



TUGAS AKHIR - KS 184822

**ANALISIS *PROPENSITY SCORE STRATIFICATION*
MENGUNAKAN KLASIFIKASI POHON PADA
PENDERITA DIABETES MELITUS TIPE 2**

**NAOMI PUSPITA HAPPY PURANTI
NRP 0621154000062**

**Dosen Pembimbing
Dr. Bambang Widjanarko Otok, S. Si., M.Si.**

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



TUGAS AKHIR - KS 184822

**ANALISIS *PROPENSITY SCORE STRATIFICATION*
MENGUNAKAN KLASIFIKASI POHON PADA
PENDERITA DIABETES MELITUS TIPE 2**

**NAOMI PUSPITA HAPPY PURANTI
NRP 0621154000062**

**Dosen Pembimbing
Dr. Bambang Widjanarko Otok, S.Si., M.Si.**

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

(Halaman ini sengaja dikosongkan)



FINAL PROJECT - KS 184822

***ANALYSIS OF PROPENSITY SCORE
STRATIFICATION USING TREE CLASSIFICATION IN
DIABETES MELITUS TYPE 2 PATIENTS***

**NAOMI PUSPITA HAPPY PURANTI
NRP 0621154000062**

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

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

(Halaman ini sengaja dikosongkan)

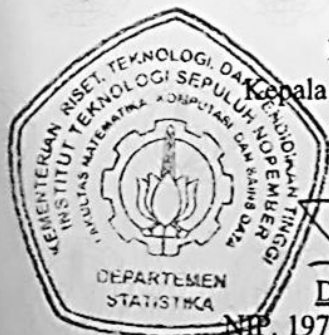
LEMBAR PENGESAHAN
ANALISIS PROPENSITY SCORE STRATIFICATION
MENGGUNAKAN KLASIFIKASI POHON PADA
PENDERITA DIABETES MELITUS TIPE 2

TUGAS AKHIR

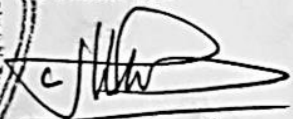
Diajukan Untuk Memenuhi Salah Satu Syarat Memperoleh Gelar
Sarjana Statistika
Pada
Program Studi Sarjana Departemen Statistika
Fakultas Matematika, Komputasi dan Sains Data
Institut Teknologi Sepuluh Nopember

Oleh :
Naomi Puspita Happy Puranti
NRP. 062115 4000 0062

Disetujui oleh Pembimbing :
Dr. Bambang Widjanarko Otok S.Si., M.Si.
NIP. 19681124 199412 1 001



Mengetahui,
Kepala Departemen Statistika
FMKSD ITS


Dr. Suhartono
NIP. 19710929 199512 1 001

SURABAYA, JULI 2019

(Halaman ini sengaja dikosongkan)

ANALISIS PROPENSITY SCORE STRATIFICATION MENGUNAKAN KLASIFIKASI POHON PADA PENDERITA DIABETES MELITUS TIPE 2

Nama Mahasiswa : Naomi Puspita Happy Puranti
NRP : 062154000062
Departemen : Statistika
**Dosen Pembimbing : Dr. Bambang Widjanarko Otok,
S.Si., M.Si.**

Abstrak

Diabetes Melitus adalah penyakit kronis, yang disebabkan karena pankreas tidak menghasilkan cukup insulin. Laporan dari Kementerian Kesehatan atau Kemenkes (2013) menunjukkan data untuk provinsi Jawa Timur menyumbang persentase prevalensi penderita DM sebesar 2,1% yang hampir 80% prevalensi kejadian DM adalah tipe 2. Penelitian ini akan menggunakan objek penelitian pasien DM tipe 2 di Rumah Sakit Umum Daerah (RSUD) kabupaten Pasuruan, Jawa Timur. Faktor-faktor yang diamati adalah Usia, Jenis Kelamin, Riwayat Genetik, Lama Menderita, Obesitas, Hipertensi dan Aktivitas Olahraga. Untuk mengatasi masalah bias akibat observasi yang tidak random dengan propensity score, didapatkan variabel confounding adalah aktivitas olahraga. Jenis propensity score yang digunakan adalah propensity score stratification dan didapati berdasarkan klasifikasi pohon terbentuk strata sejumlah 4, semua kovariat setelah dilakukan stratifikasi adalah balance, namun variabel yang dijadikan terminal node nilainya akan konstan dan tidak dapat dilakukan pengujian balance covariate pada variabel yang bersangkutan tersebut. Kinerja propensity score dievaluasi dengan Average Treatment Effect (ATE) dan didapati efek perlakuan Z terhadap Y adalah sebesar 0,573. Pendekatan PSS dalam penelitian ini berhasil mereduksi bias (PBR) sebesar 28,04%.

Kata Kunci : Diabetes Melitus tipe 2, Variabel Confounding, PSS, Klasifikasi Pohon, Uji Balance Covariate, ATE, PBR.

(Halaman ini sengaja dikosongkan)

ANALYSIS OF PROPENSITY SCORE STRATIFICATION USING TREE CLASSIFICATION IN DIABETES MELITUS TYPE 2 PATIENTS

Name : Naomi Puspita Happy Puranti
Student Number: 06211540000062
Department : Statistics
Supervisor : Dr. Bambang Widjanarko Otok, S.Si.,
M.Si.

Abstract

Diabetes mellitus is a chronic disease, which is caused by the pancreas not producing enough insulin. Reports from the Ministry of Health (2013) that data for East Java province accounted for the percentage prevalence of DM patients at 2.1%, which is almost 80%. The prevalence of DM is type 2. This study will use the object of research in DM type 2 patients at the Regional General Hospital (RSUD) in Pasuruan, East Java. The factors observed were Age, Gender, Genetic History, Duration of Suffering, Obesity, Hypertension and Sports Activities. The confounding variable was carried out with a method to overcome the problem of bias due to non-random observations is propensity score. It was found that the confounding variable was Sports Activity. The type of propensity score used is Propensity Score Stratification (PSS) and found that tree classification form 4 strata for the data, all covariates after doing this stratification are balanced, but the variables which used as terminal nodes are constant and cannot be tested for the balance of the covariates concerned. The performance of the propensity score stratification was evaluated by Average Treatment Effect (ATE) and found the effect of Z on Y was 0.573. The PSS approach in this study managed to reduce bias (PBR) by 28.04%.

Keywords: *Diabetes Mellitus Type 2, Confounding Variable, PSS, Tree Classification, Balance Covariate, ATE, PBR.*

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur penulis panjatkan atas berkat dan anugerah Tuhan Yang Maha Esa sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul **“Analisis Propensity Score Stratification Menggunakan Klasifikasi Pohon Pada Penderita Diabetes Melitus Tipe 2”** dengan lancar.

Penulis menyadari bahwa Tugas Akhir ini dapat terselesaikan tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada :

1. Kedua orang tua tercinta yang senantiasa mendoakan dan memberikan kasih sayang serta dukungan baik moril maupun material kepada penulis agar selalu berusaha dan melakukan yang terbaik.
2. Dr. Suhartono, M.Sc selaku Ketua Departemen Statistika FMKSD ITS.
3. Santi Wulan Purnami, M.Si, Ph.D selaku Ketua Program Studi Sarjana yang telah memberikan fasilitas, sarana, dan prasarana.
4. Dr. Bambang Widjanarko Otok, M.Si selaku dosen pembimbing Tugas Akhir yang telah meluangkan waktu dan dengan sangat sabar memberikan bimbingan, saran, dukungan serta motivasi selama penyusunan Tugas Akhir.
5. Dr. Sutikno, M.Si. dan Dr. Purhadi, M.Sc. selaku dosen penguji yang telah banyak memberi masukan kepada penulis.
6. Dr. Irhamah, S.Si., M.Si. selaku dosen wali yang telah banyak memberikan saran dan arahan dalam proses belajar di Departemen Statistika ITS.
7. Teman-teman Statistika ITS angkatan 2015, Vivacious, yang selalu memberikan dukungan kepada penulis selama ini.
8. Juwitasari, Mbak Ane, Mbak Gusde, Dimas, Mbak Dian dan Mas Musa, teman-teman seperjuangan TA, khususnya teman-teman yang berada di lab Lingkungan Kesehatan yang selama ini telah berjuang bersama dan saling memberikan semangat.
9. Irene Biyogant, Agnes Shalihah, Novelita Olivea, Irene Maya, Dian Rizky, Meyda Arynta, Novita Dwi Utami dan Zikrariza selaku sahabat saya yang selalu memberi bantuan dalam

proses perkuliahan di S1 Statistika ITS dan dukungan moril kepada penulis.

10. Teman-teman Para Navigator Surabaya dan divisi Persekutuan PMK-ITS 2018/2019 yang senantiasa memberi dukungan doa dan motivasi kepada penulis.
11. Semua pihak yang telah berkontribusi membuat *package* Klasifikasi Pohon beserta *paper* pendukungnya yang telah memberi pencerahan kepada penulis.
12. Semua pihak yang turut membantu dalam pelaksanaan Tugas Akhir yang tidak bisa penulis sebutkan satu persatu.

Besar harapan penulis untuk mendapatkan kritik dan saran yang membangun sehingga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang terkait.

Surabaya, Juli 2019

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
COVER PAGE	iii
LEMBAR PENGESAHAN	v
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
BAB I	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan.....	5
1.4 Manfaat.....	5
1.5 Batasan Masalah.....	6
TINJAUAN PUSTAKA	7
2.1 Variabel <i>Confounding</i>	7
2.2 (<i>Classification & Regression Trees</i>) <i>CART</i>	9
2.3 <i>Propensity Score Stratification (PSS)</i>	14
2.3 <i>Propensity Score Stratification (PSS)</i> Menggunakan (<i>Classification & Regression Trees</i>) <i>CART</i>	16
2.5 Uji Kovarian <i>Balance</i>	16
2.5.1 Uji-z	17
2.5.2 Uji-t.....	17
2.6 <i>Average Treatment Effect (ATE)</i>	18
2.7 <i>Percent Bias Reduction (PBR)</i>	19
2.8 Diabetes Melitus	20
2.9 Komplikasi Kesehatan Diabetes Melitus Tipe 2.....	22
METODOLOGI PENELITIAN	25
3.1 Sumber Data	25
3.2 Variabel Penelitian.....	25
3.3 Struktur Data.....	27
3.4 Langkah Analisis	27
3.5 Diagram Alir	28

ANALISIS DAN PEMBAHASAN	31
4.1 Karakteristik Pasien Diabetes Melitus di RSUD Kabupaten Pasuruan.....	31
4.2 Pemilihan Variabel <i>Confounding</i>	37
4.3 Estimasi Nilai <i>Propensity Score</i> menggunakan Klasifikasi Pohon	39
4.3.1 Penentuan Sampel Optimum.....	40
4.3.2 Pembentukan Pohon Klasifikasi	41
4.3.3 Analisis <i>Propensity Score Stratification</i> Berdasar Pohon Klasifikasi yang Terbentuk.....	44
4.3.4 Tingkat Ketepatan Prediksi <i>Propensity Score</i> Stratification berdasarkan Pohon Klasifikasi	45
4.4 Uji <i>Balance Covariate</i>	46
4.5 Estimasi <i>Average Treatment Effect</i> (ATE).....	47
4.6 Perbandingan <i>Percent Bias Reduction</i> (PBR) pada Setiap Strata	48
KESIMPULAN DAN SARAN.....	49
5.1 Kesimpulan	49
5.2 Saran	50
DAFTAR PUSTAKA	51
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2.1	Ilustrasi Faktor <i>Confounding</i>	7
Gambar 2.2	Kerangka Konsep Faktor-Faktor yang Berhubungan dengan Komplikasi Kesehatan pada Pasien DM menurut Darsana (2014) setelah dimodifikasi	24
Gambar 3.1	Diagram Alir Penelitian	28
Gambar 3.2	Diagram Alir Langkah Analisis (Lanjutan)	29
Gambar 4.1	Persentase Status Komplikasi Penyakit Pada Pasien DM tipe 2.	33
Gambar 4.2	Persentase Jenis Kelamin dengan Status Komplikasi Penyakit	34
Gambar 4.3	Persentase Riwayat Genetik dengan Status Komplikasi Penyakit	34
Gambar 4.4	Persentase Obesitas dengan Status Komplikasi Penyakit	35
Gambar 4.5	Persentase Status Hipertensi dengan Status Komplikasi Penyakit	35
Gambar 4.6	Persentase Aktivitas Olahraga dengan Status Komplikasi Penyakit	36
Gambar 4.7	Plot CP pada Pohon Klasifikasi yang terbentuk ...	41
Gambar 4.8	Pohon Klasifikasi yang terbentuk dari DM tipe 2	42

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

Tabel 2.1	Tabel Kontingensi.....	8
Tabel 2.2	<i>Confusion Matrix</i>	14
Tabel 3.1	Variabel Penelitian.....	25
Tabel 3.2	Struktur Data Penelitian.....	27
Tabel 4.1	Karakteristik Pasien Berdasarkan Usia dan Lama Menderita DM.....	37
Tabel 4.2	Uji <i>chi-square</i> antara variabel Jenis Kelamin dengan variabel lainnya.	38
Tabel 4.3	Uji <i>chi-square</i> antara variabel Riwayat Genetik dengan variabel lainnya.	38
Tabel 4.4	Uji <i>chi-square</i> antara variabel Obesitas dengan variabel lainnya.....	38
Tabel 4.5	Uji <i>chi-square</i> antara variabel Hipertensi dengan variabel lainnya.....	38
Tabel 4.6	Uji <i>chi-square</i> antara variabel Aktivitas Olahraga dengan variabel lainnya.....	39
Tabel 4.7	Uji <i>chi-square</i> antara variabel Komplikasi Penyakit dengan variabel lainnya.	39
Tabel 4.8	Pengambilan sampel optimum data <i>training</i> dan <i>testing</i> untuk Klasifikasi Pohon.	40
Tabel 4.9	Nilai proporsi yang didapat dari setiap strata.	44
Tabel 4.10	Nilai <i>propensity score</i> yang dari.	45
Tabel 4.11	Evaluasi Hasil Prediksi <i>Propensity Score Stratification</i> Pohon Klasifikasi Optimal	45
Tabel 4.12	Nilai Pengujian <i>Covariate Balance</i> sebelum dan sesudah dilakukan PSS.	46
Tabel 4.13	Hasil Estimasi ATE dan Standar Error <i>Propensity Score Stratification</i> berdasar Pohon Klasifikasi.	47
Tabel 4.14	Percent Bias Reduction PSS Menggunakan Pohon Klasifikasi.	48

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

Lampiran 1	Surat Pernyataan Pengambilan Data Sekunder.	55
Lampiran 2	Data Penderita DM tipe 2 di RSUD kab. Pasuruan Jawa Timur.	56
Lampiran 3	<i>Syntax</i> dan <i>Output</i> tabulasi silang antara kovariat X dengan status komplikasi penyakit.....	59
Lampiran 4	Data Penderita dengan nilai proporsi dan perhitungan <i>propensity score</i>	65
Lampiran 5	<i>Syntax dan Output</i> mencari Akurasi, Sensifitas, Spesifitas untuk Pohon Klasifikasi Optimum, Pembentukan Pohon dan Hasil Prediksi.....	70
Lampiran 6	Nilai Proporsi dan <i>Propensity Score</i> dalam Strata yang terbentuk.....	81
Lampiran 7	<i>Syntax dan Output</i> Uji <i>Balance Covariate</i> Sebelum dan Sesudah dilakukan Stratifikasi.....	84
Lampiran 8	<i>Syntax dan Output</i> ATE	99

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes Melitus (DM) adalah penyakit kronis, yang terjadi ketika pankreas tidak menghasilkan insulin yang cukup, atau ketika tubuh tidak dapat secara efektif menggunakan insulin yang dihasilkan. Hal ini menyebabkan peningkatan konsentrasi glukosa dalam darah (hiperglikemia)(WHO, 2018). Diabetes Melitus merupakan salah satu penyakit yang dianggap paling berbahaya karena penyakit ini secara perlahan dan diam-diam dapat menimbulkan masalah krisis kesehatan seperti komplikasi penyakit bahkan sampai kematian. Menurut data WHO, Indonesia berada di peringkat ke-4 tertinggi di dunia penderita Diabetes Melitus setelah India, China, dan Amerika Serikat. Pada tahun 2030 mendatang diperkirakan jumlah penderita diabetes melitus (DM) tipe 2 di Indonesia akan meningkat signifikan hingga 21,3 juta jiwa. Prevelensi orang dengan diabetes di Indonesia menunjukkan kecenderungan meningkat yaitu dari 5,7% (2007) menjadi 6,9% (2013). Prevelensi berat badan berlebih atau Obesitas merupakan salah satu faktor resiko terbesar diabetes meningkat terus dibandingkan Riskesdas 2007 dan 2013.

Menurut Riskesdas tahun 2013, Diabetes dengan komplikasi merupakan penyebab kematian tertinggi ketiga di Indonesia. Diperkirakan 2/3 orang dengan diabetes di Indonesia tidak mengetahui dirinya memiliki diabetes, dan berpotensi untuk mengakses layanan kesehatan dalam kondisi terlambat (sudah dengan komplikasi). Lebih dari 60% laki-laki dan 40% perempuan dengan diabetes meninggal sebelum berusia 70 tahun di wilayah regional Asia Tenggara (WHO, 2016). Sementara Laporan dari Kementerian Kesehatan atau Kemenkes (2013) menunjukkan data untuk provinsi Jawa Timur menyumbang persentase prevalensi penderita DM sebesar 2,1%. Nilai tersebut lebih besar dari persentase prevalensi Indonesia untuk pasien yang pernah didiagnosis menderita DM oleh dokter yaitu sebesar 1,5%. Secara umum, hampir 80% prevalensi kejadian DM adalah tipe 2 (Depkes, 2009). Mengingat semakin tingginya prevalensi kejadian DM di

Jawa Timur yang berdampak pada tingginya prevalensi komplikasi penyakit, maka perlu dilakukan penelitian untuk mengkaji lebih jauh faktor-faktor yang menyebabkan komplikasi penyakit pada pasien DM tipe 2.

Secara klinis ada dua tipe DM yang paling umum yaitu DM tipe 1 dan 2. DM tipe 1 merupakan penyakit autoimun ketika tubuh tidak dapat menghasilkan insulin atau yang dikenal sebagai *insulin-dependent* atau *childhood onset diabetes*, sedangkan DM tipe 2 disebabkan oleh gangguan pada sistem metabolisme manusia ketika insulin diproduksi, namun reseptor insulin tidak dapat merespon terhadap insulin yang dikenal dengan *non-insulin-dependent* atau *adult-onset diabetes* (Lewis, 2000). Diabetes melitus tipe 2 merupakan golongan diabetes dengan prevalensi tertinggi. Hal ini disebabkan karena beberapa faktor diantaranya faktor lingkungan dan faktor keturunan. Faktor lingkungan disebabkan karena adanya urbanisasi yang cenderung mengubah gaya hidup seseorang misalnya dalam hal makanan yang mulanya konsumsi makanan yang sehat dan bergizi menjadi konsumsi makanan yang cepat saji. Makanan cepat saji berpotensi utama untuk menimbulkan obesitas sehingga seseorang berisiko DM tipe 2. Orang yang sudah terkena obesitas memiliki risiko 4 kali lebih besar mengalami DM tipe 2 daripada orang dengan status gizi normal (WHO, 2018).

Diabetes melitus apabila tidak tertangani secara benar, maka dapat mengakibatkan berbagai macam komplikasi. Ada dua komplikasi pada DM yaitu komplikasi akut dan komplikasi kronik. Komplikasi kronik terdiri dari komplikasi makrovaskuler dan komplikasi mikrovaskuler. Penyakit jantung koroner, penyakit pembuluh darah otak, dan penyakit pembuluh darah perifer merupakan jenis komplikasi makrovaskular. Retinopati, nefropati, dan neuropati merupakan jenis komplikasi mikrovaskuler (Soewondo, 2006). Komplikasi yang terjadi pada penderita DM mengakibatkan terjadinya angka kematian dan angka kesakitan bukan hiperglikemi (Pernama, 2013). Keluhan pada penderita DM disebabkan oleh beberapa faktor diantaranya karakteristik individu meliputi jenis kelamin, umur, tingkat pendidikan, jenis pekerjaan, pendapatan, jumlah anggota keluarga, riwayat penyakit dan dapat

dipengaruhi juga dengan faktor penanganan yang meliputi diet, aktivitas fisik, terapi obat, dan pemantauan glukosa darah (Trisnawati, 2013).

Untuk mengetahui pengaruh komplikasi penyakit DM berdasarkan faktor yang mempengaruhi dapat digunakan metode (*Classification & Regression Trees*) CART, tujuan dari Klasifikasi Pohon adalah mengklasifikasikan suatu kelompok observasi kedalam subkelompok dari beberapa kelas yang diketahui. CART merupakan metode yang bisa diterapkan untuk data berjumlah besar, variabel yang sangat banyak dan melalui prosedur pemilah biner (Roger & Lewis, 2000).

Sebelum melakukan klasifikasi menggunakan Klasifikasi Pohon dilakukan penanganan terhadap variabel *confounding* dengan metode untuk mengatasi masalah bias akibat observasi yang tidak random dengan *propensity score* dimana metode ini diperkenalkan pertama kali oleh Rosenbaum dan Rubin pada tahun 1983. *Propensity score* digunakan untuk menyeimbangkan kovariat dan mengurangi bias melalui model distribusi yang dibentuk berdasarkan variabel perlakuan pada kovariat yang diamati. Setelah dilakukan estimasi, *propensity score* tersebut dapat mengurangi bias baik melalui *matching*, *stratification* dan *regression adjustment* atau kombinasi ketiganya (D'Agostino, 1998). Metode PSS merupakan suatu prosedur untuk mengelompokkan subjek ke dalam beberapa strata berdasarkan nilai estimasi *propensity score* yang telah diurutkan dari terkecil ke terbesar. Metode ini biasanya dibagi menjadi 5 strata dengan ukuran yang sama menggunakan nilai kuantil dari nilai estimasi *propensity score* (Kirklin, et al., 2013).

Analisis Klasifikasi Pohon, digunakan untuk memperkirakan hubungan antara kovariat hasil *propensity score* yang didapatkan. Metode ini menghindari membuat asumsi tentang bentuk fungsional dan interaksi seperti dalam regresi logistik standar. Dalam praktiknya di dunia kesehatan, peneliti sering dihadapkan dengan data yang hilang pada variabel yang diukur. Sedangkan data yang tidak lengkap menghalangi regresi logistik untuk dilakukan pengolahan, oleh karena itu digunakan algoritma Klasifikasi Pohon yang dianggap memungkinkan untuk data tidak

lengkap dalam dunia kesehatan yang kemudian diolah dalam pemasangan pohon dan memberikan estimasi skor kecenderungan untuk semua subjek. Kemampuan CART untuk mengakomodasi *missing value* telah digambarkan sebagai menguntungkan (Lee et al., 2010; McCaffrey et al., 2004; Moisen, 2008; Rai et al., 2017). Oleh karena itu penelitian ini adalah untuk menguji kinerja *propensity score stratification* berbasis Klasifikasi Pohon yang kemudian dihitung *Average Treatment Effect* nya dari strata yang telah terbentuk. Faktor penyebab DM tipe 2 yang digunakan pada penelitian ini antara lain Usia, Jenis Kelamin, Riwayat Genetik, Lama Menderita, Obesitas, Hipertensi dan Aktivitas Olahraga. Variabel yang diduga sebagai variabel *confounding* adalah variabel aktivitas olahraga. Variabel Aktivitas Olahraga diduga sebagai variabel *confounding* karena bagi penderita DM tipe 2 efektivitas dalam olahraga mampu menurunkan kadar glukosa darah. Penelitian ini akan menggunakan objek penelitian pasien DM tipe 2 di Rumah Sakit Umum Daerah (RSUD) Pasuruan, Jawa Timur.

