



PROSIDING SEMINAR NASIONAL METODE KUANTITATIF

SNMK
2017

PENGGUNAAN MATEMATIKA, STATISTIKA,
DAN KOMPUTER DALAM BERBAGAI DISIPLIN ILMU
UNTUK MEWUJUDKAN KEMAKMURAN BANGSA



**SEMINAR NASIONAL
METODE KUANTITATIF
2017**

PROSIDING
Seminar Nasional
Metode Kuantitatif 2017

ISBN No. 978-602-98559-3-7

Penggunaan Matematika, Statistika, dan Komputer dalam Berbagai Disiplin Ilmu
untuk Mewujudkan Kemakmuran Bangsa

Editor :

Prof. Mustofa Usman, Ph.D
Dra. Wamiliana, M.A., Ph.D.

Layout & Design :

Shela Malinda Tampubolon

Alamat :

Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung, Bandar Lampung
Jl. Prof. Dr. Sumantri Brojonegoro No. 1 Bandar Lampung
Telp. 0721-701609/Fax. 0721-702767

KATA SAMBUTAN KETUA PELAKSANA SEMINAR NASIONAL METODE KUANTITATIF 2017

Seminar Nasional Metode Kuantitatif 2017 diselenggarakan oleh Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung yang dilaksanakan pada tanggal 24 – 25 November 2017. Seminar terselenggara atas kerja sama Jurusan Matematika FMIPA, Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Unila, dan Badan Pusat Statistik (BPS).

Peserta dari Seminar dihadiri lebih dari 160 peserta dari 11 institusi di Indonesia, diantaranya : Kementerian Pendidikan dan Kebudayaan, Badan Pusat Statistik, Universitas Indonesia, Institut Teknologi Bandung, Universitas Sriwijaya, Universitas Jember, Universitas Islam Negeri Sunan Gunung Djati, Universitas Cendrawasih, Universitas Teknokrat Indonesia, Universitas Malahayati, dan Universitas Lampung. Dengan jumlah artikel yang disajikan ada sebanyak 48 artikel hal ini merefleksikan pentingnya seminar nasional metode kuantitatif dengan tema “penggunaan matematika, statistika dan computer dalam berbagai disiplin ilmu untuk mewujudkan kemakmuran bangsa”.

Kami berharap seminar ini menjadi tempat untuk para dosen dan mahasiswa untuk berbagi pengalaman dan membangun kerjasama antar ilmunan. Seminar semacam ini tentu mempunyai pengaruh yang positif pada iklim akademik khususnya di Unila.

Atas nama panitia, kami mengucapkan banyak terima kasih kepada Rektor, ketua LPPM Unila, dan Dekan FMIPA Unila serta ketua jurusan matematika FMIPA Unila dan semua panitia yang telah bekerja keras untuk suksesnya penyelenggaraan seminar ini.

Dan semoga seminar ini dapat menjadi agenda tahunan bagi jurusan matematika FMIPA Unila`

Bandar Lampung, Desember 2017

Prof. Mustofa Usman,Ph.D

Ketua Pelaksana

KEPANITIAAN

Penasehat : 1. Prof. Dr. Hasriadi Mat Akin, M.P
2. Prof. Dr. Bujang Rahman
3. Prof. Dr. Ir. Kamal, M.Sc
4. Ir. Warsono, M.Sc., Ph.D
5. Dr. Hartoyo, M.Si

Pengarah : 1. Prof. Warsito, S.Si., DEA, Ph.D
2. Prof. Dr. Sutopo Hadi, S.Si., M.Sc
3. Dian Kurniasari S.Si., M.Sc
4. Drs. Suratman Umar, M.Sc.

Penanggung Jawab : Dra. Wamiliana, M.A., Ph.D

Ketua Pelaksana : Prof. Drs. Mustofa, M.A., Ph.D

Sekretaris : Dra. Dorrah Aziz, M.Si

Bendahara : Amanto, S.Si., M.Sc

Kesekretariatan : Subian Saidi, S.Si., M.Si

Dr. Notiragayu, M.Si

- Syamsu Huda, S.I.P., M.M

- Srimiati, S.Pd

- Johan, S.P

- Riendi Ferdian, S.I.P

- Siti Marbiyah, S.Si

- Rosihin Anwar, S.Kom

- Shela Malinda T

- Della Desiyana

- Nandra Adi Prayoga

- Himatika

Seksi-seksi :

Acara : Dr. Aang Nuryaman, M.Si

Dr. Khoirin Nisa, M.Si

Drs. Rudi Ruswandi, M.Si

Drs. Eri Setiawan, M.Si

Konsumsi : Widiarti S.Si., M.Si
Dr. Asmiati, M.Si

Transportasi/akomodasi : Drs. Nusyirwan, M.Si
Agus Sutrisno, S.Si., M.Si

Perlengkapan : Drs. Tiryono R., M.Sc., Ph.D
- Agus Suroso, A.Md
- Tamrinsyah
- Supriyadi
- Drajat
- Maeda Sulistiana

Reviewer : Drs. Suharsono, M.Sc., Ph.D
- Dr. La Zakaria S.Si., M.Sc
- Dr. Muslim Ansori, S.Si., M.Si
- Dr. Ir. Netti Herawati, M.Sc

