

Pendekatan Bootstrap pada Klasifikasi Pemodelan Respon Ordinal (Ordinal Regression Model using Bootstrap Approach)

Bambang Widjanarko Otok¹⁾, M. Sjahid Akbar¹⁾, Suryo Guritno²⁾ dan Subanar²⁾

¹⁾ Staf Pengajar Jurusan Statistika FMIPA Institut Teknologi Sepuluh Nopember Surabaya

²⁾ Staf Pengajar FMIPA Universitas Gadjah Mada

ABSTRACT

The aim of the research content three part, thus a to know misclassification and model discriminant analysis with bootstrap approach, model regression ordinal with bootstrap approach, and model MARS with bootstrap approach. The data used is data of secondary related to matrix variance covariance is same and unequal that is (The data worker standard of living and banking performance). The result of this research shows that in determining distinguishing variable between groups there are difference of variable at each method. This matter because of at each method has specification either from fulfilled of assumption and also estimation its. So also at accuracy of classification between groups there is difference especially at matrix of variance covariance unequal at worker standard of living case. As a whole can be concluded that the problem accuracy of classification bootstrap approach at each method give small mistake of goodness at matrix variance covariance unequal and equal.

Keywords: classification, bootstrap, discriminant analysis, ordinal regression, MARS.

PENDAHULUAN

Metode klasifikasi merupakan bagian dari analisis statistika pada respon dengan skala pengukuran data nominal atau ordinal. Metode yang sering digunakan untuk masalah klasifikasi adalah analisis diskriminan yang dikembangkan Fisher (1936). Analisis diskriminan yang dikembangkan untuk populasi yang berdistribusi normal dengan varians-kovarians sama. Tetapi pada penerapannya analisis diskriminan sering melibatkan variabel-variabel kategorik yang tidak mengikuti pola distribusi normal, sehingga akibatnya diperoleh hasil tidak optimal (Dillon, 1978).

Sedang Krazanowski dalam Johnson & Wichern (1992), menyatakan bahwa fungsi diskriminan linear dapat digunakan dengan hasil yang kurang baik dan hasilnya sangat tergantung pada korelasi antar variabel kategorik dan kontinyu. Pendekatan parametrik yang lain untuk masalah klasifikasi adalah regresi logistik dan metode ini dalam Ripley (1996) dikenal dengan logistik diskriminan. (Agresti, 1990), analisis regresi logistik digunakan untuk analisis data respon kategorik (nominal atau ordinal) dengan variabel-variabel bebas kontinu dan kategorik. Pembentukan model terutama untuk pembedaan kelas dengan menghitung probabilitas masing-masing kelas.

Dalam membandingkan analisis diskriminan dan regresi logistik Sharma (1996) menunjukkan bahwa analisis diskriminan lebih efisien dibanding regresi logistik dalam

perhitungannya. Analisis regresi logistik adalah analisis yang digunakan untuk melihat hubungan antara variabel respon kategorik dengan variabel-variabel bebas kategorik maupun kontinu. Variabel respon dalam regresi logistik dapat berbentuk dikhotom (biner) maupun polytomous (ordinal atau nominal). Agresti (1990), metode yang sering digunakan untuk variabel respon berskala ordinal adalah dengan membentuk fungsi logit dari peluang kumulatif.

Dalam regresi logistik masalah data pengamatan biner terkelompok dengan mengasumsikan ketidak-bebasan memberikan pendugaan yang tidak tepat, sehingga perlu kajian mengenai pendugaan parameter dengan metode alternatif yang lebih tepat, salah satunya dengan pendekatan bootstrap. Portier (2001), beberapa metode non-parametrik untuk masalah klasifikasi yang berkembang adalah metode Kernel, Nearest Neighbors, Regresi Pohon (CART), Artificial Neural Network (ANN). Dalam perkembangannya, Tibshirani (2001), Abraham & Steinberg (2001), *Multivariate Adaptive Regression Splines* (MARS) adalah salah satu kelompok model statistik modern yang juga diterapkan untuk menyelesaikan permasalahan klasifikasi dan peramalan.

Untuk menduga parameter populasi, ada beberapa metode yang digunakan. Beberapa metode seperti momen, kemungkinan maksimum, bayes dan kuadrat terkecil tidak selalu memberikan hasil yang sama baiknya,

seringkali pemilihan metode pendugaan parameter populasi merupakan masalah khusus. Myers (1990), metode kuadrat terkecil, dapat dipakai jika asumsi dipenuhi, sedangkan metode robust, jackknife dan bootstrap merupakan pendekatan lain yang tidak mensyaratkan asumsi. Metode bootstrap merupakan metode pendugaan parameter yang berprinsip pengambilan sample kembali dengan pengembalian berulang kali. Beberapa peneliti yang memfokuskan pada pendugaan parameter dengan metode di atas, diantaranya Miller (1974), Efron and Gong (1983), Hinkley (1983, 1988), Efron and Tibshirani (1986), Good (1994), Edgington (1995), Manly (1997) dalam Walsh (2000).

Suatu model yang terbaik pada permasalahan klasifikasi, adalah model yang menghasilkan kesalahan klasifikasi yang minimal. Kenyataannya, masalah klasifikasi sering dijumpai pada data dengan pola sebaran yang tidak normal dan matrik varians kovarians tidak sama. Berbagai metode tradisional (parametrik) seperti analisis diskriminan dan regresi logistik telah dikembangkan untuk menyelesaikan problem ini. Salah satu yang dikembangkan dalam kerangka pemodelan non-parametrik adalah CART dan MARS, implementasi model ini tidak memerlukan adanya asumsi-asumsi yang harus dipenuhi seperti pemodelan parametrik pada umumnya, sehingga penerapannya lebih fleksibel.

Permasalahan utama dalam pemodelan baik parametrik dan nonparametrik kaitannya dengan klasifikasi adalah belum ada prosedur yang standart dalam metode pengambilan sample untuk mendapatkan model training dan testing, dugaan parameter dan uji hipotesa yang ada kaitan dengan model yang diperoleh, serta penentuan kriteria sebagai validasi kesesuaian model. Untuk itu secara khusus kajian peneliti memfokuskan pada permasalahan: bagaimana menguji parameter model pada respon ordinal untuk permasalahan klasifikasi dengan pendekatan bootstrap. Tujuan dalam penelitian ini adalah mengetahui tingkat kesalahan pengelompokan berdasarkan variabel yang mem-pengaruhi dan mengetahui variabel apa yang membedakan antara kelompok dengan beberapa metode.

Analisis Diskriminan

Analisis diskriminan adalah salah satu metode analisis multivariat yang bertujuan untuk memisahkan beberapa kelompok data yang

sudah terkelompok-kan dengan cara membentuk fungsi diskriminan. Untuk melakukan analisis diskriminan ada asumsi dasar yang mendasari perhitungan analisis tersebut yaitu: data kasus harus berasal dari dua atau lebih golongan kelompok, karena analisis diskriminan dipakai untuk interpretasi seberapa jauh kelompok yang dibedakan tersebut memang berbeda dan supaya data tersebut dapat dipergunakan untuk mengklasifikasikan variabel yang akan dibedakan secara matematis, maka karakteristik yang akan dipakai sebagai pembeda jenis datanya mempunyai skala pengukuran minimal interval. Secara teoritis tidak ada batas jumlah *discriminating* variabel sepanjang jumlah total kasus melebihi jumlah variabel. (William, 1991).

