

# MENENTUKAN PROBABILITAS QUALITAS LULUSAN PROGRAM STUDI MENGUNAKAN *LOGISTIC REGRESSION*

**Maxsi Ary, Slamet Risnanto**

Program Studi Manajemen Informatika

AMIK BSI Bandung

Jalan Sekolah Internasional No.1-6 Antapani, Bandung 40282

[maxsi.max@bsi.ac.id](mailto:maxsi.max@bsi.ac.id), [slamet.sat@bsi.ac.id](mailto:slamet.sat@bsi.ac.id)

**Abstract** – Human resources (HR) is one of the success factors in the economic field, namely how to create a human resources (HR) qualified and have the skills and highly competitive in the global competition. Educational level of the labor force that is still relatively low. The structure of education of the workforce is still dominated Indonesian basic education which is about 63.2%. The issue raised is to determine the probability of a program of study (whether or not) to see some of the ratio of the number of graduates by the number of students per class, the amount of quota size class (large or small) using logistic regression models. Data were obtained from a search result based on the amount of data the study program students and graduates in 2010 Data processing using SPSS. The results of the analysis by assessing model fit and the results will be given for each model fit. Starting with the hypothesis for assessing model fit, statistical -2LogL, Cox and Snell's R Square, Hosmer and Lemeshow's Goodness of Fit Test, and the classification table. The results of the analysis using SPSS as a tool aimed at measuring quality of graduate courses at a university, college, or academy, whether or not based on the ratio of the number of graduates and class quotas.

**Keywords:** *Quota Class, Probability, Logistic Regression*

**Abstrak** – Sumberdaya manusia (SDM) adalah salah satu faktor kesuksesan dalam bidang ekonomi, yaitu bagaimana menciptakan sumber daya manusia (SDM) yang berkualitas dan memiliki keterampilan serta berdaya saing tinggi dalam persaingan global. Tingkat pendidikan angkatan kerja yang ada masih relatif rendah. Struktur pendidikan angkatan kerja Indonesia masih didominasi pendidikan dasar yaitu sekitar 63,2%. Persoalan yang dikemukakan adalah menentukan probabilitas sebuah program studi (baik atau tidak) dengan melihat beberapa rasio jumlah lulusan dengan jumlah mahasiswa per angkatan, ukuran besarnya kuota kelas (besar atau kecil) menggunakan model logistic regression. Data diperoleh dari hasil penelusuran data program studi berdasarkan jumlah mahasiswa dan lulusan pada tahun 2010. Pengolahan data menggunakan SPSS. Hasil analisis dengan menilai model fit dan akan diberikan hasilnya untuk setiap model fit. Dimulai dengan hipotesis untuk menilai model fit, statistic -2LogL, Cox dan Snell's R Square, Hosmer and Lemeshow's Goodness of Fit Test, dan tabel klasifikasi. Hasil analisis menggunakan SPSS tersebut bertujuan sebagai alat pengukuran kualitas lulusan program studi di suatu Universitas, Perguruan Tinggi, maupun Akademi, baik atau tidaknya berdasarkan rasio jumlah lulusan dan kuota kelas.

**Kata Kunci:** *Kuota Kelas, Probabilitas, Logistic Regression*

## PENDAHULUAN

Terdapat beberapa teknik dalam statistik untuk menganalisis data. Salah satu caranya yaitu dengan menguji hipotesis nol ( $H_0$ ). Tujuan dari menganalisis data ini adalah mendapatkan informasi yang berada dalam data tersebut. Informasi yang diperoleh ini digunakan untuk menyelesaikan suatu

masalah. Permasalahan yang akan diselesaikan biasanya dinyatakan dalam hipotesis nol. Cara analisis data seperti ini disebut metode statistik hipotesis nol.

Untuk menganalisis data, diperlukan pula identifikasi pengukuran atau aturan pengukuran yang disebut skala pengukuran. Pengukuran menurut (Ghozali, 2011) merupakan suatu proses hal mana suatu

angka atau simbol diletakkan atau properti suatu stimuli sesuai dengan aturan atau prosedur yang telah ditetapkan. Misalkan orang dapat digambarkan dari beberapa karakteristik seperti usia, pendidikan, agama, jenis kelamin, pendapatan. Skala pengukuran yang sesuai dapat digunakan untuk menunjukkan karakteristik ini. Menurut (Steven, 1946) dalam buku (Ghozali, 2011) skala pengukuran dapat dikelompokkan menjadi empat yaitu, skala nominal, skala ordinal, skala interval, dan skala rasio.

Sumber daya manusia (SDM) adalah salah satu faktor kesuksesan dalam bidang ekonomi, yaitu bagaimana menciptakan SDM yang berkualitas dan memiliki keterampilan serta berdaya saing tinggi dalam persaingan global. Sehingga terdapat setidaknya dua hal penting menurut (Damanhuri) menyangkut kondisi SDM Indonesia, yaitu: pertama adanya ketimpangan antara jumlah kesempatan kerja dan angkatan kerja. Jumlah angkatan kerja nasional pada krisis ekonomi tahun pertama (1998) sekitar 92,73 juta orang, sementara jumlah kesempatan kerja yang ada hanya sekitar 87,67 juta orang dan ada sekitar 5,06 juta orang penganggur terbuka (*open unemployment*). Angka ini meningkat terus selama krisis ekonomi yang kini berjumlah sekitar 8 juta. Kedua, tingkat pendidikan angkatan kerja yang ada masih relatif rendah. Struktur pendidikan angkatan kerja Indonesia masih didominasi pendidikan dasar yaitu sekitar 63,2%. Kedua masalah tersebut menunjukkan bahwa ada kelangkaan kesempatan kerja dan rendahnya kualitas angkatan kerja secara nasional di berbagai sektor ekonomi.

