

**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**



SKRIPSI

Oleh :
BISRI MERLUARINI
24010210130071

**JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG
2014**

**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**

**Oleh :
BISRI MERLUARINI
24010210130071**

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Sains pada Jurusan Statistika FSM UNDIP**

**JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG**

2014

HALAMAN PENGESAHAN I

Judul : **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**

Nama : Bisri Merluarini

NIM : 24010210130071

Jurusan : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 23 Mei 2014 dan dinyatakan lulus pada tanggal 05 Juni 2014.

Semarang, 05 Juni 2014

Mengetahui,

Ketua Jurusan Statistika

Fakultas Sains dan Matematika



Dra. Dwi Ispriyanti, M.Si.

NIP. 195709141986032001

Panitia Penguji Ujian Tugas Akhir

Ketua,

Drs. Sudarno, M.Si.

NIP. 196407091992011001

HALAMAN PENGESAHAN II

Judul : **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**

Nama : Bisri Merluarini

NIM : 24010210130071

Jurusan : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 23 Mei 2014.

Semarang, 05 Juni 2014

Pembimbing I



Diah Safitri, S.Si., M.Si.
NIP. 197510082003122001

Pembimbing II



Abdul Hoyyi, S.Si., M.Si.
NIP. 197202022008011018

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah melimpahkan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **“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”**.

Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Jurusan Statistika Universitas Diponegoro. Tanpa adanya bantuan dari berbagai pihak, Tugas Akhir ini tidak akan berjalan dengan baik. Oleh karena itu penulis menyampaikan terimakasih kepada :

1. Ibu Dra. Dwi Ispriyanti, M.Si selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Ibu Diah Safitri, S.Si., M.Si dan Bapak Abdul Hoyyi, S.Si., M.Si selaku dosen pembimbing I dan dosen pembimbing II yang telah memberikan bimbingan dan pengarahan dalam penulisan Tugas Akhir ini.
3. Bapak dan Ibu dosen Jurusan Statistika Universitas Diponegoro yang telah memberikan ilmu yang bermanfaat.
4. Semua pihak yang tidak dapat disebutkan satu per satu yang telah mendukung penulis dalam menyelesaikan Tugas Akhir ini.

Penulis berharap Tugas Akhir ini bermanfaat bagi civitas akademika di Universitas Diponegoro khususnya Jurusan Statistika dan masyarakat umumnya.

Semarang, Juni 2014

Penulis

ABSTRAK

Metode klasifikasi 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). Penelitian ini bertujuan untuk membandingkan pengklasifikasian akreditasi Sekolah Dasar Negeri (SDN) di kota Semarang antara metode K-NN dan MARS. Penelitian ini menggunakan data akreditasi serta hasil penilaian kedelapan komponen akreditasi pada Sekolah Dasar Negeri berakreditasi A (kelompok 1) dan B (kelompok 0) di kota Semarang. Evaluasi hasil klasifikasi menggunakan uji statistik *Press's Q*, *APER*, *specificity*, dan *sensitivity*. Hasil klasifikasi terbaik metode K-NN adalah jika menggunakan $K=5$ karena menghasilkan laju *error* terkecil dan diperoleh informasi bahwa data yang tepat diklasifikasikan berjumlah 159 data dan yang salah diklasifikasikan berjumlah 9 data. Hasil klasifikasi terbaik menggunakan metode MARS adalah jika menggunakan kombinasi $BF=32$, $MI=2$, $MO=1$ karena menghasilkan nilai *Generalized Cross Validation* (GCV) terkecil dan diperoleh informasi bahwa data yang tepat diklasifikasikan berjumlah 164 data dan yang salah diklasifikasikan berjumlah 4 data. Berdasarkan hasil analisis, diperoleh nilai *Press's Q* yang menunjukkan bahwa kedua metode sudah baik atau signifikan secara statistik dalam mengklasifikasikan Sekolah Dasar Negeri (SDN) di kota Semarang berdasarkan akreditasinya. Berdasarkan perhitungan *APER*, *specificity*, dan *sensitivity* menunjukkan bahwa pengklasifikasian akreditasi Sekolah Dasar Negeri (SDN) di kota Semarang menggunakan metode MARS lebih baik dibandingkan dengan metode K-NN.

