E-ISSN 2527-9378

Jurnal Statistika Industri dan Komputasi
Volume 5, No. 1, Januari 2020, pp. 66-74

PERBANDINGAN METODE *CLASSIFICATION AND REGRESSION TREE* (CART) DAN METODE REGRESI LOGISTIK BINER DALAM MENGKLASIFIKASI STATUS WANITA BEKERJA DI KOTA KUPANG

Marina Elviana Fobia [1], Yudi Setyawan [2]
[1,2)] Jurusan Statistika, Fakultas Sains Terapan, IST AKPRIND Yogyakarta
e-mail: marinaelviana17@gmail.com

Abstrak.Pengklasifikasian merupakan salah satu metode statistik dalam mengelompokkan suatu data yang disusun secara sistematis. Pengklasifikasian suatu objek dapat dilakukan dengan dua pendekatan metode klasifkasi. Penelitan ini menggunakan Metode Classification And Regresion Tree (CART) dan Regresi Logistik Biner untuk mengklasifikasikan Status Wanita Bekerja di Kota Kupang. Diperoleh hasil bahwa analisis metode CART dan Regresi Logistik Biner tidak jauh berbeda yaitu variabel yang mempengaruhi status wanita bekerja adalah variabel umur (X_2) dan status perkawinan (X_3).Perbedaan prediksi pada kedua metode tersebut yaitu prediksi wanita bekerja pada metode CART yaitu wanita dengan umur 15-64 tahun dan memiliki status perkawinan menikah atau cerai. Sedangkan prediksi pada regresi logistik biner wanita bekerja hanya wanita dengan umur 15-64 tahun. Wanita menikah atau cerai diprediksi tidak bekerja. presentase ketepatan klasifikasi metode CART dan Regresi Logistik Biner sebesar 65,2%. Artinya tidak ada perbedaan presentase ketepatan klasifikasi antara metode CART dan regresi logistik biner.

Kata kunci: CART, Status Bekerja, Regresi Logistik Biner.

1. PENDAHULUAN

Penduduk yang besar merupakan asset pembangunan yang potensial terutama tenaga kerja wanita dimana penduduk wanita di Indonesia cukup besar yaitu lebih dari separuh jumlah penduduk Indonesia. Wanita membutuhkan pekerjaan untuk kelangsungan hidup, baik wanita yang belum menikah maupun wanita yang sudah menikah. Pentingnya pekerjaan untuk wanita yang belum menikah adalah untuk persiapan masa depan yang lebih cerah dan mandiri serta tidak bergantung pada orang tua dan calon suaminya. Pentingnya pekerjaan bagi wanita yang sudah menikah, pertama dapat membantu meningkatkan penghasilan rumah tangga dan meringankan beban suami, kedua sebagai antisipasi bagi wanita yang berstatus cerai hidup maupun cerai mati, dimana wanita memiliki keharusan untuk menafkahi anggota keluarga. Dari data yang diperoleh dari Badan Pusat Statistik Kota Kupang Provinsi Nusa Tenggara Timur jumlah wanita yang bekerja lebih banyak yaitu sebanyak 70901 jiwa dibandingkan dengan wanita yang belum atau tidak bekerja yaitu sebanyak 61162 jiwa.

Faktor-faktor yang diduga mempengaruhi seorang wanita untuk bekerja yaitu pendidikan, umur, status perkawinan, jumlah anggota rumah tangga, status dalam keluarga. Status Wanita bekerja diklasifikasikan menjadi dua yaitu bekerja dan tidak bekerja. Tujuan klasifikasi adalah untuk

mempermudah mengenali, membandingkan, dan mempelajari suatu objek tertentu dalam hal ini status wanita bekerja. Membandingkan berarti mencari persamaan dan pebedaan sifat atau ciri pada wanita bekerja di kota Kupang. Melakukan klasifikasi merupakan salah satu metode statistika untuk pengelompokan suatu data dalam susunan yang matematis. Pengklasifikasian objek dapat dilakukan dengan pendekatan parametrik dan non parametrik. Salah satu metode statistika non parametrik yang digunakan dalam pengklasifikasian adalah metode klasifikasi berstruktur pohon yang diperkenalkan oleh Breiman 1948. Analisis yang dapat digunakan untuk klasifikasi adalah metode CART dan metode regresi logistik biner.