Misalkan ada m kelompok sampel random yang masing-masing berukuran n_1, n_2, \dots, n_m dengan p variabel yang diamati, X_1, X_2, \dots, X_p . Vektor rata-rata dari m sampel tersebut, x_1, x_2, \dots, x_m dapat dianggap sebagai dugaan vector rata-rata populasi dan dugaan matriks varians-kovarians kelompok ke- i adalah C_i . Jarak Mahalanobis setiap atau suatu obyek dapat dihitung terhadap m vector rata-rata dan akan digolongkan pada suatu kelompok yang terdekat terhadap vector rata-rata. Jarak Mahalanobis antara suatu obyek x terhadap vector rata-rata kelompok ke- j , x_j , diduga oleh $(x-x_j)C^{-1}(x-x_j)$, dengan asumsi bahwa matrik varians kovarians sama, $C = \sum (n_i - 1)C_i / \sum (n_i - 1)$ sebagai dugaan varians-kovarians gabungan dari m kelompok sampel. Jadi, pengelompokkan obyek x ke kelompok ke- i , bila $(x-x_i)C^{-1}(x-x_i) = \text{minimum}\{(x-x_j)C^{-1}(x-x_j); j=1,2,\dots,m\}$ dan tentunya ada obyek yang sebenarnya tidak berasal dari kelompok tersebut. Fakta ini akan dapat digunakan untuk mengetahui sejauh mana kelompok-kelompok ini dapat dipisahkan dengan menggunakan variabel yang ada sebagai ukuran salah klasifikasi. Dan sebagai pengujian kestabilan ketepatan pengelompokan dengan cara menghitung Press's Q, yang diformulasikan sebagai berikut:

$$Pr\ ess's Q = \frac{[N - (nK)]^2}{N(K - 1)}$$

dimana: N = Jumlah Total sampel;
 n = Jumlah individu yang tepat diklasifikasikan
 K= Jumlah kelompok

Pendekatan lain dalam analisis diskriminan seperti diuraikan diatas adalah menentukan fungsi diskriminan, dimana fungsi diskriminan yang diperoleh, misal Z_1, Z_2, \dots, Z_i merupakan kombinasi linier yang dipilih sehingga Z_1 mereflesikan perbedaan terbesar antar kelompok, Z_2 mereflesikan perbedaan terbesar antar kelompok yang tidak dapat dicakup oleh Z_1 , Z_3 mereflesikan perbedaan terbesar antar kelompok yang tidak dapat dicakup oleh Z_1 dan Z_2 , dan seterusnya.

Dari analisis diskriminan ini dapat pula digunakan untuk mencari variabel-variabel asal yang dianggap dominan untuk digunakan dalam membedakan antar kelompok, salah satu yang digunakan melalui variabel secara bertatar, yaitu menambahkan variabel satu per satu yang relatif dominan ke dalam fungsi sampai suatu saat dimana penambahan variabel lainnya dianggap tidak menambah baik diskriminannya.

Regresi Ordinal

Metode yang sering digunakan untuk variabel respon berskala ordinal adalah dengan membentuk fungsi logit dari peluang kumulatif. Metode ini dikenal dengan Multinomial logit terurut atas Regresi Ordinal (McCullagh and Nelder, 1983; Agresti, 1990). Untuk memperoleh penduga kemungkinan maksimum bagi parameternya digunakan metode kuadrat terkecil terboboti secara iteratif. Sedangkan untuk menguji model regresi logistik yang dibentuk digunakan uji nisbah kemungkinan dan aturan klasifikasi berdasarkan nilai-nilai peluang klasifikasi satu individu ke dalam populasi. (Hosmer & Lemeshow, 1989)

Misalkan $\pi_j(x) = P(Y = j | X = x)$ adalah probabilitas dari $Y = j, j = 0, 1, \dots, k$ untuk $X = x$. Kelas dari model kontinu terkelompok adalah berdasarkan probabilitas kumulatif,

$$\gamma_j(x) = P(Y \geq j | X = x), \quad j = 1, 2, \dots, k \tag{1}$$

Dengan memandang Y suatu variabel diskret dari suatu sifat kontinu laten yang mendasari dan ditentukan oleh cutoff point j , maka dalam merumuskan probabilitas kumulatif γ_j dianggap alamiah. Tetapi hal ini tidak mutlak untuk mengandaikan keberadaan variabel kontinu yang mendasari.

Schmidt and Strauss (1975) dalam Upton (1978), memberikan model pilihan tingkat

hidup pekerja dengan kode 0, 1, 2, 3 dan 4 sebagai berikut:

$$P(Y = j) = \frac{e^{\beta_j X_i}}{\sum_{k=0}^4 e^{\beta_k X_k}} \tag{2}$$

Persamaan (2) dikenal disebut sebagai model multinomial logit (regresi ordinal). Persamaan yang diestimasi memberikan sekumpulan probabilitas untuk pengambil keputusan dengan karakteristik X_i . Namun sebelumnya harus dihilangkan indeterminasi (keadaan yang menyebabkan model itu tak bernilai). Dengan mendefinisikan

$\beta_1^* = \beta_1 + q$ untuk vektor q yang tidak nol, maka semua notasi yang melibatkan q dihilangkan. Normalisasi yang sesuai untuk menyelesaikan permasalahan diatas adalah dengan mengasumsikan $\beta_0 = 0$. Sehingga probabilitasnya sebagai berikut:

$$P(Y = j) = \frac{e^{\beta_j X_i}}{1 + \sum_{k=0}^4 e^{\beta_k X_i}}, \quad \text{untuk } j = 1, 2, \dots, J$$

$$P(Y = 0) = \frac{1}{1 + \sum_{k=0}^4 e^{\beta_k X_i}} \tag{3}$$

Koefisien-koefisien pada model ini sukar diinterpretasikan. Dengan menurunkan Persamaan (3), diperoleh efek marginal dari pembentuk regresi pada probabilitasnya, yaitu:

$$\frac{\partial P}{\partial x} = P_j \left[\beta_j - \sum_k P_k \beta_k \right] \tag{4}$$

Hal ini bisa dihitung dari estimasi parameter. Kesalahan standarnya dapat diestimasi, tetapi hasilnya sangat kompleks.

