

## Use of Classification and Regression Tree (CART) Method for Classification of Labor Force Participation Levels in Medan City in 2019

Makmul Ikhsan Siahaan<sup>1</sup>, Elly Rosmaini<sup>2</sup>

<sup>1</sup>Prodi Matematika, FMIPA, Universitas Sumatera Utara, Medan-Indonesia 20155

<sup>2</sup>Program Matematika, FMIPA, Universitas Sumatera Utara, Medan-Indonesia 20155

Email: <sup>1</sup>makmulikhsan@gmail.com, <sup>2</sup>ellyrosmaini@gmail.com

### ABSTRAK

*Classification and regression trees (CART)* merupakan metode klasifikasi yang menggunakan data historis untuk membangun suatu pohon keputusan. Metodologi CART mulai dikembangkan pada tahun 80-an oleh Breiman, Friedman, Olshen, Stone pada tahun 1984. Tingkat partisipasi angkatan kerja (TPAK) adalah penduduk yang termasuk bukan angkatan kerja adalah penduduk usia kerja (15 tahun dan lebih) yang masih sekolah, mengurus rumah tangga atau melaksanakan kegiatan lainnya selain kegiatan pribadi. Penduduk yang termasuk angkatan kerja adalah penduduk usia kerja (15 tahun dan lebih) yang masih sekolah, mengurus rumah tangga atau melaksanakan kegiatan lainnya selain kegiatan pribadi. Melalui analisis metode CART, diperoleh kelompok pertama yaitu angkatan kerja dengan jenis kelamin perempuan dengan tingkat pendidikan sedang sebanyak 861 (35,58%, n=2564). Kelompok kedua yaitu angkatan kerja dengan jenis kelamin perempuan dengan tingkat pendidikan (tinggi dan rendah) sebanyak 472 (18,41%, n=2564). Kelompok ketiga yaitu angkatan kerja dengan jenis kelamin laki-laki dengan status pernikahan menikah sebanyak 740 (28,86%, n=2564). Kelompok keempat yaitu angkatan kerja jenis kelamin laki-laki dengan status pernikahan belum menikah yang memiliki tingkat pendidikan rendah sebanyak 48 (1,87%, n=2564). Kelompok kelima angkatan kerja jenis kelamin laki-laki status pernikahan menikah yang memiliki tingkat pendidikan tinggi sedang sebanyak 444 (17,32%, n=2564).

**Kata kunci:** Angkatan Kerja, Badan Pusat Statistik, CART, Medan

### ABSTRACT

*Classification and regression trees (CART)* is a classification method that uses historical data to build a decision tree. The CART methodology was developed in the 80s by Breiman, Friedman, Olshen, Stone in 1984. The labor force participation rate (TPAK) is the population that is not included in the labor force, namely the working age population (15 years and over) who are still in school take care of the household or carry out other activities other than personal activities. The population included in the labor force is the population of working age (15 years and over) who are still in school, taking care of the household or carrying out other activities other than personal activities. Through the analysis of the CART method, the first group was obtained, namely the female workforce with a moderate level of education as many as 861 (33.58%, n=2564). The second group is the female workforce with education levels (high and low) as many as 472 (18.41%, n=2564). The third group is the workforce with male sex with married marital status as many as 740 (28.86%, n=2564). The fourth group is the male sex workforce with unmarried marital status who has a low level of education as many as 48 (1.87%, n=2564). The fifth group of the male sex workforce with married marital status who have a moderate level of higher education is 444 (17.32%, n=2564).

**Keywords:** Labor Force, Central Bureau of Statistics, CART, Medan

### A. Pendahuluan

Pengklasifikasian merupakan salah satu metode statistik untuk mengelompokkan atau mengklasifikasikan suatu data yang disusun secara sistematis. Masalah klasifikasi sering dijumpai dalam kehidupan sehari-hari, baik pengklasifikasian data pada bidang akademik,

kesehatan, segmentasi pasar, maupun pada bidang lainnya. Masalah-masalah tersebut dapat diselesaikan dengan metode klasifikasi, namun pada penyelesaiannya perlu diperhatikan dalam memilih metode klasifikasi yang tepat.

