

Penanganan *Data Missing* dengan Algoritma *Multivariate Imputation By Chained Equations* (MICE)

Fitria Dewi Herliani^{1✉}, Abdul Kudus²

⁽¹⁾Statistika, Universitas Islam Bandung, Indonesia

⁽²⁾ Statistika, Universitas Islam Bandung, Indonesia

DOI: xx.xxxxx/datamath.vxix.xxx (DOI address)

ABSTRACT

In the pre-processing and data analysis stages, variables with missing or incomplete values are often found. This phenomenon is called missing data. Missing data is a problem that must be addressed because it can cause the estimation results to be biased and inaccurate. To handle missing data, we can use the imputation method. This study aims to find imputation values to handle missing data in the case of Hepatocellular Carcinoma (HCC) data with multiple imputation methods, namely the Multivariate Imputation by Chained Equations (MICE) algorithm. The MICE method uses a series of regression models. The advantage of this method is that it can handle various types of data, such as continuous data using the Predictive Mean Matching (PMM) method and binary data with the logistic regression method. Overall, the MICE algorithm with three imputations and 20 iterations produced convergent final imputation values for each variable that experienced missing data in the HCC data and produced complete data.

Keywords: *Missing data; imputation; MICE; HCC.*

ABSTRAK

Dalam tahap pre processing dan analisis data seringkali ditemukan variabel yang nilainya hilang atau tidak lengkap. Fenomena tersebut disebut dengan data missing. Data Missing merupakan suatu masalah yang harus ditangani karena dapat mengakibatkan hasil pendugaan menjadi bias dan tidak akurat. Untuk menangani data missing dapat menggunakan metode imputasi. Tujuan penelitian ini adalah mencari nilai imputasi untuk menangani data missing pada kasus data Hepatocellular Carcinoma (HCC) dengan metode multiple imputation yaitu algoritma Multivariate Imputation by Chained Equations (MICE). Metode MICE menggunakan serangkaian model regresi. Kelebihan dari metode ini adalah dapat menangani berbagai jenis data, seperti data kontinu menggunakan metode Predictive Mean Matching (PMM) dan data biner dengan metode regresi logistik. Secara keseluruhan, algoritma MICE dengan 3 kali imputasi dan iterasi sebanyak 20 kali menghasilkan nilai imputasi akhir yang konvergen untuk setiap variabel yang mengalami data missing pada data HCC dan menghasilkan data yang lengkap.

Keywords: *Data Missing; Imputasi; MICE; HCC.*

PENDAHULUAN

Dalam tahap *pre processing* dan analisis data seringkali ditemukan variabel yang nilainya hilang atau tidak lengkap untuk satu atau beberapa subjek penelitian. Fenomena tersebut disebut dengan data *missing*. Data *missing* merupakan hal yang umum terjadi dalam sebuah penelitian. Data *missing* terjadi ketika nilai dari variabel yang diamati tidak dicatat untuk semua subjek dalam sampel. Data *missing* dapat disebabkan karena alat ukur yang tidak berfungsi dengan baik, kekurangan fasilitas, serta penolakan atau kehilangan responden untuk menanggapi pertanyaan tertentu sehingga tidak terisinya kuesioner oleh responden. Data *missing* merupakan suatu masalah yang harus ditangani karena dapat mengakibatkan hasil pendugaan menjadi bias dan tidak akurat. Menurut Rubin [1], mekanisme data *missing* terbagi menjadi tiga kategori, yaitu *Missing Completely at Random* (MCAR), *Missing at Random* (MAR), dan *Missing Not at Random* (MNAR).

