

**PENERAPAN METODE CART UNTUK MENGLASIFIKASIKAN  
TINGKAT PENGANGGURAN TERBUKA  
DI PROVINSI SUMATERA BARAT TAHUN 2020**

**TUGAS AKHIR**

*Untuk Memenuhi Sebagian Persyaratan Memperoleh Gelar Ahli Madya*



Oleh  
**SUCCY INDAH SARI**  
NIM 18037073

**PROGRAM STUDI DIPLOMA III STATISTIKA  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2022**

**PERSETUJUAN UJIAN TUGAS AKHIR**

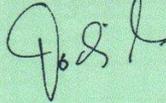
**PENERAPAN METODE CART UNTUK  
MENGKLASIFIKASIKAN TINGKAT PENGANGGURAN  
TERBUKA DI PROVINSI SUMATERA BARAT TAHUN 2020**

Nama : Succy Indah Sari  
NIM/Tahun Masuk : 18037073/2018  
Program Studi : DIII Statistika  
Jurusan : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 10 Februari 2022

Disetujui oleh:

Pembimbing Akademik



**Dodi Vionanda, M.Si. Ph.D**  
NIP. 19790611 200501 1 002

**PERSETUJUAN UJIAN TUGAS AKHIR**

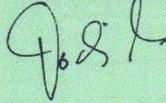
**PENERAPAN METODE CART UNTUK  
MENGKLASIFIKASIKAN TINGKAT PENGANGGURAN  
TERBUKA DI PROVINSI SUMATERA BARAT TAHUN 2020**

Nama : Sucey Indah Sari  
NIM/Tahun Masuk : 18037073/2018  
Program Studi : DIII Statistika  
Jurusan : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 10 Februari 2022

Disetujui oleh:

Pembimbing Akademik



**Dodi Vionanda, M.Si. Ph.D**  
NIP. 19790611 200501 1 002

## SURAT PERNYATAAN TIDAK PLAGIAT

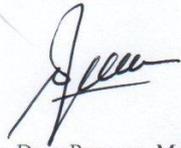
Saya yang bertanda tangan di bawah ini :

Nama : Succy Indah Sari  
NIM/TM : 18037073/2018  
Program Studi : DIII Statistika  
Jurusan : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, Tugas akhir saya dengan judul **“Penerapan Metode CART Untuk Mengklasifikasikan Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat Tahun 2020”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan. Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hokum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,  
Ketua Jurusan Statistika,



Dr. Dony Permana, M.Si.  
NIP. 19750127 200604 1 001

Saya yang menyatakan,



Succy Indah Sari  
NIM. 18037073

# **Penerapan Metode CART untuk Mengklasifikasikan Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat Tahun 2020**

**Succy Indah Sari**

## **ABSTRAK**

Masalah pengangguran sampai saat ini masih menjadi perhatian utama baik di negara maju maupun negara berkembang. Hal ini disebabkan terjadinya ketidakseimbangan jumlah penduduk angkatan kerja dengan jumlah penduduk yang bekerja. Masalah pengangguran terbuka semakin sulit karena tidak mengenal adanya tunjangan pengangguran. Akibatnya pengangguran cenderung berada dalam kemiskinan karena mereka tidak menerima penghasilan apapun serta dapat meningkatkan kriminalitas yang tinggi. Penelitian ini bertujuan untuk mengetahui model dari klasifikasi tingkat pengangguran terbuka dan faktor-faktor yang berpengaruh terhadap tingkat pengangguran terbuka di Provinsi Sumatera Barat tahun 2020.

Metode CART adalah salah satu metode atau algoritma dari teknik pohon keputusan. CART menghasilkan kumpulan data yang akurat sebagai karakteristik dalam mengklasifikasikan tingkat pengangguran terbuka. Jenis penelitian ini tergolong penelitian terapan. Variabel dalam penelitian ini terdiri dari variabel respons (Y) angkatan kerja dan variabel penjelas (X) yaitu jenis kelamin, pendidikan, usia, status dalam rumah tangga, pengalaman pelatihan kerja, status perkawinan dan tempat tinggal.

Hasil penelitian menggunakan metode CART menghasilkan sebuah diagram pohon yang terdiri dari 8 terminal node. 5 diantaranya memiliki label kelas bekerja dan 3 diantaranya memiliki label kelas tidak bekerja (pengangguran). Berdasarkan hasil terminal node tersebut diperoleh faktor-faktor yang mempengaruhi tingkat pengangguran terbuka di Provinsi Sumatera Barat Tahun 2020 yaitu status perkawinan, tempat tinggal, tingkat pendidikan, jenis kelamin, usia dan status dalam rumah tangga. Selanjutnya diperoleh persentase ketepatan model klasifikasi pada penelitian ini secara tepat sebesar 70,19%.

**Kata Kunci :** Tingkat Pengangguran Terbuka, Klasifikasi, Metode CART

## KATA PENGANTAR



Puji syukur peneliti ucapkan atas kehadiran Allah SWT, atas segala berkah, rahmat, hidayah, serta pertolongan yang diberikan-Nya kepada peneliti, sehingga peneliti bisa menyelesaikan Tugas Akhir ini dengan judul: **“Penerapan Metode CART untuk Mengklasifikasikan Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat Tahun 2020”**. Selanjutnya, shalawat beserta salam untuk nabi Muhammad SAW sebagai suri tauladan bagi seluruh umat.

Peneliti mengucapkan terimakasih kepada semua pihak yang telah memberi saran, serta ide-idenya sehingga tugas akhir ini dapat diselesaikan, yaitu:

1. Bapak Dodi Vionanda, M.Si, Ph.D pembimbing dan penasehat akademik sekaligus Koordinator Program Studi DIII Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang yang telah meluangkan waktunya dalam membimbing dan memberikan masukan-masukan yang bermanfaat bagi peneliti.
2. Bapak Dr. Syafriandi, M.Si dan Ibu Dina Fitria, M.Si dosen penguji.
3. Bapak Dr. Dony Permana, M.Si Ketua Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
4. Bapak/Ibu dosen Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
5. Teristimewa untuk orang tua dan adik yang telah mencurahkan kasih sayang, do'a, waktu dan dukungan kepada peneliti.
6. Semua sahabat terutama Fitri, Putri, Willy, Riri serta teman dan rekan-rekan Statistika 2018 yang telah banyak membantu dan memberi dukungan kepada

peneliti.

