© HAK CIPTA MILIK UNIVERSITAS ANDALAS

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
   b. Pengutipan tidak merugikan kepentingan yang wajar Unand.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Unand.

PEMBENTUKAN POHON KLASIFIKASI BINER DENGAN ALOGARITMA QUEST

SKRIPSI

YONA MALANI
07134032

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ANDALAS
PADANG
2011

TANDA PERSETUJUAN SKRIPSI

Dengan ini menyatakan bahwa :

Nama	:	Yona Malani
No. Buku Pokok	:	07 134 032
Jurusan	:	Matematika
Bidang	:	Statistika
Judul Skripsi	:	**Pembentukan Pohon Klasifikasi Biner dengan Algoritma QUEST**

telah diuji dan disetujui skripsinya sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains (S.Si) melalui ujian sarjana yang diadakan pada tanggal 27 Oktober 2011 berdasarkan ketentuan yang berlaku.

Pembimbing / Penguji

1.

Izzati Rahmi HG, M.Si
NIP. 19740928 199903 2 002

2.

Hazmira Yozza, M.Si.
NIP. 19690308 199403 2 002

Penguji

1.

Dr. Dodi Devianto
NIP. 19771227 200012 1 002

2.

Dr. Maiyastri
NIP. 19650531 199103 2 001

3.

Zulakmal
NIP. 19671108 199802 1 001

Mengetahui,
Ketua Jurusan Matematika FMIPA Unand

Dr. Syafrizal Sy
NIP. 19670807 199309 1 001

بسم الله الرحمن الرحيم

Bacalah dengan menyebut nama Tuhanmu Yang menciptakan, Dia telah menciptakan manusia dari segumpal darah, Bacalah dan Tuhanmu yang paling Pemurah, yang mengajar Manusia dengan perantaraan kalam, Dia mengajarkan kepada Manusia apa yang tidak diketahuinya. (al-alaq:1-5)

Sesungguhnya sesudah kesulitan itu ada kemudahan, maka apabila kamu telah selesai dari suatu urusan, Kerjakanlah dengan sungguh-sungguh urusan yang lain, dan hanya kepada Tuhanmulah hendaknya kamu berharap (alam nasyrah:6-8)

Great Thankfullness to Allah SWT

"Ya Allah saat ini hamba bersujud syukur atas nikmat dan karuniaMu..Berkat ridhoMu dan kemudahanMu semua rintangan dapat kulalui dengan kesabaran dan keikhlasan sehingga hamba dapat menyelesaikan skripsi ini. Dan tak henti-hentinya hamba memohon dan meminta ridho n petunjukMu agar hamba bisa selalu berada dalam bimbinganMu dan selalu menjadi hamba yg selalu bersyukur dan dekat denganMu baik didunia maupun diakhirat." Amiiinnnn.

Dengan kerendahan hati q persembahkan karya kecil q ini buat orang tua q tercinta Papa (Nasri) dan Mama (Zulma Herni), "ma, pa, terima kasih na ucapkan buat semua doa dan kasih sayang serta bimbingan yang ma2 n pa2 berikan kepada na, semoga na bisa mjd anak yang selalu bisa membahagiakan & membanggakan ma2 n pa2." My family kakak2q K' Eka, B' sil, K' Nova n K' Vera, mkc y k' buat dorongan semangat, pengertian n perhatiannya, de2kq tersayang (Roza), thx y d buat pengertian n perhatiannya, rajin2 kul y d moga cpat tamat, Seemoga qt semua mjd titipan Allah yang sll bisa membahagiakan n dibanggakan oleh ma2 n pa2. Tuk Pak Etek Defri n Oma Epi mksh buat bantuannya selama ini.

Buat pembimbingq, orgtuaq Ibu Izzati Rahmi HG, M.Si dan Ibu Hazmira Yozza, M.Si. mksh banyak bu buat bimbingan, waktu, kesabaran n limu yang bu berikan... maaf na banyak nyusahin bu...semoga na bisa menggunakan ilmu yang bu berikan buat kbaikan n bisa mjd anak yang bisa dibanggakan..Amiinn..

Orang2 terdekatq pi2, egi, yona, n ferdi thx y buat bantuan n doanya selama ni....buat pi2 n egi smoga sukses sll, buat yona n ferdi semangat y say "dgn doa n usaha serta semangat yang tinggi semua pasti bisa!!". Buat udaq tacinto alias dj alias da jon, thx bgt buat bantuanny y dj, semoga cepat menyusul dan sll semangat n optimis y & sll tersenyum.." utk kakandaq bg zoellvendi thx buat nasehat, perhatian, dorongan n bantuan utk d slama ni, smoga kanda sll bahagia n dberi yang terbaik...tak lupa buat *someone special,* yang g tau syapa n ntah dmn skrg, moga Qt dpertemukan pd saat yang tepat n kondsi yang mendukung qt...☺

Buat sobatq ana(thx y ana buat istilah kedokterannya), tika (thx y buat nasehat n konsultasinya) n wi2 (akhirnya qt bsa juga say)..wlw jarak mmisahkan qt, tpi hati sll dekat....hehehehe..

Buat uda2 n adek2q yang gokil da jack, eris, rika, dina n meta thx buat bantuan n kelucuannya.. semoga cepat menyusul n semangat!!!!

Buat anak kos-an pi2, mksh bnyk buat bantuan n pengertiannya...buat na2, mksh y d buat bantuannya, maaf y kak banyak repotin d...rajin2 kul y, moga cepat wisuda.. hehhehhe..

Buat tman n uni seperjuangan tia, novi, ni ri2, ni riza, ni mega, ni pu2t mkh y buat info n bantuannya, semoga qt dberi kelancaran utk menggapai ksuksesn...

Finally I say thank's to :

Para Pengujiq Pak Dr. Dodi devianto, Ibu Dr. Maiyastri dan Pak Zulakmal, M.Si mksh pak n bu buat saran n kritikannya, buat waktu n bantuannya...

Dosen-dosen di Jurusan Matematika yang telah berbagi ilmu dengan q, semoga q bisa menggunakannya dengan baik...

Buat ma2 cun, kak opi, Pak syamsir n Buk eli mksh buat bantuannya dalam proses seminar n sidang.....

Buat tman2 angkatan 2007 "MC Zoven" thx buat semua kebersamaan, kelucuan, kbahagiaan selama ni.. semoga qt bsa menggapai kesuksesan n sll bsa mjga komunikasi..wlw jarak memisahkan qt, tpi hati tetaplah dekat..^_^

Buat semua elemen HIMATIKA, mksh y buat keceriaan n warna kehidupan yang diberikan, smoga qt bsa mjga kekompakan n sll sejalan n SEMANGAT HIMATIKA.....

Dan kepada seluruh pihak yang telah membantu dalam pengerjaan skripsi ini, semoga ini dapat menjadi amalan bagi qt semua..Amiiiinn.

Ya Allah, janganlah Engkau jadikan hatiq condong kepada kesesatan, sesudah Engkau menunjukiq dan iringilah selalu perjalanan q dengan ridho n karuniaMu... Aminn.

Y O N A

KATA PENGANTAR

Alhamdulillah, puji syukur penulis sampaikan kehadirat Allah SWT karena berkat ridho dan izin-Nya penulis dapat menyelesaikan skripsi dengan judul **"Pembentukan Pohon Klasifikasi Biner dengan Algoritma QUEST"**. Salawat dan salam tidak lupa penulis kirimkan kepada junjungan kita Nabi Muhammad SAW yang telah membawa umat manusia dari zaman kebodohan ke zaman yang penuh ilmu pengetahuan. Skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana Sains di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Andalas Padang.

Selanjutnya, penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam penyelesaian skripsi ini, terutama kepada :

1. Ibu Izzati Rahmi HG, M.Si selaku Pembimbing I yang telah dengan sabar membantu mengarahkan penulis dalam menyelesaikan penulisan skripsi ini. Serta ilmu, ide, saran, dan nasihat yang diberikan selama proses bimbingan tugas akhir maupun selama penulis menjalani perkuliahan.
2. Ibu Hazmira Yozza, M.Si selaku Pembimbing II yang telah membantu penulis dalam penyempurnaan penulisan skripsi ini. Serta ilmu yang didapat selama penulis menjalani perkuliahan.
3. Bapak Dr. Dodi Devianto, Ibu Dr. Maiyastri, dan Bapak Zulakmal, M.Si. selaku penguji yang telah membaca, memberi masukan dan saran kepada penulis dalam penyempurnaan penulisan skripsi ini.

4. Ibu selaku Pembimbing Akademik yang telah membantu penulis dalam urusan akademik terutama dalam merancang studi agar dapat selesai tepat pada waktunya. Serta nasihat dan ilmu yang telah diberikan selama penulis menjalani proses studi.
5. Bapak Dr. Syafrizal Sy selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Andalas Padang.
6. Bapak/Ibu dosen dan karyawan/i Jurusan Matematika FMIPA Unand.

Penulis selalu terbuka terhadap sumbangan pemikiran baik kritik maupun saran, karena tidak ada sesuatu yang sempurna dalam hidup ini, sehingga proses belajar dari kehidupan itu sendiri menjadi sangat penting dalam rangka pendewasaan diri baik pikiran maupun sikap dalam menyikapi kehidupan. Besar harapan, semoga skripsi ini bermanfaat untuk sivitas akademik Universitas Andalas khususnya dan masyarakat umumnya. Amin.

Padang, Nopember 2011

Penulis

ABSTRAK

Metode klasifikasi berstruktur pohon merupakan salah satu metode klasifikasi nonparametrik yang sering digunakan untuk pengklasifikasian objek. Metode klasifikasi berstruktur pohon dapat dilakukan dengan algoritma QUEST yang menghasilkan pohon klasifikasi biner. Proses pembentukan pohon biner dengan algoritma QUEST terbagi menjadi dua tahap, yaitu penentuan peubah penyekat dan pemilihan titik penyekat. Penentuan peubah penyekat menggunakan uji Khi-Kuadrat untuk kebebasan pada peubah kategorik dan uji *F* pada peubah numerik. Pemilihan titik penyekat menerapkan analisis diskriminan kuadratik. Penerapan algoritma QUEST yang dilakukan untuk mengetahui faktor-faktor yang mempengaruhi seseorang menderita penyakit jantung menghasilkan empat kelompok, yaitu (1) kelompok dengan kondisi jantung normal, (2) kelompok dengan kondisi jantung cacat dan memiliki jumlah pembuluh darah besar yang diwarnai oleh flourosopy 1 sampai 3, (3) kelompok dengan kondisi jantung cacat, tidak memiliki jumlah pembuluh darah besar yang diwarnai oleh flourosopy dan apabila latihan dapat menginduksi angina, dan (4) kelompok dengan kondisi jantung cacat, tidak memiliki jumlah pembuluh darah besar yang diwarnai oleh flourosopy dan apabila latihan tidak dapat menginduksi angina. Kelompok (1) dan (4) termasuk pasien yang cenderung tidak terserang penyakit jantung, sedangkan kelompok (2) dan (3) termasuk pasien yang cenderung terserang penyakit jantung.

Kata kunci : *pohon klasifikasi, QUEST, biner, uji Khi-Kuadrat, uji F, analisis diskriminan kuadratik*

DAFTAR ISI

KATA PENGANTAR..... v

ABSTRAK..... vii

DAFTAR ISI..... viii

DAFTAR TABEL..... x

DAFTAR GAMBAR..... xi

DAFTAR LAMPIRAN..... xii

BAB I PENDAHULUAN

1.1 Latar Belakang..... 1

1.2 Perumusan Masalah..... 2

1.3 Pembatasan Masalah..... 3

1.4 Tujuan Penelitian..... 3

1.5 Sistematika Penulisan..... 3

BAB II LANDASAN TEORI

2.1 Skala Pengukuran Data..... 4

2.2 Pengujian Hipotesis..... 6

2.3 Uji Khi-Kuadrat untuk Kebebasan..... 7

2.4 Uji Anova F..... 10

2.5 Uji Levene F..... 11

2.6 Analisis Diskriminan Kuadratik..... 12

2.7 Metode Klasifikasi Berstruktur Pohon..... 14

BAB III METODE QUEST

3.1 Metode QUEST..... 18

3.2 Algoritma QUEST..... 18

3.2.1 Algoritma Pemilihan Peubah Penyekat..... 19

3.2.2 Algoritma Penentuan Titik Penyekat..... 20

3.2.3 Algoritma Transformasi Peubah Kategorik menjadi Peubah Numerik..... 21

3.3 Contoh Penerapan Metode QUEST..... 23

3.3.1 Kasus..... 23

3.3.2 Data..... 23

3.3.3 Penyelesaian dengan Metode QUEST..... 25

BAB IV PENUTUP

4.1 Kesimpulan..... 48

4.2 Saran..... 49

DAFTAR PUSTAKA..... 50

LAMPIRAN..... 52

DAFTAR TABEL

No.		Halaman
2.3.1	Tabel Kontingensi Dua Arah	8
2.4.1	Tabel Struktur Data ANOVA F	10
2.4.2	Tabel ANOVA F	10
2.6.1	Tabel Struktur Data Analisis Diskriminan Kuadratik	13
3.3.1	Tabel ANOVA F	26
3.3.2	Tabel Frekuensi Teramati (O_{ij})	26
3.3.3	Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_0	27
3.3.4	Tabel Tabulasi X_{12}	29
3.3.5	Tabel Tabulasi X_{12} Hasil Transformasi	32
3.3.6	Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_1	33
3.3.7	Tabel Tabulasi X_{11}	34
3.3.8	Tabel Tabulasi X_{11} Hasil Transformasi	38
3.3.9	Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_2	40
3.3.10	Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_3	42
3.3.11	Tabel Tabulasi X_8 untuk Simpul t_3	43
3.3.12	Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_4	44
3.3.13	Hasil Uji Statistika Levene F	44

DAFTAR GAMBAR

No.		Halaman
2.6.1	Pohon Klasifikasi.....	15
3.3.1	Pohon Klasifikasi Data Penyakit Jantung.....	45

DAFTAR LAMPIRAN

No. Halaman

1. Hasil SPSS 16.0 52

2. Flowchart untuk Algoritma Pemilihan Peubah Penyekat..... 53

3. Flowchart untuk Algoritma Penentuan Titik Penyekat..... 54

BAB I

PENDAHULUAN

1.1 Latar Belakang

Masalah klasifikasi (pengelompokan) sering dijumpai pada kehidupan sehari-hari, baik dibidang pendidikan, sosial, industri, kesehatan maupun perbankan. Contoh masalah klasifikasi dalam bidang pendidikan adalah klasifikasi sekolah berdasarkan akreditasi sekolah. Dalam bidang kesehatan dilakukan pengklasifikasian penyakit berdasarkan tingkat keseriusan dan bahaya yang ditimbulkan.

