Analisis Propensity Score Matching Pada Kejadian Diabetes Melitus Yang Memuat Faktor Confounding

Naflah Faulina[1,*], Khoirin Nisa[1], Dorrah Aziz[1], dan Eri Setiawan[1]

[1]Jurusan Matematika, Fakultas MIPA, Universitas Lampung
Jl. Soemantri Brojonegoro 1 Bandar Lampung

***Email korespondensi: faulinanaflah@gmail.com**

Abstrak

Faktor *confounding* dapat didefinisikan sebagai bias dalam estimasi efek faktor risiko terhadap kejadian penyakit yang ingin diteliti. Salah satu metode yang dapat menangani faktor *confounding* adalah metode *propensity score matching* yang digunakan untuk menyeimbangan data kelompok perlakuan dan kontrol dengan melihat nilai pada hasil estimasi *propensity score* menggunakan regresi logistik, kemudian melakukan analisis *matching,* lalu melakukan *post-matching*. Tujuan dari penelitian ini adalah untuk mendapatkan hasil analisis *propensity score matching* yang memuat faktor *confounding* dan mengetahui faktor-faktor risiko yang menyebabkan diabetes melitus di RSD. Mayjend HM Ryacudu Kotabumi dengan bantuan software R. Data yang digunakan adalah status diabetes melitus (Y), jenis kelamin (x_1), usia (x_2), kadar glukosa darah (x_3), tekanan darah (x_4), kadar kolesterol (x_5), kadar asam urat serum (x_6), dan obesitas (x_7). Hasil analisis mendapatkan variabel *confounding* glukosa darah menghasilkan 49 pasangan pasien yang sesuai dan variabel tekanan darah, kadar kolestrol, obesitas seimbang, sehingga estimasi *average treatment of treated* sebesar 0,513 dengan *standar error* sebesar 0,103 dan mampu mereduksi bias sebesar 57,1%. Variabel yang berpengaruh langsung terhadap status diabtes melitus adalah kadar glukosa darah dan variabel yang tidak berpengaruh langsung terhadap status diabetes melitus yaitu tekanan darah, kadar kolestrol, dan obesitas.

Kata kunci: faktor *confounding,* analisis *propensity score matching,* diabtes melitus.

Abstract

The confounding factor can be defined as a bias in estimating the effect of a risk factor on the incidence of the disease being studied. One method that can handle confounding factors is propensity score matching which is used to balance the data for the treatment and control groups by looking at the value in the estimation results of the propensity score using logistic regression, then matching analysis and post-matching. The purpose of this study was to obtain the results of a propensity score matching analysis that contained confounding factors and knowing the risk factors that cause diabetes mellitus in Hospital Mayjend HM Ryacudu Kotabumi with R software. The data used are diabetes mellitus (Y), gender(x_1), age (x_2), blood glucose levels (x_3), blood pressure (x_4), cholesterol levels (x_5), serum uric acid levels (x_6), and obesity (x_7). The results of the analysis get obtaned blood glucose confounding variables with 49 suitable patient pairs and the variables of blood pressure, cholesterol levels, and obesity were balanced, so that the estimated average treatment of treated was 0,513 with a standard error of 0,103 and was able to reduce bias by 57,1%. The variables that had a direct effect on the status of diabetes mellitus were blood glucose levels and the variables that did not directly affect the status of diabetes mellitus were blood pressure, cholesterol levels, and obesity.

Keywords: confounding factor, propensity score matching analysis, diabetes.

1. Pendahuluan

Confounding merupakan suatu situasi ketika efek faktor risiko luar lainnya bercampur dengan efek dari faktor risiko utama sehingga menimbulkan gangguan antara faktor risiko utama dan penyakit yang diteliti, menyebabkan kesimpulan penelitian yang diperoleh tidak valid karena adanya seleksi bias [1].

Salah satu metode yang dapat menangani faktor *confounding* adalah metode *propensity score.* Metode *propensity score* pertama kali diperkenalkan oleh Rosenbaum & Rubin pada tahun 1983. *Propensity score* merupakan suatu metode probabilitas bersyarat dari perlakuan tertentu yang dapat meminimalisir bias akibat faktor *confounding* ketika menggunakan data observasi dengan menyesuaikan *propensity score* berdasarkan kovariat $(\boldsymbol{x_i})$ yang sama antara kelompok perlakuan $(\boldsymbol{Z_i = 1})$ dan kontrol $(\boldsymbol{Z_i = 0})$. Analisis *propensity score* juga dapat

menghilangkan ketidakseimbangan dalam penilitian observasi [2]. Salah satu metode dari *propensity score* adalah *propensity score matching*. *Propensity Score Matching* (PSM) digunakan untuk menyeimbangan data kelompok perlakuan dan kontrol dengan melihat nilai pada hasil estimasi *propensity score* menggunakan regresi logistik dengan *Maximum Likelihood Estimation* (MLE), kemudian melakukan analisis *matching*, yaitu pencocokan berdasarkan kovariat yang diamati menggunakan *Nearest Neigbor Matching* (NNM), lalu melakukan *post-matching*, yaitu evaluasi dari *propensity score matching*, seperti uji keseimbangan kovariat, estimasi *Average Treatment of Treated* (ATT), dan *Percent of Bias Reduction* (PBR) [3].

