

**PERBANDINGAN ANALISIS KLASIFIKASI
MENGUNAKAN METODE *K-NEAREST NEIGHBOR* (K-NN)
DAN *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* (MARS) PADA
DATA AKREDITASI SEKOLAH DASAR NEGERI
DI KOTA SEMARANG**

Bisri Merluarini¹, Diah Safitri², Abdul Hoyyi³

¹Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

^{2,3}Staf Pengajar Jurusan Statistika FSM Universitas Diponegoro

ABSTRACT

Classification methods have been developed and two of the existing are K-Nearest Neighbor (K-NN) and Multivariate Adaptive Regression Spline (MARS). The purpose of this research is comparing the classification of public elementary school accreditation in Semarang city with K-NN and MARS methods. This research using accreditation data with the result of eight accreditation components in public elementary school that has A accreditation (group 1) and B accreditation (group 2) in Semarang city. To evaluate the classification method used test statistic Press's Q, APER, specificity, and sensitivity. The best classification results of the K-NN method is when using $K=5$ because it produces the smallest error rate and obtained information that the correct classification data are 159 and the misclassification data are 9. The best classification result of the MARS method is when using combination $BF=32$, $MI=2$, $MO=1$ because it produces the smallest Generalized Cross Validation (GCV) and obtained information that the correct classification data are 164 and the misclassification data are 4. Based on analyze result, Press's Q showed that both methods are good as classification or statistically significant to classify the public elementary school in Semarang city based of the accreditation. APER, specificity, and sensitivity showed that classify of public elementary school accreditation in Semarang city using MARS method is better than K-NN method.

Keywords: Classification, K-Nearest Neighbor (K-NN), Multivariate Adaptive Regression Spline (MARS), Classification evaluation

1. PENDAHULUAN

Pendidikan sebagai salah satu sektor yang paling penting dalam pembangunan nasional berfungsi semaksimal mungkin dalam upaya meningkatkan kualitas hidup masyarakat. Pemerintah terus berupaya untuk meningkatkan mutu pendidikan dengan harapan peserta didik menjadi manusia yang aktif, kreatif, dan terampil memecahkan masalahnya sendiri (Ihsan, 1996).

Salah satu upaya meningkatkan mutu pendidikan adalah dengan melakukan penilaian-penilaian terhadap mutu sekolah yang dilakukan oleh Badan Akreditasi Nasional (BAN) di bawah naungan Kementerian Pendidikan Nasional dengan memberikan nilai yang merupakan hasil penilaian berbagai komponen akreditasi. Nilai dari masing-masing komponen menentukan peringkat akreditasi sekolah yang dinyatakan dalam huruf A (sangat baik), B (baik), C (cukup), dan tidak terakreditasi (Badan Akreditasi Nasional, 2009).

Ketepatan dalam pengklasifikasian objek sangat penting, metode klasifikasi yang baik adalah metode yang menghasilkan kesalahan yang kecil (Johnson dan Wichern, 2007). Saat ini metode statistika telah sangat berkembang

dan dua diantara metode klasifikasi yang telah ada yaitu metode *K-Nearest Neighbor* (K-NN) dan metode *Multivariate Adaptive Regression Spline* (MARS). Menurut Prasetyo (2012) metode *K-Nearest Neighbor* (K-NN) merupakan metode yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain, metode K-NN merupakan metode yang cukup sederhana namun memiliki tingkat akurasi yang tinggi, sedangkan menurut Friedman (1991) metode MARS merupakan metode regresi nonparametrik multivariat dengan data dimensional tinggi yang dapat diterapkan untuk pengklasifikasian suatu subjek tertentu.

Penelitian ini menggunakan metode *K-Nearest Neighbor* (K-NN) dan metode *Multivariate Adaptive Regression Spline* (MARS) untuk mengetahui metode yang lebih baik dalam pengklasifikasian akreditasi Sekolah Dasar Negeri (SDN) di kota Semarang.

2. TINJAUAN PUSTAKA

2.1 Akreditasi

Menurut Kamus Besar Bahasa Indonesia (2008) akreditasi adalah pengakuan terhadap lembaga pendidikan yang diberikan oleh badan yang berwenang setelah dinilai bahwa lembaga itu memenuhi syarat kebakuan atau kriteria tertentu.

