

PERBANDINGAN METODE *BINARY LOGISTIC REGRESSION* DAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)* PADA PEMINATAN JURUSAN SMA (SMA NEGERI 1 SALEM)

Fitri Inriyani¹, Rochdi Wasono², Indah Manfaati Nur³

123Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

e-mail : inriyani69@gmail.com

ABSTRAK

Peminatan di Sekolah Menengah dimaksudkan untuk memberikan kesempatan terbuka bagi siswa untuk memilih mata pelajaran yang menarik dan mengembangkan potensi mereka sesuai dengan kemampuan, minat, bakat, dan kepribadian. Spesialisasi besar di sekolah menengah dipengaruhi oleh beberapa faktor. Untuk mendeteksi faktor-faktor tersebut, digunakan metode *Binary Logistic Regression* dan *Multivariate Adaptive Regression Spline (MARS)*. *Binary Logistic Regression* adalah metode yang menggambarkan hubungan antara variabel dependen dan beberapa variabel independen, dengan variabel independennya telah dikodekan 1 sebagai mewakili IPA, dan 0 sebagai mewakili IPS. Setelah dilakukan analisis Dapat diketahui ketepatan klasifikasi menggunakan metode MARS lebih tinggi dibandingkan *Binary Logistic Regression*. Dimana persentase ketepatan klasifikasi MARS adalah sebesar 76,4%, sedangkan pada ketepatan klasifikasi *Binary Logistic Regression* adalah sebesar 64,4%. Hal ini menunjukkan bahwa metode MARS lebih cocok digunakan pada kasus Peminatan Jurusan SMA dibandingkan dengan metode *Binary Logistic Regression*.

Kata Kunci: Peminatan Jurusan, Metode *Binary Logistic Regression*, Metode *Multivariate Adaptive Regression Spline (MARS)*, Ketepatan Klasifikasi.

ABSTRACT

Specialization in Middle School is intended to provide open opportunities for students to choose interesting subjects and develop their potential according to their abilities, interests, talents, and personality. Major specialization in secondary schools is influenced by several factors. To detect these factors, the *Binary Logistic Regression* and *Multivariate Adaptive Regression Spline (MARS)* methods are used. *Binary Logistic Regression* is a method that describes the relationship between the dependent variable and several independent variables, with the independent variable being coded 1 as representing IPA, and 0 as representing IPS. After analysis, it can be seen that the accuracy of classification using the MARS method is higher than the *Binary Logistic Regression*. Where the percentage of MARS classification accuracy is 76.4%, while the accuracy of the *Binary Logistic Regression* classification is 64.4%. This shows that the MARS method is more suitable in the case of Specialization in the Department of High School compared to the *Binary Logistic Regression* method.

Keywords: Department Specialization, *Binary Logistic Regression*, *Multivariate Adaptive Regression Spline (MARS)* Method, Classification Accuracy.

PENDAHULUAN Pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif membangun potensi dirinya untuk memiliki kekuatan spiritual

keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat bangsa dan negara (Undang-Undang Republik Indonesia Nomor 20 Tahun 2003, Pasal 1 Ayat 1). Di Indonesia

pendidikan dapat dibedakan berdasarkan jalur, jenjang dan jenisnya. Jalur pendidikan dibagi menjadi tiga yaitu jalur formal atau pendidikan yang diselenggarakan di sekolah-sekolah pada umumnya, jalur nonformal atau pendidikan yang didapat dari luar seperti kursus dan pendidikan anak usia dini, serta jalur informal atau jalur pendidikan keluarga dan lingkungan berbentuk kegiatan belajar secara pribadi dan mandiri. Jenjang pendidikan jalur formal ada tiga yaitu pendidikan dasar, pendidikan menengah dan pendidikan tinggi. Sedangkan jika dilihat dari jenisnya, jenis pendidikan mencakup pendidikan umum, kejuruan, akademik, profesi, vokasi, keagamaan dan khusus. Hal ini dijelaskan dalam Undang-undang Republik Indonesia Nomor 20 Tahun 2003.

