

Comparing Classification and Regression Tree and Logistic Regression Algorithms Using $5 \times 2cv$ Combined F-Test on Diabetes Mellitus Dataset

Fashihullisan, Dodi Vionanda*, Yenni Kurniawati, dan Fadhilah Fitri

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: dodi_vionanda@fmipa.unp.ac.id

Submitted : 13 Juli 2023

Revised : 08 Agustus 2023

Accepted : 16 Agustus 2023

ABSTRACT

Classification is the process of finding a model that describes and distinguishes data classes that aim to be used to predict the class of objects whose class labels are unknown. There are few algorithms in classification, such as classification trees and regression trees (CART) and logistic regression. The k-fold cross validation method has a weakness for algorithm comparison problems, it is possible at different folds to produce different error predictions, so that the results of comparing algorithm performance will also be different. And not yet known whether CART and logistic regression have same or different error rates. Until the problem to comparison of algorithms, the researcher will apply the $5 \times 2cv$ t test method and the $5 \times 2cv$ combined F test. Out of 100 iterations the 10-fold cross validation method was only consistent three times which shows that the k-fold cross validation method has poor consistency in comparing the CART algorithm and logistic regression for diabetes mellitus data. $5 \times 2cv$ combined F test and $5 \times 2cv$ t test methods that have been carried out show that $5 \times 2cv$ combined F test is better used to get conclusions from the results of a comparison of the two algorithms because just produces one decision, in contrast to $5 \times 2cv$ t test which has the possibility to get different decisions from 10 test statistics which results makes it difficult for researchers to draw conclusions in comparing the cart algorithm and logistic regression.

Keywords: Classification, K-Fold Cross Validation, $5 \times 2cv$ t Test, $5 \times 2cv$ Combined F Test



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Klasifikasi adalah proses penemuan model yang menggambarkan dan membedakan kelas data yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Kamber, 2006). Ada beberapa algoritma yang populer dalam klasifikasi, seperti *classification trees and regression trees* (CART) dan regresi logistik (Gholamian, 2013). Kinerja model yang dihasilkan dari algoritma CART dan regresi logistik dapat diukur dengan cara menghitung prediksi galat masing-masing algoritma tersebut. Prediksi galat terhadap model adalah untuk mengukur kinerja model dengan menghitung segala bentuk tingkat kesalahan prediksi pada model. Menurut Leo Breiman, metode yang paling populer digunakan untuk prediksi galat adalah *cross validation* (Breiman, 2001). *Cross validation* adalah teknik pengujian keefektifan dari model yang dibentuk dengan melakukan penyusunan ulang (*resampling*) pada data untuk membaginya menjadi dua bagian yaitu data *training* dan data *testing* (James, dkk, 2013: 176). Dalam pengujian menggunakan metode *cross validation* ini dikenal pula istilah stratifikasi. Stratifikasi adalah proses menata ulang data pengamatan untuk memastikan setiap objek pengamatan mewakili keseluruhan data baik pada data *training* maupun pada data *testing*. Stratifikasi ini dapat diterapkan pada metode pengujian *k-fold cross validation* (Refaeilzadeh, et al., 2016). Metode *k-fold* merupakan metode dalam *cross validation* dengan cara membagi data kedalam k grup dengan sampel (pengamatan) yang sama. Pada data *training* akan digunakan sebanyak k-1 sampel dan sisanya 1 grup sampel akan menjadi data *testing*.

Namun metode *k-fold cross validation* memiliki kelemahan untuk masalah perbandingan algoritma, yaitu pada *fold* yang berbeda memungkinkan untuk menghasilkan prediksi galat yang berbeda pula, sehingga akan membuat hasil perbandingan kinerja algoritma berbeda pula (Dietterich, 1998). Sehingga dalam masalah perbandingan algoritma peneliti akan menerapkan metode uji t $5 \times 2cv$ dan uji F *combined* $5 \times 2cv$, selain itu kedua metode tersebut adalah metode khusus yang dikembangkan untuk digunakan dalam permasalahan perbandingan algoritma yang ada pada klasifikasi.

Metode uji t $5 \times 2cv$ merupakan pengembangan dari *k-fold cross validation*. Metode uji t $5 \times 2cv$ memiliki tujuan yaitu untuk membandingkan dua algoritma, apakah dua algoritma memiliki *error* yang sama atau berbeda (Dietterich, 1998). Metode uji *F combined* $5 \times 2cv$ sebenarnya tidak jauh berbeda dengan uji t $5 \times 2cv$ yaitu sama-sama untuk melihat apakah dua algoritma memiliki *error* yang sama atau berbeda, hanya saja pada metode ini telah menggabungkan 10 statistik uji t kedalam statistik uji *F* untuk mendapatkan kesimpulan dari perbandingan dua algoritma (Alpaydin, 1999).