Dalam statistika, dikenal berbagai metode untuk mengklasifikasikan objek. Klasifikasi dapat dilakukan dengan menggunakan pendekatan parametrik maupun pendekatan nonparametrik. Pendekatan parametrik biasanya sangat tergantung dari asumsi-asumsi mengenai sebaran data sehingga bila asumsi tersebut tidak terpenuhi, maka hasil yang diperoleh menjadi tidak valid. Pendekatan nonparametrik digunakan untuk mengatasi keterbatasan dari pendekatan parametrik. Pendekatan ini tidak bergantung pada asumsi tertentu sehingga memberikan kemudahan dalam menganalisa data tetapi tetap mempunyai tingkat akurasi yang tinggi. Untuk mendapatkan klasifikasi yang tepat, perlu diperhatikan metode klasifikasi yang tepat.

Metode klasifikasi berstruktur pohon adalah salah satu metode klasifikasi nonparametrik yang sering digunakan. Metode ini pertama kali diperkenalkan oleh Breiman, *et. al* pada tahun 1984. Klasifikasi pada metode ini dilakukan dengan membangun sebuah pohon klasifikasi yang diperoleh melalui penyekatan

berulang terhadap sebuah himpunan data (dalam hal ini diistilahkan sebagai simpul) menjadi dua atau lebih simpul baru.

Metode klasifikasi berstruktur pohon dapat dibagi menjadi dua kelompok. Kelompok pertama menghasilkan pohon biner, yaitu sebuah pohon yang setiap simpul disekat menjadi dua simpul yang saling terpisah. CART (*Classification and Regression Trees*) merupakan algoritma yang diperkenalkan oleh Breiman, *et.al* (1984), sedangkan QUEST (*Quick, Unbiased, Efficient Statistical Trees*) merupakan algoritma yang diperkenalkan oleh Loh dan Shih (1997). CART dan QUEST merupakan contoh metode yang menghasilkan pohon seperti ini. Kelompok kedua adalah kelompok yang menghasilkan pohon non-biner, yaitu sebuah pohon yang setiap simpul disekat menjadi dua atau lebih simpul yang terpisah. CHAID (*Chi-squared Automatic Interaction Detector*) merupakan salah satu algoritma yang diperkenalkan oleh Kass (1980), sedangkan CRUISE (*Classification Rule with Unbiased Interaction Selection and Estimation*) merupakan algoritma yang diperkenalkan oleh Kim dan Loh (2001). CHAID dan CRUISE merupakan contoh metode klasifikasi yang menghasilkan pohon non-biner.

Pada penelitian ini, akan dibahas metode klasifikasi berstruktur biner menggunakan metode QUEST. Hal tersebut dilatarbelakangi karena metode QUEST ini masih relatif baru dan jarang dibahas. Salah satu hal yang menarik untuk diketahui adalah bagaimana membentuk pohon klasifikasi dengan metode QUEST dan bagaimana menerapkan metode QUEST pada data.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, maka yang menjadi permasalahan dalam penelitian adalah bagaimana cara pembentukan pohon klasifikasi biner dengan algoritma QUEST dan bagaimana menerapkan metode QUEST pada data.

1.3 Batasan Masalah

Penelitian ini hanya akan membahas pembentukan pohon klasifikasi dengan metode QUEST, tanpa dilanjutkan dengan proses pemangkasan yang biasanya dilakukan untuk memilih pohon terbaik.

1.4 Tujuan Penelitian

Tujuan penulisan tugas akhir ini adalah :

1. mempelajari bagaimana metode QUEST digunakan dalam membentuk pohon klasifikasi biner.
2. menerapkan metode QUEST dalam mengklasifikasikan pengamatan .

1.5 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Bab ini berisikan latar belakang, perumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini berisikan teori-teori yang menjadi landasan metode QUEST.

BAB III : METODE QUEST

Bab ini berisikan pembahasan metode QUEST dan penerapannya.

BAB IV : PENUTUP

Bab ini berisikan kesimpulan dari analisis dan saran.

BAB II

LANDASAN TEORI

Pada bab ini akan dibahas tentang skala pengukuran data dan metode-metode statistika yang menjadi dasar pada metode QUEST, yaitu uji khi-kuadrat untuk kebebasan, uji *F*, uji Levene, analisis diskriminan kuadratik, dan pengertian metode klasifikasi berstruktur pohon.

2.1 Skala Pengukuran Data [8]

Skala pengukuran adalah penempatan angka atau lambang untuk menyatakan suatu hasil pengamatan/pengukuran terhadap sesuatu objek. Secara umum terdapat 4 skala pengukuran, yaitu :

1. Skala nominal

   Skala nominal adalah skala pengukuran data yang digunakan untuk mengklasifikasi objek-objek dalam kelompok (kategori) yang terpisah untuk menunjukkan kesamaan atau perbedaan ciri-ciri tertentu dari objek yang diamati. Kategori (kelompok) yang ada sudah didefinisikan sebelumnya dan dilambangkan dengan kata-kata, huruf simbol atau angka.

   Data berskala nominal memiliki ciri-ciri :

   a. hanya bersifat membedakan kategori, tidak bisa diurutkan mana yang lebih tinggi dan yang lebih rendah.

   b. kategori-kategori bersifat *mutually exclusive*, artinya setiap individu hanya dikategorikan pada satu kategori dan tidak tumpang tindih.

      Contoh skala nominal adalah agama, jenis kelamin, suku bangsa, golongan darah dan sebagainya.

2. Skala ordinal

Skala ordinal memiliki semua karakteristik skala nominal, kecuali bahwa pada skala ini terdapat urutan atau peringkat antarkategori. Angka yang digunakan untuk melambangkan kategori tidak memiliki nilai absolut, hanya menunjukkan posisi sebuah kategori relatif terhadap kategori lainnya.

Contoh skala ordinal adalah status sosial yang dikategorikan menjadi status sosial atas, status sosial menengah, status sosial bawah.

3. Skala interval

Skala interval memiliki semua karakteristik skala ordinal. Perbedaannya dengan skala ordinal adalah bahwa skala ini mempunyai satuan skala. Antar angka kategori memiliki jarak yang sama. Skala ini tidak mempunyai titik nol yang sesungguhnya, yang artinya titik nol merupakan sesuatu yang bermakna ada dengan nilai nol.

Contoh skala interval adalah suhu (^{0}C).

4. Skala rasio

Skala rasio sama dengan skala interval, tetapi skala rasio mempunyai titik nol yang sesungguhnya, yang artinya titik nol merupakan nilai yang bermakna tidak ada.

Contoh skala rasio adalah berat badan, panjang, usia, lama waktu dan sebagainya.

Data dengan skala nominal dan ordinal seringkali disebut sebagai data kategorik sedangkan data dengan skala interval dan rasio biasa disebut data numerik.

2.2 Pengujian Hipotesis

Ilmu statistika adalah ilmu yang mempelajari prosedur-prosedur yang digunakan dalam pengumpulan data, penyajian, analisis dan penafsiran data. Secara umum, ilmu statistika dapat dikelompokkan menjadi dua kelompok, yaitu statistika deskriptif dan statistika inferensia. Statistika deskriptif merupakan metode-metode yang berkaitan dengan pengumpulan dan penyajian data sehingga memberikan informasi yang berguna, sedangkan statistika inferensia merupakan semua metode yang berhubungan dengan analisis sebagian data sehingga sampai pada penarikan kesimpulan mengenai keseluruhan data.

Statistika inferensia dapat dilakukan dengan beberapa metode. Salah satunya dengan pengujian hipotesis. Pengujian hipotesis adalah metode perumusan sejumlah kaidah yang akan menghasilkan suatu kesimpulan untuk menerima atau menolak suatu penyataan tertentu. Langkah-langkah pengujian hipotesis dapat dibuat seperti berikut :

1. Rumuskan hipotesis

   Hipotesis adalah pernyataan awal yang akan diuji dalam suatu pengujian hipotesis. Hipotesis awal (H_0) adalah hipotesis yang dirumuskan dengan harapan akan ditolak. Hipotesis alternatif (H_1) adalah hipotesis yang dirumuskan dengan harapan akan diterima.

2. Tetapkan taraf nyata pengujian (α)

   α merupakan galat pengujian dengan kesalahan jenis I, yaitu kesalahan karena menolak hipotesis awal yang benar.

3. Pilih statistik uji yang sesuai

   Statistik uji adalah nilai yang diambil dari contoh dan digunakan sebagai dasar menerima atau menolak hipotesis awal.

4. Tentukan titik kritis

   Titik kritis adalah suatu nilai yang menjadi batas untuk menerima atau menolak hipotesis awal.

5. Tentukan nilai statistik uji

   Nilai statistik uji diambil berdasarkan data contoh.

6. Kesimpulan

   Jika statistik uji berada pada daerah kritis maka hipotesis awal ditolak. Jika statistik uji berada pada daerah penerimaan maka hipotesis awal diterima. Penarikan kesimpulan juga dapat dilakukan dengan membandingkan nilai α dengan nilai p, yaitu jika nilai $p <$ nilai α maka hipotesis awal ditolak dan jika nilai $p >$ nilai α maka hipotesis awal diterima.

2.3 Uji Khi-Kuadrat untuk kebebasan

Apabila antara dua peubah tidak ada hubungan, maka dapat dikatakan bahwa keduanya saling bebas. Meskipun nilai salah satu peubah untuk suatu objek diketahui, ini tidak akan membantu dalam menentukan nilai peubah yang lain untuk objek yang sama[6].

Uji Khi-Kuadrat untuk memeriksa kebebasan digunakan untuk memutuskan apakah dua peubah kategorik dalam suatu kelompok saling bebas. Uji ini memiliki asumsi-asumsi sebagai berikut :

a. Data terdiri dari sebuah contoh acak sederhana berukuran n dari suatu populasi yang diminati.

b. Hasil–hasil pengamatan dalam contoh dapat diklasifikasi secara silang (*cross-classified*) menurut peubah-peubah yang diamati.

Pengklasifikasian silang dari data dengan peubah kategorik biasanya disajikan dalam tabel kontingensi dua arah atau lebih. Bila terdapat dua peubah kategorik, data disajikan dalam tabel kontingensi dua arah seperti pada tabel berikut ini.

Tabel 2.3.1 Tabel Kontingensi Dua Arah

Kategori peubah pertama	Kategori peubah kedua				
	1	2	...	c	Jumlah
1	n_{11}	n_{12}	...	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2c}	$n_{2.}$
⋮	⋮	⋮		⋮	⋮
r	n_{r1}	n_{r2}	...	n_{rc}	$n_{r.}$
Jumlah	$n_{.1}$	$n_{.2}$		$n_{.c}$	n

Tabel disusun dari r baris dan c kolom dengan r dan c masing-masing adalah banyaknya kategori dari peubah kategorik pertama dan kedua. Isi sel pada baris-i (i = 1, 2, ..., r) dan kolom ke-j (j = 1, 2, ..., c) adalah banyaknya pengamatan yang berasal dari kategori-i peubah pertama dan kategori-j peubah kedua, biasa dinotasikan dengan n_{ij}. Isi sel ini disebut juga frekuensi sel teramati yang biasa ditulis dengan notasi O_{ij}, sehingga $O_{ij} = n_{ij}$. Jumlah frekuensi teramati pada kategori ke-i peubah pertama, ditulis dengan notasi $n_{i.}$, sedangkan jumlah frekuensi teramati pada kategori ke-j peubah kedua, ditulis dengan notasi $n_{.j}$. Tabel inilah yang menjadi dasar dalam pengujian hipotesis untuk kebebasan.

Hipotesis awal (H_0) yang digunakan dalam uji ini adalah bahwa kedua peubah saling bebas. Hipotesis alternatif (H_1) pada uji ini adalah bahwa kedua

peubah tidak saling bebas. Uji khi-kuadrat ini dilakukan dengan membandingkan frekuensi teramati dengan frekuensi yang diharapkan jika H_0 benar.

Dalam menentukan frekuensi yang diharapkan pada suatu sel digunakan hukum peluang mengenai kebebasan dua kejadian seperti dinyatakan dalam dalil berikut :

Dalil Kaidah Penggandaan Khusus[13] :

Bila dua kejadian A dan B bebas, maka $P(A \cap B) = P(A)\,P(B)$.

Jika A_i adalah kejadian objek berasal dari kategori ke-i peubah pertama dan B_j adalah kejadian objek berasal dari kategori ke-j peubah kedua, maka peluang kejadian A_i dan B_j terjadi bersama adalah $P(A_i \cap B_j) = \left(\frac{E_{ij}}{n}\right)$, dengan E_{ij} adalah frekuensi yang diharapkan. Jika A_i dan B_j saling bebas, maka

$$P\left(A_i \cap B_j\right) = P(A_i)\,P(B_j) = \left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.j}}{n}\right) \quad \text{(2.3.1).}$$

Dengan demikian, $\left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.j}}{n}\right) = \frac{E_{ij}}{n}$, sehingga

$$E_{ij} = n\left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.j}}{n}\right) = \left(\frac{n_{i.}n_{.j}}{n}\right) \quad \text{(2.3.2).}$$

Dari frekuensi sel yang teramati (O_{ij}) dan frekuensi sel yang diharapkan (E_{ij}) tersebut dapat dihitung suatu statistik uji khi-kuadrat (χ^2) yang mencerminkan perbedaan antara keduanya, yang dirumuskan sebagai berikut :

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left[\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}\right] \quad \text{(2.3.3).}$$

Dalam pengambilan keputusan, H_0 ditolak pada taraf nyata α jika nilai statistik uji χ^2 hasil perhitungan lebih besar daripada nilai $\chi^2_{\alpha,r-1,c-1}$.

MILIK
UPT PERPUSTAKAAN
UNIVERSITAS ANDALAS

2.4 Uji ANOVA *F*

Uji ANOVA *F* biasa digunakan untuk membandingkan nilai tengah dari dua atau lebih kelompok contoh yang saling bebas. Ukuran contoh antara masing-masing kelompok contoh tidak harus sama, tetapi perbedaan yang besar dalam ukuran contoh dapat mempengaruhi hasil uji perbandingan nilai tengah.

Misalkan x_{ki} merupakan pengamatan ke-*i* dari kelompok ke-*k*, maka dapat disajikan struktur data seperti pada tabel berikut ini.

Tabel 2.4.1 Tabel Struktur Data ANOVA *F*

	Kelompok				
	1	2	...	K	
	x_{11}	x_{21}	...	x_{K1}	
	x_{12}	x_{22}	...	x_{K2}	
	$\vdots$	$\vdots$		$\vdots$	
	x_{1n_1}	x_{2n_2}	...	x_{Kn_K}	
Jumlah	$T_{1.}$	$T_{2.}$	...	$T_{K.}$	$T_{..}$
Nilai tengah	$\bar{x}_{1.}$	$\bar{x}_{2.}$	...	$\bar{x}_{K.}$	$\bar{x}_{..}$

Bila μ_k adalah nilai tengah dari kelompok ke-*k* (k = 1, 2, ..., K) maka hipotesis yang digunakan dalam uji ini adalah

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$

H_1 : ada $\mu_k \neq \mu_{k'}$, $k \neq k'(k, k' = 1, 2, \ldots, K)$.