Sekolah menengah Atas (SMA) merupakan pendidikan umum jalur formal jenjang sekolah menengah yang biasa ditempuh dalam waktu 3 tahun. Pada jenjang sekolah menengah atas peserta didik akan mengikuti peminatan jurusan. Menurut Sudiarto (2013), tujuan diadakannya peminatan jurusan adalah untuk memberikan peluang yang lebih terbuka kepada peserta didik untuk memilih mata pelajaran yang diminati. Mendalami mata pelajaran dan mengembangkan berbagai potensi yang dimilikinya secara fleksibel sesuai dengan kemampuan dasar umum (kecerdasan), bakat, minat dan karakteristik kepribadian tanpa dibatasi dengan sekat-sekat penjurusan yang terlalu kaku. Peminatan jurusan di SMA sangat penting karena dapat menjadi landasan dan acuan kedepan dalam meraih cita-cita serta dalam memilih jurusan di bangku perkuliahan. Umumnya pilihan minat jurusan di SMA ada tiga yaitu Ilmu Pengetahuan Alam, Ilmu Sosial, Bahasa.

Ada beberapa faktor yang mempengaruhi peserta didik dalam memilih jurusan. Menurut Syah (2009), faktor yang mempengaruhi antara lain faktor internal dan faktor eksternal. Faktor internal yaitu faktor yang datang dari dalam diri siswa itu sendiri, sedangkan faktor eksternal adalah faktor yang datang dari luar diri siswa tersebut. Untuk mengetahui faktor apa saja yang mempengaruhi peserta didik dalam pemilihan jurusan, maka perlu diadakan sebuah penelitian.

SMA Negeri 1 Salem adalah satu-satunya SMA Negeri di Kecamatan Salem Kabupaten Brebes. Terdapat dua jurusan di SMA Negeri 1 Salem yakni jurusan Ilmu pengetahuan Alam dan Jurusan Ilmu Pengetahuan Sosial. Jumlah peserta didik SMA Negeri 1 Salem kelas X sebanyak 224

siswa, kelas sebanyak XI 281 siswa, kelas XII sebanyak 257 siswa.

Dalam ilmu statistika, banyak metode yang dapat digunakan untuk mengetahui pengaruh variabel prediktor terhadap variabel responnya yang bersifat kategorik diantaranya adalah metode regresi Logistik Biner dan *Multivariate Adaptive Regression Spline* (MARS). Kedua metode tersebut termasuk dalam analisis regresi. Menurut Montgomery dan Peck (1992) analisis regresi adalah suatu analisis statistika yang memanfaatkan hubungan antara dua variabel atau lebih. Umumnya analisis ini digunakan untuk melihat hubungan dan pengaruh variabel prediktor terhadap variabel respon dimana variabel responnya berupa data kuantitatif. Namun untuk respon yang bersifat kualitatif/kategorik dapat dianalisis menggunakan metode logistik. Hosmer dan Lemeshow (2000) mengatakan bahwa metode regresi logistik adalah suatu metode analisis statistik yang mendeskripsikan hubungan antara variabel respon yang memiliki dua kategori atau lebih dan salah satu atau lebih variabel prediktor. Salah satu model regresi logistik adalah regresi logistik biner. Model regresi logistik biner merupakan metode regresi logistik yang digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel prediktor, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik.

Metode statistika lainnya yang dapat digunakan juga yaitu pengklasifikasian objek dengan melibatkan variabel respon dan kategori dengan sejumlah variabel prediktor kontinu ataupun kategori adalah metode *Multivariate Adaptive Regression Spline* (MARS). MARS merupakan salah satu model regresi nonparametrik multivariat yang dikembangkan oleh Friedman 1991.

Metode ini merupakan pengembangan dari pendekatan *Recursive Partitioning Regression* (RPR) dengan metode *Spline* (Friedman 1991). Model MARS berguna untuk mengatasi data berdimensi tinggi dan menghasilkan prediksi variabel respon yang akurat serta menghasilkan model kontinu dalam *knot* berdasarkan nilai *Generalized Cross Validation* (GCV) terkecil. Perbandingan metode regresi logistik biner dengan MARS juga sudah pernah dilakukan sebelumnya pada penelitian yang berjudul "*Ketetapan Klasifikasi Dengan Analisis Regresi Logistik dan Multivariate Adaptive Regression*

Spline (MARS) Pada Data Dengan Peubah Data Biner". (Fitrianty et al. 2013) dan "*Pemodelan terhadap Kelulusan siswa masuk Akselerasi Menggunakan Analisis Regresi Logistik biner dan Multivariate Adaptive Regression Spline (MARS)*". (Mandaku dan Mandaku, 2010).