Metode klasifikasi dapat dilakukan dengan pendekatan parametrik dan

nonparametrik. Salah satu metode klasifikasi dengan pendekatan nonparametrik yang sering digunakan adalah *Decision Tree* (Pohon keputusan). Pohon keputusan adalah suatu metode eksplorasi berstruktur pohon untuk melihat hubungan antar variabel respon dengan variabel penjelasnya. Beberapa metode yang dapat digunakan dalam metode pohon keputusan antara lain CHAID (*Chi-Squared Automatic Interaction Detection Analysis*), QUEST (*Quick, Unbiased Efficient, Statistical Tree*), CART (*Classification and Regression Tree*), dan lain-lain dimana masing-masing metode tersebut memiliki kekuatan dan kelemahannya masing-masing (Maimon and Rokach, 2010).

CART (*Classification and regression trees*) adalah salah satu metode atau algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan. CART dikembangkan untuk melakukan analisis klasifikasi pada variabel respon baik yang nominal, ordinal maupun kontinu. CART juga dapat menyeleksi variabel-variabel dan interaksi-interaksi variabel yang paling penting dalam menentukan hasil atau variabel prediktor (Breiman *et al*, 1993). Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Prinsip dari metode pohon klasifikasi ini adalah memilah seluruh amatan menjadi dua gugus amatan dan memilah kembali gugus amatan berikutnya, hingga diperoleh jumlah amatan minimum pada tiap-tiap gugus amatan berikutnya.

Penelitian dengan menggunakan klasifikasi CART pernah dilakukan oleh Nuriyah (2013) yaitu membandingkan metode CHAID dan CART dalam menentukan klasifikasi alumni UIN Sunan Kalijaga berdasarkan masa Study dan dapat disimpulkan bahwa metode CART lebih efisien dalam mengklasifikasikan alumni UIN Sunan Kalijaga, hal ini dapat dilihat dari ketepatan klasifikasi CHAID sebesar 63.3% dan ketepatan klasifikasi CART sebesar 66.6%. Selain itu, Klasifikasi CART pernah dilakukan oleh Reny Roswita Nazar (2018) dalam jurnal yang berjudul Penerapan Metode Chaid (*Chi-Squared Automatic Interaction Detection*) dan Cart (*Classification And Regression Trees*) Pada Klasifikasi Preeklampsia, menerapkan

metode CHAID dan CART dalam klasifikasi Preeklampsia diperoleh hasil bahwa ketepatan hasil klasifikasi dengan algoritma CART sebesar 74% ini lebih tinggi dibandingkan dengan metode CHAID yang sebesar 67%.

Berdasarkan uraian tersebut maka penulis tertarik untuk melakukan penelitian dengan judul penelitian “Penggunaan Metode *Classification and Regression Tree* (CART) Untuk Klasifikasi Tingkat Partisipasi Angkatan Kerja di Kota Medan Tahun 2019”.

## B. Metode Penelitian

Penelitian ini merupakan penelitian deskriptif yaitu untuk klarifikasi menggunakan metode *Classification and Regression Tree* (CART) pada Tingkat Partisipasi Angkatan Kerja di Kota Medan Tahun 2019. Variabel yang digunakan dalam penelitian ini dapat di definisikan sebagai berikut:

1. Jenis Kelamin adalah pembagian alamiah manusia berdasarkan karakteristik yang dimiliki yaitu laki-laki atau perempuan
2. Tingkat Pendidikan adalah suatu kondisi yang menunjukkan jenjang pendidikan yang ditempuh seseorang. Ada 3 kategori tingkat pendidikan sebagai berikut:
  - a. Pendidikan rendah yaitu Sekoah Dasar
  - b. Pendidikan sedang yaitu Sekolah Menengah Pertama (SMP) dan Sekolah Menengah Atas (SMA)
  - c. Pendidikan tinggi yaitu pendidikan yang ditempuh setelah Sekolah Menengah Atas (SMA) seperti Diploma, Sarjana, dan Pascasarjana
3. Pengalaman Pelatihan Kerja adalah pelatihan-pelatihan yang dilakukan untuk menambah keterampilan seseorang seperti kursus atau seminar
4. Status Pernikahan adalah status yang dimiliki seseorang apakah sudah menikah atau belum
5. Angkatan kerja yang bekerja adalah penduduk usia kerja (15 tahun ke atas) yang sedang memperoleh pendapatan atau keuntungan
6. Angkatan kerja yang tidak bekerja adalah penduduk usia kerja (15 tahun ke atas) yang tidak mempunyai pekerjaan.

