

Penerapan Metode Chi-Squared Automatic Interaction Detection (CHAID) dan Classification And Regression Trees (CART) pada Klasifikasi Status Kerja di Kabupaten Brebes

1¹⁾Khalfia Helena, 2²⁾Yuliana Susanti, 3³⁾Respatiwan

1¹⁾Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret

2²⁾ Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret

3³⁾ Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret

Email: 1¹⁾khalfiahelena14@gmail.com

Email: 2²⁾yulsusan@gmail.com

Email: 3³⁾respatiwan@staff.uns.ac.id

Abstrak – Kabupaten Brebes merupakan salah satu kabupaten di Provinsi Jawa Tengah dengan jumlah penduduk terbanyak kedua. Jumlah penduduk di Kabupaten Brebes selalu meningkat dari tahun ke tahun. Pertumbuhan penduduk yang besar akan berpengaruh terhadap meningkatnya jumlah angkatan kerja. Penduduk yang termasuk dalam angkatan kerja adalah penduduk yang berusia lebih dari 15 tahun yang terbagi menjadi dua yaitu bekerja dan pengangguran. Peningkatan jumlah angkatan kerja berdampak pada status kerja yaitu bekerja atau pengangguran. Untuk mengetahui faktor-faktor yang berpengaruh terhadap status kerja tersebut dilakukan penelitian dengan menggunakan metode klasifikasi dalam bentuk pohon yaitu CHAID dan CART. Tujuan dari penelitian ini adalah untuk mengklasifikasikan faktor-faktor yang berpengaruh terhadap status kerja dalam bentuk pohon klasifikasi menggunakan metode CHAID dan CART serta menghitung nilai akurasi. CHAID merupakan suatu teknik pohon klasifikasi yang menguji satu persatu variabel independen berdasarkan pada tingkat signifikansi statistik chi-square, sedangkan metode CART merupakan teknik pohon klasifikasi yang menggunakan metode statistik nonparametrik dengan variabel dependen berbentuk kategorik maupun kontinu. Variabel yang berpengaruh pada penelitian ini adalah jenis kelamin, umur, status perkawinan, pelatihan kerja, dan pendidikan. Hasil dari penelitian ini berupa pohon klasifikasi status kerja. Metode CHAID menghasilkan 4 segmen klasifikasi dan CART menghasilkan 5 segmen klasifikasi. Nilai akurasi yang didapat dari masing-masing metode adalah sama yaitu sebesar 83.51%.

Kata kunci : Pohon Klasifikasi, CHAID, CART, Nilai Akurasi

PENDAHULUAN

Indonesia merupakan negara kepulauan yang memiliki jumlah penduduk yang besar. Pertumbuhan penduduk yang besar akan menciptakan jumlah angkatan kerja yang besar pula di pasar tenaga kerja. Penduduk yang termasuk dalam angkatan kerja adalah penduduk yang berusia lebih dari 15 tahun yang terbagi menjadi bekerja dan pengangguran [2].

Pada tahun 2017 BPS Jawa Tengah merilis data jumlah penduduk melalui Publikasi Keadaan Angkatan Kerja [3]. Kabupaten Brebes memiliki jumlah penduduk terbanyak kedua di Provinsi Jawa Tengah setelah Kota Semarang yaitu sebesar 1.796.004 orang. Jumlah ini naik sebesar 7124 orang daripada tahun 2016. Seiring dengan peningkatan jumlah penduduk dari tahun ke tahun, maka penduduk usia kerja juga mengalami peningkatan. Hasil Survei Angkatan Kerja Nasional (Sakernas) pada bulan Agustus 2017 jumlah penduduk angkatan kerja Kabupaten Brebes sebesar 895712 orang. Dengan jumlah penduduk yang bekerja adalah sebesar 823.661 orang dan penduduk menganggur sebanyak 72.051 orang. Berdasarkan permasalahan di atas, perlu

dilakukan pengklasifikasian status kerja bahwa seseorang dikatakan pengangguran atau bekerja.

Status kerja dapat diketahui dengan melakukan klasifikasi pada faktor-faktor yang berpengaruh. Klasifikasi merupakan pengelompokan secara sistematis ke dalam kelas tertentu berdasarkan ciri-ciri yang sama [7]. Salah satu metode klasifikasi yang dapat mengklasifikasikan data status kerja dalam bentuk visual adalah pohon klasifikasi.