DAFTAR ISI

KATA SAMBUTAN	iii
KEPANITIAAN	iv
DAFTAR ISI	vi
Aplikasi Metode Analisis Homotopi (HAM) pada Sistem Persamaan Diferensial Parsial Homogen (<i>Fauzia Anisatul F, Suharsono S, dan Dorrah Aziz</i>)	1
Simulasi Interaksi Angin Laut dan Bukit Barisan dalam Pembentukan Pola Cuaca di Wilayah Sumatera Barat Menggunakan Model Wrf-Arw (<i>Achmad Rafli Pahlevi</i>)	7
Penerapan Mekanisme Pertahanan Diri (Self-Defense) sebagai Upaya Strategi Pengurangan Rasa Takut Terhadap Kejahatan (Studi Pada Kabupaten/Kota di Provinsi Lampung yang Menduduki Peringkat <i>Crime Rate</i> Tertinggi) (<i>Teuku Fahmi</i>)	18
Tingkat Ketahanan Individu Mahasiswa Unila pada Aspek Soft Skill (<i>Pitojo Budiono, Feni Rosalia, dan Lilih Muflihah</i>)	33
Metode Analisis Homotopi pada Sistem Persamaan Diferensial Parsial Linear Non Homogen Orde Satu (<i>Atika Faradilla dan Suharsono S</i>)	44
Penerapan Neural Machine Translation Untuk Eksperimen Penerjemahan Secara Otomatis pada Bahasa Lampung – Indonesia (<i>Zaenal Abidin</i>)	53
Ukuran Risiko Cre-Var (<i>Insani Putri dan Khreshna I.A.Syuhada</i>)	69
Penentuan Risiko Investasi dengan Momen Orde Tinggi $V @ R - C_v @ R$ (<i>Marianik dan Khreshna I.A.Syuhada</i>)	77
Simulasi Komputasi Aliran Panas pada Model Pengering Kabinet dengan Metode Beda Hingga (<i>Vivi Nur Utami, Tiryono Ruby, Subian Saidi, dan Amanto</i>)	83
Segmentasi Wilayah Berdasarkan Derajat Kesehatan dengan Menggunakan <i>Finite Mixture Partial Least Square</i> (Fimix-Pls) (<i>Agustina Riyanti</i>)	90
Representasi Operator Linier Dari Ruang Barisan Ke Ruang Barisan $L 3/2$ (<i>Risky Aulia Ulfa, Muslim Ansori, Suharsono S, dan Agus Sutrisno</i>)	99
Analisis Rangkaian Resistor, Induktor dan Kapasitor (RLC) dengan Metode Runge-Kutta Dan Adams Bashforth Moulton (<i>Yudandi K.A., Agus Sutrisno, Amanto, dan Dorrah Aziz</i>)	110

Representasi Operator Linier dari Ruang Barisan Ke Ruang Barisan L 13/12 (<i>Amanda Yona Ningtyas, Muslim Ansori, Subian Saidi, dan Amanto</i>)	116
Desain Kontrol Model Suhu Ruangan (<i>Zulfikar Fakhri Bismar dan Aang Nuryaman</i>)	126
Penerapan Logika Fuzzy pada Suara Tv Sebagai Alternative Menghemat Daya Listrik (<i>Agus Wantoro</i>)	135
Clustering Wilayah Lampung Berdasarkan Tingkat Kesejahteraan (<i>Henida Widyatama</i>).....	149
Pemanfaatan Sistem Informasi Geografis Untuk Valuasi Jasa Lingkungan Mangrove dalam Penyakit Malaria di Provinsi Lampung (<i>Imawan A.Q., Samsul Bakri, dan Dyah W.S.R.W.</i>)	156
Analisis Pengendalian Persediaan Dalam Mencapai Tingkat Produksi <i>Crude Palm Oil</i> (CPO) yang Optimal di PT. Kresna Duta Agroindo Langling Merangin-Jambi (<i>Marcelly Widya W., Hery Wibowo, dan Estika Devi Erinda</i>)	171
Analisis <i>Cluster Data Longitudinal</i> pada Pengelompokan Daerah Berdasarkan Indikator IPM di Jawa Barat (<i>A.S Awalluddin dan I. Taufik</i>).....	187
Indek Pembangunan Manusia dan Faktor Yang Mempengaruhinya di Daerah Perkotaan Provinsi Lampung (<i>Ahmad Rifa'i dan Hartono</i>).....	195
<i>Parameter Estimation Of Bernoulli Distribution Using Maximum Likelihood and Bayesian Methods</i> (<i>Nurmaita Hamsyiah, Khoirin Nisa, dan Warsono</i>).....	214
Proses Pengamanan Data Menggunakan Kombinasi Metode Kriptografi <i>Data Encryption Standard</i> dan <i>Steganografi End Of File</i> (<i>Dedi Darwis, Wamiliana, dan Akmal Junaidi</i>).	228
<i>Bayesian Inference of Poisson Distribution Using Conjugate A and Non-Informative Prior</i> (<i>Misgiyati, Khoirin Nisa, dan Warsono</i>).	241
Analisis Klasifikasi Menggunakan Metode Regresi Logistik Ordinal dan Klasifikasi Naïve Bayes pada Data Alumni Unila Tahun 2016 (<i>Shintia F., Rudi Ruswandi, dan Subian Saidi</i>)....	251
Analisis Model <i>Markov Switching Autoregressive</i> (MSAR) pada <i>Data Time Series</i> (<i>Aulianda Prasyanti, Mustofa Usman, dan Dorrah Aziz</i>).....	263
Perbandingan Metode Adams Bashforth-Moulton dan Metode Milne-Simpson dalam Penyelesaian Persamaan Diferensial Euler Orde-8 (<i>Faranika Latip., Dorrah Aziz, dan Suharsono S</i>).....	278
Pengembangan Ekowisata dengan Memanfaatkan Media Sosial untuk Mengukur Selera Calon Konsumen (<i>Gustafika Maulana, Gunardi Djoko Winarso, dan Samsul Bakri</i>).....	293
Diagonalisasi Secara Unger Matriks Hermite dan Aplikasinya pada Pengamanan Pesan Rahasia (<i>Abdurrois, Dorrah Aziz, dan Aang Nuryaman</i>)	308

