

PERBANDINGAN ANALISIS DISKRIMINAN DAN REGRESI LOGISTIK (Studi Kasus Klasifikasi Konsumen Berdasarkan Tempat Berbelanja di Wilayah Taman-Sidoarjo)

Maulidya.

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
dyamoly@yahoo.co.id

Drs. Hery Tri Sutanto, M.Si

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
herytrisutanto@gmail.co.id

Affiati Oktaviarina, M.Sc

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
affiatioktaviarina@yahoo.co.id

Abstrak

Berbelanja merupakan aktifitas sehari-hari yang dilakukan oleh masyarakat untuk memenuhi kebutuhan hidup. Salah satu tempat berbelanja adalah pasar. Jenis pasar menurut transaksinya dibedakan menjadi dua yaitu pasar modern dan pasar tradisional. Dengan semakin banyaknya tempat-tempat perbelanjaan, konsumen bisa menentukan sendiri dimana tempat mereka berbelanja. Mereka dapat memilih di pasar modern atau di pasar tradisional, tergantung pada kebutuhan dan daya beli. Beberapa tahun terakhir ini di Kabupaten Sidoarjo khususnya wilayah Taman, banyak dibangun supermarket maupun minimarket baru yang dapat mengancam keberadaan pasar tradisional. Hal ini mencerminkan adanya pergeseran perilaku konsumen yang dulunya hanya berbelanja di pasar tradisional, sekarang mulai beralih berbelanja di pasar modern. Pengelompokan konsumen menjadi penting untuk mengetahui komposisi konsumen berdasarkan tempat kebiasaan mereka berbelanja kebutuhan sehari-hari. Tujuan klasifikasi adalah mengelompokkan obyek penelitian berdasarkan serangkaian variabel-variabel yang disebut sebagai variabel bebas. Dalam penelitian ini digunakan dua metode yaitu analisis diskriminan dan regresi logistik. Hasil penelitian ini menunjukkan bahwa variabel yang secara signifikan mempengaruhi pengelompokan konsumen berdasarkan tempat berbelanja dengan menggunakan metode analisis diskriminan adalah variabel promosi. Sedangkan jika menggunakan regresi logistik variabel bebas yang berpengaruh adalah variabel pendidikan dan variabel promosi. Analisis diskriminan dan regresi logistik merupakan dua metode yang dapat digunakan sebagai metode pengklasifikasian objek, sehingga kedua metode tersebut dapat dibandingkan berdasarkan ketepatan pengelompokannya. Hasil perbandingan kedua metode ini menunjukkan bahwa ketepatan klasifikasi pada analisis regresi logistik lebih tinggi, yaitu 88,0 persen untuk data *learning* dan 80,0 persen untuk data *testing* jika dibandingkan dengan analisis diskriminan yang hanya sebesar 86,7 persen untuk data *learning* dan 76,0 persen untuk data *testing*.

Kata Kunci: Pasar, Analisis Diskriminan, Regresi Logistik

Abstract

Shopping is an activity that done by the society to fulfill their necessary of life. Market is one of the shopping place. Market divided in two kinds are: traditional and modern market. The consumers can determine their selves to bought in many shopping places. They also choose modern market or traditional market , depending on the needs and purchasing power . The last few years , especially region Taman in Sidoarjo area, many people built shopping center and minimarkets that threaten the existence of the traditional market. Now, many people move modern market as shoppinh place. Group of on consumers becames the important being to know the composition based on their habits in daily shopping . The purpose of classification agglomerate the objects of research based on variables are referred to as independent variables. The study used two methods discriminant analysis and logistic regressions. The result of study showed that the variables that significantly affect the grouping consumers based on place where the shop using discriminant analysis method is variable promotion. Mean while, if using logistic regression the impact for education variable and promotion variable. discriminant analysis and logistic regression are two methods that using object of classification method. The two methods can be comparing based on accuracy of agglomerate. The result of comparison these methods shows that the classification accuracy of the logistic regression analysis was higher, 88.0 percent for learning data, 80.0 percent for testing data, compared with discriminant analysis only amounted to 86.7 percent for the learning data and 76 , 0 percent for testing data.

Keywords: Market, Discriminant analysis, Logistic regression

1 PENDAHULUAN

Analisis multivariat merupakan metode statistik yang mempelajari pengolahan data dengan banyak variabel yang saling berkorelasi. Analisis multivariat dapat dibedakan menjadi dua bagian besar, yaitu analisis multivariat dependensi dan analisis interdependensi. Jika kita menghadapi analisis yang memuat satu variabel tak bebas, salah satu metode yang dapat digunakan pada analisis multivariat dependensi yaitu metode analisis diskriminan. Analisis diskriminan adalah analisis multivariat yang diterapkan untuk memodelkan hubungan antara satu variabel respon yang bersifat dikotomis atau multikotomis maupun kualitatif dengan variabel penjelas yang bersifat kuantitatif. Dalam penerapannya, fungsi diskriminan ini mengasumsikan data variabel bebas pada masing-masing kelompok berdistribusi normal multivariat dan mempunyai struktur matriks varian kovarian yang sama untuk setiap populasi dalam variabel respon. Selain analisis diskriminan ada juga metode regresi yang variabel responnya adalah data biner, dinamakan regresi logistik. Menurut (Agresti,1990) regresi logistik adalah suatu analisis regresi yang digunakan untuk menggambarkan hubungan antara variabel respon dengan sekumpulan variabel bebas, dimana variabel dependennya biner.