Semoga segala bimbingan, bantuan dan motivasi yang telah diberikan mendapat balasan oleh Allah SWT sebagai amal ibadah.

Dalam penyusunan Tugas Akhir ini peneliti telah berusaha semaksimal mungkin untuk memberikan yang terbaik, namun peneliti menyadari bahwa penulisan Tugas Akhir ini belum sempurna, karena keterbatasan pengetahuan dan kemampuan yang peneliti miliki. Untuk itu kritikan dan saran sangat diperlukan untuk penyempurnaan Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat bagi peneliti dan pembaca umumnya. Aamiin

Padang, Januari 2022

Succy Indah Sari

## DAFTAR ISI

<b>ABSTRAK</b> .....	i
<b>KATA PENGANTAR</b> .....	ii
<b>DAFTAR ISI</b> .....	iv
<b>DAFTAR TABEL</b> .....	vi
<b>DAFTAR GAMBAR</b> .....	vii
<b>DAFTAR LAMPIRAN</b> .....	viii
<b>BAB I PENDAHULUAN</b> .....	1
A. Latar Belakang Masalah .....	1
B. Batasan Masalah .....	6
C. Rumusan Masalah .....	6
D. Tujuan Penelitian .....	6
E. Manfaat Penelitian .....	7
<b>BAB II KAJIAN TEORI</b> .....	8
A. Angkatan Kerja .....	8
B. Faktor-faktor yang diduga mempengaruhi Tingkat Pengangguran Terbuka (TPT) .....	9
C. Metode CART .....	12
D. Rpart .....	18
E. Undersampling .....	19
F. Ketepatan Klasifikasi .....	20
<b>BAB III METODOLOGI PENELITIAN</b> .....	22
A. Jenis Penelitian .....	22
B. Jenis dan Sumber Penelitian .....	22
C. Populasi dan Sampel .....	22
D. Variabel Penelitian .....	23
E. Teknik Analisis Data .....	23
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	25
A. Deskripsi Data .....	25
B. Analisis Data .....	31
C. Pembahasan .....	34
<b>BAB V PENUTUP</b> .....	37
A. Kesimpulan .....	37
B. Saran .....	38

<b>DAFTAR PUSTAKA .....</b>	<b>39</b>
<b>LAMPIRAN.....</b>	<b>42</b>

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. Data Angkatan Kerja di Provinsi Sumatera Barat Tahun 2018-2020.....	2
2. Confusion Matrix .....	20
3. Variabel Penelitian .....	23
4. Proporsi Angkatan Kerja.....	32
5. Jumlah Proporsi Angkatan Kerja .....	33
6. Ketepatan Klasifikasi Pohon .....	34

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
1. Proses Undersampling.....	19
2. Deskripsi Status Angkatan .....	25
3. Kerja Berdasarkan Jenis Kelamin .....	26
4. Angkatan Kerja Berdasarkan Tingkat Pendidikan.....	27
5. Angkatan Kerja Berdasarkan Usia .....	28
6. Angkatan Kerja Berdasarkan Status dalam Rumah Tangga .....	29
7. Angkatan Kerja Berdasarkan Pengalaman Pelatihan Kerja.....	29
8. Angkatan Kerja Berdasarkan Status Perkawinan.....	30
9. Angkatan Kerja Berdasarkan Tempat Tinggal.....	31
10. Perbandingan Jumlah Angkatan Kerja Setelah Undersampling .....	32
11. Diagram Pohon Klasifikasi .....	33

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
1. Data Angkatan Kerja Provinsi Sumatera Barat Tahun 2020 .....	42
2. Karakteristik Responden Menggunakan Syntax Crosstab .....	43
3. Menampilkan Data dengan R.....	45
4. Output Proporsi Data Angkatan Kerja .....	46
5. Output rpart dan rpart.plot.....	47
6. Output Undersampling .....	48
7. Diagram Pohon CART Setelah Undersampling .....	49
8. Akurasi Prediksi .....	51

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang Masalah**

Masalah pengangguran sampai saat ini masih menjadi perhatian utama baik di negara maju maupun negara berkembang. Hal ini disebabkan terjadinya ketidakseimbangan jumlah penduduk angkatan kerja dengan jumlah penduduk yang bekerja. Menurut Suasti dan Sri (2009:185) angkatan kerja merupakan penduduk usia kerja (15 tahun keatas) yang memiliki pekerjaan, penduduk yang masih mencari pekerjaan ataupun penduduk yang tidak memiliki pekerjaan. Jika terjadi ketidakseimbangan antara jumlah penduduk angkatan kerja dengan jumlah penduduk yang bekerja hal ini akan menyebabkan adanya jumlah penduduk yang masih mencari pekerjaan atau disebut dengan pengangguran.

Tingkat pengangguran cenderung lebih tinggi pada negara berkembang. Salah satu masalah pengangguran yang perlu diselesaikan adalah masalah pengangguran terbuka. Menurut Susanto dan Pratama (2019) dalam penelitiannya pengangguran terbuka mengacu pada tenaga kerja yang berusaha mencari pekerjaan namun mereka tidak mendapatkan pekerjaan sama sekali sehingga mereka benar-benar menganggur.

Masalah pengangguran terbuka semakin sulit karena tidak mengenal adanya tunjangan pengangguran seperti Indonesia. Akibatnya pengangguran cenderung berada dalam kemiskinan karena mereka tidak menerima penghasilan apapun. Menurut Wulandari dan Marwan (2019) selain kemiskinan, pengangguran terbuka juga dapat menyebabkan tingkat kriminalitas yang lebih tinggi dan daya beli masyarakat yang lebih rendah.

Berdasarkan data Badan Pusat Statistik (BPS), di Pulau Sumatera Provinsi Sumatera Barat menduduki posisi ketiga tingkat pengangguran terbuka tertinggi setelah Provinsi Kepulauan Riau dan Provinsi Sumatera Utara. Tabel 1 menyajikan data angkatan kerja di Provinsi Sumatera Barat Tahun 2018-2020.

