

IDENTIFIKASI FAKTOR YANG MEMPENGARUHI JENIS PEKERJAAN BERDASARKAN KARAKTERISTIK PENDUDUK DI SUMATERA BARAT

Widdya Rahmalina

*Program Studi Teknik Informatika, Fakultas Teknik, Universitas Abdurrah Pekanbaru
Jl. Riau Ujung No.73 Pekanbaru
Telp. (0761) 38762
widdya.rahmalina@univrab.ac.id*

ABSTRAK

Jenis pekerjaan di Sumatera Barat diduga dipengaruhi oleh faktor daerah tempat tinggal, jenis kelamin, bidang studi, pendidikan tertinggi yang ditamatkan, dan pengalaman kerja. Untuk melihat faktor-faktor yang mempengaruhi jenis pekerjaan tersebut digunakan metode pohon klasifikasi pada *Classification & Regression Tree* (CART). Metode pohon klasifikasi digunakan untuk menduga pengaruh peubah penjelas terhadap peubah respon kategorik dimulai dari pembentukan pohon awal dengan penyekatan, penentuan penyekatan terbaik, penghentian penyekatan, pemangkasan, sampai pada pemilihan pohon terbaik dan penentuan nilai dugaan respon. Data yang digunakan pada penelitian ini adalah data 6068 orang penduduk Sumatera Barat yang telah bekerja dengan pendidikan tertinggi minimal SMU sederajat yang diperoleh dari data mentah Survei Angkatan Kerja Nasional (SAKERNAS) bulan Agustus Tahun 2007. Berdasarkan hasil analisis CART dengan metode pohon klasifikasi, diperoleh pohon klasifikasi optimum yang terdiri dari 1 simpul akar, 3 simpul dalam, dan 5 simpul akhir. Simpul akar pertama kali disekat oleh peubah bidang studi, kemudian yang masuk simpul dalam adalah peubah daerah tempat tinggal, bidang studi, dan jenis kelamin. 5 simpul akhir menyatakan 5 kelompok masyarakat yang mempunyai pekerjaan utama, dimana 2 kelompok diantaranya didominasi oleh jenis pekerjaan tenaga profesional, teknisi dan tenaga lain ybdi. Oleh karena itu dapat disimpulkan bahwa faktor-faktor yang mempengaruhi jenis pekerjaan penduduk Sumatera Barat adalah bidang studi, daerah tempat tinggal dan jenis kelamin.

Kata kunci: analisis CART, metode pohon klasifikasi.

ABSTRACT

In West Sumatra, the type of works are influenced by area of residence, gender, field of study, highest educational attainment and work experience. To investigate the factors that influence the type of works are used the tree classification method on Classification and Regression Tree (CART). Tree Classification method is used to estimate the effect of the explanatory variables on the categorical response variable started from the initial tree formation with insulation, determining the best insulation, insulation termination, trimming, to the best tree selection and determination of the estimated value response. The data used in this study is the population data of West Sumatra who had been working with the highest level education of at least equivalent high school obtained from the raw data national labor survey (SAKERNAS) in August 2007. Based on the CART analysis by the tree classification method is obtained the optimum tree classification result that consists of one root node, 3 internal node, and 5 terminal node. The root node was first blocked by a variable field of study, then which include of internal node is residential areas, field of study, and gender. 5 terminal node of the stated five families with the main work, in which two groups of which is dominated by the kind of work of professionals, technicians and other related personnel. Therefore it can be concluded that the affected factors of the type of residents work in West Sumatra is a field of study, area of residence and gender.

Keywords: CART analysis, tree classification method.

I. PENDAHULUAN

1. Latar Belakang Masalah

Pekerjaan adalah suatu rangkaian tugas yang dirancang untuk dikerjakan oleh

satu orang dan sebagai imbalan diberikan upah dan gaji menurut kualifikasi dan berat ringannya pekerjaan tersebut. Penduduk Indonesia (khususnya penduduk usia kerja) bekerja dalam berbagai jenis pekerjaan yang

sangat banyak jenis dan variasinya. Jenis pekerjaan adalah kumpulan pekerjaan yang mempunyai rangkaian tugas yang bersamaan. Jenis pekerjaan dalam Klasifikasi Baku Jenis Pekerjaan Indonesia (KBJI) adalah kumpulan pekerjaan yang cukup bersamaan tugas utamanya sehingga bisa digabungkan bersama dalam satu kelompok dalam keseluruhan sistem klasifikasi.

Ada 10 golongan pokok jenis pekerjaan menurut Klasifikasi Baku Jenis Pekerjaan Indonesia (KBJI) yaitu : (Badan Pusat Statistik, 2002)

1. Pejabat lembaga legislatif, pejabat tinggi dan manajer
2. Tenaga professional
3. Teknisi dan asisten tenaga professional
4. Tenaga tata usaha
5. Tenaga usaha jasa dan tenaga penjualan di toko dan pasar
6. Tenaga usaha pertanian dan peternakan
7. Tenaga pengolahan dan kerajinan yang berhubungan dengan itu (ybdi).
8. Operator dan perakitan mesin
9. Pekerja kasar, tenaga kebersihan, dan tenaga ybdi
10. TNI dan POLRI

Setiap orang mempunyai bakat dan minat masing-masing. Bakat dan minat itulah yang akan menuntun mereka untuk memilih bidang yang mereka tekuni. Misalnya, orang yang berbakat melukis akan memilih bidang kesenian khususnya di bidang lukis, orang yang berbakat menghitung akan memilih bidang matematik, akuntansi dan sejenisnya. Bidang studi akan mempengaruhi jenis pekerjaan seseorang. Seseorang yang mempunyai bidang studi pendidikan biasanya menjadi tenaga pengajar, seseorang yang mempunyai bidang studi hukum dan politik akan menjadi seorang politisi, membuka jasa pengacara dan sejenisnya.

Selain bidang studi, jenis kelamin juga mempengaruhi jenis pekerjaan seseorang. Seorang perempuan umumnya bekerja sebagai tenaga pengajar, tenaga penjualan, pegawai sipil, tenaga kerajinan

seperti batik, anyaman dan sebagainya. Seorang laki-laki umumnya bekerja sebagai pejabat lembaga legislatif, pejabat tinggi, manajer, teknisi, tenaga usaha jasa dan penjualan, operator dan perakitan mesin, TNI dan POLRI sampai dengan pekerja kasar.

Daerah tempat tinggal juga menjadi salah satu faktor yang mempengaruhi jenis pekerjaan seseorang. Orang yang tinggal di pedesaan cenderung mempunyai jenis pekerjaan sebagai tenaga pertanian dan peternakan. Selain itu, masih banyak faktor lain yang mempengaruhi jenis pekerjaan seseorang seperti pendidikan terakhir yang ditamatkan, bidang studi dan sebagainya.

Dari faktor diatas, ingin dilihat faktor yang lebih mempengaruhi jenis pekerjaan penduduk Sumatera Barat. Untuk melihat faktor-faktor yang mempengaruhi jenis pekerjaan tersebut bisa dilakukan dengan mengidentifikasi jenis pekerjaan berdasarkan karakteristik penduduk. Karakteristik penduduk Sumatera Barat yang diambil dari data mentah Survei Angkatan kerja Nasional (SAKERNAS) yaitu daerah tempat tinggal, jenis kelamin, pendidikan tertinggi yang ditamatkan, bidang studi, dan pengalaman kerja.

