

Analisis Pohon Berstruktur Menggunakan Metode CHAID pada Data Respons Ordinal

Dian Cahyawati
Jurusan Matematika FMIPA Universitas Sriwijaya

ABSTRAK

Eksplorasi data merupakan suatu hal yang penting dilakukan sebelum menganalisis data dengan metode lain. Salah satu metode eksplorasi untuk data respons kategorik adalah metode CHAID (Chi-square Automatic Interaction Detection). Tulisan ini berisi tentang kajian teori metode CHAID dan penerapannya yaitu menganalisis faktor-faktor yang berpengaruh pada peningkatan omset usaha anggota koperasi simpan pinjam (KSP) sebagai respons ordinal. Hasil kajian menunjukkan bahwa metode CHAID tidak hanya dapat digunakan sebagai metode eksplorasi tetapi juga mampu mengungkapkan struktur hubungan antara variabel respons kategorik dengan serangkaian variabel penjelas serta interaksi antar variabel penjelas. Struktur hubungan ini digambarkan dengan suatu pohon berstruktur.

Kata Kunci : eksplorasi data, respons ordinal, metode CHAID, pohon berstruktur

ABSTRACT

The data exploration is an important case to do before analyze the data by other method. One of the exploration method that can be used for exploring the categorical response variable is CHAID (Chi-square Automatic Interaction Detection) method. This paper contains CHAID method theory and its application. In this case the method is applied in analyze the influence factors on the turnover increases of the saving-loan cooverative organization members as an ordinal response variable. The result of this studied give an overview that CHAID method not only can be used for exploring the data but also for explaining the relationship structure between a categorical response variable and a large series of categorical explanatory variables which themselves may interact. That relationship structure is shown graphically by a structure tree.

Key Words : data exploration, ordinal response, CHAID method, structure tree.

PENDAHULUAN

Metode eksplorasi data merupakan suatu hal yang sangat penting dilakukan dalam setiap analisis statistika. Salah satu tujuannya adalah untuk mengetahui karakteristik penyebaran data amatan. Pentingnya eksplorasi data dalam setiap analisis statistika, menjadikan metode ini hampir selalu digunakan oleh para peneliti di berbagai bidang. Karena melalui eksplorasi diharapkan dapat menghasilkan gambaran penyebaran data amatan, baik hanya untuk tujuan deskripsi data ataupun untuk melakukan analisis lebih lanjut terhadap data.

Salah satu metode untuk eksplorasi data adalah metode CHAID (*Chi-square Automatic Interaction Detection*) sebagai salah satu teknik untuk menganalisis data-data kategorik dengan jumlah yang besar yaitu ratusan atau ribuan (Kass, 1982) dengan menggunakan kriteria uji *Chi-square*. Berdasarkan hal ini, dirasakan sangat perlu untuk mengkaji konsep atau teori-teori mengenai metode CHAID, penerapan dan juga interpretasinya dalam menganalisis data-data kategorik. Sehingga hasil kajian ini diharapkan dapat memberikan informasi yang

cukup jelas tentang metode CHAID. Selain itu, metode ini juga dapat digunakan oleh peneliti di berbagai bidang ilmu yang ingin mengungkapkan hubungan antara respons kategorik (nominal atau ordinal) dan variabel penjelas yang diamati.

METODE PENELITIAN

1. Metode

Penelitian ini dilakukan melalui langkah-langkah sebagai berikut:

1. Mengkaji aspek-aspek teori metode CHAID, melalui referensi yang diperoleh dari buku-buku literatur, jurnal yang sudah ada maupun penelusuran referensi dari internet.
2. Mengumpulkan data (khususnya respons berskala ordinal) yang akan dianalisis dengan metode CHAID.
3. Mempersiapkan format data amatan
4. Menerapkan metode CHAID terhadap data.
5. Menganalisis hasil metode CHAID.
6. Menginterpretasikan hasil metode CHAID

Analisis CHAID diterapkan melalui makro *Treedisc* yang ditulis dalam paket program SAS versi 6.12.

