# KETEPATAN KLASIFIKASI DENGAN MENGGUNAKAN METODE *MULTIVARIATE ADAPTIVE REGRESSION SPLINE* (MARS) PADA DATA KELOMPOK RUMAH TANGGA KABUPATEN CILACAP

Skripsi

disusun sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Program Studi Matematika

oleh

Saroful Anam
4111412029

**JURUSAN MATEMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS NEGERI SEMARANG**

**2017**

# PERNYATAAN

Saya menyatakan bahwa skripsi ini bebas plagiat, dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan perundang-undangan.

Semarang, April 2017

METERAI TEMPEL
3CD60AEF472446646
6000
ENAM RIBU RUPIAH

Saroful Anam
4111412029

UNNES
UNIVERSITAS NEGERI SEMARANG

# PENGESAHAN

Skripsi yang berjudul

Ketepatan Klasifikasi dengan Menggunakan Metode *Multivariate Adaptive Regression Spline* (MARS) pada Data Kelompok Rumah Tangga Kabupaten Cilacap

disusun oleh

Saroful Anam

4111412029

telah dipertahankan dihadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada tanggal 30 Maret 2017.

Ketua

Prof. Zaenuri, SE, M.Si, Akt.
NIP. 196412231988031001

Sekretaris

Drs. Arief Agoestanto, M.Si.
NIP. 196807221993031005

Ketua Penguji

Putriaji Hendikawati, S.Si, M.Pd, M.Sc
NIP. 198208182006042001

Anggota Penguji/Pembimbing 1

Drs. Sugiman, M.Si
NIP. 196401111989011001

AnggotaPenguji/Pembimbing 2

Dra. Sunarmi, M.Si
NIP. 195506241988032001

UNNES
UNIVERSITAS NEGERI SEMARANG

# MOTTO DAN PERSEMBAHAN

## MOTTO

Barang siapa menginginkan dunia maka dengan ilmu, dan barang siapa menginginkan akhirat maka dengan ilmu, dan barang siapa yang menginginkan keduanya maka juga dengan ilmu (HR. Bukhori & Muslim)

## PERSEMBAHAN

- Untuk kedua orang tua Ibu Musrifah dan Bapak Maryuni Alm.
- Untuk Kakak-Kakakku tersayang Mas Wahyudi & Mbak Inah Pepri Waryani
- Untuk keluarga besar tercinta
- Untuk Teman-teman Matematika angkatan 2012
- Untuk Teman-teman Pondok Pesantren Al Asror Semarang
- Untuk Teman-teman PKL BPS Provinsi Jawa Tengah
- Untuk Teman-teman KKN Alternatif "SEKAR" 2015
- Untuk Almamater tercinta Universitas Negeri Semarang (UNNES)

UNNES
UNIVERSITAS NEGERI SEMARANG

# KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT yang Maha Pengasih dan Penyayang, atas limpahan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "**Ketepatan Klasifikasi dengan Menggunakan Metode *Multivariate Adaptive Regression Spline* (MARS) pada Data Kelompok Rumah Tangga Kabupaten Cilacap**".

Penulis menyadari dalam menyelesaikan skripsi ini memperoleh banyak bantuan dan dukungan dari berbagai pihak. Untuk itu, dengan rasa hormat, penulis menyampaikan terima kasih kepada:

1. Prof. Dr. Fathur Rokhman, M.Hum, selaku Rektor Universitas Negeri Semarang;
2. Prof. Dr. Zaenuri S.E, M.Si,Akt., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang;
3. Drs. Arief Agoestanto, M.Si., selaku Ketua Jurusan Matematika FMIPA Universitas Negeri Semarang;
4. Drs. Mashuri M.Si., selaku Ketua Prodi Matematika Jurusan Matematika FMIPA Universitas Negeri Semarang;
5. Dra. Kristina Wijayanti, M.Si., selaku dosen wali rombel 2 Prodi Matematika Jurusan Matematika FMIPA Universitas Negeri Semarang;
6. Drs. Sugiman, M.Si., selaku dosen pembimbing utama, yang telah memberikan arahan dan bimbingan dalam penyelesaian skripsi ini;

7. Dra. Sunarmi, M.Si., selaku dosen pembimbing pendamping, yang telah memberikan arahan dan bimbingan dalam penyelesaian skripsi ini;
8. Putriaji Hendikawati S.Si., M.Pd., M.Sc., selaku dosen penguji yang telah memberikan masukan untuk perbaikan skripsi ini;
9. Bapak dan Ibu Dosen Jurusan Matematika serta staf Tata Usaha Universitas Negeri Semarang yang telah membekali penulis dengan ilmu selama mengikuti perkuliahan dan penulisan skripsi;
10. Orang tua dan keluarga yang selalu memberikan doa, dukungan, dan semangat;
11. Teman-teman seperjuangan di Pondok Pesantren Al asror Semarang, yang selalu menghibur, memberikan motivasi, dorongan dan do'a.
12. Semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan bantuan.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun dari pembaca.

UNNES
UNIVERSITAS NEGERI SEMARANG

Semarang, April 2017

Penulis

# ABSTRAK

Anam, Saroful. 2017. Ketepatan Klasifikasi dengan Menggunakan Metode *Multivariate Adaptive Regression Spline* (MARS) pada Data Rumah Tangga Kabupaten Cilacap. Skripsi. Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang, Pembimbing Utama Drs. Sugiman, M.Si. dan Pembimbing Pendamping Dra. Sunarmi, M.Si.

Kata kunci: Klasifikasi, Kelompok Rumah tangga, *Multivariate Adaptive Regression Spline* (MARS).

Kelompok rumah tangga berdasarkan pengeluaran per kapita dipisahkan atas dua kelompok yaitu kelompok rumah tangga miskin dan tidak miskin. Pemisahan objek ke dalam suatu kelompok tertentu digunakan metode klasifikasi. Tujuan penelitian ini adalah mengklasifikasikan dan menghitung tingkat kesalahan hasil klasifikasi kelompok rumah tangga di Kabupaten Cilacap berdasarkan faktor-faktor yang mempengaruhi tingkat kemiskinan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS).

MARS merupakan salah satu metode regresi nonparametrik yang dapat digunakan untuk data berdimensi tinggi. Untuk mendapatkan model MARS terbaik, dilakukan kombinasi nilai *Basis Function* (BF), *Maksimum Interaction* (MI), dan *Minimum Observation* (MO) dengan cara *trial and error*. Model terbaik yang dipilih adalah model dengan kombinasi BF=45, MI=3, MO=1 karena memiliki nilai GCV terkecil yaitu sebesar 0,030.

Berdasarkan variabel-variabel yang mempengaruhi kelompok rumah tangga, diperoleh hasil klasifikasi dari 37 rumah tangga dengan kategori miskin, 34 rumah tangga tepat diklasifikasikan ke dalam kategori miskin, sedangkan 3 rumah tangga salah diklasifikasikan salah dan masuk ke dalam kategori miskin. Begitu juga dari 113 rumah tangga dengan kategori tidak miskin, 113 rumah tangga tepat diklasifikasikan ke dalam kategori tidak miskin, dan tidak ada rumah tangga yang salah diklasifikasikan ke dalam rumah tangga dengan kategori miskin.

Diperoleh ketepatan hasil klasifikasi sebesar 98,00% nilai *Apparent Error Rate* (APER) sebesar 2,00% dan nilai uji *Press's Q* menunjukkan bahwa secara statistik metode MARS sudah konsisten dalam mengklasifikasi data. Sehingga untuk penelitian mengenai klasifikasi selanjutnya disarankan menggunakan metode MARS karena terbukti menghasilkan ketepatan klasifikasi yang akurat.

# DAFTAR ISI

Halaman

**HALAMAN JUDUL** ..... i

**PERNYATAAN KEASLIAN TULISAN** ..... ii

**HALAMAN PENGESAHAN** ..... iii

**MOTTO DAN PERSEMBAHAN** ..... iv

**KATA PENGANTAR** ..... v

**ABSTRAK**... ..... vii

**DAFTAR ISI** ..... viii

**DAFTAR TABEL** ..... xi

**DAFTAR GAMBAR** ..... xii

**DAFTAR LAMPIRAN**. ..... xiii

**DAFTAR SIMBOL** ..... xiv

**BAB 1 PENDAHULUAN**

1.1 Latar Belakang ..... 1

1.2 Rumusan Masalah ..... 4

1.3 Batasan Masalah ..... 4

1.4 Tujuan Penelitian ..... 5

1.5 Manfaat penelitian ..... 5

1.6 Sistematika Penulisan Skripsi ..... 6

**BAB 2 TINJAUAN PUSTAKA**

2.1 Statistik Deskriptif ..... 9

2.2 Analisis Regresi ..... 9
2.2.1 Regresi Nonparametrik ..... 11
2.2.1.1 Regresi *Spline* ..... 12
2.2.1.2 Basis *B-Spline* ..... 14
2.3 *Multivariate Adaptive Regression Spline* (MARS) ..... 18
2.3.1 Model MARS ..... 20
2.4 Klasifikasi ..... 27
2.5 Klasifikasi pada MARS ..... 27
2.6 *Apparent Error Rate* (APER) dan *Press's Q* ..... 29
2.7 Kemiskinan ..... 31
2.7.1 Garis Kemiskinan ..... 32
2.7.2 Teknik Perhitungan Garis Kemiskinan Tingkat Kabupaten/Kota ..... 32
2.7.3 Indikator Kemiskinan ..... 35
2.8 Penelitian Terdahulu ..... 36
2.9 *Software* MARS 2.0 ..... 38
**BAB 3 METODE PENELITIAN**
3.1 Studi Pustaka ..... 43
3.2 Perumusan Masalah ..... 44
3.3 Pengumpulan Data ..... 44
3.3.1 Variabel Penelitian ..... 45
3.3.2 *Software* yang digunakan ..... 45
3.4 Langkah-Langkah Analisis ..... 45
3.5 *Flowchart* (Diagram Alir Analisis Data) ..... 47

