

# Klasifikasi Pengangguran Terbuka Menggunakan CART (*Classification and Regression Tree*) di Provinsi Sulawesi Utara

Febti Eka Pratiwi dan Ismaini Zain

Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

*e-mail*: ismaini\_z@statistika.its.ac.id

**Abstrak**—Pengangguran merupakan masalah besar yang selalu dihadapi oleh negara-negara berkembang. Tingginya pengangguran disuatu negara mengakibatkan rendahnya partisipasi masyarakat dalam pertumbuhan ekonomi yang dapat berdampak pada pembangunan nasional. Badan Pusat Statistika mencatat 7,7 juta orang Indonesia tidak terlibat dalam kegiatan ekonomi nasional yang dikarenakan pengangguran. Sulawesi Utara sebagai provinsi dengan peringkat keenam pengangguran tertinggi di Indonesia menyumbangkan 1,3 persen pengangguran di Indonesia. CART (*Classification and Regression Tree*) merupakan salah satu metode yang digunakan untuk pengklasifikasian. CART dapat digunakan pada skala data kategorik maupun rasio. Data yang digunakan pada penelitian ini adalah data sekunder dari hasil survei angkatan kerja nasional (SAKERNAS) bulan Agustus 2012. Variabel respon yaitu angkatan kerja yang bekerja dan tidak bekerja (pengangguran), dengan variabel prediktor jenis kelamin, tingkat pendidikan, usia, status dalam rumah tangga, pengalaman pelatihan kerja, status perkawinan, dan klasifikasi tempat tinggal. Klasifikasi pengangguran provinsi Sulawesi Utara dengan metode CART adalah status dalam rumah tangga, jenis kelamin, usia, pendidikan terakhir, dan status perkawinan.

**Kata Kunci**—Klasifikasi, CART, Pengangguran Terbuka

## I. PENDAHULUAN

PEMBANGUNAN nasional bertujuan untuk meningkatkan kesejahteraan masyarakat. Indonesia sebagai negara berkembang melakukan banyak perubahan untuk mendukung pembangunan nasional. Hal tersebut bertujuan agar terciptanya stabilitas nasional, iklim investasi yang baik, dan pertumbuhan ekonomi yang tinggi sehingga dapat menekan angka pengangguran di Indonesia. Tolak ukur keberhasilan pembangunan nasional salah satunya dapat dilihat dengan Indeks Pembangunan Manusia (*Human Development Index*) yang merupakan indikator peningkatan kualitas hidup manusia meliputi umur panjang dan sehat, perolehan dan pengembangan pengetahuan, dan peningkatan terhadap akses untuk kehidupan yang lebih baik. Diperlukan peningkatan pendapatan masyarakat untuk menunjang Indeks Pembangunan Manusia dan Indeks Kualitas Hidup agar Indeks Pembangunan Manusia tercapai. Terdapat masalah besar yang selalu dihadapi oleh negara berkembang dan dapat mempengaruhi pembangunan nasional yaitu pengangguran. Tingginya pengangguran di suatu negara mengakibatkan rendahnya partisipasi masyarakat dalam pertumbuhan ekonomi yang dapat berdampak pada pembangunan nasional. Badan Pusat

Statistika mencatat 7,7 juta orang Indonesia tidak terlibat dalam kegiatan ekonomi nasional yang dikarenakan pengangguran. Sulawesi Utara sebagai provinsi besar ditimur Indonesia mengalami banyak permasalahan, salah satunya adalah pengangguran. Sulawesi Utara menempati peringkat keenam pengangguran tertinggi di Indonesia. Hal tersebut berarti Sulawesi Utara menyumbangkan 1,3 persen pengangguran di Indonesia. Salah satu metode yang digunakan untuk menggolongkan pengangguran berdasarkan faktor yang mempengaruhi adalah *Classification and Regression Trees* (CART). Kedua metode ini merupakan metode yang bisa diterapkan untuk data dalam jumlah besar, variabel yang sangat banyak dan melalui prosedur pemilah biner. Pada penelitian ini mengklasifikasikan pengangguran terbuka di provinsi Sulawesi Utara yang dilakukan dengan menerapkan metode CART. Data yang digunakan dalam penelitian data sekunder angkatan kerja provinsi Sulawesi Utara dari hasil survey angkatan kerja nasional (SAKERNAS) Bulan Agustus 2012. Berdasarkan latar belakang yang telah dipaparkan penelitian ini membahas mengenai klasifikasi pengangguran berdasarkan faktor yang mempengaruhi pengangguran terbuka provinsi di Sulawesi Utara dengan metode CART.