Suatu pendekatan yang populer untuk mengatasi data *missing* adalah melakukan imputasi data *missing*. Pendekatan ini dilakukan dengan cara “imputasi”, dengan memasukkan nilai data *missing* untuk subjek yang datanya hilang pada variabel tersebut. Salah satu metode yang umum digunakan adalah *multiple imputation*. *Multiple imputation* adalah perkembangan dari metode sebelumnya yaitu *single imputation*. Dalam *single imputation*, data *missing* diisi dengan menghitung nilai tunggal (misalnya, rata-rata) untuk semua data, lalu nilai tunggal (rata-rata) ini diisikan pada setiap data *missing* pada variabel tersebut. Kelemahan metode ini bersifat bias karena memperlakukan nilai imputasi sebagai nilai sebenarnya. Sedangkan dalam *multiple imputation* memiliki banyak keuntungan karena kemungkinan hasil m imputasi yang lebih dari satu dan dapat digunakan pada persentase data *missing* yang besar [2]. Salah satu metode *multiple imputation* adalah algoritma *Multivariate Imputation by Chained Equations* (MICE). Kelebihan metode MICE adalah dapat menangani variabel dengan berbagai jenis data, seperti data kontinu dan data biner [3]. Proses imputasi menggunakan regresi linier dan mengambil nilai yang diimputasi sebagai penarikan acak dari distribusi normal. Hal tersebut akan memiliki masalah jika residual dari regresi tidak berdistribusi secara normal sehingga mungkin memiliki nilai imputasi yang negatif. Untuk menangani masalah tersebut digunakan metode *Predictive Mean Matching* (PMM) untuk mengimputasi variabel kontinu. Sedangkan, untuk mengimputasi variabel biner menggunakan metode regresi logistik. Metode regresi logistik adalah suatu metode yang optimal untuk analisis regresi dengan variabel biner [4]. Metode MICE mengasumsikan bahwa data *Missing At Random* (MAR) [5].

Umumnya, dalam penelitian klinis sering terjadi data *missing*. Salah satu data klinis yaitu *Hepatocellular Carcinoma* (HCC). *Hepatocellular Carcinoma* (HCC) merupakan salah satu jenis penyakit kanker hati yang harus dihadapi di seluruh dunia. HCC merupakan penyebab keempat terserang kematian akibat kanker di seluruh dunia [6]. Hal tersebut menjadi sangat penting untuk diamati. Jika data HCC terdapat persentase data *missing*, maka harus segera ditangani. Dalam menangani masalah tersebut tidak dapat ditangani dengan cara sederhana, maka untuk menduga nilai data *missing* tersebut menggunakan metode *multiple imputation* yaitu *Multivariate Imputation by Chained Equations* (MICE).

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana melakukan imputasi data *missing* dengan menggunakan algoritma *Multivariate Imputation by Chained Equations* (MICE) pada kasus data *Hepatocellular Carcinoma* (HCC)?”. Selanjutnya, tujuan dalam penelitian ini adalah mencari nilai imputasi untuk menangani data *missing* dengan menggunakan algoritma *Multivariate Imputation by Chained Equations* (MICE) pada kasus data *Hepatocellular Carcinoma* (HCC).

METODOLOGI

Peneliti menggunakan metode *multiple imputation* dengan algoritma *Multivariate Imputation by Chained Equations* (MICE). Metode MICE mengikuti serangkaian model regresi yang dijalankan [7]. Metode MICE memerlukan spesifikasi model imputasi untuk setiap variabel yang tidak lengkap, dan membuat imputasi per variabel secara iteratif [8].

Data yang digunakan adalah data kelangsungan hidup pada 165 pasien *hepatocellular carcinoma* (HCC) yang ada di *Coimbra's Hospital and University Centre* (CHUC). Data tersebut adalah data sekunder dan variabel yang digunakan terbagi menjadi 2 jenis, yaitu: (1) Variabel kontinu: *Age at diagnosis*, *Grams of alcohol per day*, *Packs of cigarets per year*, dan *Major dimension of nodule* (cm); (2) Variabel Biner: *Gender*, *Hepatitis B Surface Antigen* (HBSA), *Hepatitis C Virus Antibody* (HCVA), *Cirrhosis*, *Obesity*, *Hemochromatosis*, dan *Class*.

Proses imputasi dilakukan dengan $m = 3$ kali imputasi dan sebanyak 20 iterasi, jika variabel yang diimputasi adalah variabel kontinu maka menggunakan *Predictive Mean Matching* (PMM) sedangkan jika variabel yang diimputasi adalah variabel biner maka menggunakan metode regresi logistik.

