

**PERBANDINGAN METODE CART DAN METODE CHAID
UNTUK PENGKLASIFIKASIAN STATUS KERJA PADA
ANGKATAN KERJA DI KABUPATEN TANAH DATAR**

SKRIPSI

*Diajukan sebagai salah satu persyaratan guna memperoleh gelar
Sarjana Statistika*



Oleh
YULIA RIZKI FAJRIATI
NIM 18337067

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

**PERBANDINGAN METODE CART DAN METODE CHAID
UNTUK PENGKLASIFIKASIAN STATUS KERJA PADA
ANGKATAN KERJA DI KABUPATEN TANAH DATAR**

SKRIPSI

*Diajukan sebagai salah satu persyaratan guna memperoleh gelar
Sarjana Statistika*



Oleh
YULIA RIZKI FAJRIATI
NIM 18337067

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

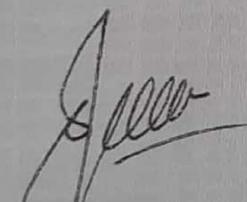
PERSETUJUAN UJIAN SKRIPSI
PERBANDINGAN METODE CART DAN METODE CHAID
UNTUK PENGKLASIFIKASIAN STATUS KERJA PADA
ANGKATAN KERJA DI KABUPATEN TANAH DATAR

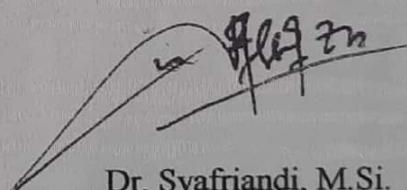
Nama : Yulia Rizki Fajriati
NIM/TahunMasuk : 18337067/2018
Program Studi : S1 Statistika
Jurusan : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 18 Februari 2022

Mengetahui,
Ketua Jurusan Statistika

Disetujui oleh:
Pembimbing


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


Dr. Syafriandi, M.Si.
NIP 19660908 199103 1 003

HALAMAN PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Yulia Rizki Fajriati
NIM/TM : 18337067/2018
Program Studi : S1 Statistika
Jurusan : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

PERBANDINGAN METODE CART DAN METODE CHAID UNTUK PENGKLASIFIKASIAN STATUS KERJA PADA ANGKATAN KERJA DI KABUPATEN TANAH DATAR

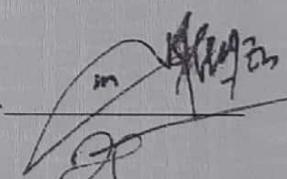
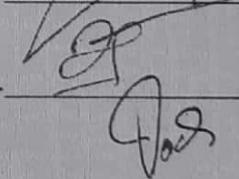
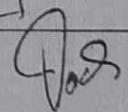
Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi
Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 18 Februari 2022

Tim Penguji

Nama

1. Ketua : Dr. Syafriandi, M.Si.
2. Anggota : Dra. Nonong Amalita, M.Si.
3. Anggota : Dodi Vionanda, M.Si., Ph.D.

1. 
2. 
3. 

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan di bawah ini :

Nama : Yulia Rizki Fajriati

NIM/TM : 18337067/2018

Program Studi : S1 Statistika

Jurusan : Statistika

Fakultas : MIPA UNP

Dengan ini menyatakan bahwa, Tugas akhir saya dengan judul **“Perbandingan Metode CART dan Metode CHAID untuk Pengklasifikasian Status Kerja pada Angkatan Kerja di Kabupaten Tanah Datar”** 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 hukum 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,

Saya yang menyatakan,


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



Yulia Rizki Fajriati
NIM. 18337067

Perbandingan Metode CART dan Metode CHAID untuk Pengklasifikasian Status Kerja Pada Angkatan Kerja di Kabupaten Tanah Datar

Yulia Rizki Fajriati

ABSTRAK

Peningkatan penduduk yang pesat membawa dampak pada tingkat pertumbuhan angkatan kerja. Salah satu kabupaten/kota di Provinsi Sumatera Barat yang mengalami peningkatan penduduk dari tahun ke tahun adalah Kabupaten Tanah Datar, tetapi jumlah angkatan kerjanya mengalami penurunan. Hal ini dapat membuat ketidakseimbangan antara jumlah penduduk dan jumlah angkatan kerja sehingga menyebabkan menurunnya kesejahteraan masyarakat. Selain menurunkan kesejahteraan masyarakat, angkatan kerja juga dapat mempengaruhi status kerja pada angkatan kerja. Permasalahan yang terjadi untuk melihat variabel yang paling dominan mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar, dan memilih metode terbaik dengan menggunakan metode CART dan metode CHAID.

Metode CART dan metode CHAID merupakan salah satu metode yang digunakan untuk pembentukan pohon klasifikasi. Data yang digunakan pada penelitian ini yaitu data sekunder dari SAKERNAS (Survei Angkatan Kerja Nasional) tahun 2021 dengan jumlah data 1057 responden. Dimana variabel yang digunakan adalah variabel pendidikan, status perkawinan, jenis kelamin, usia, dan status pelatihan kerja.

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa variabel yang paling dominan mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar adalah variabel status perkawinan dan metode terbaik untuk mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar adalah metode CART yang dilihat berdasarkan nilai *accuracy* sebesar 73,9 persen dan APER sebesar 26, 1 persen.

Kata Kunci: Angkatan Kerja, Kabupaten Tanah Datar, CART, CHAID.

Comparison of CART Method and CHAID Method for Classification of Employment Status in the Labor Force in Tanah Datar District

Yulia Rizki Fajriati

ABSTRACT

The rapid increase in population has an impact on the growth rate of the labor force. One of the regencies/cities in West Sumatra Province that has experienced an increase in population from year to year is Tanah Datar Regency, but the number of its workforce has decreased. This can create an imbalance between the population and the workforce, causing a decline in people's welfare. In addition to reducing people's welfare, the labor force can also affect the work status of the workforce. The problem that occurs is to see the most dominant variable in classifying work status in the workforce in Tanah Datar Regency, and to choose the best method using the CART method and the CHAID method.

The CART method and the CHAID method are one of the methods used to construct a classification tree. The data used in this study is secondary data from SAKERNAS (National Labor Force Survey) in 2021 with a total data of 1057 respondents. Where the variables used are education, marital status, gender, age, and job training status.

Based on the analysis that has been carried out, it can be concluded that the most dominant variable in classifying work status in the work force in Tanah Datar Regency is the marital status variable and the best method for classifying work status in the workforce in Tanah Datar Regency is the CART method which is seen based on the accuracy value of 73.9 percent and APER of 26.1 percent.

Keywords: Workforce, Tanah Datar Regency, CART, CHAID.

KATA PENGANTAR

Puji dan syukur peneliti ucapkan kepada Allah subhanahu wa ta'ala, yang telah melimpahkan rahmat dan karunia-Nya kepada peneliti, sehingga peneliti dapat menyelesaikan penyusunan Skripsi yang berjudul **“Perbandingan Metode CART dan Metode CHAID untuk Pengklasifikasian Status Kerja pada Angkatan Kerja di Kabupaten Tanah Datar”** dapat terselesaikan dengan baik.

Skripsi ini disusun untuk memenuhi salah satu persyaratan memperoleh gelar Sarjana pada Program Studi S1 Statistika Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang. Dalam penulisan Skripsi ini, peneliti banyak mengalami kendala, namun berkat bantuan, bimbingan, kerjasama dari berbagai pihak yang bersifat membangun, dan berkah dari Allah subhanahu wa ta'ala sehingga kendala-kendala yang dihadapi dapat diatasi. Dalam kesempatan ini, peneliti mengucapkan terimakasih kepada:

1. Bapak Dr. Syafriandi, M.Si, Penasihat Akademik dan Pembimbing Skripsi yang telah memberikan bimbingan dan dorongan dalam menyelesaikan Skripsi ini, sekaligus sebagai tenaga pengajar Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
2. Bapak Dodi Vionanda, M.Si., Ph.D, Dosen Penguji yang banyak memberikan masukan dan arahan demi kesempurnaan Skripsi ini, sekaligus Ketua Prodi Diploma Statistika Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.