Pada penelitian ini bertujuan untuk mendapatkan hasil analisis *propensity score matching* pada kejadian diabetes melitus yang memuat faktor *confounding* di RSD. Mayjend HM Ryacudu Kotabumi dan mengetahui faktor-faktor risiko yang menyebabkan diabetes melitus.

2. Metodologi Penelitian

Beberapa tahapan dan teori dasar yang digunakan dalam penelitian ini adalah sebagai berikut.

2.1 Faktor Confounding

Istilah confounding berasal dari bahasa latin cunfundere berarti tercampur bersama. Syarat faktor confounding merupakan faktor resiko bagi kasus yang diteliti dan mempunyai hubungan dengan variabel bebas lainnya [1].

2.2 Propensity Score

Nilai *propensity score* dari observasi $\boldsymbol{i}(i = 1,2,..,n)$ sebagai probabilitas bersyarat dari penetapan perlakuan ($\boldsymbol{Z_i = 1}$) dan kontrol ($\boldsymbol{Z_i = 0}$) dengan vektor kovariat yang diamati $\boldsymbol{x_i}$ sebagai berikut.

$$e(x_i) = P(Z_i = 1|X_i = x_i)$$

Asumsi yang diberikan untuk $\boldsymbol{X_i}$ yaitu, $\boldsymbol{Z_i}$ independen:

$$P(z_1, z_2, \dots, z_n|x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} e(x_i)^{Z_i}\{1 - e(x_i)\}^{1-Z_i}$$

dengan $0 < e(X) = P(Z = 1|X = x) < 1$ untuk setiap $x \in X$ merupakan nilai peluang bersyarat suatu kelompok berdasarkan kovariat-kovariat yang diamati. Nilai ini dapat digunakan untuk mengendalikan bias karena ketidakseimbangan kovariat yang diamati [2].

2.3 Propensity Score Matching

Langkah-langkah dalam melakukan analisis mengunakan metode *propensity score matching* adalah sebagai berikut [3].

a) Menentukan variabel *confounding.*
   Variabel *confounding* yang dinotasikan **Z** menggunakan uji *chi-square* dengan hipotesis:
   H_0 : Tidak ada hubungan yang signifikan antar variabel yang berpotensi sebagai *confounding* (**Z**) dan variabel lainnya.
   H_1 : Ada hubungan yang signifikan antar variabel yang berpotensi sebagai *confounding* (**Z**) dan variabel lainnya.
   Dapat menggunakan rumus untuk uji statistik sebagai berikut.

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{k}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

   dengan,
   O_{ij} : jumlah observasi untuk kasus-kasus yang dikategorikan dalam baris ke-i dan kolom ke-j.
   E_{ij} : jumlah kasus yang diharapkan di bawah H0 untuk dikategorikan dalam baris ke-i dan kolom ke-j dengan $E_{ij} = \frac{n_i n_j}{N}$, $i = 1,2, \dots, r$; $j = 1,2, \dots, k$ n_i jumlah baris ke-i , n_j jumlah kolom ke-j, *N* total frekuensi.
   Daerah penolakan: Tolah H_0 jika χ^2hitung $\geq$ χ^2 α; $(r-1)(k-1)$ atau *p-value* $\leq$ α [2].

b) Estimasti *propensity score* menggunakan regresi logistik dengan metode *maximum likelihood estimation.*
   Propensity score menggunakan model regresi logistik dengan variabel terikat adalah biner ($\boldsymbol{Z_i} = 1$ untuk perlakuan dan $\boldsymbol{Z_i} = 0$ untuk kontrol) dengan model sebagai berikut.

