IJSA (2025), Vol. 9, No. 1, 21-32
DOI: https://doi.org/10.29244/ijsa.v9i1p21-32

Indonesian Journal of Statistics and Its Applications
Available at https://journal-stats.ipb.ac.id/index.php/ijsa
e-ISSN: 2599-0802

Missing Value Estimation Using Fuzzy C-Means in Classification of Chronic Kidney Disease

Pendugaan *Missing Values* Menggunakan *Fuzzy C - Means* Pada Pengklasifikasian Penyakit Ginjal Kronik*

Raisa Nida Eria[1], Aam Alamudi[1‡], Itasia Dina Sulvianti[1],Pika Silvianti[1], Septian Rahardiantoro[1]

[1]Study Program on Statistics and Data Science, IPB University, Indonesia
[‡]corresponding author: aamalamudi@apps.ipb.ac.id

Copyright © 2025 Raisa Nida Eria, Aam Alamudi, Itasia Dina Sulvianti, Pika Silvianti, and Septian Rahardiantoro. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

According to the World Health Organization, chronic kidney disease ranked as the 10th leading cause of death globally in 2020. Chronic kidney disease often progresses without symptoms in its early stages, making early detection critical for effective prevention and treatment. This study aims to classify chronic kidney disease patients using historical medical data while addressing the issue of missing values. The objective of this research is to implement a classification model using the Classification and Regression Trees method and to overcome missing values in the dataset through the Fuzzy C-Means imputation method. The dataset used contains both numerical and categorical attributes with missing values. To handle this, the Fuzzy C-Means method is applied, resulting in a clustering error rate of 20.25% and a balanced accuracy of 84.80%. The classification using the Classification and Regression Trees method yields an accuracy of 97.50%, sensitivity of 100.00%, and specificity of 92.86%. Individuals are classified as having Chronic kidney disease if they meet one of the following criteria: (1) hemoglobin ≥ 13, specific gravity 1.020 or 1.025, serum creatinine < 1.3, albumin levels between 1 and 5, and sugar levels of 0, 2, 3, 4, or 5; (2) hemoglobin ≥ 13, specific gravity 1.020 or 1.025, and serum creatinine ≥ 1.3; (3) hemoglobin ≥ 13 and specific gravity of 1.005, 1.010, or 1.015; or (4) hemoglobin < 13 and red blood cell count < 5.5.

Keywords: CART, CKD, FCM, missing values.

[1] Received: Jan 2024; Reviewed: Mar 2024; Published: Jun 2025

1. Pendahuluan

Penyakit Ginjal Kronik (PGK) merupakan salah satu penyakit tidak menular (PTM) yang berdampak signifikan terhadap kualitas hidup dan angka harapan hidup masyarakat. Menurut World Health Organization (WHO), PGK menjadi penyebab kematian global yang terus meningkat dari peringkat ke-13 pada tahun 2010 menjadi peringkat ke-10 pada tahun 2020. Penderita PGK umumnya tidak menunjukkan tanda dan gejala pada tahap awal, sehingga seringkali diagnosis baru dilakukan ketika penyakit sudah berada pada tahap lanjut atau bahkan mengalami gagal ginjal. Oleh karena itu, deteksi dini menjadi sangat penting untuk mencegah komplikasi yang lebih serius (Kemenkes 2017).

Kemajuan teknologi informasi dan ketersediaan data historis kesehatan membuka peluang baru dalam mendeteksi penyakit secara dini. Salah satu sumber data terbuka yang dapat dimanfaatkan adalah *UCI Machine Learning Repository*, yang menyediakan data PGK dengan atribut campuran, yaitu peubah numerik dan kategorik. Data ini memungkinkan penerapan metode pembelajaran mesin, khususnya metode klasifikasi, untuk mengenali pola-pola tertentu pada penderita PGK.

Metode klasifikasi yang digunakan dalam penelitian ini adalah Classification and Regression Trees (CART). Metode ini memiliki keunggulan dalam menghasilkan model yang mudah dipahami dan diinterpretasikan, serta memiliki tingkat kesalahan klasifikasi yang rendah (Sartono & Syafitri, 2010). Namun, permasalahan yang sering dihadapi dalam pengolahan data medis adalah keberadaan *missing values* atau data hilang. Apabila tidak ditangani dengan tepat, *missing values* dapat menurunkan akurasi model dan menghasilkan simpulan yang bias.

