

KLASIFIKASI INDEKS PRESTASI KUMULATIF MAHASISWA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS ANDALAS MENGUNAKAN ANALISIS DISKRIMINAN BERGANDA

MITA OKTAVIANI, YUDIANTRI ASDI*, FERRA YANUAR

*Departemen Matematika dan Sains Data,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
email : oktavianimita123@gmail.com, yudiantriasdi@sci.unand.ac.id,
ferrayanuar@sci.unand.ac.id*

Diterima 7 Februari 2022 Direvisi 23 Februari 2022 Dipublikasikan 31 Juli 2022

Abstrak. Faktor yang mempengaruhi Indeks Prestasi Kumulatif (IPK) mahasiswa terbagi menjadi dua, yaitu faktor internal dan faktor eksternal. Faktor-faktor yang diduga mempengaruhi IPK, yaitu jalur masuk perguruan tinggi, usia, nilai rata-rata ujian nasional, besar biaya hidup, lama belajar, dan besar penggunaan internet. Dari faktor-faktor tersebut akan dilihat faktor yang secara signifikan mempengaruhi pengelompokan IPK mahasiswa pada masing-masing jalur masuk perguruan tinggi dan akan dilakukan prediksi pengelompokan IPK menggunakan metode analisis diskriminan berganda. Data yang digunakan adalah data primer yang diperoleh melalui penyebaran kuesioner *online* kepada mahasiswa FMIPA UNAND angkatan 2017–2019 dengan ukuran sampel sebesar 303 objek. Hasil penelitian menunjukkan bahwa tingkat keakuratan klasifikasi pada masing-masing jalur masuk perguruan tinggi lebih dari 50%, sehingga fungsi diskriminan yang diperoleh dianggap akurat dan fungsi dapat digunakan untuk mengelompokkan mahasiswa ke dalam kelompok IPK yang telah ditentukan.

Kata Kunci: Indeks Prestasi Kumulatif, Analisis Diskriminan Berganda

1. Pendahuluan

Indeks Prestasi Kumulatif (IPK) adalah akumulasi rata-rata nilai dari seluruh mata kuliah yang diikuti sejak semester pertama hingga saat dilakukan perhitungan atau evaluasi [4]. Menurut Purwanto [5], faktor-faktor yang mempengaruhi IPK terbagi menjadi dua, yaitu faktor internal yang berasal dari dalam diri dan faktor eksternal yang berasal dari luar diri. Faktor-faktor yang diduga mempengaruhi IPK yaitu jalur masuk perguruan tinggi, usia, nilai rata-rata UN, besar biaya hidup, lama belajar, dan besar penggunaan internet.

*Penulis Korespondensi

Klasifikasi merupakan teknik multivariat yang berkaitan dengan pemisahan sekumpulan objek (pengamatan) berbeda menggunakan alokasi objek baru ke dalam kelompok yang telah ditentukan sebelumnya. Salah satu metode statistika yang dapat digunakan untuk mengklasifikasikan objek yaitu analisis diskriminan.

Analisis diskriminan (*discriminant analysis*) merupakan metode statistika yang digunakan untuk mengklasifikasikan sejumlah objek berdasarkan kasus-kasus pada variabel bebas ke dalam kelompok atau kategori pada variabel terikat, dimana setiap objek menjadi anggota dari salah satu kelompok tetapi tidak ada objek yang menjadi anggota pada lebih dari satu kelompok. Pengklasifikasian kasus-kasus dapat berupa dua atau lebih kelompok [1].

2. Landasan Teori

2.1. Matriks Ragam Peragam

Matriks ragam peragam $\mathbf{S}$ adalah matriks berukuran $p \times p$ yang dapat ditulis sebagai:

$$\mathbf{S} = (s_{ik}) = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad i, k = 1, 2, \dots, p$$

dengan elemen [6]:

$$s_{ik} = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{x}_{ji} - \bar{\mathbf{x}}_i) (\mathbf{x}_{jk} - \bar{\mathbf{x}}_k). \quad (2.1)$$

Selanjutnya didefinisikan $\mathbf{S}_{gab}$ sebagai berikut

$$\mathbf{S}_{gab} = \frac{1}{\sum_{i=1}^g (n_i - 1)} \sum_{i=1}^g (n_i - 1) \mathbf{S}_i, \quad (2.2)$$

dimana n_i adalah banyaknya pengamatan pada kelompok ke- i dan $\mathbf{S}_i$ adalah matriks ragam peragam contoh pada kelompok ke- i [2].

