**Metode Regresi Logistik Biner dan**

**Metode *Chi-Squared Automatic Interaction Detection* (CHAID)**

**(Studi Kasus Kepuasan Pelanggan Pengguna Kartu Telkomsel Terhadap Pelayanan Grapari Telkomsel Di Kota Samarinda, Kalimantan Timur)**

***Binary Logistic Regression Method and***

***Chi-Squared AutomaticInteraction Detection* (CHAID) *Method***

***(Case Study Customer’sSatisfaction of Telkomsel Card Users at Grapari Telkomsel Services in Samarinda City, East Kalimantan)***

**Rizky Rizkyawandy1, Darnah Andi Nohe2, Yuki Novia Nasution3**

1Mahasiswa Program Studi Statistika FMIPA Universitas Mulawarman

2,3Dosen Program Studi Statistika FMIPA Universitas Mulawarman

Email : rizkyawandy@gmail.com1, darnah.98@gmail.com2, yuki.novia.n@gmail.com3

***Abstract***

*Binary logistic regression is a method which is used to obtain the best model that describes the relationship between the dependent variable and independent variables. CHAID method is a method which is used to divide the data into smaller based on the relationship between dependent variable and independent variables. The dependent variable consists of two categorical with nominal scale. This research observed customer is satisfaction of Telkomsel card users at Grapari Telkomsel services quality by analyzing the factors that affect customer satisfaction with service quality of Grapari Telkomsel. The referred factors are tangibles, reliability, responsiveness, assurance and emphaty. Based on the results of analysis using by binary logistic regression method derived factors that affect customer satisfaction significantly is Tangibles, Reliability and Responsiveness, the value of odds ratio is known that the chances of a customer satisfy with the physical evidence is 7.981 times greater than that not because of the physical evidence, for opportunities reliability is 19.621 times greater than that not because of the reliability and the chances of a customer satisfy with the responsiveness is 8.179 times greater than non-responsiveness. While using CHAID method derived factors that affect customer satisfaction significantly, namely reliability and responsiveness with a percentage of 98.7% and 74 people with total of customers who expressed both in reliability and responsiveness.*

*Keywords : Binary logistic method,* CHAID *method, service quality.*

**Pendahuluan**

Regresi logistik biner merupakan salah satu regresi logistik yang digunakan untuk menganalisis hubungan antara satu variabel terikat dan beberapa variabel bebas, dengan variabel terikatnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik, sedangkan variabel bebasnya dapat berupa variabel kualitatif dan kuantitatif, serta bertujuan untuk mendapatkan model terbaik dan sederhana yang menggambarkan hubungan antara variabel terikat dengan variabel-variabel bebas (Hosmer dan Lemeshow, 2009). Metode Chi-Squared Automatic Interacion Detection (CHAID) pada umumnya dikenal sebagai metode pohon klasifikasi (Classification Tree Method). Inti dari metode ini adalah membagi data menjadi kelompok-kelompok yang lebih kecil berdasarkan hubungan antara variabel terikat dan variabel bebas. Metode CHAID digunakan untuk menganalisis variabel kategorik yaitu variabel yang terdiri dari beberapa kategori (Everit dan Skrondal, 2010).

Begitu banyak penyedia jasa layanan seluler di Indonesia, salah satu penyedia jasa layanan seluler tersebut adalah PT Telkomsel (Telekomunikasi Seluler), yang merupakan operator telekomunikasi seluler GSM pertama dan terbesar di Indonesia. PT Telkomsel mulai beroperasi pada tahun 1995 dengan meluncurkan layanan pascabayar pertama di Indonesia serta di Asia yaitu kartu Halo (PT Telkomsel,http://www.telkomsel.com/about/corporate, akses 2 Februari 2015).

Berdasarkan hal-hal tersebut, maka penulis tertarik untuk menganalisis kepuasan pelayanan kepada pelanggan kartu provider Telkomsel di Grapari Telkomsel kota Samarinda, Kalimantan Timur, dengan melihat sejauh mana tingkat kepuasan pelayanan pelanggan kartu provider Telkomsel yang dituangkan dalam judul “Metode Regresi Logistik Biner dan Metode Chi-Squared Automatic Interacion Detection (CHAID) (Studi Kasus Kepuasan Pelanggan Pengguna Kartu Telkomsel Terhadap Pelayanan Grapari Telkomsel di Kota Samarinda, Kalimantan Timur) ”.