Sehubungan dengan penelitian ini, maka variabel yang diduga mempengaruhi peserta didik dalam peminatan jurusan adalah terdiri dari faktor internal yang meliputi nilai UN Matematika, nilai UN IPA, nilai UN Bahasa Inggris dan dorongan dari diri sendiri, serta faktor eksternal yang meliputi hubungan siswa dengan teman, hubungan siswa dengan guru, hubungan siswa dengan keluarga. Dengan demikian, penelitian ini dibuat dengan judul "**Perbandingan Metode Binary Logistic Regression dan Multivariate Adaptive Regression Spline (MARS) pada peminatan Jurusan SMA (Studi Kasus SMA Negeri 1 Salem)**".

TINJAUAN PUSTAKA

1. Model Regresi Logistik Biner

Analisis regresi logistik biner digunakan untuk menjelaskan hubungan antara variabel respon yang berupa data dikotomik/biner dengan variabel bebas yang berupa data berskala interval dan atau kategorik.

Regresi Logistik biner (*logistic regression*) sebenarnya sama dengan regresi berganda, hanya variabel terkaitnya merupakan variabel dummy (0 dan 1). Seperti pada penelitian ini variabel terkaitnya adalah 0 jika memilih penjurusan IPA dan 1 jika memilih penjurusan IPS. Tidak seperti *regresi logistik* biasa, *regresi logistik* biner tidak mengansumsikan hubungan antara variabel independen dan dependen secara linier.

Model yang digunakan pada *regresi logistik* biner adalah :

$$\log(P/1-p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

1.2 Estimasi Parameter

Menurut Hosmer dan Lemeshow (2000), metode estimasi yang digunakan untuk menaksir parameter pada model regresi logistik adalah metode *Maximum Likelihood Estimation* (MLE). Pada dasarnya, metode maksimum likelihood menghasilkan nilai parameter yang memaksimalkan probabilitas data observasi.

1.3 Uji Signifikansi Parameter

Menurut Hosmer dan Lemeshow (2000), uji signifikansi parameter yang digunakan adalah uji Rasio Likelihood dan uji Wald.

a. Uji Rasio Likelihood

Uji Rasio Likelihood adalah uji signifikansi parameter secara keseluruhan atau bersama-sama.

Hipotesis :

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

$$H_1 : \text{minimal ada satu } \beta_j \neq 0$$

dengan $j = 1, 2, \dots, p$.

Statistik Uji : $G = -2 \ln$

Kriteria Uji : H_0 ditolak jika $G > \chi^2_{(\alpha; p)}$

b. Uji Wald

Uji wald digunakan untuk mengetahui apakah masing-masing variabel prediktornya memiliki pengaruh terhadap model atau tidak.

Hipotesis

$$H_0 : \beta_j \neq 0$$

$$H_1 : \beta_j \neq 0, \text{ untuk } j = 1, 2, \dots, p.$$

$$\text{Statistik Uji : } W = \left(\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right)^2$$

Kriteria Uji : H_0 ditolak jika $W > \chi^2_{(\alpha; 1)}$

2. Multivariate Adaptive Regression Spline (MARS)

Multivariate Adaptive Regression Spline (MARS) adalah salah satu regresi nonparametrik yang mengkombinasi regresi *spline* dan *Recursive Partitioning regression* (RPR) yang pertama kali diperkenalkan oleh Friedman tahun 1991. Menurut Friedman (1991), regresi *spline* menghasilkan persamaan bentuk parametrik polinomial tersegmen (terbagi dalam beberapa *region*), sedangkan RPR merupakan salah satu pendekatan komputasi yang digunakan untuk data berdimensi tinggi. Itu sebabnya, metode MARS dapat digunakan pada data yang berdimensi tinggi yaitu data yang memiliki variabel prediktor $3 \leq p \leq 20$ dan memiliki ukuran sampel yang berukuran $50 \leq n \leq 2000$. Variabel respon yang diolah pada metode MARS pun dapat berbentuk kontinu atau biner (Kriener, 2007).

Menurut Nash dan Bradford (2001), beberapa hal yang perlu diperhatikan dalam menggunakan metode MARS antara lain *knot* dan fungsi basis. *Knot* adalah sebuah titik yang memisahkan akhir suatu wilayah data dengan awal suatu wilayah data yang lain. *Knot* pada MARS dipilih menggunakan *forward stepwise* dan *backward stepwise*. Penempatan *knot* tergantung pada penentuan banyaknya amatan antar *knot*. Banyaknya amatan/observasi pada masing-masing *knot* disebut sebagai *Minimum Observation* (MO).

