

**PERBANDINGAN METODE *CHI-SQUARE AUTOMATIC
INTERACTION DETECTION (CHAID)* DAN *CLASSIFICATION
AND REGRESSION TREE (CART)* DALAM MENENTUKAN
KLASIFIKASI ALUMNI UIN SUNAN KALIJAGA
BERDASARKAN MASA STUDI**

SKRIPSI

**untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-1**

Program Studi Matematika



Diajukan oleh:

Nuriyah

07610041

Kepada

Program Studi Matematika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sunan Kalijaga

Yogyakarta

2013



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi

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 : Nuriyah
NIM : 07610041
Judul Skripsi : Perbandingan Metode *Chi-Square Automatic Interaction Detection* (CHAID) dan *Classification And Regression Tree* (CART) Dalam Menentukan Klasifikasi Alumni UIN Sunan Kalijaga Berdasarkan Masa Studi

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 bidang Matematika.

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

Yogyakarta, 30 November 2012

Pembimbing

Epha Diana Supandhi, S. Si., M. Sc

NIP. 19750912 200801 2 015



Universitas Islam Negeri Sunan Kalijaga

FM-UINSK-BM-05-07/R0

PENGESAHAN SKRIPSI/TUGAS AKHIR

Nomor : UIN.02/D.ST/PP.01.1/318/2013

Skripsi/Tugas Akhir dengan judul : Perbandingan Metode *Chi - Square Automatic Interaction Detection* (CHAID) dan *Classification and Regression Tree* (CART) dalam Menentukan Klasifikasi Alumni UIN Sunan Kalijaga Berdasarkan Masa Studi

Yang dipersiapkan dan disusun oleh :
Nama : Nuriyah
NIM : 07610041
Telah dimunaqasyahkan pada : Jum'at, 15 Januari 2013
Nilai Munaqasyah : A
Dan dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga

TIM MUNAQASYAH :

Ketua Sidang

Epha Diana Supandi, S.Si. M.Sc
NIP. 19750912 200001 2 015

Penguji I

M. Farhan Qudratullah, M.Si
NIP.19790922 200801 1 011

Penguji II

Ki Haryadi, S.Si. MPH
NIDN. 0515057601

Yogyakarta, 1 Februari 2013

UIN Sunan Kalijaga

Fakultas Sains dan Teknologi

Dekan



Prof. Drs. H. Akh. Minhaji, M.A, Ph.D
NIP. 19580919 198603 1 002

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini, saya :

Nama : Nuriyah
NIM : 07610041
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul skripsi : Perbandingan Metode *Chi-Square Automatic Interaction Detection (CHAID)* dan *Classification And Regression Tree (CART)* dalam Menentukan Klasifikasi Alumni UIN Sunan Kalijaga Berdasarkan Masa Studi

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak di kemudian hari terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka saya bertanggungjawab sepenuhnya.

Yogyakarta, 04 Desember 2012



KATA PENGANTAR

Alhamdulillah Rabbil'alamin, segala puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya dengan memberikan kemudahan, kekuatan, kelancaran, dan kesehatan dalam penyusunan karya tulis skripsi yang berjudul “Perbandingan Metode *Chi-Square Automatic Interaction Detection* (CHAID) dan *Classification and Regression Tree* (CART) Dalam Menentukan Klasifikasi Alumni UIN Sunan Kalijaga Berdasarkan Masa Studi”. Penulisan skripsi ini bertujuan untuk memenuhi mata kuliah Tugas Akhir sebagai syarat untuk memperoleh derajat kesarjanaan. Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu penulis dalam penyusunan skripsi ini, baik secara langsung maupun tidak langsung, terutama penulis tujukan kepada:

1. Bapak Prof. Drs. H. Akh. Minhaji, M.A, Ph.D., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
2. Bapak Moch. Abrori, M. Kom., selaku Ketua Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Ibu Epha Diana Supandi, S. Si., M.Sc., selaku dosen pembimbing yang telah meluangkan waktu memberikan bimbingan, arahan, bantuan dan ilmu dalam menyelesaikan skripsi ini.
4. Bapak dan Ibuku tercinta, terima kasih atas keringat, air mata, semangat, senyum, do'a serta kesediaan menjadi tempatku meneduhkan jiwa dan raga selama ini. Bapak, Ibu, gelar ini ku haturkan untuk kalian.

5. Simbah kakung dan simbah putri, *matursembahnuwun*.
6. Adikku yang ku banggakan Kurniawan “wawan” Achmad, terus menatap hari esok, perjalananmu masih panjang, kejar cita-cita mu setinggi langit. Mba Bibah, yang selama ini memberi masukan.
7. Teman-teman saya di Matematika 2007. Terutama Afi, Tika, Uha, yang telah menemani malam-malam insomnia saya, bertemu mereka adalah hadiah. Nessa, Zana, Rina, Tsulus, Sulis, Dewi, Murni, Cici, Nisa, Nela... *big thanks!*
8. Agus Miftakhus Surur, mas Adit, dan mas Mahmudi, terimakasih arahannya, *gumawo...!*
9. Ama, Rita, dan Muya sahabat saya sedari SMA yang selalu siap mendengar curhatan sampah saya hingga impian-impian saya. *Maturnuwun..*
10. Teman-teman kos saya: mba Hajar, Nayla, Ratna, mba Lucy, Aida,dll.
11. Semua pihak yang telah membantu dalam penyelesaian skripsi ini, yang tidak bisa penulis sebutkan satu persatu.

Penulis menyadari masih banyak kekurangan dalam penyusunan skripsi ini, sehingga saran maupun kritik yang membangun sangat penulis harapkan agar proses berkembangnya ilmu bisa terus berlangsung. Semoga skripsi ini dapat memberi manfaat bagi kita semua.

Yogyakarta, 13 November 2012

Penulis

Nuriyah
NIM. 07610041

HALAMAN PERSEMBAHAN

Hari takkan indah tanpa mentari dan rembulan, begitu juga hidup takkan indah tanpa tujuan, harapan serta tantangan. Meski terasa berat, namun manisnya hidup justru akan terasa, apabila semuanya terlalui dengan baik, meski harus memerlukan pengorbanan.

Kupersembahkan karya kecil ini, untuk cahaya hidup, yang senantiasa ada saat suka maupun duka, selalu setia mendampingi, saat kulemah tak berdaya (Bapak dan Ibu tercinta) yang selalu memanjatkan doa kepada putri tercinta dalam setiap sujudnya. Terima kasih untuk semuanya.

Untuk ribuan tujuan yang harus dicapai, untuk jutaan impian yang akan dikejar, untuk sebuah pengharapan, agar hidup jauh lebih bermakna, karena tragedi terbesar dalam hidup bukanlah kematian tapi hidup tanpa tujuan. Teruslah bermimpi untuk sebuah tujuan, pastinya juga harus diimbangi dengan tindakan nyata, agar mimpi dan juga angan, tidak hanya menjadi sebuah bayangan semu.