2.2. Validasi Silang (Cross Validation)

Validasi silang adalah metode statistik untuk mengevaluasi dan membandingkan data dengan membagi data menjadi dua segmen, satu digunakan untuk mempelajari atau membentuk model dan yang lainnya digunakan untuk memvalidasi model. Dengan prosedur ini, data dipartisi menjadi dua bagian, bagian pertama sebagai data *training* dan bagian kedua sebagai data *testing*. Data *training* digunakan untuk membentuk model, sedangkan data *testing* digunakan untuk validasi model [1].

2.3. Analisis Diskriminan

Analisis diskriminan (*discriminant analysis*) adalah salah satu metode yang digunakan dalam analisis multivariat dengan metode dependensi, yaitu hubungan

antara variabel terikat dan variabel bebas sudah bisa dibedakan. Fungsi analisis diskriminan merupakan kombinasi linier yang memiliki bentuk sebagai berikut [1].

$$y_j = b_{0j} + b_{1j}x_1 + b_{2j}x_2 + \cdots + b_{pj}x_p \quad (2.3)$$

Keterangan:

y_j : nilai (skor) diskriminan fungsi ke- j , $j = 1, 2, \dots, s$,

x_i : variabel bebas ke- i , $i = 1, 2, \dots, p$,

b_{ij} : koefisien diskriminan variabel bebas ke- i , ke- j .

Untuk menentukan vektor koefisien fungsi diskriminan $\mathbf{b}$ digunakan Persamaan (2.4):

$$(\mathbf{B} - \lambda \mathbf{W}) \mathbf{b} = \mathbf{0}, \quad (2.4)$$

agar vektor $\mathbf{b} \neq \mathbf{0}$ merupakan solusi dari (2.4), maka haruslah

$$\det(\mathbf{W}^{-1}\mathbf{B} - \lambda \mathbf{I}) = 0. \quad (2.5)$$

Nilai eigen λ disubstitusikan ke persamaan (2.6) untuk memperoleh vektor koefisien fungsi diskriminan $\mathbf{b}$.

$$(\mathbf{W}^{-1}\mathbf{B} - \lambda \mathbf{I}) \mathbf{b} = \mathbf{0}. \quad (2.6)$$

Diketahui bahwa vektor $\mathbf{y} = [y_1, y_2, \dots, y_s]$ merupakan fungsi diskriminan yang terbentuk dan $s = \min\{g - 1, p\}$ menyatakan banyaknya fungsi diskriminan yang terbentuk. Klasifikasikan $\mathbf{x}$ ke dalam kelompok π_k jika [2]:

$$\sum_{j=1}^r (y_j - \bar{y}_{kj})^2 = \sum_{j=1}^r [\mathbf{b}_j^T (x - \bar{x}_k)]^2$$

merupakan nilai minimum, dimana:

$$\bar{y}_{kj} = \mathbf{b}_j^T \bar{x}_k, \quad j = 1, 2, \dots, s; \quad k = 1, 2, \dots, g.$$

Keterangan

y_j : skor diskriminan fungsi ke- j pengamatan,

$\bar{y}_{kj}$: rata-rata (*centroid*) fungsi ke- j pada kelompok ke- k ,

$\mathbf{b}_j$: vektor koefisien fungsi diskriminan ke- j .

2.4. Keakuratan Klasifikasi

Klasifikasi data dilakukan berdasarkan fungsi diskriminan yang terbentuk, dimana pengklasifikasian dilakukan dengan membandingkan data penelitian dan hasil klasifikasi. Jika banyaknya data untuk masing-masing kelompok yang diprediksi benar, yaitu $(n_{11} + n_{22} + \cdots + n_{gg})$, maka nilai keakuratan klasifikasi yaitu

$$Hit_{Ratio} = \frac{n_{11} + n_{22} + \cdots + n_{gg}}{n} \times 100\%, \quad (2.7)$$

dimana n adalah banyaknya data seluruhnya [2].