Mengacu pada pandangan kedua yang dikemukakan oleh (Damanhuri) tentang tingkat pendidikan yang masih didominasi pendidikan dasar, diperlukan peran perguruan tinggi untuk menanggulangi persoalan tersebut. Lebih mendasar pada peran serta program studi dalam mendukung lulusan perguruan tinggi.

*Logistic Regression* mempunyai tujuan untuk menguji apakah probabilitas terjadinya variabel terikat (*dependent variable*) dapat diprediksi dengan variabel bebasnya (*Independent Variable*). Contoh kasus sederhana dari penggunaan analisis *logistic regression* adalah:

1. Seorang auditor ingin menentukan probabilitas sebuah perusahaan bangkrut dengan melihat beberapa rasio keuangan, ukuran besarnya perusahaan (besar atau kecil).

2. Seorang dokter ingin mengetahui apakah probabilitas seorang pasien terserang penyakit jantung dapat diprediksi dari tekanan darah, kadar kolesterol, kalori yang dimakan, jenis kelamin dan gaya hidup.

Persoalan seperti contoh kasus sederhana tersebut sebenarnya dapat diselesaikan dengan analisis diskriminan. Namun demikian, asumsi multivariate normal distribution tidak dapat dipenuhi karena variabel bebas merupakan campuran antara variabel kontinu dan kategorikal. Dalam hal ini dapat di analisis dengan logistic regression karena tidak perlu asumsi normalitas data pada variabel bebasnya. Sehingga *logistic regression* dipakai pada umumnya jika asumsi *multivariate normal distribution* tidak dipenuhi.

Dalam penulisan ini, persoalan yang dikemukakan yaitu jika kita ingin menentukan probabilitas sebuah program studi baik/tidak dengan melihat beberapa rasio jumlah lulusan dengan jumlah mahasiswa per angkatan, ukuran besarnya kuota kelas (besar atau kecil). Persoalan yang dikemukakan memiliki asumsi *multivariate normal distribution* tidak dapat dipenuhi karena variabel bebas merupakan campuran antara variabel kontinu dan kategorikal.

Data diperoleh dari hasil penelusuran data program studi pada kampus Universitas BSI Tahun Akademik 2009/2010 berdasarkan jumlah mahasiswa dan lulusan dengan periode tahun yang dipilih tahun 2010. Pengolahan data menggunakan SPSS (*Statistical Package for Social Science*). Hasil analisis dengan menilai model fit dan akan diberikan hasilnya untuk setiap model fit. Dimulai dengan hipotesis untuk menilai *model fit, statistic -2LogL, Cox dan Snell's R Square, Hosmer and Lemeshow's Goodness of Fit Test*, dan tabel klasifikasi.

Hasil analisis menggunakan SPSS tersebut dapat dijadikan pengukuran kualitas lulusan program studi di suatu Universitas, Perguruan Tinggi (PT), maupun Akademi, baik atau tidaknya berdasarkan rasio jumlah lulusan dan kuota kelas.

## LANDASAN TEORI

### Skala Pengukuran

Menurut (Steven, 1946) dalam buku (Ghozali, 2011) skala pengukuran dapat dikelompokkan menjadi empat yaitu, skala nominal, skala ordinal, skala interval, dan skala rasio.

1. Skala Nominal

Skala nominal merupakan skala pengukuran yang menyatakan kategori, misalkan variabel jenis kelamin, di mana responden dapat dikelompokkan ke dalam dua kategori laki-laki dan perempuan. Kedua kelompok ini diberi kode angka 1 dan 2. Kedua angka ini berfungsi untuk label kategori saja tanpa memiliki arti apa-apa. Oleh sebab itu tidaklah tepat menghitung nilai rata-rata dan standar deviasi dari variabel jenis kelamin. Jadi uji statistic yang sesuai dengan skala nominal adalah uji statistic yang berdasarkan perhitungan seperti **modus** dan **distribusi frekuensi**.

2. Skala Ordinal

Skala ordinal tidak hanya mengkategorikan variabel kedalam kelompok, tetapi juga melakukan peringkat terhadap kategori. Misalkan responden menyatakan peringkat terhadap merk kendaraan roda dua, angka 1 untuk merk yang paling disukai, angka 2 untuk peringkat kedua, dan seterusnya. Jadi kategori antar merk tidak menggambarkan perbedaan yang sama dari ukuran atribut. Uji statistic yang sesuai dengan skala ordinal adalah **modus**, **median**, **distribusi frekuensi** dan **statistic non-parametrik** seperti **rank order correlation**. Variabel yang diukur dengan skala nominal dan ordinal umumnya disebut variabel non-parametrik atau variabel non-metrik.

