

SKRIPSI

**PERBANDINGAN ALGORITMA C5.0 DAN ALGORITMA
CLASSIFICATION AND REGRESSION TREE (CART) PADA
DECISION TREE UNTUK KLASIFIKASI DATA**

(Studi Kasus : Tingkat Pengangguran Terbuka Indonesia Menurut
Kabupaten/Kota Tahun 2020)



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

RIFDA NIDA'UL LABIBAH

NIM. 19106010017

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2023

**PERBANDINGAN ALGORITMA C5.0 DAN ALGORITMA
CLASSIFICATION AND REGRESSION TREE (CART) PADA
DECISION TREE UNTUK KLASIFIKASI DATA**

(Studi Kasus : Tingkat Pengangguran Terbuka Indonesia Menurut
Kabupaten/Kota Tahun 2020)

Skripsi

Diajukan untuk memenuhi salah satu syarat memperoleh derajat

Sarjana Matematika



Diajukan oleh

RIFDA NIDA'UL LABIBAH

NIM. 19106010017

kepada

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2023



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp :-

Kepada

Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Rifda Nida'ul Labibah

NIM : 19106010017

Judul Skripsi : *Perbandingan Algoritma C5.0 dan Algoritma Classification and Regression Tree (CART) pada Decision Tree untuk Klasifikasi Data (Studi Kasus : Tingkat Pengangguran Terbuka Indonesia Menurut Kabupaten/ Kota Tahun 2020)*

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 17 Maret 2023

Pembimbing I

Dr. Epha Diana Supandi, S.Si., M.Sc.

NIP. 19750912 200801 2 105

Pembimbing II

Muhamad Rashif Hilmi, S.Si., M.Sc.

NIP. 19920309 202012 1 001



PENGESAHAN TUGAS AKHIR

Nomor : B-930/Un.02/DST/PP.00.9/04/2023

Tugas Akhir dengan judul : PERBANDINGAN ALGORITMA C5.0 DAN ALGORITMA CLASSIFICATION AND REGRESSION TREE (CART) PADA DECISION TREE UNTUK KLASIFIKASI DATA (STUDI KASUS : TINGKAT PENGANGGURAN TERBUKA INDONESIA MENURUT KABUPATEN /KOTA TAHUN 2020)

yang dipersiapkan dan disusun oleh:

Nama : RIFDA NIDA' UL LABIBAH
Nomor Induk Mahasiswa : 19106010017
Telah diujikan pada : Kamis, 30 Maret 2023
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang
Dr. Epha Diana Supandi, S.Si., M.Sc.
SIGNED

Valid ID: 642a45fcbcb81d



Penguji I
Muhamad Rashif Hilmi, S.Si., M.Sc.
SIGNED

Valid ID: 642a2a9677954



Penguji II
Dr. Sugiyanto, S.Si., ST., M.Si.
SIGNED

Valid ID: 64253bf349a70



Yogyakarta, 30 Maret 2023
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 642a5f4dd356e

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Rifda Nida'ul Labibah
NIM : 19106010017
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 20 Maret 2023



Rifda Nida'ul Labibah

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PERSEMBAHAN

Skripsi ini saya persembahkan sebagai ungkapan terima kasih kepada dua orang paling berharga dalam hidup saya yaitu Bapak dan Mama yang selalu memanjatkan doa untuk saya dalam setiap sujudnya serta senantiasa mengisi kehidupan saya dengan kebahagiaan dan kasih sayang yang tak ternilai harganya. Semua kerja keras dan perjuangan saya dalam menyusun skripsi ini semata hanya untuk menghadirkan senyum dan kebahagiaan untuk mereka.



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

“It’s not always easy. It’s okay if it’s taking more time than you thought. Just keep going. Sabar, jangan lupa bahagia, dikuat-kuatin aja.”



KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Allah SWT yang senantiasa melimpahkan rahmat dan hidayah-Nya sehingga penulis diberikan kekuatan, kemudahan, kelancaran, dan kesehatan dalam menyelesaikan penulisan skripsi yang berjudul “Perbandingan Algoritma C5.0 dan Algoritma *Classification and Regression Tree* (CART) pada *Decision Tree* untuk Klasifikasi Data (Studi Kasus : Tingkat Pengangguran Terbuka Indonesia Menurut Kabupaten/Kota Tahun 2020)” sebagai syarat untuk memperoleh gelar Sarjana Matematika (S.Mat.) serta sebagai perwujudan partisipasi selama melaksanakan kuliah di Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.

Penulis menyadari bahwa selama proses penulisan skripsi ini tidak terlepas dari kontribusi dan bantuan dari beberapa pihak. Oleh karena itu, dalam kesempatan ini penulis ingin mengucapkan banyak terima kasih kepada pihak – pihak terkait yaitu kepada :

1. Allah SWT Sang Maha Pemberi Petunjuk yang telah memberikan kemudahan di setiap langkah yang penulis tempuh dalam menyelesaikan skripsi ini.
2. Ibu Dr. Hj. Khurul Wardati, M.Si. selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Bapak Muchammad Abrori S.Si., M.Kom. selaku Ketua Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Ibu Dr. Epha Diana Supandi S.Si., M.Sc. selaku dosen pembimbing 1 dan Bapak Muhamad Rashid Hilmi, S.Si., M.Sc. selaku dosen pembimbing 2 serta dosen penguji 1 yang telah meluangkan waktu untuk memberikan bimbingan, arahan, masukan, dan ilmu dalam penulisan skripsi ini.
5. Bapak Dr. Sugiyanto, S.Si., ST., M.Si. selaku dosen penguji 2 yang telah meluangkan waktu untuk memberi masukan dan ilmu dalam penulisan ini.
6. Bapak Muhammad Zainal Muhtadin dan Ibu Siti Ngaesaroh yang tak pernah melewatkan nama penulis disetiap doanya, senantiasa memberikan dukungan moril dan materil serta kasih sayang yang tiada henti kepada penulis.

7. Kakak-kakak, Mas Aziz dan Mbak Laras yang selalu memberikan bantuan, dukungan, dan semangat kepada penulis.
8. Sahabat “bertiga” yang selalu menjadi pendengar, penghibur, dan penyemangat bagi penulis.
9. Teman – teman seperjuangan di Matematika 2019 dan KKN 108 Sumur yang telah memberi motivasi serta menjadi tempat bertukar pikiran yang sehat dan inspiratif.
10. Seluruh pihak yang tidak dapat disebutkan satu persatu, terima kasih atas doa dan dukungan yang telah diberikan.

