partial \alpha_0 \partial \alpha_0} \right|_{\alpha^{(t)}} = - \sum_{i=1}^n b_{i0} b_{i0} \pi(\mathbf{x}_i)^{(t)} (1 - \pi(\mathbf{x}_i)^{(t)}) \\ &= - \sum_{i=1}^n \pi(\mathbf{x}_i)^{(t)} (1 - \pi(\mathbf{x}_i)^{(t)}) \\ h_{01}^{(t)} &= \left. \frac{\partial^2 l(\alpha)}{\partial \alpha_0 \partial \alpha_1} \right|_{\alpha^{(t)}} = - \sum_{i=1}^n b_{i0} b_{i1} \pi(\mathbf{x}_i)^{(t)} (1 - \pi(\mathbf{x}_i)^{(t)}) \\ &= - \sum_{i=1}^n b_{i1} \pi(\mathbf{x}_i)^{(t)} (1 - \pi(\mathbf{x}_i)^{(t)}) \\ h_{mm}^{(t)} &= \left. \frac{\partial^2 l(\alpha)}{\partial \alpha_m \partial \alpha_m} \right|_{\alpha^{(t)}} = - \sum_{i=1}^n b_{im} b_{im} \pi(\mathbf{x}_i)^{(t)} (1 - \pi(\mathbf{x}_i)^{(t)}) \\ &= - \sum_{i=1}^n b_{im} b_{im} \pi(\mathbf{x}_i)^{(t)} (1 - \pi(\mathbf{x}_i)^{(t)}) \end{aligned} \quad (2.18)$$

Matriks Hessian pada persamaan (2.16) dapat ditulis sebagai berikut.

$$\mathbf{H}(\boldsymbol{\alpha})^{(t)} = -\left\{ \mathbf{B}^T \text{diag} \left[\pi(\mathbf{x}_1)^{(t)} (1 - \pi(\mathbf{x}_1)^{(t)}), \dots, \pi(\mathbf{x}_n)^{(t)} (1 - \pi(\mathbf{x}_n)^{(t)}) \right] \mathbf{B} \right\}^{-1} \quad (2.19)$$

dengan,

$$\mathbf{B} = \begin{bmatrix} 1 & b_{11} & \cdots & b_{1M} \\ 1 & b_{21} & \cdots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & b_{n1} & \cdots & b_{nM} \end{bmatrix}$$

$\text{diag} \left[\pi(\mathbf{x}_1)^{(t)} (1 - \pi(\mathbf{x}_1)^{(t)}), \dots, \pi(\mathbf{x}_n)^{(t)} (1 - \pi(\mathbf{x}_n)^{(t)}) \right]$ merupakan $n \times n$ matriks diagonal dengan elemen utama adalah $[\pi(\mathbf{x}_i)(1 - \pi(\mathbf{x}_i))]$.

$$\begin{aligned} \mathbf{g}(\boldsymbol{\alpha})_m^{(t)} &= \left. \frac{\partial l(\boldsymbol{\alpha})}{\partial \alpha_m} \right|_{\boldsymbol{\alpha}^{(t)}} = \sum_{i=1}^n z_i b_{im} - \sum_{i=1}^n \hat{\pi}(\mathbf{x}_i) b_{im} \\ &= \sum_{i=1}^n [(z_i - \hat{\pi}(\mathbf{x}_i)) b_{im}] \end{aligned} \quad (2.20)$$

Vektor gradien pada persamaan (2.20) dapat ditulis sebagai berikut.

$$\mathbf{g}(\boldsymbol{\alpha})^{(t)} = \mathbf{B}^T (\mathbf{z} - \pi(\mathbf{x}_i)^{(t)}) \quad (2.21)$$

dengan

$$\pi(\mathbf{x}_i)^{(t)} = \frac{\exp(\hat{\boldsymbol{\alpha}}^{T(t)} \mathbf{b}_i)}{1 + \exp(\hat{\boldsymbol{\alpha}}^{T(t)} \mathbf{b}_i)}$$

Berdasarkan persamaan (2.19) dan (2.21) diperoleh estimasi parameter sebagai berikut.

$$\begin{aligned} \hat{\boldsymbol{\alpha}}^{(t+1)} &= \hat{\boldsymbol{\alpha}}^{(t)} + \left\{ \mathbf{B}^T \text{diag} \left[\pi(\mathbf{x}_1)^{(t)} (1 - \pi(\mathbf{x}_1)^{(t)}), \dots, \pi(\mathbf{x}_n)^{(t)} (1 - \pi(\mathbf{x}_n)^{(t)}) \right] \mathbf{B} \right\}^{-1} \\ &\quad \mathbf{B}^T (\mathbf{z} - \pi(\mathbf{x}_i)^{(t)}) \end{aligned} \quad (2.22)$$

Langkah-langkah iterasi Newton-Raphson adalah sebagai berikut:

1. Menentukan nilai awal dari $\hat{\alpha}$ pada saat iterasi pertama yaitu $\hat{\alpha} = 0$
2. Mulai dari iterasi pertama atau $t = 0$ dan lakukan iterasi dengan menghitung $\hat{\alpha}^{(t+1)}$.

Jika diperoleh $\|\hat{\alpha}^{(t+1)} - \hat{\alpha}^{(t)}\| \leq \Theta$ dimana Θ adalah bilangan yang sangat kecil, maka iterasi berhenti.

2.1.2 Ketepatan Klasifikasi

Ketepatan klasifikasi diperlukan untuk mengetahui pengelompokkan data yang digolongkan dengan tepat pada kelompoknya. *Apparent Error Rate* (APER) didefinisikan sebagai proporsi sampel yang tidak tepat diklasifikasikan (Johnson dan Wichern, 2007). Untuk mengetahui proporsi sampel yang tepat diklasifikasikan dapat dihitung dari nilai TAR (*Total Accuracy Rate*). Berikut ini merupakan tabel pengklasifikasian untuk respon biner.

Tabel 2.1 Klasifikasi Respon Biner

Observasi	Taksiran Observasi	
	y_0	y_1
y_0	n_{00}	n_{01}
y_1	n_{10}	n_{11}

Nilai APER dan TAR didapatkan dengan perhitungan sebagai berikut.

$$APER(\%) = \frac{n_{10} + n_{01}}{n} \times 100\% \quad (2.23)$$

dan

$$TAR(\%) = 1 - APER = 1 - \left(\frac{n_{10} + n_{01}}{n} \times 100\% \right) \quad (2.24)$$

dengan,

n = jumlah observasi

n_{00} = jumlah observasi dari y_0 yang tepat diklasifikasikan sebagai y_0

n_{11} = jumlah observasi dari y_1 yang tepat diklasifikasikan sebagai y_1

n_{01} = jumlah observasi dari y_0 yang salah diklasifikasikan sebagai y_1

n_{10} = jumlah observasi dari y_1 yang salah diklasifikasikan sebagai y_0

2.2 Propensity Score

Propensity Score merupakan probabilitas bersyarat untuk *treatment* tertentu yang diberikan dengan melibatkan kovariat yang diamati (Rosenbaum dan Rubin, 1983). Tujuan dari *propensity score* adalah untuk meningkatkan kualitas estimasi dari eksperimen non acak dengan mencoba menyeimbangkan antar kelompok melalui proses pengacakan. Untuk observasi i ($i = 1, \dots, n$), *propensity score* didefinisikan sebagai probabilitas bersyarat dari kelompok *treatment* tertentu ($Z_i = 1$) dibandingkan dengan kelompok kontrol ($Z_i = 0$) berdasarkan kovariat \mathbf{x}_i yang diamati, dimana pengacakan tidak bisa dilakukan. Secara matematis nilai *propensity score* dapat ditulis sebagai berikut.

$$e(\mathbf{x}_i) = P(Z_i = 1 | \mathbf{X}_i = \mathbf{x}_i) = \frac{\exp(f(\mathbf{x}_i))}{1 + \exp(f(\mathbf{x}_i))} \quad (2.25)$$

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*. Akibatnya, perbandingan sederhana dari hasil rata-rata antara kelompok *treatment* dan kontrol 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).

Langkah pertama yang dilakukan dalam *propensity score* adalah memilih kovariat sebagai *confounder* untuk estimasi

propensity score. Proses pemilihan kovariat pada umumnya didasarkan pada teori sebelumnya atau bukti empiris tentang hubungan antar variabel (Yanovitzky, Zanutto, dan Hornik, 2005). Terdapat empat metode *propensity score* yaitu *Propensity Score Matching* (PSM), *Propensity Score Stratification* (PSS), *Inverse Probability of Treatment Weighting Propensity Score* (IPTW), dan *Covariate Adjustment Propensity Score* (Austin, 2011).

2.2.1 Faktor *Confounding*

Confounding merupakan distorsi atau gangguan dari pengukuran pengaruh *exposure* terhadap hasil karena keterkaitannya dengan faktor *exposure* yang lain yang mempengaruhi hasil (*outcome*). *Confounder* dapat didefinisikan sebagai: (1) faktor yang berhubungan dengan variabel bebas lainnya (2) faktor risiko yang independen terhadap *outcome* (Winkelmayer dan Kurth, 2004).

Pemilihan *confounder* dapat berdasarkan pada teori sebelumnya atau bukti empiris tentang hubungan antar variabel. Untuk memeriksa hubungan antar variabel dapat digunakan uji independensi menggunakan uji *chi-square* (Agresti, 2002). Berikut hipotesis yang digunakan untuk menguji hubungan dua variabel.

H_0 : Tidak terdapat hubungan antar variabel

H_1 : Terdapat hubungan antar variabel

Statistik uji yang digunakan adalah sebagai berikut

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}; \quad \hat{\mu}_{ij} = \frac{n_{i.} \times n_{.j}}{n_{..}} \quad (2.26)$$

dengan,

I : jumlah baris

J : jumlah kolom

n_{ij} : frekuensi pengamatan baris ke- i , kolom ke- j

$\hat{\mu}_{ij}$: frekuensi harapan baris ke- i , kolom ke- j

Daerah kritis: H_0 ditolak jika $\chi^2 > \chi^2_{\alpha; (I-1)(J-1)}$ atau $p\text{-value} < \alpha$.

2.2.2 Propensity Score Menggunakan Multivariate Adaptive Regression Splines (MARS)

Salah satu metode yang dapat digunakan untuk mengestimasi nilai *propensity score* adalah *Multivariate Adaptive Regression Splines* (MARS). Estimasi *propensity score* menggunakan MARS dilakukan dengan cara memasukkan model MARS pada persamaan (2.1) ke dalam model *propensity score* pada persamaan (2.25) sehingga didapatkan estimasi nilai *propensity score* menggunakan MARS sebagai berikut.

$$\hat{e}(x_i) = \frac{\exp(\hat{f}(x_i))}{1 + \exp(\hat{f}(x_i))} \quad (2.27)$$

2.2.3 Propensity Score Stratification (PSS)

Propensity Score Stratification (PSS) merupakan pengelompokan subjek atau pengamatan ke dalam kelas-kelas yang sama berdasarkan estimasi *propensity score*. Subjek diurutkan berdasarkan estimasi *propensity score* (Austin, 2011). Menurut Cochran (1968) dalam Rosenbaum (1983) pengelompokan subjek ke dalam lima strata/subklasifikasi dapat menghilangkan 90% bias, sehingga dapat menghilangkan banyak perbedaan kovariat yang diamati antara kelompok *treatment* dan kelompok kontrol.

Tujuan pengelompokan subjek ke dalam beberapa strata adalah untuk mendapatkan kelompok strata yang tidak memiliki perbedaan antara kelompok *treatment* dan kelompok kontrol yang dikenal dengan *balance* (seimbang). *Balance* pada *propensity score* bertujuan untuk menghilangkan pengaruh dari kovariat terhadap kelompok *treatment* dan kelompok kontrol sehingga diperoleh estimasi efek *treatment* yang tepat dan tidak bias. Jika setelah dilakukan stratifikasi kovariat masih tidak seimbang maka salah satu cara yang dapat dilakukan yaitu membagi lagi subjek

ke dalam strata sampai diperoleh strata yang memenuhi asumsi *balance* kovariat (Dehejia dan Wahba, 1999).

2.3 Uji Balance Kovariat

Pengujian *balance* kovariat antar kelompok *treatment* dan kelompok kontrol dapat dilakukan dengan membandingkan variasi statistik seperti rata-rata, median, variasi, uji kesamaan dua rata-rata (uji- t), uji kesamaan dua proporsi (uji- z) pada kovariat atau menggunakan statistik Kolmogorov-Smirnov (KS) pada setiap kovariat (Li M. , 2012). Untuk menguji perbedaan dua kelompok pada data kontinu dapat digunakan uji- t , sedangkan untuk pengujian perbedaan dua kelompok pada data kategori dapat digunakan uji- z .

2.3.1 Uji t

Hipotesis yang digunakan adalah sebagai berikut (Montgomery, 2001).

$H_0 : \mu_t = \mu_c$ (Tidak ada perbedaan antara kelompok *treatment* dan kelompok kontrol)

$H_1 : \mu_t \neq \mu_c$ (Ada perbedaan antara kelompok *treatment* dan kelompok kontrol)

Statistik uji yang digunakan adalah sebagai berikut.

a. Untuk varians homogen

$$t = \frac{\bar{x}_t - \bar{x}_c}{S \sqrt{\frac{1}{n_t} + \frac{1}{n_c}}} \quad (2.28)$$

dengan,

$$S = \sqrt{\frac{(n_t - 1)S_t^2 + (n_c - 1)S_c^2}{n_t + n_c - 2}}$$

Tolak H_0 jika, $|t| > t_{\alpha/2, df}$; $df = n_t + n_c - 2$ atau $p\text{-value} < \alpha$

b. Untuk varians tidak homogen

$$t = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{S_t^2}{n_t} + \frac{S_c^2}{n_c}}} \quad (2.29)$$

dengan,

$$df = \frac{\left[\left(\frac{S_t^2}{n_t} \right) + \left(\frac{S_c^2}{n_c} \right) \right]^2}{\left[\left(\frac{S_t^2}{n_t} \right) / (n_t - 1) + \left(\frac{S_c^2}{n_c} \right) / (n_c - 1) \right]}$$

Tolak H_0 jika, $|t| > t_{\alpha/2, df}$ atau $p\text{-value} < \alpha$.

2.3.2 Uji z

Hipotesis yang digunakan pada uji-z adalah sebagai berikut.

$H_0 : p_t = p_c$ (Tidak ada perbedaan proporsi antara kelompok *treatment* dan kelompok kontrol)

$H_1 : p_t \neq p_c$ (Ada perbedaan proporsi antara kelompok *treatment* dan kelompok kontrol)

Statistik uji yang digunakan adalah sebagai berikut.

$$z = \frac{\hat{p}_t - \hat{p}_c}{\sqrt{\hat{p}(1 - \hat{p}) \left[\frac{1}{n_t} + \frac{1}{n_c} \right]}} \quad (2.30)$$

dengan,

$$\hat{p} = \frac{x_t + x_c}{n_t + n_c}; \hat{p}_t = \frac{x_t}{n_t}; \hat{p}_c = \frac{x_c}{n_c}$$

Tolak H_0 jika, $|z| > z_{\alpha/2}$ atau $p\text{-value} < \alpha$.

2.4 Estimasi Average Treatment Effect (ATE)

Setelah dilakukan uji *balance* kovariat, maka langkah selanjutnya adalah melakukan estimasi ATE ($\hat{\theta}$). Tujuan dari estimasi ATE adalah untuk mengetahui pengaruh variabel *confounding* terhadap variabel respon. Hipotesis yang digunakan adalah sebagai berikut.

$H_0 : \theta = 0$

$H_1 : \theta \neq 0$

Nilai ATE ($\hat{\theta}$) dan standar error dapat dihitung menggunakan persamaan berikut (Tu & Zhou, 2003).

$$\hat{\theta} = \sum_{k=1}^K \frac{n_{tk} + n_{ck}}{n_t + n_c} (\bar{Y}_{tk} - \bar{Y}_{ck}) \quad (2.31)$$

$$SE(\hat{\theta}) = \sqrt{\sum_{k=1}^K \left(\frac{n_{tk} + n_{ck}}{n_t + n_c} \right) \left(\frac{S_{tk}^2}{n_{tk}} + \frac{S_{ck}^2}{n_{ck}} \right)} \quad (2.32)$$

dimana $\bar{Y}_{tk} = \sum_{i=1}^{n_k} \frac{Y_{tki}}{n_{tki}}$; $\bar{Y}_{ck} = \sum_{i=1}^{n_{ck}} \frac{Y_{cki}}{n_{cki}}$; $n_t = \sum_{k=1}^K n_{tk}$; $n_c = \sum_{k=1}^K n_{ck}$

dengan,

n_t : jumlah subjek kelompok *treatment*

n_c : jumlah subjek kelompok kontrol

n_{tk} : jumlah subjek kelompok *treatment* strata ke- k

n_{ck} : jumlah subjek kelompok kontrol strata ke- k

S_{tk}^2 : varians respon untuk kelompok *treatment* strata ke- k

S_{ck}^2 : varians respon untuk kelompok kontrol strata ke- k

Berdasarkan persamaan (2.33) dan (2.34) diperoleh statistik uji untuk estimasi $\hat{\theta}$ sebagai berikut

$$t = \frac{\hat{\theta}}{SE(\hat{\theta})} \quad (2.33)$$

Daerah kritis untuk taraf signifikansi α adalah tolak H_0 jika, $|t| > t_{\alpha/2, df}$ atau $p\text{-value} < \alpha$.

2.5 Evaluasi *Propensity Score Stratification*

Kebaikan metode PSS dapat dilihat dari kemampuan metode tersebut dalam mereduksi bias untuk setiap kovariat maupun secara keseluruhan. Semakin besar kemampuan metode tersebut dalam mereduksi bias maka semakin baik metode *propensity score* yang digunakan. Bias dalam kovariat dapat didefinisikan sebagai perbedaan rata-rata antara kelompok *treatment* dan kontrol. Berikut ini merupakan perhitungan bias antara kelompok *treatment* dan kontrol (Rosenbaum dan Rubin, 1983).

$$b_{before\ strata} = E(\mathbf{x} | Z = 1) - E(\mathbf{x} | Z = 0) \quad (2.34)$$

$$b_{after\ strata} = E_s(\mathbf{x} | Z = 1) - E_s(\mathbf{x} | Z = 0)$$

dengan,

$b_{before\ strata}$: Bias sebelum stratifikasi

$b_{after\ strata}$: Bias sesudah stratifikasi

Besarnya persentase bias yang direduksi untuk masing-masing kovariat dapat dihitung dengan menggunakan persamaan berikut (Pan dan Bai, 2015).

$$PBR = \frac{b_{before\ strata} - b_{after\ strata}}{b_{before\ strata}} \times 100\% \quad (2.35)$$

2.6 Diabetes Melitus

Diabetes Melitus (DM) merupakan kelompok penyakit metabolik yang ditandai dengan keadaan hiperglikemia akibat dari gangguan sekresi insulin, kerja insulin, atau keduanya (American Diabetes Association, 2014).