Statistik uji yang digunakan adalah statistik uji *F* yang biasanya diperoleh dengan terlebih dahulu membentuk Tabel 2.4.2 seperti di bawah ini :

Tabel 2.4.2 Tabel ANOVA *F*

Sumber Keragaman	Jumlah	Derajat Bebas	Kuadrat Tengah	*F* hitung
Nilai tengah Kolom	*JKK*	$K-1$	$s_1^2 = \frac{JKK}{K-1}$	$F = \frac{s_1^2}{s_2^2}$
Galat	*JKG*	$N-K$	$s_2^2 = \frac{JKG}{N-K}$	
Total	*JKT*	$N-1$		

dengan :

$$JKK = \left(\sum_{k=1}^{K} \frac{T_{k.}^2}{n_k} - \frac{T_{..}^2}{N}\right)$$

$$JKG = \left(\sum_{k=1}^{K}\sum_{i=1}^{n_k} x_{ki}^2 - \frac{T_{..}^2}{N}\right) - \left(\sum_{k=1}^{K} \frac{T_{k.}^2}{n_k} - \frac{T_{..}^2}{N}\right)$$

$i = 1, 2, \ldots, n_k$, $k = 1, 2, \ldots, K$.

N : jumlah seluruh data

K : jumlah kelompok

n_k : ukuran contoh kelompok ke-k

$T_{k.}$: jumlah pengamatan kelompok ke-k

$T_{..}$: jumlah pengamatan seluruh data

x_{ki} : pengamatan ke-i dari kelompok ke-k.

Dalam pengambilan keputusan, H_0 ditolak pada taraf nyata α jika nilai statistik uji F hasil perhitungan lebih besar daripada nilai $F_{\alpha,K-1,N-K}$.

2.5 Uji Levene *F*

Uji Levene F digunakan untuk menguji kesamaan ragam peubah dari dua kelompok atau lebih. Hipotesis yang digunakan dalam uji ini adalah

$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$

H_1 : ada $\sigma_k^2 \neq \sigma_{k'}^2$, $k \neq k'(k, k' = 1, 2, \ldots, K)$.

Untuk mendapatkan statistik uji ini, data ditransformasikan dulu menjadi simpangan mutlaknya terhadap nilai tengah data, yaitu $y_{ki} = |x_{ki} - \bar{x}_k|$,

dengan :

$i = 1, 2, \ldots, n_k$, $k = 1, 2, \ldots, K$

y_{ki} : hasil transformasi data pengamatan ke-i dari kelompok ke-k

$\bar{x}_k$: nilai tengah contoh kelompok ke-k

x_{ki} : pengamatan ke-i dari kelompok ke-k.

Lakukan statistik uji ANOVA F pada data yang telah ditransformasi untuk mendapatkan nilai statistik uji Levene F.

Dalam pengambilan keputusan, H_0 ditolak pada taraf nyata α jika nilai statistik uji Levene F hasil perhitungan lebih besar daripada nilai $F_{\alpha,K-1,N-K}$.

2.6 Analisis Diskriminan Kuadratik [9]

Analisis diskriminan bertujuan untuk membentuk fungsi diskriminan yang mampu membedakan kelompok. Analisis ini dilakukan berdasarkan suatu perhitungan statistik terhadap objek-objek yang telah diketahui dengan jelas dan mantap pengelompokannya.

Fungsi diskriminan dapat disebut dengan fungsi pembeda. Fungsi diskriminan yang dibangun dengan asumsi bahwa kelompok-kelompok memiliki matriks ragam peragam yang sama dinamakan fungsi diskriminan linier, sedangkan fungsi yang dibangun tanpa asumsi tersebut dinamakan fungsi diskriminan kuadratik.

Misalkan $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$, maka dapat disajikan struktur data seperti pada tabel berikut ini.

Tabel 2.6.1 Tabel Struktur Data Analisis Diskriminan Kuadratik

Peubah	Kelompok			
	1	2	…	K
X_1	$\bar{x}_{11}$	$\bar{x}_{12}$	…	$\bar{x}_{1K}$
X_2	$\bar{x}_{21}$	$\bar{x}_{22}$	…	$\bar{x}_{2K}$
⋮	⋮	⋮		⋮
X_p	$\bar{x}_{p1}$	$\bar{x}_{p2}$	…	$\bar{x}_{pK}$

Bila $f_k(\boldsymbol{x})$ adalah fungsi kepekatan peluang bersama dari contoh acak yang berasal dari kelompok ke-k ($k = 1, 2, \ldots, K$). Jika contoh acak pada kelompok tersebut menyebar menirut sebaran normal multivariat, maka

$$f_k(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu_k})^t \Sigma_k^{-1}(\boldsymbol{x} - \boldsymbol{\mu_k})\right] \ldots\ldots \quad (2.6.1),$$

dengan :

$\boldsymbol{\mu_k}$ adalah vektor nilai tengah kelompok ke-k

Σ_k adalah matriks ragam peragam kelompok ke-k

p adalah banyaknya peubah.

Skor diskriminan kuadratik untuk sebuah pengamatan dengan nilai $\boldsymbol{x}^* = (x_1^*, x_2^*, \ldots, x_p^*)$ terhadap kelompok ke–k adalah

$$d_k^Q(\boldsymbol{x}^*) = -\frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(\boldsymbol{x}^* - \boldsymbol{\mu_k})^t \Sigma_k^{-1}(\boldsymbol{x}^* - \boldsymbol{\mu_k}) + \ln p_k \ldots. (2.6.2),$$

dengan :

p_k adalah peluang awal dari kelompok ke-k ($k = 1, 2, \ldots, K$).

Bila individu yang berasal dari kelompok k dinyatakan sebagai kelompok a, maka peluangnya dinotasikan menjadi P($a|k$).

Kelompokkan $\boldsymbol{x}^*$ ke-k jika

Skor kuadratik $d_k^Q(\boldsymbol{x}^*) = \max\{d_1^Q(\boldsymbol{x}^*), d_2^Q(\boldsymbol{x}^*), \ldots, d_K^Q(\boldsymbol{x}^*)\}$ (2.6.3).

MILIK
UPT PERPUSTAKAAN
UNIVERSITAS ANDALAS

Dalam prakteknya, $\boldsymbol{\mu_k}$ dan $\sum_k$ tidak diketahui, tetapi data contoh yang telah dikelompokkan secara benar tersedia untuk pengkonstruksian taksiran $\boldsymbol{\mu_k}$ dan $\sum_k$. Kuantitas contoh yang relevan untuk kelompok ke-k adalah

$\bar{\boldsymbol{x}}_k$: vektor nilai tengah contoh dari kelompok ke-k

$\mathbf{S}_k$: matriks peragam contoh dari kelompok ke-k

n_k : ukuran contoh dari kelompok ke-k.

Taksiran dari skor diskriminan kuadratik :

$$\hat{d}_k^Q(\boldsymbol{x}^*) = -\frac{1}{2}\ln|S_k| - \frac{1}{2}(\boldsymbol{x}^* - \bar{\boldsymbol{x}}_k)^t\, S_k^{-1}(\boldsymbol{x}^* - \bar{\boldsymbol{x}}_k) + \ln p_k \ldots\ldots \quad (2.6.4).$$

Kelompokkan $\boldsymbol{x}^*$ ke-k jika

$$\text{Skor kuadratik } \hat{d}_k^Q(\boldsymbol{x}^*) = \max\{\hat{d}_1^Q(\boldsymbol{x}^*), \hat{d}_2^Q(\boldsymbol{x}^*), \ldots, \hat{d}_K^Q(\boldsymbol{x}^*)\} \quad (2.6.5).$$

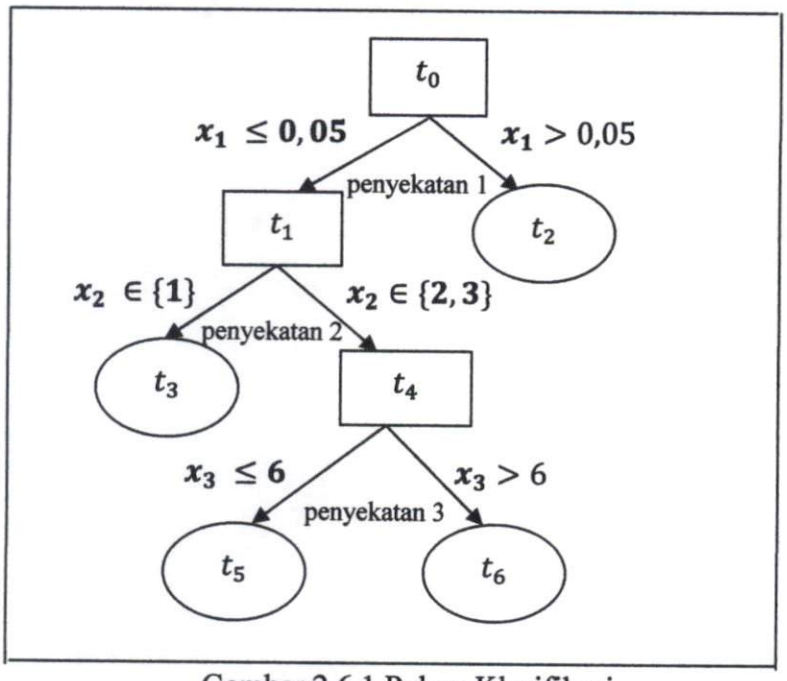
2.7 Metode Klasifikasi Berstruktur Pohon

Dalam statistika, terdapat berbagai metode yang dapat digunakan dalam menarik kesimpulan mengenai hubungan antara suatu peubah respon dengan beberapa peubah bebas. Jika peubah respon berupa data kuantitatif maka analisa mengenai hubungan peubah bebas dan respon biasanya dilakukan melalui analisis regresi biasa. Namun, bila peubah respon merupakan data kualitatif maka analisa mengenai hubungan peubah bebas dan respon salah satunya dapat dilakukan melalui teknik klasifikasi.

Metode klasifikasi berstruktur pohon merupakan metode statistika yang digunakan untuk memperkirakan keanggotaan amatan atau objek dalam kelas-kelas peubah respon kategorik, yang keanggotaannya diduga berdasarkan pengukuran terhadap satu peubah bebas atau lebih. Metode ini menghasilkan sebuah pohon klasifikasi (*classification tree*) yang dibentuk melalui penyekatan data secara berulang (rekursif) terhadap suatu himpunana data, di mana

pengelompokan dan nilai-nilai peubah bebas setiap amatan pada data contoh sudah diketahui. Setiap himpunan data dinyatakan sebagai simpul dalam pohon yang terbentuk [5].

Pohon klasifikasi dapat disajikan dalam sebuah gambar berikut:



Gambar 2.6.1 Pohon Klasifikasi

Dalam pohon klasifikasi seperti pada Gambar 2.6.1, himpunan data awal disebut simpul induk, dinotasikan sebagai t_0. Pada simpul t_0, dilakukan penyekatan sehingga terbentuk simpul t_1 dan t_2. Penyekatan dilakukan secara berulang sampai diperoleh sebuah simpul yang tidak dapat disekat lagi, yang disebut simpul akhir. Simpul yang tidak termasuk pada simpul induk dan simpul akhir disebut simpul dalam. Dapat dilihat bahwa t_1, t_4 adalah simpul dalam sedangkan t_2, t_3, t_5 dan t_6 adalah simpul akhir [4]. Pada simpul-simpul akhir ini dilakukan pendugaan respon.

Penyekatan terhadap sebuah simpul dilakukan berdasarkan nilai suatu peubah bebas (titik penyekat). Peubah bebas yang digunakan untuk menyekat sebuah simpul dinamakan sebagai peubah penyekat. Peubah penyekat dapat berupa peubah kategorik maupun peubah kontinu. Pada Gambar 2.6.1 terlihat bahwa simpul t_0 disekat berdasarkan peubah x_1 menjadi simpul t_1 bila $x_1 \leq 0{,}05$ dan simpul t_2 bila $x_1 > 0{,}05$. Simpul dalam t_1 disekat lagi berdasarkan peubah x_2 menjadi simpul t_3 bila $x_2 \in \{1\}$ dan t_4 bila $x_2 \in \{2, 3\}$. Simpul dalam t_4 disekat lagi berdasarkan peubah x_3 menjadi simpul t_5 bila $x_3 \leq 6$ dan t_6 bila $x_3 > 6$. Dalam menyekat suatu simpul, setiap peubah bebas memiliki kesempatan untuk terpilih sebagai peubah penyekat, meskipun peubah tersebut telah terpilih sebelumnya sebagai peubah penyekat simpul lain.

Proses penyekatan terhadap simpul dilakukan secara berulang sampai ditemukan salah satu dari tiga hal berikut:

a. respon di semua simpul sudah homogen nilainya
b. tidak ada lagi peubah bebas yang bisa digunakan
c. jumlah objek di dalam simpul sudah terlalu sedikit untuk menghasilkan pemisahan yang memuaskan.

Dalam pembentukan pohon klasifikasi, proses penyekatan terhadap suatu simpul dapat bersifat biner atau non biner. Pada penyekatan biner, setiap simpul hanya boleh disekat menjadi dua simpul baru, sedangkan pada penyekatan non biner setiap simpul dapat menghasilkan lebih dari dua simpul baru. Gambar 2.6.1 merupakan pohon klasifikasi dengan penyekatan biner.

Pohon yang dibentuk dari proses penyekatan tersebut dapat berukuran sangat besar. Bila pohon berukuran besar, biasanya penduga respon cenderung lebih tepat, tapi sulit diinterpretasi. Bila pohon kecil, pohon mudah diinterpretasi namun penduga respon cenderung tidak tepat. Pohon terbaik yaitu pohon yang memiliki keseimbangan antara ukuran pohon dan ketepatan penduga respon. Untuk menghasilkan pohon terbaik dilakukan pemangkasan pohon yang sudah terbentuk.

BAB III

METODE QUEST

3.1 Metode Quest

QUEST (*Quick, Unbiased, Efficient Statistical Trees*) merupakan salah satu metode yang digunakan untuk membentuk pohon klasifikasi. QUEST merupakan algoritma pemisah yang menghasilkan pohon biner yang digunakan untuk klasifikasi. Algoritma pembentukan pohon klasifikasi ini merupakan modifikasi dari analisis diskriminan kuadratik.

Pada algoritma ini, proses penyekatan dapat dilakukan pada peubah tunggal *(univariate)*. Pemilihan peubah penyekat pada QUEST menerapkan uji kebebasan khi-kuadrat untuk peubah kategorik dan uji F untuk peubah numerik. Suatu peubah dipilih sebagai peubah penyekat jika menghasilkan kelompok dengan tingkat kehomogenan peubah respon yang paling besar. Penentuan titik penyekat pada pohon klasifikasi ini dilakukan dengan menerapkan analisis diskriminan kuadratik. Pemilihan peubah dan penentuan titik penyekat dilakukan secara terpisah. Komponen dasar QUEST adalah beberapa peubah bebas yang merupakan peubah kategorik atau numerik dan peubah respon yang merupakan peubah kategorik.

