

ISBN : 978-979-17979-0-0

Prosiding

Seminar Nasional Mahasiswa S3 Matematika

REVITALISASI DAN SOSIALISASI DIRI UNTUK BERPERAN AKTIF DALAM PENINGKATAN KUALITAS PENELITIAN & PENDIDIKAN MATEMATIKA DI INDONESIA

Editor :

**Muslim Ansori
Ismail Djakaria
Dhoriva Urwatul Wutsqa
Agus Maman Abadi
M. Andy Rudhito
Umu Sa'adah
Karyati
Hasih Pratiwi**

Forum Komunikasi Mahasiswa S-3 Matematika se-Indonesia
Bekerja Sama Dengan Jurusan Matematika FMIPA UGM



Makalah Bidang Statistika

Estimating Conditional Intensity of Point Processes Models (<i>Nurtiti Sanusi, Sutawanir D., I Wayan M, Wahyu T</i>)	328
Generator Frame Wavelet Ketat dari Kelas Box Spline Multivariate (<i>Mahmud Yunus, Janny Lindiarni, Hendra Gunawan</i>)	334
Generalized Regression dalam Pendugaan Area Kecil (<i>Anang Kurnia, Khairil A. Notodiputro, Asep Saefuddin, dan I Wayan Mangku</i>)	346
Chain Ladder Method as a Gold Standard to Estimate Loss Reserves (<i>Aceng K. Mutaqin, Dumaria R. Tampubolon, Sutawanir Darwis</i>)	356
Optimalisasi Respon Ganda pada Metode Respon Permukaan (<i>Response Surface</i>) dengan Pendekatan Fungsi <i>Desirability</i> (<i>Hari Sakti Wibowo, I Made Sumertajaya, Hari Wijayanto</i>)	365
Sifat Penaksir Simulated MLE pada Respon Multinomial (<i>Jaka Nugraha</i>)	376
Estimasi Model Probit pada Respons Biner Multivariat Menggunakan MLE dan GEE (<i>Jaka Nugraha</i>)	386
Analisis Komponen Utama Temporar (<i>Ismail Djakaria</i>)	399
Pendekatan <i>Dynamic Linear Model</i> atau <i>State Space Model</i> pada Pendugaan Area Kecil (<i>Small Area Estimation</i>) (<i>Kusman Sadik dan Khairil Anwar Notodiputro</i>)	411
Pengembangan Algoritma EM untuk Data Tidak Lengkap (<i>Incomplete Data</i>) pada Model Log-Linear (<i>Kusman Sadik</i>)	422
CRUISE sebagai Metode Berstruktur Pohon (<i>Tree-Structured</i>) pada Data Non-Biner (<i>Kusman Sadik</i>)	433
Model <i>Proportional Hazards Cox</i> dengan <i>Missing Covariates</i> (<i>Nurkaromah Dwidayati, Sri Haryatmi, Subanar</i>)	447
Aplikasi Mixture Distribution dalam Pemodelan Resiko Aktuaria (<i>Adhitya Ronnie Effendie</i>)	459
Analisis Komponen Utama Probabilistik pada Data <i>Missing</i> (<i>Ismail Djakaria</i>)	464
Makalah Bidang Terapan	
Bilangan Ramsey untuk Graf Bintang S_6 dan Graf Bipartit Lengkap $K_{2,N}$, $N=2,3,4$ (<i>Isnaini Rosyida</i>)	476
H -supermagic labelings of c copies of some graphs (<i>Tita Khalis Maryati, A. N. M. Salman, Edy Tri Baskoro, Irawati</i>)	485

CRUISE sebagai Metode Berstruktur Pohon (*Tree-Structured*) pada Data Non-Biner

Kusman Sadik

Departemen Statistika, FMIPA IPB
Jl. Raya Dramaga, Kampus IPB Dramaga, Bogor
e-mail : kusmansadik@yahoo.com

Abstrak

Metode berstruktur pohon (*tree-structured methods*) pada tahun-tahun terakhir ini banyak digunakan sebagai metode analisis yang banyak digunakan dalam riset pemasaran (segmentasi pasar), kedokteran (diagnosis), ilmu komputer (menyelidiki struktur data), botani (klasifikasi), psikologi (pengambilan keputusan), dan linguistik. Pada dasarnya pohon klasifikasi (*classification tree*) merupakan metode statistika yang digunakan untuk memperkirakan keanggotaan amatan atau objek dalam kelas-kelas peubah tak bebas kategorik. Keanggotaan ini diduga dari pengukuran amatan pada satu peubah bebas atau lebih. Pohon dibentuk melalui penyekatan data sampel secara berulang atau rekursif, dimana kelas dan nilai-nilai peubah penjelas setiap amatan pada data sampel tersebut sudah diketahui. Setiap sekatan data dinyatakan sebagai node (simpul) dalam pohon yang terbentuk. CRUISE (*classification rule with unbiased interaction selection and estimation*) merupakan salah satu metode berstruktur pohon non-biner yang tidak berbias dalam pemilihan peubah. Metode ini merupakan pengembangan dari berbagai metode berstruktur pohon terutama dari FACT (*factor analysis classification trees*), QUEST (*quick, unbiased, efficient statistical tree*), dan CART (*classification and regression trees*).