## II. TINJAUAN PUSTAKA

### A. Pengangguran Terbuka

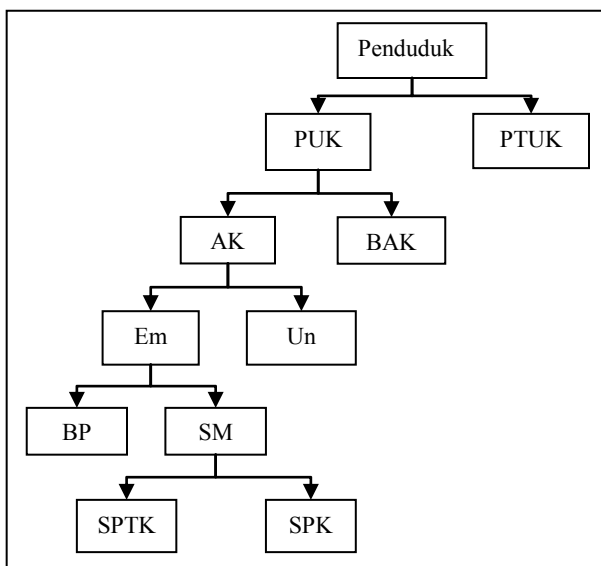
Penduduk adalah semua orang yang mendiami suatu wilayah tertentu pada waktu tertentu dan terlepas dari warga Negara atau bukan warga negara [1]. Penduduk adalah mereka yang bertempat tinggal atau berdomisili didalam suatu wilayah Negara. Jumlah penduduk yang besar disuatu negara tidak otomatis akan menjadi modal pembangunan, bahkan dapat pula justru menjadi beban dan tanggungan penduduk lainnya.

Penduduk dapat diklasifikasikan kedalam 5 level. Penduduk pada penelitian ini terbagi menjadi dua kategori dan merupakan level 2 yaitu penduduk usia kerja (PUK) dan penduduk tidak usia kerja (PTUK). Penduduk usia kerja sendiri terbagi menjadi dua kelompok, yaitu angkatan kerja (AK) dan bukan angkatan kerja (BAK). Level 4 dari pembagian penduduk adalah penduduk yang tergolong dalam angkatan kerja terbagi menjadi dua kategori, yaitu pengangguran (*Unemploy*) dan bekerja (*Employ*). Bekerja penuh (BP) dan setengah menganggur (SM) merupakan kategori dari angkatan kerja yang tergolong bekerja. Setengah menganggur dibagi

Gambar 1 Penduduk dan Tenaga Kerja

Keterangan Gambar 1 :

PUK : Penduduk Usia Kerja



- PTUK: Penduduk Tidak Usia Kerja
- AK : Angkatan Kerja
- BAK : Bukan Angkatan Kerja
- Em : *Employ* (Bekerja)
- Un : *Unemploy* (Tidak Bekerja)
- BP : Bekerja Penuh
- SM : Setengah Menganggur
- SPTK : Separuh Pengangguran Tidak Ketara
- SPK : Separuh Pengangguran Ketara

kembali menjadi dua kategori, yaitu setengah pengangguran tidak ketara (SPTK) dan setengah pengangguran ketara (SPK) [2]. Klasifikasi penduduk tersebut memperlihatkan bahwa penduduk yang merupakan angkatan kerja dan tergolong pengangguran dapat menjadi beban penduduk yang lainnya.

Pengangguran terbuka merupakan bagian dari angkatan kerja yang tidak bekerja atau sedang mencari pekerjaan (baik bagi mereka yang belum pernah bekerja sama sekali maupun yang sudah pernah bekerja), atau sedang mempersiapkan suatu usaha, mereka yang tidak mencari pekerjaan karena merasa tidak mungkin untuk mendapatkan pekerjaan dan mereka yang sudah memiliki pekerjaan namun belum memulai pekerjaan [3]. Pengangguran adalah jumlah tenaga kerja dalam perekonomian yang secara aktif mencari pekerjaan tetapi belum memperolehnya [4].

Tingkat pengangguran dapat diketahui dari hasil Survei Angkatan Kerja Nasional (SAKERNAS) yang dilakukan secara triwulan setiap tahunnya mulai tahun 2011, yaitu Februari, Mei, Agustus, dan Nopember. Usia, pendapatan rumah tangga/keluarga, tingkat pendidikan, ketrampilan dan pelatihan kerja mempengaruhi seseorang untuk bekerja [2].

**B. Classification and Regression Trees (CART)**

CART (*Classification and Regression Trees*) merupakan salah satu metode atau algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan. CART terbilang sederhana namun merupakan metode yang kuat. CART bertujuan untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian, selain itu CART digunakan untuk menggambarkan hubungan antara variabel respon (variabel dependen atau tak bebas ) dengan satu atau

lebih variabel prediktor (variabel independen atau bebas). Model pohon yang dihasilkan bergantung pada skala variabel respon, jika variabel respon data berbentuk kontinu maka model pohon yang dihasilkan adalah regression trees (pohon regresi) sedangkan bila variabel respon mempunyai skala kategorik maka pohon yang dihasilkan adalah classification trees (pohon klasifikasi) [5].