**Tabel 1. Data Angkatan Kerja di Provinsi Sumatera Barat Tahun 2018-2020**

Tahun	Jumlah Angkatan Kerja	Jumlah Bekerja	Jumlah Pengangguran	Tingkat Pengangguran
2018	2.629.104	2.480.405	148.699	5,66%
2019	2.684.389	2.540.040	144.349	5,38%
<b>2020</b>	<b>2.772.133</b>	<b>2.581.524</b>	<b>190.609</b>	<b>6,88%</b>

Sumber : *Badan Pusat Statistik*

Tabel 1 menunjukkan perkembangan jumlah penduduk angkatan kerja setiap tahunnya. Terlihat perkembangan jumlah penduduk angkatan kerja setiap tahunnya semakin meningkat. Seiring meningkatnya jumlah penduduk angkatan kerja, maka semakin tinggi ketidakseimbangan jumlah penduduk angkatan kerja dengan jumlah penduduk yang bekerja sehingga menyebabkan semakin tingginya tingkat pengangguran. Tingkat pengangguran tertinggi terjadi pada tahun 2020 dengan tingkat pengangguran sebesar 6,88%.

Kenaikan tingkat pengangguran pada tahun 2020 naik secara signifikan dari tahun sebelumnya, yaitu sebesar 1,5%. Jika dibandingkan pada tahun-tahun sebelumnya kenaikan tingkat pengangguran di Provinsi Sumatera Barat hanya berkisar 0,3%. Menurut Kementerian Keuangan Direktorat Jendral Perbendaharaan Provinsi Sumatera Barat, peningkatan tingkat pengangguran terbuka di Provinsi Sumatera Barat terjadi karena seiring dengan pandemi *Coronavirus Disease 2019* (COVID-19) dan pemberlakuan Pembatasan Sosial Berskala Besar (PSBB).

Penelitian sebelumnya yang terkait dengan tingkat pengangguran terbuka dilakukan oleh Wulandari (2021), metode statistika yang dipakai adalah metode

CHAID (*Chi-Square Automatic Interaction Detection*) dengan variabel independen yang digunakan jumlah kelamin, tingkat pendidikan, usia, status dalam rumah tangga, pengalaman pelatihan kerja, status perkawinan dan tempat tinggal. Penelitian ini disimpulkan bahwa variabel independen yang signifikan mempengaruhi tingkat pengangguran terbuka di Provinsi Sumatera Barat tahun 2020 adalah status perkawinan, tempat tinggal, pendidikan, umur, pelatihan kerja.

Penelitian lain juga dilakukan oleh Yulinda dkk (2020) menggunakan analisis statistika deskriptif tentang tingkat pengangguran terbuka (TPT) perempuan di Kabupaten Subang mengidentifikasi salah satu penyebab pengangguran adalah bonus demografi yang berdampak pada peningkatan usia kerja khususnya bagi perempuan. Pendapat serupa yang mendukung pernyataan di atas adalah Saskara dan Kaluge (2009) dalam penelitian mengenai analisis faktor yang mempengaruhi pengangguran perempuan. Penelitian ini menyatakan secara statistik, kaum perempuan menempati posisi yang kurang menguntungkan dalam berbagai aspek kehidupan dibandingkan dengan laki-laki salah satunya dalam mencari pekerjaan. Berdasarkan kedua pendapat ini secara langsung *gender*/jenis kelamin menjadi salah satu faktor penyebab terjadinya pengangguran.

Bersumber pada penelitian sebelumnya mengenai faktor yang memengaruhi tingkat pengangguran terbuka, maka didapatkan variabel dependen dan variabel independen. Variabel dependen yang diteliti untuk penelitian ini yaitu angkatan kerja yang berkategori bekerja dan tidak bekerja (pengangguran). Variabel independen yang diteliti untuk penelitian ini yaitu jenis kelamin, tingkat pendidikan, usia, status dalam rumah tangga, pengalaman pelatihan kerja, status perkawinan dan tempat tinggal.

Banyaknya faktor yang mempengaruhi tingkat pengangguran terbuka, maka pada penelitian ini akan digunakan metode klasifikasi berstruktur pohon untuk memodelkan tingkat pengangguran terbuka. Metode ini merupakan metode yang direkomendasikan untuk mengkaji hubungan antara variabel dependen dan variabel independen. Data yang dianalisis pada metode ini memiliki kompleksitas yang tinggi yaitu jumlah observasi yang besar atau jumlah peubah yang banyak.

Penelitian yang dilakukan oleh Becti dkk (2017) tentang perbandingan analisis pada pengaruh lokasi karakteristik konsumen dalam memilih minimarket dengan metode regresi logistik dan CART. Dari hasil penelitian di peroleh bahwa nilai akurasi pengklasifikasian pada model regresi logistik adalah 88% sedangkan dengan menggunakan CART diperoleh nilai akurasi sebesar 90%. Penelitian lainnya yang dilakukan oleh Suwardika (2017) tentang perbandingan ketepatan metode klasifikasi dalam mengelompokkan data hepatitis menggunakan metode *Support Vector Machine (SVM)*, *Classification and Regression Trees (CART)* dan regresi logistik biner. Berdasarkan perbandingan analisis diperoleh ketepatan metode CART memiliki hasil ketepatan klasifikasi yang lebih baik dari pada SVM dan regresi logistik biner.

Berdasarkan penelitian terdahulu, maka variabel dalam penelitian ini adalah angkatan kerja sebagai variabel dependen dan variabel independen adalah tingkat pendidikan, jenis kelamin, tempat tinggal, jenis pekerjaan, usia, status dalam rumah tangga, pengalaman pelatihan, dan status perkawinan. Berdasarkan variabel-variabel yang telah disebutkan dan analisis data yang telah dilakukan oleh penelitian terdahulu, untuk permasalahan dalam penelitian ini analisis yang cocok digunakan adalah metode CART. Menurut penelitian Becti dkk (2017) dan

Suwardika (2017) metode CART akurat untuk klasifikasi.

Menurut Kayri dan Kayri (2015) CART dikenal sebagai klasifikasi tunggal, adalah proses non-parametrik dan tidak memerlukan asumsi tentang bentuk fungsional data. Pendekatan non-parametrik tidak bergantung pada asumsi tertentu, sehingga memberikan lebih banyak fleksibilitas dalam analisis data. Namun, masih memiliki tingkat akurasi yang tinggi dan mudah diimplementasikan. Menurut Mardika dkk (2016) CART termasuk metode dengan pendekatan regresi nonparametrik yang digunakan untuk menghasilkan pohon keputusan. CART terdiri dari pohon regresi (*regression tree*) dan pohon klasifikasi (*classification tree*). Ketika peneliti menemukan pola numerik, maka menggunakan metode pohon regresi. Sedangkan jika pola kategorik menggunakan pohon klasifikasi. Pada penelitian ini, data berupa pola kategori maka hasil yang diperoleh berupa pohon klasifikasi.

