

BAB III

METODE POHON KLASIFIKASI *QUEST*

3.1 Metode Berstruktur Pohon

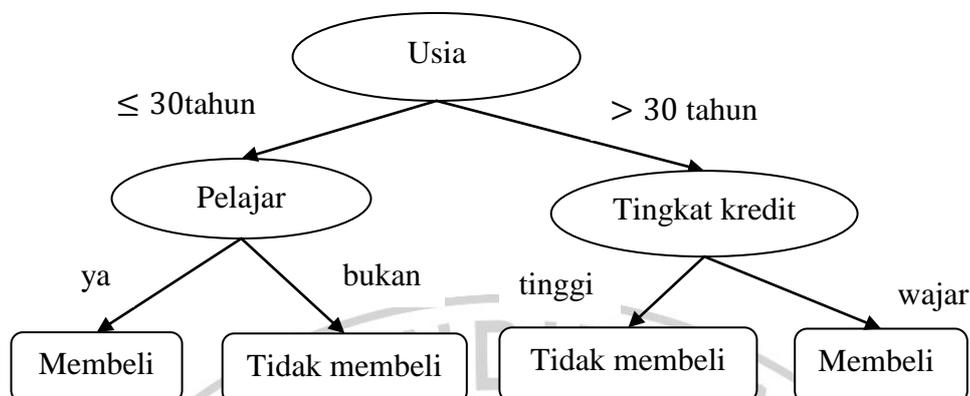
Istilah pohon dalam matematika dikenal dalam teori graf. Pertama kali konsep pohon digunakan oleh Gustav Kirchhoff (1824-1887) dalam bidang jaringan listrik. Berikutnya konsep pohon dikembangkan oleh Arthur Cayley (1821-1895). Pada tahun 1857 Cayley menggunakan konsep ini untuk menghitung banyaknya isomer-isomer yang berlainan dari C_nH_{2n+2} . Sebuah graf merupakan himpunan terhingga tak kosong yang memuat objek-objek yang disebut simpul, dan himpunan pasangan tak terurut antara simpul-simpul berlainan yang disebut sisi. Sebuah graf dikatakan pohon, jika graf tersebut merupakan graf terhubung dan tidak mengandung siklus (Kusumah, 1998:20).

Selanjutnya istilah pohon dikenal juga dalam teori keputusan, yakni pohon keputusan. Sebuah pohon keputusan merupakan sebuah penyekatan yang dinyatakan sebagai sebuah penyekatan berulang. Struktur hirarkis sebuah pohon terdiri dari simpul dan sisi. Beberapa bagian dalam sebuah pohon keputusan:

- Simpul akar, simpul yang tidak mempunyai sisi yang masuk tetapi memiliki sisi yang keluar.
- Simpul dalam, simpul yang memiliki satu sisi yang masuk dan dua atau lebih sisi yang keluar
- Simpul terminal atau simpul akhir, simpul yang mempunyai sisi yang masuk dan tidak ada sisi yang keluar. Simpul akhir dinyatakan sebagai label kelas.

Contoh :

Identifikasi pembeli komputer (dari pohon keputusan dibawah ini ternyata salah satu kelompok yang potensial adalah orang yang berusia < 30 dan pelajar)



Gambar 3.1 Contoh Pohon Keputusan

Metode berstruktur pohon merupakan salah satu teknik utama dalam data mining. Teknik ini memiliki banyak kesamaan dengan metode-metode yang lebih konvensional, seperti analisis regresi, analisis diskriminan, dan analisis kluster. Metode berstruktur pohon merupakan metodologi statistik dengan pendekatan non-parametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk variabel respon berupa nominal atau numerik.

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 diagnosa kategori penyakit berbahaya ataukah tidak), ilmu komputer (untuk menyelidiki struktur data), biologi (dalam hal klasifikasi makhluk hidup dengan ciri-ciri tertentu), psikologi (teori pengambilan keputusan), dan lain lain.