Pembandingan Metode Runge-Kutta Orde 4 dan Metode Adam-Bashfort Moulton dalam Penyelesaian Model Pertumbuhan Uang yang Diinvestasikan (<i>Intan Puspitasari, Agus Sutrisno, Tiryono Ruby, dan Muslim Ansori</i>)	328
Menyelesaikan Persamaan Diferensial Linear Orde-N Non Homogen dengan Fungsi Green (<i>Fathurrohman Al Ayubi, Dorrah Aziz, dan Muslim Ansori</i>).....	341
Penyelesaian Kata Ambigu pada Proses Pos Tagging Menggunakan Algoritma <i>Hidden Markov Model</i> (HMM) (<i>Agus Mulyanto, Yeni Agus Nurhuda, dan Nova Wiyanto</i>).....	347
Sistem Temu Kembali Citra Daun Tumbuhan Menggunakan Metode Eigenface (<i>Supiyanto dan Samuel A. Mandowen</i>)	359
Efektivitas Model <i>Problem Solving</i> dalam Meningkatkan Kemampuan Berfikir Lancar Mahasiswa pada Materi Ph Larutan (<i>Ratu Betta Rudibyani</i>).....	368
<i>The Optimal Bandwidth for Kernel Density Estimation of Skewed Distribution: A Case Study on Survival Data of Cancer Patients</i> (<i>Netti Herawati, Khoirin Nisa, dan Eri Setiawan</i>).....	380
Karakteristik Larutan Kimia Di Dalam Air Dengan Menggunakan Sistem Persamaan Linear (<i>Titik Suparwati</i>).....	389
Bentuk Solusi Gelombang Berjalan Persamaan $\Delta\Delta$ mKdV Yang Diperumum (<i>Notiragayu, Rudi Ruswandi, dan La Zakaria</i>).....	398
Pendugaan Blup Dan Eblup(Suatu Pendekatan Simulasi) (<i>Nusyirwan</i>).....	403

ANALISIS KLASIFIKASI MENGGUNAKAN METODE REGRESI LOGISTIK ORDINAL DAN KLASIFIKASI NAÏVE BAYES PADA DATA ALUMNI UNILA TAHUN 2016

Shintia Faramudhita, Rudi Ruswandi, Subian Saidi

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung

Email : shintiafaramudhita@gmail.com, ruswandi@ymail.com, subian.saidi@gmail.com

ABSTRAK

Klasifikasi adalah proses pencarian sekumpulan pola, model atau fungsi yang menggambarkan dan membedakan objek data untuk dikelompokkan ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Pada penelitian ini akan dilakukan klasifikasi tingkat kelancaran alumni unila tahun 2016 yang dibagi menjadi tiga kelas bertingkat, yaitu kelas Tidak Lancar (TL), kelas Kurang Lancar (KL) dan kelas Lancar (L). Metode analisis yang digunakan adalah Regresi Logistik Ordinal yang merupakan metode pengklasifikasi dengan teknik statistika dan Naïve Bayes yang merupakan metode pengklasifikasi dengan menggabungkan teknik statistika dan data mining. Penelitian ini bertujuan untuk mengetahui metode mana yang mempunyai tingkat akurasi yang lebih baik dalam mengklasifikasi tingkat kelancaran alumni unila tahun 2016 dalam mencari pekerjaan. Dengan melakukan sepuluh kali pengulangan klasifikasi dari masing-masing metode, didapat bahwa Regresi Logistik Ordinal mempunyai rata-rata tingkat error yang lebih kecil yaitu sebesar dibandingkan dengan Naïve Bayes.

Kata Kunci : *Klasifikasi, Regresi Logistik Ordinal, Klasifikasi Naïve Bayes*

I. PENDAHULUAN

Menurut [1] , klasifikasi adalah proses pencarian sekumpulan model, pola atau fungsi yang menggambarkan dan membedakan objek data untuk dikelompokkan ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Variabel yang digunakan dalam klasifikasi terdiri dari variabel prediktor yang merupakan faktor-faktor yang mempengaruhi atau dapat menggambarkan variabel respon. Dalam hal ini variabel respon berupa variabel kategorik baik yang mempunyai pengurutan dalam penomoran (ordinal) maupun tidak (nominal).