Penelitian ini dilakukan untuk mengetahui [1] Klasifikasi status wanita bekerja di kota Kupang menggunakan metode *Classification and Regression Tree*(CART) [2] Hubungan antara pendidikan, umur, status perkawinan, jumlah anggota dalam rumah tangga, dan status dalam keluarga terhadap status wanita bekerja di kota Kupang [3] Klasifikasi status wanita bekerja di kota Kupang menggunakan metode Regresi Logistik Biner [4] perbandingan ketepatan klasifikasi dari metode CART *(Classification And Regression Tree)* dan metode Regresi Logistik Biner.

2. METODE PENELITIAN

2.1 Bahan

Penelitian ini menggunakan data sekunder yaitu data Survei Sosial Ekonomi Nasional (SUSENAS) 2018 di Kota Kupang yang diperoleh dari Badan Pusat Statistik Provinsi Nusa Tenggara Timur. Data pada penelitan ini adalah penduduk perempuan di Kota Kupang yang berumur 15 tahun ke atas pada tahun 2018 yaitu sebanyak 733 orang. *Software* yang digunakan untuk analisis adalah *software SPSS* 18.0.

2.2 Variabel Penelitan

Variabel yang digunakan dalam penelitian ini terdiri dari 5 variabel independen yaitu tingkat pendidikan (X_1), umur (X_2), status perkawinan (X_3), Status dalam rumah tangga (X_4), jumlah anggota dalam rumah tangga (X_5), dan 1 variabel dependen berbentuk kategorik yaitu Status wanita bekerja (Y) dengan 0 untuk tidak bekerja, 1 untuk bekerja.

2.3 Metode

2.3.1 Metode Classification and Regression Tree (CART)

Metode CART *(Classification And Regression Tree)* merupakan metode statistika nonparametrik yang dikembangkan untuk keperluan analisis klasifikasi, menyatakan bahwa apabila variabel respon adalah berbentuk kontinu maka metode yang digunakan adalah metode regresi pohon *(regression tree)* yang akan menghasilkan pohon regresi. Apabila variabel respon berbentuk kategorik maka metode yang digunakan adalah metode klasifikasi pohon *(classification tree)* yang akan menghasilkan pohon klasifikasi. Tahapan pembentukan pohon klasifikasi CART terdapat 3 yaitu:

1. Pemilihan Pemilah

Terdiri dari 3 langkah perhitungan yaitu:

a. Menghitung Nilai *Indeks Gini*

$$i(t) = 1-\sum_j p^2(j|t) \quad j=(0,1) \qquad (2.1)$$

Keterangan:

$i(t)$ = Nilai indeks gini

$p(j|t)$ = Proporsi pada kelas j pada simpul *t*.

b. Mencari Nilai *Split Point*

Nilai *split-point* diperoleh dengan mencari nilai tengah dari 2 nilai atribut yang sudah diurutkan terlebih dahulu.

c. Menghitung Nilai *Goodness of Split*

Godness of split merupakan evaluasi pemilahan oleh pemilah *x* pada simpul *t* yang didefinisikan sebagai penurunan keheterogenan yang disebut nilai Impuritas (Breiman et al. 1984). dan dapat ditulis sebagai:

$$\Delta i\,(s,t) = i(t) - P_L\, i(t_L) - P_R\, i(t_R) \qquad (2.2)$$

Keterangan:

$i(t)$ = Nilai indeks gini

$P_L\, i(t_L)$ = Proporsi pengamatan dari simpul *t* menuju kiri.

$P_R\, i(t_R)$ = Proporsi pengamatan dari simpul *t* menuju kanan.

2. Penentuan Simpul Teminal

Simpul dikatakan simpul terminal ketika suatu simpul *t* mencapai batas akhir yang ditentukan sehingga tidak terdapat penurunan impuritas secara berarti. simpul *t* tidak dipilah lagi tetapi dijadikan simpul terminal dan pembentukan pohon berhenti jika tidak ada lagi peubah respon yang signifikan menunjukkan perbedaan terhadap peubah penjelas, jika pohon sudah mencapai batas nilai maksimum dari pohon spesifikasi, jika ukuran *child node* kurang dari ukuran *childnode* minimum spesifikasi.

