

**KLASIFIKASI STATUS KERJA PADA ANGKATAN KERJA KOTA
SEMARANG TAHUN 2014 MENGGUNAKAN METODE
CHAID DAN CART**



SKRIPSI

**Disusun Oleh :
NOVIE ERISKA ARITONANG
24010211140081**

**JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG**

2015

**KLASIFIKASI STATUS KERJA PADA ANGKATAN KERJA
KOTA SEMARANG TAHUN 2014 MENGGUNAKAN METODE
CHAID DAN CART**

Oleh :

NOVIE ERISKA ARITONANG

24010211140081

Diajukan Sebagai Salah Satu Syarat Memperoleh Gelar Sarjana Sains pada
Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro

**JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA**

UNIVERSITAS DIPONEGORO

SEMARANG

2015

HALAMAN PENGESAHAN I

Judul Skripsi : Klasifikasi Status Kerja pada Angkatan Kerja Kota
Semarang Tahun 2014 Menggunakan Metode CHAID
dan CART

Nama Mahasiswa : Novie Eriska Aritonang

NIM : 24010211140081

Jurusan : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 7 Desember 2015 dan dinyatakan
lulus pada tanggal 22 Desember 2015

Semarang, 22 Desember 2015

Mengetahui,

Ketua Jurusan Statistika

Fakultas Sains dan Matematika

Universitas Diponegoro



Dia. H. Dwi Ispriyanti, M. Si
NIP. 195709141986032001

Panitia Penguji Tugas Akhir

Ketua,

Triastuti Wuryandari, S.Si, M.Si
NIP. 197109061998032001


HALAMAN PENGESAHAN II

Judul Skripsi : Klasifikasi Status Kerja pada Angkatan Kerja Kota
Semarang Tahun 2014 Menggunakan Metode CHAID
dan CART

Nama Mahasiswa : Novie Eriska Aritonang

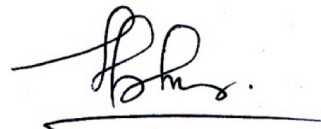
Semarang, 22 Desember 2015

Pembimbing I



Drs. Agus Rusgiyono, M.Si
NIP. 196408131990011001

Pembimbing II



Rita Rahmawati, S.Si, M.Si
NIP. 198009102005012002

KATA PENGANTAR

Puji syukur penulis ucapkan kepada Tuhan Yesus Kristus yang telah memberikan kasih dan anugerah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Klasifikasi Status Kerja pada Angkatan Kerja Kota Semarang Tahun 2014 Menggunakan Metode CHAID dan CART”.

Tugas akhir ini tidak akan selesai dengan baik tanpa adanya bantuan dan dukungan dari berbagai pihak. Ucapan terimakasih penulis haturkan kepada:

1. Ibu Dra. Dwi Ispriyanti, M.Si. selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Bapak Drs. Agus Rusgiyono, M.Si dan Ibu Rita Rahmawati, S.Si, M. Si selaku dosen pembimbing I dan dosen pembimbing II yang telah memberikan bimbingan dan pengarahan dalam penulisan Tugas Akhir ini.
3. Bapak dan Ibu Dosen Jurusan Statistika Universitas Diponegoro yang telah memberikan ilmu yang bermanfaat.
4. Semua pihak yang tidak dapat disebutkan satu per satu yang telah mendukung penulis dalam menyelesaikan tugas akhir ini.

Penulis berharap Tugas Akhir ini dapat bermanfaat bagi seluruh civitas akademika di Universitas Diponegoro khususnya Jurusan Statistika dan masyarakat pada umumnya.

