

**ANALISIS DISKRIMINAN BERGANDA DENGAN PEUBAH BEBAS
CAMPURAN KATEGORIK DAN KONTINU PADA KLASIFIKASI
INDEKS PRESTASI KUMULATIF MAHASISWA**

Nur Walidaini¹, Moch. Abdul Mukid², Alan Prahutama³, Agus Rusgiyono⁴
^{1,2,3,4}Departemen Statistika Universitas Diponegoro

e-mail: mamukid@yahoo.com

DOI: 10.14710/medstat.10.2.71-83

Article Info:

Received: 11 November 2017

Accepted: 28 Desember 2017

Available Online: 30 Desember 2017

Keywords:

*Multiple Discriminant Analysis,
Mixture Variable, Grade Point
Average*

Abstract: Multiple discriminant analysis is one of the discriminant analysis techniques where the dependent variable are grouped into more than two groups. This paper discussed how to categorize Grade Point Average (GPA) of undergraduate student of Faculty of Sciences and Mathematics Diponegoro University based on categorical and continuous independent variable including gender, internet usage, time per week for learning, average score in national examination, amount of pocket money per month and the way to enter to Diponegoro University. The GPA grouping refers to the Academic Regulations of Diponegoro University i.e. satisfactory GPA (2,00 to 2,75), very satisfactory (2,76 to 3,50) and with honors (*cum laude*) (3,51 to 4,00). By using the multiple discriminant analysis with mixture variables, the accuration of classification based on training and testing data reach to 71,875% and 41,667% respectively.

1. PENDAHULUAN

Belajar merupakan suatu proses mendapatkan suatu ilmu dari suatu sumber ilmu, dengan hasil akhir berupa prestasi belajar yang dapat diukur. Menurut Alwi (2011), prestasi belajar merupakan penguasaan pengetahuan atas keterampilan yang dikembangkan oleh mata pelajaran, lazimnya ditunjukkan dengan tes atau angka nilai yang diberikan oleh guru.

Pada jenjang pendidikan tinggi, prestasi belajar diukur dengan nilai Indeks Prestasi (IP) atau Indeks Prestasi Kumulatif (IPK). Indeks Prestasi (IP) adalah suatu nilai kredit rata-rata yang merupakan satuan nilai akhir yang menggambarkan nilai proses belajar tiap semester, atau dapat juga diartikan sebagai besaran atau angka yang menyatakan prestasi keberhasilan pada proses belajar mahasiswa dalam satu semester. Sedangkan Indeks Prestasi Kumulatif (IPK) merupakan hasil kumulatif dari IP untuk keseluruhan nilai yang didapat sebanyak total semester yang telah dijalani. IP dan IPK diukur dengan skala angka mulai dari 0 sebagai skala terendah sampai dengan 4 sebagai skala tertinggi. Mahasiswa

dengan IP atau IPK yang tinggi mengindikasikan bahwa mahasiswa tersebut mampu mengikuti perkuliahan dengan baik.

Pada dasarnya ada banyak manfaat yang didapat mahasiswa dengan IP dan IPK yang baik, diantaranya memungkinkan untuk mendapatkan beasiswa. Bahkan saat ini banyak perusahaan merekrut pegawai dengan salah satu syarat yaitu IPK mencapai nilai yang cukup tinggi yang ditetapkan perusahaan.

Berdasarkan Peraturan Akademik Undip tahun 2014, IPK diklasifikasikan menjadi 3 yaitu memuaskan (2,00-2,75), sangat memuaskan (2,76-3,50) dan dengan pujian (*cumlaude*) (3,51-4,00). Untuk mencapai IPK yang baik, banyak faktor yang mempengaruhinya. Menurut Purwanto (2002), faktor-faktor yang mempengaruhi IPK terbagi menjadi dua, yaitu faktor internal dan eksternal. Faktor internal mencakup aspek fisik, bakat, minat, kecerdasan, dan yang lainnya yang berasal dari dalam mahasiswa. Sedangkan faktor eksternal mencakup lingkungan, sarana, fasilitas, dan yang lainnya yang berasal dari luar mahasiswa. Sementara menurut Hakam (2014), faktor-faktor yang mencakup faktor internal dan eksternal yaitu uang saku, usia, nilai rata-rata UN SMA, banyak organisasi, lama penggunaan internet, dan lama belajar.

Universitas Diponegoro sebagai salah satu perguruan tinggi negeri di Indonesia diharapkan dapat menjadi salah satu instrumen pendidikan nasional yang dapat menjadi pusat penyelenggaraan dan pengembangan pendidikan tinggi serta pemeliharaan, pembinaan dan pengembangan ilmu pengetahuan, teknologi dan/atau kesenian sebagai suatu masyarakat ilmiah yang dapat meningkatkan mutu kehidupan bermasyarakat, berbangsa dan bernegara. Studi kasus dari penelitian ini dilakukan pada mahasiswa S1 angkatan 2015 Fakultas Sains dan Matematika Universitas Diponegoro (FSM Undip). Penting bagi para pemangku di pendidikan tinggi untuk membuat sebuah perkiraan tentang kisaran IPK yang akan diperoleh oleh seorang mahasiswa berdasarkan faktor-faktor internal maupun eksternal. Salah satu metode Statistika yang dapat digunakan untuk memprediksi capaian IPK mahasiswa adalah analisis diskriminan.

