

PENGGUNAAN REGRESI LOGISTIK BINER DAN *ITERATIVE DICHOTOMISER 3* (ID3) DALAM PEMBUATAN KLASIFIKASI STATUS KERJA (Studi Kasus Penduduk Kota Surakarta Tahun 2015)

Lugas Putranti Winastiti¹, Agus Rusgiyono², Diah Safitri³

¹Mahasiswa Departemen Statistika FSM Universitas Diponegoro

^{2,3}Staff Pengajar Departemen Statistika FSM Universitas Diponegoro

ABSTRACT

Discussing about the macro economy usually discuss about unemployment. Unemployment basically can not be fully eliminated. Unemployment usually symbolized with an employment status of person. In this research, two methods were used in making the classification of employment status in the population of the city of Surakarta in February 2015, the methods are binary logistic regression and *Iterative Dichotomiser 3* (ID3) Algorithm. Predictor variables used in determining employment status were age, gender, status in the household, marital status, education and work training. Comparison of the training data and testing data is 60:40. Based on calculations obtained binary logistic regression variables that significantly affect the employment status are age, gender and marital status and the accuracy using testing data is 75%, while the calculations of a decision tree using *iterative dichotomiser 3* algorithm the accuracy using testing data is 75%.

Keywords: Classification, *Iterative Dichotomiser 3* Algorithm, Binary Logistic Regression

1. PENDAHULUAN

1.1 Latar Belakang

Ilmu ekonomi mempunyai dimensi yang cukup luas. Oleh karena itu, secara garis besar ilmu ekonomi dibagi menjadi dua bagian, yaitu ekonomi makro dan ekonomi mikro. Ekonomi makro memberikan penjelasan terhadap permasalahan atau pertanyaan bagaimana faktor-faktor yang mempengaruhi pendapatan nasional, pengangguran, inflasi dan neraca pembayaran (Darwanto, 2011). Pengangguran pada dasarnya tidak bisa dihilangkan sepenuhnya, karena bagaimanapun baik dan hebatnya kemampuan suatu bangsa dalam menangani perekonomiannya, tetap saja pengangguran itu ada. Pengangguran selalu menjadi masalah, bukan saja karena pengangguran berarti pemborosan dana, akan tetapi juga memberikan dampak sosial yang tidak baik misalkan akan semakin meningkatnya tindak kriminal dan pelanggaran moral (Putong, 2013).

Metode regresi logistik biner dan *Iterative Dichotomiser 3* (ID3) merupakan metode yang dapat diterapkan untuk melakukan klasifikasi. Data yang akan diklasifikasikan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Tengah dari hasil survei angkatan kerja nasional (SAKERNAS) kota Surakarta Bulan Februari Tahun 2015.

1.2 Tujuan

Tujuan dilakukannya penelitian dari tugas akhir ini adalah:

1. Memperoleh hasil klasifikasi status kerja di Kota Surakarta tahun 2015 menggunakan metode Regresi Logistik Biner
2. Memperoleh hasil klasifikasi status kerja di Kota Surakarta tahun 2015 menggunakan metode *Iterative Dichotomiser 3* (ID3)

2. TINJAUAN PUSTAKA

2.1 Model regresi Logistik

Menurut Hosmer dan Lemeshow (2000), metode regresi yang membahas tentang hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Kasus yang sering digunakan adalah variabel respon diskrit, dengan dua atau lebih kategori.

2.1.1 Model regresi Logistik Biner

Menurut Hosmer dan Lemeshow (2000), model regresi logistik multivariabel dapat dituliskan:

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} + \beta_p x_p$$

dimana j merupakan banyaknya variabel prediktor x_j yang mempunyai k_j kategori dan $l=1, 2, \dots, k_j-1$.