3. Ibu Dra. Nonong Amalita, M.Si, Dosen Penguji yang banyak memberikan masukan dan arahan demi kesempurnaan Skripsi ini, sekaligus Sekretaris Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
4. Bapak dan Ibu Dosen Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
5. Kedua Orang Tua dan adik yang selalu mendoakan, memberi semangat, dan nasihat dalam menyelesaikan Skripsi ini.
6. Serta semua sahabat terutama Irsan, Nina, Dani, Febby, dan rekan-rekan Statistika 2018 yang telah banyak membantu dan memberi dukungan kepada peneliti.

Semoga segala bimbingan, bantuan, dan motivasi yang telah diberikan menjadi amal kebaikan dan mendapatkan balasan dari Allah subhanahu wa ta'ala. Peneliti telah berusaha semaksimal mungkin untuk memberikan yang terbaik dalam penulisan Skripsi ini, namun peneliti mengharapkan saran dan kritik yang membangun demi kesempurnaan Skripsi ini. Sehingga, peneliti berharap agar Skripsi ini dapat bermanfaat bagi semua pihak.

Padang, Februari 2022

Yulia Rizki Fajriati

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR TABEL	vi
DAFTAR GAMBAR	vii
DAFTAR LAMPIRAN	viii
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah.....	1
B. Batasan Masalah Penelitian.....	5
C. Rumusan Masalah.....	5
D. Tujuan Penelitian.....	6
E. Manfaat Penelitian.....	6
BAB II KERANGKA TEORITIS	8
A. Analisis Klasifikasi.....	8
B. Undersampling.....	22
C. Ukuran Ketepatan Klasifikasi.....	23
D. Angkatan Kerja.....	24
E. Faktor-faktor yang Mempengaruhi Status Bekerja.....	25
BAB III METODOLOGI PENELITIAN	31
A. Jenis Penelitian.....	31
B. Jenis dan Sumber Data.....	31
C. Populasi dan Sampel.....	31
D. Variabel–variabel dalam Penelitian.....	31
E. Teknik Analisis Data.....	32
BAB IV HASIL DAN PEMBAHASAN	36
A. Deskripsi Data.....	36
B. Hasil Penelitian.....	41
C. Pembahasan.....	54
BAB V KESIMPULAN	57
A. Kesimpulan.....	57
B. Saran.....	57
DAFTAR PUSTAKA	58
LAMPIRAN	61

DAFTAR TABEL

Tabel	Halaman
1. Tabel Kontingensi Dua Arah untuk Uji Chi-Square	15
2. Struktur Data Ketepatan Klasifikasi.....	24
3. Variabel Independen pada Penelitian	32
4. Data Calon Simpul Kiri dan Calon Simpul Kanan	43
5. Hasil Perhitungan Probabilitas Simpul	44
6. Hasil Perhitungan Indeks Gini	45
7. Hasil Perhitungan Goodness of Split	46
8. Tabulasi Silang Status Kerja Berdasarkan Jenis Kelamin.....	48
9. Nilai Chi-square dan Nilai P-value Variabel Independen	49
10. Nilai Chi-square dan Nilai P-value Variabel Independen Terhadap Status Kerja yang Memiliki Status Perkawinan Sudah Pernah Kawin	50
11. Nilai Chi-square dan Nilai P-value Variabel Independen Terhadap Status Kerja yang Memiliki Status Perkawinan Sudah Pernah Kawin dengan Jenis Kelamin Laki-laki	50
12. Nilai Chi-square dan Nilai P-value Variabel Independen Terhadap Status Kerja yang Memiliki Status Perkawinan Belum Pernah Kawin	51
13. Struktur Data Ketepatan Klasifikasi Metode CART.....	52
14. Struktur Data Ketepatan Klasifikasi Metode CHAID.....	53

DAFTAR GAMBAR

Gambar	Halaman
1. Perbandingan Jumlah Penduduk dan Jumlah Angkatan Kerja di Kabupaten Tanah Datar Tahun 2019 – 2021	2
2. Diagram Pohon CART	9
3. Diagram Pohon CHAID	20
4. Proses Undersampling	23
5. Deskripsi Status Kerja pada Angkatan Kerja	36
6. Deskripsi Pendidikan	37
7. Deskripsi Status Perkawinan	38
8. Deskripsi Jenis Kelamin	39
9. Deskripsi Usia	40
10. Deskripsi Status Pelatihan Kerja	41
11. Diagram Hasil Metode CART	46
12. Diagram Hasil Metode CHAID	51

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Penelitian	62
2. Hasil Sebelum Undersampling	63
3. Hasil Setelah Undersampling	64
4. Data Penelitian Setelah Undersampling	65
5. Perhitungan Pembentukan Pohon Klasifikasi Pada Metode CART Setelah Undersampling	67
6. Syntax Metode CART Setelah Undersampling	68
7. Tabulasi Silang Antara Variabel independen dengan Variabel independen Setelah Undersampling	69
8. Uji Signifikansi Variabel Independen Berdasarkan Data yang Dipisahkan Oleh Status Perkawinan Setelah Undersampling	72
9. Uji Signifikansi Berdasarkan Status Perkawinan dengan Jenis Kelamin Setelah Undersampling	76
10. Syntax Metode CHAID Setelah Undersampling	80



BAB I

PENDAHULUAN

A. Latar Belakang Masalah

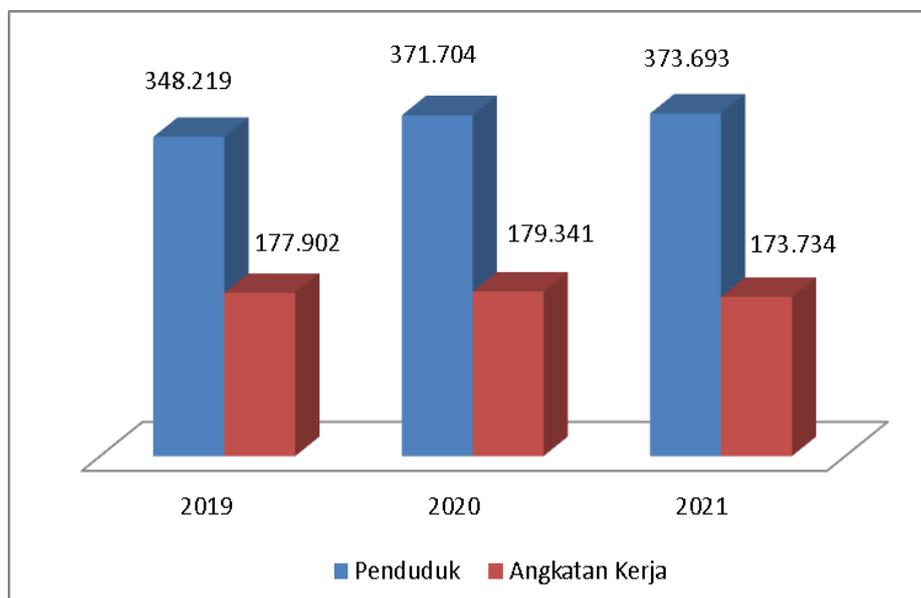
Salah satu penentu kesuksesan suatu perekonomian di bidang sumber daya manusia adalah ketersediaan tenaga kerja. Menurut Undang-Undang Nomor 13 Tahun 2003 tentang ketenagakerjaan, tenaga kerja merupakan setiap individu yang dapat melakukan pekerjaan guna menghasilkan barang atau jasa. Tenaga kerja dengan kualitas dan kuantitas yang tepat, dapat memudahkan suatu perekonomian dalam mencapai target yang optimal. Sehingga, peranan tenaga kerja dalam suatu perekonomian sangatlah penting.