Diskriminan atau biasa disebut juga sebagai metode pemisahan (*separation*) pertama kali diperkenalkan oleh Sir Ronald Fisher pada tahun 1936. Seperti pada analisis regresi, analisis diskriminan merupakan kombinasi linier dari dua atau lebih peubah bebas yang akan membedakan dua observasi atau lebih di dalam sebuah kelompok (Widarjono, 2010). Pengelompokan pada analisis diskriminan bersifat *mutually exclusive*, yaitu jika suatu objek telah masuk pada salah satu kelompok maka tidak dapat menjadi anggota dari kelompok yang lain (Jazimah, 2010). Analisis diskriminan adalah teknik multivariat yang termasuk pada *Dependence Method*, dengan ciri adanya peubah dependen dan independen. Dengan demikian, ada peubah yang hasilnya tergantung pada data peubah independen (Santoso, 2009).

Dalam analisis diskriminan akan dihasilkan suatu fungsi yang dapat membedakan antara dua kelompok atau lebih. Terbentuknya fungsi ini dikarenakan adanya pengaruh antara beberapa peubah bebas yang dapat membedakan dua atau lebih kelompok populasi yang ada terhadap peubah terikat. Pada dasarnya, fungsi diskriminan dapat digunakan untuk mendiskripsikan peubah-peubah bebas suatu observasi yang dapat membedakan dari kelompok populasi yang ada. Dengan kata lain, diskriminan merupakan suatu metode yang dapat digunakan sebagai kriteria pengelompokan yang dilakukan berdasarkan perhitungan statistik terhadap kelompok populasi. Perhitungan fungsi diskriminan akan menghasilkan suatu nilai. Nilai-nilai yang dihasilkan dari fungsi diskriminan dikenal dengan skor diskriminan (Johnson dan Winchern, 2007). Dalam analisis diskriminan, salah satu asumsi

penting yang harus dipenuhi adalah bahwa peubah-peubah bebas yang digunakan terdistribusi secara normal multivariat. Hal itu berarti bahwa peubah-peubah tersebut berjenis kontinu. Namun sering kali dalam banyak kasus tidak semua peubah-peubah penting berjenis kontinu, beberapa diantaranya berjenis kategorik.

Berdasarkan latar belakang tersebut, akan dikaji kinerja analisis diskriminan berganda dengan menggunakan peubah kategorik dan kontinu. Model tersebut digunakan untuk memprediksi kategori IPK seorang mahasiswa berdasarkan faktor-faktor eksternal maupun internal yang detail daripada faktor-faktor tersebut.

2. TINJAUAN PUSTAKA

2.1. Analisis Diskriminan Berganda

Analisis diskriminan merupakan teknik statistik yang digunakan oleh Edward I. Altman dalam penelitiannya untuk memprediksi kebangkrutan perusahaan, yang lebih dikenal dengan sebutan analisis diskriminan berganda (Altman, 1968). Misalkan \mathbf{V} adalah komponen vektor random berdimensi p yang memiliki densitas $f_i(\mathbf{v})$ pada populasi yang berbeda π_i ($i = 1, \dots, g$). Misalkan q_i adalah peluang prior dari pengambilan sebuah observasi dari π_i dan misalkan c_{ij} adalah biaya yang terkait dengan pengklasifikasian sebuah observasi ke π_i , ketika observasi sebelumnya berasal dari π_j ($i, j = 1, \dots, g$).

Pendekatan teorema pengambilan keputusan dilakukan dengan membagi ruang observasi menjadi g ruang yang *mutually exclusive* R_1, \dots, R_g , sehingga jika observasi \mathbf{v} jatuh pada daerah R_i , maka akan dialokasikan pada populasi π_i . Daerah dari klasifikasi yang meminimumkan harga harapan yang disebabkan oleh penarikan kesimpulan yang salah didefinisikan dengan menetapkan \mathbf{v} pada R_k jika

$$\sum_{i=i, i \neq k}^g q_i c_{ki} f_i(\mathbf{v}) < \sum_{i=i, i \neq j}^g q_i c_{ki} f_i(\mathbf{v}) \quad (j = 1, \dots, g; j \neq k) \quad (1)$$

Pada kasus dengan $g = 2$, hal ini mengarah pada definisi R_1 dan R_2 sebagai daerah sebagai berikut:

$$\frac{f_1(\mathbf{v})}{f_2(\mathbf{v})} \geq \frac{q_2 c_{22}}{q_1 c_{21}} \quad (2)$$

Menurut aturan Bayes, jika $c_{12} = c_{21}$, alokasikan \mathbf{v} pada populasi untuk setiap $q_i f_i(\mathbf{v})$ terbesar. Pada kasus $g = 2$, aturan ini akan menunjukkan definisi dari R_1 dan R_2 yang diberikan pada persamaan (2), tetapi dengan c_{12} dan c_{21} dihapus dari sisi kanan persamaan. Akhirnya, jika tidak menggunakan konsep peluang prior, maka aturan fungsi diskriminan akan mengalokasikan \mathbf{v} pada populasi untuk setiap $f_i(\mathbf{v})$ yang terbesar.

