

METODE QUEST DAN CHAID PADA KLASIFIKASI KARAKTERISTIK NASABAH KREDIT

Nur Faiza^{§1}, I Wayan Sumarjaya², I Gusti Ayu Made Srinadi³

¹Jurusan Matematika, FMIPA, Universitas Udayana [Email: nurfaiza_0606@yahoo.com]

²Jurusan Matematika, FMIPA, Universitas Udayana [Email: sumarjaya@gmail.com]

³Jurusan Matematika, FMIPA, Universitas Udayana [Email: srinadiigustiyumade@yahoo.co.id]

[§]Corresponding Author

ABSTRACT

This aim of this research is to find out the classification results and to compare the magnitude of misclassification of QUEST and CHAID methods on the classification of customer of Adira Kredit Elektronik branch Denpasar. QUEST (Quick, Unbiased, Efficient Statistical Trees) and CHAID (Chi-squared Automatic Interaction Detection) are nonparametric methods that produce tree diagram which is easy to interpret. The QUEST and CHAID classification methods conclude that: 1) QUEST method produces three groups which predict customers into the current category, whereas CHAID method produces four groups which also predict customer into the current category; 2) both methods generate the biggest classification accuracy for customers that current category which share similar characteristics; 3) both methods also have the same degree of accuracy in classifying customer data Adira Kredit Elektronik branch Denpasar.

Keywords: Classification, QUEST, CHAID, customer credit status.

1. PENDAHULUAN

Klasifikasi merupakan pengelompokan secara sistematis suatu objek atau benda ke dalam golongan atau pola-pola tertentu berdasarkan kesamaan ciri. Masalah klasifikasi sering dijumpai dalam kehidupan sehari-hari, baik pada bidang industri, sosial, kesehatan maupun pendidikan.

Penelitian ini membahas metode klasifikasi pohon menggunakan metode QUEST dan CHAID dengan variabel berskala kategorik, dengan tujuan untuk mengetahui hasil klasifikasi dan membandingkan besarnya kesalahan klasifikasi dari metode QUEST dan CHAID pada klasifikasi data nasabah Adira Kredit Elektronik Cabang Denpasar.

Metode QUEST merupakan algoritma yang diperkenalkan oleh Loh dan Shih [3], sebagai salah satu bentuk metode pohon

klasifikasi yang menghasilkan pohon biner. QUEST diterapkan pada data dengan variabel respon kategorik dan variabel prediktor berupa variabel kategorik maupun numerik. QUEST merupakan modifikasi dari analisis diskriminan kuadrat. Analisis diskriminan kuadrat diterapkan pada proses penentuan simpul penyekat.

Algoritma QUEST dibagi menjadi tiga bagian yaitu, algoritma pemilihan variabel penyekat, algoritma penentuan simpul penyekat, dan algoritma penghentian pembentukan pohon [3].

Langkah-langkah algoritma pemilihan variabel penyekat dengan variabel kategorik adalah sebagai berikut: pertama, untuk setiap variabel prediktor kategorik, lakukan uji khi-kuadrat dan hitung p -value. Kedua, pilih variabel prediktor yang memiliki nilai p -

value terkecil atau yang paling signifikan sebagai variabel penyekat (Loh dan Shih [3]).

Setelah terpilih variabel penyekat, langkah selanjutnya adalah menentukan simpul penyekat. Apabila variabel penyekat yang terpilih berupa variabel kategorik dan terdiri lebih dari dua kategori, maka dilakukan transformasi ke variabel numerik dan selanjutnya diterapkan analisis diskriminan kudratik untuk mencari akar dari persamaan kuadrat $ax^2 + bx + c = 0$, dengan:

$$a = s_1^2 - s_2^2$$

$$b = 2(\bar{x}_1 s_2^2 - \bar{x}_2 s_1^2)$$

$$c = (\bar{x}_2 s_1)^2 - (\bar{x}_1 s_2)^2 + 2s_1^2 s_2^2 \log \left(\frac{P(1|t)s_2}{P(2|t)s_1} \right).$$

dengan \bar{x}_1 dan s_1^2 adalah rata-rata dan ragam variabel penyekat dari pengamatan variabel respon pertama, sedangkan \bar{x}_2 dan s_2^2 adalah rata-rata dan ragam variabel penyekat dari pengamatan variabel respon kedua [3]. Jika akar yang dihasilkan adalah akar tunggal, maka akar tersebut dipilih sebagai simpul penyekat. Tetapi jika akar yang dihasilkan terdiri dari 2 akar, maka akan dipilih akar yang mendekati rata-rata dari variabel respon pertama sebagai simpul penyekat.