2.2 Klasifikasi

Klasifikasi merupakan teknik multivariat untuk memisahkan set yang berbeda dari suatu objek (observasi) dan mengalokasikan objek (observasi) baru ke dalam kelompok yang telah ditetapkan sebelumnya. Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau akan menghasilkan peluang kesalahan alokasi yang kecil (Johnson dan Wichern, 2007).

Menurut Prasetyo (2012) umumnya pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi. Tabel 1 merupakan matriks konfusi yang melakukan klasifikasi biner (dua kelas) yaitu kelas 0 dan 1.

Tabel 1. Matriks Konfusi untuk Klasifikasi Dua Kelas

f_{ij}		Kelas hasil prediksi (j)	
		Kelas = 0	Kelas = 1
Kelas asli (i)	Kelas = 0	f_{00}	f_{01}
	Kelas = 1	f_{10}	f_{11}

f_{00} adalah jumlah data dalam kelas 0 yang dipetakan secara benar ke kelas 0, f_{01} adalah jumlah data dalam kelas 0 yang dipetakan secara salah ke kelas 1, f_{11} adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1, f_{10} adalah jumlah data dalam kelas 1 yang dipetakan secara salah ke kelas 0 (Prasetyo, 2012).

2.3 *K-Nearest Neighbor* (K-NN)

Menurut Prasetyo (2012) *K-Nearest Neighbor* (K-NN) adalah metode yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain. Nilai K pada K-NN berarti K-data terdekat dari data uji. Parameter jarak

yang umum digunakan adalah jarak *Euclidean*. Menurut Han dan Kamber (2006) jarak *Euclidean* antara dua titik, katakan $\mathbf{x}_1 = (x_{11}, x_{12}, \dots, x_{1n})$ dan $\mathbf{x}_2 = (x_{21}, x_{22}, \dots, x_{2n})$ adalah:

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}$$

Metode K-NN cukup sederhana, tidak ada asumsi mengenai distribusi data dan mudah diaplikasikan. Pemilihan nilai K (jumlah data/tetangga terdekat) ditentukan oleh peneliti. Pemilihan nilai K ini bisa mempengaruhi tingkat akurasi prediksi yang dikerjakan (Santosa, 2007).

2.4 *Multivariate Adaptive Regression Spline (MARS)*

Menurut Friedman (1991) metode *Multivariate Adaptive Regression Spline* (MARS) merupakan pengembangan dari metode *Recursive Partitioning Regression* (RPR). Metode MARS berguna untuk mengatasi permasalahan data berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor sebanyak $3 \leq n \leq 20$, serta memiliki ukuran sampel sebesar $50 \leq N \leq 1000$. Model dalam metode MARS adalah sebagai berikut:

$$\hat{Y} = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [S_{km} (X_{V(k,m)} - t_{km})]_+$$

dengan:

- a_0 = koefisien konstan fungsi basis
- a_m = koefisien dari fungsi basis ke- m
- M = maksimum fungsi basis
- K_m = derajat interaksi
- S_{km} = nilainya +1 dan -1
- $X_{V(k,m)}$ = variabel prediktor
- t_{km} = nilai knot dari variabel prediktor $X_{V(k,m)}$

Menurut Friedman (1991) algoritma dalam MARS terdiri dari *forward stepwise* (langkah maju) dan *backward stepwise* (langkah mundur). Prosedur *Forward stepwise* berfungsi untuk membentuk model dengan seluruh fungsi basis yang sebelumnya telah ditentukan oleh peneliti. Karena prosedur *forward stepwise* memberikan model yang sangat kompleks dengan memasukkan seluruh fungsi basis, maka dalam prosedur *backward stepwise* untuk masing-masing iterasi dalam pengulangan algoritmanya menyebabkan fungsi basis yang tidak memiliki pengaruh signifikan akan dihapus. *Backward stepwise* akan menghasilkan model terbaik. Model terbaik ditentukan berdasarkan kriteria *Generalized Cross Validation* (GCV) minimum yang didefinisikan sebagai berikut:

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

dengan :

