

# Pengklasifikasian Status Kerja pada Angkatan Kerja di Kabupaten Tanah Datar Menggunakan Metode CART dan Metode CHAID

Yulia Rizki Fajriati<sup>1</sup>, Syafriandi<sup>2</sup>

<sup>1</sup> Prodi Statistika, Fakultas Matematika Ilmu Pengetahuan dan Alam Universitas Negeri Padang (UNP)

<sup>2</sup> Dosen Statistika, Fakultas Matematika Ilmu Pengetahuan dan Alam Universitas Negeri Padang (UNP)

---

## Article Info

### Article history:

Received February 16, 2022

Revised August 01, 2022

Accepted September 15, 2022

---

### Keywords:

WorkForce

Tanah Datar Regency

CART

CHAID

### Kata Kunci:

Angkatan Kerja

Kabupaten Tanah Datar

CART

CHAID

## ABSTRACT

Tanah Datar Regency experiencing an increase in population, but the number of workforce doesn't increase. It makes the number of population with the workforce imbalance so that the public welfare decreased. Beside decreasing in public welfare, the workforce also has impacts in their employment status. Therefore, the classification needs to be done to find out the dominant factors to identified certain characteristics from certain segment using CART and CHAID methods. The results from CART and CHAID methods has obtained that the marriage status is the most dominant variable in classifying employment status in Tanah Datar Regency. Meanwhile, the best methods to classify the employment status in workforce in Tanah Datar Regency is CART method. It can be seen from the 73,9% accuracy and 26,1% of APER

## ABSTRAK

Kabupaten Tanah Datar mengalami peningkatan jumlah penduduk, namun jumlah angkatan kerja tidak mengalami peningkatan. Hal ini membuat ketidakseimbangan antara jumlah penduduk dan jumlah angkatan kerja, sehingga menyebabkan menurunnya kesejahteraan masyarakat. Selain menurunkan kesejahteraan masyarakat, angkatan kerja juga dapat mempengaruhi status kerja pada angkatan kerja. Oleh karena itu, perlu dilakukan pengklasifikasian untuk mengetahui faktor yang paling dominan dalam mengidentifikasi karakteristik dari segmen tertentu dengan menggunakan metode CART dan metode CHAID. Berdasarkan hasil metode CART dan metode CHAID, diperoleh bahwa variabel yang paling dominan mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar adalah variabel status perkawinan, sedangkan metode terbaik untuk mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar adalah metode CART. Ini dapat dilihat berdasarkan nilai *accuracy* sebesar 73,9 persen dan *APER* sebesar 26,1 persen.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



---

## Yulia Rizki Fajriati

Yulia Rizki Fajriati

Prodi Statistika, Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Negeri Padang, Jl. Prof. Dr. Hamka, Air Tawar Barat, Padang Utara, Padang, 25171 Padang Sumatera Barat

Email: [yulifajriati@gmail.com](mailto:yulifajriati@gmail.com)

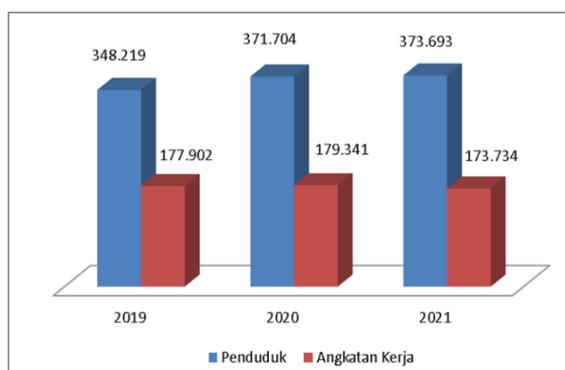
---

## 1. PENDAHULUAN

Salah satu penentu kesuksesan suatu perekonomian di bidang sumber daya manusia adalah ketersediaan tenaga kerja. Tenaga kerja merupakan semua penduduk yang telah mencapai batas usia kerja tertentu. Di Indonesia batasan usia minimal 15 tahun tanpa batasan maksimal [1]. Oleh

karena itu, mereka yang telah mencapai usia pensiun masih tetap bekerja, sehingga disebut tenaga kerja [2]. Tenaga kerja yang terdiri dari bekerja dan tidak bekerja (pengangguran) disebut angkatan kerja.