Imputasi dengan *Predictive Mean Matching* (PMM) untuk Variabel Kontinu

Untuk menggunakan PMM, digunakan metode regresi linear untuk mengestimasi variabel kontinu yang berdistribusi normal sebagai berikut:

$$\theta | \mathbf{X}; \boldsymbol{\beta} \sim N(\boldsymbol{\beta} \mathbf{X}, \sigma^2) \tag{1}$$

Misalkan $\hat{\boldsymbol{\beta}}$ adalah estimasi parameter dari pencocokan model dengan individu θ yang diamati. Misalkan V adalah estimasi matriks kovarians dari $\hat{\boldsymbol{\beta}}$, dan $\hat{\sigma}$ adalah estimasi *Root Mean-Squared Error*. Selanjutnya, menarik parameter imputasi σ^* , $\boldsymbol{\beta}^*$ dari distribusi posterior gabungan dari σ , $\boldsymbol{\beta}$. Pertama, σ^* dihitung sebagai berikut:

$$\sigma^* = \hat{\sigma} \sqrt{(n_{obs} - k) / g} \tag{2}$$

Di mana g adalah pengambilan acak dari distribusi *chi-square* dengan derajat bebas $n_{obs} - k$, dan n_{obs} adalah jumlah individu yang diamati. Kedua, $\boldsymbol{\beta}^*$ dihitung sebagai berikut:

$$\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}} + \frac{\sigma^*}{\hat{\sigma}} \mathbf{u}_1 \mathbf{V}^{1/2} \tag{3}$$

Di mana \mathbf{u}_1 adalah vektor $k \times 1$ yang berdistribusi normal baku dari pengambilan acak dari distribusi normal baku dan $\mathbf{V}^{1/2}$ adalah dekomposisi *cholesky* dari V . Selanjutnya, PMM mengidentifikasi nilai *missing* dengan menghitung selisih nilai prediksi untuk data *missing*, yaitu:

$$\eta(i, j) = |\hat{\boldsymbol{\beta}} \mathbf{X}_{obs,[i]} - \boldsymbol{\beta}^* \mathbf{X}_{miss,[j]}| \tag{4}$$

Di mana X_{obs} adalah data yang diamati dan X_{miss} adalah data *missing* dengan $i = 1, \dots, n$ dan $j = 1, \dots, n_0$. Setelah menghitung nilai tersebut, mengumpulkan selisih nilai prediksi data *missing* sebanyak n_0 himpunan. Setiap himpunan n_0 berisikan d data. Kemudian, memilih d data lengkap yang memiliki selisih jarak minimum antar nilai prediksi. Salah satu dari d dipilih secara acak untuk imputasi data *missing*.

Imputasi dengan Model Regresi Logistik untuk Variabel Biner

Menurut White et al. [9], model regresi logistik biasa digunakan untuk mengimputasi variabel biner dengan persamaan sebagai berikut:

$$\text{logit } Pr(\theta = 1 | \mathbf{X}; \boldsymbol{\beta}) = \boldsymbol{\beta} \mathbf{X} \tag{5}$$

Misalkan $\hat{\boldsymbol{\beta}}$ adalah estimasi parameter dari pencocokan model ke individu dengan θ yang diamati, dengan estimasi matriks varians-kovarians V . Misalkan $\boldsymbol{\beta}^*$ adalah hasil acak dari distribusi posterior $\boldsymbol{\beta}$, dengan pendekatan $MVN(\hat{\boldsymbol{\beta}}, V)$. Untuk setiap pengamatan yang hilang θ_i , hitung nilai p_i^* sebagai berikut:

$$p_i^* = [1 + \exp(-\boldsymbol{\beta}^* \mathbf{X}_i)]^{-1} \tag{6}$$

Kemudian tarik nilai imputasi θ_i^* sebagai berikut:

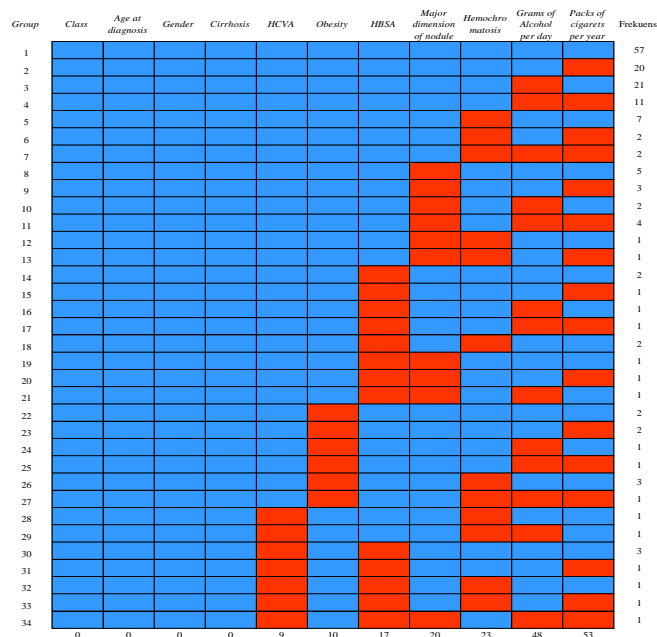
$$\theta_i^* = \begin{cases} 1, & \text{jika } u_i < p_i^* \\ 0, & \text{lainnya} \end{cases} \quad (7)$$

Dimana u_i adalah sebuah penarikan acak dari distribusi uniform (0, 1).

HASIL DAN PEMBAHASAN

Eksplorasi Data

Eksplorasi data dilakukan untuk melihat persebaran data *missing* dan statistika deskriptif pada masing-masing variabel. Pola data *missing* dapat dilihat pada gambar sebagai berikut.



Gambar 1. Pola Data Missing

Pada Gambar 1 pola yang berwarna biru adalah data lengkap dan pola yang berwarna merah adalah data *missing*. Terlihat bahwa pola data *missing* tersebar secara acak dan tidak membentuk suatu pola tertentu dengan jumlah data yang lengkap pada seluruh variabel sebanyak 57 pengamatan dan jumlah data *missing* tertinggi untuk seluruh variabel sebanyak 1 pengamatan. Selain itu, jumlah data *missing* tertinggi terdapat pada variabel *Packs of cigarettes per year* yaitu 53 pengamatan sedangkan variabel yang tidak memiliki data *missing* terdapat pada variabel *Class*, *Age at diagnosis*, *Gender*, dan *Cirrhosis*.

Berikut ini Tabel 1 menunjukkan statistika deskriptif pada setiap variabel, terlihat bahwa terdapat 2 jenis variabel yaitu variabel kontinu dan variabel biner. Secara keseluruhan, persentase data *missing* tertinggi sebesar 32,12% yaitu variabel *Packs of cigarettes per year*. Sedangkan, persentase data *missing* terendah sebesar 0% yaitu pada variabel *Age at diagnosis*, *Gender*, *Cirrhosis*, dan *Class*.

Tabel 1. Statistika Deskriptif Pada Setiap Variabel

Variabel	Range	Rata-rata atau Modus	Jumlah Data Yang Diamati	Jumlah Data Missing	Persentase Data Missing
Variabel Kontinu					
<i>Age at diagnosis</i>	20-93	64,69	165	0	0%
<i>Grams of alcohol per day</i>	0-500	71,01	117	48	29,09%
<i>Packs of cigarets per year</i>	0-510	20,46	112	53	32,12%
<i>Major dimension of nodule</i> (cm)	1,5-22	6,85	145	20	12,12%
Variabel Biner					
<i>Gender</i>	0/1	1	165	0	0%
<i>Hepatitis B Surface Antigen</i> (HBSA)	0/1	0	148	17	10,30%
<i>Hepatitis C Virus Antibody</i> (HCVA)	0/1	0	156	9	5,45%
<i>Cirrhosis</i>	0/1	1	165	0	0%
<i>Obesity</i>	0/1	0	155	10	6,06%
<i>Hemochromatosis</i>	0/1	0	142	23	13,94%
<i>Class</i>	0/1	1	165	0	0%

Penerapan Metode MICE

MICE adalah sebuah pendekatan yang bersifat fleksibel untuk menangani berbagai jenis data [3]. Langkah pertama yang dilakukan adalah menentukan model imputasi, model imputasi yang digunakan tergantung pada jenis variabel. Dalam hal ini, ketika variabel yang diimputasi merupakan variabel kontinu maka menggunakan model *Predictive Mean Matching* (PMM). Sedangkan, jika variabel yang diimputasi merupakan variabel biner maka menggunakan model regresi logistik.

