Perbandingan Klasifikasi Penjurusan Peserta Didik pada Model Diskriminan dan Regresi Logistik Multinomial

Nanda Arista Rizki^{1*}, Petrus Fendiyanto², Ainun Jariah³

^{1,2}Program Studi Pendidikan Matematika, Universitas Mulawarman

³Sekolah Menengah Atas Negeri 2 Samarinda

*nanda.arista@fkip.unmul.ac.id

Abstract

The Covid-19 pandemic, which threatens all corners of the world, has made policy on the National Examination abolished. This lead to a majoring system in SMAN 2 Samarinda just based on test scores from their origin schools exam and student's interest. The results of majoring in the school can be modeled through classification method. This study aimed to compare the accuracy of majoring prediction between discriminant models and multinomial logistic regression models. The research data was taken from origin schools exam and the result of majoring for class X in SMAN 2 Samarinda. Stages of analysis performed were descriptive statistics, discriminant analysis, multinomial logistic regression analysis, determination the best models, and interpretation for the best model. The proportion of training data and testing data applied in this study was 60:40, 70:30, 80:20, and 90:10 with resampling bootstrap B=1000. Based on the results, multinomial logistic regression model was chosen as the best model that describes the student majors system in SMAN 2 Samarinda. It is because their accuracy are higher than accuracy of the discriminant model for every possible proportion of training data and testing data. Thus, the higher the scores of Bahasa, Mathematics and Social Studies, the results of the majors tend not to be in the Language and Culture department. However, the higher the score of English, the results of the majors tend to lie in the Language and Culture department. The difference between the Social Sciences majors and the Mathematics and Natural Sciences majors than the language and culture majors, can be seen in the scores of science subjects.

Keywords: Linear Discriminant Analysis, Multinomial Logistic Regression, Student's Majoring

Abstrak

Pandemi covid-19 yang mengancam hingga seluruh penjuru dunia, membuat kebijakan mengenai Ujian Nasional dihapuskan. Hal ini berakibat terhadap sistem penjurusan di SMAN 2 Samarinda hanya berdasarkan nilai ujian sekolah dan peminatan saja. Hasil penjurusan yang dilakukan oleh sekolah dapat dimodelkan melalui metode klasifikasi. Penelitian ini bertujuan untuk membandingkan ketepatan hasil prediksi penjurusan dari diskriminan dan regresi logistik multinomial. Data yang digunakan dalam penelitian ini adalah nilai ujian sekolah dan hasil penjurusan peserta didik kelas X di SMAN 2 Samarinda. Tahapan analisis yang dilakukan adalah statistika deskriptif, pembentukan model analisis diskriminan, pembentukan model regresi logistik multinomial, penentuan model terbaik, dan interpretasi model terbaik. Proporsi pembagian data training dan data testing yang diterapkan dalam penelitian ini adalah 60:40, 70:30, 80:20, dan 90:10 dengan resampling bootstrap B=1000. Berdasarkan hasil analisis penelitian, maka model regresi logistik multinomial dipilih sebagai model terbaik yang menggambarkan kondisi penjurusan peserta didik di SMAN 2 Samarinda. Hal ini dikarenakan tingkat akurasi lebih tinggi dibandingkan pada model diskriminan untuk setiap kemungkinan dalam pembagian proporsi data training dan testing. Dengan demikian, semakin tinggi nilai mata pelajaran Bahasa Indonesia, Matematika, dan IPS, maka hasil penjurusannya cenderung tidak pada jurusan Bahasa dan budaya. Namun semakin tinggi nilai Bahasa Inggris, maka hasil penjurusannya cenderung terletak pada jurusan Bahasa dan budaya. Perbedaan antara jurusan IPS dan jurusan MIPA dibanding jurusan Bahasa dan budaya, nampak pada nilai mata pelajaran IPA.