2. Sumber Data

Data yang digunakan dalam penelitian adalah data sekunder, hasil penelitian Praonco pada tahun 2003 tentang "Pengaruh Lembaga Keuangan Koperasi Simpan Pinjam terhadap Keberhasilan Usaha Anggota" (Studi Kasus di DKI Jakarta). Metode pengambilan sampel yang digunakan adalah metode *Stratified Purposive Sampling*. Survei dilakukan terhadap 120 responden anggota koperasi simpan pinjam (KSP).

HASIL DAN PEMBAHASAN

Pada umumnya, metode eksplorasi data selalu dilakukan dalam setiap penelitian. Hal ini dikarenakan, melalui eksplorasi data akan dapat diketahui gambaran penyebaran data amatan baik hanya untuk sekedar deskripsi data maupun untuk keperluan analisis lebih lanjut terhadap data, seperti ingin mengungkapkan hubungan antara suatu variabel respons dengan serangkaian variabel penjelas.

Skala pengukuran variabel respons menentukan metode analisis yang digunakan. Apabila skala pengukurannya berupa nominal atau ordinal, metode CHAID dapat digunakan baik untuk eksplorasi maupun ingin

mengungkapkan hubungan antara variabel respons dengan variabel-variabel penjelas.

1. Metode CHAID

Metode CHAID (*Chi-square Automatic Interaction Detection*) merupakan salah satu tipe metode AID (*Automatic Interaction Detection*). AID adalah suatu teknik untuk menganalisis kelompok data berukuran besar dengan membaginya menjadi sub-sub kelompok yang tidak saling tumpang tindih (Kass, 1982). Teknik pemecahan kelompok menjadi beberapa sub kelompok dilakukan sedemikian rupa sehingga diperoleh sub-sub kelompok yang secara maksimal saling berbeda (Huba, 2001).

Proses pemecahan dan pengelompokkan dilakukan secara iteratif, dimulai dari variabel penjelas yang mempunyai hubungan paling kuat dengan respons yang ditunjukkan oleh besarnya nilai-*p* (*p-value*) berdasarkan uji Chi-square. Analisis CHAID juga akan melakukan penggabungan kategori-kategori variabel penjelas yang tidak memiliki pengaruh terhadap respons menjadi satu kategori, sehingga hasil penggabungan itu menjadi kategori yang berpengaruh terhadap respons.

a. Algoritma Metode CHAID

Proses pengelompokan dengan metode CHAID ini menggunakan *default* batas nilai- $p = 0.05$, artinya jika ada dua buah kategori/kelompok yang memiliki nilai- $p \geq 0.05$ maka kategori itu akan digabung. Sedangkan untuk pemecahan (*splitting*), menggunakan *default* nilai- $p < 0.10$, artinya jika suatu kategori memberikan nilai- p kurang dari 0.10 , maka kategori itu akan dipecah berdasarkan kategori/kelompok lain.

Melalui algoritma ini, akan dihasilkan variabel penjelas yang berpengaruh terhadap respons mulai dari yang paling tinggi keeratannya hingga yang paling rendah. Sedangkan untuk variabel penjelas yang tidak berpengaruh terhadap respons, maka variabel itu akan dikeluarkan. Pada analisis regresi, proses seperti ini dikenal dengan metode *selection forward* untuk menentukan variabel-variabel yang berpengaruh.

Secara singkat algoritma CHAID yang menghasilkan suatu pohon berstruktur adalah sebagai berikut ([http://www.cbs.wl/en/services/autimp/appendix%201-tree-\(autimp\).pdf](http://www.cbs.wl/en/services/autimp/appendix%201-tree-(autimp).pdf)) :