UNNES
UNIVERSITAS NEGERI SEMARANG

3.6 Penarikan Kesimpulan ..... 48

**BAB 4 HASIL DAN PEMBAHASAN**

4.1 Hasil ..... 49

4.1.1 Analisis Deskriptif ..... 49

4.1.1.1 Statistik Deskriptif Jumlah Pendapatan dan Pengeluaran Kepala Rumah Tangga ..... 50

4.1.2 Klasifikasi Kelompok Rumah Tangga di Kabupaten Cilacap dengan Menggunakan Metode MARS ..... 50

4.1.2.1 Pemodelan MARS ..... 51

4.1.2.2 Interpretasi Model MARS ..... 54

4.1.2.3 Pendugaan Klasifikasi dengan Model MARS ..... 64

4.1.3 Kesalahan Klasifikasi Model MARS ..... 66

4.1.4 Uji *Press's Q* ..... 67

4.2 Pembahasan ..... 68

**BAB 5 PENUTUP**

5.1 Kesimpulan ..... 71

5.2 Saran ..... 72

**DAFTAR PUSTAKA** ..... 72

**LAMPIRAN-LAMPIRAN** ..... 76

UNNES
UNIVERSITAS NEGERI SEMARANG

# DAFTAR TABEL

Tabel Halaman

Tabel 2.1.Matriks Konfusi untuk Klasifikasi Dua Kelas ..... 29

Tabel 4.1.Statistik Jumlah Pendapatan dan Pengeluaran Kepala Rumah Tangga 50

Tabel 4.2.*Trial and error* kombinasi BF, MI dan MO ..... 51

Tabel 4.3.Tingkat kepentingan variabel..... 62

Tabel 4.4.Matriks konfusi untuk klasifikasi kelompok rumah tangga..... 66

UNNES
UNIVERSITAS NEGERI SEMARANG

# DAFTAR GAMBAR

Gambar Halaman

Gambar 2.1. Menu Utama MARS 2.0 ..... 38

Gambar 2.2. Data dalam program SPSS versi 16.0 ..... 38

Gambar 2.3. Tampilan variabel *view* dalam program SPSS 16.0 ..... 39

Gambar 2.4. Memberi label untuk data yang memiliki kategori ..... 39

Gambar 2.5. Tampilan Membuka Dokumen ..... 40

Gambar 2.6. Tampilan variabel – variabel dalam *software* MARS 2.0 ..... 40

Gambar 2.7. Data dalam *software* MARS 2.0 ..... 41

Gambar 2.8. Model dalam *software* MARS 2.0 ..... 41

Gambar 2.9. *Options and Limits* dalam *software* MARS 2.0 ..... 42

Gambar 3.1. *Flowchart* Langkah-Langkah Penelitian ..... 47

Gambar 4.1. Prosentase kelompok rumah tangga di Kabupaten Cilacap ..... 49

UNNES
UNIVERSITAS NEGERI SEMARANG

# DAFTAR LAMPIRAN

Lampiran Halaman

Lampiran 1. Data hasil SUSENAS tahun 2014, Kabupaten Cilacap ..... 76

Lampiran 2. Data penelitian beserta kategorinya ..... 84

Lampiran 3. Hasil klasifikasi dengan metode MARS ..... 87

Lampiran 4. Output pengoolahan data pada *software* MARS 2.0 ..... 90

Lampiran 5. Tabel *Chi-square* ..... 94

UNNES
UNIVERSITAS NEGERI SEMARANG

# DAFTAR SIMBOL

| | |
|---|---|
| $Y$ | : data kelompok rumah tangga Kabupaten Cilacap tahun 2014 |
| $X_1$ | : data variabel jenis kelamin kepala rumah tangga |
| $X_2$ | : data variabel umur kepala rumah tangga |
| $X_3$ | : data variabel Ijazah/STTB tertinggi yang dimiliki |
| $X_4$ | : data variabel pendapatan kepala rumah tangga |
| $X_5$ | : data variabel pengeluaran perkapita rumah tangga |
| $X_6$ | : data variabel status penguasaan bangunan tempat tinggal |
| $X_7$ | : data variabel pengalaman membeli raskin |
| $X_8$ | : data variabeljumlah anggota keluarga |
| $X_9$ | : data variabel jenis atap terluas |
| $X_{10}$ | : data variabel jenis dinding terluas |
| $X_{11}$ | : data variabel jenis lantai terluas |
| $X_{12}$ | : data variabel sumber air minum utama |
| $X_{13}$ | : data variabel penggunaan tempat buang air besar |
| $X_{14}$ | : data variabel tempat pembuangan akhir tinja |
| $X_{15}$ | : data variabel bahan bakar/energi untuk memasak |
| BF | : fungsi basis (*Basis Function*) |
| MI | : maksimum Interaksi |
| MO | : minimum Observasi |
| n | : jumlah sampel |
| p | : jumlah prediktor |

$y_i$ : variabel respon pada amatan ke-$i$

$f(x_i)$ : fungsi *smooth* yang tidak diketahui

$\varepsilon_i$ : *error* ke-$i$ yang saling bebas

$a_0$ : fungsi basis induk

$K_m$ : derajat Interaksi ke-$m$

$S_{km}$ : nilai 1 atau -1 jika data berada di sebelah kanan atau kiri knot

$x_v(k,m)$ : variabel prediktor dari p dengan observasi ke-$n$

$t_{km}$ : nilai knot dari variabel prediktor $x_v(k,m)$

$f_i(x_i)$ : semua fungsi basis untuk satu variabel

$f_i(x_i, x_j)$ : semua fungsi basis untuk interaksi antar dua variabel

$f_i(x_i, x_j, x_k)$ : semua fungsi basis untuk interaksi antar tiga variabel

$\pi(x)$ : nilai peluang untuk membandingkan data masuk pada kategori 1 atau 0

$y_0$ : data dengan kategori 0

$y_1$ : data dengan kategori 1

$f_{00}$ : Jumlah subjek dari $y_0$ tepat diklasifikasikan sebagai $y_0$

$f_{01}$ : Jumlah subjek dari $y_0$ tidak tepat diklasifikasikan sebagai $y_1$

$f_{10}$ : Jumlah subjek dari $y_1$ tidak tepat diklasifikasikan sebagai $y_0$

$f_{11}$ : Jumlah subjek dari $y_1$ tepat diklasifikasikan sebagai $y_1$

# BAB 1

# PENDAHULUAN

## 1.1 Latar Belakang

Masalah kemiskinan merupakan salah satu permasalahan pemerintah yang diprioritaskan dalam menyusun strategi pembangunan setiap negara termasuk Indonesia. Meskipun pertumbuhan ekonomi negara Indonesia semakin meningkat, namun jumlah penduduk juga semakin bertambah sehingga beban ekonomi juga terus bertambah. Pesatnya pertumbuhan penduduk yang tidak seimbang dengan meningkatnya *Gross Domestic Product* (GDP), mengakibatkan semakin luasnya kesenjangan sosial yang terjadi di masyarakat. Kemiskinan timbul apabila masyarakat tidak memiliki kemampuan-kemampuan utama, pengeluaran per kapita, pendidikan yang memadai, kondisi kesehatan yang baik, rasa aman, kepercayaan diri yang tinggi dan hak seperti kebebasan berbicara (Haughton dan Khandker, 2012). Menurut Badan Pusat Statistik (2014b), Provinsi Jawa Tengah memiliki jumlah penduduk miskin terbanyak di Indonesia. Kabupaten Cilacap merupakan salah satu kabupaten terluas di Jawa Tengah dimana di kabupaten tersebut masih menjadi salah satu kabupaten miskin di Provinsi Jawa Tengah, karena kekuatan ekonominya lebih banyak ditopang oleh sektor pertanian.

Menurut Badan Pusat Statistik (2014a), kelompok rumah tangga berdasarkan pengeluaran per kapita yang sudah dikonversikan berdasarkan garis

kemiskinan terdiri dari dua kelompok yaitu kelompok rumah tangga miskin dan kelompok rumah tangga tidak miskin.

Berdasarkan Kuisioner Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2014 yang diperoleh dari Badan Pusat Statistik, faktor-faktor yang diduga mempengaruhi kelompok rumah tangga adalah jenis kelamin kepala rumah tangga, umur kepala rumah tangga, ijazah tertinggi kepala rumah tangga, pendapatan kepala rumah tangga, pengeluaran perkapita rumah tangga, status penguasaan bangunan tempat tinggal, pengalaman membeli raskin, jumlah anggota keluarga, jenis atap terluas, jenis dinding terluas, jenis lantai terluas, sumber air minum utama, penggunaan tempat buang air besar tempat pembuangan akhir tinja dan bahan bakar/energi ntuk memasak.

Dalam ilmu statistika, banyak metode yang dapat digunakan untuk mengetahui pengaruh variabel prediktor terhadap variabel responnya yang bersifat kategori. Salah satunya adalah metode *Multivariate Adaptive Regression Spline* (MARS). Metode tersebut termasuk dalam metode analisis regresi. Analisis regresi merupakan alat yang sangat terkenal dalam statistika terapan untuk penelitian bidang ekonomi sumber daya manusia, ekonometrika, pemasaran dan bidang-bidang lainnya. Keunggulan regresi adalah kemampuannya untuk membantu memberikan penjelasan secara statistik pengaruh variabel-variabel prediktor terhadap variabel respon (Bambang, 2008:53). Pada umumnya analisis regresi digunakan untuk melihat hubungan dan pengaruh variabel prediktor terhadap variabel respon, dimana variabel responnya berupa data kuantitatif.

UNNES
UNIVERSITAS NEGERI SEMARANG

Metode statistika yang dapat digunakan untuk menganalisis dan mengklasifikasikan suatu objek dan variabel respon yang bersifat kualitatif atau kategori dengan sejumlah variabel prediktor kontinu atau kategori adalah metode MARS. MARS merupakan pendekatan untuk regresi multivariate nonparametrik yang dikembangkan oleh Friedman (1991). Teknik pemodelan MARS menjadi populer karena tidak mengasumsikan dan tidak menentukan tipe khusus seperti pada hubungan linier, kuadratik, kubik di antara variabel prediktor dan respon (Budiantara, 2006).

Model MARS difokuskan untuk mengatasi permasalahan dimensi tinggi, memiliki variabel banyak, serta ukuran sampel besar sehingga diperlukan perhitungan yang rumit berdasarkan nilai *Generalized Cross Validation* (GCV) terkecil. Pada MARS, beberapa hal yang perlu diperhatikan dalam pemilihan model yang paling optimum (terbaik) adalah jika nilai GCV dari model tersebut mempunyai nilai yang paling rendah (minimum) diantara model-model yang lain (Friedman, 1990). Penelitian mengenai klasifikasi dengan menggunakan MARS sebelumnya juga pernah dilakukan oleh Ghofar (2014) pada penelitian yang berjudul "*Klasifikasi Kelulusan Mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro Menggunakan Multivariatee Adaptive Regression Spline* (MARS)". Penelitian yang lain juga dilakukan oleh Wengang (2014) pada penelitian yang berjudul "*Multivariate Adaptive Regression Splines and Neural Network Models for Prediction of Pile Drivability*". Metode MARS sangat baik untuk klasifikasi karena akan menghasilkan sedikit kesalahan klasifikasi.

UNNES
UNIVERSITAS NEGERI SEMARANG

Metode klasifikasi merupakan bagian dari analisis statistika yang bertujuan memisahkan individu atau objek ke dalam suatu kelompok sehingga dapat diketahui suatu individu berada pada kelompok tertentu. Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau akan menghasilkan peluang kesalahan alokasi yang kecil (Johnson dan Wichern, 1992).

Berdasarkan uraian di atas, maka peneliti mengangkat topik yang berjudul "*Ketepatan Klasifikasi dengan Menggunakan Metode Multivariate Adaptive Regression Spline (MARS) pada Data Kelompok Rumah Tangga Kabupaten Cilacap*".

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, maka permasalahan penelitian ini dirumuskan sebagai berikut.

1. Bagaimana hasil klasifikasi berdasarkan variabel yang mempengaruhi kelompok rumah tangga di Kabupaten Cilacap dengan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS) ?
2. Bagaimana hasil perhitungan kesalahan klasifikasi kelompok rumah tangga di Kabupaten Cilacap dengan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS) ?

## 1.3 Batasan Masalah

Batasan masalah yang dilakukan pada penelitian ini adalah sebagai berikut.

1. Data yang digunakan pada penelitian ini adalah data rumah tangga hasil Survei Sosial dan Ekonomi Nasional (SUSENAS) tahun 2014 di Kabupaten Cilacap.
2. Penelitian hanya menggunakan metode *Multivariate Adaptive Regression Spline* (MARS) untuk menganalisis dan mengklasifikasi data rumah tangga di Kabupaten Cilacap.
3. Paket program yang mendukung penelitian ini adalah *software* SPSS 16.0 dan MARS 2.0.

## 1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut.

1. Mengklasifikasi kelompok rumah tangga di Kabupaten Cilacap berdasarkan faktor-faktor yang mempengaruhi tingkat kemiskinan di Kabupaten Cilacap dengan metode *Multivariate Adaptive Regression Spline* (MARS).
2. Menghitung besar kesalahan klasifikasi kelompok rumah tangga di Kabupaten Cilacap dengan menggunakan metode MARS.

UNNES
UNIVERSITAS NEGERI SEMARANG

## 1.5 Manfaat penelitian

Manfaat yang dapat diperoleh dalam penulisan skripsi ini diantaranya.

### 1.5.1 Bagi Mahasiswa

1. Mahasiswa memperoleh pengetahuan tentang metode *Multivariate Adaptive Regression Spline* (MARS).

2. Mahasiswa memperoleh pengetahuan tentang prosedur penggunaan metode *Multivariate Adaptive Regression Spline* (MARS).
3. Mahasiswa memperoleh pengetahuan tentang bagaimana cara mengklasifikasi dan menghitung kesalahan klasifikasi data dengan menggunakan metode *Multivariate Adaptive Regression Spline* (MARS).