Semarang, Desember 2015

Penulis

ABSTRAK

Pertumbuhan angkatan kerja akan meningkat seiring dengan meningkatnya jumlah penduduk. Peningkatan jumlah angkatan kerja ini tentu saja akan berdampak pada status kerjanya, apakah bekerja atau tidak bekerja. Metode yang dapat digunakan untuk melakukan klasifikasi status kerja pada angkatan kerja adalah CHAID (*Chi-squared Automatic Interaction Detection*) dan CART (*Classification and Regression Trees*). Kedua metode ini bertujuan untuk mengetahui faktor-faktor apa saja yang mempengaruhi status kerja dari angkatan kerja. Data yang digunakan adalah data angkatan kerja Kota Semarang pada tahun 2014. Berdasarkan metode CHAID faktor-faktor yang mempengaruhi status kerja dari angkatan kerja adalah jenis kelamin, usia serta status kelengkapan pasangan hidup dengan ketepatan hasil klasifikasi sebesar 72,63%. Faktor-faktor yang mempengaruhi status kerja dari angkatan kerja dengan metode CART adalah jenis kelamin, usia, status pendidikan, serta status kelengkapan pasangan hidup dengan ketepatan hasil klasifikasi sebesar 72,79%. Berdasarkan uji proporsi masing-masing metode diperoleh bahwa kedua metode tersebut tidak berbeda signifikan dalam melakukan klasifikasi.