Kata kunci: Analisis Diskriminan Linear, Regresi Logistik Multinomial, Penjurusan Peserta Didik

1. Pendahuluan

Kurikulum 2013 dirancang dalam mempersiapkan peserta didik untuk kehidupan kehidupan generasi muda bangsa. Kurikulum ini dikembangkan berdasarkan budaya bangsa Indonesia yang beragam. Kurikulum ini juga dikembangkan atas dasar adanya kebutuhan akan perubahan rancangan dan proses pendidikan dalam rangka memenuhi dinamika kehidupan masyarakat, bangsa, dan negara, sebagaimana termaktub dalam tujuan pendidikan nasional. Dalam pelaksanaannya, kurikulum ini memberikan hak agar sekolah dapat menentukan penjurusannya sendiri. [1]

Sekolah Menengah Atas Negeri (SMAN) 2 Samarinda merupakan salah satu sekolah favorit di kota Samarinda. Sekolah ini memiliki 3 jurusan, yaitu jurusan Bahasa dan Budaya, jurusan Ilmu Pengetahuan Sosial (IPS), dan jurusan Matematika dan Ilmu Pengetahuan Alam (MIPA) dimana kuota peserta didiknya ditentukan berdasarkan banyaknya guru pengajar bidang studi. Namun peserta didik juga diperkenankan memilih mata pelajaran lintas minat dan/atau pendalaman minat. Calon peserta didik di sekolah ini diberi 3 pilihan untuk memilih jurusan yang diminatinya.

Pandemi covid-19 yang mengancam hingga seluruh penjuror dunia, terutama di kota Samarinda, membuat sistem penjurusan calon peserta didik SMAN 2 Samarinda kini berbeda dari tahun sebelumnya. Sekolah ini tidak lagi melaksanakan ujian untuk melakukan penjurusan, namun hanya berdasarkan nilai ujian sekolah dan peminatan saja, tanpa nilai Ujian Nasional (UN). Pelaksanaan UN pada Tahun 2020 dihapus sejak diberlakukannya kebijakan pemerintah terkait UN ini. Hal ini berbeda dengan sistem penjurusan untuk tahun-tahun sebelumnya, dimana penjurusan peserta didik juga menggunakan nilai Ujian Nasional sebagai pertimbangan dalam sistem penjurusan.

Kondisi hasil penjurusan yang dilakukan oleh sekolah dapat dimodelkan melalui metode klasifikasi. Hasil prediksi model klasifikasi dapat dijadikan bahan evaluasi kurikulum. Beberapa metode klasifikasi yang menarik untuk dibahas

adalah analisis diskriminan dan regresi logistik multinomial. Dua metode ini merupakan metode sederhana namun memiliki interpretasi yang baik ketika variabel dependennya memiliki lebih dari dua kategori. [2]

Hasil penelitian yang dilakukan oleh [3] mengenai analisis diskriminan linear untuk memodelkan penjurusan peserta didik, menunjukkan bahwa analisis diskriminan linier tidak cocok untuk data yang tidak memiliki variansi yang homogen. Hal ini dapat ditunjukkan dari matriks kovariansi sampel melalui uji Box's M. Selain itu, data sampel juga harus berdistribusi Normal Multivariat [4]. Formula prediktor linear untuk model diskriminan linear mirip dengan model regresi logistik tetapi lebih efisien ketika variabel independen berasal dari distribusi Normal. [5]

Studi mengenai penjurusan peserta didik juga dilakukan melalui penelitian oleh [6] menggunakan analisis diskriminan dan regresi logistik multinomial. Hasil penelitiannya menunjukkan bahwa regresi logistik multinomial mampu memberikan tingkat ketepatan klasifikasi yang lebih tinggi dibandingkan model analisis diskriminan. Kemampuan model regresi logistik multinomial sudah cukup baik dalam mengelompokkan peserta didik ke dalam jurusannya dengan ketepatan klasifikasi sebesar 70.6%.