Tenaga kerja merupakan semua penduduk yang telah mencapai batas usia kerja tertentu. Menurut Badan Pusat Statistik, tenaga kerja merupakan setiap penduduk yang masuk ke dalam usia kerja (15 tahun atau lebih). Dimana Indonesia memiliki batasan usia minimal 15 tahun, tetapi tidak memiliki batasan usia maksimal. Hal ini sudah diatur dalam Undang-Undang Nomor 25 Tahun 1997 tentang ketenagakerjaan.

Badan Pusat Statistik membagi penduduk yang merupakan tenaga kerja menjadi dua bagian yaitu angkatan kerja dan bukan angkatan kerja. Angkatan kerja terdiri dari bekerja dan tidak bekerja. Angkatan kerja merupakan penduduk usia 15 tahun atau lebih yang melakukan pekerjaan untuk memperoleh pendapatan atau keuntungan termasuk mereka yang sudah mempunyai pekerjaan tetapi seminggu yang lalu tidak bekerja karena cuti, sakit, mogok, mangkir, menunggu panen dan lain-lain. Sedangkan bukan angkatan kerja adalah penduduk usia 15 tahun atau lebih yang tidak bekerja

dan tidak mempunyai pekerjaan, karena sedang melakukan kegiatan seperti bersekolah, mengurus rumah tangga, pensiun, dan penderita cacat.

Peningkatan penduduk yang pesat membawa dampak pada tingkat pertumbuhan angkatan kerja. Salah satu kabupaten/kota di Provinsi Sumatera Barat yang mengalami peningkatan penduduk dari tahun ke tahun adalah Kabupaten Tanah Datar. Perbandingan jumlah penduduk dan jumlah angkatan kerja di Kabupaten Tanah Datar dapat dilihat pada Gambar 1.



Sumber : BPS Sumatera Barat

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

Berdasarkan Gambar 1, jumlah angkatan kerja di Kabupaten Tanah Datar tahun 2019–2021 dari tahun ke tahun mengalami perubahan, dengan jumlah penduduk di Kabupaten Tanah Datar tahun 2019–2021 dari tahun ke tahun mengalami peningkatan. Dimana jumlah angkatan paling banyak pada tahun 2020 dan paling sedikit pada tahun 2021. Sedangkan jumlah penduduk paling banyak pada tahun 2021 dan paling sedikit pada tahun 2019.

Jumlah penduduk dari tahun ke tahun yang meningkat, sedangkan jumlah angkatan kerja dari tahun ke tahun tidak meningkat, hal ini dapat membuat ketidakseimbangan jumlah penduduk dengan jumlah angkatan kerja sehingga menyebabkan menurunnya kesejahteraan masyarakat. Selain menurunkan kesejahteraan masyarakat, angkatan kerja juga dapat mempengaruhi status kerja baik bekerja maupun tidak bekerja. Dimana yang bekerja pada tahun 2021 sebanyak 95,35 persen dan tidak bekerja sebanyak 4,65 persen.

Faktor utama penduduk bekerja adalah kebutuhan ekonomi yang besar. Kebutuhan ekonomi yang besar membuat penduduk harus memiliki pendapatan untuk memenuhi kebutuhan hidupnya. Sehingga penduduk memiliki partisipasi untuk bekerja yang tinggi dipengaruhi oleh banyak faktor lain sesuai dengan kondisi masing-masing individu.

Menurut penelitian Ariska (2018), faktor-faktor yang mempengaruhi partisipasi tenaga kerja di Kabupaten Tanah Datar tahun 2017 dengan menggunakan analisis regresi probit bahwa faktor yang mempengaruhi partisipasi tenaga kerja yaitu pendidikan, kesehatan, jaminan sosial, status perkawinan, jenis kelamin, usia, status kepala rumah tangga dan status pekerjaan. Hasil analisis diperoleh bahwa faktor-faktor yang mempengaruhi partisipasi tenaga kerja di Kabupaten Tanah Datar adalah jaminan sosial, usia, status kepala rumah tangga, dan status pekerjaan dan peluang tertinggi dari faktor-faktor yang mempengaruhi partisipasi tenaga kerja adalah tenaga kerja dengan yang tidak memiliki jaminan sosial, tidak berada pada usia produktif, bukan sebagai kepala rumah tangga dan tidak memiliki status pekerjaan yang memiliki peluang sebesar 95,64%.

Menurut penelitian Aritonang dkk (2016), klasifikasi status kerja pada angkatan kerja Kota Semarang tahun 2014 menggunakan metode CHAID dan metode CART bahwa faktor yang mempengaruhi status kerja pada angkatan kerja yaitu status hubungan dalam rumah tangga, jenis kelamin, usia, kelengkapan pasangan hidup, status pendidikan, dan status pelatihan kerja. Hasil analisis diperoleh bahwa ketepatan hasil klasifikasi antara metode CART lebih baik dari pada metode CHAID. Dimana, ketepatan klasifikasi metode CART adalah sebesar 72,79% sedangkan dengan metode CHAID 72,63%. Berdasarkan uji proporsi yang dilakukan diperoleh bahwa tidak terdapat perbedaan yang signifikan dari kedua metode ini. Sehingga, dapat dikatakan bahwa kedua metode tersebut mempunyai ketepatan klasifikasi yang relatif sama.

Berdasarkan penelitian sebelumnya, variabel yang digunakan pada penelitian ini adalah variabel independen terdiri dari pendidikan, status perkawinan, jenis kelamin, usia, dan status pelatihan kerja. Sedangkan variabel dependen status kerja pada angkatan kerja terdiri dari bekerja dan tidak bekerja. Dimana skala yang digunakan bersifat kategorik yaitu skala nominal dan ordinal.

Analisis yang digunakan untuk pengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar adalah metode *Classification and Regression Tree* (CART) dan metode *Chi-square Automatic Interaction Detection* (CHAID). Kedua metode ini merupakan metode klasifikasi yang pekerjaannya menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia (Prasetyo, 2012). Sehingga memiliki tujuan yang

sama yaitu untuk pengklasifikasian. Pengklasifikasian ini bertujuan untuk menentukan faktor-faktor yang paling mampu membedakan klasifikasi objek. Pada metode CART, data di eksplorasi agar mengetahui variabel-variabel independen yang memiliki hubungan serta mengelompokkan data ke dalam kategori-kategori yang ada pada variabel dependen (Lewis, 2000). Sedangkan metode CHAID bertujuan menganalisis kumpulan data besar dan membagi data menjadi himpunan bagian yang saling eksklusif yang dapat menggambarkan variabel dependen (Gallagher, 2001).

Berdasarkan permasalahan di atas, maka penelitian ini diberi judul **“Perbandingan Metode CART dan Metode CHAID untuk Pengklasifikasian Status Kerja pada Angkatan Kerja di Kabupaten Tanah Datar”**.

B. Batasan Masalah Penelitian

Pada penelitian ini, perlu adanya batasan masalah dengan maksud untuk memperjelas arah dari suatu masalah agar tidak menimbulkan kekeliruan. Batasan masalah pada penelitian ini mencakup faktor pendidikan, status perkawinan, jenis kelamin, usia, dan status pelatihan kerja di Kabupaten Tanah Datar.

C. Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Apa variabel yang paling dominan mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar menggunakan metode CART dan metode CHAID?

2. Manakah metode terbaik untuk mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar?

D. Tujuan Penelitian

Tujuan dalam penelitian ini adalah sebagai berikut.

1. Untuk mengetahui variabel paling dominan yang dapat mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar menggunakan metode CART dan metode CHAID.
2. Untuk mengetahui manakah yang merupakan metode terbaik untuk mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar.

E. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

1. Penulis, dapat menambah ilmu pengetahuan dan pemahaman terhadap analisis CART dan CHAID.
2. Bagi pemerintahan, dapat memberikan sumbangan pemikiran atau sebagai rujukan dalam pembuatan kebijakan dan keputusan yang berhubungan dengan tenaga kerja.
3. Peneliti selanjutnya, sebagai referensi untuk mengembangkan dan memperluas cakupan hasil penelitian.