Model multinomial logit mengimpli-kasikan log-odds ratio ke-J, sebagai berikut,

$$\ln \left[\frac{P_{ij}}{P_{i0}} \right] = \beta_j X_i \tag{5}$$

Sehingga dalam menormalkan probabilitas yang lain, dapat diperoleh hasil berikut,

$$\ln \left[\frac{P_{ij}}{P_{ik}} \right] = X_i (\beta_j - \beta_k) \tag{6}$$

Estimasi ini akan bermanfaat karena odds ratio $\frac{P_j}{P_k}$, tidak tergantung pada pilihan yang lainnya. Hal ini mengikuti independensi dari unsur gangguan pada model yang asli. Estimasi dari model multinomial logit dengan Metode Newton mendapatkan solusi yang baik jika data tidak terkondisikan jelek. Log likelihood dapat dihasilkan dengan mendefinisikan masing-masing individu $d_{ij} = 1$ jika alternatif j dipilih oleh seseorang individu i, dan $d_{ij} = 0$ jika alternatif j tidak dipilih oleh seseorang individu i. Log likelihood ini merupakan generalisasi dari model probit dan logit.

$$\ln L = \sum_i \sum_{j=0}^J d_{ij} \ln P(Y_i = j) \dots\dots\dots(7)$$

Apabila kita turunkan akan memiliki bentuk yang sederhana, sebagai berikut:

$$\frac{\partial \ln L}{\partial \beta_j} = \sum_i [d_{ij} - P_{ij}] x_i, \text{ untuk } j = 1, 2, \dots, J \dots\dots\dots(8)$$

Matrik derivatif eksak kedua memiliki blok J^2 , yang masing-masing berukuran $k \times k$. Blok diagonal ke - j adalah,

$$-\sum_i P_{ij} (1 - P_{ij}) x_i x_i' \dots\dots\dots(9)$$

Blok selain diagonal ke-j adalah,

$$\sum_i (P_{ij} P_{ik}) x_i x_i' \dots\dots\dots(10)$$

Jika nilai Hessian tidak mencakup d_{ij} , maka Persamaan diatas merupakan nilai ekspektasi, dan metode Newton sama dengan metode penyekoran. Begg and Gray (1984) menyatakan bahwa metode ini dapat digunakan dengan menjumlahkan produk-produk luar dari

$$\hat{f}(x) = a_0 + \sum_{K_m=1} f_i(x_i) + \sum_{K_m=2} f_{ij}(x_i, x_j) + \sum_{K_m=3} f_{ijk}(x_i, x_j, x_k) + \dots \dots\dots(13)$$

Persamaan (13), menunjukkan bahwa penjumlahan pertama meliputi semua basis fungsi untuk satu variabel, penjumlahan kedua meliputi semua basis fungsi untuk interaksi

derivatif pertama. Tetapi, hal ini jarang dikembangkan karena bentuknya yang sangat sederhana dan karena kecekungan global dari log likelihoodnya. Hal ini juga tidak baik jika terjadi overparameter dalam model. Sementara itu data cross section secara khusus terkadang melibatkan jumlah pembentuk regresi yang banyak.

Multivariate Adaptive Regression Spline (MARS)

Model MARS selain penentuan knots yang dilakukan secara otomatis dari data, juga menghasilkan model yang kontinu pada knots. Pemilihan knots pada MARS menggunakan algoritma *forward stepwise* dan *backward stepwise* yang salah satunya didasarkan nilai Generalized Cross Validation (GCV) minimum. Model MARS hasil algoritma diatas dapat ditulis sebagai berikut:

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})] \dots\dots\dots(11)$$

dimana:

- a_0 = basis fungsi induk
- a_m = koefisien dari basis fungsi ke- m
- M = maksimum basis fungsi (*nonconstant basis fungsi*)
- K_m = derajat interaksi
- S_{km} = nilainya ± 1
- $x_{v(k,m)}$ = variabel independen
- t_{km} = nilai knots dari variabel independen $x_{v(k,m)}$

Penjabaran dari Persamaan (11) dapat disajikan sebagai berikut:

$$\begin{aligned} \hat{f}(x) = & a_0 + \sum_{m=1}^M a_m [s_{1m} \cdot (x_{v(1,m)} - t_{1m})] \\ & + \sum_{m=1}^M a_m [s_{1m} \cdot (x_{v(1,m)} - t_{1m})] [s_{2m} \cdot (x_{v(2,m)} - t_{2m})] \\ & + \sum_{m=1}^M a_m [s_{1m} \cdot (x_{v(1,m)} - t_{1m})] [s_{2m} \cdot (x_{v(2,m)} - t_{2m})] [s_{3m} \cdot (x_{v(3,m)} - t_{3m})] \\ & + \dots \dots\dots(12) \end{aligned}$$

dan secara umum Persamaan (11) dapat dituliskan sebagai berikut:

antara dua variabel, penjumlahan ketiga meliputi semua basis fungsi untuk interaksi antara tiga variabel dan seterusnya.

Misalkan $V(m) = \{v(k, m)\}_1^{K_m}$ adalah himpunan dari variabel yang dihubungkan dengan basis fungsi B_m ke- m , maka setiap penjumlahan pertama pada Persamaan (13) dapat dinyatakan sebagai:

$$f_i(x_i) = \sum_{\substack{K_m=1 \\ i \in V(m)}} a_m B_m(x_i) \dots\dots\dots(14)$$

$f_i(x_i)$ merupakan penjumlahan semua basis fungsi untuk satu variabel x_i dan merupakan spline dengan derajat $q=1$ yang merepresentasikan fungsi univariat. Setiap fungsi bivariat pada Persamaan (13) dapat ditulis sebagai:

$$f_{ij}(x_i, x_j) = \sum_{\substack{K_m=2 \\ (i,j) \in V(m)}} a_m B_m(x_i, x_j) \dots\dots\dots(15)$$

yang merepresentasikan penjumlahan semua basis fungsi dua variabel x_i dan x_j . Penambahan ini untuk menghubungkan kontribusi univariat, yang dituliskan sebagai berikut:

$$f_{ij}^*(x_i, x_j) = f_i(x_i) + f_j(x_j) + f_{ij}(x_i, x_j) \dots\dots\dots(16)$$

Untuk fungsi trivariat pada penjumlahan yang ketiga diperoleh dengan menjumlahkan semua basis fungsi untuk tiga variabel, yang dituliskan sebagai berikut:

$$f_{ij}(x_i, x_j, x_k) = \sum_{\substack{K_m=3 \\ (i,j,k) \in V(m)}} a_m B_m(x_i, x_j, x_k) \dots\dots\dots(17)$$

Penambahan fungsi univariate danbivariate mempunyai kontribusi dalam bentuk:

$$f_{ijk}^*(x_i, x_j, x_k) = f_i(x_i) + f_j(x_j) + f_k(x_k) + f_{ij}(x_i, x_j) + f_{ik}(x_i, x_k) + f_{jk}(x_j, x_k) + f_{ijk}(x_i, x_j, x_k) \dots\dots\dots(18)$$

Persamaan (13) merupakan dekomposisi dari analisis varians untuk table kontingensi, yang dikenal dengan dekomposisi ANOVA dari model MARS.

Interpretasi model MARS melalui dekomposisi ANOVA adalah merepresen-tasikan variabel yang masuk dalam model, baik

untuk satu variabel maupun interaksi antara variabel, selanjutnya merepresen-tasikan secara grafik. Penambahan aditif Persamaan (14) dapat ditunjukkan dengan membuat plot antara $f_i(x_i)$ dengan x_i sebagai salah satu model aditif. Kontribusi interaksi antara dua variabel dapat divisualisasikan dengan membuat plot antara $f_{ij}^*(x_i, x_j)$ dengan x_i dan x_j menggunakan countur plot. Model dengan interaksi yang lebih tinggi dalam visualisasi dapat dibuat dengan menggunakan plot dalam beberapa variabel fixed dengan variabel komplemen.

Bootstrap

Penggunaan metode statistik parametric biasanya mensyaratkan informasi mengenai distribusi yang harus dipenuhi dan ini sulit untuk dipenuhi. Untuk mengatasi hal ini dapat digunakan metode-metode yang tidak memerlukan asumsi ketat, salah satunya metode Bootstrap. Metode Bootstrap merupakan teknik nonparametric untuk penarikan kesimpulan (*inference*).