Pohon klasifikasi adalah metode yang menggunakan aturan untuk menentukan kelas dari suatu objek yang memiliki nilai-nilai variabel independen. Metode pohon klasifikasi yang digunakan dalam penelitian ini adalah CHAID dan CART.

Penelitian mengenai klasifikasi status kerja telah dilakukan oleh beberapa peneliti sebelumnya. Penelitian [11] membandingkan metode klasifikasi *Naïve Bayes* dan *K-Nearest Neighbor* tentang klasifikasi status kerja di Kabupaten Demak. Penelitian [12] membandingkan metode algoritma C4.5 dan *Fuzzy K-Nearest Neighbor In Every Class* (FK-NNC) mengenai klasifikasi status kerja di Kabupaten Tegal.

Penelitian tentang pohon klasifikasi juga telah dilakukan oleh beberapa peneliti. Pada penelitian [1] membandingkan metode CART dan CHAID pada klasifikasi status kerja Kota Semarang dan diperoleh CART memiliki akurasi yang lebih baik. Penelitian [10] menggunakan metode CART dan CHAID pada klasifikasi pembiakan domba karayaka dan diperoleh bahwa CHAID lebih baik daripada CART.

Pada penelitian ini diterapkan metode CHAID dan CART untuk klasifikasi status kerja di Kabupaten Brebes, kemudian menentukan tingkat akurasinya.

METODE

Data yang digunakan dalam penelitian ini adalah data status kerja di Kabupaten Brebes tahun 2017. Pada penelitian ini menggunakan data sekunder yang diperoleh dari Survei Angkatan Kerja Nasional (SAKERNAS) yang dilakukan oleh Badan Pusat Statistik Provinsi Jawa Tengah pada tahun 2017.

Metode CHAID. CHAID merupakan suatu teknik pohon klasifikasi yang menguji satu persatu variabel independen dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya [6]. Metode CHAID memiliki tiga tahap, yaitu penggabungan, pemisahan, dan penghentian [4]. Algoritma CHAID mempunyai langkah-langkah sebagai berikut:

1. Tahap penggabungan. Tahap ini dilakukan pada variabel independen yang memiliki kategori lebih dari dua.
 - (a) Membentuk tabel kontingensi untuk masing-masing variabel independen dengan variabel dependennya.
 - (b) Melakukan uji independensi *chi-square* untuk setiap pasang kategori variabel independen dan dependen dengan membentuk sub tabel kontingensi 2 x J dengan J banyaknya kategori variabel dependen.
 - (c) Menghitung statistik *chi-square* untuk setiap tabel kontingensi 2 x J kategori berpasangan untuk menentukan variabel kategori berpasangan signifikan atau tidak. Hipotesis yang diuji adalah sebagai berikut:
 H_0 : Variabel dependen dan independen saling bebas
 H_1 : Variabel dependen dan independen tidak saling bebas
 Statistik uji yang digunakan adalah sebagai berikut.

$$X^2_{hitung} = \sum_{i=1}^b \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

(1)

Untuk menghitung nilai frekuensi harapan menggunakan rumus sebagai berikut.

$$E_{ij} = \frac{n_i \cdot n_j}{N}, i = 1 \dots b \text{ dan } j = 1 \dots k$$

(2)

dengan

- O_{ij} : nilai observasi baris ke-i kolom ke-j
- E_{ij} : nilai frekuensi harapan baris ke-I kolom ke-j
- n_i : baris ke-i, $i = 1 \dots b$
- n_j : kolom ke-j, $j = 1 \dots k$
- N : total banyaknya pengamatan

Kriteria pengujian hipotesisnya adalah H_0 ditolak jika nilai $X^2_{hitung} > \chi^2_{(1-\alpha);(b-1)(k-1)}$.

- (d) Menggabungkan variabel kategori berpasangan yang tidak signifikan dengan kategori yang lain menjadi pasangan kategori gabungan dan kemudian dilanjutkan ke langkah (e). Jika semua pasangan kategori adalah signifikan maka lanjut ke langkah (f).
- (e) Melakukan uji *chi-square* kembali untuk pasangan kategori yang digabung tetap tidak signifikan maka ulangi langkah (c). Jika pasangan kategori yang digabung signifikan maka dilanjutkan ke langkah selanjutnya
- (f) Mengalikan nilai p dari hasil uji *chi-square* dengan pengali Bonferroni sehingga menghasilkan nilai p terkoreksi Bonferroni. Pengali Bonferroni untuk masing-masing tipe variabel-variabel independen adalah berbeda [6]. Pengali Bonferroni untuk masing-masing jenis variabel-variabel independen adalah sebagai berikut.