2. Metoda Penelitian

Penelitian ini merupakan penelitian kuantitatif yang bertujuan untuk membandingkan ketepatan hasil prediksi penjurusan dari model analisis diskriminan dan model regresi logistik multinomial. Data diambil dari nilai ujian sekolah dan hasil penjurusan peserta didik kelas X di Sekolah Menengah Atas Negeri (SMAN) 2 Samarinda. Populasi penelitian ini adalah data peserta didik kelas X tahun ajaran 2020/2021 yaitu sejumlah 357 peserta didik, dengan semua anggota populasi digunakan sebagai sampel. Sehingga teknik sampel yang digunakan adalah sampling jenuh, agar penelitian ini dapat menggambarkan seluruh keadaan sistem penjurusan peserta didik. Variabel independen dalam penelitian ini

adalah nilai ujian sekolah yaitu Bahasa Indonesia atau disingkat INDO (X_1), Bahasa Inggris atau disingkat EN (X_2), Matematika atau disingkat MTK (X_3), Ilmu Pengetahuan Alam atau disingkat IPA (X_4), dan Ilmu Pengetahuan Sosial atau disingkat IPS (X_5).

Tahapan analisis yang dilakukan dalam penelitian ini adalah statistika deskriptif, pembentukan model analisis diskriminan, pembentukan model regresi logistik multinomial, penentuan model terbaik, dan interpretasi model terbaik. Pengujian asumsi kenormalan dan homogenitas matriks kovariansi dilakukan ketika pembentukan model analisis diskriminan. Performa model diuji dengan melihat keakuratan prediksi. Proporsi pembagian data training dan data testing yang diterapkan dalam penelitian ini adalah 60:40, 70:30, 80:20, dan 90:10 dengan *resampling bootstrap* $B=1000$. Taraf kesalahan yang ditolerir adalah sebesar 5%. Piranti lunak yang digunakan adalah R 4.0.2.

3. Hasil Penelitian

Data persentase peminatan calon peserta didik di SMAN 2 Samarinda disajikan dalam Tabel 1. Nampak bahwa jurusan yang sangat diminati oleh calon peserta didik adalah MIPA, lalu IPS, lalu Bahasa dan budaya. Hal ini dikarenakan peminatan untuk pilihan ke 1 mayoritas adalah jurusan MIPA sebanyak 64,71%. Kemudian peminatan untuk pilihan ke 2 didominasi oleh jurusan IPS sebanyak 58,54%. Lalu jurusan Bahasa dan budaya paling banyak diminati untuk pilihan ke 3 atau sebesar 71,99%.

Tabel 1. Peminatan Peserta Didik di SMAN 2 Samarinda.

Pilihan	MIPA	IPS	Bahasa dan budaya
1	64,71%	29,13%	6,16%
2	20,17%	58,54%	21,29%
3	15,12%	12,89%	71,99%

Persentase hasil penjurusan yang dilakukan oleh sekolah disajikan dalam Tabel 2. Hal ini berdasarkan dalam hasil prioritas jurusan dari pilihan peserta didik, yaitu paling banyak adalah jurusan MIPA, lalu disusul jurusan IPS, dan terakhir adalah jurusan Bahasa dan budaya. Kuota kelas untuk

penjurusan ini juga mempertimbangkan banyaknya tenaga pengajar di SMAN 2 Samarinda.

Tabel 2. Hasil Penjurusan dari Sekolah.

MIPA	IPS	Bahasa dan budaya
59,95%	30,25%	9,80%

3.1. Metode Diskriminan

Sebelum melakukan analisis diskriminan, dilakukan pengujian untuk mengetahui apakah data berdistribusi Normal Multivariat dan matriks kovariansinya bersifat homogen. Pengujian kenormalan data menggunakan uji Shapiro-Wilk, sedangkan homogenitas matriks kovariansi menggunakan uji Box's M. Adapun statistik uji dan P value untuk setiap pengujian tersebut disajikan ke dalam Tabel 3. Berdasarkan nilai P value untuk kedua pengujian masih kurang dari 5%, maka analisis diskriminan linear kurang baik digunakan sehingga penelitian ini menggunakan analisis diskriminan kuadrat sebagai analisis alternatifnya.