Kata kunci: Angkatan kerja, Klasifikasi, CHAID, CART, Ketepatan klasifikasi

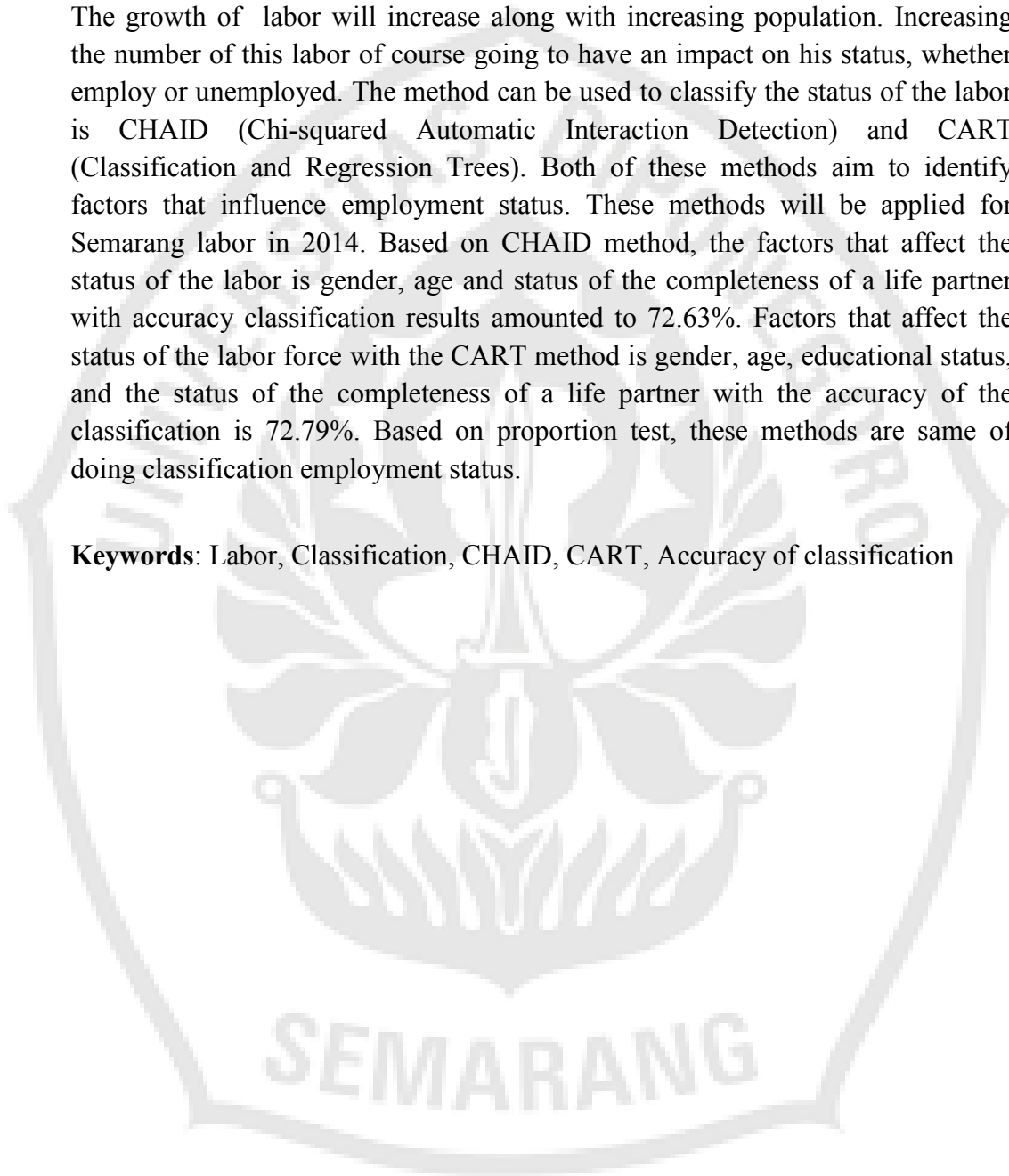


SEMARANG

ABSTRACT

The growth of labor will increase along with increasing population. Increasing the number of this labor of course going to have an impact on his status, whether employ or unemployed. The method can be used to classify the status of the labor is CHAID (Chi-squared Automatic Interaction Detection) and CART (Classification and Regression Trees). Both of these methods aim to identify factors that influence employment status. These methods will be applied for Semarang labor in 2014. Based on CHAID method, the factors that affect the status of the labor is gender, age and status of the completeness of a life partner with accuracy classification results amounted to 72.63%. Factors that affect the status of the labor force with the CART method is gender, age, educational status, and the status of the completeness of a life partner with the accuracy of the classification is 72.79%. Based on proportion test, these methods are same of doing classification employment status.