HALAMAN MOTTO

*“Setiap orang semua sama, memiliki 24 jam dalam satu hari.
Kesuksesan ditentukan oleh seberapa baiknya Anda
memanfaatkan waktu Anda”
(Tung Desem Waringin)*

*” Jangan biarkan rasa takut gagal membuatmu berhenti
mencoba”
(A Cinderella Story)*

*“Perubahan tidak akan terjadi kalau kita hanya menunggu
orang lain atau waktu lain tuk melakukannya”
(Obama)*

*“Ada tempat di dunia ini yang tidak terbuat dari batu..
Sesuatu dalam diri kita yang tidak bisa mereka dapatkan,
tidak bisa mereka sentuh. Sesuatu itu adalah.....
Harapan.”
(The Shawshank Redemptions)*

DAFTAR ISI

HALAMAN JUDUL	i
SURAT PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
SURAT PERNYATAAN KEASLIAN	iv
KATA PENGANTAR	v
HALAMAN PERSEMBAHAN	vii
HALAMAN MOTTO	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
DAFTAR SIMBOL	xvii
ABSTRAK	xix
BAB I PENDAHULUAN	1
1. 1 Latar Belakang Masalah	1
1. 2 Batasan Masalah	5
1. 3 Rumusan Masalah	5
1. 4 Tujuan Penelitian	6
1. 5 Manfaat Penelitian	6
1. 6 Tinjauan Penelitian	7
1. 7 Sistematika Penulisan	10
BAB II LANDASAN TEORI	12
2.1 Klasifikasi	12
2.2 Analisis Non-Parametrik	14
2.3 Analisa Multivariat	15
2.4 Probabilitas	16
2.5 Distribusi Probabilitas	17
2. 5. 1 Variabel Acak Diskrit	18
2. 5. 2 Variabel Acak Kontinu	18

2.6 Distribusi Peluang Bersyarat	19
2.7 Probabilitas <i>Chi-Square</i> (χ^2)	19
2.8 Uji Independensi <i>Chi-Square</i> (χ^2)	22
2.9 Skala Pengukuran	25
2.10 Penyesuaian Bonferroni (<i>Adjustment Bonferroni</i>)	26
2.11 Model Keputusan Pohon	27
2.12 <i>Binnary Recursive Partition</i>	29
2.13 <i>Missing Values</i>	29
BAB III METODE PENELITIAN	30
3.1 Jenis Penelitian dan sumber Data	30
3.2 Metode Pengumpulan Data	30
3.3 Variabel Penelitian	30
3.4 Metodologi Penelitian	31
3.5 Metode Analisis Data	31
3.6 Alat Pengolah Data	32
BAB IV PEMBAHASAN	35
4.1 Analisis CHAID	37
4.1.1 Variabel-variabel dalam CHAID	38
4.1.2 Algoritma CHAID	39
4.1.3 Koreksi Bonferroni	45
4.1.4 Diagram Pohon Algoritma CHAID	46
4.2 Analisi CART	47
4.2.1 Karakteristik dalam CART	47
4.2.2 Algoritma CART	48
4.2.3 Langkah Pembentukan Pohon CART	49
4.2.4 Diagram Pohon Algoritma CART	53
4.3 Klasifikasi Alumni UIN Sunan Kalijaga	54
BAB V STUDI KASUS	57
5. 1 Deskripsi Profil Responden	57
5. 2 Analisis Data dengan Metode CHAID	62
5.2.1 Metode Analisis Data	62

5.2.2 Interpretasi Output SPSS	62
5. 3 Analisis Data dengan Metode CART	67
5.3.1 Metode Analisis Data	67
5.2.2 Interpretasi Output SPSS	67
5. 4 Perbandingan Hasil Klasifikasi Metode CHAID dan CART ...	75
BAB VI KESIMPULAN DAN SARAN	76
6.1 Kesimpulan	76
6.2 Saran	78
DAFTAR PUSTAKA	79
CURRICULUM VITAE	81
LAMPIRAN	90

DAFTAR TABEL

Tabel 1.1:	Pemetaan Tinjauan Pustaka	7
Tabel 2.1:	Struktur Data Uji <i>Chi-square</i>	23
Tabel 2.2:	Tabel Peluang Kejadian	24
Tabel 4.1:	Tabulasi Silang IPK dengan Masa Studi	40
Tabel 4.2:	Pasangan Penggabungan Variabel	42
Tabel 4.3:	Pasangan Penggabungan Variabel	43
Tabel 4.4:	Data Jenis Sekolah, IPK, Nilai TOEC, dan Masa Studi Alumni Mahasiswa UIN Sunan Kalijaga	49
Tabel 4.5:	Daftar Calon Cabang Mutakhir Masalah Alumni UIN Sunan Kalijaga	49
Tabel 4.6:	Perhitungan Nilai Kesesuaian Untuk Calon Cabang	51
Tabel 4.7:	Pengkategorian Alumni dan Pemberian Kode Berdasarkan Jenis Sekolah	55
Tabel 4.8:	Pengkategorian Alumni dan Pemberian Kode Berdasarkan Nilai IPK	55
Tabel 4.9:	Pengkategorian Alumni dan Pemberian Kode Berdasarkan Nilai TOEC	56
Tabel 4.10:	Pengkategorian Alumni dan Pemberian Kode Berdasarkan Nilai IKLA	56
Tabel 4.11:	Pengkategorian Alumni dan Pemberian Kode Berdasarkan Nilai ICT	56
Tabel 5.1:	Ringkasan Pembentukan Model dengan Metode CHAID	62
Tabel 5.2:	Segmentasi Alumni UIN Sunan Kalijaga	65
Tabel 5.3:	Persentase Setiap Alumni Segmen Alumni UIN	66
Tabel 5.4:	Klasifikasi dari Metode CHAID	67
Tabel 5.5:	Ringkasan Pembentukan Model dengan Metode CART ...	68
Tabel 5.6:	Segmentasi Alumni UIN Sunan Kalijaga	72
Tabel 5.7:	Persentase Setiap Segmen Alumni UIN Sunan Kalijaga ...	73

Tabel 5.8:	Klasifikasi dari Metode CART	74
Tabel 6.1:	Perbandingan Hasil Klasifikasi Metode CHAID dan CART	77

DAFTAR GAMBAR

Gambar 2.1: <i>Decesion Tree</i>	13
Gambar 3.1: Diagram Alur	32
Gambar 4.1: <i>Decision Tree</i>	36
Gambar 4.2 : Diagram Pohon CHAID	46
Gambar 4.3: Diagram Pohon CART	53
Gambar 5.1: Komposisi Alumni Berdasarkan Asal Daerah	57
Gambar 5.2: Komposisi Alumni Berdasarkan nilai ICT	58
Gambar 5.3: Komposisi Alumni Berdasarkan Nilai IKLA	58
Gambar 5.4: Komposisi Alumni Berdasarkan nilai IPK	59
Gambar 5.5: Komposisi Alumni Berdasarkan Jenis Kelamin	60
Gambar 5.6: Komposisi Alumni Berdasarkan Asal Sekolah	60
Gambar 5.7: Komposisi Alumni Berdasarkan Masa Studi	61
Gambar 5.8: Komposisi Alumni Berdasarkan nilai TOEC	61
Gambar 5.9: Diagram Pohon oleh Metode CHAID pada Kasus Masa Studi	64
Gambar 5.10: Diagram Pohon oleh Metode CART pada Kasus Masa Studi	70

DAFTAR LAMPIRAN

Lampiran 1: Deskripsi Data	82
Lampiran 2: Langkah-Langkah dalam Pengujian SPSS	85
Lampiran 3: Output Pemrosesan Data dan Pembentukan Model dengan CHAID	88
Lampiran 4: Output Pemrosesan Data dan Pembentukan Model dengan CART	90