Penelitian mengenai *propensity score* dengan metode CART pernah dilakukan oleh Stone et al (1995) menggunakan pendekatan ini untuk membandingkan pasien rawat inap dan rawat jalan dengan pneumonia yang didapat masyarakat, metode *propensity score stratification* pada kasus DM yang dilakukan oleh Adityaningrum, Otok, & Fithriasari (2016) dimana hasil penelitian menunjukkan bahwa variabel yang berpengaruh secara langsung terhadap status Neuropati Diabetik Perifer (NDP) pasien DM adalah variabel Jenis Kelamin, Dislipidemia, dan Obesitas serta memberi kesimpulan bahwa strata terbentuk sebanyak 5 menunjukkan error terkecil. Selanjutnya penelitian mengenai penerapan metode *propensity score matching* pernah dilakukan oleh Silviatul Hasanah (2018) dengan hasil penelitian menunjukkan bahwa variabel yang mempengaruhi secara langsung pada pasien DM tipe 2 adalah Usia, Jenis Kelamin, Riwayat Genetik, Lama Menderita, Obesitas dan Hipertensi. Sedangkan klasifikasi penderita DM menggunakan CART juga pernah dilakukan oleh Ria Dhea (2013) didapati hasil faktor penderita DM tipe 2 yaitu riwayat keluarga, Obesitas, Usia dan Jenis Kelamin.

1.2 Rumusan Masalah

Adanya hubungan antar faktor yang mempengaruhi komplikasi kesehatan pada penderita Diabetes Melitus tipe 2 seringkali dapat menyebabkan hasil estimasi yang bias dan kesimpulan yang tidak akurat karena adanya faktor *confounding* yang tidak teridentifikasi. Oleh karena itu langkah awal dalam penelitian ini adalah bagaimana mendapatkan variabel *confounding* dalam kasus penderita DM tipe 2 yang dapat menyebabkan efek bias secara empiris. Salah satu metode yang bisa menangani bias adalah metode *Propensity Score Stratification* (PSS) yang stratanya didapatkan menggunakan Pohon Klasifikasi, Pohon Klasifikasi sendiri adalah metode statistik non parametrik yang dikembangkan untuk analisis klasifikasi untuk variabel respon kategorik. Tujuan utamanya adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Berdasarkan uraian tersebut maka penelitian ini memaparkan bagaimana hasil estimasi nilai *propensity score stratification* menggunakan metode Klasifikasi Pohon berdasarkan variabel *confounding* yang ditentukan, bagaimana hasil *Average Treatment Effect* dan berapa bias yang dapat direduksi dengan menggunakan metode PSS-Pohon Klasifikasi.

1.3 Tujuan

Berdasarkan rumusan masalah yang telah dijabarkan, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mendapatkan nilai *propensity score stratification* dengan metode Klasifikasi Pohon dalam kasus DM tipe 2 di RSUD Kabupaten Pasuruan Jawa Timur berdasarkan variabel *confounding* yang ditentukan.
2. Mendapatkan nilai (*Average Treatment Effect*) ATE dari variabel *confounding* terhadap variabel respon dan bias yang tereduksi dengan dilakukannya *propensity score stratification* yang stratanya didapatkan menggunakan klasifikasi pohon.

1.4 Manfaat

Adapun manfaat bagi bidang kesehatan maupun bagi bidang pendidikan yang diharapkan dari pelaksanaan penelitian ini antara lain sebagai berikut.

1. Memperoleh faktor-faktor yang menyebabkan komplikasi kesehatan pada kasus DM tipe 2 di RSUD Kabupaten Pasuruan Jawa Timur.
2. Memberikan wawasan keilmuan bidang statistika dalam pengembangan dan penerapan metode *propensity score stratification* serta mengetahui aplikasinya dalam bidang kesehatan.

1.5 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini antara lain:

1. Metode yang digunakan dalam estimasi nilai *propensity score stratification* adalah metode Klasifikasi Pohon.
2. Variabel *confounding* yang digunakan dalam penelitian ini adalah 1 variabel.
3. Data yang digunakan adalah data sekunder hasil rekam medis pasien DM yang diamati adalah pasien DM tipe 2 di Rumah Sakit Umum Daerah (RSUD) Kabupaten Pasuruan, Jawa Timur.

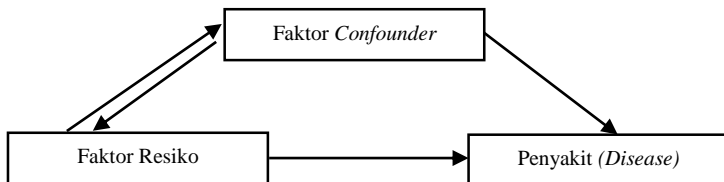
BAB II TINJAUAN PUSTAKA

Tinjauan pustaka ini membahas mengenai variabel *Confounding*, (*Classification & Regression Trees*) CART, *Propensity Score Stratification* (PSS), Uji Kovarian *Balance*, *Average Treatment Effect* (ATE), Evaluasi *Propensity Score Stratification* (PSS), Diabetes Melitus dan Komplikasi penyakit Diabetes Melitus Tipe 2.

2.1 Variabel *Confounding*

Nama *confounding* berasal dari bahasa latin “*cunfundere*” artinya *to mix together*. *Confounding* merupakan distorsi (gangguan) dalam menaksir pengaruh paparan terhadap kejadian penyakit/outcome akibat tercampurnya pengaruh faktor luar. Faktor luar (*extraneous variable*) ini dapat mempengaruhi faktor resiko dan penyakit secara bersama-sama dalam proses timbulnya penyakit, sehingga menyebabkan bias terhadap kesimpulan hasil studi. Variabel *confounding* biasanya ada jika perbandingan antara kelompok perlakuan dan kontrol tidak seimbang.

Confounding mengaburkan efek nyata dari variabel *exposure* terhadap hasil. Jager, Zoccali, MacLeod, & Dekker (2008) menjelaskan bahwa untuk menjadi potensial *confounding*, sebuah variabel perlu memenuhi ketiga kriteria berikut: (1) harus memiliki hubungan dengan penyakit tersebut, yaitu harus menjadi faktor resiko untuk penyakit, (2) harus berkaitan dengan variabel *exposure* (variabel independen), dan (3) tidak boleh menjadi efek dari *exposure*. Hubungan antara faktor *confounding* dalam suatu penyakit pada Gambar 2.1 berikut.



Gambar 2.1 *Illustrasi Faktor Confounding* (Bruce, et.al., 2008)

Pemilihan *confounder* dapat berdasarkan teori, atau berdasarkan bukti empiris tentang hubungan antar variabel. Salah

satu penghitungan secara empiris dapat digunakan uji independensi menggunakan uji *chi-square* untuk memeriksa hubungan antara dua variabel (Agresti, 1990). Hubungan antar variabel dapat dilihat di tabel kontingensi 2.1.

Tabel 2.1 Tabel Kontingensi

X ₁	X ₂		Total
	1	2	
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

Statistik Uji :

H₀: Tidak terdapat hubungan yang signifikan antar variabel.

H₁: Terdapat hubungan yang signifikan antar variabel.

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

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}; i = 1, 2, ; j = 1, 2 \quad (2.1)$$

$$\hat{\mu}_{ij} = n_{.i} \hat{\pi}_i \hat{\pi}_j = \frac{n_{i.} n_{.j}}{n_{..}} \quad (2.2)$$

$$\hat{\pi}_i = \frac{n_{i.}}{n_{..}} \quad (2.3)$$

$$\hat{\pi}_j = \frac{n_{.j}}{n_{..}} \quad (2.4)$$

dimana,

χ^2 : nilai variabel random yang distribusi sampelnya didekati oleh distribusi *chi square* dengan derajat bebas $v = (I-1)(J-1)$

i : jumlah baris

j : jumlah kolom

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

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

$\hat{\pi}_i$: Probabilitas observasi pada baris ke- i

$\hat{\pi}_j$: Probabilitas observasi pada kolom ke- j

Daerah kritik: H₀ ditolak jika $\chi^2 > \chi^2_{\alpha; df=(I-1)(J-1)}$ atau

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

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

2.2 (*Classification & Regression Trees*) CART

CART (*Classification & Regression Trees*) adalah metode statistik non parametrik yang dikembangkan untuk analisis klasifikasi, baik variabel respon kategorik maupun kontinyu, Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. CART mempunyai beberapa kelebihan seperti hasilnya lebih mudah diinterpretasikan, lebih akurat dan lebih cepat penghitungannya. Metode ini biasanya diterapkan untuk himpunan data dalam jumlah besar, variabelnya bersifat banyak dengan skala variabel campuran melalui prosedur penilaian biner.

Langkah-langkah penerapan Algoritma CART adalah sebagai berikut.

1. Pembentukan Pohon Klasifikasi

Proses pembentukan Pohon Klasifikasi terdiri atas 3 tahapan yaitu

a) Pemilihan Pemilah (*Classifier*)

Untuk membentuk Pohon Klasifikasi digunakan sampel data *Learning (L)* yang masih bersifat heterogen. Sampel tersebut akan dipilah berdasarkan aturan pemilahan dan kriteria *goodness-of-split*. Pemilihan pemilah tergantung pada jenis pohon atau lebih tepatnya tergantung pada jenis variabel responnya.

Untuk mengukur tingkat keheterogenan suatu kelas dari suatu simpul tertentu dalam Pohon Klasifikasi dikenal dengan istilah *impurity measure* $i(t)$. Ukuran ini akan membantu kita menemukan fungsi pemilah yang optimal. Fungsi kehetoregenan indeks gini $i(t)$ yang digunakan adalah sebagai berikut :

$$\text{Indeks Gini : } i(t) = \sum_{i,j=1} p(i|t) p(j|t), i \neq j \quad (2.5)$$

dimana,

$i(t)$: fungsi kehetoregenan indeks gini $i(t)$

$p(i|t)$: peluang i pada simpul t

$p(j|t)$: peluang j pada simpul t

Goodness of Split $\emptyset(s, t)$ didefinisikan sebagai penurunan keheterogenan. Kualitas ukuran dari seberapa baik pemilah s dalam menyaring data menurut kelas merupakan ukuran penurunan keheterogenan dari suatu kelas dan didefinisikan sebagai

$$\emptyset(s, t) = \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.6)$$

dimana,

$\emptyset(s, t)$: penurunan keheterogenan

$\Delta i(s, t)$: pengurangan keragaman

$p_L i(t_L)$: proporsi peluang simpul kiri

$p_R i(t_R)$: proporsi peluang simpul kanan.

pemilah yang menghasilkan nilai $\Delta i(s, t)$ atau pengurangan keragaman lebih tinggi merupakan pemilah yang lebih baik karena hal ini memungkinkan untuk mereduksi kehetoregenan secara lebih signifikan. t_L dan t_R , merupakan partisi dari simpul t menjadi dua himpunan bagian saling lepas dimana p_L dan p_R adalah proporsi masing-masing peluang simpul. Karena $t_L \cup t_R = t$ maka nilai $\Delta i(s, t)$ merepresentasikan perubahan dari keheterogenan dalam simpul t yang semata-mata disebabkan oleh pemilah s . Jika simpul yang diperoleh merupakan kelas yang tidak homogen, prosedur yang sama diulangi sampai Pohon Klasifikasi menjadi suatu konfigurasi tertentu, dan memenuhi

$$\Delta i(s^*, t_1) \max_{s \in S} \Delta i(s, t). \quad (2.7)$$

b) Penentuan Simpul Terminal

Suatu simpul t akan menjadi simpul terminal atau tidak, akan dipilah kembali bila pada simpul t tidak terdapat penurunan keheterogenan secara berarti atau adanya batasan minimum n seperti hanya terdapat satu pengamatan pada tiap simpul anak. Menurut Breiman (1993), umumnya jumlah kasus minimum dalam suatu terminal akhir adalah 5, dan apabila hal itu

terpenuhi maka pengembangan pohon dihentikan. Misalkan sebuah Pohon Klasifikasi telah terbentuk dan memiliki sekumpulan atau himpunan terminal *nodes* \tilde{T} , didefinisikan *impurity node* $I(t)$, dengan

$$I(t) = i(t)p(t) \quad (2.8)$$

didefinisikan pula *tree impurity* $I(T)$, dengan

$$I(T) = \sum_{t \in \tilde{T}} I(t) = \sum_{t \in \tilde{T}} i(t)p(t) \quad (2.9)$$

sehingga didapatkan hasil sebagai berikut

$$\Delta i(s, t) = I(t) - I(t_L) - I(t_R) \quad (2.10)$$

dimana,

\tilde{T} : himpunan terminal *nodes*

$I(t)$: *impurity node*

$I(T)$: *tree impurity*

$I(t_L)$: *impurity node* simpul kiri

$I(t_R)$: *impurity node* simpul kanan.

pemilihan split s yang memaksimalkan $\Delta i(s, t)$ ekuivalen dengan pemilihan split s yang meminimalkan *tree impurity* $I(t)$.

c) Penandaan Label Kelas

Pelabelan kelas adalah proses pengidentifikasian tiap *nodes* pada suatu kelas tertentu. Pelabelan kelas tidak hanya diberlakukan untuk *terminal nodes* saja, *nonterminal nodes* bahkan *root node* mengalami proses ini. Hal ini dikarenakan setiap *nonterminal nodes* memiliki kesempatan untuk menjadi *terminal nodes*. Sehingga proses pelabelan kelas akan terus dilakukan selama proses *splitting* masih berlanjut.

Walaupun semua *nodes* mengalami proses ini, tentu saja diantaranya yang paling membutuhkan proses ini adalah *terminal nodes*, karena *terminal nodes* adalah *nodes* yang sangat penting dalam memprediksi suatu objek pada kelas tertentu jika objek tersebut berada pada *terminal nodes* tersebut. Misalkan sebuah Pohon Klasifikasi telah terbentuk dan memiliki *terminal nodes* \tilde{T} .

Menurut Breiman, *Class assignment rule* mengidentifikasikan sebuah kelas $j \in \{1, 2, \dots, j\}$ pada setiap terminal node $t \in \tilde{T}$. Aturan pelabelan kelas, berdasarkan *misclassification cost* berbeda. Aturan itu diterangkan pada

definisi *the plurality rule* (aturan keragaman) dan berdasarkan nilai *misclassification cost* yang sama pada setiap kelas seperti berikut

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.11)$$

dimana,

$p(j|t)$: proporsi kelas j pada simpul t

$N_j(t)$: jumlah pengamatan kelas j pada simpul t

$N(t)$: jumlah pengamatan pada simpul t

j_0 : nilai dugaan kesalahan pengklasifikasian simpul t terbesar.

Proses pembentukan Pohon Klasifikasi berhenti saat terdapat hanya satu pengamatan dalam tiap tiap simpul anak atau adanya batasan minimum n , semua pengamatan dalam tiap simpul anak identik, dan adanya batasan jumlah level atau kedalaman pohon maksimal. Setelah terbentuk pohon maksimal, tahap selanjutnya adalah pemangkasan pohon untuk mencegah terbentuknya Pohon Klasifikasi yang berukuran besar dan kompleks.

2. Pemangkasan Pohon Klasifikasi

Pemangkasan dilakukan dengan jalan memangkas bagian pohon yang kurang penting sehingga didapatkan pohon optimal. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak adalah *Cost complexity minimum* (Breiman, 1993).

Sebagai ilustrasi, untuk sembarang pohon T yang merupakan sub pohon dari pohon terbesar $T_{max}(T < T_{max})$ ukuran *cost complexity* yaitu

$$R(T) = \sum_{t \in \tilde{T}} r(t)p(t) = \sum_{t \in \tilde{T}} R(t) \quad (2.12)$$

dimana,

$R(T)$ = *Tree misclassification cost*

α = Kompleksitas parameter (*complexity parameter*)

$|\tilde{T}|$ = ukuran banyaknya simpul terminal pohon T .

$R_\alpha(t)$ merupakan kombinasi linear biaya dan kompleksitas pohon yang dibentuk dengan menambahkan *cost penalty* bagi kompleksitas terhadap biaya kesalahan klasifikasi pohon. *Cost complexity pruning* menentukan suatu pohon bagian $T(\alpha)$ yang

meminimumkan $R_\alpha(t)$ pada seluruh pohon bagian, atau untuk setiap nilai α , dicari pohon bagian $T(\alpha) < T_{max}$ yang menimbulkan $R_\alpha(t)$ yaitu

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}| \quad (2.13)$$

dimana,

$R_\alpha(t)$: kombinasi linear biaya dan kompleksitas pohon

cost complexity pruning menentukan suatu pohon bagian $T(\alpha)$ yang meminimumkan $R_\alpha(t)$ pada seluruh pohon bagian, atau untuk setiap nilai α , dicari pohon bagian $T(\alpha) < T_{max}$ yang meminimumkan $R_\alpha(T)$ yaitu

$$R_\alpha(T(\alpha)) = \min_{T < T_{max}} R_\alpha(T) = \min_{T < T_{max}} [R(T) + \alpha|\tilde{T}|] \quad (2.14)$$

proses pemangkasan Pohon Klasifikasi dimulai dengan mengambil t_R yang merupakan *right child node* dan t_L yang merupakan *left child node* dari T_{max} yang dihasilkan dari *parent node* t . jika diperoleh dua *child node* dan *parent node* yang memenuhi persamaan $R(t) = R(t_R) + R(t_L)$ maka *child node* t_R dan t_L dipangkas. Hasilnya adalah pohon T_1 yang memenuhi kriteria $R(T_1) = R(T_{max})$. Proses tersebut diulang sampai tidak ada lagi pemangkasan yang mungkin terjadi. Dan diperoleh urutan sebagai berikut $\{T_1, T_2, \dots, \{t_1\}\}$, $T_1 > T_2 > \dots > \{t_1\}$. Dengan urutan α sebagai berikut $\alpha_1 = 0 < \alpha_2 < \alpha_3 < \dots < \alpha_k$.

3. Penentuan Pohon Klasifikasi Optimal

Pohon Klasifikasi yang berukuran besar akan memberikan nilai penduga pengganti paling kecil, sehingga pohon ini cenderung dipilih untuk menduga nilai respon. Tetapi ukuran pohon yang besar akan menyebabkan nilai kompleksitas yang tinggi karena struktur data yang digambarkan cenderung kompleks, sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti cukup kecil. Ada dua jenis penduga pengganti, penduga sampel uji (*test sample estimate*) dan penduga validasi silang lipat V (*cross validation V-fold estimate*). Dalam penelitian ini akan digunakan dua jenis penduga yaitu *training* dan *testing*.

4. Ketepatan Klasifikasi

Ukuran ketepatan klasifikasi dapat ditinjau dari besar akurasi (%), *sensifity* (%), dan *specificity* (%). Akurasi adalah ukuran untuk mengukur ketepatan hasil klasifikasi secara keseluruhan, semakin tinggi nilainya maka metode klasifikasi yang digunakan semakin baik. *Sensifity* menunjukkan kelas positif yang benar dibanding total data yang diklasifikasikan sebagai kelas positif. *Specificity* adalah ukuran yang menyatakan kelas negatif benar dibanding dengan total data yang diklaifikasikan sebagai kelas negatif. Untuk mempermudah hitungan maka dibuatlah dalam tabel 2.2.

Tabel 2.2 *Confusion Matrix*

	Prediksi (+)	Prediksi (-)
Aktual (+)	TP (<i>True Positives</i>)	FN (<i>False Negatives</i>)
Aktual (-)	FP (<i>False Positives</i>)	TN (<i>True Negatives</i>)

Ketiga rumus ukuran ketepatak klasifikasi dapat dituliskan sebagai berikut.

$$\text{Akurasi (\%)} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \%. \quad (2.15)$$

$$\text{Sensifity (\%)} = \frac{TP}{TP+FN} \times 100 \%. \quad (2.16)$$

$$\text{Specificity (\%)} = \frac{TN}{FP+TN} \times 100 \%. \quad (2.17).$$

2.3 *Propensity Score Stratification* (PSS)

Propensity Score Stratification (PSS) merupakan suatu metode pengelompokan subjek ke dalam kelas-kelas yang berbeda berdasarkan karakteristik yang diamati, sehingga dapat dilakukan perbandingan antara kelompok perlakuan dan kontrol untuk setiap kelas yang sama, cara ini dilakukan untuk mengontrol perbedaan sistematis antara kelompok perlakuan dan kontrol.

Seluruh sampel pengamatan akan dikelompokkan atau distratifikasi menjadi lima kelompok dengan ukuran yang sama menggunakan kuantil dari estimasi *propensity score* (Austin, 2011). Menurut Li, Graham, & Majumdar di tahun 2012, keefektifan subklasifikasi dalam menghilangkan bias pada penelitian observasional dan menemukan bahwa lima strata dapat

menghilangkan 90% bias, sehingga *propensity score* distratifikasi menjadi lima subklasifikasi dapat menghilangkan banyak perbedaan kovariat yang diamati antara kelompok perlakuan dan kelompok kontrol. Sedangkan menurut Imbens (2004), kovariat di dalam setiap subklasifikasi/strata harus seimbang. Jika *propensity score* pada tiap subklasifikasi konstan, maka distribusi kovariat antara kelompok perlakuan dan kelompok kontrol adalah sama. Bila kovariat tidak seimbang, spesifikasi *propensity score* tidak memadai dan akan ditentukan kembali, misalnya melalui penambahan bentuk dengan orde lebih tinggi atau interaksi atau dengan membagi lagi subjek ke dalam sub-kelas (Dehejja & Wahba, 1999). Penghitungan *propensity score* dapat dihitung dengan

$$e(x_j) = P(Z_i = 1 | X_j = x_j) = \frac{\exp(f(x_j))}{1 + \exp(f(x_j))} \quad (2.18)$$

dimana,

- i = mengidentifikasi nilai setiap individu pada kelompok yang terbentuk dalam kovariat $i=1, 2, \dots, I$
- j = jumlah kelompok dalam kovariat $j=0, 1$
- $(Z_i = 1)$ = probabilitas bersyarat dari kelompok *treatment*
- $(Z_i = 0)$ = probabilitas bersyarat dari kelompok kontrol
- $f(x_j)$ = nilai proporsi dari kelompok *treatment* tertentu ($Z_i = 1$) dibandingkan dengan kelompok kontrol ($Z_i = 0$)
- $e(x_j)$ = estimasi *propensity score* dari proporsi yang didapat yang dilakukan dengan cara memasukkan nilai proporsi pada persamaan.

Misalkan terdapat p variabel prediktor (X_1, X_2, \dots, X_p) mempengaruhi variabel respon Y sehingga membentuk suatu fungsi $Y=f(X_1, X_2, \dots, X_p)$. Hal ini menunjukkan bahwa masing-masing variabel prediktor (X_1, X_2, \dots, X_p) berpengaruh terhadap variabel respon Y . Dalam suatu keadaan terdapat hubungan yang terjadi antar variabel prediktor, misalkan variabel X_6 berpengaruh terhadap variabel prediktor lainnya. Variabel X_6 ini dapat disebut dengan istilah faktor *confounding* dan selanjutnya dapat dinotasikan dengan Z . Hubungan variabel Z dengan variabel prediktor X_1, X_2, \dots, X_p selain X_6 dapat dituliskan menjadi fungsi $Z=f(X_1, X_2, \dots, X_{p-1})$ sehingga variabel prediktor sebesar $p-1$ dapat

tereduksi menjadi satu skor dimensi menjadi variabel Z . Variabel Z inilah selanjutnya digunakan dalam estimasi *propensity score*. Estimasi *propensity score* dapat diestimasi menggunakan beberapa metode sesuai skala data variabel *confounding Z*. Jika variabel Z mempunyai skala nominal atau ordinal maka dapat menggunakan beberapa metode antara lain: regresi logistik, CART, MARS klasifikasi, atau SVM klasifikasi. Sedangkan jika variabel Z mempunyai skala interval atau rasio maka dapat menggunakan beberapa metode diantaranya regresi linear, MARS prediksi, atau SVR (Ernawati, 2018)

2.4 Propensity Score Stratification (PSS) Menggunakan (Classification & Regression Trees) CART

Propensity score pada umumnya diestimasi dengan menggunakan metode regresi logistik. Pada penelitian ini menggunakan estimasi dengan salah satu metode *machine learning* yaitu (*Classification & Regression Trees*) CART. Pendekatan Pohon Klasifikasi CART memiliki sejumlah keunggulan dibandingkan dengan regresi logistik, karena klasifikasi algoritma *propensity score* secara otomatis memilih variabel untuk model, algoritma juga secara otomatis mendeteksi interaksi dalam data, sehingga interaksi tersebut tidak harus ditemukan dan dimodelkan secara eksplisit seperti yang dilakukan dalam regresi logistik, simpul terminal pohon secara otomatis membentuk strata.