### 1.5.2 Bagi Pembaca

1. Dapat menambah atau memperkaya khasanah kepustakaan Jurusan Matematika.
2. Menambah topik kajian tentang metode *Multivariate Adaptive Regression Spline* (MARS).

## 1.6 Sistematika Penulisan Skripsi

Secara garis besar skripsi ini dibagi menjadi tiga bagian (bab) yaitu bagian awal skripsi, bagian isi skripsi, dan bagian akhir skripsi. Berikut ini dijelaskan masing-masing bagian skripsi.

1. Bagian awal skripsi

   Bagian awal skripsi meliputi halaman judul, pernyataan keaslian tulisan, pengesahan, motto dan persembahan, kata pengantar, abstrak, daftar isi, daftar gambar, daftar tabel, dan daftar lampiran.
2. Bagian isi skripsi

   Bagian isi skripsi secara garis besar terdiri dari lima bab, yaitu:

UNNES
UNIVERSITAS NEGERI SEMARANG

BAB 1 PENDAHULUAN

Bab ini berisi mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

BAB 2 TINJAUAN PUSTAKA

Bab ini berisi kajian teori yang mendasari dan berhubungan dengan pemecahan masalah. Teori-teori tersebut digunakan untuk memecahkan masalah yang diangkat dalam skripsi ini. Teori yang digunakan adalah Analisis Regresi, Regresi Nonparametrik, *Recursive Partitioning Regression* (RPR), Regresi Spline, *Multivariate Adaptive Regression Splines* (MARS), Klasifikasi, Klasifikasi pada MARS, *Apparent Error Rate (APER),* Uji *Press Q*, Kemiskinan, garis kemiskinan, indikator kemiskinan, SPSS versi 16.0, dan *Software* MARS 2.0 yang akan digunakan untuk membantu menyelesaikan masalah yang ada.

BAB 3 METODE PENELITIAN

Bab ini mengulas metode yang digunakan dalam penelitian yang berisi langkah-langkah yang dilakukan untuk memecahkan masalah yaitu pengumpulan data, analisis data dan kesimpulan.

BAB 4 HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi mengenai penyelesaian dari permasalahan yang diungkapkan.

BAB 5 PENUTUP

Bab ini berisi tentang simpulan dari pembahasan dan saran yang berkaitan dengan simpulan.

UNNES
UNIVERSITAS NEGERI SEMARANG

3. Bagian akhir skripsi

Bagian akhir skripsi meliputi daftar pustaka yang memberikan informasi tentang buku sumber serta literatur yang digunakan dan lampiran-lampiran yang mendukung skripsi.

UNNES
UNIVERSITAS NEGERI SEMARANG

# BAB 2

# TINJAUAN PUSTAKA

## 2.1 Statistik Deskriptif

Statistik deskriptif merupakan metode analisis yang berkaitan dengan pengumpulan dan penyajian data sehingga menghasilkan informasi yang bermanfaat. Bentuk penyajian statistik deskriptif dapat berupa tabel, grafik, diagram, histogram, dan lainnya (Walpole, 1993). Pada penelitian ini statistik deskriptif digunakan untuk memperoleh gambaran data secara umum mengenai prosentase data kelompok rumah tangga di Kabupaten Cilacap dan statistik jumlah pendapatan dan pengeluaran kepala rumah tangga.

## 2.2 Analisis Regresi

Regresi adalah suatu alat yang tujuaannya untuk memprediksi pengaruh suatu variabel atau beberapa variabel prediktor terhadap variabel respon (Iriawan, 2006:199). Analisis regresi didefinisikan sebagai kajian terhadap hubungan satu variabel atau peubah yang disebut peubah yang diterangkan (*the explained variable*) yang juga disebut sebagai peubah tergantung dengan satu atau dua variabel/peubah yang menerangkan (*the explanatory*) yang juga disebut peubah bebas.

Model regresi memiliki variabel respon $y$ dan variabel prediktor $x$. Variabel respon adalah variabel yang dipengaruhi suatu variabel prediktor.

Variabel respon sering dikenal dengan variabel dependen karena peneliti tidak bebas mengendalikannya. Kemudian, variabel prediktor digunakan untuk memprediksi nilai variabel respon dan sering disebut dengan variabel independen karena peneliti bebas mengendalikannya. Kedua variabel dihubungkan dalam bentuk persamaan matematika. Secara umum, bentuk persamaan regresi dinyatakan sebagai berikut.

$$\hat{y}_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{2.1}$$

dengan,

$\hat{y}_i$ = variabel terikat (nilai taksiran $\hat{y}$)

$x_1, x_2, \ldots, x_k$ = variabel bebas

$\beta_0, \beta_1, \ldots \beta_k$ = koefisien regresi liniear berganda

$\varepsilon$ = error (nilai sisaan)

Konstanta atau parameter $\beta_0$ yang disebut intersep adalah nilai variabel respon ketika variabel prediktor bernilai 0 (nol). Kemudian, $\beta_1, \beta_2, \ldots, \beta_k$ adalah parameter-parameter model regresi untuk variabel $x_1, x_2, \ldots x_k$. Dalam model regresi, baik variabel respon maupun variabel prediktor harus berskala kontinyu. Artinya skala data untuk kedua variabel harus rasio atau interval (Astuti, 2006:200). Pada model regresi dalam persamaan (2.1) terdapat pula istilah $\varepsilon$ adalah *error* atau sering disebut residual. Residual adalah jarak antara nilai sebenarnya dengan garis model taksiran yang dirumuskan $\varepsilon = (y - \hat{y})$. Dalam

hal ini $y$ adalah nilai sebenarnya variabel respon dan $\hat{y}$ adalah nilai taksiran variabel respon.

Analisis regresi dilakukan jika variabel independen memiliki pengaruh terhadap variabel dependen. Ada atau tidaknya pengaruh antar variabel dapat diketahui dengan mempelajari lebih jauh mengenai kedua variabel. Salah satu cara mengenalinya adalah mengetahui dari literatur lain, yang mengatakan bahwa kedua variabel yang akan diteliti memiliki hubungan.

### 2.2.1 Regresi Nonparametrik

Pendekatan nonparametrik merupakan metode pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva regresi tertentu dimana kurva regresi hanya diasumsikan *smooth* (mulus), artinya termuat di dalam suatu ruang fungsi tertentu sehingga regresi nonparametrik memiliki fleksibilitas yang tinggi karena data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh faktor subyektifitas peneliti (Eubank 1988). Penggunaan statistik nonparametrik mengabaikan asumsi-asumsi yang melandasi penggunaan metode statistika parametrik. Tidak semua pola hubungan dapat didekati dengan pendekatan parametrik, karena tidak adanya suatu informasi mengenai bentuk hubungan variabel respon dan variabel prediktor. Jika bentuk kurva tidak diketahui dan pola menyebar maka kurva regresi dapat diduga menggunakan pendekatan model regresi nonparametrik. Apalagi tes nonparametrik tidak memerlukan pengukuran seperti yang dibutuhkan untuk menyelesaikan tes parametrik, tes nonparametrik paling berlaku untuk data dalam

UNNES
UNIVERSITAS NEGERI SEMARANG

skala ordinal, dan berlaku juga untuk sampel data skala nominal. Secara umum model regresi nonparametrik dapat dituliskan sebagai berikut.

$$y_i = f(x_i) + \varepsilon_i \tag{2.2}$$

dengan,

$y_i$ : variabel respon pada amatan ke-$i$

$f(x_i)$ : fungsi *smooth* yang tidak diketahui

$\varepsilon_i$ : error ke-$i$ yang saling bebas

Fungsi regresi $f(x_i)$ hanya diasumsikan termuat dalam suatu ruang fungsi tertentu, dimana pemilihan ruang fungsi tersebut biasanya dimotivasi oleh sifat kemulusan (*smoothness*) yang dimiliki oleh fungsi $f(x_i)$ tersebut. Salah satu pendekatan dalam regresi nonparametrik adalah regresi spline. Spline mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus (Hardle, 1990).

UNNES
UNIVERSITAS NEGERI SEMARANG

***2.2.1.1 Regresi Spline***

Regresi spline merupakan salah satu metode regresi nonparametrik yang bertujuan untuk memperkecil keragaman dan mengestimasi perilaku data yang cenderung berbeda. Pendekatan spline memiliki kemampuan untuk mengatasi pola data yang menunjukkan naik turun yang tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus. Titik-titik knot ialah perpaduan

bersama yang menunjukkan terjadinya perubahan spline untuk menyesuaikan diri secara efektif terhadap karakteristik lokal dari suatu fungsi atau data. Spline merupakan potongan (*piecewise*) polinomial orde $q$ dan memiliki turunan yang kontinu dengan knot sampai orde $(q-1)$ (Friedman, 1991). Dalam spline univariat dengan K knot memiliki fungsi basis sebagai berikut.

$$\{x^j\}_1^q, \{(x - t_k)_+^q\}_1^K \tag{2.3}$$

Sehingga model *spline* dapat ditulis menjadi,

$$f(x) = \beta_0 + \beta_1 x + \cdots + \sum_{k=1}^{K} \gamma_k \ (x - t_x)_+^q \tag{2.4}$$

dimana $q \geq 1$ dan $(t_1, t_2, \ldots, t_k)$ merupakan titik-titik knot. Sedangkan fungsi dari *truncated power* $(x - t_x)_+^q$ adalah

$$(x - t_x)_+^q = \begin{cases} (x - t_x)_+^q & ; x > t_k \\ 0 & ; x \leq t_k \end{cases} \tag{2.5}$$

dimana $q$ menunjukkan orde polinomial dari fungsi spline. Di setiap titik knot, diharapkan adanya dari fungsi basis antar satu *region* dengan *region* lainnya. Oleh karena itu pada umumnya fungsi basis yang dipilih adalah berbentuk polinomial dengan turunan $(q-1)$ yang kontinu disetiap titik knot.

***2.2.1.2 Basis B-Spline***

Model regresi $y_i = g(t_i) + \varepsilon_i : i = 1, 2, \dots, n,$ merupakan residual dan $g(t_i)$ adalah kurva regresi. B-Spline mengatasi kelemahan regresi spline yaitu saat order spline tinggi, knot yang banyak dan knot yang terlalu dekat akan membentuk matriks dalam perhitungan yang hampir singular, sehingga persamaan tidak dapat diselesaikan. Apabila digunakan pendekatan kurva *spline truncated* dikatakan regresi nonparametrik, maka kurva regresi $g$ dapat ditulis sebagai berikut (Eubank, 1988).

$$g(t) = \sum_{i=1}^{m} \alpha_i t^{i-1} + \sum_{j=1}^{K} \beta_i (t - u_j)_+^{m-1} \tag{2.6}$$

dengan $u_j,\ j = 1,2, \dots, K,$ dengan $u_1 < u_2 < \cdots < u_K$ adalah titik knot dan $m\ \varepsilon\ N_o$ (integer non positif). Nilai $m$ menunjukkan derajat *spline truncated*, yang juga merupakan potongan polinomial berderajat $m$−1 dengan $m$−2 turunan kontinu di titik knot. Jika kurva regresi $g$ didekati dengan fungsi B–Spline maka $g$ dapat ditulis menjadi,

$$g(t) = \sum_{j=1}^{m+K} \gamma_j \beta_j\ - m, m(t) \tag{2.7}$$

dengan $\beta_j - m, m$ merupakan basis B-Spline.