Keywords: Labor, Classification, CHAID, CART, Accuracy of classification

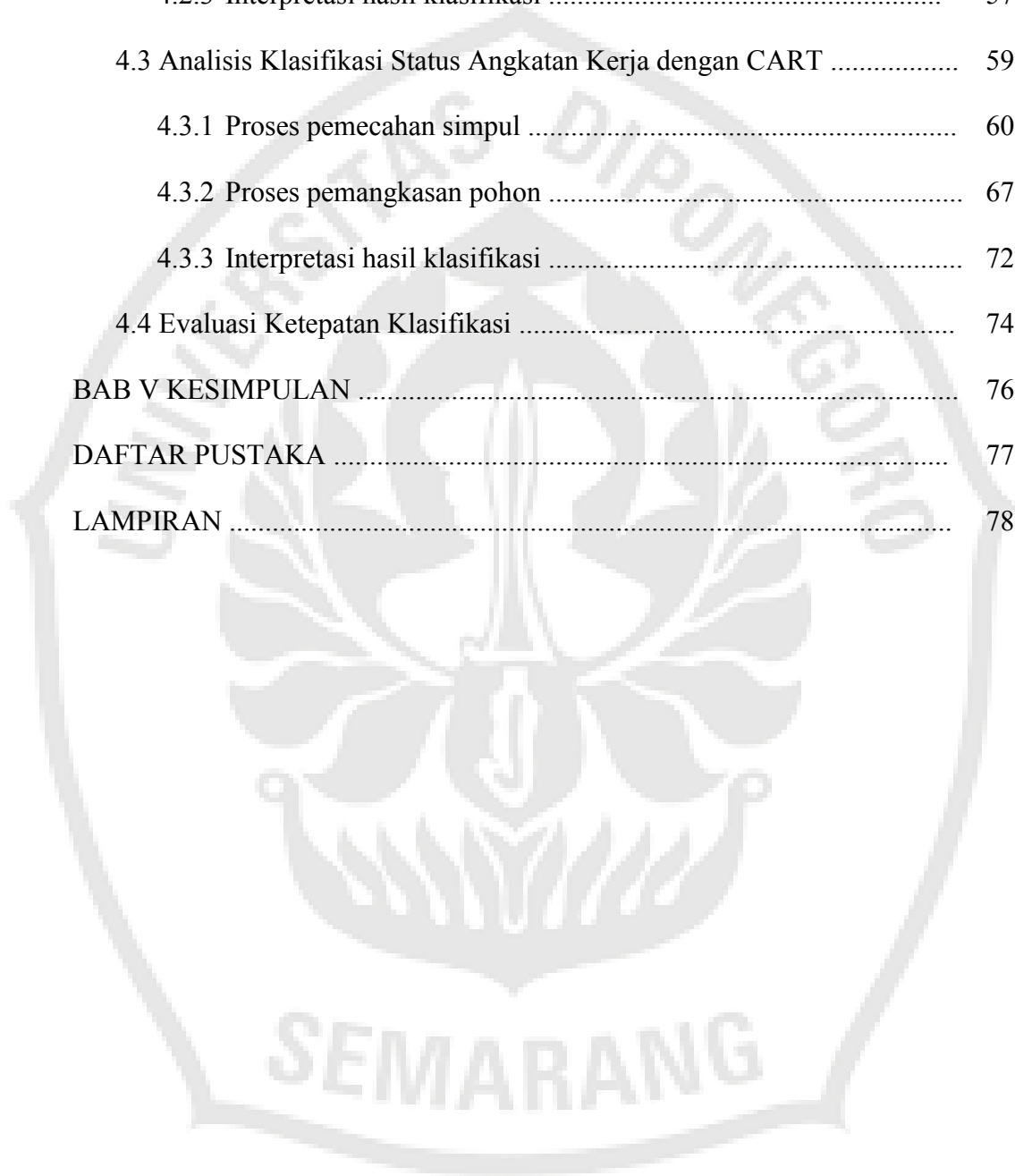


DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN I	ii
HALAMAN PENGESAHAN II	iii
KATA PENGANTAR	iv
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI	vii
DAFTAR SIMBOL	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Angkatan Kerja	5
2.2 Skala Pengukuran	8
2.3 CHAID	9
2.3.1 Konsep CHAID	9
2.3.2 Variabel-variabel dalam analisis CHAID	10

2.3.3 Uji Independensi χ^2 (<i>Chi-Square</i>)	11
2.3.4 Langkah Kerja CHAID	12
2.3.5 Koreksi Bonferroni	14
2.3.6 Struktur Pohon Klasifikasi CHAID	15
2.4 CART	17
2.4.1 Konsep CART	17
2.4.2 Struktur Pohon Klasifikasi CART	17
2.4.3 Langkah Kerja CART	19
2.5 Ukuran Kinerja Klasifikasi	24
BAB III METODOLOGI PENELITIAN	27
3.1 Sumber Data	27
3.2 Variabel Penelitian	27
3.3 Tahapan Analisis	28
3.4 Diagram Alir	30
BAB IV HASIL DAN PEMBAHASAN	34
4.1 Analisis Deskriptif Data	34
4.1.1 Status kerja menurut hubungan dalam rumah tangga	34
4.1.2 Status kerja menurut jenis kelamin	35
4.1.3 Status kerja menurut usia	36
4.1.4 Status kerja menurut kelengkapan pasangan hidup	36
4.1.5 Status kerja menurut pendidikan	37
4.1.6 Status kerja menurut pelatihan kerja	38
4.2 Analisis Klasifikasi Status Angkatan Kerja dengan CHAID	38
4.2.1 Penggabungan Kategori	39

4.2.2 Uji independensi variabel independen dengan variabel dependen	43
4.2.3 Interpretasi hasil klasifikasi	57
4.3 Analisis Klasifikasi Status Angkatan Kerja dengan CART	59
4.3.1 Proses pemecahan simpul	60
4.3.2 Proses pemangkasan pohon	67
4.3.3 Interpretasi hasil klasifikasi	72
4.4 Evaluasi Ketepatan Klasifikasi	74
BAB V KESIMPULAN	76
DAFTAR PUSTAKA	77
LAMPIRAN	78



