PEMBANDINGAN BEBERAPA PENDUGA TINGKAT KESALAHAN KLASIFIKASI PADA ANALISIS DISKRIMINAN KUADRATIK

By Khoirin Nisa

PEMBANDINGAN BEBERAPA PENDUGA TINGKAT KESALAHAN KLASIFIKASI PADA ANALISIS DISKRIMINAN KUADRATIK

(COMPARISON OF SEVERAL CLASSIFICATION ERROR RATE ESTIMATORS ON QUADRATIC DISCRIMINANT ANALYSIS)

Khoirin Nisa
Jurusan Matematika Universitas Lampung
Jl. Prof. Dr. Soemantri Brojonegoro Bandar Lampung 35145
Email : nisa@unila.ac.id

Abstract

The major objective of discriminant analysis is to classify multivariate data into different population on the basis of a training sampel for which the source populations are known. Since the primary goal of discriminant analysis is classifying data, it is important to know the probability of misclassification (which is also called: classification error rate) of the classification rule we use. In this paper we compared four methods for estimating the classification error rate through Monte Carlo simulation, the methods are the Resubstituion method, the Jackknife method, U estimator and $\overline{U}$ estimator. We set the simulation using 1000 random samples with size: n = 20, 40 and 60. The comparison of the predictions of error rate was done using the MSE (mean square error) resulted from all methods. Te result showed that the Jackknife method always performs better than the other three methods.

PENDAHULUAN

Analisis Diskriminan merupakan suatu teknik analisis data multivariat yang digunakan untuk mengklasifikasikan objek ke dalam populasi-populasi yang berbeda berdasarkan sampel latihan (*training sampel*) yang telah diketahui asal populasinya. Berdasarkan sampel tersebut, sebuah aturan pengklasifikasian dibangun dan selanjutnya digunakan untuk mengklasifikasikan objek baru ke dalam salah satu populasi. Aturan pengklasifikasian yang diperoleh merupakan sebuah fungsi yang disebut sebagai fungsi diskriminan.

Pada kasus dua populasi, analisis diskriminan dibedakan menjadi dua jenis, yaitu analisis diskriminan linier dan analisis diskriminan kuadratik. Analisis diskriminan linier digunakan jika matriks kovarian kedua populasi diasumsikan sama, sedangkan analisis diskriminan kuadratik digunakan jika matriks kovarian kedua populasi diasumsikan berbeda. Dalam penelitian yang diusulkan ini kami memfokuskan pada kasus analisis diskriminan dua populasi dengan matriks kovarian berbeda, yaitu analisis diskriminan kuadratik (ADK).

Karena tujuan utama dari analisis diskriminan adalah mengkalsifikasikan data, maka merupakan suatu hal yang sangat penting untuk mengetahui peluang kesalahan klasifikasi (*probability of misclassifications*) atau tingkat kesalahan klasifikasi (*classification error rate*). Berbagai penelitian tentang penaksiran tingkat kesalahan pada analisis diskriminan telah dilakukan selama beberapa dasawarsa. Diantaranya oleh Critchley & Vitieelo (1991), Mangku (1992), Joossens & Croux (2003), Filzmoser dkk (2004), dsb.

Permasalahan yang muncul dalam penaksiran peluang salah klasifikasi adalah ketika data training yang tersedia hanya sedikit, sehingga sampel yang digunakan untuk membangun fungsi diskriminan berukuran relatif kecil terhadap banyaknya variabel. Untuk kasus seperti ini pendugaan tingkat kesalahan klasifikasi secara konvensional tidak dapat memberikan hasil yang akurat. Untuk mengatasi masalah tersebut perlu dilakukan pendugaan dengan pendekatan teknik *resampling*, yaitu pendugaan dilakukan berdasarkan subsampel berulang yang diambil dari sampel yang sama dengan mempertimbangan seluruh kemungkinan subsampel (replikasi) yang dapat dibentuk.

Dalam tulisan ini kami membandingkan beberapa metode penduga tingkat kesalahan klasifikasi pada sampel berukuran kecil. Metode penduga yang akan kami bandingkan yaitu metode Resubstitusi, metode *Jackknife*, metode penduga U, dan metode penduga $\overline{\mathrm{U}}$. Metode Resubsitusi merupakan metode konvensional, sedangkan tiga metode lainnya merupakan metode pendugaan berbasis *resampling* dengan spesifikasi *leave one out* dan validasi silang. Pembandingan keempat metode dilakukan secara empiris dengan simulasi Monte Carlo untuk mengetahui penduga tingkat kesalahan klasifikasi yang terbaik.