3. Skala Interval

Misalkan responden menentukan peringkat terhadap merk, kemudian diminta memberikan nilai (*rate*) terhadap preferensi merk sesuai dengan skala penilaian sebagai berikut:

Nilai Skala	Preferensi
1	Preferensi Sangat Tinggi
1	Preferensi Tinggi
1	Preferensi Moderat
1	Preferensi Rendah
1	Preferensi Sangat Rendah

Jika kita berasumsi bahwa urutan kategori menggambarkan tingkat preferensi yang sama, maka kita dapat mengatakan bahwa perbedaan preferensi responden untuk dua merk motor yang mendapat peringkat 1 dan 2 adalah sama dengan perbedaan preferensi untuk dua merk lainnya yang memiliki rating 4 dan 5. Namun demikian, kita tidak dapat menyatakan bahwa preferensi responden terhadap merk yang mendapat rating 5 nilainya lima kali preferensi untuk merk yang mendapat rating 1. Uji statistik

yang sesuai untuk jenis pengukuran skala ini adalah **semua uji statistik**, kecuali yang mendasarkan pada rasio seperti **koefisien variasi**.

4. Skala Rasio

Skala rasio adalah skala interval dan memiliki nilai dasar (*based value*) yang tidak dapat dirubah. Misalkan usia memiliki nilai dasar 0 tahun. Skala rasio dapat ditransformasikan dengan cara mengalikan dengan konstanta, tetapi transformasi tidak dapat dilakukan jika dengan cara menambah konstanta. Variabel yang diukur dengan skala interval dan rasio disebut variabel **metrik**.

**METODE PENELITIAN**

**1. Metode Analisis Data**

Misalkan kita anggap memiliki data yang berisi *n* observasi dengan *p* variabel. *p* variabel ini dapat dibagi menjadi dua kelompok atau subset. Uji statistik untuk menganalisis set data seperti ini disebut **metode dependen** (*dependence method*). Metode dependen menguji ada tidaknya hubungan dua set variabel. jika peneliti atas dasar teori yang ada menyatakan bahwa satu variabel dari subset adalah variabel bebas (*independence variable*) dan variabel lainnya dari subset adalah variabel terikat (*dependence variable*) maka tujuan dari metode dependen adalah menentukan apakah variabel bebas mempengaruhi variabel terikat secara individual dan atau bersamaan. Sebaliknya jika set data yang ada tidak mungkin untuk mengelompokkan kedalam variabel bebas dan variabel terikat, maka tujuan dari jenis set data seperti ini adalah mengidentifikasi bagaimana dan mengapa variabel tersebut saling berkaitan satu sama lainnya. Metode statistik yang sesuai untuk menganalisis data set seperti ini disebut **metode interdependen** (*interdependence method*).

**a. Metode Dependen**

Metode dependen dapat dikelompokkan lagi menjadi:

- 1) Jumlah variabel bebas – satu atau lebih dari satu variabel bebas
- 2) Jumlah variabel terikat – satu atau lebih dari satu variabel terikat
- 3) Jenis skala pengukuran variabel bebas – metrik dan non-metrik
- 4) Jenis skala pengukuran variabel terikat – metrik dan non-metrik

**a) Satu Variabel terikat (metrik) dan satu variabel bebas (non-metrik)**

Metode statistika untuk satu variabel terikat dan satu variabel bebas sering disebut **metode univariat** (*univariate method*). Apabila variabel bebas merupakan ukuran non-metrik dengan kategori lebih dari dua dan variabel terikat dengan pengukuran metric, misalkan apakah ada perbedaan gaji dilihat dari tempat tinggal yang dibagi menjadi empat kategori. Alat uji untuk kasus ini adalah **Analysis of Variance** (ANOVA).

**b) Dua atau lebih Variabel terikat (metrik) dan dua atau lebih variabel bebas (non-metrik)**

Metode statistika untuk menguji lebih dari satu variabel terikat dan lebih dari satu variabel bebas adalah **Multivariate Analysis of Variance** (MANOVA). Contoh kasus yang dapat diberikan diantaranya: a) seperti apakah rata-rata gaji kepala keluarga dan total income anggota keluarga berbeda secara nyata untuk tiap region, b) seperti apakah rata-rata gaji kepala keluarga dan *total income* anggota keluarga berbeda secara nyata untuk tiap region dan ras, c) seperti apakah rata-rata gaji kepala keluarga berbeda nyata untuk tiap region dan ras.

**c) Satu Variabel Terikat (metrik) dan Satu atau Lebih Variabel Bebas (metrik)**

Metode statistic untuk menguji hubungan antara satu variabel terikat dan satu atau lebih variabel bebas adalah **regresi**. **Regresi sederhana** (*simple regression*) untuk menguji pengaruh satu variabel bebas terhadap satu variabel terikat, sedangkan untuk lebih dari satu variabel bebas disebut **regresi berganda** (*multiple regression*).

**d) Satu variabel terikat (metrik) dan lebih dari satu variabel bebas (metrik)**

Misalkan variabel terikat adalah intensitas membeli produk yang diukur dengan skala nominal. Responden diminta untuk menjawab apakah akan membeli produk (1) atau tidak akan membeli produk (0). Variabel bebasnya adalah usia, pendapatan, dan pendidikan, semuanya diukur dengan ukuran interval atau rasio, maka kita sekarang mempunyai set data dimana variabel terikatnya adalah kategori / nominal dan variabel bebasnya adalah metric / kontinyu. Persoalan uji statistic pada model seperti ini adalah menentukan apakah kedua grup yaitu pembeli dan bukan pembeli produk

secara signifikan berbeda dikaitkan dengan variabel bebas. Metode statistic untuk menjawab persoalan ini dilakukan dengan **analisis diskriminan** (*discriminant analysis*). Asumsi yang mendasari adalah data harus berdistribusi normal dan jumlah sample untuk kategori variabel terikat yaitu pembeli dan bukan pembeli harus sebanding 50% pembeli dan 50% bukan pembeli.