Cara membangun fungsi B-spline orde m dengan titik-titik knot $\alpha < u_j < \cdots < u_K < b$ adalah dengan terlebih dahulu mendefinisikan knot tambahan sebanyak 2m, yaitu $u_{-(m-1)}, \dots, u_{-1}, u_0, \dots, u_{K+m}$ dimana $u_{-(m-1)} = \cdots = u_0 = a$ dan $u_{K+1} = \dots = u_{K+m} = b$, biasanya $a$ diambil dari nilai minimum t dan $b$ diambil dari nilai maksimum t. Fungsi B-spline didefinisikan secara rekursif sebagai berikut (Botella dan Shariff, 2003).

$$\beta_{i-m}(t) = \frac{t - u_i}{u_{i+m-1} - u_i} B_{i,m-1}(t) + \frac{u_{i+m} - t}{u_{i+m} - u_{i+1}} B_{i,m-1}(t) \tag{2.8}$$

$$\beta_j(t) = \begin{cases} 1, u_j < t \leq u_{j+1} \\ 0, t < u_j, t \geq u_j \end{cases} \tag{2.9}$$

dengan m adalah derajat dari B–Spline. Untuk $m$=2 memberikan fungsi B–Spline linier $m = 3$ memberikan fungsi B–Spline kuadratik dan $m$=4 memberikan fungsi B–Spline kubik. Untuk mengestimasi koefisien γ dapat didefinisikan matriks sebagai berikut.

$$\beta(\lambda) = \left(\beta_{j,m}(t)\right)_{i=1,\dots,n_{j=-(m-1),\dots,K}} \tag{2.10}$$

atau dapat ditulis sebagai berikut.

$$\beta(\lambda) = \begin{bmatrix} B_{-(m-1),m(t_1)} & B_{-(m-2),m(t_2)} & B_{K,m(t_1)} \\ \vdots & \vdots & \vdots \\ B_{-(m-1),m(t_n)} & B_{-(m-2),m(t_n)} & B_{K,m(t_n)} \end{bmatrix} \tag{2.11}$$

Jadi ($\lambda$) adalah sebuah matriks berukuran $n$×($m$+$K$).

Sebagai gambaran untuk menjelaskan fungsi B–Spline, misalnya B–Spline linier ($m$ = 2), dengan satu (1) titik knot, pada $t$ = 5, dengan nilai $t$ minimum 1 dan nilai $t$ maksimum 10. Maka langkahnya adalah menentukan knot tambahan sebanyak $2m$, yaitu diambil dari nilai minimum 1 dan maksimum 10, sehingga knot menjadi $u_{-1}$= $u_0$=1,$u_1$=5, $u_2$ = $u_3$=10, maka matriks yang akan dibentuk adalah $\beta(\lambda)$= ($B_{-1,2}(t_i)$, $B_{0,2}(t_i)$, $B_{1,2}(t_i)$),$i$=1,2,…,$n$ yaitu sebuah matriks dengan ukuran $n$ × 3. Dengan persamaan (2.8), $B_{-1,2}(t_i)$ dapat ditulis sebagai berikut.

$$B_{-1,2}(t) = \frac{t-1}{1-1} B_{-1,1}(t) + \frac{5-t}{5-1} B_{0,1}(t) \tag{2.12}$$

$B_{-1,1}(t)$ didefinisikan bernilai 0 karena $u_{-1} = u_0$ (Eubank, 1988). Dan $B_{0,1}(t)$ akan bernilai 1 pada t bernilai $u_0 = 0$ sampai dengan $u_1 = 1$, dan bernilai 0 untuk yang lain, sehingga dapat ditulis seperti persamaan (2.9) sebagai berikut.

$$B_{-1,2}(t) = \begin{cases} \frac{5-t}{4}, 1 < t \leq 5 \\ 0, \ 5 < t \ \leq 10 \end{cases} \tag{2.13}$$

Sedangkan untuk basis $B_{0,2}(t)$ dengan menggunakan persamaan (2.11) dapat ditulis sebagai berikut.

$$B_{0,2}(t) = \frac{t-1}{5-1}, B_{0,1}(t) + \frac{10-t}{10-1} B_{1,1}(t) \tag{2.14}$$

dan dapat juga ditulis seperti persamaan (2.13) sebagai berikut.

$$B_{0,2}(t) = \begin{cases} \frac{t-1}{4}, & 1 < t \leq 5 \\ \frac{10-t}{5}, & 5 < t \leq 10 \end{cases} \tag{2.15}$$

Untuk basis $B_{1,2}(t_i)$ dapat menggunakan persamaan (2.12) dan dapat ditulis sebagai berikut.

$$B_{1,2}(t) = \frac{t-5}{10-5}, B_{1,1}(t) + \frac{10-t}{10-1} B_{2,1}(t)$$

$$B_{1,2}(t) = \frac{t-5}{10-5} B_{1,1}(t) \tag{2.16}$$

dan dapat ditulis juga seperti persamaan (2.11) sebagai berikut.

$$B_{1,2}(t) = \begin{cases} 0, & 1 < t \leq 5 \\ \frac{t-5}{5}, & 5 < t \leq 10 \end{cases} \tag{2.17}$$

Untuk kurva B–Spline kuadratik dengan 2 titik knot, misalnya pada $t$ =3 dan $t$ =7 dapat dicari dengan cara yang serupa, dengan hasil sebagai berikut.

$$B_{-1,3}(t) = \begin{cases} \frac{(t-1)(13-3t)}{12}, & 1 < t \leq 3 \\ \left[\frac{(7-t)^2}{24}\right], & 3 < t \leq 7 \\ 0, & 3 < t \leq 7 \end{cases} \tag{2.18}$$

dengan cara sama, dapat dibuat kurva B–Spline dengan berbagai $m$ dan beberapa titik knots.

## 2.3 *Multivariate Adaptive Regression Spline* (MARS)

Model MARS berguna untuk mengatasi permasalahan data berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor sebesar $3 \leq n \leq 20$. dan menghasilkan prediksi variabel respon yang akurat, serta menghasilkan model yang kontinu dalam knot berdasarkan nilai GCV terkecil (Friedman, 1991). MARS dikembangkan oleh Friedman (1991) untuk pendekatan model regresi Multivariatee nonparametrik antara variabel respon dan beberapa variabel prediktor pada *piecewise* regresi.

*Piecewise* regresi merupakan regresi yang memiliki sifat tersegmen. Apabila suatu garis regresi tidak bisa menjelaskan keseluruhan data maka beberapa garis regresi digunakan untuk menjelaskan seluruh data yang ada dari variabel yang independen. Tempat perubahan pola itulah yang dinamakan knot. Perlu dilakukan pengujian untuk mengontrol jumlah derajat bebas (DF) untuk optimalisasi knot. Sedangkan basis fungsi merupakan suatu fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor. Pada umumnya basis fungsi yang dipilih adalah berbentuk polinomial dengan turunan yang kontinu pada setiap titik knot.

Friedman (1991) menyarankan jumlah maksimum basis fungsi (BF) dua sampai empat kali jumlah variabel prediktor. Jumlah interaksi maksimum (MI) adalah satu, dua, atau tiga dengan pertimbangan jika lebih dari tiga akan

menghasilkan model yang sangat kompleks. Jarak minimum antar knot atau observasi minimum (MO) antar knot sebanyak 0, 10, 20, 50, dan 100. Sedangkan jumlah sampel yang diharuskan untuk pendekatan MARS adalah $50 \leq N \leq 1000$. MARS merupakan pengembangan dari pendekatan *Recursive Partitioning Regression* (RPR) dan rekursif. Beberapa hal yang perlu diperhatikan dalam menggunakan model MARS sebagai berikut (Nisa' dan Budiantara, 2012).

1. *Knot*

*Knot* yaitu akhir dari sebuah garis regresi (*region*) dan awal dari sebuah garis regresi (*region*) yang lain. Di setiap titik knot, diharapkan adanya kontinuitas dari fungsi basis antar satu *region* dengan *region* lainnya.

2. *Basis Function* (Fungsi Basis/BF)

*Basis Function* yaitu suatu fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor. Fungsi basis ini merupakan fungsi parametrik yang didefinisikan pada tiap *region*. Pada umumnya fungsi basis yang dipilih adalah berbentuk polinomial dengan turunan yang kontinu pada setiap titik knot. Friedman menyarankan jumlah maksimum fungsi basis (BF) adalah 2-4 kali jumlah variabel prediktornya.

UNNES
UNIVERSITAS NEGERI SEMARANG

3. *Interaction* (Interaksi)

Interaksi merupakan hasil perkalian silang antara variabel yang saling berkorelasi. Friedman membatasi jumlah maksimum interaksi (MI) yang diperbolehkan yaitu 1, 2, dan 3. Apabila terdapat lebih dari 3 interaksi, maka akan menimbulkan interpretasi model yang sangat kompleks dan sulit untuk diinterpretasikan. Maksimum interaksi (MI) yaitu untuk maksimum garis BF yang

dapat melewati knotnya. MI=1 artinya bahwa di dalam modelnya maksimum garis BF dapat melewati 1 titik knot, MI=2 artinya bahwa di dalam modelnya maksimum garis BF dapat melewati 2 titik knot, dan MI=3 artinya bahwa di dalam modelnya maksimum garis BF dapat melewati 3 titik knot.

Pemodelan MARS ditentukan berdasarkan *trial and error* untuk kombinasi BF, MI, dan MO untuk mendapatkan nilai dari parameter pemulus yang minimum. MO yaitu minimum jarak antara knot atau minimum observasi antara knot (MO) sebesar 0, 1, 2, dan 3. MO = 0 artinya bahwa di dalam modelnya jarak antara titik knot 0, MO = 1 artinya bahwa di dalam modelnya minimum jarak antara titik knot 1, MO = 2 artinya bahwa di dalam modelnya minimum jarak antara titik knot 2, dan MO = 3 artinya bahwa di dalam modelnya minimum jarak antara titik knot 3 (Nisa' dan Budiantara, 2012).

### 2.3.1 Model MARS

Menurut Friedman (1991) perhitungan komputasi sangat diperlukan dalam pengolahan data, karena akan selalu memperoleh hasil yang sesuai dengan algoritmanya. *Recursive Partitioning Regression* (RPR) merupakan salah satu dari program komputasi yang memiliki keunggulan dalam mengolah data yang berdimensi tinggi. Tujuan dari RPR adalah menggunakan data untuk mengestimasi *subregion* dan parameter yang berasosiasi pada setiap *subregion.* Pendekatan RPR dimulai dengan menghasilkan model yang memiliki sejumlah fungsi basis untuk meningkatkan kecocokan model.

UNNES
UNIVERSITAS NEGERI SEMARANG

$$\hat{f}(x) = \sum_{m=1}^{M} a_m B_m(x) \tag{2.19}$$

$a_m$ merupakan koefisien dari fungsi basis ke- $m$ dan fungsi basis $B_m$ diambil dari

$$B_m(x) = I\,[x \in R_m] \tag{2.20}$$

dimana $I$ adalah fungsi indikator yang memiliki nilai satu jika $x \in R_m$ dan bernilai 0 untuk $x \notin R_m$.