Yogyakarta, 30 Maret 2023

Penulis



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN SKRIPSI/TUGAS AKHIR.....	ii
HALAMAN PENGESAHAN.....	iii
SURAT PERNYATAAN KEASLIAN.....	iv
PERSEMBAHAN.....	v
MOTTO.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xv
DAFTAR LAMPIRAN.....	xvi
DAFTAR LAMBANG DAN SINGKATAN.....	xvii
INTISARI.....	xix
ABSTRACT.....	xx
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Metode Penelitian.....	5
1.6. Manfaat Penelitian.....	6
1.7. Tinjauan Pustaka.....	6
1.8. Sistematika Penulisan.....	10

BAB II LANDASAN TEORI	12
2.1. Data	12
2.1.1 Pengertian dan Jenis – Jenis Data	12
2.1.2 Skala Pengukuran Data	14
2.1.3 Macam – Macam Tipe Data	16
2.2. Statistika	17
2.2.1 Statistika Non Parametrik	17
2.2.2 Analisis Korelasi	19
2.3. Data Mining	20
2.3.1 Klasifikasi Data	22
2.4. Decision Tree	24
2.4.1 Algoritma C5.0	27
2.4.2 Algoritma Classification and Regression Tree (CART)	30
2.5. Pengertian dan Jenis – Jenis Pengangguran	32
2.6. Indeks Pembangunan Manusia	34
2.7. Kepadatan Penduduk	35
2.8. Zona Risiko COVID-19	36
2.9. Tingkat Pendidikan	37
BAB III METODE PENELITIAN	39
3.1. Jenis dan Sumber Data	39
3.2. Metode Pengumpulan Data	39
3.3. Variabel Penelitian	39
3.4. Metode Analisis Data	40
3.5. Alat Pengolah Data	42
BAB IV PEMBAHASAN	43

4.1	Algoritma C5.0.....	43
4.2	Algoritma Classification and Regression Tree (CART)	46
4.3	Contoh Perhitungan Sederhana Algoritma C5.0 dan Algoritma CART	53
BAB V STUDI KASUS		74
5.1	Pembagian Data Training dan Data Testing	74
5.2	Deskripsi Data	74
5.3	Algoritma C5.0.....	80
5.4	Algoritma CART.....	93
5.5	Perbandingan Tingkat Akurasi Hasil Klasifikasi	108
5.5.1	Akurasi Model Algoritma C5.0.....	108
5.5.2	Akurasi Model Algoritma Classification and Regression Tree ..	109
5.5.3	Perbandingan Tingkat Akurasi Algoritma C5.0 dan CART.....	109
5.6	Implementasi Metode Terbaik untuk Memprediksi Data	109
BAB VI PENUTUP		111
6.1	Kesimpulan.....	111
6.2	Saran.....	112
DAFTAR PUSTAKA		113
LAMPIRAN		116

DAFTAR TABEL

Tabel 1.1. Pemetaan tinjauan pustaka	6
Tabel 2.1. Tipe data R	17
Tabel 2.2. Contoh <i>confusion matrix</i>	23
Tabel 3.1. Variabel penelitian	40
Tabel 4.1. Data kualitas produk Toko A	53
Tabel 4.2. Distribusi frekuensi data contoh dengan C5.0.....	55
Tabel 4.3. Hasil perhitungan <i>node</i> pertama data contoh C5.0.....	58
Tabel 4.4. Hasil perhitungan cabang pada <i>node</i> 2 data contoh C5.0.....	59
Tabel 4.5. Pemilahan variabel kondisi dengan CART	61
Tabel 4.6. Pemilahan variabel harga dengan CART	62
Tabel 4.7. Pemilahan variabel diskon dengan CART	62
Tabel 4.8. Pemilahan variabel penjualan dengan CART	63
Tabel 4.9. Hasil perhitungan indeks gini untuk <i>node</i> 1 dengan CART.....	63
Tabel 4.10. Hasil perhitungan indeks gini untuk <i>node</i> 4 dengan CART.....	64
Tabel 5.1 : Distribusi frekuensi data <i>training</i>	75
Tabel 5.2. Koefisien korelasi variabel X terhadap Y.....	79
Tabel 5.3. Distribusi frekuensi data TPT dengan C5.0	80
Tabel 5.4. Hasil perhitungan pembentukan <i>parent node</i> dengan C5.0.....	85
Tabel 5.5. Keterangan setiap simpul pada pohon maksimal C5.0.....	88
Tabel 5.6. Tingkat pengangguran terbuka berdasarkan tingkat pendidikan perguruan tinggi yang rendah dengan C5.0.....	88
Tabel 5.7. Tingkat pengangguran terbuka berdasarkan tingkat pendidikan perguruan tinggi yang tinggi dan kepadatan penduduk dengan C5.0	89

Tabel 5.8. Tingkat pengangguran terbuka berdasarkan tingkat pendidikan perguruan tinggi yang tinggi, kepadatan penduduk, dan zona COVID-19	90
Tabel 5.9. Tingkat pengangguran terbuka berdasarkan tingkat pendidikan perguruan tinggi yang tinggi, kepadatan penduduk, dan zona COVID-19	91
Tabel 5.10. Pemilahan IPM kemungkinan pertama	93
Tabel 5.11. Pemilahan IPM kemungkinan kedua.....	94
Tabel 5.12. Pemilahan kepadatan penduduk kemungkinan pertama.....	94
Tabel 5.13. Pemilahan kepadatan penduduk kemungkinan kedua.....	95
Tabel 5.14. Pemilahan zona COVID-19 kemungkinan pertama	96
Tabel 5.15. Pemilahan zona COVID-19 kemungkinan kedua	96
Tabel 5.16. Pemilahan zona COVID-19 kemungkinan ketiga	97
Tabel 5.17. Pemilahan zona COVID-19 kemungkinan keempat	97
Tabel 5.18. Pemilahan zona COVID-19 kemungkinan kelima	98
Tabel 5.19. Pemilahan zona COVID-19 kemungkinan keenam	98
Tabel 5.20. Pemilahan zona COVID-19 kemungkinan ketiga	99
Tabel 5.21. Pemilahan tingkat pendidikan \leq SD	99
Tabel 5.22. Pemilahan tingkat pendidikan SMP - SMA kemungkinan pertama	100
Tabel 5.23. Pemilahan tingkat pendidikan perguruan tinggi kemungkinan pertama	101
Tabel 5.24. Hasil perhitungan indeks gini untuk <i>parent node</i>	101
Tabel 5.25. Keterangan setiap simpul pada pohon maksimal CART.....	103
Tabel 5.26. Penandaan label kelas.....	104
Tabel 5.27. Tingkat pengangguran terbuka berdasarkan IPM dan tingkat pendidikan perguruan tinggi dengan CART.....	105
Tabel 5.28. Tingkat pengangguran terbuka berdasarkan IPM dengan CART	106