Analisis Diskriminan Kuadratik

Misalkan terdapat himpunan n_1 buah pengamatan $X_1=\{x_{11}, x_{12}, \ldots, x_{1n_1}\}$ dalam ruang berdimensi p yang berasal dari populasi pertama $\Phi_1 \sim N_p(\mu_1, \Sigma_1)$ dan himpunan n_2 buah pengamatan $X_2=\{x_{21}, x_{22}, \ldots, x_{2n_2}\}$ dalam ruang berdimensi p yang berasal dari populasi kedua $\Phi_2 \sim N_p(\mu_2, \Sigma_2)$. Maka fungsi diskriminan kuadratik diberikan oleh :

$$Q(x) = \frac{1}{2}\ln\left[\frac{|\Sigma_2|}{|\Sigma_1|}\right] - \frac{1}{2}\left((x-\mu_1)'\Sigma_1^{-1}(x-\mu_1) - (x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)\right) \quad (1)$$

Sebuah pengamatan baru x akan diklasifikasikan ke dalam Φ_1 jika

$$Q(x) > \ln\left[\frac{c_2\pi_2}{c_1\pi_1}\right] = \theta$$

dimana c_1 merupakan biaya yang diakibatkan oleh kesalahan dalam mengklasifikasikan sebuah unit dalam Φ_1 , dan c_2 merupakan biaya yang diakibatkan oleh kesalahan dalam mengklasifikasikan sebuah unit dalam Φ_2, sedangkan π_1 merupakan peluang prior bahwa x berasal dari Φ_1 dan π_2 merupakan peluang prior bahwa x berasal dari Φ_2. Dalam berbagai studi diasumsikan bahwa biaya salah klasifikasi dan peluang prior objek dari kedua kelompok memiliki nilai yang sama, yaitu $c_1 = c_2$ dan $\pi_1 = \pi_2$, sehingga diperoleh $\theta = 0$. Dengan kata lain, keputusannya adalah sebagai berikut : klasifikasikan x ke dalam Φ_1 jika $Q(x) > 0$, klasifikasikan x ke dalam Φ_2 jika $Q(x) < 0$, dan klasifikasikan x secara sembarang ke dalam Φ_1 atau Φ_2 jika $Q(x) = 0$ (Joossens & Croux, 2003).

Tingkat Kesalahan Klasifikasi

Misalkan aturan pengklasifikasian yang digunakan adalah Q(x) seperti pada persamaan (1). Maka tingkat kesalahan klasifikasinya ditentukan oleh rumus berikut:

$$P_1 = P(Q(x) < 0 \text{ jika x berasal dari } \Phi_1 \mid T \text{ tetap})$$

$$P_2 = P(Q(x) \geq 0 \text{ jika x berasal dari } \Phi_2 \mid T \text{ tetap})$$

dengan $T = X_1 \cup X_2$. Dan tingkat kesalahan total dapat dihitung dengan menggunakan rumus :

$$P = \frac{n_1}{n_1+n_2}P_1 + \frac{n_2}{n_1+n_2}P_2 \quad (2)$$

Penduga Resubsitusi

Penduga ini diperkenalkan oleh Smith pada tahun 1997 (Mangku, 2004). Ide dasarnya adalah merealokasi setiap individu dalam sampel training T dengan menggunakan aturan Q(x,T). Misalkan kriteria perhitungannya adalah C(i,j) dengan ketentuan sebagai berikut :

$$C(i,j)=\begin{cases}0, & \text{jika } i=j\\ 1, & \text{jika } i\neq j\end{cases} \qquad \text{untuk setiap } i \text{ dan } j$$

Maka penduga Resubstitusi dapat didefinisikan sebagai :

$$\hat{P}_1(R)=\frac{1}{n_1}\sum_{j=1}^{n_1}C[1,G(x_{1j},T)] \qquad \text{dan} \qquad \hat{P}_2(R)=\frac{1}{n_2}\sum_{j=1}^{n_{21}}C[2,G(x_{2j},T)]$$

dengan $G(x_{1j},T)$ merupakan hasil pengklasifikasian suatu unit dalam Φ_1, sedangkan $G(x_{2j},T)$ merupakan hasil pengklasifikasian suatu unit dalam Φ_2.