Berbelanja merupakan suatu aktivitas yang biasa dilakukan setiap orang, salah satu tempat berbelanja adalah pasar. Salah satu penerapan kedua metode tersebut untuk mengetahui faktor-faktor yang dipertimbangkan konsumen dalam memilih tempat untuk berbelanja barang yang dibutuhkan. Secara nasional perkembangan pasar tradisional memang telah mengalami penurunan bila dibanding dengan perkembangan pasar modern, sebagaimana hasil penelitian Ac Nielson (2005) terhadap perkembangan pasar modern dan tradisional di Indonesia, pada trend pertumbuhannya pasar modern tumbuh pesat yaitu 34 % dan pasar tradisional mengalami penurunan pertumbuhan - 8,5 5%. Pasar tradisional merupakan tulang punggung perekonomian yang tak bisa dibiarkan tergerus oleh pasar modern yang semakin menjamur, karena pasar ini melibatkan jutaan pedagang yang relatif berskala kecil. Pedagang Pasar Seluruh Indonesia mempunyai anggota 24.000 pasar, yang mencakup 12,60 juta pedagang tersebar di 26 provinsi. Dapat dibayangkan, jika separuh dari jumlah pedagang ini gulung tikar karena dagangannya selalu rugi dan tidak dapat bertahan di tengah derasnya persaingan usaha yang semakin ketat dengan pasar modern, hasilnya adalah jumlah pengangguran Indonesia yang akan meningkat. Berdasarkan uraian diatas maka penulis tertarik untuk mengetahui pola hubungan antara pemilihan tempat belanja terhadap variabel-variabel yang mempengaruhinya dan ketepatan klasifikasi antara metode regersi logistik dan analisis diskriminan.

2 KAJIAN PUSTAKA

2.1 Regresi

Regresi yang berarti peramalan, penaksiran, atau pendugaan pertama kali diperkenalkan oleh Sir Francis Galton pada tahun 1877 sehubungan dengan penelitiannya terhadap tinggi manusia. Persamaan regresi adalah persamaan matematik yang memungkinkan untuk meramalkan nilai-nilai suatu variabel tak bebas dari nilai-nilai satu atau lebih variabel bebas.

2.1.1 Regresi Linier Dan Regresi NonLinier

Regresi dikatakan linier apabila hubungan antara variabel independen dan variabel dependennya adalah linier. Regresi linier dibagi menjadi dua yaitu regresi linier sederhana dan regresi linier berganda.

Regresi linier berganda adalah regresi linier yang mempunyai variabel bebas (x) lebih dari satu. Jika pada regresi linier sederhana hanya terdapat satu variabel bebas, pada regresi linier berganda terdapat dua variabel bebas atau lebih. Model regresi linier berganda adalah

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

Nonlinier yang dimaksud pada regresi non linier bisa berarti nonlinier dalam variabel ataupun nonlinier dalam parameter. Bentuk dari hubungan regresi nonlinier adalah

$$Y_i = f(X_i, \gamma) + \varepsilon$$

2.2 Regresi Logistik

Model regresi logistik biner merupakan model matematika yang dapat digunakan untuk memodelkan hubungan antara variabel bebas X dengan variabel respon Y yang bersifat biner. Variabel respon (y) mengikuti distribusi Bernoulli dengan fungsi probabilitas :

$$f(y_i) = p^{y_i} (1 - p)^{1 - y_i}, y_i = 0, 1. \text{ Distribusi dari}$$

variabel respon ini merupakan pembeda antara regresi logistik dengan regresi linear. Adapun fungsi logistik adalah sebagai berikut :

$$f(x) = \frac{1}{1 + e^{-x}}, -\infty < x < \infty$$

$$\lim_{x \rightarrow \infty} \frac{1}{1 + e^{-x}} = \frac{1}{1 + 0} = 1 \text{ dan } \lim_{x \rightarrow -\infty} \frac{1}{1 + e^{-(-x)}} = \frac{1}{1 + e^{\infty}} = \frac{1}{1 + \infty} = \frac{1}{\infty} = 0$$

Sehingga untuk x mendekati $-\infty$ maka

$$\lim_{x \rightarrow -\infty} f(x) = 0, \text{ sedangkan untuk } x \text{ mendekati } \infty, \text{ maka}$$

$$\lim_{x \rightarrow \infty} f(x) = 1. \text{ Bentuk model regresi logistik adalah}$$

(Agresti,1990) :

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

Model regresi logistik dengan k variabel bebas adalah:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}$$

Jika model ditransformasikan dengan transformasi logit, maka akan menghasilkan bentuk logit : $g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$

2.3 Pendugaan Parameter

Penaksiran parameter pada model regresi logistik dapat menggunakan metode Maksimum

Likelihood Estimator (MLE). Dengan menetapkan asumsi distribusi binomial dan setiap objek pengamatan saling independen, fungsi likelihood metode MLE merupakan fungsi linear maka untuk memperoleh taksiran parameter dilakukan proses iterasi dengan metode Newton-Raphson, dengan menentukan nilai awal dari β yaitu β^0 .