Variabel yang digunakan dalam penelitian ini meliputi variabel dependen dan variabel independen. Variabel dependen yang digunakan pada penelitian ini adalah status

tingkat partisipasi angkatan kerja di kota Medan tahun 2019. Variabel dependen ini berskala biner yaitu:

- 1 : Angkatan kerja yang bekerja
  - 0 : Angkatan kerja yang tidak bekerja
- Sedangkan variabel independen yang digunakan pada penelitian ini merupakan faktor-faktor yang mempengaruhi tingkat partisipasi angkatan kerja di kota Medan tahun 2019 yang dapat dilihat pada tabel 1.

**Tabel 1.** Variabel Penelitian

Nama Variabel	Skala	Kategori
Jenis Kelamin	Nominal	1 : Laki-Laki
		2 : Perempuan
Tingkat Pendidikan	Ordinal	1 : Rendah ( $\leq$ SD)
		2 : Sedang (SMP-SMA)
		3 : Tinggi (Sarjana)
Status Pernikahan	Nominal	1 : Menikah
		2 : Belum Menikah
Pengalaman Pelatihan Kerja	Nominal	1 : Pernah
		1 : Tidak Pernah

Analisis data yang digunakan pada penelitian ini sebagai berikut:

1. Membagi data menjadi dua bagian yaitu data *learning* dan data *testing*.
2. Membentuk pohon klasifikasi maksimal menggunakan data *learning* dengan langkah-langkah sebagai berikut:
  - a. Melakukan pemilihan pemilah berdasarkan aturan pemilihan indeks gini yang selanjutnya dievaluasi menggunakan kriteria *goodness of split*. Pertama kali dihitung indeks gini.

$$GI(t) = - \sum_{j=1}^n P(j|t) \log_2 P(j|t)$$

(Jiawei, 2012)

Fungsi Indeks Gini sebagai berikut:

$$i(t) = 1 - \sum_{j=1}^n P^2(j|t)$$

(Jiawei, 2012)

Langkah selanjutnya adalah menentukan kriteria *goodness-of-split* sebagai berikut:

$$\phi(s, t) = i(t) - P_L i(t_L) - P_R i(t_R)$$

- b. Menentukan simpul terminal atau terminal node.

$$R(t) > R(t_R) + R(t_L)$$

- c. Melakukan penandaan label kelas pada terminal node berdasarkan aturan jumlah terbanyak dari tiap kelas yang ada pada variabel respon.

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

3. Memangkas pohon berdasarkan kriteria *cost complexity minimum*.

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

Cabang yang dipangkas adalah cabang yang memiliki nilai  $g_m(t)$  terkecil, yaitu:

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

4. Memilih pohon klasifikasi optimal dengan menggunakan persamaan *cross validation V-fold estimated*.

$$R(T_t^{(v)}) = \frac{1}{N_v} \sum_1^v X(d^{(v)})$$

5. Menghitung ketepatan klasifikasi pohon hasil bentukan dengan menggunakan data *learning* dan mengevaluasinya dengan menjalankan data *testing* pada klasifikasi yang terbentuk dengan menggunakan *sensitivity*, *specificity*, dan akurasi.

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

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

$$Akurasi = \frac{n_{11} + n_{22}}{N}$$

6. Menghitung jumlah tingkat partisipasi angkatan kerja yang bekerja dan tidak bekerja berdasarkan pohon klasifikasi yang terbentuk.

**C. Hasil dan Pembahasan**

**1. Membagi data menjadi data *learning* dan data *testing***

Membagi data menjadi dua bagian yaitu data *learning* dan data *testing*. Pembagian data *learning* dan data *testing* pada penelitian ini dapat dilihat pada tabel dibawah ini :

**Tabel 2.** Perbandingan data *learning* dan data *testing*

No	Perbandingan		Akurasi
	Data <i>Learning</i>	Data <i>Testing</i>	
1	95%	5%	72,16%
2	90%	10%	71,66%
3	85%	15%	71,58%
4	80%	20%	71,78%
5	75%	25%	70,28%

Berdasarkan Tabel 2 di atas perbandingan data *learning* dan data *testing* diperoleh proporsi data *learning* dan data *testing* yang memiliki nilai akurasi tertinggi adalah perbandingan 95%:5%, sehingga dalam penelitian ini akan digunakan proporsi data *learning* dan data *testing* adalah 95%:5%, dimana data *learning* sebanyak 2564 sampel dan data *testing* sebanyak 135 sampel.

**2. Membentuk Pohon Klasifikasi Maksimal**

Membentuk pohon klasifikasi maksimal dengan melakukan pemilihan pemilah berdasarkan aturan pemilihan indeks gini yang selanjutnya dievaluasi menggunakan kriteria *goodness of split*.