Metode pohon klasifikasi sering digunakan untuk menduga pengaruh peubah penjelas terhadap peubah respon kategorik. Beberapa peneliti telah menggunakan metode tersebut dalam bidang kesehatan (Darsyah, M.Y., 2013) , (Sumartini, Siti Holis, 2015), dan dalam bidang kependudukan (Inayah, Riza, 2014).

2. Rumusan Masalah

Berdasarkan latar belakang di atas, masalah yang akan dibahas pada penelitian ini adalah bagaimana mengidentifikasi faktor-faktor yang mempengaruhi jenis pekerjaan berdasarkan karakteristik penduduk Sumatera Barat.

3. Batasan Masalah

Pada pembahasan penelitian ini batasan masalah difokuskan sebagai berikut:

- a. Penelitian ini terbatas pada data penduduk Sumatera Barat Bulan Agustus Tahun 2007.
- b. Data penduduk juga dibatasi yaitu yang telah bekerja dan memiliki pendidikan tertinggi yang ditamatkan minimal SMU sederajat.
- c. Peubah respon yakni jenis pekerjaan berdasarkan Klasifikasi Baku Jenis Pekerjaan Indonesia (KBJI) Tahun 2002.

4. Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengidentifikasi faktor-faktor yang mempengaruhi jenis pekerjaan berdasarkan karakteristik penduduk Sumatera Barat dan untuk mengetahui kelompok penduduk berdasarkan jenis pekerjaan di Sumatera Barat.

II. STUDI PUSTAKA

1. Metode CART

CART (*Classification and Regression Tree*) merupakan suatu metode yang digunakan untuk menduga hubungan antara peubah penjelas dengan peubah respon dalam suatu sistem. Adapun tujuan metode pohon klasifikasi dan pohon regresi adalah untuk mengetahui peubah penjelas mana yang berpengaruh terhadap peubah respon. Metode ini dimulai pada tahun 1960 dengan program *Automatic Interaction Detection* (AID) yang dikembangkan oleh Morgan dan Sonquist, yang ditujukan untuk

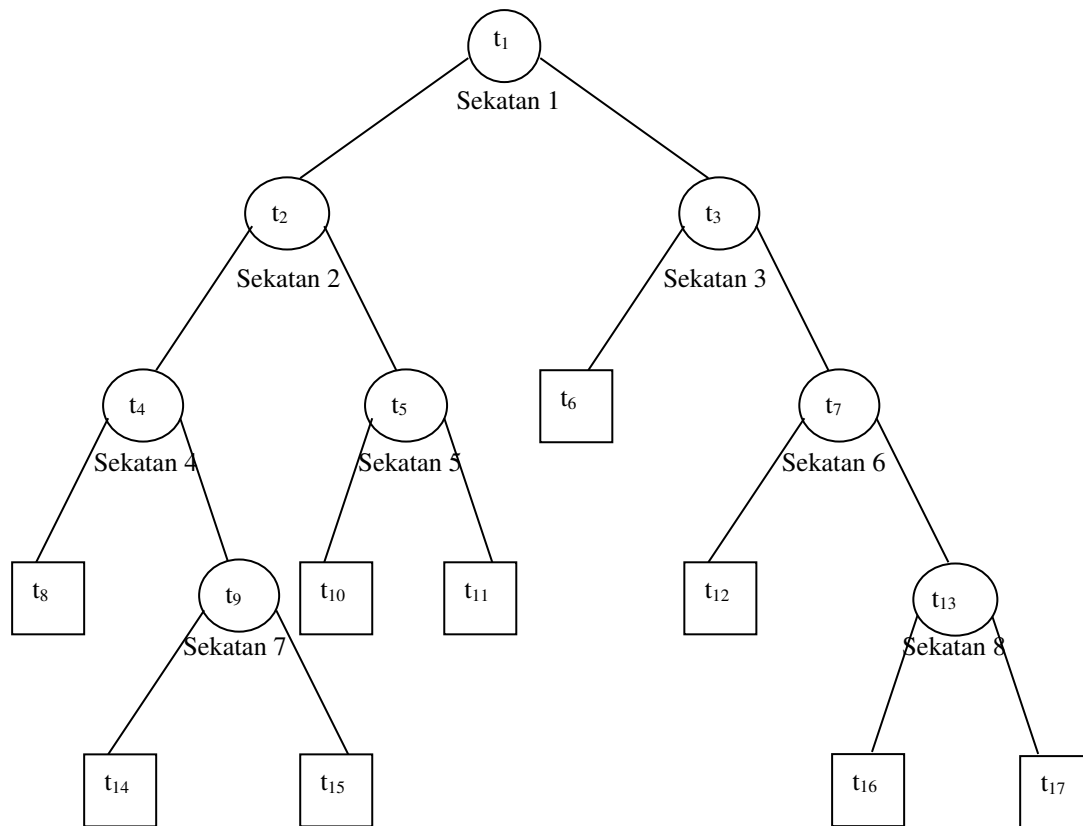
masalah-masalah regresi dan klasifikasi dengan melalui pembentukan pohon biner. Pada tahun 1973 Breiman dan Friedman secara terpisah menggunakan

metode pohon untuk masalah klasifikasi (Breiman, L, 1984).

Metode CART terdiri dari 2 cabang yaitu metode pohon klasifikasi dan pohon regresi. Metode pohon klasifikasi digunakan untuk menduga pengaruh peubah penjelas terhadap peubah respon kategorik. Algoritma lain yang dapat digunakan untuk menduga pengaruh peubah penjelas terhadap peubah respon kategorik adalah GDA (*Generalized Discriminant Function Analysis*). Sedangkan metode pohon regresi merupakan metode alternatif dari GLM (*Generalized Linear Models*) dan GRM (*Generalized Regression Models*) yang digunakan untuk menduga pengaruh peubah penjelas terhadap peubah respon kontinu.

Analisis data pada metode ini dilakukan melalui suatu prosedur iteratif, sehingga bila dilakukan secara manual, tentulah sangat rumit dan lama. Namun, dengan adanya perkembangan teknologi komputer, perhitungan dapat dilakukan dengan cepat dan mudah.

Metode CART diawali dengan melakukan penyekatan biner (*binary splitting*) terhadap keseluruhan pengamatan yang berada pada simpul akar menjadi 2 bagian yang dinamakan sebagai simpul anak. Pada tahap selanjutnya, setiap simpul anak ini akan menjadi simpul induk baru yang akan disekat lagi menjadi 2 simpul anak baru. Demikian seterusnya sampai diperoleh simpul-simpul yang tidak dapat disekat lagi. Hasil dari prosedur penyekatan pada CART ini disajikan dalam suatu diagram pohon seperti pada Gambar 1.



Gambar 1. Struktur Pohon Klasifikasi

2. Pembentukan Pohon Klasifikasi

Pembentukan pohon, baik klasifikasi maupun regresi dalam metode ini dilakukan berdasarkan penyekatan terhadap ruang peubah penjelas. Penyekatan yang dibentuk pada pohon regresi dilakukan dengan mempertimbangkan tingkat kehomogenan respon dalam setiap simpul. Sedangkan pada pohon klasifikasi, penyekatan dibentuk berdasarkan nilai perubahan *impurity* yang maksimum pada setiap simpul.

Langkah-langkah membuat pohon klasifikasi dalam metode CART adalah sebagai berikut :

1. Pembentukan pohon awal, terbagi atas :
 - a. Menentukan penyekatan setiap simpul
 - b. Menentukan penyekatan terbaik

- c. Menentukan simpul akhir atau penghentian penyekatan
2. Pemangkasan
3. Pemilihan pohon terbaik
4. Menentukan nilai dugaan setiap simpul akhir.