1. Untuk setiap variabel penjelas cari pasangan kategori yang memiliki nilai- p paling besar.
2. Bandingkan nilai- p itu dengan batas nilai- p yang telah ditetapkan, jika lebih besar maka gabungkan. Jika nilai- p terbesar masih lebih kecil dari batas nilai- p maka tidak ada kategori yang perlu digabungkan. Jika variabel penjelas hanya memiliki dua kategori, dan apabila nilai- p yang ada lebih besar dari batas nilai- p , maka ini dikeluarkan dari model. Lanjutkan proses ini sampai tidak ada lagi nilai pasangan kategori yang mempunyai nilai- p lebih besar dari batas nilai- p . Untuk kategorik nominal, penggabungan dapat dilakukan antara kategori mana saja. Tetapi pada berskala ordinal, penggabungan hanya dapat dilakukan antara kategori-kategori yang berurutan.
Apabila pada proses 2 ini terjadi penggabungan kategori kedalam suatu , atau pengurangan jumlah kategori dari c kategori menjadi r kategori, maka nilai- p yang digunakan dikalikan dengan pengganda Bonferroni (B). Pengganda Bonferroni ini tergantung pada tipe

kategoriknya yaitu nominal atau ordinal. Untuk :

I. Nominal

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

II. Ordinal

$$B = \binom{c-1}{r-1} \dots(2)$$

3. Dari semua variabel penjelas hasil proses 1 dan proses 2, cari variabel dengan nilai-p paling kecil sebagai variabel pertama yang masuk dalam model.
4. Ulangi proses 1, 2 dan 3 untuk variabel penjelas sisanya dengan menempatkan variabel penjelas itu sebagai sub populasi berdasarkan kategori yang terpilih.

b. Cara Membaca Pohon Berstruktur Hasil Metode CHAID

Hasil analisis metode CHAID berupa suatu pohon berstruktur, yaitu pohon yang memiliki “akar” (*roots*), “batang” (*branches*),

dan “daun” (*leaves*). Pohon ini dapat digambarkan mulai dari atas ke bawah, atau dari bawah ke atas, ataupun dari kiri ke kanan, atau dari kanan ke kiri (Huba, 2001). Setiap “daun” pada pohon berstruktur hasil metode CHAID diberi nomor simpul atau disebut dengan “node”.

Cara membaca pohon berstruktur, disesuaikan dengan penggambaran pohon berstruktur tersebut. Apabila struktur pohon digambarkan dari atas ke bawah, maka cara membacanya dimulai dari atas ke bawah. Demikian juga untuk pohon berstruktur yang digambarkan dengan cara yang lain, cara membacanya sama dengan cara penggambaran pohon berstruktur tersebut.

2. Penerapan Metode CHAID

a. Data Amatan

Variabel respons dan variabel penjelas kategorik yang diamati dalam data penelitian ini dapat dilihat pada Tabel 3.1. berikut :

Tabel 1. Variabel-variabel Amatan

Notasi	Nama Variabel	Skala Variabel
RESPONS		
OMSET	Peningkatan Omset: 1 = <25%, 2 = 25 - 49%, 3 = 50 - 74%, 4 = 75 - 100%, dan 5 = > 100%	Ordinal
PENJELAS		
JNSUSA	Jenis Usaha: 1= Lainnya , 2= Jasa, dan 3= Perdagangan	Nominal
LAMUSA	Lamanya Usaha: 1 = ≤ 5 tahun , 2 = 6-10 tahun, 3 = 11-15 tahun, 4 = 16-20 tahun dan 5 = >20 tahun	Ordinal
SIMPPW	Jumlah Simpanan Pokok & Wajib: 1 = ≤ 250.000, 2 = 250.001-500.000, 3 = 500.001-750.000, 4 = 750.001-1.000.000 dan 5 = >1.000.000	Ordinal
SKBSIM	Sukubunga Simpanan per bulan: 1 = < 0.3%, 2 = 0.3-0.4%, 3 = 0.41-0.5%, 4 = 0.51-0.6% dan 5 = > 0.6%	Ordinal
SMPSUK	Jumlah Simpanan Sukarela: 1 = ≤ 250.000, 2 = 250.001-500.000, 3 = 500.001-750.000, 4 = 750.001-1.000.000 dan 5 = >1.000.000	Ordinal
PINJ	Jumlah Pinjaman: 1 = ≤ 2.500.000, 2 = 2.500.001-5.000.000, 3 = 5.000.001-7.500.000, 4 = 7.500.001-10.000.000 dan 5 = >10.000.000	Ordinal
SKBPIN	Sukubunga Pinjaman per bulan: 1=1-1.5%, 2= 1.6-2%, 3= 2.1-2.5%, 4= 2.6%-3%, dan 5=>3%	Ordinal
BININS	Pembinaan Instansi : 1= belum pernah, 2 = 1 kali dalam beberapa tahun, 3 = 1 kali dalam 1 tahun, dan 4 = lebih dari 1 kali dalam 1 tahun	Ordinal
BINKEL	Pembinaan Kelompok: 1= lebih dari 2 tahun, 2 = 2 tahun sekali, 3 = 1.5 tahun sekali, 4 = 1 tahun sekali, dan 5 = kurang dari 1 tahun.	Ordinal
SIMPPW	Jumlah Simpanan Pokok & Wajib: 1 = ≤ 250.000, 2 = 250.001-500.000, 3 = 500.001-750.000, 4 = 750.001-1.000.000 dan 5 = >1.000.000	Ordinal
ANGGT	Lamanya Keanggotaan: 1= 1 tahun ,2 = 2 tahun, 3 = 3 tahun,4 = 4 tahun dan 5 = >4 tahun	Ordinal