Penduga Jackknife

Penduga Jackknife untuk pendugaan tingkat kesalahan diberikan oleh rumus berikut :

$$\hat{P}_1(JK)=\hat{P}_1(R)+(n-1)\left(R_1^+-R_{1(.)}\right)$$

$$\hat{P}_2(JK)=\hat{P}_2(R)+(n-1)\left(R_2^+-R_{2(.)}\right)$$

dengan $\hat{P}_1(R)$ dan $\hat{P}_2(R)$ merupakan penduga Resubstitusi, dan

$$R_1^+=\left(\frac{1}{n_1}\right)^2\sum_{j=1}^{n_1}\sum_{k=1}^{n_1}C[1,G(x_{1k},T_{[1j]})]$$

$$R_2^+=\left(\frac{1}{n_2}\right)^2\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}C[2,G(x_{2k},T_{[2j]})]$$

$$R_{1(.)}=\frac{1}{n_1}\sum_{j=1}^{n_1}\left(\frac{1}{n_1-1}\sum_{k\neq 1}^{n_1}C[1,G(x_{1k},T_{[1j]})]\right)$$

$$R_{2(.)}=\frac{1}{n_2}\sum_{j=1}^{n_2}\left(\frac{1}{n_2-1}\sum_{k\neq 1}^{n_{21}}C[2,G(x_{2k},T_{[2j]})]\right)$$

(Mangku, 2004)

Penduga U

Metode ini didasari pada teknik *leave-one-out* dan diperkenalkan oleh Lachenbruch pada tahun 1967 (Mangku, 2002). Ide dasar metode ini adalah menduga tingkat kesalahan klasifikasi dengan menghilangkan satu objek atau individu dari sampel training T satu persatu secara bergantian, dan pada setiap penghilangan objek tersebut dilakukan pendugaan aturan pengklasifikasian berdasarkan data yang tersisa, selanjutnya aturan ini digunakan untuk mengalokasikan objek yang dikeluarkan. Proses ini diulang sampai setiap objek telah dikeluarkan satu kali dari sampel training. Dugaan tingkat kesalahan dari Penduga U diberikan oleh proporsi dari objek yang salah diklasifikasi terhadap banyaknya objek yang dikeluarkan. Misalkan $T_{[ij]}$ menotasikan sampai training asal yang telah dikurangi oleh x_{ij} (x_{ij} dikeluarkan dari sampel) dengan i = 1, 2, dan j = 1,2,…, n_i, maka penduga U diberikan oleh :

$$\hat{P}_1(U) = \frac{1}{n_1}\sum_{j=1}^{n_1} C[1, G(x_{1j}, T_{[1j]})] \qquad \text{dan} \qquad \hat{P}_2(U) = \frac{1}{n_2}\sum_{j=1}^{n_2} C[2, G(x_{2j}, T_{[2j]})]$$

Penduga $\overline{\mathrm{U}}$

Penduga $\overline{\mathrm{U}}$ juga diperkenalkan oleh Lachenbruch dan Mickey pada tahun 1968 (Mangku, 2004), penduga ini menggabungkan keempirisan dari penduga U dan teori distribusi normal terhadap fungsi diskriminan. Misalkan $\overline{q}_1$ dan s_{q1} adalah nilai tengah dan simpangan baku dari $Q_{11}(x_{11})$, $Q_{12}(x_{12})$, $Q_{13}(x_{13})$, .., $Q_{1n_1}(x_{1n_1})$, dimana $Q_j(x_{ij}) = Q(x_{ij}, T_{[ij]})$. Dan misalkan $\overline{q}_2$ dan s_{q2} adalah nilai tengah dan simpangan baku dari $Q_{21}(x_{21})$, $Q_{22}(x_{22})$, $Q_{23}(x_{23})$, .., $Q_{1n_2}(x_{2n_2})$. Dengan mengasumsikan kenormalan terhadap masing-masing fungsi diskriminan $Q_{1j}(x_{1j})$ dan $Q_{2j}(x_{2j})$, maka tingkat kesalahan dari Penduga $\overline{\mathrm{U}}$ diberikan oleh :

$$\hat{P}_1(\overline{U}) = \psi\left(-\frac{\overline{q}_1}{s_{q1}}\right) \qquad \text{dan} \qquad \hat{P}_2(\overline{U}) = \psi\left(\frac{\overline{q}_2}{s_{q2}}\right)$$

dengan $\psi(.)$ menotasikan fungsi sebaran peluang normal baku.