Pertumbuhan angkatan kerja menjadi meningkat, apabila jumlah penduduk mengalami peningkatan. Peningkatan jumlah penduduk terjadi di Kabupaten Tanah Datar. Berdasarkan data tahun 2019-2021, jumlah penduduk di Kabupaten Tanah Datar mengalami peningkatan, namun jumlah angkatan kerjanya tidak mengalami peningkatan. Hal ini dapat dilihat pada Gambar 1.



Sumber: BPS Sumatera Barat

Gambar 1. Perbandingan Jumlah Penduduk dan Jumlah Angkatan Kerja di Kabupaten Tanah Datar Tahun 2019-2021

Ketidakseimbangan jumlah penduduk dengan jumlah angkatan kerja terjadi pada tahun 2021 di Kabupaten Tanah Datar. Ketidakseimbangan ini menyebabkan menurunnya kesejahteraan masyarakat [3]. Selain menurunkan kesejahteraan masyarakat, juga dapat mempengaruhi status kerja. Ada beberapa faktor yang mempengaruhi status kerja diantaranya pendidikan, status perkawinan, jenis kelamin, usia, dan status pelatihan kerja [4]. Sehingga, untuk mengetahui faktor yang paling dominan dalam mengidentifikasi karakteristik dari segmen tertentu, maka perlu dilakukannya pengklasifikasian terhadap status kerja.

Pengklasifikasian dilakukan dengan menggunakan metode klasifikasi. Metode klasifikasi dibedakan menjadi dua kelompok yaitu non-parametrik dan parametrik [5]. Metode ini digunakan untuk mempelajari sekumpulan data sehingga memperoleh aturan yang bisa mengklasifikasi data baru [6]. Selain itu, metode klasifikasi dapat menganalisis hubungan antara variabel independen dengan variabel dependen [7]. Ada beberapa analisis klasifikasi diantaranya *Support Vector Machine*, *AdaBoost*, *Naive Bayes*, *Constant*, *K-Nearest Neighbour* (KNN), pohon klasifikasi, dan lain-lain [8].

Salah satu metode klasifikasi dengan pendekatan non-parametrik adalah pohon klasifikasi. Pohon klasifikasi digunakan untuk mengidentifikasikan karakteristik dari segmen tertentu dan memprediksi data baru. Ada beberapa kelebihan pohon klasifikasi. Pertama, pengambilan keputusan simpel dan spesifik. Kedua, sampel yang diuji berdasarkan kriteria. Ketiga, dapat menghindari munculnya permasalahan [9]. Beberapa metode yang dapat digunakan dalam metode pohon klasifikasi yaitu *Classification and Regression Tree* (CART), *Chi-square Automatic Interaction Detection* (CHAID), metode *Quick Unbiased Efficient Statistical Tree* (QUEST), dan lain-lain [10].

Metode pohon klasifikasi yang sering digunakan adalah metode CART dan metode CHAID. Metode CART memiliki beberapa kelebihan. Pertama, mudah untuk diinterpretasikan. Kedua, lebih mudah dalam eksplorasi data. Ketiga, menghasilkan klasifikasi akhir yang sederhana [11]. Sedangkan metode CHAID juga memiliki beberapa kelebihan. Pertama, tepat sasaran. Kedua, mendefinisikan ke dalam kelas regu yang sesuai. Ketiga, bersifat interatif [12]. Berdasarkan



pemaparan di atas, perlu dilakukan pengklasifikasian status kerja pada angkatan kerja di Kabupaten Tanah Datar tahun 2021 dengan menggunakan metode CART dan metode CHAID.

Untuk membandingkan kedua metode tersebut, maka pada setiap metode dilakukan perhitungan ketepatan klasifikasi. Hal ini dilakukan untuk melihat metode terbaik dari kedua metode tersebut [13]. Perhitungan *accuracy* dan *APER* yang digunakan untuk mengukur ketepatan klasifikasi. Dimana nilai *accuracy* tertinggi dan nilai *APER* terendah yang merupakan metode terbaik [14].