3. Pembentukan Pohon Awal dengan Penyekatan

Elemen yang dibutuhkan dalam proses pembentukan pohon awal adalah :

1. Himpunan Q yang berisi pertanyaan-pertanyaan biner dalam bentuk {Apakah $x \in A$ } dan $A \subset X$
2. Kriteria kebaikan sekatan $\phi(s, t)$ yang akan dievaluasi untuk setiap sekatan s dan simpul t
3. Aturan penghentian penyekatan

4. Pendugaan respon untuk setiap amatan

Himpunan Q yang berisi pertanyaan-pertanyaan biner dalam bentuk {apakah $x \in A$ } menghasilkan himpunan S dari sekatan-sekatan s pada tiap simpul t dan misalkan $A \subset X$ dimana X adalah ruang peubah penjelas, A adalah himpunan sekatan

3.1 Aturan Penyekatan

Proses penyekatan setiap simpul dilakukan dengan cara mencari semua kemungkinan penyekatan pada setiap peubah penjelas. Kemudian dicari penyekatan terbaik dari himpunan sekatan tersebut.

Penyekatan dapat dilakukan dengan aturan sebagai berikut :(Breiman, L, 1984)

1. Tiap sekatan hanya bergantung pada nilai yang berasal dari satu peubah penjelas.
2. Untuk peubah kontinu x_m , penyekatan yang diperbolehkan adalah untuk $x_m \leq t$ dan $x_m > t$, dimana t adalah nilai tengah antara 2 amatan secara berurutan. Jika ada n ruang contoh pada peubah kontinu, maka terdapat $n-1$ penyekatan.

Contoh 1 :

Peubah respon Y dengan peubah penjelas A (kontinu dengan nilai a_1, a_2, a_3, a_4, a_5) diurutkan menaik. Peubah penjelas A memiliki kemungkinan sekatan 4, dengan penyekatan sebagai berikut :

Tabel 1. Penyekatan Ruang Contoh Kontinu

Penyekatan Ruang Contoh Berdasarkan Peubah A			
No	Nilai t	Simpul Kiri	Simpul Kanan
1	$(a_1+a_2)/2$	$\{a_1\}$	$\{a_2, a_3, a_4, a_5\}$
2	$(a_2+a_3)/2$	$\{a_1, a_2\}$	$\{a_3, a_4, a_5\}$
3	$(a_3+a_4)/2$	$\{a_1, a_2, a_3\}$	$\{a_4, a_5\}$
4	$(a_4+a_5)/2$	$\{a_1, a_2, a_3, a_4\}$	$\{a_5\}$

3. Untuk peubah kategorik x_m ordinal mempunyai taraf L maka terdapat $L-1$ penyekatan, dan x_m nominal mempunyai taraf L , maka terdapat $2^{L-1}-1$ penyekatan.

pada simpul-simpul akhir.

dan x adalah suatu amatan. Untuk jawaban “ya” maka amatan dalam simpul t dimasukkan ke sekatan A dan untuk jawaban “tidak” maka amatan dalam simpul t dimasukkan ke sekatan A^c , dimana A^c adalah komplemen dari A dalam X .

Contoh 2 :

Peubah respon Y dengan peubah penjelas B (ordinal dengan nilai terurut $\{b_1, b_2, b_3, b_4\}$), dan C (nominal dengan nilai terurut $\{c_1, c_2, c_3\}$). Kemungkinan sekatan pada B adalah 3, dan pada C adalah 3.

Tabel 2. Penyekatan Ruang Contoh Ordinal dan Nominal

Penyekatan Ruang Contoh Berdasarkan Peubah B dan C				
No	Peubah B		Peubah C	
	Simpul Kiri	Simpul Kanan	Simpul Kiri	Simpul Kanan
1	$\{b_1\}$	$\{b_2, b_3, b_4\}$	$\{c_1\}$	$\{c_2, c_3\}$
2	$\{b_1, b_2\}$	$\{b_3, b_4\}$	$\{c_2\}$	$\{c_1, c_3\}$
3	$\{b_1, b_2, b_3\}$	$\{b_4\}$	$\{c_3\}$	$\{c_1, c_2\}$

3.2 Kriteria Keباikan Sekatan (*Good of Split Criterion*)

Kemungkinan sekatan yang dapat terbentuk pada suatu simpul sangat banyak, sehingga sulit menafsirkan peubah penjelas mana yang paling mempengaruhi peubah respon. Oleh karena itu, harus dipilih sekatan terbaik diantara kumpulan sekatan dari setiap pengamatan. Pemilihan sekatan terbaik dilakukan dengan menggunakan kriteria gini (*gini criterion*) yang dihitung berdasarkan pada nilai perubahan *impurity* yang maksimum.(Breiman, L, 1984)

Misalkan $p(j|t)$ adalah peluang peubah respon kategori ke- j ($j=1,2,\dots,J$; j adalah banyak kategori peubah respon) pada simpul ke- t , sehingga $p(1|t) + p(2|t) + \dots + p(J|t) = 1$. Simpul t disekat oleh sekatan s dengan penyekatan biner menjadi simpul t_L dan t_R . Penyekatan dibentuk dari pertanyaan bahwa {Apakah

$x_i \in A$ } dan $A \subset X$, jika “ya” maka semua amatan x_i masuk ke simpul t_L dan jika “tidak” maka semua amatan x_i masuk ke simpul t_R . Misalkan proporsi pengamatan yang masuk ke simpul t_L adalah P_L dan proporsi pengamatan yang masuk pada simpul t_R adalah P_R .

Definisikan fungsi *impurity* pada simpul t , $I(t)$, sebagai fungsi dari $p(j|t)$; ($j = 1, 2, \dots, J$) yaitu :

$$I(t) = \phi(p(1|t), \dots, p(J|t)) \dots\dots\dots(1)$$

Dengan kriteria gini, fungsi *impurity* tersebut didefinisikan sebagai :

$$I(t) = \sum_{j \neq i} p(j|t)p(i|t) \dots\dots\dots(2)$$

yang dapat ditulis sebagai :

$$I(t) = 1 - \sum_j p^2(j|t) \dots\dots\dots(3)$$

Jika s adalah sekatan pada simpul t , P_L dan P_R masing-masing adalah proporsi amatan yang masuk ke simpul t_L dan t_R serta $I(t_L)$ dan $I(t_R)$ masing-masing adalah *impurity* pada simpul t_L dan t_R . Maka perubahan *impurity* akibat penyekatan t menjadi t_L dan t_R oleh sekatan s adalah :

$$\Delta I(s, t) = I(t) - P_R I(t_R) - P_L I(t_L) \dots\dots\dots(4)$$

Sekatan terbaik dari himpunan sekatan S adalah sekatan s^* yang dapat memaksimumkan perubahan *impurity* atau s^* sehingga :

$$\Delta I(s^*, t) = \text{Max}_{s \in S} \Delta I(s, t)$$

3.3

roses Penghentian Penyekatan

Setelah melakukan beberapa tahap penyekatan terhadap simpul-simpul, maka proses penyekatan akan dihentikan bila ditemui salah satu keadaan sebagai berikut : (Breiman, L, 1984)

1. bila jumlah amatan pada setiap simpul kurang dari jumlah amatan minimum yang ditetapkan.