**a. Gain Information**

Sebelum mencari nilai *gain information*, untuk memudahkan sebaiknya terlebih dahulu mencari nilai probabilitas untuk setiap simpul.

**Tabel 3.** Hasil Perhitungan probabilitas simpul

	$P_L$	$P_R$	Kelas	$P(j t_L)$	$P(j t_R)$
1	$\frac{1231}{2564} = 0,48010$	$\frac{1333}{2564} = 0,5198$	Bekerja	$\frac{833}{1231} = 0,67668$	$\frac{539}{1333} = 0,40435$
			Tidak Bekerja	$\frac{398}{1231} = 0,32331$	$\frac{794}{1333} = 0,59564$
2	$\frac{439}{2564} = 0,17121$	$\frac{2125}{2564} = 0,8287$	Bekerja	$\frac{298}{439} = 0,67881$	$\frac{1073}{2125} = 0,50494$
			Tidak Bekerja	$\frac{141}{439} = 0,3211$	$\frac{1052}{2125} = 0,49505$
3	$\frac{1724}{2564} = 0,67238$	$\frac{840}{2564} = 0,32761$	Bekerja	$\frac{871}{1724} = 0,50522$	$\frac{501}{840} = 0,59642$

	$P_L$	$P_R$	Kelas	$P(j t_L)$	$P(j t_R)$
			Tidak Bekerja	$\frac{853}{1724} = 0,49477$	$\frac{339}{840} = 0,40357$
4	$\frac{401}{2564} = 0,15639$	$\frac{2163}{2564} = 0,84360$	Bekerja	$\frac{202}{401} = 0,50374$	$\frac{1169}{2163} = 0,54045$
			Tidak Bekerja	$\frac{199}{401} = 0,49625$	$\frac{994}{2163} = 0,45954$
5	$\frac{1667}{2564} = 0,65015$	$\frac{897}{2564} = 0,349843$	Bekerja	$\frac{958}{1667} = 0,57468$	$\frac{413}{897} = 0,46042$
			Tidak Bekerja	$\frac{709}{1667} = 0,42531$	$\frac{484}{897} = 0,53957$
6	$\frac{121}{2564} = 0,047191$	$\frac{2443}{2564} = 0,95280$	Bekerja	$\frac{74}{121} = 0,61157$	$\frac{1298}{2443} = 0,53131$
			Tidak Bekerja	$\frac{47}{121} = 0,38842$	$\frac{1145}{2443} = 0,46868$

Dengan menggunakan nilai pada Tabel 3 selanjutnya untuk nilai *gain information* yang digunakan untuk mendeteksi atribut-atribut yang paling banyak memberikan informasi berdasarkan kelas tertentu. Untuk mencari nilai *gain information* untuk setiap atribut:

**Tabel 4.** Hasil Perhitungan *Gain information*

Simpul	$GI(P_L)$	$GI(P_R)$
1	0,90802	0,97344
2	0,90566	0,99993
3	0,99992	0,973
4	0,99996	0,99527
5	0,98384	0,99548
6	0,963779	0,99717

**b. Indeks Gini**

Dengan menggunakan nilai dari tabel 4. selanjutnya mencari nilai indeks gini untuk setiap simpul:

**Tabel 5.** Hasil Perhitungan indeks gini

Simpul	$i(t)$
1	0,499209
2	0,283803
3	0,440566
4	0,263873
5	0,454906

**c. Goodness of split**

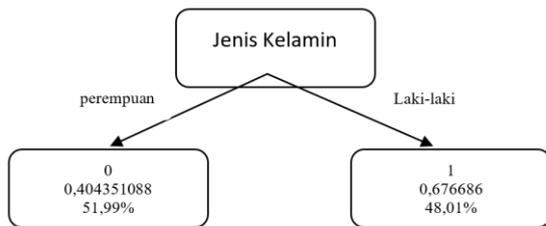
Selanjutnya dilakukan pemilahan pemilah atau calon simpul yang akan menjadi *parent node* atau simpul akar dengan menggunakan kriteria *Goodness of split*. Nilai

*goodness of split* untuk setiap calon simpul dapat dilihat pada Tabel 6.

