

**BKS PTN-BMIPA**

**2012**

**mti**

# Prosiding

**BIDANG  
MATEMATIKA**

## SEMINAR & RAPAT TAHUNAN

BKS-PTN B Tahun 2012

BIDANG ILMU MIPA

Badan Kerjasama Perguruan Tinggi Negeri  
Wilayah Barat

*Tema :*

*Peran MIPA dalam Pengembangan  
SDM dan SDA*

Hotel Madani Medan

11 - 12 Mei 2012



Penyelenggara  
FMIPA  
UNIVERSITAS  
NEGERI MEDAN



Jl. Willem Iskandar, Psr V Medan 20221

Telp. (061) 6625970 Medan

[www.semirataunimed.com](http://www.semirataunimed.com) Email: [semiratabks2012@yahoo.co.id](mailto:semiratabks2012@yahoo.co.id)

## DAFTAR ISI

HALAMA

**Kata Pengantar dari Editor**

**Kata Sambutan Ketua Panitia**

**Kata Sambutan Ketua BKS-PTN B Bidang MIPA**

**Kata Sambutan Rektor Universitas Negeri Medan**

### DAFTAR ISI

Admi Nazra	A Lower- Bound of the Number of Diffeomorphism Classes Of Real Boot Manifolds	1	8
Ahmad Iqbal Baqi	Estimasi Fertilitas Provinsi Sumatera Utara 1995-2005 Dengan Menggunakan Metoda Antar Survei	9	12
Alfirman	Pengendalian putaran Motor Stepper dengan Menggunakan Port Paralel Komputer	13	17
Asep Rusyana	Rancangan Faktorial Dengan Pengamatan Berulang Untuk Mengidentifikasi Pengaruh Mulsa Dan Jarak Tanam Terhadap Radiasi Surya Pada Kacang Kedelai	18	22
Asmara Karma	Pemakaian Transformasi Baru Elzaki dalam Menyelesaikan Persamaan Differensial	23	27
Aziskhan	Penggunaan Persamaan Diferensial geometri dalam menyelesaikan persoalan pada elektrostatika	28	31
Budi Rudianto	Penerapan Metode Graf Multi- Transformasi Pada Penyelesaian Sirkuit Elektronik	32	37
Eduward H Hutabarat	Persamaan dan Fungsi Potensial Kompleks airfoil Dalam Analisis Transformasi Joukowski	38	43
Dian Kurniasari	Model Berperingkat Tidak Penuh Pada Data Spasial Dengan Metode Dekomposisi Spektral	44	49
Dodi Devianto	Sebaran Eksponensial Terbagi Tak Hingga	50	53
Efendi	Konstruksi Model Untuk Melihat Pengaruh Bentuk Geometri Habitat Pada Perkembangan Populasi Aedes Dengan Bentuk Geometri Habitat Kerucut.	54	61
Effendi	Algorithma String Pada Bioinformatik	62	64
Evfi Mahdiyah	Analisa dan Pengembangan artificial Intelligence Markup Language (AIML) Tentang Istilah Komputer Dalam Bahasa Indonesia Menggunakan Alice chat bot	65	69
Fatayat	Penerapan Metode <i>Neural Network</i> Dalam Prediksi Persediaan Darah Pertahun Pada PMI Rumah Sakit	70	75
Johannes Kho	Perbaikan Metode Secant Steffensen Untuk Menyelesaikan Persamaan Nonlinier	76	79
Leli Deswita	Pemodelan Matematika Bagi Aliran Syaraf Batas Konveksi Bebas pada Flat Horizontal	80	83
M. D. H. Gamal	Penjadwalan Perawat Dengan Menggunakan Pemrograman Tujuan	84	92
M. Natsir	Superstruktur Umum dan Optimisasi Global Proses Desain Jaringan Air Terpadu.	93	98
Machudor Yusman M	Konstruksi Algoritma Sorting Berdasarkan Indeks Data	99	104
Nonong Amalita	Estimasi Parameter pada Distribusi Rayleigh untuk Sampel Lengkap dan Tersensor	105	110