DATA DAN METODE

Data yang digunakan dalam penelitian ini merupakan data simulasi yang dibangkitkan dengan menggunakan perangkat lunak SAS/IML. Data dua populasi masing-masing dibangkitkan sebanyak $N_1=N_2=1000$ dari distribusi normal multivariat $N_p(\mu, \Sigma)$ dengan p = 3, 4 dan 5. Dalam penelitian ini kami menggunakan 2 kasus berbeda untuk matriks varian kovarian populasi Σ. Pada kasus I, matriks Σ untuk setiap kelompok kami tetapkan berbentuk diagonal yang mengakibatkan variabel-variabel X_i saling bebas. Sedangkan pada kasus II, matriks Σ untuk setiap kelompok kami tetapkan berbentuk nondiagoal (umum) yang mengakibatkan variabel-variabel X_i saling berkorelasi. Selanjutnya dari kedua kelompok populasi diambil sampel berulang, ukuran sampel untuk setiap kelompok ditetapkan sama yaitu $n_1 = n_2$, dalam penelitian ini akan dicobakan n = 20, 40 dan 60 dengan $n = n_1 + n_2$. Pengulangan dilakukan sebanyak 1000 kali untuk setiap ukuran sampel yang digunakan. Tingkat kesalahan optimum (π) ditentukan dengan menggunakan rata-rata peluang salah klasifikasi dari data populasi yang diklasifikasikan dengan menggunakan fungsi diskriminan kuadratik yang dibangun berdasarkan sampel training.

Prosedur Simulasi

1. Membangkitkan data populasi Φ sebanyak N = 2000 yang dibagi ke dalam dua kelompok Φ_1 dan Φ_2 masing-masing berukuran $N_1 = N_2 = 1000$ dengan dimensi p = 5 dari sebaran normal dengan vektor nilai tengah dan matriks kovarian berbeda, yaitu $\Phi_1 \sim N_p(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$ dan $\Phi_2 \sim N_p(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})$.
2. Mengambil sampel training T secara acak dari Φ_1 dan Φ_2 yaitu $X_1 = \{x_{11}, x_{12}, \ldots, x_{1n_1}\}$ dan $X_2 = \{x_{21}, x_{22}, \ldots, x_{2n_2}\}$ masing-masing berukuran 10 sehingga $n = n_1 + n_2 = 20$, dengan $T = X_1 \cup X_2$.
3. Menghitung vektor nilai tengah dan matriks kovarian dari X_1 dan X_2.
4. Menduga fungsi diskriminan quadratik menurut persamaan (1) berdasarkan vektor nilai tengah dan matriks kovarian yang diperoleh pada langkah 3.
5. Mengklasifikasikan data seluruh populasi menggunakan fungsi diskriminan yang diperoleh pada langkah 4, dan menghitung banyaknya objek pada populasi 1 dan populasi 2 yang salah diklasifikasikan (*misclassified*). Selanjutnya menghitung tingkat kesalahan yang dihasilkan dan disimpan sebagai π.
6. Menghitung tingkat kesalahan klasifikasi dari fungsi diskriminan yang diperoleh dengan menggunakan penduga Resubstitusi, Jackknife, U, dan $\overline{\mathrm{U}}$. Simpan tingkat kesalahan keempat penduga berturut-turut sebagai $\hat{P}_i(R)$, $\hat{P}_i(JK)$, $\hat{P}_i(U)$ dan $\hat{P}_i(\overline{U})$.
7. Ulangi langkah 2 sampai 6 sebanyak 1000 kali.
8. Menghitung rata-rata dari π yaitu :

$$\overline{\pi} = \frac{1}{1000}\sum_{i=1}^{1000}\pi_i$$

9. Menghitung nilai Kuadrat Tengah Galat (KTG) dari penduga tingkat kesalahan masing-masing metode dengan rumus sebagai berikut :

$$\text{KTG}\,(\hat{P}_i) = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{P}_i - \overline{\pi}\right)^2$$