Menurut Rahmi, Hazmira dan Hafifatul (2017) CART terbilang sederhana namun merupakan metode yang kuat. Tujuan CART yaitu menghasilkan kumpulan data yang akurat sebagai karakteristik klasifikasi. CART juga digunakan juga untuk menggambarkan hubungan antara variabel respon (dependen/tak bebas) dengan satu atau lebih variabel prediktor (independen/bebas).

Berdasarkan latar belakang tersebut, dalam penelitian ini akan dilakukan klasifikasi tingkat pengangguran terbuka menggunakan metode CART. Untuk itu penelitian ini akan dilakukan dengan judul **“Penerapan Metode CART untuk Mengklasifikasikan Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat Tahun 2020”**.

## **B. Batasan Masalah**

Batasan ruang lingkup dalam penelitian ini adalah pengelompokan faktor-faktor yang dapat mempengaruhi pengangguran terbuka yaitu usia, tingkat pendidikan, jenis kelamin, daerah tempat tinggal, status dalam rumah tangga, pengalaman pelatihan kerja dan status perkawinan.

## **C. Rumusan Masalah**

Berdasarkan permasalahan diatas, masalah yang dibahas dapat dirumuskan sebagai berikut :

1. Bagaimana model untuk memprediksi klasifikasi Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat Tahun 2020 menggunakan metode CART?
2. Faktor apa saja yang berpengaruh signifikan terhadap klasifikasi Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat Tahun 2020 dengan menggunakan metode CART?

## **D. Tujuan Penelitian**

Berdasarkan rumusan masalah diatas, maka diperoleh tujuan penelitian ini adalah sebagai berikut:

1. Mendapatkan model pada data klasifikasi Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat Tahun 2020.
2. Untuk mengetahui faktor apa saja yang berpengaruh signifikan terhadap klasifikasi Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat Tahun 2020 dengan menggunakan metode CART.

### **E. Manfaat Penelitian**

Berdasarkan uraian pada tujuan penelitian, maka manfaat penelitian adalah sebagai berikut :

1. Bagi penulis, dapat menambah pengetahuan tentang kemampuan dalam menerapkan metode statistik yang diperoleh
2. Bagi Pembaca, dapat digunakan sebagai referensi mengenai Penerapan Metode CART untuk Mengklasifikasikan Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat Tahun 2020.
3. Bagi pemerintah, dapat memberikan solusi mengenai bagaimana cara mengurangi tingkat pengangguran di Sumatera Barat dengan melihat acuan pada hasil klasifikasi yang paling mempengaruhi berdasarkan penelitian.

## **BAB II**

### **KAJIAN TEORI**

#### **A. Angkatan Kerja**

Menurut Suasti dan Sri (2009:185), angkatan kerja adalah penduduk usia kerja (15 tahun ke atas) yang memiliki pekerjaan, penduduk yang masih mencari pekerjaan, atau penduduk yang menganggur. Angkatan kerja berperan aktif dalam kegiatan ekonomi. Partisipasi penduduk dalam kegiatan ekonomi diukur dari penduduk yang termasuk dalam pasar kerja yaitu yang bekerja atau yang mencari pekerjaan.

Menurut Badan Pusat Statistik, angkatan kerja yang digolongkan bekerja dan tidak bekerja (pengangguran) adalah:

##### **1. Angkatan kerja yang digolongkan bekerja**

Bekerja adalah kegiatan ekonomi yang dilakukan oleh seseorang dalam seminggu terakhir dengan maksud membantu untuk mendapatkan atau menghasilkan pendapatan atau keuntungan selama setidaknya satu jam (secara terus menerus). Kegiatan ini juga mencakup kegiatan pekerja tidak dibayar yang mendukung kegiatan usaha/ekonomi.

Punya pekerjaan tetapi sementara tidak bekerja adalah suatu keadaan dimana seseorang mempunyai pekerjaan tetapi untuk sementara tidak bekerja karena berbagai alasan selama seminggu yang lalu, seperti: sakit, cuti, menunggu panen, mogok dan sebagainya.

##### **2. Angkatan kerja yang digolongkan Tidak bekerja (Pengangguran)**

Pengangguran terbuka terdiri dari mereka yang tak punya pekerjaan dan mencari pekerjaan, mereka yang tak punya pekerjaan dan mempersiapkan usaha,

mereka yang tak punya pekerjaan dan tidak mencari pekerjaan, karena merasa tidak mungkin mendapatkan pekerjaan, mereka yang sudah punya pekerjaan, tetapi belum mulai bekerja. Angkatan kerja yang digolongkan tidak bekerja (pengangguran) disebut dengan pengangguran terbuka.

## **B. Faktor-faktor yang diduga mempengaruhi Tingkat Pengangguran Terbuka (TPT)**

Tingkat pengangguran terbuka dapat dipengaruhi oleh banyak faktor. Berikut ini beberapa faktor yang diduga mempengaruhi tingkat pengangguran terbuka:

### 1. Jenis Kelamin

Menurut Hungu (2016:43) jenis kelamin adalah perbedaan biologis antara perempuan dan laki-laki sejak lahir. Perbedaan biologis dan fungsi biologis pria dan wanita tidak dapat dipertukarkan, dan fungsi tersebut tetap ada pada pria dan wanita dari semua ras di bumi. Menurut Saskara dan Kaluge (2009) jenis kelamin dikelompokkan menjadi perempuan dan laki-laki. Namun dalam perkembangannya, satu pihak lebih unggul dari pihak lainnya. Perempuan berada pada posisi yang kurang menguntungkan secara statistik dibandingkan laki-laki dalam berbagai bidang kehidupan, salah satunya adalah pencarian kerja. Pada fakta lapangan yang sering kita temui saat ini, banyak sekali tenaga kerja bagian lapangan pada umumnya didominasi oleh laki-laki, sedangkan pada bagian kantor suatu perusahaan pada umumnya dapat didominasi oleh laki-laki maupun perempuan. Peluang lebih banyak dalam hal pekerjaan yaitu untuk pihak laki-laki.

### 2. Tingkat pendidikan

Tingkat Pendidikan di dunia kerja merupakan salah satu acuan untuk memenuhi suatu jabatan atau standar untuk mencapai suatu jabatan di dunia kerja.