2.3.1 Pengujian Parameter

Uji yang dapat digunakan untuk menguji signifikansi koefisien β dari model dapat menggunakan uji secara parsial maupun serentak

1. Uji Parsial

Statistik uji Wald digunakan untuk menguji parameter secara parsial (Hosmer and Lemeshow, 2000).

Rumus umum untuk uji-Wald

berdasarkan hipotesis :

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

$$\text{Statistik Uji : Wald } (w) = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (2.13)$$

daerah penolakan H_0 adalah jika $W > \chi^2_{(db,\alpha)}$ atau p-value $< \alpha$

2. Uji Serentak

Untuk mengetahui signifikansi model secara keseluruhan.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_i = 0$$

$$H_1: \text{paling sedikit ada satu } \beta_i \neq 0 \text{ (} i = 1, 2, \dots, k \text{)}$$

Statistik Uji :

$$G = -2 \ln \frac{\binom{n_1}{n}^{n_1} \binom{n_0}{n}^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}}$$

H_0 akan ditolak jika $G > X_{(v,\alpha)}$ atau p-value $< \alpha$, dengan $v = p-1$ dimana p adalah banyaknya variabel bebas.

2.4 Uji Kesesuaian Model

Uji kesesuaian model digunakan untuk menilai apakah model sesuai atau tidak. Hipotesis yang digunakan sebagai berikut.

$$H_0: \text{model sesuai}$$

$$H_1: \text{model tidak sesuai}$$

$$\hat{C}(\text{Hosmer} - \text{Lemeshow}) = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Apabila $\hat{C} < X^2_{(db,\alpha)}$ maka gagal tolak H_0 .

2.5 Odds Ratio

Odds ratio didefinisikan sebagai perbandingan dari nilai variabel sukses terhadap variabel bernilai gagal.

Untuk variabel bebas dikotomus, ada dua nilai $\pi(x)$ dan

$$\begin{aligned} \text{dua nilai } 1 - \pi \quad \psi &= \frac{\left(\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \right) / \left(\frac{1}{1 + e^{\beta_0 + \beta_1}} \right)}{\left(\frac{e^{\beta_0}}{1 + e^{\beta_0}} \right) / \left(\frac{1}{1 + e^{\beta_0}} \right)} \\ &= \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} \\ &= e^{(\beta_0 + \beta_1) - \beta_0} \\ &= e^{\beta_1} \end{aligned}$$

Jika nilai $\psi = 1$, maka antara kedua variabel tidak terdapat hubungan. Jika nilai $\psi < 1$, maka antara kedua variabel terdapat hubungan negatif dan demikian sebaliknya bila $\psi > 1$ maka kedua variabel terhadap hubungan positif.

2.6 Analisis Multivariat

Data multivariat merupakan data hasil pengukuran pada beberapa variabel yang secara umum dinotasikan sebagai x_{ij} . Notasi tersebut menunjukkan harga tertentu dari pengamatan ke $-i$ pada variabel ke $-j$ dan jika terdapat p variabel dan ada sebanyak n pengamatan, yang dapat ditunjukkan dalam bentuk matriks seperti dibawah ini, dimana x_{np} merupakan objek pengamatan ke- n untuk variabel ke- p

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

4. Matriks Varians dan Kovarians

Untuk lebih memudahkan dalam menganalisis data, varians – kovarians sampel dapat di susun dalam sebuah matriks dengan ukuran $p \times p$.

$$S = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{bmatrix} \quad (2.14)$$

Diagonal utama dari matriks menyatakan varians, dan selain diagonal utama menyatakan kovarians.

2.8 Analisis Diskriminan

Analisis diskriminan adalah teknik statistika yang dipergunakan untuk mengklasifikasikan suatu individu atau observasi ke dalam suatu kelas atau kelompok berdasarkan sekumpulan variabel-variabel (Johnson & Wichern 2007). Model umum analisis diskriminan merupakan suatu kombinasi linear yang bentuknya sebagai berikut:

$$Z_{jk} = a + W_i X_{ik} + \dots + W_n X_{nk}$$

2.9 Uji Asumsi Untuk Analisis Diskriminan

Membentuk fungsi diskriminan yang optimal diperlukan beberapa asumsi terhadap data yang digunakan. Asumsi ini antara lain adalah data pada variabel bebas seharusnya berdistribusi normal multivariat dan adanya kesamaan matriks varians - kovarians antar kelompok.

2.9.1 Pengujian Normal Multivariat

Menurut Johnson dan Wichern (2007), untuk memeriksa data apakah merupakan normal

multivariat, dapat dilihat dari plot antara d_j^2 dengan $chi-square$ $\left(\frac{j-0,5}{n}\right)$ apakah hasil plot

menggambarkan garis lurus maka data tersebut dapat dinyatakan sebagai normal multivariat. Statistik uji yang digunakan :

$$d_j^2 = [X_j - \bar{X}]' S^{-1} [X_j - \bar{X}] \quad , j = 1, 2, \dots, n$$

Hipotesis yang digunakan adalah

H_0 : data berdistribusi multivariat normal.

H_1 : data tidak berdistribusi multivariat normal.

Pemeriksaan normal multivariat dilakukan dengan algoritma sebagai berikut :

1. Menghitung d_j^2 .
2. Mendaftar d_j^2 sedemikian hingga $d_1^2 \leq d_2^2 \leq \dots \leq d_n^2$
3. Membuat plot $(d_j^2, \chi_{p; \frac{j-0,5}{n}}^2)$

Kriteria gagal tolak H_0 berarti data berdistribusi multivariat normal, kriteria pemenuhan asumsi dilakukan secara visual yaitu jika plot membentuk garis lurus berarti data dapat didekati sebaran normal (Johnson & Wichern 2007).