BAB II

KERANGKA TEORITIS

A. Analisis Klasifikasi

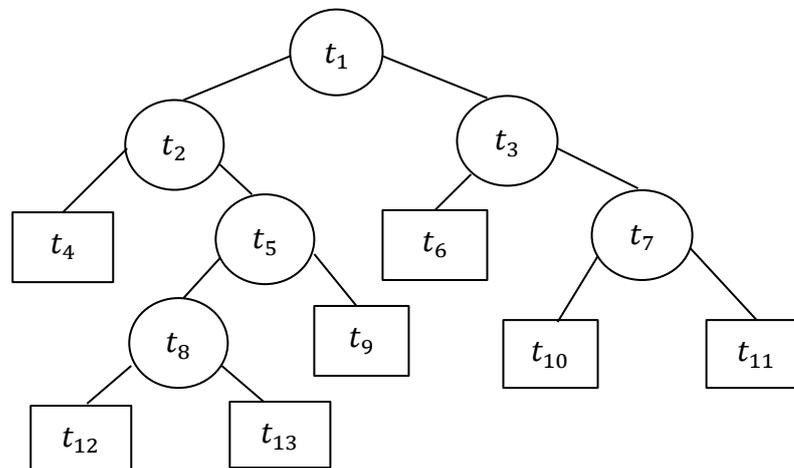
Analisis klasifikasi merupakan salah satu bagian terpenting dalam data mining, yaitu untuk mempelajari sekumpulan data sehingga memperoleh aturan yang bisa mengklasifikasi data-data baru. Menurut Suyanto (2019), klasifikasi merupakan suatu proses menyatakan suatu objek data sebagai salah satu kategori yang telah didefinisikan sebelumnya.

Berdasarkan pemaparan di atas, sejumlah teknik klasifikasi telah diusulkan dalam literatur. Beberapa metode klasifikasi tersebut diantaranya adalah metode *Classification and Regression Tree* (CART) dan metode *Chi-square Automatic Interaction Detection* (CHAID).

1. Metode CART

Metode *Classification and Regression Tree* atau biasa disebut CART dikembangkan oleh Leo Breiman, Jerome H. Friedman Richard A. Oleshen, dan Charles J. Stone pada tahun 1984. Metode CART merupakan salah satu metode dari teknik pohon keputusan. Metode CART merupakan statistik nonparametrik yang dibuat untuk topik analisis klasifikasi, baik peubah respon maupun peubah kontinu (Komalasari, 2007). Metode CART menghasilkan suatu pohon klasifikasi jika peubah responnya kategorik, dan menghasilkan pohon regresi jika peubah responnya kontinu (Breiman, 1984:4). Tujuan metode CART adalah untuk mendapatkan kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian (Komalasari, 2007). Pohon keputusan

metode CART menggunakan algoritma penyekatan rekursif secara biner (*binary recursive partitioning*) (Lewis, 2000). Dimana diagram pohon CART dapat dilihat pada Gambar 2.



Gambar 2. Diagram Pohon CART

Berdasarkan Gambar 2, t_1 merupakan simpul utama (root node), t_2 , t_3 , t_5 , t_7 , dan t_8 merupakan simpul dalam (internal nodes). Sedangkan t_4 , t_6 , t_9 , t_{10} , t_{11} , t_{12} dan t_{13} merupakan simpul terminal (terminal nodes), sehingga tidak terjadi lagi pemilihan. Perhitungan kedalaman pohon (*depth*) dimulai dari simpul utama atau t_1 yang berada pada kedalaman 1, sedangkan t_2 dan t_3 berada pada kedalaman 2. Begitu seterusnya sampai pada simpul terminal t_{12} dan t_{13} yang berada pada kedalaman 5.

a. Algoritma CART

Menurut Sumartini dan Purnami (2015), algoritma CART terdiri dari tiga tahapan, sebagai berikut:

1) Pembentukan Pohon Klasifikasi

Pembentukan pohon klasifikasi terdiri dari tiga tahapan adalah sebagai berikut:

a) Pemilihan Pemilah

Pada pemilahan pemilah, dipilah berdasarkan aturan indeks Gini yang di evaluasi menggunakan kriteria *goodness of split*. Fungsi indeks Gini adalah sebagai berikut:

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

(Breiman, 1984:19)

$P(j|t)$ adalah probabilitas kelas j pada simpul t . Penyekatan (s) pada simpul t adalah sebagai penurunan keheterogenan, sebagai berikut:

$$\Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (2)$$

Dimana:

$i(t)$ = Fungsi keheterogenan pada simpul

$i(t_R)$ = Fungsi keheterogenan pada simpul anak kanan

$i(t_L)$ = Fungsi keheterogenan pada simpul anak kiri

P_R = Probabilitas pengamatan pada simpul kanan

P_L = Probabilitas pengamatan pada simpul kiri

(Breiman, 1984:20)

Probabilitas pengamatan yang masuk pada simpul kanan dan kiri dihitung dengan rumus sebagai berikut (Mardiani, 2012):

$$P_L = \frac{\text{calon simpul kiri}}{\text{data latihan}} \quad (3)$$

$$P_R = \frac{\text{calon simpul kanan}}{\text{data latihan}} \quad (4)$$

Simpul terbaik adalah simpul yang memiliki nilai kebaikan simpul terbesar dengan rumus sebagai berikut:

$$\Delta i(s^*, t) = \max_{s \in S} \Delta i(s, t) \quad (5)$$

(Breiman, 1984:20)

b) Penentu Simpul Terminal

Suatu simpul t akan menjadi simpul terminal atau tidak, apabila pengamatan berjumlah kurang atau sama dengan 5 ($n \leq 5$) dan apabila tingkat kedalaman pohon (*depth*) sudah maksimal (Breiman, 1984:20).

c) Penandaan Label Kelas

Penandaan label kelas merupakan proses pengidentifikasian tiap simpul pada suatu kelas tertentu. Penandaan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak sebagai berikut:

$$P(j_0|t) = \max_j P(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (6)$$

Dimana:

$P(j|t)$ = Probabilitas kelas j pada node t

$N_j(t)$ = Jumlah pengamatan kelas j pada node t

$N(t)$ = Jumlah pengamatan pada node t

(Pratiwi dan Zain, 2014)

2) Pemangkasan Pohon Klasifikasi

Pemangkasan pohon digunakan untuk mencegah terbentuknya pohon klasifikasi yang besar dan kompleks, sehingga didapat ukuran pohon yang layak (Lewis, 2000). Untuk memperoleh ukuran pohon yang layak, dilakukan pemangkasan pohon berdasarkan ukuran *cost complexity minimum* menggunakan rumus sebagai berikut:

$$g_m(t) = \frac{R(t) - R(T_k)}{|T_k| - 1} \quad (7)$$

Dimana:

$g_m(t)$ = *Complexity parameter*

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

T_k = Subtree ke-k, dengan $k= 1,2,\dots,n$

$R(T_k)$ = Kesalahan pengklasifikasian pada pohon

Cabang yang dipangkas adalah cabang yang memiliki nilai $g_m(t)$ terkecil, dengan rumus sebagai berikut:

$$g_m(t_m) = \min_{t \in T_k} g_m(t) \quad (8)$$

3) Penentuan Pohon Klasifikasi Optimal

Penentuan klasifikasi yang berukuran besar akan memberikan nilai penduga pengganti paling kecil, sehingga pohon klasifikasi akan cenderung dipilih untuk menduga nilai respon. Penduga pengganti yang digunakan adalah penduga validasi silang lipat V (*cross validation V-fold estimate*) sebagai berikut:

$$R(T_t^{(v)}) = \frac{1}{N_v} \sum X(d^{(v)}) \quad (9)$$

Dimana $X(d^{(v)})$ hasil pengklasifikasian dan N_v jumlah pengamatan dalam L_v . Pohon klasifikasi optimal yang dipilih adalah pohon dengan *cross validation* yang terkecil yang artinya pohon yang memiliki tingkat kesalahan pengklasifikasian terkecil.