Beberapa pendekatan telah dikembangkan untuk mengatasi *missing values*, antara lain metode penghapusan (deletion), metode imputasi, dan metode khusus. Di antara ketiganya, metode imputasi menjadi pilihan yang paling sering digunakan karena mampu menjaga struktur dan informasi data. Salah satu metode imputasi yang terbukti efektif dalam penelitian sebelumnya adalah Fuzzy C-Means (FCM). Aristiawati *et al.* (2019) menunjukkan bahwa metode FCM memberikan hasil imputasi yang lebih baik dibandingkan metode imputasi sederhana seperti rata-rata, khususnya dalam data penyakit Paru Obstruktif Kronik (PPOK).

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada penerapan metode CART dalam klasifikasi penderita PGK dengan mengatasi *missing values* menggunakan metode imputasi Fuzzy C-Means (FCM). Diharapkan pendekatan ini dapat menghasilkan model klasifikasi yang akurat serta mempertahankan kualitas data historis yang digunakan.

Penelitian ini bertujuan untuk menerapkan metode CART untuk mengklasifikasikan penderita PGK berdasarkan data historis, mengatasi *missing values* dalam data PGK menggunakan metode imputasi Fuzzy C-Means (FCM), dan menganalisis pengaruh metode imputasi FCM terhadap akurasi model klasifikasi.

Deteksi dini PGK sering tidak terdiagnosis pada tahap awal. Dengan meningkatnya kasus PGK secara global, dibutuhkan pendekatan berbasis data yang

mampu memanfaatkan informasi historis secara optimal. Penanganan *missing values* secara tepat melalui metode imputasi yang andal seperti FCM juga menjadi kebutuhan mendesak untuk menjaga integritas data dan hasil analisis. Penelitian ini diharapkan dapat berkontribusi terhadap pengembangan sistem deteksi dini PGK berbasis data mining dan pembelajaran mesin.

Penelitian terdahulu telah banyak menerapkan metode klasifikasi dalam bidang kesehatan. Sartono dan Syafitri (2010) menyatakan bahwa metode CART efektif dalam membangun model klasifikasi yang interpretatif dengan kesalahan klasifikasi yang rendah. Namun, tantangan besar yang masih dihadapi adalah keberadaan *missing values* dalam data. Hair *et al*. (2010) menyarankan penghapusan data hanya jika proporsi *missing values* kurang dari 10%, sedangkan dalam praktiknya, proporsi yang lebih besar seringkali dijumpai. Oleh karena itu, metode imputasi menjadi solusi yang umum digunakan. Aristiawati *et al.* (2019) membuktikan bahwa metode Fuzzy C-Means (FCM) memiliki performa lebih baik dibandingkan metode imputasi nilai rata-rata dalam data medis. Oleh karena itu, kombinasi antara CART dan FCM dalam penelitian ini diharapkan mampu meningkatkan akurasi klasifikasi penderita PGK secara signifikan.

2. Metodologi

2.1 Data

Penelitian ini menggunakan data sekunder yang didapatkan dari University of California Irvine (UCI) *repository of machine learning database*. Data terdiri atas 400 amatan dengan 24 peubah bebas dan satu peubah tak bebas.Peubah bebas yang digunakan disajikan pada Tabel 1 (Dewi, 2010; Rianto & Iswari, 2017).

Tabel 1. Daftar Peubah Bebas

No	Peubah	Keterangan
1	Age	Umur pasien.
2	Bp	*Blood pressure* yaitu tekanan darah pasien dengan satuan yang dipakai adalah mm/Hg.
3	Sg	*Spesific gravity* yaitu berat jenis untuk menghitung kepekatan urine pasien. Atribut ini memiliki 5 nilai yaitu 0 = 1,005; 1 = 1,010; 2 = 1,015; 3 = 1,020; 4 = 1,025.
4	Al	*Albumin* yaitu kadar albumin atau protein pada urine pasien. Atribut ini memiliki 6 nilai yaitu 0,1,2,3,4,5.
5	Su	*Sugar* yaitu kadar gula darah pada pasien. Atribut ini memiliki 6 nilai yaitu 0,1,2,3,4,5.
6	Rbc	*Red blood cells* yaitu kadar dari banyaknya sel darah merah pada pasien, 0 jika abnormal dan 1 jika normal.
7	Pc	*Pus cell* yaitu kadar dari banyaknya sel nanah pada pasien, 0 jika *abnormal* dan 1 jika *normal*.
8	Pcc	*Pus cell clumps* yaitu keadaan yang menunjukkan adanya gumpalan sel nanah dalam darah pasien, 0 jika *notpresent* dan 1 jika *present*.
9	Ba	*Bacteria* yaitu keadaan yang menunjukkan adanya bakteri dalam darah pasien, 0 jika *notpresent* dan 1 jika *present*.