Proses imputasi dimulai dari variabel yang memiliki data *missing* terendah. Pada saat melakukan imputasi, maka data *missing* yang ada pada variabel lain diisi dengan suatu nilai yang diambil secara acak. Langkah selanjutnya yaitu membentuk $m=3$ kali imputasi sehingga menghasilkan 3 gugus data lengkap hasil imputasi dengan iterasi yang digunakan sebanyak 20 untuk setiap gugus data lengkap hasil imputasi yang dibentuk. Misalkan dalam mengimputasi variabel biner, yaitu Variabel *Hepatitis C Virus Antibody* (HCVA) menggunakan model regresi logistik dengan menjadikan variabel tersebut menjadi variabel respon dan variabel lainnya menjadi variabel prediktor sehingga diperoleh model untuk setiap imputasi sebagai berikut.

$$\begin{aligned} \text{logit}[\hat{\pi}(HCVA_1)] &= 0,6244 - 0,6557Class + 0,3800Age\ at\ diagnosis + 0,1338Gender(1) + \\ &\quad 0,1044Cirrhosis(1) + 0,2019Obesity(1) - 0,4246HBSA(1) + \\ &\quad 0,2602Major\ dimension\ of\ nodule + 0,8653Hemochromatosis(1) + \\ &\quad 1,0769Grams\ of\ alcohol\ per\ day - 0,3289Packs\ of\ cigarets\ per\ year \\ \text{logit}[\hat{\pi}(HCVA_2)] &= 1,4301 - 0,3173Class + 0,6654Age\ at\ diagnosis + 0,3772Gender(1) + \\ &\quad 0,7260Cirrhosis(1) + 0,74831Obesity(1) + 0,1616HBSA(1) + \\ &\quad 0,1868Major\ dimension\ of\ nodule - 0,0366Hemochromatosis(1) + \\ &\quad 0,5850Grams\ of\ alcohol\ per\ day + 0,4866Packs\ of\ cigarets\ per\ year \\ \text{logit}[\hat{\pi}(HCVA_3)] &= 1,8031 - 0,4234Class + 0,4158Age\ at\ diagnosis + 0,2088Gender(1) - \\ &\quad 0,6099Cirrhosis(1) - 0,5101Obesity(1) - 0,2145HBSA(1) + \\ &\quad 0,6092Major\ dimension\ of\ nodule - 0,0724Hemochromatosis(1) + \\ &\quad 0,9087Grams\ of\ alcohol\ per\ day + 0,0464Packs\ of\ cigarets\ per\ year \end{aligned}$$

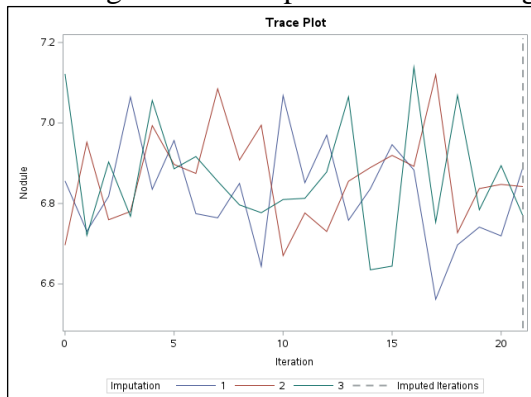
Begitu pun pada saat mengimputasi variabel kontinu, yaitu misal variabel *Major dimension of nodule* (cm) menggunakan model *Predictive Mean Matching* (PMM) dengan menjadikan variabel tersebut sebagai variabel respon sedangkan variabel lainnya menjadi variabel prediktor sehingga diperoleh model untuk setiap imputasi sebagai berikut.

$$\hat{Y}_{Major\ dimension\ of\ nodule\ (1)} = 1,1792 + 0,2099Class + 0,0002Age\ at\ diagnosis - 0,0934Gender(1) + 0,8428Cirrhosis(1) + 0,1358HCVA(1) - 0,1475Obesity(1) - 0,7005HBSA(1) + 0,1359Hemochromatosis(1) - 0,0331Grams\ of\ alcohol\ per\ day - 0,1461Packs\ of\ cigarets\ per\ year$$