DAFTAR SIMBOL

- O_{ij} : Observasi yang memenuhi sifat variabel pertama kriteria ke- i dan variabel kedua kriteria ke- j , $i=1,2,\dots,r$ dan $j=1,2,\dots,c$
- E_{ij} : Nilai harapan observasi yang memenuhi sifat variabel pertama kriteria ke- i dan variabel kedua kriteria ke- j
- n_{i+} : banyak pengamatan pada baris ke- i kriteria ke- i
- n_{+j} : banyak pengamatan pada kolom ke- j kriteria ke- j
- r : banyak kriteria pada variabel pertama
- c : banyak kriteria pada variabel kedua
- a : banyak kategori variabel independen sebelum penggabungan
- b : banyak kategori variabel independen setelah penggabungan
- $n_{Y=1}$: frekuensi variabel dependen kategori 1
- $X_{i,j}$: variabel independen ke- i kategori ke- j
- B : pengali Bonferroni
- m : nilai terendah dari kategori
- $\Delta_i(s,t)$: fungsi pengurangan keragaman pada *node* t dengan kriteria pemecah s
- $I(t)$: fungsi keragaman indeks Gini
- $p(A_i|t)$: peluang kelas i pada *node* t
- $p(A_j|t)$: peluang kelas j pada *node* t
- T : pohon klasifikasi maksimal
- T_t : cabang dari T
- \tilde{T} : banyak simpul-simpul akhir dari pohon klasifikasi