DAFTAR SIMBOL

X	: variabel independen
Y	: variabel dependen
b	: baris
k	: kolom
O_{11}	: banyaknya pengamatan dengan sifat A_1 dan B_1
O_{ij}	: banyaknya pengamatan dengan sifat A_i , dan B_j , $i=1, 2, \dots, b$, dan $j=1, 2, \dots, k$
n_i	: banyaknya pengamatan dengan sifat A_i , $i=1, 2, \dots, b$
n_j	: banyaknya pengamatan dengan sifat B_j , $j=1, 2, \dots, k$
t_L	: calon cabang kiri dari noktah keputusan t
t_R	: calon cabang kanan dari noktah keputusan t
E_{ij}	: nilai harapan untuk masing-masing sel dalam tabel <i>chi-square</i>
$\varphi(s t)$: nilai besaran kesesuaian

PERBANDINGAN METODE CHAID (*CHI-SQUARE AUTOMATIC INTERACTION DETECTION*) DAN CART (*CLASSIFICATION AND REGRESSION TREE*) DALAM MENENTUKAN KLASIFIKASI ALUMNI UIN SUNAN KALIJAGA BERDASARKAN MASA STUDI

ABSTRAK

**Oleh:
Nuriyah**

Metode klasifikasi yang sering digunakan adalah metode klasifikasi berstruktur pohon, diantaranya yaitu CART (*Classification and Regression Tree*) dan CHAID (*Chi-square Automatic Interaction Detection*). Kedua metode ini memiliki perbedaan dalam pembentukan pohon klasifikasi. CART memilih pemisah terbaiknya berdasarkan tingkat kehomogenan yang maksimum. Sedangkan CHAID menggunakan uji *Chi-Square* untuk menentukan pemisah terbaik pada setiap langkahnya. Penyusunan skripsi ini bertujuan untuk menjelaskan prosedur klasifikasi dengan metode CHAID dan metode CART. Metode CHAID dan CART diterapkan pada kasus klasifikasi alumni UIN Sunan Kalijaga berdasarkan masa studinya.

Data yang digunakan dalam penelitian adalah alumni yang diwisuda pada periode I dan II TA 2011/2012. Variabel dependennya adalah masa studi. Sedang variabel independen yang digunakan adalah: jenis kelamin, asal daerah, jenis sekolah, IPK, nilai TOEC, nilai IKLA, dan nilai ICT.

Metode CHAID menghasilkan 5 segmen dimana ada 3 variabel independen yang signifikan terhadap model, yaitu: IPK, ICT, dan asal daerah. Metode CHAID menyebutkan segmen yang memiliki ketepatan waktu masa studi terbesar adalah segmen dimana yang alumninya memiliki IPK antara $3,00 \leq \text{IPK} \leq 3,50$ dan berasal dari luar DIY. Sedangkan CART menghasilkan 6 segmen dimana ada 7 variabel independen yang signifikan terhadap model, yaitu: IPK, IKLA, ICT, jenis kelamin, TOEC, dan asal daerah. Metode CART menyebutkan segmen yang memiliki ketepatan waktu masa studi terbesar adalah alumni yang memiliki IPK antara $3,00 \leq \text{IPK} \leq 3,50$. Perbedaan dalam klasifikasi dengan metode CHAID dengan metode CART tidak jauh berbeda. Hal ini tidak terlihat pada ketepatan hasil klasifikasi. Ketepatan hasil CHAID 63,3%, sedangkan CART ketepatan hasil klasifikasinya 66,6%.

Kata kunci: CHAID (*Chi-square Automatic Interaction Detection*), CART (*Classification and Regression Tree*), klasifikasi.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengklasifikasian merupakan salah satu metode statistika untuk mengelompokkan atau mengklasifikasikan suatu data yang disusun secara sistematis. Masalah klasifikasi sering dijumpai dalam kehidupan sehari-hari. Baik itu pengklasifikasian data pada bidang akademik, kesehatan, segmentasi pasar, maupun pada bidang lainnya. Masalah tersebut dapat diselesaikan dengan metode klasifikasi. Namun, pada penyelesaian masalah klasifikasi perlu diperhatikan dalam memilih metode klasifikasi yang tepat. Sebagai contoh dalam masalah kesehatan, apabila ingin mengelompokkan pasien yang terkena penyakit diabetes dan tidak terkena diabetes. Jika mengelompokkan pasien yang terkena penyakit diabetes ke dalam kelompok pasien yang tidak terkena penyakit diabetes merupakan kesalahan yang dapat berakibat cukup fatal (Widagdo, 2010).

Dalam segmentasi pasar klasifikasi digunakan untuk membagi suatu populasi individu ke dalam kelompok atau segmen yang lebih kecil karakteristiknya satu dengan yang lain. Segmen-segmen yang terbentuk tersebut akan digunakan sebagai acuan untuk menentukan target pasar. Misalnya ada sebuah bank ingin menyalurkan produk pelayanan mereka. Bank tersebut akan mempelajari data nasabah sebelumnya dari produk pembiayaan. Dengan membagi nasabah berdasarkan status kredit lancar atau macet, maka akan bisa diketahui kelompok-kelompok nasabah pembiayaan dengan latar belakang yang berbeda pada masing-

masing kelompok. Pihak bank tinggal memutuskan kelompok nasabah mana yang akan dipilih sebagai sasaran utama pemasaran dengan mempertimbangkan status kredit. Diharapkan dengan penentuan kelompok sasaran pemasaran yang tepat, resiko kredit bisa diminimalkan.

Dalam statistika ada beberapa metode klasifikasi yang digunakan untuk melakukan klasifikasi data seperti: analisis *Chi-Square Automatic Interaction Detection* (CHAID), *Clasification and Regresion Tree* (CART), *Neural Network* (NN), *K-Nearest Neighbour* (KNN), dan *Multivariate Adaptive Regression Spline* (MARS). Masing-masing metode tersebut memiliki kelebihan dan kelemahan. Dalam penelitian ini yang akan dibahas ialah metode klasifikasi dengan metode CHAID (*Chi-Square Automatic Interaction Detection*) dibandingkan dengan algoritma *Clasification and Regresion Tree* (CART).

Metode CHAID (*Chi-Square Automatic Interaction Detection*) merupakan salah satu tipe dari metode AID (*Automatic Interaction Detection*). Metode AID adalah suatu teknik untuk menganalisis kelompok data berukuran besar dengan membaginya menjadi sub-sub kelompok yang tidak saling tumpang tindih (Kass 1982, diacu dalam Soemartojo 2002) yang diperuntukkan bagi data dengan peubah penjelas berskala ratio atau interval. Metode ini terutama dikembangkan untuk menelusuri keterkaitan struktural dalam data survei (Fielding, 1977).

Metode CHAID merupakan teknik eksplorasi nonparametrik untuk menganalisis sekumpulan data yang berukuran besar dan cukup efisien untuk menduga peubah-peubah penjelas yang paling signifikan terhadap peubah respon. Interaksi antar peubah juga dapat dideteksi melalui metode ini (Du toit *et al.*

1977), sehingga diharapkan metode CHAID dapat menjadi alternatif analisis untuk menduga faktor-faktor yang menentukan suatu respon kategorik.

Dalam menganalisis kelompok data, metode CHAID memisahkan data kedalam beberapa kelompok secara bertahap. Tahap pertama diawali dengan membagi data menjadi beberapa kelompok berdasarkan satu variabel independen yang pengaruhnya paling signifikan terhadap variabel dependent. Masing-masing kelompok yang diperoleh diperiksa secara terpisah untuk membaginya menjadi beberapa kelompok berdasarkan variabel independen dan seterusnya hingga pada akhirnya diperoleh kelompok-kelompok pengamatan yang memiliki variabel dependent dan variabel independen tertentu yang berkaitan.