Permasalahan yang muncul pada penelitian ini yaitu dimulai dengan adanya kekurangan yang dimiliki oleh metode *k-fold cross validation* yang telah dijelaskan sebelumnya dan belum diketahuinya apakah algoritma CART dan regresi logistik memiliki laju galat yang sama atau berbeda. Sehingga sejalan dengan tujuan penelitian ini yaitu untuk melihat konsistensi metode *k-fold cross validation*, melihat perbandingan kinerja algoritma *classification and regression tree* dan regresi logistik menggunakan uji t $5 \times 2cv$ dan uji *F combined* $5 \times 2cv$. Adapun manfaat yang ingin dicapai yaitu dapat memverifikasi bagaimana konsistensi dari metode *k-fold cross validation*, mendapatkan pengetahuan lebih lanjut mengenai perbandingan algoritma CART dan regresi logistik dengan menggunakan uji t $5 \times 2cv$ dan uji *F combined* $5 \times 2cv$ serta hasil perbandingan dari dua algoritma tersebut nantinya dapat berguna sebagai pedoman bagi penelitian selanjutnya dalam memilih algoritma mana yang ingin digunakan.

II. METODE PENELITIAN

A. Jenis Penelitian dan Sumber Data

Penelitian ini merupakan penelitian dasar. Penelitian dasar atau murni adalah penelitian yang bertujuan menemukan pengetahuan baru yang sebelumnya belum pernah diketahui (Suriasumantri, 1985: 2). Penelitian ini merupakan penelitian pertama yang pernah ada yaitu meneliti tentang perbandingan dua algoritma dengan menggunakan metode uji *F combined* $5 \times 2cv$ atau dengan kata lain belum ada sebelumnya yang melakukan penelitian ini. Data yang digunakan pada penelitian ini adalah data sekunder. Data sekunder adalah data yang diperoleh melalui pihak kedua yang diperoleh dari bahan publikasi, buku teks, hasil penelitian dan jurnal (Hardani, 2020:104). Data yang digunakan dalam penelitian ini bersumber dari rekam medis dan informasi dari pasien yang dilakukan oleh RSUP Dr. M. Djamil Padang.

B. Tahapan Analisis

Tahapan analisis dalam penelitian ini yaitu pertama, menyiapkan data yang akan diteliti yaitu data diabetes melitus dari RSUP Dr. M. Djamil Padang sebanyak 500 amatan dan terdiri dari 22 variabel prediktor serta 1 variabel respon. Kedua, *preprocessing* data untuk melakukan eksplorasi dan penanganan terhadap data hilang. Ketiga, menghitung hasil *repeated k-fold cross validation* untuk CART dan regresi logistik pada 100 kali perulangan. Keempat, membandingkan hasil untuk setiap *fold* pada langkah ke-3. Kelima, melakukan prediksi galat dengan menggunakan *metode k-fold cross validation* pada algoritma CART dan melakukan perulangan sebanyak 5 kali. Keenam, melakukan prediksi galat dengan menggunakan *metode k-fold cross validation* pada algoritma regresi logistik dan melakukan perulangan sebanyak 5 kali. Ketujuh, membandingkan kinerja algoritma CART dan regresi logistik menggunakan uji t $5 \times 2cv$. kedelapan, membandingkan kinerja algoritma CART dan regresi logistik menggunakan uji *F combined* $5 \times 2cv$ dan terakhir mengambil kesimpulan.

C. Classification and Regression Tree

Classification and regression tree (CART) adalah teknik pohon keputusan yang digunakan untuk klasifikasi dan prediksi kasus. Pohon keputusan dapat dianggap sebagai diagram alir yang menunjukkan jalur logis untuk menjawab serangkaian pertanyaan (Ma, 2018: 2). Pada pohon klasifikasi terdapat 3 jenis *node*, yaitu: *Root node*, merupakan *node* awal yang tidak memiliki input dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu. *Internal node*, merupakan *node* percabangan. *Node* ini hanya memiliki satu *input* dan memiliki minimal dua *output*. *Terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*. Pada CART alat yang digunakan dalam perhitungan untuk pemilihan atribut yang dijadikan *node* adalah *gini index*, jadi pada penelitian ini perhitungan *gini index* digunakan untuk pemilahan dan membentuk *node*. Adapun algoritma CART sebagai berikut. Pertama, menghitung *gini index*

$$i(t) = 1 - \sum_{j=1}^n P^2(j|t) \quad (1)$$

Dengan,

$$P(j|t) = \frac{n_j(t)}{n(t)} \quad (2)$$

Keterangan:

$i(t)$: *gini index*

$P(j|t)$: proporsi kelas j pada simpul t , dimana $j = 1, 2, 3, \dots, n$

$n_j(t)$: banyak pengamatan kelas j pada simpul t
 $n(t)$: banyak pengamatan pada simpul t

Kedua, menghitung *goodness of split*. Pemilahan untuk pemilah s pada node t maka dapat dituliskan rumus sebagai berikut:

$$\phi(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (3)$$

Keterangan:

t_L : simpul kanan atau *left node*
 t_R : simpul kiri atau *right node*
 P_L : proporsi banyaknya objek yang masuk pada simpul kanan
 P_R : proporsi banyaknya objek yang masuk pada simpul kiri

Ketiga, Melakukan *splitting* (pemilahan) dari nilai *gini index* dan *goodness of split*. *Splitting* akan berhenti jika jumlah amatan ≤ 10 dan *node* tersebut disebut *terminal node*.

D. Analisis Regresi Logistik Biner

Regresi logistik merupakan model regresi yang digunakan bila variabel responnya bersifat kualitatif. Model regresi logistik sederhana yaitu model regresi logistik untuk satu variabel prediktor X dengan variabel respon Y yang bersifat dikotomi. Nilai variabel $Y = 1$ menyatakan adanya suatu karakteristik dan $Y = 0$ menyatakan tidak adanya suatu karakteristik (Hosmer dan Lemeshow, 1989). Bentuk model regresi logistik adalah (Agresti, 2002):

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x + \dots + \beta_n x_n)} \quad (4)$$

Nilai probabilitas $\pi(x)$ berkisar antara nol dan satu. Untuk mempermudah penaksiran parameter dan untuk mendapatkan fungsi linier maka akan dilakukan transformasi menggunakan transformasi logit. Transformasi harus digunakan untuk menjamin nilai peluang $\pi(x)$ akan selalu berada pada selang $[0,1]$ (Menezes, dkk, 2017). Bentuk persamaan logit dari regresi logistik biner adalah sebagai berikut.

$$g(x) = \beta_0 + \beta_1 x + \dots + \beta_n x_n \quad (5)$$

Apabila persamaan (4) ditransformasikan kedalam bentuk transformasi logit, maka menghasilkan bentuk logit:

$$\pi(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (6)$$

E. Prediksi Galat

Prediksi galat terhadap model adalah mengukur kinerja model dengan menghitung segala bentuk tingkat kesalahan prediksi pada model. Metode paling sederhana dan paling banyak digunakan untuk prediksi galat adalah *cross validation* (Hastie, 2008: 241). CV adalah teknik pengujian keefektifan dari model yang dibentuk dengan melakukan penyusunan ulang (*resampling*) pada data untuk membaginya menjadi dua bagian yaitu data *training* yang digunakan untuk membentuk model dan data *testing* untuk menguji model (James, dkk, 2013: 176). Dalam pengujian menggunakan metode *cross validation* ini dikenal pula istilah stratifikasi. Stratifikasi adalah proses menata ulang data pengamatan untuk memastikan setiap objek pengamatan mewakili keseluruhan data baik pada data *training* maupun pada data *testing*. Stratifikasi ini dapat diterapkan pada metode pengujian *k-fold cross validation* (Refaeilzadeh, dkk, 2016). Metode *k-fold* merupakan metode dalam *cross validation* dengan cara membagi data kedalam k grup dengan sampel (pengamatan) yang sama. Pada data *training* akan digunakan sebanyak $k-1$ sampel dan sisanya 1 grup sampel akan menjadi data *testing*. Untuk menghitung rata-rata atau prediksi galat pada kelompok atau iterasi ke- j untuk $j = 1, 2, \dots, k$ adalah sebagai berikut:

$$\bar{E}_j = \frac{\sum_{i=1}^{n_{uji(j)}} I(y_i \neq \hat{y}_i)}{n} \quad (7)$$

Berdasarkan persamaan (4) maka didapatkan rumus untuk menghitung prediksi galat dengan metode *k-fold cross validation* adalah sebagai berikut:

$$\hat{E}^{CV} = \frac{\sum_{j=1}^k \bar{E}_j}{k} \quad (8)$$

(Wood, dkk, 2007)

Keterangan:

\hat{E}^{CV} : Perdiksi galat dengan metode *k-fold CV*
 k : Jumlah kelompok data
 \bar{E}_j : Rata-rata atau prediksi galat pada iterasi ke- j
 $n_{uji(j)}$: Jumlah pengamatan data uji pada iterasi ke- j

F. Uji t 5×2cv

Metode uji t 5×2cv merupakan pengembangan dari *k-fold cross validation*. Metode uji t 5×2cv memiliki tujuan yaitu untuk membandingkan dua algoritma, apakah dua algoritma memiliki *error* yang sama atau berbeda. Pada uji t 5×2cv terdapat hipotesis yang dimunculkan yaitu sebagai berikut:

H_0 : kedua algoritma memiliki *error* yang sama

H_1 : kedua algoritma memiliki *error* yang berbeda.