$$e(x_i) = P(Z_i = 1|X_i = x_i) = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k\right)}$$

$$\left(e(x_i)\right)\left(1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)\right) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$

$$\big(e(x_i)\big) + (e(x_i)\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$
$$e(x_i) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) - e(x_i)\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$
$$e(x_i) = (1 - e(x_i))(\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k))$$
$$\frac{e(x_i)}{(1 - e(x_i))} = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$
$$ln\left(\frac{e(x_i)}{1 - e(x_i)}\right) = ln(\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k))$$
$$ln\left(\frac{e(x_i)}{1 - e(x_i)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$
$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

dengan,
β_0 : konstanta
$\beta_1, \beta_2, \dots, \beta_k$: koefisien regresi
$x_1, x_2, \dots, x_k$: kovariat

c) Matching dengan metode *nearest neighbor matching.*
Metode *nearest neighbor matching* jika nilai mutlak dari *propensity score* minimum diantara semua pasangan nilai *propensity score* antara 1 dan 0, maka

$$A(P_i) = \min\big||P_1 - P_0|\big|, \qquad 0 \in I_0$$

dengan,
$A(P_i)$: memuat setiap peserta pada kelompok kontrol ($0 \in I_0$) sebagai pasangan dari peserta pada kelompok perlakuan ($1 \in I_1$).
P_1 : nilai *propensity score* dari kelompok perlakuan.
P_0 : nilai *propensity score* dari kelompok kontrol.
I_1 : himpunan peserta pada kelompok perlakuan.
I_0 : himpunan peserta pada kelompok kontrol.

d) Post-matching yaitu, menguji keseimbangan variabel ($\boldsymbol{x}_i$) dan estimasi *Average Treatment of Treated* (ATT). Statistik uji untuk estimasi $\hat{\theta}$ adalah sebagai berikut

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

$\hat{\theta}$ merupakan ATT dan $SE(\hat{\theta})$ merupakan standar *error average treatment of treated*. Daerah penolakan: Tolak H_0 jika $|T_{hit}| > T_{\frac{\alpha}{2}}$ atau $p - value \leq \alpha$ [4].

e) Percent of Bias Reduction (PBR).
Untuk melihat seberapa besar bias dapat direduksi, dengan

$$PBR = \frac{|B_{\text{sebelum PSM}} - B_{\text{setelah PSM}}|}{|B_{\text{sebelum PSM}}|} \times 100\%$$

dengan,

$$B = p_1(x_p) - p_0(x_p)$$

B merupakan selisih rata-rata kelompok perlakuan dan kontrol untuk setiap kovariat, $p_1(x_p)$ dan $p_0(x_p)$ adalah proporsi dari kovariat untuk kelompok, $B_{\text{sebelum PSM}}$ dan $B_{\text{setelah PSM}}$ merupakan perbedaan antara rata-rata perlakuan dan kontrol sebelum *propensity score* dan setelah *propensity score* [3].

2.4 Diabetes Melitus

$$X_{ij} = \sqrt{(1 - \rho^2)}\, x_{ij} + \rho x_{ip}$$

dengan $i = 1, 2, 3, \dots, n$ dan $j = 1, 2, 3, \dots, p$. Adapun x_{ij} dibangkitkan berdistribusi normal dengan μ dan σ ditentukan. Berikut adalah langkah-langkah simulasi yang dilakukan

a) Membangkitkan data $\boldsymbol{X_i} \sim N_p(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$, $\boldsymbol{X_i} = [X_1, X_2, X_3, \dots, X_p]^t$ adalah vektor pengamatan dengan $i = 1, 2, 3, \dots$ dibangkitkan secara acak yang kemudian dikonversi menjadi data multikolinearitas dengan $\rho^2 = 0.96$ dan ketentuan:
   1) Data ke-1 dibangkitkan dengan $n = 10$ yang membentuk dua klaster. Klaster pertama dengan $n = 5$ berdistribusi $X_j \sim N(0,1)$ sehingga $\boldsymbol{X_1} \sim N_4(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$ dengan $\boldsymbol{\mu_1} = [0,0,0,0]^t$ dan $\boldsymbol{\Sigma_1} = 1\mathbf{I_4}$ dimana $j =$

$1, 2, 3, 4$. Klaster kedua dengan $n = 5$ objek berdistribusi $X_j \sim N(5,2)$ sehingga $\boldsymbol{X_2} \sim N_4(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})$ dengan $\boldsymbol{\mu_2} = [5,5,5,5]^t$ dan $\boldsymbol{\Sigma_2} = 2\mathbf{I_4}$ dimana $j = 1, 2, 3, 4$ yang kemudian gabungan data $\boldsymbol{X_1}$ dan $\boldsymbol{X_2}$ dikonversi menjadi data multikolinearitas dengan menggunakan persamaan (9)