No	Peubah	Keterangan
10	Bgr	*Blood glucose random* yaitu banyaknya kadar glukosa pada darah dengan satuan yang dipakai adalah mgs/dL.
11	Bu	*Blood urea* yaitu kadar ureum dalam darah pasien dengan satuan yang digunakan yaitu mgs/dL.
12	Sc	*Serum creatinine* yaitu kadar kreatininin dalam darah pasien dengan satuan yang digunakan yaitu mgs/dL.
13	Sod	*Sodium* yaitu kadar natrium dalam darah pasien dengan satuan yang digunakan yaitu mEq/L.
14	Pot	*Potassium* yaitu kadar kalium pada pasien dengan satuan yang digunakan yaitu mEq/L.
15	Hemo	*Hemoglobin* yaitu banyaknya hemoglobin pada pasien dengan satuan yang dipakai adalah gms.
16	Pcv	*Packed cell volume* yaitu kadar sel darah merah dalam darah pasien dengan satuan yang dipakai adalah mL in volume
17	Wc	*White blood cell count* yaitu banyaknya sel darah putih pasien dengan satuan yang dipakai adalah cells/cumm.
18	Rc	*Red blood cell count* yaitu banyaknya sel darah merah pasien dengan satuan yang dipakai adalah millions/cmm.
19	Htn	*Hypertension* yaitu kondisi yang menunjukkan pasien mengalami peningkatan tekanan darah, 0 jika *no* dan 1 jika *yes*.
20	Dm	*Diabetes mellitus* yaitu kondisi yang menunjukkan pasien terkena diabetes melitus, 0 jika *no* dan 1 jika *yes*.
21	Cad	*Coronary artery disease* yaitu kondisi yang menunjukkan pasien memiliki penyakit arteri koroner, 0 jika *no* dan 1 jika *yes*.
22	Appet	*Appetite* yaitu kondisi mengenai nafsu makan pasien, 0 jika *poor* dan 1 jika *good*.
23	Pe	*Pedal edema* yaitu keadaan mengenai pasien memiliki pedal edema, 0 jika *no* dan 1 jika *yes*.
24	Ane	*Anemia* yaitu keadaan yang menunjukkan pasien mengalami penurunan sel darah merah atau hemoglobin, atau keduanya. Atribut ini memiliki 2 nilai yaitu 0 untuk *no* dan 1 jika *yes*.

Peubah tak bebas dari gugus data adalah *status penderita PGK* yang diberi kode class yaitu 0 jika *notckd* atau bukan penderita PGK dan 1 jika *ckd* atau penderita PGK.

2.2 **Metode Penelitian**

Analisis yang digunakan pada penelitian ini dengan bantuan Microsoft Excel 2010 dan R versi 4.1.1. *package* yang digunakan yaitu *fclust, caret, rpart,* dan *rpart plot*. Tahapan analisis data yang digunakan dalam penelitian ini adalah sebagai berikut:

1) Melakukan pra-proses data meliputi:
   a. Melakukan pelabelan ulang terhadap label yang telah tersedia pada data PGK.
   b. Melakukan transformasi data, pada peubah numerik akan dilakukan normalisasi data dan pada peubah kategorik akan dikonversi menjadi peubah numerik.

c. Membagi data menjadi dua bagian, yaitu *complete data* dan *missing data*.
d. Menggerombolkan *complete* data dengan menerapkan metode FCM.
e. Melakukan metode imputasi pada *missing data*.
f. Menggabungkan *complete data* dan *missing data* yang diimputasi menjadi satu.
g. Melakukan penggerombolan pada data yang telah digabung.
h. Mengevaluasi kinerja penggerombolan pada data secara keseluruhan.
i. Melakukan transformasi balik data, pada peubah numerik akan dilakukan denormalisasi data dan pada peubah kategorik akan dikonversi balik sesuai dengan kategori pada tahap 1.a.