Beberapa metode statistika yang dapat digunakan dalam klasifikasi melalui analisis data kategorik, yaitu Regresi Logistik Biner, Regresi Logistik Multinomial, Regresi Logistik Ordinal dan Model Log Linier. Pada penelitian ini data yang digunakan adalah klasifikasi tingkat kelancaran alumni Universitas Lampung (Unila) tahun 2016 dalam mendapatkan pekerjaan. Klasifikasi alumni dibagi menjadi kelompok tidak lancar, kelompok kurang lancar dan kelompok lancar berdasarkan lamanya waktu yang dibutuhkan untuk mendapatkan pekerjaan . Variabel respon yang digunakan berbentuk kategorik bertingkat atau ordinal. Sehingga metode yang dapat digunakan adalah Regresi Logistik Ordinal.

Metode pengklasifikasi lain yang dapat digunakan adalah Klasifikasi Naïve Bayes. Naïve Bayes merupakan metode penggolongan berdasarkan probabilitas sederhana dan dirancang untuk dipergunakan dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung (independen) meskipun asumsi ini tidak terpenuhi, dalam prakteknya masih berfungsi dengan baik. Naïve Bayes merupakan gabungan dari teknik statistika yang didasari oleh Teorema Bayes dan data mining *machine learning* yang mampu memberikan informasi berguna tentang teknik klasifikasi untuk menentukan alumni Unila akan bergabung dengan kelompok mana dalam mencari pekerjaan.

Karena kedua metode klasifikasi diatas memiliki perbedaan, dimana Regresi Logistik Ordinal merupakan metode klasifikasi yang menggunakan teknik statistika, sedangkan Naïve Bayes adalah metode klasifikasi yang menggabungkan teknik statistika dengan data mining. Sehingga menarik untuk mengkaji perbandingan penggunaan kedua metode tersebut pada data alumni Unila lulusan 2016. Untuk mengetahui metode mana yang mempunyai tingkat akurasi yang lebih baik. Akan tetapi, penelitian ini hanya dikhususkan untuk pengklasifikasian data Alumni Unila tahun 2016.

Berdasarkan latar belakang diatas maka penlitian ini akan menganalisis tingkat ketepatan metode Regresi Logistik Ordinal dan Naive Bayes dalam mengklasifikasi tingkat kelancaran alumni Unila tahun 2016 dalam mencari pekerjaan berdasarkan rata-rata tingkat error dari masing-masing metode.

II. TINJAUAN PUSTAKA

a. Regresi Logistik Ordinal

Regresi Logistik Ordinal merupakan salah satu metode statistika untuk menganalisis variabel respon yang mempunyai skala data ordinal yang memiliki tiga kategori atau lebih. Sedangkan variabel prediktor yang digunakan berupa data katgorik dan atau kuantitatif. Pada regresi logistik ordinal model berupa Model Logit Kumulatif (*Cumulative Logit Models*). Model logit kumulatif ini diperoleh dengan membandingkan peluang kumulatif yaitu peluang kurang dari atau sama dengan kategori respon ke-j pada i variabel prediktor yang dinyatakan dalam vektor x_i adalah $P(Y \leq j | X_i)$, dengan peluang lebih dari kategori respon ke-j pada i variabel prediktor vektor X_i $P(Y > j | X_i)$. Peluang kumulatif $P(Y \leq j | X_i)$ didefinisikan sebagai berikut :

$$\begin{aligned} P(Y \leq j | X_i) = \pi(x_i) &= \frac{e^{g_j(x)}}{1 + e^{g_j(x)}} \\ &= \frac{e^{\beta_{j0} - \sum_{i=1}^i \beta_i x_i}}{1 + e^{\beta_{j0} - \sum_{i=1}^i \beta_i x_i}} \end{aligned} \quad (1)$$

Model logit kumulatif didefinisikan dengan :

$$\begin{aligned} g_j(x) &= \ln \left[\frac{\pi_j(x)}{1 - \pi_j(x)} \right] \\ &= \beta_{j0} - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_i x_i \end{aligned} \quad (2)$$

Dengan, j adalah jumlah kategori variabel respon $j = 1, 2, \dots, J$ dan i adalah jumlah variabel prediktor. Jika terdapat kategori respon dimana $j=1,2,3$, maka Odds Ratio atau nilai peluang untuk tiap kategori respon dapat dihitung dengan menggunakan persamaan dibawah ini :

$$\pi_1(x) = \frac{e^{g_1(x)}}{1 + e^{g_1(x)}} \quad (3)$$

$$\pi_2(x) = \frac{e^{g_2(x)} - e^{g_1(x)}}{(1 + e^{g_2(x)})(1 + e^{g_1(x)})} \quad (4)$$

$$\pi_k(x) = 1 - \pi_{k-1}(x) \quad (5)$$

Untuk menentukan kelas dari suatu objek dapat dilihat dari nilai peluang kategori yang paling besar. Metode yang digunakan untuk menaksir parameter-parameter pada Regresi Logistik Ordinal adalah *Maximum Likelihood Estimation* (MLE) dengan fungsi likelihood :

$$l(\beta) = \prod_{i=1}^n [\pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}} \pi_3(x_i)^{y_{3i}}] \quad (6)$$

Dengan,

$$y_{ji} = \begin{cases} 1 & \text{untuk } y=j \\ 0 & \text{untuk } y \neq j \end{cases}$$