Secara umum penyakit DM dibagi menjadi dua tipe yaitu DM tipe 1 dan DM tipe 2. DM tipe 1 merupakan penyakit autoimun dimana sistem kekebalan tubuh secara keliru menghancurkan sel beta pada pankreas yang berfungsi memproduksi insulin. Penderita DM tipe 1 biasanya adalah anak-anak dan remaja, namun terkadang juga bisa terjadi pada orang dewasa muda (Riaz, 2009). Untuk bertahan hidup penderita DM

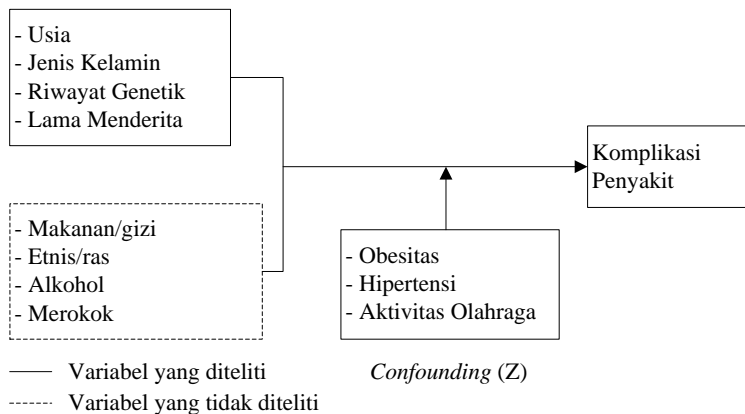
tipe 1 memerlukan pemberian insulin setiap hari sehingga dapat menormalkan kadar glukosa dalam darah. Apabila tidak dilakukan pemberian insulin maka nyawa penderita bisa terancam dan dapat berakibat fatal. Beberapa gejala yang timbul pada penderita DM tipe 1 diantaranya rasa haus dan kencing yang ekstrim, rasa lapar yang terus menerus, penurunan berat badan, perubahan penglihatan, dan kelelahan (Okur, Karantas, dan Siafaka, 2017). DM tipe 2 merupakan suatu gangguan metabolisme yang melibatkan resistensi insulin. Pada penderita DM tipe 2, pankreas masih mampu memproduksi insulin, tetapi tubuh mengalami kesulitan menggunakan hormon pengontrol glukosa ini. Akibatnya pankreas tidak dapat memproduksi insulin yang cukup untuk menanggapi kebutuhan tubuh akan hal itu. Sejauh ini, DM tipe 2 merupakan tipe DM yang paling umum, mencakup 85 hingga 95% kasus DM. Berbeda dengan DM tipe 1 penderita DM tipe 2 tidak membutuhkan pemberian insulin, tetapi hanya memerlukan obat untuk memperbaiki fungsi insulin, menurunkan gula, dan lain-lain. Penyakit DM tipe 2 membutuhkan waktu bertahun-tahun untuk berkembang. Biasanya didahului oleh prediabetes, dimana kadar glukosa darah di atas normal tetapi belum cukup tinggi untuk terdiagnosis diabetes (Riaz, 2009).

Pengontrolan kadar gula darah pada penderita DM sangat penting untuk dilakukan. Semakin buruk kontrol gula darah maka semakin mudah seseorang terkena komplikasi penyakit. Komplikasi akibat penyakit DM dapat dibagi dalam dua kelompok yaitu komplikasi akut dan komplikasi kronis. Komplikasi akut biasanya terjadi secara mendadak. Beberapa komplikasi penyakit akut akibat diabetes yaitu hipoglikemia, ketoasidosis diabetik, dan koma hiper osmolar non-ketotik. Berbeda dengan komplikasi akut, komplikasi kronis biasanya timbul secara perlahan, kadang tidak diketahui, tetapi akhirnya berangsur menjadi makin berat dan membahayakan. Beberapa komplikasi penyakit kronis yaitu kerusakan saraf (neuropati), kerusakan ginjal (*nephropathy*), kerusakan mata, penyakit

jantung, stroke, penyakit pembuluh darah perifer, gangguan pada hati, penyakit paru-paru, gangguan saluran makan, infeksi, dan penyakit kulit (Tandra, 2018).

2.7 Faktor yang Mempengaruhi Komplikasi Penyakit DM Tipe 2

Komplikasi penyakit DM tipe 2 dapat disebabkan oleh beberapa faktor. Berikut ini merupakan kerangka konsep mengenai faktor yang mempengaruhi penyakit DM tipe 2 berdasarkan penelitian yang telah dilakukan oleh Hasanah (2018).



Gambar 2.1 Kerangka Konsep Faktor yang Mempengaruhi Komplikasi Penyakit DM Tipe 2

Berikut ini merupakan teori mengenai faktor-faktor yang mempengaruhi komplikasi penyakit DM tipe 2.

1. Usia

Salah satu faktor yang mempengaruhi komplikasi pada penyakit diabetes mellitus adalah usia. Risiko untuk menderita intoleransi glukosa meningkat seiring dengan meningkatnya usia. Artinya semakin bertambah usia seseorang maka semakin besar risiko terjadinya komplikasi pada penyakit diabetes mellitus (PERKENI, 2015).

2. Jenis Kelamin

Jenis kelamin merupakan salah satu faktor risiko diabetes yang tidak dapat diubah. Berdasarkan penelitian yang dilakukan oleh (Wahyuni, 2010) diperoleh hasil bahwa jenis kelamin perempuan memiliki kecenderungan 1,39 kali untuk terkena penyakit DM dibandingkan dengan jenis kelamin laki-laki.

3. Riwayat Genetik

Riwayat genetik juga merupakan salah satu faktor risiko diabetes yang tidak dapat diubah. Jika salah satu dari orang tua terkena diabetes maka resiko untuk menderita diabetes adalah sebesar 15%. Jika kedua orangtua terkena diabetes maka resiko untuk menderita diabetes adalah sebesar 75% (Diabetes UK, 2010).

4. Lama Menderita

Menurut (Bansal, et al., 2014) risiko pasien yang menderita DM selama 6-10 tahun untuk terkena penyakit komplikasi Neuropati Diabetik Perifer (NDP) adalah sebesar 1,33 kali, sedangkan pasien yang menderita DM lebih dari 15 tahun mempunyai risiko terkena komplikasi NDP sebesar 8,03 kali. Hal ini menunjukkan bahwa semakin lama seseorang menderita DM maka risiko terjadinya komplikasi pada penyakit DM akan semakin besar.

5. Obesitas

Obesitas merupakan suatu kelainan yang ditandai oleh penimbunan jaringan lemak dalam tubuh secara berlebihan. Obesitas akan menyebabkan resistensi insulin sehingga insulin tidak dapat bekerja dengan baik dan kadar gula darah bisa naik. Berdasarkan PERKENI (2015) salah satu faktor risiko terjadinya penyakit DM adalah obesitas. Seseorang dikatakan obesitas jika Indeks Masa Tubuh (IMT) $\geq 27 \text{ kg/m}^2$ (Kemenkes, 2003).

6. Hipertensi

Hipertensi dapat menyebabkan penebalan pembuluh darah arteri sehingga diameter pembuluh darah menjadi menyempit. Penyempitan diameter tersebut akan menyebabkan proses pengangkutan glukosa dari dalam darah menjadi terganggu

sehingga akan berisiko terhadap penyakit DM. seseorang dikatakan hipertensi apabila tekanan darah ≥ 130 mmHg (sistolik) dan ≥ 80 mmHg (diastolic) (American College of Cardiology, 2017).

7. Aktivitas Olahraga

Olahraga sangat bermanfaat untuk mengontrol kadar gula darah. Pada saat tubuh melakukan aktivitas/gerakan, maka sejumlah gula akan dibakar untuk dijadikan tenaga gerak. Sehingga jumlah gula dalam tubuh akan berkurang. Sedangkan pada orang yang jarang berolahraga, maka zat makanan yang masuk ke dalam tubuh tidak dibakar, tetapi hanya akan ditimbun dalam tubuh sebagai lemak dan gula. Jika kadar gula darah tinggi maka akan menyebabkan timbulnya penyakit diabetes mellitus (Lanywati, 2001).

Halaman ini sengaja dikosongkan

BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Penelitian Unggulan Perguruan Tinggi (PUPT) oleh Dr.Bambang Widjanarko Otok, M.Si yang berjudul “Estimasi Parameter *Propensity Score Bootstrap* model *Machine Learning* dan Terapannya pada Data Kesehatan yang Memuat *Confounding*”.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari dua jenis variabel antara lain variabel prediktor dan variabel respon. Variabel-variabel tersebut disajikan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Nama Variabel	Keterangan	Skala Data
Komplikasi Penyakit (Y)	1 = tidak memiliki komplikasi penyakit 0 = memiliki komplikasi penyakit	Nominal
Usia (X ₁)	Usia pasien pada KTP	Rasio
Jenis Kelamin (X ₂)	1 = laki-laki 0 = perempuan	Nominal
Riwayat Genetik (X ₃)	1 = tidak punya riwayat genetic 0 = punya riwayat genetic	Nominal
Lama Menderita (X ₄)	Waktu dalam hitungan tahun	Rasio
Obesitas (X ₅)	1 = tidak obesitas jika <i>Body Mass Index</i> < 27,5 0 = obesitas jika <i>Body Mass Index</i> ≥ 27,5	Nominal
Hipertensi (X ₆)	1 = tidak hipertensi jika tekanan darah sistolik < 130 mmHg 0 = hipertensi jika tekanan darah sistolik ≥ 130 mmHg	Nominal
Aktivitas Olahraga (X ₇)	1 = aktivitas olahraga cukup 0 = aktivitas olahraga kurang	Nominal

3.3 Data Penelitian

Berdasarkan variabel penelitian, maka data penelitian yang digunakan dapat dilihat pada Tabel 3.2. Data penelitian lebih lengkap disajikan pada Lampiran 2.

Tabel 3.2 Struktur Data

Pasien ke-i	Variabel Respon (Y)	Variabel Prediktor (X)				
		X ₁	X ₂	X ₃	...	X ₇
1	0	47	0	0	...	0
2	1	43	1	1	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
96	0	29	0	0	...	0

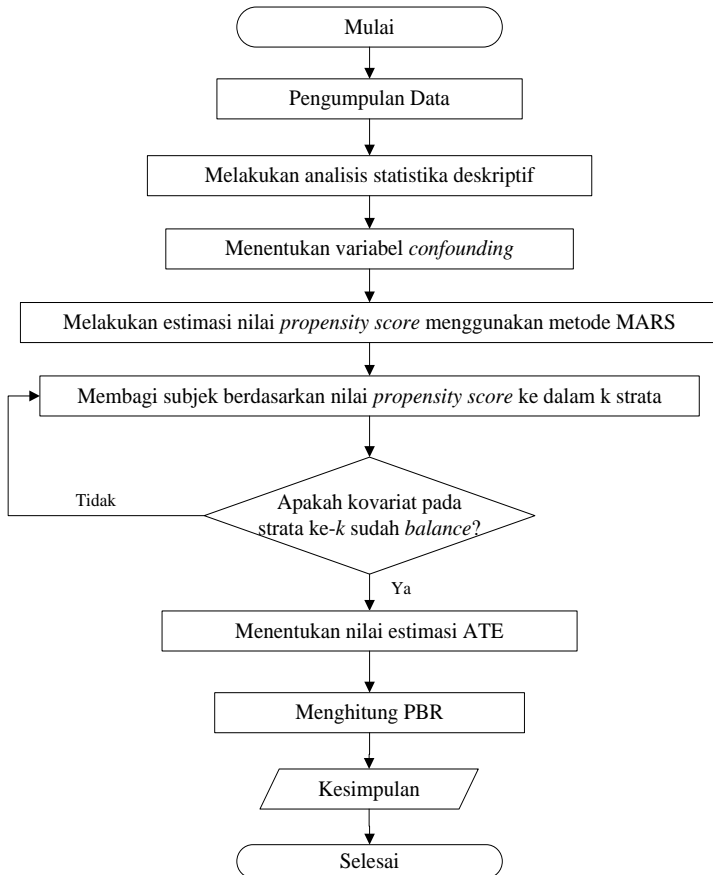
3.4 Langkah Analisis

Langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut.

1. Mengumpulkan data DM tipe 2.
2. Analisis statistika deskriptif untuk mengetahui karakteristik dari pasien DM tipe 2.
3. Menentukan variabel *confounding* menggunakan persamaan (2.26), selanjutnya variabel *confounding* dinotasikan dengan z .
4. Melakukan estimasi nilai *propensity score* dengan metode MARS menggunakan persamaan (2.27).
5. Membagi subjek berdasarkan nilai *propensity score* ke dalam k strata.
6. Melakukan uji *balance* kovariat. Jika datanya kontinu menggunakan persamaan (2.28) atau (2.29), sedangkan jika datanya kategori menggunakan persamaan (2.30).
7. Menentukan nilai estimasi ATE menggunakan persamaan (2.31).
8. Menghitung nilai PBR menggunakan persamaan (2.35).
9. Menginterpretasikan hasil analisis serta memberikan kesimpulan dan saran.

3.5 Diagram Alir Penelitian

Berikut merupakan diagram alir berdasarkan langkah analisis yang telah dijelaskan sebelumnya.



Gambar 3.1 Diagram Alir Penelitian

Halaman ini sengaja dikosongkan

BAB IV ANALISIS DAN PEMBAHASAN

Tahap awal yang dilakukan pada bab ini adalah melakukan analisis statistika deskriptif untuk mengetahui karakteristik data pasien Diabetes Melitus Tipe 2. Selanjutnya dilakukan tahapan analisis PSS MARS yang dimulai dari pemilihan variabel *confounding*, mendapatkan model MARS terbaik, menghitung nilai estimasi *propensity score*, menghitung nilai ATE, sampai menghitung besarnya bias yang direduksi.

4.1 Karakteristik Pasien Diabetes Melitus (DM) Tipe 2

Karakteristik pasien DM tipe 2 dapat dilihat dari analisis deskriptif pada masing-masing variabel. Variabel yang digunakan dalam penelitian ini terdiri dari dua yaitu variabel data kontinu dan variabel data kategori. Variabel data kontinu yang digunakan adalah variabel usia (X_1) dan lama menderita (X_4). Deskripsi variabel data kontinu dapat dilihat dari nilai rata-rata (*mean*), standar deviasi, maksimum, dan minimum. Berikut merupakan deskripsi variabel data kontinu berdasarkan ada atau tidaknya komplikasi pada pasien DM tipe 2.

Tabel 4.1 Karakteristik Pasien untuk Data Kontinu

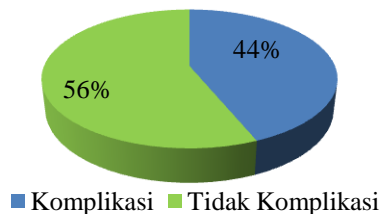
Variabel	Status Pasien	Mean	Standar Deviasi	Max	Min
Usia (Tahun)	Komplikasi	52,02	15,67	77	29
	Tidak Komplikasi	53,56	14,63	75	29
Lama Menderita (Tahun)	Komplikasi	4	2,08	7	1
	Tidak Komplikasi	4,11	1,96	7	1

Berdasarkan Tabel 4.1 diketahui bahwa rata-rata usia pasien DM tipe 2 yang tidak memiliki komplikasi hampir sama dengan pasien yang memiliki komplikasi yaitu ± 52 tahun. Adapun standar deviasi usia pasien yang memiliki komplikasi

tidak jauh berbeda dengan pasien yang tidak memiliki komplikasi. Ini berarti keragaman usia antara pasien yang memiliki komplikasi dan pasien yang tidak memiliki komplikasi hampir sama. Selain itu, diketahui juga bahwa usia paling muda pasien yang memiliki komplikasi adalah 29 tahun sedangkan yang paling tua adalah 77 tahun. Hal ini berarti komplikasi DM tipe 2 tidak hanya terjadi pada usia yang tua namun juga bisa terjadi pada usia muda.

Tabel 4.1 juga menunjukkan bahwa rata-rata lama menderita pasien DM tipe 2 yang memiliki komplikasi dan yang tidak memiliki komplikasi tidak jauh berbeda yaitu sebesar 4 tahun dan 4,11 tahun. Selain itu, standar deviasi lama menderita untuk pasien yang memiliki komplikasi juga hampir sama dengan pasien yang tidak memiliki komplikasi. Ini berarti keragaman waktu lama menderita untuk pasien yang memiliki komplikasi tidak jauh berbeda dengan pasien yang tidak memiliki komplikasi. Adapun waktu lama menderita tercepat untuk pasien yang memiliki komplikasi dan yang tidak memiliki komplikasi adalah sama yaitu sebesar 1 tahun, sedangkan waktu terlamanya adalah 7 tahun.

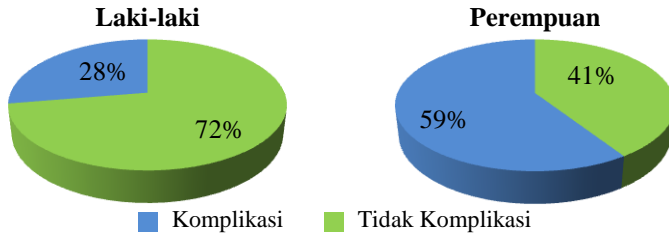
Karakteristik pasien DM tipe 2 selanjutnya dapat dilihat dari variabel data kategori. Variabel data kategori yang digunakan adalah status komplikasi (Y), jenis kelamin (X_2), riwayat genetik (X_3), obesitas (X_5), hipertensi (X_6), dan aktivitas olahraga (X_7). Berikut ini merupakan deskripsi dari variabel data kategori.



Gambar 4.1 Persentase Pasien Komplikasi dan Tidak Komplikasi

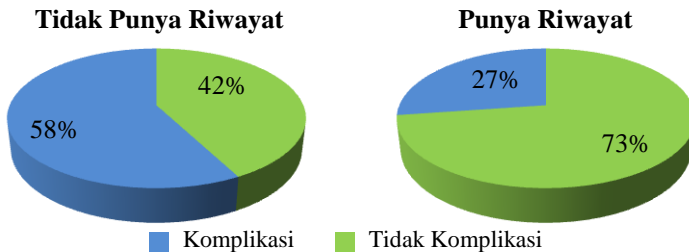
Berdasarkan Gambar 4.1 terlihat bahwa pasien DM tipe 2 yang tidak memiliki komplikasi lebih banyak daripada pasien

yang memiliki komplikasi yaitu sebesar 56%. Ini berarti sebagian besar pasien DM tipe 2 tidak memiliki komplikasi penyakit.