2. semua amatan pada simpul anak mempunyai sebaran nilai dugaan yang mengakibatkan penyekatan tidak mungkin dilanjutkan.

3. jika nilai perubahan *impurity* simpul pada setiap peubah penjelas sudah kecil dari nilai batas *impurity* minimum yang telah ditentukan. Misalkan $\beta > 0$ adalah nilai batas minimum *impurity* maka penyekatan akan berhenti jika $\text{Max}_{s \in S} \Delta I(s, t) < \beta$

4. Pemangkasan

Karena penghentian penyekatan dilakukan berdasarkan perubahan *impurity*, maka pohon yang terbentuk dapat berukuran sangat besar. Semakin banyak sekatan, semakin besar pohon yang dihasilkan dengan tingkat kesalahan yang kecil. Namun, semakin besar pohon, interpretasi menjadi semakin sulit. Sebaliknya jika sekatan sedikit, pohon yang dihasilkan tidak besar dan mudah menginterpretasikannya, tetapi tingkat kesalahannya besar. Untuk menghindari kedua kondisi ekstrim di atas maka dicari pohon dengan ukuran yang layak.

Pemilihan pohon dengan ukuran yang layak dapat dilakukan dengan langkah sebagai berikut : (Breiman, L, 1984)

1. Penentuan pohon awal yang besar
2. Secara iteratif pohon tersebut dipangkas (*pruning*) menjadi sekuen pohon yang semakin kecil. **P**
3. Pilih pohon terbaik dari sekuen tersebut dengan menggunakan contoh uji (*test sample*) atau contoh validasi silang (*cross validation sample*).

Pemangkasan pada pohon klasifikasi dilakukan dengan menggunakan ukuran biaya kompleksitas (*Cost-Complexity*) minimum. Pendefinisian pohon besar yang dibentuk dari prosedur penyekatan adalah T_{max} . Pemangkasan dilakukan bukan terhadap

T_{max} tetapi terhadap T_1 yaitu pohon terkecil yang memenuhi kondisi : $R(T_{max}) = R(T_1)$.

Untuk mendapatkan T_1 dari T_{max} dilakukan evaluasi terhadap semua t_L dan t_R yang merupakan simpul akhir dari pohon besar T_{max} . Jika $R(t) = R(t_L) + R(t_R)$, maka pangkas kedua simpul akhir tersebut sampai tidak ada lagi yang mungkin untuk diperoleh, sehingga diperoleh T_1 .

Secara umum untuk sembarang T yang merupakan sub pohon dari T_{max} dapat didefinisikan suatu ukuran biaya kompleksitas sebagai berikut :

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \dots\dots\dots(6)$$

dimana :

- $R_\alpha(T)$ = ukuran biaya kompleksitas
- $R(T)$ = nilai kesalahan pengelompokan dari pohon pada biaya kompleksitas simpul akhir
- $|\tilde{T}|$ = banyak simpul akhir pada pohon T
- α = parameter biaya kompleksitas

Untuk setiap $t \in T$ kesalahan pengelompokan pada simpul T didefinisikan :

$$r(t) = 1 - \max_j p(j|t) \dots\dots\dots(7)$$

kemudian ditulis :

$$R(t) = r(t) p(t)$$

dimana

$$p(t) = \sum_j p(j,t) = \sum_j \frac{\pi(j)N_j(t)}{N_j} = \sum_j \frac{N_j \cdot N_j(t)}{N \cdot N_j} = \sum_j \frac{N_j(t)}{N}$$

$$= \frac{N_1(t)}{N} + \frac{N_2(t)}{N} + \dots + \frac{N_j(t)}{N}$$

$$p(t) = \frac{N(t)}{N} \dots\dots\dots(9)$$

Kesalahan pengelompokan untuk semua simpul t pada pohon T adalah :

$$R(T) = \sum_{t \in \tilde{T}} R(t) \dots\dots\dots(10)$$

Proses pemangkasan minimum merupakan pemotongan jalur terlemah

(weakest link). Untuk setiap T_i yang merupakan simpul dalam dari T_1 , ukuran biaya kompleksitasnya dapat didefinisikan sebagai :

$$R_\alpha(T_i) = R(T_i) + \alpha |\tilde{T}_i| \dots\dots\dots(11)$$

Sedangkan biaya kompleksitas dari simpul dalam $t \in T_i$ dapat didefinisikan sebagai :

$$R_\alpha(\{t\}) = R(t) + \alpha \dots\dots\dots(12)$$

Selama $R_\alpha(T_i) < R_\alpha(\{t\})$, cabang T_i memiliki biaya kompleksitas lebih kecil dari simpul tunggal $\{t\}$. Tapi untuk suatu simpul t tertentu, nilai $R_\alpha(T_i)$ sama dengan nilai $R_\alpha(\{t\})$, tergantung dari nilai kritis α . Nilai kritis dari α diperoleh dengan menyamakan kedua nilai kompleksitas tersebut yang dapat didefinisikan sebagai :

$$\alpha = \frac{R(t) - R(T_i)}{|\tilde{T}_i| - 1} \dots\dots\dots(13)$$

Kemudian didefinisikan suatu fungsi $h_1(t)$, $t \in T_1$ yaitu :

$$h_1(t) = \begin{cases} \frac{R(t) - R(T_i)}{|\tilde{T}_i| - 1}, & t \notin \tilde{T}_1 \\ +\infty, & t \in \tilde{T}_1 \end{cases} \dots\dots\dots(14)$$

Sehingga jalur terlemah dalam simpul T_1 dilambangkan dengan \bar{t}_1 merupakan simpul yang memenuhi kriteria :

$$h_1(\bar{t}_1) = \min_{t \in T_1} h_1(t) \dots\dots\dots(15)$$

Nilai parameter kompleksitas yang memenuhi untuk persamaan tersebut didefinisikan sebagai :

$$\alpha_2 = h_1(\bar{t}_1) \dots\dots\dots(16)$$

Kemudian dibentuk pohon baru dengan cara memangkas cabang $T_{\bar{t}_1}$ yang merupakan cabang dari simpul dalam T_1 yang memiliki simpul utama $\{\bar{t}_1\}$. Hasil pemangkasan ini adalah T_2 yaitu :

$$T_2 = T_1 - T_{\bar{t}_1} \dots\dots\dots(17)$$

Sehingga T_2 merupakan sub pohon yang memenuhi kriteria biaya kompleksitas minimum dengan parameter kompleksitas bernilai α_2 .