Metode CHAID digunakan bila variabel dependent berskala nominal atau ordinal dengan kriteria statistik uji *chi-square* pada setiap pemisahannya. Inti dari metode ini adalah membagi data menjadi kelompok-kelompok yang lebih kecil berdasarkan keterkaitan antara variabel dependen dengan variabel independen. Metode CHAID hanya efektif bila diterapkan dengan pada data dengan pengamatan yang sangat banyak (Du Toit, S. H. C., A. G. W. Steyn & R. H. Stumph, 1986).

Dalam penelitian ini metode CHAID akan diterapkan pada data wisuda mahasiswa UIN Sunan Kalijaga periode TA 2011-2012. Variabel dependent yang digunakan adalah masa studi, dengan pengkategorian tepat waktu dan tidak tepat waktu. Yang dikatakan tepat waktu adalah alumni yang masa studinya kurang dari atau tepat 8 semester, dan yang lebih dari 8 semester disebut tidak tepat waktu. Metode CHAID nantinya akan membentuk segmentasi dari data alumni

berdasarkan hubungan antara variabel dependen yaitu masa studi dengan variabel-variabel independen yang merupakan profil dari alumni. Hasil segmentasi akan menunjukkan segmen/kelompok alumni mana saja yang memiliki kelulusan tepat waktu ataupun tidak tepat waktu. Pihak kampus bisa menggunakan hasil segmentasi ini untuk mengetahui bagaimana ciri mahasiswa yang dikhawatirkan memiliki masa studi yang tidak tepat waktu dengan melihat ciri-ciri pada segmen alumni yang lulus tidak tepat waktu. Jika sudah diketahui tipe mahasiswa mana yang dikhawatirkan lulus tidak tepat waktu, pihak kampus bisa mengambil sebuah kebijakan yang tepat sehingga angka kelulusan tidak tepat waktu bisa dikurangi. Sebagai analisis pembandingan dari metode CHAID akan digunakan algoritma CART (*Classification and Regression Tree*).

(Komalasari, 2007) CART (*Classification and Regression Tree*) adalah salah satu metode atau algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan. Metode ini dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen dan Charles J. Stone sekitar tahun 1980-an. (Breiman *et al.*, 1993), CART merupakan metodologi statistik nonparametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk variabel respon kategorik maupun kontinu. CART menghasilkan suatu pohon klasifikasi jika variabel responnya kategorik, dan menghasilkan pohon regresi jika variabel responnya kontinu.

CART dapat menyeleksi variabel-variabel dan interaksi-interaksi variabel yang paling penting dalam menentukan hasil atau variabel dependennya. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai

penciri dari suatu pengklasifikasian. Metode ini merupakan metode yang bisa diterapkan untuk himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan biner.

Karena metode CHAID (*Chi-Square Automatic Interaction Detection*) dan algoritma CART (*Classification and Regression Tree*) hampir memiliki sifat yang sama, maka pada tulisan ini akan dilakukan perbandingan performansi CHAID dan CART. Dari tulisan ini akan dicari info seberapa besar perbedaan output yang dihasilkan dari kedua metode tersebut, apakah berbeda signifikan atau tidak.

1.2 Batasan Masalah

Masalah yang akan dibahas pada tugas akhir ini dibatasi pada variabel. Variabel yang digunakan adalah jenis kelamin, asal daerah, jenis sekolah, IPK, nilai TOEC, nilai IKLA, dan nilai ICT. Data yang digunakan adalah data wisuda tahun 2011-2012.

1.3 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana prosedur analisis klasifikasi dengan metode CHAID dan CART?
2. Bagaimana hasil penerapan metode CHAID dan CART pada kasus klasifikasi alumni UIN Sunan Kalijaga berdasarkan masa studi?
3. Bagaimana hasil perbandingan klasifikasi dengan metode CHAID dengan algoritma CART pada kasus masa studi alumni mahasiswa UIN?

1.4 Tujuan Penelitian

Tujuan yang akan dicapai antara lain:

1. Menjelaskan bagaimana prosedur metode CHAID dan CART sebagai salah satu metode dalam teknik klasifikasi.
2. Menerapkan metode CHAID dan CART untuk klasifikasi alumni mahasiswa UIN Sunan Kalijaga berdasarkan masa studinya.
3. Membandingkan hasil klasifikasi metode CHAID dengan metode CART.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini antara lain:

- a. Menambah referensi tentang ilmu matematika, khususnya statistika dalam hal ini tentang multivariat dengan model *Chi-Square Automatic Interaction Detection* (CHAID) dan metode *Classification and Regression Tree* (CART)
- b. Hasil yang diperoleh dari klasifikasi dapat diterapkan oleh Universitas untuk mengetahui segmentasi alumni mahasiswa UIN Sunan Kalijaga berdasarkan masa studinya. Segmentasi mahasiswa adalah pembagian mahasiswa menjadi kelompok-kelompok yang lebih kecil yang memiliki ciri/ karakteristik yang berbeda dan mungkin memerlukan perlakuan yang berbeda pula (Permana, 2011), sehingga setelah melihat segmen dari kelompok maka yang dikhawatirkan tidak tepat waktu masa studinya, pihak dari Universitas dapat mengambil kebijakan untuk mengurangi mahasiswa yang tidak lulus tepat waktu.

1.6 Tinjauan Pustaka

Metode yang digunakan dalam penyusunan penelitian ini adalah metode studi literatur, yaitu studi yang dilakukan dengan mempelajari beberapa buku, jurnal, karya ilmiah, dan hasil penelitian sebelumnya yang berkaitan dengan penelitian ini.

Tabel 1.1 Pemetaan Tinjauan Pustaka

No	Tahun	Peneliti	Judul	Data	Institusi
1.	2006	Yohanes Sondang Kunto dan Siti Nurul Hasana	Analisis CHAID sebagai Alat Bantu Statistika untuk Segmentasi Pasar	Status Kredit dan Data Demografis Nasabah (2002-2003)	Universitas Kristen Petra, Surabaya.
2.	2006	Siti Nur Sholihah	Metode CHAID (<i>Chi-Square Automatic Interaction Detection</i>) untuk Klasifikasi Data Mining (Studi Kasus Klasifikasi Saham pada Kategori Harga Perubahan Saham di Bursa Saham Efek Kuala Lumpur (Kuala Lumpur Stock Exchange) Malaysia)	Bursa Saham Efek Kuala Lumpur	Universitas Gadjah Mada, Yogyakarta
3.	2009	Atina Chalisa	Pemodelan Credit Scoring untuk Kredit Konsumtif Bank Menggunakan Algoritma CHAID dan CART	Data kredit konsumtif Bank Perkreditan Rakyat	Universitas Gadjah Mada, Yogyakarta
4.	2011	Husein Permana	Klasifikasi dengan Metode CHAID (<i>Chi-Square Automatic Interaction Detection</i>) dan Penerapannya pada klasifikasi Alumni FMIPA UNY	Data alumni FMIPA UNY	Universitas Negeri Yogyakarta Yogyakarta

Kunto dan Hasana (2006), dalam jurnalnya yang berjudul “Analisis CHAID sebagai Alat Bantu Statistika untuk Segmentasi Pasar” menjelaskan analisis