Gambar 4.2 Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Jenis Kelamin

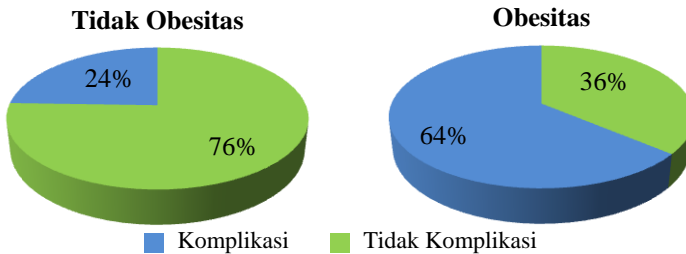
Gambar 4.2 memberikan informasi bahwa pasien laki-laki yang memiliki komplikasi lebih sedikit daripada yang tidak memiliki komplikasi yaitu sebesar 28%, sedangkan pasien perempuan yang memiliki komplikasi lebih banyak daripada yang tidak memiliki komplikasi dengan persentase 59%. Jika dibandingkan antara pasien laki-laki dan perempuan, pasien perempuan cenderung lebih banyak yang terkena komplikasi daripada pasien laki-laki.



Gambar 4.3 Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Riwayat Genetik

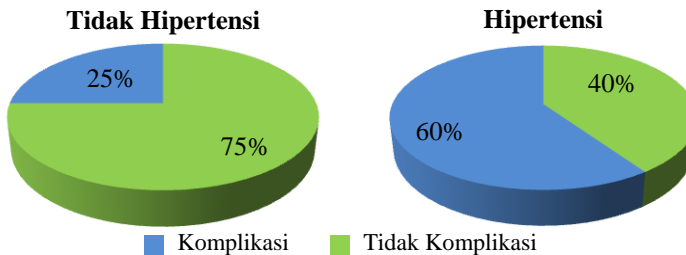
Berdasarkan Gambar 4.3 terlihat bahwa persentase pasien yang tidak mempunyai riwayat genetic sekaligus komplikasi adalah 58%. Sementara itu pasien yang mempunyai riwayat genetic sekaligus komplikasi hanya memiliki persentase sebesar 27% dibandingkan dengan yang tidak komplikasi. Hal ini berarti,

pasien yang mempunyai riwayat genetic belum tentu mudah terkena komplikasi penyakit.



Gambar 4.4 Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Obesitas

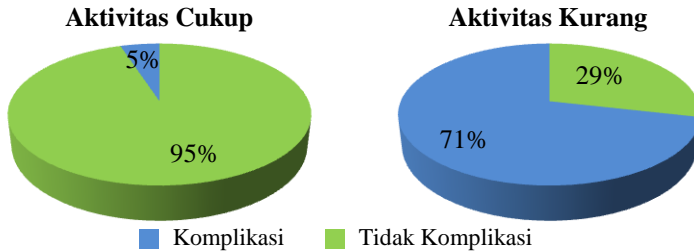
Pasien yang tidak obesitas dan tidak komplikasi memiliki persentase yang lebih besar daripada yang memiliki komplikasi yaitu sebesar 76% (Gambar 4.4). Selain itu, persentase pasien yang obesitas dan memiliki komplikasi lebih besar daripada yang tidak komplikasi. Hal ini menunjukkan bahwa pasien yang memiliki obesitas mempunyai kecenderungan untuk terkena komplikasi penyakit.



Gambar 4.5 Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Hipertensi

Gambar 4.5 memberikan informasi bahwa pasien yang tidak hipertensi sekaligus menderita komplikasi memiliki persentase yang cukup kecil yaitu sebesar 25%. Selain itu, diketahui juga bahwa sebagian besar pasien yang hipertensi memiliki kecenderungan untuk terkena komplikasi penyakit. Ini

berarti jika pasien mengalami hipertensi maka akan mudah juga terkena komplikasi.



Gambar 4.6 Persentase Pasien Komplikasi dan Tidak Komplikasi Berdasarkan Aktivitas Olahraga

Pasien yang memiliki aktivitas olahraga cukup sekaligus tidak memiliki komplikasi lebih banyak daripada pasien yang terkena komplikasi dengan persentase sebesar 95% (Gambar 4.6). Hal ini berarti aktivitas olahraga yang cukup dapat mengurangi resiko terjadinya komplikasi penyakit. Pasien yang memiliki aktivitas olahraga kurang lebih banyak menderita komplikasi daripada yang tidak komplikasi. Hal ini menunjukkan bahwa kurangnya aktivitas olahraga dapat menyebabkan komplikasi penyakit.

4.2 Penentuan Variabel *Confounding*

Langkah pertama yang dilakukan dalam analisis *propensity score stratification* adalah menentukan kovariat yang digunakan sebagai variabel *confounding*. Penentuan variabel *confounding* dapat berdasarkan teori atau bukti empiris hubungan antar variabel. Variabel *confounding* yang dipilih dalam penelitian ini yaitu kovariat yang paling banyak memiliki hubungan dengan kovariat lainnya. Untuk menguji hubungan antar variabel dapat digunakan statistik uji *chi-square* dengan taraf signifikansi $\alpha = 1\%$. Hasil pengujian hubungan antar kovariat dapat dilihat pada Tabel 4.2 (lihat Lampiran 4).

Tabel 4.2 Pengujian Hubungan Antar Variabel X

Variabel	χ^2	<i>p-value</i>	Keputusan
X ₃ dan X ₂	0,399	0,528	Gagal Tolak H ₀
X ₅ dan X ₂	4,188	0,041	Gagal Tolak H ₀
X ₆ dan X ₂	0,399	0,528	Gagal Tolak H ₀
X ₇ dan X ₂	7,061	0,008	Tolak H ₀
X ₂ dan X ₃	0,399	0,528	Gagal Tolak H ₀
X ₅ dan X ₃	0,399	0,528	Gagal Tolak H ₀
X ₆ dan X ₃	2,483	0,115	Gagal Tolak H ₀
X ₇ dan X ₃	10,147	0,001	Tolak H ₀
X ₂ dan X ₅	4,188	0,041	Gagal Tolak H ₀
X ₃ dan X ₅	0,399	0,528	Gagal Tolak H ₀
X ₆ dan X ₅	3,463	0,063	Gagal Tolak H ₀
X ₇ dan X ₅	7,433	0,006	Tolak H ₀
X ₂ dan X ₆	0,399	0,528	Gagal Tolak H ₀
X ₃ dan X ₆	2,483	0,115	Gagal Tolak H ₀
X ₅ dan X ₆	3,463	0,063	Gagal Tolak H ₀
X ₇ dan X ₆	5,543	0,019	Gagal Tolak H ₀
X ₂ dan X ₇	7,061	0,008	Tolak H ₀
X ₃ dan X ₇	10,147	0,001	Tolak H ₀
X ₅ dan X ₇	7,433	0,006	Tolak H ₀
X ₆ dan X ₇	5,543	0,019	Gagal Tolak H ₀

Berdasarkan Tabel 4.2 diketahui bahwa variabel jenis kelamin (X₂) memiliki hubungan yang signifikan dengan satu variabel lainnya yaitu aktivitas olahraga (X₇) karena memiliki nilai *p-value* < α (0,01). Variabel riwayat genetik (X₃) dan variabel obesitas (X₅) juga memiliki hubungan signifikan hanya dengan satu variabel yaitu variabel aktivitas olahraga (X₇), sedangkan variabel hipertensi (X₆) tidak memiliki hubungan yang signifikan dengan variabel lainnya. Jika dibandingkan dengan variabel sebelumnya, variabel aktivitas olahraga (X₇) merupakan variabel yang paling banyak memiliki hubungan dengan variabel lain, yaitu jenis kelamin (X₂), riwayat genetik (X₃), dan obesitas (X₅). Oleh karena itu, variabel aktivitas olahraga (X₇) dipilih menjadi variabel *confounding*.

Selanjutnya dilakukan pengujian hubungan antar variabel *confounding* (X_7) dengan variabel respon (Y) untuk mengetahui apakah terdapat hubungan antara variabel *confounding* dengan variabel respon. Berdasarkan hasil pengujian pada Lampiran 2 diketahui bahwa terdapat hubungan yang signifikan antara aktivitas olahraga dan status komplikasi penyakit DM tipe 2 karena nilai *p-value* $< \alpha$ (0,01).

4.3 Estimasi Nilai Propensity Score dengan MARS

Setelah variabel *confounding* ditentukan, langkah selanjutnya yang dilakukan adalah menghitung nilai estimasi *propensity score*. Metode yang digunakan untuk mengestimasi nilai *propensity score* dalam penelitian ini yaitu MARS. Variabel *confounding* yang telah terpilih akan dijadikan sebagai variabel respon dalam analisis MARS. Sehingga variabel aktivitas olahraga digunakan sebagai variabel respon dan selanjutnya dinotasikan dengan Z. Variabel prediktor terdiri dari 6 variabel yaitu usia (X_1), jenis kelamin (X_2), riwayat genetik (X_3), lama menderita (X_4), obesitas (X_5), dan hipertensi (X_6).

4.3.1 Pembentukan Model MARS

Pembentukan model MARS dilakukan dengan cara *trial and error* untuk semua kombinasi Basis Fungsi (BF), Maksimum Interaksi (MI), dan Minimum Observasi (MO). Jumlah maksimum basis fungsi yaitu 2 sampai 4 kali jumlah variabel prediktor. Karena terdapat 6 variabel prediktor, maka jumlah basis fungsi yang digunakan dalam penelitian ini adalah 12, 18, dan 24. Maksimum interaksi yang digunakan yaitu sebanyak 1, 2, dan 3 karena jika lebih dari 3 maka akan menghasilkan model yang sangat kompleks. Untuk minimum observasi yang digunakan adalah 0, 1, 2, 3, 5, dan 10. Berdasarkan kombinasi tersebut, banyaknya model yang terbentuk adalah sebanyak 54 model. Berikut merupakan *trial and error* pembentukan model MARS untuk setiap kombinasi BF, MI, dan MO (lihat Lampiran 5).

Tabel 4.3 *Trial and Error* Pembentukan Model MARS

Model	Kombinasi			GCV	R ²
	BF	MI	MO		
1	12	1	0	0,176317	0,274582
2	12	1	1	0,176317	0,274582
3	12	1	2	0,176317	0,274582
4	12	1	3	0,176317	0,274582
5	12	1	5	0,176317	0,274582
6	12	1	10	0,176317	0,274582
7	12	2	0	0,17096	0,296622
8	12	2	1	0,17096	0,296622
9	12	2	2	0,17096	0,296622
10	12	2	3	0,17096	0,296622
11	12	2	5	0,17096	0,296622
12	12	2	10	0,17096	0,296622
13	12	3	0	0,17096	0,296622
14	12	3	1	0,17096	0,296622
15	12	3	2	0,17096	0,296622
16	12	3	3	0,17096	0,296622
17	12	3	5	0,17096	0,296622
18	12	3	10	0,17096	0,296622
19	18	1	0	0,16642	0,315302
20	18	1	1	0,16642	0,315302
21	18	1	2	0,16642	0,315302
22	18	1	3	0,16642	0,315302
23	18	1	5	0,16642	0,315302
24	18	1	10	0,168408	0,30712
25	18	2	0	0,159446	0,343995
26	18	2	1	0,142794	0,412506
27	18	2	2	0,150894	0,37918
28	18	2	3	0,150894	0,37918
29	18	2	5	0,150894	0,37918

Tabel 4.3 *Trial and Error* Pembentukan Model MARS (Lanjutan)

Model	Kombinasi			GCV	R ²
	BF	MI	MO		
30	18	2	10	0,161101	0,337185
31	18	3	0	0,16642	0,343995
32	18	3	1	0,138174	0,431511
33	18	3	2	0,144646	0,404885
34	18	3	3	0,144921	0,403754
35	18	3	5	0,148211	0,390219
36	18	3	10	0,160682	0,338907
37	24	1	0	0,144407	0,405867
38	24	1	1	0,144407	0,405867
39	24	1	2	0,144407	0,405867
40	24	1	3	0,144407	0,405867
41	24	1	5	0,144407	0,405867
42	24	1	10	0,157353	0,352606
43	24	2	0	0,149123	0,386466
44	24	2	1	0,126029	0,481482
45	24	2	2	0,12748	0,47551
46	24	2	3	0,13	0,465142
47	24	2	5	0,130618	0,462599
48	24	2	10	0,153456	0,36864
49	24	3	0	0,14777	0,392031
50	24	3	1	0,105668	0,565254
51	24	3	2	0,120372	0,504754
52	24	3	3	0,121644	0,499524
53	24	3	5	0,127773	0,474304
54	24	3	10	0,150499	0,380806

Pemilihan model terbaik MARS ditentukan dari nilai GCV yang paling kecil. Berdasarkan Tabel 4.3 diketahui bahwa model yang memiliki nilai GCV terkecil adalah model 50 dengan basis fungsi = 24, maksimum interaksi = 3, dan minimum observasi = 1

serta nilai GCV sebesar 0,565. Model MARS terbaik yang didapatkan adalah sebagai berikut.

$$\begin{aligned}\hat{z} = & 0,367 + 0,230BF1 - 0,387BF2 + 0,385BF3 - 0,014BF4 \\ & - 0,027BF5 + 0,146BF6 + 0,104BF7 - 0,26BF8 - 0,375BF9 \\ & - 0,145BF10 + 0,018BF11 + 0,336BF12 + 0,046BF13 \\ & + 0,041BF14 - 0,012BF15 - 0,007BF16\end{aligned}$$

dengan,

$$BF1 = X_2$$

$$BF2 = X_3$$

$$BF3 = X_6$$

$$BF4 = \max(0, 58 - X_1)$$

$$BF5 = \max(0, X_1 - 58)$$

$$BF6 = \max(0, 3 - X_4)$$

$$BF7 = \max(0, X_4 - 3)$$

$$BF8 = X_3 * X_5$$

$$BF9 = X_3 * X_6$$

$$BF10 = \max(0, X_1 - 66) * X_6$$

$$BF11 = \max(0, 70 - X_1) * X_6$$

$$BF12 = \max(0, X_1 - 70) * X_6$$

$$BF13 = \max(0, 51 - X_1) * X_3 * X_5$$

$$BF14 = \max(0, X_1 - 51) * X_3 * X_5$$

$$BF15 = \max(0, 70 - X_1) * \max(0, X_4 - 4) * X_6$$

$$BF16 = \max(0, 70 - X_1) * \max(0, 4 - X_4) * X_6$$

Berdasarkan model MARS terbaik diketahui bahwa variabel usia, jenis kelamin, riwayat genetic, lama menderita, obesitas, dan hipertensi memiliki pengaruh terhadap aktivitas olahraga.

Interpretasi model untuk setiap basis fungsi adalah sebagai berikut.

1. $BF1 = X_2$

Artinya, koefisien dari BF1 akan bermakna jika jenis kelamin (X_2) pasien adalah laki-laki. Pasien DM tipe 2 yang berjenis kelamin laki-laki cenderung memiliki aktivitas olahraga yang cukup sebesar 1,259 kali dibandingkan dengan pasien perempuan.

2. $BF2 = X_3$

Artinya, koefisien dari BF2 akan bermakna jika pasien tidak mempunyai riwayat genetik (X_3). Pasien DM tipe 2 yang tidak mempunyai riwayat genetik cenderung memiliki aktivitas olahraga yang cukup sebesar 0,679 kali dibandingkan dengan pasien yang memiliki riwayat genetik.

3. $BF3 = X_6$

Artinya, koefisien dari BF3 akan bermakna jika pasien tidak memiliki hipertensi (X_6). Pasien DM tipe 2 yang tidak hipertensi cenderung memiliki aktivitas olahraga yang cukup sebesar 1,47 kali dibandingkan dengan pasien yang hipertensi.

4. $BF4 = \max(0, 58 - X_1)$

Artinya, koefisien dari BF4 akan bermakna jika usia (X_1) pasien lebih kecil dari 58 tahun. Pasien DM tipe 2 yang berusia kurang dari 58 tahun cenderung memiliki aktivitas olahraga yang cukup sebesar 0,986 kali dibandingkan dengan pasien yang berusia lebih dari 58 tahun.

5. $BF5 = \max(0, X_1 - 58)$

Artinya, koefisien dari BF5 akan bermakna jika usia (X_1) pasien lebih besar dari 58 tahun. Pasien DM tipe 2 yang berusia lebih dari 58 tahun cenderung memiliki aktivitas olahraga yang cukup sebesar 0,973 kali dibandingkan dengan pasien yang berusia kurang dari 58 tahun.

6. $BF6 = \max(0, 3 - X_4)$

Artinya, koefisien dari BF6 akan bermakna jika lama menderita (X_4) pasien lebih kecil dari 3 tahun. Pasien DM tipe 2 dengan lama menderita kurang dari 3 tahun cenderung

memiliki aktivitas olahraga yang cukup sebesar 1,157 kali dibandingkan dengan pasien yang menderita DM lebih dari 3 tahun.

7. $BF7 = \max(0, X_4 - 3)$

Artinya, koefisien dari BF7 akan bermakna jika lama menderita (X_4) pasien lebih besar dari 3 tahun. Pasien DM tipe 2 dengan lama menderita lebih dari 3 tahun cenderung memiliki aktivitas olahraga yang cukup sebesar 1,11 kali dibandingkan dengan pasien yang menderita DM kurang dari 3 tahun.

8. $BF8 = X_3 * X_5$

Artinya, koefisien dari BF8 akan bermakna jika pasien tidak memiliki riwayat genetik (X_3) dan tidak obesitas (X_5). Pasien DM tipe 2 yang tidak memiliki riwayat genetik dan tidak obesitas cenderung memiliki aktivitas olahraga yang cukup sebesar 0,771 kali dibandingkan dengan pasien yang memiliki riwayat genetic dan obesitas.

9. $BF9 = X_3 * X_6$

Artinya, koefisien dari BF9 akan bermakna jika pasien tidak memiliki riwayat genetik (X_3) dan tidak hipertensi (X_6). Pasien DM tipe 2 yang tidak memiliki riwayat genetik dan tidak hipertensi cenderung memiliki aktivitas olahraga yang cukup sebesar 0,687 kali dibandingkan dengan pasien yang memiliki riwayat genetik dan hipertensi.