CART mempunyai beberapa kelebihan dibandingkan metode klasifikasi lainnya, yaitu hasilnya lebih mudah diinterpretasikan, lebih akurat dan lebih cepat penghitungannya, selain itu CART bisa diterapkan untuk himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan biner [6]. Data *learning* digunakan untuk pembentukan pohon klasifikasi optimal sedangkan data *testing* digunakan untuk validasi model yaitu seberapa besar kemampuan model dalam memprediksi data baru. Metode CART (*Classification and Regression Tress*) memiliki kelemahan sebagai berikut :

1. CART mungkin tidak stabil dalam *decision trees* (pohon keputusan) karena CART sangat sensitif dengan data baru.
 

CART sangat bergantung dengan jumlah sampel. Jika sampel data *learning* dan *testing* berubah maka pohon keputusan yang dihasilkan juga ikut berubah.
2. Tiap pemilihan bergantung pada nilai yang hanya berasal dari satu variabel penjelas.

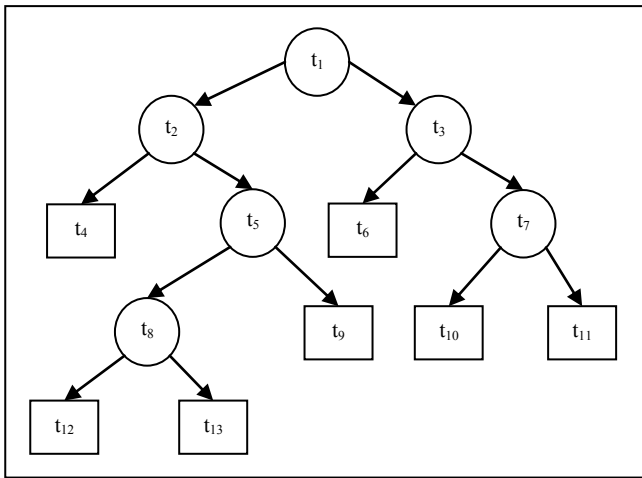
Pohon keputusan dibentuk dengan menggunakan algoritma penyekatan rekursif secara biner (*binery recursive partitioning*). Pemilahan dilakukan untuk memilah data menjadi 2 kelompok, yaitu kelompok yang masuk simpul kiri dan yang masuk simpul kanan. Pemilahan dilakukan pada tiap simpul sampai didapatkan suatu simpul terminal/akhir. Variabel yang memilah pada simpul utama adalah variabel terpenting dalam menduga kelas dari amatan.

Simpul utama (*root node*) dinotasikan sebagai  $t_1$ , sedangkan simpul  $t_2, t_3, t_5, t_7$  dan  $t_8$  disebut simpul dalam (*internal nodes*). Simpul akhir yang juga disebut sebagai simpul terminal (*terminal nodes*) adalah  $t_4, t_6, t_9, t_{10}, t_{11}, t_{12}$  dan  $t_{13}$  dimana tidak terjadi lagi pemilahan. Kedalaman pohon (*depth*) dihitung dimulai dari simpul utama atau  $t_1$  yang berada pada kedalaman 1, sedangkan  $t_2$  dan  $t_3$  berada pada kedalaman 2. Begitu seterusnya sampai pada simpul terminal  $t_{12}$  dan  $t_{13}$  yang berada pada kedalaman 5.

Langkah-langkah penerapan Algoritma CART adalah sebagai berikut :

1. Pembentukan pohon klasifikasi
 

Proses pembentukan pohon klasifikasi terdiri atas 3 tahapan, yaitu
- a. Pemilihan (*Classifier*)



Gambar 2 Struktur Pohon Klasifikasi

Sampel data *Learning (L)* yang masih bersifat heterogen digunakan untuk pembentukan pohon klasifikasi. Sampel tersebut akan dipilah berdasarkan aturan pemilahan dan kriteria *goodness-of-split* dan pemilihan pemilah tergantung pada jenis variabel responnya. Metode pemilihan pemilah menggunakan *Impurity measure i(t)* merupakan pengukuran tingkat keheterogenan suatu kelas dari suatu simpul tertentu dalam pohon klasifikasi yang dapat membantu kita menemukan fungsi pemilah yang optimal. Beberapa fungsi *Impurity measure i(t)* adalah sebagai berikut :

- i. indeks Gini :  $i(t) = \sum_{i \neq j} p(i|t)p(j|t)$
- ii. indeks Informasi  $i(t) = -\sum_j p(j|t) \log [p(j|t)]$
- iii. indeks Twoing :  $i(t) = \frac{2}{\sqrt{2}} \left[ \sum p(j|t_L) - p(j|t_R) \right]^2$
- iv. indeks Entropi :  $i(t) = N_j(t) \log [p(j|t)]$