CHAID yang diterapkan untuk menentukan segmentasi nasabah koperasi Syari'ah Al-Hidayah. Variabel-variabel yang digunakan adalah data demografis dari nasabah, yaitu: jenis kelamin, usia, pendidikan terakhir, pekerjaan, penghasilan rata-rata keluarga, dan ukuran keluarga. Hasil yang diperoleh ada empat segmen yang berbeda. Siti Nur Sholihah (2006), pada penelitian ini meneliti “Metode CHAID (*Chi-Square Automatic Interaction Detection*) untuk Klasifikasi Data Mining (Studi Kasus Klasifikasi Saham pada Kategori Harga Perubahan Saham di Bursa Saham Efek Kuala Lumpur (Kuala Lumpur *Stock Exchange*) Malaysia)” yang berisi tentang klasifikasi data mining dengan menggunakan analisis CHAID. Variabel-variabel yang digunakan adalah: ref, high, low, last, change, dan volume. Data yang dipakai adalah data transaksi dalam 1 hari yaitu pada tanggal 30 desember 2005 dengan 1323 emiten terdaftar sebagai anggota bursa efek di Kuala Lumpur. Dari hasil analisis klasifikasi yang telah dilakukan ternyata perubahan harga saham yang terjadi berkaitan erat dengan harga penutupan kemarin dan volume penjualan lembar saham. Hampir seluruh perubahan harga saham yang terjadi hanya pada level tetap.

Atina Chalisa (2009), melakukan penelitian berjudul “Pemodelan *Credit Scoring* untuk Kredit Konsumtif Bank Menggunakan Algoritma CHAID dan CART”. Dari penelitiannya diharapkan mampu memberi analisa kredit yang tujuannya untuk menilai kondisi calon debitur. Dengan analisa tersebut diharapkan dapat memperkecil resiko default, yang disebabkan ketidakmampuan debitur memenuhi kewajibannya sesuai yang disepakati sebagaimana tertuang dalam perjanjian kredit. Data yang digunakan adalah salah satu Bank Perkreditan Rakyat

di Yogyakarta dari oktober 2002 sampai desember 2007 sebanyak 1962 buah data nasabah. Variabel-variabel yang digunakan dalam penelitian ini: kolektibilitas, pendapatan, nilai kredit, nilai jaminan, jaminan, persen bunga, sifat, penggunaan, golongan, jangka waktu, cara, penjamin, jenis kredit, dan lokasi. Dari aturan klasifikasi berdasarkan algoritma CHAID, karakteristik yang mempunyai probability of default paling besar adalah debitur dengan persen bunga 1,34-1,75. Mempunyai jenis kredit berupa kredit profesi atau KPM, lokasi tempat tinggal debitur tersebut berada di Yogyakarta.

Selain itu, digunakan juga penelitian dari Husein Permana (2011) yang berjudul “Klasifikasi dengan Metode CHAID (*Chi-Square Automatic Interaction Detection*) dan Penerapannya pada klasifikasi Alumni FMIPA UNY” yang berisi tentang perbandingan metode CHAID dengan metode Biner pada kasus masa studi alumni FMIPA UNY. Tujuan yang diharapkan dari penelitian ini adalah untuk mencari hasil segmentasi mahasiswa FMIPA UNY berdasarkan masa studi yang mereka tempuh. Variabel-variabel yang digunakannya adalah: masa studi, jenis kelamin, asal daerah, jalur masuk, program studi, dan IPK semester 1. Hasil dari metode CHAID menyebutkan bahwa segmen yang kelulusan tidak tepat waktunya paling besar adalah alumni yang IP Semester 1 kurang dari sama dengan 2,50, berasal dari program studi non kependidikan dan jalur masuk non reguler. Dan alumni yang kelulusan tepat waktunya paling besar adalah alumni dengan IP Semester 1 lebih dari sama dengan 3, 51.

1.7 Sitematika Penulisan

Untuk memberikan gambaran menyeluruh mengenai metode *Chi-Square Automatic Interaction Detection* (CHAID) dan *Classification and Regression Tree* (CART), penelitian ini terdiri dari:

Bab I berisi pendahuluan, yang membahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, tinjauan pustaka dan sistematika penulisan.

Bab II berisi dasar teori yang akan digunakan sebagai dasar pembahasan dari penulisan ini meliputi: klasifikasi, analisis non-parametrik, analisis multivariat, probabilitas, distribusi probabilitas, distribusi peluang bersyarat, probabilitas *chi-square* (χ^2), uji independensi *chi-square* (χ^2), skala pengukuran, penyesuaian Bonferroni (*Bonferroni adjustment*), *binary recursive partition*, dan *missing values*.

Bab III berisi metode penelitian, yang membahas mengenai metode pengumpulan data, metode analisis data dan alat pengolah data.

Bab IV berisi penjelasan mengenai perbandingan metode *Chi-Square Automatic Interaction Detection* (CHAID) dan metode *Classification and Regression Tree* (CART).

Bab V berisi studi kasus, yaitu perbandingan metode *Chi-Square Automatic Interaction Detection* (CHAID) dengan metode *Classification and Regression Tree* (CART) yang diaplikasikan pada klasifikasi data alumni UIN sunan kalijaga tahun 2011-2012 periode I dan II berdasarkan masa studi.

Bab VI kesimpulan, berisi kesimpulan yang dapat diambil dari pembahasan permasalahan yang ada dan saran-saran yang berkaitan dengan penelitian sejenis di masa yang akan datang.

BAB VI

KESIMPULAN DAN SARAN

6.1. Kesimpulan

Berdasarkan pembahasan sebelumnya, maka dapat disampaikan beberapa kesimpulan sebagai berikut:

1. Prosedur

a. CHAID

Secara ringkas, tahapan CHAID dibagi menjadi tiga tahap:

- Penggabungan (*Merging*)
- Pemisahan (*splitting*)
- Penghentian (*stopping*)

b. CART

Tahapan CART diringkas menjadi tiga, yaitu:

- Menyusun calon cabang (*candidate split*). Penyusunan ini dilakukan terhadap seluruh variabel independen secara lengkap.
- Menilai kinerja keseluruhan calon cabang yang ada pada daftar calon cabang mutakhir dengan jalan menghitung nilai besaran kesesuaian ($\phi(s|t)$)
- Memilih calon cabang yang akan dijadikan cabang dengan cara memilih nilai $\phi(s|t)$ terbesar.

2. Hasil

a. CHAID

Hasil segmentasi dari metode CHAID menyebutkan bahwa segmen/ kelompok alumni yang kelulusan tepat waktunya paling besar adalah alumni yang memiliki IPK antara $3,01 \leq \text{IPK} \leq 3,50$ dan asal daerahnya berasal dari luar DIY. Sedangkan alumni yang kelulusan tidak tepat waktu paling banyak adalah alumni yang memiliki IPK antara $2,51 \leq \text{IPK} \leq 3,00$.

b. CART

Hasil segmentasi dari metode CART menyebutkan bahwa segmen/ kelompok alumni yang kelulusan tepat waktunya paling besar adalah alumni yang memiliki IPK antar $3,01 \leq \text{IPK} \leq 3,50$.

c. Perbandingan CHAID dengan CART

Tabel 6.1 Perbandingan hasil klasifikasi metode CHAID dengan metode
CART

Metode CHAID	Metode CART
Ketepatan hasil klasifikasi 63,3%	Ketepatan hasil klasifikasi 66,6%
Metode CHAID menghasilkan 3 variabel independen yang signifikan terhadap model, yaitu IPK, ICT, dan asal daerah	Metode CART menghasilkan 7 variabel independen yang signifikan terhadap model, yaitu IPK, IKLA, ICT, jenis kelamin, TOEC, dan asal daerah
Hasil klasifikasi metode CHAID membagi alumni UIN Sunan Kalijaga menjadi 5 segmen. Segmen yang memiliki kelulusan tepat waktu terbesar adalah alumni dengan memiliki IPK antara $3,01 \leq \text{IPK} \leq 3,50$ dan asal daerahnya berasal dari luar DIY. Sedangkan alumni yang kelulusan tidak tepat waktu paling banyak adalah alumni yang memiliki IPK antara $2,51 \leq \text{IPK} \leq 3,00$.	Hasil klasifikasi metode CART membagi alumni UIN Sunan Kalijaga menjadi 6 segmen. Segmen yang memiliki kelulusan tepat waktu terbesar adalah alumni yang memiliki IPK antar $3,01 \leq \text{IPK} \leq 3,50$. Sedangkan alumni yang kelulusan tidak tepat waktu paling banyak adalah alumni yang memiliki IPK kurang dari sama dengan $2,51 \leq \text{IPK} \leq 3,00$.

