

**KLASIFIKASI STATUS KABUPATEN/KOTA DI PULAU JAWA
MENGUNAKAN METODE C4.5, CHAID (CHI-SQUARED
AUTOMATIC INTERACTION DETECTION), DAN CART
(CLASSIFICATION AND REGRESSION TREE)**



oleh

RESPATI AYU SYARASWATI

M0112069

SKRIPSI

ditulis dan diajukan untuk memenuhi sebagian persyaratan memperoleh gelar
Sarjana Sains Matematika

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SEBELAS MARET
SURAKARTA**

2017

KLASIFIKASI STATUS KABUPATEN/KOTA DI PULAU JAWA
MENGUNAKAN METODE C4.5, CHAID (CHI-SQUARED AUTOMATIC
INTERACTION DETECTION), DAN CART (CLASSIFICATION AND
REGRESSION TREE)

SKRIPSI

RESPATI AYU SYARASWATI

NIM. M0112069

dibimbing oleh

Pembimbing I

Pembimbing II

Drs. Isnandar Slamet, M.Sc., Ph.D.
NIP. 19660328 199203 1 001

Bowo Winarno, S.Si., M.Kom.
NIP. 19810430 200812 1 001

telah dipertahankan di hadapan Dewan Penguji

dan dinyatakan telah memenuhi syarat

pada hari Selasa, 24 Januari 2017.

Dewan Penguji

Jabatan	Nama dan NIP	Tanda Tangan	Tanggal
Ketua	Dr. Dewi Retno Sari S., S.Si., M.Kom. NIP. 19700720 199702 2 001		07-02-2017
Sekretaris	Drs. Santoso Budi Wiyono, M.Si NIP. 19620203 199103 1 001		06-02-2017
Anggota Penguji	Drs. Isnandar Slamet, M.Sc., Ph.D. NIP. 19660328 199203 1 001		27-01-2017
	Bowo Winarno, S.Si., M.Kom. NIP. 19810430 200812 1 001		26-01-2017

Disahkan

di Surakarta pada tanggal 08 FEB 2017

Kepala Program Studi Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Sebelas Maret Surakarta



Di Wibowo, S.Si., M.Si.

NIP. 19681110 199512 1 001

ABSTRAK

Respati Ayu Syaraswati. 2017. KLASIFIKASI STATUS KABUPATEN/KOTA DI PULAU JAWA MENGGUNAKAN METODE C4.5, CHAID (CHI-SQUARED AUTOMATIC INTERACTION DETECTION), DAN CART (CLASSIFICATION AND REGRESSION TREE). Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Sebelas Maret.

Salah satu indikator keberhasilan perekonomian suatu wilayah dapat diukur berdasarkan tingkat pertumbuhan ekonomi. Untuk mengukurnya, dapat digunakan nilai PDRB (Produk Domestik Regional Bruto). Pulau Jawa memiliki kontribusi terhadap Produk Domestik Bruto (PDB) yang paling besar. Namun, tidak semua kabupaten/kota di Pulau Jawa memiliki kontribusi PDRB yang sama. Metode C4.5, CHAID, dan CART merupakan metode yang digunakan dalam klasifikasi dengan pendekatan nonparametrik. Metode C4.5 dan CHAID menghasilkan pohon nonbiner, sedangkan metode CART menghasilkan pohon biner.

Tujuan penelitian ini adalah menerapkan klasifikasi status kabupaten/kota di Pulau Jawa dan menentukan faktor yang berpengaruh dari klasifikasi status kabupaten/kota di Pulau Jawa. Data yang digunakan dalam penelitian ini merupakan data sekunder pada klasifikasi status kabupaten/kota dengan tipologi Klassen. Tipologi Klassen digunakan untuk mengetahui gambaran tentang pola pertumbuhan daerah. Variabel dependen dalam penelitian ini adalah status kabupaten/kota yang terbagi menjadi empat kategori kuadran, yaitu kuadran I, kuadran II, kadran III, dan kuadran IV. Variabel independen yang digunakan ada sebanyak 14 variabel.