Dalam hal ini model regresi logistiknya adalah

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1+e^{g(\mathbf{x})}}$$

2.1.2 Penaksiran Parameter Model regresi Logistik Biner

Menurut Agresti (2002), metode Newton Raphson dapat digunakan untuk mendapatkan estimasi parameter dengan langkah-langkah sebagai berikut:

1. Menentukan nilai taksiran awal $\hat{\boldsymbol{\beta}}^{(0)} = \mathbf{0}$.
2. Membentuk vektor $\mathbf{u}' = \left(\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1}, \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_2}, \dots \right)$.
3. Membentuk matriks Hessian \mathbf{H} , dengan anggota $h_{ab} = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_a \partial \beta_b}$.
4. Memasukan nilai $\hat{\boldsymbol{\beta}}^{(0)}$ ke dalam elemen-elemen vektor \mathbf{u} dan matriks \mathbf{H} sehingga diperoleh $\mathbf{u}^{(t)}(\boldsymbol{\beta}^{(0)})$ dan $\mathbf{H}^{(t)}(\boldsymbol{\beta}^{(0)})$
5. Melakukan proses iterasi dimulai dari $t=0$ pada persamaan
$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\mathbf{H}^{(t)} \boldsymbol{\beta}^{(t)})^{-1} \mathbf{u}^{(t)} \boldsymbol{\beta}^{(t)}$$
.
6. Iterasi akan konvergen jika $|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}| \leq \epsilon$, dimana ϵ bilangan positif yang sangat kecil.

2.1.3 Uji Rasio Likelihood (Uji Keseluruhan)

Hosmer dan Lemeshow (2000) mengatakan bahwa, uji rasio likelihood merupakan uji signifikansi parameter secara keseluruhan.

Hipotesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{paling sedikit salah satu dari } \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

Statistik uji:

$$G = -2 \ln \left(\frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}} \right)$$

Kriteria uji: H_0 ditolak jika nilai $G > \chi^2_{(\alpha, p)}$

2.1.4 Uji Wald (Uji Parameter Secara Individu)

Menurut Hosmer dan Lemeshow (2000), uji Wald merupakan uji signifikansi parameter untuk masing-masing variabel prediktor.

Hipotesis :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

$$\text{Statistik uji } W_j^2 = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2$$

Kriteria uji: H_0 ditolak jika nilai $W_j^2 > \chi_{(\alpha,1)}^2$

2.1.5 Uji Goodness of Fit

Hosmer dan Lemeshow (2000) mengatakan bahwa uji kesesuaian model merupakan uji yang dilakukan untuk mengetahui apakah ada perbedaan antara prediksi dan hasil observasi (model sesuai atau tidak).

Hipotesis :

H_0 : Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

H_1 : Model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi)

Statistik uji:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{(n'_k \bar{\pi}_k)(1 - \bar{\pi}_k)}$$

Dimana

g = banyaknya grup

n'_k = banyaknya subjek pada grup ke-k

$o_k = \sum_{j=1}^{c_k} y_j$, jumlah nilai variabel tak bebas pada grup ke-k

$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$, rata-rata taksiran probabilitas

m_j = banyaknya subjek pada c_k kombinasi variabel prediktor

Kriteria uji: H_0 ditolak jika nilai $\hat{C} > \chi_{(\alpha:g-2)}^2$

2.2 Algoritma ID3

Menurut Prasetyo (2014), algoritma ID3 (*iterative dichotomiser 3*) pertama kali diperkenalkan oleh Quinlan pada tahun 1986 yang digunakan untuk menginduksi pohon keputusan. Algoritma ID3 dapat bekerja dengan baik pada semua variabel yang mempunyai tipe data kategorikal (nominal atau ordinal) (Prasetyo, 2014). Entropy digunakan untuk menentukan yang manakah simpul yang akan menjadi pemecah data *training* berikutnya. Menurut Suyanto (2007), entropy dapat dirumuskan sebagai berikut:

$$\text{Entropy}(S) = \sum_i^c -p_i \log_2 p_i$$

Dimana c adalah jumlah kelas klasifikasi, p_i adalah proporsi untuk kelas i

Menurut Suyanto (2007), gain sebuah variabel A dihitung menggunakan persamaan berikut:

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Dimana :

A adalah variabel

v menyatakan suatu nilai yang mungkin untuk variabel A

Values(A) adalah himpunan nilai-nilai yang mungkin untuk variabel A

$|S_v|$ adalah ukuran sampel untuk nilai v

$|S|$ adalah ukuran seluruh sampel data

$\text{Entropy}(S_v)$ adalah *entropy* untuk sampel-sampel yang memiliki nilai

2.3 Ketepatan Klasifikasi

Menurut Johnson dan Wichern (2007), *Apparent Error Rate* (APER) dapat dihitung dengan mudah dari matriks konfusi, yang mana menunjukkan hasil sebenarnya dan hasil prediksi anggota kelompok. Matriks konfusi dapat dibentuk sebagai berikut.