Misalkan $(g)$ adalah prosedur perhitungan *lack-of-fit* dan fungsi $(x)$ dan $[\eta]$ adalah fungsi yang memiliki nilai positif, yaitu

$$H[\eta] = \begin{cases} 1\,, \eta \geq 0 \\ 0, lainnya \end{cases} \tag{2.21}$$

Sehingga prosedur regresi *stepwise forward* menghasilkan

$$B_m(x) = \prod_{k=1}^{Km} H\left[s_{km}(x_{v(km)} - t_{km})\right] \tag{2.22}$$

Menurut Friedman (1991) keterbatasan mendasar yang dimiliki oleh model RPR adalah kurangnya kontinuitas. Model yang dihasilkan terpotong dan tidak kontinu pada batas subregion. Fungsi tangga pada RPR dapat dimodifikasi sehingga menghasilkan model yang kontinu dengan memandang fungsi tangga

sebagai bentuk khusus dari basis fungsi spline. Basis fungsi terpotong (*truncanted*) dua arah adalah sebagai berikut.

$$b_q^{\pm}(x-t) = [\pm(x-t)]_+^q \tag{2.23}$$

dengan $t$ merupakan lokasi knot, $q$ merupakan orde spline dan tanda positif di bagian bawah mengindikasikan bagian positif dari argumen, jika dijabarkan fungsinya menjadi sebagai berikut.

$$(x-t)_+ = \begin{cases} (x-t), jika\ x > t \\ 0, untuk\ yang\ lain \end{cases} \tag{2.24}$$

Model MARS adalah hasil modifikasi model RPR dengan kombinasi spline. Model MARS digunakan untuk mengatasi kelemahan *Recursive Partitioning Regression* (RPR) yaitu menghasilkan model yang kontinu pada knot dan dapat mengidentifikasi adanya fungsi linear dan aditif. Hasil modifikasi model *recursive partitioning regression* (RPR) dengan kombinasi *splines* oleh Friedman (1991) adalah model *Multivariatee adaptive regression splines* (MARS) sebagai berikut.

$$f(x) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{K_m} \left[ s_{km}\left(x_{v(k,m)} - t_{km}\right)\right]_+ + \varepsilon \qquad (2.25)$$

dengan,

$\alpha_0$ = konstanta regresi dari fungsi basis

$\alpha_m$ = koefisien dari fungsi basis ke-*m*

M = maksimum fungsi basis (*nonconstant* fungsi basis)

$K_m$ = derajat interaksi

$S_{km}$ = bernilai +1 jika *knot* terletak di kanan titik *knot*, dan bernilai -1 jika *knot* terletak di kiri titik *knot*

$x_{v(k,m)}$ = variabel prediktor

$t_{km}$ = nilai *knot* dari variabel prediktor $x_{v(k,m)}$

Berdasarkan persamaan (2.25), model MARS dapat ditulis sebagai berikut.

$$f(x) = \alpha_0 + \sum_{m=1}^{M} \alpha_m B_m(x) + \varepsilon \qquad (2.26)$$

dengan $B_m(x) = \prod_{k=1}^{K_m} \left[ s_{km}\left(x_{v(k,m)} - t_{km}\right)\right]_+$

sehingga dapat dituliskan dalam bentuk matriks menjadi seperti berikut.

$$Y = B\alpha + \varepsilon \qquad (2.27)$$

dengan,

$$\boldsymbol{Y} = (Y_1, \dots, Y_N)^T, \boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_M)^T, \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_N)^T$$

$$\mathbf{B} = \begin{bmatrix} 1 & \prod_{k=1}^{K_1}\left[s_{k1}\left(x_{v1(k,1)} - t_{k1}\right)\right]_+ & \cdots & \prod_{k=1}^{K_M}\left[s_{kM}\left(x_{v1(k,M)} - t_{kM}\right)\right]_+ \\ 1 & \prod_{k=1}^{K_1}\left[s_{k1}\left(x_{v2(k,1)} - t_{k1}\right)\right]_+ & \cdots & \prod_{k=1}^{K_M}\left[s_{kM}\left(x_{v2(k,M)} - t_{kM}\right)\right]_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1}\left[s_{k1}\left(x_{vN(k,1)} - t_{k1}\right)\right]_+ & \cdots & \prod_{k=1}^{K_M}\left[s_{kM}\left(x_{vN(k,M)} - t_{kM}\right)\right]_+ \end{bmatrix}$$

Persamaan (2.25) dapat dituliskan kedalam bentuk sebagai berikut.

$$\hat{f}(x) = \alpha_0 + f_i(x_i) + f_{ij}(x_i, x_j) + f_{ijk}(x_i, x_j, x_k) + \cdots \tag{2.28}$$

Persamaan (2.28) menunjukkan bahwa penjumlahan pertama meliputi semua fungsi basis untuk satu variabel, penjumlahan kedua meliputi semua fungsi basis untuk interaksi antara dua variabel, penjumlahan ketiga meliputi semua fungsi basis untuk interaksi antara tiga variabel dan seterusnya. Misalkan V(m) = $\{v(k,m)\}_1^{Km}$ adalah himpunan dari variabel yang dihubungkan dengan fungsi basis $B_m$ ke-m, maka setiap penjumlahan pertama pada persamaan (2.28) dapat dituliskan sebagai berikut.

$$f_i(x_i) = \sum_{\substack{K_m=1 \\ i \in V(m)}} \alpha_m B_m(x_i) \tag{2.29}$$

$f_i(x_i)$ merupakan penjumlahan semua fungsi basis untuk satu variabel $x_i$ dan merupakan *spline* dengan derajat q=1 yang mempresentasikan fungsi *univariate*. Setiap fungsi *bivariate* pada persamaan (2.28) dapat dituliskan sebagai berikut.

$$f_{ij}(x_i, x_j) = \sum_{\substack{K_m=2 \\ (i,j)\in V(m)}} \alpha_m B_m(x_i, x_j) \tag{2.30}$$

dengan $B_m(x_i, x_j) = [s_{2m}(x_{i(2,m)} - t_{2m})]_+ [s_{2m}(x_{j(2,m)} - t_{2m})]_+$

Persamaan (2.30) merupakan penjumlahan semua fungsi basis dua variabel $x_i$ dan $x_j$. Untuk fungsi *trivariate* pada penjumlahan yang ketiga diperoleh dengan menjumlahkan semua fungsi basis untuk tiga variabel prediktor, yang dituliskan sebagai berikut.

$$f_{ijk}(x_i, x_j, x_k) = \sum_{\substack{K_m=3 \\ (i,j,k)\in V(m)}} \alpha_m B_m(x_i, x_j, x_k) \tag{2.31}$$

dengan,

$$B_m(x_i, x_j, x_k) = [s_{3m}(x_{i(3,m)} - t_{3m})]_+ [s_{3m}(x_{j(3,m)} - t_{3m})]_+ [s_{3m}(x_{k(3,m)} - t_{3m})]_+ \tag{2.32}$$

Persamaan (2.28) dikenal dengan dekomposisi ANOVA dari model MARS. Interpretasi model MARS melalui dekomposisi ANOVA adalah mempresentasikan variabel prediktor yang masuk ke dalam model, baik untuk satu variabel prediktor maupun interaksi lebih dari satu variabel prediktor.

Penentuan *knot* pada metode MARS dapat ditentukan secara otomatis oleh data dan menghasilkan model yang kontinu pada *knot.* Penentuan *knot* pada MARS menggunakan *forward stepwise* dan *backward stepwise*. Untuk mendapatkan model dengan fungsi basis maksimum digunakan *forward stepwise*. Sedangkan *backward stepwise* digunakan untuk memenuhi prinsip parsimoni (model yang sederhana) dengan cara mengeliminasi fungsi basis yang berkontribusi kecil sampai tidak ada fungsi basis yang dapat dieliminasi. Ukuran kontribusi pada *backward stepwise* ditentukan dengan kriteria *Generalized Cross Validation* (GCV) (Pintowati dan Otok, 2012). Bentuk fungsi GCV minimum menurut Friedman (1991) dapat didefinisikan pada persamaan berikut:

$$GCV(M) = \frac{\frac{1}{N}\sum_{i=1}^{N}\left[y_i - \hat{f}_M(x_i)\right]^2}{\left[1 - \frac{\tilde{C}(M)}{N}\right]^2} \tag{2.33}$$

dengan,

N = ukuran sampel

$y_i$ = variabel respon

$x_i$ = variabel prediktor

M = banyaknya fungsi basis (*nonconstant* fungsi basis)

$\tilde{C}(M)$ = $C(M) + dM$

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

d = nilai ketika setiap fungsi basis mencapai nilai optimal ($2 \leq d \leq 4$).

## 2.4 Klasifikasi

Klasifikasi merupakan pengelompokan obyek kedalam suatu kelas atau kategori yang telah ditentukan. Klasifikasi adalah teknik *Multivariate* untuk memisahkan set yang berbeda dari suatu objek dan dengan mengalokasikan objek 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, 1992).

Menurut Prasetyo (2012), sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar, sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi. Tabel 2.1 merupakan matriks konfusi yang melakukan klasifikasi biner (dua kelas) yaitu kelas 0 dan 1.

## 2.5 Klasifikasi pada MARS

Pada dasarnya, klasifikasi dilakukan untuk melihat seberapa besar ketepatan pengelompokan sekumpulan data untuk digolongkan dengan tepat pada kelompoknya. Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau peluang kesalahan klasifikasi yang kecil (Johnson dan Wichern, 1992).

Menurut Cox dan Snell (1989) jika variabel respon terdiri dari dua nilai, maka dapat dikatakan sebagai regresi dengan *binary response.* Sehingga didapatkan model probabilitas dengan persamaan sebagai berikut:

$$\pi(\boldsymbol{x}) = \frac{e^{\hat{f}(\boldsymbol{x})}}{1+e^{\hat{f}(\boldsymbol{x})}} \tag{2.34}$$

$$\{1-\pi(\boldsymbol{x})\} = \frac{1}{1+e^{\hat{f}(\boldsymbol{x})}} \tag{2.35}$$

dengan $\hat{f}(\boldsymbol{x}) = logit\ \pi(\boldsymbol{x})$

sehingga didapatkan $prob(Y=1) = \pi(\boldsymbol{x})$ dan $prob(Y=0) = 1-\pi(\boldsymbol{x})$

Karena variabel respon Y bersifat biner (0 dan 1) dengan variabel prediktor X sebanyak n, $\boldsymbol{x} = (x_1, \dots, x_n)$, maka model MARS untuk klasifikasi dapat dituliskan sebagai berikut,

$$logit\ \pi(\boldsymbol{x}) = ln\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{K_m} \left[s_{km}\left(x_{v(k,m)} - t_{km}\right)\right]_+ \tag{2.36}$$

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

## 2.6 *Apparent Error Rate* (APER) dan *Press's Q*

Menurut Prasetyo (2012), sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua *data set* dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Pada umumnya, pengukuran klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Jika $y_0$ dan $y_1$ merupakan subjek pengklasifikasian, maka bentuk matriks konfusi dapat dilihat pada tabel 2.1.

Tabel 2.1. Matriks Konfusi untuk Klasifikasi Dua Kelas

| Kelas Observasi (Actual Class) | Kelas Prediksi (Predicted Class) | |
|---|---|---|
| | Kelas 0 | Kelas 1 |
| Kelas 0 | $f_{00}$ | $f_{01}$ |
| Kelas 1 | $f_{10}$ | $f_{11}$ |

dengan,

$f_{00}$ = Jumlah subjek dari $y_0$ tepat diklasifikasikan sebagai $y_0$

$f_{01}$ = Jumlah subjek dari $y_0$ tidak tepat diklasifikasikan sebagai $y_1$

$f_{10}$ = Jumlah subjek dari $y_1$ tidak tepat diklasifikasikan sebagai $y_0$

$f_{11}$ = Jumlah subjek dari $y_1$ tepat diklasifikasikan sebagai $y_1$

Setiap sel $f_{ij}$ dalam matriks menyatakan jumlah data dari kelas i yang hasil prediksinya masuk ke kelas j, $f_{11}$ 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, $f_{01}$ adalah jumlah data dalam kelas 0 yang dipetakan

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

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. Perhitungan besarnya nilai APER dapat dihitung dengan menggunakan rumus,

$$APER = \frac{f_{01} + f_{10}}{f_{00} + f_{01} + f_{10} + f_{11}} \text{ x } 100\% \qquad (2.37)$$

Menurut Hair *et al* (2006), *Press's Q* merupakan statistik uji untuk mengetahui kestabilan dalam pengklasifikasian atau sejauh mana kelompok-kelompok tersebut dapat dipisahkan dengan menggunakan variabel yang ada. Perhitungan nilai *Press's Q* dapat dihitung dengan menggunakan rumus:

$$Press's\ Q = \frac{[N - (rK)]^2}{N(K - 1)} \qquad (2.38)$$

dengan,

N = ukuran sampel

r = banyaknya individu yang tepat diklasifikasikan

K = banyaknya kelompok (kategori)

Nilai dari *Press's Q* membandingkan antara banyaknya ketepatan klasifikasi dengan ukuran sampel dan banyaknya kelompok. Nilainya dibandingkan dengan nilai kritis (tabel *chi-square* dengan derajat bebas 1). Jika nilai *Press's Q* melebihi nilai kritis, maka klasifikasi dapat dianggap sudah stabil dan konsisten secara statistik.