**Tabel 6.** Tabel nilai *goodness of split*

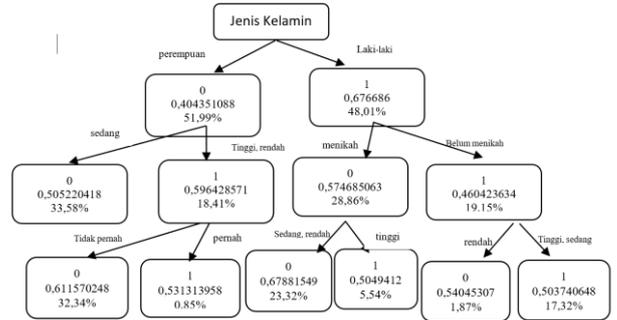
Simpul	$\phi(s, t)$	Kriteria kebaikan
1	<b>0,038697292</b>	1
2	-0,205206941	4
3	-0,053305234	3
4	-0,233361656	5
5	-0,036744782	2
6	-0,407026913	6

Berdasarkan perhitungan nilai *goodness of split* pada tabel 6 di atas untuk setiap calon simpul yang tertinggi adalah calon simpul 1 sebesar 0,038697292, sehingga calon simpul 1 ini akan menjadi *parent node*. Pada atribut jenis kelamin terdapat 2 nilai atribut yaitu laki-laki dan perempuan dan dapat dilihat pada gambar berikut:



**Gambar 1.** Nilai Atribut Laki-Laki dan Perempuan

Berdasarkan gambar di atas diperoleh bahwa *parent node* jenis kelamin yang terdiri dari node 2 yaitu laki-laki dan node 3 yaitu perempuan. Dimana dapat dilihat di atas bahwa nilai probabilitas untuk angkatan kerja laki-laki yang bekerja sebanyak 0,676686 dari total angkatan kerja laki-laki yang bekerja sebanyak 48,01% sedangkan nilai probabilitas untuk angkatan kerja perempuan yang bekerja sebanyak 0,404351088 dari total angkatan kerja perempuan sebanyak 51,99%. Proses pemilihan akan terus berlanjut secara rekursif dengan cara yang sama pada node atau simpul yang lain sampai pada simpul akhir atau terminal node. Setelah dilakukan proses rekursif terhadap node-node yang lain dan sampai pada terminal node maka diperoleh pohon keputusan sebagai gambar berikut:



**Gambar 2.** Pohon Keputusan Probabilitas Angkatan Kerja Laki-Laki dan Perempuan

Berdasarkan node akar adalah jenis kelamin yang terdiri dari perempuan dan laki-laki yang masing-masing menunjukkan node 2 dan 3 setelah mengalami proses pemecahan maka node 2 bercabang menjadi node 4 tingkat pendidikan rendah yang menjadi terminal node 5 tingkat pendidikan tinggi dan sedang. Begitu pula node 3 bercabang menjadi node 6 status pernikahan menikah dan node 7 status pernikahan belum menikah. Kemudian node 5 bercabang menjadi node 8 pengalaman pelatihan kerja tidak pernah dan node 9 pengalaman pelatihan kerja pernah. Pada terminal node 6 bercabang menjadi node 10 tingkat pendidikan sedang dan rendah dan node 11 tingkat pendidikan tinggi. Pada node 7 bercabang menjadi node 12 tingkat pendidikan rendah, dan node 13 tingkat pendidikan tinggi dan sedang.

**Tabel 7.** Struktur Pohon

Node	Nama	Keterangan
1	Jenis kelamin	Non terminal node
2	Jenis kelamin perempuan	Non terminal node
3	Jenis kelamin laki-laki	Non terminal node
4	Tingkat pendidikan sedang	Non terminal node
5	Tingkat pendidikan tinggi dan rendah	Terminal node
6	Status pernikahan belum menikah	Non terminal node
7	Status	Non terminal

	pernikahan menikah	node
8	Pengalaman pelatihan kerja tidak pernah	Terminal node
9	Pengalaman pelatihan kerja pernah	Terminal node
10	Tingkat pendidikan rendah	Terminal node
11	Tingkat pendidikan tinggi dan sedang	Terminal node
12	Tingkat pendidikan tinggi	Terminal node
13	Tingkat pendidikan sedang dan rendah	Terminal node

Berdasarkan tabel di atas dapat dilihat bahwa yang menjadi non terminal node adalah jenis kelamin, jenis kelamin perempuan, jenis kelamin laki-laki, tingkat pendidikan sedang, status pernikahan belum menikah dan status pernikahan menikah sedangkan yang menjadi terminal node adalah tingkat pendidikan tinggi dan rendah, pengalaman pelatihan kerja tidak pernah, pengalaman pelatihan kerja pernah, tingkat pendidikan rendah, tingkat pendidikan tinggi dan sedang, tingkat pendidikan tinggi dan tingkat pendidikan sedang dan rendah.