Apabila variabel bebasnya kombinasi antara metric dan nominal (non-metrik), maka asumsi normalitas multivariate tidak akan dapat dipenuhi. Penyimpangan dari asumsi normalitas multivariate mempengaruhi signifikanis uji statistic dan tingkat ketepatan klasifikasi. Jika hal ini terjadi, maka digunakan uji statistik **logistic regression**. Logistik regression tidak mensyaratkan jumlah sample untuk kategori variabel terikat.

**b. Analisis Regresi**

Istilah regresi pertama kali diperkenalkan oleh Sir Francis Galton tahun 1886. Galton berasumsi bahwa orang tua dengan badan yang tinggi, memiliki anak yang tinggi pula dan orang tua berbadan pendek akan memiliki anak yang pendek pula. Walaupun berasumsi seperti itu, dalam pengamatannya ada kecenderungan bahwa tinggi anak bergerak menuju rata-rata tinggi populasi secara keseluruhan. Inilah yang disebut dengan hukum Galton mengenai regresi universal.

Interpretasi modern mengenai regresi agak berlainan dengan versi Galton. Secara umum, analisis regresi pada dasarnya adalah studi mengenai ketergantungan variabel dependen dengan satu atau lebih variabel independen, dengan tujuan untuk mengestimasi dan/atau memprediksi rata-rata populasi atau nilai rata-rata variabel dependen berdasarkan nilai variabel independen yang diketahui (Gujarati, 2003).

Hasil analisi regresi adalah koefisien untuk masing-masing variabel independen. Koefisien ini diperoleh dengan cara memprediksi nilai variabel dependen dengan suatu persamaan. Koefisien regresi dihitung dengan dua tujuan; pertama untuk meminimumkan penyimpangan antara nilai actual dan nilai estimasi variabel dependen berdasarkan data yang ada (Tabachnick, 1996).

**c. Analisis Diskriminan**

Analisis diskriminan merupakan bentuk regresi dengan variabel terikat berbentuk non-metrik/kategori. Sebagai contoh misalkan seorang analis keuangan

ingin mengetahui variabel-variabel apa saja yang membedakan antara perusahaan sehat dan perusahaan yang mengalami kebangkrutan. Analisis keuangan tersebut juga ingin mengetahui apakah mungkin menggunakan faktor-faktor yang telah teridentifikasi tadi sebagai bentuk indek yang mampu membedakan kedua perusahaan sehat dan bangkrut. Indeks tersebut kemudian digunakan untuk meramalkan kemungkinan perusahaan akan bangkrut dikemudian hari.

Tujuan dari contoh kasus tersebut diatas sebetulnya ingin menjawab tiga hal sebagai berikut:

- 1) Mengidentifikasi variabel-variabel yang mampu membedakan antara kedua kelompok.
- 2) Menggunakan variabel-variabel yang telah teridentifikasi untuk menyusun persamaan atau fungsi untuk menghitung variabel baru atau indek yang dapat menjelaskan perbedaan antara dua kelompok.
- 3) Menggunakan variabel yang telah teridentifikasi atau indek untuk mengembangkan aturan atau cara mengelompokkan observasi di masa datang kedalam satu dari kedua kelompok.

## 2. Logistic Regression

*Logistic Regression* digunakan untuk menguji apakah probabilitas terjadinya variabel terikat dapat diprediksi dengan variabel bebasnya (Ghozali, 2011). Contoh kasus untuk logistic regression yaitu misalkan seorang auditor ingin menentukan probabilitas sebuah perusahaan bangkrut dengan melihat beberapa rasio keuangan, ukuran besarnya perusahaan (besar atau kecil). Dalam hal kasus diatas variabel bebas merupakan campuran antara variabel kontinyu dan kategori, sehingga tidak perlu asumsi normalitas data pada variabel bebasnya. Jadi *logistic regression* digunakan jika asumsi multivariate normal distribution tidak dipenuhi.

Konsep logistic regression berkaitan erat dengan probabilitas. Odds dan probabilitas memberikan informasi yang sama, tetapi dalam bentuk yang berbeda. Dari dua bentuk yang berbeda ini, dapat dirubah odds menjadi probabilitas atau sebaliknya, yaitu dengan cara sebagai berikut:

$$P(S | B) = \frac{odds(S | B)}{1 + odds(S | B)}$$

..... 1

$$odds(S | B) = \frac{P(S | B)}{1 - P(S | B)}$$

..... 2

Perhitungan odds di atas dapat dihitung nilai log naturalnya menjadi sebagai berikut:

$$Ln[odds(S | B)] = Ln(odds(S | B))$$

..... 3

$$Ln[odds(S | K)] = Ln(odds(S | K))$$

..... 4

Kedua persamaan ini (persamaan 3 dan 4) dapat digabungkan kedalam persamaan di bawah ini untuk memberikan log odds sebagai fungsi ukuran kuota kelas program studi (SIZE):

$$Ln[odds(S | SIZE)] =$$

$$Ln[odds(S | K)] + Ln[odds(S | B)]SIZE$$

..... 5

Dimana SIZE = 1 jika kuota kelas program studi besar dan SIZE = 0, jika kuota kelas program studi kecil. Jadi jelas bahwa log dari odds adalah fungsi linear dari variabel bebas SIZE dan dapat diinterpretasikan seperti koefisien pada analisis regresi. Tanda koefisien SIZE positif berarti log dari odds akan meningkat jika SIZE meningkat, dimana log dari odds kuota kelas program studi besar yang sukses lebih tinggi daripada kuota kelas program studi kecil. Persamaan *logistic regression* untuk k variabel bebas dapat dinyatakan sebagai berikut:

$$Ln[odds(S | X_1, X_2, \dots, X_3)] =$$

$$b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

..... 6

atau

$$Ln \frac{p}{1-p} = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

..... 7

di mana:

$$Odds(S | X_1, X_2, \dots, X_3) = \frac{p}{1-p}$$

..... 8

*p* adalah probabilitas program studi sukses dengan variabel bebas  $X_1, X_2, \dots, X_3$ . Model log dari odds merupakan fungsi linear dari variabel bebas dan ekuivalen dengan

persamaan multiple regression dengan log dari odds sebagai variabel terikat. Variabel bebasnya dapat berupa kombinasi variabel kontinyu maupun variabel kategori. Oleh karena log dari odds sering disebut logit maka persamaan regresinya disebut *multiple logistic regression* atau *logistic regression*.

Untuk sederhana misalkan hanya ada satu variabel bebas SIZE, maka persamaan logistic regression dapat dinyatakan sebagai:

$$\ln \frac{p}{1-p} = b_0 + b_1 \text{ SIZE}$$

$$p = \frac{1}{1 + e^{-(b_0 + b_1 \text{ SIZE})}}$$

Hubungan antara probabilitas  $p$  dan variabel bebas SIZE adalah non linear, sedangkan hubungan antara log dari odds dan variabel bebas SIZE adalah linear.

Dengan demikian interpretasi terhadap koefisien variabel bebas SIZE harus dilihat pengaruhnya terhadap log dari odds dan bukan terhadap probabilitas  $p$ . Prosedur estimasi maksimum likelihood dapat digunakan untuk menaksir parameter dan hal ini dilakukan dengan prosedur iterasi untuk mendapatkan nilai parameter.

**PEMBAHASAN**

Pada proses pengukuran ini dilakukan menggunakan model *logistic regression* dan analisis menggunakan SPSS Inc 17.0.

**Pengolahan Data Kuota Kelas**

Berikut adalah Perguruan Tinggi Universitas BSI, tanggal berdiri 19-01-2010 dan No SK PT 04/D/O/2010 tanggal SK PT 19-01-2010.

**Tabel 1**  
**.Daftar Program Studi Universitas BSI**

No	Kode	Nama Program Studi	Jenjang
1	61101	Manajemen	S-2
2	90221	Desain Interior	S-1
3	90241	Desain Komunikasi Visual	S-1
4	14201	Ilmu Keperawatan	S-1
5	70201	Ilmu Komunikasi	S-1
6	26201	Teknik Industri	S-1
7	62401	Akuntansi	D-3
8	57401	Manajemen Informatika	D-3
9	14901	Pendidikan Profesi Ners	Profesi

Sedangkan data jumlah mahasiswa dan jumlah tahun 2010 disajikan pada tabel 2. lulusan Perguruan Tinggi Universitas BSI

**Tabel 2**  
**Jumlah Mahasiswa dan Lulusan Tahun 2010**

No	Kode	Nama Program Studi	Jumlah Mahasiswa	Jumlah Lulusan
1	61101	Manajemen (S2)	484	249
2	14201	Ilmu Keperawatan	59	57
3	26201	Teknik Industri	0	0
4	70201	Ilmu Komunikasi	160	160
5	90221	Desain Interior	0	0
6	90241	Desain Komunikasi Visual	55	55
7	57401	Manajemen Informatika	0	0
8	62401	Akuntansi	0	0
9	14901	Pendidikan Profesi Ners	-	-
10	61201	Manajemen (STP) Ars Int	49	29
11	57401	Manajemen Informatika	511	390
12	57402	Komputerisasi Akuntansi	126	102
13	61405	Manajemen Perusahaan	136	105
14	63412	Sekretari	35	26
15	93401	Usaha Perjalanan Wisata	0	0
16	93402	Perhotelan	41	39
17	93403	Manajemen Perhotelan	0	0

**Model Logistic Regression**

Akan digunakan data 14 program studi dengan pembagian sebagai berikut: 5 program studi sukses dan 9 program studi tidak sukses. Tabel 3 berikut ini merupakan data ukuran program studi (SIZE) dalam bentuk

kategori yaitu program studi kuota kelas (kuota 60 per kelas) besar diberi kode 1 dan program studi kuota kelas kecil diberi kode 0. Program studi sukses (kelulusan  $\geq 80\%$ ) diberi kode 1 dan program studi tidak sukses (kelulusan  $< 80\%$ ).