t_L : simpul anak kiri

t_R : simpul anak kanan

f_{11}, f_{00} : data yang diklasifikasikan secara benar

f_{01}, f_{10} : data yang diklasifikasikan tidak tepat

P_1 : proporsi metode CHAID

P_2 : proporsi metode CART

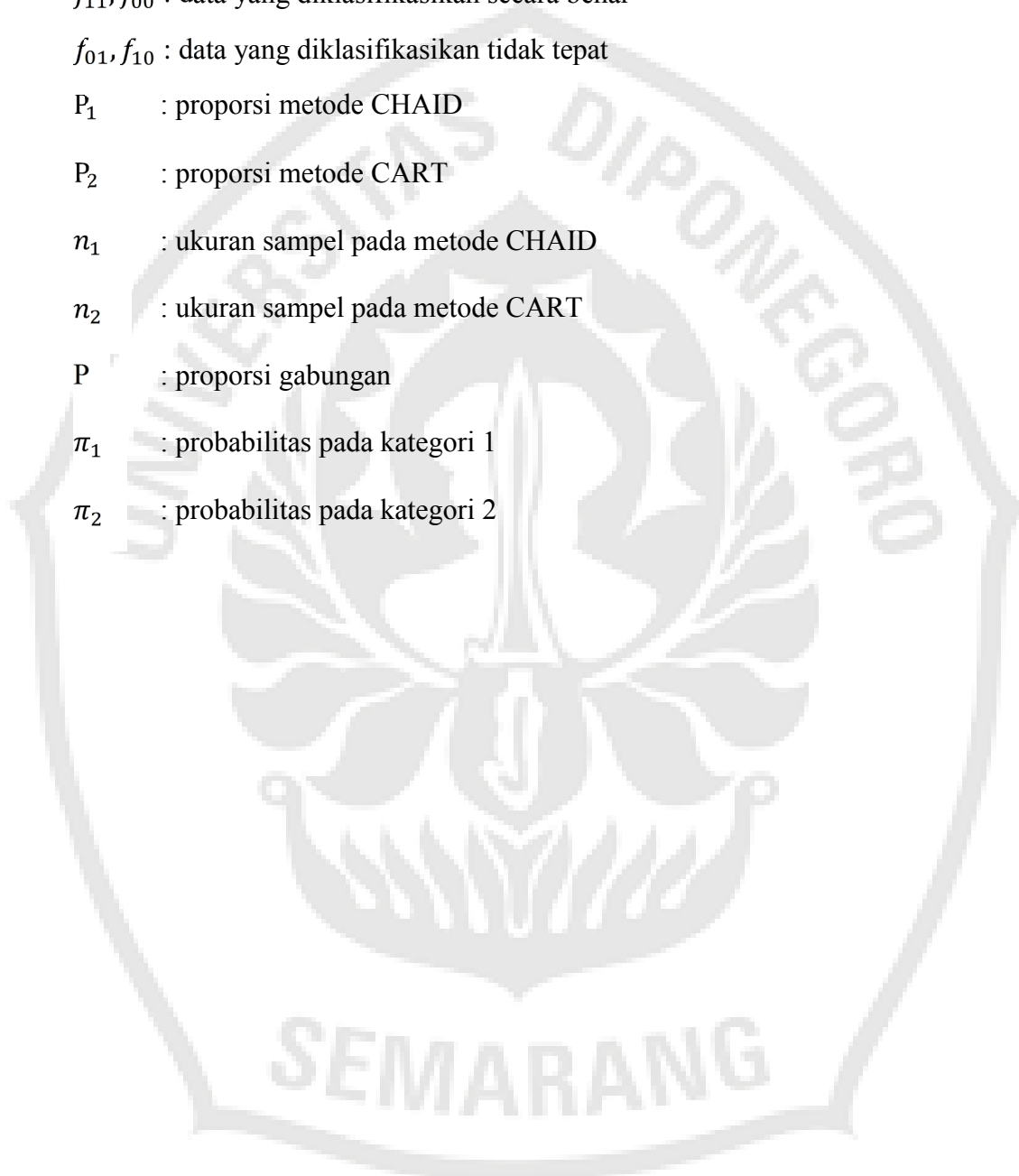
n_1 : ukuran sampel pada metode CHAID

n_2 : ukuran sampel pada metode CART

P : proporsi gabungan

π_1 : probabilitas pada kategori 1

π_2 : probabilitas pada kategori 2



DAFTAR TABEL

	Halaman
Tabel 1. Matriks konfusi untuk klasifikasi dua kelas	25
Tabel 2. Sub tabulasi silang untuk kategori 1 dan kategori 2	39
Tabel 3. Sub tabulasi silang untuk kategori 2 dan kategori 3	40
Tabel 4. Keputusan hasil penggabungan kategori	41
Tabel 5. Sub tabulasi silang untuk penggabungan kategori baru	42
Tabel 6. Tabulasi silang status hubungan dalam rumah tangga dengan status kerja	43
Tabel 7. Tabulasi silang jenis kelamin dengan status kerja	45
Tabel 8. Tabulasi silang usia dengan status kerja	46
Tabel 9. Tabulasi silang status status kelengkapan pasangan hidup dengan status kerja	47
Tabel 10. Tabulasi silang pendidikan dengan status kerja	49
Tabel 11. Tabulasi silang status pendidikan (kategori 3 dan kategori 1) dengan status kerja	50
Tabel 12. Tabulasi silang status pendidikan (kategori 3 dan kategori 2) dengan status kerja	51
Tabel 13. Tabulasi silang status pendidikan (kategori 1 dan kategori 2) dengan status kerja	52
Tabel 14. Tabulasi silang status pelatihan kerja dengan status kerja	52
Tabel 15. Uji independensi variabel independen dan variabel dependen	54
Tabel 16. Uji independensi untuk jenis kelamin laki-laki.....	55

Tabel 17. Uji independensi untuk jenis kelamin laki-laki usia produktif ..	56
Tabel 18. Uji independensi untuk jenis kelamin laki-laki usia produktif dan tidak ada pasangan	56
Tabel 19. Akurasi ketepatan klasifikasi Chaid	59
Tabel 20. Peluang kelas bekerja dan tidak bekerja	60
Tabel 21. Peluang simpul kiri dan simpul kanan	61
Tabel 22. Peluang tiap kelas pada simpul kiri dan simpul kanan	61
Tabel 23. Nilai <i>goodness of split</i> pemilah pertama	63
Tabel 24. Nilai <i>goodness of split</i> untuk jenis kelamin perempuan	64
Tabel 25. Nilai <i>goodness of split</i> untuk jenis kelamin perempuan usia produktif	65
Tabel 26. Nilai <i>goodness of split</i> untuk jenis kelamin perempuan usia produktif dan tidak ada pasangan	66
Tabel 27. Nilai <i>goodness of split</i> untuk jenis kelamin perempuan usia produktif, tidak ada pasangan, dan tidak pernah pelatihan kerja	66
Tabel 28. Nilai <i>goodness of split</i> untuk jenis kelamin perempuan usia produktif, tidak ada pasangan, tidak pernah pelatihan kerja, dan buka kepala rumah tangga	67
Tabel 29. Hasil proses pemangkasan pohon	72
Tabel 30. Akurasi ketepatan klasifikasi CART	74