Ridha Ferdhiana	Pendugaan Selang Kepercayaan Koefisien Korelasi Pearson menggunakan Metode Bootstrap	111 - 115
Riri Lestari	Batas Exercise Opsi Put Amerika	116 - 117
Sugandi Yahdin	Model Keputusan Membeli Di Pasar Tradisional Dengan Metode Regresi Logistik Biner	118 - 122
Syafruddin	Pelabelan Supersisi Ajaib Dari Suatu Graf $(n,2)$ -KITE	123 - 126
Syarifah Meurah Yuni	Model Matematika Resistensi Parasit Plasmodium falciparum Terhadap Obat Tunggal dan Obat Campuran Antimalaria	127 - 132
Yusmet Rizal	Suatu Penyajian Geometris Grup Fungsi pada Himpunan $\{1, 2, 3, 4\}$	133 - 138
Hazmira Yozza	kajian Perbandingan Beberapa metode Klasifikasi	139 - 147
Helmi	Metode Transformasi Sumudu Dalam Penyelesaian Persamaan Diferensial Parsial Linear Order Dua	148 - 156
Indrawati	Perapihan dan Proyeksi Penduduk Sumatera Selatan Berdasarkan Tingkat Fertilitas Total (Total Fertility Rate) dan Rasio Jenis Kelamin (Sex Ratio)	157 - 167
Intan Syahrini	Algoritma Genetik Untuk Masalah Optimisasi Program Non Linier Genetic Algorithm For Nonlinear Program Optimization Problem	168 - 175
Joko Risanto	Algoritma Menghitung Nilai Kesesuaian Menggunakan Metode Lickert dalam Suatu Analisa SWOT Perencanaan Srategis.	176 - 184
Marzuki	Pendugaan Model Regresi dengan Regresi Fuzzy	185 - 191
Media Rosha	PENGGUNAAN PENALARAN TRANSFORMASIONAL DALAM BERFIKIR KREATIF MATEMATIK DARI PERMASALAHAN MULTINOMIAL $(a_1 + a_2 + \dots + a_k)^n$	192 - 202
Nina Fitriyati	HISTORY MATCHING OF ONE-DIMENSIONAL HOMOGENOUS RESERVOIR PARAMETER FOR TWO INTERACTING WELLS	203 - 210
Novi Reandy Sasmita	Perbandingan Metode Fuzzy C-Means (FCM) dan Fuzzy C-Shell (FCS) Menggunakan Data Citra Satelit Quickbird (Studi Kasus Daerah Peukan Bada, Aceh Besar)	211 - 218
Pepi Novianti	Kajian Circular Descriptive Statistics Pada Data Yang Berupa Arah Dan Sudut	219 - 225
Rahma Zuhra	Kajian Tentang Integral Daniell	226 - 231
Ramya Rachmawati	Penerapan Pemrograman Dinamis Dalam Sistem Inventori	232 - 238
Riry Sriningsih	MODEL MATEMATIKA PENGARUH VAKSINASI TERHADAP PENYEBARAN FLU BURUNG PADA POPULASI UNGGAS & MANUSIA	239 - 249

Rolan Pane	Menyelesaikan Persamaan Non Linier dengan Metode Iterasi Parsial yang diturunkan Menggunakan Integral Parsial	250 - 259
Syafriandi	PEMETAAN KABUPATEN/KOTA DI SUMATERA BARAT BERDASARKAN PERSENTASE PENGUASAAN MATERI MATA PELAJARAN YANG DI-UN-KAN MENGGUNAKAN ANALISIS GEROMBOL	260 - 267
Yulita Molliq Rangkuti	KEAKURATAN METODE ITERASI VARIASI UNTUK MENYELESAIKAN MASALAH DEFLEKSI BALOK KANTILEVER DENGAN BEBAN TERDISTRIBUSI SECARA SERAGAM.	268 - 274
Yundari	Ruang Fungsi Holder	275 - 281
Zul Amry	Analisis Time Series Angka Inflasi Nasional Dengan Model Arima.	282 - 294
Zulakmal	Menentukan Solusi Persamaan Laplace Dua Dimensi Yang Mempunyai Syarat Batas Robin Dengan Metoda Dekomposisi Adomian	295 - 299
Zulfia Memi Mayasari	Pengembangan Tapis Morfologi Matematik Menggunakan Teori Ordered set dan lattice	300 - 306
Nanci Nababan	Pemodelan keputusan membeli di pasar tradisional Dengan metode regresi logistik biner (studi kasus di pasar cinde)	307 - 330
Zaiful Bahri	Perbandingan Metode Moment Invariant Hu Dan Metode Deskriptor Fourier Dalam Pengenalan Pola Karakter	331 - 340
Agus Salim	Penentuan Peluang Kesalahan Pelepasan Partikel Minyak Menggunakan Fault Tree Analysis (Fta)	341 - 351
Mulyono	Teorema Kekonvergenan pada Integral-C	352 - 358
Haposan Sirait	Tipe Penaksir Rasio Variansi Dalam Sampling Acak Sederhana	359 - 366
Arnellis	Kunci Publik Elliptic Curve System	367 - 374
T.P Nababan	Eksistensi Solusi Optimum Dalam Analisa Sistem Persediaan Tanpa Shortage*	375 - 382