Pembootstrapan bertitik tolak atas dasar analog antara sampel dan populasi dari mana sampel tersebut diambil. Penarikan kesimpulan dengan bootstrap akan mem-berikan hasil yang lebih baik apabila asumsi yang ada tidak jelas dan kurang realistik untuk diterapkan pada populasi tersebut. Metode bootstrap pertama kali dipelajari oleh Efron (1979). Metode bootstrap merupakan suatu metode penaksiran non-parametrik yang dapat menaksir parameter-parameter dari suatu distribusi, variansi dari sample median, serta dapat menaksir tingkat kesalahan (error). Pada metode bootstrap dilakukan pengambilan sampel dengan pengembalian (resampling with replacement) dari sampel data. Secara singkat, algoritma bootstrap adalah sebagai berikut: (Mooney, 1993)

Ambil sampel berukuran n , yaitu $S : x_1, x_2, \dots, x_n$

Ambil sampel kembali dari S dengan pengembalian berukuran n dan dapatkan nilai statistik $\hat{\theta}_i$ untuk sample S_i .

Lakukan langkah 2 sebanyak B . [B sebanyak 1000 – 5000]

Tentukan nilai statistik dengan bootstrap:

$$\hat{\theta}_b = B^{-1} \sum \hat{\theta}_i \text{ dan } \hat{\sigma}_\theta^* = \frac{\sum(\hat{\theta}_i - \hat{\theta}_b)^2}{(B - 1)}$$

Ketenagakerjaan

Pembangunan di Indonesia merupakan perpaduan antara berbagai kegiatan, penempatan dan pembangunan sumber daya alam maupun sumber daya manusia serta pemeliharaan maupun pemanfaatan hasilnya secara optimal. Proses gerak dan langkah pembangunan itu sendiri melibatkan unsur manusia karena merupakan fokus utama sebagai sumber daya yang efektif.

Tenaga kerja merupakan faktor yang terpenting dalam proses produksi. Sebagai sarana produksi, tenaga kerja lebih penting daripada sasaran produksi lain seperti: bahan mentah, tanah, air dan sebagainya. Karena manusialah yang menggerakkan semua sumber-sumber tersebut untuk menghasilkan barang.

Penyediaan tenaga kerja pun sifatnya terbatas karena tidak semua penduduk merupakan angkatan kerja. Hanya penduduk yang telah mencapai umur minimum tertentu baru dianggap sebagai tenaga kerja potensial atau angkatan kerja. Selain itu, tidak semua angkatan kerja terlibat dalam kegiatan ekonomi (bekerja). Jumlah angkatan kerja dalam suatu negara atau daerah pada suatu waktu tertentu tergantung dari jumlah penduduk usia kerja.

Tingkat Partisipasi Angkatan Kerja (TPAK) dipengaruhi oleh berbagai faktor demografis, sosial dan ekonomi. Faktor-faktor ini antara lain adalah umur, status perkawinan, tingkat pendidikan, daerah tempat tinggal (kota/desa), pendapatan dan agama. Pengaruh dari masing-masing faktor ini terhadap TPAK berbeda antara penduduk laki-laki dan penduduk perempuan. Pengaruh ini terhadap tingkat partisipasi laki-laki dalam angkatan kerja tidaklah begitu besar, sebab umumnya laki-laki merupakan pencari nafkah utama di dalam keluarga. Karena hampir semua laki-laki yang telah mencapai usia kerja terlibat dalam kegiatan ekonomi.

Kinerja Perbankan

Penilaian kinerja perusahaan dengan menggunakan rasio-rasio keuangan memerlukan ukuran atau standar sebagai pembanding, salah satu pendekatan yaitu membandingkan rasio-rasio perusahaan dengan pola untuk industri atau lini usaha dimana perusahaan secara dominan beroperasi, dengan

demikian perusahaan dapat membandingkan kinerja yang dimiliki dengan kinerja perusahaan lain dalam industri tersebut. Begitu pula bank dikatakan mempunyai kinerja yang baik apabila bank dikelola lebih baik dibanding bank lainnya, baik dalam mengelola resiko maupun upaya untuk mendapatkan return.

Weston and Thomas (1995), untuk menganalisis rasio-rasio keuangan: rasio-rasio harus dikelompokkan menjadi tiga kelompok, yaitu ukuran kinerja (*performance measures*), ukuran efisiensi operasi (*operating efficiency measures*), dan ukuran kebijakan keuangan (*financial policy measures*). Sedangkan Hempel *et al.*, (1991), untuk melihat kinerja perbankan digunakan rasio-rasio yang dapat memberi gambaran *risk* dan *return* (seperti: *Interest Margin, Net Margin, Return On Asset, Leverage Multiplier, Return On Equity, Liquidity Risk, Interest Rate Risk, Credit Risk, Capital Risk*). Selain itu untuk memprediksi kinerja bank di masa datang, bank harus mengetahui prestasi dan kegagalan, kekuatan dan kelemahan yang mungkin terjadi, oleh karena itu bank harus menganalisis semua aspek-aspek pokok yang berkaitan dengan kinerja termasuk pula mengukur modal yang dibutuhkan oleh bank.

Dalam memprediksi kinerja perbankan di masa yang akan datang digunakan variabel rasio keuangan, yang meliputi *credit risk, liquidity risk, interest rate risk, operational risk dan capital or solvency risk* dan beberapa rasio keuangan yang dapat secara spesifik menunjukkan kondisi perbankan di Indonesia. Masalah pengelompokan (klasifikasi) terjadi pada kinerja perbankan yang didasarkan beberapa metoda formal untuk pertimbangan situasi baru. Apabila ada kasus baru yang berhubungan dengan prosedur pengelompokan, maka kasus baru tersebut didefinisikan pada kelompok mana atas dasar variabel yang diamati.

METODE

Dalam penelitian ini digunakan data sekunder yang berkaitan dengan respon ordinal yang diambil dari:

- Data 'Tingkat Hidup Pekerja', faktor-faktor yang mempengaruhi tingkat hidup pekerja antara lain adalah umur, status perkawinan, tingkat pendidikan, daerah tempat tinggal (kota/desa), pendapatan dan agama, dalam hal

ini peubah-peubah yang terkait didefinisikan sebagai berikut: (BPS, 2004).

Y = Tingkat Hidup Pekerja

(1 = Kurang baik)

(2 = Sama baik)

(3 = Lebih baik)

X1 = Pendidikan

(1 = SD)

(2 = SMP)

(3 = SMA)

X2 = Status Pekerjaan

(1 = HL/Borongan/Kontrak)

(2 = Harian Tetap)

(3 = Bulanan)

X3 = Upah/Gaji Sebulan

(1 = kurang dari Rp 500.000)

(2 = Rp 500.001 – Rp 1.000.000)

(3 = Rp 1000.001 – Rp 1.500.000)

(4 = lebih dari Rp 1.500.000)

X4 = Status Perkawinan

(1 = Menikah)

(2 = Belum Menikah).