$p(j|t)$  adalah peluang  $j$  pada simpul  $t$ . *Goodness of Split* merupakan suatu evaluasi pemilahan oleh pemilah  $s$  pada simpul  $t$  yang didefinisikan sebagai penurunan keheterogenan dan didefinisikan sebagai

$$G(s,t) = \Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R)$$

Pemilah yang menghasilkan nilai  $\Delta i(s,t)$  lebih tinggi merupakan pemilah yang lebih baik karena hal ini memungkinkan untuk mereduksi keheterogenan secara lebih signifikan. Metode pemilahan yang sering digunakan adalah indeks Gini, hal tersebut dikarenakan lebih mudah dan sesuai untuk diterapkan dalam berbagai kasus dan mempunyai perhitungan yang sederhana dan cepat [5].

b. Penentuan Simpul Terminal

Suatu simpul  $t$  akan menjadi simpul terminal atau tidak, akan dipilah kembali bila pada simpul  $t$  tidak terdapat penurunan keheterogenan secara berarti atau adanya batasan minimum  $n$  seperti hanya terdapat satu pengamatan pada tiap simpul anak. Jumlah kasus minimum dalam suatu terminal

akhir umumnya adalah 5, dan apabila hal itu terpenuhi maka pengembangan pohon dihentikan [5].

c. Penandaan Label Kelas

Penandaan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak. Label kelas simpul terminal  $t$  adalah  $j_0$  yang memberi nilai dugaan kesalahan pengklasifikasian simpul  $t$  terbesar. Proses pembentukan pohon klasifikasi berhenti saat terdapat hanya satu pengamatan dalam tiap simpul anak atau adanya batasan minimum  $n$ , semua pengamatan dalam tiap simpul anak identik, dan adanya batasan jumlah level/kedalaman pohon maksimal.

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)}$$

dengan

$p(j|t)$  : proporsi kelas  $j$  pada simpul

$N_j(t)$  : jumlah pengamatan kelas  $j$  pada simpul  $t$

$N(t)$  : jumlah pengamatan pada simpul  $t$ .

2. Pemangkasan pohon klasifikasi

Bagian pohon yang kurang penting dilakukan pemangkasan sehingga didapatkan pohon klasifikasi yang optimal. Pemangkasan didasarkan pada suatu penilaian ukuran sebuah pohon tanpa mengorbankan kebaikan ketepatan melalui pengurangan simpul pohon sehingga dicapai ukuran pohon yang layak. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak tersebut adalah *Cost complexity minimum* [6].

Sebagai ilustrasi, untuk sembarang pohon  $T$  yang merupakan sub pohon dari pohon terbesar  $T_{max}$  ( $T < T_{max}$ ) ukuran *cost complexity* yaitu I

$$R_\alpha(t) = R(t) + \alpha |\tilde{T}| \tag{2}$$

dimana :

$R(T) = Resubstitution Estimate$  (Proporsi kesalahan pada sub pohon)  $\tag{3}$

$\alpha$  = kompleksitas parameter (*complexity parameter*)

$|\tilde{T}|$  = ukuran banyaknya simpul terminal pohon  $T$   $\tag{4}$

3. Penentuan Pohon Klasifikasi Optimal

Ukuran pohon yang besar akan menyebabkan nilai kompleksitas yang tinggi karena struktur data yang digambarkan cenderung kompleks, sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti cukup kecil. Ada dua jenis penduga pengganti, penduga sampel uji (*test sample estimate*) dan penduga validasi silang lipat  $V$  (*cross validation V-fold estimate*). Penelitian ini menggunakan penduga sampel uji (*test sample estimate*) karena *cross validation V-fold estimate* digunakan untuk jumlah sampel kecil (kurang dari 3000). Penelitian ini menggunakan penduga sampel uji (*test sample estimate*) untuk penentuan pohon optimal.

1. Penduga Sampel Uji (*test sample estimate*)

Penduga sampel uji  $L$  dibagi menjadi dua himpunan data, yaitu  $L_1$  (*learning set*) dan  $L_2$  (*testing set*). Data yang masuk kedalam kelompok  $L_1$  digunakan untuk pembentukan pohon

Tabel 1.  
Struktur Data Hasil Klasifikasi Pohon Optimal

Aktual	Prediksi		Total
	1	2	
1	n <sub>11</sub>	n <sub>12</sub>	N <sub>1</sub>
2	n <sub>21</sub>	n <sub>22</sub>	N <sub>2</sub>
Total	N <sub>1</sub>	N <sub>2</sub>	N

$$R^{ts}(T_0) = \frac{1}{N_2} \sum_{i,j} C(i|j) N_{2ij}$$

klasifikasi. Data yang masuk kedalam kelompok  $L_2$  digunakan untuk menduga  $R^{ts}(T_i)$  dimana  $N_2$  adalah jumlah amatan dalam  $L_2$ .  $R^{ts}(T_i)$  adalah total proporsi dari kesalahan klasifikasi pohon klasifikasi. Pohon optimal adalah  $T^*$  yang memenuhi kriteria  $R^{ts}(T^*) = \min R^{ts}(T_i)$ .