2. Analisis CHAID

Metode *Chi-Square Automatic Interaction Detection* atau biasa disebut dengan metode CHAID pertama kali diperkenalkan pada tahun 1980 dalam artikel G. V. Kass yang berjudul “*An Exploratory Technique for Investigating Large Quantities of Categorical Data*”. Metode ini merupakan bagian dari teknologi sebelumnya dan disebut *Automatic Interaction Detector (AID)*.

Metode CHAID menggunakan statistik *chi-square* sebagai alat utamanya. Metode CHAID merupakan metode pengklasifikasian data yang berupa data kategorik. Metode ini bertujuan untuk membagi kumpulan data menjadi sub kelompok berdasarkan variabel dependen. Sehingga metode CHAID dianggap cukup efektif untuk mengestimasi variabel independen terpenting yang berhubungan dengan variabel dependen serta dapat melihat hubungan antara variabel independen.

a. Variabel-variabel dalam Analisis CHAID

Variabel dalam analisis CHAID terdiri dari dua variabel yaitu variabel dependen dan variabel independen yang merupakan variabel kategorik. Menurut Gallagher (2000), variabel independen yang merupakan variabel kategorik pada metode CHAID dibedakan menjadi tiga, sebagai berikut:

1) Monotonik (*monotonic*)

Variabel independen yang kategorinya dapat dikombinasikan atau digabungkan ketika keduanya saling berdekatan atau mengikuti urutan aslinya (data ordinal). Contoh: tingkat pendidikan

2) Bebas (*free*)

Variabel independen yang kategorinya dapat dikombinasikan atau digabungkan ketika keduanya saling berdekatan ataupun tidak (data nominal).

Contoh: jenis kelamin

3) Mengambang (*floating*)

Variabel independen yang kategorinya dapat diperlakukan seperti monotonik, kecuali untuk kategori yang *missing value* yang dapat dikombinasikan dengan kategori manapun.

b. Uji *Chi-square* untuk Kebebasan

Uji *chi-square* dilakukan untuk memeriksa kebebasan, hal ini guna untuk memutuskan dua variabel dalam satu populasi saling bebas atau tidak saling bebas. Menurut Hatcher (2003:633-634), uji ini memiliki asumsi-asumsi sebagai berikut:

- a) Variabel dependen dan variabel independen dapat dinilai pada semua skala pengukuran dengan pengambilan sampel secara acak dari populasi.
- b) Setiap subjek harus muncul hanya dalam satu sel dari tabel klasifikasi dua arah dengan frekuensi yang diamati tidak bernilai nol. Uji *chi-square* mungkin tidak valid apabila frekuensi yang diamati salah satu sel bernilai nol.

Banyak aplikasi yang memerlukan pengujian independensi dari dua faktor. Apabila ada hubungan yang signifikan antara dua faktor, uji *chi-square* yang dapat membantu untuk memahami hubungan sebab-akibat. Hipotesis untuk uji *chi-square* adalah sebagai berikut.

H_0 : Variabel saling bebas (tidak terdapat hubungan antara variabel dependen dan variabel independen)

H_1 : Kedua variabel tidak saling bebas (terdapat hubungan antara variabel dependen dan variabel independen)

Pengklasifikasian silang dari data dengan variabel kategorik biasanya disajikan dalam tabel kontingensi. Apabila terdapat dua variabel kategorik maka data disajikan dalam tabel kontingensi dua arah yang dapat dilihat pada Tabel 1 .

Tabel 1. Tabel Kontingensi Dua Arah untuk Uji Chi-Square

	B ₁	B ₂	...	B _m	Total Baris
A ₁	n ₁₁	n ₁₂	...	n _{1m}	n _{1.}
A ₂	n ₂₁	n ₂₂	...	n _{2m}	n _{2.}
...
A _k	n _{k1}	n _{k1}	...	n _{km}	n _{k.}
Total Kolom	n _{.1}	n _{.2}	...	n _{.m}	n _{..} = n

Sumber : Baron (2013:311)

Dimana:

A_k = Kategori dari variabel dependen ke-kB_m = Kategori dari variabel independen ke-mn_{k.} = Jumlah baris ke-kn_{.m} = Jumlah kolom ke-m

n = Total baris dan kolom

Untuk mendapatkan frekuensi-frekuensi sel yang diharapkan sehingga digunakanlah hukum dasar peluang yaitu apabila dua peristiwa saling bebas, peluang untuk terjadi secara bersamaan sama dengan hasil perkalian besarnya peluang masing-masing peristiwa. Dengan demikian peluang untuk setiap sel ij sebagai berikut:

$$\hat{P} \{ x \in A_i \cap B_j \} = \left(\frac{n_i}{n} \right) \left(\frac{n_j}{n} \right)$$

$$\widehat{Exp}_{(i,j)} = n \left(\frac{n_i}{n} \right) \left(\frac{n_j}{n} \right) = \frac{(n_i)(n_j)}{n} \quad (10)$$

(Baron, 2013:311)

Dari frekuensi sel teramati dan frekuensi yang diharapkan tersebut, dilakukan perhitungan dengan statistik uji yang mencerminkan besarnya selisih antara keduanya, sebagai berikut:

$$X_{obs}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{\{Obs_{(i,j)} - \widehat{Exp}_{(i,j)}\}^2}{\widehat{Exp}_{(i,j)}} \quad (11)$$

Dimana:

k = Jumlah baris

m = Jumlah kolom

i = Indeks baris

j = Indeks kolom

$Obs_{(i,j)}$ = Frekuensi sel baris ke-i dan kolom ke-j

$\widehat{Exp}_{(i,j)}$ = Nilai harapan sel baris ke-i dan kolom ke-j

(Baron, 2013:311)

Dalam pengambilan keputusan, jika X^2 lebih besar dari X^2 tabel dengan derajat bebas $(k - 1) (m - 1)$ atau p -value kecil dari α yang berarti tolak H_0 yang menyatakan adanya hubungan signifikan antar variabel.

Metode CHAID menggunakan statistik *chi-square* dalam dua cara, sebagai berikut:

- a) *Chi-square* digunakan untuk menentukan kategori-kategori dalam sebuah variabel independen bersifat seragam dan bisa digabungkan menjadi satu.
- b) Ketika semua variabel independen sudah diringkas menjadi bentuk yang signifikan dan tidak mungkin digabung lagi, kemudian *chi-square* digunakan untuk menentukan variabel independen yang paling signifikan untuk membagi atau membedakan kategori-kategori dalam variabel dependen.
- c. Koreksi Bonferroni

Koreksi Bonferroni adalah koreksi yang digunakan pada beberapa uji statistik untuk kebebasan atau ketidakbebasan yang dilakukan pada saat yang

sama (Kunto dan Hasana, 2006). Koreksi Bonferroni biasanya digunakan dalam perbandingan berganda. Menurut Gallagher (2000), pengali Bonferroni untuk masing-masing jenis variabel independen sebagai berikut:

- 1) Variabel independen monotonik digunakan jika peubah asal berskala ordinal dengan rumus.

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

Dimana:

B = Pengali Bonferroni

c = Banyak kategori asal

r = Banyak kategori baru

- 2) Variabel independen bebas digunakan apabila peubah asal berskala nominal dengan rumus.

$$B = \sum_{i=0}^{r-1} (-1)^i \frac{\binom{c-i}{r-i}}{i!} \quad (13)$$

Dimana:

B = Pengali Bonferroni

c = Banyak kategori asal

r = Banyak kategori baru

i = Kategori baru ke-i

- 3) Variabel independen mengambang digunakan jika peubah asal berskala ordinal namun terdapat kategori yang belum dapat ditentukan urutannya dengan rumus.

$$B = \binom{c-2}{r-2} + r \binom{c-2}{r-1} \quad (14)$$

Dimana:

B = Pengali Bonferroni

c = Banyak kategori asal

r = Banyak kategori baru

d. Algoritma CHAID

Algoritma CHAID terdiri dari tiga tahapan, yaitu penggabungan (*merging*), pemisahan (*splitting*), dan penghentian (*stopping*).