**Kualitas Pelayanan**

Dalam Irawan 2008, Kualitas Pelayanan sangat bergantung pada tiga hal, yaitu sistem, teknologi dan manusia. Konsep dari service quality yang populer adalah ServQual yang diyakini mempunyai 5 dimensi yaitu tangible, reliability, responsiveness, assurance dan empathy.

1. Bukti Fisik
2. Keandalan
3. Daya Tanggap
4. Asuransi
5. Empati

**Regresi Logistik Biner**

Analisis Regresi Logistik Biner memiliki tujuan untuk mendapatkan model terbaik dan sederhana yang menggambarkan hubungan antara variabel terikat dengan variabel-variabel bebas, dengan variabel terikatnya yang bersifat dikotomi dan variabel bebasnya dapat berupa variabel kualitatif dan kuantitatif (Hosmer dan Lemeshow, 2000).

Syarat utama dalam regresi logistik biner adalah variabel terikatnya berupa variabel biner yaitu variabel diskrit dengan dua nilai. Misalnya diambil ilustrasi, bila variabel terikat tidak terjadi diberi nilai 0 dan bila variabel terikat terjadi diberi nilai 1, sedangkan variabel bebasnya dapat berupa variabel kuantitatif. Untuk nilai variabel bebas bertipe kualitatif biasanya disebut juga variabel boneka (dummy) dan dilakukan pemberian suatu angka agar dapat dianalisis. Variabel kualitatif ini berupa variabel dikotomi yaitu variabel dengan dua kategori atau dapat berupa variabel polikotomus (variabel yang memiliki lebih dari dua kategori). Untuk variabel berupa kuantitatif didefinisikan secara langsung dan biasanya disebut variabel kontinu.

**Penaksiran Parameter**

Dalam Hosmer dan Lemeshow (2000), penaksiran parameter pada model regresi logistik yang mempunyai variabel terikat dikotomi adalah menggunakan metode Maximum Likelihood Estimation (MLE). Pada dasarnya metode MLE menetapkan asumsi distribusi Bernoulli dan objek pengamatan saling bebas atau memberikan nilai taksiran parameter dengan memaksimumkan fungsi likelihood (likelihood function). Dalam bentuk persamaan matematis, persamaan logistik dinyatakan dalam bentuk berikut :



**Uji Simultan**

Menurut Hosmer dan Lemeshow (2000), uji simultan bertujuan untuk mengetahui pengaruh variabel bebas secara serentak atau simultan terhadap variabel terikat. Langkah pengujian adalah sebagai berikut:

**Hipotesis** :

H0 : β1 = β2 ...= βp = 0, (secara simultan variabel bebas tidak berpengaruh terhadap variabel terikat)

H1 : βj ≠ 0, (minimal ada satu variabel bebas yang berpengaruh secara simultan terhadap variabel terikat) dimana j = 1, 2,....p

**Statistik uji :**

Statistik uji yang digunakan adalah G *likelihood ratio*



Statistik uji G ini mengikuti distribusi *chi-square* dengan derajat bebas banyaknya parameter dalam model. Keputusan uji diperoleh dengan membandingkan G dan nilai (H0 di tolak jika nilai G ≥ (α,v)) , dengan nilai *v* adalah derajat bebas.

**Uji Parsial**

Uji parsial dilakukan dengan menguji setiap βj secara parsial akan menunjukan apakah suatu variabel bebas layak untuk masuk dalam model atau tidak.

**Hipotesis :**

H0 : βj = 0 (tidak ada pengaruh variabel bebas ke-j terhadap variabel terikat) dimana j = 1,2,...,p.

H1 : βj ≠ 0( ada pengaruh variabel bebas ke-j terhadap variabel terikat) dimana j = 1,2,...,p.