Kata kunci : *cruise methods, classification rule, unbiased interaction selection, FACT, QUEST, CART, multiway tree-structured, misclassification.*

1. Pendahuluan

1.1. Latar Belakang

Penggunaan metode berstruktur pohon (*tree-structured methods*) sebagai alternatif terhadap metode-metode konvensional telah menyebar luas di berbagai bidang dalam beberapa tahun terakhir. Bidang terapan yang menggunakan metode tersebut antara lain riset pemasaran (dalam segmentasi pasar), kedokteran (untuk diagnosis), ilmu komputer (untuk menyelidiki struktur data), botani (dalam hal klasifikasi), psikologi (teori pengambilan keputusan), dan linguistik. Hasil analisis utama metode ini adalah grafik pohon yang

memudahkan pengguna – terutama yang bukan statistisi – dalam interpretasi (Statsoft, Inc.,2002). Berdasarkan tipe peubah responnya, ada dua tipe metode berstruktur pohon, yaitu tipe klasifikasi (peubah respon kategorik) dan tipe regresi (peubah respon numerik) (Statsoft, Inc.,2002).

Struktur suatu pohon haruslah mudah dipahami agar dapat memberikan informasi yang bermanfaat. Namun hal lain yang tak kalah pentingnya adalah ada tidaknya bias dalam pemilihan peubah. Jika pemilihan peubah berbias, peubah-peubah bebas akan memiliki peluang berbeda untuk dipilih sebagai peubah pemilah (*split*). Akibatnya sulit diketahui apakah suatu peubah terpilih karena memang merupakan peubah penting (yaitu dapat menjelaskan peubah respon), atautkah hanya karena bias (Kim & Loh, 2001).

CRUISE adalah salah satu metode berstruktur pohon non-biner yang tidak berbias dalam pemilihan peubah. Selain itu, CRUISE juga menghasilkan salah klasifikasi terkecil serta pohon terkecil bila dibandingkan dengan QUEST dan CHAID oleh Faridhan (2003) yang diterapkannya pada bidang mikologi, yaitu mengidentifikasi jamur beracun. Penulis mencoba mengaplikasikan CRUISE dibidang lain, yaitu bidang gizi masyarakat.

Mahasiswa IPB yang masuk melalui jalur USMI (Ujian Seleksi Masuk IPB) adalah mahasiswa yang memiliki prestasi yang konsisten selama 3 tahun di SMU. Selain itu secara psikologis, mahasiswa tersebut berbeda dengan mahasiswa lain karena lebih dahulu mendapat informasi mengenai diterimanya masuk IPB sebelum mengetahui status kelulusannya di SMU, serta dipastikan tercatat sebagai mahasiswa IPB ketika yang lain tengah mempersiapkan diri dalam ujian SPMB (Seleksi Penerimaan Mahasiswa Baru). Oleh karena itu banyak penelitian tentang karakteristik dari mahasiswa tersebut. Dalam penelitian ini metode CRUISE digunakan untuk mengetahui status gizi dari mahasiswa USMI IPB serta hubungannya dengan faktor sosio ekonomi, wilayah dan prestasi di SMU.

1.2. Tujuan

Penelitian ini bertujuan untuk menerapkan CRUISE sebagai metode pohon klasifikasi non-biner yang tidak berbias pada analisis hubungan antara status gizi mahasiswa USMI IPB dengan faktor sosioekonomi, wilayah, dan prestasi SMU.

2. Tinjauan Pustaka

2.1. Metode Berstruktur Pohon

Metode berstruktur pohon, yaitu pohon klasifikasi dan pohon regresi, merupakan salah satu teknik utama dalam penambangan data (*data mining*). Teknik ini memiliki banyak kesamaan dengan metode-metode yang lebih konvensional, seperti analisis regresi, analisis logit/probit, analisis diskriminan, analisis gerombol, statistika non-parametrik, dan pendugaan non-linier.

Pohon klasifikasi (*classification tree*) merupakan metode statistika yang digunakan untuk memperkirakan keanggotaan amatan atau objek dalam kelas-kelas peubah tak bebas kategorik. Keanggotaan ini diduga dari pengukuran amatan pada satu peubah bebas atau lebih. Pohon dibentuk melalui penyekatan data contoh secara berulang (rekursif), dimana kelas dan nilai-nilai peubah penjelas setiap amatan pada data contoh tersebut sudah diketahui. Setiap sekatan data dinyatakan sebagai node (simpul) dalam pohon yang terbentuk.

Salah satu keuntungan penggunaan metode berstruktur pohon adalah tampilan grafisnya sehingga lebih mudah untuk diinterpretasikan. Selain itu metode tersebut juga lebih fleksibel serta mampu memeriksa pengaruh peubah bebas satu per satu (tidak memeriksa langsung seluruh peubah, seperti umumnya dilakukan metode lain) (Statsoft, Inc., 2002).