2) Data ke-2 dibangkitkan dengan $n = 20$ yang membentuk tiga klaster. Klaster pertama dengan $n = 7$, berdistribusi sama dengan klaster pertama pada poin a. Klaster kedua dengan $n = 7$, berdistribusi sama dengan klaster kedua pada poin a. Klaster ketiga dengan $n = 6$ berdistribusi $X_j \sim N(8,3)$ sehingga $\boldsymbol{X_3} \sim N_4(\boldsymbol{\mu_3}, \boldsymbol{\Sigma_3})$ dengan $\boldsymbol{\mu_3} = [8,8,8,8]^t$ dan $\boldsymbol{\Sigma_3} = 3\mathbf{I_4}$ dimana $j = 1, 2, 3, 4$ yang kemudian gabungan data $\boldsymbol{X_1}$, $\boldsymbol{X_2}$ dan $\boldsymbol{X_3}$ dikonversi menjadi data multikolinearitas dengan menggunakan persamaan (9)
3) Data ke-3 dibangkitkan dengan $n = 50$ yang membentuk empat klaster. Klaster pertama dengan $n = 12$, berdistribusi sama dengan klaster pertama pada poin a. Klaster kedua dengan $n = 12$, berdistribusi sama dengan klaster kedua pada poin a. Klaster ketiga dengan $n = 13$, berdistribusi sama dengan klaster ketiga pada poin b. Klaster keempat dengan $n = 13$ berdistribusi $X_j \sim N(10,4)$ sehingga $\boldsymbol{X_4} \sim N_4(\boldsymbol{\mu_4}, \boldsymbol{\Sigma_4})$ dengan $\boldsymbol{\mu_4} = [10,10,10,10]^t$dan $\boldsymbol{\Sigma_4} = 4\mathbf{I_4}$ dimana $j = 1, 2, 3, 4$ yang kemudian gabungan data $\boldsymbol{X_1}$, $\boldsymbol{X_2}$, $\boldsymbol{X_3}$ dan $\boldsymbol{X_4}$ dikonversi menjadi data multikolinearitas dengan menggunakan persamaan (9)
4) Data ke-4 dibangkitkan dengan $n = 100$ yang membentuk lima klaster. Klaster pertama dengan $n = 20$, berdistribusi sama dengan klaster pertama pada poin a. Klaster kedua dengan $n = 20$, berdistribusi sama dengan klaster kedua pada poin a. Klaster ketiga dengan $n = 20$, berdistribusi sama dengan klaster ketiga pada poin b. Klaster keempat dengan $n = 20$, berdistribusi sama dengan klaster keempat pada poin c. Klaster kelima dengan $n = 20$ berdistribusi $X_j \sim N(13,3)$ sehingga $\boldsymbol{X_5} \sim N_4(\boldsymbol{\mu_5}, \boldsymbol{\Sigma_5})$ dengan $\boldsymbol{\mu_5} = [13,13,13,13]^t$dan $\boldsymbol{\Sigma_5} = 3\mathbf{I_4}$ dimana $j = 1, 2, 3, 4$ yang kemudian gabungan data $\boldsymbol{X_1}$, $\boldsymbol{X_2}$, $\boldsymbol{X_3}$, $\boldsymbol{X_4}$ dan $\boldsymbol{X_5}$ dikonversi menjadi data multikolinearitas dengan menggunakan persamaan (9).

b) Standarisasi data kedalam bentuk nilai Z
c) Melakukan uji asumsi multikolinearitas (VIF)
d) Mengatasi data yang mengandung multikolinearitas menggunakan analisis komponen utama (AKU)
e) Melakukan pengklasteran dengan menggunakan metode *average linkage* dan metode Ward dengan menggunakan data hasil analisis komponen utama
f) Menghitung dan mencatat indeks Dunn pada tiap metode
g) Menghitung dan mencatat indeks RS pada tiap metode
h) Mengulang langkah 1 (satu) sampai langkah 7 (tujuh) sebanyak 1000 (seribu) kali pengulangan
i) Melakukan evaluasi indeks Dunn dan indeks RS dengan menghitung rata-ratanya
j) Analisis hasil

3. Hasil dan Pembahasan

Penelitian ini menggunakan data sekunder yaitu data rekam medis diabetes melitus yang diperoleh dari RSD. Mayjend Hm Ryacudu Kotabumi pada bulan Agustus 2020. Gambaran umum dari data yang digunakan dalam analisis 123 pasien yang melakukan pemeriksaan kesehatan mengenai status diabetes melitus (Y) beserta faktor-faktor yang mempengaruhi diabetes melitus, diantaranya jenis kelamin (x_1), usia (x_2), kadar glukosa darah (x_3), tekanan darah (x_4), kadar kolesterol (x_5), kadar asam urat serum (x_6),dan obesitas (x_7). Karakteristik pasien sebagai berikut.