2.5. Uji Normal Multivariat

Uji normal multivariat bertujuan untuk menguji apakah model analisis diskriminan ini berdistribusi normal multivariat atau tidak. Hipotesis uji adalah:

$$\begin{aligned} H_0 &: \text{data berdistribusi normal multivariat,} \\ H_1 &: \text{data tidak berdistribusi normal multivariat.} \end{aligned}$$

Statistik uji yang digunakan adalah

$$d^2_{(j)} = (\mathbf{x}_j - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}); j = 1, 2, \dots, n, \quad (2.8)$$

dan

$$q_{\frac{j-0.5}{n}(p)} = \chi^2_{\frac{n-j+0.5}{n}(p)}, j = 1, 2, \dots, n. \quad (2.9)$$

Apabila terdapat kurang dari 50% jarak $d^2_{(j)} \leq q_{(j)}$, maka tolak H_0 . Atau data diasumsikan berdistribusi normal multivariat jika plot antara $d^2_{(j)}$ dan $q_{(j)}$ membentuk garis lurus [2].

2.6. Uji Homogenitas

Uji Homogenitas atau uji kesamaan matriks ragam peragam merupakan uji yang digunakan untuk melihat apakah dua atau lebih kelompok memiliki matriks ragam peragam yang sama atau berbeda. Hipotesis dari pengujian ini adalah:

$$\begin{aligned} H_0 &: \Sigma_1 = \Sigma_2 = \Sigma_3 = \dots = \Sigma_g = \Sigma, \\ H_1 &: \text{paling tidak terdapat } i \text{ dan } j, \text{ dengan } i \neq j, \text{ sedemikian sehingga } \Sigma_i \neq \Sigma_j. \end{aligned}$$

Statistik uji yang digunakan adalah

$$C = (1 - u) M, \quad (2.10)$$

dimana

$$u = \left[\sum_{i=1}^g \frac{1}{(n_i - 1)} - \frac{1}{\sum_{i=1}^g (n_i - 1)} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right], \quad (2.11)$$

$$M = \left[\sum_{i=1}^g (n_i - 1) \right] \ln |\mathbf{S}_{gab}| - \sum_{i=1}^g [(n_i - 1) \ln |\mathbf{S}_i|], \quad (2.12)$$

dengan derajat bebas

$$v = \frac{1}{2} p(p+1)(g-1). \quad (2.13)$$

Apabila pada taraf signifikansi α diperoleh $C > \chi^2_{\alpha(v)}$, maka tolak H_0 [2].

2.7. Uji Kesamaan Vektor Rata-rata

Pengujian kesamaan vektor rata-rata ini digunakan untuk menguji apakah terdapat perbedaan yang signifikan setiap variabel bebas antar kelompok atau tidak. Hipotesis dari pengujian ini adalah:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_g,$$

$$H_1 : \text{paling tidak terdapat } i \text{ dan } j, \text{ dengan } i \neq j, \text{ sedemikian sehingga } \mu_i \neq \mu_j, \\ \text{dimana } i, j = 1, 2, \dots, g.$$

Statistik uji yang digunakan adalah

$$V = - \left[(n-1) - \frac{p+g}{2} \right] \ln \Lambda. \quad (2.14)$$

Apabila pada taraf signifikansi α diperoleh $V > \chi^2_{1-\alpha(v)}$ untuk $v = p(g-1)$, maka tolak H_0 . Jika menggunakan SPSS, kriteria uji yang digunakan yaitu jika $p\text{-value} < \alpha$ maka tolak H_0 [3].

3. Metode Penelitian

Data yang digunakan pada penelitian ini adalah data primer yang diperoleh melalui penyebaran kuisioner *online* kepada mahasiswa S1 FMIPA UNAND angkatan 2017-2019. Variabel yang digunakan dalam penelitian ini adalah kelompok IPK (y), usia (x_1), rata-rata nilai ujian nasional (x_2), biaya hidup (x_3), lama waktu belajar (x_4), dan besar penggunaan internet (x_5). Langkah-langkah analisis data yang dilakukan dalam penelitian ini adalah:

- (1) melakukan analisis deskriptif terhadap data yang digunakan pada penelitian ini;
- (2) membagi data asli menjadi data *training* dan data *testing* menggunakan metode data *splitting*;
- (3) melakukan uji distribusi normal multivariat terhadap data;
- (4) melakukan uji kesamaan matriks ragam beragam;
- (5) melakukan uji kesamaan vektor rata-rata;
- (6) membentuk fungsi diskriminan;
- (7) melakukan prediksi klasifikasi berdasarkan fungsi diskriminan; dan
- (8) menghitung tingkat keakuratan klasifikasi.