Demikian selanjutnya dengan cara yang sama tentukan jalur terlemah dalam T_2 dan seterusnya. Akan didapatkan sub pohon yang tersarang yaitu $\{T_1, T_2, T_3, \dots, \{t\}\}$ dengan hubungan $T_1 > T_2 > T_3 > \dots > \{t\}$ dan himpunan ukuran kompleksitas $\{\alpha_1, \alpha_2, \alpha_3, \dots\}$ dengan hubungan $\alpha_1 < \alpha_2 < \alpha_3 < \dots$

Jika terdapat jalur terlemah yang sama pada langkah ke- k , $k=1,2,3,\dots$, misal $h_k(\bar{t}_k) = h_k(\bar{t}'_k)$ maka pemangkasan dilakukan dengan cara :

$$T_{k+1} = T_k - (T_{\bar{t}_k} - T_{\bar{t}'_k}) \dots \dots \dots (18)$$

5. Pemilihan Pohon terbaik

Setelah dilakukan pemangkasan, maka harus dipilih pohon terbaik dari himpunan cabang pohon yang sudah dipangkas. Pemilihan pohon terbaik dapat dilakukan dengan dua cara yaitu :

1. Dugaan contoh uji (test sample estimate)

Test sample estimate dinotasikan dengan $R^{ts}(T)$. Untuk memperoleh nilai $R^{ts}(T)$ maka amatan L dibagi menjadi dua secara acak yaitu L_1 sebagai *learning sample* dan L_2 sebagai *test sample*. L_1 digunakan untuk menentukan sekuens pohon $\{T_k\}$ dengan urutan yaitu $T_1 > T_2 > \dots > \{t_1\}$ yang melalui pemangkasan. Sedangkan L_2 digunakan untuk menentukan $R^{ts}(T_k)$. Jika L_2 berukuran $N^{(2)}$, maka :

$$R^{ts}(T) = \frac{1}{N^{(2)}} \sum_{i,j} c(i|j) N_{ij}^{(2)}$$

dimana $R^{ts}(T)$ = nilai dugaan berdasarkan total kesalahan pengelompokkan pada T. Pohon terbaik adalah T_{k0} , dengan kriteria sebagai berikut :

$$R^{ts}(T_{k0}) = \min_k R^{ts}(T_k)$$

2. validasi silang (cross-validation estimates)

Untuk menentukan *cross-validated estimate* yaitu dengan *v-fold*, jika amatan L berukuran N yang dibagi secara acak menjadi V kelompok, yaitu L_1, L_2, \dots, L_v yang berukuran sama. *Learning sample* ke- v adalah $L^{(v)} = L - L_v$, $v = 1, 2, \dots, V$ yang digunakan untuk membentuk urutan pohon $\{T_k\}$ dan urutan parameter kompleksitas $\{\alpha_k\}$. Sehingga diperoleh v urutan $\{T_k\}$ dan v urutan $\{\alpha_k\}$, kemudian dengan menggunakan amatan L dapat dibentuk urutan $\{T_k\}$ dan urutan $\{\alpha_k\}$. Biaya kompleksitas minimum pada pohon dari amatan ke- L berukuran N adalah :

$$R^{cv}(T(\alpha)) = \frac{1}{N} \sum_{i,j} c(i|j) N_{ij} \dots \dots (21)$$

dimana sama dengan biaya kompleksitas minimum T_k untuk $\alpha_k \leq \alpha \leq \alpha_{k+1}$.

Definisikan $\alpha'_k = \sqrt{\alpha_k \alpha_{k+1}}$ sehingga α'_k adalah nilai tengah pada parameter biaya kompleksitas yang mengakibatkan $T(\alpha) = T_k$, maka diperoleh :

$$R^{cv}(T_k) = R^{cv}(T(\alpha'_k))$$

dimana $R^{cv}(T_k)$ = nilai dugaan pada *test sample* L_v dan pohon $T^{(v)}(\alpha'_k)$.

Pohon terbaik adalah T_{k0} , yaitu :

$$R^{cv}(T_{k0}) = \min_k R^{cv}(T_k) \dots \dots \dots (23)$$

dimana $R^{cv}(T_{k0})$ adalah nilai dugaan kesalahan pengelompokkan.

6. Penentuan Nilai Dugaan Respon

Setiap simpul dimulai dari simpul induk yang ditentukan nilai dugaannya. Nilai dugaan pada setiap simpul bergantung pada beberapa faktor, yaitu :

1. Asumsikan peluang awal setiap simpul dalam kelompok data

2. Fungsi subjek pada setiap hasil dalam kelompok data yang berakhir pada setiap simpul.

Misalkan $j \in \{1, 2, \dots, J\}$ adalah kelompok dugaan dan simpul akhir $t \in \tilde{T}$ dengan \tilde{T} adalah sub pohon T, maka nilai dugaan pada simpul akhir $t \in \tilde{T}$ dinotasikan dengan $\hat{j}(t)$. Aturan penentuan nilai dugaan, yaitu :

$$\hat{j}(t) = \max_j P(j|t)$$

Misalkan $\hat{j}(t)$ dikatakan nilai dugaan pada simpul t jika $P(j|t) = \max_i P(i|t)$ maka $\hat{j}(t) = j$. Jika dua atau lebih kelompok maksimum yang berbeda maka $\hat{j}(t)$ juga maksimum.

III. METODE PENELITIAN

Pada penelitian ini data yang diambil adalah data karakteristik penduduk Sumatera Barat yang bekerja dengan klasifikasi jenis pekerjaan yang telah ditetapkan Badan Pusat Statistik Pusat. Data ini berukuran 6068 orang penduduk yang bersumber dari data Survei Angkatan Kerja Nasional (SAKERNAS) Bulan Agustus Tahun 2007 yang diperoleh dari Badan Pusat Statistik (BPS) Propinsi Sumatera Barat. Peubah-peubah yang diamati dalam penelitian ini terdiri dari peubah respon yaitu jenis pekerjaan dan peubah penjelas yaitu daerah tempat tinggal, jenis kelamin, pendidikan tertinggi yang ditamatkan minimal SMU sederajat, jurusan pendidikan/bidang studi, dan pengalaman kerja.

Untuk memudahkan dalam mengolah data, maka peubah-peubah yang berkaitan perlu dikodekan.

Tabel 3. Pengkodean peubah

Peubah	Jenis Peubah	Nilai	Kode
Jenis Pekerjaan	Peubah Respon	Tenaga Profesional, Teknisi dan Tenaga Lain Ybdi	1
		Tenaga Kepemimpinan dan Ketatalaksanaan	2
		Pejabat Pelaksana, Tenaga Tata Usaha dan Tenaga Ybdi	3
		Tenaga Usaha Penjualan	4
		Tenaga Usaha Jasa	5
		TU Tani, Kebun, Ternak, Ikan, Hutan dan Perburuan	6
		TNI dan POLRI	7
		Tenaga Produksi Op.Alat Angkutan dan Pekerja Kasar	8
Daerah	Peubah Penjelas	Perkotaan	0
		Pedesaan	1
Jenis Kelamin	Peubah Penjelas	Laki-laki	0
		Perempuan	1
Pendidikan Yang Tertinggi Ditamatkan	Peubah Penjelas	SMU	1
		SMK	2
		Diploma I/II	3
		Akademi/Diploma III	4

		Diploma IV/S1/S2	5		
Bidang Studi	Peubah Penjelas	Agama dan Ilmu Ketuhanan	1		
		Administrasi/Manajemen Bisnis	2		
		Administrasi/Manajemen Keuangan	3		
		Administrasi Pemerintahan	4		
		Administrasi/Manajemen Perkantoran	5		
		Arsitektur dan Perencanaan Kota	6		
		Bahasa dan Sastra	7		
		Ekonomi	8		
		Hukum dan Kehakiman	9		
		Humanisme/Humaniora	10		
		Ilmu Pengetahuan Alam (IPA)	11		
		Ilmu Sosial dan Politik (IPS)	12		
		Kedokteran dan Kesehatan	13		
		Kehutanan	14		
		Kesenian dan Seni Rupa	15		
		Komunikasi Massa dan Dokumentasi	16		
		Matematika dan Ilmu Komputer	17		
		Pelayanan Jasa	18		
		Perikanan	19		
		Pertanian	20		
		Pertukangan, Kerajinan dan Industri	21		
		Peternakan	22		
		Psikologi	23		
		Teknik/Teknologi	24		
		Transportasi dan Komunikasi	25		
		Kependidikan dan Keguruan	26		
		Lainnya	27		
		IPA SMU	28		
		IPS SMU	29		
		Fisika SMU	30		
		Biologi SMU	31		
		Bahasa dan Sastra SMU	32		
		Listrik, Mesin, Grafika, Tekstil, Penerbangan dan lainnya (STM)	33		
		Ekonomi, Akuntansi, dan Tata Niaga (SMEA)	34		
		Tata Boga, Tata Busana, dan Tata Graha (SMKK)	35		
		Perhotelan, Pariwisata, Perawat dan lainnya (SMK)	36		
		Pengalaman Kerja	Peubah Penjelas	Ya	0
				Tidak	1