Tabel 5.29. Tingkat pengangguran terbuka berdasarkan tingkat pendidikan perguruan tinggi dan zona COVID-19 dengan CART	106
Tabel 5.30. <i>Confusion matrix</i> algoritma C5.0	108
Tabel 5.31. <i>Confusion matrix</i> algoritma CART	109
Tabel 5.32. Hasil prediksi tingkat pengangguran terbuka 5 kabupaten/kota tahun 2023	110



DAFTAR GAMBAR

Gambar 2.1. Ilustrasi <i>decision tree</i>	24
Gambar 2.2. Ilustrasi pohon klasifikasi CART (Sumartini 2015).....	30
Gambar 3.1. <i>Flowchart</i> atau Diagram alur penelitian	41
Gambar 4.1. Ilustrasi pohon klasifikasi dengan enam kelas.....	47
Gambar 4.2. Pembentukan cabang <i>parent node</i> data contoh dengan C5.0	58
Gambar 4.3. Hasil pohon maksimal data contoh C5.0	60
Gambar 4.4. Hasil pembentukan cabang <i>parent node</i> data contoh CART.....	64
Gambar 4.5. Hasil pohon maksimal data contoh CART	65
Gambar 5.1. <i>Pie chart</i> persentase TPT Indonesia 2020	75
Gambar 5.2. Hasil pembentukan cabang <i>parent node</i> dengan C5.0.....	86
Gambar 5.3. Hasil pembentukan pohon maksimal dengan C5.0.....	87
Gambar 5.4. Hasil pembentukan cabang <i>parent node</i> dengan CART.....	102
Gambar 5.5. Hasil pembentukan pohon maksimal dengan CART.....	103

DAFTAR LAMPIRAN

Lampiran 1. Data Tingkat Pengangguran Terbuka Indonesia Menurut Kabupaten/Kota Tahun 2020.....	116
Lampiran 2. Data <i>Testing</i> dengan Metode <i>Cross Fold Validation</i>	117
Lampiran 3: <i>Syntax</i> dan Output <i>Software R</i>	118



DAFTAR LAMBANG DAN SINGKATAN

- F_{ij} : Banyak data pada kelas i yang diprediksi menjadi kelas j
- S : Himpunan kasus
- K : Jumlah kelas pada variabel/atribut A
- p_i : Proporsi dari sampel S_i terhadap himpunan S
- A : Variabel atau atribut
- N : Jumlah kategori/partisi pada variabel/atribut A
- $|S_i|$: Jumlah kasus pada kategori ke- i
- $|S|$: Jumlah kasus dalam S
- $gini(T_1)$: Nilai indeks gini subset T_1 pada setiap variabel
- $gini(T_2)$: Nilai indeks gini subset T_2 pada setiap variabel
- b : Banyak data pada variabel T
- b_1 : Banyak data pada T_1
- b_2 : Banyak data pada T_2
- $X_{(i,j)}$: Indeks komponen ke- i dari daerah j
- $X_{(i-\min)}$: Nilai minimum dari X_i
- $X_{(i-\max)}$: Nilai maksimum dari X_i
- $freq(C_i, S)$: Banyaknya kasus pada himpunan S yang masuk ke dalam kelas C_i
- $|T_i|$: Banyaknya partisi atau subset pada himpunan T
- $|T|$: Jumlah kasus dalam himpunan T

$p(j T)$: Proporsi dari kasus milik kelas j pada T
T_L	: <i>Node</i> atau simpul kiri
T_R	: <i>Node</i> atau simpul kanan
p_L	: Proporsi banyaknya objek terhadap T_L
p_R	: Proporsi banyaknya objek terhadap T_R
A	: Label kelas
L	: Banyaknya kategori pada suatu variabel
$N(T)$: Total kasus
$N_i(T)$: Jumlah kasus kelas i pada T
$\pi(i)$: Probabilitas kelas i yang akan ditampilkan pada pohon
$R(T)$: kesalahan pengklasifikasian berdasarkan <i>node</i> T
$R(T_R)$: kesalahan pengklasifikasian berdasarkan <i>node</i> T sisi kanan
$R(T_L)$: kesalahan pengklasifikasian berdasarkan <i>node</i> T sisi kiri
j_0	: Label <i>node</i> terminal
Y	: Tingkat Pengangguran Terbuka
X_1	: Indeks Pembangunan Manusia
X_2	: Kepadatan Penduduk
X_3	: Zona COVID-19
X_4	: Tingkat Pendidikan Rendah
X_5	: Tingkat Pendidikan Menengah
X_6	: Tingkat Pendidikan Tinggi

INTISARI

PERBANDINGAN ALGORITMA C5.0 DAN ALGORITMA *CLASSIFICATION AND REGRESSION TREE (CART)* PADA *DECISION TREE* UNTUK KLASIFIKASI DATA

Oleh

RIFDA NIDA'UL LABIBAH

NIM. 19106010017

Klasifikasi data adalah salah satu bagian penting dalam analisis data yang bertujuan untuk menemukan kesamaan karakteristik data yang akan dikelompokkan dalam suatu kelas. Salah satu metode untuk melakukan klasifikasi data adalah metode *decision tree* atau pohon keputusan. Untuk membentuk *decision tree* ada berbagai algoritma yang bisa digunakan diantaranya adalah C5.0 dan *Classification and Regression Tree (CART)*. Algoritma C5.0 menggunakan perhitungan nilai *gain* dan *gain ratio*. Sedangkan Algoritma CART menggunakan perhitungan indeks gini. Algoritma C5.0 dan CART diterapkan pada studi kasus tingkat pengangguran terbuka Indonesia menurut kabupaten/kota tahun 2020. Diperoleh hasil tingkat akurasi dengan algoritma C5.0 sebesar 64.009% sedangkan tingkat akurasi dengan algoritma CART sebesar 60.019%. Sehingga dapat dikatakan algoritma C5.0 adalah metode terbaik untuk mengklasifikasikan tingkat pengangguran terbuka di Indonesia dibandingkan dengan algoritma CART.