Node terminal masing-masing menghasilkan *propensity score* yang sama untuk kasus-kasus dalam suatu simpul. *Node* terminal yang berbeda dapat menghasilkan *propensity score* yang sama nilai skor melalui serangkaian pemisahan yang berbeda maka jika ini terjadi, terminal *node* tersebut akan diciutkan untuk membentuk strata tunggal. Selanjutnya, akan dibuat vektor berisi estimasi *propensity score* masing-masing faktor dan kemudian diperiksa tumpang tindih dalam skor kecenderungan antara kelompok-kelompok dalam strata (Luellen, Shadish, & Clark, 2005).

2.5 Uji Kovarian Balance

Menurut Li, Graham, & Majumdar di tahun 2012, salah satu cara untuk menilai kualitas *Propensity Score Stratification* yaitu dengan membandingkan variasi statistik seperti rata-rata, median,

variansi, statistik uji-t, statistik uji z. Pada penelitian ini akan dilakukan uji lanjutan untuk menguji perbedaan antara kelompok perlakuan dan kontrol dengan uji-t sebagai uji beda dua kelompok data kontinu, sedangkan uji-z untuk uji beda dua kelompok data kategorik.

2.5.1 Uji-z

Uji pada data kategorik untuk perbandingan antara kelompok *treatment* dan kelompok kontrol dapat digunakan uji-z (Li, Graham, & Majumdar, 2012). Hipotesis yang digunakan untuk dua kelompok tersebut adalah sebagai berikut :

$H_0: P_t = P_c$ (Tidak terdapat perbedaan antara kelompok perlakuan dan kontrol)

$H_1: P_t \neq P_c$ (Terdapat perbedaan antara kelompok perlakuan dan kontrol)

Statistik uji dilakukan sebagai berikut :

$$Z = \frac{\hat{p}_t - \hat{p}_c}{\sqrt{\hat{p}(1-\hat{p})\left[\frac{1}{n_t} + \frac{1}{n_c}\right]}} \quad (2.19)$$

$$\hat{p} = \frac{x_t + x_c}{n_t + n_c} \quad (2.20)$$

$$\hat{p}_t = \frac{x_t}{n_t} \quad (2.21)$$

$$\hat{p}_c = \frac{x_c}{n_c} \quad (2.22)$$

dimana,

\hat{p} : proporsi gabungan

\hat{p}_t : proporsi kelompok

\hat{p}_c : proporsi kelompok kontrol

n_t : banyak observasi kelompok perlakuan

n_c : banyak observasi kelompok kontrol

Daerah kritik : H_0 ditolak jika $|z_{hitung}| > z_{\alpha/2}$ atau $p\text{-value} < \alpha$.

2.5.2 Uji-t

Sedangkan untuk data yang kontinu dapat digunakan uji-t untuk perbandingan antara kelompok *treatment* dan kelompok kontrol (Kleyman, 2009). Hipotesis yang digunakan untuk dua kelompok tersebut adalah sebagai berikut :

$H_0: \mu_t = \mu_c$ (Tidak terdapat perbedaan antara kelompok perlakuan dan kontrol)

$H_1: \mu_t \neq \mu_c$ (Terdapat perbedaan antara kelompok perlakuan dan kontrol)

Statistik uji dilakukan sebagai berikut :

$$t_1 = \frac{\bar{x}_t - \bar{x}_c}{s_{pooled} \sqrt{\frac{1}{n_t} + \frac{1}{n_c}}} \quad (2.23)$$

$$t_2 = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}}} \quad (2.24)$$

$$s_{pooled} = \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c - 2}} \quad (2.25)$$

$$df = \frac{[(S_t^2/n_t) + (S_c^2/n_c)]^2}{[(S_t^2/n_t)/(n_t - 1) + (S_c^2/n_c)/(n_c - 1)]} \quad (2.26)$$

dimana,

t_1 : Uji-t varians homogen

t_2 : Uji-t varians tidak homogen

\bar{x}_t : *mean* kelompok perlakuan

\bar{x}_c : *mean* kelompok kontrol

s_t^2 : varians kelompok perlakuan

s_c^2 : varians kelompok kontrol

s_{pooled} : simpangan baku gabungan

Daerah kritik : H_0 ditolak jika $|t_{hitung}| > t_{\alpha/2, df}$ atau $p\text{-value} < \alpha$

2.6 Average Treatment Effect (ATE)

Setelah estimasi nilai *propensity score* sudah didapatkan dan uji *balance* kovariat telah terpenuhi maka dilakukan estimasi dari setiap efek perlakuan $\hat{\theta}$ (ATE), efek perlakuan yang meliputi standar error dan statistik ujinya dapat dilakukan melalui langkah-langkah sebagai berikut (Tu & Zhou, 2002) :

1. Membagi subjek atau melakukan stratifikasi ke dalam minimal 5 subkelas yang homogen
2. Menentukan rata-rata respon dari kelompok perlakuan dan kelompok kontrol dalam subkelas yang telah ditentukan, serta banyak subjek kelompok perlakuan dan kelompok kontrol. Dapat dihitung dengan persamaan (2.26) :

$$\bar{Y}_{tk} = \sum_{i=1}^{N_{tk}} \frac{Y_{tki}}{N_{tk}} \quad (2.27)$$

$$\bar{Y}_{ck} = \sum_{i=1}^{N_{ck}} \frac{Y_{cki}}{N_{ck}} \quad (2.28)$$

$$n_t = \sum_{k=1}^K N_{tk} \quad (2.29)$$

$$n_c = \sum_{k=1}^K N_{ck} \quad (2.30)$$

3. Setelah dilakukan penghitungan seperti diatas maka kemudian dimasukkan kedalam penghitungan efek perlakuan dalam ATE dan standar error ATE sebagai berikut :

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

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

dimana,

n_t : Subjek kelompok perlakuan

n_c : Subjek kelompok kontrol

N_{tk} : Banyak subjek kelompok perlakuan strata ke-k

N_{ck} : Banyak subjek kelompok kontrol strata ke-k

s_{tk}^2 : varians respon kelompok perlakuan ke-k

s_{ck}^2 : varians respon kelompok kontrol ke-k

k : banyaknya kelompok sub-kelas dimana $k=1,2,\dots,K$.

4. Berdasarkan persamaan yang telah diperoleh maka dilakukan statistik uji estimasi menggunakan uji-t seperti berikut :

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

$\hat{\theta}$ merupakan rata-rata efek perlakuan (ATE), $SE(\hat{\theta})$ merupakan standar error ATE. Daerah kritis untuk taraf signifikansi α adalah tolak H_0 bila $|t_{hitung}| > t_{\alpha/2, df}$.

2.7 Percent Bias Reduction (PBR)

Salah satu kriteria kebaikan metode PSS dapat dilihat dari seberapa besar bias yang mampu direduksi oleh metode tersebut untuk setiap kovariat maupun bias secara keseluruhan. Perhitungan bias melalui kelompok perlakuan dan kontrol dapat digunakan rumus.

$$B = E(x|Z = 1) - E(x|Z = 0) \quad (2.34)$$

dapat juga disebut sebagai rata-rata kovariat kelompok perlakuan dan rata-rata kelompok kontrol. Setelah itu, menghitung besarnya

bias yang direduksi untuk masing-masing kovariat dapat menggunakan persamaan (2.28) sebagai berikut (Pan & Bai, 2015):

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

dengan PBR merupakan *Percent Bias Reduction, before Strata B* dan *after Strata B* masing masing merupakan bias sebelum dilakukan stratifikasi dan setelah dilakukan stratifikasi.

2.8 Diabetes Melitus

Diabetes Melitus adalah penyakit kronis, yang disebabkan karena pankreas tidak menghasilkan cukup insulin, atau ketika tubuh tidak dapat secara efektif menggunakan insulin yang dihasilkan. Akibat yang terjadi karenanya adalah peningkatan konsentrasi glukosa dalam darah (hiperglikemia) sehingga bila keadaan ini terjadi secara menahun maka akan mengenai seluruh sistem tubuh (WHO, 2018). Orang yang terkena diabetes memiliki Hiperglikemia hingga kadar glukosa puasa yang lebih tinggi dari 110 mg/dL. Glukosa difiltrasi oleh glomerulus selama kadar glukosa dalam plasma tidak melebihi 160-180mg/dL (Price & Wilson, 2006). Kejadian DM juga ditandai dengan kadar glukosa darah melebihi normal yaitu kadar glukosa darah sewaktu ≥ 200 mg/dL (Misnadiarly, 2006). Jenis diabetes menurut WHO di tahun 2018 yaitu :

1. Diabetes tipe 1 : Ditandai oleh kurangnya produksi insulin
2. Diabetes tipe 2 : Disebabkan oleh penggunaan efektif tubuh insulin. Hal ini terjadi karena berat badan berlebih dan kurangnya aktivitas fisik
3. Gestational Diabetes : Hiperglikemia yang ditemukan saat pertama kali kehamilan

Menurut Sustrani, Alam, & Hadibroto (2010), faktor risiko DM antara lain :

a) Faktor Resiko yang tidak dapat dimodifikasi

1. Faktor Usia

Manusia mengalami penurunan fungsi fisiologis yang lebih cepat setelah mencapai Usia 40 tahun pada umumnya. Pada umumnya, DM sering muncul pada orang-orang yang telah memasuki Usia diatas 40 tahun dan mereka mengalami

berat badan berlebih (Obesitas), sehingga tubuh menjadi tidak peka terhadap insulin.

2. Faktor keturunan (Riwayat Genetik)

Penyakit diabetes dapat diturunkan menurut silsilah keluarga karena *Deoxyribo Nucleic Acid* atau DNA orang yang mengalami diabetes akan ikut diinformasikan pada gen berikutnya seperti pada keturunan selanjutnya terkait penurunan produksi insulin.

3. Faktor Jenis Kelamin

Jenis Kelamin adalah perbedaan seks didapat dari sejak lahir yang dibedakan antara laki-laki dan perempuan. Baik pria maupun wanita memiliki resiko yang sama besar untuk mengidap diabetes sampai Usia dewasa awal. Setelah Usia 30 tahun wanita memiliki resiko yang lebih tinggi daripada pria (Depkes, 2009).

4. Faktor Demografi (Ras/Etnik)

Faktor-faktor demografi seperti jumlah penduduk yang meningkat, urbanisasi, kelompok penduduk berUsia diatas 40 tahun yang meningkat, dan kurangnya gizi pada suatu lingkungan tertentu dapat menjadi faktor risiko terjadinya DM.

b) Faktor Resiko yang dapat dimodifikasi

1. Faktor Obesitas

Menurut Wijayanti (2013), Obesitas merupakan keadaan patologis karena penimbunan lemak berlebih dari yang diperlukan oleh tubuh. Obesitas merupakan salah satu faktor risiko penyebab terjadinya penyakit degeneratif seperti DM, jantung koroner, dan Hipertensi. Pengembangan DM tipe 2 terjadi pada kelompok Asia dengan BMI lebih dari 27,5 (NHS, 2016).

2. Faktor Hipertensi

Hipertensi merupakan suatu "*acceleration*" pada komplikasi kardiovaskular dan memiliki pengaruh buruk pada mikroangiopati (retina, ginjal). Prevalensi Hipertensi pada DM dua kali lebih banyak dari penduduk umum. 80% penderita diabetes mengalami Hipertensi, indikator dari

terkena Hipertensi apabila tekanan darah pasien diatas 130 mmHg.

3. Gaya Hidup

Gaya hidup yang dimaksud adalah kebiasaan konsumsi makanan beresiko, merokok, konsumsi alkohol. Untuk konsumsi makanan beresiko dalam penelitian ini antara lain: jeroan, makanan yang asam, dan minuman berkafein.

4. Aktivitas Olahraga

Aktifitas fisik yang teratur dapat berperan dalam mencegah risiko DM dengan meningkatkan massa tubuh tanpa lemak dan secara bersamaan mengurangi lemak tubuh. Aktifitas fisik mengakibatkan insulin semakin meningkat sehingga kadar gula dalam darah akan berkurang. Orang yang jarang beraktifitas fisik dan jarang melakukan olahraga, zat makanan yang masuk ke dalam tubuh tidak akan dibakar tetapi akan ditimbun dalam bentuk lemak dan gula. Jika kondisi pankreas tidak adekuat dalam menghasilkan insulin dan tidak mencukupi untuk mengubah glukosa menjadi energi maka akan timbul penyakit DM (Kemenkes, 2013).

2.9 Komplikasi Kesehatan Diabetes Melitus Tipe 2

Komplikasi kesehatan DM tipe 2 dibagi kedalam dua jenis, yaitu komplikasi metabolik dan akut. Menurut Soewondo (2006), komplikasi metabolik akut terdiri dari komplikasi yang terjadi ketika kadar glukosa darah < 60 mg % (hipoglikemia) dan kadar glukosa darah > 250 mg % (hiperglikemia), sedangkan komplikasi kronik pada dasarnya terjadi pada semua pembuluh darah di seluruh bagian tubuh.

Komplikasi kronik dari diabetes melitus sendiri dapat dibagi menjadi 2 : komplikasi mikrovaskuler dan makrovaskuler. Komplikasi mikrovaskuler terdiri dari:

1. Retinopati diabetik

Pada retinopati diabetik proliferasi terjadi iskemia retina yang progresif yang merangsang neovaskularisasi yang menyebabkan kebocoran protein-protein serum dalam jumlah besar. Neovaskularisasi yang rapuh ini berproliferasi ke bagian dalam korpus vitreum yang bila tekanan meninggi saat

berkontraksi maka bisa terjadi perdarahan masif yang berakibat penurunan penglihatan mendadak. Hal tersebut pada penderita DM bisa menyebabkan kebutaan (Bilous & Donnelly, 2014).

2. Neuropati diabetik

Neuropati diabetik perifer merupakan penyakit neuropati yang paling sering terjadi. Berisiko tinggi untuk terjadinya ulkus kaki dan amputasi. Gejala yang sering dirasakan kaki terasa terbakar dan bergetar sendiri dan lebih terasa sakit di malam hari (Soewondo, 2006).

3. Nefropati diabetik

Ditandai dengan albuminuria menetap > 300 mg/24 jam atau > 200 ig/menit pada minimal 2x pemeriksaan dalam waktu 3-6 bulan. Berlanjut menjadi proteinuria akibat hiperfiltrasi patogenik kerusakan ginjal pada tingkat glomerulus. Akibat glikasi nonenzimatik dan AGE, *advanced glycation product* yang *irreversible* dan menyebabkan hipertrofi sel dan kemoatraktan mononuklear serta inhibisi sintesis nitric oxide sebagai vasodilator, terjadi peningkatan tekanan intraglomerulus dan bila terjadi terus menerus dan inflamasi kronik, nefritis yang reversible akan berubah menjadi nefropati dimana terjadi keruakan menetap dan berkembang menjadi *chronic kidney disease* (Soewondo, 2006).

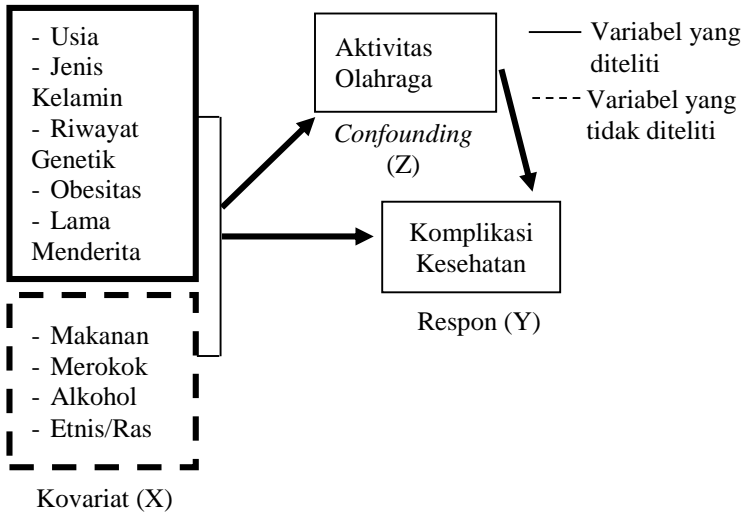
Komplikasi makrovaskular yang sering terjadi biasanya merupakan makroangiopati. Penyakit yang termasuk dalam komplikasi makrovaskular adalah :

1. Penyakit pembuluh darah jantung atau otak
2. Penyakit pembuluh darah tepi

Penyakit arteri perifer sering terjadi pada penyandang diabetes, biasanya terjadi dengan gejala tipikal intermiten atau klaudikasio, meskipun sering tanpa gejala. Terkadang ulkus iskemik kaki merupakan kelainan yang pertama muncul (Soewondo, 2006).

Menurut Bansal *et al.* (2014) resiko pasien terkena penyakit komplikasi akan semakin besar bila pasien menderita DM dengan rentang waktu yang lama. Menurut Bazzano, Serdula, & Liu (2005)

faktor-faktor risiko DM tipe 2 yang dapat dikurangi risikonya meliputi dari Obesitas, asupan alkohol, merokok, kegiatan sehari hari dan faktor diet seperti asupan lemak, serat, serta beban glikemik. Berdasarkan pendukung diatas maka diasumsikan variabel yang diduga kuat akan menjadi variabel *confounding* adalah Hipertensi dan Aktivitas Olahraga. Berikut merupakan kerangka konsep yang digunakan dalam penelitian ini.



Gambar 2.2 Kerangka Konsep Faktor-Faktor yang Berhubungan dengan Komplikasi Kesehatan pada Pasien DM menurut Darsana (2014) setelah dimodifikasi

BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yaitu data rekam medik pasien DM tipe 2 di Rumah Sakit Umum Daerah (RSUD) Kabupaten Pasuruan, Jawa Timur. Penelitian ini merupakan penelitian terapan lanjutan Dr. Bambang Otok S.Si., 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 antara lain sebagai berikut.

Tabel 3.1 Variabel Penelitian

Nama Variabel	Skala Data	Keterangan	Definisi Operasional
Usia (X_1)	Rasio	Usia pasien berdasarkan KTP (tahun)	Salah satu faktor risiko komplikasi DM Tipe 2 adalah orang yang berusia >45 tahun.
Jenis Kelamin (X_2)	Nominal	0 = Perempuan 1 = Laki-Laki	Penyebaran penyakit diabetes melitus tipe 2 pada laki-laki dan perempuan yang berbeda antara lingkungan.
Riwayat Genetik (X_3)	Nominal	0 = Punya Riwayat Genetik 1 = Tidak Punya Riwayat Genetik	Faktor resiko terdiri dari faktor risiko yang dapat diubah dan tidak dapat diubah. Faktor riwayat genetik merupakan tidak dapat diubah.

Tabel 3.1 Variabel Penelitian (Lanjutan)

Nama Variabel	Skala Data	Keterangan	Definisi Operasional
Lama Menderita DM (X ₄)	Rasio	Waktu dalam hitungan tahun	Semakin lama seseorang menderita DM maka risiko untuk terkena penyakit komplikasi semakin besar (Bansal <i>et al</i> , 2014)
Obesitas (X ₅)	Nominal	1 = Tidak Obesitas jika <i>Body Mass Index</i> < 27,5 0 = Obesitas jika <i>Body Mass Index</i> ≥ 27,5	Menurut (PERKENI, 2015) salah satu faktor risiko DM adalah Obesitas (IMT ≥ 23 kg/m ²).
Status Hipertensi (X ₆)	Nominal	1 = Tidak Hipertensi jika tekanan darah sistolik ≤ 130 mmHg 0 = Hipertensi jika tekanan darah sistolik > 130 mmHg	Penderita Hipertensi untuk pasien DM dan memiliki penyakit komplikasi seperti gagal ginjal kronis memiliki target penurunan tekanan darah tinggi mencapai nilai dibawah 130/80 mm/Hg.
Aktivitas Olahraga (X ₇)	Nominal	0 = Kurang Mengikuti Senam Diabetes 1 = Cukup Banyak Mengikuti Senam Diabetes	Aktivitas Olahraga dapat menurunkan kadar glukosa darah secara signifikan pada penderita DM.
Komplikasi Kesehatan (Y)	Nominal	0 = memiliki penyakit komplikasi 1 = Tidak memiliki penyakit komplikasi	Dikatakan mengalami komplikasi kesehatan jika orang tersebut juga menderita penyakit lainnya.

3.3 Data Penelitian

Dari variabel penelitian diatas dapat dilihat struktur data pada data penelitian ini yang disajikan pada Tabel 3.2, Struktur beserta data yang lengkap dapat dilihat di Lampiran 2.

Tabel 3.2 Struktur Data Penelitian

Pasien ke-i	Komplikasi Kesehatan (y_i)	Variabel Prediktor ($x_{i,p}$)			
		Usia	Jenis Kelamin	...	Aktivitas Olahraga
1	0	47	0	...	0
2	1	43	1	...	0
3	1	70	0	...	1
⋮	⋮	⋮	⋮	⋮	⋮
96	0	29	0	...	1

Keterangan:

n = Jumlah pasien DM tipe 2 di RSUD Kabupaten Pasuruan.

y_i = Komplikasi Kesehatan pada pasien ke- i , $i=1,2,\dots,96$; bernilai 0 untuk pasien yang memiliki komplikasi penyakit dan bernilai 1 untuk pasien yang tidak memiliki komplikasi penyakit.

$x_{i,p}$ = Nilai variabel prediktor pada pasien ke- i prediktor ke- p , $i=1,2,\dots,96$; $p=1,2,\dots,7$.

Pada penelitian ini kategori 0 digunakan untuk kelompok kontrol kegiatan dan kategori 1 untuk kelompok *treatment*.