b. Deskripsi Data

Deskripsi sederhana data sekunder yang diperoleh dari 120 responden anggota KSP secara singkat disajikan pada Tabel 3.2.

Nilai-nilai kategori awal variabel telah diformat sehingga siap untuk dianalisis dengan menggunakan metode CHAID.

Tabel 2. Deskripsi Sederhana Data Keanggotaan KSP

Variabel Penjelas	Kategori	Kategori Respons (Peningkatan Omset)					Jumlah
		< 25%	25 - 49%	50 - 74%	75 - 100%	> 100%	
JNSKSP	1	-	-	3	11	16	30
	2	-	3	22	5	-	30
	3	1	-	2	26	1	30
	4	1	1	6	15	7	30
JNSUSA	1	1	2	15	32	20	70
	2	-	-	4	8	3	15
	3	1	2	14	17	1	35
LAMUSA	1	1	-	2	13	3	19
	2	-	-	6	5	3	14
	3	-	1	4	5	4	14
	4	-	-	6	1	4	11
	5	1	3	15	33	10	62
SIMPPW	1	2	2	25	21	7	57
	2	-	1	2	3	2	8
	3	-	1	5	12	2	20
	4	-	-	-	14	5	19
	5	-	-	1	7	8	16
SKBSIM	1	2	4	28	44	12	90
	2	-	-	2	9	9	20
	3	-	-	-	1	1	2
	4	-	-	2	2	-	4
	5	-	-	1	1	2	4
SMPSUK	1	2	2	21	41	6	72
	2	-	1	6	3	-	10
	3	-	-	5	3	2	10
	4	-	1	-	1	1	3
	5	-	-	1	9	15	25
PINJ	1	2	-	12	14	1	29
	2	-	1	6	6	-	13
	3	-	-	3	1	2	6
	4	-	-	10	23	8	41
	5	-	3	2	13	13	31
SKBPIN	1	1	-	3	19	14	37
	2	-	-	1	18	2	21
	3	-	-	2	-	-	2
	4	1	4	27	20	7	59
	5	-	-	-	-	1	1
BININS	1	-	-	-	4	2	6
	2	1	2	4	14	2	23
	3	-	2	20	15	4	41
	4	1	-	9	24	16	50
BINKEL	1	1	-	4	4	1	10
	2	1	4	26	45	23	99
	3	-	-	2	6	-	8
	4	-	-	-	2	-	2
	5	-	-	1	-	-	1
ANGGT	1	1	1	15	10	6	33
	2	-	1	1	11	1	14
	3	-	-	2	11	13	26
	4	-	1	2	5	2	10
	5	1	1	13	20	2	37

3. Analisis Hasil Metode CHAID

a. Pengkategorian Ulang Hasil Metode CHAID

Seluruh variabel penjelas dengan kategori awal dapat dilihat pada Tabel 3.1., dari sebelas variabel penjelas yang dianalisis, seluruh variabel penjelas yang signifikan terhadap respons mengalami pengkategorian ulang dengan metode CHAID seperti yang ditampilkan pada Tabel 3.3. Variabel penjelas yang tidak signifikan, tidak ikut ditampilkan dalam tabel tersebut.