Kata kunci : Klasifikasi, *Decision Tree*, Algoritma C5.0, Algoritma *Classification and Regression Tree (CART)*, Tingkat Pengangguran Terbuka

ABSTRACT

COMPARISON OF C5.0 ALGORITHM AND CLASSIFICATION AND REGRESSION TREE (CART) ALGORITHM ON DECISION TREE FOR DATA CLASSIFICATION

by

RIFDA NIDA'UL LABIBAH

NIM. 19106010017

Data classification is an important part of data analysis which aims to find similarities in data characteristics that will be grouped into a class. One method for classifying data is the decision tree method. To form a decision tree there are various algorithms that can be used including C5.0 and Classification and Regression Tree (CART). The C5.0 algorithm uses the calculation of the gain value and gain ratio. While the CART algorithm uses the gini index calculation. The C5.0 and CART algorithms were applied to a case study of Indonesia's open unemployment rate by district/city in 2020. The results obtained were an accuracy rate with the C5.0 algorithm of 64,009% while the accuracy rate with the CART algorithm was 60,019%. So that it can be said that the C5.0 algorithm is the best method for classifying the open unemployment rate in Indonesia compared to the CART algorithm.

Keywords: Classification, Decision Tree, C5.0 Algorithm, Classification and Regression Tree (CART) Algorithm, Open Unemployment Rat

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi terjadi selaras dengan berkembangnya data. Sehingga adanya sebutan Era *Big Data* untuk menggambarkan zaman sekarang adalah hal yang tepat karena semua kegiatan di dunia ini selalu menghasilkan data sehingga data bertambah dengan pesatnya seiring berjalannya waktu. Untuk mengatur dan mengolah data-data tersebut diperlukan sebuah teknik atau metode yang disebut *data mining* atau analisis data. Menurut Likmund analisis data dibagi menjadi tiga tingkat berdasarkan jumlah variabel yang terlibat yaitu analisis univariat, analisis bivariat, dan analisis multivariat. Penelitian ini akan menggunakan jenis analisis data multivariat yaitu analisis yang melibatkan lebih dari dua variabel. Namun menurut Hair, pengertian analisis multivariat tidak sesederhana itu. Menurutnya, analisis multivariat tidak hanya membahas banyaknya variabel yang terlibat, namun juga membahas ada tidaknya interrelasi antar variabel dan efek masing-masing variabel yang cukup sulit untuk diinterpretasikan (Simamora 2005).

Klasifikasi data adalah salah satu bagian penting dalam proses *data mining* atau analisis data yang tujuannya adalah untuk menemukan kesamaan karakteristik data yang akan dikelompokkan dalam suatu kelompok atau kelas. Dari proses klasifikasi tersebut nantinya dapat diperoleh model atau fungsi yang dapat menjelaskan perbedaan konsep antar kelompok atau kelas data. Salah satu kegunaan dari model dan fungsi tadi adalah dapat memperkirakan data di masa depan. Klasifikasi sendiri sering digunakan untuk mengatasi permasalahan di berbagai bidang seperti ekonomi, pendidikan, kesehatan, kependudukan dll. Contoh penerapan di bidang ekonomi adalah untuk menentukan target pemasaran pelanggan karena dalam beberapa kasus dijumpai beberapa variabel dapat digunakan untuk memprediksi minat beli pelanggan. Contoh di bidang pendidikan adalah untuk menentukan klasifikasi mahasiswa berdasarkan lama masa studinya.

Contoh di bidang kesehatan sendiri ada banyak seperti memprediksi diagnosis penyakit medis. Sedangkan contoh di bidang kependudukan adalah mengklasifikasi data penduduk berdasarkan faktor penyebab atau kriteria kelas, salah satunya adalah klasifikasi data pengangguran berdasarkan faktor-faktor yang mempengaruhinya.

Dari penjelasan sebelumnya diketahui bahwa klasifikasi data bisa diterapkan di banyak bidang dengan tujuan akhir yang hampir sama yaitu memprediksi data di masa depan. Walaupun pada intinya klasifikasi mempunyai fungsi yang sama di berbagai bidang, perlu diperhatikan pula bahwa dalam memilih metode yang digunakan haruslah disesuaikan dengan kebutuhan masing-masing bidang. Klasifikasi data memiliki berbagai cara atau metode, diantaranya adalah klasifikasi dengan analisis diskriminan dan klasifikasi dengan *decision tree*. Perbedaan yang mendasar dari kedua metode klasifikasi tersebut adalah ada tidaknya asumsi yang perlu dipenuhi. Analisis diskriminan termasuk metode statistika parametrik karena ada beberapa asumsi seperti multivariat, tidak adanya multikolinearitas antar variabel independen, homogenitas varians antar kelompok dan asumsi lainnya yang harus dipenuhi. Sedangkan metode *decision tree* sendiri termasuk metode statistika non parametrik dimana tidak ada asumsi distribusi variabel independen yang harus dipenuhi dan pada penelitian ini akan menggunakan metode *decision tree*. *Decision tree* atau pohon keputusan dapat dibentuk menggunakan berbagai jenis algoritma seperti ID3, C4.5, C5.0, *Classification and Regression Tree (CART)*, dll. Penelitian ini akan membandingkan dua algoritma pada *decision tree* yaitu algoritma C5.0 dan algoritma CART.