3.4 Langkah Analisis

Sesuai dengan tujuan penelitian maka perlu melakukan langkah-langkah sebagai berikut:

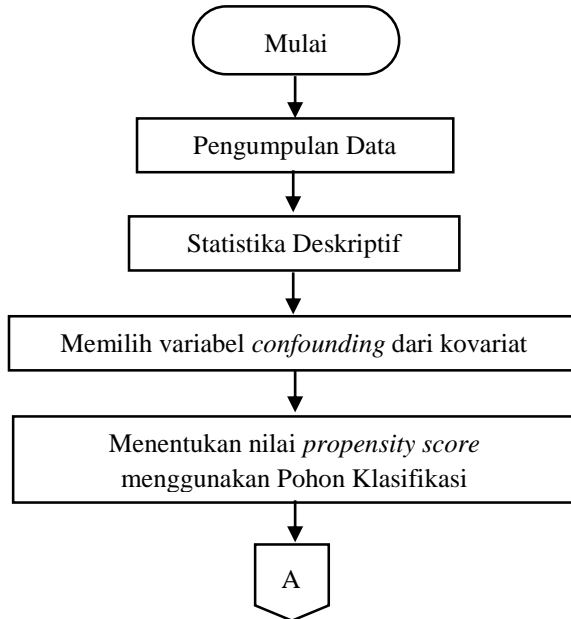
1. Mendeskripsikan karakteristik-karakteristik pasien DM tipe 2.
2. Menentukan variabel *confounding*.
 - a) Memilih variabel yang diduga menyebabkan *confounding*
 - b) Menentukan variabel *confounding* salah satunya menggunakan uji *chi-square*.
 - c) Membentuk fungsi antara variabel *confounding* dan variabel prediktor lainnya selain variabel *confounding*

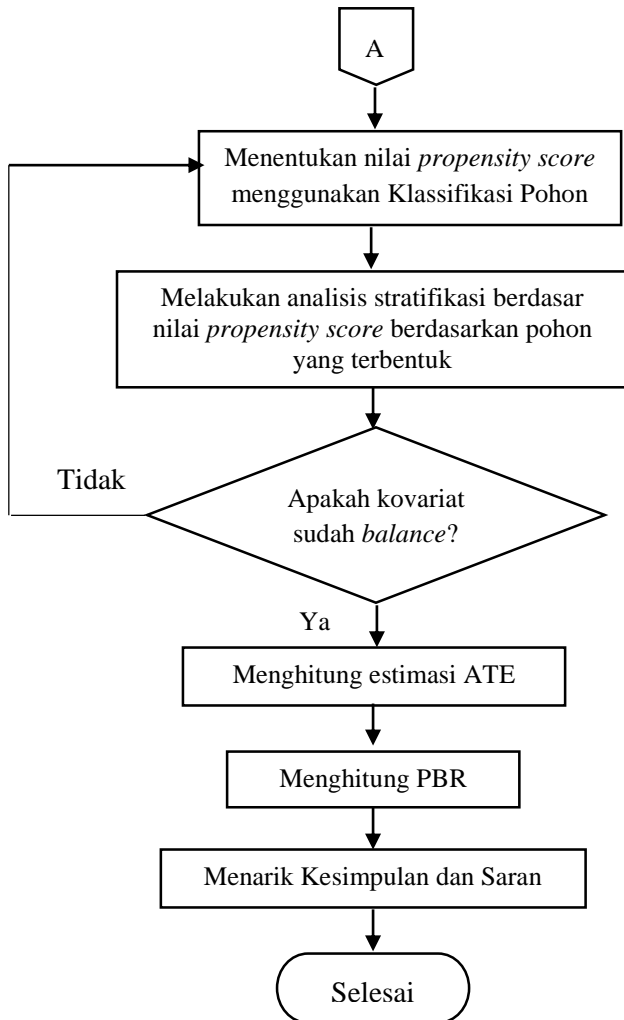
$$z_i = f(x_{1i}, x_{2i}, \dots, x_{ni}).$$

3. Menghitung nilai *propensity score* dengan menggunakan CART.
4. Melakukan analisis *stratification* dengan berdasarkan *propensity score* dari metode CART yang diperoleh.
5. Menguji apakah *propensity score* memenuhi asumsi *balance*.
 - a) Pengujian *balance* kovariat menggunakan uji-t untuk data kontinyu.
 - b) Pengujian *balance* kovariat menggunakan menggunakan uji-z untuk data kategorik.
6. Menghitung nilai estimasi *Average Treatment Effect* (ATE) dari *propensity score* untuk perlakuan dan kontrol.
7. Melakukan evaluasi kinerja PSS dengan menghitung *Percent Bias Reduction* (PBR).
8. Membuat kesimpulan dan saran.

3.5 Diagram Alir

Diagram Alir penelitian yang digunakan dalam langkah-langkah adalah sebagai berikut.





Gambar 3.1 Diagram Alir Penelitian

(Halaman ini sengaja dikosongkan)

BAB IV ANALISIS DAN PEMBAHASAN

Bab ini membahas analisis *propensity score* menggunakan variabel *confounding* secara *stratification* dengan metode Klasifikasi Pohon, kemudian setelah itu dilakukan pengecekan apakah variabel perlakuan dan kontrolnya sudah *balance* dan dilakukan evaluasi hasil sebelum dan sesudah dilakukan *stratification*.

4.1 Karakteristik Pasien Diabetes Melitus di RSUD Kabupaten Pasuruan

Analisis statistika deskriptif ini memiliki tujuan untuk memberikan gambaran (deskripsi) mengenai suatu data agar data yang tersaji menjadi mudah dipahami dan informatif. Statistika deskriptif menjelaskan berbagai karakteristik data seperti rata-rata (*mean*), simpangan baku (*standard deviation*), varians (*variance*), nilai minimum dan maksimum. Karakteristik pasien diabetes dapat dilihat secara deskriptif pada masing-masing variabel. Variabel data kontinu yang digunakan adalah variabel Usia (X_1) dan Lama Menderita (X_4) sedangkan variabel data kategorik yang digunakan adalah variabel Jenis Kelamin (X_2), Riwayat Genetik (X_3), Obesitas (X_5), Hipertensi (X_6) dan Aktivitas Olahraga (X_7). Berikut merupakan karakteristik pasien untuk data kontinu ditampilkan pada Tabel 4.1.

Tabel 4.1. Karakteristik Pasien Berdasarkan Usia dan Lama Menderita DM tipe 2.

Variabel (X)	Status Komplikasi	Mean (tahun)	Standar Deviasi (tahun)	Median (tahun)	Min (tahun)	Max (tahun)
Usia (X_1)	Komplikasi	52,88	15,03	51	29	77
	Tidak Komplikasi	53,47	14,89	51	29	77
Lama Menderita (X_4)	Komplikasi	4,06	2	4	1	7
	Tidak Komplikasi	4,06	1,99	4	1	7

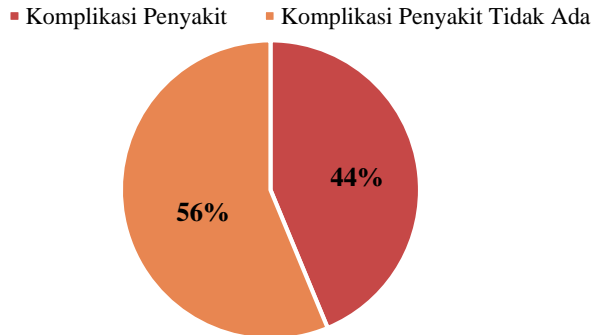
Tabel 4.1 memberi informasi bahwa hasil perhitungan statistika deskriptif menunjukkan usia dan lama menderita

memiliki nilai deviasi standar kurang dari nilai *mean*, artinya nilai *mean* tersebut telah representatif terhadap data variabel yang digunakan dalam penelitian ini karena dengan deviasi standar yang lebih kecil dari nilai *mean* menunjukkan persebaran data tidak menyebar sangat tinggi atau tidak terdapat nilai yang ekstrim yaitu nilai yang menyebabkan ada beberapa data yang *outlier* sehingga dapat dikatakan bahwa data pasien DM tipe 2 di RSUD Kabupaten pasuruan sudah homogen atau baik.

Dari perhitungan diatas pasien DM tipe 2 yang memiliki komplikasi penyakit di RSUD Kabupaten Pasuruan memiliki rata-rata usia yang lebih muda daripada yang tidak terkena komplikasi penyakit, hal ini mengindikasikan bahwa usia bukan menjadi tolok ukur pasien untuk menderita komplikasi penyakit, bukan berarti semakin tua akan banyak terkena komplikasi penyakit tetapi ada faktor-faktor yang menyebabkan mengapa usia yang lebih muda lebih banyak menderita komplikasi penyakit seperti gaya hidup yang tidak sehat, pola makan dan pengaruh etnis/ras dari pasien (Darsana,2014). Rentang usia pasien yang terkena DM tipe 2 dari yang termuda hingga tua memiliki usia 29-77 tahun, hal ini menunjukkan bahwa diabetes melitus dapat menyerang usia paruh baya hingga usia lanjut (lansia) namun juga dapat dimiliki oleh usia produktif. Median menunjukkan jika data dibagi menjadi 2, ada 50% penderita yang memiliki usia dibawah 51 tahun dan 50% sisanya memiliki usia diatas 51 tahun.

Pada Tabel 4.1 diatas juga dapat diketahui bahwa rata-rata pasien memiliki lama menderita DM tipe 2 di RSUD Pasuruan adalah selama 4,06 tahun, hal ini tergolong cukup lama, dari lama waktu ini dapat memunculkan dugaan yang dinyatakan oleh Depkes di tahun 2009 bahwa kebanyakan pasien DM sudah mengidap DM sebelum pasien tersebut mengetahui dan memeriksakan ke rumah sakit, sehingga ketika diperiksakan diagnosis yang didapat adalah bahwa pasien tersebut sudah lebih lama menderita DM dari perkiraannya. 50% pasien DM tipe 2 memiliki lama menderita kurang dari 4 tahun dan 50% sisanya lebih dari 4 tahun. Pasien DM tipe 2 di RSUD kabupaten Pasuruan paling sedikit sudah menderita penyakit ini selama 1 tahun dan paling lama sekitar 7 tahun.

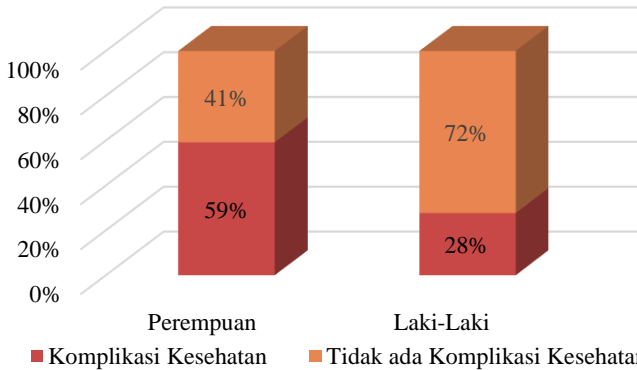
Gambaran statistika deskriptif status komplikasi penyakit pada pasien DM tipe 2 (Y) dapat ditampilkan pada Gambar 4.1.



Gambar 4.1 Persentase Status Komplikasi Penyakit Pada Pasien DM tipe 2.

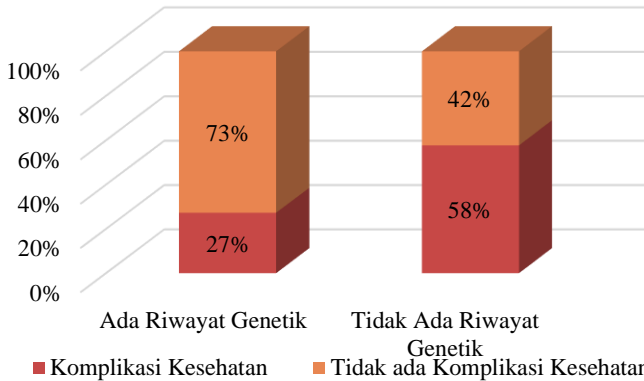
Gambar 4.1 memberi gambaran bahwa dari keseluruhan data penderita Diabetes Melitus tipe 2 di RSUD Kabupaten Pasuruan maka cenderung lebih banyak pasien yang tidak memiliki Komplikasi Penyakit. Namun selisihnya sangat sedikit jika dibandingkan dengan yang memiliki komplikasi penyakit sehingga perlu ditangani agar pasien diabetes tersebut tidak sampai terkena komplikasi penyakit, karena komplikasi penyakit yang terjadi dari waktu ke waktu dapat menyebabkan kerusakan berbagai sistem tubuh terutama syaraf dan pembuluh darah, serta mempercepat kematian (Kemenkes,2013). WHO, 2013 juga mengatakan bahwa diabetes adalah ibu dari segala penyakit karena merupakan penyakit yang dapat memengaruhi banyak organ dalam tubuh. Kadar gula darah yang tinggi pada penderita diabetes dapat menyebabkan banyak komplikasi diabetes. Diabetes memiliki hubungan dengan kadar gula darah, tekanan darah, dan kolesterol. Oleh karena itu, jika penderita diabetes yang tidak ingin menderita komplikasi diabetes, Diharuskan untuk mengontrol ketiga hal tersebut. Menerapkan pola makan sehat, melakukan olahraga teratur, berhenti merokok dan konsumsi minuman alkohol, dapat membantu untuk terhindar dari komplikasi diabetes. Untuk melihat persentase jenis kelamin pasien DM tipe 2 yang menderita

komplikasi penyakit di RSUD Kabupaten Pasuruan dapat dilihat pada Gambar 4.2.



Gambar 4.2 Persentase Jenis Kelamin dengan Status Komplikasi Penyakit.

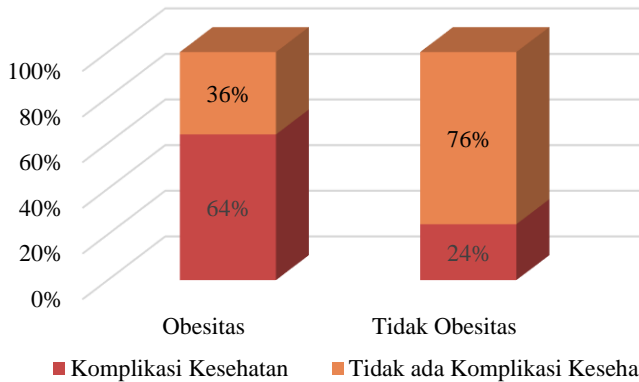
Gambar 4.2 menunjukkan bahwa pasien yang berjenis kelamin perempuan memiliki komplikasi penyakit lebih tinggi daripada pasien yang berjenis kelamin laki-laki, artinya perempuan cenderung lebih memiliki peluang untuk memiliki komplikasi penyakit daripada laki-laki.



Gambar 4.3 Persentase Riwayat Genetik dengan Status Komplikasi Penyakit.

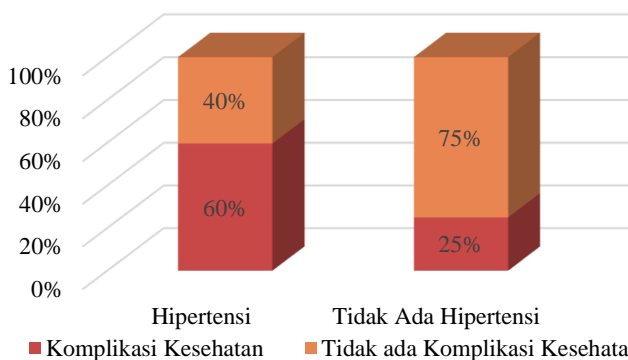
Gambar 4.3 menunjukkan bahwa pada kenyataannya di RSUD Kabupaten Pasuruan, Lebih sedikit pasien DM tipe 2 yang

memiliki riwayat genetik mengalami komplikasi kesehatan daripada yang tidak memiliki riwayat genetik, secara tidak langsung memberi kesimpulan bahwa Riwayat Genetik tidak begitu memberi pengaruh terhadap komplikasi dalam penyakit DM tipe 2.



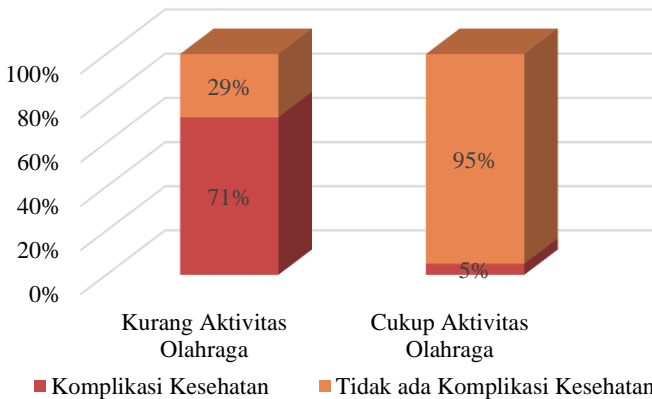
Gambar 4.4 Persentase Obesitas dengan Status Komplikasi Penyakit.

Gambar 4.4 memberi gambaran bahwa pasien DM tipe 2 yang cenderung memiliki komplikasi penyakit dan mengalami Obesitas lebih banyak daripada yang tidak memiliki Obesitas, hal ini menunjukkan bahwa obesitas cukup mempengaruhi terjadinya komplikasi penyakit dan diduga banyak memberi pengaruh untuk menimbulkan komplikasi penyakit.



Gambar 4.5 Persentase Status Hipertensi dengan Status Komplikasi Penyakit.

Gambar 4.5 menunjukkan bahwa memiliki Hipertensi akan terkena resiko besar untuk terkena komplikasi penyakit pada kasus DM tipe 2, sehingga disebut salah satu faktor yang dianggap berpengaruh terhadap status komplikasi penyakit. Terbukti dengan, pasien DM tipe 2 yang memiliki komplikasi penyakit sekaligus mengalami Hipertensi lebih banyak daripada yang tidak mengalami Hipertensi. Seorang pasien dinyatakan mengalami Hipertensi ketika hasil tekanan darah sistolik menunjukkan hasil lebih dari 130 mmHg, sedangkan hasil tekanan darah diastolik menunjukkan hasil lebih dari 80 mmHg. Pada penelitian ini, Hipertensi hanya dilihat berdasarkan hasil rekam medis untuk tekanan darah sistolik.



Gambar 4.6 Persentase Aktivitas Olahraga dengan Status Komplikasi Penyakit.

Gambar 4.6 memberi informasi bahwa bahwa salah satu dugaan timbulnya komplikasi penyakit adalah kurangnya melakukan aktivitas olahraga benar adanya, dibuktikan dengan tingginya pasien yang menderita komplikasi kesehatan karena kurang melakukan aktivitas olahraga memiliki selisih besar daripada yang melakukan aktivitas olahraga dan terkena komplikasi penyakit, padahal seperti yang diketahui aktivitas olahraga dapat menurunkan kadar glukosa darah. Dari nilai proporsi pasien DM tipe 2 tersebut dapat dilihat bahwa ketertarikan dan minat pasien DM tipe 2 terhadap olahraga masih rendah

sehingga perlu ditingkatkan lagi karena Aktivitas Olahraga cukup memberi pengaruh terhadap kesehatan pasien DM tipe 2.

4.2 Pemilihan Variabel *Confounding*

Langkah awal dalam melakukan analisis *propensity score stratification* adalah terlebih dahulu memilih variabel *confounding* yang dapat dilakukan berdasarkan teori dan menunjukkan bukti secara empiris dengan melihat hubungan antar variabel. Statistik uji yang dapat digunakan untuk melihat hubungan antar variabel adalah uji *chi-square*. Selain dilakukan pengujian hubungan antar kovariat untuk menentukan variabel *confounding*, maka sebelumnya dilakukan pengujian untuk mengetahui hubungan kovariat (X) yang berpotensi sebagai *confounding* dan variabel Komplikasi Kesehatan (Y).

Dalam analisis ini dijelaskan bahwa faktor-faktor yang memengaruhi komplikasi penyakit DM tipe 2 adalah faktor Usia (X_1), Jenis Kelamin (X_2), Riwayat Genetik (X_3), Lama Menderita (X_4), Obesitas (X_5), Hipertensi (X_6) dan Aktivitas Olahraga (X_7). Faktor-faktor tersebut merupakan faktor risiko dari kejadian komplikasi penyakit (Y) DM tipe 2. Selanjutnya dilakukan uji *chi-square* untuk melihat hubungan dari tiap variabel dan menentukan variabel mana yang disebut variabel *confounding*, variabel dapat dikatakan akan *confounding* bila tolak H_0 dengan taraf signifikansi $\alpha = 5\%$ bila dibandingkan dengan nilai *p-value* yang didapat seperti yang tercantum di dalam daftar pustaka, untuk melihat pengujian *chi-square* secara lengkap dapat dilihat pada Lampiran 4.

Uji *chi-square* yang dilakukan pada variabel Jenis Kelamin didapati ditampilkan pada Tabel 4.2.

Tabel 4.2 Uji *chi-square* antara variabel Jenis Kelamin dengan variabel lainnya.

Variabel	χ^2	<i>P-value</i>	Keputusan
$X_3 * X_2$	0,3991	0,5276	Gagal Tolak H_0
$X_5 * X_2$	4,1877	0,0407	Tolak H_0
$X_6 * X_2$	0,3991	0,5276	Gagal Tolak H_0
$X_7 * X_2$	7,0614	0,0079	Tolak H_0

Tabel 4.2 diatas memberi informasi bahwa variabel Obesitas dan Aktivitas Olahraga Tolak H_0 terhadap uji *chi-square* maka 2

variabel tersebut dapat diduga menjadi variabel *confounding*. Artinya variabel tersebut memiliki pengaruh satu sama lain dengan variabel jenis kelamin. Selanjutnya dilakukan uji *chi-square* terhadap Riwayat Genetik yang dapat ditampilkan di Tabel 4.3.

Tabel 4.3 Uji *chi-square* antara variabel Riwayat Genetik dengan variabel lainnya.

Variabel	χ^2	<i>P-value</i>	Keputusan
X ₂ *X ₃	0,3991	0,5276	Gagal Tolak H ₀
X ₅ *X ₃	0,3991	0,5276	Gagal Tolak H ₀
X ₆ *X ₃	2,4834	0,1150	Gagal Tolak H ₀
X ₇ *X ₃	10,1467	0,0014	Tolak H ₀

Tabel 4.3 memberi informasi bahwa variabel Aktivitas Olahraga Tolak H₀ terhadap uji *chi-square* maka variabel tersebut diduga menjadi variabel *confounding* terhadap Riwayat Genetik. Selanjutnya dilakukan uji *chi-square* terhadap variabel Obesitas dtampilkan pada Tabel 4.4.

Tabel 4.4 Uji *chi-square* antara variabel Obesitas dengan variabel lainnya.

Variabel	χ^2	<i>P-value</i>	Keputusan
X ₂ *X ₅	4,1877	0,0407	Tolak H ₀
X ₃ *X ₅	0,3991	0,5276	Gagal Tolak H ₀
X ₆ *X ₅	3,4633	0,0627	Gagal Tolak H ₀
X ₇ *X ₅	7,4330	0,0064	Tolak H ₀

Pada Tabel 4.4 dapat dilihat bahwa variabel Obesitas dan Aktivitas Olahraga Tolak H₀ terhadap uji *chi-square* maka 2 variabel tersebut dapat diduga menjadi variabel *confounding*. Selanjutnya dilakukan uji *chi-square* terhadap variabel Hipertensi seperti di Tabel 4.5.

Tabel 4.5 Uji *chi-square* antara variabel Hipertensi dengan variabel lainnya.

Variabel	χ^2	<i>P-value</i>	Keputusan
X ₂ *X ₆	0,3991	0,5276	Gagal Tolak H ₀
X ₃ *X ₆	2,4834	0,1150	Gagal Tolak H ₀
X ₅ *X ₆	3,4633	0,0627	Gagal Tolak H ₀
X ₇ *X ₆	5,5433	0,0186	Tolak H ₀

Tabel 4.5 memberi informasi sama seperti uji *chi-square* pada variabel Riwayat Genetik dengan kovariat lainnya yaitu variabel Aktivitas Olahraga Tolak H_0 terhadap uji *chi-square* maka variabel tersebut dapat diduga menjadi variabel *confounding*. Selanjutnya dilakukan uji *chi-square* terhadap variabel Aktivitas Olahraga dengan variabel lain dan didapati hasilnya seperti Tabel 4.6.

Tabel 4.6 Uji *chi-square* antara variabel Aktivitas Olahraga dengan variabel lainnya.