Kelebihan lain dari metode tersebut adalah sifatnya yang non-parametrik dan non-linier. Metode ini dapat mengungkapkan hubungan sederhana antar beberapa peubah pada data yang tidak diketahui bagaimana hubungan antar peubahnya, yang mungkin tak terdeteksi oleh teknik analisis lainnya.

2.2. Metode CRUISE

CRUISE (*Classification Rule with Unbiased Interaction Selection and Estimation*) diperkenalkan oleh Kim dan Loh (2001) sebagai jawaban dari metode-metode berstruktur pohon lain yang berbias dalam pemilihan peubah. CRUISE menghasilkan dua sampai J *split*, dimana J adalah banyaknya kategori peubah respon. Metode ini merupakan pengembangan dari gabungan berbagai metode berstruktur pohon, terutama dari FACT (*Factor Analysis Classification Trees*) (Loh & Vanichsetakul dalam Kim & Loh, 2001) dan QUEST (*Quick, Unbiased, Efficient Statistical Tree*) (Loh & Shih, 1997) untuk pemilihan *split*, dan CART

(*Classification and Regression Trees*) (Breiman et al, 1993) untuk pemangkasan (*pruning*). Tujuan umum metode ini serupa dengan tujuan CART dan QUEST, yaitu memperoleh pohon yang sedemikian hingga data contoh pada setiap node akhir menjadi ‘murni’ (Kim & Loh, 2002), yaitu menunjukkan salah satu kelas peubah respon.

Metode ini memiliki sifat-sifat berikut :

1. Pohon yang dihasilkan seringkali memiliki keakuratan penduga yang tinggi.
2. Kecepatan komputasi tinggi.
3. Bebas dari bias dalam pemilihan peubah.
4. Sensitif terhadap interaksi lokal antar peubah. Ini menghasilkan pohon yang lebih pendek serta pemilihannya lebih ‘pintar’.
5. Keempat sifat di atas juga berlaku untuk data yang memiliki amatan hilang.

CRUISE dapat melakukan pemilahan tunggal (*univariate split*) maupun pemilahan kombinasi linier (*linier combination split*). Pemilahan kombinasi linier lebih fleksibel dan menghasilkan keakuratan penduga yang lebih baik, sehingga node akhir yang diperoleh juga lebih sedikit. Namun interpretasi pohon tidak mudah karena pemilahan kombinasi linier lebih sulit untuk dipahami.

2.3. Pemilihan Peubah *Split*

Terdapat dua metode pemilihan peubah dalam pemilahan tunggal yang disebut metode 1D dan 2D. Ide pengembangan metode 1D diperoleh dari QUEST. Prosedur ini tak bias dalam pengertian bahwa jika penduga dan peubah respon saling bebas, setiap peubah memiliki peluang yang sama untuk terpilih. Kelemahan metode 1D terletak pada rancangannya yang ditujukan untuk mendeteksi ketakhomogenan nilai tengah dan ragam kelas pada peubah numerik. Jika sebaran kelas tersebut berbeda, metode ini menjadi tidak efektif.

Metode 2D dapat mengatasi kelemahan metode 1D dengan mendeteksi interaksi berpasangan antar peubah pada pohon klasifikasi. Hal ini menyebabkan algoritma metode 2D lebih panjang, rumit dan juga menyebabkan waktu pemrosesannya (*CPU time*) menjadi lebih lama (Faridhan, 2003).

CRUISE menangani pemilihan peubah dan pemilihan titik *split* secara terpisah. Berikut algoritma Metode 1D.

Algoritma 1 : CRUISE – Metode 1D

Misalkan α adalah taraf kepercayaan yang dipilih (harga *default* adalah 0.05).

1. Untuk penduga numerik (interval, rasio), lakukan sidik ragam satu-arah dengan kategori Y sebagai perlakuan; kemudian hitung nilai-p dari uji-F. Misalkan X_{k1} memiliki nilai-p terkecil α_1 .
2. Untuk setiap penduga kategorik (nominal, ordinal), buat tabel kontingensi dengan nilai-nilai kategori sebagai baris dan nilai-nilai kelas (kategori peubah respon) sebagai kolom, dan hitung nilai-p dari uji χ^2 . Misalkan nilai-p adalah α_2 dan peubah yang bersesuaian adalah X_{k2} .
3. Jika $\alpha_1 \leq \alpha_2$, pilih peubah numerik X_{k1} ; jika tidak, pilih peubah kategorik X_{k2} . Anggap peubah terpilih dinamakan X_{k^*} .
4. Jika $\min(\alpha_1, \alpha_2) < \alpha/K$ (koreksi Bonferroni pertama), maka pilih X_{k^*} sebagai peubah *split*.
5. Jika (4) tidak dipenuhi, hitung nilai-p untuk uji-F Levene untuk setiap peubah numerik. Misalkan $X_{k^{**}}$ memiliki nilai-p terkecil $\bar{\alpha}$.
 - (a) Jika $\bar{\alpha} < \alpha/(K+K_1)$, pilih $X_{k^{**}}$ (koreksi Bonferroni kedua), dimana K = banyaknya penduga numerik dan K_1 = banyaknya penduga kategorik.
 - (b) Jika tidak, pilih X_{k^*} .