- Kinerja Perbankan, dalam memprediksi kinerja perbankan di masa yang akan datang digunakan variabel rasio keuangan, yang meliputi *credit risk, liquidity risk, interest rate risk, operational risk dan capital or solvency risk* dan beberapa rasio keuangan (Hempel et al, 1991 dalam Hidayati, 2002), dan peubah-peubah tersebut diatas didefinisikan sebagai berikut:

Peubah Tak Bebas:

Y = Kinerja Perbankan

(1 = Sehat)

(2 = Tidak Sehat)

Peubah bebas meliputi:

Gross Profit Margin (X₁), Net Interest Margin (X₂), Earning Spread (X₃), Net Non Interest Margin (X₄), Return on Asset (X₅), Return on Equity (X₆), Net Profit Margin (X₇), Asset Utilization (X₈), Interest Margin on Loans (X₉), Operating Efficiency Ratio (X₁₀), Cash Position Indicator (X₁₁), Total Loans Over Total Assets (X₁₂), Loan to Deposit Ratio (X₁₃), Quick Ratio (X₁₄), Deposit Composition Ratio (X₁₅), Capital Adequacy Ratio (X₁₆), Internal Capital Growth Rate (X₁₇), Leverage Ratio (X₁₈), Aktiva Produktif yang diklasifikasikan (X₁₉), Penyisihan Penghapusan Aktiva Produktif (X₂₀), Ratio Non Performing Loans to Total Lonas (X₂₁), dan Rasio Total Obligasi Pemerintah dengan Total Dana (X₂₂).

Pada tahap analisis data, dari hasil pengumpulan data kemudian dilakukan

pengolahan data dengan tahapan sebagai berikut:

- Melakukan analisis deskriptif untuk mengetahui karakteristik dari sampel.
- Melakukan analisis beberapa metode (sesuai dengan tujuan khusus) dan bootstrap.
- pemilihan model terbaik yang meliputi penilaian dugaan parameter-parameter di dalam model dan kesalahan klasifikasi yang kecil dengan kriteria Press's Q
- Analisis yang dilakukan terhadap data dengan pendekatan bootstrap dengan langkah – langkah sebagai berikut :
- Mencari persentase ketepatan klasifikasi masing-masing sel pada data respon
- Resampling persentase ketepatan klasifikasi masing-masing sel pada data respon dilakukan replikasi sebanyak 1000 kali
- Menghitung rata-rata persentase ketepatan klasifikasi masing-masing sel pada data respon dari resampling.

HASIL DAN PEMBAHASAN

Penelitian ini ditekankan pada data respon antar kelompok dengan matrik varians tidak sama dan matriks varians sama.

Tingkat Hidup Pekerja dengan Analisis Diskriminan

Analisis diskriminan yang merupakan salah satu metode statistik yang sering digunakan dalam masalah pengelompokan. Analisis ini dilakukan terhadap tiga kelompok data yang sudah valid pengelompok-kannya, yaitu kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik'.

Analisis diskriminan pada kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' didasarkan pada 4 variabel. Secara univariate dengan menggunakan statistik uji F dan tingkat signifikansi 0.05, diperoleh satu variabel yang rata-ratanya berbeda antara kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' (Sig. = 0.027 < 0.05), dan variabel tersebut adalah pendidikan (X₁) Selanjutnya secara multivariate dengan prosedur stepwise (bertatar) diperoleh informasi variabel-variabel mana yang membedakan antara kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik'. Hasil pengolahan menunjukkan bahwa variabel pendidikan (X₁) sangat mempengaruhi dalam membedakan penge-lompokan terhadap kelompok tingkat hidup pekerja

'kurang baik', 'sama baik' dan 'lebih baik'. Sedangkan hubungan antara variabel pembeda dengan kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' ditunjukkan oleh korelasi kanonik yaitu sebesar 12,6 persen.

Fungsi diskriminan yang membedakan antara kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' adalah sebagai berikut :

$$F = -2.122 + 1.365 X_j$$

Fungsi diskriminan di atas mampu menerangkan ketepatan klasifikasi pada kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' sebesar 54,6%. Titik pembatas menunjukkan bahwa jika nilai-nilai variabel pembeda dimasukkan dalam fungsi diskriminan maka dengan nilai tertentu dapat dikelompokkan pada kelompok tingkat hidup.

Tingkat ketepatan prediksi model dengan analisis diskriminan untuk mengelompokkan kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' secara keseluruhan sebesar 54,6%. Untuk lebih jelas dapat dilihat pada Tabel 1.

Untuk lebih akurat mengenai presentasi ketepatan pengklasifikasian pada analisis diskriminan, berikut disajikan hasil perhitungan Press's Q:

$$Pr ess'Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[454 - (248 \times 3)]^2}{454(3 - 1)} = 92.62$$

Nilai Press'Q = 92.62 dibandingkan $\chi^2_{2,0.05} = 5.9915$, ternyata nilai Press'Q lebih besar dari nilai chi-square, sehingga keakuratan pengklasifikasian taraf hidup pekerja adalah konsisten.

Tingkat Hidup Pekerja Dengan Regresi Ordinal

Dari deskriptif diatas diduga pengelompokkan tingkat hidup pekerja pada kelompok 'kurang

baik', 'sama baik' dan 'lebih baik' dipengaruhi oleh empat variabel yaitu pendidikan (X1), status pekerjaan (X2), upah/gaji sebulan (X3) dan status perkawinan (X4). Sedangkan dengan analisis regresi ordinal, ternyata tiga variabel yaitu pendidikan (X1) dengan Sig. = 0,049 < $\alpha=0,05$, status pekerjaan (X2) dengan Sig. = 0,024 < $\alpha=0,05$, dan upah/gaji sebulan (X3) dengan Sig. = 0,030 < $\alpha=0,05$. Sedangkan variabel status perkawinan (X4) dengan Sig. = 0,084 > $\alpha=0,05$ tidak signifikan secara statistik. Hal ini menunjukkan bahwa ketiga variabel tersebut mempengaruhi pengelompokkan tingkat hidup pekerja.

Tingkat hidup pekerja yang lebih baik dipengaruhi oleh pendidikan, status pekerjaan, upah/gaji sebulan dan status perkawinan. Pendidikan SMA mempunyai tingkat hidup yang lebih baik sebesar 28 kali dibandingkan dengan pendidikan SD, sedangkan pendidikan SMP mempunyai tingkat hidup yang lebih baik sebesar 32 kali dibandingkan pendidikan SD. Pekerja dengan status pekerjaan sebagai HL/Borongan/Kontrak mempunyai tingkat hidup yang lebih baik dibanding status pekerjaan sebagai bulanan. Untuk pekerja dengan upah/gaji sebulan sebesar Rp 1.000.000 s/d Rp 1.500.000 mempunyai tingkat hidup yang lebih baik dibanding dengan upah/gaji sebulan kurang dari Rp 500.000, dan pekerja dengan status menikah mempunyai tingkat hidup yang lebih baik sebesar 1.8 kali dibandingkan dengan status yang belum menikah.

Sedangkan tingkat ketepatan prediksi model dengan analisis regresi ordinal untuk mengelompokkan tingkat hidup pekerja yang dipengaruhi empat variabel (Pendidikan (X1), Status pekerjaan (X2), Upah/Gaji Sebulan (X3) dan Status perkawinan (X4)) secara keseluruhan sebesar 54,6%. Untuk lebih jelas dapat dilihat pada Tabel 2.