**Tabel 3**  
**Pengolahan Data Jumlah Mahasiswa dan Kelulusan**

No	Kd Prodi	Nama Prodi	mhs	lulus	Lulus (%)	SIZE	Jml Kelas	KELAS
1	61101	Manajemen (S2)	484	249	51.45	0	8.07	1
2	14201	Ilmu Keperawatan	59	57	96.61	1	0.98	0
3	26201	Teknik Industri	0	0	0.00	0	0.00	0
4	70201	Ilmu Komunikasi	160	160	100.00	1	2.67	1
5	90221	Desain Interior	0	0	0.00	0	0.00	0
6	90241	Desain Komunikasi Visual	55	55	100.00	1	0.92	0
7	61201	Manajemen (STP) Ars Int	49	29	59.18	0	0.82	0
8	57401	Manajemen Informatika	511	390	76.32	0	8.52	1
9	57402	Komputerisasi Akuntansi	126	102	80.95	1	2.10	1
10	61405	Manajemen Perusahaan	136	105	77.21	0	2.27	1
11	63412	Sekretari	35	26	74.29	0	0.58	0
12	93401	Usaha Perjalanan Wisata	0	0	0.00	0	0.00	0
13	93402	Perhotelan	41	39	95.12	1	0.68	0
14	93403	Manajemen Perhotelan	0	0	0.00	0	0.00	0

Tabel 4 merupakan data program studi sukses lebih dari atau kurang dari 80% dan tidak sukses berdasarkan asumsi kelulusan

**Tabel 4**  
**Program Studi Sukses dan Tidak Sukses**

No	Prodi	SUKSES	TIDAK SUKSES
		Kode 1	Kode 0
		SIZE	KELAS
1	61101	0	1
2	14201	1	0
3	26201	0	0
4	70201	1	1
5	90221	0	0
6	90241	1	0
7	61201	0	0
8	57401	0	1
9	57402	1	1
10	61405	0	1
11	63412	0	0
12	93401	0	0
13	93402	1	0
14	93403	0	0

Dari tabel 4 dapat disusun tabel kontijensi antara tipe program studi (sukses dan tidak sukses) dan ukuran program studi (besar dan

kecil). Tabel 5 berikut ini adalah tabel kontijensi dari tipe program studi dan ukuran program studi.

**Tabel 5**  
**Kontijensi Tabel Tipe Program Studi dan Ukuran Program Studi**

Tipe Program Studi	Size		Total
	Besar	Kecil	
Sukses (S)	2	3	5
Tidak Sukses (TS)	3	6	9
Total	5	9	14

Tabel 5 memberikan beberapa kemungkinan perhitungan probabilitas sebagai berikut:

1. Probabilitas bahwa kuota kelas program studi akan sukses adalah  $P(S) = 5/14 = 0.36$
2. Probabilitas bahwa kuota kelas program studi akan sukses dan kuota kelas program studi Besar (B) adalah:  $P(S|B) = 2/5 = 0.400$
3. Probabilitas bahwa kuota kelas program studi akan sukses dan kuota kelas program studi Kecil (K) adalah:  $P(S|K) = 3/9 = 0.333$

Probabilitas kadang-kadang dinyatakan dalam istilah odds. Dari tabel 4.5 di atas dapat dihitung odds sebagai berikut:

1. Odds sebuah kuota kelas program studi akan sukses adalah  $odds(S) = 5/5 = 1$  yang berarti odds sebuah kuota kelas akan sukses atau tidak sukses adalah sama atau odds 1 lawan 1
2. Odds sebuah kuota kelas program studi akan sukses dan kuota kelas besar adalah  $odds(S|B) = 2/3 = 0.667$  yang berarti odds kuota kelas besar yang akan sukses adalah 2 banding 3 atau 0.667 banding 1
3. Odds sebuah kuota kelas program studi akan sukses dan kuota kelas kecil adalah  $odds(S|K) = 3/6 = 0.5$  yang berarti odds kuota kelas kecil yang akan sukses adalah

3 banding 6 atau 0.5 banding 1

Odds dan probabilitas memberikan informasi yang sama, tetapi dalam bentuk yang berbeda. Dari dua bentuk yang berbeda ini, dapat dirubah odds menjadi probabilitas atau sebaliknya, yaitu dengan cara sebagai berikut:

$$P(S | B) = \frac{odds(S | B)}{1 + odds(S | B)} = \frac{0.667}{1 + 0.667} = 0.40$$

$$odds(S | B) = \frac{P(S | B)}{1 - P(S | B)} = \frac{0.40}{1 - 0.40} = 0.667$$

Perhitungan odds di atas dapat dihitung nilai log naturalnya menjadi sebagai berikut:

$$\ln[odds(S | B)] = \ln(0.667) = -0.405$$

$$\ln[odds(S | K)] = \ln(0.5) = -0.693$$

Kedua persamaan ini dapat digabungkan kedalam persamaan di bawah ini untuk memberikan log odds sebagai fungsi ukuran kuota kelas program studi (SIZE):

$$\ln[odds(S | SIZE)] = -0.693 + 0.288 \text{ SIZE}$$

### 1. Hasil Pengukuran Pengolahan Data

Hasil Output SPSS: *Logistic Regression*

**Tabel 6**  
**Case Processing Summary**

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	16	100.0
	Missing Cases	0	.0
	Total	16	100.0
Unselected Cases		0	.0
	Total	16	100.0

a. *If weight is in effect, see classification table for the total number of cases.*

**Tabel 7 Dependent Variable Encoding**

Original Value	Internal Value
0	0
1	1

**Tabel 8 Categorical Variables Codings**

SIZE	Parameter coding		
	Frequency	(1)	
0	11	1.000	
1	5	.000	

**Tabel 9 Iteration History<sup>a,b,c</sup>**

Iteration		Coefficients	
		-2 Log likelihood	Constant
Step 0	1	19.880	-.750
	2	19.875	-.788
	3	19.875	-.788

- a. Constant is included in the model.
- b. Initial -2 Log Likelihood: 19.875
- c. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