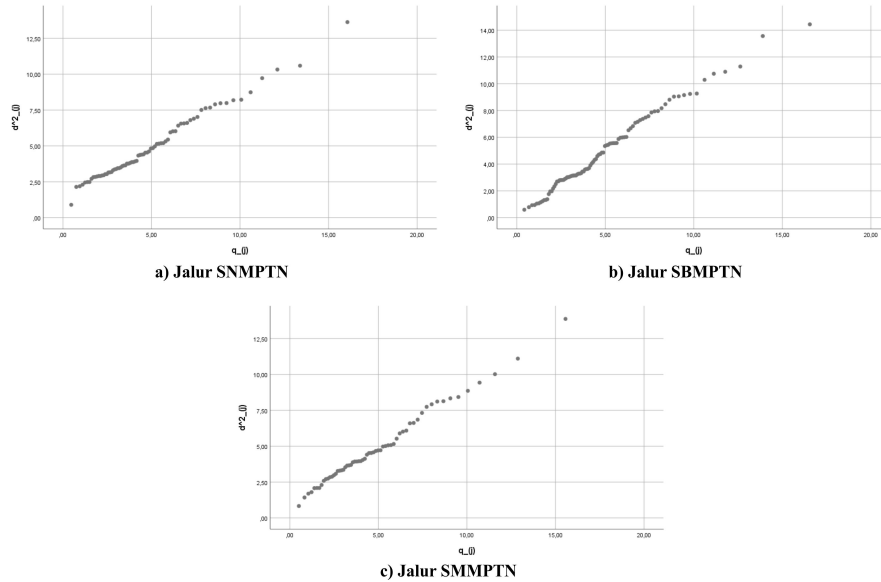
4. Hasil dan Pembahasan

4.1. Analisis Diskriminan

4.1.1. Uji Normalitas Multivariat

Diperoleh plot antara kuadrat jarak Mahalanobis (d_j^2) dan nilai kuantil normal standar (q_j) untuk jalur SNMPTN, SBMPTN dan SMMPTN seperti pada Gambar 1.

Dengan menggunakan plot Khi Kuadrat, diperoleh plot membentuk garis lurus untuk jalur SNMPTN, SBMPTN dan SMMPTN maka disimpulkan bahwa tidak



Gambar 1. Plot Khi Kuadrat

tolak H_0 . Pada taraf signifikansi $\alpha = 0,05$ data mahasiswa yang masuk di ketiga jalur berdistribusi normal multivariat.

4.1.2. Uji Homogenitas

(a) Pada jalur SNMPTN, diperoleh nilai $M = 64,2088$ dan $u = 0,1394$ maka:

$$C = (1 - u)M = 55,2589.$$

(b) Pada jalur SBMPTN, diperoleh nilai $M = 65,7744$ dan $u = 0,1137$ maka:

$$C = (1 - u)M = 58,2962.$$

(c) Pada jalur SMMPTN, diperoleh nilai $M = 72,5805$ dan $u = 0,1741$ maka:

$$C = (1 - u)M = 59,9447.$$

Karena $C \leq \chi^2_{0,05(45)}$, maka tidak tolak H_0 . Pada taraf signifikansi $\alpha = 0,05$ diketahui bahwa matriks ragam peragam antar keempat kelompok IPK tersebut sama untuk ketiga jalur masuk.

4.1.3. Uji Kesamaan Vektor Rata-rata

Dengan menggunakan SPSS, diperoleh tiga variabel yang berbeda secara signifikan untuk keempat kelompok pada jalur SNMPTN, yaitu rata-rata UN (x_2), biaya hidup (x_3), dan lama belajar (x_4) karena nilai $p\text{-value} < 0,05$. Pada jalur SBMPTN, diperoleh dua variabel yaitu biaya hidup (x_3) dan lama belajar (x_4). Sedangkan pada jalur SMMPTN, diperoleh tiga variabel yaitu rata-rata UN (x_2), biaya hidup (x_3), dan lama belajar (x_4).