DAFTAR GAMBAR

	Halaman
Gambar 1. Struktur pohon klasifikasi CHAID	16
Gambar 2. Struktur pohon klasifikasi CART	18
Gambar 3. Pohon klasifikasi T	22
Gambar 4. <i>Branch</i> T_{t_2}	22
Gambar 5. <i>Branch</i> T- T_{t_2}	23
Gambar 6. Diagram alir penelitian	30
Gambar 7. Pembentukan klasifikasi dengan analisis CHAID	32
Gambar 8. Pembentukan klasifikasi dengan analisis CART	33
Gambar 9. Persentase angkatan kerja yang bekerja dan tidak bekerja	34
Gambar 10. Status kerja menurut hubungan dalam rumah tangga	35
Gambar 11. Status kerja menurut jenis kelamin	35
Gambar 12. Status kerja menurut kelompok usia	36
Gambar 13. Status kerja menurut kelengkapan pasangan hidup	37
Gambar 14. Status kerja menurut status pendidikan	37
Gambar 15. Status kerja menurut pelatihan kerja	38
Gambar 16. Pemecahan simpul menjadi simpul kiri dan simpul kanan	62

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data angkatan kerja Kota Semarang tahun 2014	78
Lampiran 2. Pengolahan <i>software</i> SPSS untuk nilai uji chi-square masing-masing variabel independen	79
Lampiran 3. Output klasifikasi angkatan kerja dengan metode CHAID ..	
3a. Hasil pengklasifikasian dan ketepatan klasifikasi	86
3b. Pohon klasifikasi CHAID	87
Lampiran 4. Output klasifikasi angkatan kerja dengan metode CART	
4a. Hasil pengklasifikasian pada pohon klasifikasi sebelum pemangkasan	88
4b. Hasil pengklasifikasian pada pohon klasifikasi setelah pemangkasan	90

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pembangunan nasional ke arah yang lebih baik terus diupayakan oleh setiap negara tidak terkecuali di Indonesia. Dalam mendukung proses pembangunan ini banyak program yang harus dikerjakan. Banyak faktor yang mempengaruhi program-program pembangunan tidak berjalan dengan baik, salah satunya adalah mengenai ketenagakerjaan. Setiap tahun jumlah penduduk semakin bertambah dan tentu saja hal ini mempengaruhi banyak aspek.

Pertumbuhan penduduk selalu dialami oleh negara-negara berkembang, salah satunya adalah Indonesia. Indonesia mengalami tingkat pertumbuhan penduduk yang pesat dari tahun ke tahun. Perkembangan penduduk yang pesat ini membawa akibat salah satunya pada tingkat pertumbuhan angkatan kerja. Tingkat pertumbuhan angkatan kerja yang berubah mengakibatkan perubahan pada Tingkat Partisipasi Angkatan Kerja (TPAK).

Angkatan kerja dapat dibedakan menjadi penduduk yang bekerja dan tidak bekerja (pengangguran). Kegiatan bekerja didefinisikan sebagai kegiatan ekonomi dengan menghasilkan barang atau jasa yang dilakukan oleh seseorang dengan maksud memperoleh atau membantu memperoleh pendapatan atau keuntungan, paling sedikit satu jam (tidak terputus) dalam seminggu yang lalu. Kegiatan tersebut termasuk pula kegiatan pekerja tak dibayar yang membantu dalam suatu usaha atau kegiatan ekonomi. Pengangguran meliputi penduduk yang sedang

mencari pekerjaan, atau mempersiapkan suatu usaha, atau merasa tidak mungkin mendapatkan pekerjaan, atau sudah punya pekerjaan tetapi belum mulai bekerja (BPS Provinsi Jawa Tengah, 2015).