Kehidupan manusia sebagai makhluk ekonomi yang bermoral menandakan bahwa manusia tidak bisa lepas dari keinginannya untuk memenuhi kebutuhan hidupnya. Dengan kemampuan dan keterbatasan yang dimiliki, secara tidak langsung keinginan tersebut mendorong manusia untuk melakukan aktivitas yang disebut dengan bekerja. Indonesia merupakan negara dengan jumlah penduduk terbesar keempat di dunia. Hasil Sensus Penduduk SP2020 periode 2010-2020 menunjukkan bahwa jumlah penduduk Indonesia saat ini sebesar 270,20 juta jiwa atau bertambah sebesar 32,56 juta jiwa dibandingkan dengan hasil Sensus

Penduduk SP2010 periode 2000-2010. Dari ratusan juta jiwa penduduk Indonesia, jumlah angkatan kerja penduduk Indonesia per Februari 2022 adalah sebanyak 144,01 juta jiwa. Yang termasuk angkatan kerja bukan hanya orang yang memiliki pekerjaan saja, termasuk didalamnya orang yang tidak memiliki pekerjaan atau sering disebut dengan pengangguran.

Tingkat pengangguran merupakan salah satu indikator penting dalam pertumbuhan ekonomi suatu negara. Karena tolak ukur keberhasilan pembangunan ekonomi suatu negara dapat dilihat dari Indeks Pembangunan Manusia (IPM) yang menggambarkan kualitas hidup manusia dan didalamnya meliputi peningkatan terhadap pemenuhan kebutuhan pokok untuk memiliki kehidupan yang lebih baik. Untuk memenuhi kebutuhan pokok sehari-hari, masyarakat harus bekerja dan memiliki pendapatan. Dari penjelasan tersebut dapat diketahui bahwa tingkat pengangguran yang tinggi dapat menghambat pertumbuhan ekonomi yang dapat berdampak pada pembangunan nasional. Menurut data Badan Pusat Statistik (BPS) ada sebanyak 7,7 juta dari 270,20 juta penduduk Indonesia yang tidak terlibat dalam aktivitas ekonomi nasional karena pengangguran. Ada banyak faktor penyebab seseorang menjadi pengangguran. Fenomena yang dimulai pada tahun 2019 yaitu Pandemi COVID-19 adalah salah satu hal yang membuat banyak orang kehilangan pekerjaannya dan menjadi pengangguran. Menurut BPS tingkat pengangguran terbuka Indonesia pada tahun 2020 mengalami kenaikan sebesar 1,84 persen dibandingkan tahun sebelumnya. Sehingga penelitian ini dilakukan untuk mengetahui bagaimana klasifikasi tingkat pengangguran terbuka di Indonesia saat masa Pandemi COVID-19 tahun 2020.

Penelitian ini akan membahas penggunaan metode algoritma C5.0 dan algoritma CART beserta pemilihan model terbaiknya untuk dilakukan klasifikasi terhadap tingkat pengangguran terbuka di Indonesia berdasarkan beberapa variabel yang dianggap mempunyai pengaruh terhadap tingkat pengangguran terbuka di Indonesia.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah di atas, dapat diperoleh rumusan masalah yang akan diselesaikan dalam penelitian ini adalah :

1. Bagaimana cara kerja algoritma C5.0 dan algoritma CART dalam melakukan klasifikasi tingkat pengangguran terbuka di Indonesia?
2. Bagaimana hasil *decision tree* yang terbentuk dari algoritma C5.0 dan algoritma CART?
3. Bagaimana tingkat akurasi hasil algoritma C5.0 dan algoritma CART dalam mengklasifikasikan data tingkat pengangguran terbuka di Indonesia?
4. Metode atau algoritma manakah yang paling baik digunakan untuk klasifikasi data tingkat pengangguran terbuka di Indonesia?

1.3. Batasan Masalah

Masalah yang akan diselesaikan pada penelitian ini terbatas pada variabel yang digunakan. Dimana variabel yang digunakan antara lain adalah indeks pembangunan manusia, zona COVID-19, kepadatan penduduk, dan tingkat pendidikan. Data yang digunakan adalah data keadaan angkatan kerja di Indonesia yang diambil dari situs resmi Badan Pusat Statistik masing-masing provinsi. Penelitian ini juga terbatas pada *Software* yang digunakan yaitu *Rstudio*. Dan analisis pada penelitian ini hanya akan melakukan proses klasifikasi dan prediksi data tingkat pengangguran terbuka di Indonesia.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dari penelitian ini adalah sebagai berikut :

1. Untuk mengetahui cara kerja algoritma C5.0 dan algoritma CART dalam melakukan klasifikasi tingkat pengangguran terbuka di Indonesia.
2. Untuk mengetahui hasil *decision tree* yang terbentuk dari algoritma C5.0 dan algoritma CART.

3. Untuk mengetahui tingkat akurasi hasil dari algoritma C5.0 dan algoritma CART dalam mengklasifikasikan data tingkat pengangguran terbuka di Indonesia.
4. Untuk mengetahui metode atau algoritma mana yang paling baik digunakan untuk klasifikasi data tingkat pengangguran terbuka di Indonesia.

1.5. Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah studi literatur seperti buku, jurnal, dan media informasi lainnya yang dapat digunakan sebagai referensi dalam penelitian ini. Adapun langkah-langkah yang akan ditempuh dalam melakukan penelitian adalah sebagai berikut :

1. Tahap Persiapan

Pada tahap ini peneliti melakukan observasi awal untuk mengetahui gambaran permasalahan yang ada serta untuk menentukan narasumber atau informan yang tepat dan sesuai dengan tema penelitian tugas akhir. Selain melakukan observasi, pada tahap persiapan ini peneliti juga melakukan penyusunan proposal penelitian.

2. Pengumpulan Data

Pada tahap ini peneliti kembali melakukan observasi setelah observasi awal dengan mengumpulkan dokumentasi dan mengumpulkan data atau dokumen yang diperlukan secara mendalam untuk menjawab permasalahan dalam penelitian ini.

3. Analisis Data

Tahap ini diawali dengan peneliti melakukan analisis awal dengan data yang sudah diperoleh dan dirasa sudah cukup lengkap. Apabila dalam tahap analisis awal ditemukan adanya kekurangan data maka peneliti akan kembali ke mencari data yang diperlukan. Pada tahap analisis data akan diperoleh kesimpulan akhir dari penelitian yang nantinya kesimpulan tersebut akan dikembangkan bentuk penyajiannya untuk dilakukan penyusunan laporan akhir.