3.2 Algoritma QUEST

Berikut ini akan dijelaskan algoritma pembentukan pohon pada QUEST. Algoritma ini dipisah menjadi tiga bagian, yaitu algoritma pemilihan peubah penyekat, algoritma penentuan titik penyekat dan algoritma transformasi peubah kategorik menjadi peubah numerik

3.2.1 Algoritma Pemilihan Peubah Penyekat

Dalam tulisan ini akan dibahas algoritma pemilihan peubah penyekat berupa satu peubah. Dalam menentukan peubah penyekat pada suatu simpul setiap peubah memiliki kesempatan untuk terpilih sebagai peubah penyekat, meskipun peubah tersebut telah terpilih sebagai peubah penyekat untuk simpul sebelumnya.

Berikut adalah langkah-langkah pemilihan peubah penyekat :

1. Untuk setiap peubah X :
   - Jika X merupakan peubah kategorik, lakukan uji χ^2 untuk kebebasan antara peubah X dan peubah respon Y dan hitung nilai p dari pengujian tersebut.
   - Jika X merupakan peubah numerik, lakukan uji ANOVA F dan hitung nilai p dari pengujian tersebut.
2. Pilih peubah dengan nilai p terkecil.
3. Bandingkan nilai p terkecil dengan taraf α/M_1, dengan pilih taraf nyata $\alpha = 0{,}05$ dan M_1 adalah banyaknya peubah bebas.
   - Jika nilai p kurang dari α/M_1, maka pilih peubah yang bersesuaian sebagai peubah penyekat. Teruskan ke langkah (5).
   - Jika nilai p lebih dari α/M_1, teruskan ke langkah (4).
4. Untuk setiap peubah X yang numerik, maka hitung nilai p dari uji Levene untuk menguji kehomogenan ragam.
   - Pilih peubah dengan nilai p terkecil.
   - Bandingkan nilai p terkecil dari uji Levene dengan taraf $\alpha/(M_1+M_2)$, dengan M_2 adalah banyaknya peubah bebas numerik.
   - Jika nilai p kurang dari $\alpha/(M_1+M_2)$, maka pilih peubah yang bersesuaian sebagai peubah penyekat. Teruskan ke langkah (5).

- Jika nilai p lebih dari $\alpha/(M_1+M_2)$, maka peubah tersebut tidak dipilih menjadi peubah penyekat.

5. Misalkan X^* adalah peubah penyekat yang diperoleh dari langkah (3) atau (4).
   - Jika X^* merupakan peubah numerik, maka teruskan ke langkah (6).
   - Jika X^* merupakan peubah kategorik, X^* ditranformasikan ke dalam peubah *dummy*, lalu proyeksikan ke dalam koordinat diskriminan terbesarnya. Proyeksi ke dalam koordinat diskriminan terbesar dapat dilihat pada Subbab 3.2.3.
6. Lakukan analisis diskriminan kuadratik untuk menentukan titik penyekat.

3.2.2 Algoritma Penentuan Titik Penyekat

Misalkan peubah respon memiliki dua kategori. Misalkan pula bahwa X^* merupakan peubah yang terpilih untuk menyekat simpul t.

1. Definisikan $\bar{x}_0$ dan $s^2{}_0$ adalah nilai tengah dan ragam X^* dari pengamatan dengan respon 0, sedangkan $\bar{x}_1$ dan $s^2{}_1$ adalah nilai tengah dan ragam X^* dari pengamatan dengan respon 1. Misalkan $\mathrm{P}(k|t) = N_{k,t}/N_k$ merupakan peluang dari masing-masing kategori peubah respon, dengan $N_{k,t}$ adalah jumlah data pada simpul t untuk respon k dan N_k adalah jumlah data pada simpul awal untuk respon k.
2. Tentukan solusi dari persamaan

$$\mathrm{P}(0|t)s_0{}^{-1}\varphi\left(\frac{x-\bar{x}_0}{s_0}\right) = \mathrm{P}(1|t)s_1{}^{-1}\varphi\left(\frac{x-\bar{x}_1}{s_1}\right).$$

Solusi tersebut dapat ditentukan dengan menentukan akar persamaan kuadrat $ax^2 + bx + c = 0$, dengan :

$$a = s^2{}_0 - s^2{}_1$$

$$b = 2(\bar{x}_0\, s^2{}_1 - \bar{x}_1\, s^2{}_0)$$

$$c = (\bar{x}_1\, s_0)^2 - (\bar{x}_0\, s_1)^2 + 2s^2{}_0 s^2{}_1 \ln\left\{\frac{P(0|t)s^2{}_1}{P(1|t)s^2{}_0}\right\}$$

3. Simpul disekat pada titik $x^* = d$, di mana d didefinisikan sebagai berikut:

   a. Jika $\bar{x}_0 < \bar{x}_1$, maka $d = \bar{x}_0$

   b. Jika $a = 0$, maka $d = \begin{cases} \frac{\bar{x}_0+\bar{x}_1}{2} - (\bar{x}_0 - \bar{x}_1)^{-1} s^2{}_0 \ln\left\{\frac{P(0|t)}{P(1|t)}\right\}, & \bar{x}_0 \neq \bar{x}_1 \\ \bar{x}_0 & , \bar{x}_0 = \bar{x}_1 \end{cases}$

   c. Jika $a \neq 0$, maka:

   - Jika $b^2 - 4ac < 0$, maka $d = \frac{1}{2}(\bar{x}_0 + \bar{x}_1)$
   - Jika $b^2 - 4ac \geq 0$, maka :

     d adalah akar dari $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$ yang lebih mendekati nilai $\bar{x}_0$, dengan syarat menghasilkan dua simpul tak-kosong.

3.2.3 Algoritma Transformasi Peubah Kategorik menjadi Peubah Numerik.

Misalkan X adalah peubah kategorik, dengan kategori $b_1, b_2, \ldots, b_L$. Transformasi X menjadi peubah numerik ξ untuk tiap kelas X dilakukan dengan langkah-langkah sebagai berikut :

1. Transformasikan masing-masing nilai x ke vektor *dummy* L dimensi

   $$\boldsymbol{v} = (v_1, v_2, \ldots, v_L)',$$

   dengan $v_l = \begin{cases} 1 & x = b_l \\ 0 & x \neq b_l \end{cases}$, $l = 1, 2, \ldots, L$

2. Cari nilai tengah untuk X

   $$\bar{\boldsymbol{v}} = \frac{\sum_{l=1}^{L} f_l \boldsymbol{v}_l}{N_t} \qquad \bar{\boldsymbol{v}}^{(k)} = \frac{\sum_{l=1}^{L} n_l\, \boldsymbol{v}_l}{N_{k,t}}$$

   dengan :

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

$\overline{\boldsymbol{v}}^{(k)}$: rata-rata untuk semua pengamatan pada simpul t untuk kelompok ke-k

f_l : jumlah pengamatan pada simpul t untuk v_l

n_l : jumlah pengamatan pada simpul t kelompok ke-k untuk v_l

N_t : jumlah pengamatan pada simpul t

$N_{k,t}$: jumlah pengamatan pada simpul t untuk kelompok ke-k.

3. Tentukan matriks L x L berikut :

$$\overline{\boldsymbol{v}} = \frac{\sum_{l=1}^{L} f_l \boldsymbol{v_l}}{N_t} \qquad \overline{\boldsymbol{v}}^{(k)} = \frac{\sum_{l=1}^{L} n_l \boldsymbol{v_l}}{N_{k,t}}$$

$$\boldsymbol{B} = \sum_{k=1}^{K} N_{k,t} \left(\overline{\boldsymbol{v}}^{(k)} - \overline{\boldsymbol{v}}\right)\left(\overline{\boldsymbol{v}}^{(k)} - \overline{\boldsymbol{v}}\right)'$$

$$\boldsymbol{T} = \sum_{l=1}^{L} f_l(\boldsymbol{v_l} - \overline{\boldsymbol{v}})(\boldsymbol{v_l} - \overline{\boldsymbol{v}})'$$

4. Lakukan SVD dari $\boldsymbol{T} = \boldsymbol{QDQ'}$, dengan $\boldsymbol{Q}$ adalah matriks orthogonal L x L, $\boldsymbol{D}$ = diag(d_1, ..., d_L) dengan $d_1 \geq d_2 \geq \ldots \geq d_L \geq 0$.

5. Tentukan $\boldsymbol{D}^{-\frac{1}{2}} = \text{diag}(d_1^*, \ldots, d_L^*)$,

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

6. Lakukan SVD dari $\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{Q'BQD}^{-\frac{1}{2}}$, tentukan vektor eigen $\boldsymbol{a}$ yang merupakan vektor eigen yang berpadanan dengan nilai eigen terbesar.

7. Tentukan koordinat diskriminan terbesar dari $\boldsymbol{v}$, yaitu :

$$\xi = \boldsymbol{a'D}^{-\frac{1}{2}}\boldsymbol{Q'v}$$

3.3 Contoh Penerapan Metode QUEST

3.3.1 Kasus

Penerapan Metode QUEST ini akan diperlihatkan melalui kasus pengkajian faktor-faktor apa yang mempengaruhi seseorang menderita penyakit jantung. Penyakit jantung merupakan salah satu penyakit yang dapat menyebabkan kematian. Beberapa faktor yang mempengaruhi seseorang menderita penyakit jantung adalah usia, jenis kelamin, tekanan darah, kolesterol, latihan menginduksi angina, kondisi jantung dan sebagainya.

Pada usia lanjut, kemungkinan terserang penyakit jantung lebih besar. Pria beresiko terserang penyakit jantung pada usia 45 tahun, sedangkan perempuan pada usia 55 tahun. Tekanan darah yang tinggi dan kolesterol darah yang tidak normal juga dapat menyebabkan terserang penyakit jantung. Jika gula darah seseorang > 120 mg maka gula darah akan mengendap di saluran darah sehingga menghambat darah sampai ke jantung. Elektrokardiografi merupakan alat yang bisa mendeteksi serangan jantung. Elektrokardiografi juga bisa untuk mendeteksi penyakit lainnya. Bila menghasilkan kelainan gelombang, kemungkinan terserang penyakit jantung lebih besar. Apabila latihan menginduksi angina, dapat terserang penyakit jantung. Jika kondisi jantung normal, kemungkinan terserang penyakit jantung sangatlah kecil. Oleh karena itu, untuk mengetahui faktor yang lebih berpengaruh terhadap penyakit jantung dilakukan penelitian ini.

3.3.2 Data

Berikut ini disajikan data yang diperoleh untuk mengkaji faktor–faktor yang mempengaruhi seseorang menderita penyakit jantung. Peubah tak bebas Y merupakan peubah kategori dengan dua kategori, yaitu 1 bila seseorang menderita

penyakit jantung dan 0 bila seseorang tidak menderita penyakit jantung. Data ini diperoleh dari UCI Irvine Machine Learning Repository (Blake CL, Merz CJ), terdiri dari 12 peubah bebas dengan jumlah pengamatan 270.

a. Peubah numerik

- X_1 merupakan usia seseorang
- X_3 merupakan tekanan darah seseorang pada saat beristirahat
- X_4 merupakan kolesterol darah seseorang
- X_7 merupakan denyut jantung maksimum yang dicapai seseorang
- X_9 merupakan hasil *oldpeak* seseorang

b. Peubah kategorik

- X_2 merupakan jenis kelamin seseorang dengan dua kategori, yaitu perempuan dan laki-laki
- X_5 merupakan gula darah puasa seseorang dengan dua kategori, yaitu gula darah > 120 mg dan gula darah < 120 mg
- X_6 merupakan hasil elektrokardiografi seseorang pada saat beristirahat dengan tiga kategori, yaitu normal, memiliki kelainan gelombang dan adanya kemungkinan hipertrofi vertikal
- X_8 merupakan latihan menginduksi angina seseorang dengan dua kategori, yaitu ya dan tidak
- X_{10} merupakan kemiringan puncak segmen ST pada saat latihan dengan tiga kategori, yaitu condong keatas, datar dan condong kebawah
- X_{11} merupakan jumlah pembuluh darah besar yang diwarnai oleh fluorosopy dengan empat kategori, yaitu 0, 1, 2 dan 3

- X_{12} merupakan kondisi jantung dengan 3 kategori, yaitu normal, cacat tetap dan cacat sementara.

3.3.3 Penyelesaian dengan Metode QUEST

Berikut akan diilustrasikan pembentukan pohon dengan menggunakan metode QUEST untuk data penyakit jantung. Simpul awal t_0 terdiri dari 270 pengamatan dari dua kategori pengamatan yaitu pasien yang berpenyakit jantung (120 pengamatan) dan pasien yang tidak berpenyakit jantung (150 pengamatan). Pemilihan penyekat dipilih dengan melakukan uji ANOVA F bila peubah penyekat numerik dan uji χ^2 bila peubah penyekat kategorik.

Pada data ini, proses penyekatan akan dihentikan bila banyaknya data pada simpul dalam paling sedikit 50, sedangkan banyaknya data pada simpul akhir paling sedikit 20. Pilihan tersebut diambil untuk menggambarkan pohon klasifikasi dengan algoritma QUEST pada pengamatan penyakit jantung.

Sebagai ilustrasi, uji ANOVA F dilakukan pada peubah numerik X_1 dan uji χ^2 dilakukan pada peubah kategorik X_2.

Untuk peubah X_1, misalkan $T_{1.}$ merupakan jumlah pengamatan peubah X_1 pada pasien yang tidak terserang penyakit jantung, $T_{2.}$ merupakan jumlah pengamatan peubah X_1 pada pasien yang terserang penyakit jantung, dan $T_{..}$ merupakan jumlah pengamatan peubah X_1 untuk seluruh data, diperoleh :

$T_{1.}$	= 7906	n_1	= 150
$T_{2.}$	= 6791	n_2	= 120
$T_{..}$	= 14697	N	= 270

sehingga :

Tabel 3.3.1 Tabel Anova F

Sumber Keragaman	Jumlah	Derajat Bebas	Kuadrat Tengah	F hitung
Nilai tengah Kolom	1006,215	1	$s_1^2 = 1006{,}215$	$F = 12{,}65$
Galat	21314,085	*268*	$s_2^2 = 79{,}53$	
Total	22320,300	*269*		

dengan :

$$JKK = \left(\sum_{k=1}^{K} \frac{T_{k.}^2}{n_k} - \frac{T_{..}^2}{N}\right) = \left(\frac{7906^2}{150} + \frac{6791^2}{120} - \frac{14697^2}{270}\right) = 1006{,}215$$

$$JKG = \left(\sum_{k=1}^{K}\sum_{i=1}^{n_k} x_{ik}^2 - \frac{T_{..}^2}{N}\right) - \left(\sum_{k=1}^{K} \frac{T_{k.}^2}{n_k} - \frac{T_{..}^2}{N}\right)$$

$$= \left[\left(70^2 + 67^2 + \cdots + 67^2 - \frac{14697^2}{270}\right) - \left(\frac{7906^2}{150} + \frac{6791^2}{120} - \frac{14697^2}{270}\right)\right]$$

$$= (22320{,}3 - 1006{,}215) = 21314{,}085$$

Dengan bantuan Minitab diperoleh nilai-p = 0,000

Peubah X_2 terdiri dari dua kategori, yaitu perempuan bila 0 dan laki-laki bila 1.