**Statistik uji :**

W = 

Statistik Wald mengikuti distribusi normal sehingga untuk memperoleh keputusan pengujian, nilai W dibandingkan dengan nilai Zα/2 (H0 ditolak jika nilai W > Zα/2 atau p-value < α).

**Uji Kesesuaian Model**

Dalam Hosmer dan Lemeshow (2000), uji statistik *Goodness of Fit* digunakan untuk mengetahui kesesuaian model. Langkah uji *Goodness of Fit* adalah sebagai berikut:

**Hipotesis :**

H0 :Tidak ada perbedaan antara hasil pengamatan dengan nilai dugaan.

H1 : Ada perbedaan antara hasil pengamatan dengan nilai dugaan.

**Statistik uji :**

$\hat{C}$ = 

Jika H0 benar, maka distribusi statistik uji mengikuti distribusi *chi-square* dengan derajat bebas g - 2. Daerah penolakan H0 adalah jika > atau *p-value*< α.

**Penafsiran Koefisien Model Regresi Logistik**

Dalam menginterpretasikan atau menafsirkan koefisien *βj* pada regresi logistik, hal yang harus selalu diperhatikan adalah jenis variabel bebasnya, berupa dikotomi (terdiri dua kategori), polikotomus (variabelnya memiliki lebih dari dua kategori), atau kontinu. Sedangkan untuk menginterpretasikan koefisien parameter digunakan *odds ratio* (). Dengan mensubstitusikan model regresi logistik maka di peroleh nilai *odds ratio* sebagai berikut :

 =$\frac{π(1)/[1-π(1)]}{π(0)/[1-π(0)]}$

  =

  =

  = 

  = 

**Metode CHAID**

CHAID adalah singkatan dari *Chi-Squared Automatic Interaction Detection*. CHAID pertama kali diperkenalkan dalam sebuah artikel yang berjudul *“An Exploratory Technique For Investigating Large Quantities Of Categorical Data”* oleh G. V. Kass tahun 1980. CHAID merupakan salah satu tipe dari metode AID (*Automatic Iteration Detection*). Metode AID adalah suatu teknik untuk menganalisis kelompok data berukuran besar dengan membaginya menjadi sub-sub kelompok yang tidak saling tumpang tindih (Kass, 1980). Teknik pemecahan (*splitting*) kelompok menjadi beberapa sub kelompok sehingga diperoleh sub-sub kelompok yang secara maksimal berbeda.

**Algoritma CHAID**

Algoritma CHAID digunakan untuk melakukan pemisahan dan penggabungan kategori-kategori dalam variabel yang dipakai dalam analisisnya. Magdison dalam Bagozzi (1994), menerangkan bahwa langkah-langkah analisis CHAID secara garis besar dapat dibagi menjadi tiga tahap yaitu :

1. **Penggabungan**

Tahap pertama dalam algoritma CHAID adalah penggabungan (*merging*). Pada tahap ini akan diperiksa signifikansi dari masing-masing kategori variabel bebas terhadap variabel terikat.

1. **Pemisahan**

Tahap *splitting*  memilih variabel bebas yang mana akan digunakan sebagai *split node* (pemisah *node*) yang terbaik. Pemilihan dikerjakan dengan membandingkan *p-value* (dari tahap *merging*) pada setiap variabel bebas.

1. **Penghentian**

Tahap *stopping* dilakukan jika proses pertumbuhan pohon harus dihentikan jika tidak ada lagi variabel bebas yang signifikan menunjukan perbedaan terhadap variabel terikat.

**Diagram Pohon Klasifikasi CHAID**

Hasil pembentukan segmen dalam CHAID akan ditampilkan dalam sebuah diagram pohon. Diagram pohon dimulai dari *root node* (node akar) melalui tiga tahap tersebut pada setiap *node* yang terbentuk dan secara berulang. Secara umum diagram pohon CHAID (Lehmann dan Eherler, 2001).