## 2. METODE

Data yang digunakan pada penelitian ini adalah data Survei Angkatan Kerja Nasional (SAKERNAS) tahun 2021 Kabupaten Tanah Datar sebanyak 1.057 responden. Data bersumber dari BPS Kabupaten Tanah Datar. Perangkat lunak yang digunakan adalah Rstudio.

Metode CART pertama kali diperkenalkan pada tahun 1984 [15]. Metode CART merupakan salah satu metode statistika non-parametrik yang dikembangkan untuk analisis klasifikasi dalam bentuk pohon [16]. Pohon klasifikasi pada metode CART terbentuk apabila data variabel dependen kategorik [15]. Tujuan metode ini adalah untuk mendapatkan kelompok data yang akurat [17]. Proses dalam metode CART terdiri dari tiga tahapan [18]:

### a. Pembentukan pohon klasifikasi

#### 1) Pemilahan pemilah

Pemilahan pemilah adalah ketergantungan nilai yang berasal dari variabel independen [19]. Pemilahan pemilah dipilih berdasarkan aturan indeks Gini sebagai berikut [15]:

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

$P(j|t)$  adalah probabilitas kelas  $j$  pada simpul  $t$ . Untuk menghitung probabilitas dari simpul kiri dan kanan digunakan rumus sebagai berikut [20]:

$$P_L = \frac{\text{calon simpul kiri}}{\text{data latihan}}. \quad (2)$$

$$P_R = \frac{\text{calon simpul kanan}}{\text{data latihan}}. \quad (3)$$

Untuk menentukan nilai *goodness of split* adalah sebagai berikut [15]:

$$\Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R). \quad (4)$$

Dimana:

$\Delta i(s, t)$  = Nilai *goodness of split*.

$i(t)$  = Indeks Gini pada simpul.

$i(t_R)$  = Indeks Gini pada simpul kanan.

$i(t_L)$  = Indeks Gini pada simpul kiri.

$P_R$  = Probabilitas simpul kanan.

$P_L$  = Probabilitas simpul kiri.

Nilai *goodness of split* terbesar yang merupakan pemilah utama.

#### 2) Penentuan simpul terminal

Penentuan simpul terminal dapat diketahui apabila pengamatan berjumlah kurang atau sama dengan 5 ( $n < 5$ ) dan tingkat kedalaman pohon sudah maksimal.

#### 3) Penandaan label kelas

Penandaan label kelas dilakukan berdasarkan aturan jumlah terbanyak sebagai berikut [21]:

$$P(j_0|t) = \max_j P(j|t). \quad (5)$$

### b. Pemangkasan pohon klasifikasi

Pemangkasan pohon klasifikasi digunakan untuk mendapatkan ukuran pohon yang layak, dengan rumus sebagai berikut [22]:

$$g_m(t) = \frac{R(t) - R(T_k)}{|T_k| - 1}. \quad (6)$$

Dimana:

$g_m(t)$  = Complexity parameter.

$R(t)$  = Kesalahan pengklasifikasian pada node  $t$ .

$T_k$  = *Subtree* ke- $k$ .

$R(T_k)$  = Kesalahan pengklasifikasian pada pohon.

Pemangkasan dilakukan pada  $g_m(t)$  terkecil dengan rumus sebagai berikut:

$$g_m(t_m) = \min_{t \in T_k} g_m(t). \quad (7)$$

c. Penentuan pohon klasifikasi optimal

Penentuan ini dilakukan berdasarkan penduga *cross validation* sebagai berikut:

$$R(T_t^{(v)}) = \frac{1}{N_v} \sum X(d^{(v)}). \quad (8)$$

Dimana:

$X(d^{(v)})$  = Hasil pengklasifikasian.

$N_v$  = Jumlah pengamatan dalam  $L_v$ .

Nilai penduga terkecil merupakan pohon klasifikasi optimal.

Metode CHAID pertama kali diperkenalkan pada tahun 1980 [23]. Metode CHAID merupakan salah satu metode statistika non-parametrik. Metode ini menggunakan data kategorik. Tujuan metode ini adalah untuk membagi kumpulan data menjadi sub kelompok berdasarkan variabel dependen [24]. Proses dalam metode CHAID juga terdiri dari tiga tahapan [25]:

a. Penggabungan

- 1) Membentuk tabel kontingensi dua arah untuk masing-masing variabel independen dengan variabel dependen yang terdapat pada Tabel 1.