UU Sisdiknas 2003 No. 20 menyatakan tingkat pendidikan formal terbagi menjadi tiga jenjang pendidikan dimana pendidikan tersebut dimulai dari pendidikan dasar, pendidikan menengah serta pendidikan tinggi.

Berdasarkan undang-undang ini maka kategori tingkat pendidikan untuk penelitian ini dibagi menjadi tiga. Pertama, pendidikan dasar mencakup  $\leq$  SD. Kedua, pendidikan menengah mencakup SMP/SMK/MI. Ketiga, pendidikan tinggi mencakup perguruan tinggi atau dikategorikan dengan  $>$ SMA.

### 3. Usia

Usia adalah satuan yang mengukur waktu lamanya hidup atau ada suatu benda atau makhluk. Menurut Naf'an (2014:131) usia yang termasuk angkatan kerja yaitu berusia antara 15 keatas. Menurut Suasti dan Sri (2009:185) angkatan kerja merupakan penduduk usia kerja (15 tahun keatas) yang memiliki pekerjaan, penduduk yang masih mencari pekerjaan ataupun penduduk yang tidak memiliki pekerjaan.

Pembagian kelompok usia dapat dibedakan menjadi usia produktif dan usia tidak produktif. Usia produktif mencakup usia dengan rentang usia 15 tahun hingga 64 tahun. Sedangkan usia tidak produktif mencakup usia dengan rentang usia lebih dari 64 tahun.

### 4. Status dalam rumah tangga

Menurut Pratiwi dan Zain (2014) status dalam rumah tangga adalah kedudukan dan status yang dimiliki seseorang dalam keluarga atau rumah tangganya. Secara umum seorang suami berperan sebagai kepala keluarga yang bertugas mencari nafkah untuk memenuhi kebutuhan sandang, pangan dan papan. Namun, ada kalanya seorang istri dapat berperan sebagai kepala keluarga atas

alasan tertentu.

Menurut Yulianti dkk (2011) peluang kehilangan pekerjaan ketika seseorang menjadi kepala rumah tangga lebih kecil daripada peluang seseorang tidak menjadi kepala rumah tangga. Hal ini dikarenakan tanggung jawab dan kewajiban yang harus dilaksanakan sebagai kepala rumah tangga. Seorang kepala rumah tangga akan berusaha agar terus bekerja demi memenuhi kebutuhan keluarga.

#### 5. Pengalaman pelatihan kerja

Secara teori, keterampilan merupakan indikator kualitas produktivitas tenaga kerja, sehingga yang memiliki keterampilan lebih cepat menemukan pekerjaan dibandingkan yang tidak. Pengalaman dalam bentuk pelatihan profesional diperlukan untuk meningkatkan keterampilan tersebut. Menurut Pratiwi (2012) dari perspektif permintaan tenaga kerja, pengusaha cenderung memilih pekerja terdidik dan terlatih untuk meminimalkan biaya pelatihan yang dikeluarkan perusahaan saat melatih karyawan baru.

#### 6. Status Perkawinan

Menurut Undang – Undang Perkawinan Nomor 1 tahun 1974 Pasal 1, pernikahan adalah ikatan lahir batin antara seorang pria dengan seorang wanita sebagai suami istri dengan tujuan membentuk keluarga yang bahagia dan kekal berdasarkan Ketuhanan yang Maha Esa. Status perkawinan yang berstatus kawin memiliki pengangguran lebih rendah dibandingkan dengan wanita dengan status yang belum kawin. Hal ini mungkin disebabkan oleh adanya tanggungan untuk mencari nafkah setelah pernikahan.

Menurut Yulianti dkk (2011) sejalan dengan status dalam rumah tangga, seseorang yang belum kawin belum memiliki tanggung jawab baik secara

ekonomi maupun tanggung jawab dalam keluarga, namun sebaliknya seseorang yang sudah kawin memiliki tanggung jawab yang besar dan harus memiliki pekerjaan untuk menghidupi keluarga.

#### 7. Tempat Tinggal

Menurut Pratomo (2018) hubungan tempat tinggal dengan pengangguran terdapat pada 2 wilayah yaitu perkotaan dan perdesaan. Banyak orang pindah ke kota untuk mencari pekerjaan impian mereka menciptakan ketidakseimbangan antara penawaran dan permintaan tenaga kerja perkotaan. Penawaran yang jauh lebih banyak daripada permintaan tenaga kerja menyebabkan kota memiliki probabilitas lebih tinggi.

### C. Metode CART

#### 1. Pengertian CART

*Classification and Regression Trees* (CART) merupakan salah satu metode atau algoritma dari teknik pohon keputusan (*decision tree*). Metode CART mulai dikembangkan pada tahun 80-an oleh Breiman, Friedman, Olshen, Stone pada tahun 1984 merupakan metode klasifikasi yang menggunakan algoritma partisi secara biner (*binary recursive partitioning*). Menurut An (2019) CART menggambarkan hubungan antara variabel dependen dengan satu atau lebih variabel independennya sehingga menyajikan keputusan yang berbentuk pohon biner. Analisis ini menghasilkan kelompok data hierarkis mulai dari simpul akar (*root node*) dan diakhiri dengan sekelompok kecil pengamatan yang homogen.

Menurut Sumartini (2015) tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai karakteristik klasifikasi. Metode pengklasifikasian CART memiliki beberapa kelebihan. Pertama, CART

merupakan bentuk statistika nonparametrik sehingga memerlukan asumsi distribusi variabel prediktor dan uji hipotesis yang perlu dipenuhi. Kedua, CART dapat menggunakan sembarang kombinasi data kontinu/*numeric* dan kategorik. Ketiga, metode ini lebih mudah dipelajari dan pengambilan keputusan karena memberikan tampilan visual struktur data yang kompleks dan multivariabel. Keempat, hasil klasifikasi akhir mengklasifikasikan data baru secara sederhana dan efisien. Kelima, kemudahan interpretasi hasil.

## 2. Algoritma CART

Menurut Sumartini (2015) proses menganalisis metode CART untuk menghasilkan pohon klasifikasi disebut *Binary Recursive Partitioning*. Istilah *binary* berarti bahwa setiap simpul induk dibagi menjadi dua kelompok simpul anak. Istilah *recursive* mengacu pada fakta bahwa proses keputusan diulang. Dengan demikian, simpul induk yang membuat dua simpul anak dapat membuat simpul anak tambahan dengan memisahkan setiap simpul anak. Istilah *Partitioning* mengacu pada fakta bahwa kumpulan data di partisi menjadi beberapa bagian.