## KAJIAN PERBANDINGAN BEBERAPA METODE KLASIFIKASI

Hazmira Yozza, Izzati Rahmi HG  
Jurusan Matematika FMIPA Universitas Andalas

### ABSTRAK

Dalam statistika, terdapat beberapa metode yang dapat digunakan untuk mengklasifikasikan objek berdasarkan nilai pengamatan terhadap beberapa peubah. Beberapa di antaranya adalah Analisis Regresi Logistik, Analisis Klasifikasi Pohon dan Analisis Diskriminan. Pada tulisan ini, akan ditelaah perbandingan beberapa metode klasifikasi tersebut.

**Kata Kunci :** *analisis regresi logistik, analisis klasifikasi pohon, analisis diskriminan*

### ABSTRACT

There are various statistical methods used to classify objects into groups on the basis of their measurements on a set of independent variables. Some of them are Logistic regression, Discriminant Analysis and Classification Tree. In this paper, those methods are compared, based on data requirements, assumptions and result of analysis.

**Keywords :** *logistic regression, discriminant analysis, classification tree.*

### A. PENDAHULUAN

Masalah klasifikasi sering dijumpai pada kehidupan sehari-hari. Di bidang pendidikan contohnya adalah pengklasifikasian sekolah berdasarkan akreditasi sekolah. Di bidang kesehatan dilakukan pengklasifikasian penyakit berdasarkan tingkat keseriusannya. Di bidang biologi, masalah klasifikasi ditemui ketika mengklasifikasi hewan/tumbuhan ke dalam kelompok-kelompok berdasarkan kesamaan cirinya.

Dalam setiap masalah pengklasifikasian, seringkali ingin diketahui bagaimana aturan pengklasifikasi objek ke dalam kategori-kategorinya dan untuk kemudian menggunakan aturan tersebut untuk menduga keanggotaan dari suatu objek baru yang belum jelas berasal kategori yang mana. Aturan tersebut biasanya dibentuk berdasarkan pengukuran satu atau lebih variabel terhadap objek yang telah jelas kategorinya.

Beberapa metode pengklasifikasian yang sering digunakan adalah analisis diskriminan, analisis regresi logistik dan klasifikasi berstruktur pohon. Untuk mendapatkan klasifikasi yang tepat, perlu diperhatikan metode yang sesuai dengan jenis data dan tujuan analisis. Suatu hal yang menarik, adalah bagaimana perbandingan metode-metode pengklasifikasian tersebut.

Tulisan ini bertujuan untuk membandingkan beberapa metode pengklasifikasian objek yang biasa digunakan, yaitu analisis diskriminan, analisis regresi logistik dan klasifikasi berstruktur pohon. Perbandingan dibatasi pada data yang digunakan, asumsi dan hasil analisis.

## 1. LATAR BELAKANG TEORITIS

### 1.1. Analisis Regresi Logistik

Analisis regresi adalah suatu analisis statistika yang digunakan untuk mengetahui hubungan antara satu atau lebih variabel penjelas  $X_i$  ( $i=1,2,\dots,k$ ) dengan suatu variabel respons  $Y$ . Hubungan tersebut dinyatakan dalam suatu model linier sebagai berikut :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Dalam analisis regresi ini diasumsikan bahwa galat  $\varepsilon$  merupakan variabel acak yang saling bebas dari suatu sebaran  $N(0, \sigma^2)$ . Dengan demikian, variabel respons  $Y$  juga akan merupakan variabel acak  $N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k, \sigma^2)$ . Jika  $Y$  merupakan peubah kategorik, maka asumsi-asumsi tersebut terlanggar, sehingga analisis regresi biasa tidak dapat digunakan. Analisis regresi logistik adalah analisis regresi yang digunakan pada kondisi ini.

Misalkan dimiliki  $n$  pengamatan binomial dalam bentuk  $\tilde{p}_i = y_i/n_i$   $i = 1, 2, \dots, n$  dengan  $y_i$  dan  $\tilde{p}_i$  berturut-turut adalah banyak kejadian berhasil dari  $n_i$  kejadian dan peluang berhasil untuk pengamatan ke- $i$ . Model logistik linier yang dapat digunakan untuk menggambarkan ketergantungan dengan peubah-peubah penjelas  $x_1, x_2, \dots, x_k$  adalah :

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad \dots \dots \dots (1)$$

Model regresi logistic dibentuk melalui suatu transformasi logit terhadap model (1) menjadi :

$$\text{Logit } p = \log_e \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad \dots \dots \dots (2).$$

Terlihat bahwa model (2) adalah anggota dari kelas Model Linier Umum (*Generalized Linear Models*) fungsi *logit* ( $p$ ) sebagai fungsi penghubung (*link function*) (McCullagh & Nelder, 1989).