Dari persamaan di atas didapatkan fungsi *ln- Likelihood* sebagai berikut :

$$\begin{aligned} \ln l(\beta) &= \sum_{i=1}^n \{y_{1i} \ln[\pi_1(x_i)] + y_{2i} \ln[\pi_2(x_i)] + y_{3i} \ln[\pi_3(x_i)]\} \\ &= \sum_{i=1}^n \{y_{1i} \ln \left[\frac{e^{g_1(x)}}{1 + e^{g_1(x)}} \right] + y_{2i} \ln \left[\frac{e^{g_2(x)} - e^{g_2(x)}}{(1 + e^{g_2(x)})(1 + e^{g_1(x)})} \right] + y_{3i} \ln \left[1 - \frac{e^{g_2(x)}}{1 + e^{g_2(x)}} \right]\} \end{aligned} \quad (7)$$

Untuk mendapatkan nilai pendugaan parameter dari fungsi *ln-likelihood* pada regresi logistik ordinal dilakukan metode iterasi *Newton Raphson* [2]. Persamaan *Newton Raphson*.

$$\beta^{(t+1)} = \beta^t - (H^t)^{-1} u^t \quad (8)$$

Model yang telah terbentuk dari parameter-parameter yang di estimasi menggunakan *maximum likelihood estimation* selanjutnya diuji signifikansinya baik secara keseluruhan maupun secara parsial.

1. Uji Ratio Likelihood

Uji Ratio Likelihood dilakukan untuk menguji kesesuaian model dengan variabel-variabel prediktor secara keseluruhan [3].

Adapun hipotesis yang digunakan dalam uji ratio likelihood

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

H_1 : paling sedikit salah satu dari $\beta_r \neq 0$ dengan $r = 1, 2, \dots, p$

Dengan statistik uji

$$G = -2 \ln \left[\frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}} \right] \quad (9)$$

$$G = 2 \{ \sum_{i=1}^n [y_i \ln(\hat{\pi}_i) + (1-y_i) \ln(1-\hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \} \quad (10)$$

dengan :

n_1 = banyaknya observasi berkategori 1
 n_0 = banyaknya observasi berkategori 0
 n = banyaknya observasi ($n_1 + n_0$)

Statistik uji G mengikuti distribusi *chi-square*. Sehingga untuk mengambil keputusan dilakukan perbandingan dengan χ^2 tabel. Kriteria penolakan tolak H_0 jika $\chi^2_{hit} > \chi^2_{(db,a)}$

2. Uji Wald

Uji Wald dilakukan untuk mengetahui variabel-variabel prediktor mempengaruhi variabel respon secara individu dengan kata lain apakah suatu variabel prediktor layak dimasukkan kedalam model [3].

Hipotesis yang digunakan dalam uji wald :

$$H_0 : \beta_i = 0;$$

$$H_1 : \beta_i \neq 0, \text{ dengan } i = 1, 2, \dots, n$$

Dengan statistik uji

$$W_r = \left[\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \right]^2 \quad (11)$$

Statistik uji W_r mengikuti sebaran normal baku. Sehingga untuk memperoleh keputusan dilakukan perbandingan dengan distribusi normal baku (Z). Dengan kriteria pengambilan keputusan tolak H_0 Jika $W_i > Z_{\alpha/2}$.

b. Naive Bayes

Menurut [1], metode Klasifikasi Bayesian merupakan pengklasifikasian secara statistika yang memprediksi peluang anggota kelas tertentu berdasarkan database yang memenuhi syarat keanggotaan kelas tersebut. Dengan formula Naive Bayes :

$$P(Y_j|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y_j)}{P(X)} \quad (12)$$

Dengan,

$P(Y_j|X_i)$ = Probabilitas posterior kelas Y pada semua fitur vektor

$P(Y)$ = Probabilitas prior kelas Y (*prior probability*)

$P(X_i|Y_j)$ = probabilitas posterior semua fitur vektor X pada kelas Y

$\prod_{i=1}^n P(X_i|Y_j)$ = $P(X_1|Y_j) \times P(X_2|Y_j) \times \dots \times P(X_n|Y_j)$

$P(X)$ = Probabilitas X

Pada Klasifikasi Naive Bayes hasil klasifikasi ditentukan dengan melihat nilai $P(Y_j|X)$ paling besar dari setiap variabel Y.

c. Error Rate

Untuk mengetahui tingkat akurasi hasil klasifikasi, maka dilakukan uji ketepatan hasil klasifikasi dengan menggunakan APER (*Apparent Error Rate*) atau yang disebut juga dengan laju error. Untuk menghitung

nilai APER, langkah yang harus dilakukan membentuk tabel perbandingan hasil klasifikasi berdasarkan observasi dengan hasil klasifikasi berdasarkan prediksi suatu metode yang disebut dengan matrik konfusi hasil klasifikasi [1].

Tabel 1. Matrik Konfusi Hasil Klasifikasi

<i>Fij</i>		Kelas Asli (Hasil Observasi)		
		Kelas = 1	Kelas = 2	Kelas = 3
Kelas Prediksi (Hasil Prediksi)	Kelas = 1	<i>F11</i>	<i>F12</i>	<i>F13</i>
	Kelas = 2	<i>F21</i>	<i>F22</i>	<i>F23</i>
	Kelas = 3	<i>F31</i>	<i>F32</i>	<i>F33</i>

Dengan,

F11 = Jumlah alumni kelas 1 pada kelas asli dan kelas 1 pada kelas prediksi

F12 = Jumlah alumni kelas 1 pada kelas asli dan kelas 2 pada kelas prediksi

F13 = Jumlah alumni kelas 1 pada kelas asli dan kelas 3 pada kelas prediksi

.
.
.