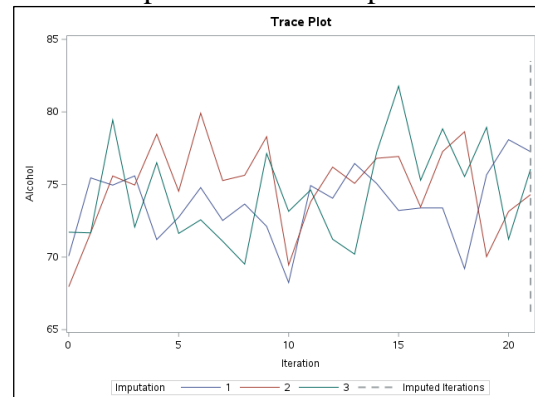
$$\hat{Y}_{Major\ dimension\ of\ nodule\ (2)} = 0.2002 + 0,3299Class + 0,1530Age\ at\ diagnosis - 0,1956Gender(1) + 0,4948Cirrhosis(1) + 0,1503HCVA(1) - 0,0976Obesity(1) - 0,0374HBSA(1) + 0,1490Hemochromatosis(1) - 0,1250Grams\ of\ alcohol\ per\ day - 0,0527Packs\ of\ cigarets\ per\ year$$

$$\hat{Y}_{Major\ dimension\ of\ nodule\ (3)} = 0,1596 + 0,1774Class + 0,1448Age\ at\ diagnosis - 0,1341Gender(1) + 0,3669Cirrhosis(1) + 0,1843HCVA(1) - 0,0752Obesity(1) - 0,1192HBSA(1) + 0,2678Hemochromatosis(1) - 0,2388Grams\ of\ alcohol\ per\ day - 0,3192Packs\ of\ cigarets\ per\ year$$

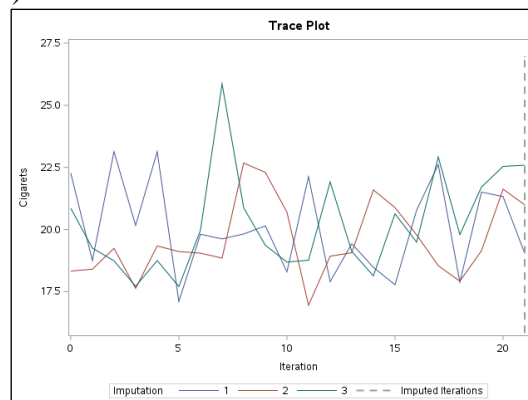
Kemudian diperoleh grafik nilai imputasi untuk setiap variabel kontinu yang bertujuan untuk melihat konvergensi hasil imputasi. Berikut ini grafik nilai imputasi untuk setiap variabel kontinu.



(a) Variabel *Major dimension of nodule* (cm)



(b) Variabel *Grams of Alcohol per day*



(c) Variabel *Packs of cigarets per year*

Gambar 2. Grafik Nilai Imputasi Pada Setiap Variabel Kontinu

Berdasarkan Gambar 2 di atas, dapat dilihat konvergensi hasil imputasi yang menunjukkan pergerakan nilai imputasi di setiap iterasi pada variabel *Major dimension of nodule* (cm), *Grams of Alcohol per day*, dan *Packs of cigarets per year*. Menurut Buuren [10], hasil imputasi dapat dikatakan konvergen apabila garis tidak membentuk suatu pola tren tertentu. Pada grafik setiap variabel tersebut terlihat bahwa grafik tidak membentuk pola tertentu sehingga hasil imputasi dengan 20 iterasi menunjukkan hasil yang konvergen.

Berikut ini diberikan hasil nilai imputasi data *missing* untuk setiap variabel. Di mana nilai imputasi akhir diperoleh dari rata-rata atau modus pada seluruh imputasi untuk setiap pengamatan pada setiap variabel yang mengalami data *missing*. Sehingga data kelangsungan hidup 165 pasien *hepatocellular carcinoma* (HCC) menjadi data yang lengkap.