Tabel 1. Matriks Konfusi

f_{ij}		Kelas Hasil Prediksi	
		Kelompok 1	Kelompok 2
Kelas asli	Kelompok 1	n_{11}	n_{12}
	Kelompok 2	n_{21}	n_{22}

$$APER = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

Berdasarkan perhitungan APER, dapat diketahui nilai akurasinya yaitu dengan $1 - APER$.

2.4 Ketenagakerjaan

Tingkat partisipasi angkatan kerja dipengaruhi oleh faktor demografis, sosial dan ekonomis. Faktor-faktor ini antara lain adalah umur, status perkawinan, tingkat pendidikan, daerah tempat tinggal (daerah kota/pedesaan), pendapatan, dan agama. Pengaruh dari masing-masing faktor ini terhadap tingkat pertumbuhan angkatan kerja berbeda antara penduduk laki-laki dan perempuan (Manning, 1984).

3. METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari hasil survei angkatan kerja nasional (Sakernas) kota Surakarta oleh Badan Pusat Statistik pada bulan Februari 2015.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah status kerja (bekerja dan tidak bekerja), jenis kelamin (laki-laki dan perempuan), umur (15-19 tahun, 20-54 tahun dan 55-64 tahun), status dalam rumah tangga (kepala keluarga dan bukan kepala keluarga), status perkawinan (belum menikah, menikah dan cerai), status pendidikan (dasar, menengah dan tinggi), dan pelatihan kerja (pernah dan tidak pernah).

3.3 Langkah-langkah Analisis Data

Metode Regresi Logistik Biner:

1. Melakukan input data
2. Membagi data menjadi data *training* dan *testing*
3. Menghitung estimasi parameter untuk mendapatkan model awal
4. Melakukan uji likelihood untuk mengetahui apakah variabel prediktor secara bersama-sama memengaruhi variabel respon atau tidak dan uji wald untuk mengetahui ada tidaknya pengaruh masing-masing variabel prediktor
5. Melakukan uji kesesuaian model untuk mengetahui sesuai atau tidaknya model
6. Menentukan model akhir dan menginterpretasikan model
7. Menguji model menggunakan data *testing* dan menghitung ketepatan klasifikasi

Metode *Iterative Dichotomiser 3* (ID3):

1. Melakukan input data *training* dengan variabel independen yang signifikan pada metode regresi logistik biner
2. Mengkonstruksikan pohon keputusan algoritma ID3 dengan menghitung nilai *entropy* dan nilai perolehan informasinya dari masing-masing variabel
3. Melakukan analisis terhadap hasil pohon keputusan yang terbentuk
4. Melakukan identifikasi data status kerja
5. Menghitung hasil ketepatan klasifikasi menggunakan data *testing*

4. HASIL DAN PEMBAHASAN

4.1 Pengklasifikasian Regresi Logistik Biner

Berdasarkan estimasi parameter tersebut, diperoleh model sebagai berikut:

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}}$$

dengan nilai $g(\mathbf{x}) = -0,958 + 1,732X_{1(1)} - 1,030X_{1(2)} - 1,325X_{2(1)} - 0,383X_{3(1)} + 1,026X_{4(1)} + 0,937X_{4(2)} - 0,111X_{5(1)} + 0,526X_{5(2)} - 0,096X_{6(1)}$

4.1.1 Uji Rasio Likelihood

Hipotesis

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

H_1 : paling sedikit salah satu dari $\beta_j \neq 0$ dengan $j = 1, 2, \dots, 5$

Taraf Signifikansi: $\alpha = 5\%$

Statistik uji

$$G = -2 \ln \left(\frac{\text{likelihood tanpa variable bebas}}{\text{likelihood dengan varibale bebas}} \right) = 133,005$$

Kriteria uji: H_0 ditolak jika nilai $G > \chi^2_{(\alpha, p)}$

Keputusan

$$H_0 \text{ ditolak karena nilai } G = 133,005 > \chi^2_{(0,05,5)} = 11,071$$

Kesimpulan

Jadi pada taraf signifikasni 5%, karena H_0 ditolak, maka dapat disimpulkan bahwa secara bersama-sama variabel prediktor mempengaruhi model