Proses pemilihan simpul secara rekursif akan terhenti apabila salah satu dari aturan penghentian tercapai. Ada empat aturan penghentian pembentukan pohon yaitu: pertama, jika simpul penyekat menjadi murni yaitu apabila semua kasus masuk ke dalam salah satu kategori variabel respon. Kedua, jika semua kasus yang terdapat di dalam simpul memiliki nilai-nilai identik untuk tiap variabel prediktor. Kedua, jika kedalaman pohon sudah mencapai kedalaman pohon maksimum yang ditetapkan. Keempat, jika banyaknya kasus yang masuk ke dalam simpul sudah mencapai jumlah minimum yang ditetapkan.

Hasil analisis dari metode QUEST akan dibandingkan dengan metode pendahulunya, yaitu metode CHAID. CHAID merupakan algoritma yang diperkenalkan oleh Kass (1980), metode klasifikasi yang lebih dulu diterapkan dan menghasilkan pohon nonbiner (Kass [2]).

CHAID diterapkan pada data dengan variabel respon dan variabel prediktor berupa variabel kategorik.

CHAID membedakan variabel prediktor kategorik menjadi tiga bentuk yang berbeda, yaitu: monotonik, bebas, dan mengambang (Gallagher [1]).

Algoritma CHAID secara lengkap yaitu sebagai berikut (Kass [2]): Pertama, untuk setiap variabel prediktor, buat tabulasi silang kategori-kategori variabel prediktor dengan kategori-kategori variabel respon. Kedua, untuk setiap tabulasi silang yang diperoleh (susun subtabel yang berukuran $2 \times d$ yang mungkin, d adalah banyaknya kategori variabel respon) cari pasangan kategori variabel prediktor yang memiliki angka uji paling kecil. Jika angka uji tidak mencapai nilai kritis, gabungkan kedua kategori tersebut menjadi satu kategori campuran. Ulangi langkah ini sampai angka uji kategori campuran melampaui nilai kritis. Ketiga, untuk setiap kategori gabungan yang terbentuk dari 3 atau lebih kategori asal, temukan pemisahan biner yang memiliki angka uji yang paling besar (pemisahan berdasarkan tipe variabel prediktor). Jika angka uji ini lebih besar dari nilai kritis, buatlah pemisahan tersebut dan kembali ke langkah 2 di atas. Keempat, untuk setiap tabulasi silang variabel prediktor yang telah digabungkan secara optimal, pilih yang memiliki angka uji paling besar. Jika angka uji ini lebih besar dari nilai kritis, bagilah data menurut kategori-kategori yang telah digabungkan dari variabel prediktor yang telah dipilih. Pada tahap ini apabila terjadi pengurangan tabel kontingensi dari tabel asal maka gunakan uji koreksi Bonferoni dan jika tidak maka lakukan uji khi-kuadrat untuk uji signifikansi. Kelima, untuk setiap pembagian data yang belum dianalisis, kembali ke langkah 1.

Dalam menyeleksi apakah seseorang layak mendapatkan pinjaman kredit, dibutuhkan suatu analisis statistika salah satunya dengan metode klasifikasi. Dalam hal ini, metode klasifikasi digunakan untuk membagi suatu nasabah ke dalam kelompok yang lebih kecil kelompoknya berdasarkan status pembayaran kredit lancar atau

macet. Kelompok-kelompok yang terbentuk tersebut akan digunakan oleh Adira Kredit Elektronik Cabang Denpasar sebagai acuan untuk menentukan sasaran utama pemasaran dengan mempertimbangkan status kredit. Dengan penentuan kelompok sasaran kredit yang tepat, diharapkan risiko kredit macet bisa diminimalkan.