10. $BF10 = \max(0, X_1 - 66) * X_6$

Artinya, koefisien dari BF10 akan bermakna jika usia (X_1) pasien lebih besar dari 66 tahun dan pasien tidak hipertensi (X_6). Pasien DM tipe 2 yang berusia lebih dari 66 tahun dan tidak hipertensi cenderung memiliki aktivitas olahraga yang cukup sebesar 0,865 kali dibandingkan dengan pasien yang berusia kurang dari 66 tahun dan hipertensi.

11. $BF11 = \max(0, 70 - X_1) * X_6$

Artinya, koefisien dari BF11 akan bermakna jika usia (X_1) pasien lebih kecil dari 70 tahun dan pasien tidak hipertensi

(X_6). Pasien DM tipe 2 yang berusia kecil dari 70 tahun dan tidak hipertensi cenderung memiliki aktivitas olahraga yang cukup sebesar 1,018 kali dibandingkan dengan pasien yang berusia lebih dari 70 tahun dan hipertensi.

12. $BF12 = \max(0, X_1 - 70) * X_6$

Artinya, koefisien dari BF12 akan bermakna jika usia (X_1) pasien lebih besar dari 70 tahun dan pasien tidak hipertensi (X_6). Pasien DM tipe 2 yang berusia lebih dari 70 tahun dan tidak hipertensi cenderung memiliki aktivitas olahraga yang cukup sebesar 1,399 kali dibandingkan dengan pasien yang berusia kurang dari 70 tahun dan hipertensi.

13. $BF13 = \max(0, 51 - X_1) * X_3 * X_5$

Artinya, koefisien dari BF13 akan bermakna jika usia (X_1) pasien lebih kecil dari 51 tahun, tidak memiliki riwayat genetic (X_3) dan tidak obesitas (X_5). Pasien DM tipe 2 yang berusia kurang dari 51 tahun, tidak memiliki riwayat genetik, dan tidak obesitas cenderung memiliki aktivitas olahraga yang cukup sebesar 1,047 kali dibandingkan dengan pasien yang berusia lebih dari 51 tahun, memiliki riwayat genetik dan obesitas.

14. $BF14 = \max(0, X_1 - 51) * X_3 * X_5$

Artinya, koefisien dari BF14 akan bermakna jika usia (X_1) pasien lebih besar dari 51 tahun, tidak memiliki riwayat genetik (X_3) dan tidak obesitas (X_5). Pasien DM tipe 2 yang berusia lebih dari 51 tahun, tidak memiliki riwayat genetik, dan tidak obesitas cenderung memiliki aktivitas olahraga yang cukup sebesar 1,042 kali dibandingkan dengan pasien yang berusia kurang dari 51 tahun, memiliki riwayat genetik dan obesitas.

15. $BF15 = \max(0, 70 - X_1) * \max(0, X_4 - 4) * X_6$

Artinya, koefisien dari BF15 akan bermakna jika usia (X_1) pasien lebih kecil dari 70 tahun, lama menderita (X_4) pasien lebih besar dari 4 tahun, dan tidak hipertensi (X_6). Pasien DM tipe 2 yang berusia kurang dari 70 tahun, lama menderita

lebih besar dari 4 tahun, dan tidak hipertensi cenderung memiliki aktivitas olahraga yang cukup sebesar 0,988 kali dibandingkan dengan pasien yang berusia lebih dari 51 tahun, lama menderita kurang dari 4 tahun, dan hipertensi.

$$16. BF16 = \max(0, 70 - X_1) * \max(0, 4 - X_4) * X_6$$

Artinya, koefisien dari BF16 akan bermakna jika usia (X_1) pasien lebih kecil dari 70 tahun, lama menderita (X_4) pasien lebih kecil dari 4 tahun, dan tidak hipertensi (X_6). Pasien DM tipe 2 yang berusia kurang dari 70 tahun, lama menderita lebih kecil dari 4 tahun, dan tidak hipertensi cenderung memiliki aktivitas olahraga yang cukup sebesar 0,993 kali dibandingkan dengan pasien yang berusia lebih dari 51 tahun, lama menderita lebih dari 4 tahun, dan hipertensi.

Berdasarkan model terbaik MARS yang telah didapatkan, maka hasil perhitungan untuk prediksi nilai Z adalah sebagai berikut.

Tabel 4.4 Prediksi Nilai Z Menggunakan MARS

Data ke-	$\hat{f}(x_i)$							
1-8	0,210	0,287	-0,243	0,388	0,077	0,249	0,283	0,538
9-16	0,587	0,159	0,635	0,652	-0,102	0,086	-0,426	0,534
17-24	1,014	0,379	-0,133	0,086	0,455	0,524	1,181	0,284
25-32	0,743	0,391	0,209	0,397	1,077	0,769	0,497	0,099
33-40	0,147	0,012	0,441	0,103	0,040	1,006	0,634	1,036
41-48	0,588	-0,114	0,685	0,871	0,293	-0,047	0,355	0,830
49-56	0,426	0,725	0,268	0,297	0,743	0,878	0,507	0,026
57-64	0,031	0,746	0,115	1,202	0,829	1,040	0,347	0,330
65-72	0,618	0,143	1,208	0,898	0,355	0,290	1,032	0,816
73-80	-0,050	0,279	0,455	-0,222	0,240	0,272	0,366	-0,062
81-88	1,109	0,057	0,355	0,837	0,325	0,061	-0,047	0,444
89-96	1,180	0,352	0,242	0,138	0,624	0,099	0,296	0,012

Hasil prediksi nilai Z pada Tabel 4.4 dapat dikategorikan dalam 2 kategori yaitu 0 dan 1. Hasil kategori prediksi lebih lengkap dapat dilihat pada Lampiran 3.

4.3.2 Ketepatan Klasifikasi Model MARS

Ketepatan klasifikasi digunakan untuk mengetahui pengelompokan data yang digolongkan dengan tepat pada kelompoknya. Berikut ini merupakan hasil dari ketepatan klasifikasi model MARS.

Tabel 4.5 Ketepatan Klasifikasi pada Model MARS

Data Aktual Z	Data Prediksi Z		Total
	Aktivitas olahraga kurang (0)	Aktivitas olahraga cukup (1)	
Aktivitas olahraga kurang (0)	53	3	56
Aktivitas olahraga cukup (1)	12	28	40
Total	65	31	96

Berdasarkan Tabel 4.5 terlihat bahwa dari 56 pasien yang memiliki aktivitas olahraga kurang, terdapat 53 pasien yang tepat diklasifikasikan pada aktivitas olahraga yang kurang dan dari 40 pasien yang memiliki aktivitas olahraga cukup, terdapat 28 pasien yang telah tepat diklasifikasikan, sehingga dapat dilakukan perhitungan akurasi dari ketepatan model MARS sebagai berikut.

$$APER(\%) = \frac{12 + 3}{96} \times 100\% = 15,6\%$$

$$TAR(\%) = 1 - APER = 1 - 0,156 = 0,844 = 84,4\%$$

Nilai TAR sebesar 84,4% menunjukkan bahwa model MARS yang diperoleh telah mampu memprediksi dengan tepat sebesar 84,4%.

4.3.3 Estimasi Nilai *Propensity Score*

Estimasi nilai *propensity score* untuk setiap subjek dapat dihitung dengan menggunakan persamaan (2.27). Berdasarkan nilai prediksi MARS pada Tabel 4.4 diperoleh perhitungan nilai *propensity score* sebagai berikut.

Tabel 4.6 Estimasi Nilai *Propensity Score* Menggunakan MARS

Data ke-	$\hat{e}(\mathbf{x}_i)$							
1-8	0,552	0,571	0,440	0,596	0,519	0,562	0,570	0,631
9-16	0,643	0,540	0,654	0,658	0,474	0,522	0,395	0,630
17-24	0,734	0,594	0,467	0,522	0,612	0,628	0,765	0,571
25-32	0,678	0,597	0,552	0,598	0,746	0,683	0,622	0,525
33-40	0,537	0,503	0,609	0,526	0,510	0,732	0,653	0,738
41-48	0,643	0,472	0,665	0,705	0,573	0,488	0,588	0,696
49-56	0,605	0,674	0,567	0,574	0,678	0,706	0,624	0,506
57-64	0,508	0,678	0,529	0,769	0,696	0,739	0,586	0,582
65-72	0,650	0,536	0,770	0,710	0,588	0,572	0,737	0,693
73-80	0,488	0,569	0,612	0,445	0,560	0,568	0,591	0,484
81-88	0,752	0,514	0,588	0,698	0,580	0,515	0,488	0,609
89-96	0,765	0,587	0,560	0,534	0,651	0,525	0,573	0,503

Berdasarkan Tabel 4.6 diketahui bahwa nilai *propensity score* yang minimum adalah sebesar 0,395 dan nilai maksimum sebesar 0,770.

4.4 Stratifikasi dan Pengujian *Balance Kovariat*

Setelah nilai *propensity score* didapatkan, maka dilakukan pembagian subjek ke dalam beberapa strata. Pembagian ini bertujuan untuk mendapatkan kelompok strata yang tidak memiliki perbedaan antara kelompok *treatment* dan kelompok kontrol yang dikenal dengan *balance* (seimbang). Tahap awal yang dilakukan dalam stratifikasi adalah mengurutkan nilai *propensity score* dari yang terkecil hingga yang terbesar dan kemudian subjek dibagi ke dalam beberapa strata. Setelah dilakukan pembagian strata, maka dilakukan pengujian *balance kovariat* pada masing-masing strata.

Pengujian *balance kovariat* pada setiap strata dibedakan menjadi dua yaitu pengujian untuk variabel data kategori dan variabel data kontinu. Untuk variabel data kategori digunakan uji z , sedangkan pada variabel data kontinu digunakan uji t . Hasil

pembagian strata dan pengujian *balance* kovariat untuk setiap strata dapat dilihat pada Tabel 4.7 (lihat lampiran 5).

Tabel 4.7 Pembagian Strata dan Pengujian *Balance* Kovariat

Banyak Strata	Strata ke- k	n		<i>p-value</i>					
		ck	tk	X1	X2	X3	X4	X5	X6
Sebelum Strata		56	40	0,518	0,008	0,001	0,283	0,006	0,019
2	1	46	3	0,847	0,979	0,979	0,314	0,979	0,959
	2	10	37	0,279	0,654	0,032	0,732	0,987	0,209

Berdasarkan Tabel 4.7 diketahui bahwa sebelum dilakukan pembagian strata terdapat tiga kovariat yang memiliki nilai *p-value* $< \alpha$ (0,01) yaitu X_2 (jenis kelamin), X_3 (riwayat genetik), dan X_5 (obesitas). Ini berarti ketiga kovariat tersebut belum *balance* atau masih terdapat perbedaan proporsi antara kelompok *treatment* dan kontrol sebelum dilakukan stratifikasi. Pada pembentukan 2 strata terlihat bahwa nilai *p-value* untuk masing-masing kovariat sudah lebih dari 0,01 sehingga dapat dikatakan bahwa semua kovariat sudah *balance*. Jika dilanjutkan pembentukan untuk 3 strata maka terdapat salah satu kelompok *treatment* atau kontrol yang memiliki jumlah anggota nol. Sehingga tidak dapat dilakukan perbandingan antara kelompok *treatment* dan kontrol.

4.5 Estimasi *Average Treatment Effect* (ATE)

Estimasi ATE bertujuan untuk mengetahui pengaruh variabel *confounding* (aktivitas olahraga) terhadap variabel respon (komplikasi penyakit). Perhitungan nilai ATE hanya dilakukan pada strata sebanyak 2 karena semua kovariat sudah memenuhi asumsi *balance* dan jumlah anggota untuk kelompok *treatment* maupun kontrol tidak ada yang berjumlah nol. Nilai ATE diperoleh menggunakan persamaan (2.31). Adapun hasil estimasi nilai ATE adalah sebagai berikut.

Tabel 4.8 Hasil Estimasi ATE dan Standar Error PSS MARS

Banyak Strata	Strata ke-k	$\hat{\theta}_k$	$\hat{\theta}$	SE ($\hat{\theta}$)	df	$ t_{hit} $	<i>p-value</i>
2	1	0,377	0,645	0,089	92	7,276	0,000
	2	0,267					

Tabel 4.8 menunjukkan hasil estimasi ATE untuk strata sebanyak 2. Pada strata 1 diperoleh hasil estimasi ATE sebesar 0,377 yang berarti peluang pasien DM tipe 2 yang memiliki aktivitas olahraga cukup terkena komplikasi penyakit adalah sebesar 0,377 kali dibandingkan dengan pasien yang memiliki aktivitas olahraga kurang. Sedangkan pada strata 2 hasil estimasi ATE yang diperoleh adalah sebesar 0,267. Artinya, peluang pasien DM tipe 2 yang memiliki aktivitas olahraga cukup terkena komplikasi penyakit adalah sebesar 0,267 kali dibandingkan dengan pasien yang memiliki aktivitas olahraga kurang. Secara keseluruhan, pengaruh yang diberikan oleh 2 strata adalah sebesar 0.645 dengan standar error sebesar 0,089.

Berdasarkan Tabel 4.8 juga terlihat bahwa nilai *p-value* yang dihasilkan adalah 0,000 lebih kecil dari α (0,01). Hal ini berarti aktivitas olahraga merupakan variabel yang berpengaruh terhadap variabel komplikasi penyakit.

4.6 Perbandingan Bias Sebelum Strata dan Sesudah Strata

Kebaikan metode PSS dapat dilihat dari seberapa besar bias yang mampu direduksi oleh metode tersebut. Dengan menggunakan persamaan (2.35) diperoleh hasil reduksi bias sebagai berikut.

Tabel 4.9 Percent Bias Reduction PSS MARS

Bias	Nilai Bias
Sebelum Strata	0,131
Setelah Strata	0,114
PBR	13,232

Berdasarkan Tabel 4.9 diketahui bahwa nilai bias sebelum dilakukan stratifikasi adalah 0,131 dan nilai bias setelah dilakukan stratifikasi 2 strata adalah 0,114. Hal ini menunjukkan bahwa setelah dilakukan pembagian subjek ke dalam 2 strata bias yang dihasilkan makin berkurang. Secara keseluruhan bias yang mampu direduksi oleh metode PSS ini adalah sebesar 13,232%.

Halaman ini sengaja dikosongkan

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan diperoleh kesimpulan sebagai berikut.

1. Hasil pengujian *chi-square* menunjukkan bahwa variabel yang terpilih sebagai variabel *confounding* (variabel yang memiliki hubungan dengan variabel prediktor sekaligus variabel respon) adalah aktivitas olahraga.
2. Model MARS terbaik yang diperoleh adalah model dengan basis fungsi = 24, maksimum interaksi = 3, dan minimum observasi = 1.

$$\begin{aligned}\hat{z} = & 0,367 + 0,230BF1 - 0,387BF2 + 0,385BF3 - 0,014BF4 \\ & - 0,027BF5 + 0,146BF6 + 0,104BF7 - 0,26BF8 - 0,375BF9 \\ & - 0,145BF10 + 0,018BF11 + 0,336BF12 + 0,046BF13 \\ & + 0,041BF14 - 0,012BF15 - 0,007BF16\end{aligned}$$

Dari model terbaik MARS diketahui bahwa variabel usia, jenis kelamin, riwayat genetic, lama menderita, obesitas dan hipertensi memiliki hubungan dengan variabel aktivitas olahraga.

3. Hasil estimasi ATE menunjukkan bahwa variabel aktivitas olahraga memiliki pengaruh yang signifikan terhadap variabel komplikasi penyakit. Sementara itu, besarnya bias yang dapat direduksi menggunakan metode PSS MARS adalah kurang dari 25%..

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan diperoleh hasil bahwa aktivitas olahraga yang cukup dapat meminimalkan risiko pada pasien DM tipe 2 untuk terkena komplikasi penyakit. Oleh karena itu pasien DM tipe 2 dianjurkan untuk beraktivitas olahraga yang cukup agar dapat mengurangi risiko terjadinya komplikasi penyakit. Beberapa saran yang dapat diberikan untuk

peneliti selanjutnya diantaranya yaitu menambah jumlah data. Tujuannya agar pada saat pembagian strata, masing-masing strata memiliki anggota yang lebih dari satu baik untuk kelompok treatment maupun kelompok kontrol. Selain itu, penelitian ini hanya dilakukan untuk melihat hubungan antar variabel *confounding* dengan variabel respon, tetapi tidak sampai melakukan pemodelan. Sehingga peneliti selanjutnya dapat melakukan pemodelan lebih lanjut antara variabel *confounding* dengan variabel respon.

DAFTAR PUSTAKA

- Adityaningrum, A. (2017). *Estimasi Propensity Score Matching Menggunakan Pendekatan Multivariate Adaptive Regression Splines (MARS) pada Kasus Diabetes Melitus (DM)*. Surabaya: Tesis Institut Teknologi Sepuluh Nopember.
- Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). United State of America: Jhon Wiley and Sons, Inc.
- Akbar, M. B. (2017). *Propensity Score Stratification Menggunakan Regresi Logistik pada Kasus Infeksi Saluran Pernapasan Akut Terhadap Balita di Puskesmas Pasuruan Jawa Timur*. Surabaya: Tugas Akhir Institut Teknologi Sepuluh Nopember.
- Akolo, I. R. (2016). *Propensity Score Stratification Bootstrap Menggunakan Regresi Logistik Biner pada Kasus Diabetes Melitus (DM)*. Surabaya: Tesis Institut Teknologi Sepuluh Nopember.
- American College of Cardiology. (2017). *Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults*. America: American College of Cardiology.
- American Diabetes Association. (2014). Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, 37, 81-90.
- Anam, S., Sugiman, & Sunarmi. (2017). Ketepatan Klasifikasi dengan Menggunakan Metode Multivariate Adaptive Regression Splines (MARS) pada Data Kelompok Rumah Tangga Kabupaten Cilacap. *Unnes Journal of Mathematics*, 6(1), 92-101.
- Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46, 399-424.