Tabel tabulasi silang antara X_2 dan Y adalah sebagai berikut :

Tabel 3.3.2 Tabel frekuensi teramati (O_{ij})

Y	X_2		Total
	0	1	
1	20	100	120
0	67	83	150
Total	87	183	270

Frekuensi sel yang diharapkan dapat dicari dengan cara sebagai berikut :

$$E_{11} = \left(\frac{n_{1.}n_{.1}}{n}\right) = \left(\frac{120.87}{270}\right) = 38{,}67$$

$$E_{12} = \left(\frac{n_{1.}n_{.2}}{n}\right) = \left(\frac{120.183}{270}\right) = 81{,}33$$

$$E_{21} = \left(\frac{n_{2.}n_{.1}}{n}\right) = \left(\frac{150.87}{270}\right) = 48{,}33$$

$$E_{22} = \left(\frac{n_{2.}n_{.2}}{n}\right) = \left(\frac{150.183}{270}\right) = 101{,}67$$

Nilai statistik uji χ^2 diperoleh sebagai berikut :

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right]$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2}\left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right]$$

$$= \frac{(20-38{,}67)^2}{38{,}67} + \frac{(100-81{,}33)^2}{81{,}33} + \frac{(67-48{,}33)^2}{48{,}33} + \frac{(83-101{,}67)^2}{101{,}67}$$

$$= 23{,}932$$

Dengan bantuan Minitab diperoleh nilai $p = 0{,}000$.

Dengan cara yang sama diperoleh nilai statistik uji dan nilai p untuk peubah lain seperti pada tabel berikut :

Tabel 3.3.3 Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_0

Peubah	Nilai Statistik Uji	Nilai p
X_1	$F = 12{,}65$	$p = 0{,}000$
X_2	$\chi^2 = 23{,}93$	$p = 0{,}000$
X_3	$F = 6{,}63$	$p = 0{,}011$
X_4	$F = 3{,}79$	$p = 0{,}053$
X_5	$\chi^2 = 0{,}07$	$p = 0{,}789$
X_6	$\chi^2 = 8{,}98$	$p = 0{,}011$
X_7	$F = 56{,}91$	$p = 0{,}000$
X_8	$\chi^2 = 47{,}47$	$p = 0{,}000$
X_9	$F = 56{,}73$	$p = 0{,}000$
X_{10}	$\chi^2 = 40{,}37$	$p = 0{,}000$
X_{11}	$\chi^2 = 62{,}86$	$p = 0{,}000$
X_{12}	$\chi^2 = 74{,}57$	$p = 0{,}000$

Bila digunakan α = 0,05, didapat α/M_1 = 0,05/12 = 0,0042. Dari tabel diatas terlihat bahwa nilai p = 0,000 < 0,0042 ditemukan pada beberapa peubah bebas. Pada dasarnya nilai ini adalah nilai pembulat ke-3 angka desimal pertama. Nilai p terkecil akan diperoleh pada peubah dengan nilai statistik uji Anova F dan uji χ^2 terbesar, yaitu pada peubah X_{12} dan peubah X_7. Karena keterbatasan alat hitung maka X_{12} dianggap memiliki nilai p terkecil dengan melihat pohon klasifikasi dengan algoritma QUEST yang dilakukan dengan software SPSS. Oleh karena itu, peubah tersebut dipilih sebagai peubah penyekat.

Peubah X_{12} adalah peubah kategorik dengan banyak kategori > 2 sehingga untuk mendapatkan titik penyekat, peubah ini ditransformasikan terlebih dahulu menjadi peubah numerik. Langkah-langkah transformasi adalah sebagai berikut :

1. Transformasikan masing-masing nilai x_{12} ke vektor peubah *dummy* 3 dimensi

   $\boldsymbol{v} = (v_1, v_2, v_3)'$,

   dengan $v_l = \begin{cases} 1 & x = b_l \\ 0 & x \neq b_l \end{cases}$, $l = 1, 2, 3$

   Dengan demikian,

   ❖ jika $x_{12} = 3$ maka $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

   ❖ jika $x_{12} = 6$ maka $\boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

   ❖ jika $x_{12} = 7$ maka $\boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

Berikut ini adalah tabel tabulasi dari transformasi peubah kategorik

Tabel 3.3.4 Tabel Tabulasi X_{12}

Y	X_{12}			Total
	$\boldsymbol{v}_1 = (1\,0\,0)'$ ($X_{12}=3$)	$\boldsymbol{v}_2 = (0\,1\,0)'$ ($X_{12}=6$)	$\boldsymbol{v}_3 = (0\,0\,1)'$ ($X_{12}=7$)	
0	119	6	25	150
1	33	8	79	120
Total	152	14	104	270

2. Cari nilai tengah untuk X_{12}

$$\bar{\boldsymbol{v}} = \frac{\sum_{l=1}^{3} f_l \boldsymbol{v}_l}{N_t} = \frac{152\begin{pmatrix}1\\0\\0\end{pmatrix} + 14\begin{pmatrix}0\\1\\0\end{pmatrix} + 104\begin{pmatrix}0\\0\\1\end{pmatrix}}{270}$$

$$= \begin{pmatrix}152\\14\\104\end{pmatrix}/270 \quad = \begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}$$

$$\bar{\boldsymbol{v}}^{(k)} = \frac{\sum_{l=1}^{3} n_l\, \boldsymbol{v}_l}{N_{k,t}}$$

$$\bar{\boldsymbol{v}}^{(0)} = \frac{119\begin{pmatrix}1\\0\\0\end{pmatrix} + 6\begin{pmatrix}0\\1\\0\end{pmatrix} + 25\begin{pmatrix}0\\0\\1\end{pmatrix}}{150} = \frac{\begin{pmatrix}119\\6\\25\end{pmatrix}}{150} = \begin{pmatrix}119/150\\6/150\\25/150\end{pmatrix}$$

$$\bar{\boldsymbol{v}}^{(1)} = \frac{33\begin{pmatrix}1\\0\\0\end{pmatrix} + 8\begin{pmatrix}0\\1\\0\end{pmatrix} + 79\begin{pmatrix}0\\0\\1\end{pmatrix}}{120} = \frac{\begin{pmatrix}33\\8\\79\end{pmatrix}}{120} = \begin{pmatrix}33/120\\8/120\\79/120\end{pmatrix}$$

3. Kalkulasikan nilai berikut dalam bentuk matriks 3 x 3

$$\boldsymbol{B} = \sum_{k=1}^{K} N_{k,t}\left(\bar{\boldsymbol{v}}^{(k)} - \bar{\boldsymbol{v}}\right)\left(\bar{\boldsymbol{v}}^{(k)} - \bar{\boldsymbol{v}}\right)'$$

$$= 150\left(\bar{\boldsymbol{v}}^{(0)} - \bar{\boldsymbol{v}}\right)\left(\bar{\boldsymbol{v}}^{(0)} - \bar{\boldsymbol{v}}\right)' + 120\left(\bar{\boldsymbol{v}}^{(1)} - \bar{\boldsymbol{v}}\right)\left(\bar{\boldsymbol{v}}^{(1)} - \bar{\boldsymbol{v}}\right)'$$

$$= 150\left(\begin{pmatrix}119/150\\6/150\\25/150\end{pmatrix} - \begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)\left(\begin{pmatrix}119/150\\6/150\\25/150\end{pmatrix} - \begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)'$$

$$+120\left(\begin{pmatrix}33/120\\8/120\\79/120\end{pmatrix}-\begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)\left(\begin{pmatrix}33/120\\8/120\\79/120\end{pmatrix}-\begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)'$$

$$=150\begin{pmatrix}311/1350\\-8/675\\-59/675\end{pmatrix}(311/1350 \quad -8/675 \quad -59/675)$$

$$+120\begin{pmatrix}-311/1080\\2/135\\59/216\end{pmatrix}(-311/1080 \quad 2/135 \quad 59/216)$$

$$=\begin{pmatrix}17{,}91 & -0{,}92 & -28{,}51\\-0{,}92 & 0{,}05 & 1{,}47\\-28{,}51 & 1{,}47 & 51{,}30\end{pmatrix}$$

$$\boldsymbol{T}=\sum_{l=1}^{3} f_l(\boldsymbol{v}_l-\overline{\boldsymbol{v}})(\boldsymbol{v}_l-\overline{\boldsymbol{v}})'$$

$$=152(\boldsymbol{v}_1-\overline{\boldsymbol{v}})(\boldsymbol{v}_1-\overline{\boldsymbol{v}})'+14(\boldsymbol{v}_2-\overline{\boldsymbol{v}})(\boldsymbol{v}_2-\overline{\boldsymbol{v}})'$$

$$+104(\boldsymbol{v}_3-\overline{\boldsymbol{v}})(\boldsymbol{v}_3-\overline{\boldsymbol{v}})'$$

$$=152\left(\begin{pmatrix}1\\0\\0\end{pmatrix}-\begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)\left(\begin{pmatrix}1\\0\\0\end{pmatrix}-\begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)'$$

$$+14\left(\begin{pmatrix}0\\1\\0\end{pmatrix}-\begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)\left(\begin{pmatrix}0\\1\\0\end{pmatrix}-\begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)'$$

$$+104\left(\begin{pmatrix}0\\0\\1\end{pmatrix}-\begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)\left(\begin{pmatrix}0\\0\\1\end{pmatrix}-\begin{pmatrix}152/270\\14/270\\104/270\end{pmatrix}\right)'$$

$$=\begin{pmatrix}-1874{,}74 & -167{,}387 & 2042{,}12\\217{,}99 & 19{,}464 & -237{,}46\\1678{,}54 & 149{,}870 & -1828{,}41\end{pmatrix}$$

4. Lakukan SVD terhadap $\boldsymbol{T}$ menjadi $\boldsymbol{T}=\boldsymbol{QDQ}'$,

   dengan :

   $\boldsymbol{Q}$ adalah matriks orthogonal yang kolomnya merupakan vektor eigen dari $\boldsymbol{T}'\boldsymbol{T}$

$\boldsymbol{D}$ adalah matriks diagonal yang merupakan akar dari nilai eigen dari $\boldsymbol{T'T}$, $d_1 \geq d_2 \geq d_3 \geq 0$.

Diperoleh :

$$\boldsymbol{Q} = \begin{pmatrix} -0{,}675042 & -0{,}737599 & -0{,}016270 \\ -0{,}060272 & 0{,}077112 & -0{,}995199 \\ 0{,}735313 & -0{,}670821 & -0{,}096510 \end{pmatrix}$$

$$\boldsymbol{D} = \begin{pmatrix} 3741{,}69 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

5. Diperoleh matriks $\boldsymbol{D}^{-\frac{1}{2}} = \begin{pmatrix} 0{,}0163481 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

6. Lakukan SVD dari $\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{Q'BQD}^{-\frac{1}{2}}$, diperoleh vektor $\boldsymbol{a}$ yang merupakan vektor eigen yang berpadanan dengan nilai eigen terbesar : $\boldsymbol{a} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

7. Diperoleh koordinat diskriminan terbesar dari $\boldsymbol{v}$, yaitu :

$$\xi = \boldsymbol{a'D}^{-\frac{1}{2}}\boldsymbol{Q'v} = (-0{,}0110032 \quad -0{,}0009824 \quad 0{,}0119856)\,\boldsymbol{v}$$

Dengan demikian, diperoleh :

❖ untuk $x_{12} = 3$, maka

$$\xi = (-0{,}0110032 \quad -0{,}0009824 \quad 0{,}0119856)\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = -0{,}0110032,$$

❖ untuk $x_{12} = 6$, maka

$$\xi = (-0{,}0110032 \quad -0{,}0009824 \quad 0{,}0119856)\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = -0{,}0009824,$$

❖ untuk $x_{12} = 7$, maka

$$\xi = (-0{,}0110032 \quad -0{,}0009824 \quad 0{,}0119856)\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0{,}0119856.$$

Dengan demikian diperoleh tabel tabulasi X_{12} yang baru seperti pada tabel 3.3.5

Tabel 3.3.5 Tabel Tabulasi X_{12} Hasil Transformasi

Y	ξ			Total
	$-0{,}0110032$ ($X_{12}=3$)	$-0{,}0009824$ ($X_{12}=6$)	$0{,}0119856$ ($X_{12}=7$)	
0	119	6	25	150
1	33	8	79	120
Total	152	14	104	270

Pada peubah X_{12} dilakukan analisis diskriminan kuadratik untuk memperoleh titik penyekat. Dari data diperoleh :

$\bar{x}_0 = -0{,}00669$ $\qquad$ $\bar{x}_1 = 0{,}00493$

$s_0^2 = (0{,}008703)^2$ $\qquad$ $s_1^2 = (0{,}010222)^2$

$\mathrm{P}(0|t) = 1$ $\qquad$ $\mathrm{P}(1|t) = 1$

Dapat dilihat bahwa $\bar{x}_0 < \bar{x}_1$ maka $d = \bar{x}_0$ sehinggga titik penyekat simpul t_0 adalah $\bar{x}_0 = -0{,}00669$. Disimpulkan bahwa X_{12} memotong di nilai $\xi = -0{,}0110032$. Dengan demikian, simpul t_0 disekat menjadi dua simpul, yaitu simpul t_1 dengan $\xi > -0{,}0110032$ dan simpul t_2 dengan $\xi \leq -0{,}0110032$.

Dengan kata lain, simpul tersebut disekat menjadi dua simpul, yaitu simpul t_1 dengan $x_{12} \in \{6, 7\}$ dan simpul t_2 dengan $x_{12} \in \{3\}$. Simpul t_1 terdiri dari 118 pengamatan dari dua kategori, yaitu pasien berpenyakit jantung (87 pengamatan) dan pasien tidak berpenyakit jantung (31 pengamatan). Simpul t_2 terdiri dari 152 pengamatan dari dua kategori, yaitu pasien berpenyakit jantung (33 pengamatan) dan pasien tidak berpenyakit jantung (119 pengamatan). Proses penyekatan dilanjutkan karena simpul t_1 dan t_2 merupakan simpul dalam.

Selanjutnya dilakukan langkah yang sama untuk mendapatkan peubah penyekat pada simpul t_1. Dari uji yang dilakukan diperoleh nilai χ^2 dan nilai F sebagai berikut :

Tabel 3.3.6 Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_1

Peubah	Nilai Statistik Uji	Nilai p
X_1	F = 0,13	p = 0,714
X_2	χ^2 = 0,89	p = 0,344
X_3	F = 2,35	p = 0,128
X_4	F = 1,65	p = 0,201
X_5	χ^2 = 0,95	p = 0,330
X_6	χ^2 = 6,36	p = 0,012
X_7	F = 10,54	p = 0,002
X_8	χ^2 = 12,64	p = 0,000
X_9	F = 12,56	p = 0,001
X_{10}	χ^2 = 9,36	p = 0,009
X_{11}	χ^2 = 23,01	p = 0,000
X_{12}	χ^2 = 2,26	p = 0,133

Dari hasil tabulasi diatas, terlihat bahwa nilai p = 0,000 < 0,0042 ditemukan pada peubah bebas kategorik X_8 dan X_{11}. Dari nilai statistik uji χ^2 terbesar diketahui bahwa nilai p terkecil adalah peubah bebas kategorik X_{11}. Oleh karena itu, peubah tersebut dipilih sebagai peubah penyekat.