Berdasarkan hasil penelitian ini diperoleh 17 aturan klasifikasi dengan metode C4.5, 7 aturan klasifikasi dengan metode CHAID, dan 5 aturan klasifikasi dengan metode CART. Terdapat beberapa faktor yang memengaruhi klasifikasi status kabupaten/kota di Pulau Jawa. Faktor yang memengaruhi klasifikasi status kabupaten/kota di Pulau Jawa dengan metode C4.5 adalah laju pertumbuhan ekonomi, pendapatan sektor pertambangan dan penggalian, pendapatan sektor listrik, gas, dan air bersih, pendapatan sektor bangunan, angka melek huruf, dan luas wilayah. Faktor yang memengaruhi klasifikasi status kabupaten/kota di Pulau Jawa dengan metode CHAID adalah laju pertumbuhan ekonomi, pendapatan sektor industri pengolahan, pendapatan sektor listrik, gas, dan air bersih, persentase penduduk miskin, serta luas wilayah. Sedangkan faktor yang memengaruhi klasifikasi status kabupaten/kota di Pulau Jawa dengan metode CART adalah laju pertumbuhan ekonomi, pendapatan sektor pertanian, pendapatan sektor listrik, gas, dan air bersih, serta pendapatan sektor industri pengolahan.

Kata kunci : *klasifikasi, C4.5, CHAID, CART*

ABSTRACT

Respati Ayu Syaraswati. 2017. CLASSIFICATION OF STATUS OF THE REGION ON JAVA ISLAND USING C4.5, CHAID (CHI-SQUARED AUTOMATIC INTERACTION DETECTION), AND CART (CLASSIFICATION AND REGRESSION TREE) METHODS. Faculty of Mathematics and Nature Sciences. Sebelas Maret University.

The indicator of region economic success can be measured by economic growth, presented by value of Gross Regional Domestic Product (GRDP). Java island has the biggest GDP contribution toward the Indonesian government, but not all of the region gives equality contribution. The C4.5, CHAID, and CART methods can be used for classifying the status of the region with nonparametric approach. C4.5 and CHAID methods are non-binary decision tree, meanwhile CART method is binary decision tree.

The purposes of this research are to apply the classification and to determine the factors that influence on classification of the region. This research use secondary data about classification of status of the region on Java island with Klassen typology. Klassen typology is used to overview the region growth pattern. The dependent variable is status of the region which is divided into four categories based on Klassen typology, i.e quadrant I, quadrant II, quadrant III, and quadrant IV. There are fourteen independent variables.

The result shows there are 17 classification rules with C4.5 method, 7 classification rules with CHAID method, and 5 classification rules with CART method. There are many factors that influence the classification of status of the region on Java Island. The factors that influence the classification of status of the region on Java Island based on C4.5 method are economic growth rate, income of mining and quarrying sector, income of electricity, gas, and water sector, income of construction sector, the literacy rate, and area. The factors that influence the classification of status of the region on Java Island based on CHAID method are economic growth rate, income of manufacturing sector, income of electricity, gas, and water sector, percentage of poor people, and area. While factors that influence the classification of status of the region on Java island based on CART method are economic growth rate, income of agriculture sector, income of electricity, gas, and water sector, and income of manufacturing sector.

Keywords : *classification, C4.5, CHAID, CART*

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas limpahan rahmatNya sehingga penulis dapat menyelesaikan skripsi ini. Ucapan terimakasih penulis sampaikan kepada

1. Drs. Isnandar Slamet, M.Sc., Ph.D. sebagai Pembimbing I yang telah memberikan arahan, bimbingan, motivasi, arahan dalam hal penulisan skripsi, pemilihan tema, pendalaman materi, dan pengambilan data,
2. Bowo Winarno, S.Si., M.Kom. sebagai Pembimbing II yang telah memberikan bimbingan, motivasi, arahan dalam membuat program, pendalaman materi, dan penyusunan alur penulisan.

Semoga skripsi ini dapat bermanfaat.