Fij = Jumlah alumni kelas *i* pada kelas asli dan kelas *j* pada kelas prediksi

Selanjutnya dapat dilakukan perhitungan nilai APER dengan formulasi sebagai berikut

$$APER = \frac{f_{12}+f_{13}+f_{21}+f_{23}+f_{31}+f_{32}}{f_{11}+f_{12}+f_{13}+f_{21}+f_{22}+f_{23}+f_{31}+f_{32}+f_{33}} \times 100\% \quad (13)$$

$$\text{Tingkat Akurasi} = 1 - \text{APER} \quad (14)$$

Suatu metode dikatakan memiliki tingkat akurasi yang baik jika mempunyai nilai APER yang kecil dan tingkat akurasi yang tinggi.

d. Uji Dua Rata-Rata

Uji hipotesis dua rata-rata digunakan mengetahui ada atau tidaknya perbedaan (kesamaan) antara dua data .

Hipotesis :

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Statistik Uji :

$$t = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{S_p^2(1/n + 1/m)}} \quad (15)$$

Dengan

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

Keputusan :

Tolak H_0 jika $t < -t_{(\alpha,df)}$ atau $t > t_{(\alpha,df)}$

III. METODOLOGI PENELITIAN

a. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data *Tracer Study* Universitas Lampung (Unila) 2016 yang diperoleh dari UPT. Pengembangan Karir dan Kewirausahaan Universitas Lampung. . Data yang digunakan adalah alumni yang mempunyai riwayat mencari pekerjaan, sedangkan yang tidak pernah mencari pekerjaan sama sekali baik melanjutkan studi maupun berwirausaha tidak dimasukkan dalam objek penelitian. Software yang digunakan yaitu Rstudio (3.3.3).

Tabel.2 Variabel Prediktor

Variabel Prediktor (X)	Kategori	
Program Studi (X_1)	$X_{1=(1)}$ = Eksak	
	$X_{1=(2)}$ = Non Eksak	
Jenjang Pendidikan (X_2)	$X_{2=(1)}$ = D3	
	$X_{2=(2)}$ = S1	
IPK (X_3)	$X_{3=(1)}$ = ≤ 2.75	
	$X_{3=(2)}$ = $>2.75 - \leq 3.5$	
	$X_{3=(3)}$ = >3.5	
Lama Study (X_4)	$X_{4=(1)}$ = ≤ 3.5 tahun	≤ 3 tahun
	$X_{4=(2)}$ = $> 3.5-4.5$ tahun	$>3 - \leq 4$ tahun
	$X_{4=(3)}$ = >4.5 tahun	>4 tahun
Cara Mencari Pekerjaan (X_5)	$X_{5=(1)}$ = Melalui Media	
	$X_{5=(2)}$ = Secara Mandiri	
	$X_{5=(3)}$ = Relasi	
Penguasaan Pengetahuan diluar Program studi (X_6)	$X_{6=(1)}$ = Sangat Tidak Menguasai	
	$X_{6=(2)}$ = Tidak Menguasai	
	$X_{6=(3)}$ = Cukup Menguasai	
	$X_{6=(4)}$ = Menguasai	
	$X_{6=(5)}$ = Sangat Menguasai	

Langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Menentukan klasifikasi awal

Pada studi kasus klasifikasi tingkat kelancaran alumni dalam mendapatkan pekerjaan dapat dibuat klasifikasi awal dengan kriteria waktu yang dibutuhkan untuk mendapatkan pekerjaan pertama :

Tabel.3 Variabel Respon

Variabel Respon (Y)	Kriteria
Tidak Lancar (TL) = 1	Mendapatkan pekerjaan dalam waktu > 12 bulan setelah wisuda
Kurang Lancar (KL) = 2	Mendapatkan pekerjaan dalam waktu > 6 bulan -12 bulan setelah wisuda
Lancar (L) = 3	Mendapatkan pekerjaan dalam ≤ 6 setelah wisuda

2. Membagi data menjadi dua, yaitu data training 75% dan data testing 25%. Dengan data yang sama dilakukan sepuluh kali pengulangan pembagian data training dan data testing, proporsi jumlah data testing data training yang sama tetapi kombinasi data yang berbeda.
3. Membuat model regresi logistik ordinal
 - a. Membentuk model awal regresi logistik ordinal dengan menggunakan data training.
 - b. Menguji signifikansi parameter secara keseluruhan dengan menggunakan Uji Ratio Likelihood
 - c. Menguji parameter secara parsial dengan Uji Wald, pengujian ini dilakukan untuk mengetahui variabel-variabel prediktor mempengaruhi variabel respon secara individu.
 - d. Pembentukan model akhir regresi logistik ordinal
 - e. Menentukan klasifikasi data testing menggunakan model akhir. Dalam regresi logistik ordinal kelas hasil prediksi adalah kelas yang memiliki nilai peluang paling tinggi.
 - f. Menghitung nilai APER dan akurasi dari model yang terbentuk.
4. Naive Bayes

Adapun tahapan klasifikasi Naive Bayes sebagai berikut :

 - a. Menghitung probabilitas awal (*prior probability*) peluang kelas Y P(Y)
 - b. Menghitung nilai probabilitas X_i bersyarat Y_j $P(X_i|Y_j)$
 - c. Menghitung *posterior probability*
 - d. Menentukan hasil klasifikasi, pada Klasifikasi Naive Bayes hasil klasifikasi ditentukan dengan melihat nilai $P(Y_j|X)$ paling besar dari setiap variabel Y dan data yang digunakan adalah data testing.
 - e. Menghitung nilai APER dan akurasi dari model yang terbentuk.
5. Membandingkan nilai error rate dari masing-masing metode, untuk mengetahui metode yang lebih baik dapat dilihat dari rata-rata akurasi yang lebih tinggi.
6. Menguji kesamaan dua rata-rata kedua metode tersebut.