Tabel 1 Karakteristik data diabetes melitus

Variabel	Skala	Kategori	
		Kontrol	Perlakuan
Status Diabetes Melitus (Y)	Nominal	0 = Tidak diabetes	1 = Diabetes tipe I,II, dan Gestational
Jenis Kelamin (x_1)	Nominal	0 = Laki-laki	1 = Perempuan
Kadar Glukosa Darah (x_2)	Nominal	0 = Rendah ($<$ 200 mg/dL)	1 = Tinggi ($\geq$ 200 mg/dL)
Usia (x_3)	Nominal	0 = $<$ 45 tahun	1 = $\geq$ 45 tahun
Tekanan Darah (x_4)	Nominal	0 = Rendah ($<$ 130 mmHg)	1 = Tinggi ($\geq$ 130 mmHg)
Kadar Kolestrol (x_5)	Nominal	0 = Rendah ($<$ 250 mg/dL)	1 = Tinggi ($\geq$ 250 mg/dL)
Kadar Asam Urat Serum (x_6)	Nominal	0 = Rendah ($<$ 7 mg/dL)	1 = Tinggi ($\geq$ 7 mg/dL)
Obesitas (x_7)	Nominal	0 = Tidak (BMI $<$ 27 kg/m^2)	1 = Ya (BMI $\geq$ 27 kg/m^2)

Berdasarkan data pada Tabel 1 tersebut, kemudian dilakukan analisis data menggunakan analisis *propensity score matching* dengan bantuan *software* R. Tahapan yang dilakukan adalah sebagai berikut.

Menentukan Variabel *Confounding*

Tahapan yang pertama dalam analisis *propensity score matching* adalah menentukan variabel *confounding* dengan menggunakan uji *chi-square* untuk mendapatkan hubungan antar variabel. Berdasarkan pengujian hipotesis yang berpotensi sebagai *confounding* (**Z**) adalah variabel kadar glukosa darah karena paling banyak menghasilkan terdapat hubungan antar variabel dengan pengujian hipotesis sebagai berikut.

Tabel 2 Pengujian hipotesis variabel kadar glukosa darah dengan variabel lainnya

Variabel	χ^2	Df	*P-value*	Keputusan
Kadar Glukosa Darah-Jenis Kelamin	0,164	1	0,685	Gagal Tolak H_0
Kadar Glukosa Darah-Usia	0,427	1	0,514	Gagal Tolak H_0
Kadar Glukosa Darah-Tekanan Darah	5,444	1	0,019	Tolak H_0
Kadar Glukosa Darah-Kadar Kolestrol	4,420	1	0,036	Tolak H_0
Kadar Glukosa Darah-Kadar AUS	0,783	1	0,376	Gagal Tolak H_0
Kadar Glukosa Darah-Obesitas	5,343	1	0,021	Tolak H_0
Kadar Glukosa Darah-Status DM	27,456	1	0,000	Tolak H_0

H_0 : Tidak terdapat hubungan yang signifikan antara variabel kadar glukosa darah dan variabel lainnya.
H_1 : Terdapat hubungan yang signifikan antara variabel kadar glukosa darah dan variabel lainnya.

Dengan taraf signifikan 0,05 dan aturan keputusan tolak H_0 jika χ^2hitung $\geq \chi^2_{\alpha;(r-1)(k-1)}$ atau *p-value* $\leq \alpha$. Menghasilkan variabel kadar glukosa darah terhadap variabel jenis kelamin, usia, dan asam urat serum *p-value* > 0,05 dan χ^2hitung < 3,84 maka gagal tolak H_0 dengan kata lain terima H_0. Sedangkan terhadap variabel tekanan darah, kadar kolestrol, obesitas dan status diabetes melitus menghasilkan *p-value* $\leq$ 0,05 dan χ^2hitung $\geq$ 3,84 maka tolak H_0 dengan kata lain terima H_1. Sehingga dapat disimpulkan tidak terdapat hubungan antara variabel kadar glukosa darah dan variabel jenis kelamin, usia, asam urat serum. Sedangkan terdapat hubungan antara variabel kadar glukosa darah dan variabel tekanan darah, kadar kolestrol, obesitas dan status diabetes melitus.

Estimasi *Propensity Score*

Estimasi nilai *propensity score* menggunakan regresi logistik biner dengan metode *Maximum Likelihood Estimation* (MLE) mendapatkan hasil sebagai berikut.

Tabel 3 Hasil estimasi *propensity score* menggunakan regresi logistik biner dengan MLE.