Variabel	χ^2	<i>P-value</i>	Keputusan
$X_2 * X_7$	7,0614	0,0079	Tolak H_0
$X_3 * X_7$	10,1467	0,0014	Tolak H_0
$X_5 * X_7$	7,4330	0,0064	Tolak H_0
$X_6 * X_7$	5,5433	0,0186	Tolak H_0

Tabel 4.6 memberi informasi bahwa variabel Aktivitas Olahraga ketika diuji *chi-square* Tolak H_0 terhadap semua variabel sehingga variabel tersebut ditetapkan menjadi variabel *confounding*. Untuk menguatkan argumen tersebut maka dilakukan uji *chi-square* terhadap variabel respon yang dapat dilihat di Tabel 4.7.

Tabel 4.7 Uji *chi-square* antara variabel Komplikasi Penyakit dengan variabel lainnya.

Variabel	χ^2	<i>P-value</i>	Keputusan
$X_2 * Y$	9,6874	0,0019	Tolak H_0
$X_3 * Y$	8,9617	0,0028	Tolak H_0
$X_5 * Y$	15,0866	0,0001	Tolak H_0
$X_6 * Y$	11,6044	0,0007	Tolak H_0
$X_7 * Y$	41,8395	0,0000	Tolak H_0

Tabel 4.7 diatas dapat menguatkan bahwa variabel Aktivitas Olahraga (X_7) adalah variabel yang dapat dijadikan *confounding* karena variabel tersebut selain memiliki hubungan yang signifikan dengan variabel komplikasi penyakit (Y) juga memiliki hubungan yang signifikan dengan variabel kovariat lainnya. Digunakan notasi (Z) untuk menandai bahwa variabel Aktivitas Olahraga adalah *confounding*.

4.3 Estimasi Nilai *Propensity Score* menggunakan Klasifikasi Pohon

Estimasi nilai *propensity score* dilakukan setelah menentukan variabel *confounding* (Z). Klasifikasi karena variabel *confounding* yang terpilih terdiri atas dua kategori yang nantinya menjadi perlakuan dan kontrol. Nilai *propensity score* menyatakan seberapa besar peluang seseorang untuk terkena suatu perlakuan berdasarkan kovariat yang diamati. Selanjutnya variabel respon yang digunakan adalah variabel *confounding* yang telah ditentukan di langkah sebelumnya yaitu variabel Aktivitas Olahraga (X_7) yang selanjutnya dinotasikan (Z) dan variabel prediktor yang digunakan adalah variabel Usia (X_1), Jenis Kelamin (X_2), Riwayat Genetik (X_3), Lama Menderita (X_4), Obesitas (X_5) dan status Hipertensi (X_6). *Output* dan perhitungan lebih jelas dalam pembentukan klasifikasi pohon ini secara lengkap dapat dilihat pada Lampiran 5.

Sebelum mendapatkan nilai *propensity score* menggunakan Pohon Klasifikasi, maka perlu didapatkan pengambilan sampel secara optimum guna keperluan pembuatan pohon yang menggunakan data *training*.

4.3.1 Penentuan Sampel Optimum

Penentuan sampel optimum dalam Pohon Klasifikasi digunakan dengan data *training*, ukuran ketepatan klasifikasi yang digunakan dapat ditinjau dari besar akurasi (%), *sensifity* (%), dan *specificity* (%) seperti yang dicantumkan pada tinjauan pustaka. Pada penelitian ini didapatkan sampel untuk data *training* dan data *testing* yang disajikan dalam Tabel 4.8.

Tabel 4.8 Pengambilan sampel optimum untuk Klasifikasi Pohon.

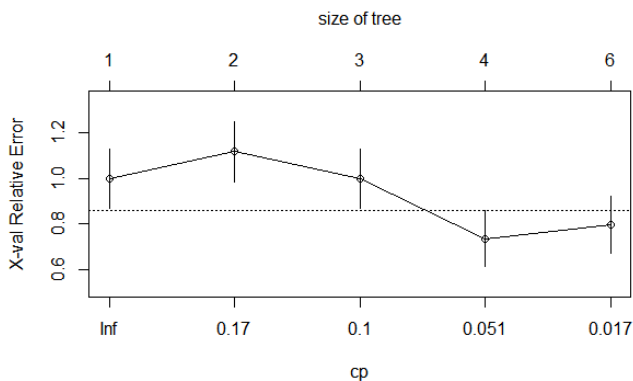
Persentase Data	Data Training	Data Testing	Akurasi	Sensifitas	Spesifitas
65% : 35%	62	34	0.7097	0.3846	0.9444
70% : 30%	67	29	0.7612	0.5714	0.8974
75% : 25%	72	24	0.7778	0.6667	0.8571
80% : 20%	77	19	0.7532	0.5000	0.9333
85% : 15%	82	14	0.7927	0.8529	0.7500

Tabel 4.8 memberi kesimpulan bahwa yang akan dipilih untuk pengambilan sampel terbaik menggunakan persentase 85% : 15%

dimana 85% merupakan data *training* dan 15% sisanya akan digunakan data *testing*, hal ini dipilih karena memiliki akurasi paling tinggi dimana akurasi adalah ukuran untuk mengukur ketepatan hasil klasifikasi secara keseluruhan, semakin tinggi nilainya maka metode klasifikasi yang digunakan akan semakin baik. Dan nilai sensitifitas yang paling tinggi juga karena *sensifity* menunjukkan kelas positif yang benar dibanding total data yang diklasifikasikan sebagai kelas positif. Kemudian juga karena nilai spesifitasnya paling rendah dimana *specifity* adalah ukuran yang menyatakan kelas negatif benar dibanding dengan total data yang diklasifikasikan sebagai kelas negative, jadi semakin rendah nilainya akan semakin baik. Ulasan lebih lengkap dapat dilihat pada lampiran.

4.3.2 Pembentukan Pohon Klasifikasi

Pembentukan Pohon Klasifikasi dilakukan setelah memilih sampel optimum yang digunakan untuk data *training*, keuntungan mencari *propensity score* dengan Pohon Klasifikasi adalah klasifikasi yang nantinya terbentuk tidak dimodelkan secara eksplisit seperti dalam regresi logistik, namun simpul terminal pohon yang terbentuk akan secara otomatis membentuk strata, sebelumnya dilakukan cek terhadap jumlah terminal optimum yang terbentuk dan didapati hasil seperti di Gambar 4.7.

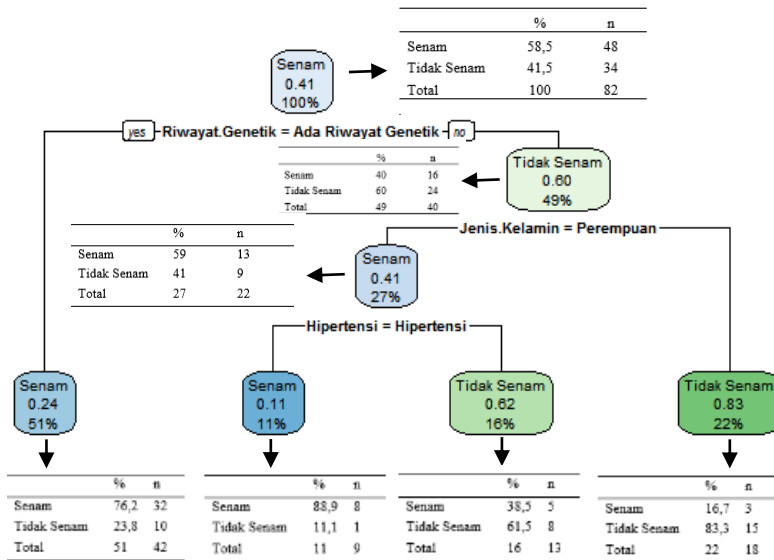


Gambar 4.7 Plot CP pada Pohon Klasifikasi yang terbentuk

Gambar 4.7 adalah gambaran yang menunjukkan bahwa nilai CP yang memiliki error paling rendah berada pada pohon yang

memiliki *terminal node* sejumlah 4 yaitu *error* nya sebesar 0,051 lebih kecil dari nilai *error* lainnya, oleh karena itu dapat disimpulkan Pohon Klasifikasi optimum yang akan terbentuk memiliki 4 terminal dan diindikasikan akan terjadi 4 strata yang terbentuk dalam Pohon Klasifikasi ini. Adapun nilai dari jumlah variabel *treatment* dan terkontrol nantinya akan dihitung berdasar proporsi dalam setiap terminal. Selanjutnya dibentuk Pohon Klasifikasi, pada awalnya pohon klaifikasi sebelum dilakukan pemangkasan menghasilkan 6 *terminal node* seperti yang bisa dilihat pada lampiran 4, namun setelah dilakukan pemangkasan untuk pohon yang lebih optimal didapatkan pohon yang tergambar seperti di Gambar 4.8.

Pohon Keputusan Optimal untuk Diabetes Melitus Tipe 2



Gambar 4.8 Pohon Klasifikasi yang terbentuk dari DM tipe 2

Gambar 4.8 memberi informasi bahwa ada 4 terminal optimal yang terbentuk atau ada 4 strata yang dapat dibentuk dalam kasus DM tipe 2 ini. Pemilah pertama adalah Riwayat Genetik yang memiliki interpretasi bahwa Riwayat Genetik memiliki nilai heterogenitas yang paling tinggi dibanding variabel lainnya yang

dapat dipakai untuk pemilah awal, nilai proporsi pasien yang memiliki Riwayat Genetik adalah 51% dan yang tidak memiliki Riwayat Genetik 49% yang kemudian dipilah lagi dengan pemilah kedua. Pemilah kedua yaitu variabel Jenis Kelamin pada *node* yang tidak memiliki Riwayat Genetik, hal ini memberi keimpulan bahwa terdapat keheterogenan yang tinggi antara pasien DM tipe 2 yang berjenis kelamin perempuan dan laki-laki. Kemudian pada *node* Jenis Kelamin perempuan dapat dibedakan lagi menggunakan variabel Hipertensi yang memiliki proporsi data 27%, artinya terdapat kelompok yang berbeda sehingga dibedakan pengklasifikasian antara perempuan yang memiliki Hipertensi dan tidak.

Pada strata 1 yang terbentuk memiliki nilai proporsi data sebesar 51% dengan kategori klasifikasi kovariat yang masuk didalamnya adalah ada Riwayat Genetik dan bila dilihat dari gambar diketahui proporsi yang lebih dominan dalam strata ini adalah proporsi yang banyak melakukan aktivitas kegiatan ditandai dengan simbol senam.

Strata 2 yang terbentuk memiliki nilai proporsi data sebesar 11% dengan kategori klasifikasi kovariat yang masuk didalamnya adalah tidak ada Riwayat Genetik, memiliki Jenis Kelamin perempuan dan memiliki Hipertensi. Nilai proporsi yang lebih dominan dalam strata ini adalah proporsi yang banyak melakukan aktivitas kegiatan ditandai dengan simbol senam.

Kelompok strata 3 yang terbentuk memiliki nilai proporsi data sebesar 16% dengan kategori klasifikasi kovariat yang masuk didalamnya adalah tidak ada Riwayat Genetik, memiliki Jenis Kelamin perempuan dan tidak memiliki Hipertensi. Dari gambar dapat diketahui bahwa proporsi yang lebih dominan dalam strata ini adalah proporsi yang kurang melakukan aktivitas kegiatan ditandai dengan simbol tidak senam.

Strata terakhir yang terbentuk adalah kelompok strata 4 dan memiliki nilai proporsi data keseluruhan sebesar 22% dengan kategori klasifikasi kovariat yang masuk didalamnya adalah tidak ada Riwayat Genetik dan berjenis Kelamin laki-laki. Dari gambar dapat diketahui bahwa proporsi yang lebih dominan dalam strata

ini adalah proporsi yang kurang melakukan aktivitas kegiatan ditandai dengan simbol tidak senam.

4.3.3 Analisis *Propensity Score Stratification* Berdasar Pohon Klasifikasi yang Terbentuk

Pada Pohon Klasifikasi nilai *propensity score* secara strata sudah terbentuk pada terminal akhirnya, dalam penghitungan berikutnya proporsi kelompok perlakuan dan kontrol yang akan menjadi nilai dari *propensity score* itu sendiri, proporsi yang terbentuk disimbolkan dengan $f(x_j)$ sedangkan nilai *propensity score* disimbolkan dengan $e(x_j)$ sehingga nilai *propensity score* akan diperoleh jika proporsi kelompok perlakuan dan kontrol dari setiap strata telah didapatkan. Hasil estimasi parameter nilai *propensity score* dengan 4 strata dari Pohon Klasifikasi yang terbentuk dapat dilihat pada Tabel 4.9.

Tabel 4.9 Nilai proporsi yang didapat dari setiap strata.

Variabel	$f(x_j)$			
	Strata 1	Strata 2	Strata 3	Strata 4
<i>Treatment</i>	0,7619	0,8889	0,3846	0,1667
Kontrol	0,2381	0,1111	0,6154	0,8333

Tabel 4.9 memberi informasi nilai proporsi data dalam variabel Aktivitas Olahraga yang terbentuk dalam 4 strata. Dari 4 strata ini dapat diketahui bahwa strata 1 berisi 41 data dengan jumlah yang banyak melakukan Aktivitas Olahraga sebanyak 31 orang atau proporsinya 0,7619 sedangkan yang sedikit melakukan Aktivitas Olahraga sebanyak 10 orang dengan proporsi 0,2381. Strata 2 berisi 9 data dengan jumlah yang banyak melakukan Aktivitas Olahraga sebanyak 9 orang atau proporsinya 0,8889 sedangkan yang sedikit melakukan Aktivitas Olahraga sebanyak 1 orang dengan proporsi 0,1111. Pada strata 3 berisi 13 data dengan jumlah yang banyak melakukan Aktivitas Olahraga sebanyak 5 orang atau proporsinya 0,3846 sedangkan yang sedikit melakukan Aktivitas Olahraga sebanyak 8 orang dengan proporsi 0,6154. Strata terakhir yang terbentuk berisi 18 data dengan jumlah yang banyak melakukan Aktivitas Olahraga sebanyak 3 orang atau proporsinya 0,1667 sedangkan yang sedikit melakukan Aktivitas Olahraga sebanyak 15 orang dengan proporsi 0,8333. Setelah

diperoleh estimasi fungsi klasifikasi dalam bentuk proporsi, maka langkah selanjutnya adalah mengestimasi nilai *propensity score* dari setiap strata dengan menggunakan rumus pada Persamaan (2.18) estimasi nilai *propensity score* dapat ditunjukkan pada Tabel 4.10.

Tabel 4.10 Nilai *propensity score* yang didapat dari setiap strata.

Variabel	$e(x_j)$			
	Strata 1	Strata 2	Strata 3	Strata 4
<i>Treatment</i>	0,6818	0,7087	0,5950	0,5416
Kontrol	0,5592	0,5277	0,6492	0,6971

Tabel 4.10 memberi kesimpulan bahwa nilai yang didapat dari variabel yang banyak melakukan aktivitas kegiatan dan sedikit melakukan aktivitas kegiatan tidak memiliki selisih yang banyak, pada strata 1 dijelaskan skor kecenderungan pasien tidak melakukan aktivitas olahraga berdasarkan kovariat riwayat genetik adalah 0,5592, pada strata 2 skor kecenderungan pasien tidak melakukan aktivitas olahraga berdasarkan kovariat riwayat genetik, jenis kelamin dan hipertensi adalah 0,5277, begitu juga interpretasi untuk strata 3 dan 4. Langkah selanjutnya adalah melihat ketepatan prediksi dari PSS berdasarkan Pohon Klasifikasi.

4.3.4 Tingkat Ketepatan Prediksi *Propensity Score Stratification* berdasarkan Pohon Klasifikasi

Pada langkah sebelumnya telah dilakukan pembagian nilai PSS-Pohon Klasifikasi ke dalam strata 1, 2, 3 dan 4. Selanjutnya dilakukan evaluasi hasil ketepatan prediksi berdasarkan nilai *propensity score* yang terbentuk secara keseluruhan dalam Pohon Klasifikasi menggunakan Akurasi, Sensifitas dan Spesifitas yang ditunjukkan dalam Tabel 4.11.

Tabel 4.11 Evaluasi Hasil Prediksi *Propensity Score Stratification* dalam Pohon Klasifikasi Optimal yang Terbentuk.

Ketepatan Klasifikasi		
Akurasi	Sensifitas	Spesifikasi
0,5714	0,3333	0,7500

Apabila dilihat tingkat akurasi secara keseluruhan, maka nilai akurasi prediksi secara keseluruhan yaitu sebesar 0,57 atau 57

persen. Hal ini menunjukkan bahwa *propensity score stratification* berdasar Pohon Klasifikasi dapat memprediksi secara benar bahwa pasien DM tipe 2 yang memiliki komplikasi penyakit dipengaruhi oleh aktivitas kegiatan.

4.4 Uji *Balance Covariate*

Langkah awal sebelum menghitung signifikansi hasil ATE adalah mengecek keseimbangan kovariat pada variabel *confounding* antara kelompok *treatment* dan kontrol. Pada penelitian ini untuk menguji perbedaan antara kelompok *treatment* dan kontrol pada masing-masing kovariat dapat menggunakan uji kesamaan dua rata-rata dan uji proporsi (persamaan 2.19 dan 2.23). Pada penelitian ini untuk menguji perbedaan antara kelompok *treatment* dan kontrol digunakan uji-z untuk variabel kategorik yaitu Riwayat Genetik, Jenis Kelamin, Obesitas dan Hipertensi, sedangkan uji-t digunakan untuk menguji kesamaan rata-rata Usia pasien dan lama pasien menderita DM tipe 2.

Pada uji kovariat *balance* ini α yang digunakan untuk uji-z dan uji-t ini masing masing adalah 1%. Proses pengujian *balance* pada masing-masing kovariat sebelum dan sesudah dilakukan strata selengkapnya dapat dilihat di Lampiran 7 dan disajikan secara ringkas di Tabel 4.12.

Tabel 4.12 Nilai Pengujian *Balance Covariate* sebelum dan sesudah dilakukan PSS.

Strata ke-k	Jumlah Data		<i>P-value</i>					
	<i>Tk</i>	<i>Ck</i>	X_1	X_2	X_3	X_4	X_5	X_6
Sebelum	56	40	0,518	0,008	0,001	0,283	0,006	0,019
1	32	10	0,382	0,010	*	0,853	0,047	0,374
2	8	1	0,382	*	*	0,853	0,571	*
3	5	8	0,537	*	*	0,755	0,725	*
4	3	15	0,930	*	*	0,672	1,000	0,527

* memiliki arti *NaN (Not a Number)* karena adanya error dalam pengujian dan nilai yang dihasilkan tidak dapat ditentukan.

Tabel 4.12 didapati bahwa sebelum dilakukan pengklasifikasian kedalam 4 strata nilai variabel yang *balance* antara kelompok *treatment* dan kontrolnya adalah variabel Usia

(X_1), Lama Menderita (X_4) dan status Hipertensi (X_6). Dari hasil diatas didapati bahwa variabel Riwayat Genetik tidak muncul nilai p -*value*nya sejak strata pertama karena variabel ini dalam Pohon Klasifikasi merupakan pemilah pertama, maka dapat disimpulkan ketika nilai p -*value* tidak muncul, atau bernilai NaN, hal ini sesuai dengan perkataan (Luellen, 2009) bahwa dalam suatu Pohon Klasifikasi yang telah terbentuk dapat dicari uji *balance* nya untuk setiap kovariat dari strata yang terbentuk, kecuali kovariat yang menjadi *node* Pohon Klasifikasi tersebut karena nilai kovariatnya akan menjadi konstan sehingga akan menjadi error bila akan dilakukan uji proporsi atau uji kesamaan rata-rata, sehingga keseimbangan kovariat ini dapat diabaikan. Namun bila dilihat semua kovariatnya setelah stratifikasi hasil p -*value* dari uji *balance* semua nilainya gagal tolak H_0 artinya semua kovariat telah *balance* dan tidak perlu dimodelkan Pohon Klasifikasi ulang.

4.5 Estimasi *Average Treatment Effect* (ATE)

Langkah selanjutnya setelah mengecek keseimbangan kovariat adalah menghitung signifikansi hasil ATE. ATE akan digunakan untuk mengetahui seberapa besar pengaruh Aktivitas Olahraga (Z) terhadap kejadian komplikasi penyakit (Y) pada saat pengaruh dari variabel lain (kovariat) sudah direduksi menjadi variabel *confounder* (Z). Variabel *confounder* pada penelitian ini adalah variabel aktivitas olahraga. Taraf signifikansi (α) yang digunakan sebesar 0,05 dan statistik uji yang digunakan dapat dilihat seperti yang ada dalam tinjauan pustaka pada persamaan rumus (2.31) setelah itu diperoleh hasil estimasi nilai *Average Treatment Effect* (ATE) secara lengkap dapat dilihat pada Lampiran 8 dan dapat ditunjukkan pada tabel 4.13.

Tabel 4.13 Hasil Estimasi ATE dan Standar Error *Propensity Score Stratification* berdasar Pohon Klasifikasi.

Strata ke-k	$\hat{\theta}_k$	$\hat{\theta}$	SE($\hat{\theta}$)	thitung	df	p - <i>value</i>
1	0,2817					
2	0,0823					
3	0,0634	0,5737	0,1159	4,9501	74	0,0000*
4	0,1463					

Tabel 4.13 memberi informasi bahwa estimasi efek perlakuan dapat diketahui pada setiap strata, pada strata 1 efek aktivitas olahraga terhadap komplikasi kesehatan adalah sebesar 0,2817 atau peluang pasien DM tipe 2 yang banyak melakukan Aktivitas Olahraga untuk tidak terkena komplikasi penyakit sebesar 0,2817 kali dibandingkan pasien yang sedikit melakukan aktivitas olahraga. Dan begitu seterusnya interpretasi untuk besarnya efek perlakuan Aktivitas Olahraga terhadap komplikasi penyakit di strata yang lain. Secara keseluruhan 4 strata memberikan efek rata-rata sebesar 0,5737 dengan standar error sebesar 0,1159. Berdasarkan strata yang terbentuk maka standar error yang cenderung kecil tersebut menunjukkan bahwa pembentukan 4 strata dapat mereduksi efek bias yang terbesar.

4.6 Perbandingan *Percent Bias Reduction (PBR)* pada Setiap Strata

Analisis *propensity score stratification* dengan Pohon Klasifikasi yang sudah terbentuk sudah dilakukan *balance* kovariat pada strata 1,2,3 dan 4, Selanjutnya, untuk menilai apakah metode PSS sudah baik atau tidak, dapat dilihat dari seberapa besar bias yang mampu direduksi oleh metode PSS tersebut, nilai tersebut dapat dihitung dan disajikan pada Tabel 4.14.

Tabel 4.14 *Percent Bias Reduction PSS* Menggunakan Pohon Klasifikasi.

Strata ke-k	Rata-Rata Kelompok		Nilai Bias	PBR
	<i>Treatment</i>	<i>Control</i>		
Sebelum Strata	0,6418	0,6027	0,0392	28,04%
Sesudah Strata	0,6684	0,6403	0,0282	

Tabel 4.14 memberikan informasi bahwa nilai bias yang didapat sebelum dilakukan stratifikasi menggunakan Pohon Klasifikasi adalah sebesar 0,0392 sedangkan nilai bias yang didapat sesudah stratifikasi adalah sebesar 0,0282. Hal ini menunjukkan bahwa nilai bias yang dihasilkan setelah stratifikasi lebih kecil dibandingkan nilai bias sebelum stratifikasi yaitu sebesar 0,0109. Berdasarkan keefektifan metode Pohon Klasifikasi, metode PSS dengan Pohon Klasifikasi mampu mereduksi bias sebesar 28,04%.