Menurut Tittonel dkk (2018) proses partisi rekursif biner menciptakan struktur pohon yang sangat besar, tetapi kembali ke pohon berukuran optimal berdasarkan tingkat kesalahan *relative* atau *misclassification error*. Tingkat kesalahan dapat ditentukan dengan validasi silang. Pohon yang terbentuk terdiri dari serangkaian simpul terminal yang mewakili jenis yang sama dari kelompok pengamatan variabel respons, dan *splitting node* yang mewakili variabel penjelas.

Beberapa tahapan metode CART dalam pembuatan pohon keputusan, yaitu sebagai berikut :

a. Pembentukan pohon keputusan

Proses pembentukan pohon keputusan terdiri atas 3 tahapan, yaitu :

1) Pemilihan Atribut/Pemilah

Data akan dipilah berdasarkan aturan pemilahan dan kriteria *goodness-of-split* dengan metode pemilah *Impurity  $i(t)$*  serta pemilihan pemilah tergantung pada jenis variabel responnya. Metode pemilihan pemilah menggunakan *Impurity  $i(t)$*  membantu menemukan fungsi pemilah yang optimal dengan mengukur derajat heterogenitas kelas pada simpul tertentu dalam pohon klasifikasi. Beberapa fungsi *Impurity measure  $i(t)$*  adalah sebagai berikut :

- a) Indeks Gini
- b) Indeks Informasi
- c) Indeks Twoing
- d) Indeks Entropi

Berdasarkan keempat fungsi *impurity  $i(t)$*  peneliti akan menggunakan fungsi indeks gini. Menurut Sumartini dan Purnami (2015) indeks Gini merupakan fungsi heterogen yang sangat mudah diterapkan pada berbagai kasus. Indeks Gini selalu mengisolasi kelas dengan anggota terbesar atau kelas terpenting dalam simpul tersebut terlebih dahulu. Atribut yang dipilih berdasarkan indeks Gini membentuk simpul atau kumpulan kelas yang disebut simpul atau *node*. Simpul akan selalu mengalami pemilihan secara rekursif sampai diperoleh *terminal nodes*.

Menurut Breiman dkk (1993) derajat *impurity* yang tinggi menunjukkan bahwa *node* tersebut belum homogen. Sementara itu, suatu node dengan derajat *impurity* yang rendah menunjukkan bahwa node tersebut sudah homogen. Apabila

class suatu objek dinyatakan dengan  $u$ ,  $u = 1; 2; \dots; m$ , dimana  $m$  adalah jumlah class untuk variabel dependen ( $y$ ). Misalkan Indeks Gini untuk mengukur impurity dari suatu partisi  $D$ , dirumuskan sebagai berikut :

$$Gini(D) = 1 - \sum_j P^2(j|t)$$

(Breiman dkk., 1993)

Keterangan :

Gini (D) : ukuran impurity dari partisi D

$j$  : jumlah class pada variabel  $y$

$P(j|t)$  : peluang class  $j$  pada node  $t$

Pembagian biner diperiksa dengan cara menjumlahkan ukuran *impurity* dari setiap partisi yang diperoleh dari pembagian tersebut. Misalkan pembagian yang dilakukan pada parameter  $A$  mempartisi  $D$  menjadi  $D_1$  dan  $D_2$ . Sehingga Indeks Gini dari pembagian pada parameter  $A$  tersebut adalah sebagai berikut :

$$Gini_A(D) = \frac{|D_1|}{|D|} \cdot Gini(D_1) + \frac{|D_2|}{|D|} \cdot Gini(D_2)$$

(Breiman dkk., 1993)

Keterangan :

$Gini_A(D)$  : impurity dari partisi  $D$  pada parameter  $A$

$Gini(D_1)$  : impurity dari partisi pertama  $D$

$Gini(D_2)$  : impurity dari partisi kedua  $D$

$D$  : partisi  $D$

$D_1$  : partisi pertama  $D$

$D_2$  : partisi kedua  $D$

Penurunan tingkat *impurity* yang dihasilkan dari suatu pembagian biner terhadap parameter  $A$  adalah sebagai berikut:

$$\Delta\text{Gini}(A) = \text{Gini}(D) - \text{Gini}_A(D)$$

(Breiman dkk., 1993)

Keterangan :

$\Delta\text{Gini}(A)$  : tingkat impurity

$\text{Gini}_A(D)$  : impurity dari partisi D pada parameter A

$\text{Gini}(D)$  : impurity dari partisi pertama D

Parameter yang memaksimalkan penurunan dari tingkat *impurity* tersebut yang akan dipilih sebagai *splitting parameter*.

## 2) Penentuan Simpul Terminal

Menurut Lemon dkk (2003:175) suatu simpul  $t$  akan menjadi simpul terminal atau tidak, akan dipilah kembali apabila pada simpul  $t$  tidak terdapat penurunan keheterogenan secara berarti atau adanya batasan minimum sebesar  $n$  seperti hanya terdapat satu pengamatan pada tiap simpul anak. 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.

## 3) Penandaan Label Kelas

Pelabelan kelas adalah proses mengidentifikasi setiap node dari kelas tertentu. Penandaan kelas dilakukan pada *terminal nodes*, *non terminal nodes*, dan *root nodes*. Namun, pelabelan paling dibutuhkan pada simpul terminal atau simpul nodes karena simpul ini penting digunakan untuk memprediksi objek dari kelas tertentu pada simpul *nodes* tersebut. Penandaan label kelas pada simpul terminal

diberi label menurut aturan bilangan tertinggi jika:

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

(Pratiwi dan Zain, 2014)

Keterangan :

$P(j|t)$  : Proporsi kelas  $j$  pada simpul  $t$

$N_j(t)$  : Jumlah Pengamatan kelas  $j$  pada simpul  $t$

$N(t)$  : Jumlah Pengamatan pada simpul  $t$

Label kelas simpul terminal  $t$  adalah  $j$  yang memberi nilai dugaan kesalahan pengklasifikasian simpul  $t$  terbesar. Proses pembentukan pohon klasifikasi berhenti apabila hanya ada satu pengamatan dalam tiap simpul anak atau adanya batas minimum  $n$ , semua pengamatan dalam simpul anak identik, dan adanya batasan jumlah level atau kedalaman pohon maksimal.