Surakarta, Januari 2017

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
ABSTRAK	iii
<i>ABSTRACT</i>	iv
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
II LANDASAN TEORI	4
2.1 Tinjauan Pustaka	4
2.1.1 Tipologi Klassen	5
2.1.2 Klasifikasi	6
2.1.3 Metode C4.5	6
2.1.4 Metode <i>CHAID</i>	8
2.1.5 Metode <i>CART</i>	11
2.1.6 Pengujian Akurasi	13
2.2 Kerangka Pemikiran	14

III METODE PENELITIAN	15
IV HASIL DAN PEMBAHASAN	18
4.1 Deskripsi Atribut Data	18
4.2 Klasifikasi Status Kabupaten/Kota Menggunakan Metode C4.5 . .	19
4.2.1 Menentukan Simpul Akar	19
4.2.2 Menentukan Simpul Cabang	20
4.2.3 Pengujian Akurasi Pohon Klasifikasi C4.5	27
4.2.4 Interpretasi Pohon Klasifikasi C4.5	28
4.3 Klasifikasi Status Kabupaten/Kota Menggunakan Metode <i>CHAID</i>	33
4.3.1 Tahap Penggabungan	33
4.3.2 Tahap Pemisahan	39
4.3.3 Tahap Penghentian	40
4.3.4 Pengujian Akurasi Pohon Klasifikasi <i>CHAID</i>	41
4.3.5 Interpretasi Pohon Klasifikasi <i>CHAID</i>	42
4.4 Klasifikasi Status Kabupaten/Kota Menggunakan Metode <i>CART</i>	44
4.4.1 Pemilihan Pemilah	45
4.4.2 Pelabelan Kelas	49
4.4.3 Proses Penghentian Pohon	50
4.4.4 Proses Pemangkasan Pohon	50
4.4.5 Pohon Klasifikasi Optimal	52
4.4.6 Pengujian Akurasi Pohon Klasifikasi <i>CART</i>	53
4.4.7 Interpretasi Pohon Klasifikasi <i>CART</i>	54
V PENUTUP	57
5.1 Kesimpulan	57
5.2 Saran	58
DAFTAR PUSTAKA	59
LAMPIRAN	

DAFTAR TABEL

2.1	Tipologi Klassen	5
4.1	Tipologi Klassen	19
4.2	Perhitungan Simpul Akar	20
4.3	Perhitungan Simpul Cabang 2.1	21
4.4	Perhitungan Simpul Cabang 3.1	22
4.5	Perhitungan Simpul Cabang 3.2	23
4.6	Perhitungan Simpul Cabang 3.3	24
4.7	Perhitungan Simpul Cabang 2.2	25
4.8	Perhitungan Simpul Cabang 3.4	26
4.9	Hasil Pengujian Data <i>Learning</i>	27
4.10	Hasil Pengujian Data <i>Testing</i>	28
4.11	Aturan Klasifikasi dengan Metode C4.5	29
4.12	Tabulasi Silang Sektor Pertanian dan Status	33
4.13	Tabulasi Silang Kategori 1 dan 2	34
4.14	Tabulasi Silang Kategori 2 dan 3	35
4.15	Tabulasi Silang Kategori 3 dan 4	36
4.16	Nilai χ^2 Tahap Penggabungan Variabel Pendapatan Sektor Pertanian	36
4.17	Tabulasi Silang Kategori Penggabungan	37
4.18	Nilai χ^2 Tahap Penggabungan	39
4.19	Variabel yang Signifikan	40
4.20	Hasil Pengujian Data <i>Learning CHAID</i>	41
4.21	Hasil Pengujian Data <i>Testing CHAID</i>	42
4.22	Aturan Klasifikasi dengan Metode <i>CHAID</i>	43

4.23	Nilai $Gini_{split}$ Sektor Pertanian	47
4.24	Pemilihan Simpul Akar	48
4.25	Perhitungan Proses Pemangkasan Simpul 4	51
4.26	Nilai Penduga Pengganti	52
4.27	Hasil Pengujian Data <i>Learning CART</i>	54
4.28	Hasil Pengujian Data <i>Testing CART</i>	54
4.29	Aturan Klasifikasi dengan Metode <i>CART</i>	55

DAFTAR GAMBAR

4.1	Pohon Klasifikasi C4.5	27
4.2	Pohon Klasifikasi <i>CHAID</i>	41
4.3	Pemilihan simpul Akar <i>CART</i>	49
4.4	<i>Terminal Node</i> 8 dan 9	50
4.5	Pohon Klasifikasi Maksimal	50
4.6	Pemangkasan Pohon Maksimal	51
4.7	Pemangkasan Pohon Optimal	53