Metode 2D merupakan pengembangan masalah klasifikasi dari pendekatan Loh (2001) dalam mendeteksi interaksi berpasangan antar peubah pada pohon klasifikasi. Algoritmanya adalah sebagai berikut:

Algoritma 2 : CRUISE-Metode 2D

Misalkan J_t = Banyaknya kelas (kategori peubah respon) pada node t ; K = banyaknya penduga numerik; dan K_1 = banyaknya peubah kategorik.

1. Uji marjinal untuk setiap peubah numerik X (interval, rasio). Misalkan X_{k1} memiliki nilai-p terkecil α_1 .
 - (a) Bagi data menjadi empat kelompok menurut kuartil contoh dari X.
 - (b) Buat tabel kontingensi $J_t \times 4$ dengan kelas sebagai baris dan kelompok sebagai kolom.

(c) Hitung statistik χ^2 Pearson dengan derajat bebas $v = 3(J_t - 1)$.

(d) Konversikan χ^2 menjadi nilai normal baku dengan transformasi Peizer-Pratt

$$z = \begin{cases} |W|^{-1} (W - 1/3) \sqrt{(v-1) \log \left(\frac{v-1}{\chi^2} \right)} + W, & v > 1 \\ \sqrt{\frac{2}{\chi^2}}, & v = 1 \end{cases}$$

dimana $W = \chi^2 - v + 1$

Misalkan $z_n = \max \{z_1, \dots, z_{k1}\}$

2. Uji marjinal untuk setiap peubah kategorik X (nominal, ordinal). Misalkan C adalah banyaknya kategori pada X.

(a) Buat tabel kontengensi $J_t \times C$ dengan kelas sebagai baris dan kategori C sebagai kolom.

(b) Hitung statistik χ^2 Pearson dengan derajat bebas $v = (J_t - 1)(C - 1)$

(c) Gunakan transformasi Peizer-Oratt pada 1(d).

Misalkan $z_c = \max \{z_{k+1}, \dots, z_k\}$.

3. Uji interaksi untuk setiap pasang peubah numerik ($X_k, X_{k'}$):

(a) Bagi ruang ($X_k, X_{k'}$) ke dalam empat kuadran menurut median contoh.

(b) Buat tabel kontingensi $J_t \times 4$ dengan kelas sebagai baris dan kuadran sebagai kolom.

(c) Hitung statistik χ^2 pearson dengan derajat bebas $v = 3(J_t - 1)$.

(d) Gunakan transformasi Peizer-Pratt.

Misalkan z_{nn} adalah nilai-z terbesar diantara $K_1(K_1 - 1)/2$ nilai-z yang ada.

4. Uji interaksi untuk setiap pasang peubah kategorik: Jika pasangan peubah memiliki kategori sebanyak C_1 dan C_2 , akan diperoleh tabel kontingensi $J_t \times C_1 \times C_2$. Misalkan z_{cc} adalah nilai-z terbesar diantara $(K - K_1)(K - K_1 - 1)/2$ nilai-z yang ada.

5. Uji interaksi untuk pasangan ($X_k, X_{k'}$) dimana X_k adalah peubah numerik dan $X_{k'}$ adalah kategorik: Jika $X_{k'}$ memiliki C kategori, bentuk tabel kontingensi $J_t \times 2C$. Misalkan z_{nc} adalah nilai-z terbesar di antara $K_1(K - K_1)/2$ nilai-z yang ada.

Misalkan f^* adalah nilai bootstrap (Lampiran 8) dan definisikan

$Z^* = \max \{f^*z_n, z_c, f^*z_{nn}, z_{cc}, z_{nc}\}$.

(a) Jika $f^*z_n = z^*$, pilih peubah numerik dengan nilai-z terbesar.

(b) Jika $z_c = z^*$, pilih peubah kategorik dengan nilai-z terbesar.

- (c) Jika $f^*z_{nn} = z^*$, pilih peubah numerik dalam pasangan yang nilai-z-nya lebih besar.
- (d) Jika $z_{cc} = z^*$, pilih peubah kategorik dalam pasangan yang nilai-z-nya lebih besar.
- (e) Jika $z_{nc} = z^*$, pilih peubah kategorik pada pasangan yang berinteraksi.

Tingkat salah klasifikasi metode 2D umumnya sama atau sedikit lebih baik dari metode 1D. Untuk data berukuran besar atau data dengan banyak peubah bebas, penerapan metode 2D memakan waktu CPU lebih lama daripada metode 1D karena algoritmanya yang lebih kompleks.

Pemilihan Titik Split

Analisis diskriminan linier dilakukan jika X terpilih merupakan peubah numerik pada algoritma 1D dan peubah tersebut tidak terpilih melalui uji Levene. Transformasi Box-Cox dapat diterapkan apabila peubah *split* terpilih melalui uji Levene (berarti ragam tak homogen) (Lampiran 7). Selain itu, transformasi Box-Cox dilakukan jika X terpilih merupakan peubah numerik pada algoritma 2D.