Pada penggunaan teori tersebut, asumsi menyatakan bahwa π_i adalah populasi normal multivariat dengan vektor mean μ_i dan matriks varian kovarian Σ_i ($i = 1, \dots, g$). Untuk lebih menyederhanakan perhitungan matematisnya, matriks varian kovarian juga diasumsikan $\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$ dan biaya yang sama yaitu $c_{ij} = c$ ($i, j = 1, \dots, g$). Pada kasus ini, harga harapan minimum dan aturan pengalokasian Bayes menunjukkan pada daerah R_j yang didefinisikan sebagai himpunan \mathbf{v} sehingga

$$u_{jk}(\mathbf{v}) > \log\left(\frac{q_k}{q_j}\right), \quad k = 1, \dots, g; k \neq j \quad (3)$$

dimana

$$u_{jk}(\mathbf{v}) = \log\{f_j(\mathbf{v})/f_k(\mathbf{v})\} = (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}^{-1} \left\{ \mathbf{v} - \frac{1}{2}(\boldsymbol{\mu}_j + \boldsymbol{\mu}_k) \right\} \quad (4)$$

$$u_{jk}(\mathbf{v}) > 0 \quad (k = 1, \dots, g; k \neq j) \quad (5)$$

Jika setiap pendekatan di atas diadopsi ketika $g > 2$, maka sebuah himpunan daerah harus dibentuk; yang masing-masing membutuhkan perbandingan dari sejumlah fungsi. Hal ini akan menjadi proses yang panjang dan salah satu yang mungkin tidak memberikan pemahaman pada struktur populasi. Pada praktiknya, parameter-parameter populasi yang disebutkan di atas tidak diketahui. Informasi yang didapat hanyalah tentang π_j yang biasanya dalam bentuk data training atau data sampel yang diketahui untuk masing-masing populasi. Krzanowski (1982) telah mereview prosedur yang mungkin untuk kasus $g = 2$. Jadi $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_g$ dan $\boldsymbol{\Sigma}$ di atas akan digantikan dengan dugaan dari data sampel, yang dinamakan vektor rata-rata sampel $\bar{\mathbf{x}}_i (i = 1, \dots, g)$ dan matrik varian kovarian S .

2.2. Analisis Diskriminan dengan Peubah Bebas Campuran Kategorik dan Kontinu

Misalkan diasumsikan bahwa vektor random \mathbf{V} dipartisi menjadi $(\mathbf{Z}'|\mathbf{X}')$ dimana \mathbf{Z} adalah vektor dari r peubah diskrit, dan \mathbf{X} adalah vektor dari p peubah kontinu. Setiap kombinasi tertentu dari komponen $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_k$, dari \mathbf{Z} mendefinisikan sebuah kondisi dari peubah random multinomial \mathbf{W} . Misalkan jumlah maksimum dari kondisi yang mungkin untuk \mathbf{W} adalah k , dan misalkan W_m menunjukkan keadaan ke- m , maka model lokasinya seperti yang diperkenalkan oleh Olkin dan Tate (1961), dan seperti yang digunakan pada analisis diskriminan oleh Krzanowski (1982), mengasumsikan bahwa distribusi bersyarat dari \mathbf{X} diberikan sebagai berikut:

$$\mathbf{X} | W_m, \pi_i \sim N(\boldsymbol{\mu}_i^m, \boldsymbol{\Sigma}), m = 1, \dots, k; i = 1, \dots, g \quad (6)$$

Diasumsikan peluang mendapatkan state W_m dari \mathbf{W} , adalah p_{im} dengan $m = 1, \dots, k$ dan $i = 1, \dots, g$. Dengan asumsi tersebut, aturan diskriminan untuk kasus $g = 2$ dinyatakan dengan menentukan kondisi W_m dari \mathbf{Z} , kemudian mengklasifikasikan \mathbf{v} pada π_1 jika nilai \mathbf{y} dari \mathbf{X}

$$(\boldsymbol{\mu}_1^m - \boldsymbol{\mu}_2^m)' \boldsymbol{\Sigma}^{-1} \left\{ \mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1^m - \boldsymbol{\mu}_2^m) \right\} \geq \log(p_{2m}/p_{1m}) \quad (7)$$

dan pada π_2 jika tidak memenuhi kondisi di atas.

Untuk memecahkan kasus $g > 2$ akan diterapkan densitas peluang yang tidak bersyarat untuk g populasi dengan menggunakan model lokasi atau modifikasinya. Hal tersebut digunakan untuk mendefinisikan daerah g yaitu R_1, \dots, R_g pada aturan perbandingan fungsi pasangan seperti persamaan (6). Krzanowski (1983) telah menemukan jarak antara campuran dua peubah populasi berdasarkan asumsi variasi tentang struktur populasi, termasuk model lokasi tertentu menggunakan model di atas.

Selanjutnya, menurut Krzanowski (1980), untuk $g > 2$ proses mangalokasikan sebuah obyek ke populasi ke j didasarkan atas skor diskriminan yang definisikan sebagai berikut:

$$\rho_j = C \cdot (p_{jm})^{1/2} \exp \left\{ -\frac{1}{4} (\mathbf{x} - \boldsymbol{\mu}_j^{(m)})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_j^{(m)}) \right\} \quad (8)$$

dimana $C = \{(2\pi)^p |\boldsymbol{\Sigma}|\}^{-1/4}$ dan \mathbf{x} adalah subvektor kontinu dari \mathbf{v} serta m adalah state multinomial yang didefinisikan dengan subvektor diskrit dari \mathbf{v} . \mathbf{v} dialokasikan pada populasi π_j jika nilai ρ_j pada persamaan (8) bernilai maksimum.