**Metodologi Penelitian**

Dalam penelitian ini data yang digunakan adalah data primer yang dikumpulkan dengan cara melakukan survey dan dengan mengunakan instrumen penelitian berupa kuesioner.Variabel penelitian yang digunakan dalam penelitian ini adalah :

1. Kepuasan pelanggan sebagai variabel terikat (Y) dengan kategori
2. Tidak puas : jika pelanggan mengatakan tidak puas dengan kode 0.
3. Puas : jika pelanggan mengatakan puas dengan kode 1.
4. *Tangible* (bukti fisik) sebagai variabel bebas (X1) dengan kategori
5. Kurang baik: jika pelanggan menjawab kurang baik < 50% dengan kode 0.
6. Baik: jika pelanggan menjawab baik ≥ 50 % dengan kode 1.
7. *Reliability* (keandalan) sebagai variabel bebas (X2) dengan kategori.
8. Kurang baik : jika pelanggan menjawab kurang baik < 50% dengan kode 0.
9. Baik: jika pelanggan menjawab baik ≥ 50 % dengan kode 1.
10. *Responsiveness* (daya tanggap) sebagai variabel bebas (X3) dengan kategori
11. Kurang baik: jika pelanggan menjawab kurang baik < 50% dengan kode 0.
12. Baik: jika pelanggan menjawab baik ≥ 50 % dengan kode 1.
13. *Assurance* (jaminan) sebagai variabel bebas (X4) dengan kategori
14. Kurang baik: jika pelanggan menjawab kurang baik < 50% dengan kode 0.
15. Baik: jika pelanggan menjawab baik ≥ 50 % dengan kode 1.
16. *Emphaty* sebagai variabel penjelas (X5) dengan kategori
17. Kurang baik: jika pelanggan menjawab kurang baik < 50% dengan kode 0.
18. Baik: jika pelanggan menjawab baik ≥ 50 % dengan kode 1.



Gambar 1. Diagram Pohon Metode CHAID

**Metode Analisis Data**

Langkah-langkah dalam menganalisis menggunakan metode Regresi Logistik Biner menggunakan α sebesar 0,05 adalah sebagai berikut :

1. Menghitung jawaban responden berdasarkan atribut pertanyaan dengan kategori tidak puas (0) jika pelanggan menjawab kurang baik < 50% dan sebaliknya puas (1) jika pelanggan menjawab baik ≥ 50%.
2. Mencari persamaan logistik untuk memperoleh model yang akan dipakai.
3. Melakukan estimasi parameter dari model yang diperoleh dengan menggunakan metode Maximum Likelihood Estimator (MLE).
4. Melakukan uji simultan terhadap parameter-parameter model regresi logistik.
5. Melakukan uji parsial untuk setiap masing-masing parameter model regresi logistik.
6. Melakukan uji kesesuaian model guna mengetahui kelayakan dari suatu model yang dihasilkan oleh regresi logistik.

Langkah-langkah dalam menganalisis menggunakan metode CHAID menggunakan α sebesar 0,05 adalah sebagai berikut :

1. Pemisahan, pada tahap ini memilih variabel bebas yang akan digunakan sebagai split node (pemisah node) yang terbaik menggunakan α sebesar 0,05. Pemilihan dikerjakan dengan membandingkan p-value pada setiap variabel bebas.
2. Pemisahan, tahap ini dilakukan jika proses pertumbuhan pohon harus dihentikan ketika pertumbuhan sudah mencapai batas kedalamannya maka pohon klasifikasi dihentikan.
3. Diagram pohon, Proses metode CHAID keseluruhan untuk pembentukan segmen akan ditampilkan dalam sebuah diagram pohon.

**Hasil dan Pembahasan**

**Analisis Regresi Logistik Biner**

**Model Regresi Logistik**

Tabel 1. Penaksiran Parameter Variabel Bebas Terhadap Variabel Terikat

|  |  |  |
| --- | --- | --- |
| Variabel |  | *P-value* |
| X1 | 2,369 | 0,076 |
| X2 | 2,980 | 0,001 |
| X3 | 1,976 | 0,034 |
| X4 | -0,437 | 0,746 |
| X5 | 1,753 | 0,499 |
| Konstan | -4,488 | 0,116 |

serta taksiran model regresi yang diperoleh adalah :



**Uji Simultan**

Berdasarkan perhitungan pada *software SPSS 18*diperoleh:

-2 Log likelihood = 37,373

Log likelihood = $-\frac{37,373}{2}=- $18,686

G =2 {(-18,686) – (89 ln(89) + 11 ln(11) – 100 ln(100))}

= 2 {(-18,686) – ( -34,651)}

= 2 {15,965}

= 31,930

maka diperoleh nilai G adalah 31,930 dimana nilai G >= 31,930 > 11,07 maka dapat diambil keputusan bahwa H0 ditolak dan dapat disimpulkan bahwa minimal ada satu variabel bebas yang berpengaruh terhadap variabel terikat.