- 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 Diabetes Mellitus in a Tertiary Care Setting. *Journal of Diabetes Investigation*, 5(6), 714-721.
- Cox, D. R., & Snell, E. J. (1989). *Analysis of Binary Data* (2nd ed.). London: Chapman & Hall/CRC.
- Dehejia, R. H., & Wahba, S. (1999). Causal Effect in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 94(448), 1053-1062.
- Diabetes UK. (2010). *Diabetes UK*. Diakses 1 Februari 2019, dari Diabetes in the UK 2010: Key Statistics on Diabetes: https://www.diabetes.org.uk/resources-s3/2017-11/diabetes_in_the_uk_2010.pdf
- Ernawati. (2018). *Propensity Score Stratification Menggunakan Support Vector Machine pada Kasus HIV/AIDS*. Surabaya: Tesis Institut Teknologi Sepuluh Nopember.
- Fahmiyah, I. (2016). *Faktor yang Mempengaruhi Kadar Glukosa Darah Puasa Pasien Diabetes Mellitus Tipe 2 di Poli Diabetes RSUD Dr. Soetomo Surabaya Menggunakan Regresi Probit Biner*. Surabaya: Tugas Akhir Institut Teknologi Sepuluh Nopember.
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1-67.
- Guo, S., & Fraser, M. W. (2010). *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks, CA: SAGE Publications.
- Hasanah, S. (2018). *Propensity Score Matching Menggunakan Support Vector Machine pada Kasus Diabetes Melitus (DM) Tipe 2*. Surabaya: Tesis Akhir Institut Teknologi Sepuluh Nopember.
- Holistic Health Solution. (2011). *Diabetes di Usia Muda*. Jakarta: Grasindo.

- International Diabetes Federation. (2017). *International Diabetes Federation*. Diakses 28 Januari 2019, dari <http://diabetesatlas.org/resources/2017-atlas.html>
- Isnaini, F. (2017). *Pemodelan Kasus Diabetes Melitus Tipe 2 di Klinik Assalam Kabupaten Banjarnegara Jawa Tengah dengan Metode Probit Biner*. Surabaya: Tugas Akhir Institut Teknologi Sepuluh Nopember.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. United States of America: Pearson Prentice Hall.
- Kemenkes. (2003). *Pedoman Praktis Memantau Status Gizi Orang Dewasa*. Jakarta: Kementerian Kesehatan RI.
- Kemenkes. (2013a). *Pusat Data Informasi Kementerian Kesehatan RI*. Jakarta: Kementerian Kesehatan RI.
- Kemenkes. (2013b). *Riset Kesehatan Dasar 2013*. Jakarta: Kementerian Kesehatan RI.
- Lanywati, E. (2001). *Diabetes Mellitus: Penyakit Kencing Manis*. Yogyakarta: Kanisus.
- 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, 148-157.
- Li, M. (2012). Using the Propensity Score Method to Estimate Causal Effects: A Review and Practical Guide. *Organizational Research Methods*, 1-39.
- Montgomery, D. C. (2001). *Design and Analysis of Experiments* (5th ed.). United States of America: John Wiley & Sons, Inc.
- Nash, M. S., & Bradford, D. F. (2001). *Parametric and Non Parametric Logistic Regression for Prediction of Presence/Absence of an Amphibian*. Las Vegas: Nevada.
- Nisa', S. F., & Budiantara, I. N. (2012). Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Splines pada Kasus Demam Berdarah Dengue (DBD). *Jurnal Sains dan Seni ITS*, 1(1), 318-323.

- Okur, M. E., Karantas, L. D., & Siafaka, P. I. (2017). Diabetes Mellitus: A Review on Pathophysiology, Current Status of Oral Medications and Future Perspectives. *Acta Pharmaceutica Scientia*, 55(1), 61-82.
- Otok, B. W., Guritno, S., Subanar, & Haryatmi, S. (2006). Bootstrap dalam MARS untuk Klasifikasi Perbankan. *Inferensi Jurnal Statistik*, 2(1).
- Pan, W., & Bai, H. (2015). *Propensity Score Analysis: Fundamentals and Developments*. New York: The Guilford Press.
- PERKENI. (2015). *Konsensus Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 di Indonesia*. Jakarta: Perkumpulan Endokrinologi Indonesia.
- Riaz, S. (2009). Diabetes Mellitus. *Scientific Research and Essay*, 4(5), 367-373.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of The Propensity Score in Observational Studies for Causal Effects. *Jurnal Biometrika*, 70(1), 41-55.
- Sa'diyah, W. U. (2017). *Boosting Multivariate Adaptive Regression Spline (MARS) pada Klasifikasi Diabetes Melitus Tipe II di Klinik As-Shafa, Sidoarjo*. Surabaya: Tugas Akhir Institut Teknologi Sepuluh Nopember.
- Sumarno. (2014). *Model Klasifikasi Menggunakan Metode Regresi Logistik dan Multivariate Adaptive Regression Splines (MARS)*. Surabaya: Tesis Institut Teknologi Sepuluh Nopember.
- Tandra, H. (2018). *Segala Sesuatu yang Harus Anda Ketahui Tentang Diabetes*. Jakarta: PT Gramedia.
- Tu, W., & Zhou, X.-H. (2003). A Bootstrap Confidence Interval Procedure for the Treatment Effect Using Propensity Score Subclassification. *UW Biostatistics Working Paper Series*, 3-20.
- Wahyuni, S. (2010). *Faktor-faktor yang Berhubungan dengan Penyakit Diabetes Melitus (DM) Daerah Perkotaan di*

- Indonesia Tahun 2007*. Jakarta: Skripsi Univesitas Islam Negeri Syarif Hidayatullah.
- WHO. (2016). *Global Report on Diabetes*. France: World Health Organizaion.
- Winkelmayer, W. C., & Kurth, T. (2004). Propensity Score: Help or Hype? *Nephrol Dial Transplant*, 19(7), 1671-1673.
- Yanovitzky, I., Zanutto, E., & Hornik, R. (2005). Estimating Causal Effects of Public Helath Education Campaigns Using Propensity Score Methodology. *Evaluation and Program Planning*, 28, 209-220.

Halaman ini sengaja dikosongkan

LAMPIRAN

Lampiran 1 Surat Izin Penggunaan Data

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini, mahasiswi Departemen Statistika FMKSD ITS.

Nama : Ane Wahyuni

NRP : 0621174500005

menyatakan bahwa data yang digunakan dalam Tugas Akhir ini merupakan data sekunder yang diambil dari Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Tahun 2017 oleh Dr. Bambang Widjanarko Otok, M.Si yang berjudul:

“Estimasi Parameter *Propensity Score Bootstrap Model Machine Learning* dan Terapannya pada Data Kesehatan yang Memuat *Confounding*”.

Surat pernyataan ini dibuat dengan sebenar-benarnya. Apabila ada pemalsuan data maka saya siap menerima sanksi sesuai dengan aturan yang berlaku.

Mengetahui,
Pembimbing Tugas Akhir



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

Surabaya, Juni 2019



Ane Wahyuni
NRP.0621174500005

Lampiran 2 Data Diabetes Melitus Tipe 2

No	Y	X1	X2	X3	X4	X5	X6	X7
1	0	47	0	0	3	0	0	0
2	1	43	1	1	1	0	0	0
3	1	70	0	1	2	1	1	0
4	1	57	1	1	2	1	1	0
5	1	51	1	1	2	1	1	0
6	0	57	0	1	5	1	1	0
7	0	77	0	1	3	1	0	0
8	0	58	1	1	4	0	1	1
9	0	72	1	1	2	1	0	1
10	0	34	1	1	1	0	0	0
11	1	72	1	0	7	1	0	1
12	1	40	1	0	6	1	0	1
13	0	31	0	1	2	0	1	0
14	0	51	0	1	6	0	1	0
15	0	39	0	1	7	0	1	0
16	1	39	1	0	5	0	0	1
17	1	58	1	0	7	0	0	1
18	1	58	1	1	3	1	1	0
19	0	66	0	1	4	0	0	0
20	0	45	0	1	1	0	0	0
21	0	34	0	1	1	1	0	0
22	0	72	0	1	6	1	0	0
23	1	58	1	0	7	1	1	1
24	1	70	1	0	5	1	1	0
25	1	64	1	1	7	1	0	1
26	1	35	0	1	4	0	1	0
27	1	31	1	1	7	1	1	0
28	1	43	1	1	1	1	0	1
29	1	48	1	0	3	1	1	1
30	1	75	0	0	4	1	1	1
31	1	57	1	1	6	1	0	1
32	1	43	1	1	4	0	0	0
33	0	68	1	1	5	0	0	0
34	0	31	0	1	7	0	0	0

Lampiran 2 Data Diabetes Melitus Tipe 2 (Lanjutan)

No	Y	X1	X2	X3	X4	X5	X6	X7
35	1	39	1	1	5	1	0	0
36	0	64	0	1	6	0	1	0
37	0	35	0	1	3	0	1	0
38	1	29	1	1	3	1	1	1
39	1	31	1	1	2	1	0	1
40	1	45	1	0	2	1	1	1
41	1	66	1	0	5	1	0	1
42	0	31	0	1	1	0	0	0
43	1	43	0	0	6	0	1	1
44	1	35	0	1	4	1	1	1
45	1	43	0	1	7	1	0	1
46	1	51	1	1	4	1	0	0
47	1	58	1	1	2	0	0	1
48	1	34	0	0	5	0	1	1
49	1	72	1	0	5	0	0	1
50	1	51	0	0	7	0	1	1
51	0	64	1	1	4	0	1	0
52	0	48	1	1	1	0	1	0
53	0	68	1	0	7	1	0	0
54	1	57	0	0	3	1	1	1
55	1	72	1	0	4	1	1	1
56	0	64	0	1	5	0	0	0
57	0	47	0	1	5	0	0	0
58	1	48	1	0	1	0	0	1
59	0	47	0	1	1	0	0	0
60	1	51	1	0	5	1	1	1
61	1	48	0	0	1	1	1	1
62	1	75	1	0	2	1	1	1
63	0	40	1	1	2	1	0	0
64	1	68	0	0	4	1	1	0
65	0	40	1	1	7	1	0	0
66	0	72	1	1	6	0	0	0
67	1	75	1	0	6	1	1	1
68	1	64	0	0	7	1	1	1

Lampiran 2 Data Diabetes Melitus Tipe 2 (Lanjutan)

No	Y	X1	X2	X3	X4	X5	X6	X7
69	1	70	0	0	6	0	0	1
70	0	47	1	1	6	1	0	0
71	1	35	1	0	3	0	1	1
72	1	39	0	0	3	0	1	1
73	1	70	0	0	4	1	1	0
74	0	57	0	1	6	0	0	0
75	1	34	0	1	1	1	0	1
76	0	77	0	1	6	0	0	0
77	1	75	1	0	4	0	0	0
78	0	58	0	1	1	0	0	0
79	1	66	0	1	5	1	1	1
80	0	39	1	1	3	0	0	0
81	1	66	1	0	7	1	1	1
82	0	29	0	0	4	1	0	0
83	0	70	0	0	6	0	0	0
84	1	45	0	0	3	1	1	1
85	1	51	1	1	3	0	1	1
86	0	77	0	0	5	0	0	0
87	1	29	0	0	3	0	0	0
88	0	40	1	0	4	0	0	0
89	1	75	1	1	6	1	1	1
90	0	57	0	0	3	1	0	0
91	0	68	0	0	2	0	0	0
92	1	70	0	0	1	1	1	0
93	0	40	0	0	6	0	1	0
94	0	29	0	0	2	1	0	0
95	0	66	0	0	2	0	0	0
96	0	29	0	0	7	0	1	0

Lampiran 3 Tabulasi Silang Antar Kovariat

a. Tabulasi Silang Antara Kovariat X dengan Jenis Kelamin (X_2)

Jenis Kelamin (X_2)	Riwayat Genetik (X_3)		Total
	Punya	Tidak Punya	
Perempuan	24	25	49
Lak-laki	20	27	47
Total	44	52	96

Jenis Kelamin (X_2)	Obesitas (X_5)		Total
	Ya	Tidak	
Perempuan	29	20	49
Lak-laki	18	29	47
Total	47	49	96

Jenis Kelamin (X_2)	Hipertensi (X_6)		Total
	Ya	Tidak	
Perempuan	25	24	49
Lak-laki	27	20	47
Total	52	44	96

Jenis Kelamin (X_2)	Aktivitas Olahraga (X_7)		Total
	Kurang	Cukup	
Perempuan	35	14	49
Lak-laki	21	26	47
Total	56	40	96

Lampiran 3 Tabulasi Silang Antar Kovariat (Lanjutan)*b.* Tabulasi Silang Antara Kovariat X dengan Riwayat Genetik (X₃)

Riwayat Genetik (X ₃)	Jenis Kelamin (X ₂)		Total
	Perempuan	Laki-laki	
Punya	24	20	44
Tidak Punya	25	27	52
Total	49	47	96

Riwayat Genetik (X ₃)	Obesitas (X ₅)		Total
	Ya	Tidak	
Punya	20	24	44
Tidak Punya	27	25	52
Total	47	49	96

Riwayat Genetik (X ₃)	Hipertensi (X ₆)		Total
	Ya	Tidak	
Punya	20	24	44
Tidak Punya	32	20	52
Total	52	44	96

Riwayat Genetik (X ₃)	Aktivitas Olahraga (X ₇)		Total
	Kurang	Cukup	
Punya	18	26	44
Tidak Punya	38	14	52
Total	56	40	96

Lampiran 3 Tabulasi Silang Antar Kovariat (Lanjutan)

c. Tabulasi Silang Antara Kovariat X dengan Obesitas (X_5)

Obesitas (X_5)	Jenis Kelamin (X_2)		Total
	Perempuan	Laki-laki	
Ya	29	18	47
Tidak	20	29	49
Total	49	47	96

Obesitas (X_5)	Riwayat Genetik (X_3)		Total
	Punya	Tidak Punya	
Ya	20	27	47
Tidak	24	25	49
Total	44	52	96

Obesitas (X_5)	Hipertensi (X_6)		Total
	Ya	Tidak	
Ya	30	17	47
Tidak	22	27	49
Total	52	44	96

Obesitas (X_5)	Aktivitas Olahraga (X_7)		Total
	Kurang	Cukup	
Ya	34	13	47
Tidak	22	27	49
Total	56	40	96

Lampiran 3 Tabulasi Silang Antar Kovariat (Lanjutan)d. Tabulasi Silang Antara Kovariat X dengan Hipertensi (X_6)

Hipertensi (X_6)	Jenis Kelamin (X_2)		Total
	Perempuan	Laki-laki	
Ya	25	27	52
Tidak	24	20	44
Total	49	47	96

Hipertensi (X_6)	Riwayat Genetik (X_3)		Total
	Punya	Tidak Punya	
Ya	20	32	52
Tidak	24	20	44
Total	44	52	96

Hipertensi (X_6)	Obesitas (X_5)		Total
	Ya	Tidak	
Ya	30	22	52
Tidak	17	27	44
Total	47	49	96

Hipertensi (X_6)	Aktivitas Olahraga (X_7)		Total
	Kurang	Cukup	
Ya	36	16	52
Tidak	20	24	44
Total	56	40	96

Lampiran 3 Tabulasi Silang Antar Kovariat (Lanjutan)

- e. Tabulasi Silang Antara Kovariat X dengan Aktivitas Olahraga (X₇)

Aktivitas Olahraga (X ₇)	Jenis Kelamin (X ₂)		Total
	Perempuan	Laki-laki	
Kurang	35	21	56
Cukup	14	26	40
Total	49	47	96

Aktivitas Olahraga (X ₇)	Riwayat Genetik (X ₃)		Total
	Punya	Tidak Punya	
Kurang	18	38	56
Cukup	26	14	40
Total	44	52	96

Aktivitas Olahraga (X ₇)	Obesitas (X ₅)		Total
	Ya	Tidak	
Kurang	34	22	56
Cukup	13	27	40
Total	47	49	96

Aktivitas Olahraga (X ₇)	Hipertensi (X ₆)		Total
	Ya	Tidak	
Kurang	36	20	56
Cukup	16	24	40
Total	52	44	96

Lampiran 4 Syntax dan Output Chi-square

a. Syntax Chi-square

```

data = read.csv("E:/Data DM.csv",header = TRUE)

Y = data$Y
X1 = data$X1
X2 = data$X2
X3 = data$X3
X4 = data$X4
X5 = data$X5
X6 = data$X6
X7 = data$X7

#Chi-square X2
tableX3X2=table(X3,X2)
tableX5X2=table(X5,X2)
tableX6X2=table(X6,X2)
tableX7X2=table(X7,X2)
ujiX3X2=chisq.test(tableX3X2,correct=FALSE)
ujiX5X2=chisq.test(tableX5X2,correct=FALSE)
ujiX6X2=chisq.test(tableX6X2,correct=FALSE)
ujiX7X2=chisq.test(tableX7X2,correct=FALSE)

variabelX2=rbind("X3*X2","X5*X2","X6*X2","X7*X2")
chisquareX2=rbind(ujiX3X2$statistic,ujiX5X2$statistic,
ujiX6X2$statistic,ujiX7X2$statistic)
pvalueX2=rbind(ujiX3X2$p.value,ujiX5X2$p.value,ujiX6X2
$p.value,ujiX7X2$p.value)
confoundX2=data.frame(variabelX2,chisquareX2,pvalueX2)

#Chi-square X3
tableX2X3=table(X2,X3)
tableX5X3=table(X5,X3)
tableX6X3=table(X6,X3)
tableX7X3=table(X7,X3)
ujiX2X3=chisq.test(tableX2X3,correct=FALSE)
ujiX5X3=chisq.test(tableX5X3,correct=FALSE)
ujiX6X3=chisq.test(tableX6X3,correct=FALSE)
ujiX7X3=chisq.test(tableX7X3,correct=FALSE)
variabelX3=rbind("X2*X3","X5*X3","X6*X3","X7*X3")
chisquareX3=rbind(ujiX2X3$statistic,ujiX5X3$statistic,
ujiX6X3$statistic, ujiX7X3$statistic)