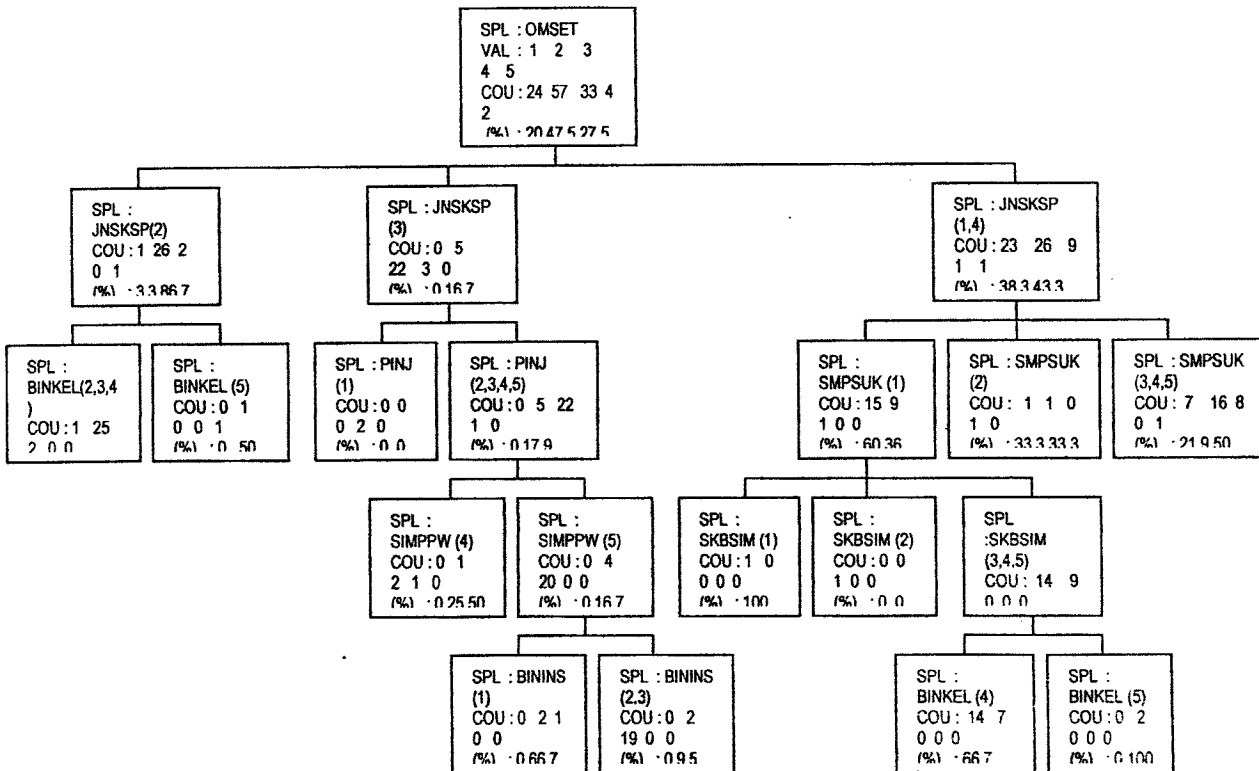
Tabel 3.3. Pengkategorian Ulang Hasil Metode CHAID