Analisis data dilakukan dengan metode CART dengan langkah-langkah sebagai berikut :

1. Pembentukan pohon awal dengan penyekatan menggunakan kriteria Gini

- (*Gini Criterion*), dengan menetapkan nilai batas *impurity* minimum 0,0001.
2. Penentuan penyekatan terbaik berdasarkan perubahan *impurity* yang maksimum.
 3. Penghentian penyekatan jika maksimum kurang dari nilai batas *impurity* minimum.
 4. Pemangkasan terhadap pohon yang berukuran besar dengan menggunakan ukuran biaya kompleksitas (*Cost-Complexity*) minimum.
 5. Pemilihan pohon terbaik dari himpunan yang terbentuk akibat pemangkasan,

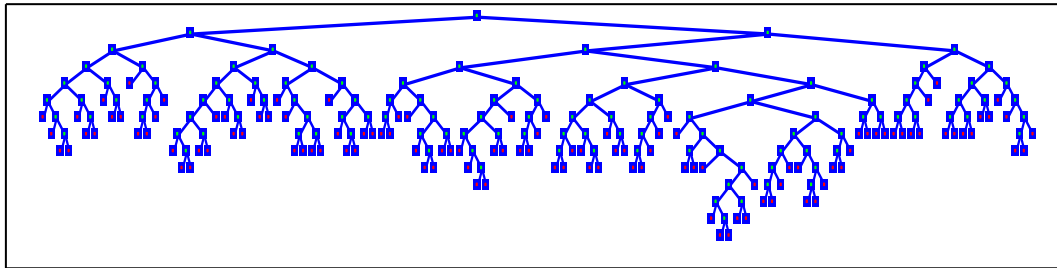
dengan menggunakan kriteria dugaan contoh uji (*test sample estimate*) atau contoh validasi silang (*cross validation sample*).

6. Penentuan nilai dugaan respon.

Pengolahan data dilakukan dengan bantuan *Software SPSS versi 16* dan *CART versi 4.0*.

IV. HASIL DAN PEMBAHASAN

Hasil analisis CART yaitu pohon klasifikasi dengan penyekatan untuk jenis pekerjaan di Propinsi Sumatera Barat dapat dilihat pada gambar berikut :



Gambar 2. Pohon klasifikasi awal

Pohon awal yang terbentuk memberikan 115 buah simpul akhir. Pohon yang terbentuk sangat besar sehingga sulit untuk menginterpretasi hasilnya. Oleh karena itu dilakukan pemangkasan, dengan tujuan agar pohon memiliki struktur sederhana dan informasi yang dibutuhkan tidak hilang. Metode pemangkasan yang digunakan adalah validasi silang lipat 10. Pada Tabel 4 memperlihatkan *Cross Validated Relative Error* yang terbentuk dari hasil pemangkasan.

Tabel 4. Kesalahan relatif validasi silang

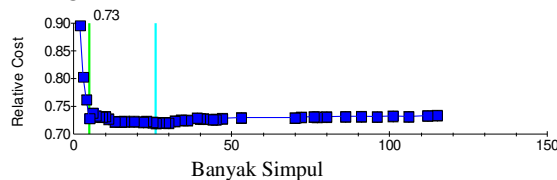
Tree Number	Terminal Nodes	Cross-Validated Relative Cost	Resubstitution Relative Cost	Complexity
1	115	0.73 ± 0.0096	0.64	-1.00
2	112	0.73 ± 0.0096	0.64	0.00011
3	106	0.73 ± 0.0096	0.65	0.00015
4	101	0.73 ± 0.0096	0.65	0.00017
5	96	0.73 ± 0.0097	0.65	0.00019
6	91	0.73 ± 0.0097	0.65	0.00020
7	86	0.73 ± 0.0097	0.65	0.00022
8	80	0.73 ± 0.0097	0.65	0.00024

Tree Number	Terminal Nodes	Cross-Validated Relative Cost	Resubstitution Relative Cost	Complexity
9	77	0.73 ± 0.0096	0.65	0.00027
10	76	0.73 ± 0.0096	0.65	0.00028
11	72	0.73 ± 0.0096	0.65	0.00030
12	70	0.73 ± 0.0097	0.65	0.00034
13	53	0.73 ± 0.0097	0.66	0.00037
14	47	0.73 ± 0.0097	0.66	0.00039
15	45	0.73 ± 0.0097	0.67	0.00040
16	44	0.73 ± 0.0097	0.67	0.00043
17	41	0.73 ± 0.0097	0.67	0.00046
18	40	0.73 ± 0.0097	0.67	0.00047
19	39	0.73 ± 0.0097	0.67	0.00049
20	36	0.73 ± 0.0098	0.67	0.00052
21	34	0.73 ± 0.0098	0.67	0.00053
22	32	0.72 ± 0.0098	0.67	0.00058
23	30	0.72 ± 0.0099	0.67	0.00062
24	29	0.72 ± 0.0099	0.67	0.00073
25	27	0.72 ± 0.0099	0.68	0.00075
26*	26	0.72 ± 0.0099	0.68	0.00076
27	25	0.72 ± 0.0099	0.68	0.00085
28	23	0.72 ± 0.0099	0.68	0.00093
29	22	0.72 ± 0.0099	0.68	0.00095
30	19	0.72 ± 0.0099	0.68	0.00096
31	16	0.72 ± 0.0099	0.69	0.00099
32	15	0.72 ± 0.0099	0.69	0.0010
33	13	0.72 ± 0.0098	0.69	0.0013
34	11	0.73 ± 0.0098	0.70	0.0017
35	10	0.73 ± 0.0096	0.70	0.0024

Tree Number	Terminal Nodes	Cross-Validated Relative Cost	Resubstitution Relative Cost	Complexity
36	9	0.73 ± 0.0095	0.70	0.0024
37	8	0.73 ± 0.0095	0.70	0.0031
38	7	0.73 ± 0.0095	0.71	0.0032
39	6	0.74 ± 0.0094	0.71	0.0035
40**	5	0.73 ± 0.0089	0.72	0.0046
41	4	0.76 ± 0.0086	0.75	0.03
42	3	0.80 ± 0.0052	0.80	0.04
43	2	0.90 ± 0.0030	0.89	0.08
44	1	1.00 ± 0.00	1.00	0.09