2.9.2 Pengujian Kesamaan Matrik Kovarian

Menurut Morisson (1967) untuk pengujian kehomogenan matriks kovarian dapat dilakukan dengan uji Box's M. Uji Box's M dapat dirumuskan sebagai berikut

$$S = \frac{1}{\sum_{i=1}^g n_i} \sum_{i=1}^g n_i S_i$$

$$M = \sum_{i=1}^g n_i \ln |S| - \sum_{i=1}^g n_i \ln |S_i|$$

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \left(\sum_{i=1}^g \frac{1}{n_i} - \frac{1}{\sum_{i=1}^g n_i} \right)$$

dengan hipotesis sebagai berikut

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k = \Sigma$$

$$H_1 : \text{minimal ada satu } \Sigma_i \neq \Sigma$$

$$i = 1, 2, \dots, k$$

Daerah penolakan untuk hipotesis nol dapat dihipotesis dengan menggunakan sebaran $chi-square$ sebagai berikut:

$$MC^{-1} > \chi_{\alpha, db}^2$$

2.10 Metode Untuk Membentuk Fungsi Diskriminan

Dalam penyusunan fungsi diskriminan ada dua metode yang dapat digunakan, yaitu estimasi simultan dan estimasi *stepwise*. Cara mengetahui variabel bebas yang dapat mendiskriminasi suatu kelompok adalah menggunakan statistik uji Wilks Lamda, yaitu

$$\Delta^* = \frac{\left| \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \right|}{\left| \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(x_{ij} - \bar{x})' \right|}$$

2.11 Fungsi Diskriminan

Fungsi diskriminan adalah kombinasi linear dari variabel-variabel yang dimiliki oleh kelompok - kelompok yang akan diklasifikasikan. Untuk dua grup dapat dilakukan menggunakan fungsi diskriminan linier Fisher, yaitu

$$Y = (\bar{x}_1 - \bar{x}_2)' S_{gab}^{-1} x$$

Nilai fungsi diskriminan Z merupakan dasar untuk menentukan suatu objek masuk kelompok yang mana dengan membandingkannya dengan rata-rata (centroid) dari nilai Z masing-masing kelompok. Jika ada dua kelompok masing-masing A dan B, maka

$$Z_{cu} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B}$$

2.13 Kriteria Perbandingan

Kriteria perbandingan teknik klasifikasi didasarkan pada kesalahan klasifikasinya yang dikenal dengan *Apparent Error Rate* (APER) merupakan nilai dari besar kecilnya jumlah observasi yang salah dalam pengklasifikasian berdasarkan suatu fungsi klasifikasi (Johnson & Wichern 2007). Adapun APER dihitung dengan terlebih dahulu membuat tabel klasifikasi sebagai berikut:

Tabel 2.2 Klasifikasi *actual* dan *predicted* group

Actual group	Predicted Group	
	π_0	π_1
π_0	n_{00}	$n_{01} = n_0 - n_{00}$
π_1	$n_{10} = n_1 - n_{11}$	n_{11}

Sumber : Johnson and Wichern (2007)

$$APER = \frac{n_{01} + n_{10}}{n_0 + n_1}$$

Sedangkan ketepatan prediksi pengelompokan secara tepat dapat menggunakan rumus *Hit Ratio*

$$HitRatio = \frac{n_{00} + n_{11}}{n_0 + n_1} \times 100\%$$

2.14 Pembagian Data Sampel

Salah satu cara untuk menghindari bias adalah membagi sampel menjadi dua bagian yaitu data *learning* yang digunakan untuk membentuk aturan klasifikasi serta pembentukan model baik model regresi logistik ataupun model analisis diskriminan dan data *testing* yang digunakan untuk mengevaluasi fungsi klasifikasi, menguji ketepatan, keakuratan, dan kehandalan dari fungsi yang telah dihasilkan.

2.15 Pasar

Pasar adalah tempat fisik dimana para pembeli dan penjual berkumpul untuk membeli dan menjual barang. Jenis pasar menurut transaksinya dibedakan menjadi pasar modern dan pasar tradisional.

Pasar tradisional dikenal sebagai pasar yang bangunannya relatif sederhana, dengan suasana yang relatif kurang menyenangkan (ruang tempat usaha sempit, sarana parkir yang kurang memadai, kurang

menjaga kebersihan pasar, dan penerangan kurang baik). Pasar modern adalah pasar yang bersifat modern dimana barang-barang diperjual belikan dengan harga pas dan layanan tersendiri.

2.16 Perilaku Konsumen

Menurut Kotler dan Keller (2009 : 166) perilaku konsumen adalah studi tentang bagaimana individu, kelompok, dan organisasi memilih, membeli, menggunakan dan bagaimana barang, jasa, ide, atau pengalaman untuk memuaskan kebutuhan dan keinginan mereka.