Variabel	$\hat{\beta}$	SE	*P-value*	OR
Intercept	-1,117	0,591	0,059	0,327
Jenis Kelamin.1 ($x_{1.1}$)	-0,320	0,414	0,439	0,726
Usia.1 ($x_{3.1}$)	0,080	0,419	0,849	1,083
Tekanan Darah.1 ($x_{4.1}$)	1,007	0,419	0,016	2,737
Kadar Kolestrol.1 ($x_{5.1}$)	0,982	0,417	0,018	2,669
Kadar AUS.1 ($x_{6.1}$)	0,368	0,419	0,381	1,445
Obesitas.1 ($x_{7.1}$)	-0,910	0,424	0,032	0,403

Berdasarkan Tabel 3 di peroleh model regresi logistik sementara,

$$e(Z) = \frac{\exp(-1{,}117 \; - 0{,}320\, x_{1.1} + 0{,}080\, x_{3.1} + 1{,}007\, x_{4.1} + 0{,}982\, x_{5.1} + 0{,}368\, x_{6.1} - 0{,}910\, x_{7.1})}{1 + \exp(-1{,}117 \; - 0{,}320\, x_{1.1} + 0{,}080\, x_{3.1} + 1{,}007\, x_{4.1} + 0{,}982\, x_{5.1} + 0{,}368\, x_{6.1} - 0{,}910\, x_{7.1})}$$

Analisis *Matching*

Analisis *matching* menggunakan metode *nearest neighbor matching* untuk mencocokkan data pada kelompok perlakuan dengan data kelompok kontrol.

Tabel 4 Hasil *matching* menggunakan metode NNM.

	Control	*Treated*
All	74	49
Matched	49	49
Unmatched	25	0
Discarded	0	0

Berdasarkan Tabel 4 dari 123 pasien hanya 74 pasien dengan kadar glukosa darah rendah (kontrol) yang dipasangkan ke pasien dengan kadar glukosa darah tinggi (perlakuan) sehingga menghasilkan 49 pasangan pasien sedangkan pasien dengan kadar glukosa darah rendah yang tidak memiliki pasangan sebanyak 25 pasien, sehingga dikeluarkan dan tidak diikutkan pada analisis selanjutnya.

Analisis *Post-Matching*

Terdapat dua langkah dalam analisis *post-matching*, yaitu melakukan uji keseimbangan dan estimasi *Average Treatment of Treated* (ATT).

a. Uji keseimbangan

Evaluasi ini untuk melihat keseimbangan variabel (x_i) pada variabel *confounding* antara kelompok perlakuan dan kontrol.

Tabel 5 Hasil uji keseimbangan

Variabel	Tahapan	*Before Matching*	*After Matching*
Jenis Kelamin (x_1)	*Mean treatment*	0,571	0,571
	Mean control	0,608	0,693
	p-value	0,689	0,011
Usia (x_3)	*Mean treatment*	0,653	0,653
	Mean control	0,594	0,448
	p-value	0,515	0,005
Tekanan Darah (x_4)	*Mean treatment*	0,673	0,673
	Mean control	0,459	0,755
	p-value	0,018	0,099
Kadar Kolestrol (x_5)	*Mean treatment*	0,469	0,469
	Mean control	0,283	0,408
	p-value	0,040	0,317
Kadar AUS (x_6)	*Mean treatment*	0,673	0,673
	Mean control	0,594	0,795
	p-value	0,376	0,011
Obesitas (x_7)	*Mean treatment*	0,265	0,265
	Mean control	0,472	0,285
	p-value	0,017	0,656

H_0 : Tidak terdapat perbedaan proporsi variabel (x_i) antara kelompok perlakuan dan kelompok kontrol.
H_1 : Terdapat perbedaan proporsi variabel (x_i) antara kelompok perlakuan dan kelompok kontrol.

Berdasarkan Tabel 5 dengan menggunakan taraf signifikansi 0,05 dan aturan keputusan tolak H_0 jika *p-value* $\leq \alpha$. Karena variabel jenis kelamin (x_1), usia (x_3), dan asam urat serum (x_6) menghasilkan *p-value after matching* $\leq 0{,}05$ maka tolak H_0 dengan kata lain terima H_1. Dapat disimpulkan bahwa variabel jenis kelamin (x_1), usia (x_3), dan asam urat serum (x_6) tidak seimbang karena terdapat perbedaan proporsi variabel jenis kelamin (x_1), usia (x_3), dan asam urat serum (x_6) antara kelompok perlakuan dan kelompok kontrol, sehingga dilakukan estimasi ulang *propensity score* tanpa variabel yang tidak seimbang.