Tabel 1. Ketepatan Klasifikasi Tingkat Hidup Pekerja dengan Analisis Diskriminan

Kelompok	Prediksi			Ketepatan Kelompok
	Kurang Baik	Sama Baik	Lebih Baik	
Kurang Baik	0	0	81	0
Sama Baik	0	0	125	0
Lebih Baik	0	0	248	100,0%
Ketepatan Keseluruhan				54,6 %

Sumber: Data diolah

Tabel 2. Ketepatan Klasifikasi Tingkat Hidup Pekerja dengan Regresi Ordinal

Kelompok	Prediksi			Ketepatan Kelompok
	Kurang Baik	Sama Baik	Lebih Baik	
Kurang Baik	0	0	81	0
Sama Baik	0	0	125	0
Lebih Baik	0	0	248	100,0 %
Ketepatan Keseluruhan				54,6 %

Sumber: Data diolah

Berdasarkan pada Tabel 2 diatas, variabel yang mempengaruhi tingkat hidup pekerja mampu menerangkan ketepatan klasifikasi pada kelompok 'kurang baik', 'sama baik' dan 'lebih baik' sebesar 54,6 persen dan kesalahan sebesar 45,4 persen. Tingkat ketepatan prediksi model untuk mengelompokkan 'kurang baik' secara keseluruhan 0 persen, 'sama baik' secara keseluruhan 0 persen dan 'lebih baik' secara keseluruhan 100 persen (248 dari 248).

Kesalahan pengklasifikasian pada metode diatas terjadi karena beberapa faktor lain selain faktor tersebut di atas, karena faktor-faktor lain seperti fasilitas kerja, budaya kerja belum dimasukkan dalam model. Untuk kriteria kestabilan pengelompokkan dengan *Press-Q* metode ini signifikan secara statistik, yang berarti ketepatan klasifikasi signifikan. Hal ini menunjukkan bahwa hasil pengelompokkan dengan metode ini lebih stabil dalam memprediksi kedepan.

Untuk lebih akurat mengenai presentasi ketepatan pengklasifikasian pada regresi ordinal, berikut disajikan hasil perhitungan *Press's Q*:

$$\text{Pr ess}'Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[454 - (248 \times 3)]^2}{454(3 - 1)} = 92.62$$

Nilai *Press'Q* = 92.62 dibandingkan $\chi^2_{2,0.05} = 5.9915$, ternyata nilai *Press'Q* lebih besar dari nilai chi-square, sehingga keakuratan pengklasifikasian taraf hidup pekerja adalah konsisten.

Tingkat Hidup Pekerja Dengan MARS

Fungsi pengelompokkan dalam mem-bedakan tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' adalah sebagai berikut :

$$\hat{f}(x) = 2.473 - 0.192 \text{ BF1}$$

dimana:

BF1 = Status pekerjaan (X2)

Dari fungsi tersebut, menunjukkan bahwa Status pekerjaan (X2) merupakan variabel yang penting pertama dengan kontribusi 100 persen dalam pengelompokkan tingkat hidup pekerja. Sedangkan tingkat ketepatan prediksi model dengan *MARS* untuk mengelompokkan tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' berdasarkan fungsi tersebut diatas, secara keseluruhan sebesar 55,5%, sedangkan pada tingkat hidup pekerja 'kurang baik' sebesar 1,23%, tingkat hidup pekerja 'sama baik' sebesar 2,4 % dan pada tingkat hidup pekerja 'lebih baik' sebesar 100%. Untuk lebih jelas dapat dilihat pada Tabel 3.

Untuk lebih akurat mengenai presentasi ketepatan pengklasifikasian pada pendekatan *MARS*, berikut disajikan hasil perhitungan *Press's Q*:

$$\text{Pr ess}'Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[454 - (252 \times 3)]^2}{454(3 - 1)} = 100.44$$

Nilai *Press'Q* = 100.44 dibandingkan $\chi^2_{2,0.05} = 5.9915$, ternyata nilai *Press'Q* lebih besar dari nilai chi-square, sehingga keakuratan pengklasifikasian taraf hidup pekerja adalah konsisten.

Tingkat Hidup Pekerja dengan Bootstrap

Ketepatan prediksi model dengan *bootstrap* sebanyak 1000 kali pada analisis diskriminan, regresi ordinal dan *MARS* untuk mengelompokkan kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' secara keseluruhan dapat dilihat pada Tabel 4, Tabel 5 dan Tabel 6..

Tabel 3. Ketepatan Klasifikasi Tingkat Hidup Pekerja dengan MARS

Kelompok	Prediksi			Ketepatan Kelompok
	Kurang Baik	Sama Baik	Lebih Baik	
Kurang Baik	1	0	80	1,23 %
Sama Baik	0	3	122	2,4 %
Lebih Baik	0	0	248	100,0 %
Ketepatan Keseluruhan				55,5 %

Sumber: Data diolah

Tabel 4. Ketepatan Klasifikasi Tingkat Hidup Pekerja dengan Bootstrap Diskriminan

Kelompok	Prediksi			Ketepatan Kelompok
	Kurang Baik	Sama Baik	Lebih Baik	
Kurang Baik	2,46 % (2)	3,70% (3)	93,84% (76)	2,46%
Sama Baik	0	5,60% (7)	94,40% (118)	5,60%
Lebih Baik	0	0	100,0% (248)	100,0%
Ketepatan Keseluruhan				56,6 %

Sumber: Data diolah

Tabel 5. Ketepatan Klasifikasi Tingkat Hidup Pekerja dengan Bootstrap Regresi Ordinal

Kelompok	Prediksi			Ketepatan Kelompok
	Kurang Baik	Sama Baik	Lebih Baik	
Kurang Baik	0	0	100,0% (81)	0,0 %
Sama Baik	0	0	100,0% (125)	0,0 %
Lebih Baik	0	0	100,0% (248)	100,0%
Ketepatan Keseluruhan				54,6 %

Sumber: Data diolah

Tabel 6. Ketepatan Klasifikasi Tingkat Hidup Pekerja dengan Bootstrap MARS

Kelompok	Prediksi			Ketepatan Kelompok
	Kurang Baik	Sama Baik	Lebih Baik	
Kurang Baik	4,94 % (4)	1,23% (1)	93,84% (76)	4,94%
Sama Baik	0	5,60% (7)	94,40% (118)	5,60%
Lebih Baik	0	0	100,0% (248)	100,0%
Ketepatan Keseluruhan				57,0 %

Sumber: Data diolah

Berdasarkan pada Tabel 4, Tabel 5 dan Tabel 6, ketepatan klasifikasi secara keseluruhan pada kelompok 'kurang baik', 'sama baik' dan 'lebih baik' pada pendekatan bootstrap MARS memberikan hasil yang lebih

baik dibandingkan dengan pendekatan yang lain yaitu sebesar 57,0%, sedangkan pendekatan yang lain sebesar 56,6 % dan 54,6%.

Kinerja Perbankan dengan Analisis Diskriminan

Analisis diskriminan yang merupakan salah satu metode statistik yang sering digunakan dalam masalah pengelompokan. Analisis ini dilakukan terhadap dua kelompok data yang sudah valid pengelompokannya, yaitu kelompok 'Kinerja Bank Sehat' dan kelompok 'Kinerja Bank Tidak Sehat'.