M = jumlah fungsi basis

x_i = variabel prediktor, $i= 1,2, \dots, N$

y_i = variabel respon, $i= 1,2, \dots, N$

N = banyaknya sampel

$C(\hat{M}) = C(M) + d.M$, nilai d terbaik berada pada interval $2 \leq d \leq 4$

$C(M) = \text{Trace} [\mathbf{B} (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T] + 1$

Menurut Andari, *et al.* (2013) model MARS untuk nilai variabel respon biner merupakan pendekatan regresi logistik, yaitu:

$$\ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [S_{km}(X_{V(k,m)} - t_{km})]_+$$

Menurut Hosmer dan Lemeshow (1989) apabila variabel respon adalah biner maka dapat dikatakan sebagai regresi respon biner. Penentuan klasifikasi pada variabel respon biner (0 dan 1) dilakukan dengan menggunakan titik potong sebesar 0,5 dengan ketentuan jika estimasi peluang melebihi 0,5 maka hasil prediksi adalah masuk ke kelas 1, jika estimasi peluang kurang dari atau sama dengan 0,5 maka hasil prediksi adalah masuk ke kelas 0.

2.5 Evaluasi Ketepatan Hasil Klasifikasi

Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi (Johnson dan Wichern, 2007). Pada penelitian ini untuk menguji hasil klasifikasi digunakan statistik uji *Press's Q*, *APER*, *sensitivity*, dan *specificity*.

Menurut Hair, *et al.* (2006) Statistik uji *Press's Q* membandingkan antara jumlah ketepatan klasifikasi dengan total pengamatan dan banyaknya kelompok. Jika nilai *Press's Q* lebih besar dari nilai *chi-square* berderajat bebas 1, maka prediksi dalam klasifikasi dianggap signifikan secara statistik. Uji statistik *Press's Q* didefinisikan sebagai berikut:

$$Press's Q = \frac{[N-(rC)]^2}{N(C-1)} \sim \chi^2_{(1)}$$

dengan:

N = ukuran sampel

r = banyaknya objek yang tepat diklasifikasikan

C = banyaknya kelompok

Menurut Johnson dan Wichern (2007) *APER* (*Apparent Error Rate*) merupakan suatu nilai yang digunakan untuk melihat peluang kesalahan dalam mengklasifikasi objek. Nilai *APER* menyatakan proporsi sampel yang salah diklasifikasikan. Mengacu pada tabel 1, nilai *APER* didefinisikan sebagai berikut:

$$APER = \frac{f_{01} + f_{10}}{f_{00} + f_{10} + f_{01} + f_{11}} \times 100\%$$

Menurut Han dan Kamber (2006) *Specificity* mengukur proporsi yang benar-benar negatif, yaitu proporsi dari kelas 0 yang teridentifikasi secara benar. *Sensitivity* mengukur proporsi benar-benar positif, yaitu proporsi dari kelas 1 yang teridentifikasi secara benar. Berdasarkan tabel 1, formula *specificity* dan *sensitivity* dapat dituliskan sebagai berikut:

$$Specificity (\%) = \frac{f_{00}}{f_{01} + f_{00}} \times 100$$

$$Sensitivity (\%) = \frac{f_{11}}{f_{10} + f_{11}} \times 100$$

3. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data akreditasi serta hasil penilaian kedelapan komponen akreditasi dari 168 Sekolah Dasar Negeri berakreditasi A dan B yang ada di Kota Semarang. Data ini diperoleh melalui situs resmi Badan Akreditasi Nasional, <http://www.ban-sm.or.id>.

Variabel dalam penelitian ini terdiri dari variabel prediktor dan variabel respon. Variabel respon pada penelitian adalah nilai akreditasi sekolah, dimana pada penelitian ini dikelompokkan kedalam dua kategori yaitu:

1. Kelompok 0, yaitu kelompok sekolah dasar negeri yang mendapatkan akreditasi B.
2. Kelompok 1, yaitu kelompok sekolah dasar negeri yang mendapatkan akreditasi A.