Untuk membentuk model regresi linier dari suatu data, terlebih dahulu harus dilakukan pendugaan terhadap  $k+1$  parameter  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  dengan menggunakan metode kemungkinan maksimum dengan fungsi kemungkinan :

$$L(\beta) = \prod_{i=1}^n \binom{n_i}{y_i} p_i^{y_i} (1-p_i)^{n_i - y_i} = \prod_{i=1}^n \binom{n_i}{y_i} \left( \frac{p_i}{1-p_i} \right)^{y_i} (1-p_i)^{n_i}$$

Fungsi kemungkinan ini tergantung pada peluang keberhasilan  $p_i$  yang dimodelkan sebagai kombinasi linier dari  $\beta$ , yang berarti bahwa fungsi kemungkinan ini merupakan fungsi dari  $\beta$ . Penduga kemungkinan maksimum bagi  $\beta$  diperoleh dengan menentukan nilai-nilai  $\beta$  yang dapat memaksimumkan fungsi kemungkinan tersebut, yaitu dengan mengevaluasi  $\partial L(\beta)/\partial \beta_i = 0$ .

Memaksimumkan nilai  $L(\beta)$  sama dengan memaksimumkan :

$$\log_e L(\beta) = \sum_{i=1}^n \left\{ \log_e \binom{n_i}{y_i} + y_i \eta_i - n_i \log_e (1 + e^{\eta_i}) \right\}$$

Turunan fungsi log kemungkinan ini terhadap  $k+1$  parameter  $\beta$  adalah :

$$\frac{\partial \log_e L(\beta)}{\partial \beta_j} = \sum_{i=1}^n y_i x_{ji} - \sum_{i=1}^n n_i x_{ji} e^{\eta_i} (1 + e^{\eta_i})^{-1} \quad j = 0, 1, 2, \dots, k$$

Evaluasi turunan ini pada nilai-nilai  $\beta$  dan menyamakannya dengan 0 akan menghasilkan  $k + 1$  persamaan non linier. Persamaan-persamaan hanya dapat diselesaikan secara numerik. Salah satunya dengan metode *scoring Fisher*.

## 1.2. Analisis Diskriminan

Analisis diskriminan adalah suatu analisis statistika yang digunakan untuk menentukan variabel yang membedakan dua atau lebih kelompok objek. Analisis ini dilakukan dengan mem-bentuk suatu fungsi diskriminan berdasarkan perhitungan statistik terhadap objek-objek yang te-lah diketahui dengan jelas dan mantap pengelompokannya. Fungsi diskriminan yang dihasilkan dapat digunakan untuk menduga kelompok dari suatu objek yang belum jelas berasal dari kelom-pok mana. Fungsi diskriminan yang dibangun dengan asumsi bahwa kelompok-kelompok memiliki ragam yang sama dinamakan fungsi diskriminan linier, sedangkan fungsi yang dibangun tanpa asumsi tersebut dinamakan fungsi diskriminan kuadratik.

Misalkan terdapat  $K$  kelompok dan dari setiap kelompok ditarik contoh acak, masing-masing berukuran  $n_1, n_2, \dots, n_k$  dan terhadap setiap objek pada contoh tersebut diukur  $p$  peubah ( $X_1, X_2, \dots, X_p$ ). Matriks ragam peragam total dari ( $X_1, X_2, \dots, X_p$ ),  $T$ , diuraikan menjadi :

$$T = W + B$$

dengan  $W$  dan  $B$  adalah matriks ragam peragam dalam kelompok dan antar kelompok.