- a. Variabel monotonik untuk data berskala ordinal.

$$M = \binom{c-1}{r-1}$$

(3)

- b. Variabel independen untuk data berskala nominal

$$M = \sum_{i=0}^{r-1} (-1)^i \frac{(r-1)^c}{i!(r-i)!}$$

(4)

- M : koreksi Bonferroni
- c : kategori variabel independen awal
- r : kategori variabel independen setelah penggabungan
- i : kategori baru ke-i.

2. Tahap pemisahan. Tahap ini dilakukan dengan memilih variabel independen yang memiliki nilai *chi-square* terbesar dan *p-value* terkoreksi Bonferroni terkecil yang akan digunakan sebagai pemisah simpul terbaik. Jika tidak ada variabel independen dengan nilai *p-value* yang $\leq \alpha$ maka tidak dilakukan pemisahan.
3. Tahap penghentian. Tahap ini dilakukan jika sudah tidak ada hubungan antara variabel independen dengan variabel dependen.

Metode CART. CART adalah metode statistik nonparametrik yang menggambarkan hubungan antara variabel dependen dengan satu atau lebih variabel independen yang dikembangkan untuk topik analisis klasifikasi. Tujuan metode CART yaitu untuk mendapatkan suatu kelompok data yang akurat sebagai pencari suatu pengklasifikasian [5]. Prinsip kerja dari metode CART disebut sebagai *binary recursive partitioning* [9]. Istilah *binary* menyatakan bahwa setiap simpul induk akan dipisah menjadi dua simpul anak. Istilah *recursive* mengacu pada proses pemisahan simpul dilakukan. Istilah *partitioning* mengacu pada data dipisah menjadi bagian-bagian atau partisipansi yang lebih kecil. Langkah-langkah metode CART adalah sebagai berikut:

1. Menyusun calon cabang yang dilakukan terhadap seluruh variabel independen.
2. Menilai kinerja keseluruhan calon cabang dengan jalan menghitung nilai besaran kesesuaian $\emptyset(s/t)$. Kesesuaian $\emptyset(s/t)$ dari calon cabang *s* pada noktah keputusan *t*, didefinisikan sebagai berikut :

$$\emptyset(s/t) = 2P_L P_R Q(s|t) \quad (6)$$

dengan

$$\emptyset(s/t) = \sum_{j=1}^{jumlah\ kategori} |P(j|t_L) - P(j|t_R)| \quad (7)$$

t_L = cabang kiri dan noktah keputusan *t*

t_R = cabang kanan dari noktah keputusan *t*

$$P_L = \frac{\text{calon cabang kiri } t_L}{\text{data latihan}} \quad (8)$$

$$P(j|t_L) = \frac{j \text{ calon cabang kiri } t_L}{\text{noktah keputusan } t} \quad (9)$$

$$P_R = \frac{\text{calon cabang kanan } t_R}{\text{data latihan}} \quad (10)$$

$$P(j|t_R) = \frac{j \text{ calon cabang kiri } t_R}{\text{noktah keputusan } t} \quad (11)$$

3. Menentukan calon cabang yang memiliki nilai kesesuaian $\emptyset(s|t)$ terbesar. Jika masih terdapat noktah keputusan, pelaksanaan algoritma dilanjutkan dengan kembali ke langkah kedua. Setelah noktah keputusan tidak ada lagi maka algoritma CART dihentikan.

Nilai Akurasi. Cara untuk mengevaluasi hasil klasifikasi dapat dilakukan dengan cara menghitung nilai akurasi. *Confusion matrix* merupakan tabel yang digunakan untuk mencatat hasil klasifikasi dan menghitung akurasi [8]. Cara menghitung *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

Kelas Hasil Prediksi		
Kelas Aktual	Kelas Positif	Kelas Negatif
Kelas Positif	<i>True Positive</i>	<i>True Negative</i>
Kelas Negatif	<i>False Positive</i>	<i>False Negative</i>

Nilai akurasi dapat dihitung dengan rumus sebagai berikut.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (12)$$

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data status kerja di Kabupaten Brebes. Jumlah total data sebanyak 473 dengan 395 penduduk berstatus bekerja dan 78 berstatus pengangguran. Faktor yang digunakan dalam penelitian ini adalah jenis kelamin, umur, status perkawinan, pelatihan kerja, dan pendidikan. Rincian kategori dari masing-masing variable ditunjukkan pada Tabel 2.