4.1.2 Uji Wald

Hipotesis :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

dengan $j = 1, 2, \dots, 5$

Taraf Signifikansi: $\alpha = 5\%$

$$\text{Statistik uji: } W_j^2 = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2$$

Kriteria uji: H_0 ditolak jika nilai $W_j^2 > \chi^2_{(\alpha, 1)}$

Tabel 2. Nilai Wald II untuk Setiap Parameter pada Model Awal

Variabel Prediktor (X_j)	Wald (W_j^2)	$\chi^2_{(0,05;1)}$	Keputusan
umur ($X_{1(1)}$)	17,586	3,841	H_0 ditolak
umur ($X_{1(2)}$)	16,641	3,841	H_0 ditolak
jenis kelamin ($X_{2(1)}$)	43,709	3,841	H_0 ditolak
Status perkawinan ($X_{4(1)}$)	7,511	3,841	H_0 ditolak
Status perkawinan ($X_{4(2)}$)	9,656	3,841	H_0 ditolak

Kesimpulan

Jadi berdasarkan Tabel 2, pada taraf signifikansi $\alpha = 5\%$, karena H_0 ditolak pada variabel umur, jenis kelamin dan status dalam rumah tangga, maka ketiga variabel tersebut yang signifikan terhadap model dan ketiga variabel tersebut yang akan diuji selanjutnya

4.1.3 Uji Goodness of Fit

Hipotesis

H_0 : Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

H_1 : Model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi)

Taraf Signifikansi: $\alpha = 5\%$

Statistik uji

$$\hat{C} = \sum_{b=1}^g \frac{(o_k - n_k \bar{\pi}_k)^2}{(n_k \bar{\pi}_k)(1 - \bar{\pi}_k)} = 4,028$$

Kriteria uji: H_0 ditolak jika nilai $\hat{C} > \chi^2_{(\alpha; g-2)}$

Keputusan

Karena $\hat{C} = 4,028 < \chi^2_{(0.05; 6)} = 12,592$ maka H_0 diterima

Kesimpulan

Jadi pada taraf signifikansi $\alpha = 5\%$ karena H_0 diterima maka, jadi model sesuai

4.1.4 Model Akhir

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

dengan nilai $g(x) = -0,972 + 2,086X_{1(1)} - 0,929X_{1(2)} - 1,499X_{2(1)} + 1,321X_{4(1)} + 1,17X_{4(2)}$

Selanjutnya model akhir ini yang akan diterapkan pada data *testing* untuk menghitung ketepatan klasifikasi

4.1.5 Ketepatan Klasifikasi

Tabel 3. Hasil klasifikasi dengan data testing

Kelas asli	Kelas Hasil Prediksi	
	Bekerja	Tidak Bekerja
Bekerja	256	21
Tidak Bekerja	79	44

$$\text{APER} = \frac{79+21}{400} \times 100\% = 25\%$$

$$\text{Akurasi} = \frac{256+44}{400} \times 100\% = 75\%$$

4.2 Pengklasifikasian Algoritma ID3

4.2.1 Pembentukan Pohon keputusan Algoritma ID3

Berikut ini adalah perhitungan mencari nilai *entropy* dan perolehan informasi pada simpul akar menggunakan data *training* dengan Algoritma ID3 untuk mengkonstruksikan pohon keputusan. Perhitungannya adalah sebagai berikut :

1. Menghitung proporsi masing-masing kelas

Tabel 4. Proporsi Masing-masing Kelas

Kelas	Jumlah	Proporsi
Bekerja	416	0,6922
Tidak Bekerja	185	0,3079
Total	601	1,00

2. Menghitung nilai entropy kelas yang disimbolkan dengan E(S)
 $E(\text{Status Kerja}) = \sum_i^c -p_i \log_2 p_i = 0,8906$
3. Menghitung frekuensi masing-masing kategori pada variabel umur berdasarkan kelasnya.

Tabel 5. Frekuensi Masing-masing Kategori pada Variabel umur Berdasarkan Kelasnya

Umur	Frekuensi		Total
	Bekerja	Tidak Bekerja	
15-19 Tahun	12	47	59
20-54 Tahun	294	80	374
55-64 Tahun	110	58	168

Tabel 6. Frekuensi Masing-masing Kategori pada Variabel jenis kelamin Berdasarkan Kelasnya

Jenis Kelamin	Frekuensi		Total
	Bekerja	Tidak Bekerja	
Laki-Laki	224	58	282
Perempuan	192	127	319

Tabel 7. Frekuensi Masing-masing Kategori pada Variabel status perkawinan Berdasarkan Kelasnya.