Dalam analisis diskriminan, dibentuk fungsi diskriminan yang merupakan kombinasi linier dari peubah-peubah asal (biasanya yang telah terkoreksi terhadap rataan) atau

$$Y = a^T X$$

Ragam dari fungsi diskriminan tersebut adalah :

$$\text{Var}(Y) = \text{Var}(a^T \tilde{X}) = a^T \text{Var}(\tilde{X})a = a^T T a = a^T W a + a^T B a$$

Fungsi tersebut dibentuk sehingga keragaman di dalam kelompok sangat kecil daripada keragaman antar kelompok atau dengan kata lain rasio antara keragaman antar kelompok dan dalam kelompok sekecil mungkin. Untuk itu,

perlu ditentukan vektor  $a$  sedemikian sehingga  $\lambda = \frac{a^T B a}{a^T W a}$  maksimum. Agar  $\lambda$  maksimum, maka haruslah :

$$\frac{\partial \lambda}{\partial a} = \frac{(2Ba)(a^T Wa) - (a^T Ba)(2Wa)}{(a^T Wa)^2} = \frac{(2Ba)}{(a^T Wa)} - \lambda \frac{(2Wa)}{(a^T Wa)} = 0$$

$$Ba - \lambda Wa = 0$$

Dengan mengalikan persamaan tersebut di sebelah kiri dengan  $W^{-1}$ , diperoleh :

$$W^{-1}Ba - \lambda a = 0 \Leftrightarrow (W^{-1}B - \lambda I)a = 0$$

Dari persamaan di atas, terlihat bahwa  $\lambda = \frac{a^T Ba}{a^T Wa}$  adalah nilai eigen dari matriks  $W^{-1}B$  dengan  $a$  merupakan vektor eigen padanannya.

Fungsi diskriminan pertama yang dibangun adalah  $Y_1 = a_1^T \tilde{X}$  dengan vektor  $a_1$  yang dapat memaksimalkan  $\lambda = \frac{a^T Ba}{a^T Wa}$ . Agar  $\lambda$  maksimum, harus dipilih  $\lambda_1$  yang merupakan nilai eigen terbesar dari  $W^{-1}B$  dengan  $a_1$  merupakan vektor eigen padanannya. Dengan memberikan batasan bahwa fungsi diskriminan pertama tidak berkorelasi dengan fungsi diskriminan kedua, ketiga dan seterusnya, maka akan diperoleh fungsi diskriminan ke- $i$  :

$$Y_i = a_i^T \tilde{X} \quad i = 1, 2, \dots, s$$

dengan  $a_i$  merupakan vektor eigen yang berpadanan dengan nilai eigen terbesar ke- $i$  dari matriks  $W^{-1}B$ . Bila  $K$  adalah banyaknya kelompok dan  $p$  adalah banyaknya peubah, maka maksimum banyaknya fungsi diskriminan yang dapat dibangun adalah  $s = \min\{K-1, p\}$ .

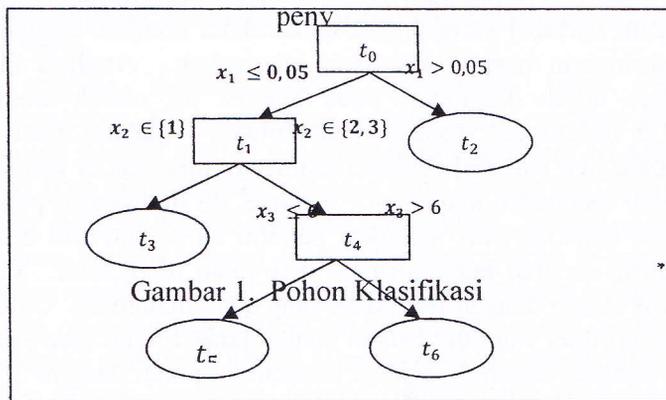
Analisis diskriminan biasa hanya dapat digunakan bila semua variabel penjelas merupakan variabel kontinu. Jika semua variabel penjelas adalah variabel diskret atau campuran, maka dapat digunakan Analisis Diskriminan Metode *Minimum Distance Probability* yang diperkenalkan oleh Villaroya, Rios dan Oller. Secara umum, metode ini dilakukan dengan membentuk suatu sekatan pada ruang contoh berdasarkan pada fungsi jarak yang sesuai, kemudian menggunakan kaidah Bayes untuk menempatkan setiap objek ke masing-masing kelompok. Metode ini juga dapat mempertimbangkan informasi mengenai peluang awal dan memberikan suatu ukuran tingkat kepercayaan pengklasifikasian untuk setiap pengamatan pada masing-masing kelas.

### 1.3. Klasifikasi Berstruktur Pohon

Metode klasifikasi berstruktur pohon merupakan metode statistika yang digunakan untuk memperkirakan keanggotaan amatan atau objek dalam kelas-kelas peubah respon kategorik, 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 himpunan data [5]. Gambar 1 adalah contoh pohon klasifikasi.