Pohon klasifikasi optimal yang telah terbentuk dilakukan evaluasi dari hasil klasifikasi. Cara untuk mengevaluasi hasil klasifikasi adalah dengan menghitung akurasi klasifikasi. *Sensitivity* menggambarkan akurasi pada sampel kelas  $i$  sedangkan *specificity* menggambarkan akurasi pada sampel kelas  $j$ . G-means dapat menggambarkan bagaimana sebuah metode klasifikasi mampu mengukur *sensitivity* dan *specificity*. Semakin besar nilai G-means menunjukkan metode klasifikasi tersebut mampu memprediksi data di tiap kelas dengan baik.

$$1 - \text{APER (Tingkat Akuransi Total)} = \frac{n_{11} + n_{22}}{N}$$

$$\text{Total Tingkat Kesalahan} = \frac{n_{11} + n_{22}}{N}$$

$$Se = \frac{n_{11}}{N_1}$$

$$Sp = \frac{n_{22}}{N_2}$$

Keterangan :

n<sub>11</sub> : jumlah observasi dari kelas 1 yang tepat diprediksi sebagai kelas 1

n<sub>22</sub> : jumlah observasi dari kelas 2 yang tepat diprediksi sebagai kelas 2

n<sub>12</sub> : jumlah observasi dari kelas 1 yang salah diprediksi sebagai kelas 2

n<sub>21</sub> : jumlah observasi dari kelas 2 yang salah diprediksi sebagai kelas 1

N<sub>1</sub> : jumlah observasi dari kelas 1

N<sub>2</sub> : jumlah observasi dari kelas 2

N : jumlah observasi

Se : *Sensitivity* (ketepatan klasifikasi kelas 1 terhadap jumlah total observasi kelas 1)

Sp : *Specificity* (ketepatan klasifikasi kelas 2 terhadap jumlah total observasi kelas 2)

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder. Data tersebut merupakan data angkatan kerja provinsi Sulawesi Utara yang berasal dari hasil survei angkatan kerja nasional (SAKERNAS) bulan Agustus 2012. Jumlah data dalam penelitian ini sebanyak 14597 responden.

#### B. Definisi Operasional Variabel

##### 1. Pengangguran

Pengangguran di Indonesia merupakan bagian dari angkatan kerja. Angkatan kerja digolongkan menjadi dua kategori, yaitu pengangguran dan bukan pengangguran.

##### 2. Jenis Kelamin

Jenis kelamin adalah kondisi biologis dengan pembagian alamiah manusia yang masing-masing menunjukkan karakteristik laki-laki atau perempuan.

Tabel 2.  
Variabel Penelitian

No	Variabel	Skala	Keterangan
1	Status (Y)	Nominal	1: Bukan Pengangguran 2: Pengangguran
2	Jenis Kelamin (X <sub>1</sub> )	Nominal	1: Laki-laki 2: Perempuan
3	Tingkat Pendidikan (X <sub>2</sub> )	Nominal	1: Rendah (≤ SD) 2: Sedang (SMP-SMA) 3: Tinggi (> SMA)
4	Usia (X <sub>3</sub> )	Nominal	1: Sekolah (10-14 tahun) 2: Produktif (15-64 tahun) 3: Tidak Produktif (> 64 tahun)
5	Status dalam Rumah Tangga (X <sub>4</sub> )	Nominal	1: Kepala Rumah Tangga 2: Bukan Kepala Rumah Tangga
6	Pelatihan Kerja (X <sub>5</sub> )	Nominal	1: Pernah 2: Tidak Pernah
7	Status Perkawinan (X <sub>6</sub> )	Nominal	1: Belum Kawin 2: Kawin
8	Klasifikasi Tempat Tinggal (X <sub>7</sub> )	Nominal	1: Perkotaan 2: Pedesaan

#### C. Identifikasi Variabel

##### 3. Tingkat Pendidikan

Tingkat pendidikan adalah suatu kondisi jenjang pendidikan tertinggi yang dimiliki oleh seseorang melalui pendidikan formal yang dipakai oleh pemerintah serta disahkan oleh departemen pendidikan.

Kategori dalam jenjang pendidikan adalah sebagai berikut:

- Pendidikan rendah adalah Sekolah Dasar
- Pendidikan sedang adalah sekolah menengah baik itu Sekolah Menengah Pertama (SMP) maupun Sekolah Menengah Atas (SMA)
- Pendidikan tinggi yang meliputi Diploma, Sarjana, dan Pascasarjana.