Kategori peubah terpilih X (merupakan peubah kategorik) diubah menjadi nilai *crimcoord* terlebih dahulu (Lampiran 6) (algoritma 1D dan 2D) sebelum dilakukan analisis diskriminan linier. Jika ada nilai X yang negatif, $2x_{(i+1)} - x_{(i)}$ ditambahkan pada nilai-nilai X, dimana $x_{(i)}$ adalah statistik tataan ke-i pada X. Setelah itu transformasi Box-Cox dapat dilakukan.

Analisis diskriminan linier dilakukan setelah X ditransformasi untuk menentukan titik *split*. Jika X kategorik, setelah titik *split* diperoleh, nilai *crimcoord* diubah kembali menjadi kategori asal.

Aturan Penghentian

Pada setiap node dilakukan pen-*split*-an berdasarkan aturan pembentukan pohon. Pen-*split*-an itu dilakukan secara berulang, yaitu pada tiap node anak hasil pen-*split*-an dilakukan pen-*split*-an ulang.

Pen-*split*-an dihentikan jika banyaknya amatan dari suatu node sudah lebih kecil dari ambang tertentu (*minimum size of each node/mindat*) dan node tersebut menjadi node akhir (terminal).

Pemangkasan Pohon (*Pruning*)

Pohon yang dibentuk dengan aturan penghentian berdasarkan ukuran minimum dari suatu simpul akhir akan berukuran sangat besar dan kompleks, sedangkan model yang diperlukan bukanlah pohon yang besar. Pohon tersebut analog dengan model regresi yang dihasilkan

melalui prosedur regresi bertatar langkah maju (*forward stepwise regression*) dimana tidak ada lagi peubah bebas yang masuk ke dalam model oleh karena tidak ada lagi F-hitung dari peubah bebas yang bisa mencapai suatu ambang yang ditetapkan.

Pemangkasan dengan menggunakan aturan *cost-complexity minimum* merupakan salah satu cara untuk menghindari terbentuknya pohon yang berukuran besar (Breiman et al, 1993). Strategi validasi silang V-lipatan (*V-fold cross-validation*) dapat menjadi jalan alternatif lain agar pohon yang terbentuk mempunyai kinerja prediksi yang baik .

Status Gizi dalam Antropometrik

Status gizi adalah keadaan seseorang yang diakibatkan oleh konsumsi, penyerapan, dan penggunaan makanan dan zat gizi dalam jangka waktu yang lama. Status gizi merupakan suatu bagian penting dari status kesehatan seseorang (Harper, Deaton, & Driskel, 1986 dalam Novikasari, 2003).

Menurut Gibson (1990), ada berbagai variasi penilaian status gizi untuk menggambarkan masing-masing tahap dalam perkembangan kekurangan gizi. Metode penilaian didasarkan pada pengukuran konsumsi makanan, laboratorium, klinis, dan antropometri. Pengukuran dimensi fisik dan komposisi tubuh digunakan dalam metode antropometri (Jelliffe et al., 1966 dalam Gibson, 1990).

Beberapa macam antropometri atau ukuran-ukuran tubuh manusia yang telah digunakan antara lain : berat badan (BB), tinggi badan (TB), lingkaran lengan atas (LLA), lingkaran kepala (LK), lingkaran dada (LD), lingkaran pinggang dan pinggul (LPP), dan lapisan lemak bawah kulit (LLBK). Di Indonesia, jenis antropometri yang banyak digunakan adalah BB dan TB. Dalam pemakaian untuk penilaian status gizi, antropometri disajikan dalam bentuk indeks, seperti : Indeks berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U), berat badan menurut tinggi badan (BB/TB), Indeks Massa Tubuh (IMT), dan sebagainya.

Gibson (1990) menjelaskan bahwa IMT dapat dihitung dengan rumus sebagai berikut :

$$IMT = \frac{\text{Berat Badan (kg)}}{[\text{Tinggi Badan (m)}]^2}$$

Bagi Remaja, indikator terbaik menurut WHO (1995) yang digunakan adalah IMT menurut umur (IMT/U). Indikator ini sudah divalidasi sebagai indikator lemak tubuh total pada persentil atas, dan juga sejalan dengan indikator yang sudah direkomendasikan untuk orang dewasa. Nilai IMT dikategorikan sesuai nilai ambang IMT/U yang disajikan pada

Tabel 1. Persentil IMT/U dari berbagai umur remaja diambil dari WHO, 1995 berdasarkan survey yang dilakukan NHANES I *United States of America* (Lampiran 1 dan 2).

Tabel 1. Klasifikasi Status Gizi berdasarkan IMT/U (Sumber WHO, 1995)

Kategori	Nilai Titik Batas
Kurus	< persentil ke-5
Normal	Persentil ke-5 sampai persentil ke-85
Gemuk	≥ persentil ke-85

3. Bahan dan Metode

3.1. Bahan

Penelitian ini menggunakan data mahasiswa baru TPB IPB melalui jalur USMI, yang diambil dari Panitia Penerima Mahasiswa Baru (PPMB) IPB, serta Klinik Konsultasi Gizi IPB.