Peubah X_{11} adalah peubah kategorik dengan 4 kategori, sehingga untuk mendapatkan titik penyekat, peubah ini ditransformasi terlebih dahulu menjadi peubah numerik. Langkah-langkah transformasi adalah sebagai berikut :

1. Transformasikan masing-masing nilai x_{11} ke vektor peubah *dummy* 4 dimensi

   $\boldsymbol{v} = (v_1, v_2, v_3, v_4)'$,

   dengan $v_l = \begin{cases} 1 & x = b_l \\ 0 & x \neq b_l \end{cases}$, $l = 1, 2, 3, 4$

   Dengan demikian,

- jika $x_{11} = 0$ maka $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

- jika $x_{11} = 1$ maka $\boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$

- jika $x_{11} = 2$ maka $\boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$

- jika $x_{11} = 3$ maka $\boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

Berikut ini adalah tabel tabulasi dari peubah kategorik X_{11}

Tabel 3.3.7 Tabel Tabulasi X_{11}

Y	X_{11}				Total
	0	1	2	3	
0	25	4	0	2	31
1	28	28	20	11	87
Total	53	32	20	13	118

2. Cari nilai tengah untuk X_{11}

$$\overline{\boldsymbol{v}} = \frac{\sum_{l=1}^{4} f_l \boldsymbol{v}_l}{N_t} = \frac{53\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + 32\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + 20\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + 13\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}}{118}$$

$$= \begin{pmatrix} 53 \\ 32 \\ 20 \\ 13 \end{pmatrix} / 118 = \begin{pmatrix} 53/118 \\ 32/118 \\ 20/118 \\ 13/118 \end{pmatrix}$$

$$\overline{\boldsymbol{v}}^{(k)} = \frac{\sum_{l=1}^{4} n_l \boldsymbol{v}_l}{N_{k,t}}$$

$$\overline{\boldsymbol{v}}^{(0)} = \frac{25\begin{pmatrix}1\\0\\0\\0\end{pmatrix} + 4\begin{pmatrix}0\\1\\0\\0\end{pmatrix} + 0\begin{pmatrix}0\\0\\1\\0\end{pmatrix} + 2\begin{pmatrix}0\\0\\0\\1\end{pmatrix}}{31} = \frac{\begin{pmatrix}25\\4\\0\\2\end{pmatrix}}{31} = \begin{pmatrix}25/31\\4/31\\0/31\\2/31\end{pmatrix}$$

$$\overline{\boldsymbol{v}}^{(1)} = \frac{28\begin{pmatrix}1\\0\\0\\0\end{pmatrix} + 28\begin{pmatrix}0\\1\\0\\0\end{pmatrix} + 20\begin{pmatrix}0\\0\\1\\0\end{pmatrix} + 11\begin{pmatrix}0\\0\\0\\1\end{pmatrix}}{87} = \frac{\begin{pmatrix}28\\28\\20\\11\end{pmatrix}}{87} = \begin{pmatrix}28/87\\28/87\\20/87\\11/87\end{pmatrix}$$

3. Kalkulasikan nilai berikut dalam bentuk matriks 4 x 4

$$\boldsymbol{B} = \sum_{k=1}^{K} N_{k,t}\left(\overline{\boldsymbol{v}}^{(k)} - \overline{\boldsymbol{v}}\right)\left(\overline{\boldsymbol{v}}^{(k)} - \overline{\boldsymbol{v}}\right)'$$

$$= 31\left(\overline{\boldsymbol{v}}^{(0)} - \overline{\boldsymbol{v}}\right)\left(\overline{\boldsymbol{v}}^{(0)} - \overline{\boldsymbol{v}}\right)' + 87\left(\overline{\boldsymbol{v}}^{(1)} - \overline{\boldsymbol{v}}\right)\left(\overline{\boldsymbol{v}}^{(1)} - \overline{\boldsymbol{v}}\right)'$$

$$= 31\left(\begin{pmatrix}25/31\\4/31\\0/31\\2/31\end{pmatrix} - \begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)\left(\begin{pmatrix}25/31\\4/31\\0/31\\2/31\end{pmatrix} - \begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)'$$

$$+ 87\left(\begin{pmatrix}28/87\\28/87\\20/87\\11/87\end{pmatrix} - \begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)\left(\begin{pmatrix}28/87\\28/87\\20/87\\11/87\end{pmatrix} - \begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)'$$

$$= \begin{pmatrix}5{,}36769 & -2{,}13558 & -2{,}54627 & -0{,}685850\\ -2{,}13558 & 0{,}84966 & 1{,}01305 & 0{,}272871\\ -2{,}54627 & 1{,}01305 & 1{,}20787 & 0{,}325346\\ -0{,}68585 & 0{,}27287 & 0{,}32535 & 0{,}087634\end{pmatrix}$$

$$\boldsymbol{T} = \sum_{l=1}^{4} f_l(\boldsymbol{v}_l - \overline{\boldsymbol{v}})(\boldsymbol{v}_l - \overline{\boldsymbol{v}})'$$

$$= 53(\boldsymbol{v}_1 - \overline{\boldsymbol{v}})(\boldsymbol{v}_1 - \overline{\boldsymbol{v}})' + 32(\boldsymbol{v}_2 - \overline{\boldsymbol{v}})(\boldsymbol{v}_2 - \overline{\boldsymbol{v}})'$$

$$+20(\boldsymbol{v}_3 - \overline{\boldsymbol{v}})(\boldsymbol{v}_3 - \overline{\boldsymbol{v}})' + 13(\boldsymbol{v}_4 - \overline{\boldsymbol{v}})(\boldsymbol{v}_4 - \overline{\boldsymbol{v}})'$$

$$= 53\left(\begin{pmatrix}1\\0\\0\\0\end{pmatrix}-\begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)\left(\begin{pmatrix}1\\0\\0\\0\end{pmatrix}-\begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)'$$

$$+ 32\left(\begin{pmatrix}0\\1\\0\\0\end{pmatrix}-\begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)\left(\begin{pmatrix}0\\1\\0\\0\end{pmatrix}-\begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)'$$

$$+ 20\left(\begin{pmatrix}0\\0\\1\\0\end{pmatrix}-\begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)\left(\begin{pmatrix}0\\0\\1\\0\end{pmatrix}-\begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)'$$

$$+ 13\left(\begin{pmatrix}0\\0\\0\\1\end{pmatrix}-\begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)\left(\begin{pmatrix}0\\0\\0\\1\end{pmatrix}-\begin{pmatrix}53/118\\32/118\\20/118\\13/118\end{pmatrix}\right)'$$

$$= \begin{pmatrix}29{,}1949 & -14{,}3729 & -8{,}9831 & -5{,}8390\\ -14{,}3729 & 23{,}3220 & -5{,}4237 & -3{,}5254\\ -8{,}9831 & -5{,}4237 & 16{,}6102 & -2{,}2034\\ -5{,}8390 & -3{,}5254 & -2{,}2034 & 11{,}5678\end{pmatrix}$$

4. Lakukan SVD terhadap $\boldsymbol{T}$ menjadi $\boldsymbol{T} = \boldsymbol{QDQ}'$,

   dengan :

   $\boldsymbol{Q}$ adalah matriks orthogonal yang kolomnya merupakan vektor eigen dari $\boldsymbol{T}'\boldsymbol{T}$

   $\boldsymbol{D}$ adalah matriks diagonal yang merupakan akar dari nilai eigen dari $\boldsymbol{T}'\boldsymbol{T}$, $d_1 \geq d_2 \geq d_3 \geq 0$.

   Diperoleh :

$$\boldsymbol{Q} = \begin{pmatrix}-0{,}804043 & 0{,}276342 & 0{,}164768 & 0{,}500001\\ 0{,}567760 & 0{,}615426 & 0{,}221132 & 0{,}500000\\ 0{,}158437 & -0{,}717234 & 0{,}458774 & 0{,}500000\\ 0{,}077848 & -0{,}174535 & -0{,}844676 & 0{,}499999\end{pmatrix}$$

$$\boldsymbol{D} = \begin{pmatrix} 41{,}679 & 0 & 0 & 0 \\ 0 & 24{,}189 & 0 & 0 \\ 0 & 0 & 14{,}826 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

5. Diperoleh matriks $\boldsymbol{D}^{-\frac{1}{2}} = \begin{pmatrix} 0{,}1549 & 0 & 0 & 0 \\ 0 & 0{,}2033 & 0 & 0 \\ 0 & 0 & 0{,}2597 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

6. Lakukan SVD dari $\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{Q}'\boldsymbol{B}\boldsymbol{Q}\boldsymbol{D}^{-\frac{1}{2}}$, diperoleh vektor $\boldsymbol{a}$ yang merupakan vektor eigen yang berpadanan dengan nilai eigen terbesar :

$$\boldsymbol{a} = \begin{pmatrix} 0{,}906268 \\ -0{,}420317 \\ 0{,}044848 \\ 0 \end{pmatrix}, \quad \boldsymbol{a}' = (0{,}906268 \quad -0{,}420317 \quad 0{,}044848 \quad 0)$$

7. Diperoleh koordinat diskriminan terbesar dari $\boldsymbol{v}$, yaitu :

$$\xi = \boldsymbol{a}'\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{Q}'\boldsymbol{v}$$

$$= (-0{,}134567 \quad 0{,}0296899 \quad 0{,}0888728 \quad 0{,}0160044)\,\boldsymbol{v}$$

Dengan demikian, diperoleh :

❖ untuk $x_{11} = 0$ maka

$$\xi_1 = (-0{,}134567 \quad 0{,}0296899 \quad 0{,}0888728 \quad 0{,}0160044)\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$= -0{,}134567,$$

❖ untuk $x_{11} = 1$ maka

$$\xi_2 = (-0{,}134567 \quad 0{,}0296899 \quad 0{,}0888728 \quad 0{,}0160044)\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$= 0{,}0296899,$$

❖ untuk $x_{11} = 2$ maka

$$\xi_3 = (-0{,}134567 \quad 0{,}0296899 \quad 0{,}0888728 \quad 0{,}0160044)\begin{pmatrix}0\\0\\1\\0\end{pmatrix}$$

$$= 0{,}0888728,$$

❖ untuk $x_{11} = 3$ maka

$$\xi_4 = (-0{,}134567 \quad 0{,}0296899 \quad 0{,}0888728 \quad 0{,}0160044)\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

$$= 0{,}0160044.$$

Dengan demikian diperoleh tabel tabulasi X_{11} yang baru seperti pada Tabel 3.3.8

Tabel 3.3.8 Tabel Tabulasi X_{11} Hasil Transformasi

Y	ξ				Total
	−0,134567 (X_{11} = 0)	0,0160044 (X_{11} = 3)	0,0296899 (X_{11} = 1)	0,0888728 (X_{11} = 2)	
0	25	2	4	0	31
1	28	11	28	20	87
Total	53	13	32	20	118

Pada peubah X_{11} dilakukan analisis diskriminan kuadratik untuk memperoleh titik penyekat. Dari data diperoleh :

$\bar{x}_0 = -0{,}10366$ $\qquad$ $\bar{x}_1 = -0{,}01130$

$s_0^2 = (0{,}06420)^2$ $\qquad$ $s_1^2 = (0{,}08901)^2$

$\mathrm{P}(0|t) = 0{,}207$ $\qquad$ $\mathrm{P}(1|t) = 0{,}725$

Titik penyekat adalah akar dari persamaan $ax^2 + bx + c = 0$.

Dengan :

$$a = s^2{}_0 - s^2{}_1 = (0{,}06420)^2 - (0{,}08901)^2 = -0{,}00380114$$

$$b = 2(\bar{x}_0\, s^2{}_1 - \bar{x}_1\, s^2{}_0)$$

$$= 2(-0{,}10366.(0{,}08901)^2 - (-0{,}01130)(0{,}06420)^2) = -0{,}0015494$$

$$c = (\bar{x}_1\, s_0)^2 - (\bar{x}_0\, s_1)^2 + 2s^2{}_0 s^2{}_1 \ln\left\{\frac{p(0|t)s^2{}_1}{p(1|t)s^2{}_0}\right\}$$

$$= (-0{,}01130.\ 0{,}06420)^2 - (-0{,}10366.\ 0{,}08901)^2$$

$$+ 2(0{,}06420)^2.(\ 0{,}010222)^2 \ln\left\{\frac{0{,}207.(0{,}08901)^2}{0{,}725.(0{,}06420)^2}\right\} = -0{,}0001239,$$

diperoleh persamaan (-0,00380114) x^2 - 0,0015494 x - 0,0001239 = 0.

Akar dari persamaan di atas adalah :

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

$$= \frac{0{,}0015494 \pm \sqrt{(-0{,}0015494)^2 - 4.(-0{,}00380114).(-\ 0{,}0001239)}}{2(-0{,}00380114)}$$

$$= \frac{0{,}0015494 \pm 0{,}00072}{-0{,}00760228}$$

$$d_1 = \frac{0{,}0015494 + 0{,}00072}{-0{,}00760228} = -0{,}29837$$

$$d_2 = \frac{0{,}0015494 - 0{,}000860214}{-0{,}00760228} = -0{,}1092.$$

Dari dua akar persamaan ini, diambil akar yang paling mendekati $\bar{x}_0 = -0{,}10366$ yaitu $d = -0{,}1092$. Disimpulkan bahwa X_{11} memotong di nilai $\xi = -0{,}134567$. Dengan demikian, simpul t_1 disekat menjadi dua simpul , yaitu simpul t_3 dengan $\xi \leq -0{,}134567$ dan simpul t_4 dengan $\xi > -0{,}134567$.

Dengan kata lain, simpul tersebut disekat menjadi dua simpul, yaitu simpul t_3 dengan $x_{11} \in \{0\}$ dan simpul t_4 dengan $x_{11} \in \{1, 2, 3\}$ Simpul t_3 terdiri dari 53 pengamatan dari dua kategori, yaitu pasien berpenyakit jantung (28 pengamatan) dan pasien tidak berpenyakit jantung (25 pengamatan). Simpul t_4 terdiri dari 65 pengamatan dari dua kategori, yaitu pasien berpenyakit jantung (59 pengamatan) dan pasien tidak berpenyakit jantung (6 pengamatan). Proses penyekatan dilanjutkan karena simpul t_3 dan t_4 merupakan simpul dalam.

Selanjutnya dilakukan langkah yang sama untuk mendapatkan peubah penyekat pada simpul t_2. Dari uji yang dilakukan diperoleh nilai χ^2 dan nilai F sebagai berikut :

Tabel 3.3.9 Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_2

Peubah	Nilai Statistik Uji	Nilai p
X_1	F = 13,20	p = 0,000
X_2	χ^2 = 7,74	p = 0,005
X_3	F = 0,58	p = 0,449
X_4	F = 2,84	p = 0,094
X_5	χ^2 = 0,04	p = 0,842
X_6	χ^2 = 6,21	p = 0,045
X_7	F = 26,55	p = 0,000
X_8	χ^2 = 13,70	p = 0,000
X_9	F = 16,42	p = 0,000
X_{10}	χ^2 = 11,87	p = 0,003
X_{11}	χ^2 = 29,20	p = 0,000

Dari hasil tabulasi diatas, terlihat bahwa nilai p = 0,000 < 0,0042 ditemukan pada beberapa peubah bebas. Nilai p terkecil akan diperoleh pada peubah dengan nilai statistik uji Anova F dan uji χ^2 terbesar, yaitu pada peubah X_{11} dan peubah X_7. Karena keterbatasan alat hitung maka X_7 dianggap memiliki nilai p terkecil dengan melihat pohon klasifikasi dengan algoritma QUEST yang dilakukan dengan software SPSS. Oleh karena itu, peubah tersebut dipilih sebagai peubah penyekat.