2) Melakukan pengujian terhadap hubungan antara peubah penjelas dan peubah respon. Pada peubah penjelas berskala numerik akan diuji menggunakan uji *t-test* dan pada peubah penjelas berskala kategorik akan diuji menggunakan *chi-square test*.
3) Membentuk model klasifikasi.
   a. Membagi data menjadi data latih dan data uji sebesar 70 : 30 dengan menggunakan *set.seed*(012).
   b. Membentuk model klasifikasi menggunakan data latih dengan menggunakan parameter biaya kompleksitas sebesar 0,001 mengikuti *default* pada *package rpart.* Jumlah minimum amatan pada simpul (*minsplit*) sebanyak 2 amatan dan kedalaman pohon maksimal (*maxdepth*) sebesar 5.
   c. Melakukan pemangkasan pada jumlah simpul dengan nilai *Complexity Parameter* (CP) yang memiliki biaya resubtitusi relatif terkecil dari *Cross Validation* (CV) untuk mendapatkan pohon klasifikasi yang optimal.
   d. Membuat tabel ketepatan klasifikasi dari hasil pohon yang dihasilkan menggunakan data uji. Hasil dari tabel ketepatan klasifikasi akan dilihat nilai akurasi, sensitivitas, dan spesifisitas. Pada penelitian ini sensitivitas menggambarkan akurasi amatan dari penderita PGK dan spesifisitas menggambarkan akurasi amatan dari bukan pendeita PGK.
4) Menginterpretasikan hasil dari pohon klasifikasi yang terbentuk untuk melihat karakteristik dari penderita PGK.

3. Hasil dan Pembahasan

3.1 Gambaran Umum Data

Data penelitian terdiri dari 24 peubah penjelas dan satu peubah respon dengan total observasi sebanyak 400. Banyaknya penderita yang terkena PGK sebesar 62,5% dan sebesar 37,5% bukan penderita PGK. Data yang digunakan mengandung *missing values*, dengan 1.012 buah pengamatan hilang.

3.2 Pra-Proses Data

Tahap awal yang dilakukan yaitu membagi data menjadi dua bagian yaitu *complete data* dan *missing data*. *Missing data* akan diimputasi menggunakan metode penggerombolan FCM. Data yang *complete* akan diterapkan metode FCM terlebih dahulu. Statistik deskriptif peubah yang digunakan dapat dilihat pada Tabel 2.

Tabel 2 menunjukkan bahwa keragaman pada peubah *white blood cell count* lebih besar dibandingkan dengan peubah *red blood cell count*. Hal tersebut diatasi dengan melakukan normalisasi data. Normalisasi mengubah skala data menjadi nilai dengan rentang 0 sampai 1. Menurut Salleh dan Samat (2017) bahwa peubah berskala kategorik akan dilakukan konversi menjadi peubah numerik dengan rentang nilai yang

sama dengan peubah numerik.

Tabel 2. Statistik deskriptif peubah yang digunakan

Kode	Minimum	Rataan	Maksimum	Simpangan baku
age	2,0	51,48	90,0	17,17
bp	50,0	76,47	180,0	13,68
bgr	22,0	148,00	490,0	79,28
bu	1,5	57,43	391,0	50,50
sc	0,4	3,07	76,0	5,74
sod	4,5	137,50	163,0	10,41
pot	2,5	4,63	47,0	3,19
hemo	3,1	12,53	17,8	2,91
pcv	9,0	38,88	54,0	8,99
wc	2.200,0	8.406,00	26.400,0	2.944,47
rc	2,1	4,70	8,0	1,03

Tahap awal dalam melakukan penggerombolan data PGK menggunakan *complete data*. Penggerombolan dilakukan dengan menetapkan *set.seed*(012) sebagai inisiasi derajat keanggotaan untuk matriks partisi awal. Penetapan *set.seed* tersebut dilakukan untuk menghasilkan nilai inisiasiasi derajat keanggotaan yang sama ketika melakukan analisis data. Selain itu, banyaknya gerombol yang digunakan sebanyak dua karena data PGK memiliki dua kelas yaitu *ckd* dan *notckd*. Keputusan suatu observasi masuk ke dalam gerombol berdasarkan nilai derajat keanggotaannya. Apabila nilai derajat keanggotaan pada satu gerombol lebih besar dibandingkan gerombol lainnya, maka observasi akan masuk ke dalam gerombol yang memiliki nilai derajat keanggotaan yang lebih besar.