Tabel 1. Tabel Kontingensi Dua Arah Untuk Uji Chi-square

	B <sub>1</sub>	B <sub>2</sub>	...	B <sub>m</sub>	Total Baris
A <sub>1</sub>	n <sub>11</sub>	n <sub>12</sub>	...	n <sub>1m</sub>	n <sub>1.</sub>
A <sub>2</sub>	n <sub>21</sub>	n <sub>22</sub>	...	n <sub>2m</sub>	n <sub>2.</sub>
...	...	...	...	...	...
A <sub>k</sub>	n <sub>k1</sub>	n <sub>k2</sub>	...	n <sub>km</sub>	n <sub>k.</sub>
Total Kolom	n <sub>.1</sub>	n <sub>.2</sub>	...	n <sub>.m</sub>	n <sub>..</sub> = n

- 2) Hitung uji statistik *chi-square* dengan rumus sebagai berikut:

$$\chi_{obs}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{\{Obs_{(i,j)} - \overline{Exp}_{(i,j)}\}^2}{\overline{Exp}_{(i,j)}}. \quad (9)$$

- 3) Untuk kategori berpasangan dilakukan hitung *p-value* berpasangan.
- 4) Setelah kategori digabungkan maka periksa kembali signifikansinya, apabila masih ada yang tidak signifikan lakukan langkah ketiga.
- 5) Hitung *p-value* terkoreksi Bonferroni berdasarkan kategori yang digabungkan, dimana pengali Bonferroni berbeda-beda pada setiap variabel independen. Apabila variabel independen monotonik, maka rumusnya sebagai berikut:

$$B = \binom{c-1}{r-1}. \quad (10)$$

Apabila variabel independen bebas, maka rumusnya sebagai berikut:

$$B = \sum_{i=0}^{r-1} (-1)^i \frac{\binom{c-i}{r-i}}{i!}. \quad (11)$$

Apabila variabel independen mengambang, maka rumusnya sebagai berikut:

$$B = \binom{c-2}{r-2} + r \binom{c-2}{r-1}. \quad (12)$$

b. Pemisahan

- 1) Memilih variabel yang paling signifikan.
- 2) Apabila *p-value* sama dengan  $\alpha$ , maka pemisah simpul menggunakan variabel independen.



c. Penghentian

- 1) Pada kasus hanya terdapat kasus yang berasal dari salah satu kategori variabel dependen.
- 2) Kedalaman pohon telah maksimal.
- 3) Jumlah kasus simpul terminal kurang dari minimal kasus untuk menjadi *parent node*.
- 4) Jumlah kasus dalam satu atau lebih simpul anak akan kurang dari jumlah minimal kasus untuk simpul anak.
- 5) Apabila tidak ada variabel independen yang signifikan.

Sebelum melakukan proses kedua metode tersebut, maka perlu dilihat deskripsi data terlebih dahulu. Apabila data seimbang, maka dapat melanjutkan analisis dari masing-masing metode. Tetapi apabila data tidak seimbang, maka perlu dilakukan proses *undersampling*. Proses *undersampling* merupakan proses pengambilan kelas mayoritas secara acak yang dilakukan tanpa pengembalian sehingga kelas mayoritas akan memiliki proporsi yang sama dengan kelas minoritas [26].

Setelah kedua metode tersebut dilakukan, maka dihitung ketepatan klasifikasi untuk setiap metode. Struktur data ketepatan klasifikasi terdapat pada Tabel 2.