6.2. Saran

Setelah membahas klasifikasi dengan metode CHAID dan CART, saran yang dapat penulis sampaikan adalah sebagai berikut:

- a. Hasil segmentasi bisa digunakan oleh pihak universitas untuk mengetahui ciri-ciri alumni mana saja yang studinya tepat waktu dan mahasiswa yang masa studinya tidak tepat waktu. Setelah diketahui ciri-ciri dari masing-masing segmen, dari pihak Universitas bisa digunakan untuk mengambil keputusan guna mengantisipasi/meminimalisir mahasiswa yang studinya tidak tepat waktu.
- b. Perlu dilakukan penelitian lanjutan dengan metode yang hampir serupa. Misalnya memakai metode *Neural Network* (NN), *K-Nearest Neighbour* (KNN), atau *Multivariate Adaptive Regression Spline* (MARS).
- c. Penerapan metode ke bidang yang lain. Misalnya kesehatan, pemasaran, dan perbankan.

DAFTAR PUSTAKA

- Breiman, L., Friedman, J., Olsen, R.A., dan Stone, C. (1984). *Classification and Regression Trees*. California: Wadsworth Belmonth.
- Chalisa, Atina. (2009). *Pemodelan Kredit Scoring untuk Kredit Konsumtif Bank Menggunakan Algoritma CHAID dan CART (Studi Kasus Bank Perkreditan Rakyat di Yogyakarta)*. Yogyakarta: Skripsi FMIPA UGM.
- Djarwanto, Drs., Ps. (2004). *Statistik Non Parametrik*. Yogyakarta: BPFE Yogyakarta.
- GSBIPB. (2013, 18 January). *Sebaran Chi-Square dan Kegunaannya*. [online]. Tersedia di <http://gsbipb.com/> (6 Februari 2013)
- Haryatmi, S. (1986). *Analisis Data Statistik*. Jakarta: Karunika Universitas Terbuka.
- Komalasari, Wieta, B. (2007). *Metode Pohon Regresi untuk Eksploratori Data dengan Peubah yang Banyak dan Kompleks*. Jurnal Informatika Pertanian Volume 16 No. 1, Juli 2007
- Kunto, Y. S. dan Hasana, S. N. (2006). *Analisis CHAID sebagai Alat Bantu Statistika Untuk Segmentasi Pasar*. Jurnal Manajemen, vol. 1 No. 2. Surabaya: Universitas Kristen Petra.
- Larose, D. T., (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: Wiley-nterscience.
- Lewis dan Roger, J. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis*. Presented at the 2000 Annual Meeting of Society for Academic Emergency Medicine of Sanfransisco. California.
- Permana , Husein. (2011). *Klasifikasi dengan Metode CHAID (Chi-Squared Automatic Interaction Detection) dan Penerapannya pada Klasifikasi Alumni S1 FMIPA UNY*. Yogyakarta: Skripsi FMIPA UNY.
- Sholihah, S. N. (2006). *Metode CHAID (Chi-Squared Automatic Interaction Detection) untuk Klasifikasi Data Mining (Studi Kasus Klasifikasi Saham pada Kategori Harga Perubahan Saham di Bursa Efek Kuala Lumpur (Kuala Lumpur Stock Exchange) Malaysia)*. Yogyakarta: Skripsi FMIPA UGM.

- Simamora, B. (2005). *Analisis Multivariat Pemasaran*. Jakarta: PT Gramedia Pustaka Utama.
- Spiegel, R. M. (1996). *Statistika Edisi Kedua*. I Nyoman Susila dan Ellen Gunawan (alih bahasa). Jakarta: Penerbit Erlangga.
- Supranto, J. (2004). *Analisis Multivariat "Arti dan Interpretasi"*. Jakarta: Rineka Cipta
- Sarwono, J. (2009). *Statistik Itu Mudah*. Yogyakarta: Andi Offset.
- Siregar, Ir. Syofian., M.M. (2010). *Statistik Deskriptif Untuk Penelitian*. Jakarta: Rajawali Pers.
- Susanto, S & Suryadi, D.(2010). *Pengantar Data Mining*. Yogyakarta: Andi Offset.
- Riawan, Yuli. (2012), "*Tabel Distribusi Normal Z, T, dan Chi Kuadrat*". <http://10111941.blog.unikom.ac.id/tabel-distribusi.52x> (Diakses 20 November 2012)
- Widagdo, K. A. (2010). *Pembentukan Pohon Klasifikasi Biner dengan Algoritma Cart (Clasification and Regression Trees) (studi Kasus Penyakit Diabetes Suku Pima Indian)*. Semarang: Skripsi Universitas Diponegoro.
- Wijaya, Tony. (2010). *Analisis Multivariat*. Yogyakarta: Universitas Atma Jaya.
- Yuzarli, A. "*Model Pengambilan Keputusan*" <http://andri88-blog.blogspot.com/2009/09/model-pengambilan-keputusan.html> (diakses pada tanggal : 22 Juni 2012)

LAMPIRAN

CURRICULUM VITAE

Nama : Nuriyah
TTL : 17 April 1989
Jenis Kelamin : Perempuan
Alamat : Jogoresan, Purwodadi, Purworejo
Nama Ayah : Sugiyarto
Nama Ibu : Rusmiyati
Riwayat Pendidikan : TK Achmad Yani Th. 1994 – 1995
SD Negeri Jogoresan Th. 1995 – 2001
MTs An-nawawi Berjan Th. 2001 – 2004
MA An-nawawi Berjan Th. 2004 – 2007
UIN Sunan Kalijaga, Yogyakarta Th. 2007 – sekarang.