d. Pelabelan Kelas

Pelabelan kelas pada terminal node berdasarkan aturan jumlah terbanyak dari tiap kelas yang ada pada variabel respon. Penandaan label kelas berguna untuk mengidentifikasi tiap node pada kelas tertentu. Tabel penandaan label kelas dapat dilihat sebagai berikut:

**Tabel 8.** Penandaan Label Kelas

Node	Nama	Label kelas
1	Jenis kelamin	1
2	Jenis kelamin perempuan	0
3	Jenis kelamin laki-laki	1
4	Tingkat pendidikan sedang	0

Node	Nama	Label kelas
5	Tingkat pendidikan tinggi dan rendah	1
6	Status pernikahan menikah	0
7	Status pernikahan belum menikah	1
8	Pengalaman pelatihan kerja pernah	1
9	Pengalaman pelatihan kerja tidak pernah	0
10	Tingkat pendidikan rendah	0
11	Tingkat pendidikan tinggi dan sedang	1
12	Tingkat pendidikan tinggi	1
13	Tingkat pendidikan sedang dan rendah	0

Berdasarkan tabel 4.16 di atas dapat dijelaskan bahwa node dengan label kelas angkatan kerja yang bekerja berturut-turut adalah 1,3,5,7,8,10 dan 12. Sedangkan untuk node yang tidak bekerja berturut-turut adalah 2,4,6,9,11 dan 13.

**3. Pemangkasan Pohon Klasifikasi**

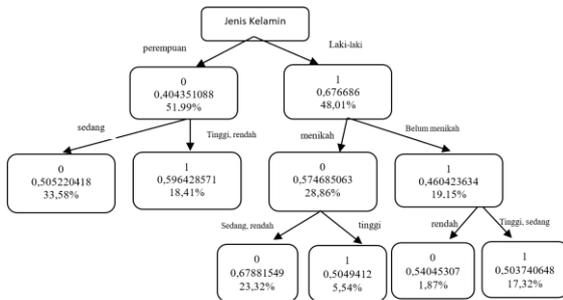
Pohon klasifikasi maksimal yang dihasilkan kemudian dilakukan pemangkasan pohon klasifikasi berdasarkan kriteria *cost complexity minimum*. Dalam mendapatkan pohon yang baik maka dapat dilakukan dengan melakukan pemangkasan pohon berdasarkan persamaan (2.9) dan (2.10). Cabang yang akan dilakukan pemangkasan adalah cabang yang memiliki nilai *complexity parameter* pada pohon T sebagai berikut :

**Tabel 9.** nilai *complexity parameter* pohon T

No	<i>complexity parameter</i>
1	0,310452418
2	0,113028257
3	0,079610073
4	0,0209459464
5	0,015243902
6	0,005807201

Berdasarkan pada tabel di atas diperoleh nilai *complexity parameter* untuk setiap split berturut-turut adalah 0,310452418,

0,113028257, 0,079610073, 0,0209459464, 0,015243902 dan 0,005807201. Maka pohon yang akan dipangkas adalah subpohon yang memiliki nilai *complexity parameter minimum* dimana ditunjukkan bahwa nilai *complexity parameter minimum* yang akan dipangkas adalah 0,005807201. Sehingga setelah dilakukan pemangkasan maka diperoleh struktur pohon sebagai berikut:



**Gambar 3.** Pohon Pemangkasan Complexity Parameter Minimum

Berdasarkan node akar adalah jenis kelamin yang terdiri dari laki-laki dan perempuan yang masing-masing menunjukkan node 2 dan 3 setelah mengalami proses pemecahan maka node 2 bercabang menjadi node 4 tingkat pendidikan rendah yang menjadi terminal node 5 tingkat pendidikan tinggi dan sedang. Begitu pula node 3 bercabang menjadi node 6 status pernikahan menikah dan node 7 status pernikahan belum menikah. Kemudian node 6 bercabang menjadi node 8 tingkat pendidikan sedang dan rendah dan node 9 tingkat pendidikan tinggi. Pada node 7 bercabang menjadi node 10 tingkat pendidikan rendah, dan node 11 tingkat pendidikan tinggi dan sedang.