Adanya kesamaan karakteristik dari metode QUEST dan CHAID yaitu dapat diterapkan pada data variabel kategorik dan dalam pemilihan variabel penyekat menggunakan uji khi-kuadrat. Sehingga penulis tertarik untuk melakukan penelitian dengan metode QUEST dan CHAID yang akan diterapkan pada data nasabah Adira Kredit Elektronik Cabang Denpasar. Penelitian ini membandingkan hasil klasifikasi dari kedua metode tersebut untuk dapat dijadikan bahan analisis perusahaan dalam menentukan calon nasabah pada masa yang akan datang.

2. METODE PENELITIAN

Penelitian ini mengambil data nasabah Adira Kredit Elektronik Cabang Denpasar sebagai study kasus. Data diperoleh dari hasil pengisian formulir aplikasi permohonan kredit nasabah Adira Kredit Elektronik Cabang Denpasar yang sebelumnya sudah pernah mengajukan kredit dan melakukan pengajuan kembali pada bulan September sampai dengan Oktober 2014.

Adapun Variabel yang digunakan dalam penelitian ini berupa variabel kategori yaitu meliputi status kredit (Y), jenis kelamin (X_1), status pernikahan (X_2), jenis pekerjaan (X_3), dan status tempat tinggal (X_4), pendidikan terakhir (X_5), jabatan nasabah (X_6), lama angsuran (X_7), dan pendapatan per bulan (X_8).

Langkah-langkah yang dilakukan dalam metode penelitian ini adalah:

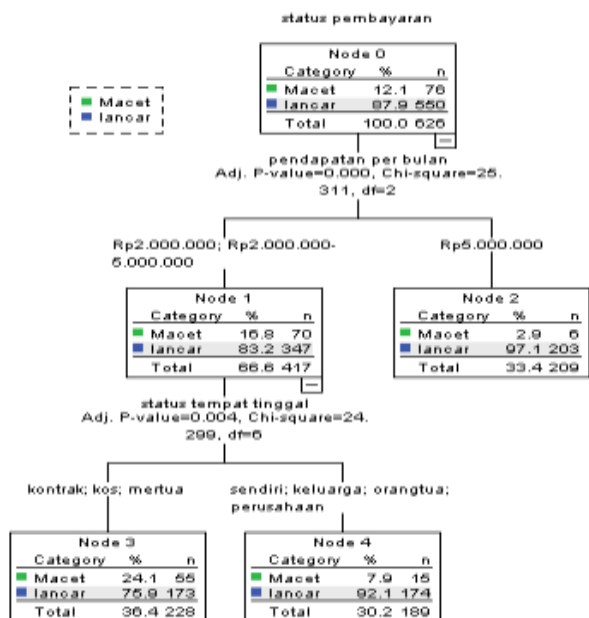
1. Mengumpulkan dan menyeleksi data.
2. Meringkas dan menyajikan data.
3. Pembentukan pohon klasifikasi.
4. Menginterpretasi hasil klasifikasi.
5. Perbandingan persentase kesalahan klasifikasi.

3. HASIL DAN PEMBAHASAN

3.1. Klasifikasi dengan Metode QUEST

Gambar 1 menunjukkan bahwa dari 626 sampel yang diteliti, hasil yang didapatkan adalah nasabah dengan status lancar sebanyak 550 orang (87,9%) dan nasabah dengan status macet sebanyak 76 orang (12,1%). Variabel pendapatan per bulan (X_8) adalah variabel prediktor yang paling signifikan, sehingga variabel tersebut dipilih sebagai variabel penyekat. Variabel pendapatan per bulan (X_8) disekat menjadi 2 simpul, yaitu simpul ke-1 dengan kategori nasabah yang memiliki pendapatan per bulan <Rp2.000.000 dan Rp2.000.000-Rp5.000.000 menjadi satu simpul, dan simpul ke-2 dengan kategori nasabah yang memiliki pendapatan per bulan >Rp5.000.000 menjadi satu simpul. Pada simpul ke-1 proses penyekatan dilanjutkan dan menjadi simpul dalam, karena variabel status tempat tinggal (X_4) merupakan variabel prediktor kedua yang signifikan untuk membagi kategori pada simpul ke-1. Sementara pada simpul ke-2 proses penyekatan dihentikan karena semua kasus yang terdapat di dalam simpul memiliki nilai-nilai identik untuk tiap variabel prediktor, sehingga simpul ke-2 menjadi simpul akhir.