Lampiran 1

Deskripsi data

Frequencies Table

		Statistics							
		Masa Studi	Jenis Kelamin	Asal Daerah	Jenis Sekolah	IPK	TOEC	KLA	ICT
N	Valid	841	841	841	841	841	841	841	841
	Missing	0	0	0	0	0	0	0	0
Mean		1.4911	1.5268	1.7015	1.4411	3.1177	1.6338	1.4721	1.7860
Median		1.0000	2.0000	2.0000	1.0000	3.0000	2.0000	1.0000	2.0000
Mode		1.00	2.00	2.00	1.00	3.00	2.00	1.00	1.00
Std. Deviation		.50022	.49958	.45785	.59496	.53379	.61062	.54291	.99252
Variance		.250	.250	.210	.354	.285	.373	.295	.985
Minimum		1.00	1.00	1.00	1.00	2.00	1.00	1.00	1.00
Maximum		2.00	2.00	2.00	3.00	4.00	3.00	4.00	5.00

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Masa Studi * Jenis Kelamin	841	100.0%	0	.0%	841	100.0%
Masa Studi * Asal Daerah	841	100.0%	0	.0%	841	100.0%
Masa Studi * Jenis Sekolah	841	100.0%	0	.0%	841	100.0%
Masa Studi * IPK	841	100.0%	0	.0%	841	100.0%
Masa Studi * TOEC	841	100.0%	0	.0%	841	100.0%
Masa Studi * KLA	841	100.0%	0	.0%	841	100.0%
Masa Studi * ICT	841	100.0%	0	.0%	841	100.0%

Masa Studi * Jenis Kelamin Crosstabulation

Count

		Jenis Kelamin		Total
		L	P	
Masa Studi	Tepat Waktu	190	238	428
	Tidak Tepat Waktu	208	205	413
Total		398	443	841

Masa Studi * Asal Daerah Crosstabulation

Count

		Asal Daerah		Total
		DIY	Luar DIY	
Masa Studi	Tepat Waktu	117	311	428
	Tidak Tepat Waktu	134	279	413
Total		251	590	841

Masa Studi * Jenis Sekolah Crosstabulation

Count

		Jenis Sekolah			Total
		SMU/SMTA	MA	SMK	
Masa Studi	Tepat Waktu	256	144	28	428
	Tidak Tepat Waktu	259	137	17	413
Total		515	281	45	841

Masa Studi * IPK Crosstabulation

Count

		IPK			Total
		2.51-3.00	3.01-3.50	>3.51	
Masa Studi	Tepat Waktu	5	277	146	428
	Tidak Tepat Waktu	71	313	29	413
Total		76	590	175	841

Masa Studi * TOEC Crosstabulation

Count

		TOEC			Total
		<= 400	401-450	>451	
Masa Studi	Tepat Waktu	159	240	29	428
	Tidak Tepat Waktu	208	175	30	413
Total		367	415	59	841

Masa Studi * ICLA Crosstabulation

Count

		IKLA				Total
		<=30	31-50	51-70	>71	
Masa Studi	Tepat Waktu	210	207	11	0	428
	Tidak Tepat Waktu	252	155	5	1	413
Total		462	362	16	1	841

Masa Studi * ICT Crosstabulation

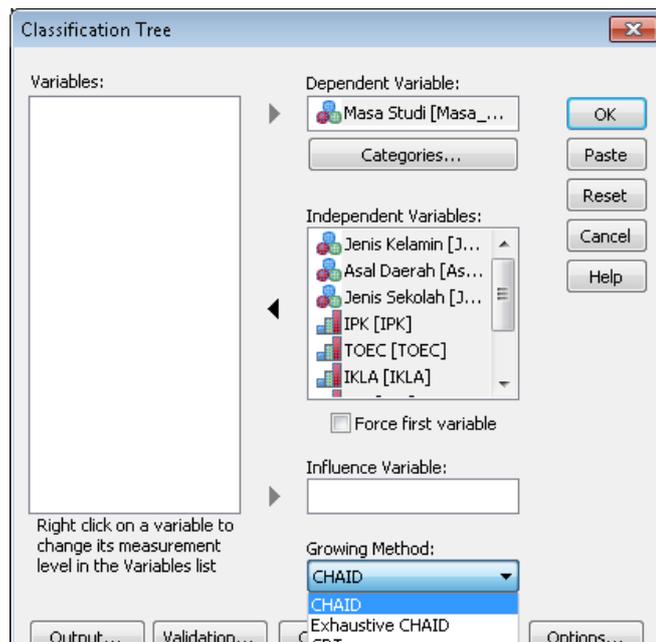
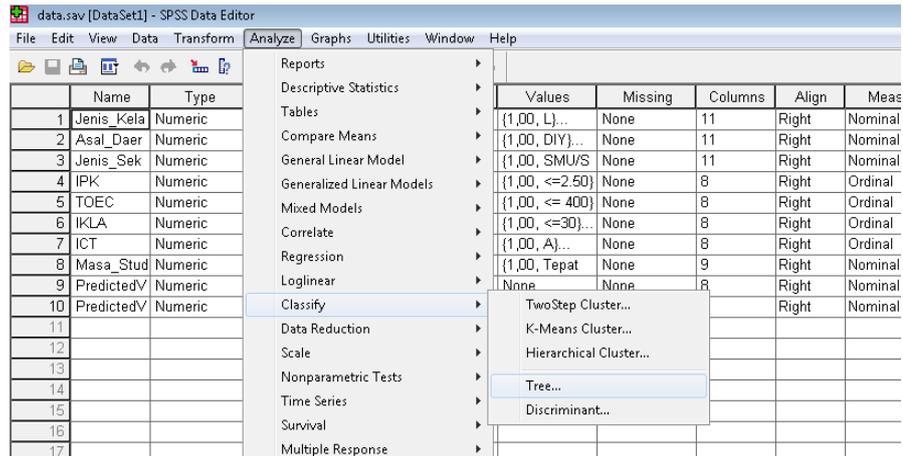
Count

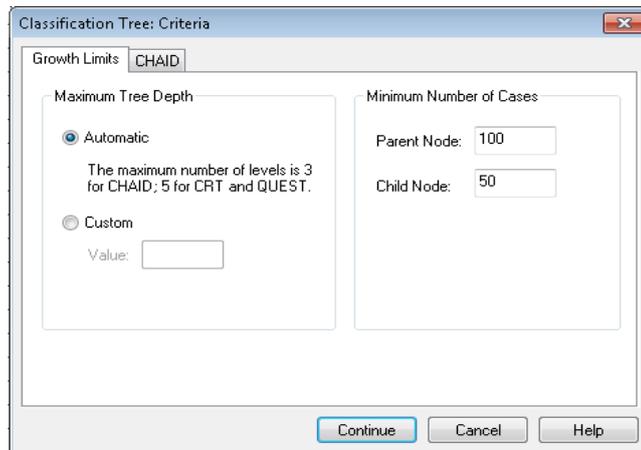
		ICT					Total
		A	B	C	D	E	
Masa Studi	Tepat Waktu	231	166	3	23	5	428
	Tidak Tepat Waktu	167	168	13	54	11	413
Total		398	334	16	77	16	841