4.1.4. Fungsi Diskriminan

(a) Pada jalur SNMPTN, diperoleh tiga fungsi diskriminan yaitu:

$$\begin{aligned}y_1 &= 0,701 - 0,042x_2 + 1,622x_3 - 0,189x_4, \\y_2 &= -11,037 + 0,111x_2 + 0,970x_3 + 0,071x_4, \\y_3 &= 0,501 - 0,065x_2 + 0,772x_3 + 0,358x_4.\end{aligned}$$

(b) Pada jalur SBMPTN, diperoleh dua fungsi diskriminan yaitu:

$$\begin{aligned}y_1 &= -1.150 + 1.578x_3 - 0.288x_4, \\y_2 &= -4.509 + 1.182x_3 + 0.279x_4.\end{aligned}$$

(c) Pada jalur SMMPTN, diperoleh tiga fungsi diskriminan yaitu:

$$\begin{aligned}y_1 &= -3.315 + 0.046x_2 - 1.222x_3 + 0.308x_4, \\y_2 &= -9.615 + 0.132x_2 + 0.251x_3 - 0.173x_4, \\y_3 &= -7.876 + 0.043x_2 + 1.326x_3 + 0.238x_4.\end{aligned}$$

4.1.5. Hasil Prediksi Klasifikasi

Hasil prediksi klasifikasi pada data *training* untuk ketiga jalur masuk dapat dilihat pada Tabel 1, Tabel 2 dan Tabel 3.

Tabel 1. Hasil Prediksi Klasifikasi Data *Training* Jalur SNMPTN

Kelompok Asli	Hasil Prediksi Kelompok				Total
	Dengan Pujian	Sangat Memuaskan	Memuaskan	Cukup Memuaskan	
Dengan Pujian	12	3	0	0	15
Sangat Memuaskan	7	18	1	3	29
Memuaskan	0	4	10	4	18
Cukup Memuaskan	0	4	6	3	13
	19	29	17	10	75

Tabel 2. Hasil Prediksi Klasifikasi Data *Training* Jalur SBMPTN

Kelompok Asli	Hasil Prediksi Kelompok				Total
	Dengan Pujian	Sangat Memuaskan	Memuaskan	Cukup Memuaskan	
Dengan Pujian	22	9	1	0	32
Sangat Memuaskan	5	16	5	2	28
Memuaskan	0	5	7	5	17
Cukup Memuaskan	0	2	4	9	15
	27	32	17	16	92

Tabel 3. Hasil Prediksi Klasifikasi Data *Training* Jalur SMMPTN

Kelompok Asli	Hasil Prediksi Kelompok				Total
	Dengan Pujian	Sangat Memuaskan	Memuaskan	Cukup Memuaskan	
Dengan Pujian	11	4	0	0	15
Sangat Memuaskan	5	14	2	2	23
Memuaskan	0	2	8	3	13
Cukup Memuaskan	0	1	3	6	10
	16	21	13	11	61

Selanjutnya hasil prediksi klasifikasi data *testing* dapat dilihat pada Tabel 4, Tabel 5 dan Tabel 6.

Tabel 4. Hasil Prediksi Klasifikasi Data *Testing* Jalur SNMPTN

Kelompok Asli	Hasil Prediksi Kelompok				Total
	Dengan Pujian	Sangat Memuaskan	Memuaskan	Cukup Memuaskan	
Dengan Pujian	5	0	0	0	5
Sangat Memuaskan	4	6	0	0	10
Memuaskan	0	2	3	1	6
Cukup Memuaskan	0	2	0	2	4
	9	10	3	3	25

Tabel 5. Hasil Prediksi Klasifikasi Data *Testing* Jalur SBMPTN

Kelompok Asli	Hasil Prediksi Kelompok				Total
	Dengan Pujian	Sangat Memuaskan	Memuaskan	Cukup Memuaskan	
Dengan Pujian	8	2	0	0	10
Sangat Memuaskan	1	5	3	0	9
Memuaskan	0	1	1	3	5
Cukup Memuaskan	0	1	1	3	5
	9	9	5	6	29

Tabel 6. Hasil Prediksi Klasifikasi Data *Testing* Jalur SMMPTN

Kelompok Asli	Hasil Prediksi Kelompok				Total
	Dengan Pujian	Sangat Memuaskan	Memuaskan	Cukup Memuaskan	
Dengan Pujian	3	1	0	0	4
Sangat Memuaskan	2	4	1	1	8
Memuaskan	0	1	1	3	5
Cukup Memuaskan	0	0	1	3	4
	5	6	3	7	21

4.1.6. Keakuratan Klasifikasi

Tingkat keakuratan klasifikasi pada data *training* yaitu sebesar 57,33% untuk jalur SNMPTN, 58,70% untuk jalur SBMPTN dan 63,93% untuk jalur SMMPTN. Sedangkan pada data *testing* yaitu sebesar 64,00% untuk jalur SNMPTN, 58,62% untuk jalur SBMPTN dan 52,38% untuk jalur SMMPTN.