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. Hasil estimasi variabel *confounding* didapati variabel Aktivitas Olahraga sebagai *counfounder*, artinya Aktivitas Olahraga merupakan faktor penentu yang keberadaanya mempengaruhi pasien untuk terjangkit komplikasi penyakit atau tidak.
2. Pohon Klasifikasi optimal yang terbentuk menggunakan pengambilan sampel dengan 85% data *training* dan 15% data *testing*, dari Pohon Klasifikasi tersebut dapat diketahui bahwa secara berturut turut faktor yang paling besar memberi pengaruh adalah Jenis Kelamin, Riwayat Genetik, dan Hipertensi, dengan ketepatan hasil prediksi 57%.
3. Empat variabel paling berpengaruh mengindikasi *terminal node* artinya bahwa jumlah strata untuk data DM tipe 2 ini berjumlah 4.
4. Estimasi *propensity score* menjelaskan bahwa pasien pada strata 4 memiliki skor kecenderungan lebih tinggi daripada strata lainnya dalam tidak melakukan aktivitas olahraga yang cukup, sedangkan skor kecenderungan terendah dalam tidak melakukan aktivitas olahraga yang cukup adalah pasien pada strata 2.
5. Signifikansi hasil *Average Treatment of Effect* adalah variabel aktivitas olahraga cukup memberi pengaruh secara signifikan terhadap timbulnya komplikasi penyakit pada pasien DM tipe 2.
6. Dari penelitian ini juga dapat diketahui bahwa 4 strata yang terbentuk dalam Klasifikasi Pohon berhasil mereduksi lebih dari seperempat total nilai bias atau kesalahan mendapatkan hasil prediksi, artinya kelompok data yang dipisah-pisahkan dalam strata menggunakan Pohon Klasifikasi tersebut cukup baik dalam menyeimbangkan atau menyamakan kelompok yang terkena komplikasi penyakit atau tidak, sehingga dapat

mengurangi nilai ketidak acakan dalam melakukan survey pasien yang terkena DM tipe 2 di RSUD Kabupaten Pasuruan.

5.2 Saran

Berdasarkan hasil analisis data dan pembahasan diperoleh hasil bahwa banyak melakukan aktivitas olahraga dapat mengurangi peluang terjangkitnya komplikasi penyakit. Oleh karena itu, untuk meminimalkan resiko terkena komplikasi penyakit, maka sebaiknya banyak melakukan aktivitas olahraga. hal lain yang dapat disarankan untuk penelitian selanjutnya, yaitu :

1. Dilakukan survei untuk penambahan data lagi terutama menambah banyaknya kelompok kontrol, agar kelompok *treatment* dan kontrol yang dipasangkan akan *balance*.
2. Melakukan penghitungan *propensity score stratification* dengan menggunakan metode *machine learning* yang lain, untuk mencari nilai reduksi bias yang lebih tinggi.
3. Pengujian kovariat *balance* dengan menggunakan pendekatan *propensity score stratification* dianggap kurang cocok bila dalam metode Klasifikasi Pohon, karena pasti akan mengabaikan kovariat yang menjadi *terminal node* sehingga nilai kovariat tersebut konstan dan error jika dilakukan pengujian, oleh karena itu dapat dicoba dengan menggunakan pendekatan lain, seperti *matching*, *covariance adjustment* atau *weighting*.

DAFTAR PUSTAKA

- Adityaningrum, A., Otok, B. W., & Fithriasari, K. (2016). *Estimasi Propensity Score Matching Menggunakan Pendekatan Multivariate Adaptive Regression Splines pada Kasus Diabetes Melitus*. Surabaya: Jurusan Statistika Institut Teknologi Sepuluh Nopember.
- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley and Sons.
- Austin, P. C. (2011). *A Tutorial and Case Study in Propensity Score Analysis: Research*, 46: pp: 135-147.
- Bansal, D., Gudala, K., Muthyala, H., Esam, H., Nayakallu, R., & Bhansali, A. (2014). Prevalence and Risk Factors of Development of Peripheral Diabetic Neuropathy in Type 2 DM in a Tertiary Care Setting. *Journal Diabetes Invest*, Vol. 5: pp. 714–721.
- Bazzano, L. A., Serdula, M., & Liu, S. (2005). Prevention of Type 2 Diabetes by Diet and Lifestyle Modification. *Journal of the American College of Nutrition*. 24, 310-319.
- Bilous, R., & Donnelly, R. (2014). *Buku Pegangan Diabetes, Edisi ke 4*. Diterjemahkan dari bahasa inggris, oleh Egi Komara Yudha. Jakarta: Bumi Medika.
- Breiman, L. (1993). *Classification and Regression Trees*. New York: CRC Press.
- Bruce, S. G., & Young, T. K. (2008). Prevalence and Risk Factors for Neuropathy in a Canadian First Nation Community. *Diabetes Care*, 31(9), 1837-1841.
- D'Agostino, R. B. (1998). Tutorial in Biostatistics Propensity score Method for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group. *Statistics in Medicine*, 17, 2265-2281.
- Darsana, I. (2014). *Korelasi Positif Kadar Asam Urat Serum Tinggi dengan Neuropati Diabetik Perifer pada Penderita DM Tipe-2 di RSUP Sanglah Denpasar*. Tesis Universitas Udayana.
- Depkes. (2009). *Tahun 2030 Prevalensi Diabetes Melitus di Indonesia Mencapai 21,3 Juta Orang*. Dipetik 15 Februari 2019, dari Kementerian Kesehatan Republik Indonesia:

<http://www.depkes.go.id/article/print/414/tahun-2030-prevalensi-diabetes-melitus-di-indonesia-mencapai-213-juta-orang.html>.

- Dehejja, RH & Wahba, S (1999). Causal Effects in Nonexperimental Studies : Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 1053-1062.
- Ernawati. (2018). *Propensity Score Stratification dengan metode SVM pada data Diabetes Melitus Tipe 2*. Tesis Statistika ITS.
- Imbens, G. (2004). Non Parametric Estimation of Average Treatment Effects Under Exogenit : A Review. *The Review of Economics and Statistics*, 86(1), 4-29.
- Jager, K., Zoccali, C., MacLeod, A., & Dekker, F. (2008). Confounding: What it is and how to deal with it. *Kidney International*, 73, 256-260.
- Kemenkes. (2013). *Riset Kesehatan Dasar 2013*. Jakarta: Badan Penelitian dan Pengembangan Kesehatan Kemenkes RI.
- Kirklin, J. K., Hanley F. L., Blackstone E. H., & Kouchoukos N. T. (2013). *Morphology, Diagnostic Criteria, Natural History, Techniques, Results, and indications in: Cardiac Surgery*. Philadelphia: Saunders Elsevier; PA 19103-2899.
- Kleyman N, Y. (2009). Testing For Covariate Balance in Comparative Studies, Dissertation Doctor of Philosophy. *Statistics*.
- Lee, B., J. Lessler, and E. Stuart. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29, 337–346.
- Luellen Jason K., Shahdish William, and Clark M.H.(2005). “Propensity Scores An Introduction and Experimental Test”. *Evaluation Review*, 29, 539-543.
- Lewis, Sharon M., Heitkemper, Margaret M., & Dirkson, Shannon R. (2000). *Medical-Surgical Nursing:Assessment, and Management of Clinical Problems, 5th ed*. St. Louis. MO: Mosby.
- Li, H., Graham D.J., & Majumdar A. (2012). The Impacts of Speed Cameras on Road Accidents: An application of Propensity

- Score Matching Methods. *Accident Analysis and Prevention*, 148-157.
- McCaffrey, D., G. Ridgeway, and A. Morral. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, 9, 403-425.
- Moisen, G. (2008). Classification and regression trees, in S. Jorgensen and B. Fath, eds. *Encyclopedia of Ecology*, 1, 582-588.
- Misnadiarly. (2006). *Diabetes Melitus Gangren, Ulcer, Infeksi, Mengenali Gejala, Menanggulangi, dan Mencegah Komplikasi*. Jakarta: Pustaka Obor Populer.
- NHS. (2016). *Type 2 Diabetes - Causes*. Dipetik 12 Februari 2019, dari NHS Choices: <http://www.nhs.uk/Conditions/Diabetestype2/Pages/Causes.aspx>.
- Pan, W., & Bai, H. (2015). *Propensity Score Analysis: Fundamentals and Developments*. New York: Guilford Press.
- Pernama, H. (2013). *Komplikasi Kronik dan Penyakit Penyerta pada Diabetes*. Bandung: Division of Endocrinology and Metabolism Department of Internal Medicine Padjadjaran University Medical School/Hasan Sadikin Hospital.
- Price, S., & Wilson, L. (2006). *Patofisiologi : Konsep Klinis Proses-Proses Penyakit* (Edisi 6, Vol. 1), Diterjemahkan dari bahasa inggris, oleh Sylvia A. Jakarta: EGC.
- Rai, D., B. Lee, C. Dalman, C. Newschaffer, G. Lewis, and C. Magnusson. (2017). Antidepressants during pregnancy and autism in offspring: population based cohort study. *BMJ*, 385-2811.
- Ria Dhea, L. (2013). Klasifikasi Penderita Diabetes Melitus dengan metode CHAID dan CART. Tugas Akhir Statistika ITS.
- Riskesdas. (2013). *Laporan Nasional*. Jakarta: Badan Penelitian dan Pengembangan Kesehatan Departemen Kesehatan RI.
- Roger, J., & Lewis, M.D. (2000). An Introduction to Classification and Regression Tree (CART) Analysis. *Journal of*

- Department of Emergency Medicine Harbor-UCLA Medical Center, 4, 221-250.*
- Rosenbaum, P., & Rubin, D. (1983). The Central Role of The Propensity Score in Observational Studies for Causal Effects. *Journal of Biometrika*, 70, 41-55.
- Silviatul, H. (2018). *Propensity Score Matching dengan metode SVM pada data HIV*. Tesis Statistika ITS.
- Soewondo, P. (2006). *Ketoasidosis diabetik*. Buku ajar ilmu penyakit dalam jilid III edisi IV. Jakarta: Penerbit FK UI.
- Stone, R. A., Obrosky D.S., Singer D. E., Kapoor W. N., Fine M. J., and the Pneumonia Patient Outcomes Research Team Investigators. (1995). Propensity score adjustment for pretreatment differences between hospitalized and ambulatory patients with community-acquired pneumonia. *Medical Care* 33, 56-66.
- Sustrani, L., Alam, S., & Hadibroto, I. (2010). *Diabetes: Informasi Lengkap untuk Penderita dan Keluarganya*. Jakarta: Gramedia Pustaka.
- Tu, W., & Zhou, X. (2002). A Bootstrap Confidence Interval Procedure for the Treatment Effect Using Propensity Score Cubclassification. *Health Services & Outcomes Research Methodology* 3, 135-147.
- Trisnawati, S.K., & Setyorogo, S. (2013). Faktor risiko Kejadian diabetes melitus tipe II di puskesmas kecamatan cengkareng Jakarta Barat Tahun 2012. Jakarta: Jurnal Ilmiah Kesehatan.
- WHO. (2016). *Global Report on Diabetes*. Geneva: WHO Library Cataloguing-in-Publication Data.
- WHO. (2018). Diabetes. Dipetik 08 Februari 2019, dari World Health Organization : <https://www.who.int/news-room/fact-sheets/detail/diabetes>.
- Wijayanti, D. N. (2013). *Analisis Faktor Penyebab Obesitas dan Cara Mengatasi Obesitas pada Remaja Putri (Studi Kasus pada Siswi SMA Negeri 3 Temanggung)*. Tesis Universitas Negeri Semarang.

Lampiran 1. Surat Pernyataan Pengambilan Data Sekunder.

SURAT PERNYATAAN

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

Nama : Naomi Puspita Happy Puranti
NRP : 06211540000062
Program Studi : S1 Statistika

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".

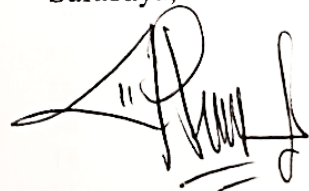
Surat pernyataan ini dibuat dengan sebenar-benarnya. Apabila ada pemalsuan data maka saya siap menerima sanksi sesuai dengan aturan yang berlaku.

Pembimbing Tugas Akhir



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

Mengetahui,
Surabaya, Juni 2019



Naomi Puspita H.P
NRP.06211540000062

Lampiran 2. Data Penderita DM tipe 2 di RSUD kab. Pasuruan.

No	Y	X1	X2	X3	X4	X5	X6
1	0	47	0	0	3	0	0
2	1	43	1	1	1	0	0
3	1	70	0	1	2	1	1
4	1	57	1	1	2	1	1
5	1	51	1	1	2	1	1
6	0	57	0	1	5	1	1
7	0	77	0	1	3	1	0
8	0	58	1	1	4	0	1
9	0	72	1	1	2	1	0
10	0	34	1	1	1	0	0
11	1	72	1	0	7	1	0
12	1	40	1	0	6	1	0
13	0	31	0	1	2	0	1
14	0	51	0	1	6	0	1
15	0	39	0	1	7	0	1
16	1	39	1	0	5	0	0
17	1	58	1	0	7	0	0
18	1	58	1	1	3	1	1
19	0	66	0	1	4	0	0
20	0	45	0	1	1	0	0
21	0	34	0	1	1	1	0
22	0	72	0	1	6	1	0
23	1	58	1	0	7	1	1
24	1	70	1	0	5	1	1
25	1	64	1	1	7	1	0
26	1	35	0	1	4	0	1
27	1	31	1	1	7	1	1
28	1	43	1	1	1	1	0
29	1	48	1	0	3	1	1
30	1	75	0	0	4	1	1
31	1	57	1	1	6	1	0
32	1	43	1	1	4	0	0
33	0	68	1	1	5	0	0
34	0	31	0	1	7	0	0
35	1	39	1	1	5	1	0
36	0	64	0	1	6	0	1
37	0	35	0	1	3	0	1

Lampiran 2. Data Penderita DM tipe 2 di RSUD kab. Pasuruan
(Lanjutan)

No	Y	X1	X2	X3	X4	X5	X6
38	1	29	1	1	3	1	1
39	1	31	1	1	2	1	0
40	1	45	1	0	2	1	1
41	1	66	1	0	5	1	0
42	0	31	0	1	1	0	0
43	1	43	0	0	6	0	1
44	1	35	0	1	4	1	1
45	1	43	0	1	7	1	0
46	1	51	1	1	4	1	0
47	1	58	1	1	2	0	0
48	1	34	0	0	5	0	1
49	1	72	1	0	5	0	0
50	1	51	0	0	7	0	1
51	0	64	1	1	4	0	1
52	0	48	1	1	1	0	1
53	0	68	1	0	7	1	0
54	1	57	0	0	3	1	1
55	1	72	1	0	4	1	1
56	0	64	0	1	5	0	0
57	0	47	0	1	5	0	0
58	1	48	1	0	1	0	0
59	0	47	0	1	1	0	0
60	1	51	1	0	5	1	1
61	1	48	0	0	1	1	1
62	1	75	1	0	2	1	1
63	0	40	1	1	2	1	0
64	1	68	0	0	4	1	1
65	0	40	1	1	7	1	0
66	0	72	1	1	6	0	0
67	1	75	1	0	6	1	1
68	1	64	0	0	7	1	1
69	1	70	0	0	6	0	0
70	0	47	1	1	6	1	0
71	1	35	1	0	3	0	1
72	1	39	0	0	3	0	1

Lampiran 2. Data Penderita DM tipe 2 di RSUD kab. Pasuruan
(Lanjutan)

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

Lampiran 3.

- a. *Syntax* tabulasi silang antara kovariat X dengan status komplikasi penyakit DM tipe-2 (Y)

```

data=read.csv("E:\\Data\\College Life\\TA\\DataDM.csv",sep=";",header=TRUE)
X2=data$X2
Y=data$Y
tableYX2=table(X2,Y)
colnames (tableYX2)=c("Komplikasi Kesehatan","Tidak ada Komplikasi Kesehatan")
rownames (tableYX2)=c("Perempuan","Laki-laki")
tableYX2

```

```

X3=data$X3
tableYX3=table(X3,Y)
colnames (tableYX3)=c("Komplikasi Kesehatan","Tidak Komplikasi Kesehatan")
rownames (tableYX3)=c("Ada Riwayat Genetik","Tidak Ada Riwayat Genetik")
tableYX3

```

```

X5=data$X5
tableYX5=table(X5,Y)
colnames (tableYX5)=c("Komplikasi Kesehatan","Tidak Ada Komplikasi Kesehatan")
rownames (tableYX5)=c("Obesitas","Tidak Obesitas")
tableYX5

```

```

X6=data$X6
tableYX6=table(X6,Y)
colnames (tableYX6)=c("Komplikasi Kesehatan","Tidak Ada Komplikasi Kesehatan")
rownames (tableYX6)=c("Hipertensi","Tidak Ada Hipertensi")
tableYX6

```

```

Z=data$Z
tableYZ=table(Z,Y)
colnames (tableYZ)=c("Ada Komplikasi Penyakit","Tidak Ada Komplikasi Penyakit")
rownames (tableYZ)=c("Kurang Aktivitas Olahraga","Cukup Aktivitas Olahraga")
tableYZ
rbind(tableYX2,tableYX3,tableYX5,tableYX6,tableYZ)

```

```

X2=data$X2
X3=data$X3
tableX2X3=table(X3,X2)
colnames (tableX2X3)=c("Perempuan","Laki-laki")
rownames (tableX2X3)=c("Ada Riwayat Genetik","Tidak Ada Riwayat Genetik")
tableX2X3

```

```

X5=data$X5
tableX2X5=table(X5,X2)
colnames (tableX2X5)=c("Perempuan","Laki-laki")
rownames (tableX2X5)=c("Obesitas","Tidak Obesitas")
tableX2X5

```

- a. *Syntax* tabulasi silang antara kovariat X dengan status komplikasi penyakit DM tipe-2 (Y) (Lanjutan)

```

X6=data$X6
tableX2X6=table(X6,X2)
colnames (tableX2X6)=c("Perempuan","Laki-laki")
rownames (tableX2X6)=c("Hipertensi","Tidak Hipertensi")
tableX2X6

```

```

Z=data$Z
tableX2Z=table(Z,X2)
colnames (tableX2Z)=c("Perempuan","Laki-laki")
rownames (tableX2Z)=c("Kurang Aktivitas Olahraga","Cukup Aktivitas Olahraga")
tableX2X7
rbind(tableX2X3,tableX2X5,tableX2X6,tableX2Z)

```

```

X3
X3=data$X3
X2=data$X2
tableX3X2=table(X2,X3)
colnames (tableX3X2)=c("Ada Riwayat Genetik","Tidak Ada Riwayat Genetik")
rownames (tableX3X2)=c("Perempuan","Laki-laki")
tableX3X2

```

```

X5=data$X5
tableX3X5=table(X5,X3)
colnames (tableX3X5)=c("Ada Riwayat Genetik","Tidak Ada Riwayat Genetik")
rownames (tableX3X5)=c("Obesitas","Tidak Obesitas")
tableX3X5

```

```

X6=data$X6
tableX3X6=table(X6,X3)
colnames (tableX3X6)=c("Ada Riwayat Genetik","Tidak Ada Riwayat Genetik")
rownames (tableX3X6)=c("Hipertensi","Tidak Hipertensi")
tableX3X6

```

```

Z=data$Z
tableX3Z=table(Z,X3)
colnames (tableX3Z)=c("Ada Riwayat Genetik","Tidak Ada Riwayat Genetik")
rownames (tableX3Z)=c("Kurang Aktivitas Olahraga","Cukup Aktivitas Olahraga")
tableX3Z
rbind(tableX3X2,tableX3X5,tableX3X6,tableX3Z)

```

```

X5
X5=data$X5
X2=data$X2
tableX5X2=table(X2,X5)
colnames (tableX5X2)=c("Obesitas","Tidak Obesitas")
rownames (tableX5X2)=c("Perempuan","Laki-laki")
tableX5X2

```

a. *Syntax* tabulasi silang antara kovariat X dengan status komplikasi penyakit DM tipe-2 (Y) (Lanjutan)

```
X3=data$X3
tableX5X3=table(X3,X5)
colnames (tableX5X3)=c("Obesitas","Tidak Obesitas")
rownames (tableX5X3)=c("Punya Riwayat Genetik","Tidak Punya Riwayat Genetik")
tableX5X3
```

```
X6=data$X6
tableX5X6=table(X6,X5)
colnames (tableX5X6)=c("Obesitas","Tidak Obesitas")
rownames (tableX5X6)=c("Hipertensi","Tidak Hipertensi")
tableX5X6
```

```
Z=data$Z
tableX5Z=table(Z,X5)
colnames (tableX5Z)=c("Obesitas","Tidak Obesitas")
rownames (tableX5Z)=c("Kurang Aktivitas Olahraga","Cukup Aktivitas Olahraga")
tableX5Z
rbind(tableX5X2,tableX5X3,tableX5X6,tableX5Z)
```

```
X6
X6=data$X6
X2=data$X2
tableX6X2=table(X2,X6)
colnames (tableX6X2)=c("Hipertensi","Tidak Hipertensi")
rownames (tableX6X2)=c("Perempuan","Laki-laki")
tableX6X2
```

```
X3=data$X3
tableX6X3=table(X3,X6)
colnames (tableX6X3)=c("Hipertensi","Tidak Hipertensi")
rownames (tableX6X3)=c("Ada Riwayat Genetik","Tidak Ada Riwayat Genetik")
tableX6X3
```

```
X5=data$X5
tableX6X5=table(X5,X6)
colnames (tableX6X5)=c("Hipertensi","Tidak Hipertensi")
rownames (tableX6X5)=c("Obesitas","Tidak Obesitas")
tableX6X5
```

```
Z=data$Z
tableX6Z=table(Z,X6)
colnames (tableX6Z)=c("Hipertensi","Tidak Hipertensi")
rownames (tableX6Z)=c("Kurang Aktivitas Olahraga","Cukup Aktivitas Olahraga")
tableX6Z
rbind(tableX6X2,tableX6X3,tableX6X5,tableX6Z)
```

- a. *Syntax* tabulasi silang antara kovariat X dengan status komplikasi penyakit DM tipe-2 (Y) (Lanjutan)

```
Z
Z=data$Z
X2=data$X2
tableZX2=table(X2,Z)
colnames (tableZX2)=c("Kurang Aktivitas Olahraga", "Cukup Aktivitas Olahraga")
rownames (tableZX2)=c("Perempuan", "Laki-laki")
tableZX2
```

```
X3=data$X3
tableZX3=table(X3,Z)
colnames (tableZX3)=c("Kurang Aktivitas Olahraga", "Cukup Aktivitas Olahraga")
rownames (tableZX3)=c("Punya Riwayat Genetik", "Tidak Punya Riwayat Genetik")
tableZX3
```

```
X5=data$X5
tableZX5=table(X5,Z)
colnames (tableZX5)=c("Kurang Aktivitas Olahraga", "Cukup Aktivitas Olahraga")
rownames (tableZX5)=c("Obesitas", "Tidak Obesitas")
tableZX5
```

```
X6=data$X6
tableZX6=table(X6,Z)
colnames (tableZX6)=c("Kurang Aktivitas Olahraga", "Cukup Aktivitas Olahraga")
rownames (tableZX6)=c("Obesitas", "Tidak Obesitas")
tableZX6
rbind(tableZX2,tableZX3,tableZX5,tableZX6)
```

- b. *Output* tabulasi silang antara kovariat Jenis Kelamin dengan status komplikasi penyakit DM tipe-2 (Y)

	Komplikasi	Tidak	Total
Perempuan	29	20	49
Laki-Laki	13	34	47
Total	42	54	96

- c. *Output* tabulasi silang antara kovariat Riwayat Genetik dengan status komplikasi penyakit DM tipe-2 (Y)

	Komplikasi	Tidak	Total
Riwayat Genetik	12	32	44
Tidak Ada	30	22	52
Total	42	54	96

- d. *Output* tabulasi silang antara kovariat Obesitas dengan status komplikasi penyakit DM tipe-2 (Y)

	Komplikasi	Tidak	Total
Obesitas	30	17	47
Tidak Obesitas	12	37	49
Total	42	54	96

- e. *Output* tabulasi silang antara kovariat Hipertensi dengan status komplikasi penyakit DM tipe-2 (Y)

	Komplikasi	Tidak	Total
Hipertensi	31	21	52
Tidak Hipertensi	11	33	44
Total	42	54	96

- f. *Output* tabulasi silang antara kovariat Aktivitas Olahraga dengan status komplikasi penyakit DM tipe-2 (Y)

	Komplikasi	Tidak	Total
Kurang	40	16	56
Cukup	2	38	40
Total	42	54	96

- g. *Output* tabulasi silang antara kovariat Aktivitas Olahraga dengan status Jenis Kelamin

	Kurang	Cukup	Total
Perempuan	35	14	49
Laki-Laki	21	26	47
Total	56	40	96

- h. *Output* tabulasi silang antara kovariat Aktivitas Olahraga dengan Riwayat Genetik

	Kurang	Cukup	Total
Punya	18	26	44
Tidak Punya	38	14	52
Total	56	40	96

- i. *Output* tabulasi silang antara kovariat Aktivitas Olahraga dengan Obesitas

	Kurang	Cukup	Total
Obesitas	34	13	47
Tidak Obesitas	22	27	49
Total	56	40	96

Lampiran 4.