Estimasi Ulang *Propensity Score* Tanpa Variabel yang Tidak Seimbang

Tabel 6 Hasil estimasi ulang *propensity score* tanpa variabel yang tidak seimbang

Variabel	$\hat{\beta}$	SE	*P-value*	OR
Intercept	-0,987	0,387	0,011	0,373
Tekanan Darah.1 ($x_{4.1}$)	0,971	0,406	0,017	2,641
Kadar Kolestrol.1 ($x_{5.1}$)	0,954	0,412	0,021	2,596
Obesitas.1 ($x_{7.1}$)	-0,919	0,417	0,028	0,399

Berdasarkan Tabel 6 di peroleh model regresi logistik,

$$e(Z) = \frac{exp(-0{,}987 + 0{,}971\, x_{4.1} + 0{,}954\, x_{5.1} - 0{,}919\, x_{7.1})}{1 + exp\,(-0{,}987 + 0{,}971\, x_{4.1} + 0{,}954\, x_{5.1} - 0{,}919\, x_{7.1})}$$

Untuk mendapatkan model regresi logistik yang linier dalam parameter, dilakukan transformasi logit pada persamaan di atas dengan misalkan, $b = -0{,}987 + 0{,}971\, x_{4.1} + 0{,}954\, x_{5.1} - 0{,}919\, x_{7.1}$.

$$e(Z) = \frac{exp(\,b)}{1 + exp(\,b)}$$
$$e(Z)[1 + exp\,(b)] = exp(\,b)$$
$$e(Z) + e(Z)exp\,(b) = exp(\,b)$$
$$e(Z) = exp(b) - e(Z)exp\,(b)$$
$$e(Z) = [1 - e(Z)]\, exp(b)$$
$$\frac{e(Z)}{1 - e(Z)} = exp(b)$$
$$ln\frac{e(Z)}{1 - e(Z)} = ln\, exp(b)$$
$$ln\frac{e(Z)}{1 - e(Z)} = b$$

$$g(x) = In\left(\frac{e(Z)}{1 - e(Z)}\right) = -0{,}987 + 0{,}971\, x_{4.1} + 0{,}954\, x_{5.1} - 0{,}919\, x_{7.1}$$

diperoleh model regresi logistik sementara sebagai berikut.

$$g(x) = -0{,}987 + 0{,}971\, x_{4.1} + 0{,}954\, x_{5.1} - 0{,}919\, x_{7.1}$$

Analisis *Post-matching* Tanpa Variabel yang Tidak Seimbang

a. Uji Keseimbangan Tanpa Variabel yang Tidak Seimbang.

Evaluasi ini untuk melihat keseimbangan variabel (x_i) pada variabel *confounding* antara kelompok perlakuan dan kontrol dengan pengujian hipotesis sebagai berikut.

Tabel 7 Hasil uji keseimbangan tanpa variabel yang tidak seimbang.

Variabel	Tahapan	*Before Matching*	*After Matching*
Tekanan Darah (x_4)	*Mean treatment*	0,673	0,673
	Mean control	0,459	0,673
	p-value	0,018	1
Kadar Kolestrol (x_5)	*Mean treatment*	0,469	0,469
	Mean control	0,283	0,469
	p-value	0,040	1
Obesitas (x_7)	*Mean treatment*	0,265	0,265
	Mean control	0,472	0,265
	p-value	0,017	1

H_0 : Tidak terdapat perbedaan proporsi variabel (x_i) antara kelompok perlakuan dan kelompok kontrol.
H_1 : Terdapat perbedaan proporsi variabel (x_i) antara kelompok perlakuan dan kelompok kontrol.

Berdasarkan Tabel 7 dengan menggunakan taraf signifikansi 0,05 dan aturan keputusan tolak H_0 jika *p-value* $\leq \alpha$. Dapat disimpulkan bahwa setiap variabel menghasilkan *p-value after matching* lebih besar dari 0,05 sehingga variabel tekanan darah (x_4), kadar kolestrol (x_5), dan obesitas (x_7) seimbang, yang artinya tidak terdapat perbedaan proporsi antara kelompok perlakuan dan kontrol. Tahapan selanjutnya menghitung *Average Treatment of Treated (ATT).*

b. *Average Treatment of Treated* Tanpa Variabel yang Tidak Seimbang.

ATT dilakukan untuk mengetahui seberapa besar pengaruh kadar glukosa darah (**Z**) terhadap status diabetes melitus (Y) pada saat pengaruh dari variabel lain sudah direduksi. Hasil ATT sebagai berikut.

Tabel 8 Hasil estimasi *Average Treatment of Treated* (ATT).

Variabel	ATT	SE	*P-value*
$(x_4),\ (x_5),\ (x_7)$	0,513	0,103	0,000

Berdasarkan Tabel 8 diperoleh estimasi ATT sebesar 0,513 dengan *standar error* sebesar 0,103. Pada uji signifikansi dengan hipotesis.