## 2.7 Kemiskinan

Kemiskinan sendiri merupakan bentuk ketidakmampuan untuk meraih kesejahteraan dipandang dari sisi ekonomi dalam memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan (SUSENAS, 2011). Badan Pusat Statistik mendefinisikan garis kemiskinan sebagai nilai rupiah yang harus dikeluarkan seseorang dalam sebulan agar dapat memenuhi kebutuhan dasar asupan kalori sebesar 2100 kkal/hari per kapita (garis kemiskinan makanan) ditambah dengan kebutuhan minuman non makanan yang merupakan kebutuhan seseorang yaitu papan, sandang, sekolah, transportasi dan kebutuhan individu Rumah Tangga dasar lainnya (garis kemiskinan non makanan).

Menurut Dermoredjo dan Simatupang (2003), kemiskinan *(poverty)* merupakan masalah yang dihadapi oleh seluruh negara, terutama di negara berkembang seperti Indonesia. Hal ini dikarenakan kemiskinan itu bersifat multidimensional.

### 2.7.1 Garis Kemiskinan

Metode yang digunakan adalah menghitung Garis Kemiskinan (GK), terdiri dari dua komponen yaitu Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM), secara formula adalah sebagai berikut:

$$GK = GKM + GKNM \tag{2.39}$$

Penghitungan Garis Kemiskinan dilakukan secara terpisah untuk daerah perkotaan dan perdesaan. Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah Garis Kemiskinan.

Garis Kemiskinan Makanan (GKM) merupakan nilai pengeluaran kebutuhan minimum makanan yang disetarakan dengan 2.100 kilokalori per kapita per hari. Paket komoditi kebutuhan dasar makanan diwakili oleh 52 jenis komoditi (padipadian, umbi-umbian, ikan, daging, telur dan susu, sayuran, kacang-kacangan, buahbuahan, minyak dan lemak, dll).

Kemiskinan Non Makanan (GKNM) adalah kebutuhan minimum untuk perumahan, sandang, pendidikan dan kesehatan. Paket komoditi kebutuhan dasar non makanan diwakili oleh 51 jenis komoditi di perkotaan dan 47 jenis komoditi di perdesaaan.

UNNES
UNIVERSITAS NEGERI SEMARANG

### 2.7.2 Teknik Perhitungan Garis Kemiskinan Tingkat Kabupaten/Kota

Menurut Badan Pusat Statistik Provinsi Jawa Tengah, untuk menghitung garis kemiskinan tingkat Kabupaten/Kota adalah sebagai berikut:

1. Menentukan penduduk referensi yaitu 20 persen penduduk yang berada di atas Garis Kemiskinan Sementara (GKS). GKS adalah GK periode lalu yang

di-inflate dengan inflasi umum (IHK). Dari penduduk referensi ini kemudian dihitung Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non-Makanan (GKNM).

2. GKM adalah nilai pengeluaran dari 52 komoditi dasar makanan yang riil dikonsumsi penduduk referensi yang kemudian disetarakan dengan 2100 kilokalori perkapita per hari. Penyetaraan nilai pengeluaran kebutuhan minimum makanan dilakukan dengan menghitung harga rata-rata kalori dari ke-52 komoditi tersebut. Formula dasar dalam menghitung GKM adalah:

$$GKM_j = \sum_{k=1}^{52} P_{jk}.Q_{jk} = \sum_{k=1}^{52} V_k \qquad (2.40)$$

dimana,

$GKM_j$ = GKM daerah j (sebelum disetarakan menjadi 2100 kilokalori).

$P_{jk}$ = harga komoditi $k$ di daerah $j$

$Q_{jk}$ = rata-rata kuantitas komoditi $k$ yang dikonsumsi di daerah $j$.

$V_{jk}$ = nilai pengeluaran untuk konsumsi komoditi $k$ di daerah $j$.

$J$ = daerah (perkotaan atau perdesaan)

GKNM merupakan penjumlahan nilai kebutuhan minimum dari komoditi-komoditi non-makanan terpilih yang meliputi perumahan, sandang, pendidikan, dan kesehatan. Nilai kebutuhan minimum per komiditi/sub kelompok non-makanan dihitung dengan menggunakan suatu rasio pengeluaran komoditi/sub kelompok tersebut terhadap total pengeluaran

komoditi/sub kelompok yang tercatat dalam data Susenas modul konsumsi. Rasio tersebut dihitung dari hasil Survei Paket Komoditi Dasar 2004 (SPKKD 2004) yang dilakukan untuk mengumpulkan data pengeluaran konsumsi rumah tangga per komiditi non makanan yang lebih rinci dibandingkan data Susenas modul konsumsi.

3. Nilai kebutuhan minimum non makanan secara matematis dapat diformulasikan sebagai berikut:

$$GKNM_p = \sum_{i-1}^{n} r_i \, v_i \qquad (2.41)$$

dimana,

$GKNM_p =$ pengeluaran minimum non-makanan atau garis kemiskinan nonmakanan daerah p

$v_i$ = nilai pengeluaran per komoditi/sub-kelompok non-makanan daerah p (dari Susenas modul konsumsi).

$r_i$ = rasio pengeluaran komoditi/sub-kelompok non-makanan menurut daerah (hasil SPKKD 2004).

$i$ = jenis komoditi non-makanan terpilih di daerah p.

$p$ = daerah (perkotaan/perdesaan).

4. GK merupakan penjumlahan dari GKM dan GKNM. Penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah GK dikategorikan sebagai penduduk miskin.

5. Selanjutnya dihitung jumlah penduduk di bawah GK untuk tingkat kabupaten/kota.

UNNES
UNIVERSITAS NEGERI SEMARANG

6. Melakukan *prorate* jumlah penduduk miskin kabupaten/kota yang diperoleh pada langkah sebelumnya, terhadap jumlah penduduk miskin provinsi yang dihasilkan dengan menggunakan data Susenas Triwulan III.
7. Dari jumlah penduduk miskin yang diperoleh dari hasil *prorate*, maka dihitung persentase penduduk miskin dan Garis Kemiskinan (GK) dari seluruh kabupaten/kota. Penghitungan GK kabupaten/kota dilakukan sebagai dasar dalam penghitungan indeks kedalaman kemiskinan/*poverty gap index* ($P_1$).

### 2.7.3 Indikator Kemiskinan

***Head Count Index* (*HCI* - $P_0$)** yaitu persentase penduduk yang berada di bawah Garis Kemiskinan (GK).

**Indeks Kedalaman Kemiskinan (*Poverty Gap Index* $P_1$)** merupakan ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan. Semakin tinggi nilai indeks, semakin jauh rata-rata pengeluaran penduduk dari garis kemiskinan.

**Indeks Keparahan Kemiskinan (*Poverty Severity Index* $P_2$)** memberikan gambaran mengenai penyebaran pengeluaran di antara penduduk miskin. Semakin tinggi nilai indeks, semakin tinggi ketimpangan pengeluaran di antara penduduk miskin.

UNNES
UNIVERSITAS NEGERI SEMARANG

## 2.8 Penelitian Terdahulu

Penelitian yang berkaitan dengan klasifikasi data menggunakan *Multivariatee Adaptive Regression Splines* (MARS) diantaranya:

1. Penelitian oleh Ghofar (2014) yang berjudul "*Klasifikasi Kelulusan Mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro Menggunakan Multivariatee Adaptive Regression Splines (MARS)*". Penelitian ini dilakukan untuk mendapatkan suatu model yang mampu mengklasifikasi data kelulusan mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro Semarang dengan menggunakan metode *Multivariatee Adaptive Regression Spline* (MARS). Model terbaik yang digunakan adalah model dengan kombinasi BF=28, MI=2, MO=1 karena memiliki nilai GCV terkecil yaitu sebesar 0,17781. Terdapat tiga variabel yang berkontribusi ke dalam model MARS yaitu variabel IPK, jurusan dan jenis kelamin. Didapatkan kesalahan klasifikasi sebesar 20,50%. Nilai uji *Press's Q* menunjukkan bahwa secara statistik metode MARS sudah konsisten dalam mengklasifikasi data kelulusan mahasiswa FSM Undip Semarang.
2. Penelitian oleh Muslikah (2015) yang berjudul "*Multivariate Adaptive Regression Spline (MARS) untuk Klasifikasi Kejadian Konstipasi terhadap Pemberian Air Susu Ibu dan Pemberian Air Susu Formula*". Tujuan dari penelitian ini yaitu untuk mengklasifikasi dan menganalisis kejadian konstipasi terhadap pemberian ASI eksklusif dan pemberian susu formula pada bayi usia 6 – 12 bulan. Diperoleh model MARS terbaik dengan

UNNES
UNIVERSITAS NEGERI SEMARANG

kombinasi Fungsi Basis (BF), Maksimum Interaksi (MI), dan Minimum Observasi (MO) secara *Trial and Error* yaitu BF = 24, MI = 2, dan MO = 0. Terdapat lima variabel yang berpengaruh terhadap model yaitu variabel frekuensi defekasi atau BAB < 5 kali (X1), pemberian makanan bayi pada usia 6 bulan (X2), Kontsipasi tinja padat (X3), Mengejan (X4) dan kesakitan (X5).

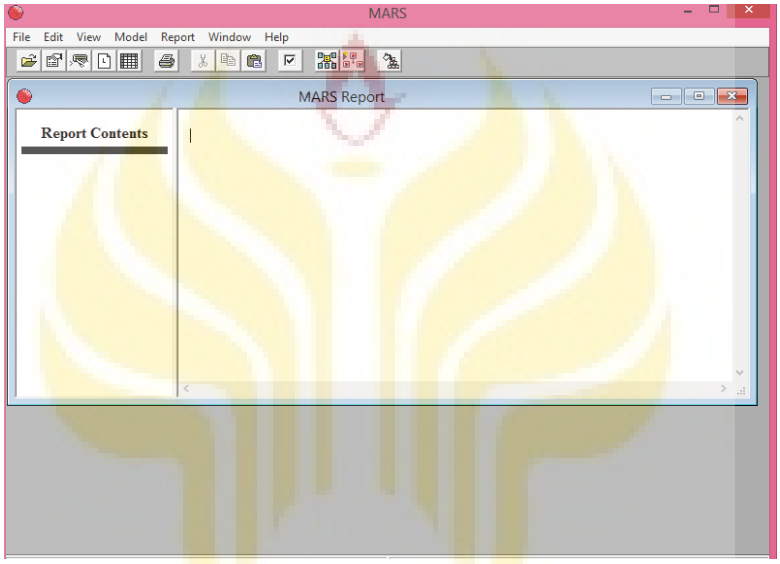
3. Penelitian oleh Kishartini (2014) yang berjudul "*Multivariate Adaptive Regression Splines (MARS) Untuk Klasifikasi Status Kerja Di Kabupaten Demak*". Penelitian ini membahas model MARS pada klasifikasi kerja di Kabupaten Demak adalah model MARS (BF=24, MI=3, dan MO=1) karena memiliki nilai GCV paling kecil yaitu sebesar 0,03451 dengan bentuk persamaan:

$$\hat{f}(x) = -0{,}526742 + (0{,}264444)BF2 + (12{,}2382)BF5 - (7{,}76719)BF15 + (4{,}96445)BF17.$$

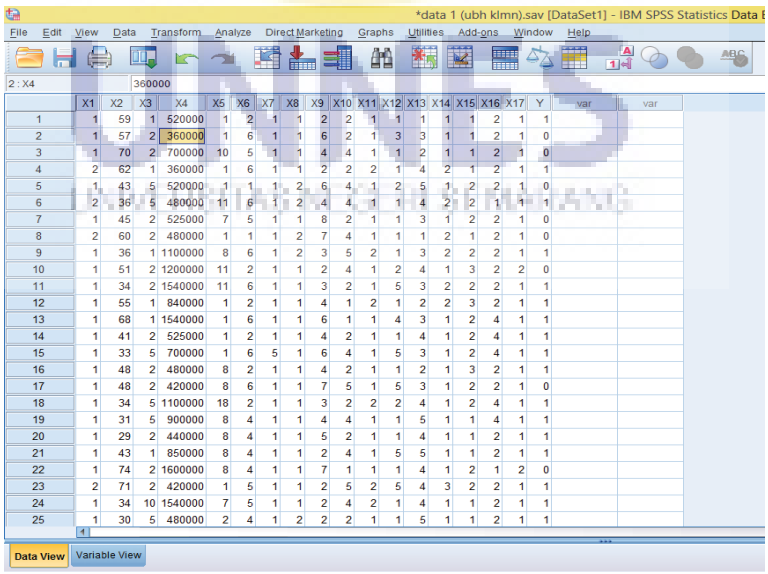
UNNES
UNIVERSITAS NEGERI SEMARANG

## 2.9 *Software* MARS 2.0

*Software Multivariate Adaptive Regression Spline* Versi 2.0 atau MARS 2.0 merupakan *software* yang membantu dalam pengerjaan dalam memodelkan *Multivariate Adaptive Regression Splines* (MARS). Adapun tampilan jendela utama dari *software* MARS 2.0 seperti pada gambar 2.1.