Lampiran 2

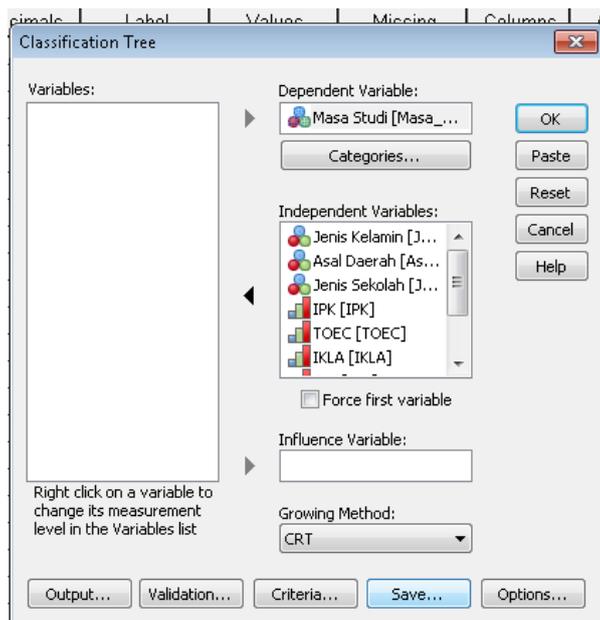
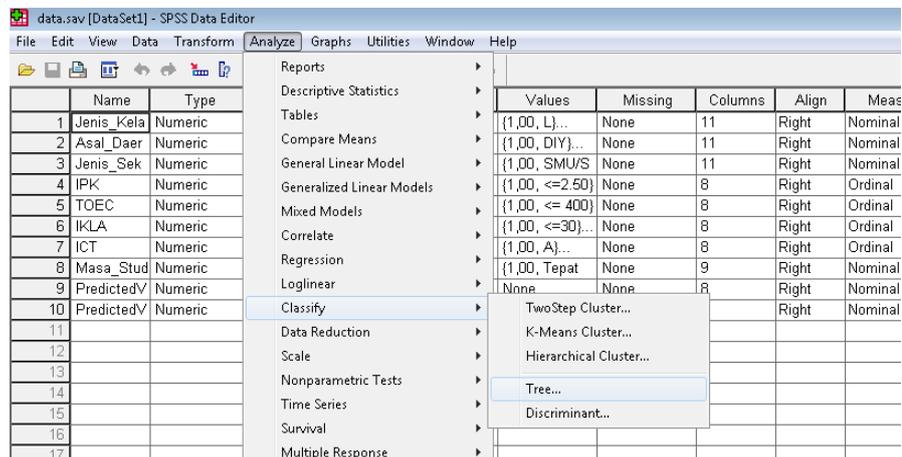
Langkah-langkah pengujian dalam SPSS

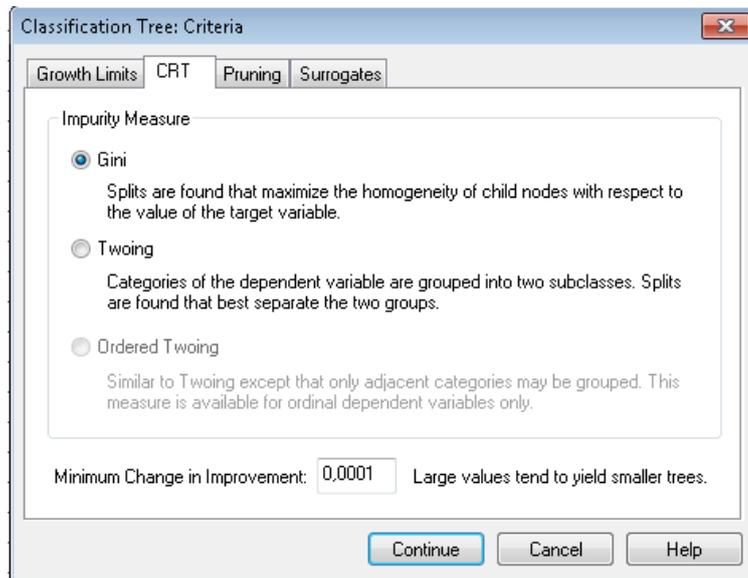
Langkah-langkah CHAID





Langkah-langkah CART





Lampiran 3

Output pemrosesan Data dan pembentukan Model dengan CHAID

Classification Tree

Model Summary

Specifications	Grow ing Method	CHAID		
	Dependent Variable	Masa Studi		
	Independent Variables	Jenis Kelamin, Asal Daerah, Jenis Sekolah, IPK, TOEC, IKLA, ICT		
	Validation	None		
	Maximum Tree Depth		3	
	Minimum Cases in Parent Node		100	
	Minimum Cases in Child Node		50	
	Results	Independent Variables Included	IPK, ICT, Asal Daerah	
		Number of Nodes		8
		Number of Terminal Nodes		5
Depth			2	

Target Category: Tepat Waktu

Gains for Nodes

Node	Node		Gain		Response	Index
	N	Percent	N	Percent		
6	117	13.9%	104	24.3%	88.9%	174.7%
7	58	6.9%	42	9.8%	72.4%	142.3%
4	508	60.4%	255	59.6%	50.2%	98.6%
5	82	9.8%	22	5.1%	26.8%	52.7%
1	76	9.0%	5	1.2%	6.6%	12.9%

Grow ing Method: CHAID

Dependent Variable: Masa Studi

Target Category: Tidak Tepat Waktu**Gains for Nodes**

Node	Node		Gain		Response	Index
	N	Percent	N	Percent		
1	76	9.0%	71	17.2%	93.4%	190.2%
5	82	9.8%	60	14.5%	73.2%	149.0%
4	508	60.4%	253	61.3%	49.8%	101.4%
7	58	6.9%	16	3.9%	27.6%	56.2%
6	117	13.9%	13	3.1%	11.1%	22.6%

Growing Method: CHAID
 Dependent Variable: Masa Studi

Risk

Estimate	Std. Error
.367	.017

Growing Method: CHAID
 Dependent Variable: Masa Studi

Classification

Observed	Predicted		
	Tepat Waktu	Tidak Tepat Waktu	Percent Correct
Tepat Waktu	401	27	93.7%
Tidak Tepat Waktu	282	131	31.7%
Overall Percentage	81.2%	18.8%	63.3%

Growing Method: CHAID
 Dependent Variable: Masa Studi

Lampiran 4

Output pemrosesan Data dan pembentukan Model dengan CART

Classification Tree**Model Summary**

Specifications	Grow ing Method	CART		
	Dependent Variable	Masa Studi		
	Independent Variables	Jenis Kelamin, Asal Daerah, Jenis Sekolah, IPK, TOEC, IKLA, ICT		
	Validation	None		
	Maximum Tree Depth		5	
	Minimum Cases in Parent Node		100	
	Minimum Cases in Child Node		50	
	Results	Independent Variables Included	IPK, IKLA, ICT, Jenis Kelamin, TOEC, Asal Daerah	
		Number of Nodes		11
		Number of Terminal Nodes		6
Depth			5	

Prior Probabilities

	Prior Probability
Masa Studi	
Tepat Waktu	.509
Tidak Tepat Waktu	.491

Priors are obtained from the training sample

Target Category: Tepat Waktu**Gains for Nodes**

Node	Node		Gain		Response	Index
	N	Percent	N	Percent		
2	175	20.8%	146	34.1%	83.4%	163.9%
10	104	12.4%	57	13.3%	54.8%	107.7%
8	268	31.9%	144	33.6%	53.7%	105.6%
9	136	16.2%	54	12.6%	39.7%	78.0%
6	82	9.8%	22	5.1%	26.8%	52.7%
3	76	9.0%	5	1.2%	6.6%	12.9%

Grow ing Method: CART

Dependent Variable: Masa Studi

Target Category: Tidak Tepat Waktu**Gains for Nodes**

Node	Node		Gain		Response	Index
	N	Percent	N	Percent		
3	76	9.0%	71	17.2%	93.4%	190.2%
6	82	9.8%	60	14.5%	73.2%	149.0%
9	136	16.2%	82	19.9%	60.3%	122.8%
8	268	31.9%	124	30.0%	46.3%	94.2%
10	104	12.4%	47	11.4%	45.2%	92.0%
2	175	20.8%	29	7.0%	16.6%	33.7%

Growing Method: CRT

Dependent Variable: Masa Studi

Risk

Estimate	Std. Error
.334	.016

Growing Method: CART

Dependent Variable: Masa Studi

Classification

Observed	Predicted		
	Tepat Waktu	Tidak Tepat Waktu	Percent Correct
Tepat Waktu	347	81	81.1%
Tidak Tepat Waktu	200	213	51.6%
Overall Percentage	65.0%	35.0%	66.6%

Growing Method: CART

Dependent Variable: Masa Studi