**Tabel 10 Iteration History<sup>a,b,c,d</sup>**

Iteration		Coefficients			
		-2 Log likelihood	Constant	SIZE(1)	KELAS
Step 1	1	18.639	.952	-1.758	-.286
	2	18.590	1.086	-1.946	-.341
	3	18.590	1.100	-1.960	-.348
	4	18.590	1.100	-1.960	-.348

- a. Method: Enter
- b. Constant is included in the model.
- c. Initial -2 Log Likelihood: 19.875
- d. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

**Tabel 11 Omnibus Tests of Model Coefficients**

Step		Chi-square	df	Sig.
Step 1	Step	1.285	2	.526
	Block	1.285	2	.526
	Model	1.285	2	.526

**Tabel 12 Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	18.590 <sup>a</sup>	.077	.108

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

**Tabel 13 Hosmer and Lemeshow Test**

Step	Chi-square	Df	Sig.
1	11.212	5	.047

**Tabel 14 Classification Table<sup>a</sup>**

	Observed	Predicted		Percentage Correct
		program_studi	0	
Step 1	program 0	10	1	90.9
	_studi 1	3	2	40.0
Overall Percentage				75.0

a. The cut value is .500

**Tabel 15**  
*Variables in the Equation*

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	SIZE(1)	-1.960	1.806	1.178	1	.278	.141
	KELAS	-.348	.381	.834	1	.361	.706
	Constant	1.100	1.777	.384	1	.536	3.005

a. Variable(s) entered on step 1: SIZE, KELAS.

Analisis:

**Menilai Model Fit**

Langkah pertama untuk menilai model fit adalah dengan menilai overall fit model terhadap data. Hipotesis untuk menilai model fit adalah sebagai berikut:

H0 : Model yang dihipotesiskan fit dengan data  
H1 : Model yang dihipotesiskan tidak fit dengan data

Statistik yang digunakan berdasarkan pada fungsi *likelihood*. *Likelihood* L dari model adalah probabilitas bahwa model yang dihipotesiskan menggambarkan data input. Untuk menguji hipotesis nol dan alternatifnya, L ditransformasikan menjadi  $-2\text{LogL}$ . Statistik  $-2\text{LogL}$  disebut juga *likelihood ratio* statistik  $\chi^2$ , dimana  $\chi^2$  distribusi dengan *degree of freedom*  $n - q$ ,  $q$  adalah jumlah parameter dalam model. Output SPSS memberikan dua nilai  $-2\text{LogL}$  yaitu satu untuk model yang hanya memasukan konstanta yaitu sebesar 19.875 dan memiliki distribusi  $\chi^2$  dengan df 15 (16-1), walaupun tidak tampak dalam output SPSS nilai  $-2\text{LogL}$  19.875 ini signifikan pada alpha 5% dan hipotesis nol ditolak yang berarti model hanya dengan konstanta saja tidak fit dengan data.  $-2\text{LogL}$  yang kedua adalah untuk model dengan konstanta dan variabel bebas SIZE dan KELAS dengan nilai  $-2\text{LogL}$  sebesar 18.590 atau memiliki distribusi  $\chi^2$  dengan df 13 (16-3).  $-2\text{LogL}$  untuk model dengan konstanta dan variabel bebas SIZE dan KELAS ternyata tidak signifikan pada alpha 5% yang berarti hipotesis nol tidak dapat ditolak dan model fit dengan data.

*Statistic -2LogL* dapat juga digunakan untuk menentukan jika variabel bebas ditambahkan kedalam model apakah secara signifikan memperbaiki model fit. Selisih  $-2\text{LogL}$  untuk model dengan konstanta saja dan  $-2\text{LogL}$  untuk model dengan konstanta dan variabel bebas didistribusikan sebagai  $\chi^2$  dengan df (selisih df kedua model). Output SPSS menunjukkan selisih kedua  $-2\text{LogL}$  sebesar 1.285 (19.875 – 18.590) dengan df 2 (15 –

13) dan angka ini signifikan secara statistik. Hal ini berarti hipotesis nol ditolak dan penambahan variabel bebas SIZE dan KELAS kedalam model memperbaiki model fit.

*Cox* dan *Snell's R Square* merupakan ukuran yang mencoba meniru ukuran  $R^2$  pada *multiple regression* yang didasarkan pada teknik estimasi *likelihood* dengan nilai maksimum kurang dari 1 (satu) sehingga sulit diinterpretasikan. *Nagelkerke's R square* merupakan modifikasi dari koefisien *Cox* dan *Snell's R Square* untuk memastikan bahwa nilainya bervariasi dari 0 sampai 1. Hal ini dilakukan dengan cara membagi nilai *Cox* dan *Snell's R Square* dengan nilai maksimumnya. Nilai *Nagelkerke's R square* dapat diinterpretasikan seperti nilai  $R^2$  pada *multiple regression*. Dilihat dari output SPSS nilai *Cox* dan *Snell's R Square* sebesar 0.077 dan nilai *Nagelkerke's R square* adalah 0.108 yang berarti variabilitas variabel dependen yang dapat dijelaskan oleh variabilitas variabel independen sebesar 10.8%.