##### 4. Usia

Usia adalah satuan waktu yang mengukur waktu keberadaan suatu benda atau makhluk, baik yang hidup maupun mati. Usia dapat dibedakan menjadi 3 kategori yaitu :

- Usia sekolah adalah usia ≤ 14 tahun
- Usia produktif adalah usia antara 15-64 tahun
- Usia tidak produktif adalah usia > 64 tahun

##### 5. Status dalam Rumah Tangga

Status dalam rumah tangga adalah status atau kedudukan yang dimiliki seseorang dalam rumah tangga atau keluarga. Objek dalam penelitian ini adalah seluruh responden pada SAKERNAS 2012.

##### 6. Pelatihan Kerja

Pelatihan kerja adalah suatu pelatihan/seminar/kursus yang menumbuhkan kreatifitas ataupun ketrampilan untuk mendukung seseorang memasuki dunia kerja atau agar dapat mandiri membuka lapangan pekerjaan.

##### 7. Status Perkawinan

Status perkawinan adalah status yang dimiliki seseorang apakah memiliki/ pernah memiliki ikatan perkawinan, baik secara formal negara, agama atau adat

##### 8. Klasifikasi Tempat Tinggal

Klasifikasi tempat tinggal (perkotaan atau pedesaan) mempengaruhi angkatan kerja untuk bekerja atau tidak [1].

D. Metode Analisis Data

Analisis CART menggunakan langkah-langkah sebagai berikut

- a. Kombinasi yang digunakan data *learning* dan data *testing*, yaitu (95 persen : 5 persen), (90 persen : 10 persen), (85 persen : 15 persen), (80 persen : 20 persen), dan (75 persen : 25 persen) dengan melihat nilai ketepatan klasifikasi data *testing* dan data *learning*.
- b. Pembentukan (*growing*) pohon klasifikasi maksimal,
- c. Pemangkasan pohon klasifikasi yang paling kecil dengan kriteria kompleksitas kesalahan (*cost complexity*) yang minimum
- d. Penentuan pohon klasifikasi optimal
- e. Validasi pohon klasifikasi optimal.

IV. PEMBAHASAN

Hasil penelitian yang dilakukan beserta analisisnya adalah sebagai berikut.

A. Karakteristik Pengangguran Sulawesi Utara 2012

Hasil survei angkatan kerja nasional (SAKERNAS) Agustus 2012 memberikan informasi bahwa dari 14597 terdapat 49,8 persen atau 7264 angkatan kerja adalah pengangguran dengan 3831 dari 7264 pengangguran memiliki pendidikan yang rendah ( $\leq$  Sekolah Dasar (SD)). Mayoritas pengangguran di provinsi Sulawesi Utara berjenis kelamin perempuan yaitu sebesar 4827 dan sebanyak 66,02 persen dari 7264 pengangguran di Provinsi Sulawesi Utara tergolong dalam usia yang produktif. Hasil SAKERNAS Provinsi Sulawesi Utara bulan Agustus 2012 menyebutkan bahwa hanya 2 persen dari 7264 pengangguran yang memperoleh pelatihan kerja.

B. Hasil Klasifikasi Pengangguran dengan Metode CART

Data *learning* digunakan untuk pembentukan pohon klasifikasi sedangkan data *testing* digunakan untuk validasi model.

Kombinasi data *learning* dan *testing* tertinggi dicapai data *learning* 95 persen dan *testing* 5 persen. Sehingga kombinasi data *learning* dan *testing* inilah yang digunakan pada analisis selanjutnya

C. Pohon Klasifikasi Maksimal

Pohon klasifikasi maksimal adalah pohon klasifikasi dengan jumlah simpul terminal terbanyak. Metode pemilih pada penelitian ini menggunakan Indeks Gini. Semua variabel prediktor masuk dalam klasifikasi pohon maksimal yang terbentuk. Variabel-variabel tersebut adalah Jenis Kelamin ( $X_1$ ), Pendidikan terakhir ( $X_2$ ), Usia ( $X_3$ ), Status dalam Rumah Tangga ( $X_4$ ), Pengalaman Pelatihan Kerja ( $X_5$ ), Status Perkawinan( $X_6$ ), dan Klasifikasi Tempat Tinggal ( $X_7$ ). Variabel status dalam rumah tangga merupakan pemilah yang memiliki peranan utama dalam pembentukan pohon maksimal dan merupakan variabel yang sangat dominan dalam pengelompokan.

Tabel 3. Perbandingan Ketepatan Klasifikasi Antar Kombinasi Data

No	Kombinasi Data (%)		Ketepatan Klasifikasi (%)	
	Learning	Testing	Learning	Testing
1	95	5	0,778	0,789
2	90	10	0,786	0,771
3	85	15	0,785	0,778
4	80	20	0,785	0,779
5	75	25	0,786	0,777

Tabel 4.