Tabel 2. Struktur Data Ketepatan Klasifikasi

Kelas Asli (i)	Kelas Hasil Prediksi (j)		Total
	Kategori 1	Kategori 2	
Kategori 1	$n_{11}$	$n_{12}$	$N_{1.}$
Kategori 2	$n_{21}$	$n_{22}$	$N_{2.}$
Total	$N_{.1}$	$N_{.2}$	N

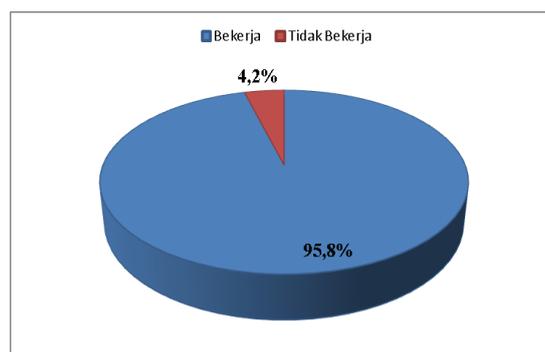
Untuk mengukur kinerja klasifikasi menggunakan rumus sebagai berikut:

$$APER = \frac{n_{12} + n_{21}}{N} \quad (13)$$

$$Accuracy = \frac{n_{11} + n_{22}}{N} \quad (14)$$

### 3. HASIL DAN PEMBAHASAN

Sebelum menganalisis data dengan metode CART maupun metode CHAID terlebih dahulu dilihat keseimbangan data yang digunakan. Pada Gambar 2 disajikan deskripsi data status kerja pada tahun 2021.



Gambar 2. Deskripsi Data Status Kerja Tahun 2021

Berdasarkan Gambar 2, responden yang bekerja berjumlah 1.013 orang dan tidak bekerja berjumlah 44 orang. Ini menunjukkan responden yang bekerja lebih banyak dari pada tidak bekerja. Sehingga terlihat bahwa data tersebut tidak seimbang. Untuk menyeimbangkan data tersebut, perlu dilakukannya proses *undersampling*.

Proses *undersampling* dilakukan menggunakan software Rstudio dengan *package* ROSE. Setelah proses *undersampling* dilakukan, maka diperoleh data yang seimbang yaitu 88 orang yang

terdiri dari bekerja 44 orang dan tidak bekerja 44 orang. Setelah data seimbang, maka selanjutnya dilakukan analisis data

Analisis pertama yang dilakukan adalah metode CART. Metode pengklasifikasian ini dimulai dengan menentukan pemilah utama. Pada Tabel 3 disajikan data calon pemilah.

Tabel 3. Data Calon Simpul Kiri dan Calon Simpul Kanan

No	Calon Simpul Kiri	Calon Simpul Kanan
1.	Pendidikan: <ul style="list-style-type: none"> <li>▪ Pendidikan dasar</li> <li>▪ Pendidikan menengah</li> </ul>	Pendidikan: <ul style="list-style-type: none"> <li>▪ Pendidikan tinggi</li> </ul>
2.	Pendidikan: <ul style="list-style-type: none"> <li>▪ Pendidikan dasar</li> <li>▪ Pendidikan tinggi</li> </ul>	Pendidikan: <ul style="list-style-type: none"> <li>▪ Pendidikan menengah</li> </ul>
3.	Pendidikan: <ul style="list-style-type: none"> <li>▪ Pendidikan menengah</li> <li>▪ Pendidikan tinggi</li> </ul>	Pendidikan: <ul style="list-style-type: none"> <li>▪ Pendidikan dasar</li> </ul>
4.	Status perkawinan: <ul style="list-style-type: none"> <li>▪ Sudah pernah kawin</li> </ul>	Status perkawinan: <ul style="list-style-type: none"> <li>▪ Belum pernah kawin</li> </ul>
5.	Jenis kelamin: <ul style="list-style-type: none"> <li>▪ Laki-laki</li> </ul>	Jenis kelamin: <ul style="list-style-type: none"> <li>▪ Perempuan</li> </ul>
6.	Usia: <ul style="list-style-type: none"> <li>▪ Produktif</li> </ul>	Usia: <ul style="list-style-type: none"> <li>▪ Tidak produktif</li> </ul>
7.	Status pelatihan kerja: <ul style="list-style-type: none"> <li>▪ Pernah</li> </ul>	Status pelatihan kerja: <ul style="list-style-type: none"> <li>▪ Tidak pernah</li> </ul>

Selanjutnya untuk memudahkan dalam menghitung indeks Gini dan kriteria *goodness of split*, maka sebaiknya dilakukan perhitungan nilai probabilitas pada setiap simpul yang disajikan pada Tabel 4.