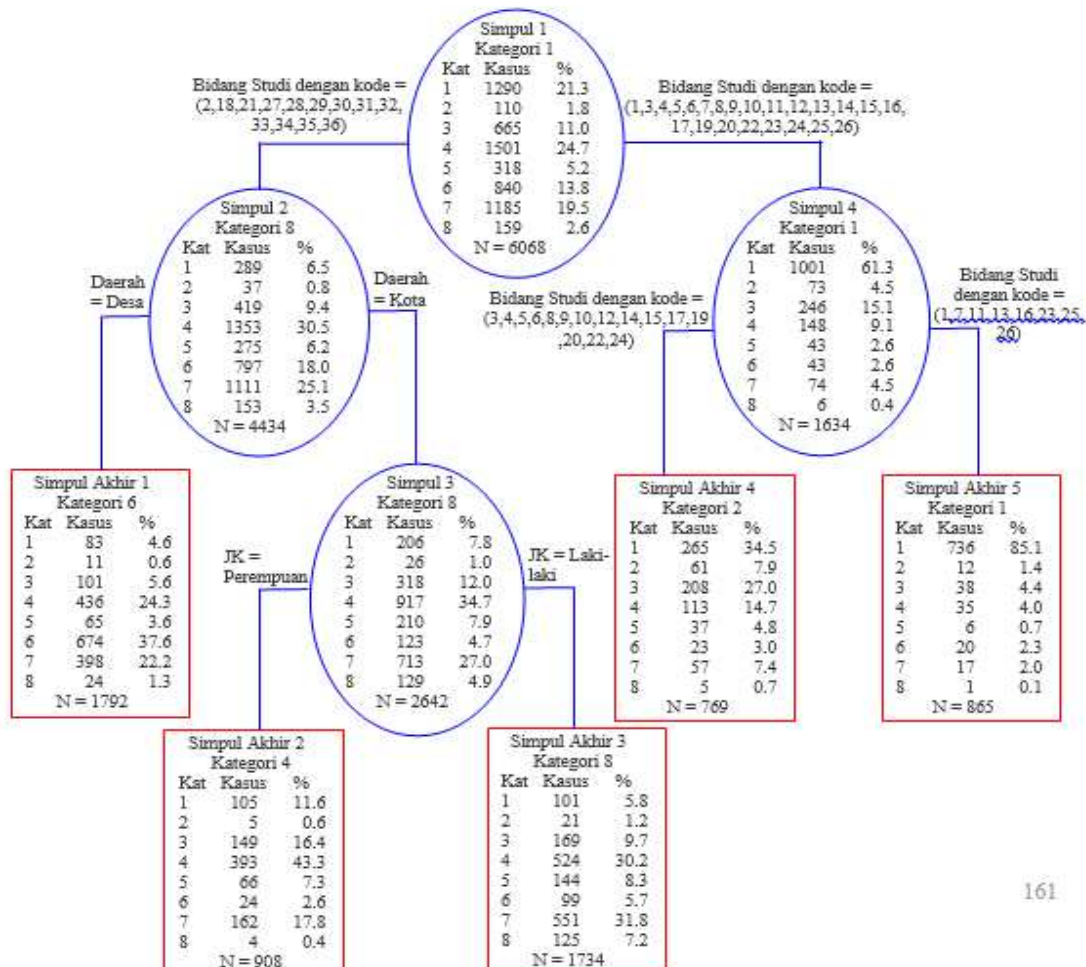
* Minimum Cost
 ** Optimal

Pada gambar di bawah ini disajikan plot hubungan antara banyak simpul akhir dengan kesalahan relatif dari pohon hasil pemangkasan.



Gambar 3. Plot hubungan jumlah simpul dengan *relative cost*

Dari Tabel 4 dan Gambar 3 dapat dilihat bahwa pohon yang memiliki kesalahan relatif terkecil adalah pohon dengan 26 simpul akhir. Dengan demikian pohon terbaik berdasarkan kriteria kesalahan relatif adalah pohon dengan 26 simpul akhir. Kesalahan untuk pohon ini berada dalam selang $0,72 \pm 0,0099$ atau berada dalam selang $0,7101$ dan $0,7299$. Pada Tabel 4 dapat diketahui bahwa masih terdapat pohon yang lebih kecil dan memiliki nilai kesalahan relatif masih berada dalam selang tersebut yaitu pohon dengan 5 simpul akhir. Dengan demikian, berdasarkan kriteria satu standar error dapat dipilih pohon dengan 5 simpul akhir. Pohon terbaik yang dipilih dapat dilihat pada gambar berikut :



Gambar 4. Pohon Terbaik dari Hasil Pemangkasan

Pohon klasifikasi optimum tersebut mempunyai 1 simpul utama (*root node*), 3 simpul dalam (*internal node*) dan 5 simpul akhir (*terminal node*). Peubah yang menyekat simpul akar adalah bidang studi, sedangkan yang masuk ke simpul dalam adalah peubah daerah tempat tinggal, bidang studi, dan jenis kelamin. Dengan demikian, faktor-faktor yang mempengaruhi jenis pekerjaan masyarakat Sumatera Barat adalah bidang studi, daerah tempat tinggal, dan jenis kelamin. Peubah penjelas lainnya seperti pendidikan tertinggi yang ditamatkan dan pengalaman kerja tidak muncul dalam pemilihan pohon terbaik (optimum). Diduga bahwa pendidikan tertinggi yang ditamatkan sangat erat hubungannya dengan bidang studi, sehingga bidang studi saja sudah cukup mewakili dalam pemilihan pohon terbaik. Pengalaman kerja tidak dapat menjadi peubah penyekat dalam pembentukan pohon optimum diduga disebabkan oleh faktor kehomogenan peubah tersebut.

Berdasarkan Tabel 3 tentang pengkodean peubah yang telah dipresentasikan pada bagian metode penelitian dan juga pada Gambar 4 juga dapat dilihat proses penyekatan sebagai berikut :

1. Tahap I

Dari 6068 responden pada simpul 1 dipecah oleh peubah bidang studi menjadi 4434 responden dengan bidang studi administrasi/manajemen bisnis; pelayanan jasa; pertukangan, kerajinan dan industri; lainnya; IPA SMU; IPS SMU; Fisika SMU; Biologi SMU; Bahasa dan Sastra SMU; Listrik, mesin, grafika, tekstil, penerbangan dan lainnya (STM); Ekonomi, akuntansi dan tata

niaga (SMEA); Tata boga, tata busana dan tata graha (SMKK); Perhotelan, pariwisata, perawat dan lainnya (SMK) sebagai simpul 2 dan 1634 responden dengan bidang studi agama dan ilmu ketuhanan; administrasi/manajemen keuangan; administrasi pemerintahan; administrasi/manajemen perkantoran; arsitektur dan perencanaan kota; bahasa dan sastra; ekonomi; hukum dan kehakiman; humanisme/humaniora; IPA; IPS; kedokteran dan kesehatan; kehumanan; kesenian dan seni rupa; komunikasi massa dan dokumentasi; matematika dan ilmu komputer; perikanan; pertanian; peternakan; psikologi; teknik/teknologi; transportasi dan komunikasi; kependidikan dan keguruan sebagai simpul 4.

2. Tahap II

Dari 4434 responden pada simpul 2 dipecah oleh peubah daerah menjadi 1792 responden di pedesaan sebagai simpul akhir 1 dan 2642 responden di perkotaan sebagai simpul 3.

3. Tahap III

Dari 2642 responden pada simpul 3 dipecah oleh peubah jenis kelamin menjadi 908 responden perempuan sebagai simpul akhir 2 dan 1734 responden laki-laki sebagai simpul akhir 3.