5. Kesimpulan

Berdasarkan analisis yang telah dilakukan pada penelitian ini, dapat diperoleh kesimpulan sebagai berikut.

- (1) Variabel yang mempengaruhi perbedaan antara masing-masing kelompok IPK yaitu
 - Pada jalur SNMPTN, variabel yang berpengaruh adalah rata-rata UN (x_2), biaya hidup (x_3), dan lama belajar (x_4).
 - Pada jalur SBMPTN, variabel yang berpengaruh adalah biaya hidup (x_3) dan lama belajar (x_4).
 - Pada jalur SMMPTN, variabel yang berpengaruh adalah rata-rata UN (x_2), biaya hidup (x_3), dan lama belajar (x_4).

(2) Fungsi diskriminan yang terbentuk adalah sebagai berikut.

– Jalur SNMPTN

$$y_1 = 0.701 - 0.042X_2 + 1.622X_3 - 0.189X_4$$

$$y_2 = -11.037 + 0.111X_2 + 0.970X_3 + 0.071X_4$$

$$y_3 = 0.501 - 0.065X_2 + 0.772X_3 + 0.358X_4$$

– Jalur SBMPTN

$$y_1 = -1.150 + 1.578X_3 - 0.288X_4$$

$$y_2 = -4.509 + 1.182X_3 + 0.279X_4$$

– Jalur SMMPTN

$$y_1 = -3.315 + 0.046X_2 - 1.222X_3 + 0.308X_4$$

$$y_2 = -9.615 + 0.132X_2 + 0.251X_3 - 0.173X_4$$

$$y_3 = -7.876 + 0.043X_2 + 1.326X_3 + 0.238X_4$$

(3) Keakuratan klasifikasi IPK mahasiswa sebagai berikut.

– Jalur SNMPTN

Keakuratan klasifikasi pada data *training* yaitu sebesar 57,33%, dan pada data *testing* yaitu sebesar 64,00%.

– Jalur SBMPTN

Keakuratan klasifikasi pada data *training* yaitu sebesar 58,70%, dan pada data *testing* yaitu sebesar 58,62%.

– Jalur SMMPTN

Keakuratan klasifikasi pada data *training* yaitu sebesar 63,93%, dan pada data *testing* yaitu sebesar 52,38%.

Karena keakuratan klasifikasi pada ketiga jalur masuk perguruan tinggi di atas 50%, maka ketepatan fungsi dianggap tinggi dan fungsi dapat digunakan untuk mengklasifikasi mahasiswa ke dalam kelompok IPK yang telah ditentukan.

6. Ucapan Terima kasih

Penulis mengucapkan terima kasih kepada Ibu Maiyastri, Ibu Lyra Yulianti, dan Bapak Dr. Ahmad Iqbal Baqi yang telah memberikan masukan dan saran sehingga artikel ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Hair, J.F., W.C. Black, B.J. Babin dan R.E. Anderson, 2009, *Multivariate Data Analysis 7th Edition*, Prentice Hall, USA
- [2] Johnson, R.A. dan D.D. Wichern, 2007, *Applied Multivariate Statistical Analysis (6th edition)*, Pearson Education Inc, USA
- [3] Nugroho, S., 2008, *Statistik Multivariat Terapan*, UNIB Press, Bengkulu
- [4] Peraturan Rektor Universitas Andalas Nomor 14 Tahun 2020 Tentang Peraturan Akademik Program Sarjana Universitas Andalas
- [5] Purwanto, N., 2002, *Psikologi Pendidikan*, PT Remaja Rosdakarya, Jakarta
- [6] Rencher, A.C. 2002. *Methods of Multivariate Analysis (2nd edition)*, A John Wiley and Sons, Inc. Publication, USA