Tabel 2. Tabulasi Kategori Setiap Variabel

Atribut	Kategori	Jumlah
Status kerja	Bekerja	395
	Pengangguran	78
Jenis Kelamin	Laki-laki	225
	Perempuan	248
Umur	Sangat produktif (15-49 tahun)	306
	Produktif (50-64 tahun)	93
	Tidak produktif (>64 tahun)	74
Perkawinan	Belum kawin	152
	Kawin	229
	Cerai	92

Pelatihan	Pernah	53
	Tidak pernah	420
Pendidikan	Dasar (SD/MI,SMP/MTS)	305
	Menengah (SMA/SMK)	131
	Tinggi (D1, D2, D3, S1, S2 dan S3)	37

Pendidikan (2,3) dan 1 44.288 0.000 H0 ditolak

Analisis Metode CHAID. Berikut ini adalah langkah-langkah dalam pembentukan pohon klasifikasi menggunakan metode CHAID. Langkah pertama yang dilakukan adalah melakukan uji independensi *chi-square* untuk setiap pasang kategori variabel independen dan dependen untuk menentukan *root node*. Hasil dari uji independen *chi-square* dapat dilihat di Tabel 3.

Tabel 3. Hasil perhitungan Uji *Chi-Square*

Variabel	Kategori	χ^2	<i>P-value</i>	Kesimpulan
Jenis Kelamin	1 dan 2	0.049	0.824	H0 diterima
	1 dan 2	31.066	0.000	H0 ditolak
Umur	1 dan 3	72.053	0.000	H0 ditolak
	2 dan 3	6.555	0.01	H0 ditolak
	1 dan 2	47.486	0.000	H0 ditolak
Perkawinan	1 dan 3	0.050	0.823	H0 diterima
	2 dan 3	31.693	0.000	H0 ditolak
Pelatihan	1 dan 2	0.277	0.599	H0 diterima
	1 dan 2	34.134	0.000	H0 ditolak
Pendidikan	1 dan 3	11.854	0.001	H0 ditolak
	2 dan 3	0.572	0.45	H0 diterima

Berdasarkan Tabel 3, variabel jenis kelamin dan pelatihan dikeluarkan dari model karena tidak saling bebas. Variabel perkawinan kategori 1 dan 3 serta variabel pendidikan kategori 2 dan 3 tidak signifikan. Variabel perkawinan dan pendidikan memiliki kategori lebih dari 2, sehingga dilakukan langkah selanjutnya yaitu tahap penggabungan. Variabel perkawinan kategori 1 dan 3 digabung dengan kategori 2, sedangkan variabel pendidikan kategori 2 dan 3 digabung dengan kategori 1, kemudian dihitung kembali uji *chi-square*nya lalu *p-value* hasil penggabungan dikali dengan pengali Bonferroni. Hasil dari penggabungan terdapat pada Tabel 4

Tabel 4. Hasil Penggabungan

Variabel	Kategori	χ^2	<i>p-value</i>	Kesimpulan
Perkawinan	(1,3) dan 2	78.397	0.000	H0 ditolak

Dari tabel di atas diperoleh variabel perkawinan memiliki nilai *chi-square* terbesar yaitu 78.397. Variabel perkawinan merupakan *root node* karena memiliki nilai *chi-square* terbesar dibandingkan variabel lainnya. Proses pemilahan terus dilakukan pada setiap simpul selama masih terdapat variabel-variabel independen yang signifikan.

- (a) Iterasi 1. *Node* perkawinan kategori menengah tinggi diklasifikasikan ke dalam grup bekerja
- (b) Iterasi 2. *Node* perkawinan kategori kawin, didapatkan nilai *chi-square* terbesar yaitu sebesar 94.902 adalah variabel pendidikan. Kategori menengah tinggi diklasifikasikan ke dalam grup bekerja, sedangkan kategori dasar dilakukan perhitungan lebih lanjut.
- (c) Iterasi 3. *Node* pendidikan kategori dasar didapatkan nilai *chi-square* terbesar yaitu sebesar 31.560 adalah variabel umur. Kategori sangat produktif, tidak produktif dan produktif diklasifikasikan ke dalam grup pengangguran.