MO yang digunakan adalah 0, 1, 2, dan 3. *Basis Function* (BF) adalah suatu fungsi yang dipisahkan oleh titik-titik *knot* yang menjelaskan hubungan antara variabel prediktor dan variabel respon. Metode MARS membentuk fungsi basis dengan prosedur *forward stepwise* dan *backward*. Friedman (1991) menyatakan bahwa jumlah BF adalah 2 sampai dengan 4 kali jumlah variabel prediktor, sedangkan jumlah *Maximum Interaction* (MI) adalah 1, 2 dan 3 dengan pertimbangan jika $MI > 3$ akan menghasilkan model yang kompleks dan interpretasinya akan hampir sama.

Menurut Friedman (1991), model untuk metode MARS adalah sebagai berikut:

$$\hat{f}(x) = \hat{a}_0 + \sum_{m=1}^M \hat{a}_m \prod_{k=1}^{K_m} [S_{km}(x_{j(k,m)} - t_{km})]_+$$

2.1 Model Terbaik

Menurut Friedman (1991), model terbaik pada MARS ditentukan berdasarkan kriteria *Generalized Cross Validation* (GCV) minimum yang didefinisikan sebagai berikut:

$$GCV(M) = \frac{\frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{\tilde{C}(M)}{n}\right]^2}$$

3. Klasifikasi

Menurut Hosmer dan Lemeshow (2000), langkah awal klasifikasi dari variabel respon biner adalah menentukan titik potong. Variabel respon yang memiliki dua kategori (biner) dapat digunakan titik potong sebesar 0,50 dengan ketentuan jika $\pi(x) \geq 0,50$, maka hasil prediksi adalah 1 dan jika $\pi(x) < 0,50$ maka hasil prediksinya adalah 0. Klasifikasi pada pendekatan analisis regresi logistik dan MARS dapat menggunakan model probabilitas $\pi(x)$, yaitu:

Untuk regresi logistik Biner

$$\pi(x_i) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}}}$$

Untuk MARS

$$\pi(x_i) = \frac{e^{\hat{a}_0 + \sum_{m=1}^M \hat{a}_m \prod_{k=1}^{K_m} [S_{km}(x_{j(k,m)} - t_{km})]_+}}{1 + e^{\hat{a}_0 + \sum_{m=1}^M \hat{a}_m \prod_{k=1}^{K_m} [S_{km}(x_{j(k,m)} - t_{km})]_+}}$$

4. Evaluasi Ketepatan Klasifikasi

Pengukuran klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*), seperti berikut:

Tabel 2.1. Matrik Konfusi Klasifikasi Dua

Hasil Observasi	Kelas	
	Prediksi	
	y_0	y_1
y_0	f_{00}	f_{01}
y_1	f_{10}	f_{11}

Dengan :

f_{00} = jumlah objek pengamatan dari y_0 dan diklasifikasikan sebagai y_0

f_{01} = jumlah objek pengamatan dari y_0 dan diklasifikasikan sebagai y_1

f_{10} = jumlah objek pengamatan dari y_1 dan diklasifikasikan sebagai y_0

f_{11} = jumlah objek pengamatan dari y_1 dan diklasifikasikan sebagai y_1

5. Apparent Error Rate (APER) dan Total Accuracy Rate (TAR)

Menurut Johnson dan Wichern (1992), *Apparent Error Rate* (APER) adalah prosedur evaluasi yang digunakan untuk melihat kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. Nilai APER menunjukkan nilai proporsi sampel yang salah diklasifikasikan pada fungsi klasifikasi.

Total Accuracy Rate (TAR) digunakan untuk menghitung ketepatan klasifikasi pada hasil pengelompokan. Nilai TAR dapat menyatakan representasi proporsi sampel yang tepat diklasifikasi-kan.

METODE PENELITIAN

1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data primer yang diperoleh dengan membagikan kuisioner pada siswa SMA Negeri 1 Salem. (2.9)

2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah nilai UN Matematika (X_1), nilai UN IPA (X_2), nilai UN Bahasa Inggris (X_4), hubungan siswa dengan teman (X_5), hubungan siswa dengan guru (X_6), hubungan siswa dengan keluarga (X_7).