Variabel prediktor pada penelitian ini adalah nilai kedelapan komponen penilaian akreditasi sekolah yang terdiri dari:

1. Komponen Standar Isi
2. Komponen Standar Proses
3. Komponen Standar Kompetensi Lulusan
4. Komponen Standar Pendidik dan Tenaga Kependidikan
5. Komponen Standar Sarana dan Prasarana
6. Komponen Standar Pengelolaan
7. Komponen Standar Pembiayaan
8. Komponen Standar Penilaian

Tahapan analisis data yang dilakukan menggunakan metode *K-Nearest Neighbor* (K-NN) adalah sebagai berikut:

- a. Menentukan data uji
- b. Menghitung jarak (kedekatan lokasi) antara data uji dengan data yang lain, parameter jarak yang digunakan adalah jarak *Euclidean*
- c. Menentukan nilai K
- d. Hitung jumlah data yang mengikuti kelas yang ada dari nilai K-tetangga tersebut
- e. Ambil kelas dengan jumlah data terbanyak yang mengikutinya sebagai kelas pemenang dan diberikan label kelas tersebut pada data uji.
- f. Menghitung klasifikasi

Tahapan analisis data yang dilakukan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS) adalah sebagai berikut:

- Mengkombinasikan banyaknya fungsi basis (*Basis Function* (BF)), *Maximum Interaction* (MI) dan *Minimum Observation* (MO). Fungsi basis yang digunakan adalah antara 2-4 kali banyaknya variabel bebas, *Maximum Interaction* (MI) yang digunakan adalah 1, 2, dan 3, serta *Minimum Observation* (MO) yang digunakan yaitu 0, 1, 2 dan 3.
- Membentuk model dengan metode pendugaan kuadrat terkecil
- Memilih model terbaik dengan nilai GCV minimal
- Menginterpretasi model terbaik yang sudah didapatkan
- Menghitung klasifikasi berdasarkan model terbaik

Setelah hasil klasifikasi dari metode *K-Nearest Neighbor* (K-NN) dan metode *Multivariate Adaptive Regression Spline* (MARS) didapatkan, langkah selanjutnya adalah membuat matriks konfusi serta menghitung akurasi klasifikasi dari kedua metode tersebut dengan menggunakan uji statistik *Press's Q*, *APER*, *sensitivity* dan *specificity* untuk menguji metode manakah yang lebih baik digunakan dalam klasifikasi akreditasi Sekolah Dasar Negeri di Kota Semarang.

4. HASIL DAN PEMBAHASAN

4.1 Metode *K-Nearest Neighbor*

Hasil laju *error* untuk percobaan nilai K pada *K-Nearest Neighbor* (K-NN) disajikan pada Tabel 2 sebagai berikut:

Tabel 2. Hasil Laju *Error* untuk Percobaan Nilai K pada K-NN

Nilai K	Laju <i>Error</i>
3	0,0655
5	0,0536
7	0,0595
9	0,0714
11	0,0833
13	0,0774

Berdasarkan Tabel 2, nilai K yang menghasilkan laju *error* terkecil adalah K=5, maka nilai K yang digunakan dalam penelitian ini adalah K=5. Berdasarkan perhitungan yang telah dilakukan, hasil klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN) dapat dilihat pada Tabel 3 berikut:

Tabel 3. Hasil Klasifikasi dengan Metode K-NN

Hasil Observasi (<i>Actual Class</i>)	Hasil Prediksi (<i>Predicted Class</i>)	
	Kelas 0 (Akreditasi B)	Kelas 1 (Akreditasi A)
Kelas 0 (Akreditasi B)	85	7
Kelas 1 (Akreditasi A)	2	74

4.2 Metode MARS

Berdasarkan kombinasi BF 16, 24, 32 dengan MI 1, 2, 3 serta MO 0, 1, 2, dan 3, model terbaik adalah jika menggunakan kombinasi BF= 32, MI= 2, dan MO= 1 karena menghasilkan nilai GCV terkecil. Model MARS terbaiknya adalah sebagai berikut:

$$Y = -0,450005 + 0,0546213 * BF1 - 0,0962844 * BF3 + 0,0360191 * BF5 \\ - 0,00313243 * BF7 - 0,00315374 * BF8 + 0,0113113 * BF9 \\ + 0,0204061 * BF10 + 0,0441561 * BF12 + 0,00466661 * BF16 \\ - 0,0186539 * BF27 + 0,000739729 * BF32$$