Tabel 4. Hasil Perhitungan Probabilitas Simpul

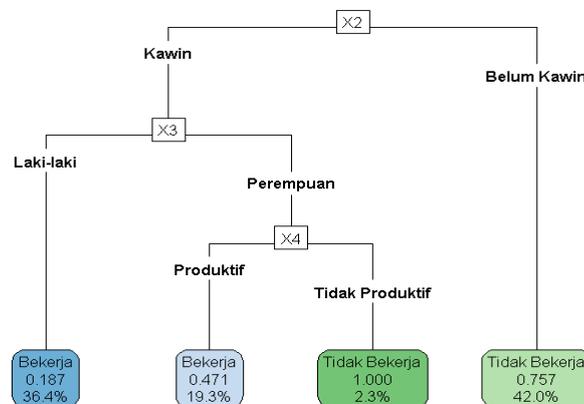
Simpul	$P_L$	$P_R$	Kelas	$P(j t_L)$	$P(j t_R)$
1	0,829	0,171	Bekerja	0,548	0,267
			Tidak Bekerja	0,452	0,733
2	0,659	0,341	Bekerja	0,517	0,467
			Tidak Bekerja	0,483	0,533
3	0,511	0,489	Bekerja	0,4	0,605
			Tidak Bekerja	0,6	0,395
4	0,580	0,420	Bekerja	0,686	0,243
			Tidak Bekerja	0,314	0,757
5	0,648	0,352	Bekerja	0,579	0,355
			Tidak Bekerja	0,421	0,645
6	0,932	0,068	Bekerja	0,5	0,5
			Tidak Bekerja	0,5	0,5
7	0,136	0,863	Bekerja	0,5	0,5
			Tidak Bekerja	0,5	0,5

Setelah memperoleh probabilitas berdasarkan simpul yang terdapat pada Tabel 4, maka selanjutnya menghitung indeks Gini dan kriteria *goodness of split* untuk menentukan *parent node* disajikan pada Tabel 5.

Tabel 5. Hasil Perhitungan Indeks Gini dan *Goodness of Split*

Simpul	$i(t)$	$\Delta i(s, t)$
1	0,283	-0,195
2	0,449	-0,049
3	0,500	-0,373
4	0,487	0,083
5	0,456	0,021
6	0,127	-0,021
7	0,236	-0,265

Berdasarkan Tabel 5, diperoleh *parent* node pada metode CART adalah calon simpul empat. Calon simpul empat ini merupakan variabel status perkawinan. Pada Gambar 3 disajikan diagram keputusan metode CART.



Gambar 3. Diagram Keputusan Metode CART

Berdasarkan Gambar 3, diperoleh empat terminal node. Kategori tidak bekerja diperoleh apabila memiliki status perkawinan belum pernah kawin atau apabila memiliki status perkawinan sudah pernah kawin, berjenis kelamin perempuan, dan usia tidak produktif. Sedangkan kategori bekerja apabila memiliki status perkawinan sudah pernah kawin yang berjenis kelamin laki-laki atau apabila memiliki status perkawinan sudah pernah kawin, jenis kelamin perempuan, yang berusia produktif

Perhitungan ketepatan klasifikasi dengan menggunakan metode CART disajikan pada Tabel 6.

Tabel 6. Hasil Ketepatan Klasifikasi Metode CART

Kelas Asli (i)	Kelas Hasil Prediksi (j)		Total
	Bekerja	Tidak Bekerja	
Bekerja	35	9	44
Tidak Bekerja	14	30	44
Total	49	39	88