3. Langkah-langkah penelitian

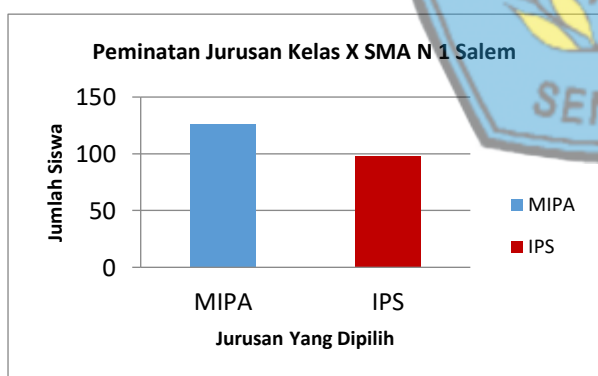
Adapun langkah-langkah penelitian ini adalah:

1. Mengumpulkan data
2. Analisis Statistika Deskriptif untuk mengetahui karakteristik dari siswa SMA Negeri 1 Salem
3. Uji Validitas dan Reliabilitas
4. Melakukan analisis Binary Logistic Regression dengan menggunakan Software R
5. Melakukan uji signifikansi Parameter
6. Menentukan model akhir dari variabel yang paling berpengaruh
7. Melakukan uji kesesuaian model
8. Klasifikasi Binary Logistic Regression
9. Menguji Ketepatan Klasifikasi
10. Melakukan analisis MARS menggunakan Software R
11. Mendapatkan model MARS terbaik untuk data berdasarkan nilai GCV terkecil
12. Mendapatkan variabel yang signifikan dari model MARS
13. Mengklasifikasikan ke dalam bentuk matriks konfusi
14. Menguji ketepatan klasifikasi

HASIL DAN PEMBAHASAN

1. Deskriptif Data

Deskripsi data dilakukan untuk mengetahui informasi awal secara umum dari data yang digunakan. Deskripsi yang dilakukan yaitu melihat jumlah peminatan jurusan kelas X di SMA Negeri 1 Salem.



Berdasarkan Gambar dapat diinterpretasikan hasil statistik deskriptif adalah siswa yang memilih jurusan MIPA sebanyak 126 siswa dan yang memilih jurusan IPS sebanyak 98 siswa.

2. Uji Validitas

Dalam penelitian ini penulis menguji validitas dan reabilitas item-item pertanyaan menggunakan Software R sehingga tidak memperhatikan item pertanyaan yang tidak valid

dan hanya menggunakan item pertanyaan yang valid. Dalam menentukan validitas daftar pertanyaan peneliti menggunakan ketentuan perbandingan nilai raw_alpha dengan r tabel ($n =$ dimana $df=n-2$, $\alpha=5\%$) Jika $raw_alpha > r$ tabel maka item pertanyaan valid. Dimana tabel diperoleh dari $df=n-2$ dan $\alpha = 5\%$ hasilnya sebesar 0.311.

Hasil uji validitas menunjukkan semua variabel nilai ujian matematika (X_1), nilai ujian IPA (X_2), nilai ujian bahasa Inggris (X_3), hubungan siswa dengan teman (X_4), hubungan siswa dengan guru (X_5), hubungan siswa dengan keluarga (X_6), motivasi dari diri sendiri (X_7) adalah Valid.

3. Uji Reliabilitas

Pengujian reliabilitas dilakukan dengan menggunakan software R maka dengan melihat nilai raw_alpha pada output software R. dengan jumlah sampel 224 responden yang diteliti. Suatu instrumen penelitian dinyatakan reliabel apabila nilai $raw_alpha > 0.60$. Hasil uji reliabilitas

Berdasarkan hasil uji Reliabilitas semua variabel penelitian tersebut dinyatakan semua variabel yang digunakan dalam penelitian adalah reliabel karena nilai ($raw_alpha = 0.87$) > 0.60 .

4. Pembentukan Model Regresi Logistik Biner

Binary logistic regression dilakukan untuk mengetahui pengaruh hubungan antar variabel peminatan jurusan terhadap variabel nilai ujian matematika (X_1), nilai ujian IPA (X_2), nilai ujian bahasa Inggris (X_3), hubungan siswa dengan teman (X_4), hubungan siswa dengan guru (X_5), hubungan siswa dengan keluarga (X_6), motivasi dari diri sendiri (X_7).