Menurut Kotler dan Keller (2009 : 214) terdapat beberapa faktor yang mempengaruhi perilaku konsumen, diantaranya adalah :

1. Faktor Budaya
2. Faktor Sosial
3. Faktor Pribadi

2.17 Bauran Pemasaran Eceran (*Retail Marketing Mix*)

Adapun pengertian dari Bauran Pemasaran (*Marketing Mix*) menurut Kotler dan Keller (2009 : 19) adalah seperangkat alat yang digunakan oleh perusahaan untuk mencapai tujuan perusahaan di pasar sasarnya. Seperangkat alat seperti yang telah disebutkan diatas terdiri dari 4 atau yang lebih dikenal dengan istilah 4 P (empat P). Keempat P tersebut menurut Mc Carthy dalam Kotler dan Keller (2009 ; 23) diantaranya adalah :

1. Produk
2. Harga
3. Lokasi
4. Promosi

3. METODE PENELITIAN

3.1 Lokasi dan Objek Penelitian

Penelitian berlokasi di Kecamatan Taman-Sidoarjo, yaitu untuk pasar modern di gerai supermarket Alfa Midi dan Master sedangkan untuk Pasar Tradisional di Kecamatan Taman dan Pasar Puspa Agro.

3.2 Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah pengunjung pasar sehingga jumlahnya tidak diketahui dengan pasti, sehingga jenis sampling yang dipilih yaitu *purposive sampling*. Teknik pengambilan sampel menggunakan *Purposive Sampling*, yaitu teknik penentuan sampel dengan mempertimbangkan karakter dan ciri – ciri yang telah ditentukan terlebih dahulu untuk membatasi sampel. Populasi dalam penelitian ini adalah pengunjung pasar sehingga jumlahnya tidak diketahui dengan pasti, maka besarnya sampel yang diperlukan sangat dipengaruhi oleh maksimum error dan derajat kepercayaan dalam populasi tersebut. Sehingga besarnya sampel dapat diketahui dengan rumus *binomunal proportions* sebagai berikut

$$n = \frac{Z^2 p(1-p)}{(\mu)^2} \quad n = \frac{(1,96)^2}{4(0,1)^2} = 96,04 \approx 100.$$

Jadi besarnya sampel yang digunakan dalam penelitian ini adalah 100 responden.

3.3 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data primer dan data sekunder.

3.4 Instrumen Pengumpul Data

Instrumen untuk mengumpulkan data penelitian ini adalah kuesioner. Sebelum digunakan untuk mendapatkan data, maka terlebih dahulu dilakukan berbagai uji untuk mengetahui apakah kuesioner itu cukup layak untuk mengambil data variabel yang diteliti, dalam hal ini menggunakan uji validitas dan uji reliabilitas untuk pengujian instrumen. Kuesioner dibagikan kepada 30 responden terlebih dahulu.

3.4.1 Uji Validitas dan Reliabilitas

Statistik uji

$$r_{xy} = \frac{n \left(\sum_{i=1}^n X_i Y_i \right) - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right)}{\sqrt{\left[n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2 \right] \left[n \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i \right)^2 \right]}}$$

Kriteria dalam menentukan validitas suatu kuesioner adalah suatu model (butir pertanyaan) dikatakan valid jika tingkat signifikansinya di bawah 0,05 atau

- a). Jika $r_{hitung} > r_{tabel}$ maka pertanyaan dinyatakan valid.
- b). Jika $r_{hitung} < r_{tabel}$ maka pertanyaan dinyatakan tidak valid.

Uji reliabilitas dalam penelitian ini menggunakan metode *Alfa Cronbach*.

Hipotesis:

H_0 : Pertanyaan menghasilkan data yang tidak reliabel.

H_1 : Pertanyaan menghasilkan data yang reliabel.

Tolak H_0 atau hasil pengukuran reliabel jika nilai *cronbach's alpha* lebih besar dari 0,60. Rumus ini ditulis sebagai berikut

$$r_{11} = \left(\frac{k}{k-1} \right) \left[1 - \frac{\sum_{b=1}^k \sigma_b^2}{\sigma_t^2} \right]$$

3.5 Variabel Penelitian

Variabel Respon (Y) = tempat belanja

Penelitian ini menggunakan satu variabel respon sebagai berikut :

0 = Pasar Tradisional (Pasar Taman, Puspa Agro)

1 =Pasar Modern, yaitu terdiri dari supermarket(Alfa Midi, Master).

Variabel Bebas (X) : Usia (X_1), Jenis kelamin (X_2) Pendidikan (X_3) Pekerjaan (X_4)Pendapatan (X_5) Produk (X_6)Harga (X_7)Promosi (X_8) Lokasi (X_9)

3.6 Prosedur Penelitian

Pemisahan penelitian menjadi dua bagian, yaitu *data learning* dan *data testing* . Komposisi pembagian data *learning* dan data *testing* pada penelitian ini adalah 75% dan 25%

3.6.1 Langkah - Langkah Untuk Mencapai Tujuan Pertama Yaitu :

1. Mendeskripsikan karakteristik konsumen.

2. Menguji Asumsi,
3. Membentuk fungsi diskriminan Menentukan nilai batas pemisah kedua kelompok.
4. Melihat nilai dari *of original grouped cases correctly classified dan cross validated grouped cases correctly classified.*
5. Uji signifikansi dilakukan untuk menganalisa model fungsi diskriminan yang terbentuk apakah cukup signifikan atau tidak.
6. Menghitung tingkat ketepatan pengelompokan hasil prediksi fungsi diskriminan.