3. Penendaan Label Kelas

Penandaan label kelas yang telah ditentukan berdasarkan jumlah kelas terbanyak pada simpul (Breiman et al. 1984). Label kelas pada simpul terminal *t* ditentukan melalui aturan jumlah terbanyak yaitu:

$$P(j_0|t) = \max_j P(j|t) = \max j \frac{Nj(t)}{N(t)} \quad (2.3)$$

Keterangan :

$P(j_0|t)$ = Proporsi kelas j pada simpul *t*

$N_j(t)$ = Jumlah pengamatan pada kelas j pada simpul *t*

$N(t)$ = Jumlah pengamatan pada simpul *t*

2.3.2 Regresi Logistik Biner

Model regresi logistik biner digunakan untuk menganalis hubungan antara satu variabel respon dan beberapa variabel bebas, dengan variabel respon berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karateristik. Menurut Hosmer dan Lemeshow (2000), bentuk model regresi logistik biner dengan variabel prediktor adalah sebagai berikut:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}} \quad (2.4)$$

Keterangan :

p :banyaknya variabel prediktor

$\pi(x)$: peluang terjadinya suatu kesuksesan

Estimasi parameter dalam regresi logistik dilakukan dengan metode *Maximum Likelihood Estimation*. Metode tersebut mengestimasi parameter β dengan cara memaksimumkan fungsi *likelihood* dan mensyaratkan bahwa data harus mengikuti distribusi tertentu. Pada regresi logistik, setiap pengamatan mengikuti distribusi Bernoulli sehingga dapat ditentukan fungsi *likelihood*

Fungsi probabilitas distribusi Bernoulli sebagai berikut:

$$f(\underline{xi}) = \pi(x_i)^{yi}(1 - \pi(x_i))^{1-yi};\ y_i = 0,1 \quad (2.5)$$

Uji signifikansi model menggunakan uji simultan (*uji likelihood ratio*), uji Parsial (uji Wald). Interpretasi koefisien menggunakan nilai *odds ratio.*

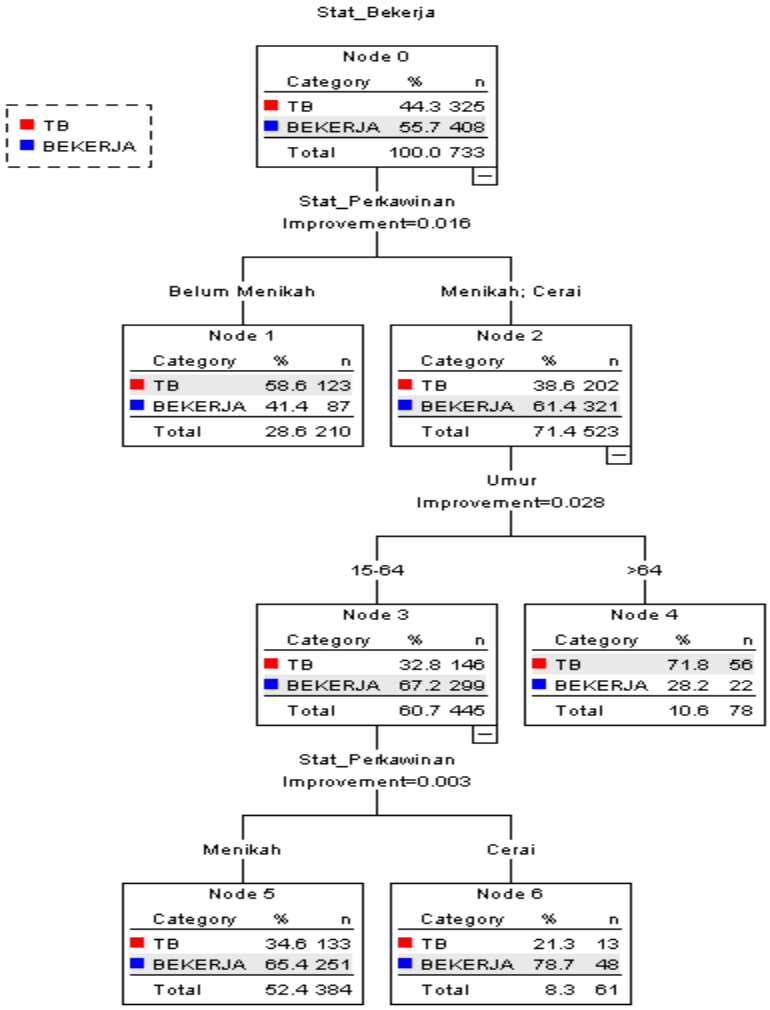
2.3.3 Ketepatan Klasifikasi Metode CART dan Regresi Logistik Biner