IV. HASIL DAN PEMBAHASAN

a. Deskripsi Data

Setelah dilakukan klasifikasi awal dengan indikator pada tabel.3 maka diperoleh bahwa jumlah alumni yang mendapatkan pekerjaan > 12 bulan sebanyak 169 alumni, Mendapatkan pekerjaan dalam waktu > 6 bulan -12 bulan setelah wisuda sebanyak 228 alumni dan Mendapatkan pekerjaan dalam ≤ 6 setelah wisuda sebanyak 709 alumni. Dengan tabulasi variabel X terhadap tabel Y pada Tabel.4

Tabel.4. Tabulasi Variabel Prediktor Terhadap Variabel Respon

Variabel Prediktor (X)		Variabel Respon (Y)		
		TL	KL	L
Program Studi (X ₁)	X ₁₌₁ Eksak	58	73	245
	X ₁₌₂ Non Eksak	111	155	464
Jenjang Pendidikan (X ₂)	X ₂₌₁ D3	25	18	51
	X ₂₌₂ S1	144	210	658
IPK (X ₃)	X ₃₌₁ ≤ 2.75	5	8	9
	X ₃₌₂ >2.75 s.d ≤ 3.5	107	138	432
	X ₃₌₃ > 3.5	57	82	268
Lama Studi (X ₄)	X ₄₌₁ Lama	34	49	143
	X ₄₌₂ Tepat Waktu	95	132	428
	X ₄₌₃ Cepat	40	47	138
Cara Mencari Pekerjaan (X ₅)	X ₅₌₁ Media	80	101	309
	X ₅₌₂ Mandiri	49	70	238
	X ₅₌₃ Relasi	40	57	162
Penguasaan Pengetahuan Diluar Bidang Studi (X ₆)	X ₆₌₁ STM	2	1	14
	X ₆₌₂ Tidak Menguasai	13	25	62
	X ₆₌₃ Cukup Menguasai	60	85	293
	X ₆₌₄ Menguasai	74	89	268
	X ₆₌₅ Sangat Menguasai	20	28	72

b. Regresi Logistik Ordinal

Salah satu contoh analisis klasifikasi tingkat kelancaran alumni Unila tahun 2016 dalam mencari pekerjaan dengan menggunakan regresi logistik ordinal.

Tabel.5 Output R

Parameter	Estimate	Std. Value	Z Value
1 2	-0.7544	0.6828	-1.105
2 3	0.4244	0.6816	0.623

Berdasarkan hasil output R 'log Lik.' -735.7309 (df=8) dapat dilihat parameter dari masing-masing variabel dengan menggunakan MLE. Selanjutnya akan dilakukan pengujian parameter secara serentak dengan ratio likelihood test menggunakan persamaan (10).

Tabel.6 Uji Ratio Likelihood Test

G	Chi Square ($\chi^2_{(0.05, 8)}$)	Df
19.8779183012	15.507	8

Berdasarkan uji ratio likelihood dapat dilihat bahwa $G (19.8779183012) > \chi^2_{(0.05, 8)} = 15.507$ maka keputusannya tolak H_0 artinya paling tidak terdapat satu variabel prediktor yang mempengaruhi model. Kemudian dilakukan pengujian parameter secara individu sebagai berikut :

Table.7 Uji Wald

Variabel	Value ($\hat{\beta}_i$)	Std. Error $\widehat{SE}(\hat{\beta}_i)$	$\left[\frac{\hat{\beta}_i}{\widehat{SE}(\hat{\beta}_i)}\right]$	$W = \left[\frac{\hat{\beta}_i}{\widehat{SE}(\hat{\beta}_i)}\right]^2$	Keputusan
X ₁	-0.14543	0.15661	-0.929	0.863041	Tidak Tolak H ₀
X ₂	0.62080	0.25246	2.459	6.046681	Tolak H ₀
X ₃	0.29365	0.14715	1.996	3.865153	Tolak H ₀
X ₄	-0.14704	0.11714	-1.255	1.5750	Tidak Tolak H ₀
X ₅	0.06980	0.09036	0.772	0.595987	Tidak Tolak H ₀
X ₆	-0.12881	0.08437	-1.527	2.331729	Tolak H ₀

Dengan menggunakan taraf signifikansi $\alpha = 0.05$ dari tabel Z, maka di peroleh nilai $Z_{w2} = 1.96$ dengan kriteria pengambilan keputusan tolak H_0 Jika $W_i > Z_{w2}$. Berdasarkan tabel diatas maka terdapat tiga variabel yang signifikan terhadap model yaitu X₂, X₃, X₆.