#### b. Pemangkasan Pohon Klasifikasi

Bagian pohon yang kurang penting dipangkas untuk mendapatkan pohon klasifikasi yang optimal. Pohon yang dibentuk dengan aturan pemilah dan kriteria *goodness of split* berukuran sangat besar karena penghentian pohon berdasarkan banyaknya amatan pada simpul terminal atau besarnya tingkat kehomogenan. Jika lebih banyak penyortiran/pemilahan dilakukan, ini dapat menyebabkan tingkat kesalahan prediksi yang rendah, tetapi sebagai hasilnya, pohon klasifikasi besar yang akan terbentuk. Pohon besar dapat menyebabkan *overfitting*, tetapi jika pengamatan pohon terbatas pada akurasi tertentu, *underfitting* dapat terjadi. Pemangkasan didasarkan pada perkiraan ukuran pohon tanpa mengorbankan akurasi dengan mengurangi node pada pohon untuk mendapatkan ukuran pohon relatif. Ukuran pemangkasan yang digunakan untuk mendapatkan ukuran pohon

yang sesuai adalah dengan *cost complexity minimum*.

$$R\alpha(T) = R(T) + \alpha|\tilde{T}|$$

(Gusriani dan Parmikanti, 2015)

Keterangan :

$R(T)$  : *Resubstitution estimate* (proporsi kesalahan pada sub pohon)

$\alpha$  : Kompleksitas parameter (*complexity parameter*)

$|\tilde{T}|$  : Ukuran banyaknya simpul terminal pohon T

### c. Penentuan Pohon Klasifikasi Optimal

Ukuran pohon yang terlalu besar akan menyebabkan nilai *cost complexity* yang tinggi karena struktur data yang digambarkan cenderung kompleks sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti yang cukup kecil. Bila  $R(T)$  dipilih sebagai penduga terbaik, maka akan cenderung dipilih pohon yang besar, sebab pohon yang semakin besar akan membuat nilai  $R(T)$  semakin kecil.

### D. Rpart

Rpart adalah singkatan dari *Recursive Partitioning and Regression Trees*. Menurut Sumartini (2015) istilah *recursive* mengacu pada fakta bahwa proses keputusan diulang. Oleh karena itu, simpul induk yang membuat dua simpul anak dapat membuat simpul anak tambahan dengan memisahkan setiap simpul anak. Istilah *Partitioning* mengacu pada fakta bahwa kumpulan data di partisi menjadi beberapa bagian. Istilah *Regression* mengacu pada hasil yang diperoleh berdasarkan nilai yang di input. Sedangkan *Trees* mengacu pada bentuk model yang dihasilkan berupa pohon. Rpart adalah paket yang disediakan pada perangkat lunak (*software*) R untuk menyelesaikan masalah klasifikasi.

Penggunaan fungsi *rpart* dapat diuraikan sebagai berikut:

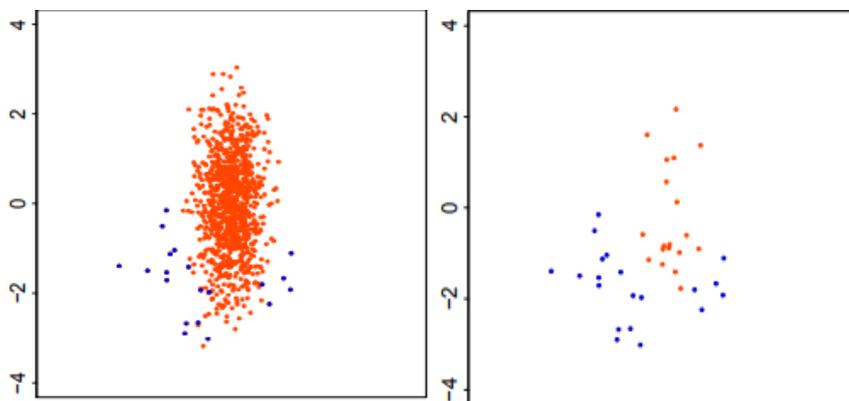
```
rpart (formula, data, weights, subset, na.action = na.rpart, method, model=  
FALSE, x = FALSE, y = TRUE, parms, control, cost, ...)
```

Fungsi *rpart.control* diperlukan untuk menentukan beberapa parameter awal dalam membangun *Classification Trees* dengan memakai fungsi *rpart* pada paket *rpart*.

```
rpart.control(minsplit = 20, minbucket = round(minsplit/3), cp = 0.01,  
maxcompete = 4, maxsurrogate = 5, usesurrogate = 2, xval =  
10, surrogatestyle = 0, maxdepth = 30, ...)
```

### E. Undersampling

*Undersampling* merupakan salah satu penyelesaian masalah ketidakseimbangan data (*imbalance*). Menurut Lunardo dkk (2014) proses *undersampling* sendiri adalah proses pengambilan kelas mayoritas secara acak (tanpa pengembalian) sehingga kelas mayoritas memiliki proporsi yang sama dengan kelas minoritas. Proses *undersampling* ditunjukkan pada Gambar 1 sebagai berikut :



**Gambar 1. Proses Undersampling**

Sumber : Lunardon, N., dkk

Pada metode CART, perlu dilakukan penanganan untuk mengatasi data yang tidak seimbang (*imbalance*). Hal ini dikarenakan data yang tidak seimbang (*imbalance*) akan menunjukkan prediksi yang buruk. Undersampling dilakukan apabila data yang diperoleh pada kelas minoritas sangat kecil dan kelas mayoritas yang sangat besar.