**Uji Parsial**

Pembentukan model regresi logistik dengan satu variabel bebas bertujuan untuk mengetahui variabel bebas mana yang berpengaruh secara parsial terhadap variabel terikat.

**Daerah Penolakan**

Tolak H0 jika nilai |W| > Zα/2 atau p-value < α.

Tabel 2. Uji Varsial Variabel X dengan

Variabel Y

|  |  |
| --- | --- |
| Wald | *P-value* |
| 3,154 | 0,076 |
| 10,470 | 0,001 |
| 4,508 | 0,034 |
| 0,105 | 0,746 |
| 0,456 | 0,499 |

Berdasarkan Tabel 2. diperoleh p-value untuk variabel X1, X2, X3, X4, dan X5 masing-masing adalah 0,076; 0,001; 0,034; 0,746; dan 0,499 maka dapat disimpulkan bahwa variabel yang berpengaruh secara parsial terhadap kepuasan pelanggan adalah Keandalan (X2) dan Daya Tanggap (X3).

**Model Terbaik Regresi Logistik Biner**

Tabel 3. Nilai Model Regresi Logistik

|  |  |  |
| --- | --- | --- |
| Variabel |  | *P-value* |
| X1 | 2,077 | 0,044 |
| X2 | 2,977 | 0,001 |
| X3 | 2,102 | 0,020 |

Serta taksiran model regresi logistik yang diperoleh adalah:



**Uji Kesesuaian Model**

Dari hasil perhitungan software SPSS 18 (lampiran 8) diketahui bahwa nilai p-value Hosmer dan Lemeshow Test adalah sebesar 0,261 > α = 0,05 maka H0 diterima. Sehingga dapat disimpulkan tidak ada perbedaan antara hasil pengamatan dengan nilai dugaan atau model regresi logistik tersebut layak digunakan.

**Interpretasi Model**

Tabel 4 menjelaskan bahwa :

1. Pelanggan yang beralasan datang ke Grapari Telkomsel karena pengaruh Bukti Fisik (X1) akan menyatakan puas terhadap pelayanan Grapari Telkomsel Kota Samarinda Sebesar 7,981 kali lebih besar daripada pelanggan yang beralasan datang ke Grapari Telkomsel Kota Samarinda bukan karena adanya pengaruh Bukti Fisik (X1).
2. Pelanggan yang beralasan datang ke Grapari Telkomsel Kota Samarinda karena pengaruh Keandalan (X2) akan menyatakan puas terhadap pelayanan Grapari Telkomsel Kota Samarinda sebesar 19,621 kali lebih besar dari pada pelanggan yang beralasan datang ke Grapari Telkomsel Kota Samarinda bukan karena adanya pengaruh Keandalan (X2).
3. Pelanggan yang beralasan datang ke Grapari Telkomsel Kota Samarinda karena pengaruh Daya Tanggap (X3) akan menyatakan puas terhadap pelayanan Grapari Telkomsel Kota Samarinda sebesar 8,719 kali lebih besar dari pada pelanggan yang beralasan datang ke Grapari Telkomsel Kota Samarinda bukan karena adanya pengaruh Daya Tanggap (X3).