Hasil penggerombolan pada *complete data* yaitu gerombol pertama memiliki 115 anggota dan gerombol kedua memiliki 39 anggota. Penggerombolan pada *complete data* akan dievaluasi dengan menghitung banyaknya ketidaktepatan dari hasil penggerombolan yang dilakukan. Banyaknya ketidaktepatan dalam penggerombolan dapat dilakukan dengan cara membandingkan hasil penggerombolan dengan dua jenis kelas data PGK. Hasil ketidaktepatan penggerombolan pada *complete data* yaitu sebanyak 4 anggota.

Tahap selanjutnya adalah menduga *missing values* dengan memanfaatkan informasi yang terdapat dalam penggerombolan *complete data* yaitu nilai pusat gerombol. *Missing values* yang telah diimputasi digabungkan kembali dengan *complete* data dan akan dilakukan penggerombolan. Hasil ketidaktepatan penggerombolan sebanyak 81 anggota. Berdasarkan hal tersebut, bahwa penggerombolan yang dilakukan menghasilkan persentase ketidaktepatan sebesar 20,25%. Menurut Ibarra *et al.* (2021) bahwa evaluasi kinerja metode penggerombolan lainnya yaitu dengan mengukur akurasi berimbang (*balanced accuracy*). Adapun hasil dari penggerombolan pada data secara keseluruhan ditunjukkan pada Tabel 3.

Tabel 3 menunjukkan bahwa sensitivitas yang dihasilkan sebesar 67,60% dan spesifisitas sebesar 100%. Nilai akurasi berimbang merupakan nilai rata-rata dari sensitivitas dan spesifisitas. Nilai akurasi berimbang yang diperoleh sebesar 83,80%. Nilai akurasi berimbang semakin mendekati nilai satu, maka kinerja metode

penggerombolan semakin baik atau mampu menggerombolkan dengan baik.

Tabel 3: Ketepatan hasil penggerombolan.

Gerombol aktual	Gerombol data secara keseluruhan		Total
	0	1	
0	150	81	231
1	0	169	169
Total	150	250	400

Tahap selanjutnya adalah melakukan transfomasi balik, pada peubah numerik dilakukan proses denormalisasi dan melakukan konversi balik pada peubah kategorik sesuai dengan kategori awal. Misalkan pada peubah sugar (su) memiliki nilai 0, maka akan dikonversi menjadi 0. Peubah sugar akan dikonversi menjadi 1 jika $0 < su \leq 0{,}2$. Peubah sugar akan dikonversi menjadi 2 jika $0{,}2 < su \leq 0{,}4$. Peubah sugar akan dikonversi menjadi 3 jika $0{,}4 < su \leq 0{,}6$. Peubah sugar akan dikonversi menjadi 4 jika $0{,}6 < su \leq 0{,}8$. Peubah sugar akan dikonversi menjadi 5 jika $0{,}8 < su \leq 1$.

3.3 Hubungan Peubah Penjelas dan Peubah Respon

Pada tahapan ini akan dilakukan pengujian terhadap peubah penjelas menggunakan uji kebebasan *chi-square* yang diduga memiliki hubungan dengan peubah respon yaitu *classification*. Pengaruh peubah penjelas yang berskala kategorik yaitu *spesific gravity* (sg), *albumin* (al), *sugar* (su), *red blood cells* (rbc), *pus cell* (pc), *pus cell clumps* (pcc), *bacteria* (ba), *hypertension* (htn), *diabetes mellitus* (dm), *coronary artery disease* (cad), *appetite* (appet), *pedal edema* (pe), dan *anemia* (ane). Hipotesis untuk masing-masing pasangan peubah respon dan peubah penjelas adalah sebagai berikut:

H_0 : Tidak ada hubungan antara peubah *status PGK* dengan peubah *bebas*.

H_1 : Ada hubungan antara peubah *status PGK* dengan peubah *bebas*.

Peubah bebas yang digunakan pada masing-masing uji disajikan pada Tabel 4.

Tabel 4. Hasil uji *chi-square* pada uji kebebasan dari peubah penjelas terhadap peubah respon

Hipotesis Nol		Peubah	χ^2	*p-value*
status PGK dengan	spesific gravity	sg	74,717	0,000
status PGK dengan	albumin	al	212,880	0,000
status PGK dengan	spesific sugar	su	71,054	0,000
status PGK dengan	red blood cells	rbc	30,167	0,000
status PGK dengan	pus cell	pc	78,944	0,000
status PGK dengan	pus cell clumps	pcc	26,397	0,000
status PGK dengan	bacteria	ba	11,663	0,000
status PGK dengan	hypertension	htn	136,930	0,000
status PGK dengan	diabetes mellitus	dm	122,600	0,000
status PGK dengan	coronary artery disease	cad	20,581	0,000
status PGK dengan	appetite	appet	60,836	0,000
status PGK dengan	pedal edema	pe	54,338	0,000
status PGK dengan	anemia	ane	40,492	0,000