Variabel penjelas	Kategori Awal	Kategori Baru Hasil CHAID
JNSKSP	4 kategori	3 kategori, yaitu 1. KSP 013 2. KSP Kodanua 3. gabungan kategori 1,4 (Seroja & Husnayin)
		2 kategori, yaitu 1. ≤ 1 juta 2. > 1 juta
SKBSIM	5 kategori	3 kategori, yaitu 1. < 0,3% 2. 0,3 – 0,4% 3. gabungan 3,4,5 (> 0,4%)

SMPSUK	5 kategori	3 kategori, yaitu 1. < 250 ribu 2. 250.001 – 500 ribu 3. gabungan 4,5 (> 500 ribu)
		2 kategori, yaitu 1. ≤ 2,5 juta 2. gabungan 2,3,4,5 (> 2,5 juta)
PINJ	5 kategori	2 kategori, yaitu 1. belum pernah 2. gabungan 2,3 (≤ 1 tahun sekali)
		2 kategori, yaitu 1. gabungan 1,2,3,4 (≥ 1 tahun sekali) 2. < 1 tahun sekali
BININS	4 kategori	2 kategori, yaitu 1. < 250 ribu 2. 250.001 – 500 ribu 3. gabungan 4,5 (> 500 ribu)
		2 kategori, yaitu 1. ≤ 2,5 juta 2. gabungan 2,3,4,5 (> 2,5 juta)
BINKEL	5 kategori	2 kategori, yaitu 1. belum pernah 2. gabungan 2,3 (≤ 1 tahun sekali)
		2 kategori, yaitu 1. gabungan 1,2,3,4 (≥ 1 tahun sekali) 2. < 1 tahun sekali

b. Pohon Berstruktur Hasil Metode CHAID

Sebanyak sebelas variabel penjelas yang dianalisis dengan metode CHAID, dihasilkan hubungan berstruktur antar variabel-variabel penjelas yang signifikan terhadap respons. Hubungan ini digambarkan dalam suatu pohon berstruktur sebagaimana terlihat pada Gambar 1. berikut ini



Keterangan:

OMSET adalah Peningkatan Omset:

(1=<25%, 2= 25-49%, 3=50-74%, 4= 75-100%, 5=>100%)

JNSKSP adalah Jenis KSP:

(2=KSP 013, 3= Kodanua dan 1, 4 = Seroja / Husnayin)

BINKEL adalah Pembinaan Kelompok:

(1,2,3,4= ≥ 1 tahun sekali dan 5= < 1 tahun sekali)

PINJ adalah Jumlah Pinjaman:

(1= <2.500.000 dan 2,3,4,5=>2.500.000)

SIMPPW adalah Jumlah Simpanan Pokok & Wajib:

(4=≤1.000.000, dan 5=>1.000.000)

BININS adalah Pembinaan Instansi:

(1= belum pernah dan 2,3= ≤1 kali setahun)

SMPSUK adalah Jumlah Simpanan Sukarela:

(1=<250.000, 2=250.001-500.000, dan 3,4,5=>500.000)

SKBSIM adalah Sukubunga Simpanan per bulan:

(1=<0.3%, 2= 0.3-0.4% dan 3,4,5= >0.4%)

Gambar 1. Pohon Berstruktur Peningkatan Omset Usaha Anggota KSP

4. Interpretasi Hasil Metode CHAID

Ada tujuh variabel penjelas yang berpengaruh terhadap peningkatan omset usaha anggota KSP baik secara langsung maupun tidak langsung dapat dilihat secara berturut-turut pada pohon berstruktur Gambar 1, yaitu: *Jenis KSP, Pembinaan Kelompok, Jumlah Pinjaman, Jumlah Simpanan Pokok dan Wajib, Pembinaan Instansi, Jumlah Simpanan Sukarela, Sukubunga Simpanan*

Gambar 1. memperlihatkan bahwa faktor utama yang berpengaruh terhadap peningkatan omset usaha anggota adalah Jenis KSP. Dari empat jenis KSP yang ada, diperoleh tiga kategori baru yang saling berbeda secara statistik yaitu KSP 013, KSP Kodanua, dan gabungan KSP Seroja dan KSP Husnayin.