Nilai simpul terminal, *test sets relative cost*, *resubstitution raltive cost*, dan *complexity*

Tree Number	Terminal Nodes	Test Set Relative Cost	Resubstitution Relative Cost	Complexity
1*	40	0.439 ± 0.030	0.430	0.000
2	36	0.439 ± 0.030	0.430	1.00E-005
3	27	0.431 ± 0.030	0.430	2.75E-005
4	21	0.434 ± 0.030	0.430	4.49E-005
5	20	0.444 ± 0.030	0.430	6.28E-005
6	17	0.444 ± 0.030	0.431	8.18E-005
7	13	0.428 ± 0.030	0.432	0.000138
8	12	0.428 ± 0.030	0.432	0.000158
9	10	0.425 ± 0.029	0.435	0.000818
10**	8	0.417 ± 0.029	0.443	0.002
11	7	0.465 ± 0.031	0.453	0.005
12	6	0.480 ± 0.031	0.471	0.009
13	5	0.513 ± 0.032	0.491	0.010
14	2	0.637 ± 0.031	0.569	0.013
15	1	1.000 ± 0.000	1.000	0.215

Keterangan :

\* : Pohon Klasifikasi Maksimal

\*\* : Pohon Klasifikasi Optimal

D. Pemangkasan Pohon Klasifikasi Maksimal (Pruning)

Proses pemangkasan dilakukan saat proses pembentukan pohon klasifikasi maksimal telah terbentuk. Pemangkasan dilakukan berdasarkan aturan *cost complexity minimum* dan menggunakan penduga sampel uji (*test sample estimate*). Ukuran pohon yang besar akan menyebabkan nilai *complexity* yang tinggi karena struktur data yang digambarkan cenderung kompleks walaupun memberikan nilai penduga yang sangat kecil, sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai pengganti yang cukup kecil. Klasifikasi pohon optimal yang terbentuk pada penelitian ini adalah 8 simpul terminal (pohon klasifikasi yang terbentuk dapat dilihat pada *draft TA*).

Pohon klasifikasi maksimal menghasilkan *resubstitution relative cost* sebesar 0,439 dan kompleksitas parameter (*complexity parameter*) sebesar 0,000. *Relative cost* yang dimiliki oleh pohon klasifikasi maksimal adalah sebesar 0,439±0,030. Sedangkan nilai penduga pengganti (*resubstitution relative cost*) pada pohon klasifikasi optimal adalah sebesar 0,443, kompleksitas parameter (*complexity parameter*) sebesar 0,002, dan *relative cost* sebesar 0,417±0,029. Sehingga nilai *relative cost* pohon maksimal lebih kecil dari nilai *relative cost* pohon optimal.

Variabel-variabel yang masuk kedalam pohon klasifikasi optimal adalah 5 variabel. Variabel tersebut adalah Jenis Kelamin ( $X_1$ ), Pendidikan terakhir ( $X_2$ ), Usia ( $X_3$ ), Status dalam Rumah Tangga ( $X_4$ ), dan Status Perkawinan( $X_6$ ).

Tabel 5.  
Skor Variabel Prediktor dalam Pohon Klafifikasi Maksimal

Variabel	Score (%)
X <sub>4</sub>	100.00
X <sub>1</sub>	69.51
X <sub>3</sub>	49.39
X <sub>6</sub>	12.61
X <sub>2</sub>	9.38
X <sub>5</sub>	0.00
X <sub>7</sub>	0.00

Tabel 6.  
Karakteristik Pengangguran Terbuka Provinsi Sulawesi Utara

No.	Simpul Terminal	Karakteristik
1	2	a. Kepala rumah tangga b. Usia tidak produktif (> 64 tahun)
2	6	a. Bukan kepala rumah tangga b. Usia produktif (15 - 64 tahun) c. Memiliki pendidikan sedang (SMP – SMA) d. Laki-laki e. Belum menikah
3	7*	a. Bukan kepala rumah tangga b. Usia produktif (15 – 64 tahun) c. Tidak berpendidikan tinggi ( ≤ SMA) d. Laki-laki
4	8*	a. Bukan kepala rumah tangga b. Usia produktif (15 – 64 tahun)

Keterangan :

(\*) Simpul terminal dengan jumlah pengangguran tertinggi

Tabel 7.

Observasi	Prediksi		Total	Kesalahan/ Misklasifikasi (%)	
	1	2			
Data	1	4615	2344	6959	33,68
Learning	2	735	6173	6908	10,64
Tingkat Akurasi Total (%)					77,80
Total Tingkat Kesalahan (%)					22,20
Sensitivity (%)					86,26
Specificity (%)					72,48

Keterangan :

1 : Bukan Pengangguran

2 : Pengangguran

3 Pohon Klasifikasi

Pohon klasifikasi optimal yang terbentuk menghasilkan 8 simpul terminal. Klasifikasi angkatan kerja yang tergolong kedalam kategori pengangguran terbentuk dalam 4 simpul terminal. Delapan puluh persen pengangguran terklasifikasi pada simpul terminal 7 dan simpul terminal 8.