1) Penggabungan (*merging*)

Tahapan penggabungan dilakukan untuk memeriksa nilai kebenaran pada suatu hipotesis yang diterima maupun ditolak dari masing-masing kategori variabel dependen terhadap variabel independen. Sehingga setiap variabel dependen akan digabungkan dalam kategori-kategori yang tidak terdapat hubungan antara variabel dependen dengan variabel independen dengan tahapan sebagai berikut:

- a) Bentuk tabel kontingensi dua arah untuk masing-masing variabel independen dengan variabel dependen.
- b) Hitung statistik *chi-square* untuk setiap pasang kategori yang dapat dipilih dan digabung menjadi satu, untuk menguji kebebasannya dalam sebuah sub tabel kontingensi $2 \times J$ yang dibentuk oleh sepasang kategori tersebut dengan variabel dependennya yang mempunyai sebanyak J kategori.
- c) Untuk masing-masing nilai *chi-square* yang berpasangan, dilakukan perhitungan nilai *p-value* berpasangan secara bersamaan. Diantara pasangan-pasangan yang tidak signifikan, gabungkan sebuah pasangan kategori yang paling mirip (pasangan yang mempunyai nilai *chi-square*

berpasangan terkecil dan *p-value* terbesar) menjadi sebuah kategori tunggal.

- d) Setelah digabung dengan kategori lain dalam variabel independen periksa kembali signifikasinya. Apabila masih ada pasangan yang belum signifikan, ulangi langkah ketiga.
- e) Hitung nilai *p-value* terkoreksi Bonferroni berdasarkan tabel yang telah digabungkan.

2) Pemisahan (*splitting*)

Pemisahan dilakukan pada variabel independen untuk memilih pembagian titik terbaik. Pemisahan dilakukan dengan membandingkan nilai *p-value* (dari tahap penggabungan) pada setiap variabel independen.

Langkah-langkah pemisahan (*splitting*) dilakukan sebagai berikut:

- a) Pilih variabel independen yang memiliki nilai *p-value* terkecil (paling signifikan).
- b) Apabila nilai *p-value* kurang dari sama dengan α , maka pemisah simpul (*split node*) menggunakan variabel independen. Apabila tidak ada variabel independen dengan nilai *p-value* yang signifikan, tidak dilakukan pemisahan dan simpul ditentukan sebagai terminal simpul (simpul akhir).

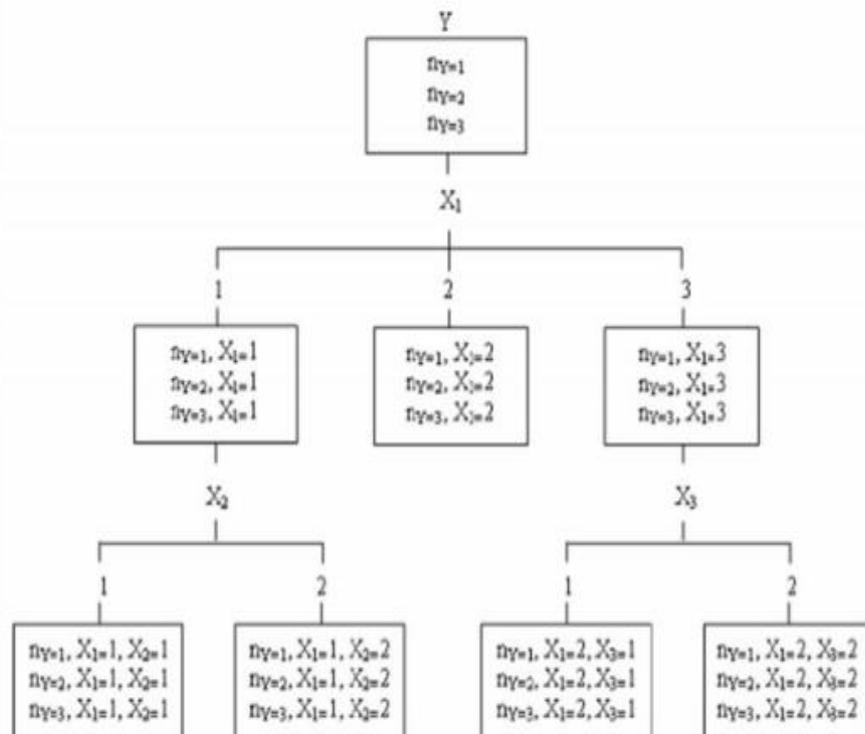
3) Penghentian (*stopping*)

Tahap penghentian dilakukan ketika proses pertumbuhan pohon harus dihentikan. Menurut Rokach (2008:72), penyekatan simpul dalam pembentukan pohon harus dihentikan dikarenakan sebagai berikut.

- a) Pada simpul hanya terdapat kasus yang berasal dari salah satu kelas variabel dependen.

- b) Kedalaman pohon maksimal telah tercapai.
 - c) Jumlah kasus di simpul terminal kurang dari jumlah minimal kasus untuk menjadi simpul orang tua (*parent node*).
 - d) Apabila simpul disekat, jumlah kasus dalam satu atau lebih simpul anak akan kurang dari jumlah minimal kasus untuk simpul anak, maka proses penyekatan simpul berhenti.
 - e) Apabila semua variabel independen mempunyai nilai signifikansi lebih besar dari nilai α yang ditentukan, maka simpul tidak dapat disekat.
- e. Diagram Pohon

Hasil yang terbentuk di bagian tengah CHAID akan ditampilkan dalam diagram pohon. Secara umum diagram pohon CHAID dapat dilihat pada Gambar 3 (Lehmann dan Eherler, 2001).



Gambar 3. Diagram Pohon CHAID

Berdasarkan Gambar 3, pohon klasifikasi pada metode CHAID mengikuti aturan “dari atas ke bawah” (*Top-down stopping rule*), dimana pohon disusun dari grup induk yang berdasarkan kriteria tertentu. Pada kedalaman pertama, dilakukan pembagian sampel dengan X_1 sebagai variabel independen terbaik berdasarkan variabel dependen pada uji *chi-square*. Setiap simpul berisi informasi tentang frekuensi variabel Y sebagai variabel dependen yang merupakan bagian dari subgrup hasil berdasarkan kategori yang disebutkan (X_1). Pada kedalaman kedua (node X_2 dan X_3) merupakan pembagian X_1 (untuk node ke-1 dan ke-3). Dengan cara yang sama, sampel dibagi lagi dengan variabel independen lainnya, yaitu X_2 dan X_3 , kemudian subgrup pada node ke-4, 5, 6, dan 7 (Lehmann dan Eherler, 2001).

Setiap simpul tersebut juga menampilkan persentase responden untuk setiap kategori dari variabel dependen, dan dapat menunjukkan jumlah total responden untuk setiap simpul. Sehingga, diagram pohon analisis CHAID akan berisi sebagai berikut:

- 1) Simbol yang menerangkan tentang kategori tertentu seperti kategori-kategori yang digabungkan.
 - 2) Sebuah ringkasan data dari variabel dependen dalam kelompok tersebut.
 - 3) Ukuran sampel untuk kelompok “n”.
3. Perbandingan Metode CART dan Metode CHAID

Metode CART dan metode CHAID memiliki beberapa kelebihan. Menurut Timofeev (2004:19-24), kelebihan metode CART sebagai berikut:

- a. Metode CART merupakan metode nonparametrik, yang berarti tidak adanya asumsi distribusi variabel independen yang perlu di penuhi.