```

Lampiran 4 Syntax dan Output Chi-square (Lanjutan)

```
pvalueX3=rbind(ujiX2X3$p.value, ujiX5X3$p.value, ujiX6X3
$p.value, ujiX7X3$p.value)
confoundX3=data.frame(variabelX3, chisquareX3, pvalueX3)

#Chi-square X5
tableX2X5=table(X2, X5)
tableX3X5=table(X3, X5)
tableX6X5=table(X6, X5)
tableX7X5=table(X7, X5)
ujiX2X5=chisq.test(tableX2X5, correct=FALSE)
ujiX3X5=chisq.test(tableX3X5, correct=FALSE)
ujiX6X5=chisq.test(tableX6X5, correct=FALSE)
ujiX7X5=chisq.test(tableX7X5, correct=FALSE)

variabelX5=rbind("X2*X5", "X3*X5", "X6*X5", "X7*X5")
chisquareX5=rbind(ujiX2X5$statistic, ujiX3X5$statistic,
ujiX6X5$statistic, ujiX7X5$statistic)
pvalueX5=rbind(ujiX2X5$p.value, ujiX3X5$p.value, ujiX6X5
$p.value, ujiX7X5$p.value)
confoundX5=data.frame(variabelX5, chisquareX5, pvalueX5)

#Chi-square X6
tableX2X6=table(X2, X6)
tableX3X6=table(X3, X6)
tableX5X6=table(X5, X6)
tableX7X6=table(X7, X6)
ujiX2X6=chisq.test(tableX2X6, correct=FALSE)
ujiX3X6=chisq.test(tableX3X6, correct=FALSE)
ujiX5X6=chisq.test(tableX5X6, correct=FALSE)
ujiX7X6=chisq.test(tableX7X6, correct=FALSE)

variabelX6=rbind("X2*X6", "X3*X6", "X5*X6", "X7*X6")
chisquareX6=rbind(ujiX2X6$statistic, ujiX3X6$statistic,
ujiX5X6$statistic, ujiX7X6$statistic)
pvalueX6=rbind(ujiX2X6$p.value, ujiX3X6$p.value, ujiX5X6
$p.value, ujiX7X6$p.value)
confoundX6=data.frame(variabelX6, chisquareX6, pvalueX6)

#Chi-square X7
tableX2X7=table(X2, X7)
tableX3X7=table(X3, X7)
```

Lampiran 4 Syntax dan Output Chi-square (Lanjutan)

```

tableX5X7=table(X5,X7)
tableX6X7=table(X6,X7)
ujiX2X7=chisq.test(tableX2X7,correct=FALSE)
ujiX3X7=chisq.test(tableX3X7,correct=FALSE)
ujiX5X7=chisq.test(tableX5X7,correct=FALSE)
ujiX6X7=chisq.test(tableX6X7,correct=FALSE)

variabelX7=rbind("X2*X7","X3*X7","X5*X7","X6*X7")
chisquareX7=rbind(ujiX2X7$statistic,ujiX3X7$statistic,
ujiX5X7$statistic,ujiX6X7$statistic)
pvalueX7=rbind(ujiX2X7$p.value,ujiX3X7$p.value,ujiX5X7
$p.value, ujiX6X7$p.value)
confoundX7=data.frame(variabelX7,chisquareX7,pvalueX7)

#Chi-square Y
tableX2Y=table(X2,Y)
tableX3Y=table(X3,Y)
tableX5Y=table(X5,Y)
tableX6Y=table(X6,Y)
tableX7Y=table(X7,Y)
ujiX2Y=chisq.test(tableX2Y,correct=FALSE)
ujiX3Y=chisq.test(tableX3Y,correct=FALSE)
ujiX5Y=chisq.test(tableX5Y,correct=FALSE)
ujiX6Y=chisq.test(tableX6Y,correct=FALSE)
ujiX7Y=chisq.test(tableX7Y,correct=FALSE)

variabelY=rbind("X2*Y","X3*Y","X5*Y","X6*Y","X7*Y")
chisquareY=rbind(ujiX2Y$statistic,ujiX3Y$statistic,uji
X5Y$statistic,ujiX6Y$statistic,ujiX7Y$statistic)
pvalueY=rbind(ujiX2Y$p.value,ujiX3Y$p.value,ujiX5Y$p.v
alue, ujiX6Y$p.value, ujiX7Y$p.value)
confoundY=data.frame(variabelY,chisquareY,pvalueY)

confoundX2
confoundX3
confoundX5
confoundX6
confoundX7
confoundY

```


Lampiran 4 Syntax dan Output Chi-square (Lanjutan)

b. Output Chi-square

```

> confoundX2
  variabelX2 X.squared   pvalueX2
1      X3*X2 0.399066 0.52757202
2      X5*X2 4.187680 0.04071878
3      X6*X2 0.399066 0.52757202
4      X7*X2 7.061398 0.00787625
> confoundX3
  variabelX3 X.squared   pvalueX3
1      X2*X3 0.399066 0.527572018
2      X5*X3 0.399066 0.527572018
3      X6*X3 2.483447 0.115049888
4      X7*X3 10.146653 0.001445634
> confoundX5
  variabelX5 X.squared   pvalueX5
1      X2*X5 4.187680 0.040718779
2      X3*X5 0.399066 0.527572018
3      X6*X5 3.463333 0.062743721
4      X7*X5 7.432988 0.006403892
> confoundX6
  variabelX6 X.squared   pvalueX6
1      X2*X6 0.399066 0.52757202
2      X3*X6 2.483447 0.11504989
3      X5*X6 3.463333 0.06274372
4      X7*X6 5.543257 0.01855203
> confoundX7
  variabelX7 X.squared   pvalueX7
1      X2*X7 7.061398 0.007876250
2      X3*X7 10.146653 0.001445634
3      X5*X7 7.432988 0.006403892
4      X6*X7 5.543257 0.018552027
> confoundY
  variabelY X.squared   pvalueY
1      X2*Y 9.687406 1.855353e-03
2      X3*Y 8.961705 2.756975e-03
3      X5*Y 15.086574 1.026910e-04
4      X6*Y 11.604396 6.579612e-04
5      X7*Y 41.839456 9.908366e-11

```

Lampiran 5 Syntax dan Output MARS

a. Syntax trial and error Kombinasi BF,MI, dan MO

```

library(earth)
data = read.csv("E:/Data DM.csv")
Y = data$Y
X1 = data$X1
X2 = data$X2
X3 = data$X3
X4 = data$X4
X5 = data$X5
X6 = data$X6
Z = data$X7

mars12.1.0=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=1,
minspan=0, endspan=0, penalty=-1, data)
mars12.1.1=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=1,
minspan=1, endspan=1, penalty=-1, data)
mars12.1.2=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=1,
minspan=2, endspan=2, penalty=-1, data)
mars12.1.3=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=1,
minspan=3, endspan=3, penalty=-1, data)
mars12.1.5=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=1,
minspan=5, endspan=5, penalty=-1, data)
mars12.1.10=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=1,
minspan=10, endspan=10, penalty=-1, data)

mars12.2.0=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=2,
minspan=0, endspan=0, penalty=-1, data)
mars12.2.1=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=2,
minspan=1, endspan=1, penalty=-1, data)
mars12.2.2=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=2,
minspan=2, endspan=2, penalty=-1, data)
mars12.2.3=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=2,
minspan=3, endspan=3, penalty=-1, data)
mars12.2.5=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=2,
minspan=5, endspan=5, penalty=-1, data)
mars12.2.10=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=2,
minspan=10, endspan=10, penalty=-1, data)

mars12.3.0=earth(Z~X1+X2+X3+X4+X5+X6, nk=12, degree=3,
minspan=0, endspan=0, penalty=-1, data)

```

Lampiran 5 Syntax dan Output MARS (Lanjutan)

```

mars12.3.1=earth (Z~X1+X2+X3+X4+X5+X6, nk=12, degree=3,
minspan=1, endspan=1, penalty=-1, data)
mars12.3.2=earth (Z~X1+X2+X3+X4+X5+X6, nk=12, degree=3,
minspan=2, endspan=2, penalty=-1, data)
mars12.3.3=earth (Z~X1+X2+X3+X4+X5+X6, nk=12, degree=3,
minspan=3, endspan=3, penalty=-1, data)
mars12.3.5=earth (Z~X1+X2+X3+X4+X5+X6, nk=12, degree=3,
minspan=5, endspan=5, penalty=-1, data)
mars12.3.10=earth (Z~X1+X2+X3+X4+X5+X6, nk=12, degree=3,
minspan=10, endspan=10, penalty=-1, data)

mars18.1.0=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=1,
minspan=0, endspan=0, penalty=-1, data)
mars18.1.1=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=1,
minspan=1, endspan=1, penalty=-1, data)
mars18.1.2=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=1,
minspan=2, endspan=2, penalty=-1, data)
mars18.1.3=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=1,
minspan=3, endspan=3, penalty=-1, data)
mars18.1.5=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=1,
minspan=5, endspan=5, penalty=-1, data)
mars18.1.10=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=1,
minspan=10, endspan=10, penalty=-1, data)

mars18.2.0=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=2,
minspan=0, endspan=0, penalty=-1, data)
mars18.2.1=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=2,
minspan=1, endspan=1, penalty=-1, data)
mars18.2.2=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=2,
minspan=2, endspan=2, penalty=-1, data)
mars18.2.3=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=2,
minspan=3, endspan=3, penalty=-1, data)
mars18.2.5=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=2,
minspan=5, endspan=5, penalty=-1, data)
mars18.2.10=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=2,
minspan=10, endspan=10, penalty=-1, data)

mars18.3.0=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=3,
minspan=0, endspan=0, penalty=-1, data)
mars18.3.1=earth (Z~X1+X2+X3+X4+X5+X6, nk=18, degree=3,
minspan=1, endspan=1, penalty=-1, data)

```

Lampiran 5 Syntax dan Output MARS (Lanjutan)

```

mars18.3.2=earth(Z~X1+X2+X3+X4+X5+X6,nk=18,degree=3,
minspan=2,endspan=2,penalty=-1,data)
mars18.3.3=earth(Z~X1+X2+X3+X4+X5+X6,nk=18,degree=3,
minspan=3,endspan=3,penalty=-1,data)
mars18.3.5=earth(Z~X1+X2+X3+X4+X5+X6,nk=18,degree=3,
minspan=5,endspan=5,penalty=-1,data)
mars18.3.10=earth(Z~X1+X2+X3+X4+X5+X6,nk=18,degree=3,
minspan=10,endspan=10,penalty=-1,data)

mars24.1.0=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=1,
minspan=0,endspan=0,penalty=-1,data)
mars24.1.1=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=1,
minspan=1,endspan=1,penalty=-1,data)
mars24.1.2=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=1,
minspan=2,endspan=2,penalty=-1,data)
mars24.1.3=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=1,
minspan=3,endspan=3,penalty=-1,data)
mars24.1.5=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=1,
minspan=5,endspan=5,penalty=-1,data)
mars24.1.10=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=1,
minspan=10,endspan=10,penalty=-1,data)

mars24.2.0=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=2,
minspan=0,endspan=0,penalty=-1,data)
mars24.2.1=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=2,
minspan=1,endspan=1,penalty=-1,data)
mars24.2.2=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=2,
minspan=2,endspan=2,penalty=-1,data)
mars24.2.3=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=2,
minspan=3,endspan=3,penalty=-1,data)
mars24.2.5=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=2,
minspan=5,endspan=5,penalty=-1,data)
mars24.2.10=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=2,
minspan=10,endspan=10,penalty=-1,data)

mars24.3.0=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=3,
minspan=0,endspan=0,penalty=-1,data)
mars24.3.1=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=3,
minspan=1,endspan=1,penalty=-1,data)
mars24.3.2=earth(Z~X1+X2+X3+X4+X5+X6,nk=24,degree=3,
minspan=2,endspan=2,penalty=-1,data)

```

Lampiran 5 Syntax dan Output MARS (Lanjutan)

```

mars24.3.3=earth (Z~X1+X2+X3+X4+X5+X6, nk=24, degree=3,
minspan=3, endspan=3, penalty=-1, data)
mars24.3.5=earth (Z~X1+X2+X3+X4+X5+X6, nk=24, degree=3,
minspan=5, endspan=5, penalty=-1, data)
mars24.3.10=earth (Z~X1+X2+X3+X4+X5+X6, nk=24, degree=3,
minspan=10, endspan=10, penalty=-1, data)

nk =c(mars12.1.0$nk, mars12.1.1$nk, mars12.1.2$nk,
mars12.1.3$nk, mars12.1.5$nk, mars12.1.10$nk,
mars12.2.0$nk, mars12.2.1$nk, mars12.2.2$nk,
mars12.2.3$nk, mars12.2.5$nk, mars12.2.10$nk,
mars12.3.0$nk, mars12.3.1$nk, mars12.3.2$nk,
mars12.3.3$nk, mars12.3.5$nk, mars12.3.10$nk,
mars18.1.0$nk, mars18.1.1$nk, mars18.1.2$nk,
mars18.1.3$nk, mars18.1.5$nk, mars18.1.10$nk,
mars18.2.0$nk, mars18.2.1$nk, mars18.2.2$nk,
mars18.2.3$nk, mars18.2.5$nk, mars18.2.10$nk,
mars18.3.0$nk, mars18.3.1$nk, mars18.3.2$nk,
mars18.3.3$nk, mars18.3.5$nk, mars18.3.10$nk,
mars24.1.0$nk, mars24.1.1$nk, mars24.1.2$nk,
mars24.1.3$nk, mars24.1.5$nk, mars24.1.10$nk,
mars24.2.0$nk, mars24.2.1$nk, mars24.2.2$nk,
mars24.2.3$nk, mars24.2.5$nk, mars24.2.10$nk,
mars24.3.0$nk, mars24.3.1$nk, mars24.3.2$nk,
mars24.3.3$nk, mars24.3.5$nk, mars24.2.10$nk)

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)

minspan = c(0, 1, 2, 3, 5, 10, 0, 1, 2, 3, 5, 10, 0, 1, 2, 3, 5, 10,
0, 1, 2, 3, 5, 10, 0, 1, 2, 3, 5, 10, 0, 1, 2, 3, 5, 10,
0, 1, 2, 3, 5, 10, 0, 1, 2, 3, 5, 10, 0, 1, 2, 3, 5, 10)

GCV = c(mars12.1.0$gcv, mars12.1.1$gcv, mars12.1.2$gcv,
mars12.1.3$gcv, mars12.1.5$gcv, mars12.1.10$gcv,
mars12.2.0$gcv, mars12.2.1$gcv, mars12.2.2$gcv,
mars12.2.3$gcv, mars12.2.5$gcv, mars12.2.10$gcv,
mars12.3.0$gcv, mars12.3.1$gcv, mars12.3.2$gcv,
mars12.3.3$gcv, mars12.3.5$gcv, mars12.3.10$gcv,
mars18.1.0$gcv, mars18.1.1$gcv, mars18.1.2$gcv,
mars18.1.3$gcv, mars18.1.5$gcv, mars18.1.10$gcv,

```

Lampiran 5 Syntax dan Output MARS (Lanjutan)

```

mars18.2.0$gcv,mars18.2.1$gcv,mars18.2.2$gcv,
mars18.2.3$gcv,mars18.2.5$gcv,mars18.2.10$gcv,
mars18.1.0$gcv,mars18.3.1$gcv,mars18.3.2$gcv,
mars18.3.3$gcv,mars18.3.5$gcv,mars18.3.10$gcv,
mars24.1.0$gcv,mars24.1.1$gcv,mars24.1.2$gcv,
mars24.1.3$gcv,mars24.1.5$gcv,mars24.1.10$gcv,
mars24.2.0$gcv,mars24.2.1$gcv,mars24.2.2$gcv,
mars24.2.3$gcv,mars24.2.5$gcv,mars24.2.10$gcv,
mars24.3.0$gcv,mars24.3.1$gcv,mars24.3.2$gcv,
mars24.3.3$gcv,mars24.3.5$gcv,mars24.3.10$gcv)

RSq = c(mars12.1.0$rsq, mars12.1.1$rsq,
mars12.1.2$rsq,mars12.1.3$rsq,mars12.1.5$rsq,
mars12.1.10$rsq,mars12.2.0$rsq,mars12.2.1$rsq,
mars12.2.2$rsq,mars12.2.3$rsq,mars12.2.5$rsq,
mars12.2.10$rsq,mars12.3.0$rsq,mars12.3.1$rsq,
mars12.3.2$rsq,mars12.3.3$rsq,mars12.3.5$rsq,
mars12.3.10$rsq,mars18.1.0$rsq,mars18.1.1$rsq,
mars18.1.2$rsq,mars18.1.3$rsq,mars18.1.5$rsq,
mars18.1.10$rsq,mars18.2.0$rsq,mars18.2.1$rsq,
mars18.2.2$rsq,mars18.2.3$rsq,mars18.2.5$rsq,
mars18.2.10$rsq,mars18.3.0$rsq,mars18.3.1$rsq,
mars18.3.2$rsq,mars18.3.3$rsq,mars18.3.5$rsq,
mars18.3.10$rsq,mars24.1.0$rsq,mars24.1.1$rsq,
mars24.1.2$rsq,mars24.1.3$rsq,mars24.1.5$rsq,
mars24.1.10$rsq,mars24.2.0$rsq,mars24.2.1$rsq,
mars24.2.2$rsq,mars24.2.3$rsq,mars24.2.5$rsq,
mars24.2.10$rsq,mars24.3.0$rsq,mars24.3.1$rsq,
mars24.3.2$rsq,mars24.3.3$rsq,mars24.3.5$rsq,
mars24.3.10$rsq)

#trial and error kombinasi BF,MI,MO
hasil=data.frame(nk, degree, minspan, GCV, RSq)
hasil

#model terbaik MARS
model=summary(mars24.3.1)
model

```

Lampiran 5 Syntax dan Output MARS (Lanjutan)

b. Output trial and error Kombinasi BF, MI, dan MO

```
> hasil
```

	nk	degree	minspan	GCV	RSq
1	12	1	0	0.1763168	0.2745822
2	12	1	1	0.1763168	0.2745822
3	12	1	2	0.1763168	0.2745822
4	12	1	3	0.1763168	0.2745822
5	12	1	5	0.1763168	0.2745822
6	12	1	10	0.1763168	0.2745822
7	12	2	0	0.1709599	0.2966220
8	12	2	1	0.1709599	0.2966220
9	12	2	2	0.1709599	0.2966220
10	12	2	3	0.1709599	0.2966220
11	12	2	5	0.1709599	0.2966220
12	12	2	10	0.1709599	0.2966220
13	12	3	0	0.1709599	0.2966220
14	12	3	1	0.1709599	0.2966220
15	12	3	2	0.1709599	0.2966220
16	12	3	3	0.1709599	0.2966220
17	12	3	5	0.1709599	0.2966220
18	12	3	10	0.1709599	0.2966220
19	18	1	0	0.1664197	0.3153019
20	18	1	1	0.1664197	0.3153019
21	18	1	2	0.1664197	0.3153019
22	18	1	3	0.1664197	0.3153019
23	18	1	5	0.1664197	0.3153019
24	18	1	10	0.1684083	0.3071203
25	18	2	0	0.1594456	0.3439952
26	18	2	1	0.1427936	0.4125061
27	18	2	2	0.1508939	0.3791796
28	18	2	3	0.1508939	0.3791796
29	18	2	5	0.1508939	0.3791796
30	18	2	10	0.1611010	0.3371846
31	18	3	0	0.1664197	0.3439952
32	18	3	1	0.1381744	0.4315109
33	18	3	2	0.1446460	0.4048852
34	18	3	3	0.1449210	0.4037537
35	18	3	5	0.1482107	0.3902188
36	18	3	10	0.1606823	0.3389073
37	24	1	0	0.1444074	0.4058667
38	24	1	1	0.1444074	0.4058667