Pohon klasifikasi yang terbentuk dapat digunakan untuk mengklasifikasikan status kerja sebagai berikut:

Klasifikasi 1: orang yang status perkawinannya belum kawin atau cerai diklasifikasikan ke dalam bekerja.

Klasifikasi 2: orang yang status perkawinannya kawin dengan pendidikan menengah atau tinggi diklasifikasikan ke dalam bekerja.

Klasifikasi 3: orang yang status perkawinannya kawin dengan pendidikan dasar serta umur produktif atau tidak produktif diklasifikasikan ke dalam pengangguran.

Klasifikasi 4: orang status perkawinannya kawin dengan pendidikan dasar serta umur sangat produktif diklasifikasikan ke dalam pengangguran.

Pohon klasifikasi yang telah dihasilkan diuji tingkat ketepatan atau akurasi dalam mengklasifikasi data status kerja. Untuk menguji nilai akurasi dapat dilakukan dengan menggunakan *confusion matrix* seperti pada Tabel 5.

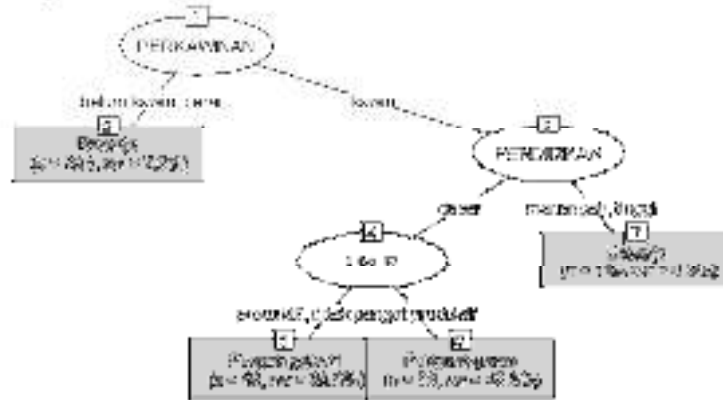
Tabel 5. *Confusion matrix* CHAID

Kelas Aktual	Kelas Hasil Prediksi	
	Bekerja	Pengangguran
Bekerja	350	45
Pengangguran	7	71

untuk memprediksi status kerja secara tepat yaitu 83.51%.

$$\text{Akurasi} = \frac{350+45}{350+45+7+71} \times 100 = 83.51\%$$

Berdasarkan Tabel 5, menunjukan bahwa dari 473 data secara keseluruhan terdapa 421 data dengan klasifikasi benar, sehingga diperoleh tingkat akurasi



Gambar 1. Pohon klasifikasi metode CHAID

Analisis Metode CART. Berikut ini adalah langkah-langkah dalam pembentukan pohon klasifikasi menggunakan metode CART. Langkah pertama yang dilakukan adalah membentuk calon cabang (*candidate split*) terhadap seluruh variabel independen sehingga terbentuk calon cabang mutakhir seperti pada Tabel 6 di bawah ini.

Tabel 6. Penyusunan calon cabang

No	Calon cabang kiri	Calon cabang kanan
1	Laki-laki	Perempuan
2	Perempuan	Laki-laki
3	Sangat produktif	Produktif, tidak produktif
4	Produktif	Sangat produktif, tidak produktif
5	Tidak produktif	Sangat produktif, produktif
6	Belum kawin	Kawin, cerai
7	Kawin	Belum kawin, cerai
8	Cerai	Kawin, belum kawin
9	Pernah	Tidak pernah
10	Tidak pernah	Pernah
11	Dasar	Menengah, tinggi
12	Menengah	Dasar, tinggi
13	Tinggi	Dasar, menengah

Langkah kedua yang dilakukan yaitu menghitung nilai *candidate split purity left* P_L dan *purity right* P_R . Hasil perhitungan P_L dan P_R terdapat dalam Tabel 7. Tabel Perhitungan P_L dan P_R

No	P_L	P_R
1	0.4757	0.5243
2	0.5243	0.4757
3	0.6469	0.3531
4	0.1966	0.8034

5	0.1564	0.8436
6	0.3214	0.6786
7	0.4841	0.5159
8	0.1945	0.8055
9	0.1121	0.8879
10	0.8879	0.1121
11	0.6448	0.3552
12	0.277	0.723
13	0.0782	0.9218