- b. Metode CART tidak hanya memberikan klasifikasi tetapi juga estimasi algoritma probabilitas kesalahan pengklasifikasian.
- c. Metode CART lebih mudah dalam hal eksplorasi dan pengambilan keputusan pada struktur data yang kompleks dan multivariabel, karena struktur data dapat dilihat secara visual.
- d. Metode CART menghasilkan klasifikasi akhir yang sederhana dan dapat mengklasifikasikan data baru secara efisien.
- e. Metode CART mudah untuk diinterpretasikan hasilnya.

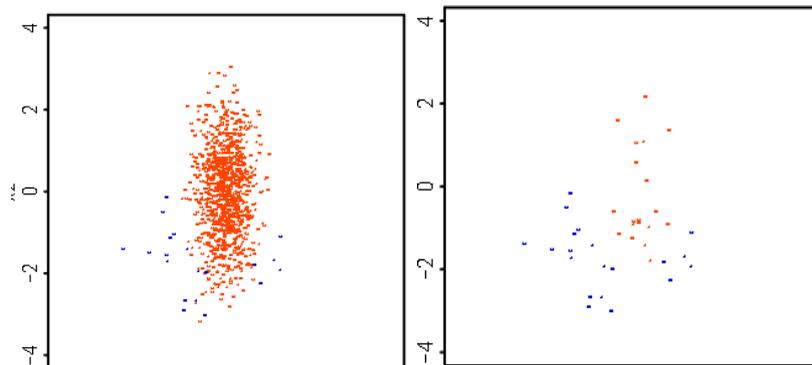
Sedangkan menurut Gallagher (2000), kelebihan metode CHAID sebagai berikut:

- a. Algoritma CHAID memungkinkan data untuk didefinisikan ke dalam kelas regu yang sesuai, sehingga menjamin bahwa kelompok tersebut terdefinisi berdasarkan pada populasi.
- b. Pembagian dilakukan berdasarkan konteks dari semua faktor, interaksi dari semua faktor secara otomatis akan tepat pada sasaran.
- c. Algoritma CHAID bersifat interatif, karena memberikan urutan variabel seperti yang diharapkan.

B. Undersampling

Undersampling merupakan salah satu penyelesaian masalah ketidakseimbangan data. *Undersampling* dilakukan apabila data yang diperoleh pada kelas minoritas sangat kecil dan kelas mayoritas yang sangat besar. Menurut Lunardo dkk (2014), proses *undersampling* merupakan proses pengambilan kelas mayoritas secara acak yang dilakukan tanpa pengembalian

sehingga kelas mayoritas memiliki proporsi yang sama dengan kelas minoritas. Proses *undersampling* dapat dilihat pada Gambar 4.



Gambar 4. Proses Undersampling

Pada metode CART dan metode CHAID, perlu dilakukan penanganan untuk mengatasi data yang tidak seimbang. Hal ini dikarenakan data yang tidak seimbang akan menunjukkan prediksi yang buruk. Sehingga perlu dilakukannya proses *undersampling*.

C. Ukuran Ketepatan Klasifikasi

Ukuran ketepatan klasifikasi digunakan untuk mengetahui apakah data diklasifikasi dengan benar atau tidak (Agresti, 2002). Menurut Johnson dan Winchern (2007:598), beberapa cara yang biasa digunakan untuk mengukur ketepatan klasifikasi dengan melalui perhitungan *sensitivity*, *specifity*, *Apparent Error Rate* (APER) dan *accuracy*. *Sensitivity* digunakan untuk mengukur ketepatan dari kejadian yang diinginkan. *Specifity* digunakan untuk mengukur ketepatan dari kejadian yang tidak diinginkan. APER digunakan untuk mengukur ketepatan klasifikasi yang diprediksi tidak benar. Sedangkan *accuracy* digunakan untuk mengukur ketepatan klasifikasi yang diprediksi benar. Struktur data ketepatan klasifikasi dapat dilihat pada Tabel 2.

Tabel 2. Struktur Data Ketepatan Klasifikasi

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

Sumber: Johnson dan Winchern (2007:598)

Dimana:

n_{11} = Jumlah item 1 yang diklasifikasikan dengan benar

n_{12} = Jumlah item 1 yang diklasifikasikan dengan salah

n_{22} = Jumlah item 2 yang diklasifikasikan dengan benar

n_{21} = Jumlah item 2 yang diklasifikasikan dengan salah

Untuk mengukur kinerja klasifikasi menggunakan rumus sebagai berikut.

$$Sensitivity = \frac{n_{11}}{N_{1.}} \quad (15)$$

$$Specificity = \frac{n_{22}}{N_{2.}} \quad (16)$$

$$APER = \frac{n_{12} + n_{21}}{N} \quad (17)$$

$$Accuracy = \frac{n_{11} + n_{22}}{N} \quad (18)$$

D. Angkatan Kerja

Menurut Suasti dan Sri (2009:185), angkatan kerja adalah penduduk usia kerja (15 tahun atau lebih) 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 terbagi menjadi 2 bagian yaitu bekerja dan tidak bekerja.

1. Bekerja

Bekerja adalah kegiatan ekonomi yang dilakukan oleh seseorang dalam seminggu terakhir dengan maksud untuk memperoleh pendapatan atau keuntungan. Kegiatan ini juga termasuk mereka yang sudah mempunyai pekerjaan tetapi seminggu yang lalu tidak bekerja karena cuti, sakit, mogok, mangkir, menunggu panen dan lain-lain.

2. Tidak bekerja

Pengangguran terbuka adalah 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, dan mereka yang sudah punya pekerjaan, tetapi belum mulai bekerja. Angkatan kerja yang digolongkan tidak bekerja disebut dengan pengangguran terbuka.

E. Faktor-faktor yang Mempengaruhi Status Bekerja

Menurut Raviantor (1985:14), ada beberapa faktor yang mempengaruhi seseorang untuk bekerja, salah satunya yaitu pendidikan.

Status perkawinan merupakan faktor yang mempengaruhi seseorang untuk bekerja. Status perkawinan yang dimiliki seseorang akan mempengaruhi produktivitas kerjanya. Dengan perkawinan akan meningkatkan tanggung jawab seseorang untuk memenuhi kebutuhan keluarganya.

Jenis kelamin dapat mempengaruhi seseorang untuk bekerja. Menurut Simanjuntak (2001:40), laki-laki dianggap sebagai pencari nafkah utama bagi

keluarga karena memiliki tanggung jawab yang besar kepada keluarganya. Tetapi pada umumnya perempuan termotivasi bekerja untuk membantu menghidupi keluarganya.

Semakin bertambah usia seseorang dapat mempengaruhi tingkat pekerjaan yang akan diperolehnya (Simanjuntak, 2001:48). Sehingga, dapat disimpulkan bahwa usia mempengaruhi seseorang untuk bekerja. Menurut Badan Pusat Statistik, usia dalam komposisi penduduk terbagi menjadi dua golongan, yaitu usia produktif dan usia tidak produktif. Usia produktif adalah penduduk yang berusia antara 15 sampai 64 tahun, sedangkan usia tidak produktif adalah penduduk berusia dibawah 15 tahun dan usia diatas 64 tahun.

faktor lain yang mempengaruhi seseorang untuk bekerja adalah status pelatihan. Status pelatihan memiliki tujuan yaitu untuk meningkatkan keterampilan seseorang dalam bekerja. sehingga status pelatihan memiliki manfaat diantaranya untuk meningkatkan kualitas dan kuantitas produktifitas kerja (Simamora, 1999:349).

Selain itu, faktor lain yang mempengaruhi status kerja pada angkatan kerja adalah status pekerjaan. Menurut BPS status pekerjaan adalah jenis kedudukan seseorang dalam melakukan pekerjaan di suatu unit usaha/kegiatan. Pada tahun 2001 status pekerjaan dibedakan menjadi 7 kategori yaitu: berusaha sendiri, berusaha dibantu buruh tak bayar, berusaha dibantu buruh dibayar, buruh/karyawan/pegawai, pekerja bebas di pertanian, pekerja bebas di non pertanian, pekerja keluarga/tidak dibayar.