Lampiran 5 Syntax dan Output MARS (Lanjutan)

39	24	1	2	0.1444074	0.4058667
40	24	1	3	0.1444074	0.4058667
41	24	1	5	0.1444074	0.4058667
42	24	1	10	0.1573526	0.3526064
43	24	2	0	0.1491229	0.3864659
44	24	2	1	0.1260286	0.4814823
45	24	2	2	0.1274802	0.4755101
46	24	2	3	0.1300001	0.4651424
47	24	2	5	0.1306184	0.4625986
48	24	2	10	0.1534555	0.3686401
49	24	3	0	0.1477702	0.3920312
50	24	3	1	0.1056675	0.5652537
51	24	3	2	0.1203723	0.5047538
52	24	3	3	0.1216436	0.4995236
53	24	3	5	0.1277733	0.4743041
54	24	3	10	0.1504985	0.3808062

c. Output Model Terbaik MARS

```

> model
Call:earth(formula=Z~X1+X2+X3+X4+X5+X6,
            data=data,degree=3,nk=24,minspan=1,
            endspan=1,penalty=-1)

            coefficients
(Intercept)      0.36668200
X2                0.22951756
X3               -0.38700167
X6                0.38458948
h(58-X1)         -0.01427744
h(X1-58)         -0.02709154
h(3-X4)          0.14607698
h(X4-3)          0.10442660
X3 * X5          -0.26018770
X3 * X6          -0.37540353
h(X1-66) * X6   -0.14504092
h(70-X1) * X6   0.01789179
h(X1-70) * X6   0.33587345
h(51-X1) * X3 * X5 0.04623988
h(X1-51) * X3 * X5 0.04146442
h(70-X1) * h(X4-4) * X6 -0.01200116
h(70-X1) * h(4-X4) * X6 -0.00704308

```


Lampiran 5 Syntax dan Output MARS (Lanjutan)

```
Selected 17 of 17 terms, and 6 of 6 predictors
Termination condition: Reached nk 24
Importance: X3, X2, X6, X1, X5, X4
Number of terms at each degree of interaction: 1 7 5 4
GCV 0.1056675      RSS 10.14408      GRSq 0.5652537      RSq
0.5652537
```

d. Syntax Prediksi Model MARS dan Ketepatan Klasifikasi

```
#prediksi model MARS
pred_Z=predict(mars24.3.1)
colnames(pred_Z)=c("Z Pred")
pred_Z_c=predict(mars24.3.1,type="class",thresh=0.58)
colnames(pred_Z_c)=c("Z Pred Kls")
prediksi=cbind(pred_Z,pred_Z_c)
round(prediksi,digits=4)

#ketepatan klasifikasi
tabel_akurasi=table(Z,pred_Z_c)
aper=sum(Z=="1"&pred_Z_c=="0",Z=="0"&pred_Z_c=="1")/
sum(tabel_akurasi)
tar=1-aper

tabel_akurasi
aper
tar
```

e. Output Prediksi Model MARS dan Ketepatan Klasifikasi

```
> round(prediksi,digits=4)
      Z Pred  Z Pred Kls      Z Pred  Z Pred Kls
[1,]  0.2096      0      [2,]  0.2872      0
[3,] -0.2427      0      [4,]  0.3883      0
[5,]  0.0766      0      [6,]  0.2486      0
[7,]  0.2828      0      [8,]  0.5375      0
[9,]  0.5866      1     [10,]  0.1587      0
[11,]  0.6346      1     [12,]  0.6525      1
[13,] -0.1021      0     [14,]  0.0861      0
[15,] -0.4262      0     [16,]  0.5338      0
[17,]  1.0139      1     [18,]  0.3786      0
[19,] -0.1326      0     [20,]  0.0862      0
[21,]  0.4551      0     [22,]  0.5242      0
```

Lampiran 5 Syntax dan Output MARS (Lanjutan)

[23,]	1.1812	0	[24,]	0.2844	0
[25,]	0.7432	1	[26,]	0.3911	0
[27,]	0.2089	0	[28,]	0.3969	0
[29,]	1.0767	1	[30,]	0.7691	1
[31,]	0.4968	0	[32,]	0.0995	0
[33,]	0.1471	0	[34,]	0.0119	0
[35,]	0.4415	0	[36,]	0.1029	0
[37,]	0.0402	0	[38,]	1.0062	1
[39,]	0.6344	1	[40,]	1.0364	1
[41,]	0.5883	1	[42,]	-0.1137	0
[43,]	0.6854	1	[44,]	0.8708	1
[45,]	0.2930	0	[46,]	-0.0465	0
[47,]	0.3553	0	[48,]	0.8295	1
[49,]	0.4258	0	[50,]	0.7249	1
[51,]	0.2676	0	[52,]	0.2965	0
[53,]	0.7430	1	[54,]	0.8780	1
[55,]	0.5074	0	[56,]	0.0260	0
[57,]	0.0315	0	[58,]	0.7456	1
[59,]	0.1148	0	[60,]	1.2016	1
[61,]	0.8294	1	[62,]	1.0403	1
[63,]	0.3467	0	[64,]	0.3305	0
[65,]	0.6184	1	[66,]	0.1432	0
[67,]	1.2075	1	[68,]	0.8978	1
[69,]	0.3549	0	[70,]	0.2902	0
[71,]	1.0321	1	[72,]	0.8163	1
[73,]	-0.0496	0	[74,]	0.2787	0
[75,]	0.4551	0	[76,]	-0.2218	0
[77,]	0.2401	0	[78,]	0.2718	0
[79,]	0.3663	0	[80,]	-0.0621	0
[81,]	1.1093	1	[82,]	0.0571	0
[83,]	0.3549	0	[84,]	0.8369	1
[85,]	0.3246	0	[86,]	0.0608	0
[87,]	-0.0474	0	[88,]	0.4436	0
[89,]	1.1801	1	[90,]	0.3524	0
[91,]	0.2418	0	[92,]	0.1382	0
[93,]	0.6242	1	[94,]	0.0987	0
[95,]	0.2960	0	[96,]	0.0124	0

Lampiran 5 *Syntax dan Output MARS (Lanjutan)*

```
> tabel_akurasi
  pred_Z_c
Z    0    1
0  53    3
1  12   28

> aper
[1] 0.15625

> tar
[1] 0.84375
```

Lampiran 6 Syntax dan Output Estimasi Nilai Propensity Score

a. Syntax Estimasi Nilai Propensity Score

```
#estimasi nilai propensity score
data$ps=exp(pred_Z)/(1+exp(pred_Z))
round(data$ps,digits=4)
```

b. Output Estimasi Nilai Propensity Score

```
> data$ps
 [1,] 0.5522      [2,] 0.5713      [3,] 0.4396
 [4,] 0.5959      [5,] 0.5192      [6,] 0.5618
 [7,] 0.5702      [8,] 0.6312      [9,] 0.6426
[10,] 0.5396     [11,] 0.6535     [12,] 0.6576
[13,] 0.4745     [14,] 0.5215     [15,] 0.3950
[16,] 0.6304     [17,] 0.7338     [18,] 0.5935
[19,] 0.4669     [20,] 0.5215     [21,] 0.6118
[22,] 0.6281     [23,] 0.7652     [24,] 0.5706
[25,] 0.6777     [26,] 0.5966     [27,] 0.5520
[28,] 0.5979     [29,] 0.7459     [30,] 0.6833
[31,] 0.6217     [32,] 0.5248     [33,] 0.5367
[34,] 0.5030     [35,] 0.6086     [36,] 0.5257
[37,] 0.5100     [38,] 0.7323     [39,] 0.6535
[40,] 0.7382     [41,] 0.6430     [42,] 0.4716
[43,] 0.6649     [44,] 0.7049     [45,] 0.5727
[46,] 0.4884     [47,] 0.5879     [48,] 0.6963
[49,] 0.6049     [50,] 0.6737     [51,] 0.5665
[52,] 0.5736     [53,] 0.6776     [54,] 0.7064
[55,] 0.6242     [56,] 0.5065     [57,] 0.5079
[58,] 0.6782     [59,] 0.5287     [60,] 0.7688
[61,] 0.6962     [62,] 0.7389     [63,] 0.5858
[64,] 0.5819     [65,] 0.6498     [66,] 0.5357
[67,] 0.7699     [68,] 0.7105     [69,] 0.5878
[70,] 0.5720     [71,] 0.7373     [72,] 0.6935
[73,] 0.4876     [74,] 0.5692     [75,] 0.6118
[76,] 0.4448     [77,] 0.5597     [78,] 0.5675
[79,] 0.5906     [80,] 0.4845     [81,] 0.7520
[82,] 0.5143     [83,] 0.5878     [84,] 0.6978
[85,] 0.5804     [86,] 0.5152     [87,] 0.4882
[88,] 0.6091     [89,] 0.7650     [90,] 0.5872
[91,] 0.5602     [92,] 0.5345     [93,] 0.6512
[94,] 0.5247     [95,] 0.5735     [96,] 0.5031
```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat

a. *Syntax* Stratifikasi Data

```
breakvals2=quantile(data$ps,prob=seq(from=0,to=1,by=1/
2),na.rm=T)
data$strata2=cut(data$ps,breaks=breakvals2,labels=c('1
','2'),include.lowest=TRUE)
breakvals3=quantile(data$ps,prob=seq(from=0,to=1,by=1/
3),na.rm=T)
data$strata3=cut(data$ps,breaks=breakvals3,labels=c('1
','2','3'),include.lowest=TRUE)
breakvals4=quantile(data$ps,prob=seq(from=0,to=1,by=1/
4),na.rm=T)
data$strata4=cut(data$ps,breaks=breakvals4,labels=c('1
','2','3','4'),include.lowest=TRUE)
breakvals5=quantile(data$ps,prob=seq(from=0,to=1,by=1/
5),na.rm=T)
data$strata5=cut(data$ps,breaks=breakvals5,labels=c('1
','2','3','4','5'),include.lowest=TRUE)
```

b. *Syntax* Uji *Balance* Kovariat Sebelum Strata

```
##Untuk X1
beftx1=t.test(X1[Z=="0"],X1[Z=="1"])
p.value_X1=beftx1$p.value
p.value_X1

##Untuk X2
X2_1 = X2[(X2=="0" & Z=="0")]
X2_2 = X2[(X2=="1" & Z=="0")]
X2_3 = X2[(X2=="0" & Z=="1")]
X2_4 = X2[(X2=="1" & Z=="1")]
X2tes=as.table(rbind(c(length((X2_1)),length((X2_2))),
c(length((X2_3)),length((X2_4)))))
colnames(X2tes) = rownames(X2tes) = c("0", "1")
names(dimnames(X2tes)) = c("Z", "X2")
X2_0 = length((X2_1))
n2_0 = length((X2_1))+length((X2_2))
X2_1 = length((X2_3))
n2_1 = length((X2_3))+length((X2_4))
z.propX2 = function(X2_0,X2_1,n2_0,n2_1)
{
numeratorX2=((X2_0)/(n2_0))-((X2_1)/(n2_1))
p.commonX2=((X2_0)+(X2_1))/((n2_0)+(n2_1))
```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```

denominatorX2=sqrt(p.commonX2*(1-
p.commonX2)*(1/(n2_0)+1/(n2_1)))
z.prop.risX2=numeratorX2/denominatorX2
return(z.prop.risX2)
}
zX2=z.propX2(X2_0,X2_1,n2_0,n2_1)
zX2
p.value_X2=2*pnorm(-abs(zX2))
p.value_X2

##Untuk X3
X3_1 = X3[(X3=="0" & Z == "0")]
X3_2 = X3[(X3=="1" & Z == "0")]
X3_3 = X3[(X3=="0" & Z == "1")]
X3_4 = X3[(X3=="1" & Z == "1")]
X3tes=as.table(rbind(c(length((X3_1)),length((X3_2))),
c(length((X3_3)),length((X3_4)))))
colnames(X3tes) = rownames(X3tes) = c("0", "1")
names(dimnames(X3tes)) = c("Z", "X3")
X3_0 = length((X3_1))
n3_0 = length((X3_1))+length((X3_2))
X3_1 = length((X3_3))
n3_1 = length((X3_3))+length((X3_4))
z.propX3 = function(X3_0,X3_1,n3_0,n3_1)
{
numeratorX3=((X3_0)/(n3_0))-((X3_1)/(n3_1))
p.commonX3=((X3_0)+(X3_1))/((n3_0)+(n3_1))
denominatorX3=sqrt(p.commonX3*(1-
p.commonX3)*(1/(n3_0)+1/(n3_1)))
z.prop.risX3=numeratorX3/denominatorX3
return(z.prop.risX3)
}
zX3=z.propX3(X3_0,X3_1,n3_0,n3_1)
zX3
p.value_X3=2*pnorm(-abs(zX3))
p.value_X3

##Untuk X4
beftx4=t.test(X4[Z=="0"],X4[Z=="1"])
p.value_X4=beftx4$p.value
p.value_X4

```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```
##Untuk X5
X5_1 = X5[(X5=="0" & Z == "0")]
X5_2 = X5[(X5=="1" & Z == "0")]
X5_3 = X5[(X5=="0" & Z == "1")]
X5_4 = X5[(X5=="1" & Z == "1")]
X5tes=as.table(rbind(c(length((X5_1)),length((X5_2))),
c(length((X5_3)),length((X5_4)))))
colnames(X5tes) = rownames(X5tes) = c("0", "1")
names(dimnames(X5tes)) = c("Z", "X5")
X5_0 = length((X5_1))
n5_0 = length((X5_1))+length((X5_2))
X5_1 = length((X5_3))
n5_1 = length((X5_3))+length((X5_4))
z.propX5 = function(X5_0,X5_1,n5_0,n5_1)
{
  numeratorX5=((X5_0)/(n5_0))-((X5_1)/(n5_1))
  p.commonX5=((X5_0)+(X5_1))/((n5_0)+(n5_1))
  denominatorX5=sqrt(p.commonX5*(1-
p.commonX5)*(1/(n5_0)+1/(n5_1)))
  z.prop.risX5=numeratorX5/denominatorX5
  return(z.prop.risX5)
}
zX5=z.propX5(X5_0,X5_1,n5_0,n5_1)
zX5
p.value_X5=2*pnorm(-abs(zX5))
p.value_X5

##Untuk X6
X6_1 = X6[(X6=="0" & Z == "0")]
X6_2 = X6[(X6=="1" & Z == "0")]
X6_3 = X6[(X6=="0" & Z == "1")]
X6_4 = X6[(X6=="1" & Z == "1")]
X6tes=as.table(rbind(c(length((X6_1)),length((X6_2))),
c(length((X6_3)),length((X6_4)))))
colnames(X6tes) = rownames(X6tes) = c("0", "1")
names(dimnames(X6tes)) = c("Z", "X6")
X6_0 = length((X6_1))
n6_0 = length((X6_1))+length((X6_2))
X6_1 = length((X6_3))
n6_1 = length((X6_3))+length((X6_4))
```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```

z.propX6 = function(X6_0,X6_1,n6_0,n6_1)
{
  numeratorX6=((X6_0)/(n6_0))-((X6_1)/(n6_1))
  p.commonX6=((X6_0)+(X6_1))/((n6_0)+(n6_1))
  denominatorX6=sqrt(p.commonX6*(1-
  p.commonX6)*(1/(n6_0)+1/(n6_1)))
  z.prop.risX6=numeratorX6/denominatorX6
  return(z.prop.risX6)
}
zX6=z.propX6(X6_0,X6_1,n6_0,n6_1)
zX6
p.value_X6=2*pnorm(-abs(zX6))
p.value_X6

summaryBALANCE=cbind(c(p.value_X1,p.value_X2,p.value_X
3,p.value_X4,p.value_X5,p.value_X6))
colnames(summaryBALANCE)=c("p-value")
rownames(summaryBALANCE)=c("X1", "X2", "X3", "X4", "X5", "X
6")
round(summaryBALANCE,digits=3)

```

c. *Syntax* Uji *Balance* Kovariat Untuk 2 Strata

```

ckX11=length(X1[data$strata2=="1"&Z=="0"]);ckX11
#jumlah kelompok control S1
tkX11=length(X1[data$strata2=="1"&Z=="1"]);tkX11
#jumlah kelompok treatment S1
ckX12=length(X1[data$strata2=="2"&Z=="0"]);ckX12
#jumlah kelompok control S2
tkX12=length(X1[data$strata2=="2"&Z=="1"]);tkX12
#jumlah kelompok treatment S2

##Uji Balance untuk X1
#strata 1
tX11=t.test(X1[data$strata2=="1"&Z=="0"],X1[data$strat
a2=="1"&Z=="1"])
bafX11=tX11$p.value
bafX11

```


Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```

#strata2
tX12=t.test(X1[data$strata2=="2"&Z=="0"],X1[data$strat
a2=="2"&Z=="1"])
bafX12=tX12$p.value
bafX12

BalStraX1 = cbind(c(bafX11,bafX12))
colnames(BalStraX1)=c("p-valueX1")
rownames(BalStraX1)=c("S1", "S2")
BalStraX1

##Uji Balance untuk X2
#strata 1
X21_1=X2[(X2=="0"&data$strata2=="1"&Z=="0")]
X21_2=X2[(X2=="1"&data$strata2=="1"&Z=="0")]
X21_3=X2[(X2=="0"&data$strata2=="1"&Z=="1")]
X21_4=X2[(X2=="1"&data$strata2=="1"&Z=="1")]

X2tes1=as.table(rbind(c(length((X21_1)),length((X21_2)
)),c(length((X21_3)),length((X21_4))))
colnames(X2tes1)=rownames(X2tes1)=c("0", "1")
names(dimnames(X2tes1)) = c("Z", "X2")
X2_01=length((X21_1))
n2_01=length((X21_1))+length((X21_2))
X2_11=length((X21_3))
n2_11=length((X21_3))+length((X21_4))
z.propX21 = function(X2_01,X2_11,n2_01,n2_11)
{
  numeratorX21=((X2_01)/(n2_01))-((X2_11)/(n2_11))
  p.commonX21=((X2_01)+(X2_11))/((n2_01)+(n2_11))
  denominatorX21=sqrt(p.commonX21*(1-
  p.commonX21)*(1/(n2_01)+1/(n2_11)))
  z.prop.risX21=numeratorX21/denominatorX21
  return(z.prop.risX21)
}
zX21=z.propX21(X2_01,X2_11,n2_01,n2_11)
p.value_X21=2*pnorm(-abs(zX21))
p.value_X21

#strata 2
X22_1=X2[(X2=="0"&data$strata2=="2"&Z=="0")]