Selanjutnya menghitung $P(j|t_L)$ dan $P(j|t_R)$ untuk kemungkinan bekerja dan pengangguran seperti pada Tabel 8 berikut:

Tabel 8. Tabel Perhitungan $P(j|t_L)$ dan $P(j|t_R)$

No	Status	$P(j t_L)$	$P(j t_R)$
1	Pengangguran	38/225	40/248
	Bekerja	187/225	208/248
2	Pengangguran	40/248	38/225
	Bekerja	208/248	187/225
3	Pengangguran	19/306	59/167
	Bekerja	287/306	108/167
4	Pengangguran	25/93	53/380
	Bekerja	69/93	327/380
5	Pengangguran	34/74	44/399
	Bekerja	40/74	355/399
6	Pengangguran	4/152	74/321
	Bekerja	148/152	247/321
7	Pengangguran	72/229	6/244

	Bekerja	157/229	238/244
8	Pengangguran	2/92	76/381
	Bekerja	90/92	305/381
9	Pengangguran	1/53	77/420
	Bekerja	52/53	343/420
10	Pengangguran	77/420	1/53
	Bekerja	343/420	52/53
11	Pengangguran	76/305	2/168
	Bekerja	229/305	166/168
12	Pengangguran	2/131	76/342
	Bekerja	129/131	266/342
13	Pengangguran	0/37	78/436
	Bekerja	37/37	358/436

Dari perhitungan di atas kemudian menghitung nilai kesesuaian. Nilai kesesuaian didapat dari perkalian antara $2P_L P_R$ dan

$Q(s|t)$. Perhitungan iterasi pertama dapat dilihat pada Tabel 9.

Tabel 9. Nilai kesesuaian untuk calon cabang

No	$Q(s t)$	$2P_L P_R$	$\phi(s t)$
1	0.4988	0.0152	0.0076
2	0.4988	0.0152	0.0076
3	0.4568	0.5824	0.2661
4	0.3159	0.2587	0.0817
5	0.2639	0.6984	0.1843
6	0.4362	0.4084	0.1781
7	0.4995	0.5796	0.2895
8	0.3133	0.3555	0.1114
9	0.199	0.3289	0.0655
10	0.199	0.3289	0.0655
11	0.4581	0.4746	0.2174
12	0.4005	0.4139	0.1658
13	0.1442	0.3578	0.0516

Berdasarkan perhitungan tabel di atas diperoleh nilai kesesuaian terbesar adalah sebesar 0.2895 yaitu variabel perkawinan dengan cabang kiri kawin dan cabang kanan belum kawin dan cerai. Untuk menentukan cabang selanjutnya dilakukan perhitungan kembali seperti diatas tanpa melibatkan variabel yang telah terpilih.

(a) Iterasi 1. *Node* perkawinan kategori kawin didapatkan nilai kesesuaian terbesar adalah variabel pendidikan kategori dasar yaitu sebesar 0.5976. Kategori dasar sebagai cabang kiri mengklasifikasikan ke dalam grup pengangguran, sedangkan kategori menengah dan tinggi sebagai cabang kanan mengklasifikasikan ke dalam grup bekerja.

(b) Iterasi 2. *Node* pendidikan kategori dasar didapatkan nilai kesesuaian terbesar adalah variabel umur kategori sangat produktif yaitu sebesar 0.5184. Kategori sangat produktif sebagai cabang kiri mengklasifikasikan ke dalam grup pengangguran, sedangkan kategori produktif dan tidak produktif sebagai cabang kanan juga mengklasifikasikan ke dalam grup pengangguran.

(c) Iterasi 3. *Node* umur kategori sangat produktif didapatkan nilai kesesuaian terbesar adalah variabel jenis kelamin kategori laki-laki yaitu sebesar 0.9405. Kategori laki-laki sebagai cabang kiri mengklasifikasikan ke dalam grup bekerja, sedangkan kategori perempuan sebagai cabang kanan mengklasifikasikan ke dalam grup pengangguran. Berikut ini gambar pohon klasifikasi status kerja menggunakan metode CART dapat dilihat di Gambar 2.

Pohon klasifikasi yang terbentuk dapat digunakan untuk mengklasifikasikan status kerja sebagai berikut:

Klasifikasi 1: Orang yang status perkawinannya belum kawin atau cerai diklasifikasikan ke dalam bekerja.