4.1 Uji Signifikansi Parameter Secara Serentak

Pada analisis *binary logistic regression* yang pertama dilakukan adalah melakukan pengujian estimasi parameter secara serentak, yang digunakan untuk mengetahui pengaruh semua variabel independen secara keseluruhan terhadap peminatan jurusan.

<i>P-Value</i>	α
0.0002427	0.1

Berdasarkan Tabel 4.1 dapat diketahui ($P-Value = 0.0003114$) $< (\alpha = 0.1)$ sehingga diperoleh keputusan tolak H_0 minimal ada satu variabel nilai UN Matematika (X_1), nilai UN IPA (X_2), nilai UN Bahasa Inggris (X_3),

hubungan siswa dengan teman (X_4), hubungan siswa dengan guru (X_5), hubungan siswa dengan keluarga (X_6), motivasi diri sendiri (X_7) yang signifikan terhadap peminatan jurusan (Y).

4.2 Uji Signifikansi Parameter Secara Parsial

Setelah dilakukan pengujian estimasi parameter secara serentak, maka dilanjutkan dengan pengujian estimasi parameter secara parsial untuk mengetahui pengaruh dari masing-masing variabel independen terhadap peminatan jurusan. Sampai didapatkan dilakukan model akhir menggunakan variabel yang berpengaruh pada model. Model akhir yang didapat adalah:

$$\hat{g}(x) = 0.1418 + 0.0809(X_2) + 0.0360(X_{7(2)})$$

5. Klasifikasi Regresi Logistik Biner

Hasil pengklasifikasian regresi logistik biner adalah :

Tabel 4.4 Hasil Klasifikasi *Binary Logistic Regression*

Hasil Observasi	Hasil Prediksi	
	Kelas 0 (Jurusan MIPA)	Kelas 1 (Jurusan IPS)
Kelas 0 (Jurusan MIPA)	38	60
Kelas 1 (jurusan IPS)	16	110

Pada Tabel terlihat bahwa pada hasil klasifikasi peminatan jurusan SMA menggunakan metode *Binary Logistic Regression* terdapat 108 siswa jurusan IPS yang secara benar dipetakan ke kelas IPS, 18 adalah jumlah data kelas IPS yang dipetakan secara salah ke kelas MIPA, 62 adalah jumlah data kelas MIPA yang dipetakan secara salah ke kelas IPS, 36 adalah jumlah data kelas MIPA yang dipetakan secara salah ke kelas IPS.

6. Menuji Ketepatan Klasifikasi

Setelah mengetahui tabel klasifikasi dari metode *Binary Logistic Regression* maka selanjutnya menghitung nilai *APER* dan *TAR* sebagai berikut :

$$APER(\%) = \frac{16 + 60}{224} \times 100\% = 33.92\%$$

$$TAR(\%) = 1 - APER = 1 - 33.92\% = 66.08\%$$

Jadi berdasarkan pada nilai *APER* tingkat kesalahan klasifikasi adalah sebesar 35.71%, sehingga kesuksesan pengklasifikasian sebesar 64.29% berdasarkan nilai *TAR*.

7. Pemodelan MARS

Untuk mendapatkan model terbaik, perlu dilakukan *trial and error* terhadap kombinasi Basis Function (BF), *Maximum Interaction* (MI) dan *Minimum Observation* (MO) dengan melihat nilai GCV terkecil, didapat hasil kombinasi BF, MI dan MO yang menghasilkan nilai GCV terkecil adalah kombinasi BF = 28, MI = 1 dan MO = 2. dengan melihat nilai GCV terkecil yang dihasilkan menggunakan software R sebesar 0.2134

dengan model yang terbentuk adalah :

$$\hat{f}(x) = -0.446 - 0.016 * BF1 + 0.080 * BF2 + 0.071 * BF3 - 0.0130 * BF4 - 0.146 * BF5 + 0.543 * BF6 - 0.443 * BF7 + 1.146 * BF8 - 1.671 * BF9 + 0.160 * BF10 - 0.558 * BF11 - 0.187 * BF12$$

8. Klasifikasi Model MARS

Dengan model probabilitas yang telah terbentuk, maka didapatkan hasil klasifikasi untuk metode MARS yaitu :