3.6.2 Langkah - Langkah Untuk Mencapai Tujuan Kedua Yaitu

- 1 Menentukan model regresi logistik antara variabel respon dengan masing- masing variabel independen secara individu.
- 2 Menentukan model regresi logistik antara variabel respon dengan variabel-variabel independen secara berganda.
- 3 Melakukan pembentukan model dengan memasukkan variabel yang signifikan pada regresi logistik berganda.
- 4 Menguji kesesuaian model untuk mengetahui apakah model regresi logistik yang telah didapat sudah sesuai atau tidak.
- 5 Menginterpretasikan model regresi logistik yang telah diperoleh.

3.6.3 Langkah - Langkah Untuk Mencapai Tujuan Ketiga Yaitu

Langkah yang digunakan adalah membandingkan presentase obyek yang mampu dikelompokkan secara tepat dari masing – masing metode. Metode yang menghasilkan nilai ketepatan klasifikasi yang lebih besar dianggap lebih handal dalam mengelompokkan obyek. Perbandingan juga dilakukan pada data untuk validasi.

4 PEMBAHASAN DAN PENERAPAN

4.1 Uji Validitas Variabel

Untuk mendapatkan data primer, peneliti menyebarkan kuesioner kepada konsumen sebanyak 100 responden. Sebelum melakukan penelitian lebih lanjut peneliti terlebih dahulu melakukan survei pendahuluan kepada 30 konsumen dengan beberapa pertanyaan untuk menguji tingkat validitas. Nilai tabel Product Moment dapat diperoleh dimana $df = n-2$, dalam hal ini adalah jumlah sampel yang digunakan $(n)=30$, maka besarnya $df=30-2=28$. Dengan α 0,05 nilai r adalah 0,374. Berdasarkan hasil penelitian bahwa semua pertanyaan pada variabel produk, harga ,lokasi, dan promosi dinyatakan valid.

4.2 Uji Reliabilitas Variabel

Tabel 4.5 Hasil Uji Reliabilitas

No.	Item	Cronbach's Alpha	Keterangan
1	X _{6_1}	0,769	Reliabel
2	X _{6_2}	0,750	Reliabel
3	X _{6_3}	0,777	Reliabel

4	X _{7_1}	0,772	Reliabel
5	X _{7_2}	0,775	Reliabel
6	X _{7_3}	0,756	Reliabel
7	X _{8_1}	0,766	Reliabel
8	X _{8_2}	0,727	Reliabel
9	X _{8_3}	0,770	Reliabel
10	X _{9_1}	0,763	Reliabel
11	X _{9_2}	0,775	Reliabel
12	X _{9_3}	0,759	Reliabel

Sumber : Diolah oleh peneliti

Berdasarkan uji reliabilitas sesuai Tabel 4.5 dapat dikatakan bahwa seluruh item reliabel. Ini bisa dilihat dari keseluruhan item pertanyaan memiliki *cronbach alpha* diatas 0,60. Sehingga dapat ditarik kesimpulan bahwa pertanyaan penelitian ini reliabel.

4.4.1 Uji Multinormal

Uji multinormal dilakukan untuk mengetahui apakah data yang digunakan sudah mengikuti distribusi multinormal. Kriteria pemenuhan asumsi dilakukan secara visual yaitu jika plot q-q membentuk garis lurus berarti data dapat didekati sebaran normal (Johnson & Wichern 2007). Persebaran datanya dapat dilihat pada Lampiran IV.

4.4.2 Uji Kesamaan Matrik Varian Kovarian

Untuk menguji kesamaan matrik varian kovarian digunakan uji Box's M. Hasil pengujian Box's M sebesar 62,381 dan nilai p-value sebesar 0,164 yang lebih besar dari signifikansi 0,05 maka terima H_0 yang berarti grup *covariance matrices* adalah sama. Hal ini berarti data diatas sudah memenuhi asumsi analisis diskriminan, dengan demikian fungsi diskriminan yang terbentuk sudah sesuai untuk mengelompokkan konsumen.

4.4.3 Pembentukan Fungsi Diskriminan

Fungsi diskriminan linier Fisher membuat semacam persamaan regresi dan membaginya berdasarkan kode grup. Sehingga fungsi diskriminan linier Fisher yang terbentuk berdasarkan output pada tabel *Classification Function Coefficients* dapat mencirikan setiap kelompok adalah sebagai berikut: Fungsi diskriminan untuk yang berbelanja di pasar tradisional adalah

$$Z(0) = -14,861 + 3,638X_9$$

Sedangkan Fungsi diskriminan untuk yang berbelanja di pasar modern adalah

$$Z(1) = -28,947 + 5,137X_9$$

Fungsi diskriminan linier berdasarkan *Z score unstandardized* yang terbentuk adalah

$$Z = -6,406 + 0,683X_9$$

Hasil skor tersebut selanjutnya akan dibandingkan dengan *cut off score*, untuk mengetahui apakah konsumen masuk ke kelompok layak atau tidak layak. Karena banyaknya sampel antar kelompok 1 dan 0 tidak sama, maka menggunakan persamaan:

$$Z_{cu} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B}$$

Pada Tabel *Function at Group Centroid* didapatkan angka centroid pada kelompok 0 adalah -1,082 dan untuk

kelompok 1 adalah 1,112. Dengan demikian perhitungan Z_{cu} didapatkan nilai sebesar

$$Z_{cu} = \frac{38.1,112 + 37. -1,082}{38 + 37} = 0,0297$$

4.5.1 Model Regresi Logistik Individu (Dengan Masing-Masing Variabel Bebas)

Pembentukan model regresi logistik secara individu bertujuan untuk mengetahui variabel bebas mana yang berpengaruh secara individu terhadap variabel respon, sebelum dilakukan pemodelan antara variabel respon dengan variabel-variabel bebas secara bersama-sama. Untuk itu, perlu dilakukan pengujian signifikansi parameter sebagai koefisien dari variabel bebas pada masing-masing model individu.