Pada peubah X_7 dilakukan analisis diskriminan kuadratik untuk memperoleh titik penyekat. Dari data diperoleh :

$\bar{x}_0 = 159{,}5$ $\qquad$ $\bar{x}_1 = 138{,}7$

$s_0^2 = (18{,}98)^2$ $\qquad$ $s_1^2 = (25{,}59)^2$

$P(0|t) = 0{,}793$ $\qquad$ $P(1|t) = 0{,}275$

Titik penyekat adalah akar dari persamaan $ax^2 + bx + c = 0$.

Dengan :

$$a = s^2{}_0 - s^2{}_1 = (18{,}98)^2 - (25{,}59)^2 = -295$$

$$b = 2(\bar{x}_0\, s^2{}_1 - \bar{x}_1\, s^2{}_0)$$

$$= 2(159{,}5.(25{,}59)^2 - 138{,}7.(18{,}98)^2) = 108966$$

$$c = (\bar{x}_1\, s_0)^2 - (\bar{x}_0\, s_1)^2 + 2s^2{}_0 s^2{}_1 \ln\left\{\frac{p(0|t)s^2{}_1}{p(1|t)s^2{}_0}\right\}$$

$$= (138{,}7.18{,}98)^2 - (159{,}5.25{,}59)^2$$

$$+2(360{,}24)^2(654{,}85)^2 \ln\left\{\frac{0{,}793.(25{,}59)^2}{0{,}275.(18{,}98)^2}\right\} = -8947475\,,$$

diperoleh persamaan $(-295)x^2 + 108966x - 8947475 = 0$

Akar dari persamaan di atas adalah :

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

$$= \frac{-108966 \pm \sqrt{(108966)^2 - 4.(-295).(-8947475)}}{2(-295)}$$

$$= \frac{-108966 \pm 36463}{-590}$$

$$d_1 = \frac{-108966 + 36463}{-590} = 123{,}05$$

$$d_2 = \frac{-108966 - 36463}{-590} = 246{,}82.$$

Dari dua akar persamaan ini, diambil akar yang paling mendekati $\bar{x}_0 = 159{,}5$ yaitu $d = 123{,}05$. Disimpulkan bahwa X_7 memotong di nilai $d = 123{,}05$. Dengan demikian, simpul t_2 disekat menjadi dua simpul, yaitu

simpul t_7 dengan $d \leq 123{,}05$ dan simpul t_8 dengan $d > 123{,}05$. Simpul t_7 terdiri dari 18 pengamatan dari dua kategori, yaitu pasien berpenyakit jantung (9 pengamatan) dan pasien tidak berpenyakit jantung (9 pengamatan). Simpul t_8 terdiri dari 134 pengamatan dari dua kategori, yaitu pasien berpenyakit jantung (24 pengamatan) dan pasien tidak berpenyakit jantung (110 pengamatan).

Pada data ini digunakan pilihan dengan banyaknya data pada simpul dalam paling sedikit 50, sedangkan banyaknya data pada simpul akhir paling sedikit 20. Pada simpul t_2 yang disekat menjadi simpul t_7 dan t_8 , banyaknya data pada simpul t_7 adalah 18 pengamatan sehingga t_7 tidak dapat dijadikan simpul dalam maupun simpul akhir. Oleh karena itu, simpul t_2 tidak dapat disekat lagi. Sehingga pada simpul t_2 penyekatan dihentikan.

Selanjutnya dilakukan langkah yang sama untuk mendapatkan peubah penyekat pada simpul t_3. Dari uji yang dilakukan diperoleh nilai χ^2 dan nilai F sebagai berikut :

Tabel 3.3.10 Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_3

Peubah	Nilai Statistik Uji	Nilai p
X_1	F = 2,91	p = 0,094
X_2	χ^2 = 0,12	p = 0,736
X_3	F = 0,43	p = 0,514
X_4	F = 0,22	p = 0,640
X_5	χ^2 = 0,31	p = 0,580
X_6	χ^2 = 1,55	p = 0,213
X_7	F = 6,19	p = 0,016
X_8	χ^2 = 12,21	p = 0,000
X_9	F = 5,80	p = 0,020
X_{10}	χ^2 = 3,16	p = 0,206
X_{12}	χ^2 = 4,81	p = 0,028

Dari hasil tabulasi diatas, dapat terlihat bahwa peubah bebas X_8 memiliki nilai p terkecil, yaitu 0,000 < 0,0042. Oleh karena itu, peubah tersebut dipilih menjadi peubah penyekat. Diperoleh hasil tabulasi X_8 :

Tabel 3.3.11 Tabel Tabulasi X_8 untuk Simpul t_3

Y	X_8		Total
	0	1	
0	20	5	25
1	9	19	28
Total	29	24	53

Dari tabel di atas, dapat dilihat bahwa peubah ini merupakan peubah kategorik dengan dua kategori sehingga titik penyekat dapat diperoleh langsung. Dengan demikian, simpul t_3 disekat menjadi dua simpul, yaitu simpul t_5 dan simpul t_6. Dengan kata lain, simpul tersebut disekat menjadi dua simpul, yaitu simpul t_5 dengan $x_8 \in \{1\}$ dan simpul t_6 dengan $x_8 \in \{0\}$. Simpul t_5 terdiri dari 24 pengamatan dari dua kategori, yaitu pasien berpenyakit jantung (19 pengamatan) dan pasien tidak berpenyakit jantung (5 pengamatan). Simpul t_6 terdiri dari 29 pengamatan dari dua kategori, yaitu pasien berpenyakit jantung (9 pengamatan) dan pasien tidak berpenyakit jantung (20 pengamatan). Simpul t_5 dan t_6 memiliki banyak data besar dari 20 dan kecil dari 50. Sehingga pada simpul t_5 dan t_6 penyekatan dihentikan.

Selanjutnya dilakukan langkah yang sama untuk mendapatkan peubah penyekat pada simpul t_4. Dari uji yang dilakukan diperoleh nilai χ^2 dan nilai F sebagai berikut :

Tabel 3.3.12 Hasil Uji Statistika Pemilihan Peubah Penyekat Simpul t_4

Peubah	Nilai Statistik Uji	Nilai p
X_1	F = 0,05	p = 0,826
X_2	χ^2 = 0,93	p = 0,335
X_3	F = 3,25	p = 0,076
X_4	F = 2,16	p = 0,147
X_5	χ^2 = 1,27	p = 0,260
X_6	χ^2 = 7,71	p = 0,005
X_7	F = 0,72	p = 0,398
X_8	χ^2 = 1,12	p = 0,290
X_9	F = 4,97	p = 0,029
X_{10}	χ^2 = 5,66	p = 0,059
X_{11}	χ^2 = 3,03	p = 0,220
X_{12}	χ^2 = 0,80	p = 0,372

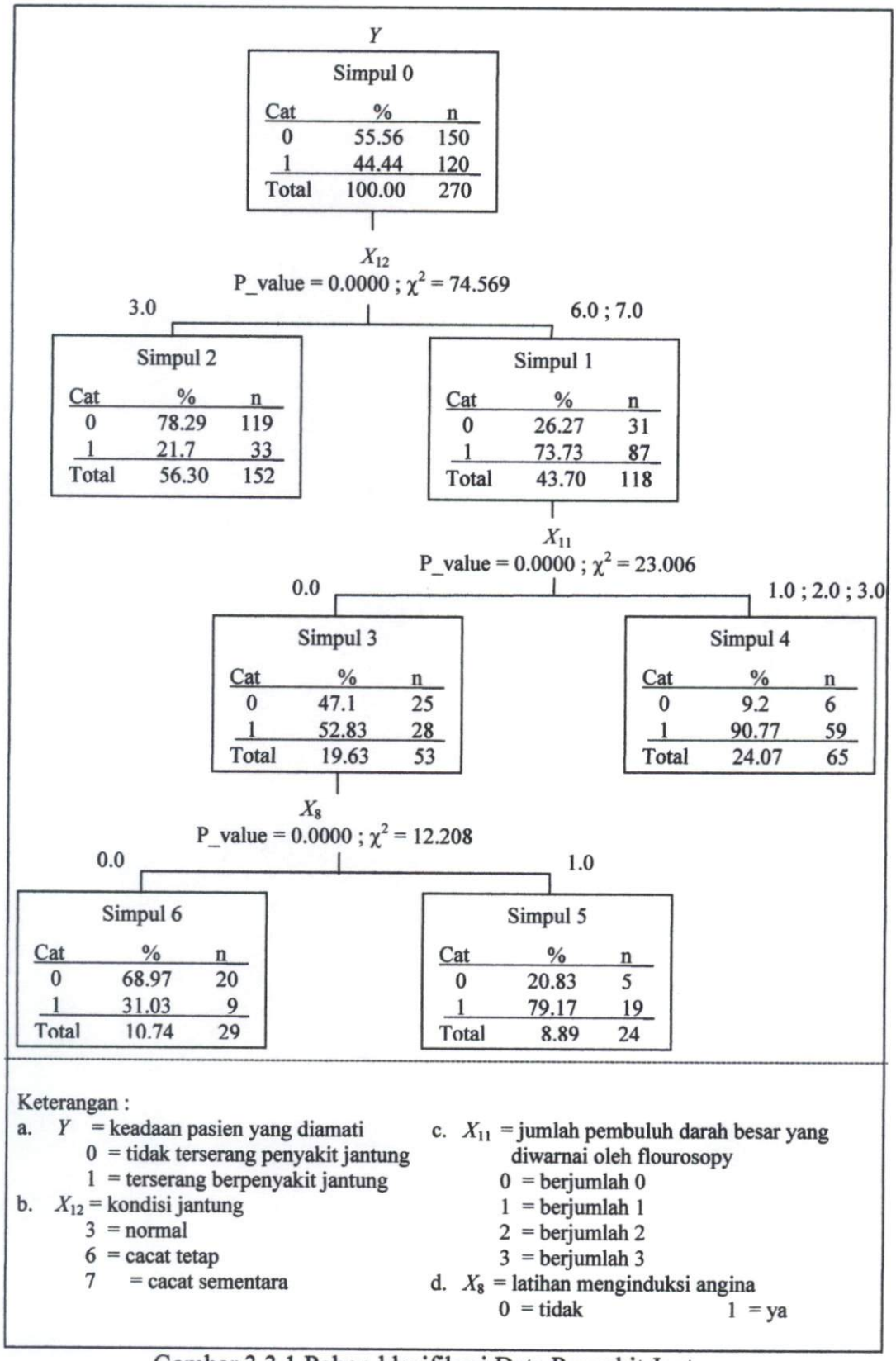
Dari hasil tabulasi diatas, dapat terlihat bahwa peubah bebas X_6 memiliki nilai p terkecil, yaitu 0,005 > 0,0042. Oleh karena itu, untuk peubah numerik dilakukan uji Levene F. Dari uji yang dilakukan diperoleh nilai χ^2 dan nilai Levene sebagai berikut :

Tabel 3.3.13 Hasil Uji Statistika Levene F

Peubah	Nilai Statistik Uji Levene F	Nilai p
X_1	1,021	p = 0,316
X_3	0,006	p = 0,940
X_4	0,000	p = 0,984
X_7	0,662	p = 0,419
X_9	1,640	p = 0,205

Bila M_2 = 5, didapat $\alpha/(M_1 + M_2)$ = 0,05/(12+5) = 0,00294. Dari tabel di atas terlihat bahwa nilai p terkecil, yaitu 0,205 > 0,00294, sehingga simpul t_4 tidak disekat.

Berdasarkan langkah-langkah sebelumnya diperoleh pohon klasifikasi sebagai berikut :



Gambar 3.3.1 Pohon klasifikasi Data Penyakit Jantung

Dari hasil klasifikasi di atas, dapat dilihat bahwa faktor yang mempengaruhi seseorang menderita penyakit jantung adalah kondisi jantung (X_{12}), jumlah pembuluh darah besar yang diwarnai oleh flourosopy (X_{11}) dan latihan menginduksi angina (X_8). Terdapat tujuh simpul yang terdiri dari satu simpul induk (simpul 0), dua simpul dalam (simpul 1 dan simpul 3) dan empat simpul akhir.

Pada simpul 2, 4, 5, dan 6 proses penyekatan dihentikan, sehingga menjadi empat simpul akhir (empat kelas). Dalam menduga respon simpul akhir, dipilih peubah respon yang presentasinya terbesar. Simpul 2 dan simpul 6 untuk respon dengan kategori 0 yaitu tidak terserang penyakit jantung, sedangkan simpul 4 dan simpul 5 untuk respon dengan kategori 1 yaitu terserang penyakit jantung. Dapat dikatakan terdapat empat kelas dengan pengelompokan sebagai berikut :

1. Kelas pertama

   Kelas pertama merupakan pasien yang kondisi jantungnya normal. Pasien yang termasuk dalam kelas ini cenderung tidak terserang penyakit jantung.

2. Kelas kedua

   Kelas kedua merupakan pasien yang kondisi jantungnya cacat atau cacat sementara dan memiliki jumlah pembuluh darah besar yang diwarnai oleh flourosopy 1 sampai 3. Pasien yang termasuk dalam kelas ini cenderung terserang penyakit jantung.

3. Kelas ketiga

   Kelas ketiga merupakan pasien yang kondisi jantungnya cacat tetap atau cacat sementara, tidak memiliki jumlah pembuluh darah besar yang diwarnai oleh

flourosopy dan apabila latihan dapat menginduksi angina. Pasien yang termasuk dalam kelas ini cenderung terserang penyakit jantung.

4. Kelas keempat

   Kelas keempat merupakan pasien yang kondisi jaantungnya cacat tetap atau cacat sementara, tidak memiliki jumlah pembuluh darah besar yang diwarnai oleh flourosopy dan apabila latihan tidak dapat menginduksi angina. Pasien yang termasuk dalam kelas ini cenderung tidak terserang penyakit jantung.

Dapat disimpulkan, kondisi jantung merupakan faktor yang paling signifikan mempengaruhi seseorang menderita penyakit jantung dibandingkan dengan jumlah pembuluh darah besar yang diwarnai oleh flourosopy dan latihan menginduksi angina.

BAB IV

PENUTUP

4.1 Kesimpulan

Metode QUEST merupakan metode yang menghasilkan pohon klasifikasi biner. Metode ini merupakan metode yang menerapkan pemilihan peubah penyekat dan penentuan titik penyekat dilakukan secara terpisah. Pemilihan peubah penyekat dapat dilakukan dengan melakukan uji kebebasan Khi–Kuadrat untuk peubah kategorik dan uji Anova F untuk peubah numerik. Penentuan titik penyekat dilakukan dengan menerapkan analisis diskriminan kuadratik.