Status Perkawinan	Frekuensi		Total
	Bekerja	Tidak Bekerja	
Belum Menikah	58	63	121
Menikah	306	111	417
Cerai	52	11	63

4. Hitung proporsi masing-masing kategori berdasarkan kelasnya :

Tabel 8. Proporsi Kategori pada Variabel Umur

Umur	Bekerja	Tidak Bekerja
15-19 Tahun	0,2034	0,7966
20-54 Tahun	0,7861	0,2139
55-64 Tahun	0,6548	0,3452

Tabel 9. Proporsi Kategori pada Variabel Jenis Kelamin

Jenis Kelamin	Bekerja	Tidak Bekerja
Laki-Laki	0,7943	0,2057
Perempuan	0,6019	0,3981

Tabel 10. Proporsi Kategori pada Variabel Status Perkawinan

Status Perkawinan	Bekerja	Tidak Bekerja
Belum Menikah	0,4793	0,5207
Menikah	0,7338	0,2662
Cerai	0,8254	0,1746

5. Menghitung nilai *entropy* pada variabel

$$E(15-19 \text{ Tahun}) = -\left(\frac{12}{59}\right) \times \log_2\left(\frac{12}{59}\right) - \left(\frac{47}{59}\right) \times \log_2\left(\frac{47}{59}\right) = 0,7287$$

$$E(20-54 \text{ Tahun}) = -\left(\frac{294}{374}\right) \times \log_2\left(\frac{294}{374}\right) - \left(\frac{80}{374}\right) \times \log_2\left(\frac{80}{374}\right) = 0,7489$$

$$E(55-64 \text{ Tahun}) = -\left(\frac{110}{168}\right) \times \log_2\left(\frac{110}{168}\right) - \left(\frac{58}{168}\right) \times \log_2\left(\frac{58}{168}\right) = 0,9297$$

$$E(\text{Laki-Laki}) = -\left(\frac{224}{282}\right) \times \log_2\left(\frac{224}{282}\right) - \left(\frac{58}{282}\right) \times \log_2\left(\frac{58}{282}\right) = 0,7331$$

$$E(\text{Perempuan}) = -\left(\frac{192}{319}\right) \times \log_2\left(\frac{192}{319}\right) - \left(\frac{127}{319}\right) \times \log_2\left(\frac{127}{319}\right) = 0,96983$$

$$E(\text{Belum Menikah}) = -\left(\frac{58}{121}\right) \times \log_2\left(\frac{58}{121}\right) - \left(\frac{63}{121}\right) \times \log_2\left(\frac{63}{121}\right) = 0,9988$$

$$E(\text{Menikah}) = -\left(\frac{306}{417}\right) \times \log_2\left(\frac{306}{417}\right) - \left(\frac{111}{417}\right) \times \log_2\left(\frac{111}{417}\right) = 0,8359$$

$$E(\text{Cerai}) = -\left(\frac{52}{63}\right) \times \log_2\left(\frac{52}{63}\right) - \left(\frac{11}{63}\right) \times \log_2\left(\frac{11}{63}\right) = 0,6681$$

6. Menghitung nilai perolehan informasi

$$\text{Gain (status kerja, Umur)} = 0,8906 - \left(\frac{59}{601}\right) 0,7287 - \left(\frac{374}{601}\right) 0,7489 - \left(\frac{168}{601}\right) 0,9297 = 0,0932$$

$$\text{Gain (Status Kerja, Jenis Kelamin)} = 0,8906 - \left(\frac{282}{601}\right) 0,7331 - \left(\frac{319}{601}\right) 0,9698 = 0,03187$$

$$\text{Gain (Status Kerja, Status Perkawinan)} = 0,8906 - \left(\frac{121}{601}\right) 0,9988 - \left(\frac{417}{601}\right) 0,8359 = 0,0395$$

Berikut ini adalah hasil perhitungan mencari nilai *entropy* dan perolehan informasi dari semua variabel untuk menentukan simpul akar.

