

MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) UNTUK KLASIFIKASI STATUS KERJA DI KABUPATEN DEMAK

Kishartini¹, Diah Safitri², Dwi Ispriyanti³

¹Mahasiswa Jurusan Statistika FSM UNDIP

^{2,3}Staf Pengajar Jurusan Statistika FSM UNDIP

ABSTRACT

Unemployment is one of the issues relating to economic activities, public relations and also the problems of humanity. Unemployment also occur in Demak and factors suspected as the cause of unemployment in Demak: gender, area of residence, age, status in the household, marriage status and education. Demak BPS records the number of people looking for work (unemployed) as many as 226.228 people, or 29,55% of the working age population. MARS (Multivariate Adaptive Regression Splines) is one of the methods used for classification. MARS is used for high-dimensional data, which is data that has a number of predictor variables for $3 \leq v \leq 20$ data used in this study is a secondary data from national labor force survey (SAKERNAS) in 2012. To get the best MARS models performed with by combining Maximum Base Function (BF), Minimal Observation (MO), and Maximum Interaction (MI) by trial and error. MARS model is used to classify employment status in Demak are MARS models (BF =24, MI=3, MO=1).

Keywords: Unemployment, Classification, MARS

1. PENDAHULUAN

Dalam perencanaan pembangunan, data ketenagakerjaan memegang peranan penting. Tanpa tenaga kerja program pembangunan tidaklah mungkin dapat dilaksanakan. Ketersediaan data ketenagakerjaan yang semakin lengkap dan tepat akan memudahkan pemerintah dalam membuat rencana pembangunan mengingat jumlah dan komposisi tenaga kerja selalu mengalami perubahan seiring dengan berlangsungnya proses demografi. Penduduk angkatan kerja dibagi menjadi dua kategori yaitu bukan pengangguran dan pengangguran terbuka adalah mereka yang tidak bekerja dan saat ini sedang aktif mencari pekerjaan, termasuk juga mereka yang pernah bekerja atau sekarang sedang dibebaskan tugas sehingga menganggur dan sedang mencari pekerjaan (Badan Pusat Statistik Kabupaten Demak, 2012). Salah satu metode untuk mengklasifikasikan angkatan kerja di Kabupaten Demak yaitu metode MARS. MARS pertama kali dikembangkan oleh Friedman (1991) untuk menyelesaikan data berdimensi tinggi. Sedangkan tujuan dari penelitian ini yaitu melakukan pemodelan MARS pada faktor-faktor yang mempengaruhi status kerja di Kabupaten Demak serta mengklasifikasi status kerja berdasarkan model MARS terbaik menggunakan metode GCV.

2. TINJAUAN PUSTAKA

Menurut BPS Kabupaten Demak (2012), angkatan kerja adalah penduduk usia kerja yang selama seminggu yang lalu mempunyai pekerjaan, baik yang bekerja maupun sementara tidak bekerja karena suatu sebab, seperti menunggu panen, pegawai yang sedang cuti dan sejenisnya. Di samping itu mereka tidak mempunyai pekerjaan tetapi sedang mencari atau mengharap pekerjaan juga masuk dalam angkatan kerja, Beberapa faktor yang mempengaruhi status kerja adalah status rumah tangga, jenis kelamin, umur, status perkawinan, daerah tempat tinggal dan pendidikan.

Metode *Multivariate Adaptive Regression Splines* (MARS) dapat digunakan untuk mengatasi data berdimensi tinggi yaitu data dengan variabel prediktor (X_p) sebanyak $3 \leq v \leq 20$ (Friedman, 1991). Menurut Pintowati dan Otok (2012), beberapa hal yang perlu diperhatikan dalam menggunakan model MARS yaitu:

- a. 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 basis fungsi antar satu region dengan region lainnya.
- b. Fungsi basis (BF), 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 kotinyu pada setiap titik knot. Friedman menyarankan banyaknya maksimum fungsi basis (BF) adalah 2-4 kali banyaknya variabel prediktornya. Banyaknya maksimum interaksi (MI) yang digunakan adalah 1,2, atau 3. Jika $MI > 3$ akan dihasilkan model yang semakin kompleks dan model akan sulit untuk diinterpretasi. Minimum jarak antara knot atau minimum observasi (MO) yang digunakan sebesar 0, 1, 2, dan 3.