Karena secara parsial hanya tiga parameter variabel yang signifikan, maka dibutuhkan pembentukan model baru dengan menggunakan variabel yang signifikan tersebut, sehingga model logit yang terbentuk sebagai berikut :

$$g_1(x) = \left(\frac{\pi_1(x)}{1-\pi_1(x)} \right) = -0.5897 - 0.50272(x_2) - 0.30652(x_3) + 0.14366(x_6)$$

$$g_2(x) = \left(\frac{\pi_2(x)}{1-\pi_2(x)} \right) = 0.5855 - 0.50272(x_2) - 0.30652(x_3) + 0.14366(x_6)$$

Untuk melakukan penerapan dari model diatas maka dihitung peluang dari masing-masing kelas dengan persamaan (3), (4) dan (5). Untuk ketepatan klasifikasi salah satu data testing menggunakan Regresi logistik Ordinal dapat menggunakan matrik konfusi sebagai berikut :

Tabel.8 Matrik Konfusi Regresi Logistik Ordinal Pertama

Kelas Prediksi	Kelas Observasi		
	Y ₁ = Tidak Lancar	Y ₂ = Kurang Lancar	Y ₃ = Lancar
Y ₁ = Tidak Lancar	0	0	0
Y ₂ = Kurang Lancar	0	0	0
Y ₃ = Lancar	45	54	177

Diperoleh tingkat error dan tingkat akurasi dari Regresi Logistik Ordinal :

$$APPER = \frac{45+54}{276} \times 100\% = 0.3586\% = 0.3586$$

$$\text{Akurasi} = 1 - \text{APPER} = 1 - 0.35507 = 0.6413$$

Selanjutnya dilakukan pengulangan klasifikasi sebanyak sepuluh kali dengan kombinasi data yang berbeda pada testing dan training.

c. Klasifikasi Naive Bayes

Dalam klasifikasi Naive Bayes yang pertama dilakukan menghitung P(Y), P(X_i|Y_j) dan P(X) dengan data training yang sama seperti Regresi Logistik Ordinal, kemudian menghitung P(Y_j|X) dengan data testing, sehingga diperoleh hasil klasifikasi naive Bayes sebagai berikut :

Tabel.9 Matrik Konfusi Naive Bayes Klasifikasi Pertama

Kelas Prediksi	Kelas Observasi		
	Y ₁ = Tidak Lancar	Y ₂ = Kurang Lancar	Y ₃ = Lancar
Y ₁ = Tidak Lancar	4	3	9
Y ₂ = Kurang Lancar	0	0	0
Y ₃ = Lancar	41	51	168

Tingkat error dan Tingkat akurasi dari Klasifikasi Naive Bayes pertama :

$$APPER = \frac{41+51+3+9}{276} \times 100\% = 37.68\% = 0.3768$$

$$\text{Akurasi} = 1 - \text{APPER} = 1 - 0.3768 = 0.6232$$

Selanjutnya dilakukan pengulangan klasifikasi sebanyak sepuluh kali dengan kombinasi data yang berbeda pada testing dan training.

d. Error Rate

Setelah dilakukan sepuluh kali pengulangan analisis pada Regresi Logistik Ordinal dan Naive Bayes (Lampiran .1) menghasilkan data error rate sebagai berikut :

Tabel.10 Perbandingan Tingkat Error Kedua Metode

Analisis Ke-	Regresi Logistik Ordinal	Naive Bayes
1	0.3586	0.3768
2	0.2935	0.2971
3	0.4094	0.434782
4	0.3732	0.39492
5	0.3514	0.3768
6	0.4239	0.45289
7	0.3768	0.39855
8	0.3297	0.35144
9	0.3696	0.37318
10	0.3225	0.3369
Rata-Rata	0.36086	0.37890664

Untuk mengetahui apakah terdapat perbedaan dalam penggunaan kedua metode dalam mengklasifikasi tingkat kelancaran alumni Unila tahun 2016 dalam mencari pekerjaan maka dilakukan Uji Kesamaan dua rata-rata pada tingkat error diatas.

Hipotesis :

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Dengan menggunakan persamaan (17) diperoleh t hitung sebesar -0.20087 dengan derajat bebas 18, berdasarkan tabel $-2.306 < t_{\alpha/2, n+m-2} < 2.306$. Maka dapat diambil keputusan tidak tolak H_0 artinya tidak ada perbedaan dari penggunaan kedua metode dalam mengklasifikasi tingkat kelancaran alumni dalam mencari pekerjaan.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan diatas maka dapat ditarik kesimpulan sebagai berikut

- Klasifikasi tingkat kelancaran alumni Unila tahun 2016 dalam mencari pekerjaan dengan menggunakan Naive Bayes mempunyai rata-rata tingkat error lebih besar yaitu 0.37890664 dibandingkan dengan metode Regresi Logistik Ordinal sebesar 0.36086.
- Hasil dari uji dua rata-rata menunjukkan tidak ada perbedaan secara signifikan penggunaan kedua metode dalam mengklasifikasi tingkat kelancaran alumni unila tahun 2016 dalam mencari pekerjaan,

KEPUSTAKAAN

- [1] Han,J., Kamber, M., Jian,P., 2012. *Data Mining Concepts and Techniques, Third Edition*. California : Morgan Kaufman.
- [2] Agresti, Alan. 2002. *An Introduction Categorical Data Analysis, Second Edition*. New Jersey : John Wiley and Sons Inc.
- [3] Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression, Second Edition*. Canada : John Wiley and Sons Inc.