10. Ulangi langkah 2 sampai 9 untuk ukuran sampel n=40 dan n=60.

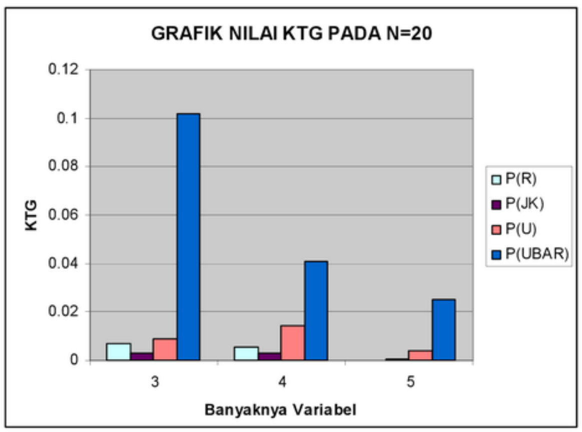
HASIL DAN PEMBAHASAN

Berdasarkan hasil simulasi dengan replikasi sebanyak 1000 kali diperoleh nilai Kuadrat Tengah Galat (KTG) untuk kasus I (variable saling bebas) pada Tabel 1. Berdasarkan nilai KTG pada Tabel 1 terlihat bahwa metode Jckknife selalu memiliki nilai KTG terkecil pada setiap ukuran sampel dan dimensi variabel yang dicobakan. Hal ini menunjukkan bahwa dugaan metode *Jackknife* lebih baik dibandingkan dengan metode lainnya. Dan dapat disimpulkan pula bahwa penambahan variabel dalam fungsi diskriminan memperkecil nilai KTG dugaan peluang setiap metode kecuali nilai dugaan peluang metode $\overline{\mathbf{U}}$. Dugaan peluang metode $\overline{\mathbf{U}}$ terlihat tidak konsisten terhadap penambahan jumlah sampel dan jumlah variabel.

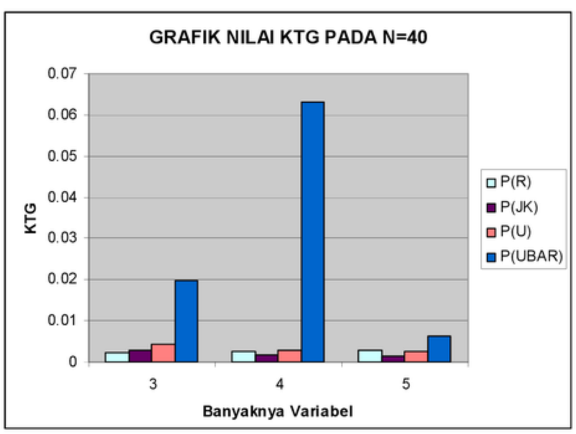
Tabel 1. Nilai Kuadrat Tengah Galat Dugaan Tingkat kesalahan Klasifikasi pada Kasus I

N	P	P(R)	P(JK)	P(U)	P($\overline{U}$)
20	3	0.00704233	0.0030118	0.0086734	0.10180579
	4	0.00517914	0.0031858	0.0141949	0.0409362
	5	0.00012387	0.0003671	0.0038283	0.02490689
40	3	0.00258658	0.002984	0.0042256	0.01971229
	4	0.00222873	0.0016072	0.0028919	0.06297178
	5	0.00298366	0.0013127	0.0027095	0.00622901
60	3	0.00145397	0.0021326	0.0029131	0.00203167
	4	0.00127465	0.0011532	0.0016902	0.16534854
	5	0.0012244	0.0007106	0.0010728	0.05260636

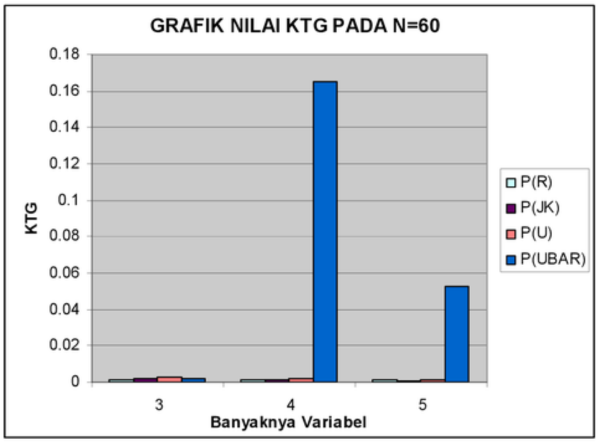
Untuk lebih jelas dalam pembandingan nilai KTG setiap metode, data nilai KTG pada Tabel 1 disajikan dalam beberapa grafik berikut :