Menurut Friedman (1991) model umum MARS dapat ditulis:

$$f(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v_N(k,m)} - t_{km})]_+ \quad (1)$$

dengan:

a_0 = basis fungsi induk

a_m = koefisien dari fungsi basis ke-m

M = banyaknya maksimum fungsi basis

K_m = derajat interaksi pada fungsi basis ke m

s_{km} = nilainya 1 atau -1 jika data berada di sebelah kanan atau kiri titik knot

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

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

Berdasarkan fungsi regresi nonparametrik, model MARS dinyatakan dalam persamaan berikut:

$$y_i = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{p_n(k,m)} - t_{km})] + \varepsilon_i \quad (2)$$

Dari model MARS pada persamaan (2) dalam bentuk matriks dapat ditulis sebagai berikut:

$$\underline{Y} = \underline{B}\underline{a} + \underline{\varepsilon} \quad (3)$$

dengan:

$$\underline{Y} = (y_1, y_2, y_3, \dots, y_N)^T$$

$$\underline{a} = (a_0, a_1, a_2, \dots, a_M)^T$$

$$\mathbf{B} = \begin{pmatrix} 1 & \prod_{k=1}^{K_1} [s_{1m} \cdot (x_{v_1(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_M} [s_{Mm} \cdot (x_{v_1(M,m)} - t_{Mm})] \\ 1 & \prod_{k=1}^{K_1} [s_{1m} \cdot (x_{v_2(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_M} [s_{Mm} \cdot (x_{v_2(M,m)} - t_{Mm})] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} [s_{1m} \cdot (x_{v_N(1,m)} - t_{1m})] & \dots & \prod_{k=1}^{K_M} [s_{Mm} \cdot (x_{v_N(M,m)} - t_{Mm})] \end{pmatrix}$$

Sedangkan estimasi modelnya diperoleh dari persamaan (3) yaitu:

$$\underline{\hat{Y}} = \mathbf{B} \underline{\hat{a}}$$

$$\underline{\hat{Y}} = \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \underline{Y}$$

Menurut Friedman (1991), yang perlu diperhatikan dalam pemilihan model MARS terbaik adalah jika nilai *generalized cross-validation* (GCV) dari model tersebut mempunyai nilai yang paling rendah (minimum) diantara model-model yang lain.

$$\text{GCV}(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{\hat{C}(M)}{N}\right]^2} \quad (4)$$

Dimana:

- y_i = variabel respon
- x_i = variabel prediktor
- N = banyaknya pengamatan
- $\hat{f}_M(x_i)$ = nilai taksiran variabel respon pada M fungsi basis di x_i
- M = banyaknyamaksimal fungsi basis
- $\hat{C}(M)$ = $C(M) + d \cdot M$
- $C(M)$ = $\text{Trace} [B (B^T B)^{-1} B^T] + 1$; \mathbf{B} adalah matriks dari M fungsi basis
- d = nilai ketika setiap fungsi basis mencapai optimasi ($2 \leq d \leq 4$)

Menurut Agresti (1990), analisis Regresi Logistik adalah analisis yang digunakan untuk melihat hubungan antara variabel respon kategorik dengan variabel-variabel prediktor kategorik maupun kontinyu. Pada Regresi Logistik dikhotomus, respon Y hanya terdiri dari dua kategori (misalnya : 1 dan 0).

Menurut Friedman (1991) model regresi logistik MARS dibentuk dengan mengganti fungsi $\hat{f}(x)$, maka model logistik MARS:

$$p = \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}}$$

dengan: $\hat{f}(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_+$

sehingga:

$$\log\left(\frac{p}{1-p}\right) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_+$$

Menurut Johnson dan Wichern (1992), evaluasi prosedur klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. Ukuran yang digunakan dalam menghitung ketepatan klasifikasi pada hasil pengelompokan digunakan *apparent error rate* (APER). Nilai APER menyatakan representasi proporsi sampel yang salah diklasifikasikan. Dalam respon biner penentuan kesalahan klasifikasi dapat dihitung dari tabel klasifikasi berikut:

| Hasil observasi (actual class) | Hasil Prediksi (predicted class) | |
|-----------------------------------|----------------------------------|------------|
| | Kelompok 1 | Kelompok 2 |
| Kelompok 1 | n_{11} | n_{12} |
| Kelompok 2 | n_{21} | n_{22} |

- n_{11} = banyaknya observasi kelompok 1 yang tepat diklasifikasikan sebagai kelompok 1.
- n_{12} = banyaknya observasi kelompok 1 yang salah diklasifikasikan sebagai kelompok 1.
- n_{21} = banyaknya observasi kelompok 1 yang salah diklasifikasikan sebagai kelompok 1.
- n_{22} = banyaknya observasi kelompok 2 yang tepat diklasifikasikan sebagai kelompok 2.