Tabel 4. Kontribusi Variabel X Terhadap Variabel Y

|  |  |
| --- | --- |
| Variabel | Exp() |
| X1 | 7,981 |
| X2 | 19,621 |
| X3 | 8,179 |

**Metode CHAID**

Telah diketahui bahwa dalam penelitian ini menggunakan 5 variabel bebas dan dalam analisis ini menggunakan skala Guttman, maka hanya memiliki nilai dua kategorik, sehingga tidak terjadi tahap penggabungan.Proses metode CHAID dapat di klasifikasikan dengan diagram pohon klasifikasi CHAID (CHAID Classification Tree).Terlihat jelas pada Gambar 3. node-node serta hasil uji Chi-Square yang didapat pada setiap variabel bebas yang berperan mempengaruhi kepuasan pelanggan (Y), yaitu dimana terdapat variabel keandalan (X2), dan daya tanggap (X3) yang mempengaruhi kepuasan pelanggan (Y).

Dari Tabel 6 dapat dilihat bahwa nilai presentase terbesar dan jumlah pelanggan terbesar yang merasa tidak puas pada pelayanan Grapari Telkomsel Kota Samarinda ada pada segmen ke-3 yaitu, golongan pelanggan yang menyatakan kurang baik pada keandalan (X2) dengan jumlah pelanggan 8 dan presentase sebesar 8%. Sedangkan nilai presentase terbesar dan jumlah pelanggan terbesar yang merasa puas pada pelayanan Grapari Telkomsel Kota Samarinda ada pada segmen ke-2 yaitu, golongan pelanggan yang menyatakan baik pada keandalan (X2), dan menyatakan baik pada daya tanggap (X3) dengan jumlah pelanggan 74 dan presentase sebesar 74%. Maka dapat diketahui bahwa pelanggan yang menyatakan puas terhadap pelayanan Grapari Telkomsel Kota Samarinda lebih banyak dibandingkan dengan pelanggan yang menyatakan tidak puas terhadap pelayanan Grapari Telkomsel Kota Samarinda.



Gambar 3. Diagram Pohon Klasifikasi Hasil Analisis CHAID

Tabel 5. Segmentasi Kepuasan Pelanggan Pengguna Kartu Telkomsel

|  |  |
| --- | --- |
| Segmen | Karakteristik |
| Segmen ke-1 | Golongan pelanggan yang menyatakan baik pada keandalan (X2), dan kurang baik pada daya tanggap (X3). |
| Segmen ke-2 | Golongan pelanggan yang menyatakan baik pada keandalan (X2), dan baik pada daya tanggap(X3), |
| Segmen ke-3 | Golongan pelanggan yang menyatakan kurang baik pada keandalan (X2). |

Tabel 6. Presentase Setiap Segmen

|  |  |  |
| --- | --- | --- |
| Segmen | Tidak Puas | Puas |
|  |  |  |  |
| 1 | 2 | 2% | 8 | 8% |
| 2 | 1 | 1% | 74 | 74% |
| 3 | 8 | 8% | 7 | 7% |

**Kesimpulan**

Berdasarkan hasil perhitungan dan pembahasan yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut :

1. Faktor-faktor yang mempengaruhi kepuasan pelanggan (Y) pengguna kartu Telkomsel secara signifikan menggunakan metode Regresi Logistik Biner adalah bukti fisik (X1), keandalan (X2) , serta daya tanggap (X3).
2. Faktor-faktor yang mempengaruhi kepuasan pelanggan (Y) pengguna kartu Telkomsel secara signifikan menggunakan metode Chi-Square Automatic Iteraction Detection (CHAID) adalah keandalan (X2) , serta daya tanggap (X3).

**Daftar Pustaka**

Bagozzi, R.P. 1994. *Advanced Methods of Marketing Research*. U.K :Blackwell Publishers Ltd. Oxford.

Everit, B. S & Skrondal, A. 2010. *The Cambridge Dictionary of Statistics Fourth Edition*. Cambridge : Cambridge University Press.

Hosmer, D. W. and Lemeshow, S. 2000. *Applied Logistic Regression*. U.S New York : John Wiley and Sons. Inc.

http://www.telkomsel.com/about/corporate, di akses pada tanggal 2 Februari 2015.

Irawan, Handi, 2008. *Sepuluh Prinsip KepuasanPelanggan*. Jakarta: Media Elex Komputindo Kelompok Gramedia.

Kass GV. 1980. *An Explotary Technique for Investigating Large Quantities of Categorical Data*. App. Statist (29):119-127.