Analisis diskriminan pada kelompok Kinerja Bank sehat dan tidak sehat didasarkan pada 22 variabel. Secara univariate dengan menggunakan statistik uji F dan tingkat signifikansi 0,05%, diperoleh 5 variabel yang rata-ratanya berbeda antara Kinerja Bank Sehat dan Kinerja Bank Tidak Sehat ($p\text{-value} < 0,0$). Variabel-variabel tersebut adalah $X_1, X_2, X_4, X_5, X_7, X_9, X_{10}, X_{11}, X_{15}, X_{16}, X_{17}, X_{18}, X_{19}$ dan X_{21} . Selanjutnya secara multivariate dengan prosedur stepwise (bertatar) diperoleh informasi variabel-variabel mana yang membedakan antara Kinerja Bank Sehat dan Kinerja Bank Tidak Sehat. Hasil pengolahan menunjukkan bahwa variabel-variabel X_1, X_9 dan X_{21} sangat mempengaruhi dalam membedakan pengelompokan terhadap Kinerja Bank Sehat dan Kinerja Bank Tidak Sehat. Sedangkan hubungan antara variabel pembeda dengan dua kelompok (Kinerja Bank Sehat dan Kinerja Bank tidak sehat) ditunjukkan oleh korelasi kanonik yaitu sebesar 0,565.

Fungsi diskriminan yang membedakan antara Kinerja Bank Sehat dan Kinerja Bank Tidak Sehat adalah sebagai berikut :

$$F = -0.451 + 1.911 X_1 + 3.542 X_9 - 2.034 X_{21}$$

Fungsi diskriminan di atas mampu menerangkan ketepatan klasifikasi pada Kinerja Bank sehat dan Kinerja Bank Tidak sehat sebesar 83,6%. Sehingga fungsi diskriminan dapat digunakan sebagai prediksi untuk kedepan. Sebagai hasilnya jika titik pembatas yang ditentukan berdasarkan fungsi diskriminan yang diperoleh dan dengan memasukkan nilai rata-rata variabel pembeda pada masing-masing kelompok (Kinerja Bank Sehat dan Kinerja Bank Tidak Sehat). Maka titik pembatas tersebut menunjukkan bahwa jika nilai-nilai variabel pembeda dimasukkan dalam

fungsi diskriminan, dengan nilai diskriminan lebih besar dari (-0,062) dimasukkan dalam kelompok Kinerja Bank Sehat, sedangkan bila lebih kecil dari (-0,062) dimasukkan dalam kelompok Kinerja Bank Tidak Sehat.

Tingkat ketepatan prediksi model dengan analisis diskriminan untuk mengelompokkan Kinerja Bank Sehat dan Kinerja Bank Tidak Sehat secara keseluruhan sebesar 83,6%, sedangkan pada Kinerja Bank Sehat sebesar 78,3% dan pada Kinerja Bank Tidak Sehat sebesar 90,0%. Untuk lebih jelas dapat dilihat pada Tabel 7.

Untuk lebih akurat mengenai presentasi ketepatan pengklasifikasian pada analisis diskriminan dapat dibandingkan hasil perhitungan Press's Q sebagai berikut:

$$\text{Press'Q} = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[110 - (92 \times 2)]^2}{110(2 - 1)} = 49.782$$

Nilai Press'Q = 49.782 dibandingkan $\chi^2_{1,0,05} = 3.841$, ternyata nilai Press'Q lebih besar dari nilai chi-square, sehingga keakuratan pengklasifikasian Kinerja Bank adalah konsisten.

Kinerja Perbankan dengan Regresi Ordinal

Dari 21 variabel yang mempengaruhi pengelompokan kinerja perbankan, ternyata dengan analisis regresi ordinal terdapat 7 variabel yang mempengaruhi. Tujuh variabel tersebut yaitu *Gross Profit Margin* (X_1) dengan Sig. = 0,001 < $\alpha=0,05$, *Net Interest Margin* (X_2) dengan Sig. = 0,000 < $\alpha=0,05$, *Net Non Interest Margin* (X_4) dengan Sig. = 0,000 < $\alpha=0,05$, *Return on Asset* (X_5) dengan Sig. = 0,000 < $\alpha=0,05$, *Return on Equity* (X_6) dengan Sig. = 0,001 < $\alpha=0,05$, *Net Profit Margin* (X_7) dengan Sig. = 0,000 < $\alpha=0,05$ dan *Aktiva Produktif yang diklasifikasikan* (X_{19}) dengan Sig. = 0,008 < $\alpha=0,05$. Sedangkan variabel yang lain tidak signifikan secara statistik.

Sedangkan tingkat ketepatan prediksi model dengan analisis regresi ordinal untuk mengelompokkan kinerja perbankan secara keseluruhan sebesar 95,5%. Untuk lebih jelas dapat dilihat pada Tabel 8 berikut.

Tabel 7. Ketepatan Klasifikasi Kinerja Perbankan dengan Analisis Diskriminan

Klasifikasi Kinerja Bank Sebenarnya	Klasifikasi Kinerja Bank Prediksi		Ketepatan Kelompok
	Kinerja Bank Sehat	Kinerja Bank Tidak Sehat	
Kinerja Bank Sehat	47	13	78,3 %
Kinerja Bank Tidak Sehat	5	45	90,0 %
Ketepatan Keseluruhan			83,6 %

Sumber: Data diolah

Tabel 8. Ketepatan Klasifikasi Kinerja Perbankan dengan Regresi Ordinal

Klasifikasi Kinerja Bank Sebenarnya	Klasifikasi Kinerja Bank Prediksi		Ketepatan Kelompok
	Kinerja Bank Sehat	Kinerja Bank Tidak Sehat	
Kinerja Bank Sehat	58	2	96,7 %
Kinerja Bank Tidak Sehat	3	47	94,0 %
Ketepatan Keseluruhan			95,5 %

Sumber: Data diolah

Untuk lebih akurat mengenai presentasi ketepatan pengklasifikasian pada analisis diskriminan dapat dibandingkan hasil perhitungan Press's Q sebagai berikut:

$$Pr ess'Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[110 - (95 \times 2)]^2}{110(2 - 1)} = 90,91$$

Nilai Press'Q = 90,91 dibandingkan $\chi^2_{1,0,05} = 3,841$, ternyata nilai Press'Q lebih besar dari nilai chi-square, sehingga keakuratan pengklasifikasian Kinerja Bank adalah konsisten.

Kinerja Perbankan dengan MARS

Pengelompokkan dengan pendekatan MARS selain diperoleh ketepatan klasifikasi juga diketahui fungsi pembeda. Adapun fungsi pengelompokkan dalam membedakan 'Kinerja Bank Sehat' dan 'Kinerja Bank Tidak Sehat' adalah sebagai berikut:

$$\hat{f}(x) = 1.013 + 1.932 BF1 + 9.595 BF2 - 9.663 BF3$$

dimana: BF1 = $(0,148 + X_9)$, BF2 = $(X_1 - 0,881)$

BF3 = $(X_{21} - 0,962)$
 X_1 = Gross Profit Margin
 X_9 = Interest Margin on Loans
 X_{21} = Non Performing Loans to Total Loans

Persamaan tersebut, menunjukkan bahwa X_1 = Gross Profit Margin merupakan variabel yang penting pertama dengan kontribusi 100 persen dalam penge-lompokkan kinerja bank, selanjutnya X_9 = Interest Margin on Loans merupakan variabel yang penting kedua dengan kontribusi 15,75 persen, dan X_{21} merupakan variabel yang tidak memberikan kontribusi dalam membedakan pengelompokkan kinerja bank.