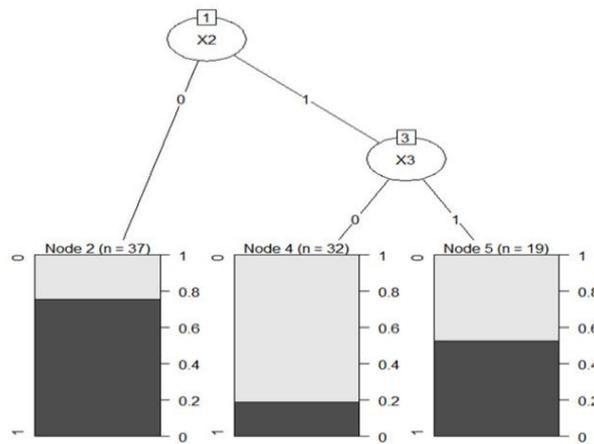
Berdasarkan Tabel 6, diperoleh nilai *accuracy* sebesar 73,9 persen dengan tingkat kesalahan *accuracy* (*APER*) sebesar 26,1 persen. Hal ini dapat dikatakan bahwa diagram keputusan metode CART mampu mengklasifikasikan data baru sebesar 73,9 persen dengan tingkat kesalahan 26,1 persen.

Analisis kedua yang dilakukan adalah metode CHAID. Metode pengklasifikasian ini dimulai dengan uji statistika *chi-square*. Hasil uji *chi-square* disajikan pada Tabel 7.

Tabel 7. Uji *Chi-square* Variabel Independen

Variabel	<i>Chi-square</i>	<i>P-value</i>
Pendidikan	5,284	0,071
<b>Status perkawinan</b>	<b>16,835</b>	<b>0,000</b>
Jenis kelamin	4,034	0,045
Usia	0,000	1,000
Status pelatihan kerja	0,000	1,000

Berdasarkan Tabel 7, diperoleh variabel yang paling signifikan yang merupakan pemilah utama adalah variabel status perkawinan. Pada Gambar 4 disajikan diagram keputusan metode CHAID.



Gambar 4. Diagram Keputusan Metode CHAID

Berdasarkan Gambar 4, diperoleh tiga terminal node. Pada node dua kategori tidak bekerja apabila memiliki status perkawinan belum pernah kawin sebanyak 37 orang. Pada node empat kategori bekerja apabila memiliki status perkawinan sudah pernah kawin dan berjenis kelamin laki-laki sebanyak 32 orang. Pada node lima kategori tidak bekerja apabila memiliki status perkawinan sudah pernah kawin dan berjenis kelamin perempuan sebanyak 19 orang.

Perhitungan ketepatan klasifikasi dengan menggunakan metode CHAID disajikan pada Tabel 8.

Tabel 8. Struktur Data Ketepatan Klasifikasi Metode CHAID

Kelas Asli (i)	Kelas Hasil Prediksi (j)		Total
	Bekerja	Tidak Bekerja	
Bekerja	26	18	44
Tidak Bekerja	6	38	44
Total	32	56	88

Berdasarkan Tabel 8, diperoleh nilai *accuracy* sebesar 73,7 persen dengan tingkat kesalahan *accuracy* (*APER*) sebesar 27,3 persen. Hal ini dapat dikatakan bahwa diagram keputusan metode CHAID mampu mengklasifikasikan data baru sebesar 73,7 persen dengan tingkat kesalahan 27,3 persen.



#### 4. Kesimpulan

Berdasarkan hasil dan pembahasan diperoleh kesimpulan bahwa variabel status perkawinan merupakan variabel yang paling dominan mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar menggunakan metode CART dan metode CHAID. Sedangkan metode terbaik untuk pengklasifikasian status kerja pada angkatan kerja di Kabupaten Tanah Datar adalah metode CART. Dimana hasil *accuracy* yang diperoleh metode CART sebesar 73,9 persen, sedangkan hasil *accuracy* metode CHAID sebesar 73,7 persen.

#### Ucapan Terimakasih

Ucapan terimakasih penulis ucapkan kepada orangtua dan adik yang selalu memberikan do'a, semangat dan nasihat serta kepada Bapak Syafriandi yang selalu membimbing penulis hingga dapat menyelesaikan jurnal ini.