Pada praktiknya, parameter populasi tidak diketahui dan akan diduga dari data sampel. Misalkan $n_{1m}, n_{2m}, \dots, n_{gm}$ adalah banyaknya amatan yang berada di sel m dari populasi $\pi_1, \pi_2, \dots, \pi_g$ dan misalkan pula $\mathbf{x}_{ji}^{(m)}$ adalah vektor peubah-peubah kontinu terkait dengan amatan ke j di sel m dari sampel populasi π_i . Kemudian jika

$$\bar{\mathbf{x}}_i^{(m)} = \frac{\sum_{j=1}^{n_{im}} \mathbf{x}_{ji}^{(m)}}{n_{im}}$$
 maka penduga-penduga kemungkinan maksimum (*maximum likelihood*)

dari parameter-parameter populasi $p_{im}, \boldsymbol{\mu}_i^{(m)}$ dan $\boldsymbol{\Sigma}$ diberikan oleh

$$\hat{p}_{im} = \frac{n_{im}}{n_i}$$

$$\hat{\boldsymbol{\mu}}_i^{(m)} = \bar{\mathbf{x}}_i^{(m)} \tag{9}$$

$$\mathbf{S}_{pooled} = \frac{1}{(n_1 + n_2 + \dots + n_g - 2k)} \sum_{i=1}^g \sum_{m=1}^k \sum_j^{n_{im}} \left(\mathbf{x}_{ji}^{(m)} - \bar{\mathbf{x}}_i^{(m)} \right) \left(\mathbf{x}_{ji}^{(m)} - \bar{\mathbf{x}}_i^{(m)} \right)' \tag{10}$$

dengan $i = 1, 2, \dots, g$ dan $m = 1, 2, \dots, k$

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data primer yang diperoleh menggunakan kuesioner dari 120 orang mahasiswa S1 Fakultas Sains dan Matematika Universitas Diponegoro angkatan 2015, berupa 30 observasi IPK memuaskan, 62 observasi IPK sangat memuaskan dan 28 observasi IPK *cumlaude*. Dalam hal ini IPK yang diperoleh sampai dengan semester 4. Sampel diambil dari tiap departemen dengan penyampelan acak sederhana, yaitu 20 mahasiswa dari departemen Matematika, 14 mahasiswa dari departemen Statistika, 11 mahasiswa dari departemen Teknik Informatika, 18 mahasiswa dari departemen Fisika, 31 mahasiswa dari departemen Kimia, dan 26 mahasiswa dari departemen Biologi.

Pada tulisan ini digunakan 6 peubah bebas, yaitu jenis kelamin (X_1) berupa data kualitatif berskala nominal dengan jenis kelamin laki-laki sebagai kategori 1 dan jenis kelamin perempuan sebagai kategori 2, penggunaan internet per bulan (X_2), jumlah waktu belajar per minggu (X_3), rata-rata nilai UN SMA (X_4), jumlah uang saku per minggu (X_5), yang masing-masing berupa data kuantitatif berskala rasio, dan jalur masuk ke universitas (X_6) berupa data kualitatif berskala nominal, dengan SNMPTN sebagai kategori 1, SBMPTN sebagai kategori 2 dan UM sebagai kategori 3, serta klasifikasi IPK sebagai peubah dependen (Y) yang berupa data kualitatif berskala ordinal, yaitu IPK 2,00-2,75 sebagai IPK memuaskan, IPK 2,76-3,50 sebagai IPK sangat memuaskan, dan IPK 3,51-4,00 sebagai IPK *cumlaude*.

Sebanyak 80% data pada masing-masing klasifikasi IPK digunakan untuk membentuk model diskriminan (data *training*) dan 20% digunakan untuk validasi kemampuan model diskriminan (data *testing*). Untuk masing-masing kelompok IPK, digunakan data *training* sebanyak 22 observasi yang masuk pada klasifikasi IPK memuaskan, 54 observasi yang masuk pada klasifikasi IPK sangat memuaskan dan 20

observasi yang masuk pada klasifikasi IPK *cumlaude*. Sedangkan sisanya, yaitu sebanyak 24 observasi dari masing-masing klasifikasi IPK digunakan untuk data *testing*, dengan rincian yaitu sebanyak 8 observasi yang masuk pada kategori IPK memuaskan, 8 observasi yang masuk pada kategori IPK sangat memuaskan dan 8 observasi yang masuk pada kategori IPK *cumlaude*.

Metode statistika yang digunakan untuk menganalisis data dalam tulisan ini adalah analisis diskriminan berganda dengan peubah bebas campuran antara kategorik dan kontinu, untuk memisahkan tiap kelompok IPK mahasiswa berdasarkan peubah dependen, dengan bantuan *Software Ms. Excel dan SPSS 16.0*.

4. HASIL DAN PEMBAHASAN

Pada bagian ini menampilkan hasil-hasil analisis yang diperoleh meliputi analisis deskriptif, pengujian asumsi, dan model diskriminan berganda untuk peubah bebas campuran.

4.1 Analisis Deskriptif

Analisis deskriptif digunakan untuk mengetahui karakteristik mahasiswa S1 angkatan 2015 FSM Undip yang masuk dalam kelompok klasifikasi IPK tertentu. Analisis tersebut dilakukan terhadap peubah-peubah yang berjenis kontinu. Ringkasan statistiknya ada di Tabel 1.