Kata kunci: Klasifikasi, *K-Nearest Neighbor* (K-NN), *Multivariate Adaptive Regression Spline* (MARS), Evaluasi klasifikasi

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

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
KATA PENGANTAR	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR LAMPIRAN	x
DAFTAR SIMBOL	xi
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
BAB II TINJAUAN PUSTAKA	
2.1 Akreditasi	5
2.2 Klasifikasi	8
2.3 <i>K-Nearest Neighbor</i> (K-NN)	9
2.4 Multivariate Adaptive Regression Spline (MARS).....	11
2.4.1 Anova Dekomposisi	14
2.4.2 Algoritma MARS	15

2.4.3 Klasifikasi pada MARS	16
2.5 Evaluasi Ketepatan Hasil Klasifikasi	17
BAB III METODOLOGI PENELITIAN	
3.1 Sumber Data	20
3.2 Variabel Penelitian	20
3.3 Metode Analisis Data	21
BAB IV HASIL DAN PEMBAHASAN	
4.1 Metode <i>K-Nearest Neighbor</i>	24
4.1.1 Pendugaan Klasifikasi Menggunakan Metode K-NN	24
4.1.2 Hasil Klasifikasi Menggunakan <i>K-Nearest Neighbor</i>	26
4.2 Metode MARS	27
4.2.1 Pemodelan MARS	27
4.2.2 Interpretasi Model MARS	29
4.2.3 Pendugaan Klasifikasi Menggunakan Metode MARS.....	31
4.2.4 Hasil Klasifikasi Menggunakan MARS.....	37
4.3 Evaluasi Ketepatan Klasifikasi	38
BAB V KESIMPULAN	42
DAFTAR PUSTAKA	43
LAMPIRAN	45

DAFTAR TABEL

	Halaman
Tabel 1. Matriks Konfusi untuk Klasifikasi Dua Kelas	9
Tabel 2. Hasil Laju <i>Error</i> untuk Percobaan nilai K pada K-NN	25
Tabel 3. Hasil Klasifikasi dengan Metode K-NN	27
Tabel 4. Hasil Klasifikasi dengan Metode MARS	37
Tabel 5. Perhitungan APER untuk Metode K-NN	39
Tabel 6. Perhitungan APER untuk Metode MARS	39
Tabel 7. Perhitungan <i>Specificity</i> dan <i>Sensitivity</i> untuk Metode K-NN	40
Tabel 8. Perhitungan <i>Specificity</i> dan <i>Sensitivity</i> untuk Metode MARS	41

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data Akreditasi serta Hasil Penilaian kedelapan Komponen akreditasi dari 168 Sekolah Dasar Negeri berakreditasi A dan B yang ada di Kota Semarang	45
Lampiran 2. <i>Syntax</i> Metode K-NN Menggunakan <i>software</i> Matlab R2009a.....	46
Lampiran 3. <i>Output</i> Hasil Klasifikasi Menggunakan K-NN menggunakan <i>software</i> Matlab R2009a	47
Lampiran 4. Hasil Perhitungan Jarak <i>Euclidean</i> dengan Menggunakan Data Uji adalah Data ke 1 dan Telah Dilakukan Pengurutan dari Terkecil hingga Terbesar	53
Lampiran 5. Hasil Perhitungan Jarak <i>Euclidean</i> dengan Menggunakan Data Uji adalah Data ke 93 dan Telah Dilakukan Pengurutan dari Terkecil hingga Terbesar	54
Lampiran 6. Output dari Metode <i>Multivariate Adaptive Regression Spline</i> (MARS) Menggunakan <i>Software Salford Predictive Modeler 7</i> (SPM7)	55
Lampiran 7. Nilai GCV untuk seluruh pengombinasian BF, MI, dan MO.....	71
Lampiran 8. Tabel <i>Chi-Square</i>	72