Variabel status tempat tinggal (X_4) disekat menjadi 2 simpul, yaitu simpul ke-3 dengan kategori nasabah yang status tempat tinggal milik mertua, kontrak, dan kos menjadi satu kelompok, dan simpul ke-4 dengan kategori nasabah yang status tempat tinggal milik sendiri, keluarga, orang tua, dan perusahaan menjadi satu kelompok. Pada simpul ke-3 dan simpul ke-4 proses penyekatan dihentikan karena semua kasus yang terdapat di dalam simpul ke-3 dan simpul ke-4 memiliki nilai-nilai identik untuk tiap variabel prediktor, sehingga simpul ke-3 dan simpul ke-4 menjadi simpul akhir. Dengan demikian, analisis dengan metode QUEST menghasilkan tiga simpul akhir.



Gambar 1. Diagram Pohon Klasifikasi QUEST untuk Data Nasabah

Hasil metode QUEST berdasarkan Gambar 1 mendapatkan 3 kelompok, yang diringkaskan dalam Tabel 1 dan Tabel 2 berikut:

Tabel 1. Hasil Klasifikasi Dari Diagram Pohon QUEST

Kelompok 1	Nasabah yang memiliki pendapatan per bulan <Rp2.000.000 dan Rp2.000.000-Rp5.000.000 dengan status tempat tinggal milik mertua, kontrak, dan kos.
Kelompok 2	Nasabah yang memiliki pendapatan per bulan <Rp2.000.000 dan Rp2.000.000-Rp5.000.000 dengan status tempat tinggal milik sendiri, keluarga, orang tua.
Kelompok 3	Nasabah dengan pendapatan per bulan >Rp5.000.000.

Tabel 2. Status Kredit Nasabah Pada Setiap Kelompok

Kelompok	Lancar		Macet		Prediksi status kredit
	N	%	N	%	
1	173	75,9	55	24,1	Lancar
2	174	92,1	15	7,9	Lancar
3	203	97,1	6	2,9	Lancar

Sumber: Data diolah, 2015

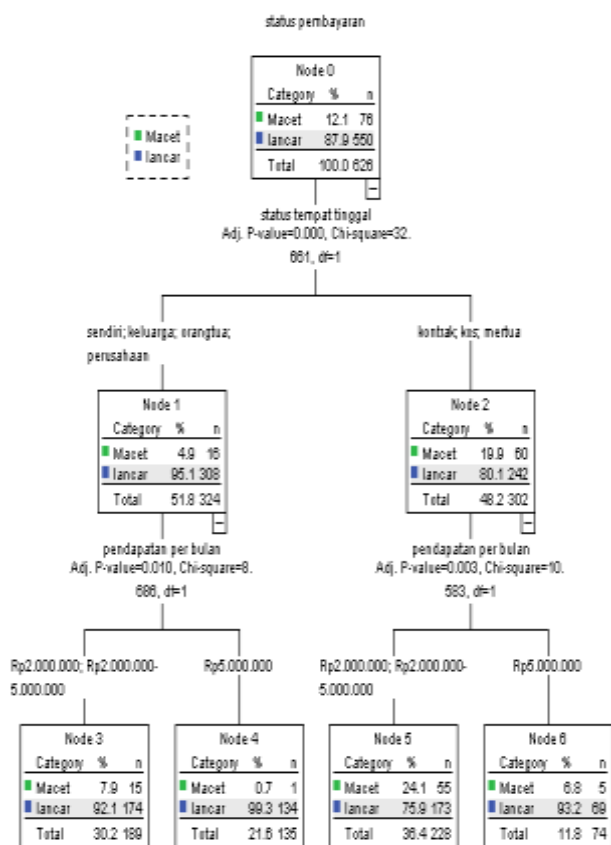
3.2. Klasifikasi dengan Metode CHAID

Gambar 2 menunjukkan bahwa variabel status tempat tinggal (X_4) adalah variabel prediktor yang paling signifikan terhadap variabel status kredit (Y). Setelah dikoreksi dengan pengali Bonferroni diperoleh nilai khi-kuadrat variabel status tempat tinggal (X_4) sebesar 32,661, sehingga variabel tersebut dipilih sebagai variabel penyekat.