Tabel 1 Statistik Deskriptif untuk Peubah-Peubah Kontinu

Peubah	Kriteria IPK					
	Memuaskan		Sangat Memuaskan		<i>Cumlaude</i>	
	Rata-rata	Simpangan Baku	Rata-rata	Simpangan Baku	Rata-rata	Simpangan Baku
Penggunaan Internet (GB)	22,733	2,690	22,290	3,091	21,071	3,150
Lama Waktu Belajar (Jam)	20,100	1,788	20,073	1,280	20,286	2,386
Rata-rata Nilai UN SMA	83,231	4,287	82,031	4,148	84,832	4,410
Banyak Uang Saku (Rupiah/Minggu)	172166,667	19013,758	174677,419	23589,227	169464,286	22125,036

Berdasarkan Tabel 1, diketahui bahwa jumlah waktu penggunaan internet terbesar adalah mahasiswa dengan IPK memuaskan, jumlah waktu belajar terbesar adalah mahasiswa dengan IPK *cumlaude*, rata-rata UN SMA terbesar adalah mahasiswa dengan IPK *cumlaude* dan jumlah uang saku terbesar adalah mahasiswa dengan IPK sangat memuaskan.

Peubah bebas kategorik yang digunakan dalam tulisan ini adalah jenis kelamin yang terdiri dari dua kategori dan jalur masuk ke perguruan tinggi yang terdiri dari tiga kategori, sehingga sel yang terbentuk sebanyak 6 sel. Sel pertama adalah jenis kelamin laki-laki dan jalur masuk SNMPTN, sel kedua jenis kelamin laki-laki dan jalur masuk SBMPTN, sel ketiga jenis kelamin laki-laki dan jalur masuk UM, sel keempat jenis kelamin perempuan dan jalur masuk SNMPTN, sel kelima jenis kelamin perempuan dan jalur masuk

SBMPTN, dan sel keenam jenis kelamin perempuan dan jalur masuk UM. Tabel berikut adalah proporsi klasifikasi IPK mahasiswa berdasarkan jenis kelamin dan jalur masuk.

Berdasarkan jalur masuk, diketahui bahwa sebanyak 42,5% mahasiswa masuk melalui jalur SNMPTN, 32,5% melalui jalur SBMPTN dan sisanya 25% melalui jalur UM. Sedangkan berdasarkan jenis kelamin, 52,5% adalah mahasiswa laki-laki dan 47,5% adalah mahasiswa perempuan. Asumsi-asumsi yang harus dipenuhi dalam model diskriminan ini adalah bahwa peubah bebas terdistribusi secara normal multivariat dengan matrik varian kovarian di setiap sel sama.

Tabel 2. Tabel Kontingensi

Sel	Kriteria IPK			Total
	Memuaskan	Sangat Memuaskan	Cumlaude	
Laki-laki, SNMPTN	8	14	6	28
Laki-laki, SBMPTN	5	6	5	16
Laki-laki, UM	5	10	4	19
Perempuan, SNMPTN	4	14	5	23
Perempuan, SBMPTN	5	13	5	23
Perempuan, UM	3	5	3	11
Total	30	62	28	120

Vektor rata-rata untuk berbagai kategori IPK ditampilkan di Tabel 3. Informasi ini berguna untuk proses pendugaan parameter-parameter populasi normal multivariat.

Tabel 3. Vektor Rata-Rata untuk berbagai Kategori IPK

Sel	Vektor Rata-Rata Kategori IPK memuaskan	Vektor Rata-Rata Kategori IPK Sangat memuaskan	Vektor Rata-Rata Kategori IPK Cumlaude
Laki-Laki, SNMPTN	$\bar{x}_1^{(1)} = \begin{pmatrix} 23 \\ 20,5714 \\ 84,1414 \\ 174285,7143 \end{pmatrix}$	$\bar{x}_2^{(1)} = \begin{pmatrix} 23,5385 \\ 19,6539 \\ 83,0069 \\ 179615,3846 \end{pmatrix}$	$\bar{x}_3^{(1)} = \begin{pmatrix} 22 \\ 22,2 \\ 86,948 \\ 178000 \end{pmatrix}$
Laki-Laki, SBMPTN	$\bar{x}_1^{(2)} = \begin{pmatrix} 25 \\ 19,33333 \\ 84,4967 \\ 181666,6667 \end{pmatrix}$	$\bar{x}_2^{(2)} = \begin{pmatrix} 22,75 \\ 19,75 \\ 78,505 \\ 192500,0000 \end{pmatrix}$	$\bar{x}_3^{(2)} = \begin{pmatrix} 23 \\ 18,6667 \\ 79,4 \\ 168333,3333 \end{pmatrix}$
Laki-Laki, UM	$\bar{x}_1^{(3)} = \begin{pmatrix} 22,3333 \\ 19,6667 \\ 83,0233 \\ 191666,6667 \end{pmatrix}$	$\bar{x}_2^{(3)} = \begin{pmatrix} 24,875 \\ 19,25 \\ 84,6675 \\ 175625,0000 \end{pmatrix}$	$\bar{x}_3^{(3)} = \begin{pmatrix} 21,5 \\ 17,5 \\ 84,5 \\ 187500,0000 \end{pmatrix}$