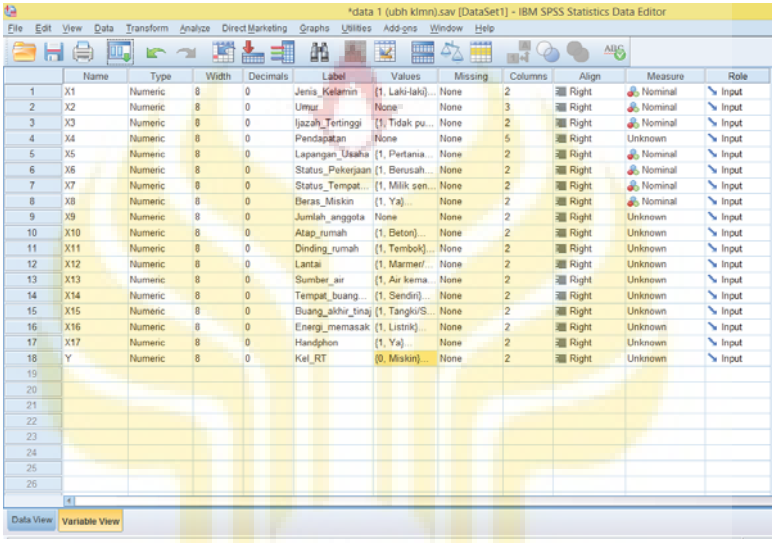



Gambar 2.1. Menu Utama MARS 2.0



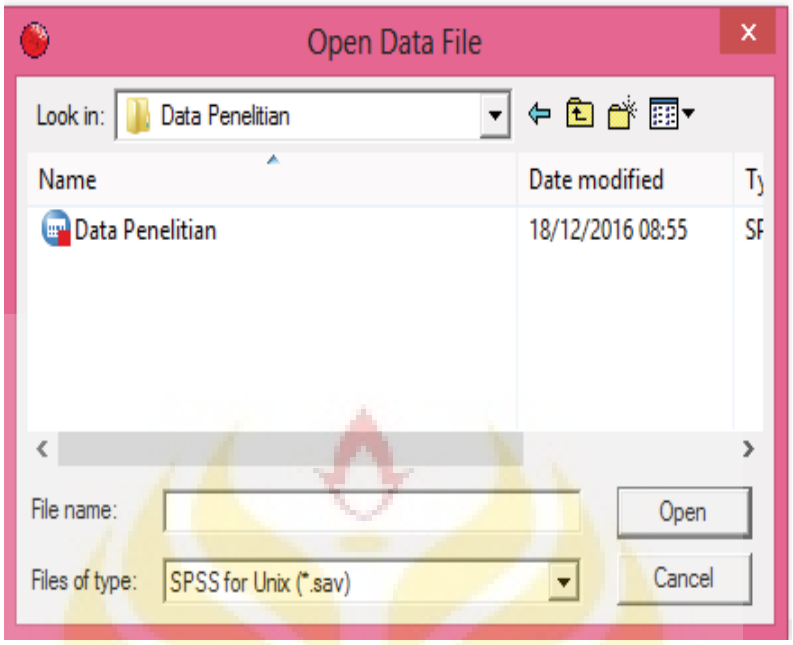

Gambar 2.2. Data dalam program SPSS versi 16.0

Gambar 2.2 merupakan data penelitian dalam program SPSS versi 16.0. Sebelum data di olah menggunakan *software* MARS 2.0 data harus disimpan terlebih dahulu dalam program SPSS. Kemudian variabel – variabel yang mempunyai tipe data kategori terlebih dahulu diberi label seperti dalam gambar 2.3 sebagai berikut.



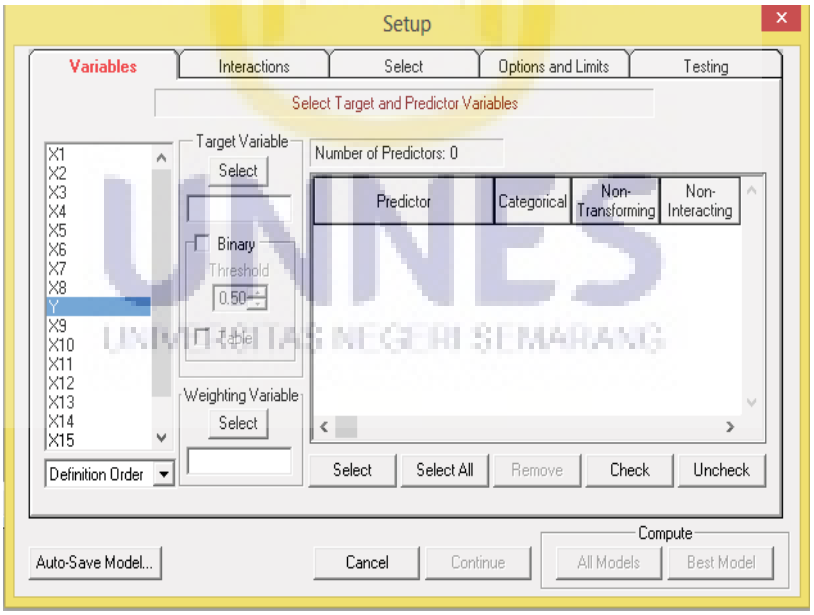

Gambar 2.3. Tampilan variabel view dalam program SPSS 16.0

Gambar 2.4. Memberi label untuk data yang memiliki kategori

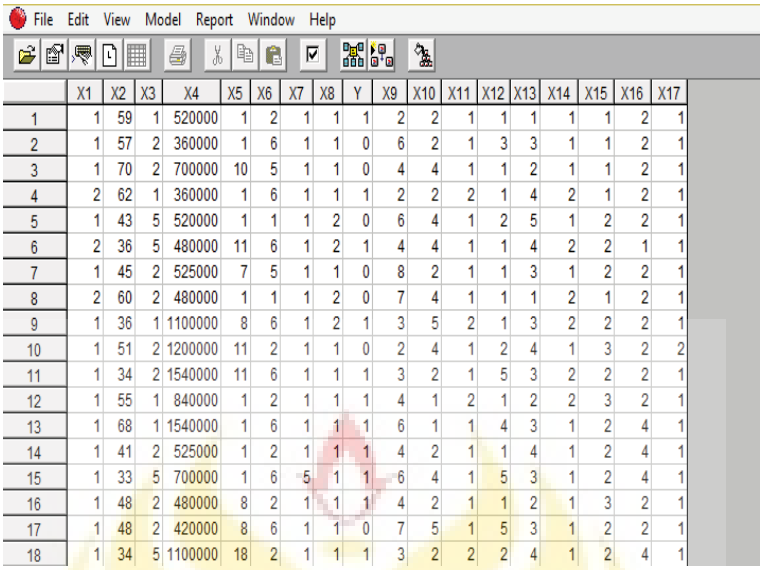



Gambar 2.5. Tampilan Membuka Dokumen

Gambar 2.5 merupakan tampilan membuka dokumen yang akan diproses menggunakan program MARS 2.0.