#### REFERENSI

- [1] P. Simanjuntak, *Komposisi Penduduk dan Tenaga Kerja*, Jakarta: Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia, 2001.
- [2] P. Simanjuntak, *Pengantar Ekonomi Sumber Daya Manusia*, Jakarta: Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia, 2005.
- [3] Dinas Tenaga Kerja, "Masalah Tenaga Kerja dan Angkatan Kerja Di Indonesia 2019", 2019.
- [4] Aritonang dkk, "Klasifikasi Status Kerja pada Angkatan Kerja Kota Semarang Tahun 2014 Menggunakan Metode CHAID dan Metode CART", *Jurnal Gaussian*, 5(1): 183-192, 2016.
- [5] R. Schowengerdt, *Techniques for Image Processing and Classification in Remote Sensing*, Academic Press, 2012.
- [6] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*, Bandung: Penerbit Informatika, 2019.
- [7] D. Eherler dan T. Lehmann, *Responder Profiling with CHAID and Regression Tree*. 2001.
- [8] I. Oktanisa dan A. Supianto, "Perbandingan Teknik Klasifikasi dalam Data Mining Untuk Bank Direct Marketing", *Jurnal Teknologi Informatika dan Ilmu Komputer (JTIK)*, 5(5): 567-576, 2018.
- [9] N. Ratniasih, "Konversi Data Training tentang Pemilahan Kelas Menjadi Bentuk Pohon Keputusan dengan Teknik Klasifikasi", *Jurnal Eksplorasi Informatika*, 4(2): 145-154, 2016.
- [10] L. Rokach dan O. Maimon, *Data Mining with Decision Trees*, Singapore: World Scientific, 2008.
- [11] R. Timofeev, *Classification and Regression Tree Theory and Application*, Jerman: Humboldt University, 2004.
- [12] C. Gallagher, M. Howard, and L. Joice, "An Iterative Approach to Classification Analysis", *Journal of Applied Statistics*, 238-280, 2000.
- [13] A. Agresti, *Categorical Data Analysis*, John Wiley and Sons Inc, New York, 2002.
- [14] R. Johnson, dan D. Winchem, *Applied Multivariate Statistical Analysis 6 ed*, New Jersey: Prentice Hall, Englewood Clift, 2007.
- [15] L. Breiman, H. Friedman, A. Olshen, dan J. Stone, *Classification and Regression Tree*, New York: Chapman and Hall, 1984.
- [16] M. Ghiasi dkk, "Decision Tree-Based Diagnosis of Coronary Artery Disease: CART Model", *Comput Methods Program Biomed*, vol. 192, p. 105400, 2020.
- [17] WB. Kolmasari, "Metode Pohon Regresi Untuk Eksploratori Data dengan Peubah yang Banyak dan Kompleks", *Informatika Pertanian*, 16(1): 967-980, 2007.
- [18] S. Sumartini, dan S. Pumami, "Pengguna Metode Classification and Regression Tree (CART) untuk Klasifikasi Rekrutasi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya", *Jurnal Sains dan Seni ITS*, 4(2): 2337-2520, 2015.
- [19] A. Dessy, "Perbandingan Ketepatan Klasifikasi Antara Metode Regresi Logistik dan Klasifikasi Pohon", *Jurnal Kependidikan dan Kemasyarakatan*, 10(1): 59-77, 2012.
- [20] Mardiani, "Penerapan Klasifikasi dengan Algoritma CART untuk Prediksi Kuliah bagi Mahasiswa Baru", *Seminar Nasional Aplikasi Teknologi Informasi 2012*, ISSN: 1907-5022, 2012.
- [21] F. Pratiwi, dan I. Zain, "Klasifikasi Pengangguran Terbuka Menggunakan CART di Provinsi Sulawesi Utara", *Jurnal Sains dan Seni Pomits*, 3(1): 2337-3520, 2014.
- [22] R. Lewis, "An Introduction to Classification and Regression Tree (CART) Analysis", *Presented at the 2000 Annual Meeting of Society for Academic Emergency Medicine in San Fransisco*, California, 2000.
- [23] V. Kass, "An Exploratory Tehnique for Investigating Large Quantities of Categorical Data", *Applied Statistic*, 29(2): 119-127, 1980.
- [24] L. Hatcher, *Step by Step Basic Statistics Using SAS*, North California: SAS Publishing, 2003.
- [25] M. Baron, *Probability and Statistics for Computer Scientists 2 ed*, University of Texas at Dallas Richardson, USA: CRC Press, 2013.
- [26] N. Lunardon, G. Menardi, dan N. Torelli, ROSE: A Package for Binary Imbalanced Learning, *The R Journal*, 6(1): 2073-4859, 2014.