Peubah status gizi sebagai peubah respon diukur berdasarkan IMT/U sehingga terbagi ke dalam tiga kategori, yaitu: 1 = kurus, 2 = normal, 3 = gemuk. Peubah-peubah penjelas yang dimasukkan adalah sebagai berikut :

Z1	Jenis Kelamin
Z2	Asal SMU
Z3	Pendidikan Ayah
Z4	Pendidikan Ibu
Z5	Pekerjaan Ayah
Z6	Pekerjaan Ibu
Z7	Penghasilan Orang Tua
Z8	Biaya Hidup
Z9	Daya Listrik
Z10	Nilai Ujian Nasional
Z11	Nilai Rapor MIPA

3.2. Metode

Data dianalisis menggunakan algoritma CRUISE dengan pemilahan tunggal (metode 1D dan 2D). Pen-*split*-an dihentikan jika banyaknya amatan dari suatu node sudah lebih kecil

dari 1% dari total data. Apabila pohon yang dihasilkan besar dan kompleks, maka dilakukan pemangkasan (*pruning*). *Software* yang digunakan adalah program komputer CRUISE (Kim & Loh, 2000) ver. 1.09 dan allCLEAR ver. 4.5.

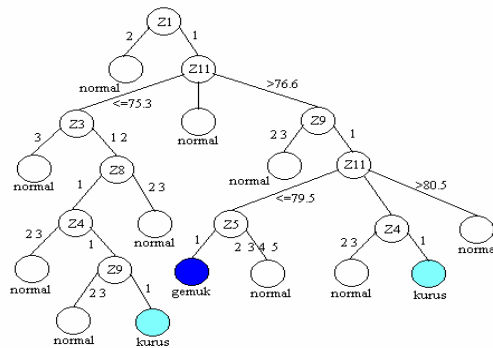
4. Hasil dan Pembahasan

Dalam tulisan ini amatan hilang (*missing value*) disisihkan, sehingga data yang dianalisis sebanyak 1811 (94.6%) dari total 1914 orang, yang terdiri dari 745 laki-laki, 1066 perempuan, dengan umur berkisar antara 15-20 tahun. Pengkategorian peubah respon berdasarkan IMT/U menghasilkan 175 (7.9%) kurus, 1523 (84.1%) normal dan 113 (6.2%) gemuk.

Agar mendapatkan pohon yang relatif sederhana dan logis maka dibutuhkan pengurangan peubah serta penyederhanaan kategori dalam analisis CRUISE.

Setelah dilakukan penyederhanaan peubah, terdapat 11 peubah, 9 peubah kategorik dan 2 peubah numerik. Pada peubah kategorik, peubah fakultas dikeluarkan dalam analisis karena dianggap tidak berhubungan dengan status gizi. Lalu peubah-peubah kategorik tersebut dilakukan penyederhanaan kategori. Untuk peubah numerik dapat diringkas menjadi dua peubah, peubah UAN (Z10) dan MIPA(Z11). Peubah MIPA didapat dari rata-rata nilai rapor Biologi, Fisika, Kimia dan Matematika.

Seperti sebelumnya, CRUISE 1D dan 2D tanpa pemangkasan menghasilkan pohon yang sangat kompleks. CRUISE 1D menghasilkan pohon 308 node dengan 170 terminal, dan salah klasifikasi 0.1485. CRUISE 2D menghasilkan pohon 253 node dengan 142 terminal dan salah klasifikasi 0.1541. Setelah dilakukan pemangkasan, CRUISE 1D dan 2D menghasilkan pohon yang lebih sederhana.



Gambar 1. Dendrogram CRUISE 1D peubah terpilih

CRUISE 1D menghasilkan pohon (Gambar 1) 23 node dengan 13 terminal dan salah satu klasifikasi 0.1552. Pohon tersebut menghasilkan peubah jenis kelamin (Z1), nilai rapor MIPA (Z11), pendidikan ayah (Z3), daya listrik (Z9), biaya hidup (Z8), pendidikan ibu (Z4), dan pekerjaan ayah (Z5) sebagai peubah split. Sedangkan peubah asal SMU (Z2), pekerjaan ibu (Z6), penghasilan orang tua (Z7), dan Nilai Ujian Nasional (Z10) tidak dapat membedakan status gizi.

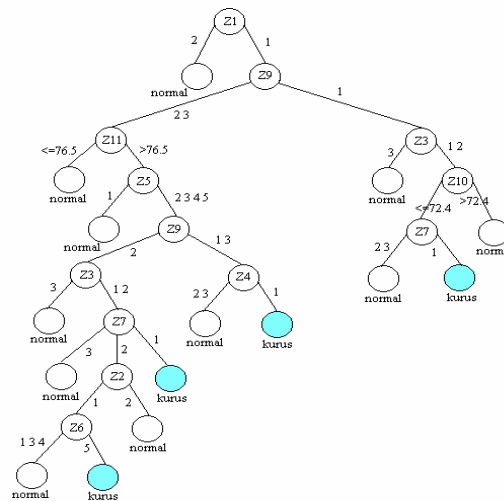
Berdasarkan Gambar 1, untuk membedakan status gizi, pertama-tama dilihat jenis kelamin. Jenis kelamin perempuan diklasifikasikan ke dalam status gizi normal. Sedangkan laki-laki harus melihat peubah nilai MIPA. Apabila nilai MIPA berkisar 75.3 – 76.6 maka diklasifikasikan ke dalam status gizi normal. Jika nilai MIPA ≤ 75.3 maka harus melihat peubah pendidikan ayah. Apabila pendidikan ayah di atas SLTA maka diklasifikasikan ke dalam status gizi normal, selainnya melihat peubah biaya hidup. Apabila biaya hidup tinggi maka diklasifikasikan ke dalam status gizi normal, selainnya melihat peubah pendidikan ibu. Apabila pendidikan ibu $\leq SD$ harus melihat peubah daya listrik, selainnya normal. Apabila daya listrik < 900 watt maka diklasifikasikan ke dalam status kurus selainnya normal.