Perempuan, SNMPTN	$\bar{x}_1^{(4)} = \begin{pmatrix} 19,66667 \\ 19 \\ 73,7133 \\ 158333,3333 \end{pmatrix}$	$\bar{x}_2^{(4)} = \begin{pmatrix} 22,6154 \\ 19,9231 \\ 81,9823 \\ 170769,2308 \end{pmatrix}$	$\bar{x}_3^{(4)} = \begin{pmatrix} 22,25 \\ 20 \\ 86,615 \\ 143750,0000 \end{pmatrix}$
Perempuan, SBMPTN	$\bar{x}_1^{(5)} = \begin{pmatrix} 23,7500 \\ 18,2500 \\ 81,6875 \\ 177500 \end{pmatrix}$	$\bar{x}_2^{(5)} = \begin{pmatrix} 18,5833 \\ 21,5 \\ 81,4 \\ 165833,3333 \end{pmatrix}$	$\bar{x}_3^{(5)} = \begin{pmatrix} 17,5 \\ 21,75 \\ 86,0375 \\ 161250,0000 \end{pmatrix}$
Perempuan, UM	$\bar{x}_1^{(6)} = \begin{pmatrix} 18 \\ 21 \\ 82,725 \\ 162500 \end{pmatrix}$	$\bar{x}_2^{(6)} = \begin{pmatrix} 23 \\ 20,25 \\ 80,7625 \\ 178750,0000 \end{pmatrix}$	$\bar{x}_3^{(6)} = \begin{pmatrix} 21,5 \\ 20,5 \\ 84,15 \\ 187500,0000 \end{pmatrix}$

4.2. Uji Asumsi Normal Multivariat dan Kesamaan Matriks Varian Kovarian

Pengujian secara formal terhadap peubah-peubah kontinu apakah memenuhi asumsi normal multivariat dilakukan dengan menghitung korelasi antara kuantil distribusi Chi-Kuadrat dengan jarak tegeneralisasi, dengan hipotesis sebagai berikut:

- Hipotesis
 H_0 : data berdistribusi normal multivariat
 H_1 : data tidak berdistribusi normal multivariat
- Taraf Signifikansi
 $\alpha = 5\%$
- Statistik Uji

$$r_Q = \frac{\sum_{j=1}^n (d_{(j)} - \bar{d})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^n (d_{(j)} - \bar{d})^2} \sqrt{\sum_{j=1}^n (q_{(j)} - \bar{q})^2}}$$

$$r_Q = \frac{\sum_{j=1}^{120} (d_{(j)} - \bar{d})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^{120} (d_{(j)} - \bar{d})^2} \sqrt{\sum_{j=1}^{120} (q_{(j)} - \bar{q})^2}}$$

$$r_Q = \frac{1064,73919}{34,98753 \cdot 30,61785} = 0,9939$$

dengan $d_{(j)}$ adalah jarak tegeneralisasi dan $q_{(j)}$ adalah kuantil atas distribusi Chi-Kuadrat $(q_p((j-0,5)/n))$.

- Kriteria Penolakan

H_0 ditolak jika $r_Q < r$ tabel. Dari Johnson dan Winchern (2007) diperoleh nilai r tabel = 0,9793 sehingga $r_Q = 0,9939 > r$ tabel = 0,9793. Oleh karena itu dapat disimpulkan bahwa pada taraf signifikansi 5% H_0 diterima dan data observasi berdistribusi normal multivariat.

Asumsi selanjutnya yang harus dipenuhi adalah asumsi kesamaan matriks varian kovarian. Pengujian ini dapat dilakukan dengan uji Box's M melalui pendekatan distribusi

Chi-Kuadrat dengan derajat bebas sebesar $p(p + 1)(g - 1)/2$. Hipotesis untuk uji kesamaan matriks varian kovarian adalah (Johnson dan Winchern, 2007):

a. Hipotesis

$$H_0: \Sigma_1^{(1)} = \Sigma_1^{(2)} = \Sigma_1^{(3)} = \dots = \Sigma_3^{(6)}$$

H_1 : terdapat paling sedikit dua matriks varian kovarian berbeda antar kelompok

b. Taraf Signifikansi

$$\alpha = 5\%$$

c. Statistik Uji

Uji Box's M:

$$M = (n - g) \log |S_{pooled}| - \sum_{i=1}^g \sum_{m=1}^k (n_i - 1) \log |S_i^{(m)}|, \text{ dengan syarat } |S_i^{(m)}| > 0$$

S_{pooled} diperoleh seperti persamaan (10)

$$S_i^{(m)} = \frac{\sum_{j=1}^{n_{im}} \left(\mathbf{x}_{ji}^{(m)} - \mathbf{x}_i^{(m)} \right) \left(\mathbf{x}_{ji}^{(m)} - \mathbf{x}_i^{(m)} \right)'}{n_{im} - 1}$$

$S_i^{(m)}$ adalah matrik varian kovarian pada sampel ke-i di sel ke-m dengan $i = 1, 2, 3$ dan $m = 1, 2, \dots, 6$

$$S_{pooled} = \begin{pmatrix} 7,8791 & -1,449 & 0,3052 & 23763,0907 \\ -1,449 & 2,7763 & 1,6724 & -935,2741 \\ 0,3052 & 1,6724 & 18,6173 & -4231,7833 \\ 23763,0907 & -935,2741 & -4231,7833 & 500525747,8684 \end{pmatrix}$$

p = banyaknya peubah kontinu, g = banyaknya kategori IPK, n = banyak data, $n_1 = 22, n_2 = 54, n_3 = 20$

d. Kriteria Penolakan

H_0 ditolak pada taraf signifikansi α apabila nilai p value $< \alpha$. Diperoleh nilai Box's M sebesar 30,062 dengan nilai p value 0,371. H_0 diterima karena p value (0,371) $> \alpha$ (0,05) sehingga dapat disimpulkan bahwa pada taraf signifikansi 5% matriks varian kovarian pada setiap sel sama.