Ketepatan prediksi pada metode CART dapat diketahui dari tabel klasfikasi. Cabang *(node)* yang digunakan dalam pengklasifikasian respon adalah cabang-cabang yang mengakhiri pengelompokan atau cabang-cabang terakhir yang disebut simpul terminal. Suatu simpul terminal dengan kelas domain menunjukkan nilai prediksi pada analisis CART. Sedangkan dalam regresi logistik biner ketepatan klasifikasi ditentukan berdasarkan nilai peluang dari model yang terbentuk dan dibandingkan dengan nilai *cut point* (c).

3.HASIL DAN PEMBAHASAN

3.1 Metode Classification and Regression Tree(CART)

Pohon Klasifikasi CART yang terbentuk adalah



Pohon klasifikasi CART yang terbentuk diperoleh jumlah simpul sebanyak 7 simpul yaitu 1 simpul asal, 2 simpul dalam dan 4 simpul terminal dengan kedalaman pohon klasifikasi adalah 3, dan ukuran *child node* minimum 50. Proses pemecahan simpul berhenti karena pertumbuhan pohon sudah mencapai kedalaman 3. Hal ini dapat dilihat pada simpul 5 dan 6 yang tidak dipilah lagi karena batas kedalaman pohon klasifikasi sudah mencapai kedalaman 3. Hasil klasifikasi dan kemungkinan bekerja atau tidak bekerja yang diperoleh yaitu Wanita bekerja adalah wanita dengan umur 15-64 tahun dan memiliki status perkawinan menikah atau cerai. Sedangkan wanita yang tidak bekerja adalah wanita dengan umur lebih dari 64 tahun dan belum minikah. Prediksi ketepatan klasifikasi CART adalah 65,2%.

3.2 Analisis Regresi Logistik Biner

a. Uji Simultan (*uji likelihood ratio*)

	χ^2	DF	χ^2 7;0.05	*P-value*
Model	74,799	7	14,067	0,000

Diperoleh nilai χ^2 = 74,799 > χ^2 tabel (14,067) atau *P-value* 0,000 < 0,05 maka H_0 ditolak. Dapat disimpulkan bahwa terdapat minimal satu variabel independen berpengaruh signifikan terhadap variabel dependen.

b. Uji Parsial (uji Wald)

	B	Wald	DF	*P-value*
Umur (X_2)	1,755	33,900	1	0,000
Status Perkawinan (X_3) (1)	-1,489	25,454	1	0,000
Status Perkawinan (X_3) (2)	-0,544	4,004	1	0,045
Constan	-0,589	5,163	1	0,023

Variabel umur(X_2) dan variabel srtatus perkawinan (X_3) memiliki nilai W_j > χ^2(0.05,1)=3,841 atau nilai *P-value* < 0,05 maka dapat dikatakan bahwa kedua variabel tersebut berpengaruh terhadap status wanita bekerja. Model regresi yang terbentuk adalah: $\hat{\pi}(x) = \frac{e^{-0,589+1,755X2-1,489Xs(1)-0,544Xs(2)}}{1+e^{-0,589+1,755X2-1,489Xs(1)-0,544Xs(2)}}$

Klasifikasi regresi logistik biner diperoleh berdarkan nilai peluang yang terbentu dari model. Hasil klasifikasi diperoleh dengan membandingkan nilai peluang dengan nilai *cut point (c)*=0,5. Apabila nilai peluang pada sebuah kategori kurang dari nilai 0,5 maka diprediksi tidak bekerja. Hasil prediksi yang diperoleh yaitu wanita dengan umur 15-64 tahun diprediksi bekerja, sedangkan wanita dengan umur lebih dari 64 tahun dan wanita yang menikah atau cerai diprediksi tidak bekerja. Hasil presentase ketepatan klasifikasi yang terbentuk adalah 65,2%.