Tabel 2. Data Hepatocellular Carcinoma (HCC) Yang Sudah Lengkap

Variabel	Pengamatan Ke-						
	1	2	3	4	5	...	165
Variabel Kontinu							
<i>Age at diagnosis</i>	67	62	78	77	76	...	52
<i>Grams of Alcohol per day</i>	137	0	50	40	100	...	50
<i>Packs of cigarets per year</i>	15	0	50	30	30	...	28
<i>Major dimension of nodule</i> (cm)	4	2	13	16	9	...	9
Variabel Biner							
<i>Gender</i>	1	0	1	1	1	...	1
<i>Hepatitis B Surface Antigen</i> (HBSA)	0	0	1	0	1	...	0
<i>Hepatitis C Virus Antibody</i> (HCVA)	0	1	0	0	0	...	1
<i>Cirrhosis</i>	1	1	1	1	1	...	1
<i>Obesity</i>	0	0	0	0	0	...	0
<i>Hemochromatosis</i>	1	0	0	0	0	...	0
<i>Class</i>	1	1	1	0	1	...	0

KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, maka dapat disimpulkan bahwa metode *Multivariate Imputation by Chained Equations* (MICE) dapat digunakan untuk mencari imputasi data *missing* pada kasus data *Hepatocellular Carcinoma* (HCC). Metode MICE dengan jumlah imputasi $m = 3$ dan iterasi sebanyak 20 kali menunjukkan hasil imputasi yang cukup baik dan konvergen untuk setiap variabel yang mengalami data *missing*. Dengan melakukan imputasi menggunakan *Predictive Mean Matching* (PMM) untuk variabel kontinu dan model regresi logistik untuk variabel biner diperoleh hasil imputasi akhir dengan menghitung rata-rata atau modus dari seluruh imputasi untuk setiap pengamatan pada variabel yang mengalami data *missing*, sehingga diperoleh data yang lengkap untuk setiap variabel pada kasus data *Hepatocellular Carcinoma* (HCC).

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih karena penelitian ini dapat terlaksana dengan baik berkat bantuan dari berbagai pihak, khususnya Allah SWT, keluarga, Bapak Abdul Kudus, M.Si., Ph.D. selaku dosen pembimbing, dan pihak lainnya yang tidak bisa disebutkan satu per satu.

DAFTAR PUSTAKA

- [1] D. B. Rubin, "Inference and Missing Data," *Source: Biometrika*, vol. 63, no. 3, pp. 581–92, 1976.
- [2] P. C. Austin, I. R. White, D. S. Lee, and S. van Buuren, "Missing Data in Clinical Research: A Tutorial on Multiple Imputation," *Canadian Journal of Cardiology*, vol. 37, no. 9, pp. 1322–1331, Sep. 2021, doi: 10.1016/j.cjca.2020.11.010.
- [3] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained

- equations: what is it and how does it work?,” *Int J Methods Psychiatr Res*, vol. 20, no. 1, pp. 40–49, 2011.
- [4] N. Hajarisman, “Analisis Data Kategorik,” Bandung: Program Studi Statistika, Universitas Islam Bandung, 2009.
- [5] S. Van Buuren and K. Groothuis-Oudshoorn, *Flexible multivariate imputation by MICE*. Leiden: TNO, 1999.
- [6] Kementerian Kesehatan Republik Indonesia, “Kementerian Kesehatan Republik Indonesia Nomor HK.01.07/MENKES/1355/2022 Tentang Pedoman Nasional Pelayanan Kedokteran Tata Laksana Karsinoma Sel Hati Pada Dewasa” Jakarta, 2022.
- [7] S. Van. Buuren and K. Groothuis-Oudshoorn, “Journal of Statistical Software mice: Multivariate Imputation by Chained Equations in R,” vol. 45, 2011, [Online]. Available: <http://www.jstatsoft.org/>
- [8] S. Van Buuren, “Multiple imputation of discrete and continuous data by fully conditional specification,” *Stat Methods Med Res*, vol. 16, no. 3, pp. 219–242, 2007.
- [9] I. R. White, P. Royston, and A. M. Wood, “Multiple imputation using chained equations: Issues and guidance for practice,” *Stat Med*, vol. 30, no. 4, pp. 377–399, Feb. 2011, doi: 10.1002/sim.4067.
- [10] S. Van Buuren, *Flexible Imputation of Missing Data*. New York: CRC Press, 2012.