4. Tahap IV

Dari 1634 responden pada simpul 4 dipecah oleh peubah bidang studi menjadi 769 responden dengan bidang studi administrasi/manajemen keuangan; administrasi pemerintahan; administrasi/manajemen perkantoran; arsitektur dan perencanaan kota; bahasa dan sastra; ekonomi; hukum dan kehakiman; humanisme/humaniora; IPS; kehumanan; kesenian dan seni rupa;

matematika dan ilmu komputer; perikanan; pertanian; peternakan; teknik/teknologi sebagai simpul akhir 4 dan 865 responden dengan bidang studi agama dan ilmu ketuhanan; IPA; kedokteran dan kesehatan; komunikasi massa dan dokumentasi; psikologi; transportasi dan komunikasi; kependidikan dan kejuruan sebagai simpul akhir 5.

Berdasarkan hasil analisis data pada gambar 3 diperoleh 5 simpul akhir yang berarti terdapat 5 kelompok masyarakat yang mempunyai pekerjaan utama dengan pendidikan tertinggi yang ditamatkan minimal SMU sederajat yaitu :

- a. Simpul akhir 1 : Kelompok masyarakat dengan bidang studi administrasi/manajemen bisnis; pelayanan jasa; pertukangan, kerajinan dan industri; lainnya; IPA SMU; IPS SMU; Fisika SMU; Biologi SMU; Bahasa dan Sastra SMU; Listrik, mesin, grafika, tekstil, penerbangan dan lainnya (STM); Ekonomi, akuntansi dan tata niaga (SMEA); Tata boga, tata busana dan tata graha (SMKK); Perhotelan, pariwisata, perawat dan lainnya (SMK) di daerah pedesaan didominasi oleh jenis pekerjaan tenaga usaha tani, kebun, ternak, ikan, hutan dan perburuan sebesar 37,6%.
- b. Simpul akhir 2 : Kelompok masyarakat dengan bidang studi administrasi/manajemen bisnis; pelayanan jasa; pertukangan, kerajinan dan industri; lainnya; IPA SMU; IPS SMU; Fisika SMU; Biologi SMU; Bahasa dan

Sastra SMU; Listrik, mesin, grafika, tekstil, penerbangan dan lainnya (STM); Ekonomi, akuntansi dan tata niaga (SMEA); Tata boga, tata busana dan tata graha (SMKK); Perhotelan, pariwisata, perawat dan lainnya (SMK) berjenis kelamin perempuan didominasi oleh jenis pekerjaan tenaga usaha penjualan sebesar 43,3%.

- c. Simpul akhir 3 : Kelompok masyarakat dengan bidang studi administrasi/manajemen bisnis; pelayanan jasa; pertukangan, kerajinan dan industri; lainnya; IPA SMU; IPS SMU; Fisika SMU; Biologi SMU; Bahasa dan Sastra SMU; Listrik, mesin, grafika, tekstil, penerbangan dan lainnya (STM); Ekonomi, akuntansi dan tata niaga (SMEA); Tata boga, tata busana dan tata graha (SMKK); Perhotelan, pariwisata, perawat dan lainnya (SMK) berjenis kelamin laki-laki didominasi oleh jenis pekerjaan TNI dan POLRI sebesar 31,8%.
- d. Simpul akhir 4 : Kelompok masyarakat dengan bidang studi administrasi/manajemen keuangan; administrasi pemerintahan; administrasi/manajemen perkantoran; arsitektur dan perencanaan kota; bahasa dan sastra; ekonomi; hukum dan kehakiman; humanisme/humaniora; IPS; kehutanan; kesenian dan seni rupa; matematika dan ilmu komputer; perikanan; pertanian; peternakan; teknik/teknologi

didominasi oleh jenis pekerjaan tenaga profesional, teknisi, dan tenaga lain ybdi sebesar 34,5%.

- e. Simpul akhir 5 : Kelompok masyarakat dengan bidang studi agama dan ilmu ketuhanan; IPA; kedokteran dan kesehatan; komunikasi massa dan dokumentasi; psikologi; transportasi dan komunikasi; kependidikan dan keguruan didominasi oleh jenis pekerjaan tenaga profesional, teknisi, dan tenaga lain ybdi sebesar 85,1%.

Pada pohon terbaik dari hasil pemangkasan terdapat kesalahan pengelompokan (*misclassification*) yang dapat dilihat pada Tabel 5 berikut :

Tabel 5. Kesalahan pengelompokan

Kategori	N Kategori	N Kesalahan Pengelompokan	Persentase Error	Peluang
1	1290	554	42.95	0.43
2	110	49	44.55	0.45
3	665	665	100	1
4	1501	1108	73.82	0.74
5	318	318	100	1
6	840	166	19.76	0.2
7	1185	1185	100	1
8	159	34	21.38	0.21

Berdasarkan Tabel 5 dapat dilihat bahwa dari 1290 responden dengan kategori 1 terjadi kesalahan pengelompokan sebanyak 554 responden. Dari 110 responden dengan kategori 2 terjadi kesalahan pengelompokan sebanyak 49 responden. Dari 1501 responden dengan kategori 4 terjadi kesalahan pengelompokan sebanyak 1108 responden. Dari 840 responden dengan kategori 6 terjadi

kesalahan pengelompokan sebanyak 166 responden. Dari 159 responden dengan kategori 8 terjadi kesalahan pengelompokan sebanyak 34 responden. Pada kategori 3, 5 dan 7, semua responden terjadi kesalahan pengelompokan.

V. KESIMPULAN

Jenis pekerjaan di Sumatera Barat diduga dipengaruhi oleh faktor daerah tempat tinggal, jenis kelamin, bidang studi, pendidikan tertinggi yang ditamatkan, dan pengalaman kerja. Dari kelima faktor tersebut yang telah diidentifikasi menggunakan metode pohon klasifikasi pada CART (*Classification & Regression Trees*) diperoleh 3 faktor yang mempengaruhi jenis pekerjaan penduduk Sumatera Barat yaitu bidang studi, daerah tempat tinggal dan jenis kelamin. Selain itu, diperoleh pohon klasifikasi optimum dengan 5 simpul akhir yang berarti terdapat 5 kelompok masyarakat yang mempunyai pekerjaan utama, dimana 2 kelompok diantaranya didominasi oleh jenis pekerjaan tenaga profesional, teknisi dan tenaga lain yang berkaitan.

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2002). *Klasifikasi Baku Jenis Pekerjaan Indonesia (KBJI)*. Jakarta-Indonesia: Badan Pusat Statistik.
- Darsyah, M.Y. (2013). Menakar tingkat akurasi support vector machine study kasus kanker payudara. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 1(1).
- Sumartini, Siti Holis. (2015). Penggunaan Metode Classification and Regression Trees (CART) Untuk

- Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya. *Jurnal Sains dan Seni ITS*, 4(2).
- Inayah, Riza. (2014). Klasifikasi Rumah Tangga Miskin di Kabupaten Jombang Berdasarkan Faktor-faktor yang Mempengaruhi dengan Pendekatan CART (Classification and Regression Trees). *Jurnal Sains dan Seni ITS*, 3(2).
- Badan Pusat Statistik. (2004). *Statistik Kesejahteraan Rakyat (Welfare Statistics)*. Jakarta-Indonesia: Badan Pusat Statistik.
- Breiman L, Friedman J, Olshen R, dan Stone C. (1993). *Classification and Regression Trees*. New York–London : Chapman Hall.
- Lewis, R J. (2000). *An Introduction to Classification And Regression Trees (CART) Analysis*. Torrance-California : Departement of Emergency Medicine Harbor-UCLA Medical Center.
- Susanti, Reni. (2008). Pohon Klasifikasi untuk Menduga Faktor-faktor yang Mempengaruhi Keberhasilan Siswa di SPMB. *Tesis-S2*, tidak diterbitkan.