1. Pendidikan

Tenaga kerja merupakan salah satu penentu kesuksesan suatu perekonomian di bidang sumber daya manusia (SDM). Salah satu investasi di bidang SDM adalah pendidikan dan latihan. Pendidikan dan latihan merupakan faktor penting dalam pengembangan SDM (Sumarsono, 2009:92). Keduanya dapat menambah pengetahuan serta meningkatkan keterampilan dalam bekerja.

Menurut Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 Bab VI Pasal 14-19 tentang Sistem Pendidikan Nasional, Di Indonesia jenjang pendidikan adalah sebagai berikut.

- a. Pendidikan Dasar, berusia 7 sampai 15 tahun. Pada tahap ini terbagi menjadi dua tingkat yaitu Sekolah Dasar (SD/MI) dan Sekolah Menengah Pertama (SMP/MTs).
- b. Pendidikan Menengah, lanjutan dari pendidikan dasar yaitu Sekolah Menengah Atas atau Kejuruan (SMA/SMK).
- c. Pendidikan Tinggi, lanjutan dari pendidikan menengah yaitu pendidikan tinggi. Pendidikan tinggi terdiri dari program diploma (D1, D2, D3), S1, S2 dan S3.

2. Status Perkawinan

Menurut LDFEUI (2011:155), status perkawinan merupakan suatu karakteristik demografi yang terdiri dari aspek sosial, ekonomi, biologis, hukum, dan agama. Status perkawinan dapat berubah-ubah baik dari status belum pernah kawin menjadi sudah pernah kawin. Perubahan status perkawinan ini dapat menyebabkan perubahan tempat tinggal atau migrasi,

perubahan partisipasi angkatan kerja, atau perubahan pendidikan (Adioetomo dan Samosir, 2010:51).

3. Jenis Kelamin

Menurut Sasana dan Herawati (2013), salah satu faktor yang mempengaruhi produktivitas kerja adalah jenis kelamin. Jenis kelamin memiliki kaitan yang erat dengan produktivitas. Tingginya tingkat produktivitas laki-laki dari pada perempuan dapat membuat laki-laki memiliki peluang lebih tinggi memperoleh pekerjaan. Menurut Shon (2015), perempuan memiliki peran ganda sehingga tidak memiliki banyak waktu untuk melakukan pekerjaan diluar pekerjaan rumah tangga.

4. Usia

Menurut Simanjuntak (2001:48), semakin bertambah usia seseorang akan mempengaruhi tingkat pekerjaan yang akan dilakukannya. Semakin dewasa seseorang dapat meningkatkan keterampilan dalam bidang tertentu dan kekuatan fisik. Sehingga, pekerja yang bekerja pada sektor formal yang banyak mengandalkan kemampuan fisik akan sangat terpengaruh oleh variabel usia.

Namun di sisi lain, pada usia yang tidak lagi produktif, keterampilan dan kekuatan fisik seseorang akan menurun. Hal ini sesuai dengan kenyataan bahwa pada usia tersebut, banyak orang yang pensiun atau secara fisik sudah tidak mampu lagi untuk bekerja, sehingga akan sangat berpengaruh terhadap tingkat pekerjaannya.

5. Status Pelatihan Kerja

Menurut Dessler (2010:280), pelatihan merupakan suatu kegiatan yang dibutuhkan karyawan baru untuk melakukan pekerjaannya. Pelatihan dan

pengembangan (*training and development*) memiliki tujuan yaitu untuk meningkatkan kompetensi karyawan serta kinerja organisasi. Sehingga dapat disimpulkan bahwa pelatihan merupakan suatu kegiatan untuk karyawan bertujuan untuk meningkatkan keterampilan sesuai bidang pekerjaan yang akan dilakukan karyawan tersebut.

Menurut Simamora (1999:349), beberapa manfaat dari pelatihan kerja diantaranya sebagai berikut.

- a. Mengurangi waktu belajar karyawan untuk mencapai standar kinerja yang dapat diterima.
- b. Membentuk sikap, loyalitas, dan kerjasama yang lebih menguntungkan.
- c. Memenuhi kebutuhan perencanaan sumber daya manusia.
- d. Mengurangi frekuensi dan biaya kecelakaan kerja.
- e. Membantu karyawan dalam peningkatan dan pengembangan pribadi mereka.



BAB V

KESIMPULAN

A. Kesimpulan

Kesimpulan dari penelitian ini adalah sebagai berikut.

1. Variabel yang paling dominan mengklasifikasi status kerja pada angkatan kerja di Kabupaten Tanah Datar dengan menggunakan metode CART dan metode CHAID adalah variabel status perkawinan
2. Ketepatan klasifikasi yang dihasilkan metode CART lebih baik dibandingkan metode CHAID. Dimana hasil ketepatan klasifikasi metode CART adalah sebesar 73,9 persen dengan kesalahan klasifikasi sebesar 26,1 persen, sedangkan hasil ketepatan klasifikasi metode CHAID adalah sebesar 73,7 persen dengan kesalahan klasifikasi sebesar 26,3 persen. Berdasarkan hasil ketepatan klasifikasi yang dihasilkan bahwa tidak terdapat perbedaan yang signifikan dari kedua metode ini. Jadi, kedua metode mempunyai ketepatan yang relatif sama.

B. Saran

Saran dari penelitian ini adalah sebagai berikut:

1. Hasil pengelompokkan ini bisa dijadikan acuan untuk melihat karakteristik status kerja pada angkatan kerja di Kabupaten Tanah Datar.
2. Bagi yang akan melakukan penelitian selanjutnya dapat menambahkan faktor-faktor yang berpengaruh terhadap status kerja pada angkatan kerja.

DAFTAR PUSTAKA

- Adioetomo SM dan Samosir OB. 2010. *Dasar-dasar Demografi edisi 2*. Jakarta: Penerbit Salemba Empat.
- Agresti, A. 2002. *Categorical Data Analysis*. John Wiley and Sons, Inc., New York.
- Ariska, T. dan Nasution, M. L. 2019. Faktor-faktor yang Mempengaruhi Partisipasi Tenaga Kerja di Kabupaten Tanah Datar Tahun 2017 dengan Menggunakan Analisis Regresi Probit. *Seminar Nasional Edusainstek*, 533-542.
- Aritonang dkk. 2016. Klasifikasi Status Kerja pada Angkatan Kerja Kota Semarang Tahun 2014 Menggunakan Metode CHAID dan Metode CART. *Jurnal Gaussian*, 5(1): 183-192.
- Baron, M. 2013. *Probability And Statistics For Computer Scientists*. 2 ed. University of Texas at Dallas Richardson, USA: CRC Press.
- Breiman L., Friedman J. H Olshen R.A & Stone C.J. 1984. *Classification and Regression Tree*. New York: Chapman And Hall.
- Dessler, Gary. 2010. *Manajemen Sumber Daya Manusia Edisi Kesepuluh*. Jakarta Barat. PT Indeks.
- Eherler, D., Lehmann, T. 2001. *Responder Profiling with CHAID and Dependency Analysis*.
- Gallagher, Cecily A, Howard M Monroe, and Joyce L Fish. 2000. "An Iterative Approach to Classification Analysis". *Journal of Applied Statistics*, 238-280.
- Hatcher, L. 2003. *Step by Step Basic Statistics Using SAS*. North California: SAS Publishing.
- Johnson, R. A. dan Winchern, D. W. 2007. *Applied Multivariate Statistical Analysis*. 6 ed. New Jersey: Prentice Hall, Englewood Clifft.
- Kass, G. V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29(2): 119-127.
- Kolmasari WB. 2007. Metode Pohon Regresi untuk Eksploratori Data dengan Peubah yang Banyak dan Kompleks. *Informatika Pertanian*, 16(1): 967-980.