#### F. Ketepatan Klasifikasi

Hasil dari proses klasifikasi akan diperoleh kelas yang benar dan salah. Untuk menghitung ketepatan ketepatan klasifikasi (akurasi) dapat disajikan dalam bentuk tabel *confusion matrix*. *Confusion matrix* merupakan suatu tabel yang berisi jumlah data uji yang salah diklasifikasikan dan jumlah data uji yang benar diklasifikasikan (Johnson dan Wichern, 2007). *Confusion matrix* untuk klasifikasi ditunjukkan pada Tabel 2 sebagai berikut:

**Tabel 2. Confusion Matrix**

Klasifikasi Kelas	Kelas Prediksi	
	1	0
1	TP	FN
0	FP	TN

Dari tabel diperoleh nilai akurasi menggunakan persamaan sebagai berikut :

$$\text{Akurasi} = \left( \frac{TP+TN}{TP+FN+FP+TN} \right) \times 100\% \quad (7)$$

(Johnson dan Wichern, 2007)

Keterangan :

TP : Jumlah subjek dari variabel Y kategori 1 yang tepat diprediksikan sebagai variabel Y kategori 1

FN : Jumlah subjek dari variabel Y kategori 0 yang salah diprediksikan sebagai variabel Y kategori 1

FP : Jumlah subjek dari variabel Y kategori 1 yang salah diprediksikan sebagai variabel Y kategori 0

TN : Jumlah subjek dari variabel Y kategori 0 yang tepat diprediksikan sebagai variabel Y kategori 0

## BAB V

### PENUTUP

#### A. Kesimpulan

Berdasarkan analisis yang telah dilakukan, maka dapat diperoleh kesimpulan sebagai berikut :

1. Faktor – faktor yang mempengaruhi tingkat pengangguran terbuka di Provinsi Sumatera Barat Tahun 2020 secara signifikan adalah status perkawinan, tempat tinggal, tingkat pendidikan, umur, jenis kelamin dan status dalam rumah tangga.
2. Node yang dihasilkan pada diagram pohon klasifikasi diperoleh sebanyak 8 terminal node dengan label kelas sebagai berikut :
  - a. *Node 1* memiliki label kelas bekerja apabila tingkat pendidikan responden  $\leq$ SD atau SMP-SMA dengan tempat tinggal perdesaan dan status perkawinan sudah pernah kawin.
  - b. *Node 2* memiliki label kelas bekerja berjenis kelamin laki-laki, tingkat pendidikan  $>$ SMA dengan tempat tinggal perdesaan dan status perkawinan sudah pernah kawin.
  - c. *Node 3* memiliki label kelas tidak bekerja (pengangguran) berjenis kelamin perempuan, tingkat pendidikan  $>$ SMA dengan tinggal di perdesaan dan status perkawinan sudah pernah kawin.
  - d. *Node 4* memiliki label bekerja apabila responden berusia kategori usia tidak produktif ( $>$ 64 tahun) dengan tempat tinggal di perkotaan dan status perkawinan sudah pernah kawin.
  - e. *Node 5* memiliki label kelas bekerja apabila responden tingkat pendidikan

>SMA atau  $\leq$ SD dengan kategori usia produktif (15-64 tahun), bertempat tinggal di perkotaan dengan status perkawinan sudah pernah kawin.

f. *Node 6* memiliki label kelas bekerja apabila responden memiliki status dalam rumah tangga sebagai kepala keluarga, berpendidikan tingkat menengah (SMP-SMA), kategori usia adalah usia produktif (15-64 tahun) dengan tempat tinggal di perkotaan dan status perkawinan sudah pernah kawin.

g. *Node 7* memiliki label kelas tidak bekerja (pengangguran) apabila responden memiliki status dalam rumah tangga bukan sebagai kepala keluarga, berpendidikan tingkat menengah (SMP-SMA), kategori usia adalah usia produktif (15-64 tahun) dengan tempat tinggal di perkotaan dan status perkawinan sudah pernah kawin.

h. *Node 8* memiliki label kelas tidak bekerja (pengangguran) dengan responden status perkawinan belum pernah kawin.

## **B. Saran**

Adapun saran dalam penelitian ini adalah sebagai berikut :

1. Karena nilai akurasi yang diperoleh tidak cukup besar, maka bagi peneliti selanjutnya disarankan menggunakan metode klasifikasi yang lain, antara lain CHAID, SVM, Random Forest dan metode klasifikasi lainnya agar nilai akurasi yang diperoleh lebih besar.
2. Untuk penelitian selanjutnya, dapat dipertimbangkan penambahan faktor-faktor yang berpengaruh terhadap tingkat pengangguran terbuka.

## DAFTAR PUSTAKA

- AN, A. N. (2019). Penerapan Metode CART (Classification And Regression Tree) pada Tingkat Partisipasi Angkatan Kerja di Kota Makassar (Doctoral dissertation, Universitas Islam Negeri Alauddin Makassar).
- Badan Pusat Statistik Sumatera barat. (2020). Data Jumlah Angkatan Kerja, Jumlah Tenaga Kerja, Jumlah Pengangguran dan Tingkat Pengangguran Tahun 2018-2020.
- Bekti, Rokhana D. (2017). Analisis Pengaruh Lokasi Karakteristik Konsumen dalam Memilih Minimarket dengan Metode Regresi Logistik dan CART. *Media Statistika* 10 (2), 119-130.
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1993). *Classification and Regression Trees*. Chapman And Hall: New York.
- Gusriani, N., & Parmikanti, K. (2015). Klasifikasi Ketepatan Masa Studi Mahasiswa FMIPA Unpad Angkatan 2001-2006 dengan Menggunakan Metode Classification and Regression Trees (CART). *Jurnal Matematika Integratif*, 11(1), 7-14.
- Hungu. (2007): *Demografi Kesehatan Indonesia*. Jakarta, Grasindo.
- Johnson, R.A. dan Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis. Sixth Edition*. Pearson Education Inc: USA.
- Kayri M, Kayri I. (2015). The Comparison of Gini and Twoing Algorithms in Terms of Predictive Ability and Misclassification Cost in Data Mining : An Empirical Study. *International Journal of Computer Trends and Technology (IJCTT)*, 27 (1), 21-30
- Lemon, SC, Roy, J., Clark, MA, Friedmann, PD, & Rakowski, W. (2003). Klasifikasi dan analisis pohon regresi dalam kesehatan masyarakat: tinjauan metodologis dan perbandingan dengan regresi logistik. *Sejarah kedokteran perilaku* , 26 (3), 172-181.
- Lunardon. N., Menardi G, and Torelli. N., (2014). ROSE: A Package for Binary Imbalanced Learning. *The R Journal* Vol. 6/1.
- Mardika Z.W, Mukid M.A, Yasin H. (2016). Pembentukan pohon klasifikasi biner dengan algoritma classification and regression trees : Studi kasus: kredit macet di PD. BPR-BKK Purwokerto Utara). *Jurnal Gaussian*, 5(3), 583-592
- Naf'an. (2014). *Ekonomi Makro Tinjauan Ekonomi Syariah*. Yogyakarta : Graha Ilmu.