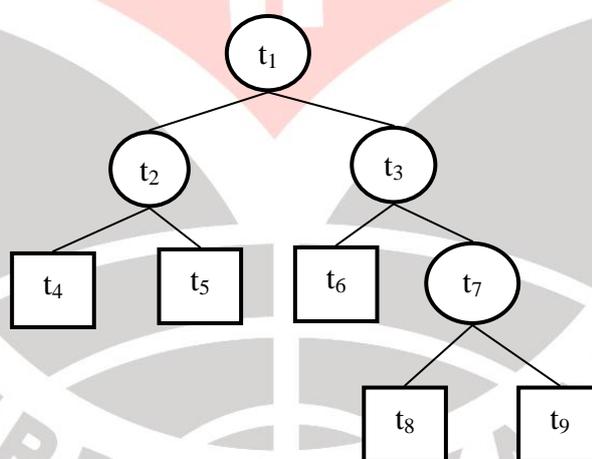
Kelinearan hubungan antara variabel respon dan prediktor seringkali menjadi kendala dalam penggunaan metode-metode konvensional (parametrik). Metode berstruktur pohon sering digunakan sebagai alternatif bila beberapa asumsi pada metode parametrik tidak dapat dipenuhi. Hasil analisis utama metode berupa grafik pohon yang memudahkan pengguna terutama yang bukan statistisi dalam hal interpretasi data.

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma QUEST (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah

Metode berstruktur pohon dapat dibagi kedalam dua kelompok. Kelompok pertama, metode yang menghasilkan pohon biner yaitu sebuah pohon yang setiap simpul disekat menjadi dua simpul yang saling terpisah. Kelompok kedua, metode yang menghasilkan pohon non-biner yaitu sebuah pohon yang setiap simpul disekat menjadi dua atau lebih simpul yang terpisah.

Kelompok pohon biner yaitu CART (*Classification and Regression Tree*) yang diperkenalkan oleh Breiman *et.al.* tahun 1984 dan QUEST (*Quick, Unbiased, Efficient Statistical Tree*) yang diperkenalkan oleh Loh dan Shih tahun 1997. Kelompok pohon non-biner, diantaranya CHAID (*Chi-Squared Automatic Interaction Detection*) yang diperkenalkan oleh Kass tahun 1980, FACT (*Factor Analysis Classification Tree*) yang diperkenalkan oleh Loh & Vanichestakul tahun 1988, dan CRUISE (*Classification Rule with Unbiased Iteration Selected and Estimation*) yang diperkenalkan oleh Kim & Loh tahun 2001. Berikut akan ditunjukkan ilustrasi visualisasi dua jenis pohon tersebut.

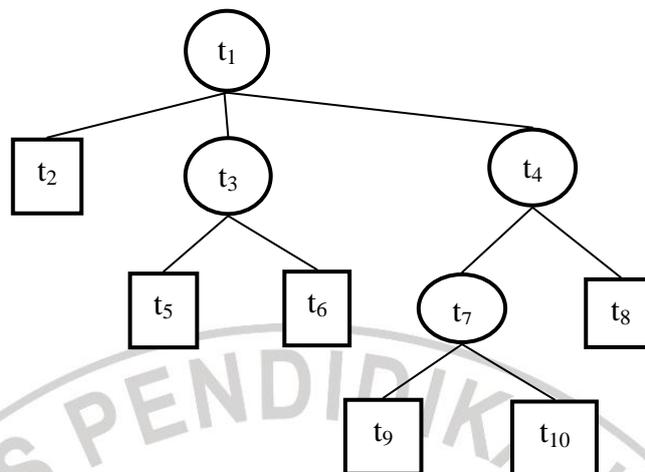


Gambar 3.2 Ilustrasi Pohon Biner

Pada Gambar 3.2 ditunjukkan struktur pohon biner yang memiliki satu simpul akar yang dinyatakan dengan t_1 yang mengandung semua gugus data. Simpul dalam adalah simpul yang bisa disekat menjadi simpul anak. Karena masih bisa disekat maka simpul ini dilambangkan dengan lingkaran yaitu t_2 , t_3 , t_7 . Sedangkan simpul akhir dilambangkan dengan kotak yaitu t_4 , t_5 , t_6 , t_8 , t_9 .