4. Penyusunan Laporan Penelitian

Pada tahap terakhir ini peneliti akan menyajikan hasil penelitian yang sebelumnya sudah dianalisis dalam bentuk laporan. Laporan ini nantinya akan diujikan dan ditinjau oleh penguji serta dilakukan revisi apabila ada kesalahan atau kekurangan pada laporan tersebut.

1.6. Manfaat Penelitian

Manfaat dari penelitian ini adalah :

- a. Memberikan referensi tambahan dalam ilmu matematika di bidang statistika mengenai algoritma C5.0 dan algoritma *Classification and Regression Tree* (CART).
- b. Hasil analisis yang diperoleh dengan klasifikasi dapat digunakan oleh pemerintah untuk dijadikan sebagai acuan dalam menyusun kebijakan bagi pengangguran sesuai dengan klasifikasinya. Hasil analisis ini juga bisa digunakan untuk dijadikan referensi utama untuk orang lain yang sedang melakukan penelitian terkait tingkat pengangguran.

1.7. Tinjauan Pustaka

Berikut adalah pemetaan tinjauan pustaka yang digunakan sebagai referensi penelitian ini.

Tabel 1.1. Pemetaan tinjauan pustaka

No	Peneliti, Institusi (Tahun)	Objek Peneliti	Metode	Hasil
1.	Apriani Candra Wijaya; Nelly Astuti Hasibuan; Putri Ramadhani, STMIK Budi Darma (2018)	Pendapatan Masyarakat Kelurahan Mesjid Kecamatan Medan Kota	Klasifikasi data dengan Algoritma C5.0	Pohon keputusan dari kasus yang diangkat menunjukkan bahwa ada beberapa atribut yang mempengaruhi penentuan masyarakat yang

Tabel 1.1. Lanjutan

No	Peneliti, Institusi (Tahun)	Objek Peneliti	Metode	Hasil
				mendapatkan BLT. Atribut yang memiliki nilai <i>gain</i> paling tinggi adalah atribut jumlah anggota keluarga sehingga atribut tersebut menjadi <i>parent node</i> .
2.	Dede Wintana; Hikmatulloh; Nurul Ichsan; Jajang Jaya Purnama; Ami Rahmawati, Universitas Bina Sarana Informatika (2019)	Penerima Manfaat Keluarga Harapan (PKH) Desa Sukamaju, Kecamatan Kadudampit	Klasifikasi data dengan Algoritma C5.0	Pohon keputusan yang dihasilkan oleh teknik klasifikasi Algoritma C5.0 pada kasus ini memiliki akar utama yaitu atribut yang memiliki nilai <i>gain</i> tertinggi. Atribut tersebut adalah kepemilikan anak sekolah dengan perolehan nilai <i>gain</i> sebesar 0,512716784.
3.	Reni Pratiwi; Memi Nor Hayati; Surya Prangga, Universitas Mulawarman (2020)	Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan	Klasifikasi data dengan Algoritma C5.0 dan Algoritma <i>Classification</i>	Rata- rata tingkat akurasi algoritma C5.0 sebesar 79,17%, sedangkan tingkat akurasi CART 84,63%. Atau dengan kata lain Algoritma CART

Tabel 1.1. Lanjutan

No	Peneliti, Institusi (Tahun)	Objek Peneliti	Metode	Hasil
		Muara Ancalong Tahun 2019	<i>and</i> <i>Regression</i> <i>Tree (CART)</i>	lebih baik untuk klasifikasi data rata-rata pendapatan masyarakat Desa Teluk Baru Kecamatan Muara Ancalong tahun 2019 dibandingkan dengan Algoritma C5.0.
4.	Risma Yulistiani; Nanda Cahaya Putra; Qahtan Said; Iin Ernawati, Universitas Pembangunan Veteran Jakarta (2020)	Tingkat Pengangguran Terbuka Indonesia	Klasifikasi data dengan Algoritma <i>Classification</i> <i>and</i> <i>Regression</i> <i>Tree (CART)</i>	Berdasarkan fungsi <i>recall</i> , <i>precision</i> , <i>accuracy</i> dan <i>nilai error</i> , pohon keputusan yang dihasilkan memiliki tingkat akurasi sebesar 91,17%. Pola yang ditemukan menunjukkan atribut yang paling berpengaruh adalah rata- rata lama sekolah dengan nilai <i>gain</i> sebesar 96,87%.
5.	Rifda Nida'ul Labibah, UIN Sunan Kalijaga (2022)	Tingkat Pengangguran Terbuka	Klasifikasi data dengan Algoritma C5.0 dan Algoritma <i>Classification</i>	Akan dibandingkan tingkat akurasi antara Algoritma C5.0 dan Algoritma CART untuk klasifikasi data tingkat pengangguran terbuka

Tabel 1.1. Lanjutan

No	Peneliti, Institusi (Tahun)	Objek Peneliti	Metode	Hasil
			<i>and Regression Tree (CART)</i>	Indonesia menurut kabupaten/kota tahun 2020

Penelitian dengan menggunakan metode klasifikasi algoritma C5.0 pernah dilakukan oleh Apriani Candra Wijaya; Nelly Astuti Hasibuan; Putri Ramadhani dari STMIK Budi Darma pada tahun 2018 yang menggunakan Algoritma C5.0 untuk mengklasifikasi data pendapatan masyarakat Kelurahan Mesjid Kecamatan Medan Kota dan diperoleh hasil bahwa Pohon keputusan dari kasus yang diangkat menunjukkan bahwa ada beberapa atribut yang mempengaruhi penentuan masyarakat yang mendapatkan BLT. Atribut yang memiliki nilai *gain* paling tinggi adalah atribut jumlah anggota keluarga sehingga atribut tersebut menjadi *parent node*.

Penelitian dengan Algoritma C5.0 juga pernah dilakukan oleh Dede Wintana; Hikmatulloh; Nurul Ichsan; Jajang Jaya Purnama; Ami Rahmawati dari Universitas Bina Sarana Informatika pada tahun 2019 yang melakukan klasifikasi data dengan Algoritma C5.0 untuk menentukan penerima manfaat program keluarga harapan (PKH) di Desa Sukamaju Kecamatan Kadudampit dan diperoleh hasil bahwa atribut yang memiliki nilai *gain* tertinggi adalah kepemilikan anak sekolah dengan perolehan nilai *gain* sebesar 0,512716784.