Tabel 3. Pengujian Asumsi Kenormalan dan Homogenitas Matriks Kovariansi.

Shapiro-Wilk		Box's M	
Statistik uji	P value	Statistik uji	P value
0.51557	<2.2e-16	94.364	1.41e-08

Sifat non linear yang melekat pada diskriminan kuadratik membuat model ini tidak dapat diinterpretasi melalui proyeksi. Bentuk kuadratik pada model diskriminan ini ditunjukkan pada pembatas daerah klasifikasi antar dua nilai ujian mata pelajaran yang dibentuk dari fungsi polinom berderajat dua yang dapat dilihat pada Gambar 1(a) sampai Gambar 1(j). Daerah klasifikasi jurusan MIPA, IPS, dan Bahasa dan budaya berturut-turut berwarna merah muda, biru laut, dan putih.

(a) Pasangan nilai EN dan INDO

(e) Pasangan nilai MTK dan EN

(b) Pasangan nilai MTK dan INDO

(f) Pasangan nilai IPA dan EN

(c) Pasangan nilai IPA dan INDO

(g) Pasangan Nilai IPS dan EN

(d) Pasangan nilai IPS dan INDO

(h) Pasangan nilai IPA dan MTK

(i) Pasangan nilai IPS dan MTK

(j) Pasangan nilai IPS dan IPA

Gambar 1. Visualisasi model diskriminan kuadratik

Selanjutnya, model diskriminan ini digunakan untuk memprediksi data testing sesuai pembagian data. Nilai akurasi yang diperoleh untuk model diskriminan kuadratik disajikan ke dalam Tabel 4. Nilai akurasi tertinggi terletak pada pembagian data 90:10, yaitu sebesar 0,6469.

Tabel 4. Akurasi Model Diskriminan Kuadratik Berdasarkan Pembagian Data.

Proporsi			
60:40	70:30	80:20	90:10
0,6429	0,6438	0,6430	0,6469

3.2. Model Regresi Logistik Multinomial

Model selanjutnya yang digunakan sebagai model perbandingan adalah model regresi logistik multinomial. Fungsi logit untuk jurusan IPS adalah

$$\frac{P(Y = IPS)}{P(Y = BdB)} = -16,63 + 0,01x_1 - 0,10x_2 + 0,22x_3 - 0,08x_4 + 0,18x_5$$

dan fungsi logit untuk jurusan MIPA adalah

$$\frac{P(Y = MIPA)}{P(Y = BdB)} = -36,77 + 0,04x_1 - 0,07x_2 + 0,31x_3 + 0,09x_4 + 0,11x_5$$

dengan BdB menyatakan jurusan Bahasa dan budaya sebagai atribut dasar yang dipilih oleh piranti lunak R. Hasil pengujian model menggunakan uji Hosmer and Lemeshow disajikan ke dalam Tabel 5. Dengan keyakinan 95% bahwa model regresi logistik sudah mampu menjelaskan data penjurusan yang ditunjukkan dengan nilai *P value* lebih dari 5%.

Tabel 5. Uji Hosmer-Lemeshow.

Statistik uji	<i>P value</i>
26,157	0,05186

Interpretasi model regresi logistik ini lebih mudah menggunakan *odds ratio* yang disajikan ke dalam Tabel 6. Semakin tinggi nilai mata pelajaran Bahasa Indonesia, Matematika, dan IPS, maka hasil penjurusannya cenderung tidak pada jurusan Bahasa dan budaya yang ditunjukkan dari nilai *odds ratio* yang lebih dari 1. Namun semakin tinggi nilai Bahasa Inggris, maka hasil penjurusannya cenderung terletak pada jurusan Bahasa dan budaya. Perbedaan antara jurusan IPS dan jurusan MIPA dibanding jurusan Bahasa dan budaya, nampak pada nilai mata pelajaran IPA. Semakin rendah nilai IPA, maka hasil penjurusannya cenderung pada jurusan IPS dibanding jurusan Bahasa dan budaya. Sedangkan semakin tinggi nilai IPA, maka hasil penjurusannya cenderung pada jurusan MIPA dibanding jurusan Bahasa dan budaya. Dengan demikian, jika jurusan Bahasa dan budaya sebagai atribut dasarnya, maka jurusan ini cenderung diisi oleh peserta didik yang memiliki nilai Bahasa Inggris yang tinggi namun nilai IPA yang rendah.