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma QUEST (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah



Gambar 3.3 Ilustrasi Pohon Non-biner

Pada Gambar 3.3 ditunjukkan struktur pohon yang memiliki satu simpul akar yang dinyatakan dengan t_1 yang mengandung semua gugus data. Dari simpul akar diperoleh hasil penyekatan berupa 3 simpul baru, yaitu t_2 , t_3 , t_4 . Karena terdapat simpul yang menghasilkan penyekatan lebih dari dua simpul, maka struktur pohon seperti ini dinamakan pohon non-biner. Seperti pada pohon biner, simpul dalam pada pohon ini dilambangkan oleh lingkaran yaitu t_3 , t_4 , t_7 dan simpul akhir dilambangkan dengan kotak yaitu t_2 , t_5 , t_6 , t_8 , t_9 , t_{10} .

Berdasarkan tipe peubah responnya, ada dua tipe metode berstruktur pohon, yaitu pohon klasifikasi jika peubah responnya kategorik, dan pohon regresi jika peubah responnya numerik. Salah satu keuntungan penggunaan metode berstruktur pohon adalah tampilan grafisnya sehingga lebih mudah untuk diinterpretasikan. Selain itu metode tersebut juga lebih fleksibel karena mampu memeriksa pengaruh variabel prediktor satu persatu (tidak memeriksa secara langsung keseluruhan variabel yang pada umumnya digunakan pada metode-metode konvensional) (Lewis, 2000).

Sartono (Berk, 2008) mengemukakan bahwa meskipun metode pohon telah banyak memberikan hasil memuaskan diberbagai bidang terapan, adapun kelemahan pohon klasifikasi dan regresi yaitu sifatnya yang tidak stabil. Jika diambil sampel berbeda dari populasi yang sama, kemungkinan diperoleh pohon dengan bentuk yang berbeda.

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma QUEST (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah

3.2 Pohon Klasifikasi

Sebuah pengklasifikasian berstruktur pohon merupakan sebuah pohon keputusan yang digunakan untuk memprediksi sebuah kelas variabel dari satu atau lebih variabel. Pohon klasifikasi (*classification tree*) merupakan metode statistika yang digunakan untuk memperkirakan keanggotaan objek dalam kelas-kelas variabel prediktor kategorik. Keanggotaan ini ditaksir dari pengukuran objek pada satu variabel prediktor atau lebih. Pohon ini dibentuk melalui penyekatan secara berulang (rekursif), dimana kelas dan nilai-nilai variabel prediktor setiap objek telah diketahui. Setiap penyekatan pada pohon yang dibentuk dinyatakan sebagai node atau simpul.

Pada metode ini objek dinyatakan sebagai vektor \mathbf{x} yang merupakan ukuran untuk karakteristik tertentu. Sebagaimana dikemukakan oleh Breiman (Susanti, 2007: 25) bahwa misalkan terdapat k buah karakteristik, maka vektor \mathbf{x} dapat dituliskan sebagai $\mathbf{x} = \{x_1, x_2, \dots, x_k\}$. Definisikan \mathfrak{X} sebagai ruang pengamatan yang memuat semua vektor yang mungkin, atau dengan kata lain \mathfrak{X} merupakan himpunan dari vektor-vektor yang ada. Misalkan terdapat N buah vektor, maka \mathfrak{X} dapat ditulis sebagai $\mathfrak{X} = \{x_1, x_2, \dots, x_N\}$. Suatu objek dikelompokkan kedalam kelas/kelompok yang diberi nomor $1, 2, \dots, j$ dan misalkan c adalah himpunan dari kelas-kelas tersebut sehingga dapat ditulis $c = \{1, 2, \dots, j\}$.