Gambar 1. Grafik Nilai KTG pada n=20 untuk kasus I

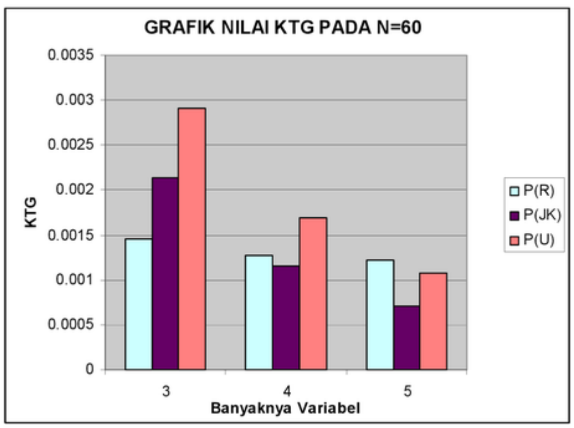


Gambar 2. Grafik Nilai KTG pada n=40 untuk kasus I



Gambar 3a. Grafik Nilai KTG pada n=60 untuk kasus I

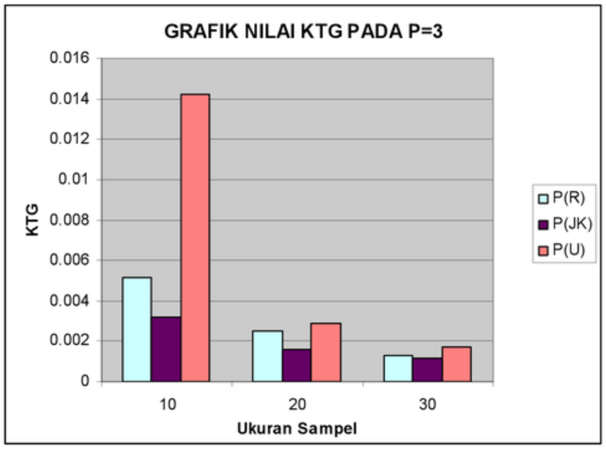
Pada gambar-gambar di atas terlihat bahwa nilai KTG yang dihasilkan metode *Jackknife* selalu lebih kecil dibandingkan metode lainnya pada setiap ukuran sampel. Meskipun pada p=3 nilai KTG metode Resubstitusi lebih kecil dari metode *Jackknife*, namun semakin banyak variabel yang digunakan tidak menurunkan nilai KTG metode Resubstitusi secara signifikan, padahal analisis diskriminan merupakan analisis multivariat yang dalam prakteknya seringkali melibatkan banyak variabel. Dengan tujuan untuk mempermudah pembandingan, grafik nilai KTG pada ukuran sampel n=60 (Gambar 3a) disajikan ulang pada Gambar 3b. berikut tanpa melibatkan nilai KTG dugaan peluang metode $\overline{\mathbf{U}}$.



Gambar 3b. Grafik Nilai KTG pada n=60 untuk kasus I

9

Berdasarkan gambar di atas dapat dilihat bahwa semakin banyak variabel yang digunakan dalam analisis diskriminan, maka penduga *Jackknife* semakin lebih baik dibandingkan metode lainnya. Sedangkan untuk pembandingan nilai KTG berdasarkan jumlah variabel juga dapat terlihat bahwa metode *Jackknife* lebih baik. Sebagai contoh, berikut ini disajikan grafik pembandingan nilai KTG setiap metode pada jumlah variabel p=3.