Dalam metode uji t 5×2cv data yang tersedia dipartisi menjadi dua *2 fold* serta masing-masing data dilakukan perulangan sebanyak 5 kali. Misalkan pada algoritma A dan B, data yang dimiliki dibagi menjadi data *training* dan data *testing*. Maka akan didapatkan empat *error*: $p_A^{(1)}$ dan $p_B^{(1)}$ untuk *fold* 1 dan $p_A^{(2)}$ dan $p_B^{(2)}$ untuk *fold* 2. Adapun statistik uji yang digunakan yaitu uji t sebagai berikut:

$$t_i^{(j)} = \frac{p_i^{(j)}}{\sqrt{\frac{1}{5} \sum_{i=1}^5 s_i^2}}, \quad j = 1, 2 \text{ dan } i = 1, 2, 3, 4, 5 \quad (9)$$

dimana:

$p_i^{(j)} = p_{Ai}^{(j)} - p_{Bi}^{(j)}$: selisih *error* algoritma A dan B pada *fold* ke-j pada perulangan ke-i

$p_{Ai}^{(j)}$: *error* pada *fold* ke-j dan perulangan ke-i pada algoritma A

$p_{Bi}^{(j)}$: *error* pada *fold* ke-j dan perulangan ke-i pada algoritma B

$\bar{p} = (p_i^{(1)} + p_i^{(2)})/2$: rata-rata selisih *error* pada perulangan ke-i

$s_i^2 = (p_i^{(1)} - \bar{p})^2 + (p_i^{(2)} - \bar{p})^2$: varians pada perulangan ke-i

dan untuk kriteria penolakan: tolak H_0 jika t hitung > t tabel (Dietterich, 1998).

G. Uji F Combined 5×2cv

Metode yang dikembangkan oleh Alpaydin memiliki nama yaitu uji F *combined 5×2cv*, jadi metode ini menggabungkan 10 statistik uji t yang didapatkan dari dua algoritma pada satu statistik uji F. Dimana 10 statistik uji tersebut merupakan hasil perhitungan menggunakan rumus pada persamaan (9), karena pada rumus tersebut terdapat 2 *fold* dan 5 perulangan maka banyak hasil perhitungannya adalah sebanyak 10 hasil. Metode ini sebenarnya tidak jauh berbeda dengan uji t 5×2cv yaitu sama-sama untuk melihat apakah dua algoritma memiliki *error* yang sama atau berbeda, hanya saja pada metode ini menggabungkan 10 statistik uji t kedalam statistik uji F untuk mendapatkan kesimpulan dari perbandingan dua algoritma. Pada uji F *combined 5×2cv* terdapat hipotesis yang dimunculkan yaitu sebagai berikut:

H_0 : kedua algoritma memiliki *error* yang sama

H_1 : kedua algoritma memiliki *error* yang berbeda

adapaun statistiknya sebagai berikut:

$$F = \frac{\sum_{i=1}^5 \sum_{j=1}^2 p_i^{(j)2}}{2 \sum_{i=1}^5 s_i^2}, \quad j = 1, 2 \text{ dan } i = 1, 2, 3, 4, 5 \quad (10)$$

dimana:

$p_i^{(j)} = p_{Ai}^{(j)} - p_{Bi}^{(j)}$: selisih *error* algoritma A dan B pada *fold* ke-j pada perulangan ke-i

$p_{Ai}^{(j)}$: *error* pada *fold* ke-j dan perulangan ke-i pada algoritma A

$p_{Bi}^{(j)}$: *error* pada *fold* ke-j dan perulangan ke-i pada algoritma B

$\bar{p} = (p_i^{(1)} + p_i^{(2)})/2$: rata-rata selisih *error* pada perulangan ke-i

$s_i^2 = (p_i^{(1)} - \bar{p})^2 + (p_i^{(2)} - \bar{p})^2$: varians pada perulangan ke-i

dan untuk kriteria penolakan: tolak H_0 jika F hitung > F tabel (Alpaydin, 1999).

H. Penanganan Data Hilang

Penanganan data hilang dapat menggunakan paket *mice()* pada *software*. MICE adalah singkatan dari *Multivariate Imputasi via Chained Equations* dan dapat menangani data yang hilang dari tipe MAR dan MNAR. MICE dapat melakukan *imputation* atau menghitung nilai pengganti untuk amatan data yang hilang menggunakan bantuan algoritma yang ada pada *classification*. Sementara itu untuk data hilang bertipe MCAR disarankan untuk dilakukan *deletion* (Yadav, 2018).