Mahasiswa berjenis kelamin laki-laki, memiliki nilai MIPA > 76.6 , dan memiliki daya listrik ≥ 900 watt diklasifikasikan ke dalam status gizi normal, selainnya harus melihat peubah nilai rapor MIPA. Apabila nilai rapor MIPA ≤ 79.5 dan ayah tidak bekerja diklasifikasikan ke dalam status gemuk. Apabila nilai rapor MIPA berkisar 79.5 – 80.5 dan pendidikan ibu $\leq SD$ maka diklasifikasikan ke dalam status kurus, selainnya normal.

Mahasiswa berjenis kelamin perempuan cenderung normal karena lebih mementingkan citra tubuhnya dibanding dengan laki-laki. Terdapat sedikit keanehan pada satu terminal yang mengidentifikasi mahasiswa gemuk pada ayah yang tidak bekerja. Dengan melihat peubah penghasilan orang tua yang tinggi (\geq Rp. 2.5 juta), terlihat indikasi adanya salah pengkodean data dimana mahasiswa yang tidak mengisi pekerjaan ayah dikategorikan ke dalam ayah yang tidak bekerja. Selain itu, mahasiswa yang memiliki ibu berpendidikan rendah, biaya hidup rendah, dan dengan rumah berdaya listrik rendah cenderung kurus karena secara ekonomi mereka kurang mampu dalam memenuhi gizi keluarganya serta minimnya pengetahuan sang ibu terhadap gizi.

CRUISE 2D menghasilkan pohon (Gambar 6) 28 node dengan 15 terminal dan salah satu klasifikasi 0.1563. Peubah splitnya adalah jenis kelamin (Z1), daya listrik (Z9), nilai rapor

MIPA (Z11), pendidikan ayah (Z3), pekerjaan ayah (Z5), nilai ujian nasional (Z10), pendidikan ibu (Z4), penghasilan orang tua (Z7), asal SMU (Z2), dan pekerjaan ibu (Z6). Peubah biaya hidup (Z8) tidak terpilih sebagai peubah split. Dari 16 terminal yang terbentuk, 4 diantaranya mengklasifikasikan kurus dan sisanya mengklasifikasikan status gizi normal. Pohon tersebut tidak berhasil mengklasifikasikan status gizi gemuk.



Gambar 2. Dendrogram CRUISE 2D peubah terpilih

Berdasarkan Gambar 2, mahasiswa TPB IPB berstatus gizi kurus mempunyai profil, yaitu:

1. Jenis kelamin laki-laki, daya listrik antara 900 dan 1300 watt, nilai rapor MIPA > 76.5 , Ayah bekerja, pendidikan ayah \leq SLTA, penghasilan orang tua berkisar 2.5 juta sampai 5 juta, asal SMU dari Jawa, dan pekerjaan ibu pegawai swasta.
2. Jenis kelamin laki-laki, daya listrik antara 900 dan 1300 watt, nilai rapor MIPA > 76.5 , Ayah bekerja, pendidikan ayah \leq SLTA, dan penghasilan orang tua lebih kecil 2.5 juta.
3. Jenis kelamin laki-laki, daya listrik ≥ 1300 watt, nilai rapor MIPA > 76.5 , Ayah bekerja, dan pendidikan ibu \leq SD.
4. Jenis kelamin laki-laki, daya listrik < 900 watt, pendidikan ayah \leq SLTA, Nilai Ujian Nasional ≤ 72.4 , dan penghasilan orang tua < 2.5 juta.

Mahasiswa TPB IPB juga dapat diklasifikasikan sebagai status gizi normal ke dalam 11 cara.