Tabel 4 memberikan kesimpulan yang sama untuk seluruh peubah penjelas yaitu menolak H_0 pada taraf nyata 5%. Hal ini berarti peubah *classification* memiliki hubungan dengan peubah *spesific gravity*, *albumin*, *sugar*, *red blood cells*, *pus cell*, *pus cell clumps*, *bacteria*, *hypertension*, *diabetes mellitus*, *coronary artery disease*, *appetite*, *pedal edema*, dan *anemia*.

Selanjutnya, untuk melihat beda antara rataan peubah penjelas berskala numerik yaitu *age*, *blood pressure* (bp), *blood glucose random* (bgr), *blood urea* (bu), *serum creatinine* (sc), *sodium* (sod), *potassium* (pot), *hemoglobin* (hemo), *packed cell volume* (pcv), *white blood cell count* (wc), *red blood cell count* (rc) dengan peubah respon digunakan uji beda rata-rata. Hipotesis uji beda dua rata-rata antara peubah penjelas dan *classification* adalah sebagai berikut:

$$H_0 : \mu_1 - \mu_2 = 0$$

Artinya, rata-rata untuk peubah penjelas dari penderita PGK dan bukan penderita PGK adalah sama.

$$H_1 : \mu_1 - \mu_2 \neq 0$$

artinya, rata-rata untuk peubah penjelas dari penderita PGK dan bukan penderita PGK adalah tidak sama.

Tabel 5. Hasil uji-t pada uji beda dua rata-rata dari peubah penjelas terhadap peubah respon.

Peubah	t-hitung	*p-value*
age	60,292	0,000
bp	113,530	0,000
bgr	39,379	0,000
bu	23,060	0,000
sc	8,794	0,000
sod	288,260	0,000
pot	28,836	0,000
hemo	76,272	0,000
pcv	87,990	0,000
wc	67,032	0,000
rc	64,364	0,000

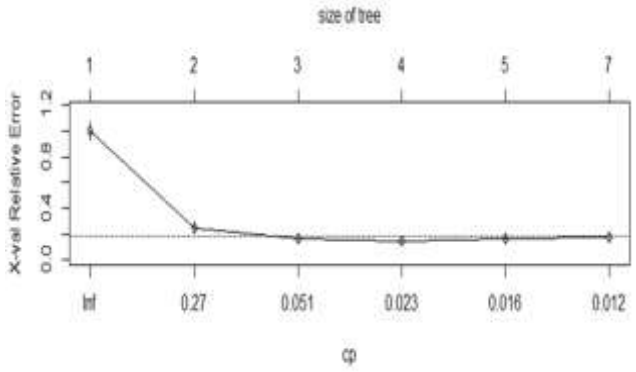
Tabel 5 menunjukkan bahwa hasil uji beda dua rata-rata untuk semua peubah penjelas memberikan kesimpulan yang sama yaitu tolak h_0 pada taraf nyata 5%. Hal ini menunjukkan bahwa rata-rata peubah *age*, *blood pressure*, *blood glucose random*, *blood urea*, *serum creatinine*, *sodium*, *potassium*, *hemoglobin*, *packed cell volume*, *white blood cell count*, *red blood cell count* dari penderita PGK dan bukan penderita PGK adalah tidak sama.

3.4 Analisis Model Klasifikasi

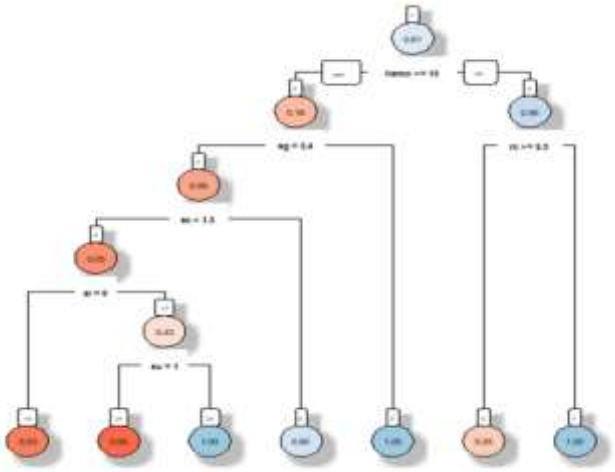
Pohon klasifikasi (CART) yang terbentuk menggunakan *set.seed*(012), sehingga akan menghasilkan pohon klasifikasi yang sama disetiap percoban. Pembentukan pohon klasifikasi diawali dengan menentukan pemilah terbaik sebagai pemilah utama. Hasil pohon klasifikasi yang terbentuk sebanyak 6 simpul pemilah dan 7 simpul terminal. Pohon klasifikasi yang dihasilkan dapat dilihat pada Gambar 2.