III. HASIL DAN PEMBAHASAN

A. Preprocessing

1. Eksplorasi Data Hilang

Data yang digunakan dalam penelitian ini bersumber dari rekam medis dan informasi dari pasien yang dilakukan oleh RSUP Dr. M. Djamil Padang terdiri 500 amatan, 22 variabel prediktor dan 1 variabel respon. Namun 7 dari 22 variabel prediktor terdapat data hilang, sehingga perlunya dilakukan eksplorasi terhadap data hilang tersebut. Pada Tabel 1 dapat dilihat jumlah amatan dan persentase terhadap data hilang dari 7 variabel tersebut.

Tabel 1. Jumlah Amatan dan Persentase Data Hilang pada Variabel Prediktor

Variabel Prediktor	Jumlah Data Hilang	Persentase
Kadar HB (mg/dl)	27	5,4%
Kadar Kolesterol (mg/dl)	126	25,2%
GDS (mg/dl)	23	4,6%
GDP (mg/dl)	73	14,6%
GD2PP (mg/dl)	64	12,8%
Nyeri	6	1,2%
Penurunan Berat Badan	1	0,2%

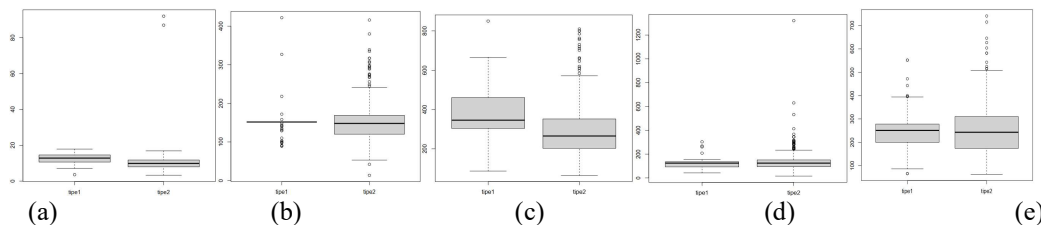
Dari Tabel 1 terlihat bahwa persentase data hilang semua variabel tidak lebih dari 80% persen, artinya pada data hilang tersebut dapat dilakukan *imputation*. Data hilang bisa saja terjadi karena kerusakan alat pengukuran, rekap medis yang tidak lengkap maupun dari petugas RS tersebut yang tidak menjalankan tugasnya dengan semestinya. Data hilang tentunya menghambat proses analisis, oleh karena itu perlunya dilakukan penanganan terhadap data hilang tersebut.

2. Penanganan data hilang

Sebelum dilakukannya penanganan terhadap data hilang, perlu diketahui tipe data hilang tersebut. Uji Little merupakan uji yang bisa digunakan untuk menentukan apakah data hilang bertipe MCAR atau bukan. Hasil Uji Little menggunakan paket *mcar_test()* pada *software* R menghasilkan *p-value* sebesar $0,0002 < \alpha = 0,05$ maka H_0 ditolak, sehingga dapat disimpulkan bahwa data hilang yang terdapat pada gugus data diabetes melitus bukan bertipe MCAR, dengan kata lain data hilang bertipe MAR atau MNAR. Penanganan data hilang bertipe MAR atau MNAR dapat dilakukan dengan *imputation* menggunakan paket *mice()* pada *software* R. MICE dapat melakukan *imputation* menggunakan algoritma yang ada pada *classification* atau dengan kata lain menghitung nilai pengganti dilakukan dengan cara memprediksinya dengan bantuan algoritma yang ada pada *classification*, dimana *imputation* tersebut bekerja dengan cara menghitung nilai pengganti yang paling memungkinkan dengan bantuan berdasarkan informasi yang didapatkan dari hasil klasifikasi, pada penelitian ini peneliti akan menggunakan algoritma CART.

B. Deskripsi Data

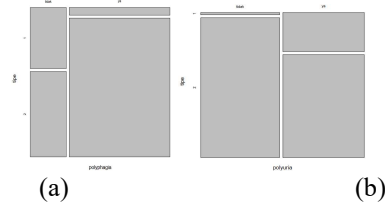
Setelah dilakukan penanganan terhadap data hilang, deskripsi data dapat dilihat melalui visualisasi data untuk mengetahui karakteristik dari variabel dengan tipe data numerik maupun kategorik yang ada pada gugus data diabetes melitus. Penyajian visualisasi data dengan tipe numerik dapat dilihat pada Gambar 1 sebagai berikut.