Terdapat sedikit keanehan dalam hasil pohon CRUISE 2D, yaitu mahasiswa dengan ayah tidak bekerja cenderung memiliki status gizi normal. Apabila dilakukan pengecekan

data, ternyata mahasiswa dalam terminal tersebut memiliki orang tua yang berpenghasilan antara Rp 2.5 juta sampai Rp 5 juta. Hal ini menunjukkan adanya indikasi kesalahan pengkodean data, dimana mahasiswa yang tidak mengisi pekerjaan ayah dikategorikan ke dalam ayah yang tidak bekerja. Selain itu profil mahasiswa kurus pertama yang mengklasifikasikan pada mahasiswa yang memiliki ayah berpenghasilan rendah karena merupakan pegawai negeri dan ibu bekerja sebagai pegawai swasta untuk menopang ekonomi keluarga (penghasilan orang tua, biaya hidup, dan daya listrik kategori menengah/sedang) diperkirakan merupakan mahasiswa yang kurang mendapat perhatian ibu karena ibu terlalu sibuk bekerja. Selain itu pula penghasilan orang tua sedang dan pendidikan ayah rendah mengidentifikasi sebuah keluarga dimana kedua orang tua sibuk bekerja dan tidak begitu memperhatikan gizi anak-anaknya. Selain itu pula mahasiswa yang berasal dari keluarga berstatus ekonomi menengah kebawah cenderung kurus. Hal ini ditandai dengan profil mahasiswa kurus, yaitu penghasilan orang tua yang rendah serta pendidikan ayah yang rendah.

CRUISE 2D tidak berhasil mengklasifikasikan mahasiswa gemuk. Hal ini bisa disebabkan oleh beberapa faktor, yaitu: peubah-peubah bebas yang memang tidak bisa membedakan mahasiswa gemuk dengan status normal, data mahasiswa gemuk terlalu sedikit, dan kesenjangan proporsi yang terlalu mencolok antara status gizi gemuk dengan status gizi yang lain. Seperti sebelumnya, strategi validasi silang 10-lipatan menghasilkan pohon satu terminal dengan gizi normal sebagai dugaan peubah respon.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Dalam menganalisis hubungan status gizi mahasiswa USMI IPB dengan faktor sosioekonomi, wilayah dan prestasi di SMU, CRUISE 1D dan 2D tanpa pemangkasan menghasilkan pohon yang sangat rumit dan tidak efektif. Pemangkasan dengan menggunakan aturan *cost-complexity minimum* menghasilkan pohon yang lebih sederhana. Namun seringkali tidak logis dalam interpretasi. Penyeleksian dan penyederhanaan kategori dapat menghasilkan pohon yang lebih sederhana dan logis dalam interpretasi.

CRUISE 1D dan 2D menghasilkan pohon yang tidak berbeda, baik sebelum maupun sesudah penyederhanaan peubah pada pemilihan peubah split pertama sampai ke tiga atau ke

empat. CRUISE 1D dan 2D berbeda pada pemilihan peubah split ke lima dan seterusnya karena CRUISE 2D lebih menekankan pada deteksi interaksi antar peubah bebas.

Secara umum berdasarkan analisis CRUISE, peubah yang paling bisa membedakan status gizi adalah jenis kelamin. Peubah nilai rapor MIPA merupakan faktor prestasi di SMU yang paling dapat membedakan status gizi. Pada faktor sosioekonomi, peubah pendidikan ibu, pendidikan ayah dan daya listrik lebih dapat membedakan status gizi dibandingkan dengan peubah lainnya. Ternyata faktor wilayah tidak dapat membedakan status gizi mahasiswa USMI IPB. Pohon yang paling optimal dari segi kesederhanaan struktur pohon dan kelogisan adalah CRUISE 1D dengan peubah yang sudah disederhanakan.

4.2. Saran

Perlu pula dilakukan penelitian yang lebih dalam untuk melihat model hubungan yang terjadi pada faktor sosioekonomi dan prestasi terhadap status gizi. Dengan banyaknya kelemahan seperti hasil pohon yang sangat rumit, pengulangan peubah split, dan juga pohon dengan tanpa peubah split, diperlukan penelitian lain dari sisi metodologi dan teori untuk mengatasi masalah tersebut.

Daftar Pustaka

- Faridhan, Y.E.** 2003. Metode Klasifikasi Berstruktur Pohon dengan Algoritma CRUISE, QUEST, dan CHAID. Tesis. Jurusan Statistika FMIPA IPB, Bogor.
- Gibson.** 1993. *Nutritional Assessment A Laboratory Manual*. Oxford University Press, Oxford.
- Kim, H, Loh W.-Y.** 2000. CRUISE User Manual. Revised ed. *Technical Report 989*. Dept. of Statistics, Univ. Of Wisconsin, Madison.
- Kim, H, Loh W.-Y.** 2001. Classification Trees with Unbiased Multiway *Splits*. *J. Am. Statist. Assoc.* 96:590–604.
- Loh, W.-Y, Shih Y.-S.** 1997. *Split Selection Methods for Classification Trees*. *Statistica Sinica* 7: 815–840.
- Novikasari, Mifta.** 2003. Perubahan Berat Badan dan Status Gizi Mahasiswa Jalur USMI tahun 2002 pada Empat Bulan Pertama di IPB. Jurusan Gizi Masyarakat dan Sumberdaya Keluarga Fakultas Pertanian IPB, Bogor.
- Statsoft, Inc.** 2002. *Electronic Statistics Textbooks*. Statsoft, Tulsa, OK. <http://www.statsoftinc.com/textbook/stathome.html>. [Mei 2002]
- WHO.** 1995. *Physical Status: The Use and Interpretation of Anthropometry*. Report of a WHO Expert Committee. WHO, Geneva