Dalam pohon klasifikasi seperti pada Gambar 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 disekek 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 atau kombinasi dari peubah. Peubah bebas yang digunakan untuk menyekat sebuah simpul dinamakan sebagai peubah penyekat. Peubah penyekat dapat berupa peubah kategorik maupun peubah kontinu. Pada Gambar 11 terlihat bahwa simpul  $t_0$  disekek 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  $t_1$  disekek lagi berdasarkan peubah  $x_2$  menjadi simpul  $t_3$  bila  $x_2 \in \{1\}$  dan  $t_4$  bila  $x_2 \in \{2,3\}$ . 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, atau (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 disekek menjadi dua simpul baru, sedangkan pada penyekatan non biner setiap simpul dapat menghasilkan lebih dari dua simpul baru. Gambar 1 merupakan pohon klasifikasi dengan penyekatan biner.

Terdapat beberapa algoritma yang dapat digunakan untuk menyekat suatu simpul. Dua di antaranya adalah metode CART dan QUEST

## 2. PEMBAHASAN

Dalam tulisan ini, akan dibandingkan metode-metode klasifikasi tersebut dalam beberapa sudut pandang

### 3.1. Data

Pada semua analisis yang dibandingkan, data terdiri dari satu peubah respons berupa data kategorik, dan beberapa variabel penjelas. Analisis diskriminan biasa mensyaratkan bahwa semua variabel penjelas minimal diukur dalam skala selang. Hal ini adalah karena analisis ini didasarkan pada matriks ragam peragam variabel di dalam dan antar kelompok, dimana matriks ini hanya akan dapat diperoleh untuk data dalam skala selang ataupun rasio.. Bila terdapat satu atau lebih variabel yang diukur dalam skala nominal ataupun ordinal, maka analisis diskriminan biasa tidak dapat digunakan. Analisis diskriminan yang dimungkinkan untuk digunakan pada kondisi ini adalah analisis diskriminan metode MDP (Minimum distance Probability). Analisis diskriminan ini dapat digunakan baik jika variabel penjelas semuanya merupakan variabel numerik atau kategorik atau campuran keduanya. Analisis ini didasarkan pada matriks kemiripan ataupun ketakmiripan, sehingga peneliti harus hati-hati dalam menentukan konsep kemiripan atau ketakmiripan yang akan digunakan. Pemilihan konsep tersebut harus sesuai dengan jenis data yang akan dianalisis. Untuk data numerik, konsep ketakmiripan yang digunakan adalah jarak Euclid atau jarak mahalanobis. Untuk data dengan variabel penjelas berskala nominal dengan 2 kategorik, dapat digunakan ukuran *simple matching*, *jackard*, *dice* dan lain-lain.

Analisis regresi logistic juga memungkinkan untuk digunakan bila variabel penjelasnya diukur dalam berbagai skala pengukuran, nominal, ordinal, selang ataupun rasio. Namun sebagaimana pada analisis regresi linier biasa, untuk variabel yang diukur dalam skala nominal ataupun ordinal, maka variabel ini harus ditransformasi terlebih dahulu menjadi satu atau lebih variabel dummy.

Untuk metode klasifikasi berstruktur pohon yang penyekatannya dilakukan berdasarkan satu variabel (univariate split), untuk peubah penjelas dengan skala nominal, suatu hal yang menarik adalah bahwa setiap transformasi yang dilakukan terhadap variabel tersebut, akan menghasilkan penyekatan yang sama dan memberikan kelompok dugaan yang sama untuk suatu objek. Demikian juga halnya dengan variabel yang diukur dengan skala ordinal. Setiap transformasi monotonik (yaitu transformasi yang dapat mempertahankan urutan antar nilai dari variabel tersebut) juga akan menghasilkan dugaan kelompok yang sama bagi objek. Dengan demikian metode klasifikasi pohon dapat digunakan untuk setiap skala pengukuran. Bila digunakan metode CART, perbedaannya hanyalah pada penetapan aturan dalam menentukan penyekatan terbaik antara variabel nominal dengan variabel ordinal, selang dan kategori. Bila digunakan metode QUEST, perbedaannya juga terletak pada aturan pencarian sekatan terbaik antara variabel nominal dan ordinal dan variabel selang dan rasio.