Gambar 1. (a) *Boxplot* Kadar HB, (b) *Boxplot* Kadar Kolesterol, (c) *Boxplot* GDS, (d) *Boxplot* GDP, (e) *Boxplot* GD2PP

Berdasarkan Gambar 1 diketahui bahwa penderita DM tipe 1 maupun tipe 2 memiliki Kadar HB, Kadar Kolesterol, GDS, GDP dan GD2PP yang bervariasi. Namun ketika melihat perbedaan DM tipe 1 dan tipe 2, terlihat

bahwa semua variabel yang terdapat pada penderita DM tipe 2 memiliki variansi yang lebih besar dari pada penderita DM tipe 1. Selain dari visualisasi untuk data tipe numerik, deskripsi data pada variabel dengan tipe data kategorik juga dapat dilihat melalui visualisasi data, seperti Gambar 2 yang menampilkan visualisasi variabel yang paling penting dari semua variabel dengan tipe data kategorik yaitu variabel *polyphagia* dan *polyuria* yang disajikan menggunakan *mosaic plot* sebagai berikut.



Gambar 2. (a) *Mosaic Plot Polyphagia*, (b) *Mosaic Plot Polyuria*

Pada Gambar 2 merupakan perbandingan pasien DM tipe 1 dan tipe 2 yang mengalami *polyphagia* dan *polyuria*. Dan terlihat bahwa penderita DM tipe 2 lebih cenderung mengalami *polyphagia* dan *polyuria* dari pada penderita DM tipe 1, namun dapat dilihat juga bahwa pasien DM tipe 1 maupun tipe 2 kebanyakan mengalami *polyphagia* dan *polyuria* untuk kasus data penderita DM di RSUP Dr. M. Djamil Padang pada tahun 2020-2021.

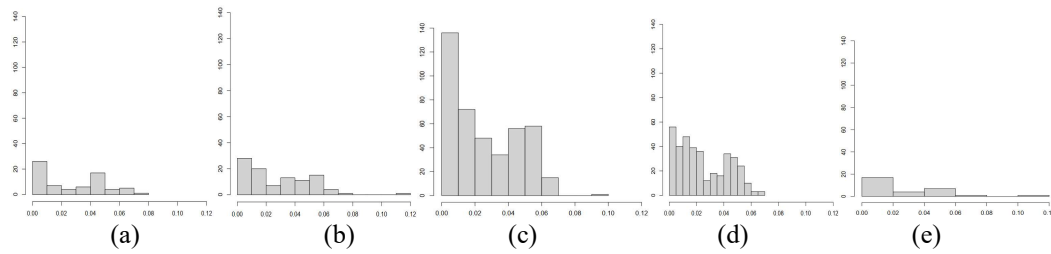
C. K-Fold Cross Validation

Pada bagian ini dilakukan pembagian data menjadi data *training* untuk membentuk model dan *testing* untuk mendapatkan prediksi galat CART dan regresi logistik pada data penderita diabetes melitus di RSUP Dr. M. Djamil Padang tahun 2020-2021 dengan menggunakan metode *k-fold cross validation*. Selain itu untuk melihat konsistensi terhadap metode *k-fold cross validation* dalam membandingkan CART dan regresi logistik, maka dilakukan perulangan sebanyak 100 kali atau disebut juga sebagai *repeated k-fold cross validation* dengan jumlah *fold* sebanyak 10 *fold*. Berikut ini adalah rekap dari perbandingan hasil prediksi galat CART dan regresi logistik menggunakan *10-fold cross validation* dengan 100 kali perulangan pada data diabetes melitus yang dilihat dari hasil perbandingan jumlah galat terkecil yang dihasilkan antara CART dan regresi logistik dan dari hasilnya muncul lima kejadian yang terjadi pada perbandingan jumlah galat terkecil tersebut yang dilihat dari 10 *fold* untuk 100 kali perulangan serta dilihat jumlah frekuensi untuk setiap kejadian yang terjadi.

Tabel 2. Perbandingan Jumlah Prediksi Galat Terkecil yang Dihasilkan antara CART dan Regresi Logistik

Kejadian	CART	Regresi Logistik	Frekuensi
1	6	4	7
2	7	3	10
3	8	2	42
4	9	1	38
5	10	0	3
	Σ		100

Berdasarkan Tabel 2 terlihat bahwa metode *k-fold cross validation* belum konsisten digunakan untuk membandingkan CART dan regresi logistik, hal ini terbukti dengan prediksi galat terkecil yang dihasilkan diantara kedua algoritma belum konsisten atau dengan kata lain ada kejadian yang menghasilkan tidak semua *fold* yang menghasilkan prediksi galat terkecil hanya pada satu algoritma saja, seperti contoh pada kejadian ke-1 sampai ke-4 tidak hanya CART yang menghasilkan prediksi galat terkecil tetapi galat terkecil juga dihasilkan oleh regresi logistik pada *fold* yang berbeda pula. Namun ada kejadian yang menunjukkan *k-fold cross validation* sudah konsisten yaitu pada kejadian ke-5 yang mana pada 10 *fold* CART selalu menghasilkan prediksi galat terkecil dan kejadian tersebut hanya terjadi sebanyak 3 kali dari 100 kali perulangan. Gambaran konsistensi dari metode *k-fold cross validation* dapat dilihat melalui histogram untuk setiap kejadian yaitu sebagai berikut.