Hasil Observasi	Hasil Prediksi	
	Kelas 0 (Jurusan MIPA)	Kelas 1 (Jurusan IPS)
Kelas 0 (Jurusan MIPA)	60	38
Kelas 1 (jurusan IPS)	15	111

Pada Tabel 4.6 terlihat bahwa pada hasil klasifikasi peminatan jurusan SMA menggunakan metode MARS terdapat 111 siswa jurusan IPS yang secara benar dipetakan ke kelas IPS, 15 adalah jumlah data kelas IPS yang dipetakan secara salah ke kelas MIPA, 38 adalah jumlah data kelas MIPA yang dipetakan secara salah ke kelas IPS, 60 adalah jumlah data kelas MIPA yang dipetakan secara salah ke kelas IPS.

9. Menguji Ketepatan Klasifikasi MARS

Setelah mengetahui tabel klasifikasi dari metode MARS maka selanjutnya menghitung nilai *APER* dan *TAR* sebagai berikut :

$$APER(\%) = \frac{15 + 38}{224} \times 100\% = 23,6\%$$

$$TAR(\%) = 1 - APER = 1 - 23,6\% = 76,4\%$$

Jadi berdasarkan pada nilai *APER* tingkat kesalahan klasifikasi adalah sebesar 23,6%, sehingga kesuksesan pengklasifikasian sebesar 76,4% berdasarkan nilai *TAR*.

10. Perbandingan Ketepatan Klasifikasi *Binary Logistic Regression* dan MARS

Setelah mendapatkan model dan tabel klasifikasi dari kedua metode maka selanjutnya membandingkan nilai *TAR*, Perbandingannya dapat dilihat dari Tabel 4.8 berikut :

Tabel 4.7 Perbandingan Nilai *APER* dan *TAR* metode *Binary Logistic Regression* dan MARS

	<i>Binary Logistic Regression</i>	MARS
APER	33.92 %	23.6%
TAR	66.08 %	76.4%

Dari tabel dapat kita lihat ketepatan klasifikasi menggunakan metode MARS lebih tinggi dibandingkan *Binary Logistic Regression*. Dimana persentase ketepatan klasifikasi MARS adalah sebesar 76,4%, sedangkan pada ketepatan klasifikasi *Binary Logistic Regression* adalah sebesar 64,4%. Hal ini menunjukkan bahwa metode MARS lebih cocok digunakan pada kasus Peminatan *Jurusan SMA* dibandingkan dengan metode *Binary Logistic Regression*.

KESIMPULAN

1. Pemodelan Peminatan *Jurusan di SMA Negeri 1 Salem* adalah sebagai berikut :

- Pemodelan dengan menggunakan *Binary Logistic Regression* menghasilkan model:

$$\hat{g}(x) = 0.1418 + 0.0809(X_2) + 0.0360X_{7(2)}$$

- Sedangkan jika menggunakan *Multivariate Adaptive Regression Spline* menghasilkan model :

$$\hat{f}(x) = -0.446 - 0.016 * BF1 + 0.080 * BF2 + 0.071 * BF3 - 0.0130 * BF4 - 0.146 * BF5 + 0.543 * BF6 - 0.443 * BF7 + 1.146 * BF8 - 1.671 * BF9 + 0.160 * BF10 - 0.558 * BF11 - 0.187 * BF12$$

2. Dapat diketahui ketepatan klasifikasi menggunakan metode MARS lebih tinggi dibandingkan *Binary Logistic Regression*. Dimana persentase ketepatan klasifikasi MARS adalah sebesar 76,4%, sedangkan pada ketepatan klasifikasi *Binary Logistic Regression* adalah sebesar 64,4%. Hal ini menunjukkan bahwa metode MARS lebih cocok digunakan pada kasus Peminatan *Jurusan SMA* dibandingkan dengan metode *Binary Logistic Regression*.