Klasifikasi 2: Orang yang status perkawinannya kawin dan pendidikan menengah atau tinggi diklasifikasikan ke dalam bekerja.

Klasifikasi 3: Orang yang status perkawinannya kawin, pendidikan dasar, dan usia produktif atau tidak produktif diklasifikasikan ke dalam pengangguran.

Klasifikasi 4: Orang yang status perkawinannya kawin, pendidikan dasar, dan usia sangat produktif, jenis kelamin laki-laki diklasifikasikan ke dalam bekerja.

Klasifikasi 5: Orang yang status perkawinannya kawin, pendidikan dasar, dan usia sangat produktif, jenis kelamin perempuan diklasifikasikan ke dalam pengangguran.

Pohon klasifikasi yang telah dihasilkan diuji tingkat ketepatan atau akurasi dalam mengklasifikasi data status kerja. Untuk menguji nilai akurasi dapat dilakukan dengan menggunakan *confusion matrix* seperti pada Tabel 10.

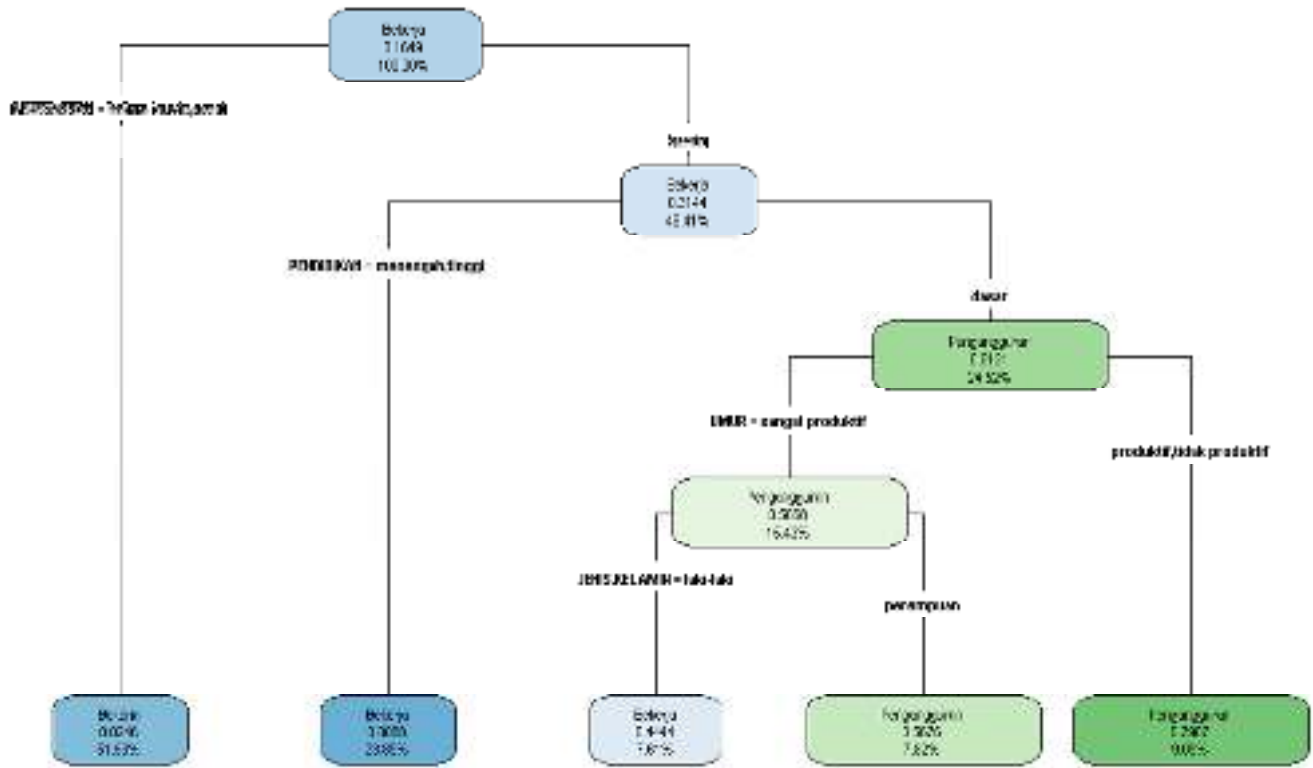
Tabel 10. *Confusion matrix* CART

Kelas Aktual	Kelas Hasil Prediksi	
	Bekerja	Pengangguran
Bekerja	370	25
Pengangguran	23	55

$$\text{Akurasi} = \frac{370+25}{370+25+23+55} \times 100\% = 83.51\%$$

benar, sehingga diperoleh tingkat akurasi untuk memprediksi status kerja secara tepat yaitu 83.51%.