Nilai APER menurut Johnson dan Wichern (1992) dapat dihitung sebagai berikut:

$$APER(\%) = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\%$$

Menurut Hair et al. (2010), statistik uji untuk mengetahui kekuatan pemisahan dari matriks klasifikasi dibandingkan dengan suatu model peluang adalah statistik Press's Q. Nilai Press's Q dibandingkan dengan nilai kritis (nilai chi-square dengan derajat bebas 1 pada tingkat kepercayaan yang diinginkan). Jika nilai Press's Q lebih besar dari nilai kritis maka klasifikasi dianggap sudah baik secara statistik.

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

dengan: N = banyaknya total sampel
n = banyaknya observasi yang tepat diklasifikasikan
K = banyaknya kelompok

3. METODE PENELITIAN

Data yang digunakan adalah data sekunder dari survei angkatan kerja triwulan tiga tahun 2012 BPS Demak dengan objek penelitiannya adalah rumah tangga. Variabel yang digunakan yaitu variabel respon (Y) yaitu pengangguran (1) dan bukan pengangguran (2) dan terdapat enam variabel prediktor (X) yaitu Status Rumah Tangga (X_1), Jenis Kelamin (X_2), Umur (X_3), Status Perkawinan (X_4), Daerah Tempat Tinggal (X_5) dan Pendidikan (X_6). Software yang digunakan yaitu SPM 7.0. Sedangkan metode analisis yang dilakukan yaitu:

1. Mencari data SAKERNAS ke Badan Pusat Statistik Demak.
2. Menentukan kemungkinan maksimum jumlah fungsi basis (BF), maksimum fungsi basis yang diperbolehkan adalah sebanyak dua sampai empat kali dari banyaknya variabel prediktor yang digunakan. Dalam menentukan model status kerja di Kabupaten Demak digunakan 6 variabel prediktor sehingga kemungkinan maksimum jumlah BF adalah 12, 18 dan 24.
3. Menentukan jumlah maksimum interaksi, dalam penelitian ini jumlah maksimum interaksi (MI) yaitu 1, 2, dan 3. Karena apabila terdapat lebih dari 3 interaksi, maka akan menimbulkan interpretasi model yang sangat kompleks.
4. Menentukan minimum observasi antara knot dengan cara *trial and error* karena belum ada landasan atau batasan yang tetap untuk penentuan minimum observasi antara knot.
5. Melakukan penaksiran parameter untuk setiap kombinasi BF, MO dan MI menggunakan metode kuadrat terkecil (*Ordinary Least Square*).
6. Menentukan model Status Kerja di Kabupaten Demak yang terbaik berdasarkan pada nilai GCV minimum.
7. Menginterpretasikan model status kerja di Kabupaten Demak dan variabel-variabel yang berpengaruh di dalam model tersebut.
8. Menguji keakuratan model status kerja di Kabupaten Demak dengan metode MARS (ketepatan klasifikasi) yang terbentuk dari data dengan menggunakan APER serta menghitung kestabilan klasifikasi dengan statistik uji *Press's Q*.

4. HASIL DAN PEMBAHASAN

4.1. Model MARS

Berdasarkan kombinasi fungsi basis (BF), maksimum interaksi (MI), dan minimum observasi (MO) diperoleh beberapa model *Multivariate Adaptive Regression Splines* (MARS) status kerja di Kabupaten Demak. Jumlah variabel prediktor yang digunakan dalam pemodelan status kerja di Kabupaten Demak adalah 6 variabel prediktor, sehingga nilai BF yang digunakan adalah 12, 18, dan 24. Langkah selanjutnya yaitu menentukan model MARS berdasarkan nilai GCV terkecil.