Data hasil survei angkatan kerja nasional oleh BPS Provinsi Jawa Tengah tahun 2014 mencatat, jumlah angkatan kerja di Jawa Tengah Agustus 2014 sebesar 17,55 juta orang, bertambah sekitar 23 ribu orang dibanding angkatan kerja Agustus 2013 sebesar 17,52 juta orang. Peningkatan jumlah angkatan kerja ini tentu saja akan berdampak pada status kerjanya, apakah bekerja atau tidak bekerja (pengangguran).

Penciptaan lapangan kerja diharapkan menjadi solusi atas dampak peningkatan jumlah angkatan kerja. Menurut Widiyanti (1987), lapangan kerja merupakan kunci dari kesejahteraan individu dalam satu masyarakat. Menurut Bakir dan Manning (1984), penciptaan kesempatan kerja merupakan sasaran kebijakan yang harus dilakukan oleh pemerintah. Ketersediaan data angkatan kerja yang di dalamnya terdapat pengelompokan penduduk bekerja dan tidak bekerja dapat membantu pemerintah mengambil tindakan yang efektif.

Menurut Prasetyo (2012), klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Klasifikasi dalam ilmu statistika dapat dilakukan dalam berbagai metode. Metode-metode tersebut diantaranya adalah CHAID dan CART. Metode CHAID dan CART ini merupakan metode eksplorasi data berbentuk pohon keputusan. Beberapa penelitian terkait dengan metode ini yaitu, Pratiwi dan Zain (2014) menyatakan bahwa jenis kelamin, pendidikan terakhir, usia, status dalam rumah tangga dan status perkawinan mempengaruhi pengangguran terbuka di

provinsi Sulawesi Utara. Dhea (2013) melakukan penelitian menggunakan metode CART dan CHAID untuk klasifikasi pasien penderita Diabetes Melitus.

Metode CHAID merupakan algoritma pohon klasifikasi yang pertama kali dikenalkan pada sebuah artikel yang berjudul "*An Exploratory Technique for Investigating Large Quantities of Categorical Data*" oleh Dr. G. V. Kass (1980). Menurut Gallagher *et al.* (2000), CHAID merupakan suatu teknik iteratif yang menguji variabel-variabel independen secara individual yang digunakan dalam klasifikasi dan menyusunnya pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya. Metode CART dikembangkan oleh Leo Breiman, Jerome H. Freidman, Richard A. Olshen, dan Charles J. Stone pada tahun 1980-an. Metode CART merupakan suatu metodologi statistik untuk analisis klasifikasi, baik untuk variabel dependen kategorik maupun kontinu. Metode CART akan menghasilkan pohon klasifikasi bila variabel dependennya kategorik dan menghasilkan pohon regresi bila variabel dependennya kontinu. Pada metode CART ini, data akan dieksplorasi untuk mengetahui variabel-variabel independen yang berpengaruh dan mengelompokkan data tersebut ke dalam kategori-kategori yang ada pada variabel dependen. Sama halnya dengan metode CART, cara kerja CHAID bertujuan untuk menduga variabel-variabel independen yang signifikan terhadap variabel respon atau dependennya.

Berdasarkan uraian tersebut, pengklasifikasian dengan menggunakan metode CHAID dan CART akan diterapkan pada penulisan tugas akhir ini menggunakan data angkatan kerja Kota Semarang tahun 2014.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Melakukan klasifikasi status kerja dari angkatan kerja Kota Semarang dengan menggunakan metode CHAID dan CART.
2. Mengetahui ketepatan klasifikasi dari metode CHAID dan CART.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah data hasil Survei Angkatan Kerja Nasional Tahun 2014 yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Tengah dengan variabel-variabel yang digunakan adalah status hubungan dalam rumah tangga, jenis kelamin, usia, status perkawinan, status pendidikan, dan status pelatihan kerja.

1.4 Tujuan

Tujuan dari penelitian tugas akhir ini adalah sebagai berikut:

1. Mendapatkan hasil klasifikasi dari metode CHAID dan CART
2. Mengetahui ketepatan klasifikasi dari metode CHAID dan CART pada data angkatan kerja Kota Semarang tahun 2014