Sedangkan tingkat ketepatan prediksi model dengan MARS untuk menge-lompokkan Kinerja Bank Sehat dan Kinerja Bank Tidak Sehat berdasarkan fungsi pembeda diatas, secara keseluruhan sebesar 89,1%, sedangkan pada Kinerja Bank Sehat sebesar 85,0% dan pada Kinerja Bank Tidak Sehat sebesar 94,0%. Untuk lebih jelas dapat dilihat pada Tabel 9 berikut.

Tabel 9. Ketepatan Klasifikasi Kinerja Perbankan dengan MARS

Klasifikasi Kinerja Bank Sebenarnya	Klasifikasi Kinerja Bank Prediksi		Ketepatan Kelompok
	Kinerja Bank Sehat	Kinerja Bank Tidak Sehat	
Kinerja Bank Sehat	51	9	85,0 %
Kinerja Bank Tidak Sehat	3	47	94,0 %
Ketepatan Keseluruhan			89,1 %

Sumber: Data diolah

Berdasarkan Tabel 9, dapat ditunjukkan akurasi mengenai presentasi ketepatan pengklasifikasian dengan perhitungan Press's Q sebagai berikut:

$$\text{Press'Q} = \frac{[N - (nK)]^2}{N(K-1)} = \frac{[110 - (98 \times 2)]^2}{110(2-1)} = 67.23$$

Nilai Press'Q = 64.38 dibandingkan $\chi^2_{1,0.05} = 3.841$, ternyata nilai Press'Q lebih besar dari nilai *chi-square* tabel ($\chi^2_{1,0.05}$), sehingga keakuratan pengklasifikasian Kinerja Bank dengan pendekatan MARS adalah konsisten. Selain itu juga ditunjukkan bahwa prediksi kinerja Bank Sehat sebesar 68.75 kali dari kinerja Bank Tidak Sehat.

Kinerja Perbankan dengan Bootstrap

Ketepatan prediksi model dengan *bootstrap* sebanyak 1000 kali pada analisis diskriminan, regresi ordinal dan MARS untuk mengelompokkan kelompok tingkat hidup pekerja 'kurang baik', 'sama baik' dan 'lebih baik' secara keseluruhan dapat dilihat pada Tabel 10, Tabel 11 dan Tabel 12. Berdasarkan pada Tabel 10, Tabel 11 dan Tabel 12, ketepatan klasifikasi secara keseluruhan pada kelompok 'kinerja bank sehat dan 'kinerja bank tidak sehat' pada pendekatan bootstrap regresi ordinal mem-berikan hasil yang lebih baik dibandingkan dengan pendekatan yang lain yaitu sebesar 96,4%, sedangkan pendekatan yang lain sebesar 90,0 % dan 86,4 %.

Tabel 10. Ketepatan Klasifikasi Kinerja Perbankan dengan Bootstrap Diskriminan

Klasifikasi Kinerja Bank Sebenarnya	Klasifikasi Kinerja Bank Prediksi		Ketepatan Kelompok
	Kinerja Bank Sehat	Kinerja Bank Tidak Sehat	
Kinerja Bank Sehat	85,0 % (56)	15,0 % (4)	85,0 %
Kinerja Bank Tidak Sehat	12,0 % (6)	88,0 % (44)	88,0 %
Ketepatan Keseluruhan			86,4 %

Sumber: Data diolah

Tabel 11. Ketepatan Klasifikasi Kinerja Perbankan dengan Bootstrap Diskriminan

Klasifikasi Kinerja Bank Sebenarnya	Klasifikasi Kinerja Bank Prediksi		Ketepatan Kelompok
	Kinerja Bank Sehat	Kinerja Bank Tidak Sehat	
Kinerja Bank Sehat	98,3 % (59)	1,7 % (1)	98,3 %
Kinerja Bank Tidak Sehat	6,0 % (3)	94,0 % (47)	94,0 %
Ketepatan Keseluruhan			96,4 %

Sumber: Data diolah

Tabel 12. Ketepatan Klasifikasi Kinerja Perbankan dengan Bootstrap Diskriminan

Klasifikasi Kinerja Bank Sebenarnya	Klasifikasi Kinerja Bank Prediksi		Ketepatan Kelompok
	Kinerja Bank Sehat	Kinerja Bank Tidak Sehat	
Kinerja Bank Sehat	94,17 % (57)	5,83 % (3)	94,17 %
Kinerja Bank Tidak Sehat	15,97 % (8)	84,03 % (42)	84,03 %
Ketepatan Keseluruhan			90,0 %

Sumber: Data diolah

KESIMPULAN

Berdasarkan analisis dan pembahasan dapat ditarik suatu kesimpulan yaitu dalam menentukan variabel pembeda antar kelompok terdapat perbedaan variabel pada masing-masing metode. Hal ini dikarenakan pada masing-masing metode mempunyai kekhasan baik dari terpenuhinya asumsi maupun estimasinya. Begitu juga pada ketepatan klasifikasi antar kelompok terdapat perbedaan khususnya pada matriks varians kovarians yang tidak sama pada kasus tingkat hidup pekerja.

Secara keseluruhan dapat disimpulkan bahwa untuk masalah ketepatan klasifikasi pendekatan bootstrap pada masing-masing metode memberikan kesalahan yang kecil baik pada matriks varians kovarians yang tidak sama maupun matriks varians kovarians yang sama.

DAFTAR PUSTAKA

- Agresti, A., 1990. *Categorical Data Analysis*, John Wiley and Sons, New York.
- Buja, A., Duffy, D., Hastie, T. and Tibshirani, R., 2001. *Multivariate adaptive regression splines* (with discussion).
- Dillon, W.R. and M. Goldstein, 1978. On the Performance of Some Multinomial Classification Rules, *Journal of American Statistical Association*, **73**, pp.305-313
- Friedman, J.H., 1990. Estimating functions of mixed ordinal and categorical variables using multivariate adaptive regression splines. *Technical Report LCS 107*, Statistics Department, Stanford University.
- Friedman, J.H., 1991. *Multivariate Adaptive Regression Splines (With Discussion)*. Stanford California 94309.
- Friedman, J.H. and Silverman, B.W., 1989. Flexible parsimony smoothing and additive modeling. *Technometrics*, **31**, 3 – 39.
- Hair J.F., Rolph E. Anderson, Ronald L. Tatham, William C. Black, 1998. *Multivariate Data Analysis*. Fifth Edition, Prentice Hall International, Inc.
- Johnson R.A. and Wichern D.W., 1992. *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, New Jersey.
- Otok, B.W., Guritno, S. and Subanar, 2004. *Analisis Diskriminan dan MARS untuk Klasifikasi Perbankan di Indonesia*. Seminar FKMS3MI ke II, UGM, Yogyakarta.
- Portier, K.M, 2001. *Multivariate Statistical Methods*, STA4702/5701.
- Sharma, S., 1996. *Applied Multivariate Techniques*, John Wiley & Sons, Inc.