Kategori pertama, dari keanggotaan KSP 013. Pada kelompok ini, satu-satunya faktor yang berpengaruh terhadap peningkatan omset adalah Pembinaan Kelompok. Hal ini menunjukkan bahwa tidak ada variabel penjelas yang berinteraksi dengan Pembinaan Kelompok dalam hal pengaruhnya terhadap peningkatan omset anggota KSP 013. Hampir seluruh anggota KSP 013 (86,7%) termasuk ke dalam respons kategori pertama, yaitu memiliki kenaikan omset yang rendah (25-49%), mereka berasal

dari anggota yang baru mendapatkan pembinaan kelompok hanya satu kali dalam waktu yang lebih dari satu tahun. Hal ini memberikan informasi bagi koperasi 013 untuk mempertimbangkan peningkatan pembinaan kelompok anggotanya, yang diharapkan akan dapat membantu mereka untuk mendapatkan omset lebih tinggi.

Kategori kedua, yaitu keanggotaan KSP Kodanua, dan Kategori ketiga yaitu gabungan KSP Seroja dan Husnayin, dapat diinterpretasikan dengan cara yang sama seperti pada kategori pertama. Dibaca dan diinterpretasikan dari atas ke bawah karena penggambaran pohon berstrukturnya dibuat dari atas ke bawah.

Apabila dilihat secara keseluruhan, pohon berstruktur pada Gambar 1. memberikan karakteristik yang benar-benar berbeda antar kelompoknya, sedangkan dalam satu kelompok relatif homogen. Secara umum dapat disimpulkan bahwa risiko peningkatan omset usaha anggota KSP dipengaruhi oleh: *Jenis KSP*, besarnya *Jumlah Pinjaman*, besarnya *Jumlah Simpanan*, *Tingkat Sukubunga*, serta mendapatkan *Pembinaan Instansi* maupun *Pembinaan Kelompok*.

KESIMPULAN DAN SARAN

1. Kesimpulan

Berdasarkan hasil penelitian, dapat diambil beberapa kesimpulan sebagai berikut

1. Eksplorasi data dengan metode CHAID, tidak hanya dapat mendeskripsikan data awal, tetapi juga memberikan informasi adanya hubungan antara variabel penjelas dengan respons.
2. Pengkategorian ulang dengan metode CHAID diperlukan untuk menghasilkan kategori/kelompok baru yang lebih sesuai dengan data amatan.
3. Metode CHAID sesuai untuk diterapkan pada data amatan yang memiliki respons kategorik.
4. Jenis KSP merupakan variabel yang paling berpengaruh, dengan kata lain faktor utama yang berpengaruh terhadap tingkat peningkatan omset adalah Jenis KSP.
5. Hasil metode CHAID memperlihatkan bahwa anggota KSP Kodanua pada umumnya memiliki risiko tingkat keberhasilan usaha yang lebih tinggi daripada KSP 013 dan Seroja atau Husnayin.

2. Saran

Bagi yang berminat untuk melanjutkan penelitian ini, disarankan untuk mencoba membentuk model regresi bagi data respons ordinal, baik data yang melalui eksplorasi dengan metode CHAID maupun tidak.

DAFTAR PUSTAKA

- Agresti, A, 1990, "*Categorical Data Analysis*", John Wiley & Sons, New York.
- Huba, G.J, 2001, "CHAID", <http://www.themeasurement.com/definitions/CHAID.htm>
- Kass, G.V, 1982, "*Automatic Interaction Detection (AID) Techniques*", *Encyclopedia of Statistical Sciences* Vol 1. Ed. Kots, Z. & Johnson, N.L. John Wiley & Son, New York.
- Ryan, T.P, 1997, "*Modern Regression Models*", John Wiley & Sons, New York.
- SAS Institute, Inc, 1990, "*SAS/STAT User's Guide, Version 6, 4th edition, Volume 2*", Cary NC:SAS Institute, Inc.
- Snedker, K., Glynn, P., & Wang, C, 2002, "*Ordered / Ordinal Logistic egression with SAS and Stata*", <http://www.indiana.edu/~statmath/>