Prosentase kesalahan pengklasifikasian angkatan kerja yang tergolong dalam bekerja untuk data *learning* adalah sebesar 33,68 persen dan dengan ketepatan pengklasifikasiannya sebesar 77,80 persen. Untuk angkatan kerja yang dikategorikan sebagai pengangguran terdapat sebanyak 6908 orang, diduga sebanyak 735 orang salah diklasifikasikan dengan prosentase kesalahan pengklasifikasian sebesar 10,64 persen dan dengan ketepatan klasifikasi sebesar 89,36 persen.

Dapat disimpulkan untuk nilai total akurasi pada data *learning* sebesar 77,80 persen artinya pohon klasifikasi yang terbentuk mampu memprediksi dengan tepat pengamatan sebesar 77,80 persen. Tingkat akurasi pada angkatan kerja yang tergolong bekerja ditunjukkan oleh nilai *sensitivity* sebesar 86,26 persen dan tingkat akurasi pada angkatan kerja yang tergolong pengangguran ditunjukkan oleh nilai *specificity* yaitu sebesar

Tabel 8.  
Ketepatan Klasifikasi Data *Testing*

Observasi	Prediksi		Total	Kesalahan/ Misklasifikasi (%)	
	1	2			
Data	1	258	116	374	31
Testing	2	38	318	356	10,67
Tingkat Akurasi Total (%)					78,90
Total Tingkat Kesalahan (%)					21,1
Sensitivity (%)					87,16
Specificity (%)					73,27

Keterangan :

1 : Bukan Pengangguran

2 : Pengangguran

72,48 persen.

#### 4. Validasi Pohon Klasifikasi

Ketepatan klasifikasi pohon optimal yang dihasilkan adalah sebesar 78,90 persen. Nilai *sensitivity* pada data testing adalah 87,16 persen yang berarti tingkat akurasi pada angkatan kerja yang tergolong bekerja adalah sebesar 87,16 persen. Tingkat akurasi pada angkatan kerja yang tergolong pengangguran adalah sebesar 73,27 persen yang ditunjukkan oleh nilai *specificity*. Tingkat akurasi total antara data *learning* dan data *testing* seimbang sehingga dapat dikatakan klasifikasi pohon optimal yang terbentuk sudah baik.

### V. KESIMPULAN DAN SARAN

Faktor yang mempengaruhi pengangguran terbuka di provinsi Sulawesi Utara dengan menggunakan CART yaitu Jenis Kelamin, Pendidikan terakhir, Usia, Status dalam Rumah Tangga, dan Status Perkawinan. Metode CART pada penelitian ini memiliki ketepatan klasifikasi 78,90 persen sehingga diperlukan penelitian dengan menggunakan metode yang berbeda (SVM ataupun CHAID). Hal tersebut bertujuan agar didapatkan ketepatan klasifikasi yang lebih tinggi. Perlu dilakukan penambahan variabel lain, missal pendapatan dan klasifikasi kabupaten atau kota agar dapat lebih mewakili karakteristik pengangguran di provinsi Sulawesi Utara.

### DAFTAR PUSTAKA

- [1] Suryana. 2000. Ekonomi Pembangunan Problematika dan Pendekatan, Bandung: Salemba Empat
- [2] Widodo, S T. 1990, Indikator Ekonomi, Yogyakarta: Kanisius.
- [3] Sistem Informasi Rujukan Statistik. 2011. Survei Angkatan Kerja Nasional (SAKERNAS) [Online]. Available : diakses tanggal 17 Januari 2014
- [4] Sadono, Sukirno . 2004. Makro Ekonomi Edisi Ketiga. Jakarta : PT. Raja Grafindo Persada
- [5] Breiman L., Friedman J.H Olshen R.A & Stone C.J. 1984. Classification And Regression Tree. New York, NY: Chapman And Hall
- [6] Lewis, M.D dan 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 [Online]. Available: diakses tanggal 28 September 2013
- [7] Badan Pusat Statistika (BPS) Provinsi Sulawesi Utara. 2013. Keadaan Ketenagakerjaan Sulawesi Utara Tahun 2012 [Online]. Available: diakses tanggal 30 September 2013
- [8] Walpole, R E. 1995. Pengantar Metode Statistika Edisi Keempat, Institut Teknologi Bandung: Bandung.
- [9] Yuniarto. 2009. Klasifikasi Angkatan Kerja Provinsi Bengkulu Menggunakan Metode CART dan Regresi Logistik [Tugas Akhir], Surabaya: Institut Teknologi Sepuluh Nopember.