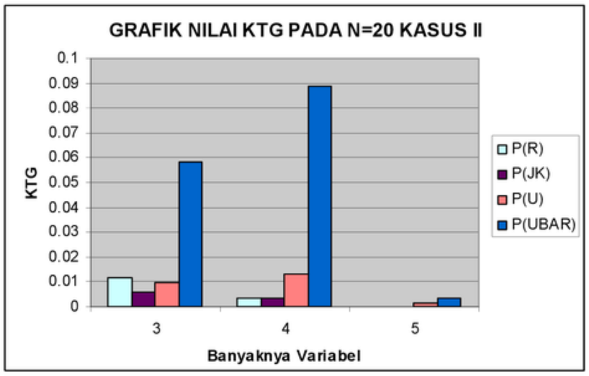
Gambar 4. Grafik Nilai KTG pada p=3 untuk kasus I

Untuk kasus ke dua (variabel saling berkorelasi) diperoleh nilai KTG dugaan peluang setiap metode sebagai berikut :

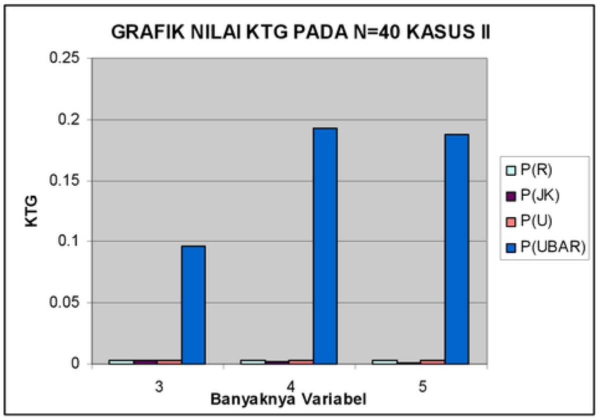
Tabel 2. Nilai Kuadrat Tengah Galat Dugaan Peluang pada Kasus II

N	P	P(R)	P(JK)	P(U)	P($\overline{U}$)
20	3	0.011561	0.005587	0.009538	0.058406
	4	0.003449	0.003628	0.013258	0.088785
	5	2.38E-05	0.000129	0.001233	0.003369
40	3	0.003473	0.003483	0.002744	0.096312
	4	0.003188	0.002012	0.002756	0.192393
	5	0.003236	0.001262	0.002663	0.187276
60	3	0.002134	0.003051	0.001552	0.212811
	4	0.001601	0.001492	0.001493	0.163298
	5	0.001496	0.00081	0.001228	0.104493

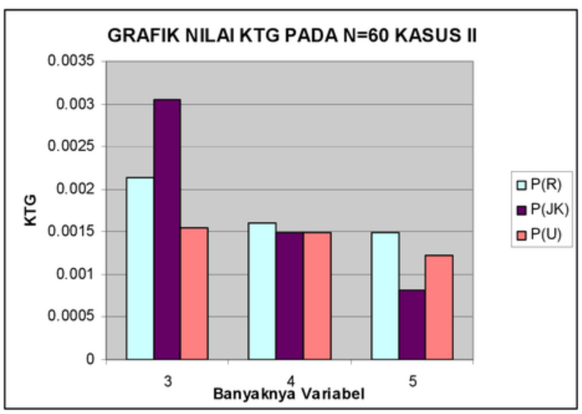
Pada Tabel 2 di atas juga terlihat bahwa metode *Jackknife* selalu memiliki nilai KTG terkecil pada setiap ukuran sampel dan dimensi variabel yang dicobakan. Hal ini menunjukkan bahwa dugaan metode *Jackknife* lebih baik dibandingkan dengan metode lainnya. Dan seperti pda kasus I, pada kasusu II ini dapat disimpulkan pula bahwa penambahan variabel dalam fungsi diskriminan memperkecil nilai KTG dugaan peluang setiap metode kecuali nilai dugaan peluang metode $\overline{U}$. Dugaan peluang metode $\overline{U}$ terlihat tidak konsisten terhadap penambahan jumlah sampel dan jumlah variabel. Untuk lebih jelas dalam pembandingan nilai KTG setiap metode, data nilai KTG pada Tabel 2 disajikan dalam beberapa grafik berikut :



Gambar 5. Grafik Nilai KTG pada N=20 untuk kasus II



Gambar 6. Grafik Nilai KTG pada N=40 untuk kasus II



Gambar 7. Grafik Nilai KTG pada N=60 untuk kasus II

Pada Gambar 5 – Gambar 7 di atas terlihat bahwa nilai KTG yang dihasilkan metode *Jackknife* selalu lebih kecil dibandingkan metode lainnya kecuali pada ukuran sampel n=60 dan p=3, namun semakin banyak variabel yang digunakan maka semakin kecil nilaiKTG metode *Jackknife* sebagaimana pada kasus I. Dengan demikian maka metode *Jackknife* memberikan nilai dugaan peluang salah klasifikisasi yang lebih baik dibandingkan metode lainnya.