3.3 Perbandingan Hasil Klasifikasi CART dan Regresi Logistik Biner

Hasil analisis dari metode CART dan Regresi Logistik Biner dapat dilihat pada tabel 3.3

Tabel 3.3 Hasil analisis Metode CART dan Regresi Logistik Biner

Nomor	Metode CART	Regresi Logistik
1.	Variabel yang mempengaruhi status wanita bekerja adalah umur(X_2) dan status perkawinan (X_3)	Variabel yang mempengaruhi status wanita bekerja adalah umur(X_2) dan status perkawinan (X_3)
2.	Wanita bekerja adalah wanita dengan umur 15-64 tahun dan memiliki status perkawinan menikah atau cerai.	Wanita bekerja adalah wanita dengan umur 15-64 tahun.
3.	Wanita tidak bekerja adalah wanita dengan umur ≥65 tahun dan belum menikah.	Wanita tidak bekerja adalah wanita dengan umur ≥65 tahun dan memiliki status perkawinan menikah atau cerai.
4.	Presentase ketepatan klasifikasi sebesar 65,2%	Presentase ketepatan klasifikasi sebesar 65,2%

Interpretasi hasil analisis:

Dari Tabel 3.3 dapat dilihat bahwa presentase ketepatan klasifikasi dari metode CART dan Regresi Logistik Biner sebesar 65,2% sehingga tidak dapat dibandingkan untuk melihat metode manakah yang memberikan hasil ketepatan klasifikasi paling baik dalam mengklasifikasikan status wanita bekerja di Kota Kupang. Namun dapat diinterpretasikan bahwa hasil analisis dari metode CART dan Regresi Logistik Biner tidak jauh berbeda yaitu variabel yang mempengaruhi status wanita bekerja adalah variabel umur (X_2) dan status perkawinan (X_3).Perbedaan prediksi pada kedua metode tersebut adalah prediksi wanita bekerja pada metode CART yaitu wanita dengan umur 15-64 tahun dan memiliki status perkawinan menikah atau cerai. Sedangkan prediksi pada regresi logistik biner wanita bekerja hanya wanita dengan umur 15-64 tahun. Wanita menikah atau cerai diprediksi tidak bekerja.

4. PENUTUP

4.1 Kesimpulan

Setelah dilakukan analisis pada data menggunakan metode CART dan Regresi Logistik Biner diperoleh kesimpulan bahwa:

1. Metode CART memprediksi bahwa wanita yang bekerja adalah wanita dengan umur 15-64 tahun dan memiliki status perkawinan menikah atau cerai. Sedangkan wanita yang diprediksi tidak bekerja adalah wanita yang berusia lebih dari 64 tahun dan wanita yang belum menikah. Presentase ketepatan klasifikasi sebesar 65,2%.

2. Dari 5 variabel yang diduga mempengaruhi status wanita bekerja di kota Kupang terdapat 2 varaibel yang berpengaruh terhadap status wanita bekerja yaitu variabel umur (X_2) dan status perkawinan (X_3)
3. Regresi logistik biner memprediksi bahwa wanita yang bekerja adalah wanita yang berumur 15-64 tahun, sedangkan wanita yang diprediksi tidak bekerja adalah wanita yang berusia lebih dari 64 tahun dan memiliki status perkawinan menikah atau cerai. Presentase Ketepatan klasifikasi sebesar 6,52%