4.3 Diskriminan Berganda dengan Peubah Bebas Campuran Kategorik dan Kontinu

Model diskriminan dengan peubah bebas campuran kategorik dan kontinu dapat dibentuk setelah asumsi normal multivariat dan kesamaan matriks varian kovarian terpenuhi. Untuk membangun model diskriminan, dilakukan perhitungan dengan persamaan pada aturan lokasi. Model diskriminan dengan peubah bebas campuran kategorik dan kontinu ini akan dimiliki oleh masing-masing sel. Untuk mendapatkan model tersebut, akan dicari peluang sebuah amatan di sel m yang berasal dari populasi j , dinotasikan dengan p_{jm} . Nilai dugaan yang diperoleh terdapat pada Tabel 6.

Tabel 6. Nilai Dugaan p_{jm}

M	Sel	Proporsi		
		IPK Memuaskan (j=1)	IPK Sangat Memuaskan (j=2)	IPK Cumlaude (j=3)
1	Laki-laki, SNMPTN	0,2667	0,2258	0,2143
2	Laki-laki, SBMPTN	0,1667	0,0968	0,1786
3	Laki-laki, UM	0,1667	0,1613	0,1429
4	Perempuan, SNMPTN	0,1333	0,2258	0,1786
5	Perempuan, SBMPTN	0,1667	0,2097	0,1786
6	Perempuan, UM	0,1000	0,0806	0,1071

Selanjutnya, berdasarkan matriks varian kovarian seperti pada Tabel 3, 4, dan 5, penduga kemungkinan maksimum (*maximum likelihood*) bagi Σ adalah:

$$S_{pooled} = \frac{1}{(n_1 + n_2 + n_3 - 2k)} \sum_{i=1}^3 \sum_{m=1}^k \sum_j^{n_m} (\mathbf{x}_{ji}^{(m)} - \bar{\mathbf{x}}_i^{(m)})(\mathbf{x}_{ji}^{(m)} - \bar{\mathbf{x}}_i^{(m)})'$$

$$S_{pooled} = \begin{pmatrix} 7,8791 & -1,449 & 0,3052 & 23763,0907 \\ -1,449 & 2,7763 & 1,6724 & -935,2741 \\ 0,3052 & 1,6724 & 18,6173 & -4231,7833 \\ 23763,0907 & -935,2741 & -4231,7833 & 500525747,8684 \end{pmatrix}$$

Aturan klasifikasi secara umum adalah:

$$\hat{p}_j = C. (\hat{p}_{jm})^{1/2} \exp \left\{ -\frac{1}{4} (\mathbf{x} - \bar{\mathbf{x}}_j^{(m)})' \hat{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_j^{(m)}) \right\}$$

dimana $C = \{[(2\pi)^p |\hat{\Sigma}|]^{-1/4}\}$, \mathbf{x} adalah subvektor kontinu dari \mathbf{v} , dengan \mathbf{v} adalah data yang akan diklasifikasikan IPKnya dan m adalah state multinomial yang didefinisikan oleh subvektor diskrit dari \mathbf{v} . \mathbf{v} dialokasikan pada populasi π_j jika nilai ρ_j pada persamaan tersebut bernilai maksimum. Karena terdapat 3 kategori IPK, maka ρ_j dihitung dari ρ_1 hingga ρ_3 . Berikut adalah formula untuk setiap ρ_j .

Untuk kategori IPK memuaskan (j = 1):

$$\hat{\rho}_1 = C. (\hat{p}_{1m})^{1/2} \exp \left\{ -\frac{1}{4} (\mathbf{x} - \bar{\mathbf{x}}_1^{(m)})' \hat{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1^{(m)}) \right\}$$

Untuk kategori IPK sangat memuaskan (j = 2):

$$\hat{\rho}_2 = C. (\hat{p}_{2m})^{1/2} \exp \left\{ -\frac{1}{4} (\mathbf{x} - \bar{\mathbf{x}}_2^{(m)})' \hat{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2^{(m)}) \right\}$$

Untuk kategori IPK cumlaude (j = 3):

$$\hat{\rho}_3 = C. (\hat{p}_{3m})^{1/2} \exp \left\{ -\frac{1}{4} (\mathbf{x} - \bar{\mathbf{x}}_3^{(m)})' \hat{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_3^{(m)}) \right\}$$

4.4. Prediksi

Proses prediksi terhadap kategori IPK dari seorang mahasiswa menggunakan model diskriminan dengan peubah campuran dilakukan melalui dua tahap. Tahap pertama dengan cara mengidentifikasi terlebih dahulu lokasi sel dari mahasiswa tersebut. Misalkan seorang mahasiswa laki-laki yang diterima di UNDIP melalui jalur SNMPTN dengan jumlah penggunaan internet per minggu (x_2) = 25 Gb, jumlah waktu belajar per minggu (x_3) = 20 jam, rata-rata nilai UN SMA (x_4) = 81,95, jumlah uang saku per minggu (x_5) = 160.000 maka mahasiswa tersebut berada di sel 1 dengan dengan vektor amatan peubah kontinu $\mathbf{x}' = (25 \ 20 \ 81,95 \ 160.000)$. Langkah kedua adalah menghitung skor diskriminan untuk masing-masing kategori IPK, yaitu:

$$\hat{p}_1 = C. (\hat{p}_{11})^{1/2} \exp \left\{ -\frac{1}{4} (\mathbf{x} - \bar{\mathbf{x}}_1^{(1)})' \hat{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1^{(1)}) \right\} = 8,4345x10^{-5}$$

$$\hat{p}_2 = C. (\hat{p}_{21})^{1/2} \exp \left\{ -\frac{1}{4} (\mathbf{x} - \bar{\mathbf{x}}_2^{(1)})' \hat{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2^{(1)}) \right\} = 7,2401x10^{-5}$$

$$\hat{p}_3 = C. (\hat{p}_{31})^{1/2} \exp \left\{ -\frac{1}{4} (\mathbf{x} - \bar{\mathbf{x}}_3^{(1)})' \hat{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_3^{(1)}) \right\} = 3,5932x10^{-5}$$

Berdasarkan hasil perhitungan tersebut, karena nilai \hat{p}_1 lebih besar dari yang lainnya, maka mahasiswa dengan karakteristik seperti di atas diklasifikasikan pada IPK kategori 1 yaitu memuaskan.