Metode ini dapat diterapkan pada bidang kesehatan. Salah satunya untuk mengetahui klasifikasi seseorang menderita penyakit jantung berdasarkan faktor-faktor yang mempengaruhinya. Dapat disimpulkan terdapat empat klasifikasi dengan pengelompokan sebagai berikut :

1. Kelas pertama

   Kelas pertama merupakan pasien yang kondisi jantungnya normal. Pasien yang termasuk dalam kelas ini cenderung tidak terserang penyakit jantung.

2. Kelas kedua

   Kelas kedua merupakan pasien yang kondisi jantungnya cacat atau cacat sementara dan memiliki jumlah pembuluh darah besar yang diwarnai oleh flourosopy 1 sampai 3. Pasien yang termasuk dalam kelas ini cenderung terserang penyakit jantung.

3. Kelas ketiga

Kelas ketiga merupakan pasien yang kondisi jantungnya cacat tetap atau cacat sementara, tidak memiliki jumlah pembuluh darah besar yang diwarnai oleh flourosopy dan apabila latihan dapat menginduksi angina. Pasien yang termasuk dalam kelas ini cenderung terserang penyakit jantung.

4. Kelas keempat

Kelas keempat merupakan pasien yang kondisi jaantungnya cacat tetap atau cacat sementara, tidak memiliki jumlah pembuluh darah besar yang diwarnai oleh flourosopy dan apabila latihan tidak dapat menginduksi angina. Pasien yang termasuk dalam kelas ini cenderung tidak terserang penyakit jantung.

4.2 Saran

Pada kesempatan ini, penerapan metode QUEST dilakukan dengan pengolahan data peubah bebas kategorik dan numerik tanpa pemangkasan. Metode ini dapat dilanjutkan dengan pemangkasan. Metode ini juga dapat dibandingkan dengan metode pohon klasifikasi biner lainnya.

DAFTAR PUSTAKA

[1] Anonim. 2009. Jurnal Ilmiah Geomatika. Badan Koordinasi Survei dan Pemetaan Nasional. *http://repository.ipb.ac.id/bitstream/handle/123456789/37892/GEOMATIKA%20B-5.pdf?sequence=1* Mei 2011.

[2] Anonim. QUEST Algorithm. http://support.spss.com/productsext/spss/documentation/statistics/algorithms/14.0/TREE-QUEST.pdf Juni 2011

[3] Black CL, Merz CJ. 1998. UCI Repository of machine learning database http://www.ics.uci.edu/~mlearn/MLRepository.html. Juni 2011.

[4] Breiman, et.al. 1984. *Classification and Regression Tree*. Chapman & Hall, New York.

[5] Faridhan, Y.E. 2003. Metode Klasifikasi Berstruktur Pohon Dengan Algoritma CRUISE, QUEST, dan CHAID. *http://repository.ipb.ac.id/bitstream/handle/123456789/8036/2003yef.pdf?sequence =4* Mei 2011.

[6] Lestari, R.D. 2005. Identifikasi Rumah Tangga Miskin di Kota Padang Menggunakan Metode CHAID. *Skripsi-S1*, tidak diterbitkan Sembiring, R. K. *Analisis Regresi*. ITB. Bandung.

[7] Loh W-Y, Shih Y-S. 1997. Split Selection Methods for Classfication Trees. Stastistica Sinica7 ; 815-840. http://www.stat.wisc.edu/~loh/ Juni 2011.

[8] Martono, N. 2010. *Statistik Sosial Teori dan Aplikasi Program SPSS*. Gava Media. Yogyakarta.

[9] Rachmatin, D dan K. Sawitri. Tanpa tahun. Penerapan Prosedur Lachenbruch Pada Kasus Quadratic Discriminant Analysis. *http://file.upi.edu/Direktori/FPMIPA/JUR._PEND._MATEMATIKA/196909291994122-DEWI_RACHMATIN/MAKALAH_PROSIDING_SEMNAS_2009/Penerapan Prosedur_Lachenbruch.pdf* Mei 2011.

[10] Sartono, Bagus dan Syafitri, Utami Dyah. 2010. Metode Pohon Gabungan : Solusi Pilihan Untuk Mengatasi Kelemahan Pohon Regresi dan Pohon Klasifikasi Tunggal. Forum Statistika dan Komputasi. Bogor.

[11] Sembiring, R. K. Analisis Regresi. ITB. Bandung.

[12] Siegel, S. 1985. *Statistika Nonparametrik Untuk Ilmu-Ilmu Sosial*. PT Gramedia, Jakarta.

[13] Walpole, R. E. 1992. Pengantar Statistika Edisi Ke-3. PT Gramedia Pustaka Utama, Jakarta.

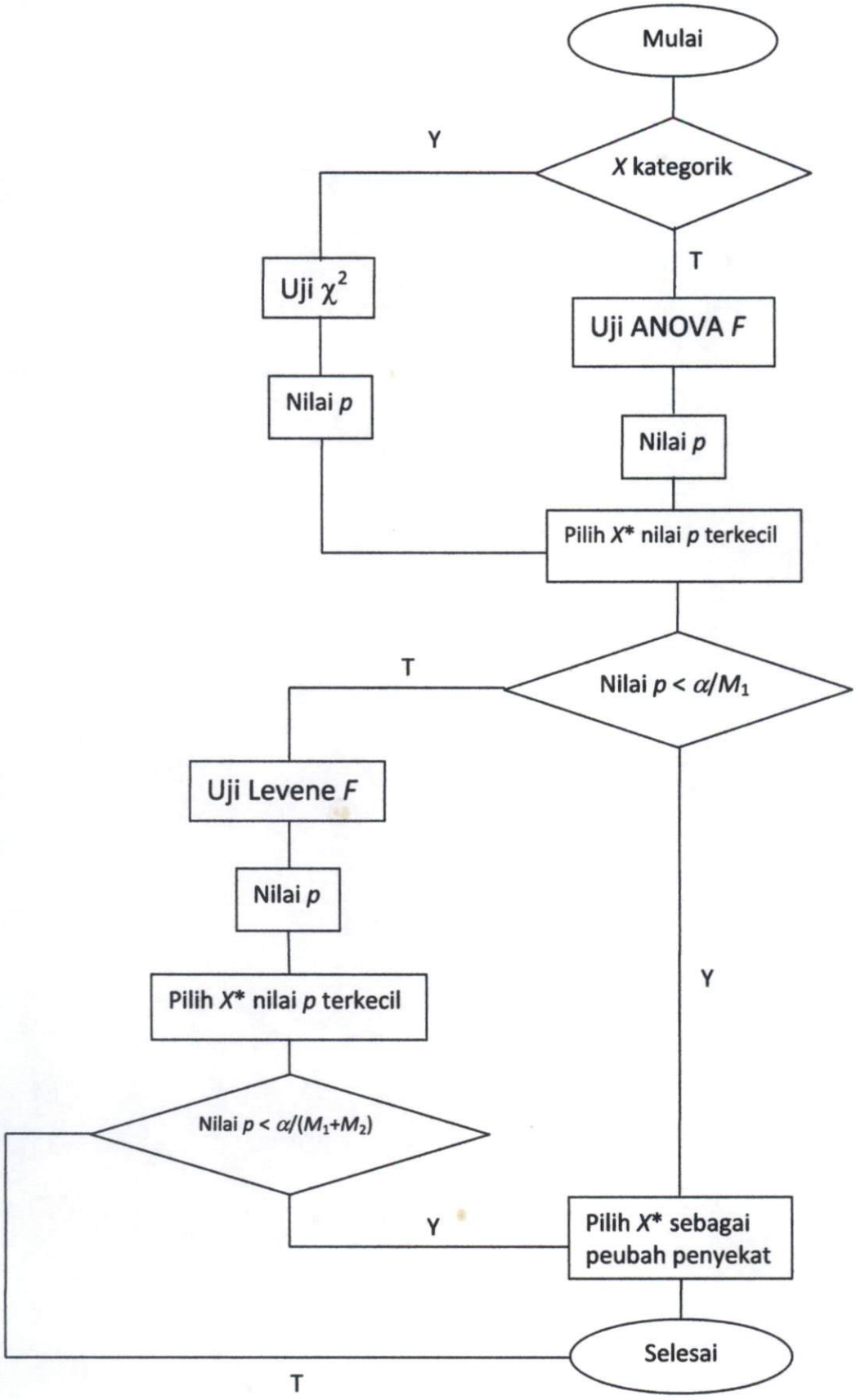
[14] Widagdo, K.A. 2010. Pembentukan Pohon Klasifikasi Biner Dengan Algoritma CART. *http://www.google.co.id/url?sa=t&source=web&cd=7&ved=0CEcQFjAG&url=http%3A%2F%2Feprints.undip.ac.id%2F20740%2F&rct=j&q=pohon%20klasifikasi%20&ei=7GzHTZmXK4iIuAOc1JiTAQ&usg=AFQjCNEqU2zN3TyqCWIRidZ-3w4YLaQv5g&cad=rja* Mei 2011.

Lampiran 1. Hasil SPSS 16.0

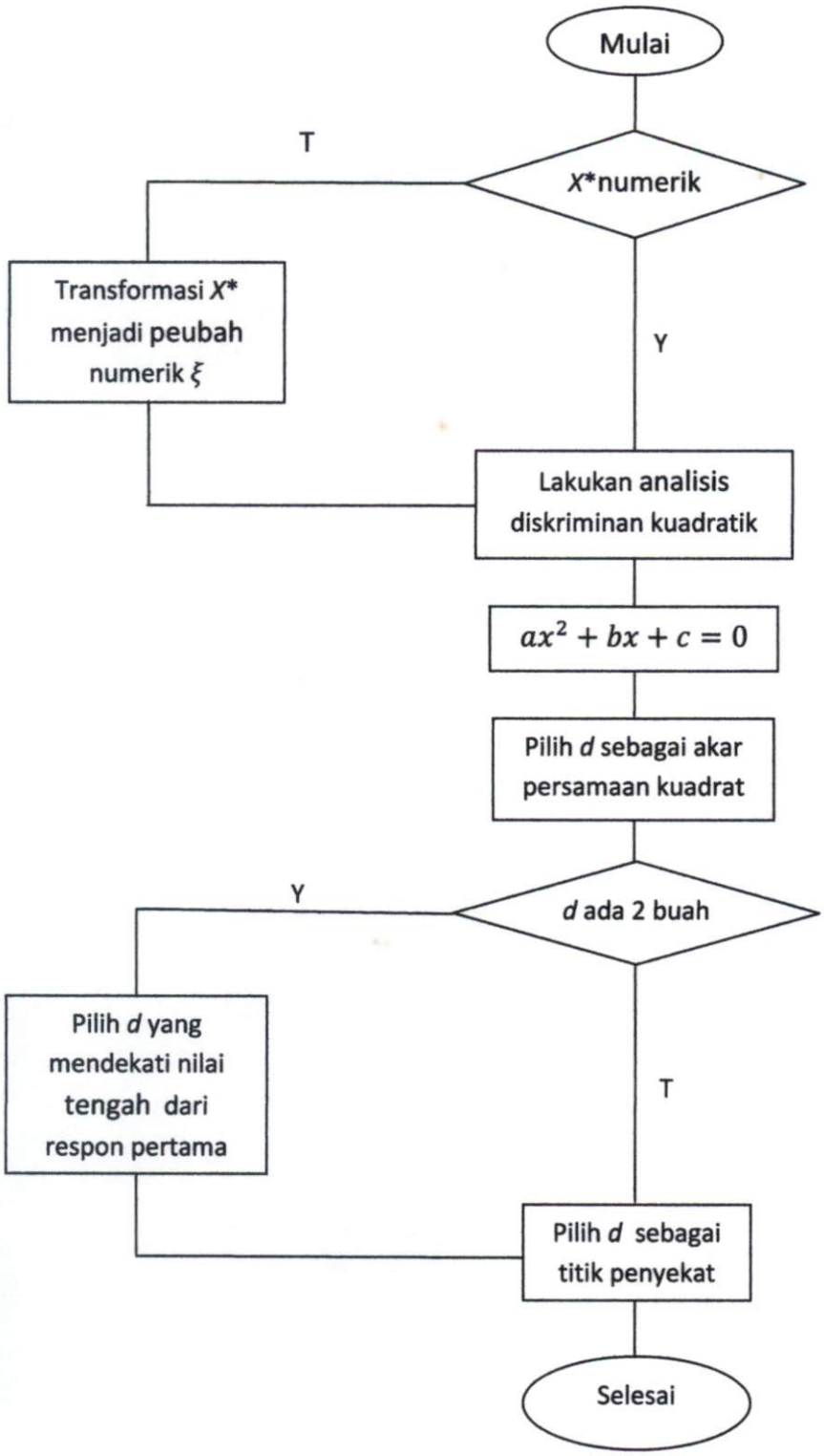
Model Summary

Specifications	Growing Method	QUEST
	Dependent Variable	Y
	Independent Variables	X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12
	Validation	None
	Maximum Tree Depth	5
	Minimum Cases in Parent Node	50
	Minimum Cases in Child Node	20
Results	Independent Variables Included	X12, X8, X7, X2, X9, X10, X11, X1, X4, X3
	Number of Nodes	7
	Number of Terminal Nodes	4
	Depth	3

Lampiran 2. Flowchart untuk algoritma pemilihan peubah penyekat



Lampiran 3. Flowchart untuk algoritma penentuan titik penyekat



RIWAYAT HIDUP PENULIS

Penulis bernama Yona Malani, dilahirkan di Kota Padang pada tanggal 10 Maret 1989. Anak keempat dari lima bersaudara, buah hati dari pasangan Nasri dan Zulma Herni. Penulis memulai pendidikannya di SD No. 28 Rawang Timur pada tahun 1996. Pada tahun 2001, penulis melanjutkan pendidikannya di SMP Negeri 20 Padang. Pada tahun 2004, penulis melanjutkan pendidikannya di SMA Negeri 2 Padang dan tamat pada tahun 2007. Pada tahun yang sama, penulis di terima menjadi mahasiswa Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Andalas melalui jalur Seleksi Penerimaan Mahasiswa Baru (SPMB).

Selama di bangku perkuliahan penulis pernah mengikuti magang di UKM FKI Rabbani Universitas Andalas tahun 2007. Penulis pernah menjadi pengurus HIMATIKA periode 2009/2010 pada Divisi Kesejahteraan dan Pengabdian Masyarakat dan penulis menjadi sekretaris acara Pekan Seni Bermatematika VII (PSB VII) se-Sumatera yang diadakan pada tanggal 24-27 Februari 2010 serta aktif di berbagai kegiatan HIMATIKA. Untuk syarat meraih gelar Sarjana Sains (S.Si) di Jurusan Matematika FMIPA UNAND, penulis pernah mengikuti Kuliah Kerja Nyata (KKN) di Nagari Lurah Ampalu, Kabupaten Padang Pariaman pada bulan Juli s/d Agustus 2010.