Tabel 6. Odds Ratio Terhadap Atribut Jurusan Bahasa dan Budaya.

Jurusan	Nilai mata pelajaran				
	INDO	EN	MTK	IPA	IPS
IPS	1.01	0.90	1.23	0.92	1.20
MIPA	1.04	0.93	1.37	1.09	1.11

Selanjutnya, model regresi logistik multivariat ini digunakan untuk memprediksi data testing. Nilai akurasi tertinggi untuk memprediksi data testing ketika menggunakan model regresi logistik multinomial dapat dilihat pada Tabel 7. Nilai akurasi tertinggi untuk model ini terletak pada proporsi 90:10, yaitu sebesar 0,6769.

Tabel 7. Akurasi Model Regresi Logistik Multinomial Berdasarkan Pembagian Data.

Proporsi			
60:40	70:30	80:20	90:10
0,6678	0,6706	0,6740	0,6769

4. Kesimpulan

Berdasarkan hasil analisis penelitian, maka model regresi logistik multinomial merupakan model terbaik yang menggambarkan kondisi penjurusan peserta didik di SMAN 2 Samarinda. Hal ini dikarenakan regresi logistik multinomial memiliki tingkat akurasi lebih tinggi untuk setiap kemungkinan dalam pembagian proporsi data. Dimana nilai akurasi tertinggi untuk model ini adalah sebesar 0,6769 yang terletak pada proporsi 90:10. Dengan demikian, semakin tinggi nilai mata pelajaran Bahasa Indonesia, Matematika, dan IPS, maka hasil penjurusannya cenderung tidak pada jurusan Bahasa dan budaya. Namun semakin tinggi nilai Bahasa Inggris, maka hasil penjurusannya cenderung terletak pada jurusan Bahasa dan budaya. Perbedaan antara jurusan IPS dan jurusan MIPA dibanding jurusan Bahasa dan budaya, nampak pada nilai mata pelajaran IPA.

5. Saran

Diperlukan adanya penelitian lebih lanjut mengenai analisis diskriminan non linear lainnya seperti model diskriminan kubik, sehingga dapat meningkatkan akurasi prediksi dari data testing.

6. Daftar Pustaka

- [1] Permendikbud Nomor 36 Tahun 2018, Perubahan Atas Peraturan Menteri Pendidikan Dan Kebudayaan Nomor 59 Tahun 2014 Tentang Kurikulum 2013 Sekolah Menengah Atas/Madrasah Aliyah, 2018.
- [2] I. Narsky, F. C. Porter, *Statistical Analysis Techniques in Particle Physics: Fits, Density Estimation and Supervised Learning*. New York: John Wiley & Sons, 2004.
- [3] N. A. Rizki, Wasono, Y. Nasution, "The Exploring of Student's School Majoring Data at Madrasah Aliyah Negeri in Samarinda Using Linear Discriminant Analysis Models," in *The 2nd International Conference on Mathematics, Science and Computer Science*, 2019, pp. 1–8.
- [4] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*. New York: John Wiley & Sons, 2004.
- [5] A. Agresti, *An Introduction to Categorical Data Analysis*. New York: John Wiley & Sons, 2019.
- [6] N. Ponto, "Pengkajian Pembentukan Model Klasifikasi Dalam Pengelompokkan Jurusan Siswa Di SMA", in *Jurnal Matematika Dan Pembelajarannya*, Vol. 2 No. 1, 2016