dengan :

$$\begin{aligned} BF1 &= \max(0, \text{STANDAR_PENGELOLAAN} - 82); \\ BF2 &= \max(0, 82 - \text{STANDAR_PENGELOLAAN}); \\ BF3 &= \max(0, \text{STANDAR_KL} - 90); \\ BF4 &= \max(0, 90 - \text{STANDAR_KL}); \\ BF5 &= \max(0, \text{STANDAR_ISI} - 82); \\ BF6 &= \max(0, 82 - \text{STANDAR_ISI}); \\ BF7 &= \max(0, \text{STANDAR_TPDK} - 84) * BF1; \\ BF8 &= \max(0, 84 - \text{STANDAR_TPDK}) * BF1; \\ BF9 &= \max(0, \text{STANDAR_PEMBIAYAAN} - 64); \\ BF10 &= \max(0, \text{STANDAR_SDP} - 94) * BF6; \\ BF12 &= \max(0, \text{STANDAR_KL} - 83) * BF2; \\ BF16 &= \max(0, \text{STANDAR_PROSES} - 95) * BF4; \\ BF27 &= \max(0, 66 - \text{STANDAR_SDP}) * BF1; \\ BF28 &= \max(0, \text{STANDAR_KL} - 67); \\ BF32 &= \max(0, \text{STANDAR_TPDK} - 59) * BF28 \end{aligned}$$

Nilai GCV terkecilnya adalah :

$$GCV = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{y}_M(x_i)]^2}{\left[1 - \frac{C(\hat{M})}{N}\right]^2} = \frac{\frac{1}{168} (8,28195)}{\left[1 - \frac{30,09676}{168}\right]^2} = 0,07316$$

Berdasarkan model yang telah diperoleh dengan metode MARS. Hasil klasifikasi dapat dilihat pada Tabel 4 berikut :

Tabel 4. Hasil Klasifikasi dengan Metode MARS

Hasil Observasi (<i>Actual Class</i>)	Hasil Prediksi (<i>Predicted Class</i>)	
	Kelas 0 (Akreditasi B)	Kelas 1 (Akreditasi A)
Kelas 0 (Akreditasi B)	89	3
Kelas 1 (Akreditasi A)	1	75

4.3 Evaluasi Ketepatan Klasifikasi

Ukuran yang digunakan untuk mengevaluasi ketepatan hasil klasifikasi akreditasi Sekolah Dasar Negeri di Kota Semarang adalah *Press'Q*, APER (*Apparent Error Rate*), *specificity*, dan *sensitivity*.

1. Statistik uji *Press's Q*

Berdasarkan hasil klasifikasi menggunakan metode *K-Nearest Neighbor* (KNN), nilai statistik uji *Press's Q* yang didapat adalah sebesar:

$$Press's Q = \frac{[N-(rC)]^2}{N(C-1)} = \frac{[168-(159.2)]^2}{168(2-1)} = 133,929$$

Berdasarkan hasil klasifikasi menggunakan metode MARS, nilai statistik uji *Press's Q* yang didapat adalah sebesar :

$$Press's Q = \frac{[N-(rC)]^2}{N(C-1)} = \frac{[168-(164.2)]^2}{168(2-1)} = 152,381$$

Berdasarkan hasil perhitungan statistik *Press's Q* untuk masing-masing metode adalah 133,929 dan 152,381 lebih besar dari $\chi^2_{(0,05;1)} = 3,84$ maka dapat dikatakan pengklasifikasian Sekolah Dasar Negeri (SDN) di Kota Semarang berdasarkan akreditasi menggunakan metode MARS dan metode *K-Nearest Neighbor* (K-NN) signifikan secara statistik.