**4. Memilih Pohon Klasifikasi Optimal**

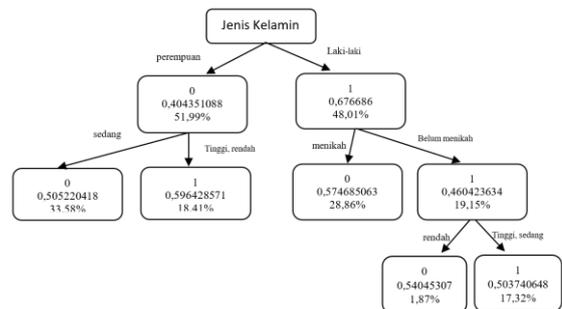
Setelah melakukan pemangkasan pohon maka akan dilakukan pemilihan pohon optimal dengan menggunakan penduga validasi silang lipat V (*cross validation V-fold estimated*) karena ukuran data yang kecil yaitu kurang dari 3000 data.

**Tabel 10.** Tabel proporsi kesalahan *V-fold estimated*

V-fold	Proporsi kesalahan klasifikasi
1	0,36544462
2	0,31437855
3	0,40435
4	0,19864865

V-fold	Proporsi kesalahan klasifikasi
5	0,45121951

Berdasarkan tabel 4.18 di atas dapat dilihat bahwa nilai proporsi kesalahan untuk 5 fold berturut-turut adalah 0,36544462, 0,31437855, 0,40435, 0,19864865, 0,45121951. Sehingga diperoleh pohon optimal dengan nilai proporsi kesalahan terkecil adalah 0,19864865. Setelah dilakukan pemangkasan maka diperoleh struktur pohon sebagai berikut :



**Gambar 4.** Pohon Pemangkasan Nilai Proporsi Kesalahan

Berdasarkan node akar adalah jenis kelamin yang terdiri dari laki-laki dan perempuan yang masing-masing menunjukkan node 2 dan 3 setelah mengalami proses pemecahan maka node 2 bercabang menjadi node 4 tingkat pendidikan rendah yang menjadi terminal node 5 tingkat pendidikan tinggi dan sedang. Begitu pula node 3 bercabang menjadi node 6 status pernikahan menikah dan node 7 status pernikahan belum menikah. Pada node 7 bercabang menjadi node 9 tingkat pendidikan rendah, dan node 10 tingkat pendidikan tinggi dan sedang.

**5. Menghitung Ketepatan Klasifikasi Untuk Data Prediksi**

Berikut merupakan tabel ketepatan klasifikasi untuk data prediksi :

**Tabel 11.** Ketepatan Klasifikasi Untuk Data Prediksi

Observasi	Prediksi		Total
	Angkatan kerja yang bekerja	Angkatan kerja yang tidak bekerja	
Angkatan kerja yang bekerja	46	26	72
Angkatan	24	39	63

Observasi	Prediksi		Total
	Angkatan kerja yang bekerja	Angkatan kerja yang tidak bekerja	
Kerja yang tidak bekerja			
Total	70	65	135

Berdasarkan tabel di atas dapat dijelaskan bahwa untuk angkatan kerja yang bekerja berdasarkan prediksi sebesar 46 orang sedangkan angkatan kerja yang tidak bekerja berdasarkan prediksi sebesar 39 orang. Untuk perhitungan akurasi, *sensitivity* dan *specificity* pada data *testing* maka diperoleh hasil berikut:

$$\text{Akurasi} = \frac{46+39}{135} \times 100\% = 62,96\%$$

$$\text{Sensitivity} = \frac{46}{72} \times 100\% = 63,89\%$$

$$\text{Specificity} = \frac{39}{63} \times 100\% = 61,90\%$$

Dari perhitungan yang dilakukan, nilai akurasi yang digunakan untuk mengukur tingkat ketepatan klasifikasi untuk data prediksi sebesar 62,96% maka dapat dikatakan bahwa pohon optimal yang terbentuk mampu mengklasifikasikan data baru sebesar 62,96%. Sedangkan untuk nilai *Sensitivity* yang digunakan untuk mengukur klasifikasi pada kelas angkatan kerja yang bekerja sebesar 63,89%. Dan *specificity* yang digunakan untuk mengukur ketepatan klasifikasi pada kelas angkatan kerja yang tidak bekerja sebesar 61,90%.

## D. Kesimpulan dan Saran

### 1. Kesimpulan:

Berdasarkan pembahasan pada penerapan metode CART (Classification and Regression Tree) pada klasifikasi tingkat partisipasi angkatan kerja di kota Medan tahun 2019:

- Kelompok pertama yaitu angkatan kerja dengan jenis kelamin perempuan dengan tingkat pendidikan sedang sebanyak 861 (33,58%, n=2564). Dimana jumlah angkatan kerja yang bekerja sebanyak 313 (36,35%, n=861) sedangkan angkatan kerja yang tidak bekerja sebanyak 548 (63,65%, n=861).