Gambar 3. (a) Histogram Kejadian ke-1, (b) Histogram Kejadian ke-2, (c) Histogram Kejadian ke-3, (d) Histogram Kejadian ke-4, (e) Histogram Kejadian ke-5

Gambar 3 merupakan histogram selisih prediksi galat CART dengan regresi logistik untuk setiap kejadian yang mana pada setiap kejadian memiliki jumlah prediksi galat yang dihasilkan berbeda-beda sesuai dengan jumlah frekuensi kejadiannya dan setiap satu frekuensi terdiri dari 10 prediksi galat yang didapatkan dari 10 *fold*. Sebaran data dari semua histogram terlihat bahwa jumlah kejadian yang paling sering terjadi yaitu pada kejadian ke-3 dan yang paling jarang terjadi yaitu pada kejadian ke-5. Dari Tabel 2 sebelumnya juga sudah terlihat bahwa pada kejadian ke-1 sampai ke-4 ada prediksi galat terkecil yang dihasilkan oleh CART dan ada juga yang dihasilkan oleh regresi logistik, sehingga hal ini yang menunjukkan bahwa metode *k-fold cross validation* belum konsisten. Sementara pada kejadian ke-5, untuk setiap *fold* prediksi galat regresi logistik selalu lebih besar dari pada CART, sehingga menunjukkan pada kejadian ini metode *k-fold cross validation* konsisten. Kesimpulannya adalah dari sekian banyak perulangan yaitu 100 perulangan, kejadian yang paling banyak terjadi yaitu kejadian yang menunjukkan bahwa metode *k-fold cross validation* belum konsisten.

D. Uji t 5×2cv

Sebelumnya telah dilakukan pembagian data, mendapatkan model dan hasil prediksi galat CART serta regresi logistik. Selanjutnya Dalam masalah perbandingan algoritma metode pertama yang dilakukan yaitu metode uji t 5×2cv digunakan untuk melihat apakah algoritma CART dan regresi logistik memiliki galat yang sama atau berbeda. Pada uji t 5×2cv menghasilkan 10 keputusan yang dapat dilihat pada Tabel 3 sebagai berikut.

Tabel 3. Hasil Uji t 5×2cv (t Tabel = 2.571)

Fold	Perulangan	t Hitung	Keputusan
Fold 1	1	2.1475	Gagal Tolak H_0
	2	3.0871*	Tolak H_0
	3	2.1568	Gagal Tolak H_0
	4	6.0001*	Tolak H_0
	5	4.2871*	Tolak H_0
Fold 2	1	1.9870	Gagal Tolak H_0
	2	2.5225	Gagal Tolak H_0
	3	5.5688*	Tolak H_0
	4	0.9480	Gagal Tolak H_0
	5	0.7530	Gagal Tolak H_0

Berdasarkan Tabel 3 terlihat bahwa hasil dari uji t 5×2cv untuk *fold* dan perulangan yang berbeda ada yang t hitung < t tabel, maka keputusannya yaitu gagal tolak H_0 bahwa algoritma CART dan regresi logistik memiliki laju galat yang sama dan ada juga yang t hitung > t tabel, maka keputusannya yaitu tolak H_0 bahwa algoritma CART dan regresi logistik memiliki galat yang berbeda. Artinya pada uji t 5×2cv belum konsisten dalam menghasilkan keputusan sehingga dapat menyulitkan peneliti dalam mengambil kesimpulan dalam membandingkan algoritma CART dan regresi logistik. Namun dari 10 hasil uji t 5×2cv, keputusan yang paling banyak adalah gagal tolak H_0 , sehingga peneliti menyimpulkan keputusan yang akan diambil yaitu gagal tolak H_0 bahwa algoritma CART dan regresi logistik memiliki galat yang sama.