DAFTAR PUSTAKA

- Agresti, A. 1990. *Categorical Data Analysis*. John Wiley and Son, Florida.
- Anonim. 2012. *Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 – Sisdiknas*. [online]. Tersedia: <http://www.unpad.ac.id/wp-content/uploads/2012/10/UU20-2003-Sisdiknas.pdf> (diakses 10 juli 2019).
- Ayu, Fanny. 2017. *Pemodelan Status Bekerja Ibu Rumah Tangga Menggunakan Model Multilevel dengan Respon Biner* (Tesis). Surabaya. Institut Teknologi Sepuluh Nopember
- Danardono. 2015. *Analisis Data Lomhitudinal*. Yogyakarta. Gadjah Mada University.
- Friedman, J. H. 1991. *Multivariate Adaptive Regression Splines*. *The Annals of Statistics*, Vol. 19.
- Hair J.F., Anderson R.E., Babin B.J, Black W.C. 2010. *Multivariate Data Analysis 7th Edition* Pearson Education. Prentice Hall,inc.
- Hastie, T., Tibshirani, R. And Friedman, J. H. 2009. *The Elements of Statistical Learning: Data Mining. Inference and Prediction Second Edition*. New York: Springer Science Bussiness Media.
- Hosmer, D. W., and Lemeshow S. 2000. *Applied Logistic Regression*. United States of American: Sons Inc.
- Homster, D. W., Lemeshow S. 2000. *Applied Logistic Regression. Second Edition*. New Jersey: Prentice-Hall.Inc.
- Homster, D. W., and Lemeshow, S. 1989. *Applied Logistic Regression*. John Willey, New York.
- Kishartini, et al 2014. *Multivariate Adaptive Regression Spline (MARS) Untuk Klasifikasi*

- Status Kerja Di Kabupaten Demak* (Jurnal). Universitas Diponegoro Semarang.
- Kriner, M. 2007. *Survival Analysis with Multivariate Adaptive Regression Spline* (Disertasi). Universitas Munchen, Louisiana.
- Kurniawan, Deny. 2008. *Tabel Distribusi* (Copyright). [online]. Tersedia: <https://ineddeni.wordpress.com/category/download-tabel-distribusi-statistika/> (diakses 14 September 2019)
- Merluarini, B., Safitri, D., & Hoyyi, A. 2014. Perbandingan Analisis Klasifikasi Menggunakan Metode K-Nearest Neighbor (K-NN) dan *Multivariate Adaptive Regression Spline* (MARS) pada Data Akreditasi Sekolah Dasar Negeri di Kota Semarang, *Jurnal Gaussian*, 3(3), htm. 313-332.
- Montgomery, D.C. and Peck, E.A. 1992. *Introduction to Linear Regression Analysis, 2nd Edition*. John Wiley & Sons, New York.
- Nash, M. S. and Bradford, D. F. 2001. *Parametric and Non Parametric Logistic Regression for Prediction of Precense/Absence of an Amphibian*. [online]. Tersedia: <http://www.epa.gov/esd/land-sci/pdf/0081eb02.pdf> (diakses 12 juli 2019)
- Otok, B.W, et al. 2007. *Pendekatan Bootstrep Pada Klasifikasi Pemodelan Respon Ordinal*. Universitas Gajah Mada, Yogyakarta
- Pintowati, W. dan Otok, B.W. 2012. *Pemodelan Kemiskinan Propinsi Jawa Timur dengan Pendekatan Mutivariate Adaptive*. *Jurnal Sains dan Seni ITS* Vol.1 No.1
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Penerbit ANDI Yogyakarta
- Rosadi, Dedi. 2011. *Analisis Ekonometrika & Runtun Waktu Terapan*. Yogyakarta. ANDI
- Sudiarto, A. 2013. *Pedoman Arah Peminatan Draf 2*. [online]. Tersedia: <http://www.akursudianto.com/wp-content/uploads/2013/06/pedoman-arrah-peminatan-draf-2.pdf> (diakses 15 juli 2019).
- Sugiarto, D.S. 2000. *Metode Statistika*. Jakarta: Gramedia Pustaka Utama.
- Syah, M. 2009. *Psikologi Pendidikan dengan Pendekatan Baru*. Bandung: Remaja Rosdakarya.
- Utomo, Setyo. 2009. *Model Regresi Logistik Untuk Menunjukkan Pengaruh Pendapatan Perkapita, Tingkat Pendidikan, dan Status Pekerjaan Terhadap Status Giji Masyarakat Kota Surakarta* (Skripsi). Surakarta: Universitas Sebelas Maret
- Wahyuningrum, S. 2008. *Pendekatan MARS untuk Ketetapan Klasifikasi Desa/Kelurahan Miskin di Kalimantan Timur* (Tesis). Surabaya: Institut Teknologi Sepuluh Nopember