Menurut Breiman (Susanti, 2007: 26), suatu pengklasifikasian merupakan fungsi $d(\mathbf{x})$ yang didefinisikan pada \mathfrak{X} sedemikian rupa sehingga untuk setiap \mathbf{x} , $d(\mathbf{x}) = j$ untuk suatu $j \in c$. Pengklasifikasian tidak dibentuk secara acak, melainkan berdasarkan pada pengalaman yang telah lalu/data sebelumnya. Dalam pembentukan klasifikasi yang sistematis, data sebelumnya dinamakan dengan *learning sample*.

3.3 Metode Pohon Klasifikasi QUEST

QUEST merupakan kepanjangan dari *Quick, Unbiased and Efficient Statistical Tree*. *QUEST* merupakan sebuah algoritma pohon keputusan biner untuk analisis klasifikasi dan data mining yang dikembangkan oleh Wei-Yin Loh

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma QUEST (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah

(*University of Wisconsin-Madison*) dan Yu-Shan Shih (*National Chung Cheng University, Taiwan*) pada tahun 1997 dari metode pohon klasifikasi non-biner FACT yang memiliki kecepatan tinggi (Loh & Vanichsetakul, 2000). Algoritma ini memiliki kecepatan dalam hal komputasi (*Quick*), menghasilkan pemilihan variabel bebas yang tak bias (*Unbiased*), dan efisien (*Efficient*) untuk data kompleks, yakni variabel prediktor terdiri dari kategorik dan numerik.

Tujuan dari *QUEST* mirip dengan algoritma CART yang dijelaskan dalam buku *Classification and Regression Trees* oleh Breiman, Friedman, Olshen dan Stone pada tahun 1984. Perbedaan yang utamanya adalah

- *QUEST* menggunakan teknik pemilihan variabel berdasarkan standar yang ditetapkan.
- *QUEST* menggunakan teknik penyekat pengganti untuk mengatasi nilai pengamatan yang hilang.
- *QUEST* dapat dengan mudah mengatasi variabel prediktor kategori dengan banyak kategori dengan mentransformasi.

QUEST merupakan modifikasi analisis diskriminan kuadratik rekursif sebagai alternatif bagi metode-metode berstruktur pohon lain yang menggunakan pendekatan *exhaustive search*. Dalam metode *QUEST*, dapat ditetapkan taraf kepercayaan untuk simpul penyekat. Sebuah variabel bebas tidak dapat digunakan sebagai simpul penyekat, bila taraf signifikansinya kurang dari atau sama dengan nilai yang ditetapkan yakni minimal 0 dan maksimal 1. Umumnya taraf signifikansi yang digunakan adalah 0,05.

Algoritma FACT melakukan pemilihan variabel penyekat dan simpul penyekat dilakukan secara terpisah, ini juga yang dilakukan pada algoritma *QUEST*. Pada setiap variabel numerik hitung statistik-F pada uji ANOVA dan untuk setiap variabel kategori hitung χ^2 pada uji kebebasan dua variabel. Variabel yang memiliki kelompok dengan tingkat kehomogenan yang paling besar dipilih sebagai variabel penyekat. Analisis diskriminan kuadratik diterapkan pada proses pemilihan simpul penyekat. Jika variabel penyekat yang terpilih berupa variabel

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma QUEST (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah

kategorik, maka dilakukan transformasi ke peubah numerik dan selanjutnya diterapkan analisis diskriminan kuadrat.

3.4 Algoritma *QUEST*

Menurut Loh dan Shih (1997), algoritma *QUEST* dibagi menjadi tiga bagian yakni, algoritma pemilihan variabel prediktor penyekat, algoritma penentuan simpul penyekat, dan algoritma penghentian pembentukan pohon.