- Kelompok kedua yaitu angkatan kerja dengan jenis kelamin perempuan dengan tingkat pendidikan (tinggi dan rendah) sebanyak 472 (18,41%, n=2564). Dimana jumlah angkatan kerja yang bekerja sebanyak 226 (47,88%, n=472) sedangkan angkatan kerja yang tidak bekerja sebanyak 246 (52,12%, n=472).
- Kelompok ketiga yaitu angkatan kerja dengan jenis kelamin laki-laki dengan status pernikahan menikah sebanyak 740 (28,86%, n=2564). Dimana jumlah angkatan kerja yang bekerja sebanyak 593 (80,13%, n=740) sedangkan angkatan kerja yang tidak bekerja sebanyak 147 (19,86%, n=740).
- Kelompok keempat yaitu angkatan kerja jenis kelamin laki-laki dengan status pernikahan belum menikah yang memiliki tingkat pendidikan rendah sebanyak 48 (1,87%, n=2564). Dimana dari jumlah tersebut angkatan kerja yang bekerja sebanyak 33 (70,21%, n=48) sedangkan angkatan kerja yang tidak bekerja sebanyak 15 (29,79%, n=48).
- Kelompok kelima angkatan kerja jenis kelamin laki-laki status pernikahan menikah yang memiliki tingkat pendidikan tinggi sedang sebanyak 444 (17,32%, n=2564). Dimana dari jumlah tersebut angkatan kerja yang bekerja sebanyak 207 (46,74%, n=444) sedangkan angkatan kerja yang tidak bekerja sebanyak 237 (53,26%, n=444).

### 2. Saran

Berdasarkan pembahasan dalam penerapan metode CART (Classification and Regression Tree) pada klasifikasi tingkat partisipasi angkatan kerja di kota Medan tahun 2019 :

- Penelitian yang penulis lakukan hanya menggunakan metode CART (Classification and Regression Tree) dengan bantuan software Microsoft Excel dan SPSS. Kedepannya penulis berharap penelitian ini dapat dikembangkan dengan mengkombinasikan metode CART

dengan metode lain dan menggunakan software lainnya

2. Penelitian ini hanya berfokus pada klasifikasi tingkat partisipasi kerja di kota Medan berdasarkan kriteria angkatan kerja, jenis kelamin, tingkat pendidikan, status pernikahan dan pengalaman pelatihan kerja. Kedepan penulis berharap penelitian dapat dikembangkan dengan menambahkan kriteria penilaian lainnya sehingga proses klasifikasi menjadi lebih beragam

#### **E. Daftar Pustaka**

- Badan Pusat Statistik. 2018. "Kota Medan dalam Angka 2018".
- Breiman L., Friedman J.H Olshen R.A & Stone C.J. (1984). *Classification And Regression Tree*. New York, NY: Chapman And Hall
- Jatmiko, Y. (2019). "Analisis Perbandingan Kinerja CART Konvensional, Bagging, dan Random Forest pada Klasifikasi Objek: Hasil dari Dua Simulasi". Padjajaran: Media Statistika.
- Komariah, N. (2008). "Classification and Regression Tree (CART) analysis Pada Penderita Skizofrenia di RSJKO Soeprapto Daerah Bengkulu". Universitas Bengkulu: Jurusan Matematika Fakultas MIPA.
- 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
- Pratiwi dan Eka. (2014). "Klasifikasi Pengangguran Terbuka Menggunakan CART (*Classification and Regression Tree*) di Provinsi Sulawesi Utara". Surabaya: Jurnal Sains dan Seni POMITS Vol.3, No.1.
- Mardiani. (2012). "Penerapan Klasifikasi dengan Algoritma CART untuk Prediksi Kuliah Bagi Mahasiswa Baru". Palembang: STMIK MDP Jurusan Sistem Informasi.
- Nuriyah. (2013). "Perbandingan Metode *Chi-Square Automatic Interaction Detection (CHAID)* dan *Classification and Regression Tree (CART)* dalam Menentukan Klasifikasi Alumni UIN Sunan Kalijaga Berdasarkan Masa Studi". Yogyakarta: UIN Sunan Kalijaga Program Sarjana Fakultas Sains dan Teknologi.
- Sumartini, S H. (2015). "Penggunaan Metode Classification and Regression Trees (CART) untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya". Surabaya: Jurnal Sains dan Seni ITS Vol. 4, No.2.