H_0 : Variabel *confounding* (**Z**) kadar glukosa darah tidak signifikan
H_1 : Variabel *confounding* (**Z**) kadar glukosa darah signifikan
Taraf signifikansi : 0,05
Statistik uji : $T = \frac{\hat{\theta}}{SE(\hat{\theta})}$
Aturan keputusan : Tolak H_0 jika *p-value* $\leq \alpha$.
Keputusan : Karena p-value (0,000) $\leq$ 0,05 maka tolak H_0
Kesimpulan : Variabel *confounding* (**Z**) kadar glukosa darah signifikan.

Hal ini menunjukkan bahwa kadar glukosa darah berpengaruh signifikan terhadap status diabetes melitus dengan pengaruh sebesar 0,513 yang artinya, peluang pasien memiliki kadar glukosa darah tinggi untuk terdiagnosis diabetes melitus adalah 0,513 kali lebih besar dibandingkan pasien yang memiliki kadar glukosa darah rendah.

***Percent of Bias Reduction* (PBR)**

Tabel 9 Hasil *Percent of Bias Reduction* (PBR)

Variabel	*Before Matching*			*After Matching*			*Percent of Bias Reduction*
	Means Treated	*Means Control*	*Mean Difference*	*Means Treated*	*Means Control*	*Mean Difference*	
distance	0,473	0,349	0,124	0,473	0,419	0,053	57,1%
$X_{4.0}$	0,326	0,540	-0,214	0,326	0,346	-0,020	90,4%
$X_{4.1}$	0,673	0,459	0,214	0,673	0,653	0,020	90,4%
$X_{5.0}$	0,530	0,716	-0,185	0,530	0,591	-0,061	67,0%
$X_{5.1}$	0,469	0,283	0,185	0,469	0,408	0,061	67,0%
$X_{7.0}$	0,734	0,527	0,207	0,734	0,571	0,163	21,3%
$X_{7.1}$	0,265	0,473	-0,207	0,265	0,428	-0,163	21,3%

Berdasarkan Tabel 9 dengan tekanan darah (x_4), kadar kolestrol (x_5), dan obesitas (x_7) mendapatkan PBR sebesar 57,1% yang artinya analisis *propensity score matching* mampu mereduksi bias 57,1%..

4. Kesimpulan

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan bahwa analisis klaster metode *average linkage* dan Ward pada data yang mengandung multikolinearitas dapat diatasi dengan analisis komponen utama. Berdasarkan nilai indeks Dunn, metode *average linkage* memberikan hasil yang lebih baik dibandingkan metode Ward dalam pengklasteran data. Ukuran indeks Dunn berlandaskan pada fakta bahwa klaster yang terpisah itu biasanya memiliki jarak antar klaster yang besar dan diameter intra klaster yang kecil, yang berarti klaster-klaster

yang dibentuk oleh metode *average linkage* memiliki jarak antar klaster yang lebih besar dan diameter intra klaster yang lebih kecil dibandingkan metode Ward. Berdasarkan nilai indeks RS, metode Ward memberikan hasil yang lebih baik dibandingkan metode *average linkage* dalam pengklasteran data. Indeks RS mengukur apakah karakteristik antar klaster saling berbeda, yang berarti klaster yang terbentuk dengan menggunakan metode Ward memiliki karakteristik yang lebih berbeda dibanding dengan metode *average linkage*

Daftar Pustaka:

[1] Usman, H. dan Nurdin, S. 2013. Aplikasi Teknik Multivariate untuk Riset Pemasaran. PT. Raja Grafindo Persada, Jakarta.
[2] Hair, J.F., Black, W.C., Babin, B.J. dan Anderson, R.E. 2014. *Multivariate Data Analysis*. 7th Edition. Pearson Education Limited, England.
[3] Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan*. UPP STIM YKPN, Yogyakarta.
[4] Supranto. 2004. *Analisis Multivariat: Arti dan Interpretasi*. Rineka Cipta, Jakarta.
[5] Johnson, R. dan Wichern, D. 2007. *Applied Multivariate Analysis*. 6th Edition. Prentice Hall Inc., New Jersey
[6] Satoto, B.D., Khotimah, B.K. dan Muhammad, A. 2015. Pengelompokan Tingkat Kesehatan Masyarakat Menggunakan *Shelf Organizing Maps* dengan *Cluster Validation Idb* dan *I-Dunn*. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi) 2015.*
[7] Sharma, S. 1996. *Applied Multivariate Techniques*. A John Wiley & Sons, Inc., Canada.
[8] Kibria, B. dan Muniz, G. 2009. On Some Ridge Regression Estimator: An Emprical Comparison. *Communication in Statistics – Simulation and Computation*. **38**: 621-630.