Terkait dengan ukuran data, analisis diskriminan biasa sangat sensitive terhadap rasio ukuran contoh terhadap banyaknya variabel penjelas. Banyak studi yang menyarankan untuk menggunakan rasio 20 pengamatan per satu variabel

penjelas. Meskipun dalam prakteknya rasio ini cukup sulit untuk dicapai, namun perlu dicatat bahwa hasil analisis akan tidak stabil dengan berkurangnya ukuran contoh relatif terhadap banyaknya variabel penjelas. Ukuran contoh minimum yang disarankan adalah 5 pengamatan per satu variabel penjelas. Satu hal penting lainnya yang harus diperhatikan adalah banyaknya pengamatan untuk masing-masing kelompok. Sebagai patokan, setiap kelompok harus memiliki minimal 20 pengamatan. Namun meskipun setiap kelompok telah memiliki lebih dari 20 pengamatan, ukuran relative dari setiap kelompok juga harus dipertimbangkan. Jika ukuran kelompok bervariasi cukup besar, maka hal ini akan mempengaruhi pendugaan fungsi diskriminan dan juga pengklasifikasian objek yang dilakukan berdasarkan fungsi diskriminan tersebut. Pada saat pengklasifikasian objek ke dalam salah satu kelompok, maka kelompok dengan ukuran contoh lebih besar akan memiliki kesempatan yang lebih besar menjadi kelompok dugaan bagi objek tersebut. (Hair, 1998)

Untuk analisis regresi logistic disarankan untuk menggunakan menggunakan minimum 10 pengamatan untuk setiap variabel penjelas yang digunakan. Contoh dengan jumlah pengamatan kurang dari 10 pengamatan per variabel penjelas akan menghasilkan penduga yang berbias bagi koefisien regresi, dalam arah positif maupun negatif (Peduzzi *et.al*, 1996). Nemez *et.al* (2009) juga menyatakan bahwa analisis regresi logistic cenderung akan overestimate dalam menduga koefisien regresi dan nilai odds rasio jika banyaknya pengamatan kurang dari 500. Dengan memperbesar ukuran contoh, maka kondisi overestimate dapat dikurangi dan dugaan odds ratio secara asimptotik akan mendekati odds ratio dari populasi. Dalam sebuah penelitian, kondisi overestimate yang disebabkan oleh kecilnya ukuran contoh tidak mempengaruhi interpretasi hasil.

## 2.2.Asumsi

Secara umum, terdapat terdapat dua pendekatan dalam metode pengklasifikasian objek tersebut, pendekatan parametrik dan pendekatan non parametrik. 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.

Secara umum, pendekatan non-parametrik digunakan bilamana pendekatan parametric tidak dapat digunakan karena tidak terpenuhinya asumsi atau terdapat kondisi "data sakit" yang mempengaruhi hasil analisis. Namun jika semua asumsi yang diperlukan terpenuhi, maka biasanya disarankan untuk tetap menggunakan pendekatan parametric ini

Analisis regresi logistik dan analisis diskriminan menggunakan pendekatan parametrik dalam pengklasifikasian objek. Dalam analisis regresi logistik biner diasumsikan bahwa peubah respons merupakan peubah acak binomial.

Dalam analisis diskriminan diasumsikan bahwa data merupakan suatu sampel acak dari sebaran normal ganda. Terpenuhi atau tidaknya asumsi kenormalan data ini dapat diperiksa dengan menggunakan histogram frekuensi

dari data ataupun melalui pengujian-pengujian kenormalan. Namun demikian, penyimpangan terhadap asumsi ini biasanya tidak terlalu mempengaruhi hasil analisis, dalam arti bahwa pengujian statistik yang dilakukan tetap dapat dipercaya.

Selain itu, jika digunakan analisis diskriminan linier, maka diasumsikan juga bahwa matriks ragam peragam dari variabel-variabel juga sama untuk semua kelompok. Penyimpangan yang tidak terlalu banyak dari asumsi ini biasanya tidak terlalu mempengaruhi hasil analisis. Namun sebelum mengambil kesimpulan akhir dari analisis ini, terutama untuk menganalisis hal-hal yang sangat penting, disarankan untuk tetap memeriksa kesamaan asumsi mengenai matriks ragam peragam ini. Berbagai pengujian dapat dilakukan untuk tujuan ini. Bila asumsi ini tidak terpenuhi dan hasil analisis meragukan, lakukan analisis diskriminan sekali lagi dengan mengeluarkan satu atau dua kelompok yang tidak terlalu penting. Analisis diskriminan kuadrat juga dapat dilakukan pada kondisi ini. Namun jika keseluruhan interpretasi yang diperoleh memuaskan, tentu saja hasil analisis diskriminan tersebut dapat diterima. Namun perlu diingat, beberapa pengujian mengenai matriks ragam peragam ini sangat sensitive terhadap asumsi kenormalan sebaran data. Asumsi mengenai kesamaan matriks ragam peragam ini tidak diperlukan jika analisis dilakukan dengan analisis diskriminan kuadrat.