Variabel status tempat tinggal (X_4) disekat menjadi 2 simpul, yaitu simpul ke-1 dan simpul ke-2. Pada simpul ke-1 terjadi penggabungan antara kategori status tempat tinggal milik sendiri, keluarga, orang tua, dan perusahaan karena jumlah nasabah berstatus lancar dan macet pada status tempat tinggal tersebut hampir seimbang atau hal ini berarti bahwa kategori status tempat tinggal milik sendiri, keluarga, orang tua, dan perusahaan memenuhi syarat kesignifikanan khi-kuadrat untuk bisa digabung menjadi satu simpul kategori campuran. Pada simpul ke-2 juga terjadi penggabungan antara kategori status tempat tinggal milik mertua, kos, dan kontrak menjadi satu simpul kategori campuran. Pada simpul ke-1 dan simpul ke-2 proses penyekatan dilanjutkan dan merupakan simpul dalam.

Variabel pendapatan per bulan (X_8) merupakan variabel prediktor kedua yang signifikan untuk menyekat kategori pada simpul ke-1 dan simpul ke-2. Variabel pendapatan per bulan (X_8) pada simpul ke-1 disekat menjadi 2 simpul, yaitu simpul ke-3 dan simpul ke-4. Pada simpul ke-3 terjadi penggabungan antara kategori nasabah yang memiliki pendapatan per bulan <Rp2.000.000 dan Rp2.000.000-Rp5.000.000 menjadi satu simpul kategori campuran dan simpul ke-4 dengan kategori nasabah yang memiliki pendapatan per bulan >Rp5.000.000 menjadi satu simpul. Pada simpul ke-3 dan simpul ke-4 proses penyekatan dihentikan karena tidak ada lagi variabel prediktor yang signifikan untuk membagi kategori pada simpul ke-3 dan simpul ke-4, sehingga simpul ke-3 dan simpul ke-4 menjadi simpul akhir. Selanjutnya untuk simpul ke-2 disekat menjadi 2 simpul, yaitu simpul ke-5 dan simpul ke-6. Pada simpul ke-5 terjadi penggabungan antara kategori nasabah yang

memiliki pendapatan per bulan <Rp2.000.000 dan Rp2.000.000-Rp5.000.000 menjadi satu simpul kategori campuran dan simpul ke-6 dengan kategori nasabah yang memiliki pendapatan per bulan >Rp5.000.000 menjadi satu simpul. Pada simpul ke-5 dan simpul ke-6 proses penyekatan dihentikan karena tidak ada lagi variabel prediktor yang signifikan untuk membagi kategori pada simpul ke-5 dan simpul ke-6, sehingga simpul ke-5 dan simpul ke-6 menjadi simpul akhir. Dengan demikian, analisis dengan metode CHAID menghasilkan empat simpul akhir.



Gambar 2. Diagram Pohon Klasifikasi CHAID untuk Data Nasabah

Hasil metode CHAID berdasarkan Gambar 2 mendapatkan 4 kelompok, yang diringkas dalam Tabel 3 dan Tabel 4.

Tabel 3. Hasil Klasifikasi Dari Diagram Pohon CHAID

Kelompok 1	Nasabah yang memiliki status tempat tinggal milik sendiri, keluarga, orang tua dan perusahaan dengan pendapatan per bulan <Rp2.000.000 dan Rp2.000.000-Rp5.000.000.
Kelompok 2	Nasabah yang memiliki status tempat tinggal milik sendiri, keluarga, orang tua dan perusahaan dengan pendapatan per bulan >Rp5.000.000.
Kelompok 3	Nasabah yang memiliki status tempat tinggal milik mertua, kontrak, dan kos dengan pendapatan per bulan <Rp2.000.000 dan Rp2.000.000-Rp5.000.000.
Kelompok 4	Nasabah yang memiliki status tempat tinggal milik mertua, kontrak, dan kos dengan pendapatan per bulan >Rp5.000.000

Tabel 4. Status Kredit Nasabah Pada Setiap Kelompok

Kelompok	Lancar		Macet		Prediksi status kredit
	N	%	N	%	
1	174	92,1	15	7,9	Lancar
2	134	99,3	1	0,7	Lancar
3	173	75,9	55	24,1	Lancar
4	69	93,2	5	6,8	Lancar

Sumber: Data diolah, 2015

3.3. Perbandingan Klasifikasi dengan Metode QUEST dan CHAID

Gambar 1 dan 2 menunjukkan bahwa metode QUEST dan CHAID memiliki kedalaman pohon yang sama yaitu sebanyak 2, namun memiliki jumlah simpul dan simpul akhir yang berbeda. Metode QUEST memiliki 5 simpul dengan 3 simpul akhir, sementara CHAID memiliki 7 simpul dengan 4 simpul akhir.