Tabel 1. Hasil Seleksi Model MARS Menggunakan GCV

| No | BF | MI | MO | GCV | No | BF | MI | MO | GCV |
|----|----|----|----|---------|------|----|----|----|---------|
| 1 | 12 | 1 | 0 | 0,03510 | 19 | 18 | 2 | 2 | 0,03484 |
| 2 | 12 | 1 | 1 | 0,03510 | 20 | 18 | 2 | 3 | 0,03484 |
| 3 | 12 | 1 | 2 | 0,03510 | 21 | 18 | 3 | 0 | 0,03492 |
| 4 | 12 | 1 | 3 | 0,03528 | 22 | 18 | 3 | 1 | 0,03459 |
| 5 | 12 | 2 | 0 | 0,03498 | 23 | 18 | 3 | 2 | 0,03492 |
| 6 | 12 | 2 | 1 | 0,03481 | 24 | 18 | 3 | 3 | 0,03492 |
| 7 | 12 | 2 | 2 | 0,03498 | 25 | 24 | 1 | 0 | 0,03511 |
| 8 | 12 | 2 | 3 | 0,03498 | 26 | 24 | 1 | 1 | 0,03511 |
| 9 | 12 | 3 | 0 | 0,03490 | 27 | 24 | 1 | 2 | 0,03582 |
| 10 | 12 | 3 | 1 | 0,03456 | 28 | 24 | 1 | 3 | 0,03511 |
| 11 | 12 | 3 | 2 | 0,03490 | 29 | 24 | 2 | 0 | 0,03486 |
| 12 | 12 | 3 | 3 | 0,03503 | 30 | 24 | 2 | 1 | 0,03469 |
| 13 | 18 | 1 | 0 | 0,03568 | 31 | 24 | 2 | 2 | 0,03486 |
| 14 | 18 | 1 | 1 | 0,03511 | 32 | 24 | 2 | 3 | 0,03486 |
| 15 | 18 | 1 | 2 | 0,03511 | 33 | 24 | 3 | 0 | 0,03461 |
| 16 | 18 | 1 | 3 | 0,03543 | **34 | 24 | 3 | 1 | 0,03451 |
| 17 | 18 | 2 | 0 | 0,03484 | 35 | 24 | 3 | 2 | 0,03461 |
| 18 | 18 | 2 | 1 | 0,03468 | 36 | 24 | 3 | 3 | 0,03461 |

Keterangan: **) adalah model terbaik

Pada nomor model ke 34 yaitu BF = 24, MI = 3, MO = 1 memiliki nilai GCV terkecil. Model yang diperoleh dari kombinasi BF = 24, MI = 3, MO = 1 Berikut model ke 18:

$$\hat{f}(x) = 0,990706 - 0,897261*BF_1 + 0,817139*BF_4 - 0,467527*BF_5 \\ - 0,0428334*BF_7 - 0,864971*BF_9 + 0,752997*BF_{19} - 0,335211*BF_{21} \\ - 0,775606*BF_{22} - 0,325682*BF_{23}$$

Dengan:

$$\begin{aligned} BF_1 &= (X_4 = 1) & BF_{19} &= (X_6 = 2) * BF_1 \\ BF_4 &= \max(0, 19 - X_3) * BF_1 & BF_{20} &= (X_6 = (1,3)) * BF_1 \\ BF_5 &= (X_1 = 1) * BF_1 & BF_{21} &= \max(0, X_3 - 18) * BF_{20} \\ BF_7 &= (X_5 = 2) * BF_1 & BF_{22} &= \max(0, 18 - X_3) * BF_{20} \\ BF_9 &= (X_6 = 2) * BF_4 & BF_{23} &= \max(0, X_3 - 20) * BF_{20} \end{aligned}$$

Penjelasan basis fungsi

$$1. BF_1 = (X_4 = 1) = \begin{cases} (X_4 = 1), & \text{jika } X_4 = 1 \\ 0, & \text{lainnya} \end{cases}$$

Artinya koefisien BF₁ akan bermakna jika nilai X₄ = 1, tetapi jika nilai X₄ ≠ 1 maka koefisien BF₁ tidak bermakna sehingga nilainya 0. Bermaknanya BF₁ akan mengurangi angka pengangguran sebesar 0,897261.

$$2. BF_4 = \max(0, 19 - X_3) * BF_1 = \begin{cases} (19 - X_3)(X_4 = 1), & \text{jika } X_3 < 19 \text{ dan } X_4 = 1 \\ 0, & \text{jika } X_3 \geq 19 \text{ atau } X_4 \text{ lainnya} \end{cases}$$