1. Algoritma Pemilihan Variabel Penyekat

- a. Untuk setiap variabel prediktor X numerik, lakukan uji ANOVA F yang menguji semua kelas berbeda dari variabel terikat Y memiliki rata-rata yang sama dari X , dan hitung nilai p -value berdasarkan statistik uji F. Untuk setiap variabel prediktor X kategori, lakukan uji chi-kuadrat dari Y dan X yang bebas dan hitung p -value berdasarkan statistik uji chi-kuadrat.
- b. Cari prediktor dengan p -value terkecil dan notasikan dengan X^* .
- c. Bandingkan p -value tersebut dengan taraf α/M_1 , dimana $\alpha \in (0,1)$ merupakan taraf signifikansi dan M_1 merupakan banyaknya variabel prediktor .
 - Jika p -value kurang dari α/M_1 , maka variabel yang bersesuaian dengan X^* dipilih sebagai variabel penyekat. Kemudian lanjutkan ke langkah (e)
 - Jika p -value lebih dari α/M_1 , maka lanjutkan ke langkah (d)
- d. Untuk setiap variabel prediktor X yang numerik,
 - Hitung statistik Lavene F untuk menguji homogenitas varians dari X untuk kelas yang berbeda dari Y , dan hitung p -value dari pengujian tersebut.
 - Cari prediktor yang memiliki p -value terkecil dan notasikan dengan X^{**}
 - Bandingkan p -value dari uji Lavene tersebut dengan taraf $\frac{\alpha}{M_1+M_2}$, dimana M_2 merupakan banyaknya variabel prediktor numerik.

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma *QUEST* (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah

- Jika X^{**} kurang dari $\frac{\alpha}{M_1+M_2}$, maka variabel yang bersesuaian dengan X^{**} dipilih sebagai variabel penyekat.
 - Jika X^{**} lebih dari $\frac{\alpha}{M_1+M_2}$, maka variabel tersebut tidak dipilih menjadi variabel penyekat.
- e. Misalkan X^* adalah variabel penyekat yang diperoleh pada langkah (c) atau (d), lanjutkan ke algoritma penentuan simpul penyekat.

2. Algoritma Penentuan Simpul Penyekat

Setelah terpilih variabel penyekat, langkah selanjutnya adalah menentukan simpul penyekat. Algoritma penentuan simpul penyekat dijelaskan sebagai berikut:

a. Prediktor Penyekat Kategorik

Jika variabel prediktor X yang terpilih adalah nominal dan biner, maka simpul penyekat jelas. Namun jika variabel prediktor X yang terpilih adalah nominal dengan lebih dari dua kategori, *QUEST* pertama-tama mentransformasi ini kedalam variabel numerik (sebut saja ξ) dengan menetapkan koordinat diskriminan yang paling besar untuk mengategorikan prediktor. *QUEST* kemudian menggunakan algoritma pemilihan simpul penyekat untuk prediktor numerik pada ξ untuk menentukan simpul penyekat.

Langkah pemetaan kategori nominal kedalam nilai koordinat diskriminan (dengan kata lain, variabel nominal ditransformasi menjadi variabel numerik) dilakukan untuk mendefinisikan jarak dan urutan nilai-nilai yang telah ditransformasi. Hal ini ditujukan untuk memaksimalkan perbedaan antara kategori variabel respon (Loh & Shih, 1997). *QUEST* menggunakan analisis diskriminan kuadratik yang telah dimodifikasi untuk menangani varians yang tidak homogen antar kategori peubah respon.

Langkah Transformasi Variabel Kategorik Menjadi Variabel Numerik

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma *QUEST* (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah

Misalkan X merupakan variabel prediktor kategori dengan kategori b_1, b_2, \dots, b_I . X akan ditransformasi menjadi variabel numerik ξ untuk setiap kelas X dengan langkah-langkah sebagai berikut :

1. Transformasi setiap nilai x dari X kedalam sebuah vektor dummy I -dimensi

$$\mathbf{v} = (v_1, v_2, \dots, v_I), \text{ dimana } v_i = \begin{cases} 1 & x = b_i \\ 0 & x \neq b_i \end{cases}, i = 1, 2, \dots, I$$