Tahap selanjutnya adalah melakukan pemangkasan pohon untuk mendapatkan

nilai pohon yang sederhana atau pohon optimal dengan menggunakan nilai CP dan CV. Pohon optimal dihasilkan dengan memangkas pohon keputusan pada simpul dengan nilai CP yang menghasilkan nilai kesalahan relatif (*xerror*) yang paling kecil. Nilai kesalahan relatif yang paling kecil sebesar 0,1 dengan nilai CP 0,012. Kurva CP pada metode CART disajikan pada Gambar 1. Hal tersebut menunjukkan bahwa pohon yang terbentuk diawal tidak perlu dipangkas karena pohon yang terbentuk dapat dikatakan sudah optimal.



Gambar 1. Plot CP pada model CART



Gambar 2. Pohon klasifikasi CART

Pohon klasifikasi yang terbentuk terdiri atas 7 simpul terminal, dengan melibatkan peubah *hemoglobin* (hemo), *spesific gravity* (sg), *serum creatinine* (sc), *albumin* (al), *sugar* (su), dan *red blood cell count* (rc). Simpul terminal yang terbentuk menghasillkan 3 simpul dengan label penderita PGK dan 4 simpul dengan label bukan penderita PGK. Pada penelitian ini akan mengunakan *default threshold* pada pohon klasifikasi biner sebesar 0,5 (Chen et al., 2006).

Tahap selanjutnya adalah menguji ketepatan hasil klasifikasi dalam mengklasifikasikan penderita PGK. Pengujian ketepatan hasil klasifikasi dilakukan dengan menggunakan data uji pada pohon klasifikasi optimal. Ketepatan hasil pohon klasifikasi ditunjukkan pada Tabel 6.

Tabel 6 menunjukkan banyaknya pengamatan yang salah klasifikasi sebanyak 3 yaitu pengamatan diklasifikasikan sebagai *ckd* yang seharusnya *notckd*. Nilai akurasi menunjukkan ketepatan hasil klasifikasi secara keseluruhan yaitu sebesar 97,50%.
Nilai sensitivitas menunjukkan ketepatan hasil klasifikasi pada kelas positif yaitu penderita PGK sebesar 100,00% dan spesifisitas menunjukkan ketepatan hasil klasifikasi pada kelas negatif yaitu bukan penderita PGK sebesar 92,86%.

Tabel 6. Ketepatan hasil pohon klasifikasi

Kelas aktual	Gerombol data secara keseluruhan		Ketepatan (%)
	Notckd (0)	*Ckd* (1)	
Notckd (0)	39	3	92,86
Ckd (1)	0	78	100,00
Akurasi			97,50

3.5 Interpretasi Pohon Klasifikasi

Karakteristik individu penderita PGK atau bukan penderita PGK berdasarkan pohon klasifikasi yang terbentuk dapat dilihat pada Tabel 7.

Tabel 7 Karakteristik penderita PGK

Simpul terminal	Kategori	N	Karakteristik
16	Tidak	103	*Hemoglobin* lebih dari atau sama dengan 13; *spesific gravity* 1,020 atau 1,025; *serum creatinine* kurang dari 1,3; dan *albumin* 0.