DAFTAR SIMBOL

f_{11}	: jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1
f_{10}	: jumlah data dalam kelas 1 yang secara salah dipetakan ke kelas 0
f_{01}	: jumlah data dalam kelas 0 yang secara salah dipetakan ke kelas 1
f_{00}	: jumlah data dalam kelas 0 yang secara benar dipetakan ke kelas 0
$d(\mathbf{x}_1, \mathbf{x}_2)$: jarak <i>Euclidean</i> antara dua titik
N	: ukuran sampel
n	: jumlah variabel prediktor
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
$X_{V(k,m)}$: variabel prediktor ke v , pilahan ke- k dan <i>subregion</i> ke- m
t_{km}	: nilai knot dari variabel prediktor $X_{V(k,m)}$
\mathbf{Y}	: matriks y berukuran $n \times 1$
\mathbf{a}	: matriks a berukuran $m \times 1$
$\boldsymbol{\varepsilon}$: matriks <i>error</i> berukuran $n \times 1$
\mathbf{B}	: matriks fungsi basis berukuran $n \times m$
$b_q^\pm(x - t)$: fungsi <i>truncated</i> fungsi basis
$y_i(x_i)$: penjumlahan semua fungsi basis untuk satu peubah prediktor
$y_{i,j}^*(x_i, x_j)$: penjumlahan semua fungsi basis untuk dua peubah prediktor

- $y_{i,j,k}^*(x_i, x_j, x_k)$: penjumlahan semua fungsi basis untuk tiga peubah prediktor
- B_m : fungsi basis ke- m
- $GCV(M)$: *Generalized Cross Validation* (kriteria penentuan model terbaik)
- x_i : variabel prediktor
- y_i : variabel respon
- $C(M)$: trace dari matriks $[\mathbf{B} (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T] + 1$
- $C(\hat{M})$: $C(M) + d.M$
- d : unsur pengali fungsi basis, nilai d berada pada interval $2 \leq d \leq 4$
- $P(Y=1 | X)$: probabilitas dari Y jika diketahui X
- C : banyaknya kelompok
- r : jumlah data yang tepat diklasifikasikan
- $\chi_{(1)}^2$: nilai dari tabel *chi square* berderajat bebas 1

BAB I

PENDAHULUAN

1.1 Latar Belakang

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).

Pandangan sebagian besar masyarakat menganggap bahwa sekolah yang baik adalah sekolah negeri dengan akreditasi A atau B. Nilai dari akreditasi sekolah negeri inilah yang menjadi salah satu pertimbangan dalam memilih sekolah. Begitu pula dalam menentukan sekolah dasar, orang tua cenderung memilihkan anaknya sekolah dasar negeri yang memiliki rekam jejak (*track record*) yang baik.

Penyusunan instrumen akreditasi sekolah mengacu pada Standar Nasional Pendidikan yang telah ditetapkan Mendiknas dan dikembangkan oleh Badan Standar Nasional Pendidikan (BSNP). Berdasarkan acuan Permendiknas tersebut,

Badan Akreditasi Nasional (BAN) menentukan delapan komponen penilaian terhadap akreditasi suatu sekolah yaitu: standar isi, standar proses, standar kompetensi lulusan, standar pendidik dan tenaga kependidikan, standar sarana dan prasarana, standar pengelolaan, standar pembiayaan, serta standar penilaian pendidikan (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.

Berdasarkan penjelasan tersebut, penulis tertarik 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.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, penulis tertarik menggunakan metode *K-Nearest Neighbor* (K-NN) dan metode *Multivariate Adaptive Regression Spline* (MARS) untuk membandingkan pengklasifikasian akreditasi berdasarkan komponen-komponen penilaiannya.

1.3 Batasan Masalah

Penelitian ini hanya dibatasi pada data akreditasi Sekolah Dasar Negeri (SDN) di kota Semarang yang status akreditasinya ditetapkan pada tahun 2010 hingga 2012 dan diklasifikasikan menjadi dua kelompok, yaitu kelompok sekolah dengan akreditasi A dan kelompok sekolah dengan akreditasi B.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengetahui hasil klasifikasi akreditasi Sekolah Dasar Negeri (SDN) di kota Semarang menggunakan metode *K-Nearest Neighbor* (K-NN) dan metode *Multivariate Adaptive Regression Spline* (MARS).
2. Mengetahui akurasi klasifikasi akreditasi Sekolah Dasar Negeri (SDN) di kota Semarang menggunakan uji statistik *Press's Q*, *APER*, *specificity*, dan *sensitivity* untuk metode *K-Nearest Neighbor* (K-NN) dan metode *Multivariate Adaptive Regression Spline* (MARS)

3. Membandingkan pengklasifikasian akreditasi Sekolah Dasar Negeri (SDN) di kota Semarang antara metode *K-Nearest Neighbor* (K-NN) dan *Multivariate Adaptive Regression Spline* (MARS)