2. APER (*Apparent Error Rate*), *Specificity* dan *Sensitivity*

Berdasarkan hasil klasifikasi menggunakan metode *K-Nearest Neighbor* (KNN), besarnya nilai APER, *specificity*, dan *sensitivity* ditunjukkan pada Tabel 5, sebagai berikut:

Tabel 5. Perhitungan APER, *Specificity*, dan *Sensitivity* untuk Metode K-NN

Hasil Observasi (<i>Actual Class</i>)	Hasil Prediksi (<i>Predicted Class</i>)	
	Kelas 0 (Akreditasi B)	Kelas 1 (Akreditasi A)
Kelas 0 (Akreditasi B)	85	7
Kelas 1 (Akreditasi A)	2	74
$\text{APER} = \frac{7+2}{85+7+2+74} \times 100\% = 5,357\%$ $\text{Specificity (\%)} = \frac{85}{7+85} \times 100 = 92,39\%$ $\text{Sensitivity (\%)} = \frac{74}{2+74} \times 100 = 97,37\%$		

Berdasarkan hasil klasifikasi menggunakan metode MARS, besarnya nilai APER, *specificity*, dan *sensitivity* ditunjukkan pada Tabel 6, sebagai berikut :

Tabel 6. Perhitungan APER, *Specificity*, dan *Sensitivity* untuk Metode MARS

Hasil Observasi (<i>Actual Class</i>)	Hasil Prediksi (<i>Predicted Class</i>)	
	Kelas 0 (Akreditasi B)	Kelas 1 (Akreditasi A)
Kelas 0 (Akreditasi B)	89	3
Kelas 1 (Akreditasi A)	1	75
$\text{APER} = \frac{3+1}{89+3+1+75} \times 100\% = 2,381\%$ $\text{Specificity (\%)} = \frac{89}{3+89} \times 100 = 96,74\%$ $\text{Sensitivity (\%)} = \frac{75}{1+75} \times 100 = 98,68\%$		

5. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dipaparkan sebelumnya, maka dapat diambil kesimpulan sebagai berikut:

1. Berdasarkan hasil perhitungan statistik *Press's Q* untuk metode K-NN adalah 133,929 dan metode MARS adalah 152,381 lebih besar dari $\chi^2_{(0,05;1)} = 3,84$ maka dapat dikatakan pengklasifikasian Sekolah Dasar Negeri (SDN) di Kota Semarang berdasarkan akreditasi menggunakan metode MARS dan metode K-NN signifikan secara statistik.
2. Berdasarkan perhitungan APER, *specificity*, dan *sensitivity* menunjukkan bahwa pengklasifikasian menggunakan metode MARS lebih baik dibandingkan dengan metode K-NN dalam mengklasifikasikan Sekolah Dasar Negeri di Kota Semarang berdasarkan akreditasinya.

6. DAFTAR PUSTAKA

- Andari, S, Purnami. S.W., Otok. B.W., 2013. *Smooth Support Vector Machine dan Multivariate Adaptive Regression Spline untuk Mendiagnosis Kanker Payudara*. Jurnal Statistika, Vol. 1, No. 2, November 2013. Surabaya
- [BAN]. Badan Akreditasi Nasional. *Hasil Akreditasi*.
<http://www.ban-sm.or.id> [diakses 02 Desember 2013]
- [BAN]. Badan Akreditasi Nasional. 2009. *Kebijakan dan Pedoman Akreditasi Sekolah/Madrasah*.
<http://www.ban-sm.or.id> [diakses 26 Desember 2013]
- Friedman. J. H. 1991. *Multivariate Adaptive Regression Splines*. The Annals of Statistics, Vol. 19
- Hair. J. F., Black. W. C., Babin. B. J., Anderson R. E. 2006. *Multivariate Data Analysis. Seventh Edition*. Pearson Education Prentice Hall. Inc
- Han, J., Kamber, M. 2006. *Data Mining: Concepts and Techniques*. San Fransisco: Morgan Kaufmann Publisers
- Hosmer. D. W., Lemeshow. S. 1989. *Applied Logistic Regression*. John Wiley & Sons, Inc
- Ihsan. F. 1996. *Dasar-dasar Kependidikan*. Jakarta: PT Rineka Cipta
- Johnson. R. A., Wichern. D. W. 2007. *Applied Multivariate Statistical Analysis. Sixth Edition*. New Jersey: Prentice Hall International. Inc
- [KBBI]. Kamus Besar Bahasa Indonesia. 2008.
<http://kbbi.web.id/akreditasi> [diakses 26 Desember 2013]
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Penerbit ANDI Yogyakarta
- Santosa, B. 2007. *Data Mining Terapan*. Yogyakarta: Graha Ilmu.