Dari Tabel 10, menunjukan bahwa dari 473 data secara keseluruhan terdapa 425 data dengan klasifikasi



Gambar 2 . Pohon Klasifikasi Metode CAR

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan bahwa penerapan metode CHAID pada klasifikasi status kerja dengan faktor yang mempengaruhi antara lain umur, status perkawinan, dan pendidikan memiliki struktur pohon klasifikasi dengan 4 tingkatan. Hasil pengujian akurasi yang diperoleh sebesar 83.51%.

Penerapan metode CART pada klasifikasi status kerja dengan faktor yang mempengaruhi adalah jenis kelamin, umur, status perkawinan, dan pendidikan memiliki struktur pohon klasifikasi dengan 5 tingkatan. Hasil pengujian akurasi yang diperoleh sebesar 83.51%.

SARAN

Penelitian ini dapat dikembangkan lebih lanjut dengan cara menambahkan variabel lain untuk mendapatkan nilai akurasi yang lebih baik .

DAFTAR PUSTAKA

- [1] Aritonang, N.E., Rusgiyono, A., dan Rahmawati, R. 2016. Klasifikasi Status Kerja Pada Angkatan Kerja Kota Semarang Tahun 2014 Menggunakan Metode CHAID dan CART. *Jurnal Gaussian* Vol 5 No.1 183 - 192.
- [2] Badan Pusat Statistik (BPS). 2014. *Indikator Kesejahteraan Rakyat Welfare Indicator 2014*. Badan Pusat Statistik Provinsi Jawa Tengah, Semarang.
- [3] Badan Pusat Statistik (BPS). 2017. *Publikasi Keadaan Angkatan Kerja Provinsi Jawa Tengah Agustus 2017*. Badan Pusat Statistik Provinsi Jawa Tengah, Semarang.
- [4] Bagozzi, R. P. 1994. *Advanced Methods of Marketing Research*. Blackwell Publishers Ltd, Oxford.
- [5] Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1993. *Classification And Regression Trees*, Chapman & Hall, New York.
- [6] Gallagher, C. A. 2000. An Iterative Approach to Identification Detection and Classification Analysis. *Journal of Applied Statistics*, Vol. 1, No. 4, 1-5.
- [7] Hamandoko dan Tairas, J. 1999. *Pengantar Klasifikasi Persepuluhan Dewey*. BPK Gunung Mulia, Jakarta.
- [8] Han, Jiwei and Michelin Kamber. 2001. *Data Mining : Concepts and Techniques*. Inteligent Databases Systems Research Lab, School of Computing Science, Simon Fraser University.
- [9] Lewis, M.D and Roger, J. 2000. *An Introduction to Classification and Regression Tree (CART) Analysis*. Presented at the 2000 Annual Meeting of Society For Academy Emergency Medicine in San Fransisco, California
- [10] Olfaz M, Tirink C, and Önder H. 2019. Use of CART and CHAID Algorithms In Karayaka Sheep Breeding. *Kafkas Univ Vet Fak Derg*, 25 (1): 105-110.
- [11] Putri, R. E, Suparti, dan Rahmawati, R. 2014. Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor pada Analisis Data Status Kerja di Kabupaten Demak Tahun 2012. *Jurnal Gaussian*, Vol 3, No 4, 831 – 838.
- [12] Tyas, A.E, Ispriyanti, D, dan Sudarno. 2015. Ketepatan Klasifikasi Status Kerja di Kota Tegal Menggunakan Algoritma C4.5 dan Fuzzy K-Nearest Neighbor In Every Class (FK-NNC). *Jurnal Gaussian*, Vol 4, No 4, 735-744.



PROSIDING

SEMINAR NASIONAL SAINS DAN ENTREPRENEURSHIP VI TAHUN 2019
"Tingkatkan Sains dalam Pembelajaran untuk Meningkatkan 3M (Manajemen, Marketing & Business) dan Berakademi
dalam 40 Rende/Entrepreneurship"

Semarang, 21 Agustus 2019