Artinya koefisien BF₄ akan bermakna jika nilai X₃ < 19 dan X₄ = 1, tetapi jika nilai X₃ ≥ 19 atau X₄ ≠ 1 maka koefisien BF₄ tidak bermakna sehingga nilainya 0.

$$3. BF_5 = (X_1 = 1) * BF_1 = \begin{cases} (X_1 = 1)(X_4 = 1), & \text{jika } X_1 = 1 \text{ dan } X_4 = 1 \\ 0, & \text{jika } X_1 \text{ atau } X_4 \text{ lainnya} \end{cases}$$

Artinya koefisien BF₅ akan bermakna jika nilai X₁ = 1 dan X₄ = 1, tetapi jika nilai X₁ = 2 atau X₄ ≠ 1 maka koefisien BF₄ tidak bermakna sehingga nilainya 0.

$$4. BF_7 = (X_5 = 2) * BF_1 = \begin{cases} (X_5 = 2)(X_4 = 1), & \text{jika } X_5 = 1 \text{ dan } X_4 = 1 \\ 0, & \text{jika } X_5 \text{ atau } X_4 \text{ lainnya} \end{cases}$$

Artinya koefisien BF₅ akan bermakna jika nilai X₅ = 2 dan X₄ = 1, tetapi jika nilai X₅ = 1 atau X₄ ≠ 1 maka koefisien BF₄ tidak bermakna sehingga nilainya 0.

$$5. BF_9 = (X_6 = 2) * BF_4$$

$$= \begin{cases} (X_6 = 2)(19 - X_3)(X_4 = 1), & \text{jika } X_6 = 2, X_3 < 19 \text{ dan } X_4 = 1 \\ 0, & \text{jika } X_3 \geq 19, X_6 \text{ atau } X_4 \text{ lainnya} \end{cases}$$

Artinya koefisien BF₉ akan bermakna jika nilai X₆ = 2, X₃ < 19 dan X₄ = 1, tetapi jika nilai X₆ ≠ 2, X₃ ≥ 19 atau X₄ ≠ 1 maka koefisien BF₄ tidak bermakna sehingga nilainya 0.

$$6. BF_{19} = (X_6 = 2) * BF_1 = \begin{cases} (X_6 = 2)(X_4 = 1), & \text{jika } X_6 = 2 \text{ dan } X_4 = 1 \\ 0, & \text{jika } X_6 \text{ atau } X_4 \text{ lainnya} \end{cases}$$

Artinya koefisien BF₉ akan bermakna jika nilai X₆ = 2 dan X₄ = 1, tetapi jika nilai X₆ ≠ 2 atau X₄ ≠ 1 maka koefisien BF₄ tidak bermakna sehingga nilainya 0.

$$7. BF_{21} = \max(0, X_3 - 18) * BF_{20}$$

$$= \begin{cases} (X_3 - 18)(X_6 = 1)(X_4 = 1), & \text{jika } X_3 \geq 18, X_6 = 1 \text{ dan } X_4 = 1 \\ (X_3 - 18)(X_6 = 3)(X_4 = 1), & \text{jika } X_3 \geq 18, X_6 = 3 \text{ dan } X_4 = 1 \\ 0, & \text{jika } X_3 < 18, X_6 \text{ atau } X_4 \text{ lainnya} \end{cases}$$

Artinya koefisien BF₉ akan bermakna jika nilai X₃ ≥ 18, X₆ = (1,3) dan X₄ = 1, tetapi jika nilai X₃ < 18, X₆ = 2 atau X₄ ≠ 1 maka koefisien BF₄ tidak bermakna sehingga nilainya 0.

$$8. BF_{22} = \max(0, 18 - X_3) * BF_{20}$$

$$= \begin{cases} (18 - X_3)(X_6 = 1)(X_4 = 1), & \text{jika } X_3 < 18, X_6 = 1 \text{ dan } X_4 = 1 \\ (18 - X_3)(X_6 = 3)(X_4 = 1), & \text{jika } X_3 < 18, X_6 = 3 \text{ dan } X_4 = 1 \\ 0, & \text{jika } X_3 \geq 18, X_6 \text{ atau } X_4 \text{ lainnya} \end{cases}$$