Secara keseluruhan, ketepatan klasifikasi model ini pada kasus IPK mahasiswa Fakultas Sains dan Matematika UNDIP dapat dilihat di Tabel 7 dan 8. Dari Tabel 7 dan 8, diketahui bahwa nilai akurasi klasifikasi untuk data training sebesar 71,875% sedangkan untuk data testing sebesar 41,667%.

Tabel 7. Tabel Konfusi Klasifikasi Data Training

Observasi Aktual	Prediksi			Total
	IPK Memuaskan	IPK Sangat Memuaskan	IPK <i>Cumlaude</i>	
IPK Memuaskan	20	2	0	22
IPK Sangat Memuaskan	9	37	8	54
IPK <i>Cumlaude</i>	5	3	12	20
Total	34	42	20	96

Tabel 8. Tabel Konfusi Klasifikasi Data Testing

Observasi Aktual	Prediksi			Total
	IPK Memuaskan	IPK Sangat Memuaskan	IPK <i>Cumlaude</i>	
IPK Memuaskan	4	4	0	8
IPK Sangat Memuaskan	5	3	0	8
IPK <i>Cumlaude</i>	3	2	3	8
Total	12	9	3	24

5. KESIMPULAN

Model diskriminan dengan peubah campuran antara kategorik dan kontinu menjadi alternatif yang baik bagi para statistisi maupun praktisi yang sering kali menemukan kendala ketika melakukan analisis data dengan menggunakan model diskriminan yang melibatkan beberapa peubah kategorik. Walaupun sedikit rumit tetapi keberadaan model seperti ini sangat dibutuhkan seiring dengan semakin meningkatnya pemanfaatan alat-alat statistika di berbagai bidang baik kesehatan, maupun bisnis, pendidikan. Dalam kasus IPK mahasiswa di Fakultas Sains dan Matematika Undip untuk mahasiswa angkatan 2015, penelitian ini menemukan bahwa mahasiswa dengan IPK *cumlaude* memiliki karakteristik penggunaan internet yang lebih rendah dari kategori IPK yang lainnya, yaitu hanya 21,0714 GB per minggu. Selain itu mahasiswa dengan IPK *cumlaude*, waktu untuk belajar cenderung lebih lama dan nilai rata-rata UN SMA cenderung lebih tinggi dari kategori IPK yang lainnya. Sebaliknya jumlah uang saku mahasiswa dengan kategori IPK *cumlaude* cenderung lebih sedikit, yaitu hanya Rp. 169.464,- per minggu. Prediksi kategori IPK menggunakan model diskriminan dengan peubah campuran menunjukkan hasil yang belum baik, yaitu hanya mencapai 71,875% untuk data *training* sedangkan untuk data *testing* sebesar 41,667%.

DAFTAR PUSTAKA

- Altman, E. I. 1968. Financial Ratio. Discriminant Analysis and The Prediction of Corporate Bankruptcy. *Journal of Finance*. Vol. XXIII No. 4. pp. 589-609
- Alwi, H. 2011. *Kamus Besar Bahasa Indonesia*. Jakarta: Gramedia Pustaka Utama
- Hakam, M. 2014. *Analisis Jalur Terhadap Faktor-faktor yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa Statistika Undip*. Skripsi. Semarang: FSM Undip
- Jazimah, R. 2010. *Analisis Diskriminan Beda Dua Kelompok dengan Metode Fisher dan Penerapannya*. Skripsi. Yogyakarta: FKIP. UNY
- Johnson, R. A. dan Winchern, D. W. 2007. *Applied Multivariat Statistical Analysis, 6th Edition*. New Jersey: Prentice Hall, Englewood Cliffs
- Krzanowski, W.J. 1980. Multiple Discriminant Analysis in The Presence of Mixed Continuous and Categorical Data. *Computation and Mathematics with Applications*. 12A. pp. 179-185
- Krzanowski, W.J. 1982. Mixtures of Continuous and Categorical Variables in Discriminant Analysis: A Hypothesis Testing Approach. *Biometrics*. 38. pp. 991-1002
- Krzanowski, W.J. 1983. Distance Between Population Using Mixed Continuous and Categorical Variables. *Biometrics*. 70. pp. 235-243
- Olkin, I. dan R.F. Tate. 1961. Multivariate Correlation Models with Mixed Discrete and Continuous Variables. *Ann. Math. Stat.* 32. pp. 448-465
- Universitas Diponegoro. 2014. *Peraturan Rektor Universitas Diponegoro No. 209/PER/UN7/2012 tentang Peraturan Akademik Bidang Pendidikan Universitas Diponegoro*. Semarang: Undip
- Purwanto, N. 2002. *Psikologi Pendidikan*. Jakarta: PT Remaja Rosdakarya
- Santoso, S. 2009. *Statistika Multivariat*. Jakarta: PT. Elex Media Komputindo

Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan*. Yogyakarta: UPP STIM YKPN.