*Hosmer and Lemeshow's Goodness of Fit Test* menguji hipotesis nol bahwa data empiris cocok atau sesuai dengan model (tidak ada perbedaan antara model dengan data sehingga model dapat dikatakan fit). Jika nilai *Hosmer and Lemeshow's Goodness of Fit Test* sama dengan atau kurang dari 0.05, maka hipotesis nol ditolak yang berarti ada perbedaan signifikan antara model dengan nilai observasinya sehingga *Goodness Fit model* tidak baik karena model tidak dapat memprediksi nilai observasinya. Jika nilai *Hosmer and Lemeshow's Goodness of Fit Test* lebih besar dari 0.05, maka hipotesis nol tidak dapat ditolak dan berarti model mampu memprediksi nilai observasinya atau dapat dikatakan model dapat diterima karena cocok dengan data observasinya. Tampilan output SPSS menunjukkan bahwa besarnya nilai *statistic Hosmer and Lemeshow's Goodness of Fit Test* 11.212 dengan probabilitas signifikan 0.047 yang nilainya kurang dari 0.05. dengan

demikian dapat disimpulkan bahwa model tidak dapat diterima.

Tabel klasifikasi 2 x 2 menghitung nilai estimasi yang benar dan salah. Pada kolom merupakan dua nilai prediksi dari variabel dependen dan hal ini sukses (1) dan tidak sukses (0), sedangkan pada baris menunjukkan nilai observasi sesungguhnya dari variabel dependen sukses (1) dan tidak sukses (0). Pada model yang sempurna, maka semua kasus akan berada pada diagonal dengan tingkat ketepatan peramalan 100%. Jika model logistic mempunyai homoskedastisitas, maka prosentase yang benar akan sama untuk kedua baris.

Hasil SPSS menunjukkan bahwa pada kolom, prediksi program studi yang sukses ada 3 program studi. Sedangkan pada baris, hasil observasi sesungguhnya yang sukses hanya 2 program studi. Jadi ketepatan model ini adalah 2/3 atau 40%.

## PENUTUP

### Kesimpulan

Hasil analisa dan pembahasan probabilitas quota kelas dan lulusan program studi menggunakan metode *logistic regression* terdapat beberapa hal yang dapat disimpulkan, yaitu:

1. Probabilitas quota kelas dan lulusan program studi dapat diselesaikan dengan asumsi variabel bebas merupakan campuran antara variabel kontinyu dan kategorikal. Sehingga kasus tersebut dapat di analisis dengan *logistic regression*, karena tidak perlu asumsi normalitas data pada variabel bebasnya.
2. Hasil analisis menggunakan SPSS diperoleh pengukuran kualitas lulusan program studi Universitas BSI tahun akademik 2009/2010 berdasarkan rasio jumlah lulusan dan quota kelas menggunakan pendekatan model *logistic regression* tidak dapat menerima model sesuai perhitungan nilai *statistic Hosmer and Lemeshow's Goodness of Fit Test*, dan keputusan model 40% untuk quota kelas dan lulusan.

### Saran

Diperlukan model fit baru dengan rasio quota kelas dan lulusan untuk menentukan kualitas lulusan. Saran penulis yaitu model quota kelas dengan nilai quota per kelas < 30 dan asumsi kelulusan  $\geq 97\%$ .

## UCAPAN TERIMA KASIH

Ucapan terima kasih kami sampaikan yang setulusnya dalam proses penelitian ini kepada:

1. Direktorat Pendidikan Tinggi (DIKTI) untuk program Desentralisasi Penelitian Dosen Pemula (PDP) tahun anggaran 2014.
2. Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) BSI Bandung.

## REFERENSI

- [1] Ary, M. (2012). Menentukan Kualitas Lulusan dengan Probabilitas Quota Kelas dan Lulusan Program Studi Menggunakan Pendekatan Model Logistic Regression. *Paradigma Jurnal Komputer dan Informatika Akademi Bina Sarana Informatika*, 34-45.
- [2] Damanhuri, D. S. (n.d.). *Dunia Esai*. Retrieved September 12, 2012, from Kumpulan esai berbahasa Indonesia: [http://www.duniaesai.com/index.php?option=com\\_content&view=article&id=108:sdm-indonesia-dalam-persaingan-global&catid=37:ekonomi&Itemid=93](http://www.duniaesai.com/index.php?option=com_content&view=article&id=108:sdm-indonesia-dalam-persaingan-global&catid=37:ekonomi&Itemid=93).
- [3] Ghozali, I. (2011). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 19*. Semarang: Badan Penerbit Universitas Diponegoro.
- [4] Gujarati, D. (2003). *Basic Econometrics*. New York: Mc-Grawhill.
- [5] Steven, S. (1946). On The Theory of Scales of Measurement. *Science*, 103.
- [6] Tabachnick, B. (1996). *Using Multivariate Statistics*. New York: Harper Collin.

## Tentang Penulis

**Maxsi Ary, S.Si., S.Kom., M.Kom**, memperoleh gelar Sarjana Saint (S.Si), Jurusan Matematika Fakultas MIPA Universitas Islam Bandung (UNISBA), lulus tahun 2005. Memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK Jabar, lulus tahun 2010. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta, lulus tahun 2011. Saat ini menjadi Dosen di AMIK BSI Bandung, ASM BSI Bandung, AKPAR BSI Bandung, Universitas BSI dan STP Ars Internasional.

**Slamet Risnanto, ST., M.Kom.**, memperoleh gelar Sarjana Teknik (ST), Jurusan Teknik Informatika STMIK Indonesia Mandiri, lulus tahun 2002. Memperoleh gelar Magister Ilmu Komputer (M.Kom) 2010 Konsentrasi Rekayasa Sistem Informasi pada STMIK LIKMI. Saat ini menjadi Dosen di Universitas BSI, STTB Mandala, Universitas Widyatama dan Universitas Sangga Buana.