Artinya koefisien BF₉ akan bermakna jika nilai X₃ < 18, X₆ = (1,3) dan X₄ = 1, tetapi jika nilai X₃ ≥ 18, X₆ = 2 atau X₄ ≠ 1 maka koefisien BF₄ tidak bermakna sehingga nilainya 0.

$$9. BF_{23} = \max(0, X_3 - 20) * BF_{20}$$

$$= \begin{cases} (X_3 - 20)(X_6 = 1)(X_4 = 1), & \text{jika } X_3 \geq 20, X_6 = 1 \text{ dan } X_4 = 1 \\ (X_3 - 20)(X_6 = 3)(X_4 = 1), & \text{jika } X_3 \geq 20, X_6 = 3 \text{ dan } X_4 = 1 \\ 0, & \text{jika } X_3 < 20, X_6 \text{ atau } X_4 \text{ lainnya} \end{cases}$$

Artinya koefisien BF₉ akan bermakna jika nilai X₃ ≥ 20, X₆ = (1,3) dan X₄ = 1, tetapi jika nilai X₃ < 20, X₆ = 2 atau X₄ ≠ 1 maka koefisien BF₄ tidak bermakna sehingga nilainya 0.

4.2 Pengklasifikasian MARS

Tabel 2. Klasifikasi model MARS

| Hasil Observasi | Hasil Prediksi | |
|--|----------------|--------------------|
| | Pengangguran | Bukan Pengangguran |
| Pengangguran | 46 | 10 |
| Bukan Pengangguran | 218 | 1102 |
| $APER (\%) = \frac{10+218}{46+10+218+1102} \times 100\% = 16,57\%$ | | |

Pengelompokan status kerja di Kabupaten Demak terbagi menjadi dua yaitu pengangguran (1) dan bukan pengangguran (2). Tabel 2 menunjukkan hasil klasifikasi pengangguran di Kabupaten Demak dari model MARS terbaik. Tingkat kesalahan klasifikasi berdasarkan nilai APER sebesar 16,57%, sehingga kesuksesan pengklasifikasian sebesar 83,43%. Evaluasi selanjutnya adalah menguji kestabilan model dengan statistik uji Press's Q. nilai statistik uji Press's Q adalah:

$$\begin{aligned} \text{Press's Q} &= \frac{[N-(nK)]^2}{N(K-1)} \\ &= \frac{[1376-(1148(2))]^2}{1376(2-1)} \end{aligned}$$

$$= 615,1162791$$

Nilai Press's Q adalah 615,1162791 kemudian dibandingkan dengan nilai chi-square derajat bebas 1 ($\chi^2_{(1,0.05)} = 3,841$). nilai Press's Q yang diperoleh jauh lebih besar dari $\chi^2_{(1,0.05)}$, sehingga dapat dikatakan pengklasifikasian pengangguran di Kabupaten Demak sudah baik secara statistik.

5. KESIMPULAN

Model MARS terbaik untuk mengklasifikasi pengangguran di Kabupaten Demak berdasarkan nilai GCV terkecil yaitu adalah model MARS(BF=24, MI=3, MO=1) karena memiliki nilai GCV paling kecil yaitu sebesar 0,03451. Kesalahan klasifikasi dari model MARS dapat dilihat dari besarnya nilai APER yaitu sebesar 16,57%, sehingga kesuksesan pengklasifikasian sebesar 83,43%.

6. DAFTAR PUSTAKA

- Agresti, A. 1990. *Categorical Data Analysis*. John Wiley and Son, Florida.
- Badan Pusat Statistik Kabupaten Demak, 2012. *Profil Ketenagakerjaan Kabupaten Demak Tahun 2012*. Badan Pusat Statistik. Demak
- Friedman, J.H. 1991. *Multivariate Adaptive Regression Splines*. *The Annals of Statistics*, Vol. 19, No. 1.
- Hair J.F., Anderson R.E., Babin B.J, Black W.C. 2010. *Multivariate Data Analysis 7th Edition Pearson Education*. Prentice Hall, inc
- Johnson, R.A. dan Wichern D.W. (1992). *Applied Multivariate Statistical Analysis 6th Edition*, Prentice Hall International Inc. New Jersey
- Pintowati, W. dan Otok, B.W. 2012. Pemodelan Kemiskinan di Propinsi Jawa Timur dengan Pendekatan Multivariate Adaptive. *Jurnal Sains dan Seni ITS* Vol.1 No.1.