Tabel 11. Nilai Perolehan Informasi pada Simpul Akar

Variabel	Gain
Umur	0,0932
Jenis kelamin	0,0319
Status Perkawinan	0,0395

Berdasarkan Tabel 11, dapat diketahui bahwa variabel umur adalah variabel dengan nilai perolehan informasi terbesar dengan nilai 0,0932, maka variabel umur merupakan pengklasifikasi terbaik dan diletakkan sebagai pemilah. Selanjutnya pembentukan cabang untuk mengkonstruksikan simpul anak berdasarkan informasi pada simpul akar yang sekaligus sebagai simpul induk. Informasi yang dapat diperoleh dari simpul akar adalah :

1. Umur dijadikan sebagai simpul akar karena mempunyai nilai gain terbesar.
2. Sub-himpunan pada variabel terpilih adalah 15-19 tahun, 20-54 tahun dan 55-64 tahun.

4.2.2 Analisis Pohon Keputusan

Hasil Algoritma ID3 untuk mengidentifikasi status kerja pada penduduk kota surakarta bulan Februari tahun 2015 dengan variabel jenis kelamin, umur, dan status perkawinan. Berikut ini informasi yang dapat diperoleh dari hasil klasifikasi menggunakan Algoritma ID3:

1. Pada penelitian ini terbentuk simpul sebanyak 9 simpul.

2. Simpul daun merupakan kelas yang terbentuk. Pada penelitian ini terbentuk sebanyak 6 simpul daun, artinya terdapat 6 karakteristik status kerja penduduk kota Surakarta tahun 2015.
3. Variabel umur merupakan variabel yang terpilih sebagai simpul akar karena mempunyai nilai perolehan informasi terbesar.

4.2.3 Ketepatan Klasifikasi

Tabel 12. Hasil klasifikasi dengan data testing

Kelas asli	Kelas Hasil Prediksi	
	Bekerja	Tidak Bekerja
Bekerja	256	21
Tidak Bekerja	79	44

Berdasarkan Tabel nilai akurasi Algoritma ID3 pada data *testing* adalah sebagai berikut :

$$APER = \frac{79+21}{400} \times 100\% = 25 \%$$

$$Akurasi = \frac{256+44}{400} \times 100\% = 75 \%$$

5. KESIMPULAN

Kesimpulan yang diperoleh dari penelitian ini adalah:

1. Berdasarkan hasil klasifikasi status kerja dengan menggunakan metode regresi logistik biner didapatkan tiga variabel yang signifikan yaitu umur, jenis kelamin dan status perkawinan. Dalam regresi logistik biner menggunakan data *testing* didapatkan hasil akurasi sebesar 75%.
2. Konstruksi pohon keputusan yang terbentuk menggunakan algoritma ID3 menghasilkan pohon dengan banyak simpul mencapai 9 simpul dimana 6 diantaranya adalah simpul daun dan variabel umur terpilih menjadi simpul akar dalam pembuatan pohon keputusan. Berdasarkan hasil klasifikasi status kerja dengan menggunakan ID3 menggunakan data *testing* didapatkan hasil akurasi sebesar 75%.
3. Dapat diketahui bahwa nilai akurasi hasil klasifikasi dari kedua metode mempunyai nilai yang sama.

DAFTAR PUSTAKA

- [1] Agresti, A. 2002. *Categorical Data Analysis*. New York: John Wiley & Sons
- [2] Darwanto. 2011. *Ekonomi Mikro I*. Semarang: Badan Penerbit Universitas Diponegoro.
- [3] Hosmer, D. W. and Lemeshow S., 2000. *Applied Logistic Regression*. United States of America: Sons Inc.
- [4] Johnson, R. A and Wichern. D. W., 2007. *Applied Multivariate Statistical Analysis. Sixth Edition*. New Jersey: Prentice Hall International, Inc.
- [5] Manning, C. 1984. *Angkatan Kerja dan Kesempatan Kerja di Indonesia Dewasa Ini dalam Zainab Bakir dan Chris Manning (Editor). Angkatan Kerja di Indonesia, Partisipasi, Kesempatan, dan Pengangguran*. Jakarta: CV. Rajawali.
- [6] Prasetyo, E. 2014. *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- [7] Putong, I. 2013. *Economics, Pengantar Mikro dan Makro*. Jakarta: Mitra Wacana Media
- [8] Suyanto. 2007 . *Artificial Intelligence Searching, Reasoning, Planning dan Learning*. Bandung: Penerbit Informatika.