Selain itu, penelitian yang membahas tentang perbandingan metode C5.0 dan CART pernah dilakukan oleh Reni Pratiwi dkk (2020) yang membandingkan metode C5.0 dan CART untuk mengklasifikasi data sosial kepala keluarga masyarakat Desa Teluk Baru Kecamatan Muara Ancalong tahun 2019 dan diperoleh kesimpulan bahwa tingkat akurasi metode CART lebih tinggi dari metode C5.0.

Sedangkan penelitian tentang klasifikasi tingkat pengangguran terbuka pernah dilakukan oleh Risma Yulistiani; Nanda Cahaya Putra; Qahtan Said; Iin Ernawati dari Universitas Pembangunan Veteran Jakarta tahun 2020 yang melakukan klasifikasi dan prediksi tingkat pengangguran terbuka di Indonesia menggunakan Algoritma CART dan diperoleh hasil bahwa berdasarkan fungsi *recall*, *precision*, *accuracy* dan *nilai error*, pohon keputusan yang dihasilkan memiliki tingkat akurasi sebesar 91,17%. Pola yang ditemukan menunjukkan atribut yang paling berpengaruh adalah rata-rata lama sekolah dengan nilai *gain* sebesar 96,87%.

Penelitian ini dilakukan oleh Rifda Nida'ul Labibah yang akan membahas perbandingan antara algoritma C5.0 dengan algoritma CART untuk klasifikasi data tingkat pengangguran terbuka di Indonesia menurut kabupaten/kota tahun 2020.

1.8. Sistematika Penulisan

Berikut ini adalah sistematika penulisan yang berisi gambaran mengenai pembahasan setiap bab dalam penelitian ini :

BAB I PENDAHULUAN

Bab ini berisi uraian mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, manfaat penelitian, tinjauan pustaka dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi uraian mengenai teori-teori yang digunakan dalam pembahasan penulisan penelitian ini.

BAB III METODE PENELITIAN

Bab ini berisi uraian mengenai metode penelitian yang meliputi pengumpulan data, metode analisis data, alat dan bahan yang digunakan untuk pengolahan data.

BAB IV PEMBAHASAN

Bab ini berisi uraian mengenai metode algoritma C5.0 dan metode *Classification and Regression Tree* (CART).

BAB V STUDI KASUS

Bab ini berisi uraian mengenai perbandingan metode algoritma C5.0 dan metode *Classification and Regression Tree* (CART) yang diaplikasikan pada klasifikasi data tingkat pengangguran terbuka Indonesia menurut kabupaten/kota tahun 2020.

BAB VI PENUTUP

Bab ini berisi kesimpulan yang menjawab rumusan masalah di atas serta saran dari penulis kepada pembaca.

BAB VI

PENUTUP

6.1 Kesimpulan

Berdasarkan hasil pembahasan dan analisis yang dilakukan pada studi kasus, dapat diperoleh kesimpulan sebagai berikut :

1. Cara Kerja Algoritma C5.0 dan Algoritma CART

Kedua algoritma menerapkan proses perhitungan berulang untuk melakukan pemilahan pada setiap *node* atau simpul dan prosesnya akan berhenti ketika tidak ada lagi proses pemilahan. Nilai yang dihitung pada algoritma C5.0 adalah *gain ratio* sedangkan pada algoritma CART menggunakan nilai indeks gini.

2. Hasil *Decision Tree*

Hasil *decision tree* yang terbentuk dari Algoritma C5.0 menjelaskan bahwa variabel yang berpengaruh terhadap tingkat pengangguran terbuka di Indonesia adalah kepadatan penduduk, zona risiko COVID-19, dan tingkat pendidikan perguruan tinggi dimana tingkat pendidikan tinggi memiliki pengaruh yang paling kuat karena terpilih menjadi *parent node*. Model ini menghasilkan lima kelas tingkat pengangguran terbuka di Indonesia berdasarkan ketiga variabel yang mempengaruhi tersebut.

Hasil *decision tree* yang terbentuk dari Algoritma CART menjelaskan bahwa variabel yang berpengaruh terhadap tingkat pengangguran terbuka di Indonesia adalah nilai IPM, zona risiko COVID-19, dan tingkat pendidikan tinggi dimana nilai IPM memiliki pengaruh yang paling kuat karena terpilih menjadi *parent node*. Model ini menghasilkan empat kelas tingkat pengangguran terbuka di Indonesia berdasarkan ketiga variabel yang mempengaruhi tersebut.

3. Tingkat Akurasi

Dari hasil perhitungan tingkat akurasi menggunakan tabel *confusion matrix*, menunjukkan ketepatan klasifikasi pada data testing dengan Algoritma C5.0 adalah sebesar 64.009% yang artinya pohon maksimal yang terbentuk dari Algoritma C5.0 dapat mengklasifikasikan data tingkat pengangguran terbuka di Indonesia sebanyak 64.009%. Sedangkan tingkat akurasi pohon dari Algoritma CART adalah sebesar 60.019% atau dapat dikatakan bahwa pohon maksimal yang terbentuk dari Algoritma CART dapat mengklasifikasikan data tingkat pengangguran terbuka di Indonesia sebanyak 60.019%.

4. Metode Terbaik

Hasil tingkat akurasi Algoritma C5.0 dan Algoritma CART berturut-turut adalah 64.009% dan 60.019%. Sehingga dapat dikatakan bahwa Algoritma C5.0 merupakan metode yang terbaik untuk mengklasifikasikan tingkat pengangguran terbuka Indonesia menurut kabupaten/kota tahun 2020 dibandingkan dengan Algoritma CART.

6.2 Saran

Berdasarkan hasil analisis klasifikasi yang telah dilakukan, saran yang dapat penulis sampaikan adalah pada penelitian selanjutnya, penulis berharap peneliti selanjutnya dapat menggunakan metode lain untuk mendapatkan tingkat akurasi yang lebih baik dalam mengklasifikasikan tingkat pengangguran terbuka Indonesia. Dapat juga dilakukan penelitian lebih lanjut mengenai kedua algoritma tersebut untuk diterapkan pada kasus lain atau dilakukan perbandingan dengan algoritma lain.