   Model umum regresi logistik yang terbentuk adalah

   $$\pi(x)=\frac{e^{-0.589+1.755\,X2-1.489X3(1)-0.544X3(2)}}{1+e^{-0.589+1.755\,X2-1.489X3(1)-0.544X3(2)}}$$
4. presentase ketepatan klasifikasi dari metode CART dan Regresi Logistik Biner sebesar 65,2% sehingga tidak dapat dibandingkan untuk melihat metode manakah yang memberikan hasil ketepatan klasifikasi paling baik dalam mengklasifikasikan status wanita bekerja di Kota Kupang. Namun terdapat perbedaan yaitu pada hasil prediksi dari kedua metode tersebut. prediksi wanita bekerja pada metode CART yaitu wanita dengan umur 15-64 tahun dan memiliki status perkawinan menikah atau cerai. Sedangkan prediksi pada regresi logistik biner wanita bekerja hanya wanita dengan umur 15-64 tahun. Wanita menikah atau cerai diprediksi tidak bekerja.

4.2 Saran

Kepada para pengambil kebijakan pemerintahan agar memperluas lagi lapangan pekerjaan bagi wanita-wanita usia produktif yaitu umur 15-64 tahun karena dengan terbukanya lapangan pekerjaan maka dapat meningkatkan jumlah wanita yang bekerja atau dengan kata lain dapat mengurangi jumlah pengangguran di kota Kupang, dan meningkatkan mutu pendidikan bagi wanita usia produktif yang belum menikah sebagai modal untuk bekerja di kemudian hari.

UCAPAN TERIMAKASIH

Dalam penyusunan tulisan ini, banyak pihak yang telah memberikan dukungan kepada penelitian ini. Peneliti menyampaikan terima kasih kepada Institut Sains & Teknologi AKPRIND Yogyakarta yang telah memberikan fasilitas sarana dan prasarana dalam pelaksanaan penelitian, khususnya di Labolatorium Statistika serta kepada Bapak/Ibu Dosen Jurusan Statistika IST AKPRIND Yogyakarta atas arahan dan bimbingannya.

DAFTAR PUSTAKA

Agresti, A. (1990). *Categorical Data Analysis.* New York: Jhon Willey and Sons.

Breiman, L. (1984). *Classification and Regresion Tree.* New York: Chapman and Hall.

Hosmer DW & Lameshow (1989). *Aplied Regresi Logistic*. New York: Jhon Willey and Sons.

Jiwadiana, I. G. (2015). Klasifikasi Karakteristik Kecelakaan Lalu Lintas di Kota Denpasar Dengan Pendekatan CART. *E-Jurnal Matematika*, 1-6.

Mujahi, M. (2016). Perbandingan Metode Regresi Logistik Biner dan Metode Backpropogation Dalam Menentukan Model Terbaik Untuk Klasifikasi Pengguna KB. *Jurnal Gaussian*, 1-10.

Musa, M. (2017). *Perbandingan Metode Regresi Logistik Biner dan Chi-Square Automath Interaction Dalam Mengklasifikasi Karakteristik Perokok di Indonesia* . Yogyakarta: Institut Sains & Teknologi AKPRIND Yogyakarta .

Noeryanti. (2015). *Petunjuk Praktikum MetodeStatistika 2.* Yogyakarta: Institut Sains & Teknologi Akprind Yogyakarta.

Rajagukguk, N. (2015). Perbandingan Metode Klasifikasi Regresi Logistik Biner dan Metode Naive Bayes Pada Status Penggunaan KB di Kota Tegal Tahun 2014. *Jurnal Gaussian*, 336-373.

Setyawan, Y. (2013). *Statistika Dasar*. Yogyakarta: Institut Sains & Teknologi AKPRIND Yogyakarta.

Siahaan, D. (2016). Aplikasi CART Dalam Bidang Pendidikan Studi Kasus : Predikat Kelulusan Mahasiswa S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman. *Jurnal Eksponensial*, 1-10.

Silaban, S. K. (2019). *Penerapan Metode CART Dalam Klasifikasi Status HIV di Rumah Sakit Tiom.* Yogyakarta: Institut Sains & Teknologi AKPRIND Yogyakarta.

Wisna, A. (2018). *Perbandingan Model Logit Biner dan Model Logit Probit Biner Pada Faktor-Faktor Yang Mempengaruhi Wanita Bekerja di Kota Yogyakarta.* Yogyakarta: Institut Sains & Teknologi AKPRIND Yogyakarta.