Simpul terminal	Kategori	N	Karakteristik
34	Tidak	4	*Hemoglobin* lebih dari atau sama dengan 13; *spesific gravity* sebesar 1,020 atau 1,025; *serum creatinine* kurang dari 1,3; *albumin* sebesar 1 atau 2 atau 3 atau 4 atau 5; dan *sugar* yaitu 1.
35	Ya	3	*Hemoglobin* lebih dari atau sama dengan 13; *spesific gravity* sebesar 1,020 atau 1,025; *serum creatinine* kurang dari 1,3; *albumin* sebesar 1 atau 2 atau 3 atau 4 atau 5; dan *sugar* yaitu 0 atau 2 atau 3 atau 4 atau 5.
9	Ya	5	*Hemoglobin* lebih dari atau sama dengan 13; *spesific gravity* sebesar 1,020 atau 1,025; dan *serum creatinine* lebih dari atau sama dengan 1,3.
5	Ya	10	*Hemoglobin* lebih dari atau sama dengan 13 dan *spesific gravity* sebesar 1,005 atau 1,010 atau 1,015.
6	Tidak	4	*Hemoglobin* kurang dari 13 dan *red blood cell count* kurang dari 5,5.
7	Ya	151	*Hemoglobin* kurang dari 13 dan *red blood cell count* lebih dari atau sama dengan 5,5.

4. Simpulan dan Saran

Berdasarkan hasil dan pembahasan yang diuraikan pada bab sebelumnya, maka dapat disimpulkan bahwa Imputasi dengan menggunakan metode FCM memiliki kemampuan menggerombolkan data PGK dengan tingkat kesalahan penggerombolan sebesar 20,25% dan nilai akurasi berimbang sebesar 83,80%. Seseorang menderita PGK yaitu jika memiliki (1) *hemoglobin* lebih dari atau sama dengan 13; *spesific gravity* sebesar 1,020 atau 1,025; *serum creatinine* kurang dari 1,3; *albumin* sebesar 1 atau 2 atau 3 atau 4 atau 5; dan nilai *sugar* yaitu 0 atau 2 atau 3 atau 4 atau 5, (2) *hemoglobin* lebih dari atau sama dengan 13; *spesific gravity* sebesar 1,020 atau 1,025; dan *serum creatinine* lebih dari atau sama dengan 1,3, (3) *hemoglobin* lebih dari atau sama dengan 13 dan *spesific gravity* sebesar 1,005 atau 1,010 atau 1,015, (4) *hemoglobin* kurang dari 13 dan *red blood cell count* kurang dari 5,5.

Daftar Pustaka

Aristiawati K, Siswantining T, Sarwinda D, Soemartojo SM. 2019. Missing values imputation based on fuzzy c-means algorithm for classification of Chronic Obstructive Pulmonary Disease (COPD). *AIP Conf Proc*. 2192(1):0600031-060007.

Chen JJ, Tsai CA, Moon H, Ahn H, Young JJ, Chen CH. 2006. Decision threshold adjustment in class prediction. SAR and QSAR in Environmental Research. 17(3):337–352.

Dewi IGAPA. 2010. Hubungan antara *quick of blood* (Qb) dengan adekuasi hemodialisis pada pasien yang mejalani terapi hemodialisis di ruang HD BRSU daerah Tabanan Bali [tesis]. Depok: Universitas Indonesiaa. 2nd ed. Florida (US): CRC Pr.

Dua D and Graff C. 2019. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, California: University of California, School of Information and Computer Science.

Hair JF, Black WC, Babin BJ, Anderson RE . 2010. Multivariate data analysis. 7th Edition. New York: Pearson

Ibarra MMA, Villuendas-Rey Y, Lytras MD, Yáñez-Márquez C, Salgado-Ramírez JC. 2021. Classification of diseases using machine learning algorithms: a comparative study. Mathematics. 9(15). doi:10.3390/math9151817.

[KEMENKES RI] Kementerian Kesehatan Republik Indonesia. 2017. Situasi penyakit ginjal kronis. Jakarta: Pusat Data dan Informasi. hal. 1-12.

Rianto R. 2016. Rancangan bangun aplikasi pendeteksi penyakit ginjal kronis dengan menggunakan algoritma C4.5 [skripsi]. Tangerang: Universitas Multimedia Nusantara.

Salleh MNM, Samat NA. 2017. FCMPSO: an imputation for missing data features in heart disease classification. IOP Conference Series: Materials Science and Engineering. 226.

Sartono B, Syafitri UD. 2010. Metode pohon gabungan: solusi pilihan untuk mengatasi

kelemahan pohon regresi dan klasifikasi tunggal. Forum Statistika dan Komputasi. 15(1):1-7.

Shofa UA. 2022. Perbandingan CART dan SMOTE CART dalam mengklasifikasikan kebutuhan KB tidak terpenuhi di Indonesia [skripsi]. Bogor: Institut Pertanian Bogor.

[WHO] World Health Organization. 2020. The top 10 causes of death [Online]. Tersedia: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (Diakses: 14 Juni 2021).