- a. *Syntax chi-square* antara variabel yang berpotensi sebagai *confounding* dengan kovariat X lainnya (Lanjutan)

```
#Chi Square
data=read.csv("E:\\Data\\College Life\\TA\\DataDM.csv",sep=";",header=TRUE)
=====
#Chisquare X2
tableX3X2=table(X3,X2)
testX3X2=chisq.test(tableX3X2,correct=FALSE)
tableX5X2=table(X5,X2)
testX5X2=chisq.test(tableX5X2,correct=FALSE)
tableX6X2=table(X6,X2)
testX6X2=chisq.test(tableX6X2,correct=FALSE)
tableZX2=table(Z,X2)
testZX2=chisq.test(tableZX2,correct=FALSE)
variabelX2=rbind("X3*X2","X5*X2","X6*X2","Z*X2")
xsquareX2=rbind(testX3X2$statistic, testX5X2$statistic, testX6X2$statistic,
testZX2$statistic)
pvalueX2=rbind(testX3X2$p.value, testX5X2$p.value, testX6X2$p.value,
testZX2$p.value)
confoundX2=data.frame(variabelX2,xsquareX2,pvalueX2)
=====
#Chisquare X3
tableX2X3=table(X2,X3)
testX2X3=chisq.test(tableX2X3,correct=FALSE)
tableX5X3=table(X5,X3)
testX5X3=chisq.test(tableX5X3,correct=FALSE)
tableX6X3=table(X6,X3)
testX6X3=chisq.test(tableX6X3,correct=FALSE)
tableZX3=table(Z,X3)
testZX3=chisq.test(tableZX3,correct=FALSE)
variabelX3=rbind("X2*X3","X5*X3","X6*X3","Z*X3")
xsquareX3=rbind(testX2X3$statistic, testX5X3$statistic, testX6X3$statistic,
testZX3$statistic)
pvalueX3=rbind(testX2X3$p.value, testX5X3$p.value, testX6X3$p.value,
testZX3$p.value)
confoundX3=data.frame(variabelX3,xsquareX3,pvalueX3)
=====
#Chisquare X5
tableX2X5=table(X2,X5)
testX2X5=chisq.test(tableX2X5,correct=FALSE)
tableX3X5=table(X3,X5)
testX3X5=chisq.test(tableX3X5,correct=FALSE)
tableX6X5=table(X6,X5)
testX6X5=chisq.test(tableX6X5,correct=FALSE)
tableZX5=table(Z,X5)
testZX5=chisq.test(tableZX5,correct=FALSE)
```

- a. *Syntax chi-square* antara variabel yang berpotensi sebagai *confounding* dengan kovariat X lainnya (Lainnya)

```

variabelX5=rbind("X2*X5","X3*X5","X6*X5","Z*X5")
xsquareX5=rbind(testX2X5$statistic, testX3X5$statistic, testX6X5$statistic,
testZX5$statistic)
pvalueX5=rbind(testX2X5$p.value, testX3X5$p.value, testX6X5$p.value,
testZX5$p.value)
confoundX5=data.frame(variabelX5,xsquareX5,pvalueX5)

```

```

#Chisquare X6
tableX2X6=table(X2,X6)
testX2X6=chisq.test(tableX2X6,correct=FALSE)
tableX3X6=table(X3,X6)
testX3X6=chisq.test(tableX3X6,correct=FALSE)
tableX5X6=table(X5,X6)
testX5X6=chisq.test(tableX5X6,correct=FALSE)
tableZX6=table(Z,X6)
testZX6=chisq.test(tableZX6,correct=FALSE)
variabelX6=rbind("X2*X6","X3*X6","X5*X6","Z*X6")
xsquareX6=rbind(testX2X6$statistic, testX3X6$statistic, testX5X6$statistic,
testZX6$statistic)
pvalueX6=rbind(testX2X6$p.value, testX3X6$p.value, testX5X6$p.value,
testZX6$p.value)
confoundX6=data.frame(variabelX6,xsquareX6,pvalueX6)

```

```

#Chisquare X7
tableX2Z=table(X2,Z)
testX2Z=chisq.test(tableX2Z,correct=FALSE)
tableX3Z=table(X3,Z)
testX3Z=chisq.test(tableX3Z,correct=FALSE)
tableX5Z=table(X5,Z)
testX5Z=chisq.test(tableX5Z,correct=FALSE)
tableX6Z=table(X6,Z)
testX6Z=chisq.test(tableX6Z,correct=FALSE)
variabelZ=rbind("X2*Z","X3*Z","X5*Z","X6*Z")
xsquareZ=rbind(testX2Z$statistic, testX3Z$statistic, testX5Z$statistic,
testX6Z$statistic)
pvalueZ=rbind(testX2Z$p.value, testX3Z$p.value, testX5Z$p.value, testX6Z$p.value)
confoundZ=data.frame(variabelZ,xsquareZ,pvalueZ)

```

```

#Chisquare Y
tableX2Y=table(X2,Y)
testX2Y=chisq.test(tableX2Y,correct=FALSE)
tableX3Y=table(X3,Y)
testX3Y=chisq.test(tableX3Y,correct=FALSE)
tableX5Y=table(X5,Y)

```


- a. *Syntax chi-square* antara variabel yang berpotensi sebagai *confounding* dengan kovariat X lainnya (Lanjutan)

```
testX5Y=chisq.test(tableX5Y,correct=FALSE)
tableX6Y=table(X6,Y)
testX6Y=chisq.test(tableX6Y,correct=FALSE)
tableZY=table(Z,Y)
testZY=chisq.test(tableZY,correct=FALSE)
variabelY=rbind("X2*Y","X3*Y","X5*Y","X6*Y","Z*Y")
xsquareY=rbind(testX2Y$statistic, testX3Y$statistic, testX5Y$statistic,
testX6Y$statistic,
testZY$statistic)
pvalueY=rbind(testX2Y$p.value, testX3Y$p.value, testX5Y$p.value, testX6Y$p.value,
testZY$p.value)
confoundY=data.frame(variabelY,xsquareY,pvalueY)
```

- b. *Output chi-square* antara variabel yang berpotensi sebagai *confounding* dengan kovariat X lainnya

```
> 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    Z*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    Z*X3 10.146653  0.001445634
> confoundX5
  variabelX5 X.squared  pvalueX5
1   X2*X5  4.187680  0.040718779
2   X3*X5  0.399066  0.527572018
```

b. *Output chi-square* antara variabel yang berpotensi sebagai *confounding* dengan kovariat X lainnya (Lanjutan)

```
3 X6*X5 3.463333 0.062743721
4 Z*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 Z*X6 5.543257 0.01855203
> confoundZ
variabelZ X.squared pvalueZ
1 X2*Z 7.061398 0.007876250
2 X3*Z 10.146653 0.001445634
3 X5*Z 7.432988 0.006403892
4 X6*Z 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 Z*Y 41.839456 9.908366e-11
```

Lampiran 5.

- a. *Syntax* mencari Akurasi, Sensifitas, Spesifitas untuk Pohon Klasifikasi Optimum, Pembentukan Pohon dan Hasil Prediksi

```

#File Data
data=read.csv("E:\\Data\\College Life\\TA\\DataFix.csv",sep=";",header=TRUE)
attach(data)
=====
#MenetapkanSeed
set.seed(1234)
=====
#Membuat Sampel
Senam=which(data[,9]=="Senam")
DataSenam=data[Senam,]
SampelSenam=round(0.70*nrow(DataSenam))
indexSenam=sample(seq_len(nrow(DataSenam)),size = SampelSenam)
=====
TidakSenam=which(data[,9]=="Tidak Senam")
DataTidakSenam=data[TidakSenam,]
SampelTidakSenam=round(0.70*nrow(DataTidakSenam))
indexTidakSenam=sample(seq_len(nrow(DataTidakSenam)),size =
SampelTidakSenam)
=====
#DataTrain
data_trainSenam=DataSenam[indexSenam,]
data_trainTidakSenam=DataTidakSenam[indexTidakSenam,]
data_train=rbind(data_trainSenam,data_trainTidakSenam)
=====
#DataTest
data_testSenam=DataSenam[-indexSenam,]
data_testTidakSenam=DataTidakSenam[-indexTidakSenam,]
data_test=rbind(data_testSenam,data_testTidakSenam)
=====
#Descriptive Pohon
fit <- rpart(Aktivitas.Olahraga ~ Usia+ Jenis.Kelamin + Riwayat.Genetik + Lama.Sakit
+ Obesitas + Hipertensi,
method="class", data=data_train)
printcp(fit) # display the results
plotcp(fit) # visualize cross-validation results
summary(fit) # detailed summary of splits
print(fit)
=====
#Membuat Pohon
rpart.plot(fit, uniform=TRUE,
main="Pohon Keputusan untuk Diabetes Melitus Tipe 2")
text(fit, use.n=TRUE, all=TRUE, cex=.8)
=====
#Pemangkas Pohon
pfit<- prune(fit, cp=fit$cpable[which.min(fit$cpable[,"xerror"]),"CP"])
=====

```

- a. *Syntax* mencari Akurasi, Sensifitas, Spesifitas untuk Pohon Klasifikasi Optimum, Pembentukan Pohon dan Hasil Prediksi (Lanjutan)

```
#Pohon yang sudah dipangkas
rpart.plot(pfit, uniform=TRUE,
           main="Pohon Keputusan Optimal untuk Diabetes Melitus Tipe 2")
text(pfit, use.n=TRUE, all=TRUE, cex=.8)
=====
#Pada Data Testing
hasil=predict(fit, data_test, type="class")
Datahasil=data.frame(hasil)
levels(data_test$Aktivitas.Olahraga)
HasilPrediksi=confusionMatrix(data = hasil, data_test$Aktivitas.Olahraga, positive =
levels(data_test$Aktivitas.Olahraga)[2])
```

- b. *Output* Akurasi, Sensifitas, Spesifitas untuk Pohon Klasifikasi Optimum

```
Sampel 65%
Confusion Matrix and Statistics
Reference
Prediction  Senam Tidak Senam
Senam      34      16
Tidak Senam  2      10

Accuracy : 0.7097
95% CI : (0.5805, 0.818)
No Information Rate : 0.5806
P-Value [Acc > NIR] : 0.025125
Kappa : 0.3557
McNemar's Test P-Value : 0.002183
Sensitivity : 0.3846
Specificity : 0.9444
Pos Pred Value : 0.8333
Neg Pred Value : 0.6800
Prevalence : 0.4194
Detection Rate : 0.1613
Detection Prevalence : 0.1935
Balanced Accuracy : 0.6645
'Positive' Class : Tidak Senam
```

b. *Output* Akurasi, Sensifitas, Spesifitas untuk Pohon Klasifikasi Optimum (Lanjutan)

Test
Confusion Matrix and Statistics
Reference
Prediction Senam Tidak Senam
Senam 15 10
Tidak Senam 5 4

Accuracy : 0.5588
95% CI : (0.3789, 0.7281)
No Information Rate : 0.5882
P-Value [Acc > NIR] : 0.7017
Kappa : 0.0377
Mcnemar's Test P-Value : 0.3017
Sensitivity : 0.2857
Specificity : 0.7500
Pos Pred Value : 0.4444
Neg Pred Value : 0.6000
Prevalence : 0.4118
Detection Rate : 0.1176
Detection Prevalence : 0.2647
Balanced Accuracy : 0.5179
'Positive' Class : Tidak Senam

Sampel 70%

Confusion Matrix and Statistics
Reference
Prediction Senam Tidak Senam
Senam 35 12
Tidak Senam 4 16

Accuracy : 0.7612
95% CI : (0.6414, 0.8569)
No Information Rate : 0.5821
P-Value [Acc > NIR] : 0.001696
Kappa : 0.4885
Mcnemar's Test P-Value : 0.080118
Sensitivity : 0.5714
Specificity : 0.8974
Pos Pred Value : 0.8000
Neg Pred Value : 0.7447
Prevalence : 0.4179
Detection Rate : 0.2388
Detection Prevalence : 0.2985
Balanced Accuracy : 0.7344

b. *Output* Akurasi, Sensifitas, Spesifitas untuk Pohon Klasifikasi Optimum (Lanjutan)

'Positive' Class : Tidak Senam

Test

Confusion Matrix and Statistics

Reference

Prediction Senam Tidak Senam

Senam 12 6

Tidak Senam 5 6

Accuracy : 0.6207

95% CI : (0.4226, 0.7931)

No Information Rate : 0.5862

P-Value [Acc > NIR] : 0.4296

Kappa : 0.2084

Mcnemar's Test P-Value : 1.0000

Sensitivity : 0.5000

Specificity : 0.7059

Pos Pred Value : 0.5455

Neg Pred Value : 0.6667

Prevalence : 0.4138

Detection Rate : 0.2069

Detection Prevalence : 0.3793

Balanced Accuracy : 0.6029

'Positive' Class : Tidak Senam

Sampel 75%

Confusion Matrix and Statistics

Reference

Prediction Senam Tidak Senam

Senam 36 10

Tidak Senam 6 20

Accuracy : 0.7778

95% CI : (0.6644, 0.8673)

No Information Rate : 0.5833

P-Value [Acc > NIR] : 0.0004238

Kappa : 0.534

Mcnemar's Test P-Value : 0.4532547

Sensitivity : 0.6667

Specificity : 0.8571

Pos Pred Value : 0.7692

Neg Pred Value : 0.7826

Prevalence : 0.4167

Detection Rate : 0.2778

Detection Prevalence : 0.3611

Balanced Accuracy : 0.7619

b. *Output* Akurasi, Sensifitas, Spesifitas untuk Pohon Klasifikasi Optimum (Lanjutan)

'Positive' Class : Tidak Senam

Test

Confusion Matrix and Statistics

Reference

Prediction Senam Tidak Senam

Senam 9 3

Tidak Senam 5 7

Accuracy : 0.6667

95% CI : (0.4468, 0.8437)

No Information Rate : 0.5833

P-Value [Acc > NIR] : 0.2702

Kappa : 0.3333

Mcnemar's Test P-Value : 0.7237

Sensitivity : 0.7000

Specificity : 0.6429

Pos Pred Value : 0.5833

Neg Pred Value : 0.7500

Prevalence : 0.4167

Detection Rate : 0.2917

Detection Prevalence : 0.5000

Balanced Accuracy : 0.6714

'Positive' Class : Tidak Senam

Sampel 80%

Confusion Matrix and Statistics

Reference

Prediction Senam Tidak Senam

Senam 42 16

Tidak Senam 3 16

Accuracy : 0.7532

95% CI : (0.6418, 0.8444)

No Information Rate : 0.5844

P-Value [Acc > NIR] : 0.001497

Kappa : 0.4603

Mcnemar's Test P-Value : 0.005905

Sensitivity : 0.5000

Specificity : 0.9333

Pos Pred Value : 0.8421

Neg Pred Value : 0.7241

Prevalence : 0.4156

Detection Rate : 0.2078

Detection Prevalence : 0.2468

Balanced Accuracy : 0.7167

b. *Output* Akurasi, Sensifitas, Spesifitas untuk Pohon Klasifikasi Optimum (Lanjutan)

'Positive' Class : Tidak Senam

Test

Confusion Matrix and Statistics

Reference

Prediction Senam Tidak Senam

Senam 8 6

Tidak Senam 3 2

Accuracy : 0.5263

95% CI : (0.2886, 0.7555)

No Information Rate : 0.5789

P-Value [Acc > NIR] : 0.7586

Kappa : -0.024

Mcnemar's Test P-Value : 0.5050

Sensitivity : 0.2500

Specificity : 0.7273

Pos Pred Value : 0.4000

Neg Pred Value : 0.5714

Prevalence : 0.4211

Detection Rate : 0.1053

Detection Prevalence : 0.2632

Balanced Accuracy : 0.4886

'Positive' Class : Tidak Senam

Sampel 85%

Confusion Matrix and Statistics

Reference

Prediction Senam Tidak Senam

Senam 36 5

Tidak Senam 12 29

Accuracy : 0.7927

95% CI : (0.6889, 0.8743)

No Information Rate : 0.5854

P-Value [Acc > NIR] : 6.025e-05

Kappa : 0.5854

Mcnemar's Test P-Value : 0.1456

Sensitivity : 0.8529

Specificity : 0.7500

Pos Pred Value : 0.7073

Neg Pred Value : 0.8780

Prevalence : 0.4146

Detection Rate : 0.3537

Detection Prevalence : 0.5000

Balanced Accuracy : 0.8015

b. *Output* Akurasi, Sensifitas, Spesifitas untuk Pohon Klasifikasi Optimum (Lanjutan)

```

'Positive' Class : Tidak Senam

Test
Confusion Matrix and Statistics
      Reference
Prediction  Senam Tidak Senam
Senam       6      4
Tidak Senam 2      2
Accuracy : 0.5714
95% CI : (0.2886, 0.8234)
No Information Rate : 0.5714
P-Value [Acc > NIR] : 0.6105
Kappa : 0.087
McNemar's Test P-Value : 0.6831
Sensitivity : 0.3333
Specificity : 0.7500
Pos Pred Value : 0.5000
Neg Pred Value : 0.6000
Prevalence : 0.4286
Detection Rate : 0.1429
Detection Prevalence : 0.2857
Balanced Accuracy : 0.5417
'Positive' Class : Tidak Senam

```

c. *Output* CART

```

Classification tree:
rpart(formula = Aktivitas.Olahraga ~ Usia + Jenis.Kelamin + Riwayat.Genetik +
Lama.Sakit + Obesitas + Hipertensi, data = data_train, method = "class")

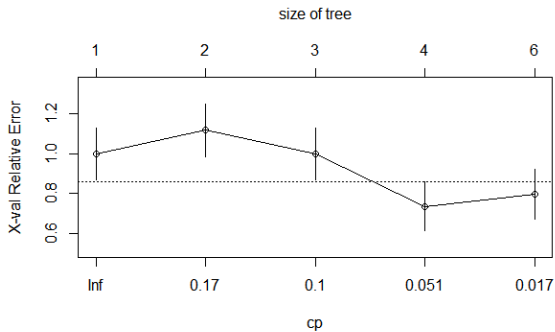
Variables actually used in tree construction:
[1] Hipertensi  Jenis.Kelamin  Riwayat.Genetik  Usia

Root node error: 34/82 = 0.41463

n= 82

      CP nsplit rel error  xerror  xstd
1 0.235294  0 1.00000 1.00000 0.13121
2 0.117647  1 0.76471 1.11765 0.13281
3 0.088235  2 0.64706 1.00000 0.13121
4 0.029412  3 0.55882 0.73529 0.12261
5 0.010000  5 0.50000 0.79412 0.12516

```

c. *Output CART*

```
> summary(fit) # detailed summary of splits
```

```
Call:
```

```
rpart(formula = Aktivitas.Olahraga ~ Usia + Jenis.Kelamin + Riwayat.Genetik +  
Lama.Sakit + Obesitas + Hipertensi, data = data_train, method = "class")  
n= 82
```

	CP	nsplit	rel error	xerror	xstd
1	0.23529412	0	1.0000000	1.0000000	0.1312122
2	0.11764706	1	0.7647059	1.1176471	0.1328105
3	0.08823529	2	0.6470588	1.0000000	0.1312122
4	0.02941176	3	0.5588235	0.7352941	0.1226088
5	0.01000000	5	0.5000000	0.7941176	0.1251634

```
Variable importance
```

Jenis.Kelamin	Riwayat.Genetik	Usia	Hipertensi	Lama.Sakit
27	22	19	15	9
Obesitas				
9				

```
Node number 1: 82 observations, complexity param=0.2352941  
predicted class=Senam expected loss=0.4146341 P(node)=1  
class counts: 48 34  
probabilities: 0.585 0.415  
left son=2 (42 obs) right son=3 (40 obs)
```

```
Primary splits:
```

```
Riwayat.Genetik splits as LR, improve=5.366783, (0 missing)  
Jenis.Kelamin splits as RL, improve=4.780488, (0 missing)  
Hipertensi splits as LR, improve=2.697223, (0 missing)  
Obesitas splits as LR, improve=2.280728, (0 missing)  
Usia < 32.5 to the left, improve=1.862717, (0 missing)
```

```
Surrogate splits:
```

```
Usia < 65 to the left, agree=0.646, adj=0.275, (0 split)
```

c. Output CART (Lanjutan)

```

Hipertensi splits as LR, agree=0.585, adj=0.150, (0 split)
Lama.Sakit < 2.5 to the left, agree=0.561, adj=0.100, (0 split)
Jenis.Kelamin splits as LR, agree=0.549, adj=0.075, (0 split)
Obesitas splits as LR, agree=0.549, adj=0.075, (0 split)
Node number 2: 42 observations, complexity param=0.02941176
predicted class=Senam expected loss=0.2380952 P(node) =0.5121951
class counts: 32 10
probabilities: 0.762 0.238
left son=4 (19 obs) right son=5 (23 obs)
Primary splits:
Jenis.Kelamin splits as RL, improve=2.3868370, (0 missing)
Obesitas splits as LR, improve=1.4325400, (0 missing)
Usia < 57.5 to the left, improve=0.5952381, (0 missing)
Hipertensi splits as LR, improve=0.2861722, (0 missing)
Lama.Sakit < 4.5 to the right, improve=0.2169188, (0 missing)
Surrogate splits:
Obesitas splits as LR, agree=0.690, adj=0.316, (0 split)
Usia < 37 to the left, agree=0.619, adj=0.158, (0 split)
Lama.Sakit < 1.5 to the left, agree=0.571, adj=0.053, (0 split)

Node number 3: 40 observations, complexity param=0.1176471
predicted class=Tidak Senam expected loss=0.4 P(node) =0.4878049
class counts: 16 24
probabilities: 0.400 0.600
left son=6 (22 obs) right son=7 (18 obs)
Primary splits:
Jenis.Kelamin splits as RL, improve=3.5636360, (0 missing)
Hipertensi splits as LR, improve=1.5838380, (0 missing)
Usia < 71 to the left, improve=1.1220780, (0 missing)
Lama.Sakit < 4.5 to the left, improve=0.8000000, (0 missing)
Obesitas splits as LR, improve=0.3929825, (0 missing)
Surrogate splits:
Lama.Sakit < 4.5 to the left, agree=0.650, adj=0.222, (0 split)
Usia < 71 to the left, agree=0.625, adj=0.167, (0 split)
Obesitas splits as LR, agree=0.625, adj=0.167, (0 split)

Node number 4: 19 observations
predicted class=Senam expected loss=0.05263158 P(node) =0.2317073
class counts: 18 1
probabilities: 0.947 0.053

Node number 5: 23 observations, complexity param=0.02941176
predicted class=Senam expected loss=0.3913043 P(node) =0.2804878
class counts: 14 9
probabilities: 0.609 0.391
left son=10 (13 obs) right son=11 (10 obs)