2. Hitung rata-rata keseluruhan kelas j dari \mathbf{v}

$$\bar{\mathbf{v}} = \frac{\sum_{i=1}^I f_i v_i}{N_t} \quad (3.1)$$

$$\bar{\mathbf{v}}^{(j)} = \frac{\sum_{i=1}^I n_i v_i}{N_{j,t}} \quad (3.2)$$

dengan

v_i : pengamatan ke- i

$\bar{\mathbf{v}}$: rata-rata untuk semua pengamatan pada simpul t

$\bar{\mathbf{v}}^{(j)}$: rata-rata untuk semua pengamatan pada simpul t kelompok respon j

f_i : jumlah pengamatan pada simpul t untuk v_i

n_i : jumlah pengamatan pada simpul t kelompok respon j untuk v_i

N_t : jumlah pengamatan pada simpul t

$N_{j,t}$: jumlah pengamatan pada simpul t kelompok respon j

3. Hitung matriks $i \times i$ berikut

$$\mathbf{B} = \sum_{j=1}^J N_{j,t} (\bar{\mathbf{v}}^{(j)} - \bar{\mathbf{v}})(\bar{\mathbf{v}}^{(j)} - \bar{\mathbf{v}})' \quad (3.3)$$

$$\mathbf{T} = \sum_{i=1}^I f_i (\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})' \quad (3.4)$$

4. Lakukan *singular value decomposition* pada \mathbf{T} untuk memperoleh $\mathbf{T} = \mathbf{Q}\mathbf{D}\mathbf{Q}'$, dimana \mathbf{Q} adalah sebuah matriks ortogonal, $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_l)$ sehingga $d_1 \geq d_2 \geq \dots \geq d_l \geq 0$. Misalkan $\mathbf{D}^{-1/2} = \text{diag}(d_1^*, d_2^*, \dots, d_l^*)$

$$\text{dimana } d_i^* = \begin{cases} d_i^{-1/2} & \text{jika } d_i > 0 \\ 0 & \text{lainnya} \end{cases}$$

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma QUEST (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah

5. Lakukan *singular value decomposition* pada $D^{-1/2}Q'BQD^{-1/2}$ untuk memperoleh vektor eigen \mathbf{a} yang berhubungan dengan nilai eigen yang terbesar.
6. Koordinat diskriminan terbesar dari \mathbf{v} merupakan proyeksi

$$\xi = \mathbf{a}' D^{-1/2} Q' \mathbf{v} \quad (3.5)$$

b. Prediktor Penyekat Numerik

Variabel respon Y terdiri dari dua kategori, yaitu A dan B , dan variabel prediktor X terpilih berupa numerik. Langkah penentuan prediktor penyekat sebagai berikut :

1. Misalkan \bar{x}_A dan s_A^2 merupakan rata-rata dan varians X^* dari pengamatan variabel respon kategori pertama, sedangkan \bar{x}_B dan s_B^2 merupakan rata-rata dan varians X^* dari pengamatan variabel respon kategori kedua. Misalkan $P(j|t) = \frac{N_{j,t}}{N_j}$ merupakan peluang dari masing-masing kategori pada variabel respon, dengan N_j adalah jumlah pengamatan pada simpul awal kelompok respon j .
2. Jika $\min(s_A^2, s_B^2) = 0$, maka urutkan kedua kelas dengan urutan varians yang meningkat kemudian notasikan varians tersebut dengan $s_1^2 \leq s_2^2$, dan bersesuaian dengan rata-rata nya \bar{x}_1, \bar{x}_2 . Misalkan ε merupakan bilangan positif yang sangat kecil nilainya, sebut saja $\varepsilon = 10^{-12}$. Jika $\bar{x}_1 < \bar{x}_2$, maka $d = \bar{x}_1(1 + \varepsilon)$, selain itu $d = \bar{x}_1(1 - \varepsilon)$.
3. Jika $\min(s_A^2, s_B^2) \neq 0$, maka analisis diskriminan kuadratik diterapkan untuk menghitung simpul penyekat d .