DAFTAR PUSTAKA

- Arifin, Danung. 2020. "Pengelompokan Kriteria Risiko COVID-19 Di Daerah Berdasarkan Zonasi Warna." *Badan Nasional Penanggulangan Bencana*.
- Breiman, Leo, Jerome H Friedman, Richard A Olshen, and Charles J Stone. 1998. *Classification and Regression Trees*. New York: Chapman and Hall.
- Finaka, Andean W. 2019. "Indeks Pembangunan Manusia Terus Meningkatkan."
- Gatiningsih dan Sutrisno, Eko. 2017. *Kependudukan Dan Ketenagakerjaan*. Sumedang: Fakultas Manajemen Pemerintahan IPDN.
- Hamidah, Ida. 2019. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer "Aplikasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus : Jurusan Teknik Komputer-UNIKOM)." Universitas Komputer Indonesia.
- Hutasoit, Imelda. 2017. "Pengantar Ilmu Kependudukan." : 192.
- Jansson, Johan. 2016. "Decision Tree Classification of Products Using C5.0 and Prediction of Workload Using Time Series Analysis." 30 HP Stockholm, Sverige.
- Kintamani DH, Ida. 2011. "Analisis Sumber Daya Manusia Pendidikan Tinggi." *Jurnal Pendidikan dan Kebudayaan* 17(4): 404–18.
- Mardiana, Linda, Dadan Kusnandar, and Neva Satyahadewi. 2022. "Analisis Diskriminan Dengan K Fold Cross Validation Untuk Klasifikasi Kualitas Air Di Kota Pontianak." *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya* 11(1): 97–102.
- Mardika, Zulfa Wahyu et al. 2016. "Pembentukan Pohon Klasifikasi Biner Dengan Algoritma CART (Classification And Regression Trees)(Studi Kasus: Kredit Macet Di PD. BPR-BKK Purwokerto Utara)." *Jurnal Gaussian* 5(3): 583–92.
- Marhaeni, Harmawanti, Sri Yati, and Bambang Tribudhi. 2007. *Indeks*

Pembangunan Manusia 2006-2007. Jakarta: Badan Pusat Statistik.

Nuryadi, Tutut Dewi Astuti, Endang Sri Utami, and M. Budiantara. 2017. *Dasar-Dasar Statistik Penelitian*. Bantul: Sibuku Media.

Pratiwi, Reni, Memi Nor Hayati, and Surya Prangga. 2020. “Perbandingan Klasifikasi Algoritma C5.0 Dengan Classification and Regression Tree (Studi Kasus : Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019).” *BAREKENG: Jurnal Ilmu Matematika dan Terapan* 14(2): 273–84.

Quinlan, J R. 1993. *5 Morgan Kaufmann C4.5: Programs For Machine Learning*. California: Morgan Kaufmann Publishers.

Ratih, Ariadni. 2015. “Implementasi Metode Pohon Keputusan Untuk Klasifikasi Data Dengan Nilai Fitur Yang Tidak Pasti.” (Computer Science): 1–7.

Ratniasih, Ni Luh. 2016. “Konversi Data Training Tentang Pemilihan Kelas Menjadi Bentuk Pohon Keputusan Dengan Teknik Klasifikasi.” *Eksplora Informatika*: 145–54.

Rokach, Lior; Maimon, Oded. 2014. *Data Mining With Decision Trees Theory and Applications*. 2nd ed. Singapore.

Rosidi, Muhammad. 2020. “Panduan Lengkap Analisis Statistika Menggunakan R Commander.” https://bookdown.org/moh_rosidi2610/Rcmdr-book/tipe-dan-struktur-data.html.

Santoso, Singgih. 2010. *Statistik Non Parametrik Konsep Dan Aplikasi Dengan SPSS*. Jakarta: PT Alex Media Komputindo.

Saputra, Nurihsan Detrinawansa. 2021. 14 *Frontiers in Neuroscience* “Penggunaan Metode Classification and Regression Tree (CART) Dalam Mengklasifikasikan Pasien Penderita DBD Di Rumah Sakit Anwar Makkatutu Kabupaten Bantaeng.”

Sari, Afrianti Yusghina. 2020. 2 *Jurnal Ekonomi* Volume 18, Nomor 1 Maret 201

“Penerapan Metode Classification and Regression Trees Pada Klasifikasi Kelayakan Peserta BPJS-PBI Di Kelurahan Meranti Pandak Kota Pekanbaru.”

Setio, Panji Bimo Nugroho, Dewi Retno Sari Saputro, and Bowo Winarno. 2020. “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5.” *PRISMA, Prosiding Seminar Nasional Matematika 3*: 64–71.

Simamora, Bilson. 2005. *Analisis Multivariat Pemasaran*. Jakarta: PT Gramedia Pustaka Utama.

Sumartini, Siti Holis. 2015. “Penggunaan Metode Classification and Regression Trees (CART) Untuk Klasifikasi Rekurensi Pasien Kanker Serviks Di RSUD Dr. Soetomo Surabaya.” Institut Teknologi Sepuluh Nopember.

Supandi, Epha Diana. 2020a. *MODUL PRAKTIKUM METODE STATISTIKA DENGAN R*. Sleman: UIN Sunan Kalijaga Yogyakarta.

———. 2020b. *Statistika Dan Terapannya*. Bandung: PT Refika Aditama.

Syafril. 2010. *Statistika*. Padang: Sukabina Press.

Terang, R.G., and R. Syafriharti. 2017. “Karakteristik Pergerakan Berdasarkan Kepadatan Penduduk Untuk Tujuan Bekerja Di Kota Bandung.” *Jurnal Wilayah dan Kota 07(01)*: 1–9.

Utomo, Dito Putro, Pahala Sirait, and Roni Yunis. 2020. “Reduksi Atribut Pada Dataset Penyakit Jantung Dan Klasifikasi Menggunakan Algoritma C5.0.” *Jurnal Media Informatika Budidarma 4(4)*: 994–1006.

Wintana, Dede et al. 2019. “Klasifikasi Penentuan Penerima Manfaat Program Keluarga Harapan (Pkh) Menggunakan Algoritma C5.0 (Studi Kasus : Desa Sukamaju, Kec. Kadudampit).” *Kumpulan jurnaL Ilmu Komputer (KLIK) 06(03)*: 254–64.