Kesimpulan: Faktor-faktor yang berpengaruh secara individu terhadap pemilihan tempat berbelanja konsumen adalah produk (X_6), harga (X_7), dan promosi (X_9). Hal ini terlihat dari nilai W pada masing-masing variabel tersebut lebih besar dari nilai $\chi^2_{(df,\alpha)} = 3,841$.

Sehingga keputusannya tolak H_0 . Langkah selanjutnya adalah menginterpretasikan model tersebut.

4.5.2 Model Regresi Logistik Berganda

Meskipun dari pemodelan regresi logistik individu telah diketahui bahwa variabel produk, harga, dan, promosi berpengaruh secara signifikan terhadap variabel respon, namun belum diketahui bagaimana hubungan antara variabel lain. Karena jika ternyata hubungannya sangat erat, maka dimungkinkan salah satu variabel akan menjadi tidak signifikan lagi pengaruhnya terhadap variabel respon. Oleh karena itu, perlu dilakukan pemodelan regresi logistik dengan memasukkan semua variabel secara bersama-sama untuk memeriksa ada atau tidaknya hubungan antara variabel tersebut. Untuk mengetahui apakah parameter-parameter model telah signifikan atau tidak maka dilakukan langkah-langkah pengujian hipotesis sebagai berikut

1. Uji Serentak

Dari hasil analisis diperoleh nilai G_{hitung} adalah 73,467 sedangkan $\chi^2_{(14;0,05)} = 23,6848$. Karena nilai $G_{hitung} > \chi^2_{14;0,05}$ maka tolak H_0 , yang berarti paling tidak ada satu variabel yang berpengaruh atau mempunyai kontribusi terhadap model.

2. Uji parsial

Untuk mengetahui pengaruh yang diberikan setiap variabel bebas, perlu dilakukan pengujian parameter secara parsial menggunakan uji Wald. Hasil uji signifikansi parameter regresi logistik secara parsial dengan sembilan variabel bebas yang dapat diketahui bahwa secara individu variabel yang signifikan berpengaruh terhadap pemilihan tempat berbelanja konsumen adalah variabel pendidikan (X_3)(2) dan variabel promosi (X_9), hal ini terlihat dari nilai W pada masing – masing variabel bebas X_3 dan X_9 lebih besar dari nilai $\chi^2_{(df,\alpha)} = 3,841$ atau nilai dari $p_value < \alpha = 0,05$.

4.6 Pembentukan Model

$$\pi(x) = \frac{\exp(-15,009 + 1,923X_3(2) + 1,474X_9)}{1 + \exp(-15,009 + 1,923X_3(2) + 1,474X_9)}$$

Model regresi logistik ditulis dalam bentuk logit menjadi $g(x) = (-15,009 + 1,923X_3(2) + 1,474X_9)$

4.7 Uji Kesesuaian Model

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} = 3,329$$

Daerah penolakan $H_0: \hat{C} > \chi^2_{(0,05;6)} = 12,5916$ atau nilai $p_value > \alpha$. Dengan demikian H_0 diterima. Kesimpulannya bahwa model sesuai atau tidak terdapat perbedaan yang nyata antara observasi dengan prediksi model.

4.9 Perbandingan Hasil kalsifikasi Analisis Diskriminan dan Regresi Logistik

Hasil ketepatan klasifikasi data *learning* untuk analisis diskriminan adalah 86,7%. Sedangkan untuk peluang kesalahan klasifikasi digunakan ukuran *Apparent Error Rates* (APER).

$$APER = \frac{n_{01} + n_{10}}{n_0 + n_1} = \frac{5 + 5}{38 + 37} = 0,13$$

Presentase kebenaran keseluruhan yang dihasilkan pada data *testing* sebesar 76,0%, sedangkan data *learning* sebesar 86,7%. Oleh karena itu model dapat dikatakan sudah valid. Nilai APER pada data *testing* adalah sebesar

$$APER = \frac{n_{01} + n_{10}}{n_0 + n_1} = \frac{2 + 4}{9 + 16} = 0,24$$

2. Hasil Klasifikasi Pada Regresi Logistik

Hasil ketepatan klasifikasi data *learning* pada regresi logistik adalah 88%, sehingga

$$APER = \frac{n_{01} + n_{10}}{n_0 + n_1} = \frac{4 + 5}{38 + 37} = 0,12$$

Presentase ketepatan klasifikasi secara keseluruhan pada data *testing* juga cukup besar yaitu 80%. Nilai APER pada data *testing* adalah sebesar

$$APER = \frac{n_{01} + n_{10}}{n_0 + n_1} = \frac{3 + 2}{9 + 16} = 0,20$$

Hasil perbandingan presentase ketepatan klasifikasi antara analisis regresi logistik dan analisis diskriminan menunjukkan bahwa analisis regresi logistik lebih baik untuk digunakan. Hal ini dikarenakan ketepatan klasifikasi regresi logistik sebesar 80,0 persen dan analisis diskriminan sebesar 76,0 persen. Regresi logistik juga memberikan selisih yang paling minimum antara data *learning* dan data *testing*. Jadi, metode analisis regresi logistik dapat memberikan hasil klasifikasi yang lebih baik daripada analisis diskriminan.