Analisis diskriminan kuadratik tradisional menaksir fungsi kepadatan dari kelas dengan fungsi kepadatan normal dengan rata-rata dan varians ditaksir dari sampel. Misalkan $\bar{x}^{(j)}$ dan s^2 merupakan rata-rata dan varians sampel kelas untuk kelas ke- j ($j = 1, 2$). Misalkan $\phi(x) = (2\pi)^{-1/2} \exp(-\frac{x^2}{2})$ merupakan fungsi kepadatan normal baku. Analisis diskriminan kuadratik menyekat sumbu X menjadi tiga interval yaitu $(-\infty, d_1)$, (d_1, d_2) dan (d_2, ∞) , dimana d merupakan akar-akar dari persamaan

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma QUEST (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah

$$P(A|t)s_A^{-1}\phi\left(\frac{x-\bar{x}_A}{s_A}\right) = P(B|t)s_B^{-1}\phi\left(\frac{x-\bar{x}_B}{s_B}\right) \quad (3.6)$$

Kedua ruas diberi log untuk memperoleh persamaan kuadrat $ax^2 + bx + c = 0$, dimana

$$a = s_A^2 - s_B^2 \quad (3.7)$$

$$b = 2(\bar{x}_A s_B^2 - \bar{x}_B s_A^2) \quad (3.8)$$

$$c = (\bar{x}_B s_A)^2 - (\bar{x}_A s_B)^2 + 2s_A^2 s_B^2 \log\left(\frac{P(A|t)s_B}{P(B|t)s_A}\right) \quad (3.9)$$

Jika $a = 0$ dan $\bar{x}_A \neq \bar{x}_B$, hanya terdapat satu akar

$$x = \frac{(\bar{x}_A + \bar{x}_B)}{2} - \frac{s_A^2}{(\bar{x}_A + \bar{x}_B)} \log\left(\frac{P(A|t)}{P(B|t)}\right) \quad (3.10)$$

Persamaan tidak memiliki akar jika $a = 0$ dan $\bar{x}_A = \bar{x}_B$.

4. Sebuah simpul dipisah pada $X^* = d$, dimana d didefinisikan sebagai berikut :

a. Jika $a = 0$, maka

$$d = \begin{cases} x = \frac{(\bar{x}_A + \bar{x}_B)}{2} - \frac{s_A^2}{(\bar{x}_A + \bar{x}_B)} \log\left(\frac{P(A|t)}{P(B|t)}\right), & \bar{x}_A \neq \bar{x}_B \\ \bar{x}_A, & \bar{x}_A = \bar{x}_B \end{cases} \quad (3.11)$$

b. Untuk $a \neq 0$

Jika $b^2 - 4ac < 0$, maka

$$d = \frac{(\bar{x}_A + \bar{x}_B)}{2} \quad (3.12)$$

Jika $b^2 - 4ac \geq 0$, maka definisikan

$$d = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (3.13)$$

yang lebih mendekati \bar{x}_A , dengan syarat menghasilkan dua simpul yang tak kosong. *QUEST* hanya menggunakan satu dari kedua akar persamaan tersebut, yaitu akar yang nilainya paling mendekati rata-rata sampel dari tiap kelas.

3. Algoritma Penghentian Pembentukan Pohon

Riski Sulistiawati Handayani, 2013

Penerapan Algoritma *QUEST* (Quick, Unbiased, Efficient, Statistical Tree) Pada Pembentukan Pohon Klasifikasi Minat Masyarakat Kabupaten Purwakarta Terhadap Bank Syariah

Proses penyekatan diulang terhadap simpul anak. Proses rekursif dilakukan sampai simpul tidak bisa disekat lagi dengan aturan penghentian proses pembentukan pohon ditentukan sebagai berikut :

1. Jika sebuah simpul menjadi murni, yaitu semua objek/kasus masuk kedalam kelas variabel terikat yang sama pada simpul tersebut, maka simpul tidak akan disekat.
2. Jika semua objek/kasus dalam sebuah simpul memiliki nilai-nilai identik untuk tiap variabel prediktor, maka simpul tidak akan disekat.
3. Jika kedalaman pohon pada saat tersebut mencapai nilai batas kedalaman pohon maksimum yang ditetapkan, maka proses pertumbuhan pohon akan berhenti.
4. Jika penyekat dari simpul menghasilkan simpul anak yang ukuran simpulnya kurang dari nilai ukuran simpul anak minimum yang ditetapkan, simpul tidak akan disekat.