```

c. *Output* CART (Lanjutan)

Primary splits:

Usia < 54 to the left, improve=1.54113700, (0 missing)
 Hipertensi splits as LR, improve=0.28985510, (0 missing)
 Obesitas splits as LR, improve=0.09937888, (0 missing)
 Lama.Sakit < 3.5 to the right, improve=0.03227931, (0 missing)

Surrogate splits:

Lama.Sakit < 4.5 to the left, agree=0.652, adj=0.2, (0 split)

Node number 6: 22 observations, complexity param=0.08823529
 predicted class=Senam expected loss=0.4090909 P(node) =0.2682927
 class counts: 13 9
 probabilities: 0.591 0.409
 left son=12 (9 obs) right son=13 (13 obs)

Primary splits:

Hipertensi splits as LR, improve=2.70474000, (0 missing)
 Usia < 61.5 to the right, improve=1.06371400, (0 missing)
 Lama.Sakit < 4.5 to the left, improve=0.20779220, (0 missing)
 Obesitas splits as LR, improve=0.03807304, (0 missing)

Surrogate splits:

Usia < 31.5 to the left, agree=0.682, adj=0.222, (0 split)
 Lama.Sakit < 2.5 to the left, agree=0.636, adj=0.111, (0 split)
 Obesitas splits as LR, agree=0.636, adj=0.111, (0 split)

Node number 7: 18 observations
 predicted class=Tidak Senam expected loss=0.1666667 P(node) =0.2195122
 class counts: 3 15
 probabilities: 0.167 0.833

Node number 10: 13 observations
 predicted class=Senam expected loss=0.2307692 P(node) =0.1585366
 class counts: 10 3
 probabilities: 0.769 0.231

Node number 11: 10 observations
 predicted class=Tidak Senam expected loss=0.4 P(node) =0.1219512
 class counts: 4 6
 probabilities: 0.400 0.600

Node number 12: 9 observations
 predicted class=Senam expected loss=0.1111111 P(node) =0.1097561
 class counts: 8 1
 probabilities: 0.889 0.111

Node number 13: 13 observations
 predicted class=Tidak Senam expected loss=0.3846154 P(node) =0.1585366
 class counts: 5 8

c. Output CART (Lanjutan)

probabilities: 0.385 0.615

> print(fit)

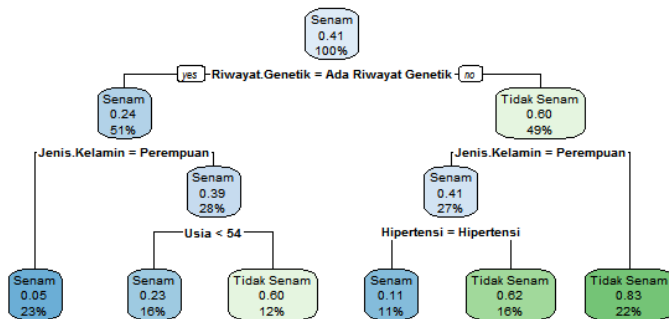
n= 82

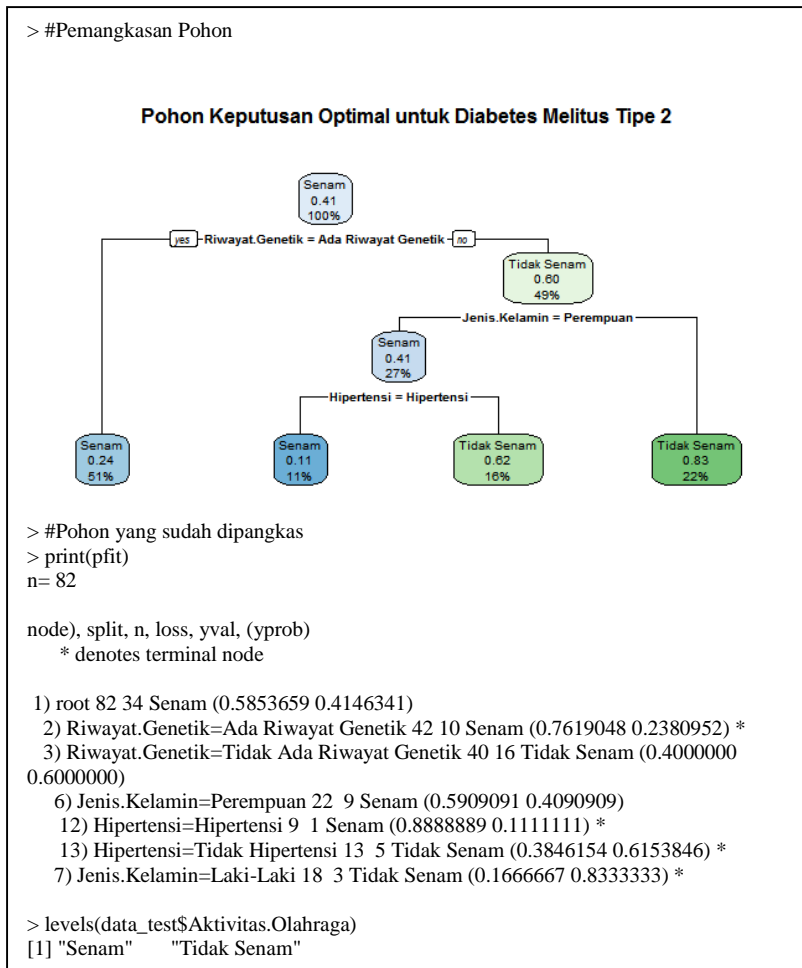
node), split, n, loss, yval, (yprob)

* denotes terminal node

- 1) root 82 34 Senam (0.58536585 0.41463415)
 - 2) Riwayat.Genetik=Ada Riwayat Genetik 42 10 Senam (0.76190476 0.23809524)
 - 4) Jenis.Kelamin=Perempuan 19 1 Senam (0.94736842 0.05263158) *
 - 5) Jenis.Kelamin=Laki-Laki 23 9 Senam (0.60869565 0.39130435)
 - 10) Usia<=54 13 3 Senam (0.76923077 0.23076923) *
 - 11) Usia>=54 10 4 Tidak Senam (0.40000000 0.60000000) *
 - 3) Riwayat.Genetik=Tidak Ada Riwayat Genetik 40 16 Tidak Senam (0.40000000 0.60000000)
 - 6) Jenis.Kelamin=Perempuan 22 9 Senam (0.59090909 0.40909091)
 - 12) Hipertensi=Hipertensi 9 1 Senam (0.88888889 0.11111111) *
 - 13) Hipertensi=Tidak Hipertensi 13 5 Tidak Senam (0.38461538 0.61538462) *
 - 7) Jenis.Kelamin=Laki-Laki 18 3 Tidak Senam (0.16666667 0.83333333) *
- > #Membuat Pohon

Pohon Keputusan untuk Diabetes Melitus Tipe 2



c. *Output* CART (Lanjutan)

Lampiran 6. Nilai Proporsi dan *Propensity Score* dalam Strata.

Y	X1	X2	X3	X4	X5	X6	Z	Strata	$\hat{f}(x)$	$\hat{e}(x)$
0	77	0	0	3	1	0	1	1	0,762	0,682
0	47	0	0	1	0	0	1	1	0,762	0,682
0	40	1	0	2	1	0	1	1	0,762	0,682
1	58	1	0	3	1	1	1	1	0,762	0,682
0	47	0	0	5	0	0	1	1	0,762	0,682
0	35	0	0	3	0	1	1	1	0,762	0,682
0	66	0	0	4	0	0	1	1	0,762	0,682
0	57	0	0	6	0	0	1	1	0,762	0,682
0	39	1	0	3	0	0	1	1	0,762	0,682
0	39	0	0	7	0	1	1	1	0,762	0,682
0	34	1	0	1	0	0	1	1	0,762	0,682
0	31	0	0	2	0	1	1	1	0,762	0,682
0	58	0	0	1	0	0	1	1	0,762	0,682
0	57	0	0	5	1	1	1	1	0,762	0,682
1	43	1	0	1	0	0	1	1	0,762	0,682
0	47	1	0	6	1	0	1	1	0,762	0,682
0	31	0	0	1	0	0	1	1	0,762	0,682
0	72	0	0	6	1	0	1	1	0,762	0,682
1	51	1	0	4	1	0	1	1	0,762	0,682
0	64	0	0	6	0	1	1	1	0,762	0,682
0	40	1	0	7	1	0	1	1	0,762	0,682
0	64	0	0	5	0	0	1	1	0,762	0,682
1	57	1	0	2	1	1	1	1	0,762	0,682
1	51	1	0	2	1	1	1	1	0,762	0,682
1	43	1	0	4	0	0	1	1	0,762	0,682
1	35	0	0	4	0	1	1	1	0,762	0,682
0	45	0	0	1	0	0	1	1	0,762	0,682
0	72	1	0	6	0	0	1	1	0,762	0,682
0	51	0	0	6	0	1	1	1	0,762	0,682
1	31	1	0	7	1	1	1	1	0,762	0,682
0	68	1	0	5	0	0	1	1	0,762	0,682
0	31	0	0	7	0	0	1	1	0,762	0,682
1	57	1	0	6	1	0	0	1	0,238	0,559
1	35	0	0	4	1	1	0	1	0,238	0,559
0	72	1	0	2	1	0	0	1	0,238	0,559
0	77	0	0	3	1	0	1	1	0,762	0,682

Lampiran 6. Nilai Proporsi dan *Propensity Score* dalam Strata
(Lanjutan)

Y	X1	X2	X3	X4	X5	X6	Z	Strata	$\hat{f}(x)$	$\hat{e}(x)$
1	43	1	0	1	1	0	0	1	0,238	0,559
0	58	1	0	4	0	1	0	1	0,238	0,559
1	51	1	0	3	0	1	0	1	0,238	0,559
1	58	1	0	2	0	0	0	1	0,238	0,559
1	29	1	0	3	1	1	0	1	0,238	0,559
1	64	1	0	7	1	0	0	1	0,238	0,559
1	75	1	0	6	1	1	0	1	0,238	0,559
1	29	0	1	4	1	0	1	2	0,889	0,709
0	29	0	1	2	1	0	1	2	0,889	0,709
0	47	0	1	3	0	0	1	2	0,889	0,709
1	29	0	1	3	0	0	1	2	0,889	0,709
0	77	0	1	5	0	0	1	2	0,889	0,709
0	66	0	1	2	0	0	1	2	0,889	0,709
0	68	0	1	2	0	0	1	2	0,889	0,709
0	70	0	1	6	0	0	1	2	0,889	0,709
1	70	0	1	6	0	0	0	2	0,111	0,528
1	68	0	1	4	1	1	1	3	0,385	0,595
1	70	0	1	4	1	1	1	3	0,385	0,595
0	29	0	1	7	0	1	1	3	0,385	0,595
0	40	0	1	6	0	1	1	3	0,385	0,595
1	70	0	1	1	1	1	1	3	0,385	0,595
1	43	0	1	6	0	1	0	3	0,615	0,649
1	57	0	1	3	1	1	0	3	0,615	0,649
1	39	0	1	3	0	1	0	3	0,615	0,649
1	75	0	1	4	1	1	0	3	0,615	0,649
1	34	0	1	5	0	1	0	3	0,615	0,649
1	45	0	1	3	1	1	0	3	0,615	0,649
1	48	0	1	1	1	1	0	3	0,615	0,649
1	51	0	1	7	0	1	0	3	0,615	0,649
0	40	1	1	4	0	0	1	4	0,167	0,542
0	68	1	1	7	1	0	1	4	0,167	0,542
1	72	1	1	7	1	0	0	4	0,833	0,697
1	48	1	1	1	0	0	0	4	0,833	0,697
1	58	1	1	7	0	0	0	4	0,833	0,697
1	66	1	1	5	1	0	0	4	0,833	0,697

Lampiran 6. Nilai Proporsi dan *Propensity Score* dalam Strata
(Lanjutan)

Y	X1	X2	X3	X4	X5	X6	Z	Strata	$\hat{f}(x)$	$\hat{e}(x)$
1	72	1	1	5	0	0	0	4	0,833	0,697
1	39	1	1	5	0	0	0	4	0,833	0,697
1	40	1	1	6	1	0	0	4	0,833	0,697
1	70	1	1	5	1	1	1	4	0,167	0,542
1	48	1	1	3	1	1	0	4	0,833	0,697
1	75	1	1	6	1	1	0	4	0,833	0,697
1	72	1	1	4	1	1	0	4	0,833	0,697
1	75	1	1	2	1	1	0	4	0,833	0,697
1	35	1	1	3	0	1	0	4	0,833	0,697
1	58	1	1	7	1	1	0	4	0,833	0,697
1	51	1	1	5	1	1	0	4	0,833	0,697
1	66	1	1	7	1	1	0	4	0,833	0,697

Lampiran 7.

- a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi.

Sebelum

```

data=read.csv("E:\\Data\\College Life\\TA\\DataDM.csv",sep=";",header=TRUE)
attach(data)
##Untuk X1
bftx1=t.test(X1[Z=="0"],X1[Z=="1"])
p.value_X1=bftx1$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))
  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)
)))

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

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
bftx4=t.test(X4[Z=="0"],X4[Z=="1"])
p.value_X4=bftx4$p.value
p.value_X4

```

```

##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

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

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))
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_X3,p.value_X4,p.value_
X5,p.value_X6))
colnames(summaryBALANCE)=c("p-value")
rownames(summaryBALANCE)=c("X1","X2","X3","X4","X5","X6")
round(summaryBALANCE,digits=3)

```

Sesudah (Strata 1)

```

strata1=read.csv("E:\\Data\\College Life\\TA\\strata1.csv",sep=";",header=TRUE)
attach(strata1)
##Untuk X1
bftx1=t.test(X1[Z=="0"],X1[Z=="1"])
p.value_X1=bftx1$p.value

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

##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))
  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))

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

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
bftx4=t.test(X4[Z=="0"],X4[Z=="1"])
p.value_X4=bftx4$p.value
p.value_X4

##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")]

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```
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))
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_X3,p.value_X4,p.value_
X5,p.value_X6))
colnames(summaryBALANCE)=c("p-value")
rownames(summaryBALANCE)=c("X1", "X2", "X3", "X4", "X5", "X6")
round(summaryBALANCE,digits=3)
```

Sesudah (Strata 2)

```
strata2=read.csv("E:\\Data\\College Life\\TA\\strata2.csv",sep=";",header=TRUE)
attach(strata2)
```

```
##Untuk X1
```

```
befx1=t.test(X1[Z=="0"],X1[Z=="1"])
```

```
p.value_X1=befx1$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")]
```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

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))
  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)

```


a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

p.value_X3=2*pnorm(-abs(zX3))
p.value_X3

##Untuk X4
bftx4=t.test(X4[Z=="0"],X4[Z=="1"])
p.value_X4=bftx4$p.value
p.value_X4

##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")

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```
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))
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_X3,p.value_X4,p.value_X5,p.value_X6))
colnames(summaryBALANCE)=c("p-value")
rownames(summaryBALANCE)=c("X1","X2","X3","X4","X5","X6")
round(summaryBALANCE,digits=3)
```

Sesudah (Strata 3)

```
strata3=read.csv("E:\\Data\\College Life\\TA\\strata3.csv",sep=";",header=TRUE)
attach(strata3)
##Untuk X1
bftx1=t.test(X1[Z=="0"],X1[Z=="1"])
p.value_X1=bftx1$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))
```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

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))
  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
bftx4=t.test(X4[Z=="0"],X4[Z=="1"])
p.value_X4=bftx4$p.value

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (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))
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))
```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

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_X3,p.value_X4,p.value_
X5,p.value_X6))
colnames(summaryBALANCE)=c("p-value")
rownames(summaryBALANCE)=c("X1","X2","X3","X4","X5","X6")
round(summaryBALANCE,digits=3)

```

Sesudah (Strata 4)

```

strata4=read.csv("E:\\Data\\College Life\\TA\\strata4.csv",sep=";",header=TRUE)
attach(strata4)
##Untuk X1
bftx1=t.test(X1[Z=="0"],X1[Z=="1"])
p.value_X1=bftx1$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))
  denominatorX2=sqrt(p.commonX2*(1-p.commonX2)*(1/(n2_0)+1/(n2_1)))
  z.prop.risX2=numeratorX2/denominatorX2
  return(z.prop.risX2)
}

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

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
bftx4=t.test(X4[Z=="0"],X4[Z=="1"])
p.value_X4=bftx4$p.value
p.value_X4

##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)
)))

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```

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)
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))

```

a. *Syntax Uji Kovariat Balance* Sebelum dan Sesudah dilakukan Stratifikasi (Lanjutan)

```
p.value_X6

summaryBALANCE=cbind(c(p.value_X1,p.value_X2,p.value_X3,p.value_X4,p.value_
X5,p.value_X6))
colnames(summaryBALANCE)=c("p-value")
rownames(summaryBALANCE)=c("X1","X2","X3","X4","X5","X6")
round(summaryBALANCE,digits=3)
```

b. *Output Uji Kovariat Balance*

	Sebelum	Strata 1	Strata 2	Strata 3	Strata 4
	p-value	p-value	p-value	p-value	p-value
X1	0.518	X1 0.382	X1 0.382	X1 0.537	X1 0.930
X2	0.008	X2 0.010	X2 NaN	X2 NaN	X2 NaN
X3	0.001	X3 NaN	X3 NaN	X3 NaN	X3 NaN
X4	0.283	X4 0.853	X4 0.853	X4 0.755	X4 0.672
X5	0.006	X5 0.047	X5 0.571	X5 0.725	X5 1.000
X6	0.019	X6 0.374	X6 NaN	X6 NaN	X6 0.527

Lampiran 8.

a. Syntax ATE.

```

dataate=read.csv("E:\\Data\\College Life\\TA\\dataate.csv",sep=";",header=TRUE)
attach(dataate)
=====
Y.QT0 <- split(dataate$Y[dataate$Z==0], dataate$Strata[dataate$Z==0])
Y.QT1 <- split(dataate$Y[dataate$Z==1], dataate$Strata[dataate$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
nc3=length(Y.QT0[[3]]) #n control untuk strata 3
nc4=length(Y.QT0[[4]]) #n control untuk strata 4
nt1=length(Y.QT1[[1]]) #n treatment untuk strata 1
nt2=length(Y.QT1[[2]]) #n treatment untuk strata 2
nt3=length(Y.QT1[[3]]) #n treatment untuk strata 3
nt4=length(Y.QT1[[4]]) #n treatment untuk strata 4
=====
nc=nc1+nc2+nc3+nc4 # n control
nt=nt1+nt2+nt3+nt4 # n treatment
=====
theta1=abs(((nt1+nc1)*(K[1]))/(nt+nc)) # ate untuk strata 1
theta2=abs(((nt2+nc2)*(K[2]))/(nt+nc)) # ate untuk strata 2
theta3=abs(((nt3+nc3)*(K[3]))/(nt+nc)) # ate untuk strata 3
theta4=abs(((nt4+nc4)*(K[4]))/(nt+nc)) # ate untuk strata 3
=====
theta_hat=(theta1+theta2+theta3+theta4) #ATE_final
theta_hat
theta1
theta2
theta3
theta4
=====
### mencari SE_ATE #####
Svar_treat=sapply(Y.QT1, var) # varian treat tiap strata Svar_treat
Svar_control=sapply(Y.QT0, var) # varian control tiap strata Svar_control
=====
Vt1=((1/nt1))*Svar_treat[1] # varians treatment strata 1
Vt2=((1/nt2))*Svar_treat[2] # varians treatment strata 2
Vt3=((1/nt3))*Svar_treat[3] # varians treatment strata 3
Vt4=((1/nt4))*Svar_treat[4] # varians treatment strata 4
=====
Vc1=((1/nc1))*Svar_control[1] # varians control strata 1
Vc2=0 # karena hanya ada 1 variabel jadi tidak bisa dihitung varians control strata 2
Vc3=((1/nc3))*Svar_control[3] # varians control strata 3
Vc4=((1/nc4))*Svar_control[4] # varians control strata 4

```

Lampiran 8.a. *Syntax* ATE.

```

var1=Vt1+Vc1 # var strata 1
var2=Vt2+Vc2 # var strata 2
var3=Vt3+Vc3 # var strata 3
var4=Vt4+Vc4 # var strata 4
=====
Var_theta1=((nt1+nc1)^2)*(var1) # var_ate untuk strata 1
Var_theta2=abs(((nt2+nc2)^2)*(var2)) # var_ate untuk strata 2
Var_theta3=((nt3+nc3)^2)*(var3) # var_ate untuk strata 3
Var_theta4=((nt4+nc4)^2)*(var4) # var_ate untuk strata 4
Var_theta_hat=(Var_theta1+Var_theta2+Var_theta3+Var_theta4)/((nt+nc)^2)
=====
#VarATE_final
se_theta_hat=abs(sqrt(Var_theta_hat)) # se ATE se_theta_hat
=====
## uj t stat ##
t=(theta_hat)/(se_theta_hat)
t
=====
## calculating estimate of df ##
numstrata=4
df = length((dataate$Z)) - (2*numstrata)
df
=====
## calculating p-value for t test ##
p_value=pt(-abs(t),df)
p_value

```

b. *Output* ATE.

```

> theta_hat 0.5737805
> theta1 0.2817073
> theta2 0.08231707
> theta3 0.06341463
> theta4 0.1463415

t
4.950191

df
74

p_value
2.26364e-06

```

BIODATA PENULIS



Penulis dengan nama lengkap Naomi Puspita Happy Puranti dilahirkan di Kabupaten Nganjuk Jawa Timur pada 17 Maret 1997. Penulis menempuh pendidikan formal di SDN Payaman 3 Nganjuk, SMPN 1 Nganjuk, dan SMAN 2 Nganjuk. Kemudian penulis diterima sebagai Mahasiswa Departemen Statistika ITS melalui jalur SNMPTN pada tahun 2015. Selama masa perkuliahan, penulis aktif di berbagai organisasi salah satunya adalah Himpunan Mahasiswa Statistika ITS 2015/2016 dan pernah menjadi ketua pelaksana *Explore Statistics* ITS 2016 yang merupakan kegiatan memperkenalkan, menjelajahi dan mengetahui lebih dalam tentang Jurusan Statistika ITS. Penulis juga aktif dalam organisasi bersama dosen-dosen Tim Pembinaan Kerohanian Kristen (TPKK) yaitu Persekutuan Mahasiswa Kristen (PMK-ITS), tahun ketiga penulis tergabung di PMK sebagai staff konseptor divisi Persekutuan 2017/2018, tahun berikutnya penulis diamanahi menjadi ketua divisi Persekutuan PMK-ITS 2018/2019. Selama menjalani masa perkuliahan, penulis juga aktif mengikuti kepanitiaan di ITS, seperti Pasar Malam Minggu ITS (PAMITS), Pekan Raya Statistika (PRS) dan Indonesia *Statistics Conference and Olympiad* (ISCO) yang diadakan oleh Ikatan Perstatistikaan Indonesia (ISI). Penulis juga pernah melakukan magang di PT.Astra Daihatsu Motor dan mengerjakan *project* tentang *customer profiling*, saat pelaksanaan magang penulis diberi kesempatan untuk berpartisipasi dalam GIIAS (Gaikindo International Auto Show) di ICE-BSD Tangerang 2018. Apabila pembaca ingin memberi kritik dan saran serta diskusi lebih lanjut mengenai Tugas Akhir ini, maka pembaca dapat menghubungi penulis melalui nomor telepon 081554875841 dan email naomizomz@gmail.com.