Pohon yang dihasilkan metode QUEST adalah pohon biner, karena QUEST hanya menghasilkan dua ketegori baru pada tiap penyekatan. Metode CHAID juga menghasilkan pohon biner, namun dengan alasan yang berbeda. Ini karena *p-value* pasangan-pasangan kategori variabel prediktor selalu lebih besar dari nilai $\alpha = 0,05$, sehingga pada akhirnya hanya tertinggal

dua kategori baru pada tiap penyekatan.

Pada metode QUEST dan CHAID persentase terbesar nasabah yang berstatus lancar memiliki karakteristik yang hampir sama. Metode QUEST dengan persentase 97,1%, yaitu nasabah dengan pendapatan per bulan >Rp5.000.000. Sementara untuk metode CHAID dengan persentase 99,3%, yaitu nasabah yang memiliki status tempat tinggal milik sendiri, keluarga, orang tua dan perusahaan dengan pendapatan per bulan >Rp5.000.000. Dan kesalahan klasifikasi terbesar dari kedua metode tersebut memiliki karakteristik yang sama, yaitu nasabah yang memiliki status tempat tinggal milik mertua, kontrak, dan kos dengan pendapatan per bulan <Rp2.000.000 dan Rp2.000.000-Rp5.000.000 dengan persentase 24,1%.

Hasil klasifikasi menunjukkan bahwa metode QUEST dan CHAID cukup baik digunakan untuk mengklasifikasi nasabah Adira Kredit Elektronik Cabang Denpasar, hal ini dilihat dari persentase keakuratan klasifikasi untuk status kredit dengan metode QUEST maupun CHAID, yaitu sebesar 87,9%. Kedua metode tersebut memiliki tingkat akurasi yang sama, hal ini dilihat dari kesalahan klasifikasi yang sama besar yaitu 12,1% seperti pada Tabel 5.

Tabel 5. Perbandingan Kesalahan Klasifikasi

Metode	Kesalahan klasifikasi	Keakuratan klasifikasi
QUEST	12,1%	87,9%
CHAID	12,1%	87,9%

4. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pembahasan yang dilakukan maka diperoleh kesimpulan: (1) Semua kelompok yang dihasilkan oleh metode QUEST dan metode CHAID menghasilkan kelompok dengan prediksi nasabah masuk ke dalam kategori lancar; (2) Pada metode QUEST dan CHAID persentase keakuratan klasifikasi terbesar nasabah yang berstatus lancar memiliki karakteristik yang hampir sama; (3) Metode QUEST dan CHAID memiliki tingkat akurasi yang sama, hal ini dilihat dari kesalahan klasifikasi dalam mengklasifikasikan data nasabah Adira Kredit Elektronik Cabang

Denpasar yang sama besar yaitu: 12,1%.

Saran yang diberikan yaitu, untuk penelitian berikutnya metode QUEST dapat dikembangkan untuk variabel respon berupa variabel kategori yang memiliki lebih dari dua kategori dan untuk variabel prediktor dapat digunakan variabel numerik, serta dapat dilanjutkan dengan proses pemangkasan yang dilakukan untuk memilih pohon terbaik.

DAFTAR PUSTAKA

- [1] Gallagher, C.A. 2000. *An Iterative Approach to Classification Analysis*. www.casact.org/library/ratemaking/90dp237.pdf [Februari 2014].
- [2] Kass, G.V. 1980. An Exploratory Technique for Investigating Large Quantities of categorical Data. *Applied Statistics*, Vol. 29 (2):119-127. <http://www4.stat.ncsu.edu/~dickey/Analytics/Datamine/Reference%20Papers/kass80.pdf> [Februari 2014].
- [3] Loh W.Y. dan Shih, Y.S. 1997. Split Selection Methods for Classification Trees. *Statistica Sinica*7; 815-840. <http://www3.stat.sinica.edu.tw/statistica/oldpdf/A7n41.pdf> [Agustus 2014].