5. KESIMPULAN DAN SARAN

1. Hasil analisis regresi logistik menunjukkan bahwa variabel bebas yang secara signifikan berpengaruh terhadap pemilihan tempat berbelanja adalah pendidikan (X_3) dan promosi (X_9). Model yang didapatkan untuk menggambarkan pola hubungan antara variabel respon dan bebas adalah sebagai berikut:

$$\pi(x) = \frac{\exp(-15,009 + 1,923X_3(2) + 1,474X_9)}{1 + \exp(-15,009 + 1,923X_3(2) + 1,474X_9)}$$

2. Hasil perbandingan presentase ketepatan klasifikasi antara analisis diskriminan dan regresi logistik menunjukkan bahwa analisis regresi logistik lebih baik untuk digunakan, hal ini terlihat dari presentase ketepatan pada regresi logistik untuk data *learning* sebesar 88,0% dan untuk data *testing* sebesar 80,0%, sedangkan untuk analisis diskriminan ketepatan presentase untuk data *learning* adalah sebesar 86,7% dan untuk data *testing* sebesar 76,0%.

5.2 Saran

Bagi peneliti selanjutnya penting untuk mengembangkan penelitian ini dengan menggunakan variabel yang berbeda dan lebih banyak variabel bebas yang dipergunakan, sehingga penemuan penelitian selanjutnya dapat menghasilkan penemuan-penemuan baru dengan harapan bisa melengkapi hasil penelitian ini.

DAFTAR PUSTAKA

- [1] Agresti, Alan. 1990. *Categorical Data Analysis*, John Wiley and Sons, Inc, New York.
- [2] Hair, JF. 2010. *Multivariate Data Analysis*. New Jersey : Prentice Hall, Inc.
- [3] Hosmer, D.W.dan Lemeshow, S. 2000. "*Applied Logistic Regression*", 2nd Edition, John Wiley and Sons, Inc., New York.
- [4] Indra, P. 2010. *Faktor-faktor Yang Mempengaruhi Resiko Penyebab Kanker Payudara Dengan Pendekatan Regresi Logistik* [Online]. Termuat di: <http://digilib.its.ac.id/public/ITS-Undergraduate-13440-Paper.pdf>. [Diakses : 25 Agustus 2013]
- [5] Johnson RA, Wichern DW. 2007. *Applied Multivariate Statistical Analysis*. New Jersey : Prentice Hall, Inc.
- [6] Kitbunrungrat, Krieng 2012. *Comparison Logistic Regression and Discriminant Analysis in classification groups for Breast Cancer* [Online]. Termuat di: http://paper.ijcsns.org/07_book/201205/20120518.pdf . [Diakses : 9 maret 2013]
- [7] Kurniati, Elvy,dkk. *Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Respon Biner Pada Pengklasifikasian Kejadian Bayi dengan Berat Badan Lahir Rendah (BBLR)* [Online]. Termuat di: <http://dc347.4shared.com/doc/cUfpzEa/preview.html> . [Diakses 9 maret 2013]
- [8] Kotler Philip, Keller Kevin . 2009. *Manajeme pemasaran edisi 12*. Jakarta : PT indeks.
- [9] Morrison DF. 1967. *Multivariate Statistical Methods*. New York Mc.Graw-Hill, Inc.2013]
- [10] Permana, Aringga. 2010. *Analisis Diskriminan Dua Kelompok Untuk Mengetahui Seseorang Kecanduan Facebook*. Skripsi S1 Jurusan Matematika FMIPA UNESA.
- [11] Revitalisasi pasar tradisional sebagai upaya menunjang aktivitas ekonomi masyarakat Sidoarjo [Online].termuat di: <http://www.asparindo.com/berita-> . [Diakses 3 april 2013]
- [12] Rohmawati, Shofi 2010. *PERBANDINGAN ANALISIS REGRESI LOGISTIK DAN DISKRIMINAN PADA KASUS KECANDUAN HANDPHONE (Studi Kasus SMA Kompleks Surabaya)* [Online]. Termuat di: <http://isjd.pdii.lipi.go.id/admin/jurnal/11208539543.pdf> . [Diakses: 9 Maret 2013]
- [13] Setiati, Dwi 2012. *Fungsi Pasar* [Online]. Termuat di: <http://.wordpress.com/2012/06/05/fungsi-pasar/> . [Diakses: 14 mei 2013]
- [14] Sudaryanto. 2012. *Analisis Diskriminan Marketing Mix Terhadap Keputusan Pembelian Harian Pagi Radar Jember* [Online]. Termuat di jurnal.unej.ac.id/index.php/BISMA/article/download/48/25 [Diakses: 13 Agustus 2013]
- [15] Supranto J. 2004. *Analisis Multivariat : Arti dan Interpretasi*. Jakarta : PT Asdi Mahasatya.