```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```

X22_2=X2[(X2=="1"&data$strata2=="2"&Z=="0")]
X22_3=X2[(X2=="0"&data$strata2=="2"&Z=="1")]
X22_4=X2[(X2=="1"&data$strata2=="2"&Z=="1")]

X2tes2=as.table(rbind(c(length((X22_1)),length((X22_2)
)),c(length((X22_3)),length((X22_4))))
colnames(X2tes2)=rownames(X2tes2)=c("0", "1")
names(dimnames(X2tes2)) = c("Z", "X2")
X2_02=length((X22_1))
n2_02=length((X22_1))+length((X22_2))
X2_12=length((X22_3))
n2_12=length((X22_3))+length((X22_4))
z.propX22 = function(X2_02,X2_12,n2_02,n2_12)
{
  numeratorX22=((X2_02)/(n2_02))-((X2_12)/(n2_12))
  p.commonX22=((X2_02)+(X2_12))/((n2_02)+(n2_12))
  denominatorX22=sqrt(p.commonX22*(1-
  p.commonX22)*(1/(n2_02)+1/(n2_12)))
  z.prop.risX22=numeratorX22/denominatorX22
  return(z.prop.risX22)
}
zX22=z.propX22(X2_02,X2_12,n2_02,n2_12)
p.value_X22=2*pnorm(-abs(zX22))
p.value_X22

BalStraX2 = cbind(c(p.value_X21,p.value_X22))
colnames(BalStraX2)=c("p-valueX2")
rownames(BalStraX2)=c("S1", "S2")
BalStraX2

##Uji Balance untuk X3
#strata 1
X31_1=X3[(X3=="0"&data$strata2=="1"&Z=="0")]
X31_2=X3[(X3=="1"&data$strata2=="1"&Z=="0")]
X31_3=X3[(X3=="0"&data$strata2=="1"&Z=="1")]
X31_4=X3[(X3=="1"&data$strata2=="1"&Z=="1")]

X3tes1=as.table(rbind(c(length((X31_1)),length((X31_2)
)),c(length((X31_3)),length((X31_4))))
colnames(X3tes1)=rownames(X3tes1)=c("0", "1")
names(dimnames(X3tes1)) = c("Z", "X3")

```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```

X3_01=length((X31_1))
n3_01=length((X31_1))+length((X31_2))
X3_11=length((X31_3))
n3_11=length((X31_3))+length((X31_4))
z.propX31 = function(X3_01,X3_11,n3_01,n3_11)
{
  numeratorX31=((X3_01)/(n3_01))-((X3_11)/(n3_11))
  p.commonX31=((X3_01)+(X3_11))/((n3_01)+(n3_11))
  denominatorX31=sqrt(p.commonX31*(1-
  p.commonX31)*(1/(n3_01)+1/(n3_11)))
  z.prop.risX31=numeratorX31/denominatorX31
  return(z.prop.risX31)
}
zX31=z.propX31(X3_01,X3_11,n3_01,n3_11)
p.value_X31=2*pnorm(-abs(zX31))
p.value_X31

#strata 2
X32_1=X3[(X3=="0"&data$strata2=="2"&Z=="0")]
X32_2=X3[(X3=="1"&data$strata2=="2"&Z=="0")]
X32_3=X3[(X3=="0"&data$strata2=="2"&Z=="1")]
X32_4=X3[(X3=="1"&data$strata2=="2"&Z=="1")]

X3tes2=as.table(rbind(c(length((X32_1)),length((X32_2)
)),c(length((X32_3)),length((X32_4))))
colnames(X3tes2)=rownames(X3tes2)=c("0", "1")
names(dimnames(X3tes2)) = c("Z", "X3")
X3_02=length((X32_1))
n3_02=length((X32_1))+length((X32_2))
X3_12=length((X32_3))
n3_12=length((X32_3))+length((X32_4))
z.propX32 = function(X3_02,X3_12,n3_02,n3_12)
{
  numeratorX32=((X3_02)/(n3_02))-((X3_12)/(n3_12))
  p.commonX32=((X3_02)+(X3_12))/((n3_02)+(n3_12))
  denominatorX32=sqrt(p.commonX32*(1-
  p.commonX32)*(1/(n3_02)+1/(n3_12)))
  z.prop.risX32=numeratorX32/denominatorX32
  return(z.prop.risX32)
}

```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```

zX32=z.propX32(X3_02,X3_12,n3_02,n3_12)
p.value_X32=2*pnorm(-abs(zX32))
p.value_X32

BalStraX3 = cbind(c(p.value_X31,p.value_X32))
colnames(BalStraX3)=c("p-valueX3")
rownames(BalStraX3)=c("S1", "S2")
BalStraX3

##Uji Balance untuk X4
#strata 1
tX41=t.test(X4[data$strata2=="1"&Z=="0"],X4[data$strata2=="1"&Z=="1"])
bafX41=tX41$p.value
bafX41

#strata2
tX42=t.test(X4[data$strata2=="2"&Z=="0"],X4[data$strata2=="2"&Z=="1"])
bafX42=tX42$p.value
bafX42

BalStraX4 = cbind(c(bafX41,bafX42))
colnames(BalStraX4)=c("p-valueX1")
rownames(BalStraX4)=c("S1", "S2")
BalStraX4

##Uji Balance untuk X5
#strata 1
X51_1=X5[(X5=="0"&data$strata2=="1"&Z=="0")]
X51_2=X5[(X5=="1"&data$strata2=="1"&Z=="0")]
X51_3=X5[(X5=="0"&data$strata2=="1"&Z=="1")]
X51_4=X5[(X5=="1"&data$strata2=="1"&Z=="1")]

X5tes1=as.table(rbind(c(length((X51_1)),length((X51_2)),
),c(length((X51_3)),length((X51_4))))
colnames(X5tes1)=rownames(X5tes1)=c("0", "1")
names(dimnames(X5tes1)) = c("Z", "X5")
X5_01=length((X51_1))
n5_01=length((X51_1))+length((X51_2))

```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```

X5_11=length((X51_3))
n5_11=length((X51_3))+length((X51_4))
z.propX51 = function(X5_01,X5_11,n5_01,n5_11)
{
  numeratorX51=((X5_01)/(n5_01))-((X5_11)/(n5_11))
  p.commonX51=((X5_01)+(X5_11))/((n5_01)+(n5_11))
  denominatorX51=sqrt(p.commonX51*(1-
  p.commonX51)*(1/(n5_01)+1/(n5_11)))
  z.prop.risX51=numeratorX51/denominatorX51
  return(z.prop.risX51)
}
zX51=z.propX51(X5_01,X5_11,n5_01,n5_11)
p.value_X51=2*pnorm(-abs(zX51))
p.value_X51

#strata 2
X52_1=X5[(X5=="0"&data$strata2=="2"&Z=="0")]
X52_2=X5[(X5=="1"&data$strata2=="2"&Z=="0")]
X52_3=X5[(X5=="0"&data$strata2=="2"&Z=="1")]
X52_4=X5[(X5=="1"&data$strata2=="2"&Z=="1")]

X5tes2=as.table(rbind(c(length((X52_1)),length((X52_2)
)),c(length((X52_3)),length((X52_4))))
colnames(X5tes2)=rownames(X5tes2)=c("0", "1")
names(dimnames(X5tes2)) = c("Z", "X5")
X5_02=length((X52_1))
n5_02=length((X52_1))+length((X52_2))
X5_12=length((X52_3))
n5_12=length((X52_3))+length((X52_4))
z.propX52 = function(X5_02,X5_12,n5_02,n5_12)
{
  numeratorX52=((X5_02)/(n5_02))-((X5_12)/(n5_12))
  p.commonX52=((X5_02)+(X5_12))/((n5_02)+(n5_12))
  denominatorX52=sqrt(p.commonX52*(1-
  p.commonX52)*(1/(n5_02)+1/(n5_12)))
  z.prop.risX52=numeratorX52/denominatorX52
  return(z.prop.risX52)
}
zX52=z.propX52(X5_02,X5_12,n5_02,n5_12)
p.value_X52=2*pnorm(-abs(zX52))
p.value_X52

```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```

BalStraX5 = cbind(c(p.value_X51,p.value_X52))
colnames(BalStraX5)=c("p-valueX5")
rownames(BalStraX5)=c("S1", "S2")
BalStraX5

##Uji Balance untuk X6
#strata 1
X61_1=X6[(X6=="0"&data$strata2=="1"&Z=="0")]
X61_2=X6[(X6=="1"&data$strata2=="1"&Z=="0")]
X61_3=X6[(X6=="0"&data$strata2=="1"&Z=="1")]
X61_4=X6[(X6=="1"&data$strata2=="1"&Z=="1")]

X6tes1=as.table(rbind(c(length((X61_1)),length((X61_2)
)),c(length((X61_3)),length((X61_4))))
colnames(X6tes1)=rownames(X6tes1)=c("0", "1")
names(dimnames(X6tes1)) = c("Z", "X6")
X6_01=length((X61_1))
n6_01=length((X61_1))+length((X61_2))
X6_11=length((X61_3))
n6_11=length((X61_3))+length((X61_4))
z.propX61 = function(X6_01,X6_11,n6_01,n6_11)
{
  numeratorX61=((X6_01)/(n6_01))-((X6_11)/(n6_11))
  p.commonX61=((X6_01)+(X6_11))/((n6_01)+(n6_11))
  denominatorX61=sqrt(p.commonX61*(1-
p.commonX61)*(1/(n6_01)+1/(n6_11)))
  z.prop.risX61=numeratorX61/denominatorX61
  return(z.prop.risX61)
}
zX61=z.propX61(X6_01,X6_11,n6_01,n6_11)
p.value_X61=2*pnorm(-abs(zX61))
p.value_X61

#strata 2
X62_1=X6[(X6=="0"&data$strata2=="2"&Z=="0")]
X62_2=X6[(X6=="1"&data$strata2=="2"&Z=="0")]
X62_3=X6[(X6=="0"&data$strata2=="2"&Z=="1")]
X62_4=X6[(X6=="1"&data$strata2=="2"&Z=="1")]

X6tes2=as.table(rbind(c(length((X62_1)),length((X62_2)
)),c(length((X62_3)),length((X62_4))))

```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat (Lanjutan)

```

colnames(X6tes2)=rownames(X6tes2)=c("0", "1")
names(dimnames(X6tes2)) = c("Z", "X6")
X6_02=length((X62_1))
n6_02=length((X62_1))+length((X62_2))
X6_12=length((X62_3))
n6_12=length((X62_3))+length((X62_4))
z.propX62 = function(X6_02,X6_12,n6_02,n6_12)
{
  numeratorX62=((X6_02)/(n6_02))-((X6_12)/(n6_12))
  p.commonX62=((X6_02)+(X6_12))/((n6_02)+(n6_12))
  denominatorX62=sqrt(p.commonX62*(1-
  p.commonX62)*(1/(n6_02)+1/(n6_12)))
  z.prop.risX62=numeratorX62/denominatorX62
  return(z.prop.risX62)
}
zX62=z.propX62(X6_02,X6_12,n6_02,n6_12)
p.value_X62=2*pnorm(-abs(zX62))
p.value_X62

BalStraX6 = cbind(c(p.value_X61,p.value_X62))
colnames(BalStraX6)=c("p-valueX6")
rownames(BalStraX6)=c("S1", "S2")
BalStraX6

###Summary Balance Strata 2
summary.balance2=cbind(c(p.value_X1,BalStraX1),c(p.val
ue_X2,BalStraX2),c(p.value_X3,BalStraX3),c(p.value_X4,
BalStraX4),c(p.value_X5,BalStraX5),c(p.value_X6,BalStr
aX6))
colnames(summary.balance2)=c("X1", "X2", "X3", "X4", "X5",
"X6")
rownames(summary.balance2)=c("bef strata",
"strata 1", "strata 2")
round(summary.balance2,digits=3)

###Summary Balance Keseluruhan
round(summaryBALANCE,digits=3)
round(summary.balance2,digits=3)

```

Lampiran 7 Stratifikasi dan Pengujian *Balance* Kovariat
(Lanjutan)d. *Output* Uji *Balance* Kovariat Sebelum Strata dan Sesudah Strata

```
> round(summaryBALANCE,digits=3)
  p-value
X1  0.518
X2  0.008
X3  0.001
X4  0.283
X5  0.006
X6  0.019
> round(summary.balance2,digits=3)
      X1      X2      X3      X4      X5      X6
bef strata 0.518 0.008 0.001 0.283 0.006 0.019
strata 1   0.847 0.979 0.979 0.314 0.979 0.959
strata 2   0.279 0.654 0.032 0.732 0.987 0.209
```


Lampiran 8 Estimasi Nilai ATE dan PBR untuk 2 Strata

a. *Syntax* Estimasi Nilai ATE dan PBR Untuk 2 Strata

```

### Estimasi ATE Manual Strata 2 ###
Y.QT0=split(Y[Z==0], data$strata2[Z==0])
Y.QT1=split(Y[Z==1], data$strata2[Z==1])
Y.QT0
Y.QT1
P=sapply(Y.QT1, mean) # estimates for trt=1
S=sapply(Y.QT0, mean) # estimates for trt=0
K=sapply(Y.QT1, mean)-sapply(Y.QT0, mean)
nc1=length(Y.QT0[[1]]) #n control untuk strata 1
nc2=length(Y.QT0[[2]]) #n control untuk strata 2
nt1=length(Y.QT1[[1]]) #n treat untuk strata 1
nt2=length(Y.QT1[[2]]) #n treat untuk strata 2
nc=nc1+nc2 # n control
nt=nt1+nt2 # n treatment
theta1=((nt1+nc1)*(K[1]))/(nt+nc) # ate untuk strata 1
theta2=((nt2+nc2)*(K[2]))/(nt+nc) # ate untuk strata 2
theta_hat=(theta1+theta2) #ATE_final
theta_hat
theta1
theta2

### Mencari SE ATE Strata 2 ###
Svar_treat=sapply(Y.QT1, var) #var treat stiap strata
Svar_treat
Svar_control=sapply(Y.QT0, var) #var control stiap
strata
Svar_control
Vt1=((1/nt1))*Svar_treat[1] #var treat strata 1
Vt2=((1/nt2))*Svar_treat[2] #var treat strata 2
Vc1=((1/nc1))*Svar_control[1] #var control strata 1
Vc2=((1/nc2))*Svar_control[2] #var control strata 2
var1=Vt1+Vc1 # var strata 1
var2=Vt2+Vc2 # var strata 2
Var_theta1=((nt1+nc1)^2)*(var1) #var_ate utk strata 1
Var_theta2=((nt2+nc2)^2)*(var2) #var_ate utk strata 2
Var_theta_hat=(Var_theta1+Var_theta2)/((nt+nc)^2)
#VarATE_final
se_theta_hat=sqrt(Var_theta_hat) # se ATE
se_theta_hat

```

Lampiran 8 Estimasi Nilai ATE dan PBR untuk 2 Strata

```

## uj t stat
t=(theta_hat)/(se_theta_hat)
t

## calculating estimate of df
numstrata=2
df = length((Z)) - (2*numstrata)
df

## calculating p-value for t test
p_value=pt(-abs(t),df)
p_value

### Menghitung PBR ###
## calculate PBR overall
pst=data$ps[data$ps & Z==1]
psc=data$ps[data$ps & Z==0]
pst.bar=mean(pst)
psc.bar=mean(psc)
B.ps.before=pst.bar - psc.bar
B.ps.before
##masuk ps after strata 1
pst.1=data$ps[data$ps & Z==1 & data$strata2==1]
psc.1=data$ps[data$ps & Z==0 & data$strata2==1]
pst.1.bar=mean(pst.1)
psc.1.bar=mean(psc.1)
B1.ps.after=pst.1.bar-psc.1.bar
B1.ps.after
##masuk ps after strata 2
pst.2=data$ps[data$ps & Z==1 & data$strata2==2]
psc.2=data$ps[data$ps & Z==0 & data$strata2==2]
pst.2.bar=mean(pst.2)
psc.2.bar=mean(psc.2)
B2.ps.after=pst.2.bar-psc.2.bar
B2.ps.after
B.ps.after2=(B1.ps.after+B2.ps.after)
B.ps.after2
PBR_2=(abs(B.ps.before-B.ps.after2)/(B.ps.before))*100
PBR_2
summary.bias=cbind(c(B.ps.before,B.ps.after2, PBR_2))
rownames(summary.bias)=c("bias sebelum
strata","bias setelah strata","PBR")

```

Lampiran 8 Estimasi Nilai ATE dan PBR untuk 2 Strata

```
colnames(summary.bias)=c("2 strata")
summary.ate=rbind(theta1,theta2,theta_hat,se_theta_hat
,t,df,p_value)

round(summary.ate,digits=3)
round(summary.bias,digits=3)
```

b. *Output* Estimasi Nilai ATE dan PBR Untuk 2 Strata

```
> round(summary.ate,digits=3)
              1
theta1        0.377
theta2        0.267
theta_hat     0.645
se_theta_hat  0.089
t             7.276
df           92.000
p_value       0.000

> round(summary.bias,digits=3)
              2 strata
bias sebelum strata    0.131
bias setelah strata   0.114
PBR                   13.232
```

Halaman ini sengaja dikosongkan

BIODATA PENULIS



Penulis dengan nama lengkap Ane Wahyuni biasa dipanggil Ane lahir di Pekan Kamis pada tanggal 1 Juni 1995. Penulis merupakan anak pertama dari dua bersaudara dari Wal Asri dan Yusnida. Pendidikan formal yang telah ditempuh penulis yaitu TK Islam Al-Azhar Bukittinggi (2000-2001), SD Islam Al-Azhar Bukittinggi (2001-2007), MTs Negeri 1 Bukittinggi (2007-2010), SMA Negeri 1 Bukittinggi (2010-2013), Diploma III Statistika FMIPA Universitas Negeri Padang (2013-2017). Pada tahun 2017 penulis melanjutkan studi Lintas Jalur S1 di Statistika FMKSD Institut Teknologi Sepuluh Nopember. Pada tahun 2019 penulis menyelesaikan Tugas Akhir yang berjudul “*Propensity Score Stratification (PSS) Menggunakan Multivariate Adaptive Regresi Spline (MARS) pada Kasus Diabetes Melitus Tipe 2*”. Bagi pembaca yang ingin memberikan saran dan kritik mengenai Tugas Akhir ini dapat melalui email anewahyuni1@gmail.com.