E. Uji F Combined 5×2cv

Metode selanjutnya yang digunakan untuk melihat perbandingan algoritma yaitu metode uji F *combined* 5×2cv yang memiliki perbedaan dengan uji t 5×2cv. pada uji F *combined* 5×2cv ini hanya menghasilkan 1 keputusan dan uji F *combined* 5×2cv juga digunakan untuk melihat apakah algoritma CART dan regresi logistik memiliki galat yang sama atau berbeda. Dari hasil uji F *combined* 5×2cv, didapatkan F hitung sebesar $2.3192 <$ dari F tabel sebesar 4,74, maka keputusannya yaitu gagal tolak H_0 bahwa algoritma CART dan regresi logistik memiliki galat yang sama. Dari metode uji F *combined* 5×2cv dan uji t 5×2cv yang telah dilakukan terlihat bahwa uji F *combined* 5×2cv lebih baik digunakan untuk mendapatkan kesimpulan dari hasil perbandingan dua algoritma dikarenakan hanya menghasilkan 1 keputusan saja, berbeda dengan uji t 5×2cv yang memiliki kemungkinan untuk mendapat keputusan yang berbeda dari 10 statistik uji yang dihasilkan.

IV. KESIMPULAN

Dari 100 kali perulangan metode *10-fold cross validation* hanya konsisten sebanyak 3 kali yang mana hal ini menunjukkan bahwa metode *k-fold cross validation* memiliki konsistensi yang tidak baik dalam membandingkan algoritma CART dan regresi logistik untuk data diabetes melitus. Selain itu, pada uji t 5×2cv belum konsisten dalam menghasilkan keputusan sehingga dapat menyulitkan peneliti untuk mengambil kesimpulan dalam membandingkan dua algoritma. Namun hasil dari 10 hasil uji t 5×2cv, keputusan yang paling banyak adalah gagal tolak H_0 , sehingga peneliti menyimpulkan keputusan yang akan diambil yaitu gagal tolak H_0 bahwa algoritma CART dan regresi logistik memiliki galat yang sama. Sedangkan pada uji F *combined* 5×2cv, hasil yang didapatkan yaitu F hitung sebesar $2.3192 <$ dari F tabel sebesar 4,74, maka keputusannya yaitu gagal tolak H_0 bahwa algoritma CART dan regresi logistik memiliki galat yang sama. Dan terlihat bahwa uji F *combined* 5×2cv hanya menghasilkan 1 keputusan saja, sehingga bisa disimpulkan bahwa uji F *combined* 5×2cv lebih baik digunakan untuk mendapatkan kesimpulan dari hasil perbandingan dua algoritma. Serta berdasarkan kekurangan dari metode *k-fold cross validation* dan uji t 5×2cv, dalam masalah perbandingan algoritma peneliti menyarankan untuk menggunakan metode uji F *combined* 5×2cv.

DAFTAR PUSTAKA

- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley and Sons, Inc. New York.
- Alpaydin, Ethem. (1999). *Combined 5x2 cv F Test for Comparing Supervised Classification Learning Algorithms*. *Neural Computation*, 11, 1885-1892.
- Breiman, L. (2001). Random Forests, *Machine Learning* 45 (1), 5-32.
- Dietterich, T.G. (1998). *Approximate Statistical Test for Comparing Supervised Classification Learning Algorithms*. *Neural Computation*, 10, 1895-1923.
- Gholamian, M. Siami, dan Z. Hajimohammadi. (2013). *Credit Scoring In Banks and Financial Institutions Via Data Mining Techniques: A Literature Review*, *J. AI Data Mining Journal AI Data Min.*, Vol. 1, No. 2, pp. 119–129.
- Hosmer, D.W., dan S. Lemeshow. (2000). *Applied Logistic Regression*. Edisi ke-2. John Wiley and Sons Inc, Canada.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. California: Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York: Springer.
- Li, Cheng. (2013). *Little's test of missing completely at random*. *The Stata Journal*, 13(4), 795-809.
- M. Kamber. (2006). *Data Mining Concepts and Techniques Second Edition*. San Francisco: Morgan Kaufman.
- Ma, Xin. (2018). *Using Classification and Regression Tree A Practical Primer*. Charlotte, NC: Information Age Publishing.
- Menezes, F. S., Liska, G. R., Cirillo, M. A., Vivanco, M. J. (2017). *Data Classification with Binary Response through the Boosting Algorithm and Logistic Regression*. *Expert Systems with Application*, Vol 69: 63-65.
- Refaeilzadeh, P., Tang, L. & Liu, H. (2016). *Cross Validation*. *Encyclopedia of Database Systems*. New York: Springer. DOI 10.1007/978-1-4899-7993-3_565-2.

Suriasumantri, S (1985), *Filsafat Ilmu*. Sinar harapan, Jakarta.

Wood, I., Vixxcher, P., & Mengersen, K. (2007). *Classification Based Upon Gene Expression Data: Bias and Precision of Error Rates*. *Bioinformatics*, 23(11), 1363-1370.

Yadav, M. L., & Roychoudhury, B. (2018). *Handling missing values: A study of popular imputation packages in R*. *Knowledge-Based Systems*, 160, 104-118.