Metode klasifikasi pohon merupakan pendekatan non-parametrik untuk mengklasifikasikan objek. Metode ini sangat banyak digunakan karena metode ini fleksibel dalam penggunaan. Tidak diperlukan asumsi apapun mengenai sebaran data dalam metode ini. Oleh karenanya, metode ini hanya dianggap sebagai metode eksplorasi saja.

### 2.3. Hasil analisis

Pada umumnya, dalam setiap metode pengklasifikasian, terdapat dua tujuan utama analisis. Pertama adalah untuk mendapatkan suatu aturan untuk pengelompokan objek yang dibangun berdasarkan pengamatan satu atau lebih variabel yang dilakukan terhadap dan kedua menggunakan aturan tersebut untuk mengklasifikasikan suatu objek baru yang belum diketahui berasal dari kelompok yang mana. Dalam mendapatkan aturan tersebut, biasanya ingin diketahui pula peubah-peubah yang berpengaruh terhadap pengelompokan tersebut.

Pada analisis diskriminan linier, aturan pengklasifikasian objek tersebut diperoleh dengan membangun satu atau lebih fungsi yang dinamakan fungsi diskriminan yang merupakan kombinasi linier dari peubah-peubah yang diamati. Dalam prosesnya, dapat ditentukan peubah mana yang secara signifikansi menjadi pembeda antar kelompok. Selanjutnya berdasarkan skor diskriminannya, suatu objek dapat diperkirakan keanggotaan dari suatu kelompok tertentu.

Berbeda dengan analisis diskriminan biasa, aturan pengelompokan pada metode MDP tidak diperoleh melalui fungsi-fungsi diskriminan. Aturan tersebut diperoleh dengan terlebih dahulu membentuk sekatan pada ruang contoh dimana pembentukan sekatan tersebut didasarkan pada fungsi jarak yang telah dipilih. Selanjutnya, dengan menggunakan kaidah keputusan Bayes, ditentukan ke dalam subpopulasi (kelompok) mana masing-masing sekatan tersebut diperkirakan berasal. Berbeda dengan analisis diskriminan biasa, pada analisis diskriminan

dengan metode ini, tidak dapat ditentukan variabel mana saja yang secara signifikans menjadi pembeda antar kelompok tersebut.

Hampir serupa dengan analisis diskriminan, pada analisis regresi logistic, aturan pengelompokan tersebut diperoleh dengan membangun sebuah model yang menggambarkan hubungan antara semua variabel penjelas dengan peubah bebas Y. Namun berbeda dengan analisis regresi biasa, pada analisis regresi logistic ini,

hubungan tersebut dinyatakan dalam hubungan non linier yaitu  $p = \frac{1}{1 + e^{-\eta_i}}$  dengan  $\eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$  atau bila dilakukan transformasi, akan

diperoleh model  $\log_e \frac{p}{1-p} = \text{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$ .

Sebagaimana analisis regresi biasa, dapat diuji variabel yang secara signifikans menentukan nilai logit tersebut. Keanggotaan dari suatu objek baru dapat diduga dengan menduga nilai logit(p) dari objek tersebut yang dihitung berdasarkan model yang diperoleh. Selanjutnya, dengan menentukan suatu cutting point bagi p, misalkan c (biasanya diambil c=0.5), maka dapat diduga ke dalam kelompok mana seharusnya objek tersebut dimasukkan.

#### DAFTAR PUSTAKA

- Collet, D. 1998. *Modelling Binary Data*. Chapman & Hall, Canada
- Gaspersz, V. 1995. *Teknik Analisis dalam Penelitian Percobaan Jilid 2*. Penerbit Tarsito, Bandung
- Hair, JF, RO Anderson, RL Tatham and WC Black. 1998. *Multivariate Data Analysis*. 5<sup>th</sup> ed. Prentice Hall, New Jersey
- Kleinbaum, DG. 1994. *Logistic Regression : A Self Learning Text*. Springer, New York
- McCullagh, P and JA Nelder. 1989. *Generalized Linear Models 2<sup>nd</sup> edition*. Chapman and Hall, London
- Nemes S, Jonasson JM, Genell A, Steineck G. 2009 Bias in odds ratios by logistic regression modelling and sample size. *BMC Medical Research Methodology* 9:56
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996). "A simulation study of the number of events per variable in logistic regression analysis". *J Clin Epidemiol* 49 (12): 1373-9.
- Myers, RH. 1990. *Classical and Modern Regression with Application 2<sup>nd</sup> edition*. PWS Kent Publishing Company, Boston
- Ryan, TP. 1997. *Modern Regression Methods*. John Wiley & Sons, Inc., New York