SIMPULAN

Berdasarkan uraian di atas maka dapat diambil kesimpulan bahwa metode Jackknife merupakan metode penduga terbaik dibandingkan metode lainnya, baik untuk kasus I (variabel saling bebas) maupun kasus II (variabel saling berkorelasi). Dan dapat disimpulkan juga bahwa semakin banyak variabel yang digunakan akan semakin menurunkan tingkat kesalahan klasifikasi atau *error rate* dari *seluruh* metode.

UCAPAN TERIMAKASIH

Terimakasih penulis ucapkan kepada Lembaga Penelitian UNILA yang telah mendanai penelitian ini melalui DIPA PNBP tahun 2008.

DAFTAR PUSTAKA

Critchley & Vitieelo, C. 1991. The influence of Observations on Misclassification Probability Estimate in Linear Discriminant Analysis. *Biometrica*, Vol. 78.

Filzmoser, P., Joossens, K. & Croux, C. 2004. Multiple Group Linear Discriminant Analyisis : Robustness and Error rate. *The Canadian Journal of Statistics*. Vol 29. hal 473-492.

Joossens, K. dan Croux, C. 2003. *Empirical comparison of the classification performance of robust linear and quadratic discriminant analysis*, manuscript, http: // www. Econ. kuleuven. ac. be/ k. joossens

Mangku, I. W. 1992. *Error Rate Estimation in Discriminant Analysis : another Look at Bootstrap and Other Empirical Technique*. Tesis master tidak dipublikasikan, Curtin University of Technology, Perth Australia.

Mangku, I.W. 2002. Discriminant Function and their Misclassification Error. *Jurnal Matematika dan Aplikasinya*. Vol 1, No.2.

Mangku, I.W. 2004. Estimating the Probability of Misclassification in Two-Groups Discriminant Analysis. *Jurnal Matematika dan Aplikasinya*. Vol 3, No.1.

PEMBANDINGAN BEBERAPA PENDUGA TINGKAT KESALAHAN KLASIFIKASI PADA ANALISIS DISKRIMINAN KUADRATIK

ORIGINALITY REPORT

7%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.econ.kuleuven.ac.be Internet	46 words — 2%
2	"Theory and Applications of Recent Robust Methods", Springer Science and Business Media LLC, 2004 Crossref	28 words — 1%
3	vdocuments.mx Internet	27 words — 1%
4	es.scribd.com Internet	17 words — 1%
5	id.123dok.com Internet	13 words — < 1%
6	www.dbs.informatik.uni-muenchen.de Internet	12 words — < 1%
7	Herlinda Rama Danti, Yuyun Fitriana, Agus Muhammad Hariri, Purnomo Purnomo. "KEANEKARAGAMAN ARTHROPODA PADA PERTANAMAN TOMAT DENGAN SISTEM PERTANAMAN BERBEDA DI KABUPATEN TANGGAMUS, LAMPUNG", Jurnal Agrotek Tropika, 2018 Crossref	11 words — < 1%
8	nicofergiyono.blogspot.com Internet	9 words — < 1%

9	id.scribd.com Internet	9 words — < 1%
10	www.antikatastasikoufomaton.gr Internet	9 words — < 1%
11	repository.ung.ac.id Internet	8 words — < 1%
12	Elis KARTIKA, LIZAWATI LIZAWATI, Hamzah HAMZAH. "Respons Tanaman Jarak Pagar Terhadap Mikoriza Indigenous dan Pupuk P di Lahan Bekas Tambang Batu Bara", Biospecies, 2018 Crossref	8 words — < 1%
13	Lars Diening. "Other Spaces of Differentiable Functions", Lecture Notes in Mathematics, 2011 Crossref	8 words — < 1%
14	sediaobatdiabetes.blogspot.com Internet	8 words — < 1%
15	econ.kuleuven.be Internet	8 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		