Gambar 2.6. Tampilan variabel – variabel dalam *software* MARS 2.0

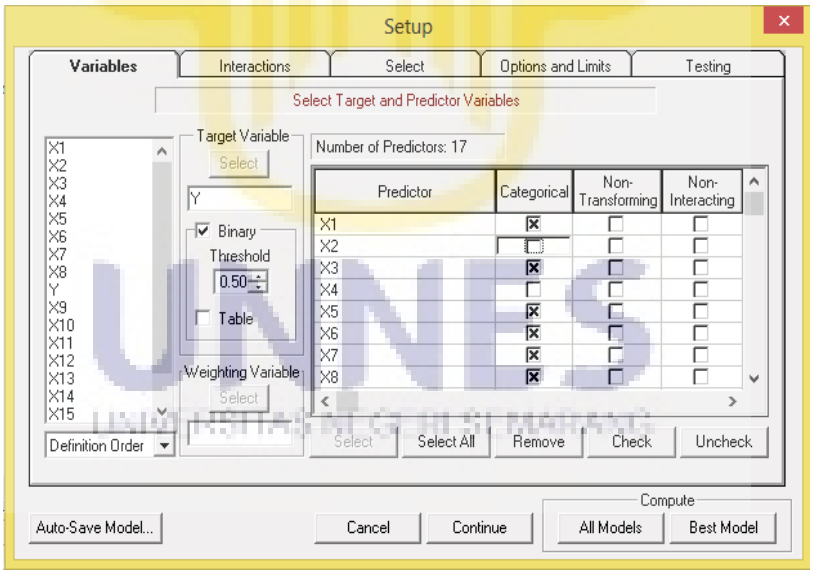


File Edit View Model Report Window Help

| | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | Y | X9 | X10 | X11 | X12 | X13 | X14 | X15 | X16 | X17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 59 | 1 | 520000 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 |
| 2 | 1 | 57 | 2 | 360000 | 1 | 6 | 1 | 1 | 0 | 6 | 2 | 1 | 3 | 3 | 1 | 1 | 2 | 1 |
| 3 | 1 | 70 | 2 | 700000 | 10 | 5 | 1 | 1 | 0 | 4 | 4 | 1 | 1 | 2 | 1 | 1 | 2 | 1 |
| 4 | 2 | 62 | 1 | 360000 | 1 | 6 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 4 | 2 | 1 | 2 | 1 |
| 5 | 1 | 43 | 5 | 520000 | 1 | 1 | 1 | 2 | 0 | 6 | 4 | 1 | 2 | 5 | 1 | 2 | 2 | 1 |
| 6 | 2 | 36 | 5 | 480000 | 11 | 6 | 1 | 2 | 1 | 4 | 4 | 1 | 1 | 4 | 2 | 2 | 1 | 1 |
| 7 | 1 | 45 | 2 | 525000 | 7 | 5 | 1 | 1 | 0 | 8 | 2 | 1 | 1 | 3 | 1 | 2 | 2 | 1 |
| 8 | 2 | 60 | 2 | 480000 | 1 | 1 | 1 | 2 | 0 | 7 | 4 | 1 | 1 | 1 | 2 | 1 | 2 | 1 |
| 9 | 1 | 36 | 1 | 1100000 | 8 | 6 | 1 | 2 | 1 | 3 | 5 | 2 | 1 | 3 | 2 | 2 | 2 | 1 |
| 10 | 1 | 51 | 2 | 1200000 | 11 | 2 | 1 | 1 | 0 | 2 | 4 | 1 | 2 | 4 | 1 | 3 | 2 | 2 |
| 11 | 1 | 34 | 2 | 1540000 | 11 | 6 | 1 | 1 | 1 | 3 | 2 | 1 | 5 | 3 | 2 | 2 | 2 | 1 |
| 12 | 1 | 55 | 1 | 840000 | 1 | 2 | 1 | 1 | 1 | 4 | 1 | 2 | 1 | 2 | 2 | 3 | 2 | 1 |
| 13 | 1 | 68 | 1 | 1540000 | 1 | 6 | 1 | 1 | 1 | 6 | 1 | 1 | 4 | 3 | 1 | 2 | 4 | 1 |
| 14 | 1 | 41 | 2 | 525000 | 1 | 2 | 1 | 1 | 1 | 4 | 2 | 1 | 1 | 4 | 1 | 2 | 4 | 1 |
| 15 | 1 | 33 | 5 | 700000 | 1 | 6 | 5 | 1 | 1 | 6 | 4 | 1 | 5 | 3 | 1 | 2 | 4 | 1 |
| 16 | 1 | 48 | 2 | 480000 | 8 | 2 | 1 | 1 | 1 | 4 | 2 | 1 | 1 | 2 | 1 | 3 | 2 | 1 |
| 17 | 1 | 48 | 2 | 420000 | 8 | 6 | 1 | 1 | 0 | 7 | 5 | 1 | 5 | 3 | 1 | 2 | 2 | 1 |
| 18 | 1 | 34 | 5 | 1100000 | 18 | 2 | 1 | 1 | 1 | 3 | 2 | 2 | 2 | 4 | 1 | 2 | 4 | 1 |

Gambar 2.7. Data dalam *software* MARS 2.0

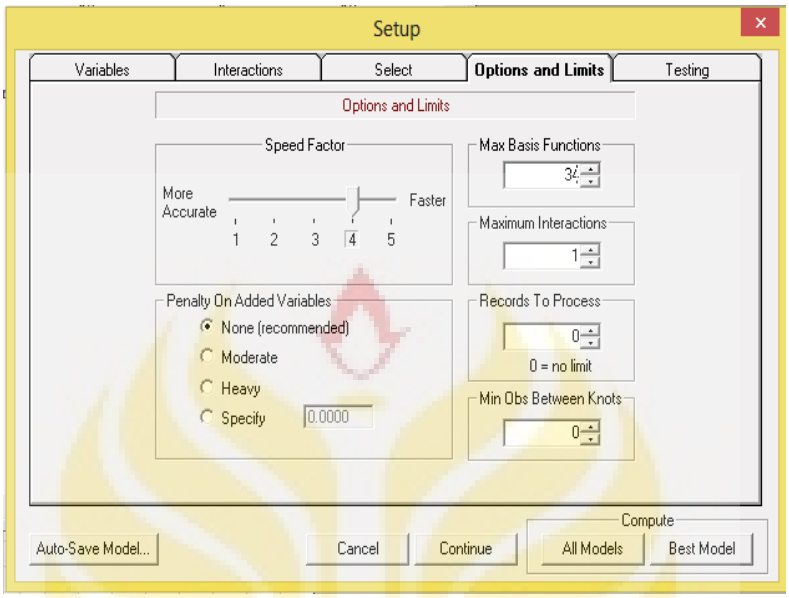
Gambar 2.7 merupakan tampilan data penelitian yang sudah di diinputkan dan akan diolah menggunakan *software* MARS 2.0




Gambar 2.8. Model dalam *software* MARS 2.0

Gambar 2.8 merupakan tampilan jendela setup sebelum proses pengolahan data. Pilih *Target Variable* yaitu Y, centang *Binary,* kemudian klik *Select All*, Pada

*Categorical* beri tanda silang variabel – variabel yang termasuk data kategori. Kemudian klik *Options and Limits*.




Gambar 2.9. Options and Limits dalam software MARS 2.0

Gambar 2.9 Menu *Options and Limits* untuk melakukan *trial and error* dengan cara menginputkan nilai pada Fungsi Basis (BF), Maksimum Interaksi (MI), dan Minimum Observasi (MO).

.

UNNES
UNIVERSITAS NEGERI SEMARANG

# BAB 5

# PENUTUP

## 5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, bahwa dengan menggunakan data kelompok rumah tangga di Kabupaten Cilacap tahun 2014 sejumlah 150 data diperoleh kesimpulan sebagai berikut:

1. Berdasarkan variabel-variabel yang mempengaruhi kelompok rumah tangga di Kabupaten Cilacap, diperoleh hasil klasifikasi dari 37 rumah tangga dengan kategori miskin, 34 rumah tangga tepat diklasifikasikan ke dalam kategori miskin, sedangkan 3 rumah tangga salah diklasifikasikan dan masuk ke dalam kategori tidak miskin. Begitu juga dari 113 rumah tangga dengan kategori tidak miskin, 113 rumah tangga tepat diklasifikasikan ke dalam kategori tidak miskin, dan tidak ada rumah tangga yang salah diklasifikasikan ke dalam rumah tangga dengan kategori miskin.
2. Diperoleh ketepatan hasil klasifikasinya sebesar 98,00%, nilai *Apparent Error Rate* (APER) sebesar 2,00% dan nilai uji *Press's Q* menunjukkan bahwa secara statistik metode MARS sudah konsisten dalam mengklasifikasi data kelompok rumah tangga Kabupaten Cilacap.

UNNES
UNIVERSITAS NEGERI SEMARANG

## 5.2 Saran

Berdasarkan simpulan di atas peneliti memberikan saran sebagai berikut.

1. Metode MARS dapat dipertimbangkan untuk mengklasifikasi data karena telah terbukti menghasilkan ketepatan klasifikasi yang akurat.
2. Penelitian selanjutnya dapat menambahkan variabel-variabel prediktor yang lain yang dapat mempengaruhi kelompok rumah tangga.

UNNES
UNIVERSITAS NEGERI SEMARANG

# DAFTAR PUSTAKA


Astuti, S. P. 2006. *Mengolah Data Statistik dengan Mudah Menggunakan Minitab 14*. Yogyakarta: Andi.

Badan Pusat Statistik. 2014a. *Indikator Kesejahteraan Rakyat 2014*. Jakarta: Badan Pusat Statistik.

Badan Pusat Statistik. 2014b. *Perhitungan Dan Analisis Kemiskinan Makro Indonesia Tahun 2014*. Jakarta: Badan Pusat Statistik.

Badan Pusat Statistik. *Data dan Informasi Kemiskinan Jawa Tengah 2010 – 2014*. Jawa Tengah: Badan Pusat Statistik.

Badan Pusat Statistik. *Data dan Informasi Kemiskinan Jawa Tengah 2010-2014*. Jawa Tengah: Badan Pusat Statistik Provinsi Jawa Tengah.

Bambang, S. 2008. *Analisis Regresi Terapan dengan SPSS*. Yogyakarta: Graha Ilmu.

Botella, O, dan Shariff, K. dalam N. Budiantara, S. Guritno, B. W. Otok, dan F. Suryadi, "Pemodelan B-Spline dan MARS pada Nilai Ujian Masuk Terhadap IPK Mahasiswa Jurusan Disain Komunikasi Visual UK Petra Surabaya," Jurnal Teknik Industri vol 8, Surabaya (2006).

Budiantara, N, S. Guritno, Otok, B.W. dan Suryadi, F. "Pemodelan B-Spline dan MARS pada Nilai Ujian Masuk Terhadap IPK Mahasiswa Jurusan Disain Komunikasi Visual UK Petra Surabaya," Jurnal Teknik Industri vol 8, Surabaya (2006).

Cox, D.R. and Snell, E.J. 1989. Cox, D.R. and Snell, E.J. 1989. *Analysis of Binary Data.* Second Edition. Chapman and Hall, London.

Dermoredjo, S. K., dan Pantjar Simatupang. 2003, *Produksi Domestik Bruto, Harga, dan Kemiskinan, Media Ekonomi dan Keuangan Indonesia*, Hal. 191 - 324, Vol. 51, No. 3.

Eubank, R. L, dalam Ayu, E. D. Pendekatan *Multivariate Adaptive Regression SPLINES* (MARS) pada Pemodelan Penduduk Miskin di Indonesia Tahun 2008-2012. *Prosiding Seminar Nasional Matematika*. Jember: Universitas Jember.

Friedman, J.H. 1991. *Multivariate Adaptive Regression Splines.* The Annals of Statistics, Vol. 19 No. 1.

Ghofar, Y.R. Safitri, D. Rusgiyono, A. 2014. Klasifikasi Kelulusan Mahasiswa Fakultas SAINS dan Matematika Universitas Diponegoro Menggunakan *Multivariate Adaptive Regression Spline* (MARS). *Jurnal Gaussian*. 3(4): 839-848.

Hardle, W. dalam Ayu, E. D. Pendekatan *Multivariate Adaptive Regression SPLINES* (MARS) pada Pemodelan Penduduk Miskin di Indonesia Tahun 2008-2012. *Prosiding Seminar Nasional Matematika*. Jember: Universitas Jember.

Haughton, J. dan Khandker, S. R. 2012. *Pedoman Tentang Kemiskinan dan Ketimpangan*. Jakarta: Salemba Empat.

Hosmer. D. W., Lemeshow. S. dalam Bisri Merluarini. 2014. Perbandingan Analsisis Klasifikasi menggunakan Metode *K-Nearst Neighbour* (K-NN) dan *Multivariate Adaptive Regression Splines* (MARS) pada Data Akreditasi Sekolah Dasar Negeri di Kota Semarang. *Jurnal Gaussia,* 3(3): 313 – 322.

Iriawan, N. 2006. *Mengolah Data Statistik dengan Mudah Menggunakan Minitab 14*. Yogyakarta: Andi.

Johnson, R.A. and Wichern, D.W.1992. *Applied Multivariatee Statistical Analysis*. Prentice Hall, New Jersey.

Kishartini, Safitri, Ispriyanti, D. 2014. Multivariate Adaptive Regression Spline (MARS) untuk Klasifikasi Status Kerja di Kabupaten Demak. *Jurnal Gaussian*. 3(4): 711-718.

Muslikah, R. dan Yamin, D. 2015. *Multivariate Adaptive Regression Spline* (MARS) untuk Klasifikasi Kejadian Konstipasi terhadap Pemberian Air Susu Ibu dan Pemberian Air Susu Formula. *Jurnal Statistika*, 3(2).

Nisa', Shofa F. dan Budiantara, N. 2012. Analisis Survival dengan Pendekatan *Multivariatee Adaptive Regression Splines* pada Kasus Demam Berdarah *Dengue* (DBD). *Jurnal Sains dan Seni*, 1(1): 318-323.

Otok, B.W. dan Pintowati, W. dalam Muslikhah, R. 2015. *Multivariate Adaptive Regression Spline* (MARS) untuk Klasifikasi Kejadian Konstipasi terhadap Pemberian Air Susu Ibu dan Pemberian Air Susu Formula. *Jurnal Statistika*, 3(2).

Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: Penerbit Andi Yogyakarta.

Provinsi Jawa Timur," Survei Sosial Ekonomi Nasional tahun 2010 Provinsi Jawa Timur,"BPS Provinsi Jawa Timur, Surabaya (2011).

Walpole, R.E. 1993. *Pengantar Statistika Edisi ke-3*. Jakarta: PT. Gramedia Pustaka Utama.

Wengang, Z. 2014. *Multivariate Adaptive Regression Splines and Neural Network Models for Prediction of Pile Drivability*. *Geoscience Frontier* 7 (2016) 45-52.

UNNES
UNIVERSITAS NEGERI SEMARANG